

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

METİN MADENCİLİĞİ TEKNİKLERİ İLE SOSYAL MEDYA GÖNDERİLERİNİN
ANALİZ EDİLMESİ: ANKİLOZAN SPONDİLİT HASTALIĞI ÖRNEĞİ

ERTUĞRUL GÜMÜŞSU

AKILLI SİSTEMLER MÜHENDİSLİĞİ ANABİLİM DALI

SAMSUN

2019

Her hakkı saklıdır.

TEZ ONAYI

Ertuğrul GÜMÜŞSU tarafından hazırlanan “METİN MADENCİLİĞİ TEKNİKLERİ İLE SOSYAL MEDYA GÖNDERİLERİNİN ANALİZ EDİLMESİ: ANKİLOZAN SPONDİLİT HASTALIĞI ÖRNEĞİ” adlı tez çalışması 08/07/2019 tarihinde aşağıdaki jüri tarafından Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Akıllı Sistemler Mühendisliği Anabilim Dalı’nda **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Danışman Dr. Öğretim Üyesi Naci MURAT
Ondokuz Mayıs Üniversitesi
Endüstri Mühendisliği Anabilim Dalı

Jüri Üyeleri

Başkan Prof. Dr. Sermin ELEVLİ
Ondokuz Mayıs Üniversitesi
Endüstri Mühendisliği Anabilim Dalı

Üye Dr. Öğretim Üyesi Naci MURAT
Ondokuz Mayıs Üniversitesi
Endüstri Mühendisliği Anabilim Dalı

Üye Dr. Öğretim Üyesi Haydar KOÇ
Çankırı Karatekin Üniversitesi
İstatistik Anabilim Dalı

Yukarıdaki sonucu onaylarım. .../.../2019

.....

Prof. Dr. Bahtiyar ÖZTÜRK
Enstitü Müdürü

ETİK BEYAN

Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

08/07/2019

Ertuğrul GÜMÜŞSU

ÖZET

Yüksek Lisans Tezi

METİN MADENCİLİĞİ TEKNİKLERİ İLE SOSYAL MEDYA GÖNDERİLERİNİN ANALİZ EDİLMESİ: ANKİLOZAN SPONDİLİT HASTALIĞI ÖRNEĞİ

Ertuğrul GÜMÜŞSU

Ondokuz Mayıs Üniversitesi

Fen Bilimleri Enstitüsü

Akıllı Sistemler Mühendisliği Anabilim Dalı

Danışman: Dr. Öğretim Üyesi Naci MURAT

Teknoloji kullanımının her geçen gün artması ile birlikte sosyal medya ve internet kullanımı hızla yaygınlaşmaktadır. Özellikle sosyal medya kullananların sayısının artmasıyla insanlar duygu ve düşüncelerini, gündemle ilgili görüş ve fikirlerini sosyal medya araçlarını kullanarak paylaşmaktadır. En çok kullanılan sosyal medya araçlarından biri olan Facebook gibi sosyal ağ uygulamalarında milyonlarca içerik paylaşılmaktadır. Paylaşılan bu içerikler yapısal olmayan veri türleridir. Bu yapısal olmayan verilerden anlamlı bilgilerin elde edilebilmesi ve çeşitli analizlerin yapılabilmesi için yapısal hale dönüştürülmesi gerekmektedir. Yapısal olmayan metin içerikli verilerin yapısal hale dönüştürülmesinde ve analiz edilmesinde metin madenciliği yöntemleri kullanılmaktadır. Bu çalışmada Facebook'ta yer alan "Ankilozan Spondilit Hasta Derneği" grubunda Ankilozan Spondilit hastaları tarafından paylaşılan metinsel içerikler metin madenciliği yöntemleri kullanılarak analiz edilmiştir. Farklı kategoriler altında veriler toplanmıştır. Ağırlıklandırma yöntemleri olarak terim frekansı, ikili ağırlıklandırma ve TFxIDF (Terim frekansı x Ters metin frekansı) ağırlıklandırma yöntemleri kullanılmıştır. Sınıflandırma işlemleri için K-en yakın komşu, SMO (Sequential minimal optimization), karar ağaçları ve Naive Bayes algoritması kullanılmıştır. Terim ağırlıklandırma yöntemlerine göre sınıflandırma algoritmalarının başarı oranları ölçülmüştür. En başarılı sonuç %53,33 doğruluk oranı ile ikili ağırlıklandırma yönteminde SMO algoritmasının sağladığı gözlenmiştir. Her kategoriye temsil edecek anlamlı bilgi çıkarımları yapılmış ve bilgi çıkarımları çeşitli görsel grafikler ile sunulmuştur. Kategorilere ait ikili ve üzeri kelime kullanımlarını tespit etmek amacıyla birliktelik analizi yapılmıştır. Birliktelik analizi işlemlerinde Apriori algoritması uygulanmış ve sonuçlar çizelgeler halinde verilmiştir. Kategorileri temsil eden anahtar kelimelerin yıllara göre farklılık gösterdiği tespit edilmiştir. Kategorilere ait anahtar kelimeleri göstermek amacıyla kelime bulutları oluşturulmuştur.

Mayıs 2019, 231 sayfa

Anahtar Kelimeler: Metin madenciliği, Metin sınıflandırma, Birliktelik analizi, Ankilozan Spondilit, Sosyal medya analizi

ABSTRACT

Master's Thesis

ANALYSIS OF SOCIAL MEDIA POSTS WITH TEXT MINING TECHNIQUES: THE CASE OF ANKYLOSING SPONDYLITIS DISEASE

Ertuğrul GÜMÜŞSU

Ondokuz Mayıs University
Graduate School of Sciences
Department of Intelligent Systems Engineering

Supervisor: Asst. Prof. Dr. Naci MURAT

With the increasing use of technology, social media and internet usage is rapidly widespread. Especially with the increase in the number of social media users, people share their feelings and thoughts, opinions and ideas about the agenda by using social media tools. Millions of content are shared in social networking applications such as Facebook, one of the most widely used social media tools. These shared contents are non-structural data types. It is necessary to extract meaningful information from these non-structural data and to be transformed into a structural format in order to make various analyzes. Text mining methods are used for the transformation and analysis of non-structural text-based data. In this study, the textual contents shared by patients with Ankylosing Spondylitis in “The Ankylosing Spondylitis Patient Association” group on Facebook were analyzed by using text mining methods. Data was collected under different categories. Term frequency, binary weighting and TFxIDF (Term frequency x Inverse document frequency) weighting methods were used as weighting methods. K-nearest neighbors, SMO (Sequential minimal optimization), decision trees and Naive Bayes algorithm were used for classification operations. Performance ratios of classification algorithms were measured according to the term weighting methods. The most successful result was obtained by SMO algorithm in binary weighting method with an accuracy rate of 53.33%. Significant information extractions to represent each category were made and the these were presented with various visual graphs. An association analysis was performed to determine binary and above word usage of the categories. Apriori algorithm was applied in association analysis and the results were given as tables. It has been found that the keywords representing the categories differ according to years. Word clouds have been created to show the keywords belonging to the categories.

May 2019, 231 pages

Key Words: Text mining, Text classification, Association analysis, Ankylosing Spondylitis, Social media analysis

ÖNSÖZ VE TEŞEKKÜR

Yüksek lisans eğitimim boyunca yardım ve desteklerini benden esirgemeyen, güler yüzlülüğünü, samimiyetini her defasında yansıtan, sabır ve özveri ile bana yol gösteren, öğrencisi olmaktan gurur duyduğum ve her alanda kendisini rehber edindiğim Sayın Dr. Öğretim Üyesi Naci MURAT'a sonsuz teşekkür ederim.

Yüksek Lisans ders döneminde kendilerinden ders aldığım, bilgi ve tecrübelerinden faydalandığımda değerli hocalarım Sayın Prof. Dr. Birol ELEVLİ'ye, Prof. Dr. Sermin ELEVLİ'ye, Prof. Dr. Mehmet Serhat ODABAŞ'a ve Doç. Dr. Fatih YAPICI'ya teşekkür ederim.

Maddi ve manevi desteklerini hiçbir zaman benden esirgemeyen, eğitimim boyunca birçok meşakkate katlanan ailemin her bir ferdine teşekkür ederim.

Mayıs 2019, Samsun

Ertuğrul Gümüşsu

İÇİNDEKİLER DİZİNİ

ÖZET	i
ABSTRACT.....	ii
ÖNSÖZ VE TEŞEKKÜR.....	iii
İÇİNDEKİLER DİZİNİ.....	iv
SİMGELER VE KISALTMALAR.....	vi
ŞEKİLLER DİZİNİ.....	viii
ÇİZELGELER DİZİNİ.....	xii
1. GİRİŞ.....	1
2. METİN MADENCİLİĞİ.....	5
2.1. Metin Madenciliğinin Tanımı.....	5
2.2. Metin Madenciliğinin Tarihsel Gelişimi.....	6
2.3. Metin Madenciliğinin Veri Madenciliği ile İlişkisi.....	6
2.4. Doğal Dil İşleme.....	7
2.5. Metin Madenciliğinin Uygulama Alanları.....	8
2.6. Metin Madenciliği İçin Kullanılan Araçlar.....	12
2.7. Literatür Taraması.....	14
3. METİN MADENCİLİĞİ SÜRECİ.....	27
3.1. Veri Kümesinin Belirlenmesi ve Toplanması.....	28
3.2. Veri Ön İşlem Aşaması.....	29
3.2.1. Dönüştürme.....	30
3.2.2. Tarama ve işaretleme.....	30
3.2.3. Kök bulma.....	31
3.2.4. Durak kelimeleri çıkarma.....	32
3.2.5. Terim ağırlıklandırma.....	32
3.2.6. Terim ayıklama.....	37
3.3. Vektör Uzay Modeli ve Vektör Oluşturma.....	37
3.4. Benzerliklerin Hesaplanması.....	38
3.5. Özellik Seçimi.....	39
3.5.1. Bilgi kazancı.....	39
3.5.2. N-Gram.....	40
3.5.3. Ki-Kare.....	42
4. METİN SINIFLANDIRMA.....	43
4.1. Metin Sınıflandırma Algoritmaları.....	44
4.1.1. Karar ağaçları.....	44
4.1.2. Destek vektör makinesi.....	46
4.1.3. Ardışık minimal optimizasyon algoritması (SMO).....	48
4.1.4. K-en yakın komşu (K-nearest neighbors).....	48
4.1.5. Naive Bayes.....	50
4.1.6. Yapay Sinir Ağları.....	51
4.2. Metin Sınıflandırma Performans Ölçütleri.....	52
4.2.1. Doğruluk.....	53
4.2.2. Kesinlik.....	54
4.2.3. Duyarlılık.....	54
4.2.4. Özgüllük.....	54
4.2.5. Negatif Prediktif Değer.....	54
4.2.6. F-Ölçeği.....	55
4.2.7. ROC Eğrisi.....	55
5. METİN KÜMELEME.....	56

6. BİRLİKTELİK ANALİZİ	58
7. UYUM ANALİZİ VE FRIEDMAN TESTİ.....	61
8. ANKİLOZAN SPONDİLİT	62
9. SOSYAL MEDYA	65
9.1. Twitter	67
9.2. Facebook	68
9.3. Youtube	69
9.4. Instagram	69
9.5. Forum sayfaları.....	69
9.6. Blog	70
10. UYGULAMA	71
10.1. Araştırmanın Amacı ve Problemin Tanımlanması	71
10.2. Kullanılan Araçlar.....	72
10.3. Yöntem.....	72
10.4. Verilerin Toplanması	73
10.5. Veri Ön İşlem Aşaması.....	76
10.6. Sınıflandırma Sonuçları	77
10.7. Şikayet Kategorisine ait Sonuçlar.....	81
10.7.1. Şikayet gönderi sonuçları	81
10.7.2. Şikayet yorum sonuçları	109
10.8. Doktor Kategorisine Ait Sonuçlar	114
10.9. Tahlil Kategorisine Ait Sonuçlar	130
10.9.1. Tahlil gönderi sonuçları.....	130
10.9.2. Tahlil yorum sonuçları	140
10.10. Tedavi Kategorisine Ait Sonuçlar.....	145
10.10.1. Tedavi gönderi sonuçları	145
10.10.2. Tedavi yorum sonuçları.....	159
10.11. İlaç Kategorisine Ait Sonuçlar.....	168
10.12. Diğer Hastalıklar Kategorisine Ait Sonuçlar	180
10.13. Rapor Kategorisine ait Sonuçlar	190
10.14. Kişisel Deneyim ve Diyet Kategorisine ait Sonuçlar	201
10.15. Spor Kategorisine Ait Sonuçlar	206
11. TARTIŞMA VE SONUÇ	220
KAYNAKLAR	223
ÖZGEÇMİŞ	

SİMGELER VE KISALTMALAR

SİMGELER

j : Doküman sayısı

i : Terim sayısı

w_{ij} : i teriminin j dokümanındaki ağırlığı

n_i : i teriminin bulunduğu doküman sayısı

n_{i-} : i terimini içeren ancak kümede bulunmayan doküman sayısı

tf_{ij} : i teriminin j dokümanındaki frekansı

F_i : i teriminin toplam frekans sayısı

D : Doküman sayısı

df_i : i terimi içeren doküman sayısı

KISALTMALAR

AS : Ankilozan Spondilit

DVM : Destek Vektör Makineleri

SMO : Ardışık Minimal Optimizasyon

MM : Metin madenciliği

KNN : K-En Yakın Komşu

YSA : Yapay Sinir Ağı

CRF : Koşullu Rastgele Alanlar

ROC : İşlem Karakteristik Eğrisi

AUC : Eğri Altındaki Alan

TP : Doğru Pozitif

FN : Yanlış Negatif

FP : Yanlış Pozitif

TN : Doğru Negatif

TF : Terim Frekansı
IDF : Ters Doküman Frekansı
ICF : Ters Kategori Frekansı
IG : Bil Kazancı
LTC : Logaritmik Terim Bileşeni
TFC : Terim Frekansı Bileşeni
NPD : Negatif Prediktif Değer



ŞEKİLLER DİZİNİ

Şekil 2.1. Metin madenciliği işlemlerinde birlikte çalışılan alanlar	5
Şekil 2.2. Metin madenciliği ve veri madenciliği arasındaki ilişki	7
Şekil 2.3. Metin kategorizasyonu	9
Şekil 2.4. Gruplara göre dokümanları organize etme	10
Şekil 2.5. Bilgi çıkarımı	11
Şekil 2.6. Bilgi getirimi	11
Şekil 3.1. Metin madenciliği süreci	28
Şekil 3.2. Doküman terim matrisi	38
Şekil 4.1. Karar ağaçları şeması	45
Şekil 4.2. Destek vektör ve hiper düzlem seçimi	47
Şekil 4.3. Yinelemeli (recurrent) yapay sinir ağı mimarisi (Irmak vd, 2012)	52
Şekil 4.4. İleri beslemeli yapay sinir ağı mimarisi (Irmak vd, 2012)	52
Şekil 5.1. Doküman koleksiyonu kümelenmesi	56
Şekil 9.1. Web teknolojilerinin gelişim dönemleri (Sönmez, 2017)	65
Şekil 9.2. Türkiye’de en çok kullanılan sosyal medya platformları	67
Şekil 9.3. 2019 yılına ait Türkiye’deki Twitter kullanıcı sayısı	68
Şekil 9.4. 2019 yılına ait Türkiye’deki Facebook kullanıcı sayısı	68
Şekil 9.5. 2019 yılına ait Türkiye’deki Instagram kullanıcı sayısı	69
Şekil 10.1. Verilere ait toplam kullanıcı sayıları	75
Şekil 10.2. Text2ArffV5 yazılımı arayüzü	77
Şekil 10.3. SMO algoritması sonucu elde edilen hataların grafiksel gösterimi	80
Şekil 10.4. Sınıflara ait uyum analizi grafiği	81
Şekil 10.5. Şikayet kategorisine ait gönderileri temsil eden anahtar kelimelerin kelime bulutu ile gösterimi	83
Şekil 10.6. 2012 yılı erkek şikayet bölgeleri	84
Şekil 10.7. 2013 yılı erkek şikayet bölgeleri	85
Şekil 10.8. 2014 yılı erkek şikayet bölgeleri	86
Şekil 10.9. 2015 yılı erkek şikayet bölgeleri	87
Şekil 10.10. 2016 yılı erkek şikayet bölgeleri	88

Şekil 10.11. 2017 yılı erkek şikayet bölgeleri	89
Şekil 10.12. 2018 yılı erkek şikayet bölgeleri	90
Şekil 10.13. 2012 yılı kadın şikayet bölgeleri	92
Şekil 10.14. 2013 yılı kadın şikayet bölgeleri	93
Şekil 10.15. 2014 yılı kadın şikayet bölgeleri	94
Şekil 10.16. 2015 yılı kadın şikayet bölgeleri	95
Şekil 10.17. 2016 yılı kadın şikayet bölgeleri	96
Şekil 10.18. 2017 yılı kadın şikayet bölgeleri	97
Şekil 10.19. 2018 yılı kadın şikayet bölgeleri	98
Şekil 10.20. 'Ağrı' kelimesinin kullanım sıklığının cinsiyete ve toplam kullanıma göre mevsimsel değişimi	99
Şekil 10.21. 'Ağrı' kelimesinin kullanım sıklığının cinsiyete ve toplam kullanıma göre aylık değişimi	100
Şekil 10.22. Şikayet bölgeleri ve yıllara göre cinsiyet grupları için basit uyum analizi sonucu	102
Şekil 10.23. Knime'de birliktelik için oluşturulan model yapısı	105
Şekil 10.24. Şikayet kategorisine ait yorumları temsil eden anahtar kelimelerin kelime bulutu ile gösterimi	111
Şekil 10.25. Doktor kategorisine ait gönderileri temsil eden anahtar kelimelerin kelime bulutu ile gösterimi	115
Şekil 10.26. 2012 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	116
Şekil 10.27. 2013 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	117
Şekil 10.28. 2014 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	117
Şekil 10.29. 2015 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	118
Şekil 10.30. 2016 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	118
Şekil 10.31. 2017 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	119
Şekil 10.32. 2018 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller	119
Şekil 10.33. 2012 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	120
Şekil 10.34. 2013 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	120
Şekil 10.35. 2014 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	121
Şekil 10.36. 2015 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	121
Şekil 10.37. 2016 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	122
Şekil 10.38. 2017 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	122

Şekil 10.39. 2018 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller	123
Şekil 10.40. Doktor kategorisini temsil eden anahtar kelimeler ve yıllar için basit uyum analizi sonucu	125
Şekil 10.41. Tahlil kategorisine ait gönderiler için bulunan toplam 80 kelimeyi gösteren kelime bulutu	132
Şekil 10.42. Tahlil kategorisini temsil eden anahtar kelimeler ve yıllar için basit uyum analizi sonucu	134
Şekil 10.43. Tahlil kategorisine ait yorumlar için bulunan toplam 207 kelimeyi gösteren kelime bulutu	141
Şekil 10.44. Erkek kullanıcılar tarafından 2012, 2013, 2014 ve 2015 yıllarında en çok konuşulan alternatif tedavi yöntemleri	146
Şekil 10.45. Erkek kullanıcılar tarafından 2016, 2017 ve 2018 yıllarında en çok konuşulan alternatif tedavi yöntemleri	147
Şekil 10.46. Kadın kullanıcılar tarafından 2012, 2013, 2014 ve 2015 yıllarında en çok konuşulan alternatif tedavi yöntemleri	148
Şekil 10.47. Kadın kullanıcılar tarafından 2016, 2017 ve 2018 yıllarında en çok konuşulan alternatif tedavi yöntemleri	149
Şekil 10.48. Alternatif tedavi yöntemlerinin kelime bulutu ile gösterimi	150
Şekil 10.49. Alternatif tedavi yöntemleri ve yıllar için basit uyum analizi sonucu	152
Şekil 10.50. 2012 ve 2013 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri	160
Şekil 10.51. 2014 ve 2015 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri	161
Şekil 10.52. 2016 ve 2017 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri	162
Şekil 10.53. 2018 yılında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri ..	163
Şekil 10.54. Tedavi kategorisine ait yorumlar için bulunan toplam 207 kelimeyi gösteren kelime bulutu	166
Şekil 10.55. 2012 ve 2013 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri	169
Şekil 10.56. 2014 ve 2015 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri	170
Şekil 10.57. 2016 ve 2017 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri	170
Şekil 10.58. 2018 yılında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri	171
Şekil 10.59. 2012 ve 2013 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri	172
Şekil 10.60. 2014 ve 2015 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri	172
Şekil 10.61. 2016 ve 2017 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri	173
Şekil 10.62. 2018 yılında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri	173

Şekil 10.63. Kullanıcılar tarafından en çok kullanılan ilaç isimleri	174
Şekil 10.64. Gönderilerde isimi geçen 60 farklı ilacın kelime bulutu ile gösterimi	174
Şekil 10.65. Kullanılan ilaçlar ve yıllar için basit uyum analizi sonucu	176
Şekil 10.66. Erkek kullanıcılar tarafından yıllara göre diğer hastalık isimlerinin kullanımı	181
Şekil 10.67. Kadın kullanıcılar tarafından yıllara göre diğer hastalık isimlerinin kullanımı	182
Şekil 10.68. Konuşulan hastalıklar ve yıllar için basit uyum analizi sonucu	184
Şekil 10.69. Rapor kategorisine ait gönderiler için bulunan toplam 80 kelimeyi gösteren kelime bulutu	192
Şekil 10.70. Rapor kategorisine ait kelimeler ve yıllar için basit uyum analizi sonucu	194
Şekil 10.71. Kişisel deneyim ve diyet kategorisine ait gönderiler için bulunan toplam 242 kelimeyi gösteren kelime bulutu	202
Şekil 10.72. Spor kategorisine ait gönderiler için bulunan toplam 255 kelimeyi gösteren kelime bulutu	208
Şekil 10.73. 2012 ve 2013 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri	209
Şekil 10.74. 2014 ve 2015 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri	209
Şekil 10.75. 2016 ve 2017 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri	210
Şekil 10.76. 2018 yılında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri	210
Şekil 10.77. 2012 ve 2013 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri	211
Şekil 10.78. 2014 ve 2015 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri	211
Şekil 10.79. 2016 ve 2017 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri	212
Şekil 10.80. 2018 yılında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri	212
Şekil 10.81. Spor kategorisine ait kelimeler ve yıllar için basit uyum analizi sonucu	214

ÇİZELGELER DİZİNİ

Çizelge 1.1. Yapısal ve yapısal olmayan veri türlerine ait örnekler	3
Çizelge 2.1. Metin madenciliğinin ilgilendiği problemler	9
Çizelge 2.2. Farklı dillerdeki dokümanlar için yapılan bazı çalışmalar	17
Çizelge 2.3. Çalışma için incelenen yüksek lisan tezleri.....	22
Çizelge 3.1. Örnek kelime kökü.....	31
Çizelge 3.2. İkili ağırlıklandırma yöntemine ait örnek sonuçlar	33
Çizelge 3.3. Frekansa göre ağırlıklandırma yöntemine ait örnek sonuçlar	34
Çizelge 3.4. TFXIDF ağırlıklandırma yöntemine ait örnek sonuçlar	35
Çizelge 3.5. 'Ankilozan' kelimesinin n-gramları	41
Çizelge 3.6. Örnek bir cümlede örtüşen n-gram grupları	41
Çizelge 3.7. Örnek bir cümlede örtüşmeyen n-gram grupları	42
Çizelge 4.1. Karışıklık matrisi.....	53
Çizelge 6.1. Ardışık ve paralel algoritmalar.....	59
Çizelge 8.1. Ankilozan spondilit için 1984 modifiye New York kriterleri (Çoban, 2016)	63
Çizelge 10.1. Kullanılan araçlar ve yapılan işlemler.....	72
Çizelge 10.2. Çalışmada kullanılan yöntemler ve kullanım amaçları	73
Çizelge 10.3. Kategoriler ve toplam veri sayısı	74
Çizelge 10.4. Cinsiyete göre gönderi ve yorum verilerine ait kullanıcı sayıları.....	75
Çizelge 10.5. Sınıflandırma işlemlerinde kullanılan kategori isimleri ve veri sayıları	78
Çizelge 10.6. Ağırlıklandırma yöntemlerine göre sınıflandırma yöntemlerinin başarı oranları.....	79
Çizelge 10.7. SMO yöntemi sonucunda elde edilen karışıklık matrisi (Confusion Matrix).....	79
Çizelge 10.8. Şikayet kategorisine ait gönderilerde yıllara göre kullanıcılar tarafından en çok kullandıkları 50 kelime	81
Çizelge 10.9. Şikayet kategorisine ait gönderi için Friedman testi sonuçları	101
Çizelge 10.10. Boyutların yıllara göre cinsiyet gruplarını açıklama oranları.....	103
Çizelge 10.11. Boyutların şikayet bölgelerini açıklama oranları	103
Çizelge 10.12. Şikayet kategorisine ait gönderilerde ikili kelime kullanımlarını gösteren birliktelik kuralları.....	106
Çizelge 10.13. Şikayet kategorisine ait gönderilerde üç ve üzeri kullanımları gösteren birliktelik kuralı	108

Çizelge 10.14. Şikayet kategorisine ait yorumlarda yıllara göre en çok kullanılan 50 kelime.....	110
Çizelge 10.15. Şikayet kategorisine ait yorumlarda ikili kelime kullanımlarını gösteren birliktelik kuralları.....	112
Çizelge 10.16. Şikayet kategorisine ait yorumlarda üç ve üzeri kelime kullanımları gösteren birliktelik kuralı.....	114
Çizelge 10.17. Doktor kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime.....	114
Çizelge 10.18. Doktor kategorisine ait gönderiler için Friedman testi sonuçları	123
Çizelge 10.19. Doktor kategorisine ait boyutların yılları açıklama oranları.....	125
Çizelge 10.20. Boyutların doktor anahtar kelimelerini açıklama oranları.....	126
Çizelge 10.21. Doktor gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralları	127
Çizelge 10.22. Doktor gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralları	128
Çizelge 10.23. Tahlil kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime	131
Çizelge 10.24. Tahlil kategorisine ait gönderiler için Friedman testi sonuçları	133
Çizelge 10.25. Tahlil kategorisine ait boyutların yılları açıklama oranları	134
Çizelge 10.26. Boyutların tahlil anahtar kelimelerini açıklama oranları	135
Çizelge 10.27. Tahlil kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları.....	137
Çizelge 10.28. Tahlil kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı	138
Çizelge 10.29. Tahlil kategorisine ait yıllara göre yorumlarda en çok kullanılan 50 kelime	140
Çizelge 10.30. Tahlil kategorisine ait yorumlar için ikili kelime kullanımlarını gösteren birliktelik kuralları.....	142
Çizelge 10.31. Tahlil kategorisine ait yorumlar için üç ve üzeri kullanımları gösteren birliktelik kuralları	143
Çizelge 10.32. Tedavi kategorisine ait gönderiler için Friedman testi sonuçları.....	150
Çizelge 10.33. Tedavi kategorisine ait boyutların yılları açıklama oranları.....	152
Çizelge 10.34. Boyutların alternatif tedavi yöntemlerini açıklama oranları.....	153
Çizelge 10.35. Tedavi kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları.....	155
Çizelge 10.36. Tedavi kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı	156
Çizelge 10.37. Tedavi yorum kategorisine ait alternatif tedavi yöntemleri için ikili kelime kullanımlarını gösteren birliktelik kuralları.....	164

Çizelge 10.38. Tedavi yorum kategorisine ait alternatif tedavi yöntemleri için üç ve üzeri kullanımları gösteren birliktelik kuralı	165
Çizelge 10.39. Tedavi kategorisine ait yıllara göre yorumlarda en çok kullanılan 50 kelime.....	165
Çizelge 10.40. Tahlil kategorisine ait yorumlar için ikili kelime kullanımlarını gösteren birliktelik kuralı	167
Çizelge 10.41. Tahlil kategorisine ait yorumlar için üç ve üzeri kullanımları gösteren birliktelik kuralı	168
Çizelge 10.42. Kullanılan ilaçlara ait Friedman testi sonuçları.....	175
Çizelge 10.43. İlaç kategorisine ait boyutların yılları açıklama oranları.....	177
Çizelge 10.44. Boyutların ilaçları açıklama oranları.....	177
Çizelge 10.45. Kullanılan ilaçlara ait ikili kullanımları gösteren birliktelik kuralları.....	179
Çizelge 10.46. Kullanılan ilaçlara ait üç ve üzeri kullanımları gösteren birliktelik kuralları.....	179
Çizelge 10.47. Kullanılan ilaçlara ait Friedman testi sonuçları.....	183
Çizelge 10.48. Diğer hastalıklar kategorisine ait boyutların yılları açıklama oranları	185
Çizelge 10.49. Boyutların diğer hastalıkları açıklama oranları	185
Çizelge 10.50. Konuşulan hastalıklara ilişkin ikili kelime kullanımlarını gösteren birliktelik kuralları	187
Çizelge 10.51. İlaç gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralı	188
Çizelge 10.52. Rapor kategorisine ait gönderiler için yıllara göre en çok kullanılan 50 kelime	191
Çizelge 10.53. Rapor kategorisini temsil eden kelimeler için Friedman testi sonuçları.....	193
Çizelge 10.54. Rapor kategorisine ait boyutların yılları açıklama oranları	194
Çizelge 10.55. Boyutların rapor kategorisine ait kelimeleri açıklama oranları	195
Çizelge 10.56. Rapor gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralı	197
Çizelge 10.57. Rapor gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralları	198
Çizelge 10.58. Kişisel deneyim kategorisine ait gönderilerde en çok kullanılan 50 kelime	201
Çizelge 10.59. Kişisel deneyim ve diyet kategorisini temsil eden kelimeler için Friedman testi sonuçları	203
Çizelge 10.60. Kişisel deneyim ve diyet gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralları	204
Çizelge 10.61. Kişisel deneyim ve diyet gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralı	205
Çizelge 10.62. Spor kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime	207
Çizelge 10.63. Spor kategorisini temsil eden kelimeler için Friedman testi sonuçları	213

Çizelge 10.64. Spor kategorisine ait boyutların yılları açıklama oranları	215
Çizelge 10.65. Boyutların spor kategorisine ait kelimeleri açıklama oranları.....	215
Çizelge 10.66. Spor kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları.....	217
Çizelge 10.67. Spor kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı	218



1. GİRİŞ

Bilimsel arařtırmalarda genellikle geleneksel veri toplama yöntemleri kullanılmaktadır. Kullanılan başlıca klasik yöntemler mülakat yöntemi, gözlem yöntemi ve anket yöntemidir. Mülakat yöntemi, belli bir amaç için iki veya daha fazla kiři arasında yapılan tartıřmalardır. Niřin sorusuna cevap aranılan durumlarda iyi bir veri toplama yöntemidir (Altunıřık vd, 2007). Gözlem yöntemi, sonuçların ve verilerin gözlem yoluyla toplanmasıdır. Gözlem tekniğinin en belirgin özelliđi, gözlenen olguların kendi dođal ortamları içerisinde yer almasıdır. Anket yöntemi, belli bir probleme, amaca veya arařtırmaya yönelik çeřitli kiřilerden bilgi almak, onların düşünce ve tecrübelerinden faydalanmak amacıyla yapılan bir dizi sorulardan oluřan veri toplama şeklidir (řen, 2005).

Anket yöntemi ile bir arařtırmaya yönelik çok sayıda veriyi kolaylıkla toplamak, analiz etmek mümkündür. Ancak anket yönteminin bazı dezavantajları vardır. Anketlerdeki sınırlılıklardan ve kısıtlardan dolayı anketlerden derinlemesine bilgi edinme imkanı yoktur. Anket tekniđi yapısal özelliklerinden ötürü sınırlı bilgiler sağlayabilir. Anket gönderilen ve yapılan kiřiler genellikle eğitim görmüş kiřiler olmasına rağmen kiřilerin, anket sorularını aynı derecede anlayamamaları veya verdikleri cevaplarda açık olmamaları ve rasgele cevap verdikleri düşünülebilir. Arařtırmacının bunu kontrol etme imkanı yoktur. Anket sonuçlarının işlenmesi ve analiz edilmesi istatistiksel bilgisayar programları aracılıđıyla yapılıyor olsa da anket formunun tasarlanması, pilot çalışmanın yapılması ve anketlerin bilgisayara girilmesi fazla zaman almaktadır (Altunıřık vd, 2007). Anket uygulamaları bir zorunluluk içermediđi için soruların cevapsız kalma ve verilen cevapların tutarsız olma olasılıđı vardır. Ayrıca anket formunda bazı sınırlılıklar vardır. Bunlar; (a) arařtırılan veya merak edilen sorunların cevaplarının esnek olmaması, (b) anket yapılan hedef kitlenin duyarlı olduđu bazı konularda çoktan seçmeli sorular kiřilerin hoşuna gitmeyebilir ve sorulara önyargılı olabilir, (c) yapılan anketlerde cevaplanmayan ve yarım kalan sorular olabilir, (d) diđer bir sınırlılık ise zaman sorunudur. Soruların artmasına bađlı olarak cevap süresinin artması, kiřilerin düşünmeden cevap verme olasılıđını hatta ankete hiç cevap vermemesi ihtimalini artırır, (e) ayrıca kiřilerin duygularını,

düşüncelerini ve deneyimlerini anlamak için anketler yetersiz kalmaktadır (Büyüköztürk, 2005).

Günümüzde geleneksel veri toplama yöntemlerinin yanı sıra sosyal ağ paylaşımlarından da yararlanılmaya başlanmıştır. İnternet kullanımının hızla arttığı günümüzde insanların, duygularını, düşüncelerini, deneyimlerini, gündemle ilgili görüşlerini yazılı ve görsel olarak paylaştıkları çeşitli sosyal medya uygulamaları ve siteleri mevcuttur. Bu platformlar üzerinden çeşitli verilere ulaşmak için yazılımlar ve araçlar geliştirilmiştir. İnternet üzerinden veri toplamak için geliştirilen bu araçlar ve yazılımlar sayesinde çok kısa sürede birçok konu hakkında binlerce veya milyonlarca veri toplamak mümkündür. Klasik veri toplama yöntemlerine göre zaman, maliyet ve veri boyutu açısından çok avantajlıdır. Sadece toplanan veriler, bilgi kirliliği, kullanıcı hatası ve eksiklerinden kaynaklanan paylaşımlar olduğundan dolayı bilgiye ulaşmayı zorlaştırmaktadır (Kaşıkçı ve Gökçen, 2014). Yapısal olmayan nitelikteki bu veri türlerinden istenilen bilgilere ve sonuçlara ulaşmak amacıyla yapısal hale dönüştürmek için metin madenciliği teknikleri kullanılır.

Veriler, yapısal veri (structured data) ve yapısal olmayan veri (unstructured data) olmak üzere ikiye ayrılır. Yapısal veriler, daha önceden belirli formatta tanımlanmış, sistemli bir şekilde işlenmiş ve düzenlenmiş, bilgisayar programları tarafından kullanılmaya hazır bilgilerdir. Yapısal olmayan veriler ise belirli bir formatta tanımlanmamış ve çeşitli programlar tarafından doğrudan kullanılmayan verilerdir. Metinsel içerikler ve dokümanlar yapısal olmayan veri türleridir. Bu metinsel verilerin metin madenciliği yöntemleri kullanılarak analiz edilmesiyle bu metinleri temsil eden kavramlar ve özellikler bulunur. Bulunan özellikler sayesinde metinlerin tamamı özetlenebilir ve sınıflandırma işlemleri yapılabilir. Sosyal medya uygulamalarında paylaşılan metinsel içeriklerde yapısal olmayan veri grubuna girdiğinden anlamlı bilgiler çıkarmak ve çeşitli analizler yapabilmek için öncelikle metin madenciliği yöntemlerinin kullanılması gerekir (Dolgun vd, 2014). Yapısal ve yapısal olmayan verilere ait örnekler Çizelge 1.1’de verilmiştir.

Çizelge 1.1. Yapısal ve yapısal olmayan veri türlerine ait örnekler

Yapısal olmayan veri örnekleri	Yapısal veri örnekleri
Pdf, word, text gibi metin dosyaları	Access veri kaynakları
Epostalar	Elektronik işlem tabloları
Web üzerinde tutulan log dosyaları	Finans verileri
Sosyal medya paylaşımları	Sensör verileri

En yaygın kullanılan sosyal medya uygulamalarından biri olan Facebook'ta, 17.07.2018 tarihinden itibaren 9858 üyesi bulunan 'Ankilozan Spondilit Hasta Derneği' grubunda kullanıcılar hastalıkla ilgili çeşitli deneyim ve bilgilerini paylaştıkları için geniş bir bilgi kaynağı oluşturmaktadır. Grubun üye sayısının fazla olması, aktif olarak her gün hastalıkla ilgili çeşitli konularda gönderilerin paylaşılması ve gruba AS hastalarının dışında başka kullanıcılar dahil edilmediğinden dolayı çalışma için bu hastalık seçilmiştir. Grupta paylaşılan içerikler yapısal olmayan veri kaynaklarıdır. Bu çalışmanın amacı yapısal olmayan bu verileri metin madenciliği yöntemlerini kullanarak yapısal hale dönüştürerek en çok şikayet edilen hususları, kullanılan tedavi yöntemlerini, en çok doktor tavsiyesi sorulan illeri, kullanılan ilaçları, rapor, tahlil, spor hakkındaki bilgileri, AS'nin diğer hastalıklarla ilişkisini ve kişisel deneyimleri belirlemektir. Metin sınıflandırma algoritmalarının terim frekansı (TF), ters doküman frekansı (TFxIDF) ve binary ağırlıklandırma yöntemlerine göre sınıflandırma performansları ölçülerek literatüre kazandırılacaktır.

Bu tez çalışmasında veriler şikayet, doktor, ilaç, tedavi, rapor, tahlil, kişisel deneyim ve diyet, diğer hastalıklar ve spor kategorileri başlıkları altında toplanmıştır. İlk olarak TF, TFxIDF ve binary ağırlıklandırma yöntemlerine göre Naive Bayes, K-en yakın komşu, karar ağaçları ve SMO algoritmalarının sınıflandırma performansları karşılaştırılmıştır. Oluşturulan her bir kategori için çeşitli bilgi çıkarımları yapılmış, sonuçlar çeşitli infografikler ve çizelgeler halinde sunulmuştur. Yapılan bu bilgi çıkarımları ile, en çok belirlenen şikayet bölgeleri sayesinde doktorlar hastalığın belirtisi ile bulunan bu şikayet yerleri ile ilişkini incelemek için çeşitli çalışmalar yapmasına olanak sağlayabilir. En çok doktor tavsiyesi sorulan illere göre Sağlık Bakanlığı tarafından bu çıkarımlar kullanılabilir ve en çok doktor tavsiyesi sorulan illerde Romatoloji bölümlerinin sayısı artırılabilir. En çok kullanılan ilaç isimleri ilaç firmaları, en çok konuşulan alternatif tedavi yöntemleri ve spor aktiviteleri hastalar

tarafından çeşitli amaçlar için kullanılabilir. İkili ve üzeri kelime kullanımlarını tespit etmek amacıyla birliktelik analizi işlemleri yapılmıştır. Sonuçlar çizelgeler halinde çalışmada verilmiştir. Ayrıca sınıfları temsil eden anahtar kelimelerin yıllara göre farklılık gösterip göstermediği test edilmiş ve test sonuçları çizelgeler halinde verilmiştir.



2. METİN MADENCİLİĞİ

2.1. Metin Madenciliğinin Tanımı

Metin madenciliği (MM), büyük boyuttaki belgelerin analizi ve metin tabanlı verilerden enformasyon çıkarmak amacıyla metinlerin analiz edilmesi süreci olarak tanımlanabilir (Dolgun vd, 2014; Yıldız, 2017). Başka bir tanımda metin madenciliği, önceden bilinmeyen ve önemli olan bilgilerin çıkarılması amacıyla çok sayıda dokümanı analiz eden bir teknoloji olarak tanımlanmaktadır (Göker ve Tekedere, 2017). Ayrıca MM, çeşitli yöntemler kullanılarak yapısal olmayan verilerin işlenmesini, işlenen verilerin bilgisayar ortamında incelenmesi için farklı gösterim formlarına dönüştürülerek anlamlı bilgilerin çıkarılmasını sağlayan yoğun bir süreçtir (Seçkin, 2011). Metin madenciliği işlemlerinde birçok alan ile birlikte çalışılmaktadır. Birlikte kullanılan alanlar Şekil 2.1’de verilmiştir.



Şekil 2.1. Metin madenciliği işlemlerinde birlikte çalışılan alanlar

Yapay zeka (Artificial intelligence, AI), insan tarafından yapılan bir görev ve davranışın bilgisayar veya bilgisayar kontrollü bir makine tarafından yapılma

yeteneğidir (Kayıkçı ve Bozkurt, 2018). Amaç insan gibi düşünen, öğrenen ve bir şeyler meydana getiren makineler yapmaktır. Tavlama benzetimi (Simulated Annealing), uzman sistemler (Expert Systems), bilgisayarlı görme (Computer Vision), konuşma tanıma (Speech Recognition), doğal dil işleme (Natural Language Processing), yapay sinir ağları, (ANN: Artificial Neural Networks), bulanık mantık (Fuzzy Logic), genetik algoritmalar (Genetic Algorithms), karınca algoritması (Ant Algorithms), kaotik modelleme, hibrid (karma) sistemler başlıca yapay zeka tekniklerini oluşturmaktadır (Erdal, 2018).

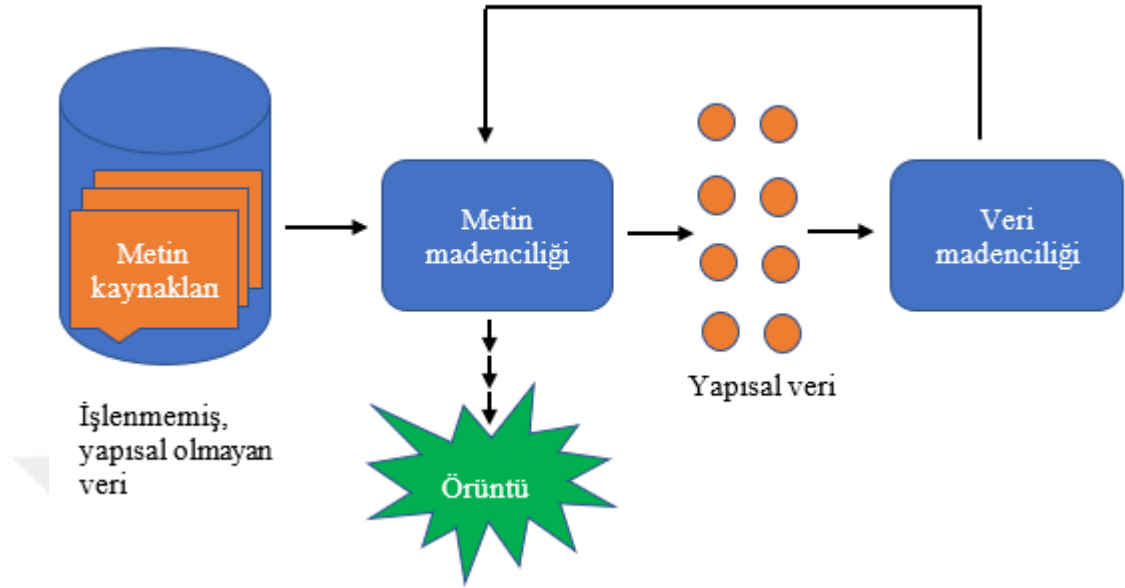
2.2. Metin Madenciliğinin Tarihsel Gelişimi

Metin veri madenciliği (Text Data Mining) ve metin veritabanlarından bilgi keşfi (Knowledge Discovery from Textual Databases) olarak adlandırılan metin madenciliği, 1960'lı yıllarda bilgi erişim sistemleri ve doğal dilin işlenmesine yönelik yapılan çalışmalar sonucu ortaya çıkmıştır. Metinlere erişim, metinlerden bilgi çıkarımı, metin kategorizasyonu, metinlerin yapısal hale dönüştürülmesine yönelik çalışmalar, 1990'lı yıllarda oldukça yaygınlaşmıştır (Oğuz vd, 2007). 1990'lı yıllardan sonra internetin ve kişisel bilgisayarların hızlıca yaygın hale gelmesiyle bilgiye erişim hızla artmaya başlamıştır. Bu durum beraberinden bilgi patlamasını meydana getirmiş ve istenilen bilgiye ulaşmayı zorlaştırmıştır. Bu nedenle istenilen bilgiye erişim sorununu çözmek için yeni yöntemler geliştirilmeye başlanmıştır. Özellikle 2000'li yıllardan sonra metin madenciliğine olan ilgi artmaya başlamıştır (Haltaş vd, 2015).

2.3. Metin Madenciliğinin Veri Madenciliği ile İlişkisi

Metin madenciliği veri madenciliğinin bir alt dalı olarak ele alınır ve aralarında interaktif bir ilişki söz konusudur (Seçkin, 2011). Bu ilişki Şekil 2.2 'de gösterilmiştir. Veri madenciliği kısaca veritabanlarından bilgi keşfi yapılması işlemi olarak tanımlanabilir (Hark vd, 2017). Veri madenciliği birçok yöntem ve yaklaşımı ham veri üzerine uygulayarak anlamlı bilgiler elde edilmesini sağlar. Veri madenciliğinde kullanılan veritabanları düzgün yapısal verileri barındırır. Metin madenciliği işlemlerinde ise kullanılan veriler yapısal olmayan metinsel verileri içermektedir. Yapısal olmayan verilere veri madenciliği yöntemlerini uygulayabilmek için metin madenciliği yöntemleri ile yapısal hale getirilmesi gerekir (Yıldız ve Demirci, 2017). Metin madenciliği yöntemleri ile bu metinleri temsil eden özellikler bulunur ve bu

özellikler nümerik sayılara dönüştürülerek veri madenciliği algoritmalarının uygulanabilmesi için hazır hale getirilir (Gürbüz ve Özyirmidokuz, 2018).



Şekil 2.2. Metin madenciliği ve veri madenciliği arasındaki ilişki

Yapısal hale dönüştürülen verilere sınıflandırma, kümeleme ve birliktelik analizi modelleri uygulanabilir. Örneğin veri madenciliği işlemlerinde birliktelik analizi, kullanıcı tarafından tanımlanan nesnelerin birlikte bulunma sıklıkları ve nesnelerin birbiriyle olan ilişkilerini inceler. Metin madenciliği işlemlerinde ise metin içerisinde kelimelerin birbirini takip etme sıklıkları incelenir (Silahtaroglu ve Demircan, 2013).

2.4. Doğal Dil İşleme

Metin madenciliği işlemlerinde sıklıkla kullanılan doğal dil işleme (Natural Language Processing, NLP), insana özgü dillerin bilgisayara tanıtılması, işlenmesi ve doğru bir şekilde analiz edilmesini sağlayan bir çalışma alanıdır (Başkaya ve Aydın, 2017). Doğal dil işleme karmaşık bir süreçtir. Sadece metin içerisindeki kelimeleri bulmak ve gruplandırmak yeterli değildir. Kelimelerin aldığı önek ve sonekler ile birlikte cümle içindeki anlamlarının da bilinmesi gerekir. Örneğin “sıcak“ (su, hava, kanlı) kelimesi birçok yerde farklı anlamda kullanılmış olabilir. Özellikle kelimelerin hangi anlamda kullanıldığını anlamada yapay zeka ve dilbilim yaklaşımları sıklıkla kullanılmaktadır (Ergün, 2016).

Doğal dil işlemede iki yaklaşım söz konusudur. Birinci yaklaşımda metnin yazıldığı dilin gramatik yapısı dikkate alınarak metin bir bütün olarak analiz edilir. İkinci yaklaşım ise “Bags of words” olarak adlandırılır. Bu yaklaşımda metinler parçalanarak incelenir, gramer yapısı dikkate alınmaz. Metinler cümlelere, cümleler kelimelere, kelimeler ise köklerine ayırarak köklerin metin içerisinde geçme sıklıklarına göre analizler yapılır (Çeliksü, 2017). Bunun için analiz edilecek metinlerin yazıldığı dilin özelliklerinin çok iyi bilinmesi gerekir. Örneğin Türkçe yapısal olarak eklemeli diller grubunda yer alır. Kelime kökleri değişmez ve sonuna aldığı ekler ile yeni kelimeler türetilir. Ayrıca bu ekler zaman, hal ve aitlik gibi durumları belirtmeye de yarar. Eklenen ekler kelimenin cümle içerisindeki anlam ve görev değişikliğine yol açtığı gibi kelimenin cümle dizimindeki yerini de değiştirebilir. Dilin tüm bu karışık özellikleri bilgisayar destekli doğal dil işlemede belli problemlere neden olmaktadır (Savaşan, 2011).

Günümüzde doğal dil işleme yazılımlarının birçoğu İngilizce gibi sık kullanılan diller için geliştirilmiştir. Bu dillerin yapısı Türkçeden farklı olduğu için geliştirilen bu doğal dil işleme sistemleri Türkçe için kullanılmaz. Kullanıldığı zaman hatalı sonuçlar elde edilir (Savaşan, 2011). Türkçe için İstanbul Teknik Üniversitesi Bilgisayar Bilişim Fakültesi’nde Doğal Dil İşleme Grubu’nun geliştirdiği İTÜ Doğal Dil İşleme Yazılım Zinciri ve Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği bölümü Kemik Doğal Dil İşleme Grubu’nun geliştirdiği Java tabanlı Kemik isimli sözlük programları mevcuttur (Çeliksü, 2017).

Doğal dil işleme çalışmalarının temel amacı metinlerin sayısal olarak ifade edilmesinden çok doğal dillerin yapı ve işlevlerinin bilgisayara tanıtılması, bilgisayar tarafından doğal dillin kullanılmasını sağlamak ve bilgisayar ile insanlar arasındaki iletişimin sağlamaktır (Seçkin, 2011).

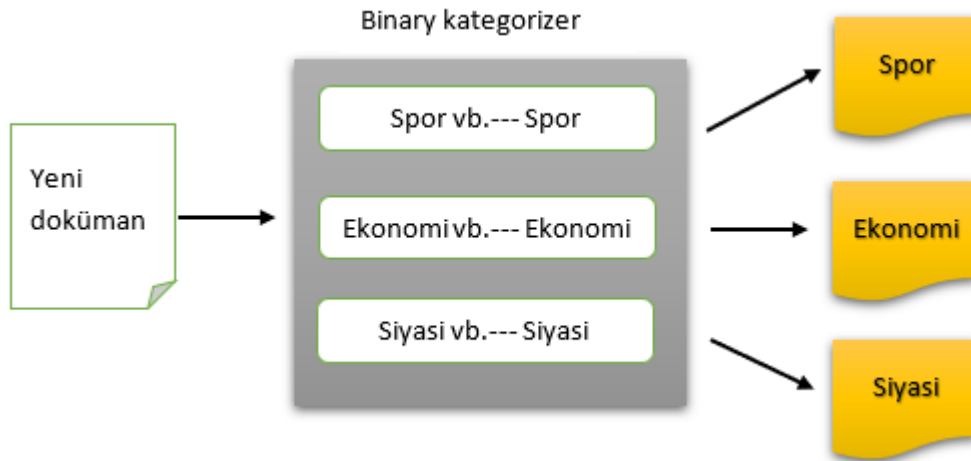
2.5. Metin Madenciliğinin Uygulama Alanları

Metin madenciliği işlemlerinde genelde Çizelge 2.1 verilen problemlerle ilgilenilir. Bunlardan bazıları Çizelge 2.1’de verilmiştir.

Çizelge 2.1. Metin madenciliğinin ilgilendiği problemler

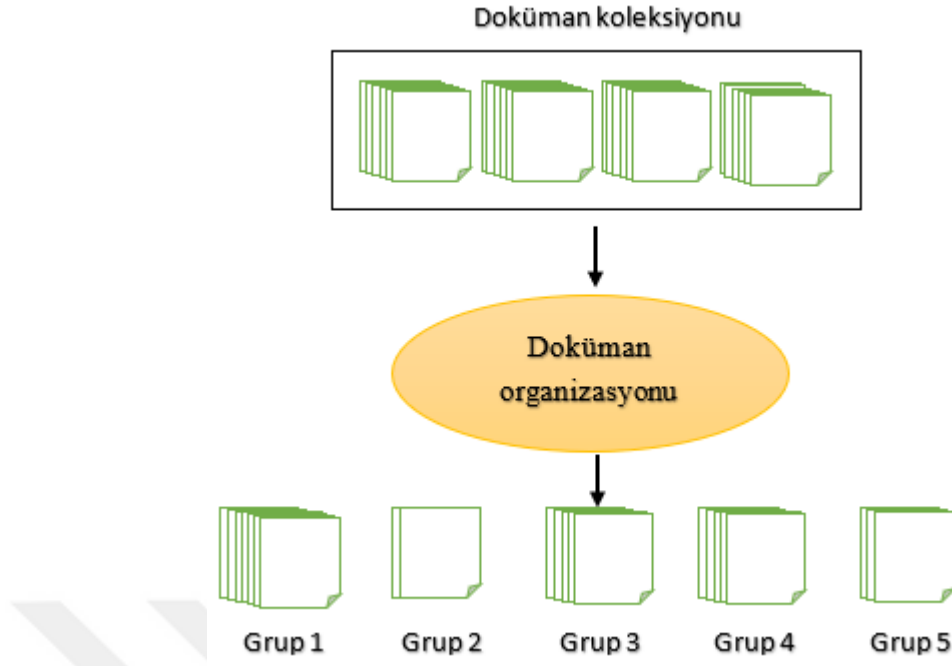
Metin madenciliğinin ilgilendiği problemler	Örnekler
Sınıflandırma	Belgelerin kategorizasyonu, belgelerin sıralanması, spam uyarı algılama
Kümeleme	Dokümanların kümelenmesi, belgelerin benzerliği, kelime gruplarının elde edilmesi
Web madenciliği	Web yapı madenciliği, web içerik madenciliği, web kullanım madenciliği
Doğal dil işleme	Kelime köklerin bulunması, kelimenin sözlükteki doğru halinin tespit edilmesi, konuşma parçalarının etiketlenmesi
Bilgi ve kavram çıkarma	Bilgi çıkarma, ilişki çıkarma, eş dizin, kelime birlikteliği, duygu analizi
Bilgi getirimi	Arama optimizasyonu, indeks çevirme

Sınıflandırma, önceden belirlenmiş sınıflara yeni bir kaydın atanması işlemidir. Eposta spam mesajlarının filtrelenmesi, belgelerin sınıflandırılması gibi birçok alanda uygulanmaktadır (Haltaş vd, 2015). Sınıflandırma yöntemine ait örnek gösterim Şekil 2.3'te verilmiştir.



Şekil 2.3. Metin kategorizasyonu

Kümeleme, sınıfı belli olmayan veya herhangi bir kategoriye atanmamış dokümanların herhangi bir ön bilgi olmadan gruplanması işlemidir. Kümeleme yönteminde birbirine benzer olan dokümanlar aynı küme içerisine atanır (Tunalı ve Bilgin, 2012). Kümeleme yöntemine ait örnek gösterim Şekil 2.4'te verilmiştir.

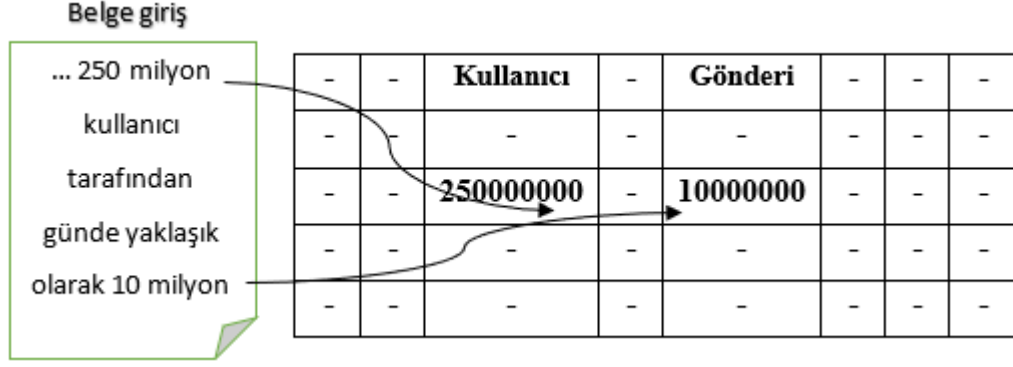


Şekil 2.4. Gruplara göre dokümanları organize etme

Web madenciliği, yapısal olmayan web verilerinin analizinde kullanılmaktadır. Web yapı madenciliği, web içerik madenciliği ve web kullanım madenciliği olmak üzere üç gruba ayrılır. Yapı madenciliğinde web siteleri ve sayfaları arasındaki ilişkiler ve bağlantılar incelenmektedir. İçerik madenciliğinde faydalı bilgiler çıkarmak amacıyla web sitelerinin içerikleri incelenir. Web kullanım madenciliğinde ise web kayıt kullanıcı bilgileri, oturum ve hareket bilgileri incelenerek kullanıcıların davranış ve tutumları tespit edilir (Dolgun vd, 2014; Ergün vd, 2013).

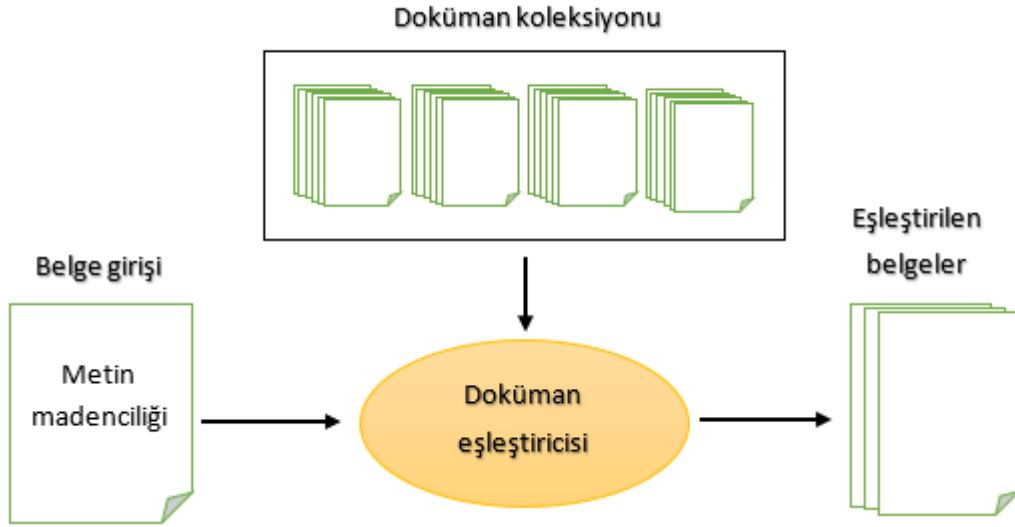
Doğal dil işleme, insanlara özgü olan dillerin bilgisayara tanıtılması, işlenmesi ve analiz edilmesiyle ilgilenen bir çalışma alanıdır. Konuşma etiketleme, intihal tespiti gibi alanlarda sıklıkla kullanılmaktadır. Metin madenciliği ön işlem aşamasında önemli bir yere sahiptir (Başkaya ve Aydın, 2017).

Bilgi ve kavram çıkarma, yapılandırılmamış metinsel veriler yapısal hale dönüştürüldüğünde uygulama alanına ilişkin faydalı bilgilerin otomatik olarak çıkarılmasıdır (Seçkin, 2011). Bilgi çıkarımına ilişkin örnek Şekil 2.5'te verilmiştir.



Şekil 2.5. Bilgi çıkarımı

Bilgi getirimi, yapısal olmayan bir veri kaynağından istenilen yapıya göre bilginin elde edilmesidir. Arama motorlarında sıkça kullanılmaktadır. Arama motorlarında yapılan her bir arama bilgi getirimi kabul edilebilir (Şeker, 2010). Bilgi getirimine ilişkin örnek Şekil 2.6’da verilmiştir.



Şekil 2.6. Bilgi getirimi

Duygu analizi, dokümanlarda geçen duygu, tutum ve görüş gibi öznel ifadelerin çıkarılmasını amaçlar. Makine öğrenmesi yöntemleri kullanılarak duygu analizi işlemleri gerçekleştirilir. Metinler olumlu/olumsuz, pozitif/negatif gibi ikili sınıfa ayrılması hedeflenir (Şeker, 2015; Yigit vd, 2017). Duygu analizi ile kişilerin bir durum, olay, hizmet, ürün, kurum, reklam, film gibi benzeri konular hakkındaki duygu ve düşünceleri tespit edilmeye çalışılır (Kasaba ve Yıldıztepe, 2016).

Metin madenciliği birçok sektör ve alanda uygulanmaktadır. Nicandro vd (2015) güvenlik uygulamaları alanında; Lee ve Park (2018) istihbarat alanında; Kaşıkçı ve Gökçen (2014) internet – yazılım alanında; Kılınç vd (2016) akademik uygulamalarda ve bilimsel araştırma alanlarında; Tsai ve Wang (2017) finans alanında; João vd (2016) pazarlama alanında; Yiğit vd (2017) müşteri ilişkileri yönetimi alanında; Akyel ve Seçkin (2012) hile denetimi alanında; Fluck ve Hofmann-Apitius (2014) biyoloji alanında; Kılıç vd (2015) basın yayın sektörü alanında; Kırmacı ve Oğul (2015) yazar tanıma sistemleri alanında; İlhan vd (2008) soru cevap sistemleri alanında; Badal vd (2015) sağlık ve medikal araştırmaları alanlarında metin madenciliği ile ilgili çalışmalar yapmıştır.

2.6. Metin Madenciliği İçin Kullanılan Araçlar

Metin madenciliğinin tüm işlem ve aşamalarında çeşitli bilgisayar yazılımları kullanılmaktadır. Metin madenciliğinin ilk aşaması olan veri toplama aşamasında bile dijital ortamlardan veri toplamak için yazılımlara ihtiyaç duyarız. Örneğin, Twitter’den veri çekmek istenildiğinde Knime programının araçlarından biri olan Twitter API kullanılarak Twitter’den rahatlıkla veri toplanabilir. Dijital ortamlardaki bu veriler, kopyala yapıştır yöntemiyle toplandığında ne kadar çok vakit alacağı aşıkardır.

Metinlerin sayılaştırılması, kelime köklerinin bulunması ve kelimelerin ağırlıklandırılması ile metinleri yapısal olmayan yapıdan yapısal hale dönüştürülmesinde doğal dil işleme için geliştirilen yazılımlar kullanılır. Bu aşamada Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği bölümü Kemik Doğal Dil İşleme Grubu’nun geliştirdiği java tabanlı Text2ArffV5 yazılımı kullanılabilir. Text2arffV5 yazılımı ile metinlerin kelime kökleri bulunabilir, 12 farklı ağırlıklandırma yöntemi ile kelime kökleri ağırlıklandırılabilir, kelime tipleri ve kelime ekleri çıkarılabilir, 2’li ve 3’lü n-gramlar oluşturulabilir, kelime ve kökler filtrelenebilir, birliktelik matrisi oluşturulabilir, kavram genelleştirme ve kümeleme işlemleri yapılabilir. Tunalı ve Bilgin (2012) Türkçe metinlerin önışlemeden geçirilmesi amacıyla Java tabanlı PRETO aracı geliştirilmiştir. Bu araç kelime köklerini bulma, durak sözcük filtreleme, istatistiksel terim filtreleme ve n-gram oluşturma özelliklerine sahiptir (Tunalı ve Bilgin, 2012; Yurt, 2015).

Metin önişlemleri yapıldıktan sonra yapısal formata dönüşen verileri analiz etmek için WEKA, KNIME, RAPID MINER, SPSS MODELER ve R dili gibi araçlar kullanılabilir. WEKA (Waikato Environment for Knowledge Analysis), birçok makine öğrenmesi algoritmalarını bünyesinde barındıran, Java programlama dili ile geliştirilmiş, açık kaynak kodlu ve GNU (GNU's Not Unix) lisanslı bir yazılımdır (Seçkin, 2011; Yurt, 2015). WEKA kendine özgü Arff (Attribute Relationship File Format) dosya yapısına sahiptir. Veri madenciliği, veri analizi, veri görselleştirme ve iş zekası gibi işlemler yapılabilir. Veri madenciliği işlemlerinden sınıflandırma, kümeleme ve birliktelik analizi yapılabilir.

KNIME veri analizi ve veri madenciliği işlemlerinde kullanılan Java tabanlı bir programdır. İşlemlerin yapılacağı bir görsel arayüz bulunmaktadır. Görsel arayüz yapılacak işleme göre düğümler birbirine bağlanır. İşleme uygun bir akış şeması elde edildikten sonra düğümler çalıştırılır ve sonuçlar elde edilir.

Diğer bir Java tabanlı yazılım ise RAPID MINER'dır. Metin madenciliği, veri madenciliği ve makine öğrenme işlemleri yapılabilir. Literatürlerde geçen birçok algoritmayı uygulayabilmektedir. 22 adet farklı dosya formatındaki veriyi işleme özelliğine sahiptir (Bastem ve Şeker, 2017).

Açık kaynak kodlu programlama dili olan R dili, Yeni Zelanda Auckland Üniversitesi'nden Ross Ihaka ve Robert Gentleman tarafından geliştirilmiştir (Doğan vd, 2017). İlk sürümü 2000 yılında yayınlanmıştır. Lineer ve non-linear modelleme, klasik istatistikler, zaman-seri analizleri, sınıflandırma, kümeleme, grafik işlemleri gibi alanlarda kullanılabilir. Özellikle veri analistleri için geniş grafik özellikleri sağlamaktır (Hark vd, 2017). Ayrıca kullanıcılar tarafından geliştirilen paketler sayesinde birçok alanda kolaylık sağlamaktadır. Karmaşık bir kod yapısına sahip olmadığı için kolayca öğrenilebilir. R dili ve R studio araçları ücretsizdir ve bu araçlar kullanılarak programlar geliştirilebilir (Kılınç vd, 2016).

Bunların dışında metin madenciliği işlemlerinde kullanılan birçok program vardır. Ancak bunları çoğu İngilizce destekli yazılımlardır. Bunların bazıları şu şekildedir; Angoss, Attensity, AUTINDEX, Autonomy, Averbis, Basis Technology, Clarabridge, ClearForest, Complete Discovery Source, Endeca Technologies, Expert System S.p.A., FICO Score, GATE, General Sentiment, IBM SPSS Text Analytics, IBM Intelligent Miner Data Mining Suite, Intellexer, Inxight, LanguageWare,

Language Computer Corporation, Lexalytics, LexisNexis, LingPipe, Luminoso, Mathematica, MATLAB Text Analytics Toolbox, Medallia, Megaputer Intelligence, NetOwl, RapidMiner, SAS Text Miner and Teragram, Semantria, Smartlogic – Semaphore, Spy-EM, STATISTICA Text Miner, Sysomos, Textalytics, Vantage Point, Vivisimo/Clusty, WordStat, Xpresso (Ergün, 2016; Seçkin, 2011).

2.7. Literatür Taraması

Metin madenciliği yöntemleri ve problemleri ile ilgili birçok akademik çalışma yapılmıştır. Türkiye’de Türkçe kaynaklı metinlerin analiz edilmesiyle ilgili birçok çalışmalar mevcuttur. Metin madenciliği kullanılarak yapılan toplam 63 makale bu çalışma için incelenmiştir. Türkiye’de yapılan çalışmaların çoğu metin sınıflandırma yöntemiyle ilgilidir.

Çalış vd (2012), reklam içerikli epostaların metin madenciliği yöntemleri ile sınıflandırmak amacıyla çalışma yapmıştır. Çalışmada sınıflandırma algoritmaları olarak destek vektör makinesi, K-en yakın komşu ve Naive Bayes algoritmaları kullanılmıştır. K-en yakın komşu algoritması ile %96,5 doğrulukta sınıflandırma başarısına ulaşılmıştır.

Kaşıkçı ve Gökçen (2014), metin madenciliği yöntemleri ile e-ticaret sitelerinin belirlenmesi hedeflenmiştir. Sınıflandırma algoritmaları olarak K-en yakın komşu ve Naive Bayes sınıflandırma algoritmaları kullanılmıştır. Naive Bayes algoritması ile %85,30 doğrulukta sınıflandırma başarısı elde edilmiştir. Doğal dil işleme kütüphaneleri olarak Zemberek ve Jsoup kullanılmıştır. Çalışma sonucunda bir uygulama geliştirilmiştir.

Yıldız (2017), metin madenciliğinde anahtar kelime seçimi için bir çalışma yapmıştır. Bu çalışmada ön işlem aşamasından sonra anahtar kelimeleri belirlemek amacıyla TFxIDF ağırlıklandırma yöntemini, ki-kare istatistiği ve bilgi kazancı yöntemlerini kullanmıştır.

Göker ve Tekedere (2017), Fatih Projesi’ne ilgili görüşleri metin madenciliği yöntemleri kullanarak değerlendirmek amacıyla bir sınıflandırma çalışması yapmıştır. Bu çalışmada Naive Bayes, K-NN, karar ağaçları (J48), SMO ve RBF Network algoritmaları uygulanmıştır. Test amacıyla k-katlı çapraz doğrulama yöntemi

kullanılmış ve %88,73 başarı oranı ile SMO algoritmasının daha başarılı olduğu tespit edilmiştir.

Kılınç vd (2016), bilimsel makale tasnifi için bir çalışma yapmıştır. Sınıflandırma algoritmaları olarak Naive Bayes, karar ağaçları (J48) ve K-NN yöntemleri kullanılmıştır. En yüksek başarı oranı K-NN algoritması (%96,67) ile elde edilmiştir.

Karaca ve Görgünoğlu (2012), köşe yazılarının analize hazır hale getirilmesi amacıyla ColumnREADY adında bir uygulama yazılımı geliştirilmiştir. Türkçe doğal dil işleme için Zemberek kütüphanesi kullanılmıştır. Ağırlıklandırma yöntemi olarak TFxIDF yöntemi kullanılmıştır.

Kılıç vd (2015), online haber sitelerinde yer alan son dakika haberlerini sınıflandırmak amacıyla bir çalışma yapmıştır. Sınıflandırma işlemlerinde Naive Bayes, SMO ve C4.5 algoritmaları kullanılmıştır. En başarılı sonuç %82 başarı oranı ile Naive Bayes yönteminden elde edilmiştir. Başka bir çalışmada Kılıç vd (2015), TESDF (Ters eksik sınıf doküman frekansı) ve SADF (Sınıflar arası doküman frekansı) olmak üzere iki yeni metin ağırlıklandırma yöntemi geliştirilmiştir. TFxIDF yöntemine dayanan bir ağırlıklandırma yöntemidir. Kelime köklerini bulmak için Zemberek kütüphanesi kullanılmıştır. Sınıflandırma yöntemleri olarak C4.5, Naive Bayes ve SMO algoritmaları uygulanmıştır. Bu iki yeni ağırlıklandırma yöntemi TFxIDF yöntemine göre daha başarılı sonuçlar vermiştir. Üç ağırlıklandırma yöntemine göre en iyi sonuç Naive Bayes sonucunda elde edilmiştir.

Tunalı ve Bilgin (2012), farklı kök bulma yöntemlerinin Türkçe metinlerin kümelenmesinde etkili olup olmadığını test etmek amacıyla bir çalışma yapmıştır. Farklı kök bulma yöntemleri olarak Zemberek, ek çıkarıcı, sabit önek 3, sabit önek 5 ve sabit önek 7 yöntemleri baz alınmıştır. Kümeleme yöntemi olarak Küresel K-means algoritması kullanılmıştır. Çalışma sonucunda Zemberek ve sabit önek 5 yöntemlerinin diğer kök bulma yöntemlerine göre kümeleme işlemlerinde daha iyi sonuçlar verdiği gözlenmiştir.

Karaca vd (2015), sınıflandırma işlemlerinde benzerlik hesaplama yöntemlerini değerlendirmek amacıyla bir çalışma yapmıştır. Çalışmada köşe yazılarını sınıflandırmak için KNN algoritması ile toplam 15 farklı benzerlik hesaplama yöntemi

kullanılmıştır. Cosine, Pearson Correlation, Bray Curits, Tanimoto ve Dice benzerlik tekniklerinde %100 doğrulukta sınıflandırma başarısı elde edilmiştir.

Açıkalın ve Bayazıt (2016), Türkçe metinlerin sınıflandırılmasında ön işlem aşamasının önemi belirlemek amacıyla bir çalışma yapmıştır. Kelime köklerini belirlemek amacıyla Zemberek kütüphanesi kullanılmıştır. Çalışmada sınıflandırma algoritmaları olarak Naive Bayes, Destek Vektör Makineleri ve Rastgele Orman yöntemleri kullanılmıştır. Ön işlem uygulayarak yapılan sınıflandırma işlemlerinde, ön işlemsiz olarak yapılan sınıflandırma işlemlerine göre %2,26 ile %4,94 arasında daha iyi sonuç elde edilmiştir.

Ayata vd (2017), Türkçe yazılmış tweetlerin olumlu ve olumsuz olarak sınıflandırmak amacıyla duygu analizi yapılmıştır. Sınıflandırma işlemlerinde destek karar makinesi ve Rastgele Orman algoritmaları kullanılmıştır.

Meral ve Dirı (2014), tarafından Twitter üzerinde bir duygu analizi çalışması yapılmıştır. Veriler olumlu, olumsuz ve nötr olarak 9 başlık altında toplanmıştır. Doğal dil işleme olarak kelime tabanlı yöntem için Zemberek kütüphanesi ve n-gram modeli kullanılmıştır. Sınıflandırma yöntemleri olarak Rastgele Orman, Naive Bayes ve Destek Vektör Makinesi kullanılmıştır. Test için 10 katlı çapraz doğrulama yöntemi kullanılmıştır.

Yıldız ve Demirci (2017), kurumsal epostaların sınıflandırılması için bir sistem geliştirilmiştir. Kelime köklerini belirlemek amacıyla Zemberek kütüphanesi kullanılmıştır. Çalışmada sınıflandırma yöntemleri olarak Naive Bayes, Naive Bayes Multinomial, SVM, YSA, AdaBost, J48 ve JRIP algoritmaları kullanılmıştır. 10 katlı çapraz doğrulama yöntemi kullanılmıştır. En başarılı sonuç %95,70 oranında Naive Bayes Multinomial ile elde edilmiştir.

Başkaya ve Aydın (2017), ekonomi, politika, spor ve sağlık kategorilerine ait haber metinlerinin metin madenciliği yöntemleri ile sınıflandırılması amaçlanmıştır. Kelime köklerinin bulunmasında Zemberek kütüphanesi kullanılmıştır. Öznitelik seçimlerinde n-gramlar kullanılmıştır. Ağırlık yöntemi olarak TF ve TFxIDF ağırlık yöntemleri kullanılmıştır. Ağırlık yöntemlerine ve öznitelik seçimlerine göre sınıflandırma işlemleri yapılmıştır. Sınıflandırma algoritmaları olarak Naive Bayes, Destek Vektör Makineleri, J48 ve Rastgele Orman yöntemleri kullanılmıştır. TF

ağırlıklandırma yöntemin TFxIDF yöntemine göre daha başarılı sonuçlar elde edilmiştir.

İngilizce gibi farklı dillerdeki dokümanlar için yapılan bazı çalışmalar Çizelge 2.2’de verilmiştir. Çizelgede yapılan akademik çalışmaların çalışma başlığı, metin madenciliğinin ilgilendiği problem, kullanılan yöntemler ve hangi alan için uygulandığına dair bilgiler verilmiştir.

Çizelge 2.2. Farklı dillerdeki dokümanlar için yapılan bazı çalışmalar

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Park vd, 2018)	Kümeleme	K-Means, Çoklu görselleştirme teknikleri	Psikiyatri	Metin madenciliği ve görselleştirme yaklaşımı ile Reddit’te yer alan üç online ruh sağlığı topluluğundaki tematik benzerliğin, farklılığın ve üyeliğin incelenmesi
(Jimeno Yepes vd, 2015)	Adlandırılmış varlık tanıma (named entity recognition,)	WSD (word sense disambiguation), Stanford NER (named entity recognition, CRF (conditional random fields) modeli	Sağlık	Twitter’da hastalıkları, ilaçları ve belirtileri tanımlama
(Korkontzelos vd, 2016)	Duygu analizi	CRF modeli	Biyomedikal	Tweetlerden ve forum yazılarından advers ilaç reaksiyonlarının çıkarılmasında duygu analizinin etkisinin analiz edilmesi

Çizelge 2.2. (devam)

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Sarker ve Gonzalez, 2015)	Sınıflandırma ve doğal dil işleme	Naive Bayes, Destek vektör makineleri, maksimum entropi	Biyomedikal	Advers ilaç reaksiyon tespiti için çoklu korpus eğitimi ile taşınabilir otomatik metin sınıflandırma
(Okuhara vd, 2018)	Kümeleme ve sınıflandırma	Japonca metinleri destekleyen niceliksel içerik analizi için bir program olan KH Coder kullanılmıştır. Hiyerarşik kümeleme, ağ analizi ve Jaccard benzerlik katsayısı	Sağlık bilimleri	Japon Pro- ve Anti-HPV aşılama web siteleri içeriklerinin metin madenciliği ile analizi
(Zhou ve Fu, 2018)	İlişki çıkarımı	Kosinüs benzerliği, terim ağırlıklandırma için doküman frekansı (IDF) ve anahtar kelimeler için ceza ağırlıklandırma (PWK- penalty weights for keywords) yöntemi kullanılmıştır	Tıp	PubMed'in metin madenciliğine dayanan gen hastalığı birliği üzerine araştırması

Çizelge 2.2. (devam)

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Amrit vd, 2017)	Sınıflandırma	Naive Bayes, rastgele orman ve destek vektör makinesi	Sağlık bilimleri	Metin madenciliği ve makine öğrenimi aracılığıyla çocuk istismarını belirleme
(Goh ve Ubeynarayana, 2017)	Sınıflandırma	Destek vektör makinesi (SVM), lineer regresyon (LR), rasgele orman (RF), K-en yakın komşu (KNN), karar ağacı (DT) ve Naive Bayes (NB)	İş sağlığı ve güvenliği	İnşaat kazası anlatılarının sınıflandırılması: Bir metin madenciliği değerlendirilmesi
(Basaldella vd, 2017)	Varlık tanıma	OGER varlık tanıyıcı, yapay sinir ağları ve koşullu rasgele alanlar (CRF)	Biyomedikal	Hibrid yaklaşımı kullanarak biyomedikal alanda varlık tanıma
(Karystianis vd, 2018)	Sınıflandırma	Kelime türeten kurallara dayanan bilgi odaklı yaklaşım, yapay sinir ağının kullanıldığı makine öğrenimi ve sinir ağı ile ilk iki yöntemi birleştiren hibrit yaklaşımı	Psikiyatri	Psikiyatrik değerlendirme notlarından semptom şiddetinin otomatik olarak incelenmesi

Çizelge 2.2. (devam)

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Papanikol aou vd, 2016)	Adlandırılmış varlık tanıma ve kümeleme	Adlandırılmış Varlık Tanıma (NER), TextQuest algoritması, TF-IDF ağırlıklandırma, benzerlik hesabı için Tanimoto katsayısı, Pearson katsayısı ve basit kosinüs benzerliği, kümelemek için (Markov Cluster Algorithm) (MCL) Markov kümeleme algoritması	Biyomedikal	Drugquest-İlaç asosiyasyonu keşfi için bir metin madenciliği iş akışı
(Badal vd, 2015)	Sınıflandırma	Destek vektör makinesi	Tıp	Protein yerleştirme için metin madenciliği (Badal vd, 2015)
(Murugesan vd, 2017)	Adlandırılmış varlık tanıma	MIRA (margin-infused relaxed algorithm) algoritmasını kullanarak model entegrasyonu sağlanmış, öğrenme ve etiketleme için iki yönlü CRF (conditional random fields) kullanılmıştır.	Tıp	BCC-NER: Gen/Protein özelliğini tanımlamak amacıyla varlık etiketleyicisi olarak adlandırılan iki yönlü bağlamsal ipuçları

Çizelge 2.2. (devam)

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Chen vd, 2017)	Sınıflandırma	Ağırlıklandırma yöntemi TFxIDF, sınıflandırma algoritmaları SVM, Naive Bayes ve karar ağaçları (J48) algoritması	Sosyal medya	Youtube videolarının duygu sınıflandırılması
(He vd, 2017)	Duygu analizi	Ağırlıklandırma yöntemi olarak terim frekansı (TF), sınıflandırma işlemi olarak Google Prediction API	Müşteri ilişkileri yönetimi	Online otel görüşlerinin analiz edilmesinde sosyal medya analitiği uygulaması
(Christensen vd, 2017)	Sınıflandırma ve bilgi çıkarımı	Destek vektör makineleri ve kısmi en küçük kareler yöntemi	Gıda kalitesi	Online topluluk verilerinin incelenmesi: Online topluluklar içerisindeki fikirlerin doğası
(Mukherjee ve Bala, 2017)	Sınıflandırma	Terim ağırlıklandırma olarak terim frekansı, sınıflandırma için Naive Bayes ve Maksimum entropi yöntemleri, test için K-katlı çapraz doğrulama	Sosyal medya	Kullanıcıların stiline dayalı olarak mikroblog metinlerin cinsiyet sınıflandırması

Çizelge 2.2. (devam)

Yazar (lar)/ Yıl	Problem	Kullanılan yöntemler	Alan	Çalışma başlığı
(Nokhbeh Zaeem vd, 2017)	Adlandırılmış varlık tanıma ve bilgi çıkarımı	Gizli Markov modelleri, Maksimum Entropi Markov modelleri ve Koşullu Rastgele Alanlar (CRF)	Güvenlik	Kimlik hırsızlığı hikayelerinin metin madenciliği ile analizi sonucunda kimlik tehdidi davranışlarının modellenmesi ve analizi

Metin madenciliği ilgili yapılan tezleri incelemek amacıyla YÖK Tez Merkezinde yapılan çalışmalar taranmıştır. Toplamda 23 tane yüksek lisans tezi incelenmiştir. İncelenen tezler ile ilgili bilgiler Çizelge 2.3’te verilmiştir.

Çizelge 2.3. Çalışma için incelenen yüksek lisan tezleri

Yazar/Yıl	Kullanılan yöntemler	Anabilim dalı	Tez başlığı
(Çeliksi, 2017)	Uzaklık ölçüsü için Öklit Uzaklığı, k-means kümeleme algoritması	İstatistik	Yabancı dizilerin altyazı ve Twitter yorumlarının metin madenciliği ile incelenmesi
(Türkmenoğlu, 2015)	Karar destek makineleri (KDM), karar ağaçları, Naive Bayes	Bilgisayar Mühendisliği	Türkçe metinlerde duygu analizi
(Amanet, 2017)	Yalın Bayes, karar ağaçları, K- en yakın komşu ve destek vektör makineleri	İstatistik ve bilgisayar bilimleri	Türkçe sosyal medya metinlerinde duygu analizi

Çizelge 2.3. (devam)

Yazar/Yıl	Kullanılan yöntemler	Anabilim dalı	Tez başlığı
(Yurt, 2015)	Naive Bayes	Bilgisayar mühendisliği	Türkçe metinlerde duygu analizi
(Savaşan, 2011)	Kelime bulutu	Bilgisayar mühendisliği	Türkçe içeriklerden otomatik etiket bulutu oluşturma
(Gündüz, 2013)	Bilgi Kazanımı, Ki-kare istatistiği, karşılık bilgi özellik seçim yöntemleri, kelime bulutu	Bilgisayar mühendisliği	Borsa İstanbul (B1st) 100 Endeksi yönünün ekonomi haberleri ile tahmin edilmesi
(Sönmez, 2017)	Konu Modelleme yöntemlerinden biri olan Gizli Dirichlet Tahsisi	Endüstri Mühendisliği	Çevrimiçi yorumların metin madenciliği ile analizi: İstanbul'daki alışveriş merkezleri üzerine bir çalışma
(Canhası, 2007)	Destek vektör makineleri ve SOM algoritması	Bilgisayar mühendisliği	Edebiyat eserlerinin web verilerine dayanarak sınıflandırılması
(Talebi, 2013)	Naive Bayes, Destek vektör makineleri ve K-en yakın komşu	Bilgisayar mühendisliği	Facebook'da yorum madenciliği kullanarak kişilerin cinsiyet, yaş ve eğitim düzeylerinin tanımlanması

Çizelge 2.3. (devam)

Yazar/Yıl	Kullanılan yöntemler	Anabilim dalı	Tez başlığı
(Seçkin, 2011)	Naive Bayes, destek vektör makinesi, K-en yakın komşuluk algoritması ve karar ağaçları algoritmaları, kümeleme için K-means yöntemi	İşletme	Metin madenciliğinde kullanılan yöntemlerin karşılaştırılması: siyasi parti liderlerinin grup genel toplantı konuşmaları ile bir uygulama
(Çelikyay, 2010)	Apriori algoritması	Matematik bilgisayar	Metin madenciliği yöntemiyle Türkçe’de en sık kullanılan ve birbirini takip eden harflerin analizi ve birliktelik kuralları
(Pilavcılar, 2007)	Naive Bayes ve K-en yakın komşu	Matematik Mühendisliği	Metin madenciliği ile metin sınıflandırma
(Gönültaş, 2010)	Kümeleme için K-means, sosyal ağ analizinde CONCUR algoritması	Bilgisayar mühendisliği	E-posta listelerinde metin kümeleme ve sosyal ağ analizi uyumu
(Kuzucu, 2015)	Naive Bayes	Bilgisayar mühendisliği	Müşteri memnuniyeti belirlemek için metin madenciliği tabanlı bir yazılım aracı
(Varol, 2011)	Naive Bayes, Multilayer Perceptron, SMO, Rotation Forest, Decision Table, Part, Random Forest, Rep Tree, Ng-İnd	Elektronik bilgisayar eğitimi	Metin madenciliği yöntemlerini kullanarak Türkçe dokümanlarda tür ve yazar tanıma

Çizelge 2.3. (devam)

Yazar/Yıl	Kullanılan yöntemler	Anabilim dalı	Tez başlığı
(Sancar, 2016)	Multinomial Naive Bayes	Bilgisayar mühendisliği	Metin madenciliği kullanılarak talep tanıma ve yönlendirme sistemi
(Döven, 2013)	Kosinüs (cosine) ve jaccard benzerlik algoritmaları	Bilgi teknolojileri	Metin madenciliği ile dokümanlar arasındaki benzerliklerin bulunması
(Kaşıkçı, 2014)	Naive Bayes ve K-en yakın komşu sınıflandırma algoritmaları	Yönetim bilişim sistemler	Metin madenciliği ile e-ticaret sitelerinin belirlenmesi
(Karaca, 2012)	Naive Bayes Multinomial ve Multivariate, K-en yakın komşu	Bilgisayar mühendisliği	Metin madenciliği yöntemi ile haber sitelerindeki köşe yazılarının sınıflandırılması
(Taha, 2011)	K-Means ve K-Median algoritmaları	Bilgisayar bilimleri	Metin madenciliği ile doküman demetleme
(Moghaddam, 2014)	Destek vektör makineleri, Naive Bayes, karar ağaçları ve doğrusal ayırıcı analizi	Bilgisayar mühendisliği	Etkin sınıflandırma için genetik algoritma tabanlı öznelik alt küme seçimi
(Kaya, 2010)	Destek vektör makinesi ve K-en yakın komşu	Bilgisayar Mühendisliği	Finansal haber makaleleri kullanılarak hisse senetleri fiyat değişimlerinin tahmin edilmesi
(Nanğır, 2013)	Naive Bayes, destek vektör makineleri, çoklu sınıflandırıcı algoritması	Bilgisayar Mühendisliği	Türk dili için çoklu sınıflandırıcı yöntemler ile duygu sınıflandırma

Genel olarak literatür çalıřmaları incelendiđinde yurtdıřı çalıřmalarında metin madenciliđi tıp ve sađlık çalıřmalarında sıklıkla kullanılmaktadır. Türkiye’de ise sađlık alanında, metin madenciliđi ile ilgili yapılan çalıřma sayısı yok denecek kadar azdır. Ayrıca Ankilozan Spondilit Hastalıđı ile ilgili birçok çalıřma yapılmıřtır. Ancak hiçbir çalıřmada metin madenciliđi yöntemi uygulanmamıřtır. Özellikle ÷lkemizde böyle bir tez çalıřması yoktur. AS hastalarının sosyal medya üzerinde paylařmıř oldukları gönderilerin metin madenciliđi yöntemleri ile analiz edilmesiyle ilgili yapılan bu çalıřma ÷lkemizde bir ilk olacaktır. Ayrıca metin madenciliđinin Türkiye’de sađlık alanında uygulanmasına örnek bir çalıřma olacaktır.

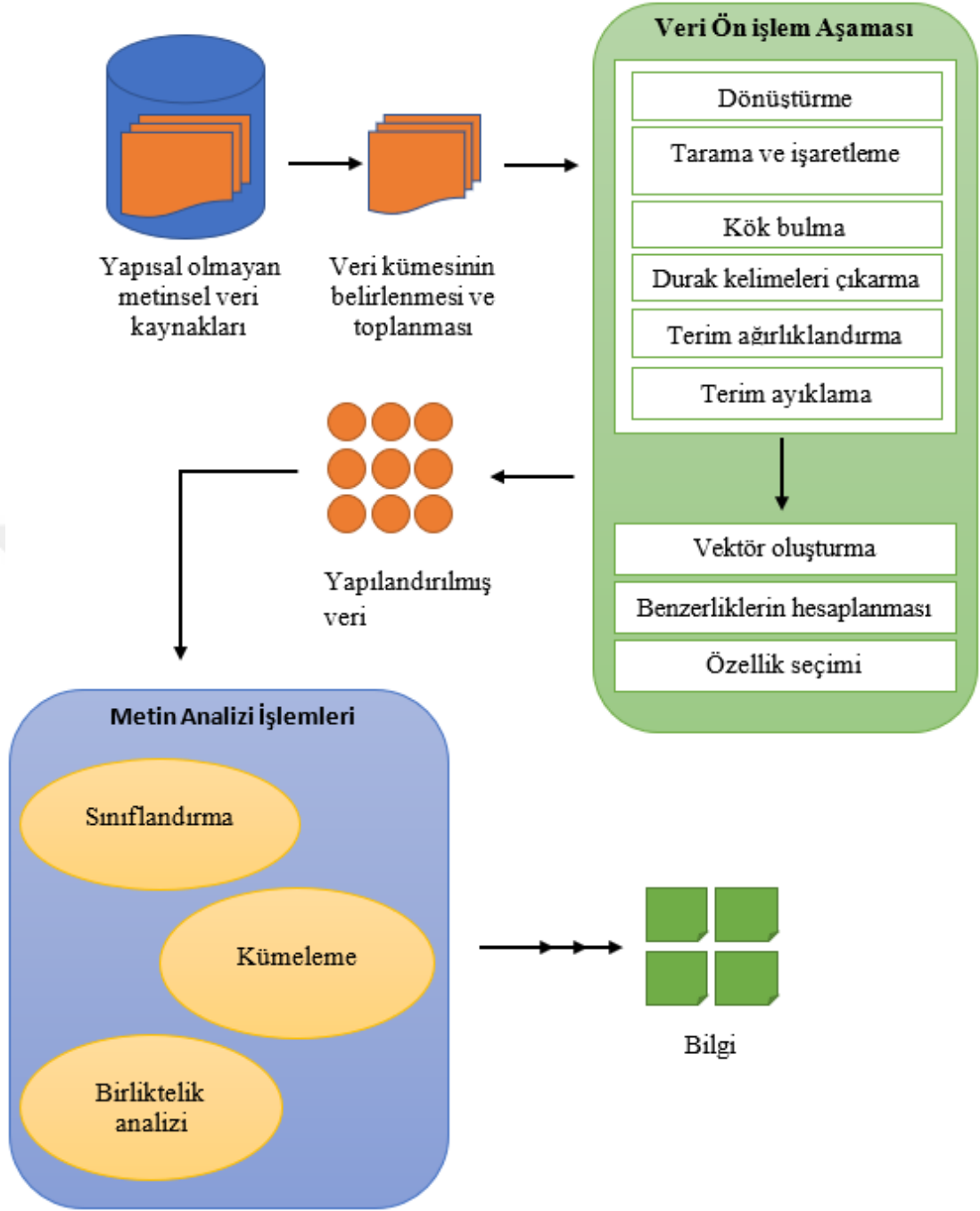


3. METİN MADENCİLİĞİ SÜRECİ

Metin madenciliği süreci temelde üç aşamadan oluşur. Birinci aşama yapılacak olan çalışmanın amacına göre verilerin toplanması ve metin ön işlemlerin uygulanması aşamasıdır. Bu aşamada ilk önce toplanan metinsel verilere dönüştürme, tarama ve işaretleme işlemleri uygulanır. Daha sonra doğal dil işleme araçları ile kelime kökleri bulunur ve çeşitli ağırlıklandırmalar yapılır. Gereksiz kelimeler çıkarılır ve terim ayıklama işlemi yapılır.

İkinci aşamada terim doküman matrisi oluşturulur. Doküman matrisi bulunan kelime kökleri veya terimler ile dokümanlar arasındaki ilişkiyi sayısal olarak gösterir. Terim doküman matrisinde sütunlar bulunan özellikleri, satırlar ise dokümanları gösterir. Matris içerisindeki sayısal değerler ise bir özelliğin o belge içerisindeki ağırlık değerini ifade eder. Matris boyutunu azaltmak için de çeşitli özellik seçimi yöntemleri kullanılır.

Üçüncü aşamada ise yapısal hale dönüştürülen verilerden bilgi çıkarımında bulunmak için sınıflandırma, kümeleme, birliktelik analizi ve çeşitli bilgi çıkarım yöntemleri kullanılır. Bu aşama sonunda ise yapılan çalışmaların amacına yönelik bilgiler elde edilmiş olur. Metin madenciliği süreci aşamaları Şekil 3.1’de ayrıntılı olarak gösterilmiştir (Seçkin, 2011).



Şekil 3.1. Metin madenciliği süreci

3.1. Veri Kümesinin Belirlenmesi ve Toplanması

Veri kümesinin belirlenmesi aşaması, araştırılan konuya yönelik bilgi erişim sistemlerini kullanarak metin veri setinin oluşturulmasıdır (Oğuz vd, 2007). Toplanacak veriler ele alınan probleme uygun bir biçimde toplanmalıdır. Metin madenciliği hangi alana yönelik uygulanacaksa o alanla ilgili gerekli dokümanların

toplanması gerekir (Bayer, 2011). Veri toplama süreci, veriyi anlama süreci olarak ifade edilir (Göker ve Tekedere, 2017).

Veri toplama süreci;

- 1) Başlangıç verilerinin toplanması
- 2) Toplanan verinin tanımlanması ve ihtiyaçları karşılayıp karşılayamayacağını değerlendirilmesi
- 3) Çalışmanın gerçekleştirilebilmesi için veri anlamında eksiklerin tespit edilmesi
- 4) Çalışmada kullanılacak olan veriler tam mı, eksik mi, hata var mı şeklinde, tanımlanabilir (Göker ve Tekedere, 2017).

Genellikle metin madenciliğinde kullanılan veri kaynakları, işletme dokümanları, müşteri yorumları, web sayfaları, sosyal medya paylaşımları, metin ve xml dosyaları şeklindedir (Ergün vd, 2013). Toplanan verinin fazla olması sınıflandırma başarısını ve kaliteli bilgi çıkarımını artırır (Bayer, 2011). Ayrıca oluşturulan sınıflar arasındaki veri sayılarının birbirinden farklı olması, dokümanların kısalığı ve bir doküman içinde çok farklı konulardan bahsedilmesi sınıflandırma başarısını düşürmektedir (Karaca vd, 2015).

3.2. Veri Ön İşlem Aşaması

Yapısal olmayan nitelikteki metin dokümanları doğal dil kullanılarak oluşturulur. Metin dokümanları üzerinde veri madenciliği tekniklerinin uygulanabilmesi için yapısal olmayan veriyi yapısal formata dönüştüren veri ön işlem adımına ihtiyaç duyulur (Tunalı ve Bilgin, 2012). Veri ön işlem aşamaları, veriyi uygun formata getirmek ve veri seti üzerinde çeşitli analizler yapabilmek için yapılan işlemler dizisidir (Göker ve Tekedere, 2017). Ön işlem; veri üzerinde bulunan problemleri çözmek, verinin doğal yapısını öğrenerek daha anlamlı ve kaliteli analiz yapabilmek ve veriden daha anlamlı bilgi üretebilmek gibi nedenlerden dolayı yapılır (Kaşıkçı ve Gökçen, 2014). Veri ön işlem aşamasında metni kelimelere ayırma, kelimelerin anlamsal değerlerini bulma (isim, sıfat, zamir, fiil, zarf), kelimeleri köklerine ayırma, kelime köklerini belirleme, dokümanları fazlalıklardan arındırma, gereksiz bilgileri çıkarma, yazım kurallarına uygunluğu tespit etme, ekleri ve genel kelimeleri çıkarma, önek ve sonekleri ayırma, kelime anlamlarını belirleme, kelimelerin doküman içerisinde sayısal olarak ifade edilmesi gibi kelimelerle ilgili işlemler yapılır (Oğuz

vd, 2007). Böylelikle hatalı sonuçlara neden olacak faktörler ön işlem aşaması ile ortadan kaldırılmış olur. Her veri seti için ön işlem aşaması aynı formatta değildir, verinin formatına göre farklılık gösterebilir (Karaca vd, 2015). Örneğin web ortamından alınan verilerden HTML etiketi temizlenirken, Twitter üzerinden alınan bir veriden “@” karakteriyle başlayan kullanıcı adları veya “#” karakteriyle başlayan hashtagler temizlenebilir.

Yapılan araştırmalar sonucunda ön işlem uygulanarak yapılan sınıflandırıcı modellerin hepsi, ön işlemsiz olarak yapılan modellerden ortalama %2,26 ile %4,94 arasında daha iyi sonuç vermiştir (Açıkalin ve Bayazıt, 2016). Ayrıca işlem süresini de kısaltmıştır. Veri ön işlem teknikleri, dönüştürme, tarama ve işaretleme, kök bulma, durak ve gereksiz kelimeleri çıkarma, terim ağırlıklandırma, terim ayıklama, vektör oluşturma ve benzerliklerin hesaplanması şeklinde sıralanabilir (Aşlıyan ve Günel, 2010; Kaşıkçı ve Gökçen, 2014).

3.2.1. Dönüştürme

İnternette elde edilen dokümanlar genellikle HTML (Hyper Text Markup Language) ve XML (Extensible Markup Language) gibi çeşitli etiketlerde tutulduğundan dolayı bunları düz metin hale dönüştürmek için dokümanlar HTML ve XML etiketlerinden temizlenir (Kaşıkçı ve Gökçen, 2014). Sosyal ağlar üzerinden elde edilen verilerde veri içerisinde yer alan mentionlar (@), hashtagler (#) ve linkler silinir. Ayrıca metin içerisinde yer alan metin olmayan çift tırnak (“), nokta (.), ünlem (!), virgül (,), tek tırnak (‘) gibi noktalama işaretleri, (*), (+), (/) gibi çeşitli karakterler, his simgeleri ve rakamlar metin içerisinden çıkarılır (Başkaya ve Aydın, 2017; Çoban vd, 2015).

3.2.2. Tarama ve işaretleme

Bir doküman veya metin içerisindeki kelimelerin tamamının elde edilmesi için işaretleme işlemi yapılır (Canhası, 2007). Yapısal olmayan veriler ilk olarak cümlelere, daha sonra da kelimelere parçalanır (Gürbüz ve Özyirmidokuz, 2018). Ayrıca metin belgeleri bölüm, paragraf, cümle, kelime ve hecelere ayrılabilir. En çok kullanılan işlem kelimelere ayırma işlemidir (Haltaş vd, 2015).

3.2.3. Kök bulma

Kök bulma işlemi, parçalanmış kelimelerin doküman içerisindeki sıklıklarına bakılırken aynı kökten gelen kelimelerin farklı ek almış hallerinin aynı kelime olarak algılanması için yapılır (Kaşıkçı ve Gökçen, 2014). Bu sayede kelimeler köke indirgenerek aynı kelime kökleri ile temsil edilmesi sağlanmış olur. Ayrıca hem özellik sayısı azalmış olacak hem de aynı kelime köküne sahip kelimelerin frekansı daha doğru bir şekilde hesaplanmış olacaktır (Haltaş vd, 2015).

Kök işlemlerinin uygulanabilmesi için metinde kullanılan dilin özelliklerinin çok iyi bilinmesi gereklidir. Bu bağlamda, Türkçe bitişken bir dildir. Her bir kelime, bir kökten ve kökün sonuna eklenmiş eklerden oluşur. Kelimelerin aldığı ekler, hem kelimenin anlamını hem de türünü değiştiren yapım ekleri ve kelimenin anlamını değiştirmeyen ancak türünü değiştirebilen çekim ekleri olmak üzere ikiye ayrılır. Türkçede kelime türleri isim, sıfat, zamir, zarf, edat, bağlaç, ünlem ve eylem olmak üzere 8 kısma ayrılır (Ergün vd, 2013). Yapılan çalışmanın amacına göre metin içerisindeki bu türler belirlenmeye çalışılır. Örneğin; Çizelge 3.1’de ‘yap’ kelime kökünün farklı ek almış halleri verilmiştir.

Çizelge 3.1. Örnek kelime kökü

Kelime	Kök
Yapıyorum	Yap
Yapacaktım	Yap
Yaptım	Yap
Yaparım	Yap

Türkçe için kullanılan kök bulma yöntemleri 3 başlık altında toplanabilir. Birincisi sözlük tabanlı kök bulma yöntemidir (Açıkalın ve Bayazıt, 2016). Türkçe metinlerin kök ve eklerini bulmak için Zemberek kütüphanesi kullanılır. Zemberek, Türk dilleri ve Türkçe için geliştirilmiştir açık kaynaklı, bağımsız bir Doğal Dil İşleme kütüphanesidir. Zemberek kelimelerin köklerini bulmak için kök ve ek sözlüğü kullanan sözlük tabanlı bir kök bulma yöntemidir (Tunalı ve Bilgin, 2012).

İkinci kök bulma yöntemi, ek çıkaran kök bulucu (affix stripping stemmer) olarak adlandırılır ve sözlük tabanlı kök bulma yöntemlerinden farklı olarak

Türkçe'nin kural tabanlı yapısından yararlanır (Tunalı ve Bilgin, 2012). Bu kök bulma yöntemi sözcüklerin sonlarındaki ekleri sondan başa doğru çıkararak kökleri bulur (Açıkalın ve Bayazıt, 2016).

Kullanılan son kök bulma yöntemi ise, sabit önek (fixed prefix stemming) yöntemi olup temel olarak sözcüklerin ilk n karakterini sözcüğün kökü olarak kabul eder. Hızlı ve basit bir yöntemdir. n karakter ve daha kısa olan sözcükler olduğu gibi kabul edilir (Tunalı ve Bilgin, 2012).

Yapılan çalışmalarda, Türkçe metinlerin sınıflandırılmasında kök bulmanın etkisinin çok az olduğu ve kök bulmanın kümeleme kalitesinde her zaman belirgin bir iyileşme sağlamadığı tespit edilmiştir. Ancak doküman-terim matrisinin boyutunda sağladığı yüksek indirgeme oranı nedeniyle kümeleme çalışmalarında kullanılması tavsiye edilmiştir.

3.2.4. Durak kelimeleri çıkarma

Edat (gibi, için, ancak, vb.), bağlaç (ve, ile, fakat vb.) ve zamir (ben, sen, o vb.) gibi metin içerisinde çok sık geçen ve tek başına bir anlam ifade etmeyen gereksiz kelimeler (stop words) metinden çıkarılması işlemidir (Haltaş vd, 2015). Durak kelimeler, bir dilde sıklıkla kullanılan ve tek başına bir anlam ifade etmeyen kelime listesidir. Her dilin kendine özgü durak kelimeleri var olmakla birlikte, standart bir liste yoktur (Çoban vd, 2015). Genellikle, durak kelimeler baz alınarak durak kelime listesi oluşturulur ve bu durak kelime listesine göre gereksiz kelimeler çıkarılır (Yıldız, 2017). Etkisiz kelimelerin metinden atılması, sınıflandırmada kullanılacak olan özniteliklerin boyutunu azaltacağından öznitelik seçme yöntemi olarak da düşünülebilir (Gündüz, 2013). Ayrıca metin içerisindeki bir kelimenin diğer kelimelere kıyasla çok veya az defa kullanılması kelimeyi ayrıştırıcı yapmadığından durak kelimesi olarak görülebilir (Canhası, 2007).

3.2.5. Terim ağırlıklandırma

Terim ağırlıklandırma ile sözcüklerin kendileri değil onları temsil eden sayısal değerleri kullanılır. Ağırlıklandırma işlemine kelimelerin doküman üzerindeki etkisi olarak da ifade edilebilir. Ayrıca yapısal olmayan veriler ağırlıklandırma ile tam yapısal hale dönüşmüş olur (Karaca vd, 2015). Ağırlıklandırma işlemi yapılırken

çeşitli ağırlıklandırma yöntemlerinden faydalanılır. Bunlar; binary ağırlıklandırma, frekansa göre ağırlıklandırma (TF), terim frekansı x ters doküman frekansı (TFxIDF), normalizasyon (TFC), logaritmik terim bileşeni (LTC), CDT Metodu ve TFRF Metodu gibi yöntemlerdir (Amasyalı vd, 2012; Canhası, 2007; Seçkin, 2011).

3.2.5.1. İkili (Binary) ağırlıklandırma metodu

Bir kelimenin doküman içerisinde geçmesi ya da geçmemesini temel alan bir ağırlıklandırma yöntemidir (Canhası, 2007). Doküman içerisinde geçmesi durumunda 1, geçmemesi durumunda ise 0 değerini alır. İkili ağırlıklandırma formülü,

$$w_{ij} = \begin{cases} 1, & \text{eğer } i \text{ terimi } j \text{ dokümanında geçiyor ise} \\ 0, & \text{diğer durumlarda} \end{cases} \quad (3.1)$$

şeklinde verilebilir. Örneğin, aşağıda D₁, D₂, D₃ ve D₄ şeklinde ifade edilen dokümanları temsil eden terimlerin ikili ağırlıklandırma yöntemine göre ağırlıklık değerleri Çizelge 3.2’de verilmiştir.

D₁: İki yıldır Ankilozan Spondilit hastasıyım. Her sabah uyandıgımda belimde tutulmalar oluyor. Şu an belim çok kötü. Sizlerde aynı durumla karşılaşıyor musun?

D₂: Hacamat ve sülük tedavisi uygulayıp faydasını gören var mı? Alternatif tedavi yöntemlerin iyileşme sürecine katkı sağlıyor mu?

D₃: Düzenli spor ve egzersizler yapmak ağrılarınızı azaltacaktır. Özellikle Ankilozan Spondilit hastalarına yüzmeyi tavsiye ederim. Yüzme eklem ağrılarınızı etki edecektir.

D₄: Samsun ilinde tavsiye edeceğiniz iyi bir romatoloji doktoru var mı? Tavsiyelerinizi bekliyorum.

Çizelge 3.2. İkili ağırlıklandırma yöntemine ait örnek sonuçlar

Dokümanlar	Terimler						
	bel	tutulma	hacamat	tedavi	egzersiz	doktor	yüzme
D ₁	1	1	0	0	0	0	0
D ₂	0	0	1	1	0	0	0
D ₃	0	0	0	0	1	0	1
D ₄	0	0	0	0	0	1	0

3.2.5.2. Frekansa göre ağırlıklandırma

Kelimelerin metin içerisinde kaç defa geçtiğini gösteren basit bir yöntemdir (Seçkin, 2011). Terim frekansı $T_j = (w_{1j}, w_{2j}, \dots, w_{ij})$ şeklinde ifade edilir. Burada T vektörü, j doküman sayısını, i terimlerin toplam sayısını ve w kelime frekansını ifade etmektedir (Haltaş vd, 2015). Örneğin, D_1, D_2, D_3 ve D_4 dokümanlarını temsil eden terimlerin frekans ağırlıklandırma yöntemine göre ağırlık değerleri Çizelge 3.3'te verilmiştir.

Çizelge 3.3. Frekansa göre ağırlıklandırma yöntemine ait örnek sonuçlar

Dokümanlar	Terimler						
	bel	tutulma	hacamat	tedavi	egzersiz	doktor	yüzme
D_1	2	1	0	0	0	0	0
D_2	0	0	1	2	0	0	0
D_3	0	0	0	0	1	0	2
D_4	0	0	0	0	0	1	0

3.2.5.3. Klasik TFXIDF ağırlıklandırma

$TFxIDF$ ağırlıklandırma, eğer bir kelime az sayıda dokümanda kullanılıyorsa, kelimenin o dokümanın sınıfının belirlenmesinde önemli olduğu, eğer bir kelime çok sayıda dokümanda kullanıyor ise kelimenin ayırt edici özelliğinin az olduğu düşüncesine dayanarak oluşturulmuştur (Seçkin, 2011). TF , terim frekansını yani terimlerin bulunduğu metin dokümanı içerisinde kaç kez kullanıldığını, IDF ise aynı terimin tüm dokümanlar arasında kaç kez kullanıldığını gösteren ters doküman frekansıdır (Cingiz ve Diri, 2012). Bu iki değerle çarpımıyla $TFxIDF$ ağırlıklandırma değerleri elde edilmiş olur. j . doküman, i . terim için oluşturulan $TFxIDF$ formülü,

$$TF_{ij} = \frac{F_{ij}}{\sum_j F_i} \quad (3.2)$$

$$IDF_i = \log\left(\frac{D}{df_i}\right) \quad (3.3)$$

$$Ağırlıklandırma = TF_{ij} \times IDF_i \quad (3.4)$$

şeklinde verilmiştir.

Burada F_{ij} ; i teriminin j dokümanında kaç kez geçtiğini, F_i ; i teriminin toplam frekans sayısını, df_i ; i terimini içeren doküman sayısını, D ; doküman sayısını

göstermektedir (Yıldız, 2017). Örneğin, D₁, D₂, D₃ ve D₄ dokümanlarını temsil eden terimlerin TFxIDF ağırlıklandırma yöntemine göre ağırlık değerleri Çizelge 3.4'te verilmiştir.

Çizelge 3.4. TFxIDF ağırlıklandırma yöntemine ait örnek sonuçlar

Dokümanlar	Terimler						
	bel	tutulma	hacamat	tedavi	egzersiz	doktor	yüzme
D ₁	1,204	0,602	0	0	0	0	0
D ₂	0	0	0,602	1,204	0	0	0
D ₃	0	0	0	0	0,602	0	1,204
D ₄	0	0	0	0	0	0,602	0

3.2.5.4. CDT metodu (Categorical term descriptor)

Bu ağırlıklandırma metodu Bong vd (2005) tarafından sunulmuştur. *TFxIDF* ağırlıklandırmasının bir dokümana değil de kategoriye uygulanması ile ortaya çıkmıştır (Canhası, 2007). *CDT* metodu formülü,

$$CDT_{tc} = CTF_{tc} \times IDF_{tc} \times ICF_t \quad (3.5)$$

$$ICF = \log \frac{|C|}{CF_t} \quad (3.6)$$

şeklinde ifade edilir.

CTF, kategori temelli terim frekansıdır. *IDF*, ters doküman frekansı ve *ICF*, ters kategori frekansıdır.

3.2.5.5. TFRF metodu (Relevance frequency)

Man vd (2005) tarafından terim frekansı ve ilgi frekansının birleştirilmesiyle oluşturulmuş bir ağırlıklandırma yöntemidir. Bu yeni ağırlıklandırma yöntemine TFRF adı verilmiştir. TFRF metodu,

$$w_{ij} = tf_{ij} \times rf \quad (3.7)$$

formülü ile hesaplanır. tf_{ij} , bilindiği üzerine i teriminin j dokümanındaki frekansıdır. rf kısmı ise

$$rf = \log(1 + \frac{n_i}{n_i^-}) \quad (3.8)$$

formülü ile hesaplanır. Burada n_i ; i teriminin bulunduğu doküman sayısını, n_i^- ; ise i terimini içeren ancak kümede bulunmayan doküman sayısını ifade etmektedir (Canhası, 2007).

3.2.5.6. Normalizasyon (Term frequency component) (TFC)

Bir metnin uzun olması, daha kısa olan metinlerden çok daha önemli olduğu ya da konu ile alakalı olduğu anlamını taşımaz. Bu nedenden dolayı uzun ve kısa metinlerin ağırlıklandırılması yapılırken kelime frekans değerleri ve metinde geçen toplam kelime sayısı dikkate alınarak normalizasyonu sağlanır. Bu yöntemin TFxIDF ağırlıklandırmadan farkı metinde geçen toplam kelime sayısını da dikkate alıyor olmasıdır (Seçkin, 2011).

$$w_{ij} = \frac{f_{ij} \times \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{k=1}^M \left[\log(f_{ij}+1) \times \log\left(\frac{N}{n_k}\right) \right]^2}} \quad (3.9)$$

f_{ij} ; i terimin j . dokümandaki frekans değeri, n_i ; i terimin bulunduğu doküman sayısı, N ; toplam doküman sayısı, M ; metin içerisinde geçen toplam kelime sayısı, n_k ; metin içerisinde yer alan k . terimin bulunduğu doküman sayısını ifade eder.

3.2.5.7. Logaritmik terim bileşeni (LTC)

Normalizasyon yönteminin değiştirilmiş halidir. Frekanslardaki büyük değişikliklerin etkisini azaltmak amacıyla ham frekans yerine logaritma kullanılır (Seçkin, 2011).

$$w_{ij} = \frac{\log(f_{ij}+1) \times \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{k=1}^M \left[\log(f_{ij}+1) \times \log\left(\frac{N}{n_k}\right) \right]^2}} \quad (3.10)$$

3.2.6. Terim ayıklama

Terim ayıklama işleminde, frekansı eşik değerin altında olan kelimelerin kaldırılmasıdır. Düşük frekanslı kelimelerin, metin içeriğinde daha az öneme sahip olduğu varsayılarak bu kelimeler çıkarılır. Basit boyut azaltma yöntemlerinden biridir (Haltaş vd, 2015). Dokümanlarda çok sık kullanılan kelimelerin ayırım güçleri çok az olduğu için maksimum geçiş sayısına göre de filtre uygulanabilir (Amasyalı vd, 2012). Kelime içerisinde tekrar eden harflerin bire indirgenmesi amacıyla metin normalizasyonu yapılabilir (Çoban vd, 2015). Örneğin; kelime “selammmmm” ise bu kelime “selam” şekline dönüştürülebilir. Minimum kelime uzunluğu belirlenerek de terim ayıklama yapılabilir (Kılınç vd, 2015). Örneğin; kelime uzunluğu minimum 3 olarak belirlenmişse metin içerisinde bir ve iki harf uzunluğuna sahip kelimeler elenecektir.

3.3. Vektör Uzay Modeli ve Vektör Oluşturma

Gerard Salton vd (1975) tarafından bilim dünyasına kazandırılan vektör uzay modelinde her bir kelime veri uzayında bir nokta olarak ifade edilmektedir (Bayer, 2011). Verinin sahip olduğu farklı özellikler noktanın koordinatları olarak gösterilmektedir (Çalış vd, 2012). Dokümanların sınıflandırılmasında sıklıkla kullanılan özellik vektör uzayı kelimelerin dokümanlardaki geçme sıklıklarına dayanmaktadır. Kısaca her bir doküman içindeki kelimelerin dokümanlardaki frekanslarının hesaplanması ile oluşturulur (Yıldız, 2017). Vektör uzay modelinde terim ağırlıklandırma yöntemleri kullanılır. Sıklıkla binary, terim frekansı ve TFxIDF ağırlıklandırma yöntemleri kullanılır (Karaca ve Görgünoğlu, 2012). Ağırlıklar hesaplandıktan sonra Doküman Terim Matrisi (Document Term Matrix-DTM) oluşturulur (Kılınç vd, 2016). Şekil 3.2’de terim matrisi gösterilmiştir. Matris, n adet kelime ve d adet doküman olmak üzere $n \times d$ boyutunda bir matris olarak ifade edilebilir. Matrisin boyutu arttıkça üzerinde hesaplama yapabilmek için daha fazla bilgi işlem gücü ve hafıza gerekir (Güven vd, 2006). Ayrıca doğru bilgiye ulaşma ihtimali de azalacaktır (Bayer, 2011). Özellik çıkarımı yöntemleri ile boyut indirgemesi yapılabilir (Seçkin, 2011). Vektör uzay modeli iki büyük kısıta sahiptir. Birincisi vektör uzayındaki boyut sayısının büyük olması doküman vektörleri arasındaki yakınlığın hesaplanmasını zorlaştırmaktadır. İkincisi olarak eşanlamlı-

çok anlamlı kelimelerin, dokümanlar arasında olan ilişkiyi yok, olmayan ilişkiyi var gibi göstermesidir (Güven vd, 2007).

$$\begin{matrix}
 & t_1 & t_2 & t_3 & \dots & t_n \\
 D_1 & (w_{11} & w_{12} & w_{13} & \dots & w_{1n}) \\
 D_2 & (w_{21} & w_{22} & w_{23} & \dots & w_{2n}) \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 D_d & (w_{d1} & w_{d2} & w_{d3} & \dots & w_{dn})
 \end{matrix}$$

Şekil 3.2. Doküman terim matrisi

Doküman vektörleri hem eğitim hem de test verisi için kullanılır ve benzerlik hesaplama işlemleri bu vektörler üzerinden gerçekleştirilir (Karaca vd, 2015).

3.4. Benzerliklerin Hesaplanması

Test dokümanının hangi sınıfa ait olduğu, eğitim dokümanı vektörü ile test dokümanı vektörü arasındaki ilişkiye ve ilişkinin seviyesine bağlıdır. Bu aşamada eğitim doküman vektörleri ile test vektörleri arasındaki benzerlik ve uzaklık ölçülür (Karaca vd, 2015). Buradaki benzerlik anlamsal benzerlik değil karakter seviyesindeki benzerliktir. Yakın kelime komşulukları ve kelime tekrarlarını bulmak amaçlanır. İntihal tespitinde, web üzerinde yer alan ve aynı olan web sayfalarının dizinler üzerinde farklı sayfalarmış gibi yer almasını engellemek amacıyla, reklam e-postalarını etili kılmak için kullanıcı profilleri arasında seçim yapılırken bunlara benzer birçok çalışmada dokümanlar arasındaki benzerliklerden yararlanır (Hark vd, 2017). Ayrıca metin sınıflandırma ve metin kümeleme gibi metin madenciliği yöntemlerinin temelini oluşturmaktadır (Keleş ve Özel, 2017).

Metin belgeleri arasındaki uzaklığı hesaplamak amacıyla metin eşleştirme algoritmalarından faydalanılır. Metin eşleştirme algoritmaları yaklaşık metin eşleştirme ve tam metin eşleştirme olmak üzere 2'ye ayrılabilir. Tam metin eşleştirme algoritmaları iki metin değerinin bire-bir eşleşip eşleşmediğini belirleyen karşılaştırma yöntemleridir. Tam metin eşleştirme algoritmalarına Boyer Moore, Karp-Rabin, Morris Pratt, Quick Search, Brute Force, Shift Or, Apostolico-Giancarlo, Turbo-BM ve Skip Search algoritmaları örnek olarak verilebilir (Keleş ve Özel, 2017).

Yaklaşık metin eşleştirme algoritmaları belirli bir metin içinde yaklaşık alt metin eşleştirmelerini bulma ve yaklaşık olarak örüntüyle eşleşen sözlük metinlerini bulma olarak iki işlem üzerinde yoğunlaşır. Yaklaşık metin eşleştirme algoritmaları, DNA dizilerini eşleştirmede, metinlerde yazım denetimi yapma, yanlış yazılmış sözcükler için olası önerileri bulmaya gibi birçok alan için yararlıdır. Literatürde yaklaşık metin eşleştirme yöntemi için birçok algoritma geliştirilmiştir. Bu algoritmaların bazıları; Jaro, Levenshtein, Öklit Uzaklığı, Jaccard Benzerliği, Q-Gram Uzaklığı, Kosinüs Benzerliği, Jaro Winkler, Blok Uzaklığı, Chapman Uzunluk Sapması (Chapman Length Deviation), Chapman Eşleme Soundex (Chapman Matching Soundex), Chapman Ortalama Uzunluk (Chapman Mean Length), Chapman Sıralı İsim Bileşik Benzerliği (Chapman Ordered Name Compound Similarity), Eşleme Katsayısı (Matching Coefficient), Smith Waterman, Smith Waterman Gotoh, Smith Waterman Gotoh Windowed Affine, Soundex gibi algoritmalar örnek verilebilir (Keleş ve Özel, 2017). Uzaklık/benzerlik ölçüsü olarak en çok kullanılan algoritmalar kosinüs benzerliği ve öklit uzaklığıdır.

3.5. Özellik Seçimi

Metin sınıflandırma işlemleri yapılırken bütün dokümanlarda geçen ve metni oluşturan tüm kelimelerin çalışmaya dahil edilmesi çalışma zamanı arttırır (Karaca vd, 2015). Ayrıca bütün kelimelerin dahil edilmesi sınıflandırma başarısını da etkileyecektir. Metinleri oluşturan kelimelerin tamamının alınması yerine metni temsil ettiği düşünülen kelimeler seçilmelidir. Düşük boyutlu özellik vektörleri ile daha başarılı sonuçlar elde edilebilir (Karaca vd, 2015). Özellik seçimi, çalışma zamanının düşürülmesi ve daha başarılı sınıflandırma performansı elde etmek amacıyla büyük boyutlu veri kümesinin daha küçük boyutlara indirgenerek ifade edilmesini sağlar (Haltaş vd, 2015). Özellik seçiminde ilgisiz ve gereksiz veriler silinir (Göker ve Tekedere, 2017). Bazı özellik çıkarım yöntemleri; N-gram metodu, korelasyon-tabanlı özellik çıkarımı, bilgi kazancı, ki-kare gibi yöntemler örnek verilebilir.

3.5.1. Bilgi kazancı

Shannon'un bilgi teoresini temel alan bilgi kazanımı, değişken değer ölçüsü olarak ifade edilir (Gündüz, 2013; Haltaş vd, 2015). Terim azaltmada sıklıkla kullanılan bir tekniktir (Yıldız, 2017). Bil kazancı, kelimenin metin içerisinde var olup olmasına

dayanarak sınıf tahmininde kullanılacak bilgi bit sayısını ölçerek hesaplanır (Canhası, 2007). Bilgi kazanım formülü 3.11'deki gibidir.

$$IG = \frac{a}{N} \log \frac{axN}{(a+c)(a+b)} + \frac{b}{N} \log \frac{bxN}{(b+d)(a+b)} + \frac{c}{N} \log \frac{cxN}{(a+c)(c+d)} + \frac{d}{N} \log \frac{dxN}{(b+d)(c+d)} \quad (3.11)$$

N ; toplam doküman sayısını,

a ; x terimini içeren kategori içerisindeki doküman sayısını,

b ; x terimini içermeyen kategori içerisindeki doküman sayısı,

c ; x terimini içeren kategori dışındaki doküman sayısı,

d ; x terimini içermeyen kategori dışındaki doküman sayısını göstermektedir (Yıldız, 2017).

3.5.2. N-Gram

N-gram metodu, bilgisayar tarafından bir metnin hangi dilde yazıldığını tespit etmek amacıyla kullanılır (Güven vd, 2006). Bu işlemi kelimeleri oluşturan harflerin veya metinleri oluşturan kelimelerin yan yana gelme örüntülerine bakarak yapar. N-gram modeli, önceki n elemanlı sıralamanın olma olasılığına dayanarak sıradaki olayın olma olasılığını tahmin etmeye çalışır ve $n-1$. sıradan daha önceki kelimeler ile bağımsızlık ilkesine dayanır (Amanet, 2017). Genellikle konuşma tanıma gibi problemlerde ve harf sıralamalarının tahmininde kullanılır (Türkmenoğlu, 2015). Ayrıca yazar tanıma, el yazısı tanıma, optik karakter tanıma, makine çevirisi gibi çalışmalarda kullanılmıştır (Seçkin, 2011). Metin içerisindeki kelimeler tek tek ele alındığında bir anlam ifade edemeyebilir. Ancak n-gramlar, bileşik kelimeler ve deyimler gibi yan yana geldiklerinden dolayı daha fazla anlamlı bilgiler içermektedir (Türkmenoğlu, 2015).

Örtüşen n-gram ve örtüşmeyen n-gram olmak üzere ikiye ayrılabilir. Genellikle unigram (1-gram), bigram (2-gram), trigram (3-gram) halleri kullanılır. Unigram kendisinden önceki 0 kelime sırasına, bigram kendisinden önceki 1 kelime sırasına, trigram kendisinden önceki 2 kelime sırasına bağlıdır (Amanet, 2017). Kısaca N-gramlar, kendisinden önceki $n-1$ kelime sırasına bağlıdır. Örneğin, 'ankilozan' kelimesinin n-gramlarına ait sonuçlar Çizelge 3.5'te verilmiştir.

Çizelge 3.5. 'Ankilozan' kelimesinin n-gramları

n-gramlar	Ankilozan
2-gram	an nk ki il lo oz za an
3-gram	ank nki kil ilo loz oza zan
4-gram	anki nkil kilo iloz loza ozan

Örtüşen ve örtüşmeyen n-gram gruplarına ait örnek metin ve sonuçlar Çizelge 3.6'da ve Çizelge 3.7'de verilmiştir.

Çizelge 3.6. Örnek bir cümlede örtüşen n-gram grupları

Metin	Alanı ile ilgili karşılaşılan sorunları araştırma yöntemlerini kullanarak çözümleyebilme
Unigramlar	'Alanı', 'ile', 'ilgili', 'karşılaşılan', 'sorunları', 'araştırma', 'yöntemlerini', 'kullanılarak', 'çözümleyebilme'
Bigramlar	'Alanı ile', 'ile ilgili', 'ilgili karşılaşılan', 'karşılaşılan sorunları', 'sorunları araştırma', 'araştırma yöntemlerini', 'yöntemlerini kullanarak', 'kullanarak çözümleyebilme'
Trigramlar	'Alanı ile ilgili', 'ile ilgili karşılaşılan', 'ilgili karşılaşılan sorunları', 'karşılaşılan sorunları araştırma', 'sorunları araştırma yöntemlerini', 'araştırma yöntemlerini kullanılarak', 'yöntemlerini kullanarak çözümleyebilme'
N-gramlar (n=4)	'Alanı ile ilgili karşılaşılan', 'ile ilgili karşılaşılan sorunları', 'ilgili karşılaşılan sorunları araştırma', 'karşılaşılan sorunları araştırma yöntemlerini', 'sorunları araştırma yöntemlerini kullanarak', 'araştırma yöntemlerini kullanarak çözümleyebilme'

Çizelge 3.7. Örnek bir cümlede örtüşmeyen n-gram grupları

Metin	Alanı ile ilgili karşılaşılan sorunları araştırma yöntemlerini kullanarak çözümleyebilme
Unigramlar	'Alanı', 'ile', 'ilgili', 'karşılaşılan', 'sorunları', 'araştırma', 'yöntemlerini', 'kullanılarak', 'çözümleyebilme'
Bigramlar	'Alanı ile', 'ilgili karşılaşılan', 'sorunları araştırma', 'yöntemleri kullanarak'
Trigramlar	'Alanı ile ilgili', 'karşılaşılan sorunları araştırma', 'yöntemlerini kullanarak çözümleyebilme'
N-gramlar (n=4)	'Alanı ile ilgili karşılaşılan', 'sorunları araştırma yöntemlerini kullanarak'

3.5.3. Ki-Kare

Ki-kare istatistiği hem özelliğin sınıfta yer alması hem de özelliğin sınıfta olmadığı durum olasılığını değerlendirir (Cingiz ve Diri, 2012). Uygulanabilirliği basit ve maliyet hesabı düşüktür (Yıldız, 2017). Özellik seçiminde sıklıkla kullanılmaktadır. Her sınıf için her bir özellik ile o sınıf arasındaki Ki-Kare istatistiği hesaplanmaktadır (Gündüz, 2013).

$$Ki - Kare = Nx \frac{(axd - bxc)^2}{(a+c)(b+d)(a+d)(c+d)} \quad (3.12)$$

Ki-Kare formülünde N ; toplam doküman sayısını, a ; x terimini içeren kategori içerisindeki doküman sayısını; b ; x terimini içermeyen kategori içerisindeki doküman sayısı, c ; x terimini içeren kategori dışındaki doküman sayısı, d ; x terimini içermeyen kategori dışındaki doküman sayısını göstermektedir (Yıldız, 2017).

4. METİN SINIFLANDIRMA

İnternetin kullanımının yaygınlaşmasıyla, online kütüphanelere, online metin ve makalelere, e-postalara vb. bilgilere erişilebilirlik artmıştır. Buna bağlı olarak bilgi yığınları oluşmuş ve bu bilgi yığınları içerisinde bilgiye ulaşmak zorlaşmıştır. İstenilen bilgiye kolayca ve hızlı bir şekilde ulaşabilmek için birbiriyle ilişkili olan bilgileri bulup aynı grup içinde toplamak gerekir. Bu bilgi karmaşıklığını önlemek için de, bu bilgi yığınlarını sınıflandırmak gerekir (Seçkin, 2011).

Sınıflandırma, sınıf değerleri önceden bilinen veri setinden elde edilen model ile sınıfı bilinmeyen yeni verinin sınıfının tahmin edilmesidir. Sınıflandırma işleminde etiket değeri önceden bilinen bir eğitim kümesine ihtiyaç vardır (Çalış vd, 2012). Sınıflandırmanın amacı yeni bir nesnenin özelliklerini belirleyerek ve bu nesnenin daha önceden tanımlanmış sınıf setlerinden birine atamasının yapılmasını sağlamaktır (Kaşıkçı ve Gökçen, 2014). Manuel olarak yapılan sınıflandırma işlemlerinde sınıflandırma yapan uzmanların vermiş olduğu kararlara göre sınıflandırma sonuçları değişebilmektedir (Karaca vd, 2015). Metin sınıflandırmanın uygulama alanları; metin bulup getirme ve kütüphane organizasyonu gibi alanlarda destek sağlayan ‘dokümanlara sınıf ataması yapmak’, mesajları, haberleri ve diğer metin dizisi halindeki bilgileri alıcılara ulaştırmak, doğal dil işleme sistemlerinin bir parçası olarak ‘ilgisiz metinleri ve metin parçalarını filtrelemek’, ‘metinleri, sınıf bazlı işleme mekanizmalarına yönlendirmek’ veya ‘sınırlı şekillerde bilgi edinimini sağlamak’ vb. amaçlar için kullanılır (Kaşıkçı ve Gökçen, 2014).

Metin sınıflandırma temel olarak beş aşamadan oluşur (Göker ve Tekedere, 2017);

- 1- Veri kümesinin toplanması
- 2- Veri ön işlem aşaması
- 3- Özellik seçimi
- 4- Vektör uzay modeli seçimi
- 5- Sınıflandırma

Kullanılan sınıflandırma algoritmasına ve benzerlik hesabına göre sınıflandırma işlemi yapılmaktadır. Sınıflandırma işlemi eğitim dokümanı ile test dokümanı

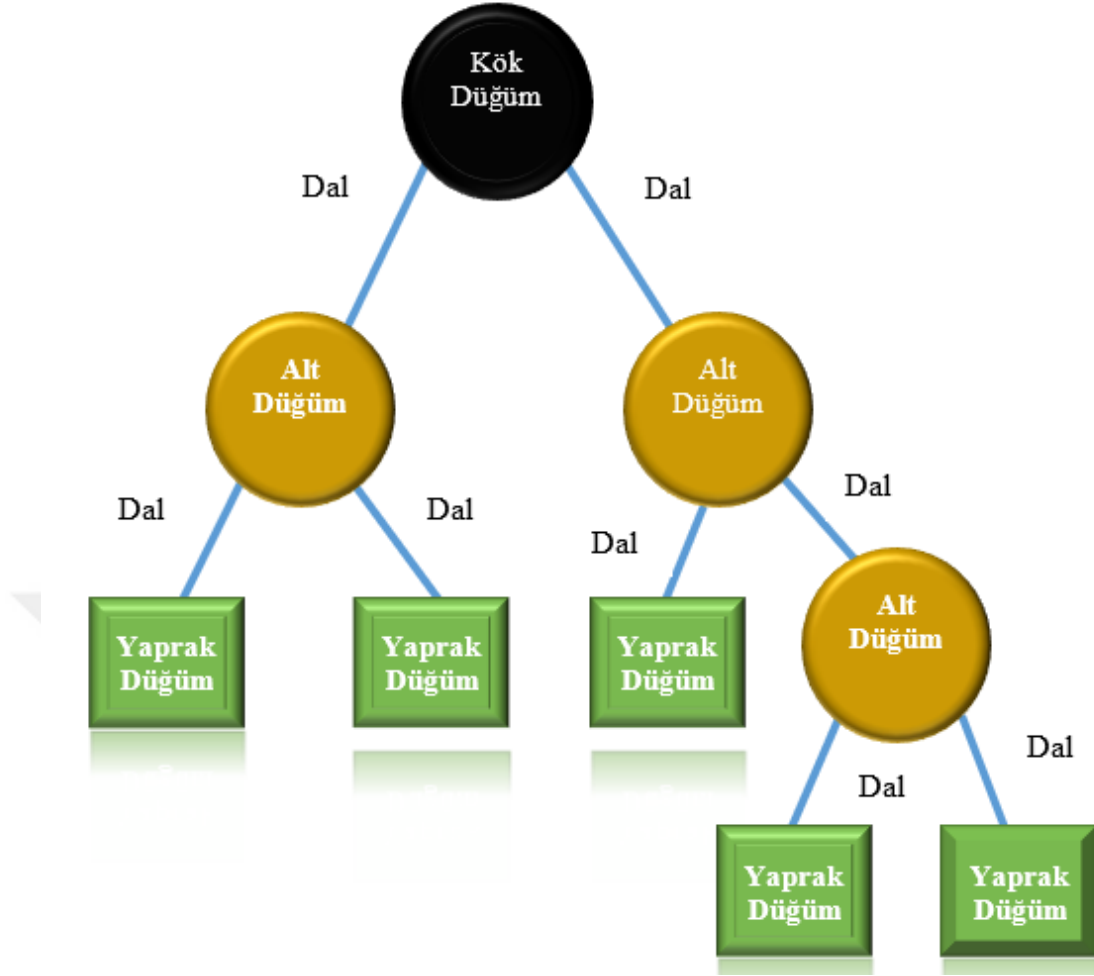
arasındaki yakınlığı esas alır. Yani test dokümanı hangi eğitim dokümanına yakınsa o eğitim dokümanın bulunduğu sınıfa ataması yapılır (Karaca vd, 2015). Sınıflandırma işlemlerinde kullanılan başlıca algoritmalara; karar ağaçları, destek vektör makineleri, K-en yakın komşu, Naive bayes, yapay sinir ağları örnek verilebilir.

4.1. Metin Sınıflandırma Algoritmaları

4.1.1. Karar ağaçları

Karar ağaçları sınıflandırma işlemlerinde en çok kullanılan yöntemlerden biridir. Köklerden dallara doğru gidilmesiyle kurallar belirlenerek ağaç oluşturulur. Karar ağaçlarında uygulanan birçok algoritma vardır. Kullanılan algoritmaya göre ağacın yapısı değişmektedir (Çeliksü, 2017). Değişen ağaç yapısı ise farklı sınıflandırma sonuçları verebilmektedir (Bayer, 2011).

Şekil 4.1’de görüldüğü gibi karar ağaçları kök düğüm, alt düğüm ve yaprak düğüm olmak üzere üç düğümden meydana gelmektedir. Kök düğüm alt düğümleri, alt düğümler ise yaprak düğümlerini oluşturur. Ağacın her düğümü verinin bir özelliğini ifade etmektedir (Amanet, 2017). Ağacın yaprakları sınıf etiketlerini ve yapraklara giden kollar ise özellikler üzerinde yapılan işlemleri göstermektedir. Son yapraklar hariç bütün düğümler belli bir özelliğe ait kuralları ifade etmektedir. Bu kurallara göre karar aşamasından sonra seçilecek dallar belirlenir. Son yapraklara sınıf etiketlerini taşımaktadır. Düğümler arasındaki bağlantılar, özelliklere göre izlenen yolu gösterdiği için karar ağaçlarında özellik seçimi önemlidir. Bu özellik seçimi kazanım oranıyla hesaplanabilmektedir (Türkmenoğlu, 2015).



Şekil 4.1. Karar ağaçları şeması

Bir sistemdeki belirsizliğin ölçüsü olan entropi kurallarına dayalı olarak karar ağaçlarının dalları oluşturulur. Entropi temelli olarak, sürekli ve kesikli değerler ile çalışabilen bir karar ağacı algoritması olan C4.5 algoritması (Onan ve Korukoğlu, 2015) da dallanma ölçütü olarak kazanç oranını kullanmaktadır. Entropi hesabı,

$$Entropi(C) = - \sum_{i=1}^n \log_2(p_i) \quad (4.1)$$

formülüyle yapılır.

C 'nin sınıf olduğu varsayılarak sınıflara ait entropi hesaplanır. Verilerin ham halinin entropi değerleri ile her bir alt kümenin entropilerinin toplamı arasındaki farka göre kazanç değeri hesaplanmış olur. Kazanım formülü,

$$Kazanım(B, C) = Entropi(C) - \sum_{i=1}^n \frac{|C_i|}{|C|} Entropi(C_i) \quad (4.2)$$

şeklinde verilir.

C kümesi için B özelliğın belirlenmesi aşamasında ayırma bilgisinden faydalanılır. Ayırma bilgisi formülü,

$$\text{Ayırma Bilgisi}(B) = - \sum_{i=1}^n \frac{|C_i|}{|C|} \log_2 \left\{ \frac{|C_i|}{|C|} \right\} \quad (4.3)$$

Kazanım oranı ile ağaç yapısını oluşturmak için hangisinin kök düğümü olacağı belirlenir. Kazanım oranı değeri büyük olan özellik kök düğüm olarak seçilir. Bu işlem yapraklar bulunana kadar bu şekilde devam etmektedir (Kılıç vd, 2015).

$$\text{Kazanım Oranı} = \frac{\text{Kazanım}(B,C)}{\text{Ayırma Bilgisi}(B)} \quad (4.4)$$

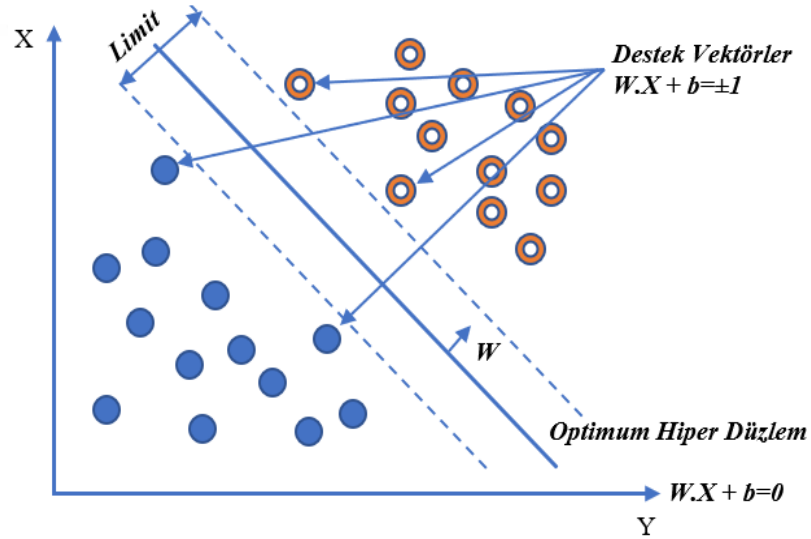
4.1.2. Destek vektör makinesi

İlk kez 1995 yılında Vapnik tarafından sınıflandırma ve regresyon tipi problem çözümleri için önerilen destek vektör makinesi (DVM-Support Vector Machine (SVM)), sınıfları birbirinden ayıracak optimal ayırma hiper düzleminin elde edilmesi amacıyla dayanır. Basit bir ifadeyle farklı sınıflara ait destek vektörleri arasındaki uzaklığı maksimize etmeye çalışır. Güçlü istatistiksel teoriler üzerine kurulmuş bir makine öğrenimi algoritması olan DVM, sınıflandırma konusunda oldukça basit ve etkili yöntemlerden birisidir (Kasaba ve Yıldıztepe, 2016). Genellikle, DVM ile sınıflandırmada $(-1, +1)$ şeklinde sınıf etiketleriyle iki gruba ait verileri sınıflandırmak için bir düzlemde iki sınıf arasında bir sınır çizilir (Amanet, 2017; Kasaba ve Yıldıztepe, 2016). Bu sınır, her iki sınıfın elemanlarının birbirine en uzak olduğu yer olmalıdır. Destek vektör makinesi bu sınırın nasıl çizileceğini belirlemektedir (Kasaba ve Yıldıztepe, 2016). DVM, hem doğrusal hem de doğrusal olmayan verilerin sınıflandırılmasında kullanılır (Onan ve Korukoğlu, 2015). Doğrusal verilerin sınıflandırılması durumunda sınıflar ayırıcı aşırı düzlem ile ayrılır. Bu durumda, DVM bu aşırı düzlemin her iki örnek sınıfa eşit uzaklıkta olmasını amaçlamaktadır (Kasaba ve Yıldıztepe, 2016). Doğrusal olmayan verilerin sınıflandırılmasında, doğrusal olmayan bir eşleme kullanarak, ilk olarak orijinal eğitim verilerini daha yüksek bir boyuta taşır. Sonra ise bu yeni yüksek boyut içerisinde doğrusal bir ayırıcı hiper düzlem bulmaya çalışır (Talebi, 2013). Bu hiper düzlem sayesinde doğrusal olmayan veriler sınıflandırılmış olur.

DVM sınıflandırılması başlıca 3 adımdan oluşmaktadır (Seçkin, 2011).

- i. Eğitim dokümanlarını oluşturan terimlerin özellik vektörü olarak ifade edilmesi (Seçkin, 2011). Tek vektör ile sadece iki sınıf belirlenebilir (Canhası, 2007).
- ii. Özellik vektörlerinin Kernel fonksiyonları kullanılarak özellik uzayına eşlenmesi (Seçkin, 2011)
- iii. Sınıfları birbirinden en iyi şekilde ayıran n-boyutlu hiper düzlemin oluşturulması

Eğitim için kullanılan D adet dokümandan oluşan verinin $\vartheta = \{x_i, y_i\}, i = 1, 2, \dots, D$ olduğunu varsayarsak x_i özellik vektörünü, $y_i \in \{-1, 1\}$ ise sınıf etiketlerini ifade eder. Dokümanları sınıflara ayırabilecek sonsuz sayıda düzlem oluşturulabilir ancak amaç bilinmeyen sınıflama hatasını en küçük yapacak hiper düzlemini oluşturmaktır. Şekil 4.2’de görüldüğü üzere $f(\vec{x}) = \vec{w}^T \vec{x} + b \geq 1$ durumu, birinci sınıf $y_i = 1$ ve $f(\vec{x}) = \vec{w}^T \vec{x} + b \leq -1$ durumuysa ikinci sınıfı $y_i = -1$ ifade eder (Arslan vd, 2017). Burada x n-boyutlu bir uzayı, w ağırlık vektörünü, b ise eğilim değerini gösterir (Amanet, 2017).



Şekil 4.2. Destek vektör ve hiper düzlem seçimi

$\lambda = 2/\bar{w}^2$ formülü iki sınır arasındaki uzaklığı ifade eder. Amaç λ değerinin maksimum yapmaktır. Buna bağlı sınırlamaysa $y_i(\vec{w}^T \vec{x} + b) - 1 \geq 0, y_i \in \{-1, 1\}$ şeklindedir (Arslan vd, 2017).

Destek vektör makinesi ile çoklu sınıf verilerinin sınıflandırılması amacıyla Kneer ve arkadaşları tarafından bire-karşı-bir yöntemi önerilmiştir. Bu yöntem ile n adet sınıf için n adet destek vektör makinesi oluşturulur. i . destek vektör makinesi, i sınıfındaki verileri birinci sınıf verisi olarak kullanılırken, diğer sınıflardan gelen verilerin hepsini ikinci sınıfın verisiymiş gibi kabul eder. Kendi verisine $+1$ etiketi verirken, diğer sınıflara ait tüm verilere -1 etiketi değerini verir. Eğitim bu şekilde n adet destek vektör makinesi için devam eder (Seçkin, 2011).

4.1.3. Ardışık minimal optimizasyon algoritması (SMO)

Ardışık minimal optimizasyon algoritması (SMO-Sequential Minimal Optimization), destek vektörlerini kullanan denetimli bir algoritmadır. DVM'lerinin eğitim sırasında ortaya çıkan ikinci dereceden problemlerin çözümünde kullanılır. SMO algoritmasında amaç her adımda minimum optimizasyon sonucuna ulaşmaktır. Ayrıca çok terimli problemlerin çözümünde daha az bellek kullanır ve daha hızlı bir şekilde çözümü gerçekleştirir.

Langrange çarpanlarının lineer bir eşitlik kısıtlamasına uyması gerektiğinden dolayı DVM'nin ikinci dereceden problemleri için ikili Langrange çarpanları kullanılmaktadır. SMO her bir aşamada, ikili Langrange çarpanlarını seçer, optimize etmek amacıyla bu çarpanlar için en uygun değerleri bulur. SVM'yi bulunan yeni uygun değerleri yansıtacak biçiminde günceller (Kılıç vd, 2015).

4.1.4. K-en yakın komşu (K-nearest neighbors)

Makine öğrenim algoritmalarının içerisinde en çok bilinen ve kullanılan algoritmalarından biri olan K-en yakın komşu (KNN) algoritması, eğitim verisini kullanarak öğrenmiş olduğu modeli test verisinde kullanarak sınıf etiketi ataması yapar (Çalış vd, 2012; Kılıç vd, 2016). KNN algoritması, sınıflandırılmak istenen örnekler ile sınıflandırılmış örnekler arasındaki benzerliğe dayalı olarak gerçekleşen bir yöntemdir (Onan ve Korukoğlu, 2015). Yeni gelen bir dokümanı sınıflandırmak için onun en yakınında bulunan k komşuya bakarak dokümanın sınıf etiketini belirler. Sınıflandırma için ilk olarak k değeri belirlenir. Veri setinin büyüklüğüne göre en sık kullanılan k değerleri 3, 5, 7'dir (Akyel ve Seçkin, 2012). k değerleri belirlendikten sonra eğitim kümesinde yer alan tüm dokümanlara göre yeni dokümanın uzaklıkları

hesaplanır. Aralarında minimum uzaklığa sahip dokümanların bulunduğu sınıf etiketlerine bakarak yeni dokümanın sınıf etiketi ataması yapılır (Çalış vd, 2012). Yeni gelen doküman ile diğer dokümanlar arasındaki uzaklıklar hesaplanırken genellikle Kosinüs benzerliği ve Oklid uzaklığı kullanılır (Akyel ve Seçkin, 2012). K-en yakın komşu algoritmasının avantajları; gürültülü eğitim dokümanlarına karşı dirençlidir, eğitim dokümanları fazla odluğunda etkilidir ve uygulanabilirliği basittir. Bir sorgu örneğinin tüm eğitim örneklerine olan uzaklığı hesaplamak gerektiğinden hesaplama maliyeti yüksektir, en iyi sonuçlara ulaşmak için hangi uzaklık tipinin kullanılacağı konusunda açık değildir ve k parametresine ihtiyaç duyulması gibi durumlar, KNN'nin dezavantajlarına örnek verilebilir (Pilavcılar, 2007)

4.1.4.1. Öklid uzaklık ölçüsü

Sınıflarda bulunan örnekler $X = (X_1, X_2, \dots, X_n)$ ve sınıflandırılmak istenen dokümanlar da $Y = (Y_1, Y_2, \dots, Y_n)$ şeklinde ifade edildiği düşünülerek iki vektör arasındaki uzaklık,

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.5)$$

formülü yardımıyla hesaplanır (Seçkin, 2011).

4.1.4.2. Kosinüs benzerlik ölçüsü

Eğitim vektörleri ile test vektörleri arasındaki ilişki aşağıdaki formül yardımı ile hesaplanabilir (Seçkin, 2011).

$$d_i = (wd_{i1}, wd_{i2}, \dots, wd_{ij}) \quad (4.6)$$

$$\text{sim}(d_i, q) = \cos\theta \quad (4.7)$$

$$\text{sim}(d_i, q) = \frac{d_i \cdot q}{|d_i||q|} = \frac{\sum_j W_{ij}xW_{qj}}{\sqrt{\sum_j w_{ij}^2} \sqrt{\sum_j w_{qj}^2}} \quad (4.8)$$

Burada; W_{ij} terimin doküman içerisindeki ağırlığı, d_i eğitim doküman vektörünü, q ise sınıf bulunması istenen vektördür. Benzerlik oranı 1'e en yakın olan n tane vektörün kategorisinden çok olanı dokümana atanır (Seçkin, 2011).

4.1.5. Naive Bayes

Veri madenciliği ve makine öğrenme probleminde sıklıkla kullanılan Bayes teoremine dayanan olasılıksal tabanlı bir sınıflandırma algoritmasıdır. Basit yapısı, hesaplama etkinliği ve yüksek sınıflandırma performansı ile birçok sınıflandırma çalışmasında başarıyla uygulanmıştır (Onan ve Korukoğlu, 2015). Naive Bayes sınıfları belli olan verileri kullanarak yeni gelen verinin hangi sınıfa ait olacağını bulmaya çalışır (Salur vd, 2017). Bir doküman ya da verinin hangi sınıfa ait olduğunu bulmak amacıyla koşullu olasılıkları hesaplanır. Klasik koşullu olasılık hesabı;

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4.9)$$

$P(A|B)$; B olayının olduğu bilindiğinde A olayının olma olasılığı

$P(B|A)$; A olayının olduğu bilindiğinde B olayının olma olasılığı

$P(B)$ ve $P(A)$; A ve B olaylarının ön olasılıkları

Dokümanların hangi sınıflara ait olduğunu hesaplamak amacıyla koşullu olasılık şu şekilde oluşturulur. Her biri ayrı bir özneliğe karşılık gelmek üzere bir metni temsil eden öznelikleri $X = \{x_1, x_2, \dots, x_d\}$ şeklinde ifade edilir. Olası sınıflarda $C = \{c_1, c_2, \dots, c_j\}$ şeklinde ifade edilir. Burada X özellik vektörünü, C ise sınıfları ifade eder (Kırmacı ve Oğul, 2015). Bayes kurallarına dayalı olarak formül aşağıdaki gibidir.

$$P(C_j|x_1, x_2, \dots, x_d) = \frac{P(C_j)P(x_1, x_2, \dots, x_d|C_j)}{P(x_1, x_2, \dots, x_d)} \quad (4.10)$$

Payda özelliklere bağlı olup, sınıf değişkenine bağlı olmadığından dolayı sabittir. Pay ise her x 'in diğer x 'lerden karşılıksız bağımsız olduğu koşullu bağımsızlık varsayımına göre aşağıdaki gibi yazılabilir.

$$P(C_j|x_1, x_2, \dots, x_d) = P(C_j)P(x_1|C_j)P(x_2|C_j)P(x_3|C_j) \dots P(x_d|C_j) \quad (4.11)$$

En yüksek olasılığa sahip kategori sınıf olarak seçilir (Özalp vd, 2013).

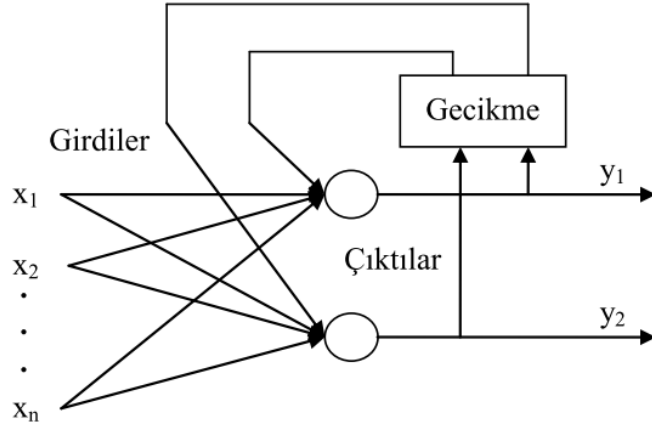
$$\operatorname{argmax}_c P(C = c) \prod_{i=1}^d P(x_i|C = c) \quad (4.12)$$

Bilinen bir sınıf için terim olasılıklarını hesaplama yöntemine göre Naive Bayes çok terimli (multinomial) ve çok değişkenli (multivariate) olmak üzere iki gruba ayrılır. Çok terimli yöntemde terimlerin metin içerisinde kaç kere geçtiği dikkate alınırken, çok değişkenli yöntemde terimlerin metin içerisinde var olup olmadıklarına bakılır (Seçkin, 2011).

4.1.6. Yapay Sinir Ağları

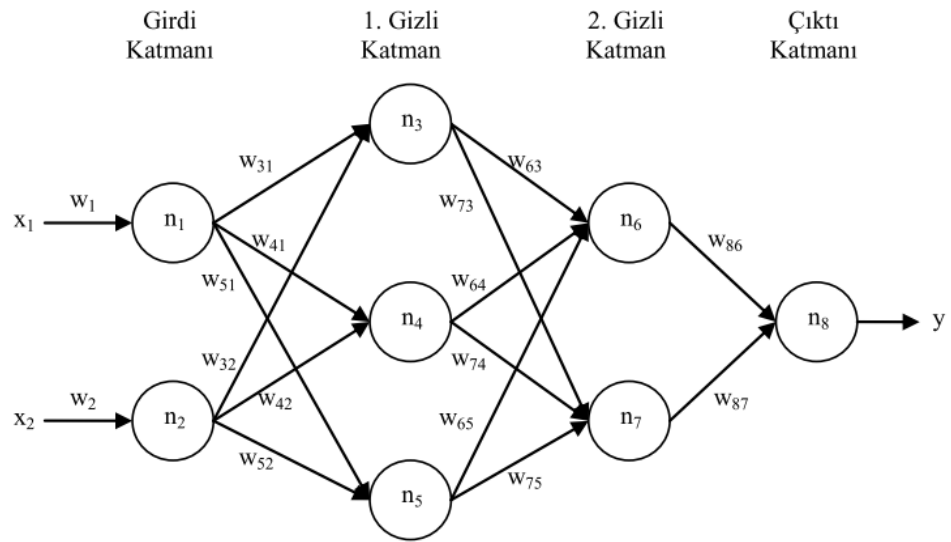
İlk olarak 1940'ların başında insan beyninin fonksiyonundan esinlenerek ortaya çıkan yapay sinir ağları (Artificial Neural Networks-ANN) (YSA), deneme yoluyla genelleme yapabilme, veriden öğrenebilme, sınırsız sayıda değişkenle işlem yapabilme gibi farklı özelliklere sahip bir makine öğrenmesi yöntemidir (Arslan vd, 2017). Beyindeki sinirlerin çalışmasından yola çıkılarak sistemlere öğrenme, hatırlama, bilgiler arasında ilişkiler oluşturma gibi çeşitli yetenekleri kazandırmayı amaçlar (Çeliksü, 2017). YSA'nın en küçük birimine nöron adı verilir. Nöronlar birleşerek katmanları, katmanlar ise bir ağı oluşturmaktadır. Girdi katmanı ve çıktı katmanı olmak üzere en az iki katman bulundurmak zorundadır. Girdi katmanı, amaçlanan çözüme ait verilerin ağ tarafından okunmasını sağlayan ve özellik sayısı kadar nöron içeren katmandır. Ağ tarafından işlenerek üretilen sonuçların dışarıya aktarıldığı katman ise çıktı katmanıdır. Çıktı katmanı tek bir nöron içerebildiği gibi çalışmanın amacına göre sınıf sayısı kadar da nöron katmanı içerebilmektedir (Arslan vd, 2017). Girdi ve çıktı katmanları arasında gizli katman yer almaktadır. Bu katman, girdi ve çıktı arasında meydana gelen birtakım hesaplamaları birleştirerek sonuçlandırır (Bayer, 2011).

Katmanlar arasındaki bağlantıların tiplerine göre yapay sinir ağları, ileri beslemeli (feedforward) ve yinelemeli (recurrent) olmak üzere ikiye ayrılır. YSA'da döngüsel yol oluşturan geri besleme (feedback) hattı varsa, yinelemeli ağ olarak ifade edilir. Senkronizasyonu sağlamak amacıyla gecikme bileşeni de kullanılmaktadır. İleri beslemeli yapıya göre daha küçük bir mimariye sahiptirler (Irmak vd, 2012). Şekil 4.3'te yinelemeli yapay sinir ağı mimarisi gösterilmiştir.



Şekil 4.3. Yinelemeli (recurrent) yapay sinir ağı mimarisi (Irmak vd, 2012)

Yapılarındaki esneklik, iyi temsil yetenekleri ve geniş sayıda öğrenme algoritmaları ile yapay sinir ağlarında en yaygın kullanılan ileri beslemeli yapay sinir ağı mimarisi bir girdi katmanı, bir veya daha fazla gizli katmanı ve bir de çıktı katmanı olan ağlardır. Bazı kaynaklarda ileri beslemeli yapay sinir ağları çok katmanlı yapay sinir ağları olarak da adlandırılmaktadır (Irmak vd, 2012). Şekil 4.4'te ileri beslemeli yapay sinir ağı mimarisi gösterilmiştir.



Şekil 4.4. İleri beslemeli yapay sinir ağı mimarisi (Irmak vd, 2012)

4.2. Metin Sınıflandırma Performans Ölçütleri

Metin sınıflandırma algoritmasının performansı ölçmek için çeşitli kriterler uygulanır. Bu kriterler doğruluk (accuracy), kesinlik (precision), duyarlılık (sensitivity), özgüllük

(Spesifite), negatif prediktif deęer (NPD), F-ölçeęi, ROC (Receiver Operating Characteristics) eęrisi gibi kriterlerdir (Amanet, 2017; Arslan vd, 2017; Haltaş vd, 2015; Talebi, 2013). Modelin başarısı doęru veya yanlış sınıflara atanan örnek sayılarıyla ilgilidir (Haltaş vd, 2015). Bu bilgiler karışıklık matrisi ile gösterilir. karışıklık matrisi Çizelge 4.1’de gösterilmiştir.

Çizelge 4.1. Karışıklık matrisi

Karışıklık matrisi		Tahmin edilen sınıf	
		<i>Sınıf=C₁ (Pozitif)</i>	<i>Sınıf=C₂ (Negatif)</i>
Doęru sınıf	Sınıf=C ₁	<i>TP (Doęru Pozitif)</i>	<i>FN (Yanlış Negatif)</i>
	Sınıf=C ₂	<i>FP (Yanlış Pozitif)</i>	<i>TN (Doęru Negatif)</i>

TP (Doęru Pozitif); doęru olarak sınıflandırılmış pozitif etiketlenen örneklerin sayısı
FN (Yanlış Negatif); yanlış olarak sınıflandırılmış negatif etiketlenen örneklerin sayısı
FP (Yanlış Pozitif); yanlış olarak sınıflandırılmış pozitif etiketlenen örneklerin sayısı
TN (Doęru Negatif); doęru olarak sınıflandırılmış negatif etiketlenen örneklerin sayısını ifade etmektedir.

4.2.1. Doğruluk

Sınıflandırma başarısının ölçülmesinde kullanılan en yaygın ve basit yöntem doğruluk oranıdır. Doğruluk oranı, doęru sınıflandırılmış örneklerin sayısının ($TP + TN$), toplam örnek sayısına oranını ($TP + TN + FN + FP$), ifade eder (Amanet, 2017). Sınıflandırma sürecinde kullanılan veri setinin dengesiz sınıf dağılımına sahip olması durumunda doğruluk oranının kullanılması yanıltıcı sonuçlar verebilmektedir. Örneęin; pozitif sınıftaki örnek sayısının negatif sınıftaki örnek sayısından fazla olması durumunda bütün test örnekleri pozitif olarak sınıflandırdığımız zaman sınıflandırma işleminin doğruluk oranı yüksek çıkacaktır. Örnek sayılarının birbirlerinden baskın olduğunda sadece doğruluk oranının deęerlendirme ölçüsü olarak kullanılması uygun deęildir. Doğruluk oranıyla birlikte dięer başarı ölçütlerinin kullanılması da gereklidir (Gündüz, 2013).

$$Doğruluk = \frac{TP+TN}{TP+TN+FN+FP} \quad (6.1)$$

4.2.2. Kesinlik

Kesinlik ölçütü pozitif olarak sınıflandırılmış örnekleri temel alır. Pozitif olarak sınıflandırılmış olan örneklerden kaç tanesinin gerçekten pozitif olduğunu ölçer (Gündüz, 2013). Doğru olarak sınıflandırılmış pozitif örnek sayısının toplam pozitif örnek sayısına olan oranıdır (Amanet, 2017).

$$Kesinlik = \frac{TP}{TP+FP} \quad (6.2)$$

4.2.3. Duyarlılık

Duyarlılık ölçütü de kesinlik ölçütü gibi pozitif olarak sınıflandırılmış örnekleri temel alır. Toplam pozitif örnek sayısından kaç tanesinin doğru olarak tanındığını ölçer (Gündüz, 2013).

$$Duyarlılık = \frac{TP}{TP+FN} \quad (6.3)$$

4.2.4. Özgüllük

Özgüllük, doğru olarak sınıflandırılmış negatif etiketli örnekleri esas alır. Toplam negatif örnek sayısından kaç tanesinin doğru olarak sınıflandırıldığını ölçer.

$$Özgüllük = \frac{TN}{TN+FP} \quad (6.4)$$

4.2.5. Negatif Prediktif Değer

NPD ölçütü negatif olarak sınıflandırılmış örnekleri temel alır. Negatif olarak sınıflandırılmış olan örneklerden kaç tanesinin gerçekten negatif olduğunu ölçer. Doğru olarak sınıflandırılmış negatif örnek sayısının toplam negatif örnek sayısına olan oranıdır.

$$NPD = \frac{TN}{TN+FN} \quad (6.5)$$

4.2.6. F-Ölçeği

F-ölçütü kesinlik ve duyarlılık ölçütlerinin birleşiminden oluşmuştur. Kesinlik ve duyarlılık değerlerinin ağırlıklı harmonik ortalamasının alınması sonucu hesaplanır (Gündüz, 2013). F-ölçeğinin değerlerinin yüksek olması sınıflandırma performansının iyi olduğunu ifade eder.

$$F = \frac{2 \times \text{Kesinlik} \times \text{Duyarlılık}}{TP + TN + FN + FP} \quad (6.6)$$

4.2.7. ROC Eğrisi

Genellikle dengesiz dağılımlı sınıflandırmaların performansını ölçmek için kullanılır. ROC eğrisinin altında kalan alan (Area Under Curve-AUC) ROC puanı olarak adlandırılır. ROC eğrisi sınıflandırma eşik değerine göre doğru pozitiflerinin sayısının, yanlış pozitiflerin bir fonksiyonu olarak çizilmesiyle oluşur. ROC eğrisinin altında kalan alan '1' olduğunda pozitiflerin iyi bir şekilde negatiflerden ayrıldığı anlamına gelirken '0' olduğunda ise herhangi bir pozitif bulunamadığı anlamı taşır (Amanet, 2017). AUC değerinin formülü,

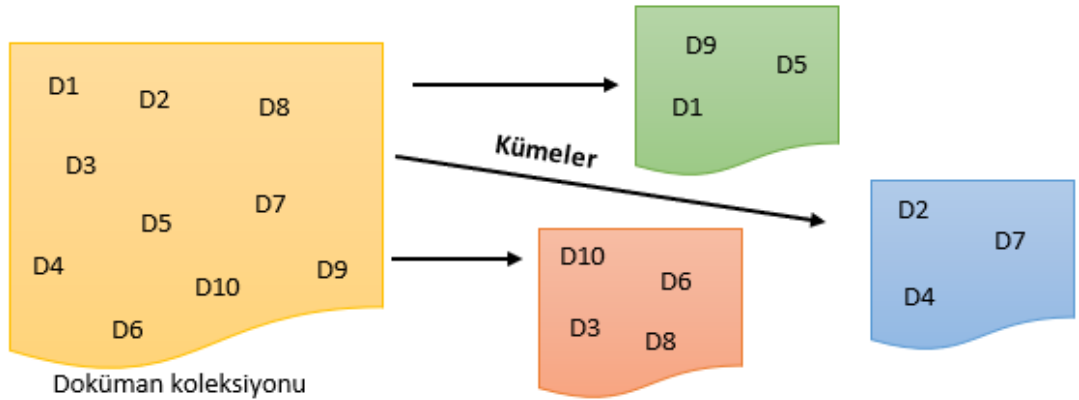
$$AUC = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (6.7)$$

şeklinde ifade edilir.

5. METİN KÜMELEME

Metin madenciliği yöntemlerinden biri olan metin kümeleme denetimsiz öğrenme modellerinden biridir. Kümeleme yönteminde herhangi bir sınıflandırma yapılmamış dokümanları gruplandırmak amacıyla dokümanlar arasındaki benzerlik ve uzaklıklara göre gruplara ayrılmasıdır (Tunalı ve Bilgin, 2012). Sınıflandırma işlemlerinde yeni bir dokümanın hangi sınıfa ait olduğu bellidir ama kümeleme işlemlerinde belli değildir. Kısaca kümelemedeki temel amaç dokümanları benzerliklerine göre alt gruplara ayırmaktır. Sınıflandırmada olduğu gibi ortak özellikleri olan dokümanlar tek bir grupta toplanır. Alt kümelere ayırmada bulunan kurallar yardımıyla diğer bir dokümanın hangi kümeye ait olduğu bulunur (Canhası, 2007).

Kümeleme işlemi sonucunda elde edilen gruplar arasında uzaklık ne kadar fazla ise kümeleme performansı o kadar iyidir. Aynı küme içerisindeki dokümanlar benzer konu içeriğine sahipken farklı kümelerdeki dokümanlar farklı konu içeriğine sahiptir. Örneğin, Şekil 5.1 doküman koleksiyonlarının kümeleneşine örnek olarak verilebilir.



Şekil 5.1. Doküman koleksiyonu kümeleneşini

Kümeleme işlemleri metin önışlem aşamasında boyut indirgeme (terim ayıklama) yöntemi olarak da kullanılabilir. Bu sayede yakın anlamlı kullanılan kelimeleri bir kavram altında toplayıp tek bir özellik haline getirebilir (Amasyalı vd, 2012). Örneğin; 'gül', 'papatya', 'menekşe' gibi çiçek isimleri 'çiçek' kavramı altında tek bir özellik olarak ifade edilebilir. Metin kümeleme işlemlerinde kullanılan birçok denetimsiz

öğrenme algoritmaları vardır. K-means, hiyerarşik kümeleme ve SOM (Self Organizing Maps) yöntemleri sıklıkla kullanılmaktadır.



6. BİRLİKTELİK ANALİZİ

Veri madenciliği tekniklerinden biri olan birliktelik kuralı (association rule), verilerin birlikte olma durumlarını inceleyerek, eş zamanlı ve olayların birbirine bağlı olarak birlikte geçme durumlarını ortaya koyan bir yöntemdir (Ateş ve Karabatak, 2017). Nesnelerin veya verilerin birlikte bulunma durumlarını inceler. Temel amacı; veriler arasındaki kuralları ortaya çıkarmaktır (Alaeddinoğlu vd, 2014). Birçok kullanım alana sahiptir. Özellikle pazar sepet analizinde ve müşterilerin alışveriş alışkanlıklarını belirleme sıkça kullanılmaktadır (Erpolat, 2012). Pazar sepeti analizi, müşterilerin alışveriş alışkanlıklarını ortaya koymak, birlikte satılan ürünlerini belirlemek ve market düzeninin sağlanması için kullanılır.

İlk olarak 1993 yılında, Agrawal, Imiellinski ve Swami tarafından birliktelik kuralının matematiksel modeli oluşturulmuştur (Öztürk ve Tanrısevdi, 2017). Bu modelde $I = \{i_1, i_2, \dots, i_m\}$ nesnelere veya ürün kodlarını, D ; tüm hareketleri, T ; ürünlerin her bir hareketinin ifade eder. TID ise her harekete ait olan tek bir belirteci gösterir (Alaeddinoğlu vd, 2014). Birliktelik kuralı $A_1, A_2, \dots, A_m \rightarrow B_1, B_2, \dots, B_n$ gibi tanımlanabilir.

A_m ve B_n ürün ve nesnelere gösterir. Bu kural, A_1, A_2, \dots, A_m ürün veya nesnelere meydana geldiğinde, belli bir sıklıkta B_1, B_2, \dots, B_n ürün veya nesnelere de aynı olaya ve hareket içerisinde yer aldığı ifade etmektedir. Birliktelik kuralları destek ve güvenilirlik değerleri üretilir. Bu değerlerin eşik değerleri kullanıcı tarafından belirlenir. Birliktelik kuralı $A \rightarrow B$ olarak ifade edildiğinde,

$$\text{Destek}(A \rightarrow B) = \frac{D \text{ içinde } A \text{ ve } B' \text{ nin beraber bulunduğu durum sayısı}}{D; \text{ tüm hareketlerin sayısı}} \quad (7.1)$$

$$\text{Güven}(A \rightarrow B) = \frac{D \text{ içinde } A \text{ ve } B' \text{ nin beraber bulunduğu durum sayısı}}{D \text{ içinde sadece } A' \text{ nin bulunduğu durum sayısı}} \quad (7.2)$$

destek ve güven değerleri şeklinde hesaplanır (Alaeddinoğlu vd, 2014).

Destek, tüm hareketler içinde ilgili ürün kümesini içeren hareketlerin yüzdesidir (Özçakır ve Çamurcu, 2007).

Güven, A 'yı içeren durumların, B 'yi de içermeye yüzdesidir.

İki ürün veya nesne arasında oluşturulan kuralın önemli olabilmesi için destek ve güven kriterlerinin yüksek olması gerekir. Fakat her zaman bu kriterlerin yüksek olması oluşturulan bağıntının önemli olduğu anlamı taşımaktadır. Bundan dolayı kuralın ne derecede önemli ve ilginç olduğunu belirlemek için lift değeri kullanılmaktadır.

$$Lift(A \rightarrow B) = \frac{Destek(A \cap B)}{Destek(A) \times Destek(B)} \quad (7.3)$$

Lift değerinin '1' olması ilginçliğin olmadığına '1'den büyük veya '1'den küçük olması durumunda ilginçliğin arttığını belirtmektedir (Ateş ve Karabatak, 2017).

Birliktelik analizi için birçok algoritma kullanılmaktadır. Birliktelik kuralında kullanılan algoritmalar ardışık ve paralel olmak üzere ikiye ayrılır. Ardışık algoritmalar, ürün kümelerinin oluşturulduğu ve sayıldığı mantıksal algoritmalar. Büyük ürün kümelerinin paralellik sağlayarak oluşturulan algoritmalar paralel algoritmalar (Erpolat, 2012). Ardışık algoritmalar ile paralel ve dağıtılmış algoritmalar Çizelge 6.1'de verilmiştir.

Çizelge 6.1. Ardışık ve paralel algoritmalar

Ardışık algoritmalar	Paralel ve dağıtılmış algoritmalar
AIS	CD (Sayım Dağılımı)
SETM	PDM (Paralel VM)
Apriori	CCPD (Ortak Aday Bölünmüş Veri Tabanı)
Apriori-TID	DD (Veri Dağılımı)
Apriori-Hybrid	IDD (Akıllı Veri Dağılımı)
OCD (Sıra dışı Aday Belirleme)	HPA (Bağlantı Kurallarının Çırpı Temelli Paralel Madenciliği)
Bölümleme Tekniği	PAR (Paralel Bağlantı Kuralları)
Örnekleme Tekniği	DMA (Dağıtılmış Madencilik Algoritması)
CARMA (Sürekli Bağlantı Kuralı Madenciliği)	Candidate Distribution (Aday Dağılımı)
DIC (Dinamik Nesne Sayımı)	SH (Çarpık Taşıma)
FP-Growth	HD (Hibrid Dağılımı)

Birliktelik işlemlerinde en çok kullanılan algoritma, ardışık algoritmalar içerisinde yer alan Apriori algoritmasıdır (Erpolat, 2012). Apriori algoritması “Agrawal” ve “Srikant” tarafından geliştirilmiş ve yaygın nesnelerin bir önceki bilgilerini kullanmasından dolayı “prior (önceki)” kelimesinden meydana gelmiştir (Bayer, 2011). Veri setinde sık geçen geniş nesne kümelerinin belirlenmesinde kullanılır. Bir nesne kümesinin kullanıcı tarafından belirlenen minimum destek sınırının üstünde bir destek seviyesine sahip olması geniş nesne kümeleri olarak adlandırılır. İlk olarak algoritma tarafından yapılan taramada nesne kümelerinin destek değerleri hesaplanarak kullanıcı tarafından belirlenen minimum destek değeri ile karşılaştırılır. Karşılaştırma durumuna göre her bir nesnenin geniş olup olmadığı incelenir. Bundan sonraki her tarama işleminde bir önceki tarama işleminde belirlenen geniş nesne kümelerinden başlar ve tekrardan yeni geniş nesne kümeleri oluşturur. Burada oluşan yeni geniş nesne kümelerine aday nesne kümeleri ismi verilir. Taramanın sonunda hangi nesne kümesinin geniş nesne kümesi olduğu kontrol edilir. Bir sonraki tarama işlemi bu şekilde devam eder ve veritabanının sonuna kadar nesne kümelerinin destek değeri hesaplanır (Erpolat, 2012). Bu tarama işlemi başka yeni geniş nesne kümeleri oluşturulmayana kadar devam eder. Bir başka ifadeyle oluşturulan nesne kümelerinin hepsinin destek seviyeleri kullanıcı tarafından girilen eşik değerinden küçük oluncaya kadar devam etmektedir. Tıp, eğitim, internet gibi alanlarda sıkça uygulanmaktadır (Bayer, 2011).

7. UYUM ANALİZİ VE FRIEDMAN TESTİ

Uyum analizi (Correspondence analysis; CA), kategorik deęişkenlerin yer aldığı veri matrisinde deęişkenler arasındaki ilişkileri analiz eden ve bunları iki boyutlu uzayda noktalarla gösterilmesine olanak sağlayan çok deęişkenli açıklayıcı istatistiksel bir tekniktir. Uyum analizi işlemlerinde çapraz (kontenjans) tablolar $r \times c$ şeklinde bir matris ile ifade edilmektedir. Burada r , satırları c ise sütunları ifade etmektedir. Kontenjans tabloları satır ve sütun deęişkenleri arasındaki ilişkinlerin, benzerliklerin ve farklılıkların yorumlanmasına imkan tanır. Uyum analizi genellikle nominal deęişkenler için uygulanmaktadır. İki yönlü kontenjans tabloların analizinde basit uyum analizi, çok yönlü kontenjans tabloların analizinde çoklu uyum analizi kullanılmaktadır. Uyum analizindeki temel amaç deęişkenleri iki boyutlu uzayda detaylı olarak göstermek ve deęişkenler arasındaki ilişkilerin kolay bir şekilde yorumlanmasını sağlamaktır (Çakmak vd, 2012; Gün ve Erdem, 2014; Cangür vd, 2005).

Friedman testi, ilişkili örneklem tek yönlü varyans analizinin non-parametrik karşılığıdır. Verilerin normal dağılmadığı, zaman sayısının 3 ve üzeri olduğu ilişkili deęişkenlerin dağılımları arasında fark olup olmadığını test etmek için kullanılır. Kısaca bir örneklemin tekrarlı ölçümleri arasında farkın olup olmadığını test eder. Friedman testi için kurulan H_0 hipotezleri farklılığın olmadığını, alternatif hipotezler ise farklılığın olduğunu ifade etmektedir.

8. ANKİLOZAN SPONDİLİT

Aksiyel iskeleti etkileyen, eklemlerin tutulumuna ve bel ağrısına yol açan ve fonksiyonel bozulmaya sebep olan Ankilozan Spondilit (AS), sakroiliak eklemler, omurga, periferik eklemler ve entezis bölgeleri ile karakterize edilen ve Spondiloartrit (SpA) hastalıkları içerisinde yer alan iltihaplı romatizmal bir hastalıktır. Eklem ve omurga hareketliliğini etkileyen AS, fiziksel engelliğe yol açabilir ve yaşam kalitesinde azalmaya sebep olabilir (Atasoy, 2015). Hastalığın bireylerde ortalama başlama yaşı 24'tür ve 10 yaşından önce ve 45 yaşından sonra hastalığın meydana gelme ihtimali çok azdır. Hastaların %15'inde, AS 16 yaşından önce başlar. AS tanısı koyabilmek için radyografik değişikliklerin gelişmesine bağlı olarak 8-10 yıl geçmesi lazımdır (Bülbul, 2015). AS prevalansı dünyada ortalama %0.9 iken Türkiye'de %0.49'dur (Aksu, 2015). Ancak HLA-B27 pozitifliği, prevalansı etkilemektedir (Çakır, 2015). Erkeklerde hastalık yaklaşık olarak 2-3'e 1 oranında kadınlardan daha sık görülür. İzmir'de yapılan çalışma sonucunda E/K oranı 1,2 olarak bulunmuştur (Karataş, 2015).

AS hastaların %75'inde, klinik belirtileri Calin vd tarafından (1977) tanımlanan inflamatuvar bel ağrısı hastalığının ilk belirtisidir. Calin kriterlerine göre en tipik klinik belirtiler; 40 yaşın altındaki bireyleri etkilemesi, sinsi bir başlangıcının olması, üç aydan daha uzun süren tanımlanmamış bel ağrısının olması, sabah tutukluğu olması ve ağrının dinlenme ile artması ve egzersiz ile azalması sayılabilir. Bu belirtilerden en az dördünü taşıyan hastaya inflamatuvar bel ağrısı tanısı konulabilmektedir (Atasoy, 2015). AS hastasının ilk yakınması göğüs ağrısı olabilir. Öksürme ve aksırmayla ağrı belirginleşir, buna plöritik göğüs ağrısı denir (Özen, 2017). Öksürme ve ani hareketler sonrasında bel ağrısında artış görülmektedir. Yaklaşık olarak hastaların %20'sinde, kalça ve omuz tutulumları görülür (Kul, 2016). Ankilozan Spondilit'in başlangıç halinde ve seyrinde hastalarda yorgunluk, uyku bozukluğu, iştahsızlık, kilo kaybı, subfebril ateş gibi belirtiler meydana gelebilir. Sabahları ilk kalkışta ortaya çıkan topuk ağrısı ve ayakların üstüne basamama AS'ye özgü bir bulgu olan entezit sonucunda gelişen önemli bir yakınmadır. Ayrıca entezit topuğun arka kısmında, diz, dirsek, el bileği, kalça eklemlerinin çevresinde veya vücudun diğer bölgelerinde de ağrı ve şişliğe sebep olabilir. Akut anterior üveit AS'nin en yaygın ekstra-artikular tutulumu

olup bazı hastalarda üveit hastalığının ilk bulgusu olabilir ve üveitli hastalarda gözde kızarıklık, ağrı, bulanık görme, lakrimasyonda artış, fotofobi ve miyozis görülmektedir (Kaya, 2016). Doğru tedavi edilmediğinde anterior veya posterior sineşiler, katarakt, glokom, maküler ödem gibi komplikasyonlar ortaya çıkabilir (Kul, 2016). Ankilozan spondilitte hastalarında meydana gelen kardiyak tutulum sonucunda aort yetmezliği, asendan aortit, aort dilatasyonu, iletim blokları, mitral yetmezlik, sol ventrikül disfonksiyonu, perikardit ve kardiyomiyopati gelişebilir (Özen, 2017).

AS sınıflandırma kriterleri olarak 1984 modifiye New York kriterleri kullanılır (Çoban, 2016). Bu kritere göre tanı koyulmaktadır. 1984 modifiye New York kriterleri Çizelge 8.1’de verilmiştir.

Çizelge 8.1. Ankilozan spondilit için 1984 modifiye New York kriterleri (Çoban, 2016)

Klinik Kriterler

1. Üç ay veya daha uzun süren, dinlenme ile geçmeyip, egzersiz ile düzelen ben ağrısı ve tutukluğu
2. Lomber omurga hareketlerinin sagittal ve frontal planlarda kısıtlılığı
3. Göğüs ekspansiyonunun yaş ve cinse göre düzeltilmiş normal değerlere göre kısıtlanması

Radyolojik Kriterler

1. Bilateral grade 2-4 sakroiliit
 2. Unilateral grade 3-4 sakroiliit
-

Kesin AS: Bir radyolojik kriter ve klinik kriterlerden biri,

Olası AS: Tek başına üç klinik kriter veya bir radyolojik kriter

Ankilozan Spondilitin tedavisinde kullanılan Non-Farmakolojik tedaviler, Farmakolojik tedaviler ve cerrahi tedaviler ile ağrı ve sabah sertliği semptomlarını azaltmak, çeşitli fonksiyonları yeniden kazandırmak, normal postürü, kas kuvvetini korumak ve arttırmak, kontraktürleri, deformiteleri ve osteoporozu önlemek, göğüs ekspansiyonunu koruyup solunum kapasitesinin arttırmak amaçlanmaktadır (Karataş, 2015; Özen, 2017).

Non-Farmakolojik tedaviler; çeşitli fiziksel tedavi yöntemleri ve hastalara yüzmek, hızlı yürümek, bisiklete binmek, kürek çekme, masa tenisi, voleybol oynamak gibi çeşitli aktiviteler önerilir. Kum, güneş ve kaplıca tedavileri de AS için kullanılan diğer tedavilerdir. Hastaların fazla güç gerektiren sporlardan kaçınması gerekir (Özen, 2017). Torasik kifoz gelişimini azaltmaya yönelik sert yatak ve ince yastıkta yatmak önerilir (Karataş, 2015). Osteoporoz gelişim riskine karşılık D vitamini ve kalsiyum miktarı yüksek olan ürünlerle beslenmelidir (Elbey, 2015).

Farmakolojik tedaviler; AS'nin tedavisinde kullanılacak ilaçlar hastalığın ilerleme evrelerine göre steroid olmayan anti inflamatuvar ilaçlar, antiromatizmal ilaçlar, analjezik ilaçlar, kortikosteroidler, anti-Tümör nekrosis faktör (TNF) olarak gruplandırılır (Atasoy, 2015). Steroid olmayan anti inflamatuvar ilaçlar ağrısı, sabah tutukluğu ve eklem katılığı olan AS'li hastalarda ilaç tedavisi olarak ilk sırada kullanılır (Elbey, 2015).

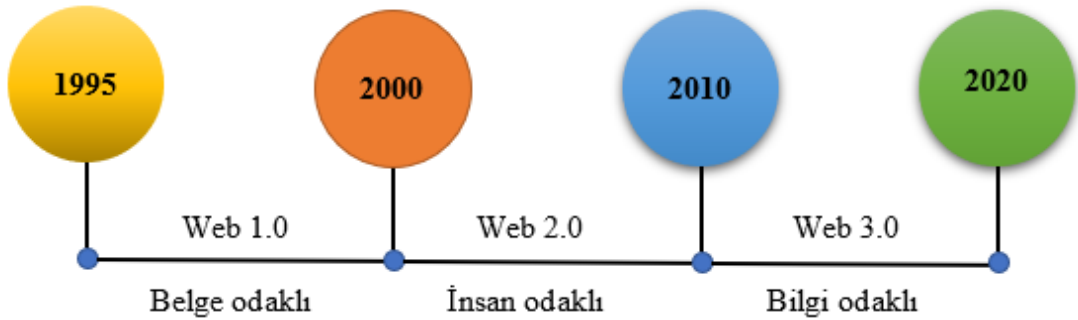
Cerrahi tedaviler; AS hastalığı ilerlemiş olan kişilerde fonksiyon kayıplarının düzeltilmesinde kullanılır. Kalça eklemlerin düzeltilmesinde sıklıkla kullanılmaktadır (Atasoy, 2015). Ayrıca vertebranın osteotomisi şiddetli spinal deformiteyi düzeltmek amacıyla cerrahi müdahaleler uygulanır (Elbey, 2015).

9. SOSYAL MEDYA

Günümüzde internet kullanımının hızla yaygınlaşması ve web teknolojilerinin gelişmesiyle birlikte insanların duygu ve düşüncelerini paylaştıkları sosyal medya platformları artmaya başlamıştır. Bu platformlar çocuk, genç, yaşlı fark etmeksizin tüm kesimler tarafından kullanılmaktadır. Kullanıcılar bu sayede kendi duygu ve düşüncelerini resim, yazı veya video şeklinde paylaşabilmektedir. Zaman ve konum fark etmeksizin dünyanın herhangi bir yerindeki kullanıcı ile görüntülü veya görüntüsüz olarak iletişim kurabilmektedir. Kısaca sosyal medyayı insanların paylaşım ve iletişim sınırı olmadan özgürce duygu ve düşüncelerini paylaştıkları dijital medya platformları olarak tanımlayabiliriz (Sönmez, 2017).

Global rekabetin hızla arttığı dönemimizde firmalar bu platformlar sayesinde ürünlerini tanıtabilmekte ve ürün hakkındaki yorum ve görüşlere rahatlıkla ulaşabilmektedir. Özellikle bu platformlar firmaların kendi ürünlerini tanıtımını ve pazarlamasını yapabilecek ortamlar sunabilmektedir. Ayrıca bu platformlar firmaların geniş kitlelere ulaşabilmesine de olanak sağlar.

Sosyal medya platformları web teknolojilerinin gelişmesiyle yaygınlaşmaya başlamıştır. Şekil 9.1’de web teknolojilerinin gelişim dönemleri verilmiştir. İlk web teknolojisi olarak 1995’de web 1.0 geliştirilmiştir. İnternetin ilk dönemleri olarak adlandırılır. Bu dönemde insanlar araştırmak istediği bilgiyi web sayfaları üzerinde araştırır, bulur ve istediği bilgiyi bulduktan sonra sayfadan ayrılırdı. Web 1.0’da kullanıcılar sadece okuyucuydu, web sayfalarında herhangi bir yorum yapma, içerik üretme veya katkı sağlama gibi imkanları yoktur.

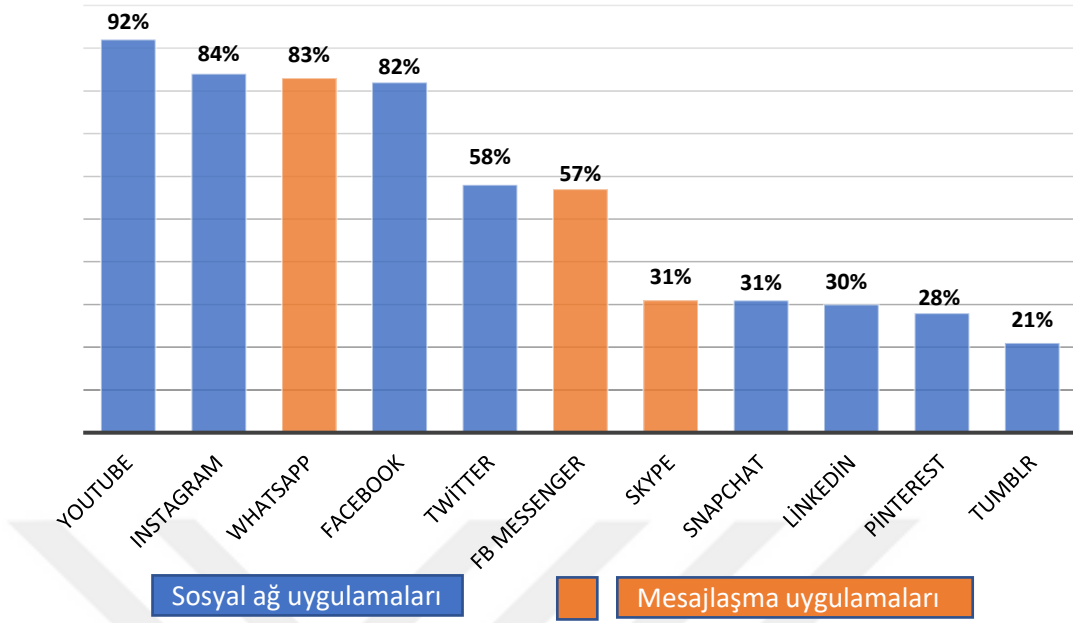


Şekil 9.1. Web teknolojilerinin gelişim dönemleri (Sönmez, 2017)

Kullanıcılar Web 2.0'nın gelişmesiyle internet hem katılımcı hem de okuyucu olarak yer almaya başladılar (Cingiz ve Diri, 2012). Bu teknoloji kullanıcılar tarafından internette içeriklerin oluşturulmasına ve bu içeriklerin diğer kullanıcılar tarafından erişimine imkan sağlar. Sosyal ağ siteleri, web tabanlı özgür ansiklopediler ve bloglar gibi sosyal iletişim araçları insanlarla tanıştırmıştır.

Web 2.0'dan sonra web 3.0 geliştirilmiştir. İnternet kullanımının insanların elinden çıktığı ve kontrolün artık yazılımların elinde olduğu bir dönemdir. Kısaca “yapay zekalı” web olarak adlandırılabilir. Web 3.0 veritabanları ile ilişkili olan, bir kullanıcının neyi sevip sevmediğini anlayan, buna göre içerik üreten ve sunan, bizim ne istediğimiz anlayan bir teknoloji diyebiliriz (Sönmez, 2017).

Her yıl “We are social” ve “Hootsui” tarafından sosyal medya istatistikleri hazırlanmaktadır. “Dijital in 2019 in Turkey” adıyla 31 Ocak tarihinde yayınlanan sosyal medya istatistiklerine bakıldığında 82,4 milyon nüfuslu ülkemizin 52 milyonu aktif sosyal medya kullanıcısıdır. Bir önceki yıla göre aktif sosyal medya kullanıcı sayısında %2'lik bir artış yaşanmıştır. Ülkemizde teknoloji kullanımına bakıldığında yetişkinlerin %98'nin telefon kullandığı ve bunların %77'sinin akıllı telefon kullandığı görülmektedir. İnsanlar günde ortalama 7 saat vakitlerini internette geçirmektedir. Günde ortalama 2 saat 46 dakika sosyal medyada ve 3 saat 9 dakika televizyon başında zaman harcamaktadır. 59,3 milyon internet kullanıcısının 56,3 milyonu telefonlarından internete bağlanmaktadır. Ülkemizde en çok kullanılan medya platformları Şekil 9.2 'de verilmiştir.



Şekil 9.2. Türkiye’de en çok kullanılan sosyal medya platformları

Ülkemizde en çok kullanılan sosyal medya uygulamalarına bakıldığında 52 milyon sosyal medya kullanıcısının %92’si Youtube, %82’si Facebook, %84’ü Instagram ve %58’i Twitter kullanmaktadır (Bayrak, 2019). Bu istatistikler sosyal medya araçlarının ne kadar sık kullanıldığını açık bir şekilde göstermektedir.

9.1. Twitter

Twitter 2006 yılında Jack Dorsey ve arkadaşları tarafından kurulmuştur (Anonim, 2019b). Twitter, kullanıcılar görüş ve fikirlerini, fotoğraflarını, ilgi alanlarıyla ilgili paylaşımlar ve günlük durumlarını takipçileri ile paylaştıkları bir sosyal ağ uygulamasıdır (Cingiz ve Diri, 2012). Gönderiler 280 karakter sayısı ile sınırlıdır ve gönderilere tweet adı verilir.

Hashtag (#) yardımıyla bir konu hakkında etiketler oluşturulup o konu hakkında gönderiler paylaşılabilir. Karakter sayısının kısıtlı olması kullanıcılar tarafından yeni yazım şekillerinin ortaya çıkmasına sebep olmuştur. Örneğin çoğu kelimeler kısaltılarak yazılmaktadır. Örneğin “selam” “slm” şeklinde yazılabilir. Kendisine haz bu yazım şekillerinden dolayı, analiz edilecek tweetlerin işlenmesi zorlaşmaktadır. Twitter’in API üzerinden tweetlere erişim olanağı sağlaması ve popüler bir sosyal ağ

uygulaması olması arařtırmacıların iřtahını kabartmaktadır. Birçok alanda yapılacak olan alıřmalar için iyi bir veri kaynađı oluřturmaktadır. 2019 yılına ait Trkiye’deki Twitter kullanıcı sayısı Őekil 9.3’te verilmiřtir.



Őekil 9.3. 2019 yılına ait Trkiye’deki Twitter kullanıcı sayısı

“We Are Social” ve “Hootsuite” tarafından 31 Ocak tarihinde yayınlanan “Digital in 2019” raporuna gre Twitter 250,8 milyon kullanıcıya sahiptir (Bayrak, 2019).

9.2. Facebook

Mark Zuckerberg tarafından 2004 yılında kurulmuřtur. Facebook kullanıcıların bařka kullanıcılarla iletiřim kurmasını ve bilgi alıřveriřinde bulunmalarını amalayan bir sosyal ađ sitesidir (Anonim, 2019a). Facebook’ta kullanıcıların kiřisel bilgilerinin, arkadař listesinin, paylařımlarının, beđendiđi sayfaların ve ye olduđu grupların yer aldıđı profil sayfaları bulunmaktadır (Hablemitođlu ve Yıldırım, 2012). Kullanıcılar bařkalarına mesaj gnderebilir ve grntl olarak iletiřime geebilmektedir. Gruplara katılabilir ve grupların amacına ynelik bu ortamlarda paylařım yapabilirler. Fotođraftan dosyaya kadar her trl ierik paylařımını yapılabilir. 2019 yılına ait Trkiye’deki Facebook kullanıcı sayısı Őekil 9.4’te verilmiřtir.



Őekil 9.4. 2019 yılına ait Trkiye’deki Facebook kullanıcı sayısı

9.3. Youtube

Youtube 2005 yılında Jawed Karim, Steve Chen ve Chad Hurley tarafından kurulmuştur. 2006 yılında Google tarafında satın alınmıştır (Anonim, 2019c). Ülkemizde en çok kullanılan sosyal medya sitesidir. Video paylaşım sitesidir. Youtube aracılığıyla her gün milyonlarca video paylaşılmakta ve izlenmektedir (Dondurucu ve Ulucay, 2015).

Teknik olarak bir ücret talep edilmeksizin kullanıcılar tarafından video paylaşılabilir ve görüntülenebilir. Kullanıcılar tarafından kendi videolarını yayımlayacakları kanallar oluşturmaya izin verir. Videolar diğer sosyal medya sitelerinde paylaşılabilir. Videolar hakkında yorumlar yapılabilmektedir. Hem amatörler hem de profesyoneller tarafından içerikler üretilebilmektedir (Mutlu ve Bazarcı, 2017).

9.4. Instagram

Instagram 2010 yılında kurulmuş ve 2012'de Facebook tarafından satın alınmasıyla büyük kullanıcı kitlesine ulaşmıştır. Kullanıcılar tarafından fotoğraf ve video paylaşmak için kullanılır. Paylaşım yapılmadan önce fotoğraflara çeşitli dijital filtrele uygulanabilir. Paylaşımlarla ilgili yorumlar ve beğeniler yapılabilir. Paylaşımların diğer sosyal ağ hesaplarında görülmesine de olanak sağlar (Yeniçıktı, 2016). 2019 yılına ait Türkiye'deki Instagram kullanıcı sayısı Şekil 9.5'te verilmiştir.



Şekil 9.5. 2019 yılına ait Türkiye'deki Instagram kullanıcı sayısı

9.5. Forum sayfaları

Forum sayfaları genellikle bir konu hakkında tartışmaların, fikir, bilgi veya tecrübe paylaşımının yer aldığı ortamlardır. Forum sayfalarında genellikle belli etiketler altında tartışmalar başlatılabilir. Forumlarda etkili üye olabilmek için paylaşılan

sorulara cevap verilmesi veya yararlı bilgilerin paylaşılması gerekir. Forumları diğer sosyal medya araçlarından ayıran en önemli farkı bir konuya yönelik binlerce forum sayfasının olmasıdır (Dondurucu ve Ulucay, 2015).

9.6. Blog

Bloglar, kullanıcıların günlük, haftalık, aylık gibi belli zamanlarda düşüncelerini yazdıkları, içerik paylaştıkları ve güncelledikleri web siteleridir. İnternet günlüğü olarak da tanımlanmaktadır. Genellikle metinsel paylaşımların yapıldığı bloglar tarihlere, kategorilere ve etiketlere göre arşivlenebilmektedir. Tüm yayınlar süresiz olarak arşivlenir. Blog yazarları blogger olarak adlandırılır.

Teknik olarak blogların sayfa tasarımı değiştirilebilir, farklı blog araçlarıyla etkileşime girilebilir, iki veya daha fazla yazarla birlikte ortaklaşa kullanılabilir (Erben ve Salı, 2015). Bloglarda yayınlar hakkında yorumlar yapılabilir ve bu yorumlar herkes tarafından görülebilir (Vural ve Bat, 2009).

10. UYGULAMA

10.1. Araştırmanın Amacı ve Problemin Tanımlanması

Dijital ortamların kullanımının artmasıyla bu ortamlarda kayıt altına alınan veriler artmaktadır. Kullanıcılar tarafından sağlanan bu veriler yapısal olmayan verilerdir. Bu veriden anlamlı bilgilerin elde edilebilmesi ve analiz edilebilmesi için yapısal forma dönüştürülmesi gerekir.

Özellikle kullanıcı sayısı oldukça fazla olan sosyal medya sitelerinde her gün milyonlarca veri paylaşılmaktadır. Bu veriler doğru bir şekilde analiz edildiği takdirde çok farklı bilgilere ulaşılabilmektedir.

Bu çalışmada Facebook'ta 10000'e yakın üyeye sahip 'Ankilozan Spondilit Hasta Derneği' adındaki grupta paylaşılan gönderiler analiz etmek ve hastalıkla ilgili anlamlı bilgi çıkarımlarında bulunmak hedeflenmiştir. Bu grupta her gün AS ile ilgili farklı konularda çeşitli gönderiler paylaşılmaktadır. Üye sayısının ve gönderi paylaşım sayısının fazla olması AS hastalığı için devasa bir veri kaynağı oluşturmaktadır. Bu veriler analiz edildiği takdirde hastalık ve hastalar hakkında çok çeşitli bilgilere ulaşılabilmektedir. Bu verileri tek tek okumak ve bunlardan çeşitli kategoriler altında bilgi çıkarımları yapmak oldukça zordur. Bu metinsel verilerin kolay bir şekilde yapısal forma dönüştürülmesi ve çeşitli analizlerin yapılabilmesi için metin madenciliği yöntemleri kullanılır.

Bu tez çalışması ile AS hakkında paylaşılan gönderiler şikayet, doktor, tedavi, tahlil, ilaç, kişisel deneyim ve diyet, spor, rapor ve diğer hastalıklar kategorileri altında toplanan verileri metin madenciliği yöntemleri ile yapısal hale dönüştürerek analiz etmek, sınıflandırmak ve kategorilere ait bilgi çıkarımları yapmak amaçlanmıştır. Metin sınıflandırma algoritmalarından Naive Bayes, KNN, SMO ve karar ağaçları algoritmaları terim frekansı (TF), ters doküman frekansı (TFxIDF) ve binary ağırlıklandırma yöntemlerine göre sınıflandırma performansları ölçülerek literatüre kazandırılacaktır. Ayrıca ülkemizde metin madenciliği yöntemlerinin sağlık alanında uygulanmasına yönelik çalışma sayısı yok denilecek kadar azdır. Bu çalışma sayesinde metin madenciliğinin sağlık alanına uygulanması ile geniş kapsamlı bilgiler elde edilebileceği vurgulanmaktadır.

10.2. Kullanılan Araçlar

Çalışmada kullanılan bilgisayar araçları ve bu araçlar ile yapılan işlemler Çizelge 10.1’de verilmiştir.

Çizelge 10.1. Kullanılan araçlar ve yapılan işlemler

Kullanılan araçlar	Yapılan işlemler
Text2ArffV5	Metinsel verilerin işlenmesinde, kelime köklerinin bulunmasında ve ağırlıklandırılmasında kullanılmıştır.
Excel	Verilerin toplanması ve düzenlenmesinde, bilgi çıkarımlarının yapılmasında, grafik oluşturmada kullanılmıştır.
WEKA	Sınıflandırma işlemleri ve sınıflandırma algoritmalarının uygulanmasında
Knime	Birliktelik analizi işlemleri
R	Kelimeler ve sınıflar arasındaki ilişkileri analiz etmede kullanılmıştır.

Bu kadar fazla araç kullanılmasının en büyük nedeni Türkçe metinleri analiz etmek için doğal dil işlemenin ve metin madenciliği yöntemlerinin birlikte uygulandığı bir yazılımın olmamasıdır. Metin madenciliği işlemlerinde kullanılan çoğu yazılım İngilizce metinler için kullanılmaktadır.

10.3. Yöntem

Çalışma kullanılan yöntemler ve bu yöntemlerin kullanım amaçları Çizelge 10.2’de verilmiştir. Çizelgenin ilk sütunu yöntemlerin kullanım amaçlarını, ikinci sütun kullanılan yöntemleri ve üçüncü sütun ise yöntemler hakkında genel bilgilerin hangi bölümde anlatıldığını göstermektedir. Metin madenciliğiyle ilgili yapılan çalışmalarda sıklıkla bu yöntemler kullanıldığı için bu çalışma için de bu yöntemler kullanılmıştır.

Çizelge 10.2. Çalışmada kullanılan yöntemler ve kullanım amaçları

Amaç	Yöntem	Bölüm
Kelime köklerinin bulunması	Sözlük tabanlı kök bulma yöntemi (Zemberek Kütüphanesi)	Bölüm 3.2.3
Metinleri temsil eden terimlerin ağırlıklandırılması	İkili ağırlıklandırma, terim frekansı (TF), TFxIDF	Bölüm 3.2.5
Gönderilerin sınıflandırılması	Naive Bayes, SMO, KNN ve karar ağaçları	Bölüm 4.1
Sınıflandırma algoritmalarının performans değerlerinin ölçülmesi	F ölçeği ve doğruluk oranı	Bölüm 4.2
İkili ve üzeri kelime kullanım birlikteliklerinin belirlenmesi	Apriori algoritması	Bölüm 6
Kategorileri temsil eden kelimeler ve yıllar arasındaki ilişkinin iki boyutlu uzayda gösterilmesi	Uyum analizi	Bölüm 7
Kategorileri temsil eden kelimelerin yıllara göre farklılık gösterip göstermediğinin test edilmesi	Friedman testi	Bölüm 7

10.4. Verilerin Toplanması

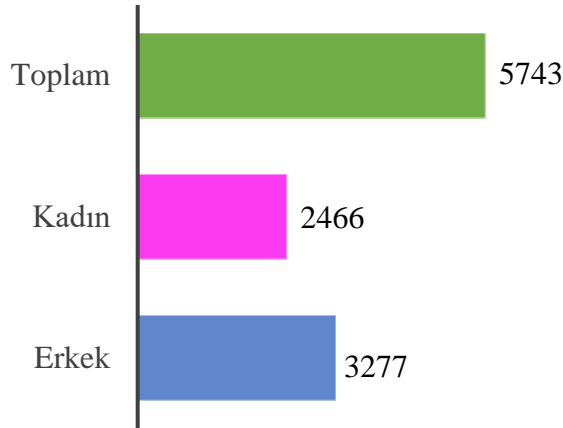
Veriler, Facebook'ta 'Ankilozan Spondilit Hasta Derneği' adıyla yer alan grupta paylaşılan gönderiler ve bu gönderilere yapılan yorumlardan oluşmaktadır. 01.01.2012 ve 30.09.2018 tarihleri arasında kullanıcılar tarafından paylaşılan gönderiler ve yorumlar Romatoloji Uzman doktoruna danışılarak belli kategori başlığı altında Excel'de toplanmıştır.

Kategorilere ait toplam veri sayısı ve kategoriler ile ilgili açıklamalar Çizelge 10.3'te verilmiştir. En yüksek veri adeti şikayet sınıfına aittir. Gönderiler ve yorumlar olmak üzere toplam 153747 adet veri toplanmıştır.

Çizelge 10.3. Kategoriler ve toplam veri sayısı

Kategoriler	Açıklama	Gönderi	Yorum	Toplam	
1	Şikayet	AS hakkında yapılan şikayetleri ve. AS hastalarında görülen belirtileri kapsayan gönderileri ifade eder.	1476	29648	31124
2	Rapor	Hastaların engelli raporu ile ilgili yapmış oldukları paylaşımlar, sosyal imkanlar, ehliyet, askerlik ve iş ile ilgili yapılan gönderileri ifade eder.	556	10775	11331
3	İlaç	Hastaların tedavi süresince kullandıkları ilaçları ve bu ilaç hakkında yapılan yorumları ifade eder.	1338	23379	24717
4	Tedavi	Genel cerrahi ve alternatif tedavi yöntemlerine ait gönderileri temsil eder.	707	13569	14276
5	Doktor	Hastaların doktor ve hastane tavsiyesi almak için yaptıkları gönderileri içerir.	692	11084	11776
6	Spor	Hastaların yaptıkları spor faaliyetlerine ve yapılan etkinliklere ait gönderilerden oluşur.	855	14412	15267
7	Tahlil	Hastaların tahlil ve tahlil sonuçlarıyla ilgili yaptıkları gönderilerden oluşur.	334	5282	5616
8	Kişisel deneyim ve diyet	Hastaların beslenme, hastalıkla ilgili tecrübe ve deneyimlerini kapsayan gönderileri temsil eder.	395	9770	10165
9	Diğer hastalıklar	Hastaların AS'nin dışındaki diğer hastalıklarıyla ilgili yapmış olduğu gönderileri içerir.	406	7125	7531
10	Sınıflandırılmayan	Herhangi bir sınıfa atanması yapılmayan AS ile ilgisi olmayan verileri temsil eder.	1952	19992	21944
Toplam			8711	145036	153747

Veri kümesi toplam 5743 farklı kullanıcı tarafından paylaşılan gönderi ve yorumlardan oluşmaktadır. Verilere ait kullanıcı sayısı Şekil 10.1’de gösterilmiştir.



Şekil 10.1. Verilere ait toplam kullanıcı sayıları

Çizelge 10.3 ve Çizelge 10.4 göre şikayet kategorisini temsil eden toplanan 1476 tane veri, 452’si kadın ve 400’ü erkek olmak üzere toplam 852 farklı kullanıcı tarafından paylaşılmıştır. Aynı şekilde 1476 gönderi için yapılan toplam 29648 yorum, 1814’ü erkek ve 1487’si kadın olmak üzere toplam 3301 farklı kullanıcı tarafından yapılmıştır.

Çizelge 10.4. Cinsiyete göre gönderi ve yorum verilerine ait kullanıcı sayıları

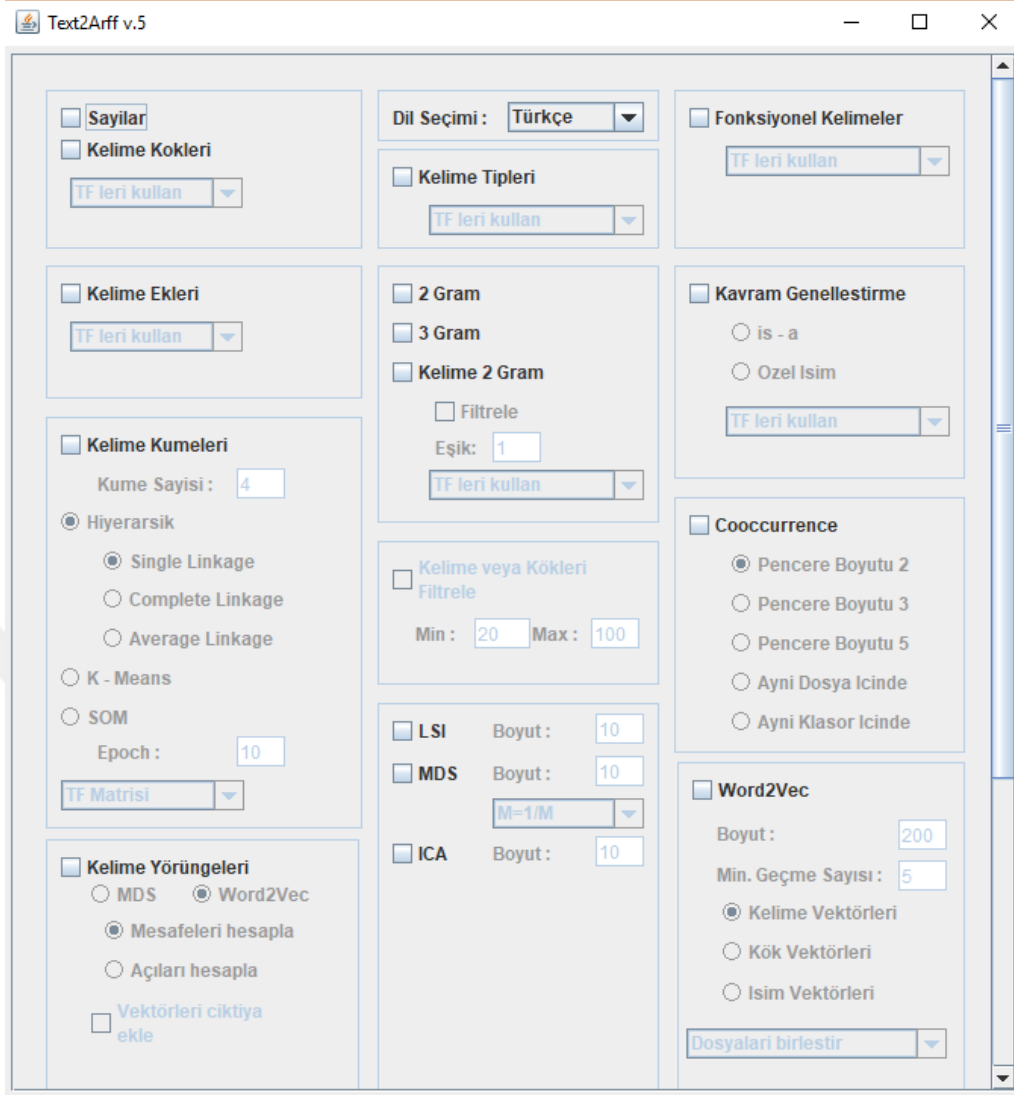
Kategoriler	Gönderi			Yorum		
	Erkek	Kadın	Toplam	Erkek	Kadın	Toplam
Şikayet	452	400	852	1814	1487	3301
Rapor	276	91	367	1123	480	1603
İlaç	463	291	754	1627	1153	2780
Tedavi	319	201	520	1305	1040	2345
Doktor	259	269	528	1069	999	2068
Spor	230	167	397	1088	771	1859
Tahlil	141	113	254	593	517	1110
Kişisel deneyim ve diyet	172	132	304	952	827	1779
Diğer hastalıklar	155	159	314	785	781	1566
Sınıflandırılmayan	377	243	620	1225	989	2214

10.5. Veri Ön İşlem Aşaması

Toplanan veriler yapısal olmayan metinsel verilerdir. Bu verilerin analiz edilebilmesi için yapısal hale dönüştürülmesi gerekir. Veri ön işlem aşamaları YTÜ Kemik Doğal Dil İşleme grubu tarafından geliştirilen Text2ArffV5 yazılımı kullanılarak yapılmıştır. Text2ArffV5 yazılımına ait arayüz Şekil 10.2’de görüldüğü gibidir.

Text2Arff yazılımı kullanılarak kelimelerin kökleri bulma, kelime ağırlıklandırma işlemleri yapma, kelime n-gramları oluşturma, fonksiyonel kelimeler ile istenmeyen kelimeler listeden çıkarma, kelime kümeleme ve kelime filtreleme işlemleri yapılabilmektedir (Can ve Amasyalı, 2016).

Yazılım kullanılarak her bir kategoriye ait metinlerin kelime kökleri belirlenmiştir. Kelime ağırlıklandırma yöntemi olarak ise TF, TFxIDF ve binary ağırlıklandırma yöntemleri uygulanmıştır. Yazılım çıktı sonucunu Weka’ya arff dosya formatında vermektedir. Bulunan kelime köklerinin boyutunu azaltmak amacıyla bir anlam ifade etmeyen durak kelimeler ve gereksiz kelimeler çıkarılmıştır. Boyut azaltma işlemleri Excel’de makro geliştirilerek yapılmıştır.



Şekil 10.2. Text2ArffV5 yazılımı arayüzü

10.6. Sınıflandırma Sonuçları

Yapılan sınıflandırma işlemlerinde ne kadar iyi sınıflandırılıp sınıflandırılmadığını kontrol etmek amacıyla K-en yakın komşu, SMO, karar ağaçları ve Naive Bayes olmak üzere 4 yöntem kullanılmıştır. Sınıf isimleri ve işlemde kullanılan veri sayısı Çizelge 10.5'te verilmiştir. Sınıflandırma işlemlerinde kategorilere ait gönderi verileri kullanılmıştır.

Çizelge 10.5. Sınıflandırma işlemlerinde kullanılan kategori isimleri ve veri sayıları

Sınıflar	Eğitim veri sayısı	Metin ön işlem sonucunda bulunan kelime sayısı
Deneyim	300	
Diğer hastalıklar	300	
Doktor	300	
İlaç	300	
Rapor	300	1402
Spor	300	
Tahlil	300	
Tedavi	300	
Şikayet	300	

Eğitim verisi olarak her kategoriye ait 300 eğitim verisi olmak üzere toplam 2700 gönderi verisi kullanılmıştır. Metin ağırlıklandırma yöntemi olarak terim frekansı, TFxIDF ve binary ağırlıklandırma yöntemleri kullanılmıştır. Ön işlem sonucunda toplam 1.402 tane kelime kökü elde edilmiştir.

Sınıflandırma işlemlerinde Weka yazılımı kullanılmıştır. Test Sınıflandırma yöntemlerinin performans oranlarının belirlenmesinde K katmanlı çapraz doğrulama (K fold cross validation) yöntemi kullanılmıştır.

Sınıflandırma algoritmaları olarak K-en yakın komşu, SMO, karar ağaçları ve Naive Bayes kullanılmıştır. Üç farklı ağırlıklandırma yöntemine göre bu algoritmaların başarı oranları hesaplanmıştır. Ağırlıklandırma yöntemlerine göre sınıflandırma yöntemlerinin başarı oranları Çizelge 10.6'da verilmiştir. Kullanılan sınıflandırma algoritmalarından ortalama olarak en başarılı sonuç SMO yönteminin sağladığı görülmektedir. K-en yakın komşu yönteminde k değeri artıkça başarının düştüğü gözlemlenmiştir. Farklı ağırlıklandırma yöntemlerine göre yapılan sınıflandırma işlemlerinde de en başarılı sonuç %53.33 doğruluk oranı ile binary ağırlıklandırma yönteminde SMO algoritmasının sağladığı gözlenmiştir.

Çizelge 10.6. Ağırlıklandırma yöntemlerine göre sınıflandırma yöntemlerinin başarı oranları

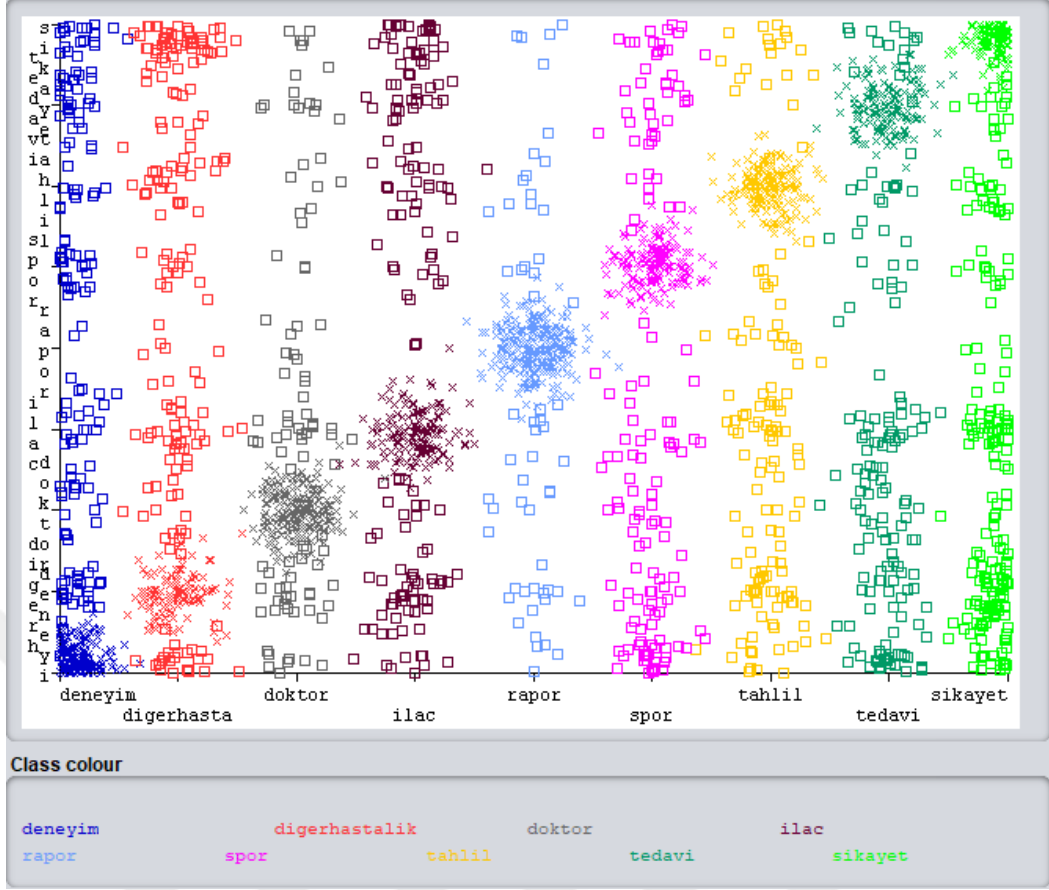
	TF		TFxIDF		Binary		
	F-ölçüsü	Doğruluk oranı (%)	F-ölçüsü	Doğruluk oranı (%)	F-ölçüsü	Doğruluk oranı (%)	
k=1	0.30	30.96	0.30	31.00	0.33	34.22	
K-en yakın komşu	k=3	0.24	25.04	0.24	25.19	0.27	27.63
	k=5	0.22	23.85	0.22	23.96	0.24	25.00
SMO	0.52	51.15	0.52	51.15	0.54	53.33	
Karar ağaçları-J48	0.46	46.78	0.46	46.78	0.47	46.96	
Naive Bayes	0.41	42.33	0.41	42.07	0.50	49.81	

SMO yöntemi kullanılarak yapılan sınıflandırma sonucunda elde edilen karışıklık matrisi Çizelge 10.7’de verilmiştir. Çizelgeye göre deneyim kategorisini temsil eden 300 gönderinin 151’i doğru sınıflandırılmışken geri kalan gönderiler ise farklı kategorilerde yanlış olarak sınıflandırılmıştır.

Çizelge 10.7. SMO yöntemi sonucunda elde edilen karışıklık matrisi (Confusion Matrix)

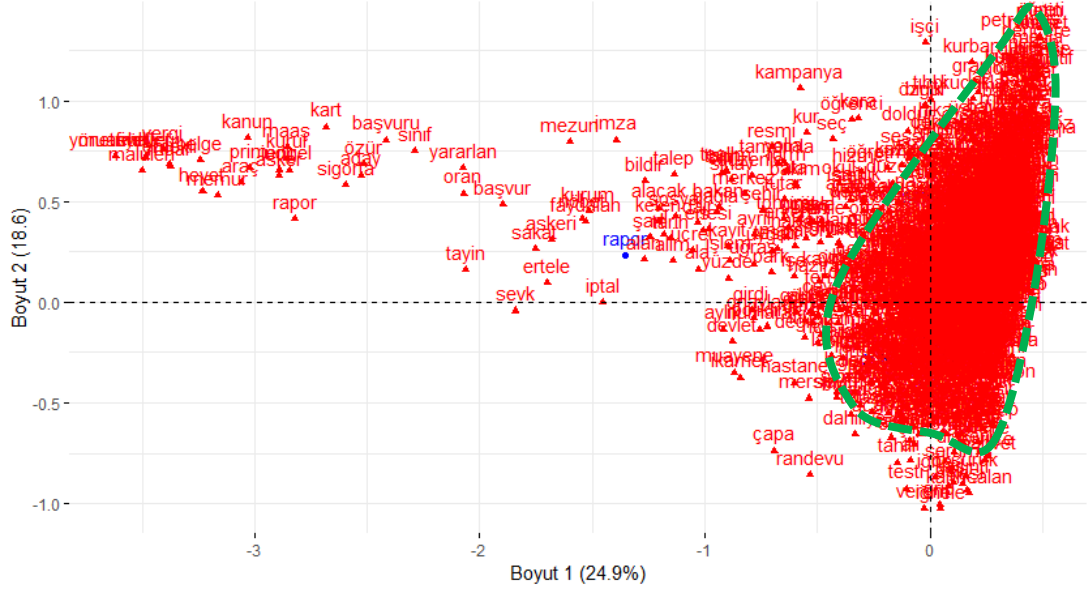
n=2700		Tahmin edilen sınıf								
		Deneyim	Diğer hastalık	Doktor	İlaç	Rapor	Spor	Tahlil	Tedavi	Şikayet
Doğru sınıf	Deneyim	151	27	15	20	5	23	11	27	21
	Diğer hastalık	25	118	19	29	7	13	30	13	46
	Doktor	9	27	204	24	9	4	7	11	5
	İlaç	10	37	15	150	2	12	23	24	27
	Rapor	5	12	8	10	243	7	9	1	5
	Spor	29	28	14	20	6	156	10	22	15
	Tahlil	10	37	12	32	14	8	167	8	12
	Tedavi	33	22	25	33	2	14	12	144	15
	Şikayet	25	61	10	43	4	14	20	16	107

Şekil 10.3’te 9 farklı sınıf, dokuz farklı renk ile gösterilmektedir. Grafikte ‘x’ ile gösterilenler doğru sınıflandırılmış dokümanları, kare ile gösterilenler ise yanlış sınıflandırılmış dokümanları göstermektedir. Örneğin deneyim kategorisinde bulunan mavi bir kare bu dokümanın deneyim sınıfında olduğunu ancak yanlış şekilde sınıflandırılarak başka bir sınıfa dahil edildiğini göstermektedir.



Şekil 10.3. SMO algoritması sonucu elde edilen hataların grafiksel gösterimi

Şekil 10.4'te sınıfları temsil eden 1402 kelime ve sınıflar arasındaki ilişkiyi iki boyutlu olarak göstermektedir. Boyut 1 değişimin %24.9'unu açıklarken boyut 2 ise %18.6'sını açıklamaktadır. Grafikte kırmızı yazılanlar kullanılan kelimeleri, mavi ile yazılı olanlar ise sınıfları temsil etmektedir. Yeşil çizgi arasında kalan alan incelendiğinde sınıflar arası kullanılan ortak kelimelerin çok olduğunu göstermektedir. Sadece rapor kategorisini temsil eden kelimelerin ayırt ediciliği daha fazladır.



Şekil 10.4. Sınıflara ait uyum analizi grafiği

10.7. Şikayet Kategorisine ait Sonuçlar

10.7.1. Şikayet gönderi sonuçları

Gönderi verilerine veri ön işlem uygulanmasından sonra gönderileri temsil eden toplam 2504 adet kelime kökü bulunmuştur ve terim frekansı ağırlıklandırma yöntemi ile ağırlıklandırılmıştır. 2'den aşağı sıklıkta geçen kelimeler ve gereksiz kelimeler elenmiştir. Kategoriye temsil eden toplam 214 tane anahtar kelime belirlenmiştir. Hastalar tarafından yıllara göre en çok kullanılan kelimeler Çizelge 10.8'de verilmiştir.

Çizelge 10.8. Şikayet kategorisine ait gönderilerde yıllara göre kullanıcılar tarafından en çok kullandıkları 50 kelime

Yıl	Kelimeler
2012	ağrı, sırt, boyun, sol, önce, taraf, uyan, çek, sabah, sağ, şiş, baş, kalça, gece, yürü, bacak, bel, hareket, uyu, yan, kes, otur, atak, dur, göğüs, gör, kan, yat, fena, hamile, kemik, parmak, ağız, ayak, diz, gergin, göz, hava, nefes, topuk, uyuş, acı, aft, böbrek, bölge, dön, işaret, kaburga, kalk, kilo

Çizelge 10.8. (devam)

Yıl	Kelimeler
2013	ağrı, kalça, sağ, bel, sol, çek, boyun, göz, gece, sabah, ayak, hareket, gör, önce, kes, kemik, baş, göğüs, kalk, sırt, şiş, uyu, acı, romatizma, bacak, ağız, aft, kol, taraf, yan, kan, böbrek, uyku, yürü, otur, dur, şiddet, ateş, diş, el, eklem, yat, fena, nefes, omurga, vücut, kalp, kas, atak, diz
2014	ağrı, boyun, gece, sırt, bacak, ayak, yan, kalça, bel, baş, sağ, yağmur, çek, sabah, önce, uyan, uyu, yürü, yağ, hareket, kemik, gir, kes, acı, hava, otur, şiddet, nefes, atak, kramp, uyku, eklem, yat, diz, sürekli, dönem, kalk, şiş, taraf, kas, bat, omuz, doğru, mide, nem, arka, ense, sol, göğüs, kol
2015	ağrı, omurga, eklem, hareket, gör, kalça, azal, sabah, şiddet, bel, uyu, iltihap, sırt, ayak, göz, uyku, vücut, yan, baş, sağ, sol, diş, alerji, gece, çek, önce, hava, tutuk, kemik, nefes, şiş, uyan, atak, dönem, yürü, taraf, akciğer, boyun, göğüs, kol, bölge, dur, reaksiyon, bacak, yağmur, kes, acı, arka, sıcak, romatizma
2016	ağrı, ayak, boyun, sırt, sabah, hareket, gece, kalça, bel, sol, önce, kalk, uyku, yürü, bacak, diz, vücut, yan, şiddet, sağ, atak, dur, kan, uyu, çek, hava, şiş, uyan, kol, omuz, dil, mevsim, göğüs, topuk, parmak, kaburga, tutuk, acı, sıcak, gir, sürekli, el, ateş, göz, kemik, yat, bat, yüksek, hamile, eklem
2017	ağrı, ayak, bel, bacak, önce, şiş, sağ, gece, parmak, boyun, çek, kemik, sırt, sabah, kalça, acı, el, kas, sol, yürü, vücut, göz, uyu, kalk, baş, taraf, yan, göğüs, nefes, üstün, yorgun, uyku, atak, dur, otur, gör, hareket, diz, uyan, topuk, kaburga, sürekli, eklem, hava, kol, kes, iltihap, bölge, çocuk, omuz
2018	ağrı, bel, ayak, önce, yan, kemik, acı, baş, sol, boyun, kalça, sırt, yürü, şiş, bacak, sağ, göz, gör, şiddet, sabah, parmak, otur, çek, dur, eklem, vücut, taraf, göğüs, sürekli, kes, azal, hareket, bük, kalk, diş, gece, el, dil, romatizma, kas, diz, uyan, yat, nefes, uyku, topuk, iltihap, bölge, kambur, uyu

Kullanıcılar tarafından kullanılan 214 tane anahtar kelimenin frekans değerleri baz alınarak kelime bulutu oluşturulmuştur. Kelime bulutu gösteriminde kelimeler frekans değerlerine göre boyutlandırılır ve renklendirilir. Frekans değeri büyük olan kelimeler daha büyük ve belirgin renkte görünmektedir. Şekil 10.6 incelendiğinde 'ağrı', 'bel', 'ayak', 'sırt', 'sol', 'boyun' gibi kelimelerin gönderi metnlerinin içerisinde daha çok kullanıldığı görülmektedir.

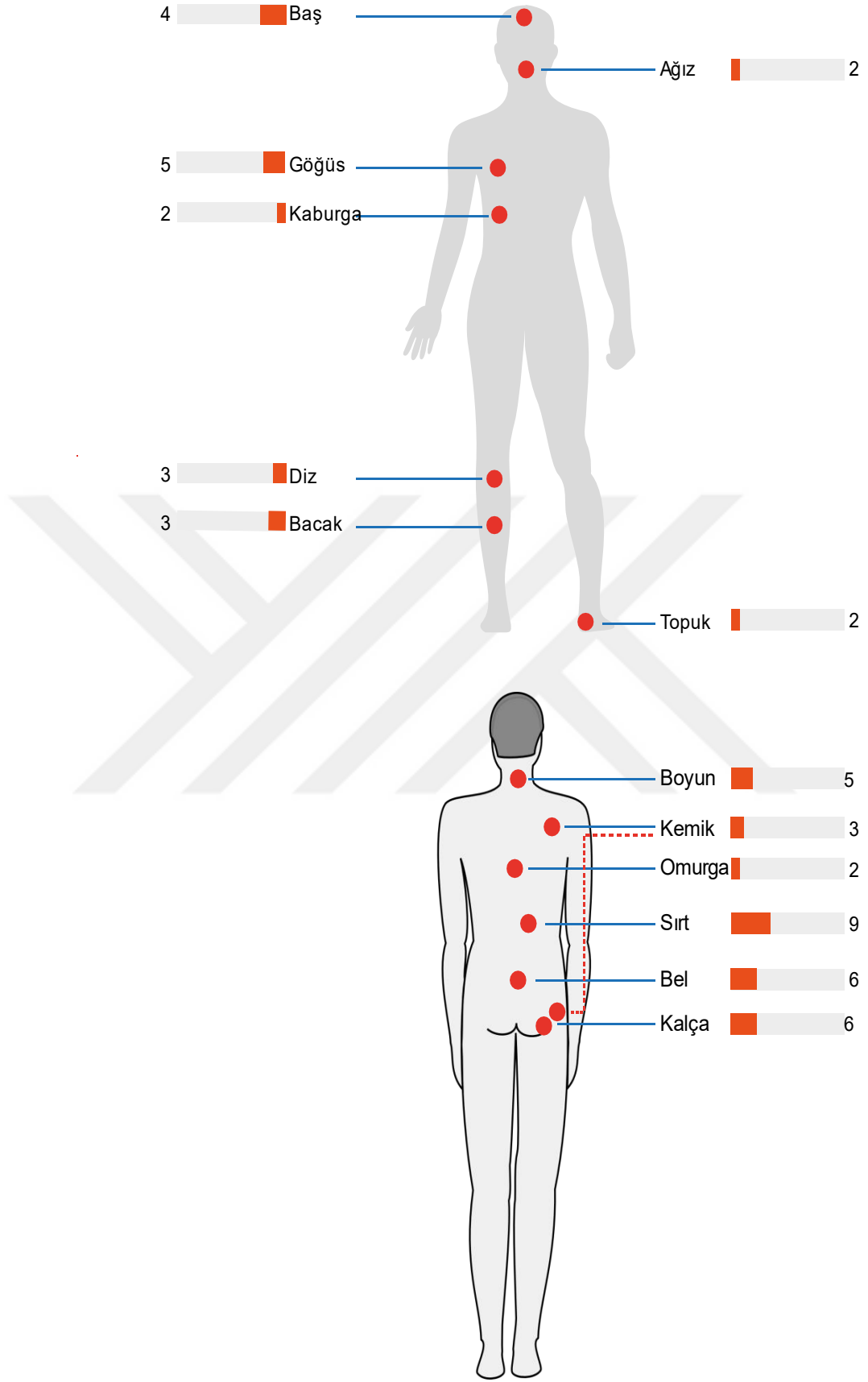


Şekil 10.5. Şikayet kategorisine ait gönderileri temsil eden anahtar kelimelerin kelime bulutu ile gösterimi

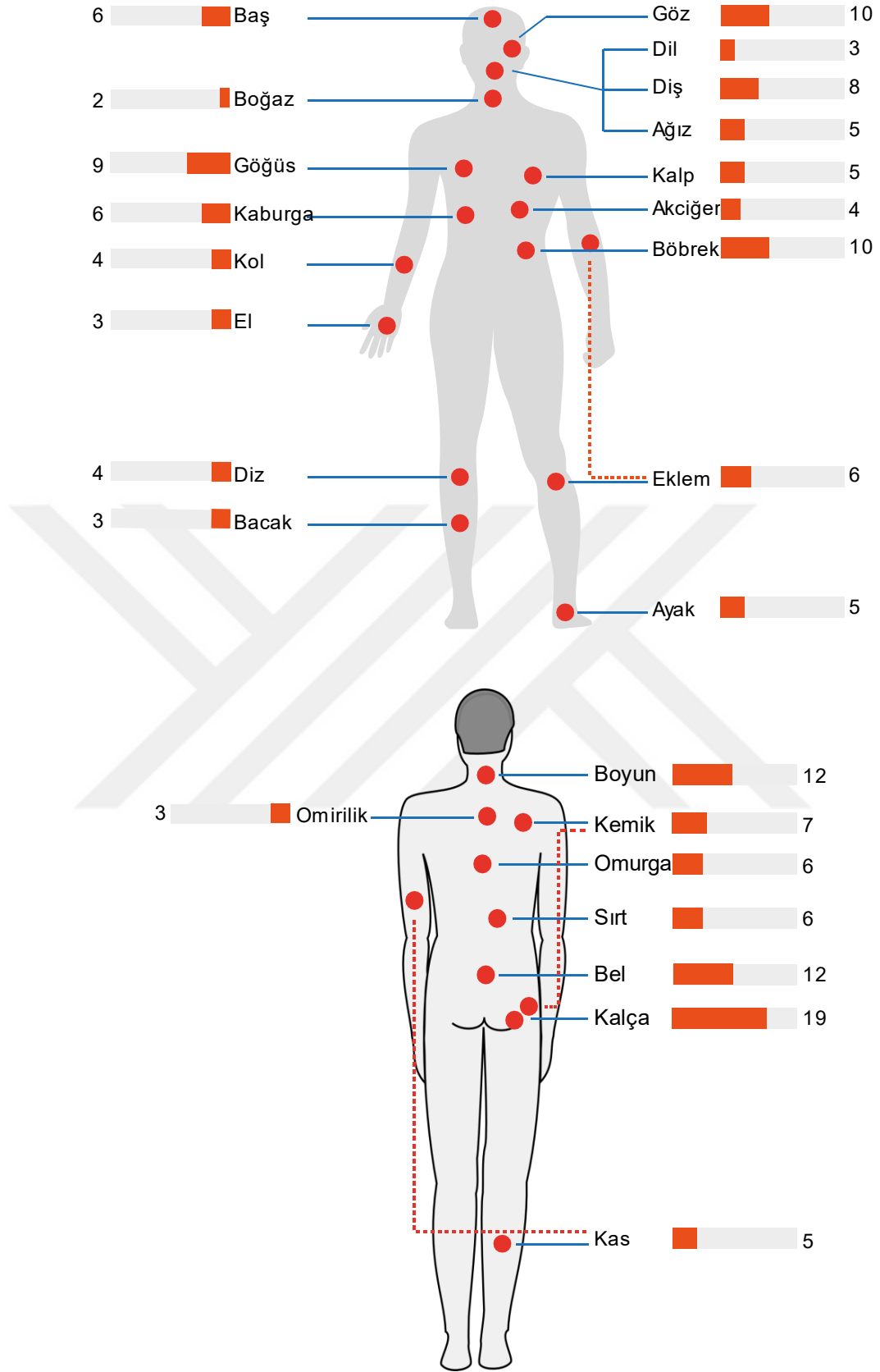
Bulunan anahtar kelimeler arasında hastalara ait toplam 46 şikayet bölgesi belirlenmiştir. Şikayet bölgeleri yıllara göre kadın ve erkek hastalar olmak üzere ayrı ayrı infografik şekillerle gösterilmiştir

Şekil 10.6–Şekil 10.12 arası sırayla 2012–2018 yılları arasında erkek kullanıcılar tarafından en çok konuşulan şikayet bölgelerini göstermektedir. Yuvarlak kırmızı şekiller şikayet bölgelerini, sayılar sıklık değerlerini göstermektedir.

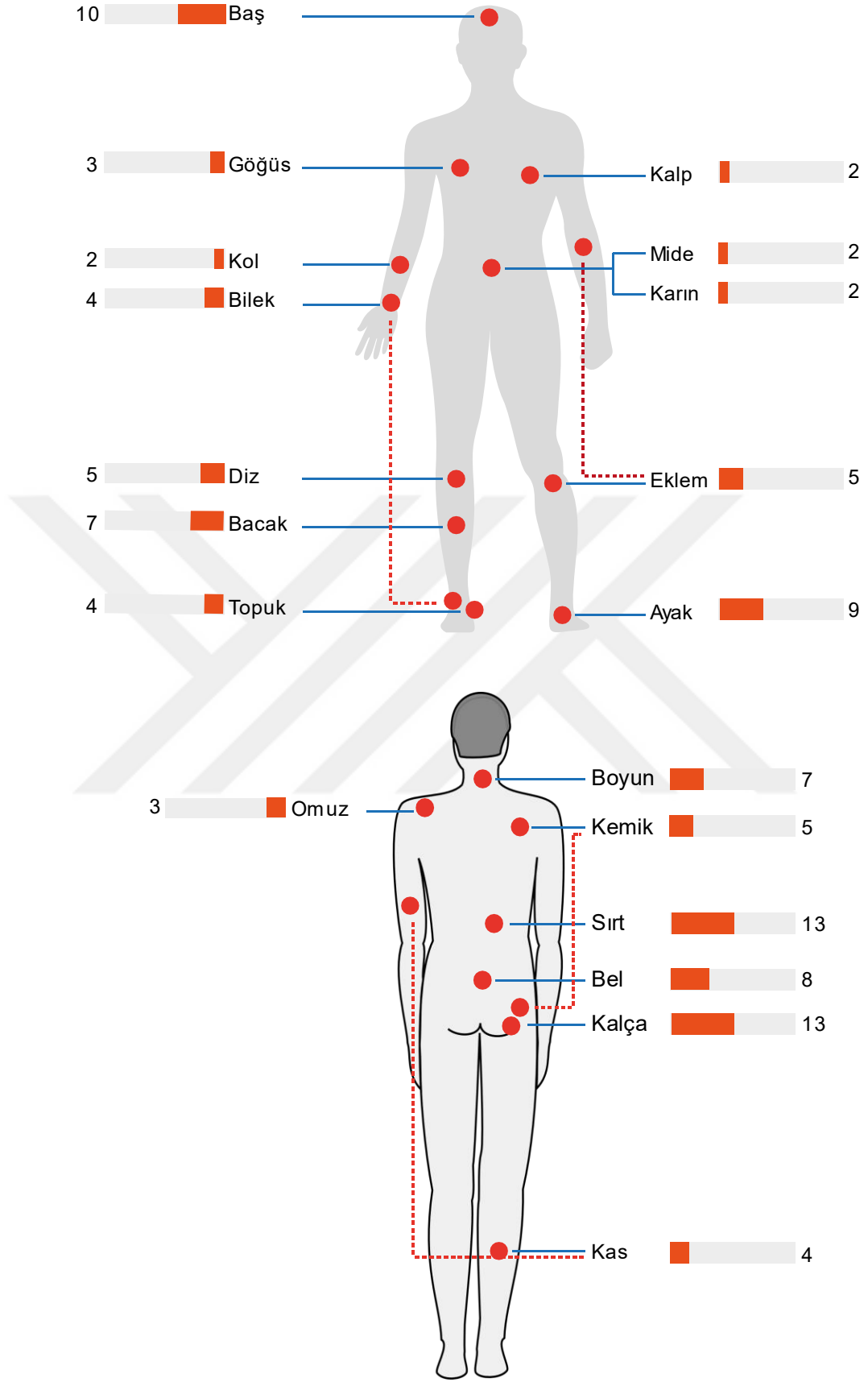
Erkek şikayet bölgeleri incelediğinde 2012 yılında toplam 13, 2013 yılında toplam 25, 2014 yılında toplam 19, 2015 yılında toplam 21, 2016 yılında toplam 23, 2017 yılında toplam 24 ve 2018 yılında toplam 30 bölge belirlenmiştir. En çok şikayet 2012 yılında sırt bölgesinden, 2013 yılında kalça bölgesinden, 2014 yılında sırt ve kalça bölgesinden, 2015 yılında sırt bölgesinden, 2016 yılında boyun bölgesinden, 2017 yılında bacak bölgesinden ve 2018 yılında bel bölgesinden gelmiştir.



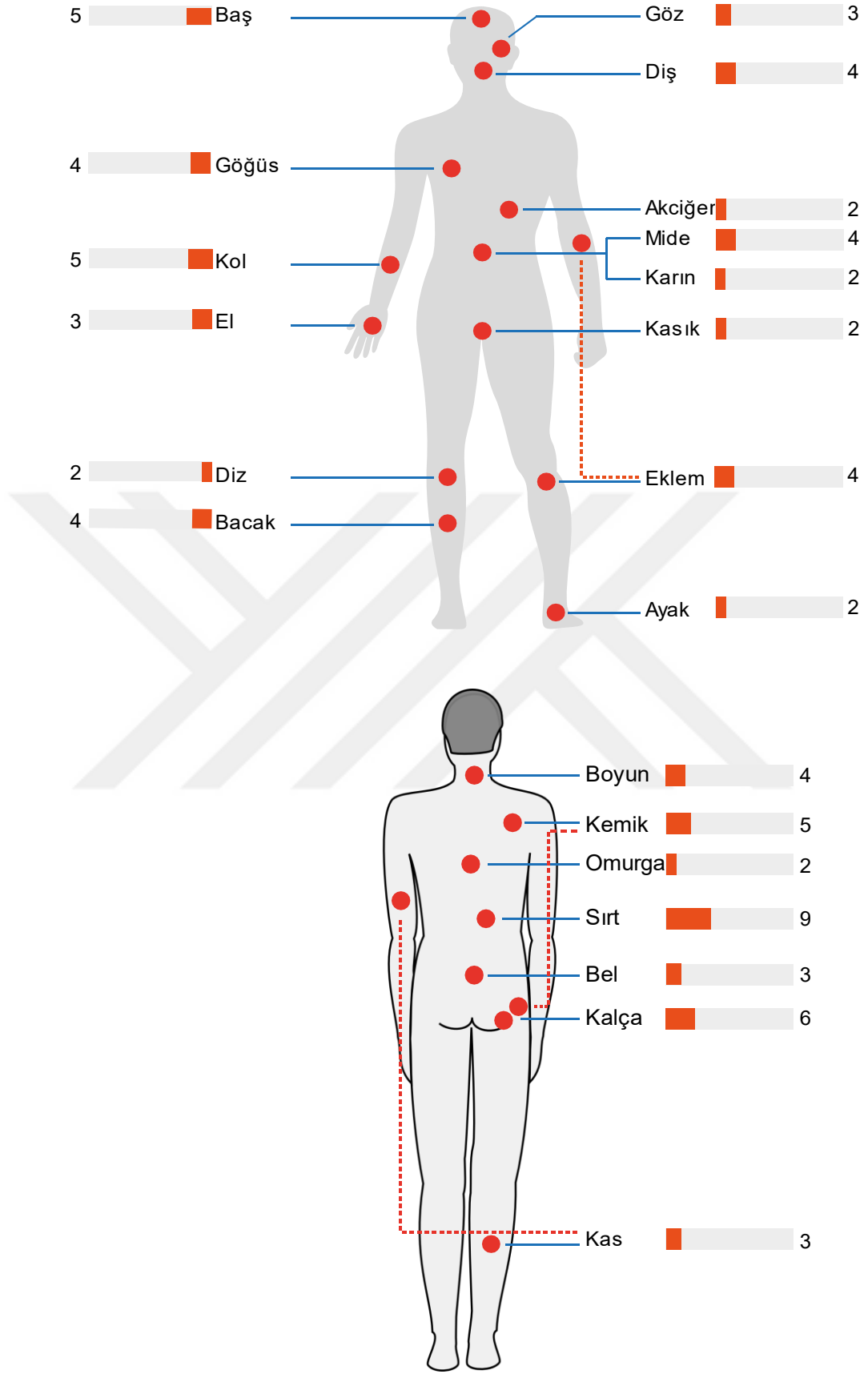
Şekil 10.6. 2012 yılı erkek şikayet bölgeleri



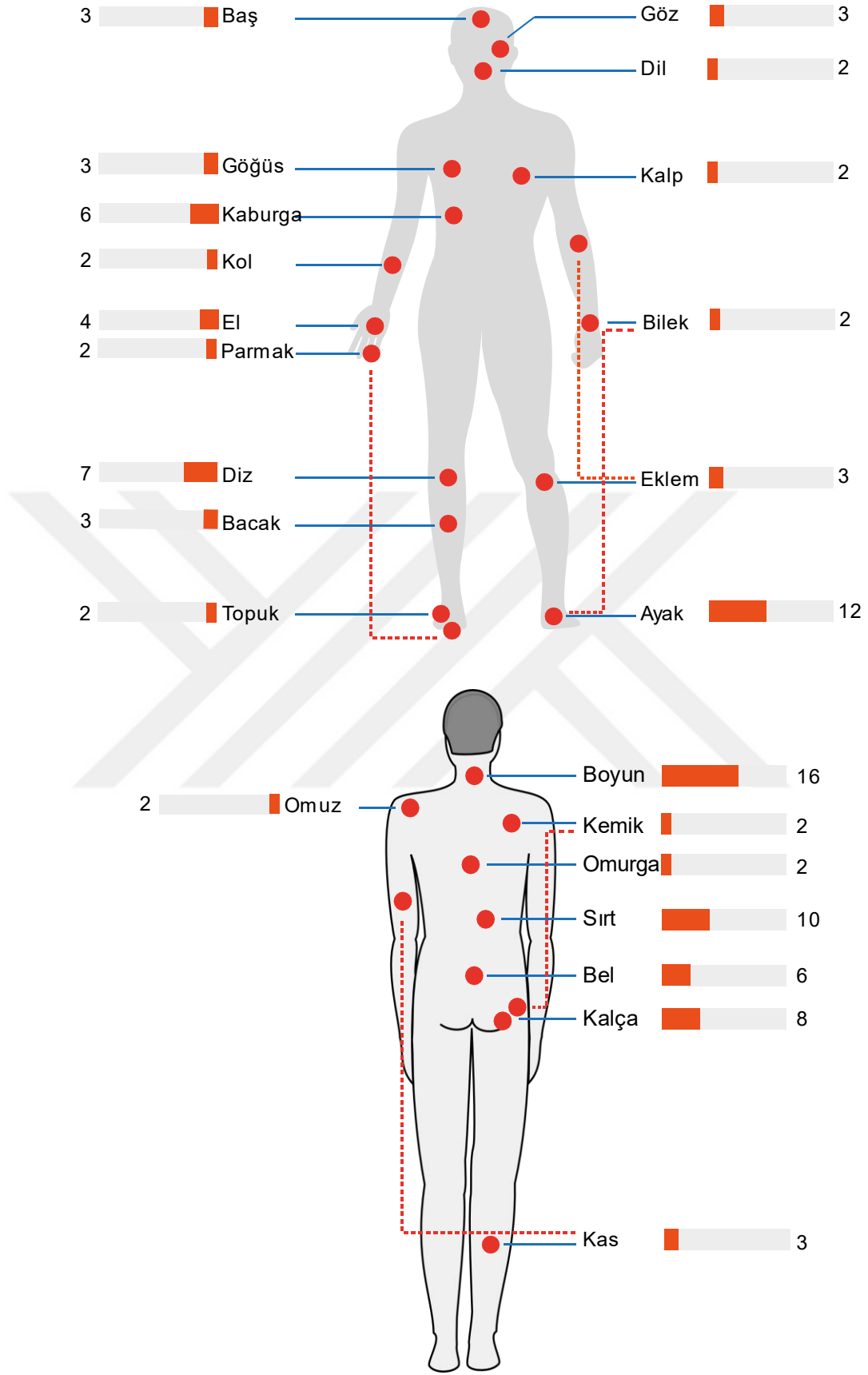
Şekil 10.7. 2013 yılı erkek şikayet bölgeleri



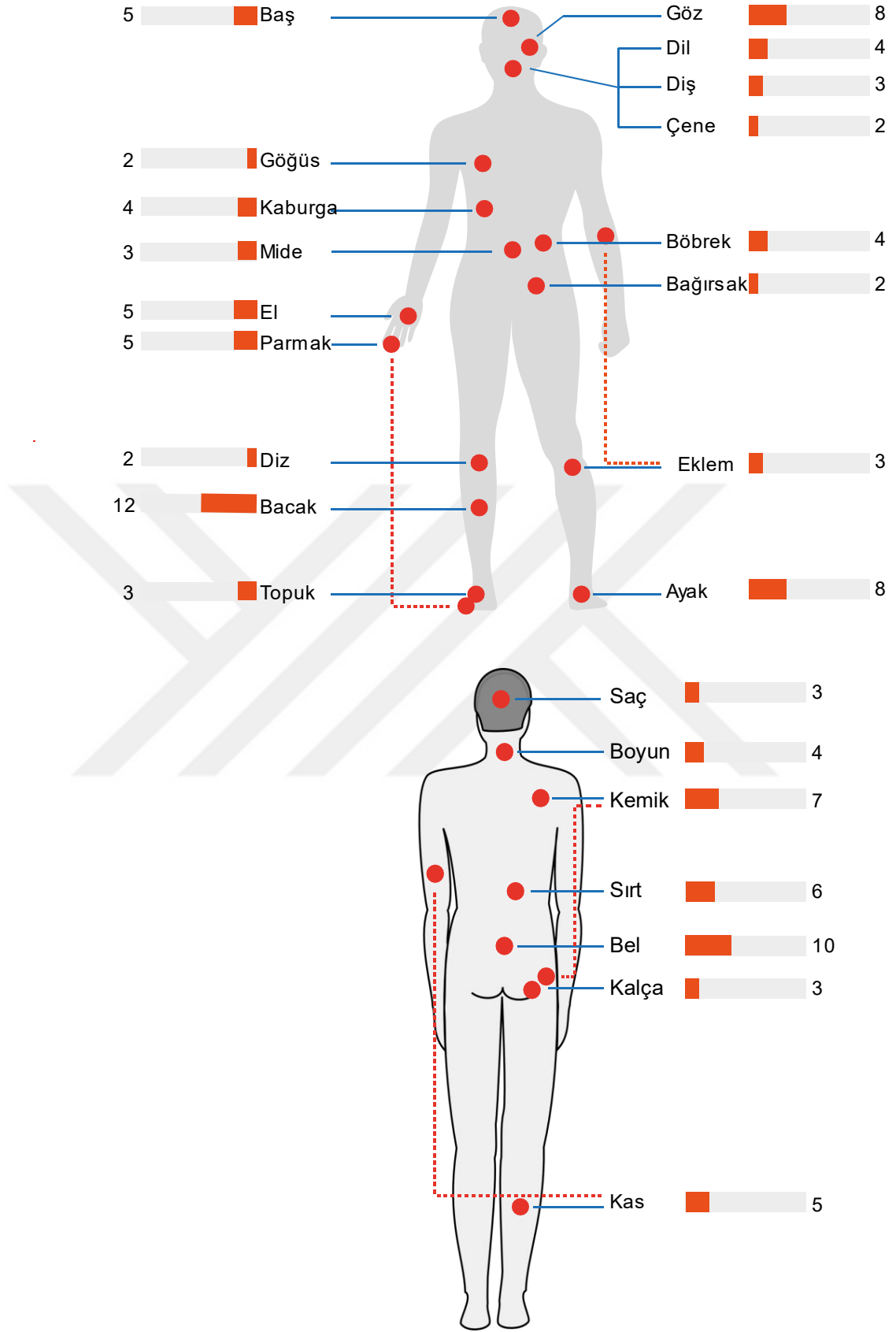
Şekil 10.8. 2014 yılı erkek şikayet bölgeleri



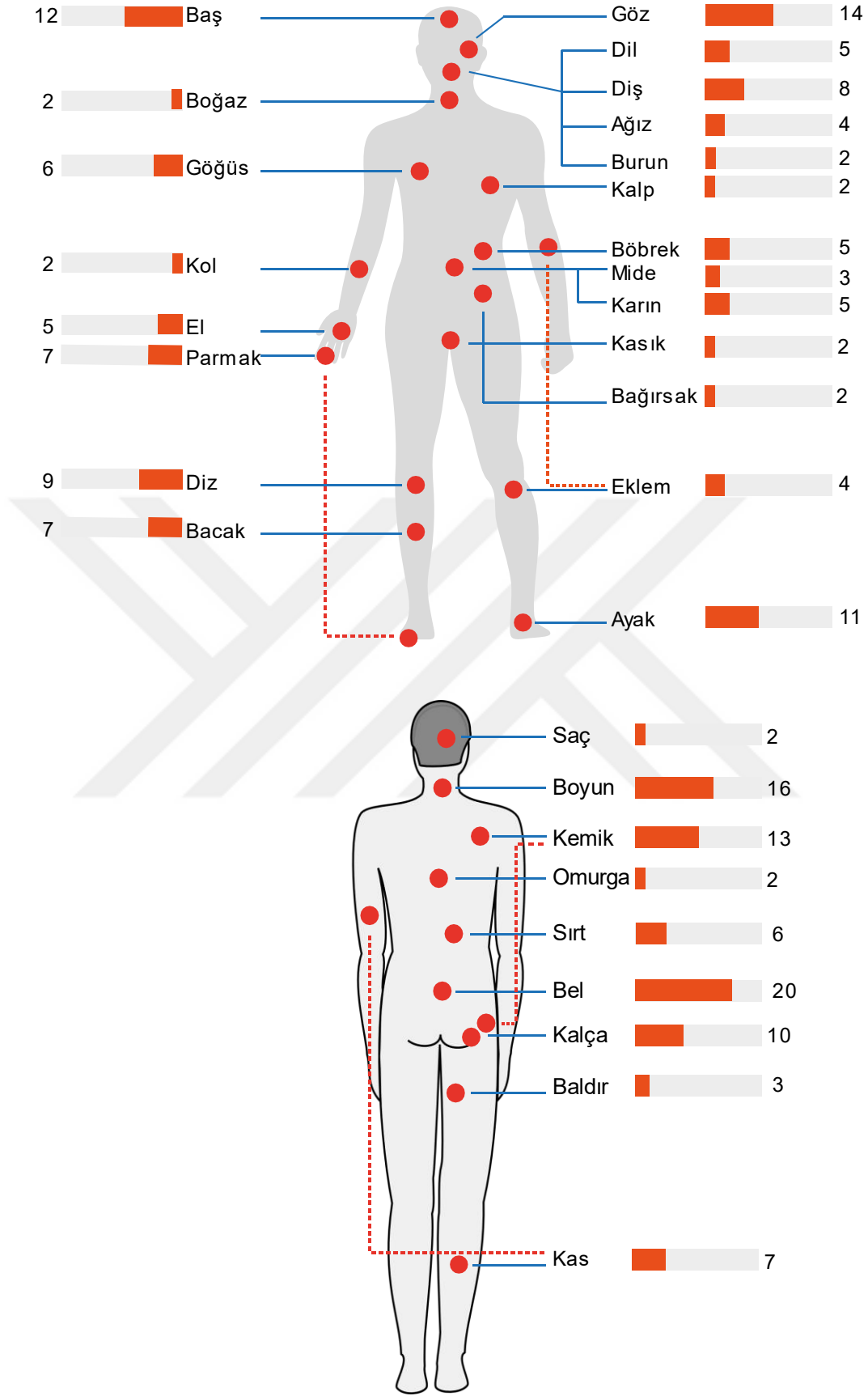
Şekil 10.9. 2015 yılı erkek şikayet bölgeleri



Şekil 10.10. 2016 yılı erkek şikayet bölgeleri



Şekil 10.11. 2017 yılı erkek şikayet bölgeleri

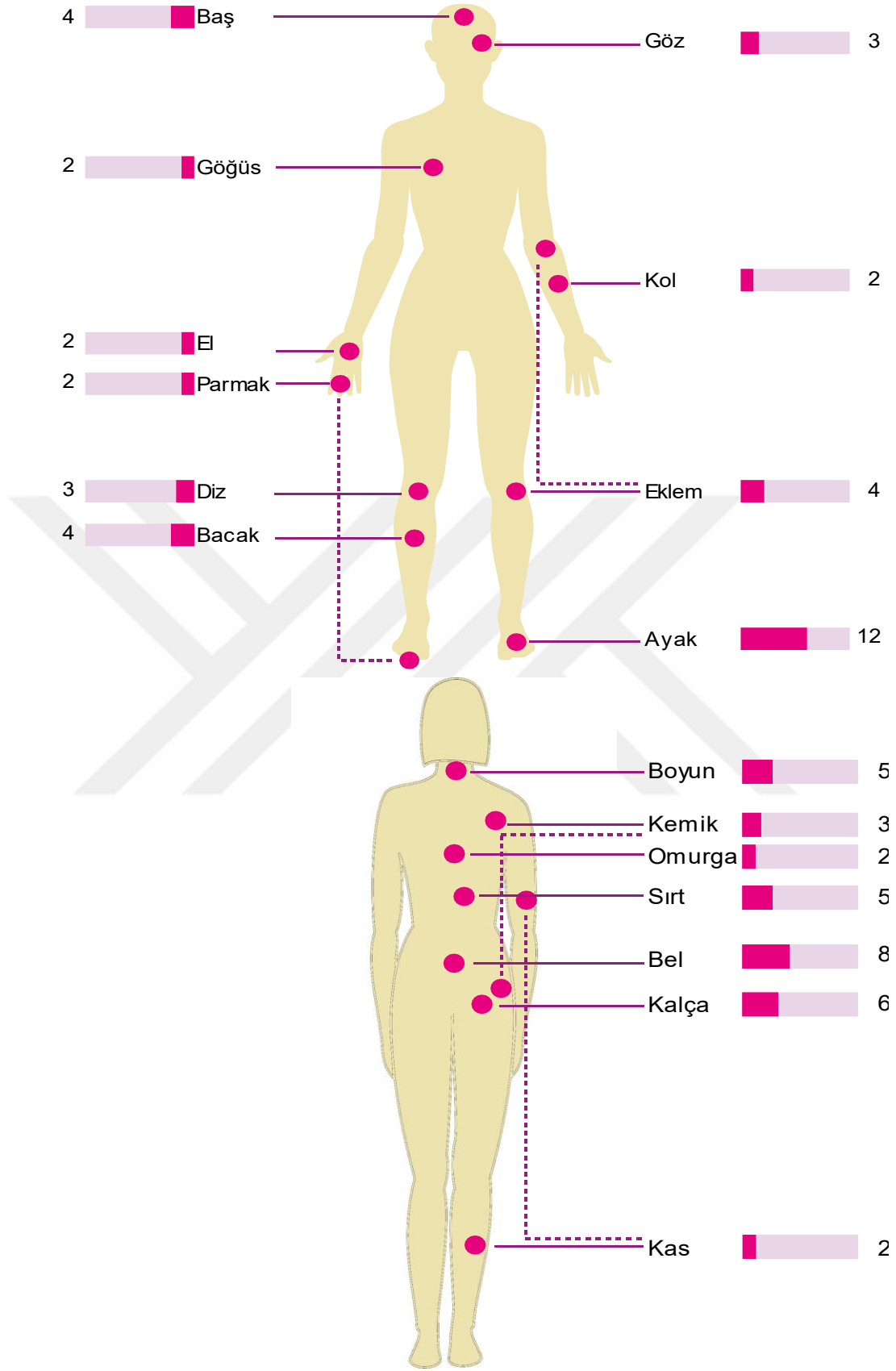


Şekil 10.12. 2018 yılı erkek şikayet bölgeleri

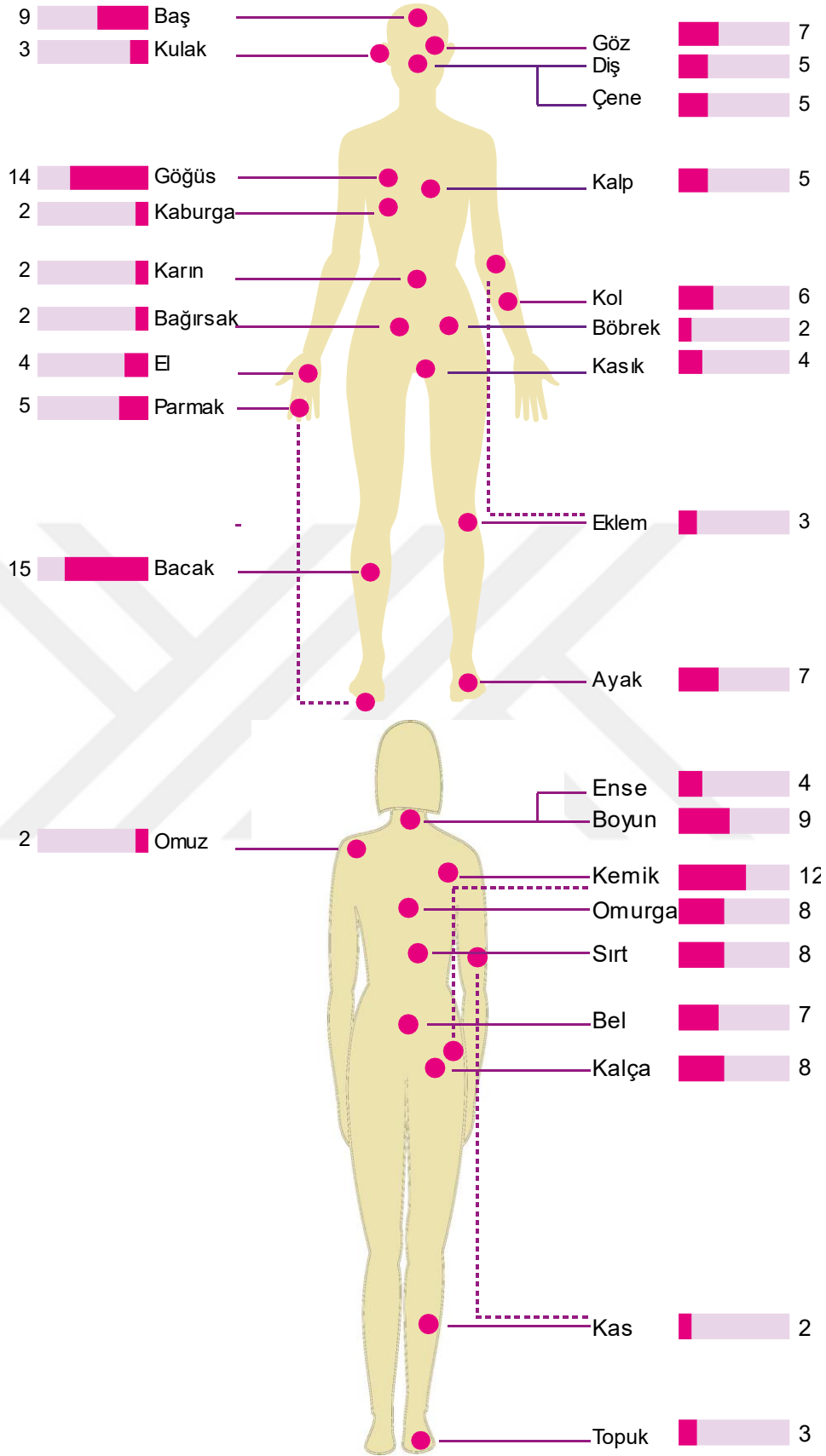
Şekil 10.13–Şekil 10.19 arası sırayla 2012–2018 yılları arasında kadın kullanıcılar tarafından en çok konuşulan şikayet bölgelerini göstermektedir.

Kadın şikayet bölgeleri incelediğinde 2012 yılında toplam 17, 2013 yılında toplam 28, 2014 yılında toplam 25, 2015 yılında toplam 30, 2016 yılında toplam 23, 2017 yılında toplam 30 ve 2018 yılında toplam 30 bölge belirlenmiştir. En çok şikayet 2012 yılında ayak bölgesinden, 2013 yılında bacak bölgesinden, 2014 yılında bel, 2015 yılında bilek bölgelerinden, 2016, 2017 ve 2018 yıllarında kemik bölgelerinden gelmiştir.

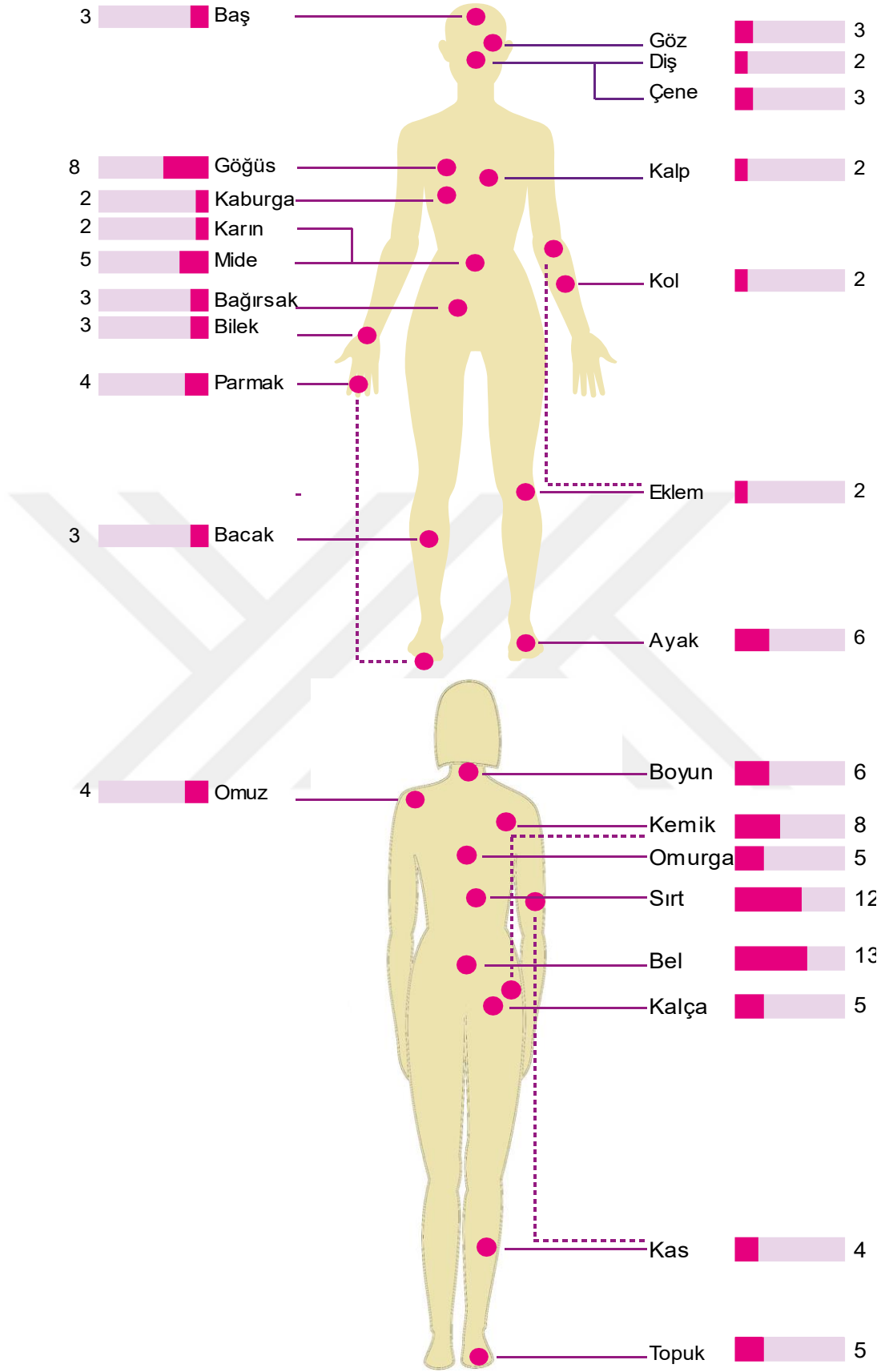




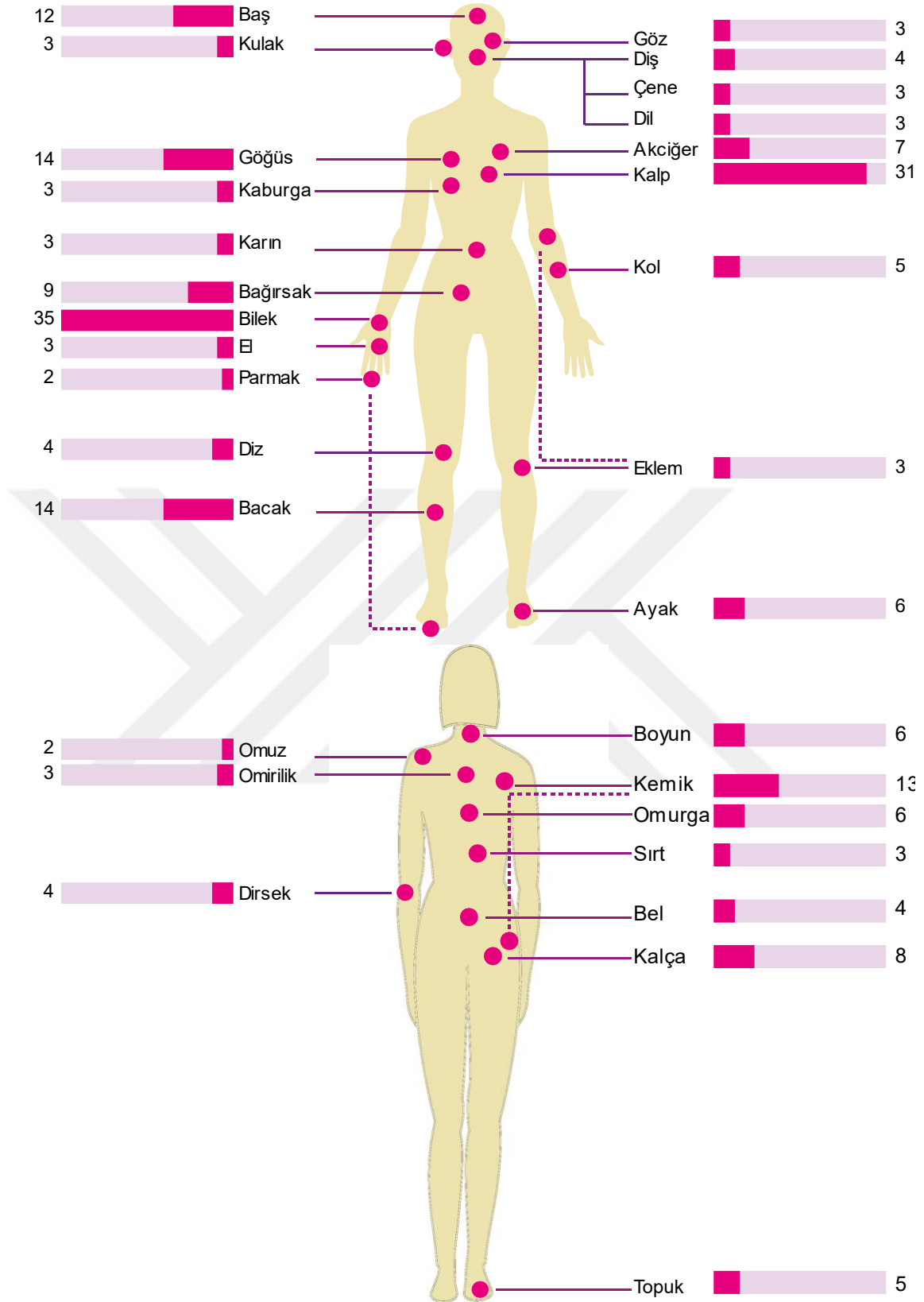
Şekil 10.13. 2012 yılı kadın şikayet bölgeleri



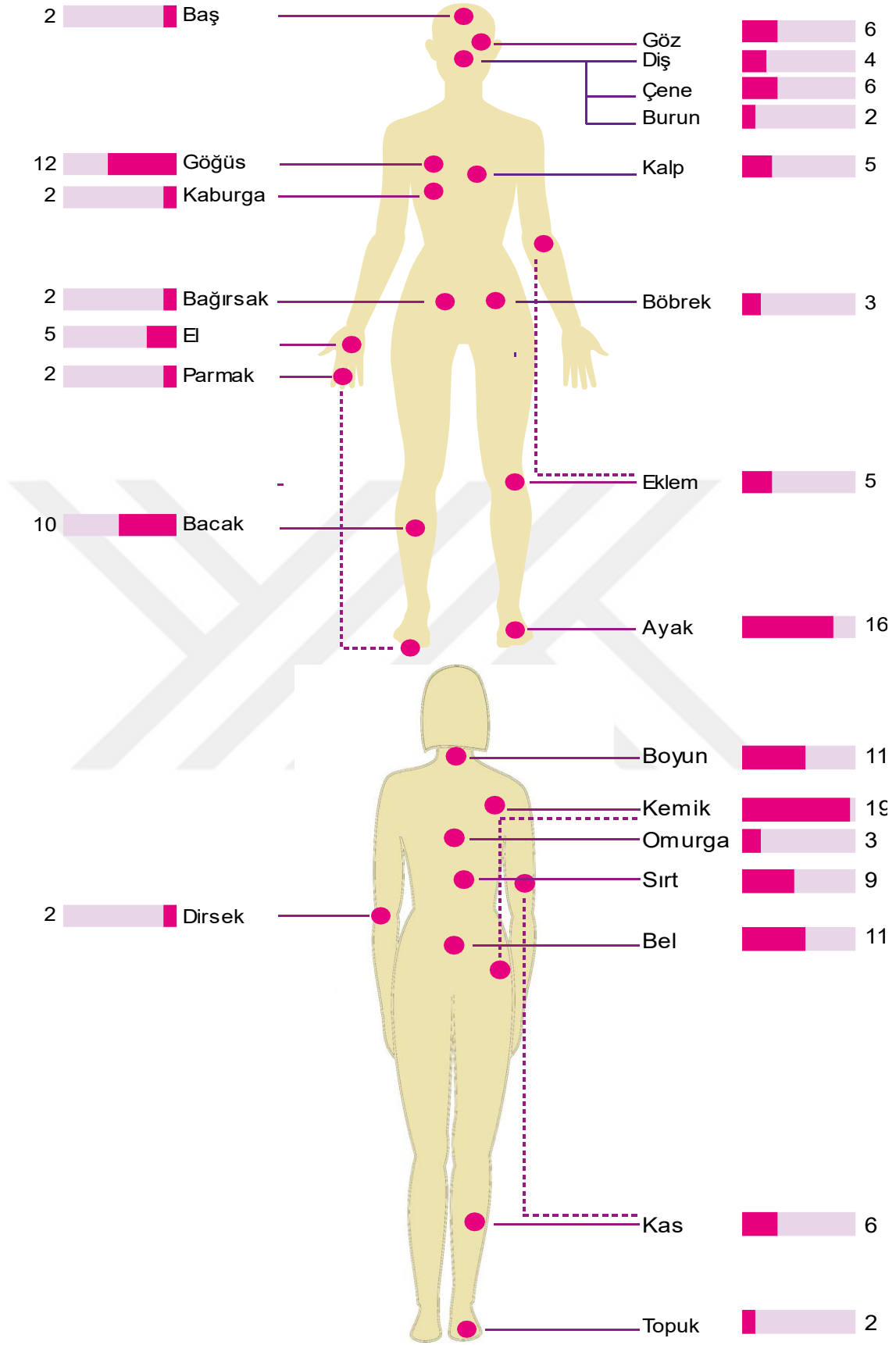
Şekil 10.14. 2013 yılı kadın şikayet bölgeleri



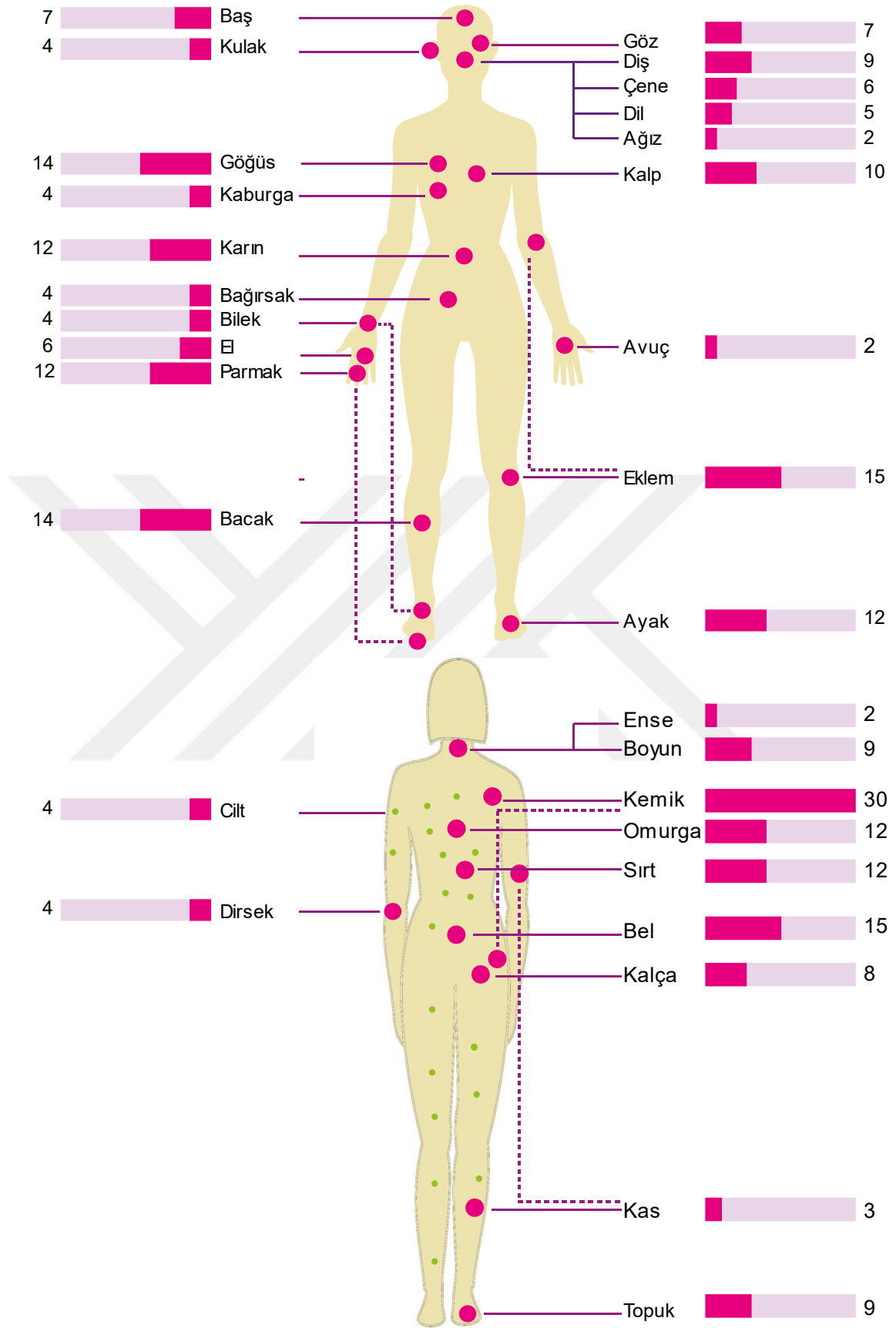
Şekil 10.15. 2014 yılı kadın şikayet bölgeleri



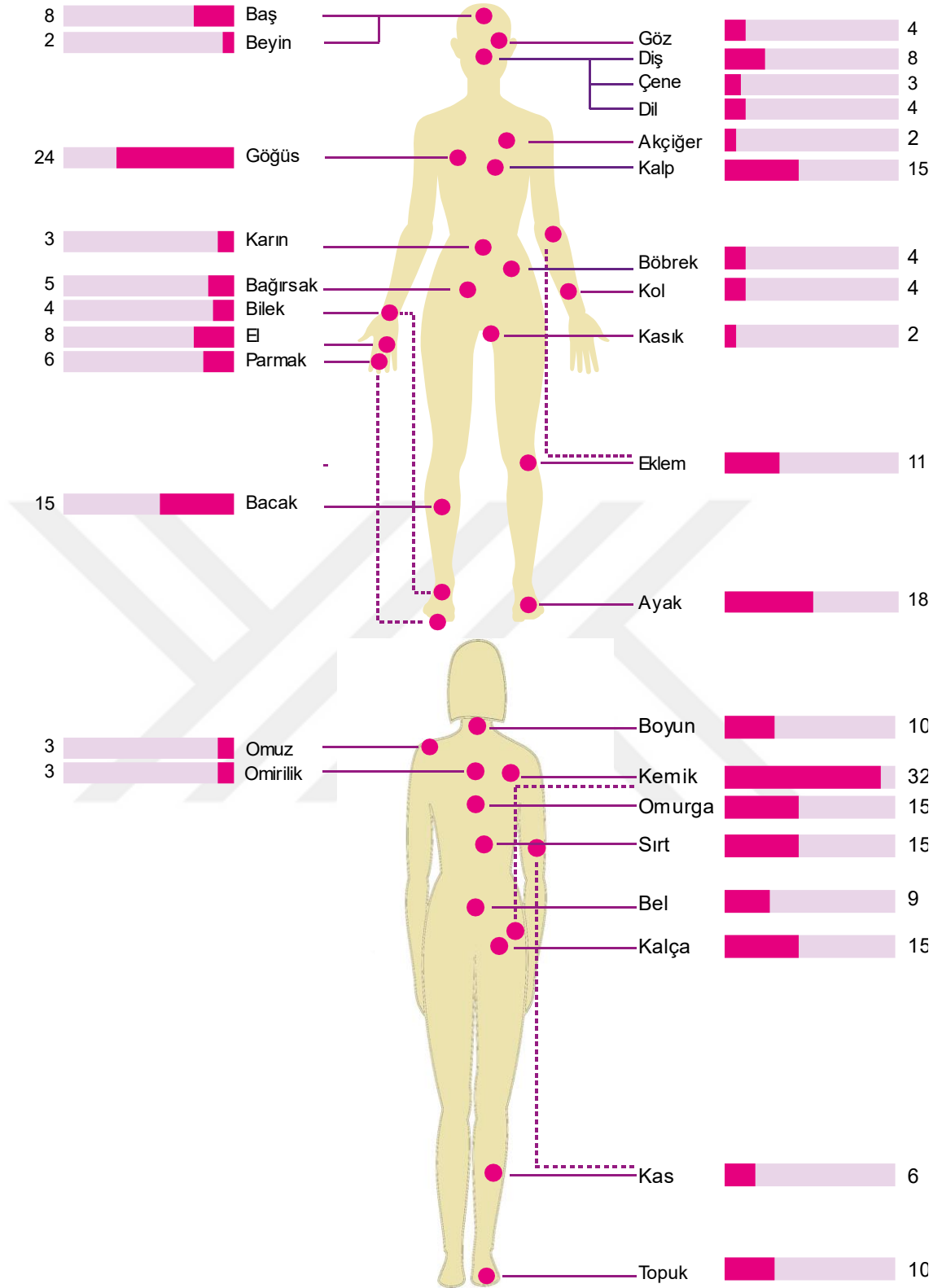
Şekil 10.16. 2015 yılı kadın şikayet bölgeleri



Şekil 10.17. 2016 yılı kadın şikayet bölgeleri



Şekil 10.18. 2017 yılı kadın şikayet bölgeleri

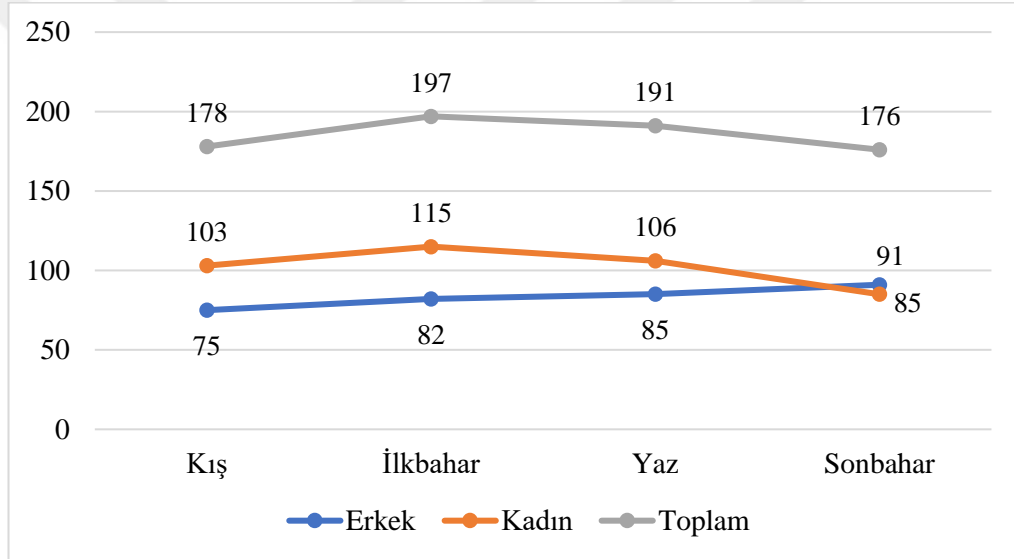


Şekil 10.19. 2018 yılı kadın şikayet bölgeleri

Binary ağırlıklandırma yönteminde yararlanılarak “ağrı” kelimesinin geçtiği gönderiler tespit edilmiştir. Bu tespit sonucunda kadın ve erkek kullanıcılarına göre

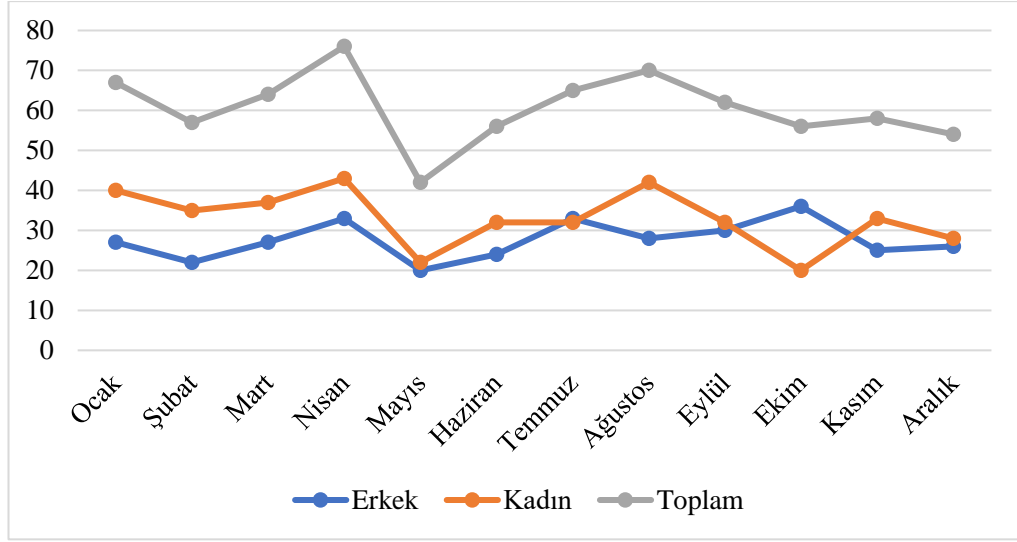
ađrı Őikayetinin mevsimsel ve aylık olarak deđiŐimini çizgi grafiđi ile gösterilmiŐtir. Buradaki çizgi grafiklerini göstermedeki amaç ađrı kelimesinin kadın ve erkek hastaları tarafından en çok hangi mevsimde ve ayda kullanıldığını belirlemek ve göstermektir. Őekil 10.20, ađrı kelimesinin cinsiyete ve toplam kullanıma göre mevsimsel deđiŐimini gösterirken, Őekil 10.21 ise ađrı kelimesinin cinsiyete ve toplam kullanıma göre aylık deđiŐimini göstermektedir.

Őekil 10.20'ye göre ađrı ile ilgili paylaŐımlar en çok ilkbahar mevsiminde yapılmıŐken en az Őikayet ise sonbahar döneminde yapılmıŐtır. Kadın kullanıcılarında kış mevsiminden ilkbahara geçiŐte ađrı ile ilgili Őikayetler artarken, ilbahardan yaz, yazdan sonbahara geçiŐlerde azalmalar görölmektedir. Erkeklerde ise kış mevsiminden sonbahara dođru ađrı ilgili Őikayetlerin arttığı görölmektedir.



Őekil 10.20. 'Ađrı' kelimesinin kullanım sıklığının cinsiyete ve toplam kullanıma göre mevsimsel deđiŐimi

Őekil 10.21'de ađrı ile ilgili toplam en çok Őikayet Nisan ayı içerisinde yapılmıŐtır. En az Őikayet ise Mayıs ayı içerisinde yapıldığı görölmektedir. Erkek kullanıcıları tarafından ađrı kelimesi en çok Ekim ayı içerisinde kullanılmaktayken en az paylaŐım Mayıs içerisinde yapılmıŐtır. Kadın kullanıcılarında ise en çok ađrı ile ilgili Őikayet Nisan ayı içerisinde yapılmıŐken, en az Őikayet ise Ekim ayı içerisinde görölmektedir.



Şekil 10.21. ‘Ağrı’ kelimesinin kullanım sıklığının cinsiyete ve toplam kullanıma göre aylık değişimi

Şikayet kategorisine ait gönderiler için ilişkili örnekleme test sonuçları

Şikayet kategorisini temsil eden anahtar kelimelerin yıllara göre farklılık gösterip göstermediğini test etmek amacıyla Friedman testi uygulanmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Şikayet kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermemektedir.

H_1 : Şikayet kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermektedir.

Test sonuçları Çizelge 10.9’da verilmiştir. Test sonucunda ($p < 0.05$) H_0 hipotezi reddedilir. Buna göre, şikayet kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermektedir. 2012 yılı 2013 – 2018’den, 2013 yılı 2018’den, 2014 yılı 2017 ve 2018’den, 2015 yılı 2018’den, 2016 yılı 2017 ve 2018’den, 2017 yılı 2018’den farklılık göstermektedir.

Çizelge 10.9. Şikayet kategorisine ait gönderi için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	224	2.68 ± 7.94	1.00 (0.00-111.00) ^f	2.62		
2013	224	5.24 ± 13.14	2.00 (0.00-183.00) ^{be}	4.24		
2014	224	4.12 ± 10.76	2.00 (0.00-151.00) ^{cde}	3.64		
2015	224	5.38 ± 12.98	3.00 (0.00-178.00) ^{bde}	4.16	196.226	<0.001
2016	224	4.30 ± 11.03	2.00 (0.00-150.00) ^{cde}	3.75		
2017	224	5.70 ± 14.70	3.00 (0.00-206.00) ^b	4.43		
2018	224	7.20 ± 17.38	3.00 (0.00-238.00) ^a	5.15		

a-f; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Şikayet kategorisine ait gönderiler için uyum analizi sonuçları

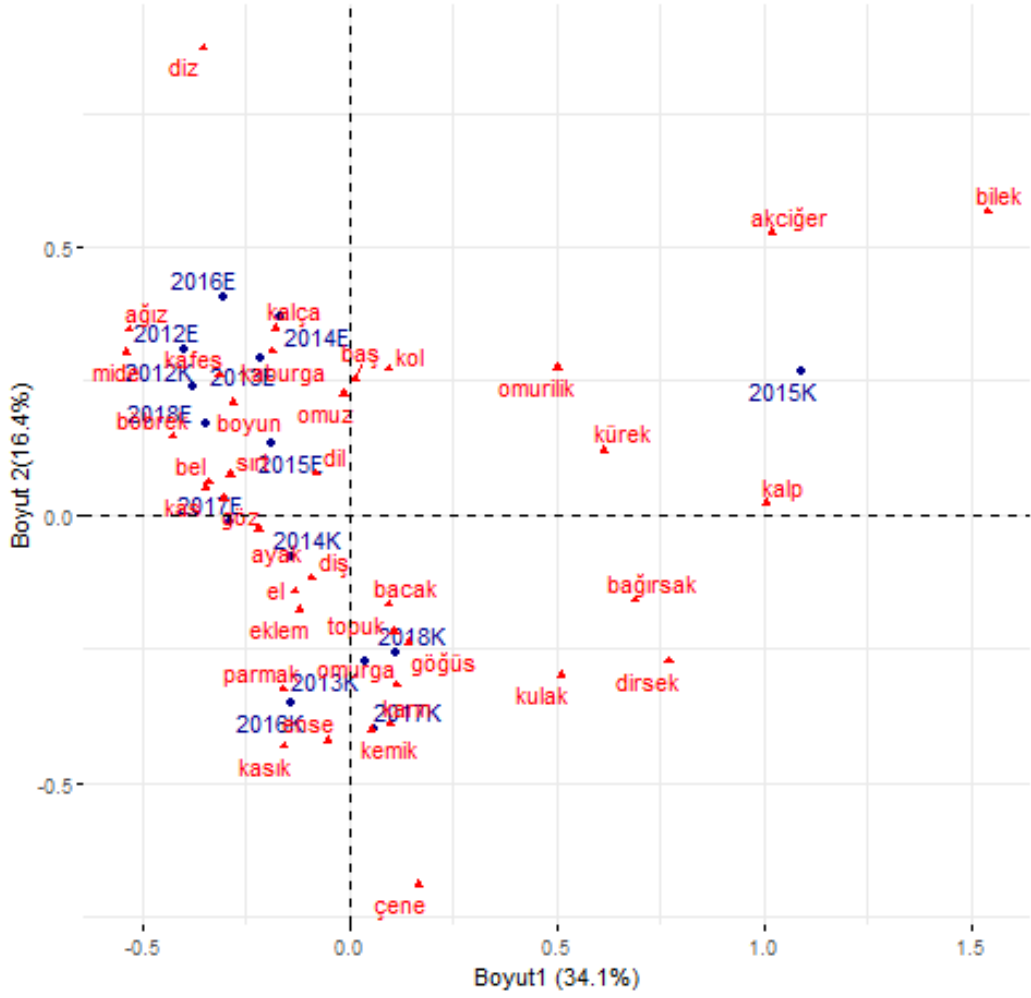
Şikayet bölgeleri ve yıllara göre cinsiyet grupları arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Şikayet bölgeleri ve yıllara göre cinsiyet grupları arasında ilişki yoktur

H_1 : Şikayet bölgeleri ve yıllara göre cinsiyet grupları arasında ilişki vardır

R programında ‘şikayet’ adında 14×38 bir matris oluşturulmuştur. 38, 10 ve üzeri sıklıkta geçen şikayet bölgelerini temsil etmektedir. 14 ise yıllara göre cinsiyet gruplarının sayısını göstermektedir. Test sonucunda $p<0.001$ olduğundan dolayı %99.9 güvenle H_0 hipotezi reddedilir. Sonuç olarak %99.9 güvenle şikayet bölgeleri ve yıllara göre cinsiyet grupları arasında bir ilişki vardır.

Şekil 10.22’de uyum analizi sonucu, boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin (inertianın) %34.1’ini açıklarken, ikinci boyut ise %16.4’ünü açıklamaktadır. Ayrıca yıllara göre erkek kullanıcıları etrafında benzer şikayet bölgelerinin daha çok yoğunlaştığı görülmektedir. Yıllara göre kadın kullanıcılar arasında dağınık bir yapı vardır. Şikayet bölgelerine göre yıllara ait kadın kullanıcıları, erkek kullanıcılarına göre daha değişken bir yapıya sahiptir.



2012E – 2018E; 2012 – 2018 yılları arası erkekleri, 2012K – 2018K; 2012 – 2018 yılları arası kadınları ifade eder.

Şekil 10.22. Şikayet bölgeleri ve yıllara göre cinsiyet grupları için basit uyum analizi sonucu

Bu boyutların şikayet bölgelerini ve yıllara göre kullanıcı gruplarını hangi oranda açıkladığı Çizelge 10.10 ve Çizelge 10.11’de verilmiştir. Çizelge 10.10’da 2018E, 2012K ve 2015K boyut 1, 2012E, 2014E, 2016E, 2016K, 2017K ve 2018K boyut 2, 2013E ve 2014K boyut 3, 2015E ve 2013K boyut 4, 2017E ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.10. Boyutların yıllara göre cinsiyet gruplarını açıklama oranları

Yıllara göre cinsiyet kategorileri	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012E	0.198	0.119	0.093	0.023	0.059
2013E	0.092	0.170	0.552	0.002	0.102
2014E	0.066	0.323	0.243	0.062	0.000
2015E	0.056	0.029	0.010	0.390	0.143
2016E	0.142	0.250	0.167	0.050	0.119
2017E	0.211	0.000	0.003	0.079	0.216
2018E	0.334	0.080	0.123	0.086	0.174
2012K	0.352	0.138	0.087	0.046	0.003
2013K	0.004	0.179	0.006	0.446	0.002
2014K	0.047	0.014	0.266	0.036	0.094
2015K	0.935	0.056	0.000	0.005	0.003
2016K	0.052	0.317	0.013	0.242	0.030
2017K	0.012	0.547	0.001	0.071	0.007
2018K	0.067	0.362	0.000	0.037	0.200

Çizelge 11'e göre ağız, akciğer, bağırsak, bel, bilek, boyun ve dirsek bölgeleri boyut 1, bacak, baş, çene, diz ve eklem bölgeleri boyut 2, ayak, böbrek ve diş bölgeleri boyut 3, dil ise büyük oranda boyut 4 tarafından açıklanmaktadır.

Çizelge 10.11. Boyutların şikayet bölgelerini açıklama oranları

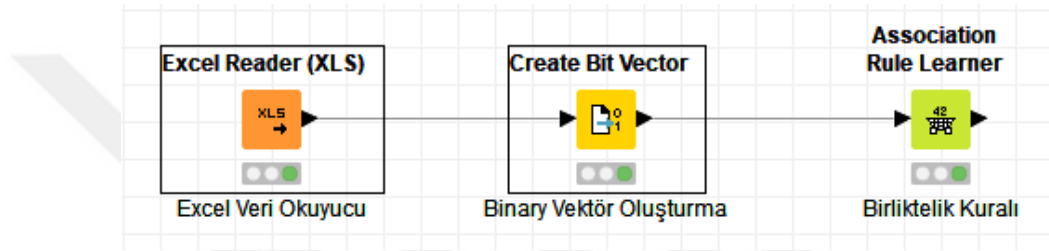
Şikayet Bölgeleri	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
ağız	0.252	0.107	0.220	0.035	0.008
akciğer	0.493	0.132	0.196	0.014	0.003
ayak	0.171	0.002	0.204	0.196	0.052
bacak	0.054	0.169	0.030	0.012	0.050
bağırsak	0.729	0.038	0.000	0.033	0.048
baş	0.001	0.357	0.002	0.146	0.117
bel	0.550	0.018	0.008	0.093	0.090
bilek	0.844	0.114	0.013	0.014	0.010
boyun	0.323	0.180	0.022	0.042	0.108

Çizelge 10.11. (devam)

Şikayet Bölgeleri	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
böbrek	0.179	0.021	0.582	0.005	0.007
çene	0.033	0.571	0.054	0.015	0.000
dil	0.017	0.015	0.167	0.323	0.083
dirsek	0.367	0.046	0.004	0.134	0.074
diş	0.050	0.080	0.738	0.010	0.024
diz	0.112	0.680	0.013	0.037	0.000
eklem	0.076	0.157	0.001	0.011	0.032
el	0.096	0.110	0.017	0.063	0.003
ense	0.001	0.084	0.093	0.325	0.067
göğüs	0.115	0.330	0.000	0.104	0.196
göz	0.251	0.003	0.372	0.056	0.150
kaburga	0.051	0.136	0.002	0.028	0.113
kafes	0.211	0.145	0.090	0.248	0.024
kalça	0.112	0.428	0.019	0.154	0.151
kalp	0.952	0.000	0.003	0.014	0.021
karın	0.012	0.195	0.011	0.005	0.144
kas	0.447	0.010	0.001	0.040	0.065
kasık	0.014	0.104	0.017	0.544	0.090
kemik	0.013	0.814	0.001	0.034	0.005
kol	0.014	0.115	0.020	0.576	0.011
kulak	0.264	0.091	0.022	0.007	0.003
kürek	0.267	0.010	0.000	0.405	0.057
mide	0.156	0.049	0.036	0.083	0.441
omurga	0.050	0.392	0.005	0.135	0.175
omurilik	0.213	0.065	0.439	0.096	0.057
omuz	0.000	0.049	0.437	0.068	0.005
parmak	0.078	0.322	0.005	0.051	0.175
sırt	0.275	0.019	0.416	0.112	0.011
topuk	0.041	0.172	0.342	0.015	0.005

Şikayet kategorisine ait gönderi birliktelik analizi sonuçları

Kullanıcıların şikayet gönderilerinde hangi anahtar kelimeleri birlikte kullandığını tespit etmek için birliktelik analizi yapılmıştır. Birliktelik kurallarına bulmak için Apriori algoritması uygulanmıştır. Şekil 10.23, birliktelik kuralı için Knime’de oluşturulmuş model yapısını göstermektedir. ‘Excel Reader’ operatörü Excel’deki verilerin okunmasını, ‘Create Bit Vector’ verilerin vektör olarak ‘1’ ve ‘0’ şeklinde ifade edilmesini ve hangi parametreler için birliktelik yapılacağını belirlenmesi sağlamaktadır. ‘Association Rule Learner’ operatörü ise birliktelik için girilen destek ve güven değerlerine göre birliktelik kurallarını oluşturmaktadır.



Şekil 10.23. Knime’de birliktelik için oluşturulan model yapısı

Minimum destek değeri 0.01 ve minimum güven değeri 0.70 olarak belirlenerek operatörler çalıştırılmıştır. Çalışma sonucunda toplam 92 kural bulunmuştur. Çizelge 10.12, ikili kullanımlar için birliktelik kurallarını gösterirken Çizelge 10.13 ise 3 ve üzeri kullanımlara ait birliktelik kurallarını göstermektedir. Kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır. Lift değerinin büyük olması bulunan kuralın ilginçliğini göstermektedir.

Çizelge 10.12’de en ilginçlik, Lift değeri 19.2246 olan ‘kapak → diz’ kuralında görülmüştür. ‘kapak’ ve ‘diz’ kelimesi tüm gönderilerin %1.23’ünde birlikte kullanılmıştır. Ayrıca ‘kapak’ kelimesinin kullanıldığı gönderilerin %77.78’inde diz kelimesi de kullanılmaktadır.

Lift değeri 17.4253 olan ‘kafes → göğüs’ kuralında, ‘kafes’ ve ‘göğüs’ kelimesi tüm gönderilerin %2.11’inde birlikte kullanılmıştır. Ayrıca ‘kafes’ kelimesinin kullanıldığı gönderilerin %88.89’unda ‘göğüs’ kelimesi de kullanılmaktadır.

Lift değeri 13.1192 olan ‘kürek → kemik’ kuralında, ‘kürek’ ve ‘kemik’ kelimesi tüm gönderilerin %1.06’sında birlikte kullanılmıştır. Ayrıca ‘kürek’

kelimesinin kullanıldığı gönderilerin %92.31’inde ‘kemik’ kelimesi de kullanılmaktadır.

Lift değeri 7.9510 olan ‘taban → ayak’ kuralında, ‘taban’ ve ‘ayak’ kelimesi tüm gönderilerin %1.06’sında birlikte kullanılmıştır. Ayrıca ‘taban’ kelimesinin kullanıldığı gönderilerin %92.31’inde ‘ayak’ kelimesi de kullanılmaktadır.

Çizelge 10.12. Şikayet kategorisine ait gönderilerde ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[kapak]	→ diz	0.0123	0.7778	19.2246
[kafes]	→ göğüs	0.0211	0.8889	17.4253
[kürek]	→ kemik	0.0106	0.9231	13.1192
[taban]	→ ayak	0.0106	0.9231	7.9510
[böbrek]	→ ağrı	0.0141	1.0000	1.6820
[kes]	→ ağrı	0.0457	0.9286	1.5618
[azal]	→ ağrı	0.0255	0.9062	1.5243
[şiddet]	→ ağrı	0.0475	0.8710	1.4649
[omurga]	→ ağrı	0.0220	0.8621	1.4500
[hamile]	→ ağrı	0.0106	0.8571	1.4417
[mevsim]	→ ağrı	0.0106	0.8571	1.4417
[mide]	→ ağrı	0.0150	0.8500	1.4297
[eklem]	→ ağrı	0.0334	0.8444	1.4203
[tahlil]	→ ağrı	0.0132	0.8333	1.4016
[kaburga]	→ ağrı	0.0211	0.8276	1.3920
[topuk]	→ ağrı	0.0264	0.8108	1.3637
[kalça]	→ ağrı	0.0756	0.8037	1.3518
[bel]	→ ağrı	0.0853	0.8017	1.3483
[tavan]	→ ağrı	0.0106	0.8000	1.3456
[otur]	→ ağrı	0.0352	0.7843	1.3192
[kafes]	→ ağrı	0.0185	0.7778	1.3082
[göğüs]	→ ağrı	0.0396	0.7759	1.3050
[sirt]	→ ağrı	0.0800	0.7583	1.2755
[sabah]	→ ağrı	0.0651	0.7551	1.2700
[uyu]	→ ağrı	0.0431	0.7538	1.2679

Çizelge 10.12. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[kalk]	→	ağrı	0.0396	0.7500	1.2615
[hava]	→	ağrı	0.0290	0.7500	1.2615
[dönem]	→	ağrı	0.0185	0.7500	1.2615
[hareket]	→	ağrı	0.0493	0.7467	1.2559
[omuz]	→	ağrı	0.0176	0.7407	1.2459
[fena]	→	ağrı	0.0150	0.7391	1.2432
[çek]	→	ağrı	0.0563	0.7356	1.2373
[gör]	→	ağrı	0.0361	0.7321	1.2314
[bilek]	→	ağrı	0.0114	0.7222	1.2147
[yatak]	→	ağrı	0.0114	0.7222	1.2147
[boyun]	→	ağrı	0.0651	0.7184	1.2084
[yat]	→	ağrı	0.0246	0.7179	1.2076
[taraf]	→	ağrı	0.0308	0.7143	1.2014
[kas]	→	ağrı	0.0264	0.7143	1.2014
[önce]	→	ağrı	0.0704	0.7080	1.1908
[çocuk]	→	ağrı	0.0106	0.7059	1.1873
[diş]	→	ağrı	0.0167	0.7037	1.1836
[doğru]	→	ağrı	0.0185	0.7000	1.1774

Çizelge 10.13'te üç ve üzeri kullanımları gösteren kurallar içerisinde en ilginçlik, Lift değeri 17.7365 olan 'ağrı, kafes → göğüs' kuralında görülmüştür. 'ağrı, kafes' kelime grubu ve 'göğüs' kelimesi tüm gönderilerin %1.67'sinde birlikte kullanılmıştır. Ayrıca 'ağrı, kafes' kelimelerinin kullanıldığı gönderilerin %90.48'inde göğüs kelimesi de kullanılmaktadır.

Lift değeri 1.6820 olan 'bel, kes → ağrı' kuralında, 'bel, kes' kelime grubu ve 'ağrı' kelimesi tüm gönderilerin %1.06'sında birlikte kullanılmıştır. Ayrıca 'bel, kes' kelimelerinin kullanıldığı gönderilerin %100'ünde ağrı kelimesi de kullanılmaktadır.

Lift değeri 1.1873 olan 'gece, uyu → ağrı' kuralında, 'gece, uyu' kelime grubu ve 'ağrı' kelimesi tüm gönderilerin %1.06'sında birlikte kullanılmıştır. Ayrıca 'gece, uyu' kelimelerinin kullanıldığı gönderilerin %70.59'unda ağrı kelimesi de kullanılmaktadır.

Çizelge 10.13. Şikayet kategorisine ait gönderilerde üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar		Sonuç	Destek	Güven	Lift
[ağrı, kafes]	→	göğüs	0.0167	0.9048	17.7365
[bel, kes]	→	ağrı	0.0106	1.0000	1.6820
[sırt, çek]	→	ağrı	0.0123	1.0000	1.6820
[kalça, önce]	→	ağrı	0.0202	1.0000	1.6820
[şiddet, sırt]	→	ağrı	0.0114	1.0000	1.6820
[bel, sabah]	→	ağrı	0.0150	0.9444	1.5885
[bel, bacak]	→	ağrı	0.0141	0.9412	1.5830
[kalk, ayak]	→	ağrı	0.0123	0.9333	1.5698
[bel, önce]	→	ağrı	0.0220	0.9259	1.5574
[diz, bel]	→	ağrı	0.0106	0.9231	1.5526
[sol, şiddet]	→	ağrı	0.0106	0.9231	1.5526
[sağ, kalça, sol]	→	ağrı	0.0106	0.9231	1.5526
[kalça, çek]	→	ağrı	0.0106	0.9231	1.5526
[kalça, otur]	→	ağrı	0.0106	0.9231	1.5526
[hareket, önce]	→	ağrı	0.0106	0.9231	1.5526
[bel, çek]	→	ağrı	0.0132	0.8824	1.4841
[önce, boyun]	→	ağrı	0.0132	0.8824	1.4841
[sırt, önce]	→	ağrı	0.0132	0.8824	1.4841
[sırt, sabah]	→	ağrı	0.0123	0.8750	1.4717
[boyun, çek]	→	ağrı	0.0123	0.8750	1.4717
[sağ, kalça]	→	ağrı	0.0237	0.8710	1.4649
[sağ, kemik]	→	ağrı	0.0114	0.8667	1.4577
[önce, çek]	→	ağrı	0.0114	0.8667	1.4577
[göğüs, sırt]	→	ağrı	0.0114	0.8667	1.4577
[bel, ayak]	→	ağrı	0.0106	0.8571	1.4417
[şiş, sabah]	→	ağrı	0.0106	0.8571	1.4417
[kalça, yürü]	→	ağrı	0.0106	0.8571	1.4417
[sabah, önce]	→	ağrı	0.0141	0.8421	1.4164
[kalça, bel]	→	ağrı	0.0229	0.8387	1.4107
[bel, boyun]	→	ağrı	0.0211	0.8276	1.3920
[otur, kalk]	→	ağrı	0.0123	0.8235	1.3851
[kalça, boyun]	→	ağrı	0.0114	0.8125	1.3666
[önce, ayak]	→	ağrı	0.0150	0.8095	1.3616

Çizelge 10.13. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[bel, sırt]	→	ağrı	0.0220	0.8065	1.3564
[sol, bacak]	→	ağrı	0.0106	0.8000	1.3456
[gör, bel]	→	ağrı	0.0106	0.8000	1.3456
[sol, hareket]	→	ağrı	0.0106	0.8000	1.3456
[sırt, boyun]	→	ağrı	0.0176	0.8000	1.3456
[göğüs, kafes]	→	ağrı	0.0167	0.7917	1.3315
[hareket, boyun]	→	ağrı	0.0123	0.7778	1.3082
[kalk, sabah]	→	ağrı	0.0176	0.7692	1.2938
[sabah, ayak]	→	ağrı	0.0114	0.7647	1.2862
[kalça, sol]	→	ağrı	0.0167	0.7600	1.2783
[sağ, boyun]	→	ağrı	0.0106	0.7500	1.2615
[bel, hareket]	→	ağrı	0.0123	0.7368	1.2393
[kalça, sırt]	→	ağrı	0.0141	0.7273	1.2232
[gece, sabah]	→	ağrı	0.0114	0.7222	1.2147
[topuk, ayak]	→	ağrı	0.0114	0.7222	1.2147
[gece, uyu]	→	ağrı	0.0106	0.7059	1.1873

10.7.2. Şikayet yorum sonuçları

28865 adet yorum verilerine veri ön işlem uygulanmasından sonra gönderileri temsil eden toplam 1837 adet kelime kökü bulunmuştur ve terim frekansı ağırlıklandırma yöntemi ile ağırlıklandırılmıştır. 11'den aşağı sıklıkta geçen kelimeler ve gereksiz kelimeler elenmiştir. Ön işlem aşaması sonucunda toplam 424 tane anahtar kelime belirlenmiştir. Hastalar tarafından yıllara göre en çok kullanılan kelimeler Çizelge 10.14'te verilmiştir. Yorumlarda sıklıkla ağrı kelimesi çok kullanılmaktadır. Ayrıca yorumlarda 'yap', 'et', 'başla' gibi öneri kelimeleri sıklıkla geçmektedir.

Çizelge 10.14. Şikayet kategorisine ait yorumlarda yıllara göre en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	ağrı, yap, doktor, kullan, et, git, ilaç, hastalık, başla, hastane, yan, gör, kalça, iç, acil, nefes, hareket, göz, boyun, kes, sabah, baş, egzersiz, mide, doğru, hamile, sırt, gece, çocuk, yara, ağız, atak, el, acı, dön, yat, bacak, ayak, çek, gir, yürü, dönem, kan, zarar, aft, bel, böbrek, hava, çıktı, aç
2013	ağrı, yap, et, kullan, doktor, ilaç, git, hastalık, başla, yan, gör, boyun, kalça, kes, ayak, egzersiz, hastane, acil, baş, hareket, bel, göz, iç, gece, çek, sabah, atak, dur, iğne, diş, acı, sırt, dönem, eklem, çocuk, kalk, el, yat, kan, sol, kas, spor, vücut, şiş, dil, kemik, fizik, nefes, topuk, doğru
2014	ağrı, yap, et, kullan, doktor, hastalık, git, ilaç, başla, boyun, egzersiz, ayak, bel, yan, hareket, kalça, kes, acil, gör, sabah, eklem, çek, yat, spor, hastane, iç, iğne, gece, acı, dönem, sırt, kalk, hava, yürü, nem, kemik, otur, sol, nefes, topuk, mide, dil, uyku, baş, atak, bas, dur, diş, ameliyat, el
2015	ağrı, yap, et, kullan, doktor, hastalık, ilaç, git, başla, boyun, yan, hareket, kes, gör, uyu, gece, iç, sabah, sıcak, egzersiz, ayak, bel, dil, çek, kas, yürü, iğne, nefes, uyku, kalça, acil, acı, hava, sol, spor, baş, stres, sırt, vücut, dön, eklem, topuk, dur, anne, kemik, sağ, dönem, nem, atak, yatak
2016	ağrı, yap, et, kullan, doktor, git, ilaç, hastalık, başla, acil, gör, spor, egzersiz, boyun, hareket, ayak, kes, bel, kalça, iğne, gece, sabah, iç, kas, vücut, çek, baş, kemik, dil, atak, dur, uyu, hava, sol, yan, yürü, nefes, uyku, eklem, vitamin, hastane, doğru, bacak, sıcak, kan, sağ, soğuk, sırt, topuk, taraf
2017	ağrı, yap, kullan, doktor, et, hastalık, ilaç, git, başla, ayak, kalça, bel, boyun, gör, hareket, yan, spor, egzersiz, kes, iğne, sabah, acil, iç, gece, çek, göz, baş, kan, eklem, sol, kemik, sırt, şiş, yürü, el, sağ, atak, dur, topuk, çocuk, anne, vitamin, doğru, çıktı, vücut, sürekli, dil, nefes, hastane, yat
2018	ağrı, yap, doktor, et, kullan, hastalık, git, ilaç, başla, ayak, gör, spor, kes, kalça, bel, egzersiz, yan, iğne, göz, hareket, sabah, iç, boyun, çek, eklem, yürü, hastane, kilo, atak, romatizma, acil, baş, vücut, yat, dur, dönem, çıktı, sürekli, sırt, kemik, çocuk, gece, şiş, el, kan, uyku, dil, ağır, vitamin, sol

Belirlenen 424 anahtar kelimenin hepsini frekans değerlerine göre göstermek amacıyla kelime bulutu oluşturulmuştur. Şikayet yorum kategorisine ait kelime bulutu Şekil 10.24'te gösterilmiştir. Kelime bulutunu incelendiğinde 'ağrı', 'git', 'yap', 'et', 'kullan' gibi kelimeler yorumlar içerisinde sıklıkla geçtiği için büyük boyutlarda gösterilmiştir.



Şekil 10.24. Şikayet kategorisine ait yorumları temsil eden anahtar kelimelerin kelime bulutu ile gösterimi

Şikayet kategorisine ait yorum birliktelik analizi sonuçları

Şikayet yorumlarında en çok birlikte kullanılan kelimeleri tespit amacıyla, boyut azaltama işlemi yapılmıştır. Toplamda içerisinde 2’den aşağı kelime kökü bulunduran yorumlar analize dahil edilmemiştir. 12103 yorum analiz için kullanılmıştır.

Minimum güven değeri 0.15 ve minimum destek değeri 0.01 olarak model çalıştırılmıştır. Apriori algoritması tarafından toplam 59 kural bulunmuştur. Bu kurallara Çizelge 10.15 ve Çizelge 10.16’da gösterilmiştir. Çizelge 10.15 ikili kelime kullanımlarına ait birliktelik kurallarını göstermektedir. Çizelge 10.16 ise 3 ve üzeri kullanımlara ait birliktelik kurallarını göstermektedir. Kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.15’te 3.4105 lift değeri ile en ilginç kural ‘ilaç → kullan’ kuralıdır. ‘ilaç’ ve ‘kullan’ kelimesi tüm gönderilerin %4.25’inde birlikte kullanılmıştır. Ayrıca ‘ilaç’ kelimesinin kullanıldığı gönderilerin %42.69’unda ‘kullan’ kelimesi de kullanılmaktadır.

Lift değeri 2.9342 olan ‘egzersiz → yap’ kuralında, ‘egzersiz’ ve ‘yap’ kelimesi tüm gönderilerin %1.89’unda birlikte kullanılmıştır. Ayrıca ‘egzersiz’ kelimesinin kullanıldığı gönderilerin %48.41’inde ‘yap’ kelimesi de kullanılmaktadır.

Lift değeri 2.7477 olan ‘spor → yap’ kuralında, ‘spor’ ve ‘yap’ kelimesi tüm gönderilerin %1.45’inde birlikte kullanılmıştır. Ayrıca ‘spor’ kelimesinin kullanıldığı gönderilerin %45.34’ünde ‘yap’ kelimesi de kullanılmaktadır.

Çizelge 10.15. Şikayet kategorisine ait yorumlarda ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[ilaç] →	kullan	0.0425	0.4269	3.4105
[kullan] →	ilaç	0.0425	0.3393	3.4105
[egzersiz] →	yap	0.0189	0.4841	2.9342
[spor] →	yap	0.0145	0.4534	2.7477
[iğne] →	kullan	0.0105	0.3387	2.7055
[git] →	doktor	0.0354	0.3523	2.5035
[doktor] →	git	0.0354	0.2513	2.5035
[kes] →	ağrı	0.0270	0.6135	2.4258
[kes] →	kullan	0.0124	0.2814	2.2482
[sırt] →	ağrı	0.0149	0.5625	2.2241
[hareket] →	et	0.0116	0.3004	2.1592
[sağ] →	ağrı	0.0104	0.5040	1.9928
[çek] →	ağrı	0.0152	0.4855	1.9196
[bel] →	ağrı	0.0195	0.4826	1.9082
[hareket] →	yap	0.0119	0.3090	1.8728
[başla] →	ilaç	0.0140	0.1819	1.8287
[kalça] →	ağrı	0.0206	0.4594	1.8165
[başla] →	kullan	0.0164	0.2131	1.7027
[başla] →	ağrı	0.0313	0.4080	1.6131
[yürü] →	ağrı	0.0106	0.4076	1.6118
[ilaç] →	hastalık	0.0152	0.1528	1.4530
[ilaç] →	doktor	0.0203	0.2043	1.4521
[hastalık] →	et	0.0211	0.2003	1.4397
[et] →	hastalık	0.0211	0.1514	1.4397
[sabah] →	ağrı	0.0129	0.3594	1.4212
[iğne] →	ağrı	0.0105	0.3387	1.3391
[ilaç] →	yap	0.0219	0.2201	1.3339
[boyun] →	ağrı	0.0143	0.3359	1.3282

Çizelge 10.15. (devam)

Gruplar	Sonuç	Destek	Güven	Lift
[ayak] →	ağrı	0.0153	0.3292	1.3016
[hastalık] →	yap	0.0222	0.2113	1.2807
[başla] →	doktor	0.0136	0.1776	1.2623
[ilaç] →	et	0.0173	0.1736	1.2476
[hastalık] →	doktor	0.0184	0.1752	1.2450
[doktor] →	kullan	0.0216	0.1538	1.2290
[kullan] →	doktor	0.0216	0.1729	1.2290
[ilaç] →	ağrı	0.0308	0.3098	1.2249
[doktor] →	et	0.0240	0.1703	1.2239
[et] →	doktor	0.0240	0.1722	1.2239
[hareket] →	ağrı	0.0114	0.2961	1.1709
[kullan] →	ağrı	0.0366	0.2924	1.1562
[git] →	et	0.0158	0.1572	1.1298
[başla] →	yap	0.0140	0.1830	1.1090
[doktor] →	yap	0.0254	0.1809	1.0961
[yap] →	doktor	0.0254	0.1542	1.0961
[egzersiz] →	ağrı	0.0107	0.2727	1.0783
[yap] →	ağrı	0.0439	0.2659	1.0513
[ağrı] →	yap	0.0439	0.1735	1.0513
[et] →	yap	0.0240	0.1722	1.0437
[yan] →	ağrı	0.0108	0.2636	1.0422
[hastalık] →	ağrı	0.0277	0.2632	1.0405
[gör] →	ağrı	0.0122	0.2606	1.0303
[et] →	ağrı	0.0356	0.2559	1.0120
[git] →	yap	0.0160	0.1597	0.9677
[kullan] →	yap	0.0198	0.1584	0.9601
[git] →	ağrı	0.0231	0.2305	0.9112
[doktor] →	ağrı	0.0320	0.2272	0.8985

Çizelge 10.16'ya göre üç ve üzeri birliktelik için toplam 3 tane kural bulunmuştur. Lift değeri 4.0618 olan 'ağrı, kullan → ilaç' kuralında görülmüştür. 'ağrı, kullan' kelime grubu ve 'ilaç' kelimesi tüm gönderilerin %1.48'inde birlikte kullanılmıştır. Ayrıca 'ağrı, kullan' kelimelerinin kullanıldığı gönderilerin %40.41'inde ilaç kelimesi de kullanılmaktadır.

Çizelge 10.16. Şikayet kategorisine ait yorumlarda üç ve üzeri kelime kullanımları gösteren birliktelik kuralı

Gruplar		Sonuç	Destek	Güven	Lift
[ağrı, kullan]	→	ilaç	0.0148	0.4041	4.0618
[ilaç, ağrı]	→	kullan	0.0148	0.4799	3.8338
[ilaç, kullan]	→	ağrı	0.0148	0.3482	1.3770

10.8. Doktor Kategorisine Ait Sonuçlar

Doktor kategorisine ait toplam 680 adet gönderiye veri ön işlem aşaması uygulandıktan sonra gönderileri temsil eden toplam 1421 adet kelime kökü bulunmuştur. Bu kelimeler terim frekansı ve binary ağırlıklandırma yöntemleri ile sayısallaştırılmıştır. 5'ten aşağı frekans değerine sahip olan kelimeler elenmiştir. Tek başına bir anlam ifade etmeyen gereksiz kelimeler de listeden çıkarılmıştır. Eleme sonucunda toplam 96 anahtar kelime belirlemiştir. Kullanıcılar tarafından doktor kategorisinde yıllara göre en çok kullanılan 50 kelime Çizelge 10.17'de verilmiştir.

Çizelge 10.17. Doktor kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	hastane, doktor, randevu, hastalık, öner, durum, sistem, hasta, onkoloji, cerrahi, tedavi, muayene, üniversite, uzman, hoca, araştırma, devlet, bölüm, tanı, iyi, ağrı, hangi, gün, yakın, özel, yarın, eğitim, sonuç, acil, kullan, ilaç, teşhis, kontrol, memnun, takip, fizik, protez, tekrar, nöroloji, son, öneri, sıra, saat, karar, numara, şiddet, geç, hekim, ameliyat, alan
2013	doktor, hastane, tedavi, tavsiye, hasta, iyi, randevu, hoca, üniversite, ağrı, bilgi, öner, sistem, uzman, muayene, yardım, ilgi, kan, ilaç, son, tahlil, hastalık, bölüm, hangi, gün, özel, teşhis, kontrol, takip, tıp, araştırma, yakın, yarın, eğitim, kullan, geç, acil, memnun, öneri, hekim, konu, fakülte, hafta, poliklinik, paylaş, numune, durum, sonuç, protez, sıra
2014	doktor, hastane, tavsiye, iyi, ağrı, tedavi, randevu, öner, muayene, yardım, konu, hasta, üniversite, uzman, hangi, acil, kullan, hoca, hastalık, gün, ilaç, teşhis, kontrol, araştırma, memnun, ilgi, takip, tıp, yakın, fakülte, durum, sağlık, devlet, fizik, bilgi, bölüm, özel, yarın, eğitim, hafta, protez, dal, bel, süreç, tekrar, nöroloji, ücret, dönem, varit, pozitif
2015	doktor, hastane, gün, hoca, iyi, randevu, muayene, tavsiye, ağrı, hangi, öner, yardım, bilgi, konu, uzman, hasta, üniversite, tekrar, son, tedavi, teşhis, araştırma, yakın, aile, ihtiyaç, hastalık, kontrol, ilgi, tıp, sağlık, devlet, yarın, hafta, göz, saat, izin, geç, dahiliye, bilgilen, acil, fakülte, bölüm, özel, eğitim, ücret, dönem, tahlil, öneri, poliklinik, tanı

Çizelge 10.17. (devam)

Yıl	Kelimeler
2016	doktor, hastane, iyi, hoca, randevu, öner, muayene, hangi, ağrı, tavsiye, tedavi, devlet, hasta, bilgi, ilgi, yardım, sıra, gün, araştırma, aile, hastalık, hafta, saat, acil, tahlil, son, kullan, konu, kontrol, geç, sevk, uzman, üniversite, yarın, eğitim, fizik, varit, ilaç, durum, kan, tekrar, dahiliye, bölüm, bel, asistan, hareket, memnun, takip, teşhis, ihtiyaç
2017	doktor, hastane, randevu, hoca, tavsiye, iyi, tedavi, hangi, araştırma, bölüm, muayene, öner, ağrı, hasta, gün, bilgi, kontrol, eğitim, yardım, devlet, fizik, ameliyat, üniversite, bel, memnun, özel, sonuç, tahlil, kullan, ilaç, öneri, hastalık, uzman, teşhis, yakın, karar, tanı, hafta, konu, sevk, tıp, ikamet, sağlık, ilgi, aile, saat, acil, tekrar, takip, alan
2018	doktor, hastane, iyi, randevu, tavsiye, öner, ağrı, tedavi, bilgi, hoca, yardım, hasta, gün, hangi, devlet, fizik, ilaç, muayene, özel, hastalık, uzman, kan, araştırma, bölüm, teşhis, eğitim, üniversite, tahlil, kullan, ameliyat, tanı, hafta, acil, hekim, sevk, alan, geç, dahiliye, sistem, bel, memnun, sonuç, yakın, konu, ilgi, tekrar, durum, poliklinik, numara, kemik

Bulunan 96 adet anahtar kelimelerin tamamını kullanım sıklığına göre göstermek için kelime bulutu kullanılmıştır. Doktor gönderi kategorisine ait kelime bulutu Şekil 10.25'te gösterilmiştir. Kelime bulutundaki 'doktor', 'hastane', 'randevu', 'iyi', 'hoca' gibi büyük boyuttaki kelimelerin doktor sınıfına ait gönderilerde sık kullanıldığını gözükmektedir.

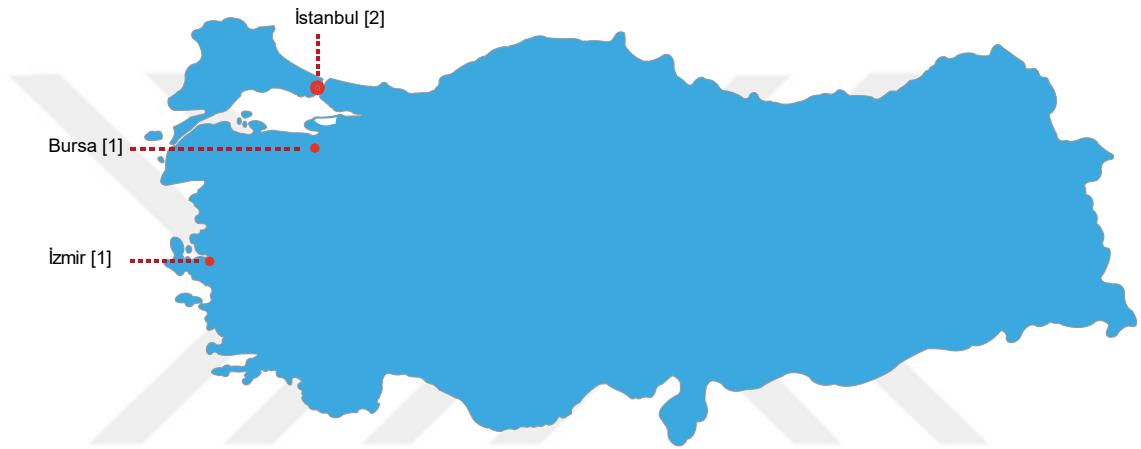


Şekil 10.25. Doktor kategorisine ait gönderileri temsil eden anahtar kelimelerin kelime bulutu ile gösterimi

Gönderilerde kullanıcıların hangi illerde daha çok doktor tavsiyesi sorduklarını belirlemek amacıyla gönderilerde en çok kullanılan illerin isimleri ve sıklıkları

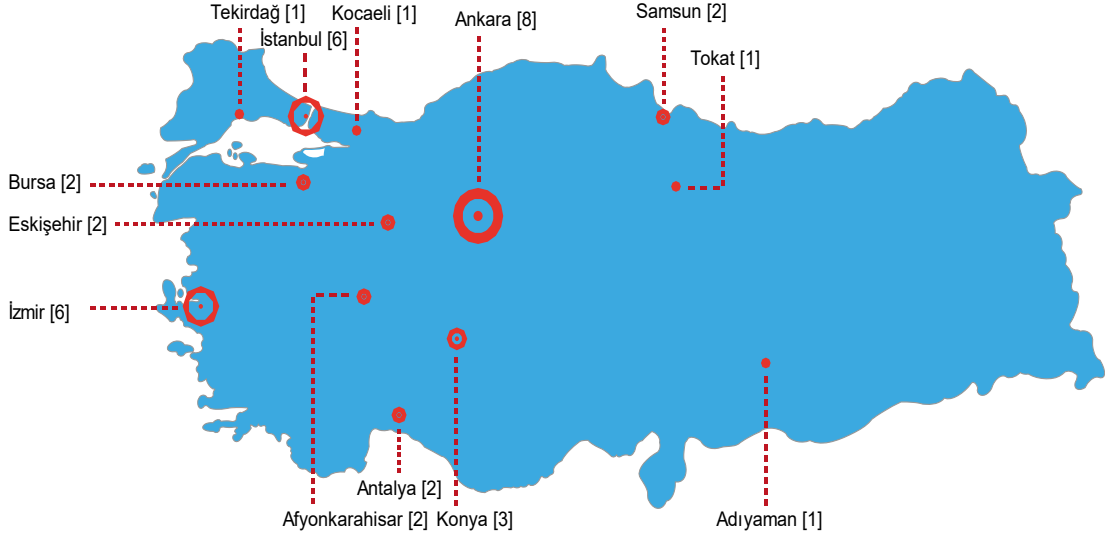
belirlenmiştir. Oluşturulan haritalarda kırmızı yuvarlaklar illerin bulunduğu yerleri göstermektedir. Ayrıca illerin kullanım sıklığı arttıkça balonların boyutu büyümektedir. Köşeli parantez içerisindeki sayılar illerin gönderilerde kaç kez geçtiğini göstermektedir.

Şekil 10.26'da erkek kullanıcılar tarafından üç il için doktor tavsiyesi sorulmuştur. Bu iller İstanbul, İzmir ve Bursa illeridir. En çok İstanbul için sorulmuştur. Frekans değerlerinin küçük olmasının sebebi grubun 2012 yılında kurulmuş olması ve bu konuda yeterli paylaşımların olmamasından kaynaklanmaktadır



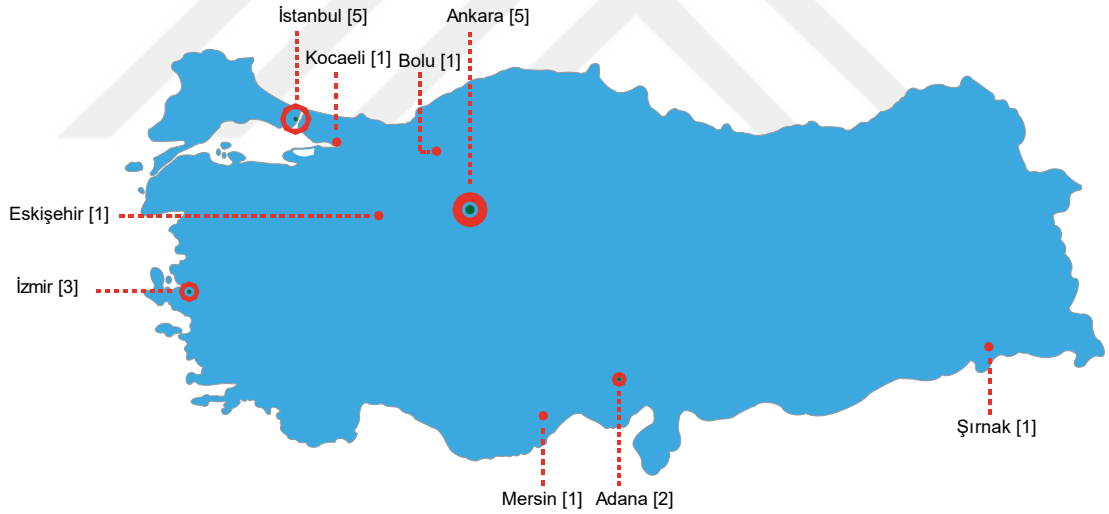
Şekil 10.26. 2012 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.27'de erkek kullanıcılar tarafından 2013 yılında toplam 13 il için doktor tavsiyesi sorulmuştur. En çok doktor tavsiyesi Ankara, İstanbul, İzmir ve Konya ili için sorulmuştur.



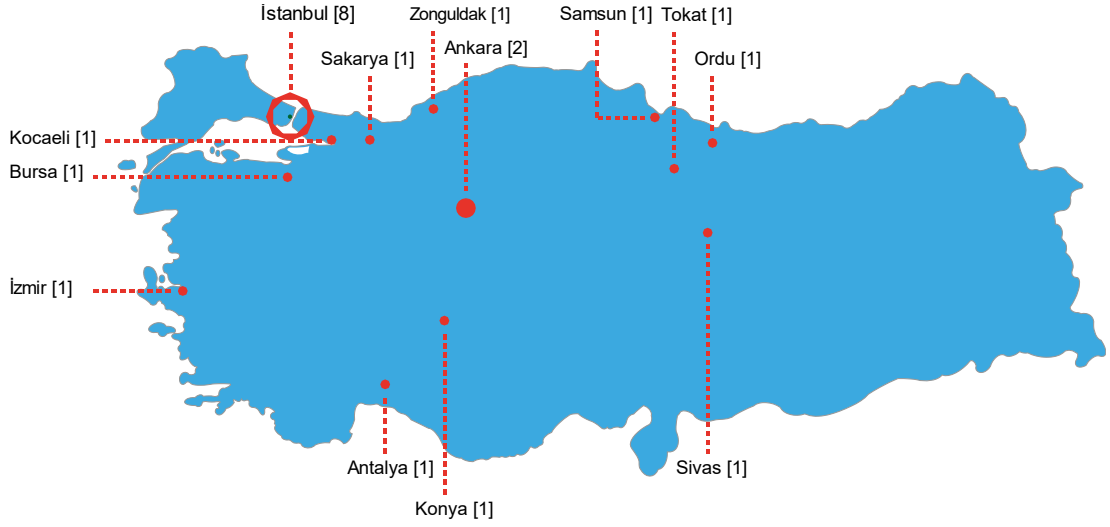
Şekil 10.27. 2013 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.28'e göre 2014 yılında erkek kullanıcılar tarafından toplam 9 il için doktor tavsiyesi sorulmuştur. En çok Ankara, İstanbul ve İzmir ili için doktor tavsiyesi sorulmuştur.



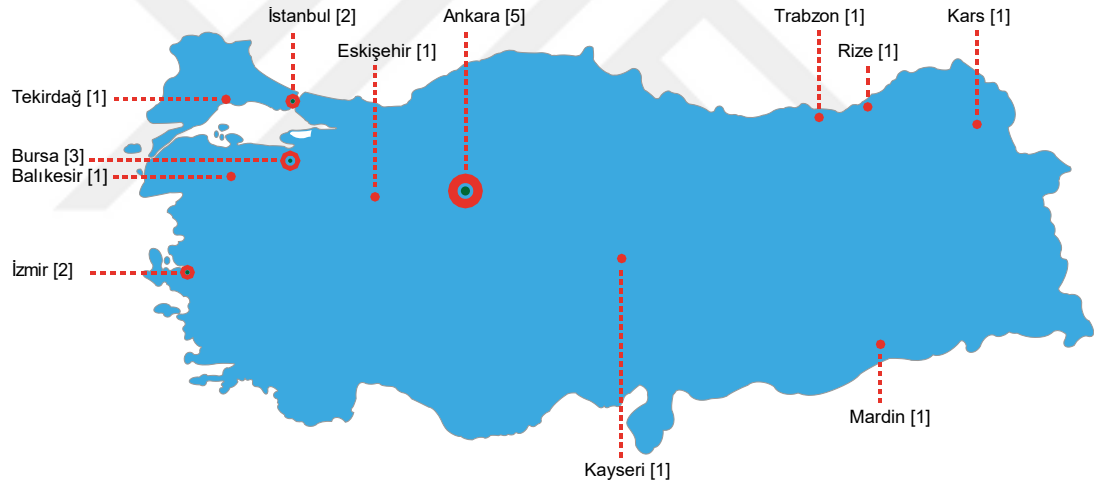
Şekil 10.28. 2014 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.29'da erkek kullanıcılar tarafından 2015 yılında toplam 13 il için doktor tavsiyesi sorulmuştur. İstanbul dışındaki diğer illerden fazla bir doktor tavsiyesi ile ilgili paylaşım yapılmamıştır. Erkek kullanıcılar tarafından bu dönemde en çok İstanbul ili için doktor tavsiyesi öğrenilmeye çalışılmıştır.



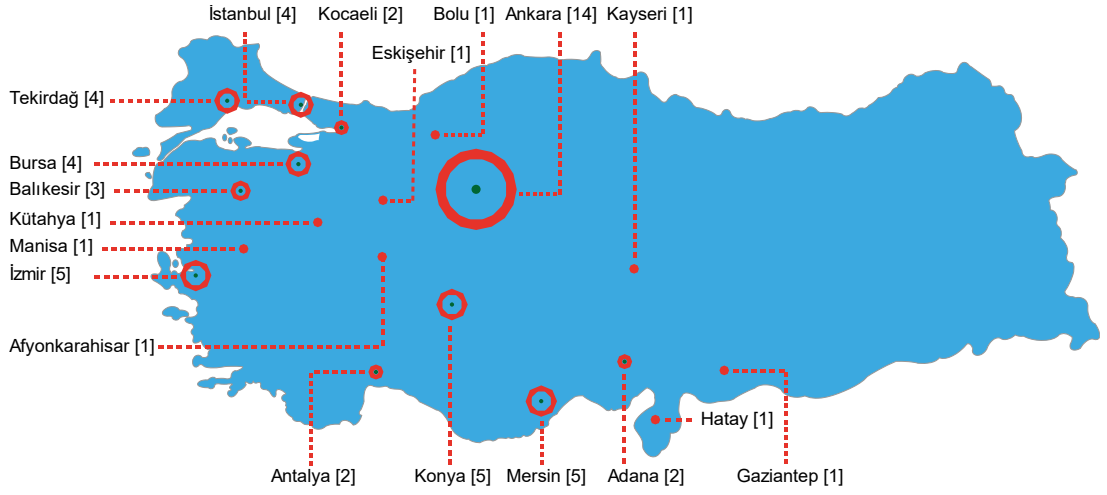
Şekil 10.29. 2015 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.30'da 2016 yılında erkek kullanıcılara tarafından toplam 12 il için doktor tavsiyesi sorulmuştur. Diğer illere göre Ankara ve Bursa ön plana çıkmaktadır.



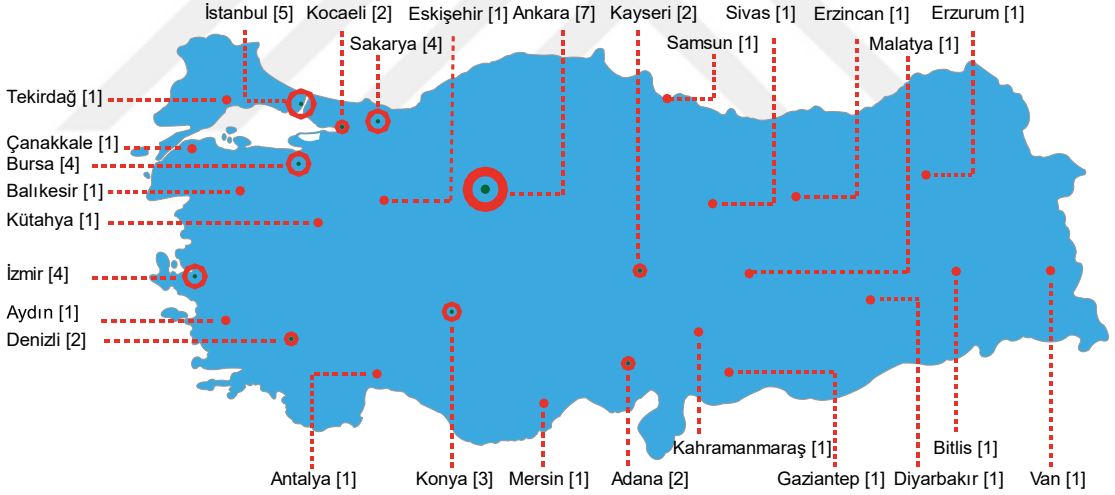
Şekil 10.30. 2016 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.31'e göre 2017 yılı erkek kullanıcılar tarafından toplam 19 yıl için doktor tavsiyesi sorulmuştur. En çok Ankara, İzmir ve Konya illeri için doktor tavsiyesi sorulmuştur.



Şekil 10.31. 2017 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.32'ye göre 2018 yılında erkek kullanıcılar tarafından toplam 28 il için doktor tavsiyesi sorulmuştur. En çok Ankara, İstanbul, Bursa, İzmir ve Sakarya illeri için sorulmuştur.



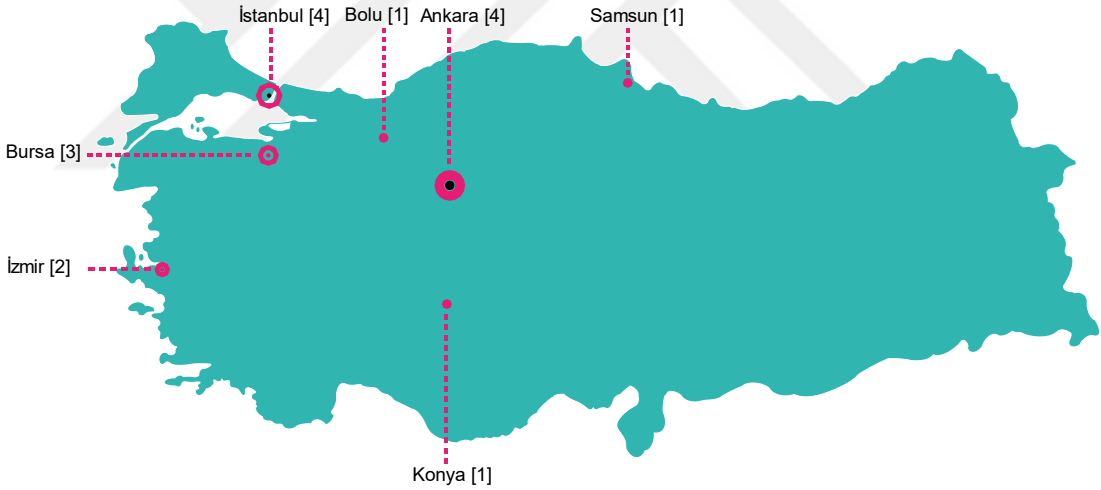
Şekil 10.32. 2018 yılı erkek kullanıcıların en çok doktor tavsiyesi sorduğu iller

Kadın kullanıcılar için de aynı haritalar oluşturulmuştur. Erkekler için oluşturulan haritalarada farkı haritanın ve yuvarlak şekillerin rengi farklıdır. 2012 yılı kadın kullanıcılarına ait Şekil 10.33'te toplam 6 il için doktor tavsiyesi sorulmuştur. En çok İzmir ve Ankara ili için doktor tavsiyesi sorulmuştur.



Şekil 10.33. 2012 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.34'e göre 2013 yılında kadın kullanıcılara tarafından toplam 7 ilin ismi geçmektedir. Bu dönemde en çok Ankara, İstanbul ve Bursa illeri için doktor tavsiyesi sorulmuştur.



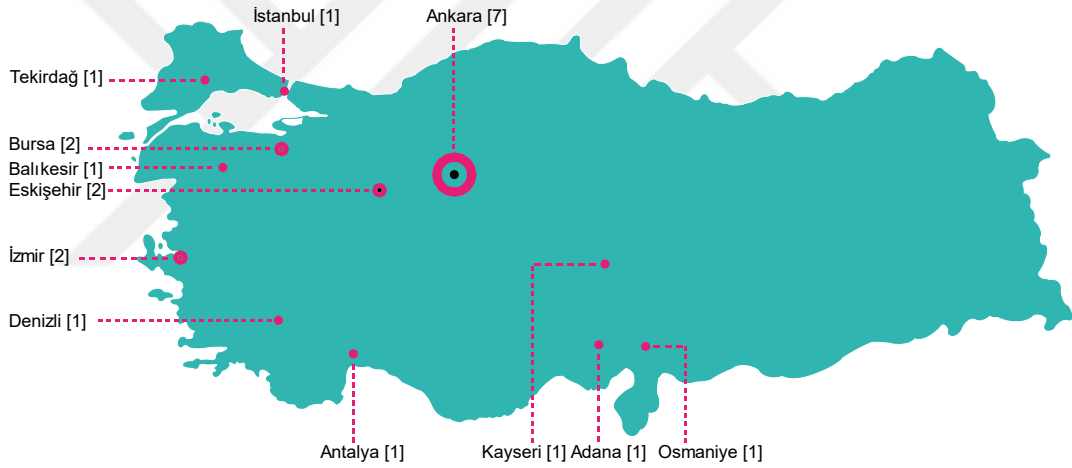
Şekil 10.34. 2013 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.35'e göre 2014 yılında kadın kullanıcılar tarafından toplam 8 il için doktor tavsiyesi sorulmuştur. En çok Ankara ve İstanbul illerin ismi geçmektedir.



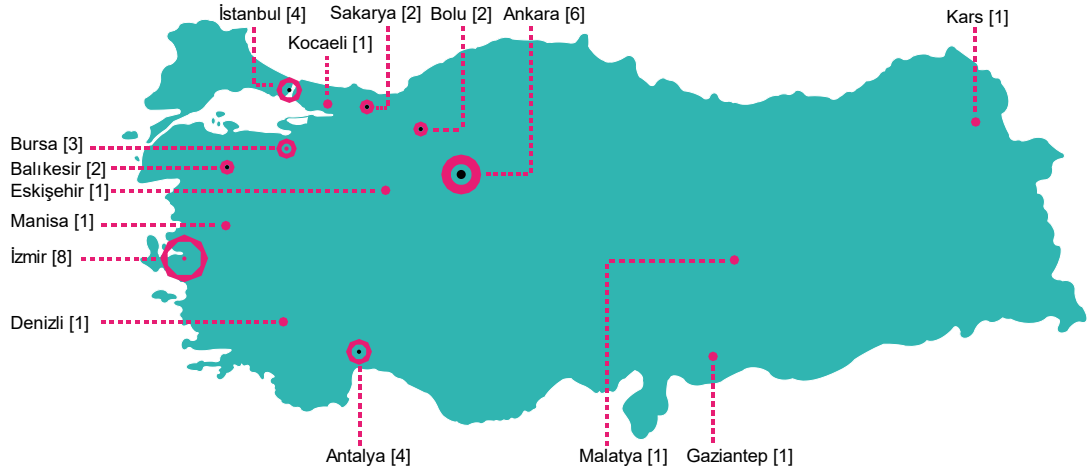
Şekil 10.35. 2014 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.36'da kadın kullanıcılar tarafından 2015 yılında toplam 12 yıl için doktor tavsiyesi sorulmuştur. En çok Ankara ili için sorulmuştur.



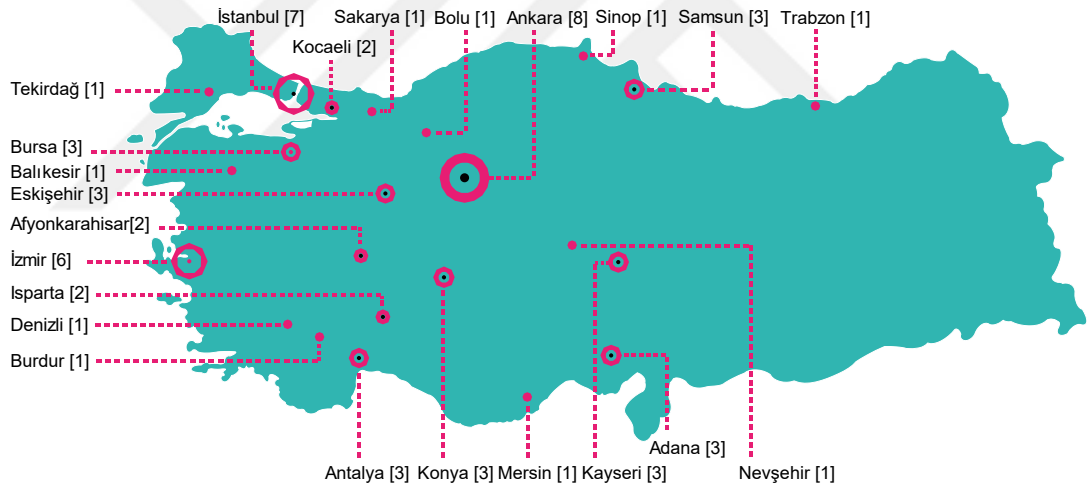
Şekil 10.36. 2015 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.37'de kadın kullanıcıları tarafından 2016 yılı için toplam 15 il için doktor tavsiyesi sorulmuştur. En çok İzmir, Ankara, İstanbul ve Antalya illeri için sorulmuştur.



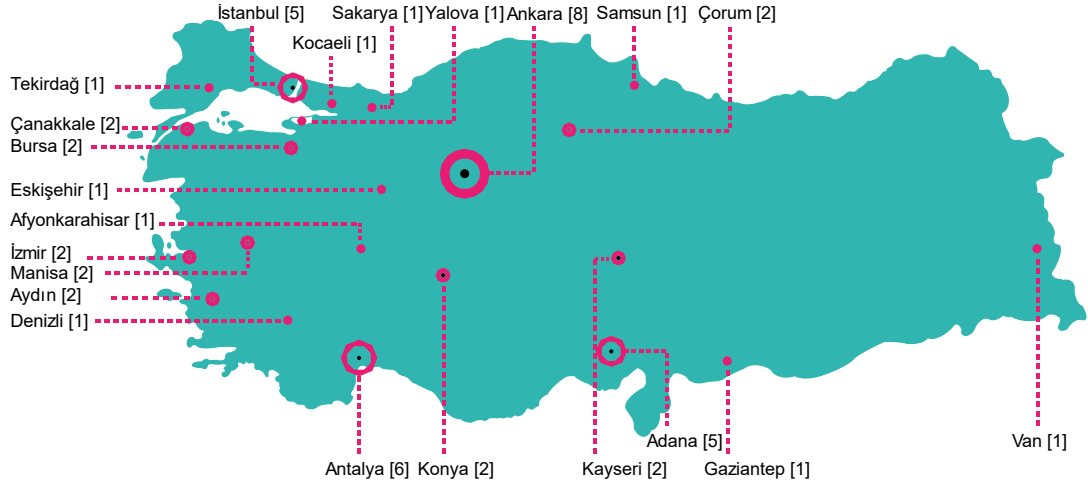
Şekil 10.37. 2016 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.38'e göre 2017 yılı kadın kullanıcılar tarafından toplam 23 il için doktor tavsiyesi öğrenilmeye çalışılmıştır. Bunlar arasından en çok Ankara, İstanbul ve İzmir illeri için sorulmuştur.



Şekil 10.38. 2017 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Şekil 10.39'a göre 2018 yılında kadın kullanıcılar tarafından toplam 22 için doktor tavsiyesi sorulmuştur. En çok Ankara, Antalya, İstanbul ve Adana illeri için doktor tavsiyesi sorulmuştur.



Şekil 10.39. 2018 yılı kadın kullanıcıların en çok doktor tavsiyesi sorduğu iller

Doktor kategorisine ait gönderiler için ilişkili örneklem test sonuçları

Doktor kategorisini temsil eden anahtar kelimelerin yıllara göre farklılık gösterip göstermediğini test etmek amacıyla Friedman testi uygulanmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Doktor kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermemektedir.

H_1 : Doktor kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermektedir.

Çizelge 10.18. Doktor kategorisine ait gönderiler için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	96	1.59±3.12	1.00 (0.00-21.00) ^e	2.26		
2013	96	3.98±5.85	2.00 (0.00-42.00) ^{cd}	3.86		
2014	96	3.48±6.19	2.00 (0.00-48.00) ^c	3.49		
2015	96	3.76±6.03	2.00 (0.00-40.00) ^{bc}	3.82	127.87	<0.001
2016	96	5.67±9.84	3.00 (0.00-67.00) ^{abd}	4.60		
2017	96	7.51±13.38	3.00 (0.00-100.00) ^a	5.09		
2018	96	6.19±11.40	3.00 (0.00-90.00) ^a	4.88		

a-e; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Test sonuçları Çizelge 10.18’de verilmiştir. Test sonucunda ($p<0.05$) H_0 hipotezi reddedilir. Buna göre, doktor kategorisine ait gönderileri temsil eden anahtar kelimeler yıllara göre farklılık göstermektedir. 2012 yılı 2013-2018’den, 2013 yılı 2017 ve

2018'den, 2014 yılı 2016-2018'den, 2015 yılı 2017 ve 2018'den farklılık göstermektedir.

Doktor kategorisine ait gönderiler için uyum analizi sonuçları

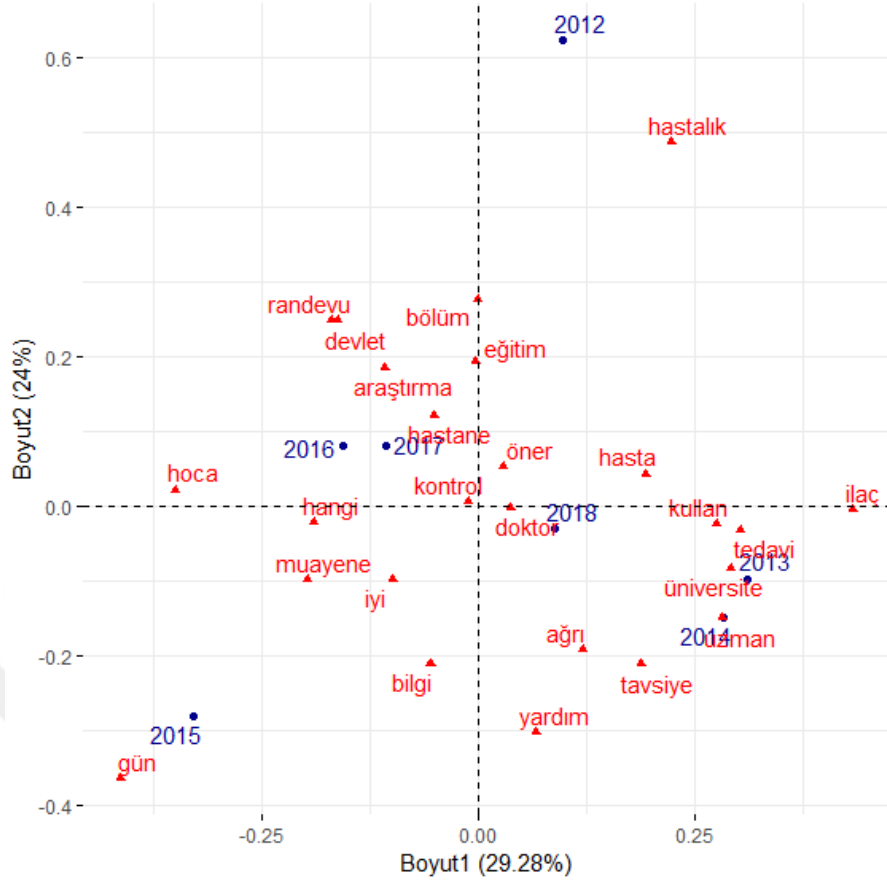
Doktor kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Doktor kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişki yoktur

H_1 : Doktor kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişki vardır

R programında 7x25 boyutunda bir matris oluşturulmuştur. 25, en çok kullanılan 25 anahtar kelimeyi temsil etmektedir. 7 ise toplam yılların sayısını göstermektedir. Test sonucunda $p < 0.001$ 'den dolayı %99.9 güvenle H_0 hipotezi reddedilir. Sonuç olarak %99.9 güvenle doktor kategorisini temsil eden anahtar kelimeler ve yıllar arasında bir ilişki vardır.

Şekil 10.40'da uyum analizi sonucu, boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %29.28'ini açıklarken, ikinci boyut ise %24'ünü açıklamaktadır. Ayrıca 'gün' kelimesinin 2015 yılında ve 'hastalık' kelimesini 2012 yılı içerisinde daha yoğun bir şekilde kullanıldığı görülmektedir.



Şekil 10.40. Doktor kategorisini temsil eden anahtar kelimeler ve yıllar için basit uyum analizi sonucu

Bu boyutların anahtar kelimeleri ve yılları hangi oranda açıkladığı Çizelge 10.19 ve Çizelge 10.20’de verilmiştir. Çizelge 10.19’da 2013, 2014 ve 2015 boyut 1, 2012 boyut 2, 2017 boyut 3, 2016 boyut 4, 2018 ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.19. Doktor kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012	0.016	0.653	0.172	0.159	0.000
2013	0.628	0.061	0.086	0.079	0.015
2014	0.518	0.142	0.036	0.000	0.245
2015	0.420	0.305	0.102	0.171	0.001
2016	0.273	0.076	0.083	0.393	0.087
2017	0.134	0.080	0.757	0.006	0.000
2018	0.121	0.013	0.191	0.266	0.364

Çizelge 10.20’de hoca, tedavi, muayene, gün, hangi, hasta, uzman, üniversite, ilaç ve kullan kelimeleri boyut 1, randevu, ağrı, yardım ve hastalık kelimeleri boyut 2, hastane, tavsiye, öner, araştırma, bölüm, kontrol ve eğitim kelimeleri boyut 3, doktor, iyi ve devlet kelimeleri boyut 4, bilgi kelimesi ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.20. Boyutların doktor anahtar kelimelerini açıklama oranları

Anahtar Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
doktor	0.144	0.000	0.001	0.358	0.005
hastane	0.055	0.311	0.627	0.000	0.007
randevu	0.310	0.680	0.010	0.000	0.000
hoca	0.719	0.003	0.130	0.011	0.022
iyi	0.102	0.101	0.133	0.647	0.006
ağrı	0.157	0.396	0.102	0.027	0.180
tavsiye	0.276	0.346	0.348	0.000	0.013
tedavi	0.632	0.007	0.312	0.007	0.006
muayene	0.479	0.115	0.060	0.082	0.220
öner	0.021	0.070	0.606	0.237	0.055
gün	0.370	0.286	0.056	0.242	0.031
hangi	0.692	0.008	0.093	0.054	0.117
hasta	0.591	0.030	0.043	0.044	0.062
bilgi	0.018	0.270	0.002	0.048	0.543
yardım	0.039	0.783	0.063	0.059	0.031
araştırma	0.108	0.327	0.421	0.032	0.003
hastalık	0.093	0.443	0.354	0.108	0.001
bölüm	0.000	0.265	0.505	0.073	0.095
uzman	0.387	0.107	0.222	0.273	0.000
üniversite	0.480	0.038	0.007	0.410	0.028
devlet	0.132	0.314	0.192	0.350	0.001
kontrol	0.001	0.000	0.662	0.092	0.227
ilaç	0.666	0.000	0.002	0.165	0.166
eğitim	0.000	0.315	0.516	0.020	0.122
kullan	0.503	0.004	0.005	0.241	0.241

Doktor kategorisine ait gönderi birliktelik analizi sonuçları

Toplam veri dosyası; 680'dir. 2'den aşağı anahtar kelime barındıran metin dosyaları elenmiştir. Analizde toplam 550 metin dosyası kullanılmıştır. Doktor kategorisine ait gönderi paylaşımlarında kullanıcıların birlikte kullandığı kelimeleri veya kelime gruplarını bulmak için birliktelik analizi yapılmıştır. Apriori algoritması kullanılmıştır. Minimum destek değeri %1 ve minimum güven değeri %70 olarak belirlenmiş ve algoritma çalıştırılmıştır. Algoritma sonucunda toplam 78 tane kural bulunmuştur. Bulunan kurallar Çizelge 10.21 ve Çizelge 10.22'de verilmiştir. Bu çizelgede yer alan kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.21'de 29.1176 Lift değeri ile en ilginç kural 'fakülte → tıp' kuralıdır. 'fakülte' ve 'tıp' kelimesi tüm gönderilerin %1.64'ünde birlikte kullanılmıştır. Ayrıca 'fakülte' kelimesinin kullanıldığı gönderilerin %90'ında 'tıp' kelimesi de kullanılmaktadır.

Lift değeri 11.4675 olan 'eğitim → araştırma' kuralında, 'eğitim' ve 'araştırma' kelimesi tüm gönderilerin %4.73'ünde birlikte kullanılmıştır. Ayrıca 'eğitim' kelimesinin kullanıldığı gönderilerin %89.66'sında 'araştırma' kelimesi de kullanılmaktadır.

Lift değeri 5.5733 olan 'fizik → tedavi' kuralında, 'fizik' ve 'tedavi' kelimesi tüm gönderilerin %3.45'inde birlikte kullanılmıştır. Ayrıca 'fizik' kelimesinin kullanıldığı gönderilerin %76'sında 'tedavi' kelimesi de kullanılmıştır.

Çizelge 10.21. Doktor gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[fakülte]	→ tıp	0.0164	0.9000	29.1176
[eğitim]	→ araştırma	0.0473	0.8966	11.4675
[fizik]	→ tedavi	0.0345	0.7600	5.5733
[sistem]	→ randevu	0.0236	0.8125	4.4688
[dahiliye]	→ randevu	0.0127	0.7000	3.8500
[devlet]	→ hastane	0.0545	0.7692	2.5487
[araştırma]	→ hastane	0.0600	0.7674	2.5427

Çizelge 10.21. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[sevk]	→	Hastane	0.0164	0.7500	2.4849
[eğitim]	→	hastane	0.0382	0.7241	2.3993
[dahiliye]	→	hastane	0.0127	0.7000	2.3193
[ihtiyaç]	→	doktor	0.0145	0.8889	1.8174
[ikamet]	→	doktor	0.0127	0.8750	1.7890
[kullan]	→	doktor	0.0364	0.7692	1.5728
[memnun]	→	doktor	0.0273	0.7500	1.5335
[ilaç]	→	doktor	0.0345	0.7308	1.4941
[öner]	→	doktor	0.0982	0.7297	1.4920
[öneri]	→	doktor	0.0182	0.7143	1.4604

Çizelge 10.22’de üç ve üzeri birliktelik kurallarına göre Lift değeri 22 olan ‘tedavi, hastane, araştırma → fizik’ kuralında ‘tedavi, hastane, araştırma’ kelime grubu ve ‘fizik’ kelimesi tüm gönderilerin %1.09’unda birlikte kullanılmıştır. Ayrıca ‘tedavi, hastane, araştırma’ kelimelerinin kullanıldığı gönderilerin %100’ünde fizik kelimesi de kullanılmaktadır.

Lift değeri 16.2562 olan ‘hoca, araştırma’ → eğitim’ kuralında ‘hoca, araştırma’ kelime grubu ve ‘eğitim’ kelimesi tüm gönderilerin %1.09’unda birlikte kullanılmıştır. Ayrıca ‘hoca, araştırma’ kelimelerinin kullanıldığı gönderilerin %85.71’inde eğitim kelimesi de kullanılmaktadır.

Lift değeri 13.7931 olan ‘hastane, randevu, eğitim’ → eğitim’ kuralında ‘hastane, randevu, eğitim’ kelime grubu ve ‘eğitim’ kelimesi tüm gönderilerin %1.45’inde birlikte kullanılmıştır. Ayrıca ‘hastane, randevu, eğitim’ kelimelerinin kullanıldığı gönderilerin %72.73’ünde eğitim kelimesi de kullanılmaktadır.

Çizelge 10.22. Doktor gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralları

Gruplar		Sonuç	Destek	Güven	Lift
[tedavi, hastane, araştırma]	→	fizik	0.0109	1.0000	22.0000
[hoca, araştırma]	→	eğitim	0.0109	0.8571	16.2562
[hastane, randevu, eğitim]	→	eğitim	0.0145	0.7273	13.7931

Çizelge 10.22. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[tedavi, hastane, eğitim]	→	araştırma	0.0200	1.0000	12.7907
[hoca, eğitim]	→	araştırma	0.0109	1.0000	12.7907
[hastane, eğitim]	→	araştırma	0.0145	1.0000	12.7907
[doktor, hastane, eğitim]	→	araştırma	0.0364	0.9524	12.1816
[doktor, eğitim]	→	araştırma	0.0109	0.8571	10.9635
[doktor, sağlık]	→	araştırma	0.0127	0.7778	9.9483
[fizik, bölüm]	→	Gün	0.0109	0.7500	6.9915
[hastane, fizik]	→	tedavi	0.0109	0.8571	6.2857
[doktor, fizik]	→	tedavi	0.0127	0.7778	5.7037
[dahiliye, hastane]	→	tedavi	0.0164	0.7500	5.5000
[bilgi, hastane, araştırma]	→	randevu	0.0109	0.8571	4.7143
[doktor, hastane, araştırma]	→	randevu	0.0109	0.8571	4.7143
[randevu, devlet]	→	randevu	0.0145	0.7273	4.0000
[randevu, araştırma, eğitim]	→	hastane	0.0182	1.0000	3.3133
[tedavi, araştırma, eğitim]	→	hastane	0.0200	1.0000	3.3133
[tedavi, araştırma]	→	hastane	0.0109	1.0000	3.3133
[bilgi, randevu, araştırma]	→	hastane	0.0127	1.0000	3.3133
[bilgi, araştırma]	→	hastane	0.0109	1.0000	3.3133
[ilaç, araştırma]	→	hastane	0.0127	1.0000	3.3133
[iyi, araştırma]	→	hastane	0.0109	1.0000	3.3133
[randevu, araştırma]	→	hastane	0.0109	1.0000	3.3133
[üniversite, araştırma]	→	hastane	0.0309	0.9444	3.1292
[doktor, randevu, araştırma]	→	hastane	0.0164	0.9000	2.9819
[araştırma, bölüm]	→	hastane	0.0145	0.8889	2.9451
[muayene, araştırma]	→	hastane	0.0127	0.8750	2.8991
[dahiliye, randevu]	→	hastane	0.0127	0.8750	2.8991
[tedavi, devlet]	→	hastane	0.0109	0.8571	2.8399
[doktor, araştırma, eğitim]	→	hastane	0.0109	0.8571	2.8399
[doktor, araştırma]	→	hastane	0.0109	0.8571	2.8399
[doktor, eğitim]	→	hastane	0.0200	0.7857	2.6033
[araştırma, eğitim]	→	hastane	0.0127	0.7778	2.5770
[hasta, araştırma]	→	hastane	0.0364	0.7692	2.5487

Çizelge 10.22. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[sevk, doktor]	→	hastane	0.0109	0.7500	2.4849
[gün, araştırma]	→	hastane	0.0109	0.7500	2.4849
[durum, hastane]	→	hastane	0.0109	0.7500	2.4849
[öner, ilaç]	→	doktor	0.0109	1.0000	2.0446
[ilaç, kullan]	→	doktor	0.0109	1.0000	2.0446
[hastane, randevu, iyi]	→	doktor	0.0145	0.8889	1.8174
[iyi, acil]	→	doktor	0.0127	0.8750	1.7890
[tavsiye, yakın]	→	doktor	0.0127	0.8750	1.7890
[yakın, iyi]	→	doktor	0.0109	0.8571	1.7525
[ilaç, hastalık]	→	doktor	0.0109	0.8571	1.7525
[konu, uzman]	→	doktor	0.0109	0.8571	1.7525
[iyi, konu]	→	doktor	0.0109	0.8571	1.7525
[hastane, kullan]	→	doktor	0.0164	0.8182	1.6729
[muayene, iyi]	→	doktor	0.0145	0.8000	1.6357
[yardım, iyi]	→	doktor	0.0145	0.8000	1.6357
[tedavi, iyi]	→	doktor	0.0218	0.8000	1.6357
[hasta, ilaç]	→	doktor	0.0200	0.7857	1.6065
[tavsiye, acil]	→	doktor	0.0127	0.7778	1.5903
[yardım, öner]	→	doktor	0.0109	0.7500	1.5335
[gün, sağlık]	→	doktor	0.0109	0.7500	1.5335
[hangi, öner]	→	doktor	0.0109	0.7500	1.5335
[randevu, iyi]	→	doktor	0.0145	0.7273	1.4870
[tavsiye, gün]	→	doktor	0.0218	0.7059	1.4433
[ağrı, kullan]	→	doktor	0.0127	0.7000	1.4312
[iyi, ilaç]	→	doktor	0.0127	0.7000	1.4312

10.9. Tahlil Kategorisine Ait Sonuçlar

10.9.1. Tahlil gönderi sonuçları

Doktor kategorisine ait toplam 334 adet gönderiye veri ön işlem aşaması uygulandıktan sonra gönderileri temsil eden toplam 1097 adet kelime kökü bulunmuştur. Bu kelimeler terim frekansı, TFxIDF ve binary ağırlıklandırma

yöntemleri ile ağırlıklandırılmıştır. 5'ten aşağı frekans değerine sahip olan kelimeler elenmiştir. Tek başına bir anlam ifade etmeyen gereksiz kelimeler de listeden çıkarılmıştır. Eleme sonucunda toplam 80 anahtar kelime belirlemiştir. Kullanıcılar tarafından tahlil kategorisinde yıllara göre en çok kullanılan 50 kelime Çizelge 10.23'te verilmiştir.

Çizelge 10.23. Tahlil kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	yorum, crp, tahlil, ağrı, yap, sonuç, sedim, testi, eklem, kan, doktor, iyi, çıktı, hastalık, ppd, hastane, tedavi, disk, ağır, kontrast, emar , ilaç, normal, hasta, negatif, pozitif, yüksek, teşhis, hlab27, kontrol, kas, hafif, değer, test, röntgen, muayene, kemik, boyun, çıkar, ayak, iltihap, özel, genetik, göğüs, bilgi, verem, tanı, film, sakroiliak, yüksel
2013	yorum, tahlil, kan, ağrı, yap, doktor, crp, sonuç, sedim, emar , testi, iyi, çıktı, ilaç, değer, muayene, ppd, hastane, tedavi, hasta, verem, film, iğne, hastalık, normal, pozitif, yüksek, röntgen, boyun, göğüs, bilgi, rapor, düşük, akciğer, negatif, teşhis, hlab27, test, genetik, yüksel, yeni, idrar, b12, eklem, disk, kontrol, kas, çıkar, iltihap, özel
2014	kan, yorum, testi, tahlil, yap, doktor, sonuç, emar , kontrol, çıktı, ilaç, normal, crp, eklem, değer, hastane, verem, film, bilgi, ağrı, iyi, hasta, iğne, negatif, teşhis, b12, kas, kemik, sakroiliak, kol, muayene, ppd, tedavi, hastalık, pozitif, yüksek, röntgen, rapor, hlab27, yüksel, yeni, çıkar, iltihap, bel, hoca, etki, doktora, bozuk, kötü, seviye
2015	doktor, kan, yorum, emar , testi, ağrı, pozitif, tahlil, çıktı, ilaç, değer, verem, bilgi, iyi, yap, sonuç, normal, tedavi, sinir, test, kontrol, crp, negatif, sakroiliak, hlab27, yeni, tanı, yardım, romatizma, eklem, hastane, hasta, iğne, ppd, yüksek, röntgen, iltihap, bel, doktora, kötü, boyun, düşük, akciğer, fitik, ayak, kalça, protein
2016	ağrı, yorum, crp, yap, doktor, emar , değer, iyi, sonuç, negatif, çıktı, ilaç, sedim, kan, normal, hasta, testi, tedavi, hlab27, hastane, tahlil, yüksek, verem, yeni, iğne, röntgen, iltihap, düşük, ayak, teşhis, yüksel, kronik, sakroileit, pozitif, kontrol, tanı, yardım, kalça, muayene, hoca, test, romatizma, eklem, ppd, bel, doktora, kötü, akciğer, film, b12
2017	emar , kan, testi, ağrı, sonuç, çıktı, teşhis, yorum, crp, doktor, hlab27, ppd, yap, tedavi, pozitif, tanı, hastalık, bilgi, sakroiliak, hastane, test, doktora, bulgu, negatif, sedim, normal, hasta, tahlil, yüksek, yeni, sakroileit, hoca, özel, temiz, kalça, kemik, çıkar, boyun, rapor, genetik, değer, iyi, ilaç, verem, röntgen, düşük, kronik, kontrol, eklem, bel
2018	emar , crp, tahlil, sonuç, doktor, ağrı, yorum, sedim, yap, kan, bilgi, çıktı, değer, normal, yüksek, bel, hasta, rapor, testi, ilaç, kol, teşhis, hastalık, kalça, yüksel, pozitif, tanı, sakroiliak, test, negatif, sakroileit, hoca, temiz, iyi, film, taş, yardım, idrar, protein, hastane, doktora, kemik, eklem, şiddet, randevu, böbrek, bozuk, ppd, tedavi, bulgu

Belirlenen 80 anahtar kelimeyi frekans değerlerine göre göstermek amacıyla kelime bulutu oluşturulmuştur. Şekil 10.41'deki kelime bulutuna göre 'yorum',

Çizelge 10.24. Tahlil kategorisine ait gönderiler için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	80	2.84±3.42	2.00 (0.00-13.00) ^{bd}	4.00		
2013	80	2.84±3.33	2.00 (0.00-15.00) ^{bd}	4.01		
2014	80	1.91±2.62	1.00 (0.00-15.00) ^{cd}	3.20		
2015	80	1.59±2.02	1.00 (0.00-9.00) ^c	2.96	73.60	<0.001
2016	80	3.03±3.58	1.50 (0.00-18.00) ^b	4.33		
2017	80	2.56±2.56	1.50 (0.00-11.00) ^{bd}	4.11		
2018	80	5.14±6.10	3.00 (0.00-31.00) ^a	5.39		

a-d; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Tahlil kategorisine ait gönderiler için uyum analizi sonuçları

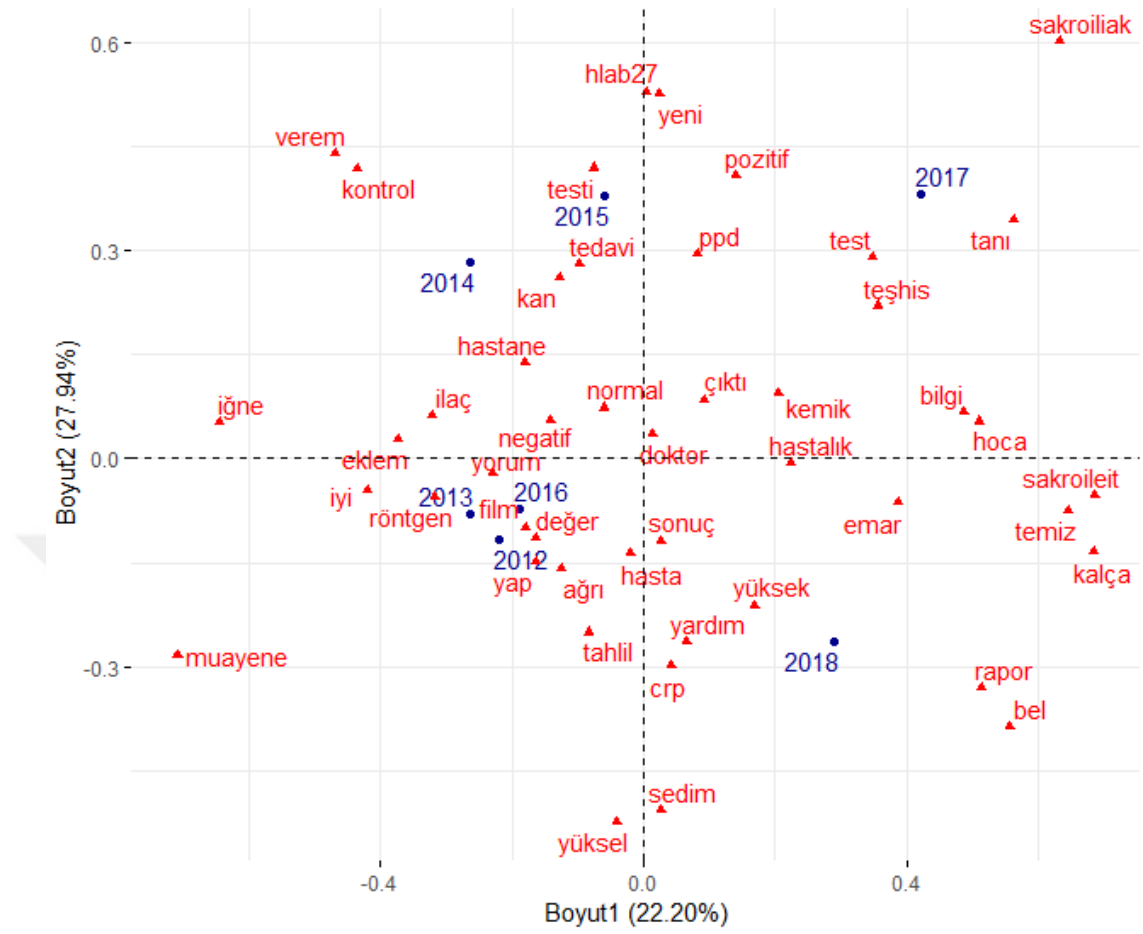
Tahlil kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Tahlil kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişki yoktur

H_1 : Tahlil kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişki vardır

R programında 7×47 boyutunda bir matris oluşturulmuştur. 47, toplamda 10 üzeri sıklıkta geçen anahtar kelime sayısını göstermektedir. 7 ise toplam yılların sayısını göstermektedir. Test sonucunda $p<0.001$ 'ten dolayı %99.9 güvenle H_0 hipotezi reddedilir. Sonuç olarak %99.9 güvenle tahlil kategorisini temsil eden anahtar kelimeler ve yıllar arasında bir ilişki vardır.

Şekil 10.42'de uyum analizi sonucu, boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %22.20'ini açıklarken, ikinci boyut ise %27.94'ünü açıklamaktadır. yorum, negatif, film, değer, röntgen, yap ve eklem gibi kelimelerin 2012, 2013 ve 2016 yılları etrafında, yüksel, rapor, bel, sedim, kalça, temiz kelimelerin 2018 yılı etrafında, sakroiliak, tanı, test kelimelerinin 2017 yılı etrafında yoğunlaştığı görülmektedir. verem, kontrol, hlab27, testi, tedavi, kan ve hastane kelimelerinin ise 2014 ve 2015 yılları etrafında yoğunlaştığı gözükmektedir.



Şekil 10.42. Tahlil kategorisini temsil eden anahtar kelimeler ve yıllar için basit uyum analizi sonucu

Bu boyutların anahtar kelimeleri ve yılları hangi oranda açıkladığı Çizelge 10.25 ve Çizelge 10.26'da verilmiştir. Çizelge 10.25'te 2013, 2017 ve 2018 boyut 1, 2015 boyut 2, 2014 boyut 3, 2012 ve 2016 büyük oranda boyut 4 tarafından açıklanmaktadır.

Çizelge 10.25. Tahlil kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4
2012	0.180	0.051	0.248	0.381
2013	0.389	0.036	0.002	0.000
2014	0.169	0.195	0.388	0.122
2015	0.009	0.356	0.092	0.248
2016	0.158	0.024	0.192	0.473
2017	0.436	0.355	0.161	0.021
2018	0.464	0.387	0.144	0.000

Çizelge 10.26’da yorum, emar, yap, çıktı, ilaç, iyi, bilgi, teşhis, verem, tanı, sakroiliak, bel, rapor, röntgen, hoca, iğne, muayene, kalça, sakroileit, temiz, yorum ve emar kelimeleri boyut 1, crp, tahlil, testi, sedim, hasta, pozitif, yüksek, yeni ve yüksel kelimeleri boyut 2, ağrı, kan, doktor, hastane, tedavi, ppd, hlab27, kontrol ve film kelimeleri boyut 3, sonuç, değer, negatif, hastalık, eklem, kemik ve yardım kelimeleri boyut 4, normal ve test kelimeleri ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.26. Boyutların tahlil anahtar kelimelerini açıklama oranları

Anahtar Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
yorum	0.926	0.008	0.003	0.005	0.057
emar	0.696	0.019	0.202	0.076	0.006
ağrı	0.118	0.189	0.481	0.210	0.002
kan	0.075	0.316	0.434	0.081	0.002
crp	0.016	0.763	0.088	0.040	0.066
tahlil	0.038	0.338	0.147	0.316	0.134
doktor	0.003	0.019	0.370	0.328	0.107
yap	0.416	0.342	0.058	0.182	0.002
sonuç	0.018	0.347	0.031	0.368	0.179
testi	0.025	0.762	0.009	0.163	0.040
çıktı	0.409	0.342	0.033	0.025	0.190
sedim	0.002	0.738	0.163	0.070	0.008
değer	0.119	0.059	0.064	0.731	0.027
ilaç	0.521	0.019	0.194	0.224	0.041
iyi	0.554	0.007	0.122	0.220	0.002
normal	0.042	0.063	0.320	0.090	0.483
bilgi	0.542	0.010	0.365	0.001	0.064
hasta	0.006	0.273	0.150	0.175	0.121
hastane	0.238	0.139	0.333	0.042	0.061
negatif	0.076	0.011	0.236	0.360	0.316
tedavi	0.037	0.306	0.531	0.054	0.068
pozitif	0.052	0.433	0.016	0.039	0.256
yüksek	0.342	0.540	0.085	0.018	0.003
teşhis	0.317	0.121	0.249	0.132	0.043
hastalık	0.127	0.000	0.292	0.538	0.038

Çizelge 10.26. (devam)

Anahtar Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
ppd	0.012	0.154	0.373	0.305	0.148
hlab27	0.000	0.424	0.541	0.024	0.006
eklem	0.168	0.001	0.000	0.555	0.079
kontrol	0.221	0.204	0.231	0.112	0.212
verem	0.341	0.299	0.124	0.156	0.074
test	0.291	0.202	0.045	0.016	0.306
tanı	0.576	0.216	0.134	0.048	0.024
film	0.069	0.021	0.323	0.143	0.018
sakroiliak	0.488	0.440	0.010	0.047	0.002
bel	0.427	0.205	0.340	0.025	0.001
rapor	0.358	0.146	0.186	0.026	0.096
röntgen	0.639	0.020	0.133	0.092	0.057
yeni	0.001	0.525	0.044	0.357	0.013
yüksel	0.004	0.706	0.048	0.082	0.073
hoca	0.572	0.006	0.005	0.011	0.084
iğne	0.479	0.003	0.072	0.147	0.027
muayene	0.497	0.080	0.012	0.010	0.192
kalça	0.706	0.027	0.001	0.222	0.035
kemik	0.098	0.021	0.001	0.453	0.426
sakroileit	0.518	0.003	0.124	0.128	0.150
temiz	0.774	0.011	0.121	0.027	0.004
yardım	0.008	0.136	0.056	0.425	0.030
yorum	0.926	0.008	0.003	0.005	0.057
emar	0.696	0.019	0.202	0.076	0.006

Tahsil kategorisine ait gönderi birliktelik analizi sonuçları

Toplam gönderi sayısı 334'tür. Bulunan anahtar kelimelerinden 2'den aşağı kelime barından gönderiler elenmiştir. Analizde toplam 279 gönderi kullanılmıştır. Birliktelik kurallarını çıkarmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.015 ve minimum güven değeri 0.7 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 66 kural bulunmuştur. Bu kurallar Çizelge 10.27 ve Çizelge 10.28'de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.27’de 5.3654 Lift değeri ile en ilginç kural ‘şiddet → ağrı’ kuralıdır. ‘şiddet’ ve ‘ağrı’ kelimesi tüm gönderilerin %2.15’inde birlikte kullanılmıştır. Ayrıca ‘şiddet’ kelimesinin kullanıldığı gönderilerin %100’ünde ‘ağrı’ kelimesi de kullanılmaktadır.

Lift değeri 4.6947 olan ‘ayak → ağrı’ kuralında, ‘ayak’ ve ‘ağrı’ kelimesi tüm gönderilerin %2.51’inde birlikte kullanılmıştır. Ayrıca ‘ayak’ kelimesinin kullanıldığı gönderilerin %87.50’sinde ‘ağrı’ kelimesi de kullanılmaktadır.

Lift değeri 4,2923 olan ‘kalça → ağrı’ kuralında, ‘kalça’ ve ‘ağrı’ kelimesi tüm gönderilerin %2.87’sinde birlikte kullanılmıştır. Ayrıca ‘kalça’ kelimesinin kullanıldığı gönderilerin %80’inde ‘ağrı’ kelimesi de kullanılmıştır.

Lift değeri 3.9997 olan ‘sedim → crp’ kuralında, ‘sedim’ ve ‘crp’ kelimesi tüm gönderilerin %14.70’inde birlikte kullanılmıştır. Ayrıca ‘sedim’ kelimesinin kullanıldığı gönderilerin %93.18’inde ‘crp’ kelimesi de kullanılmıştır.

Çizelge 10.27. Tahlil kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar		Sonuç	Destek	Güven	Lift
[şiddet]	→	ağrı	0.0215	1.0000	5.3654
[ayak]	→	ağrı	0.0251	0.8750	4.6947
[kalça]	→	ağrı	0.0287	0.8000	4.2923
[sedim]	→	crp	0.1470	0.9318	3.9997
[bulgu]	→	ağrı	0.0251	0.7000	3.7558
[ppd]	→	test	0.0502	0.7000	3.6849
[yüksel]	→	crp	0.0323	0.8182	3.5119
[sakroiliak]	→	emar	0.0394	0.8462	3.4214
[kalça]	→	emar	0.0287	0.8000	3.2348
[sakroileit]	→	emar	0.0287	0.8000	3.2348
[boyun]	→	emar	0.0179	0.7143	2.8882
[bulgu]	→	emar	0.0251	0.7000	2.8304

Çizelge 10.28’de üç ve üzeri birliktelik kurallarına göre Lift değeri 12.2368 olan ‘çıktı, pozitif → hlab27’ kuralında ‘çıktı, pozitif’ kelime grubu ve ‘hlab27’ kelimesi tüm gönderilerin %1.79’unda birlikte kullanılmıştır. Ayrıca ‘çıktı, pozitif’

kelimelerinin kullanıldığı gönderilerin %83.33'ünde 'hlab27' kelimesi de kullanılmaktadır.

Lift değeri 6.3409 olan 'crp, sonuç, tahlil → sedim' kuralında 'crp, sonuç, tahlil' kelime grubu ve 'sedim' kelimesi tüm gönderilerin %1.79'unda birlikte kullanılmıştır. Ayrıca 'crp, sonuç, tahlil' kelimelerinin kullanıldığı gönderilerin %100'ünde 'sedim' kelimesi de kullanılmaktadır.

Lift değeri 13.7931 olan 'doktor, hlab27 → çıktı' kuralında 'doktor, hlab27' kelime grubu ve 'çikti' kelimesi tüm gönderilerin %1.79'unda birlikte kullanılmıştır. Ayrıca 'doktor, hlab27' kelimelerinin kullanıldığı gönderilerin %83.33'ünde 'çikti' kelimesi de kullanılmaktadır.

Çizelge 10.28. Tahlil kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar		Sonuç	Destek	Güven	Lift
[çikti, pozitif]	→	hlab27	0.0179	0.8333	12.2368
[crp, sonuç, tahlil]	→	sedim	0.0179	1.0000	6.3409
[doktor, hlab27]	→	çikti	0.0179	0.8333	5.6707
[crp, kan, tahlil]	→	sedim	0.0287	0.8889	5.6364
[crp, tahlil]	→	sedim	0.0466	0.8667	5.4955
[doktor, hareket]	→	ağrı	0.0179	1.0000	5.3654
[kalça, doktor]	→	ağrı	0.0179	1.0000	5.3654
[hasta, crp]	→	sedim	0.0179	0.8333	5.2841
[crp, sonuç, ağrı]	→	sedim	0.0179	0.8333	5.2841
[crp, kan, sonuç]	→	sedim	0.0179	0.8333	5.2841
[crp, sonuç]	→	sedim	0.0358	0.8333	5.2841
[kalça, emar]	→	ağrı	0.0251	0.8750	4.6947
[hareket, ağrı]	→	doktor	0.0179	1.0000	4.6500
[crp, yorum]	→	sedim	0.0179	0.7143	4.5292
[crp, kan, ağrı]	→	sedim	0.0179	0.7143	4.5292
[negatif, doktor]	→	ağrı	0.0179	0.8333	4.4712
[hastane, emar]	→	ağrı	0.0179	0.8333	4.4712
[doktor, bulgu]	→	ağrı	0.0179	0.8333	4.4712
[çikti, tahlil]	→	kan	0.0179	0.8333	4.4712

Çizelge 10.28. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[crp, kan, sonuç]	→	ağrı	0.0179	0.8333	4.4712
[crp, sonuç, ağrı]	→	kan	0.0179	0.8333	4.4712
[sedim, çıktı]	→	crp	0.0179	1.0000	4.2923
[hasta, sedim]	→	crp	0.0179	1.0000	4.2923
[sedim, hastane]	→	crp	0.0179	1.0000	4.2923
[sedim, yap]	→	crp	0.0179	1.0000	4.2923
[sedim, yorum]	→	crp	0.0179	1.0000	4.2923
[sedim, emar]	→	crp	0.0179	1.0000	4.2923
[sedim, kan, ağrı]	→	crp	0.0179	1.0000	4.2923
[sedim, sonuç, ağrı]	→	crp	0.0179	1.0000	4.2923
[sedim, kan, sonuç]	→	crp	0.0179	1.0000	4.2923
[sedim, sonuç, tahlil]	→	crp	0.0179	1.0000	4.2923
[sedim, değer]	→	crp	0.0287	1.0000	4.2923
[sedim, kan, tahlil]	→	crp	0.0287	1.0000	4.2923
[sedim, sonuç]	→	crp	0.0358	1.0000	4.2923
[sedim, tahlil]	→	crp	0.0466	1.0000	4.2923
[crp, sedim, kan]	→	tahlil	0.0287	0.7273	3.9786
[emar, yorum]	→	yap	0.0215	0.7500	3.9481
[sedim, kan]	→	crp	0.0394	0.9167	3.9346
[crp, kan, ağrı]	→	sonuç	0.0179	0.7143	3.9076
[negatif, ağrı]	→	doktor	0.0179	0.8333	3.8750
[sedim, ağrı]	→	crp	0.0323	0.9000	3.8631
[hasta, ilaç]	→	kan	0.0179	0.7143	3.8324
[emar, bel]	→	ağrı	0.0179	0.7143	3.8324
[emar, bulgu]	→	ağrı	0.0179	0.7143	3.8324
[emar, çıktı]	→	ağrı	0.0179	0.7143	3.8324
[kan, sonuç, ağrı]	→	crp	0.0179	0.8333	3.5769
[kalça, ağrı]	→	emar	0.0251	0.8750	3.5380
[yap, yorum]	→	emar	0.0215	0.8571	3.4658
[bulgu, ağrı]	→	doktor	0.0179	0.7143	3.3214
[hastane, ağrı]	→	crp	0.0179	0.7143	3.0659

Çizelge 10.28. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[değer, tahlil]	→	crp	0.0179	0.7143	3.0659
[hastane, ağrı]	→	emar	0.0179	0.7143	2.8882
[bulgu, ağrı]	→	emar	0.0179	0.7143	2.8882
[bel, ağrı]	→	emar	0.0179	0.7143	2.8882

10.9.2. Tahlil yorum sonuçları

Toplam 5217 yorum ön işlem aşamasından geçirilmiştir. Yorumları temsil eden 2850 adet kelime kökü bulunmuştur. Frekans değeri 10'dan aşağı olan kelimeler elenmiştir. Ayrıca belli başına anlam ifade etmeyen durak kelimeleri ve gereksiz kelimeler silinmiştir. Bu işlemler sonrasında yorumları temsil eden 207 kelime kökü belirlenmiştir. Yıllara göre en çok kullanılan 50 kelime Çizelge 10.29'da verilmiştir.

Çizelge 10.29. Tahlil kategorisine ait yıllara göre yorumlarda en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	iyi, yap, ağrı, kullan, doktor, hastalık, ilaç, git, değil, sonuç, çık, sedim, tedavi, crp, normal, yüksek, hastane, başla, çocuk, değer, etki, yardım, çıkar, moral, kötü, eklem, boyun, randevu, negatif, gör, hasta, tahlil, pozitif, insan, yüzde, mide, göz, hoca, devam, acil, sıkıntı, yaş, emar, fazla, ppd, ameliyat, çıktı, doktora, kemik, iğne
2013	yap, iyi, ağrı, ilaç, hastalık, değil, kullan, doktor, git, tedavi, hasta, sonuç, hastane, başla, teşhis, emar, kan, çık, hlab27, testi, değer, gör, tahlil, hoca, devam, negatif, pozitif, ppd, verem, crp, normal, sıkıntı, etki, moral, doktora, yaş, fizik, spor, film, sedim, yüksek, çocuk, çıkar, acil, iğne, rapor, bilgi, egzersiz, yardım, fazla
2014	kullan, doktor, yap, ağrı, git, ilaç, iyi, kan, negatif, değil, emar, testi, başla, hastalık, normal, pozitif, teşhis, etki, sonuç, hasta, kontrol, ppd, tedavi, hareket, bel, çık, tahlil, verem, crp, doktora, gör, sedim, yeni, tanı, hoca, kol, değer, muayene, çıktı, insan, sakroiliak, hastane, yüksek, kötü, kanser, aşı, sıkıntı, spor, acil, dönem
2015	ilaç, kullan, yap, doktor, ağrı, iyi, emar, sonuç, başla, kan, hastalık, verem, git, çıktı, testi, bel, çık, hasta, hastane, sol, değil, crp, negatif, kalça, normal, tedavi, tahlil, muayene, kötü, pozitif, teşhis, sakroiliak, sıcak, sistem, fıtık, ppd, gör, yüksek, ilaçla, boyun, ayak, atak, tüberküloz, vitamin, etki, doktora, kafa, yorum, moral, yardım
2016	ağrı, doktor, kullan, iyi, yap, hastalık, git, ilaç, crp, normal, başla, emar, sedim, çık, değil, yüksek, sonuç, hasta, negatif, egzersiz, çıktı, kan, hastane, spor, acil, tahlil, teşhis, tedavi, gör, devam, tansiyon, testi, pozitif, değer, tutu, yeni, boyun, uyu, tavsiye, etki, temiz, iltihap, azal, düşük, doktora, uzman, iğne, hoca, yürü, genetik

Çizelge 10.29. (devam)

Yıl	Kelimeler
2017	emar , ağrı, yap, git, doktor, ilaç, hastalık, sonuç, kullan, iyi, değil, normal, teşhis, tedavi, testi, kalça, crp, pozitif, başla, sedim, çık, gör, yüksek, negatif, kan, eklem, çıktı, besle, hastane, hasta, değer, doktora, genetik, hareket, tahlil, tutu, tanı, sakroiliak, yaş, kriter, etki, çıkar, şikayet, fazla, devam, boyun, iltihap, iğne, romatizma, kemik
2018	ağrı, yap, kullan, emar , doktor, crp, ilaç, iyi, hastalık, değil, sedim, yüksek, git, normal, bel, teşhis, tedavi, değer, sonuç, çık, kalça, kan, hasta, hastane, başla, eklem, tahlil, çıktı, gör, negatif, fazla, ayak, kemik, bilgi, tanı, sakroiliak, şikayet, iltihap, fayda, testi, pozitif, hoca, yağ, fitik, etki, vücut, özel, düşük, hareket, romatizma

Belirlenen 207 anahtar kelimeyi frekans değerlerine göre göstermek amacıyla kelime bulutu oluşturulmuştur. Şekil 10.47'deki kelime bulutuna göre 'ağrı', 'emar', 'kullan', 'iyi', 'yap' gibi kelimelerin tahlil kategorisine ait yorumlarda sıklıkla geçtiği görülmektedir.



Şekil 10.43. Tahlil kategorisine ait yorumlar için bulunan toplam 207 kelimeyi gösteren kelime bulutu

Tahlil kategorisine ait yorumlar için birliktelik kuralı sonuçları

Gönderiler için yapılan toplam yorum sayısı 5217'dir. Bulunan anahtar kelimelerinden 2'den aşağı kelime barındıran yorumlar elenmiştir. Analizde toplam 2316 yorum kullanılmıştır. Birliktelik kurallarını oluşturmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.002 ve minimum güven değeri 0.5 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 71 kural bulunmuştur. Bu kurallar Çizelge 10.30 ve Çizelge 10.31'de verilmiştir. Çizelgelerde verilen kurallar Lif değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.30'da 177.6687 lift değeri ile en ilginç kural 'sistem → bağışık' kuralıdır. 'sistem' ve 'bağışık' kelimesi tüm gönderilerin %0.21'inde birlikte kullanılmıştır. Ayrıca 'sistem' kelimesinin kullanıldığı gönderilerin %57.89'unda 'bağışık' kelimesi de kullanılmaktadır.

Lift değeri 71.9847 olan 'savaş → verem' kuralında, 'savaş' ve 'verem' kelimesi tüm gönderilerin %0.36'sında birlikte kullanılmıştır. Ayrıca 'savaş' kelimesinin kullanıldığı gönderilerin %70.37'sinde 'verem' kelimesi de kullanılmaktadır.

Lift değeri 36.2292 olan 'üniversite → hastane' kuralında, 'üniversite' ve 'hastane' kelimesi tüm gönderilerin %0.23'ünde birlikte kullanılmıştır. Ayrıca 'üniversite' kelimesinin kullanıldığı gönderilerin %75'inde 'hastane' kelimesi de kullanılmıştır.

Lift değeri 32.2037 olan 'devlet → hastane' kuralında, 'devlet' ve 'hastane' kelimesi tüm gönderilerin %0.23'ünde birlikte kullanılmıştır. Ayrıca 'devlet' kelimesinin kullanıldığı gönderilerin %66.67'sinde 'hastane' kelimesi de kullanılmıştır.

Çizelge 10.30. Tahlil kategorisine ait yorumlar için ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[sistem]	→ bağışık	0.0021	0.5789	177.6687
[bağışık]	→ sistem	0.0021	0.6471	177.6687
[savaş]	→ verem	0.0036	0.7037	71.9847
[üniversite]	→ hastane	0.0023	0.7500	36.2292
[devlet]	→ hastane	0.0023	0.6667	32.2037
[fitık]	→ bel	0.0044	0.5610	31.8110
[ppd]	→ testi	0.0058	0.5263	25.6616
[gen]	→ testi	0.0031	0.5161	25.1649
[sıcak]	→ hasta	0.0021	0.5789	24.9617
[crp]	→ sedim	0.0215	0.5957	21.5833
[sedim]	→ crp	0.0215	0.7778	21.5833
[fizik]	→ tedavi	0.0040	0.5385	19.7828
[doktora]	→ git	0.0069	0.5294	12.8462
[temiz]	→ emar	0.0048	0.6410	12.2499

Çizelge 10.30. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[röntgen]	→	emar	0.0036	0.6129	11.7125
[şiddet]	→	ağrı	0.0031	0.8000	11.5933
[sırt]	→	ağrı	0.0042	0.7333	10.6272
[sakroiliak]	→	emar	0.0059	0.5167	9.8734
[yorum]	→	yap	0.0031	0.6154	9.4704
[tutuk]	→	ağrı	0.0029	0.6000	8.6950
[azal]	→	ağrı	0.0038	0.5882	8.5245
[bacak]	→	ağrı	0.0027	0.5833	8.4535
[kalça]	→	ağrı	0.0088	0.5349	7.7514
[bel]	→	ağrı	0.0092	0.5217	7.5609

Çizelge 10.31’de üç ve üzeri birliktelik kurallarına göre Lift değeri 55.9327 olan ‘testi, verem → ppd’ kuralında ‘testi, verem’ kelime grubu ve ‘ppd’ kelimesi tüm gönderilerin %0.21’inde birlikte kullanılmıştır. Ayrıca ‘testi, verem’ kelimelerinin kullanıldığı gönderilerin %61.11’inde ‘ppd’ kelimesi de kullanılmaktadır.

Lift değeri 52.17 olan ‘emar, eklem→ sakroiliak’ kuralında ‘emar, eklem’ kelime grubu ve ‘sakroiliak’ kelimesi tüm gönderilerin %0.23’ünde birlikte kullanılmıştır. Ayrıca ‘emar, eklem’ kelimelerinin kullanıldığı gönderilerin %60’sında ‘sakroiliak’ kelimesi de kullanılmaktadır.

Lift değeri 38.9857 olan ‘fitık, ağrı→ bel’ kuralında ‘fitık, ağrı’ kelime grubu ve ‘bel’ kelimesi tüm gönderilerin %0.21’inde birlikte kullanılmıştır. Ayrıca ‘fitık, ağrı’ kelimelerinin kullanıldığı gönderilerin %68.75’inde ‘bel’ kelimesi de kullanılmaktadır.

Çizelge 10.31. Tahlil kategorisine ait yorumlar için üç ve üzeri kullanımları gösteren birliktelik kuralları

Gruplar		Sonuç	Destek	Güven	Lift
[testi, verem]	→	ppd	0.0021	0.6111	55.9327
[emar, eklem]	→	sakroiliak	0.0023	0.6000	52.1700
[fitık, ağrı]	→	bel	0.0021	0.6875	38.9857
[negatif, testi]	→	pozitif	0.0023	0.6316	37.8730

Çizelge 10.31. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[crp, normal, yüksek]	→	sedim	0.0021	0.9167	33.2101
[verem, ppd]	→	testi	0.0021	0.6471	31.5487
[crp, değil]	→	yüksek	0.0021	0.7333	30.8532
[ppd, yap]	→	testi	0.0023	0.6000	29.2542
[testi, pozitif]	→	negatif	0.0023	0.6000	28.4564
[sirt, ağrı]	→	bel	0.0021	0.5000	28.3533
[crp, normal]	→	sedim	0.0054	0.7778	28.1782
[sedim, yüksek, ağrı]	→	crp	0.0023	1.0000	27.7500
[sedim, emar]	→	crp	0.0036	0.9500	26.3625
[sedim, ağrı]	→	crp	0.0052	0.9310	25.8362
[crp, kullan]	→	sedim	0.0023	0.7059	25.5735
[normal, sedim, yüksek]	→	crp	0.0021	0.9167	25.4375
[normal, sedim]	→	crp	0.0054	0.9032	25.0645
[crp, emar]	→	sedim	0.0036	0.6786	24.5841
[crp, değer]	→	sedim	0.0027	0.6667	24.1528
[sedim, kullan]	→	crp	0.0023	0.8571	23.7857
[sedim, değer]	→	crp	0.0027	0.8235	22.8529
[crp, kan]	→	sedim	0.0021	0.6111	22.1400
[sedim, yüksek]	→	crp	0.0063	0.7857	21.8036
[normal, yüksek]	→	sedim	0.0023	0.6000	21.7375
[crp, yüksek]	→	sedim	0.0063	0.6000	21.7375
[crp, ağrı]	→	sedim	0.0052	0.5625	20.3789
[crp, yüksek, ağrı]	→	sedim	0.0023	0.5455	19.7614
[doktor, doktora]	→	git	0.0025	0.7647	18.5557
[değil, yüksek]	→	crp	0.0021	0.6471	17.9559
[sedim, kan]	→	crp	0.0021	0.6471	17.9559
[normal, yüksek]	→	crp	0.0023	0.6000	16.6500
[yüksek, ağrı]	→	crp	0.0042	0.5789	16.0658
[git, kullan]	→	ilaç	0.0029	0.7895	15.1981
[bel, sirt]	→	ağrı	0.0021	1.0000	14.4917
[doktor, ağrı, kullan]	→	ilaç	0.0023	0.6667	12.8339

Çizelge 10.31. (devam)

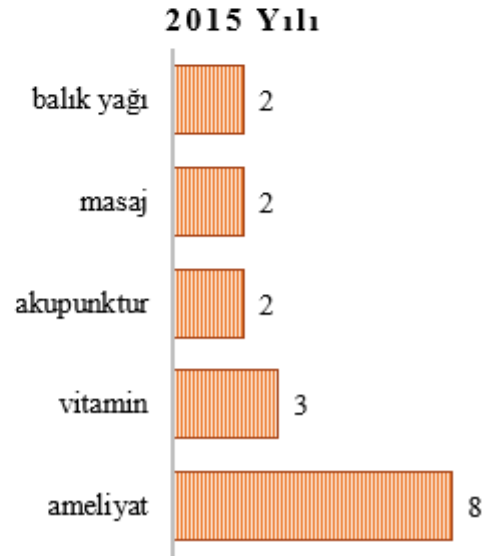
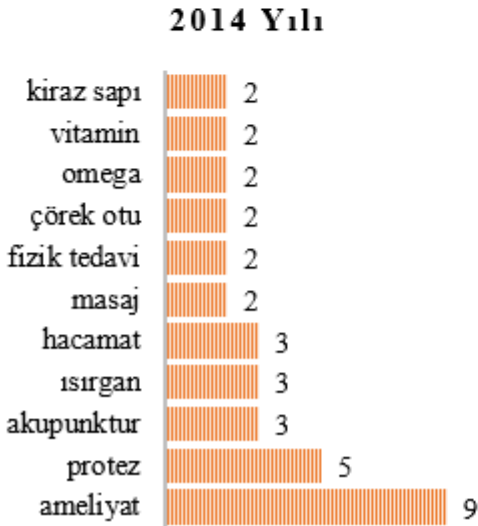
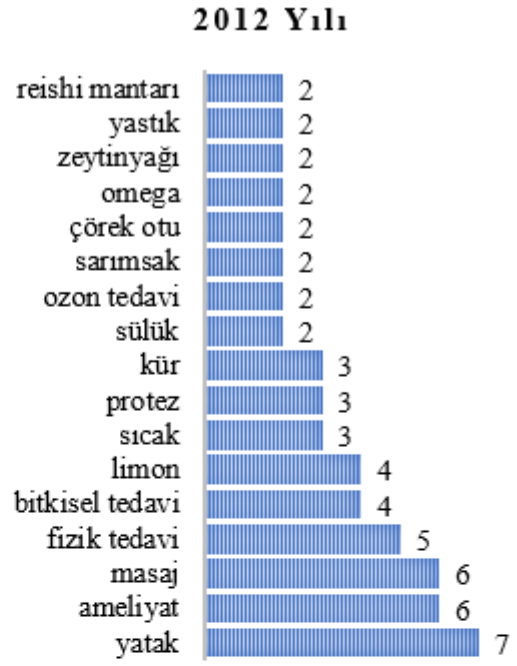
Gruplar		Sonuç	Destek	Güven	Lift
[kalça, bel]	→	ağrı	0.0033	0.7727	11.1981
[hastalık, kullan]	→	ilaç	0.0023	0.5455	10.5005
[doktor, ilaç, ağrı]	→	kullan	0.0023	0.5714	10.3512
[git, sonuç]	→	doktor	0.0023	0.5714	10.1399
[git, emar]	→	doktor	0.0033	0.5667	10.0554
[sakroiliak, eklem]	→	emar	0.0023	0.5217	9.9704
[başla, kullan]	→	ilaç	0.0040	0.5122	9.8602
[ilaç, ağrı]	→	kullan	0.0052	0.5294	9.5901
[ilaç, hastalık]	→	kullan	0.0023	0.5217	9.4511
[sonuç, ağrı]	→	doktor	0.0021	0.5000	8.8724
[doktor, ilaç, kullan]	→	ağrı	0.0023	0.5000	7.2458
[çık, doktor]	→	ağrı	0.0021	0.5000	7.2458

10.10. Tedavi Kategorisine Ait Sonuçlar

10.10.1. Tedavi gönderi sonuçları

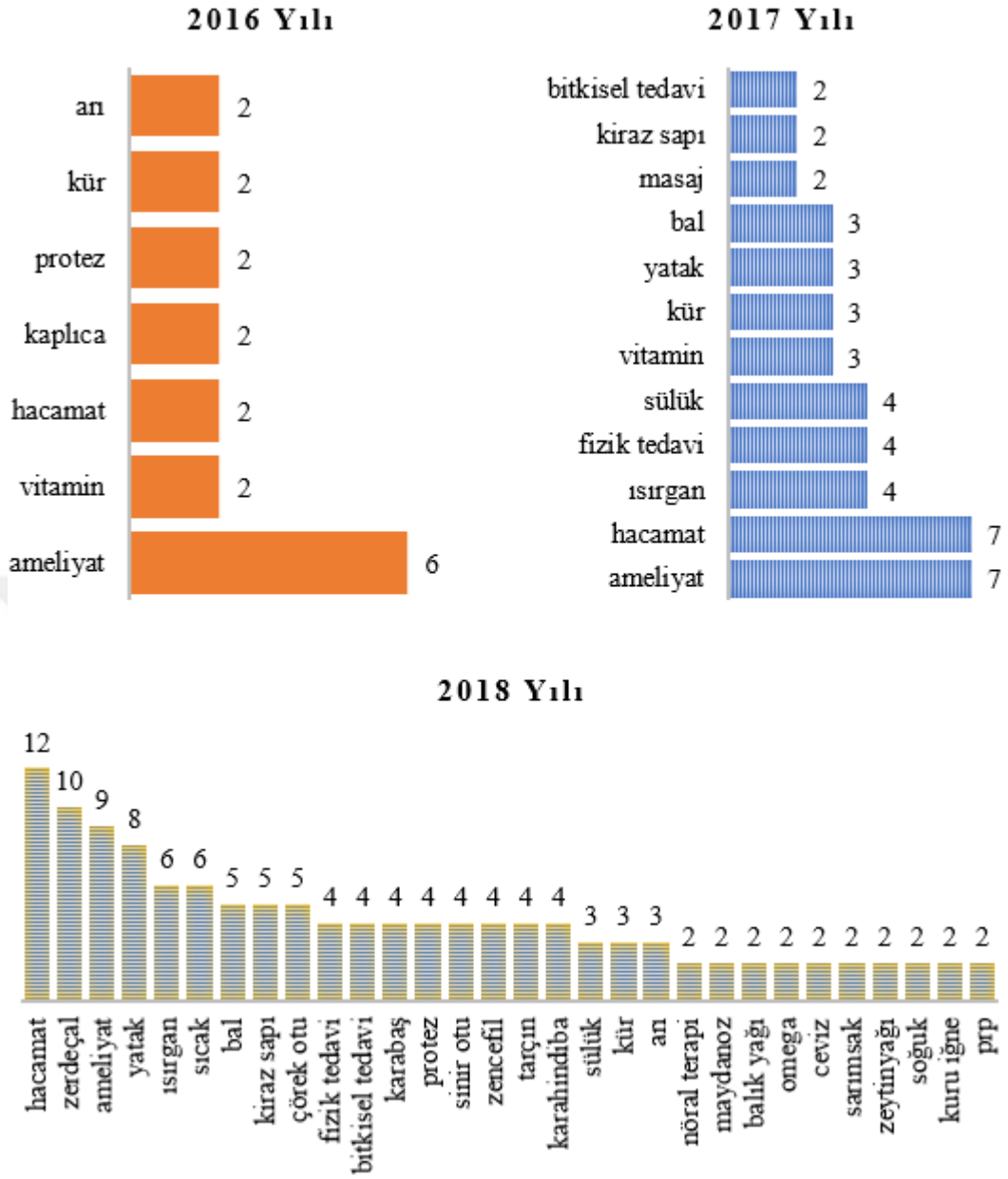
Tedavi kategorisine ait toplam 707 gönderi analiz edilmiştir. Gönderilerde toplam 87 farklı alternatif tedavi yönteminden bahsedilmiştir. Tedavi yöntemlerinin paylaşılan gönderilerde var olup olmadıklarını esas alarak binary ağırlıklandırma yöntemi ile ağırlıklandırılmıştır. Yıllara göre kadın ve erkek kullanıcıları tarafından en çok konuşulan tedavi yöntemleri grafikler ile gösterilmiştir.

Şekil 10.44, 2012, 2013, 2014 ve 2015 yıllarında erkek kullanıcılar tarafından en çok bahsedilen alternatif tedavi yöntemlerini göstermektedir. 2012 yılında kaplıca ve fizik tedavi yöntemleri ön plana çıkmaktadır. 2013 yılında yatak, ameliyat ve masaj, 2014 yılında ameliyat ve protez, 2015 yılında ise ameliyat ve vitamin yöntemleri erkek kullanıcılar tarafında sıkça bahsedilen tedavi yöntemleridir.



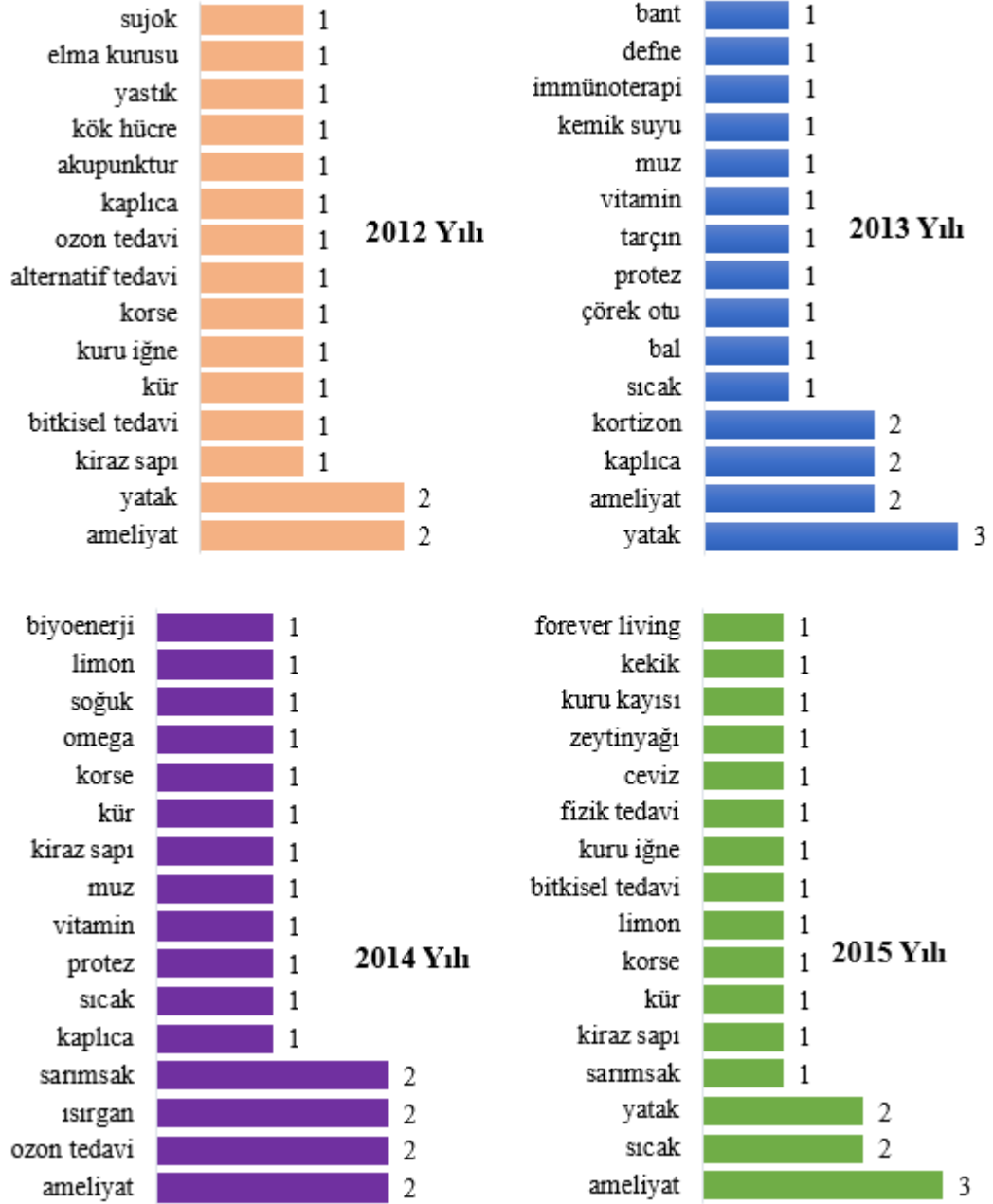
Şekil 10.44. Erkek kullanıcılar tarafından 2012, 2013, 2014 ve 2015 yıllarında en çok konuşulan alternatif tedavi yöntemleri

Şekil 10.45’de 2016 yılında ameliyat, 2017 yılında hacamat ve ameliyat, 2018 yılında ise hacamat, zerdeçal ve ameliyat erkek kullanıcılar tarafından sıkça bahsedilmektedir.



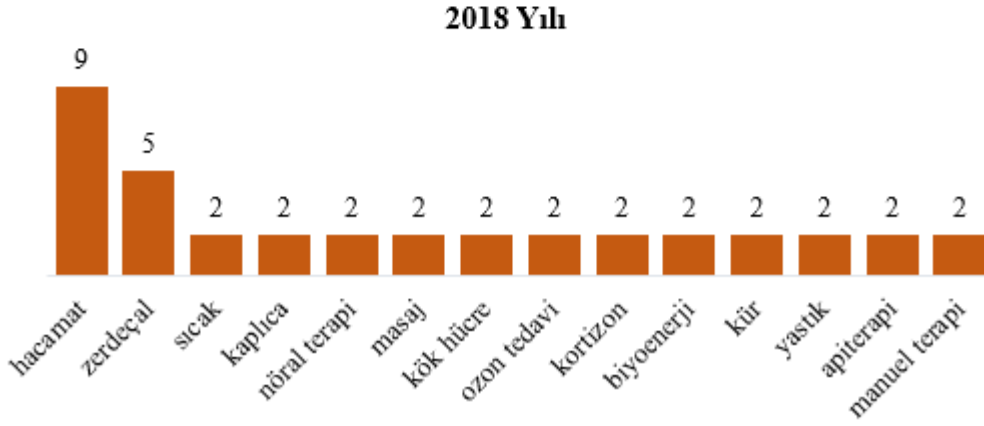
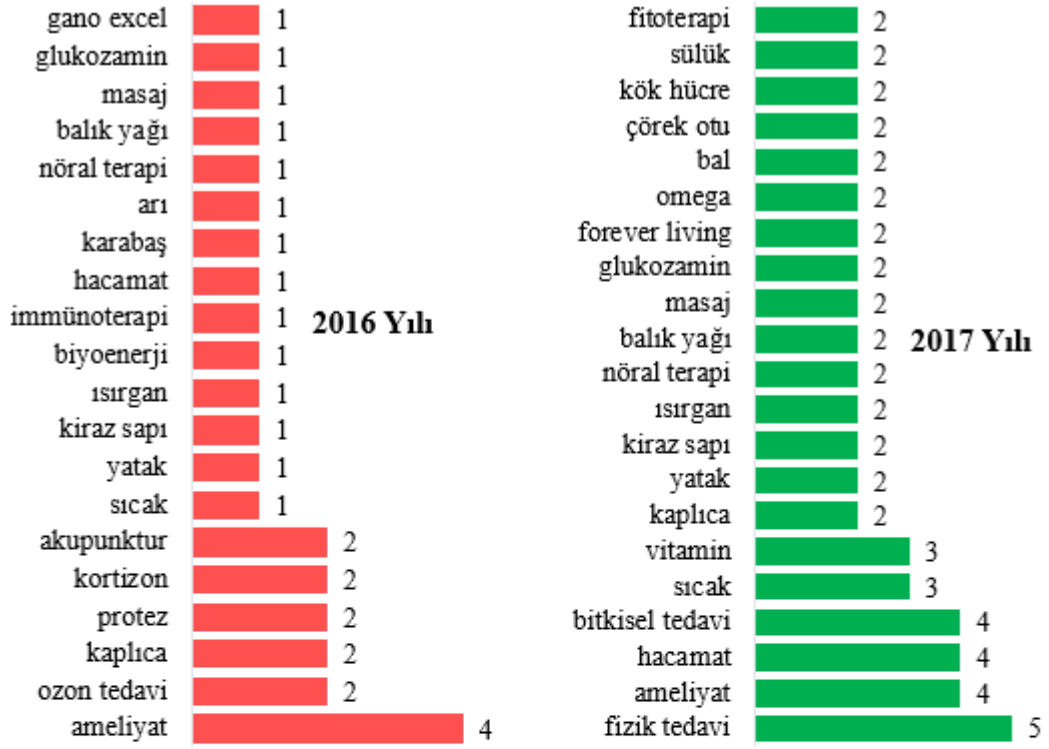
Şekil 10.45. Erkek kullanıcılar tarafından 2016, 2017 ve 2018 yıllarında en çok konuşulan alternatif tedavi yöntemleri

Şekil 10.46'da, kadın kullanıcılar tarafında 2012 yılında yatak ve ameliyat, 2013 yılında yatak, ameliyat, kaplıca ve kortizon, 2014 yılında ameliyat, ozon tedavisi, sarımsak ve ısırgan, 2015 yılında ise ameliyat, sıcak ve yatak alternatif tedavi yöntemlerinden bahsedilmektedir.



Şekil 10.46. Kadın kullanıcılar tarafından 2012, 2013, 2014 ve 2015 yıllarında en çok konuşulan alternatif tedavi yöntemleri

Şekil 10.47’de, kadın kullanıcılar tarafından 2016 yılında ameliyat, ozon tedavisi, kaplıca, protez, kortizon ve akupunktur, 2017 yılında fizik tedavisi, ameliyat, hacamat ve bitkisel tedavi, son olarak 2018 yılında ise hacamat ve zerdeçal yöntemlerinden bahsedilmektedir.



Şekil 10.47. Kadın kullanıcılar tarafından 2016, 2017 ve 2018 yıllarında en çok konuşulan alternatif tedavi yöntemleri

Şekil 10.48’de hakkında çeşitli gönderiler paylaşılan 87 farklı tedavi yöntemi kelime bulutu ile gösterilmiştir. Kelime bulutuna göre yatak, ameliyat, fizik tedavi, hacamat, ısırgan gibi çeşitli yöntemler ön plana çıkmaktadır.

2013 yılı 2018'den, 2014 yılı 2017 ve 2018'den, 2015 yılı 2017 ve 2018'den, 2016 yılı 2017 ve 2018'den farklılık göstermektedir.

Tedavi kategorisine ait gönderiler için uyum analizi sonuçları

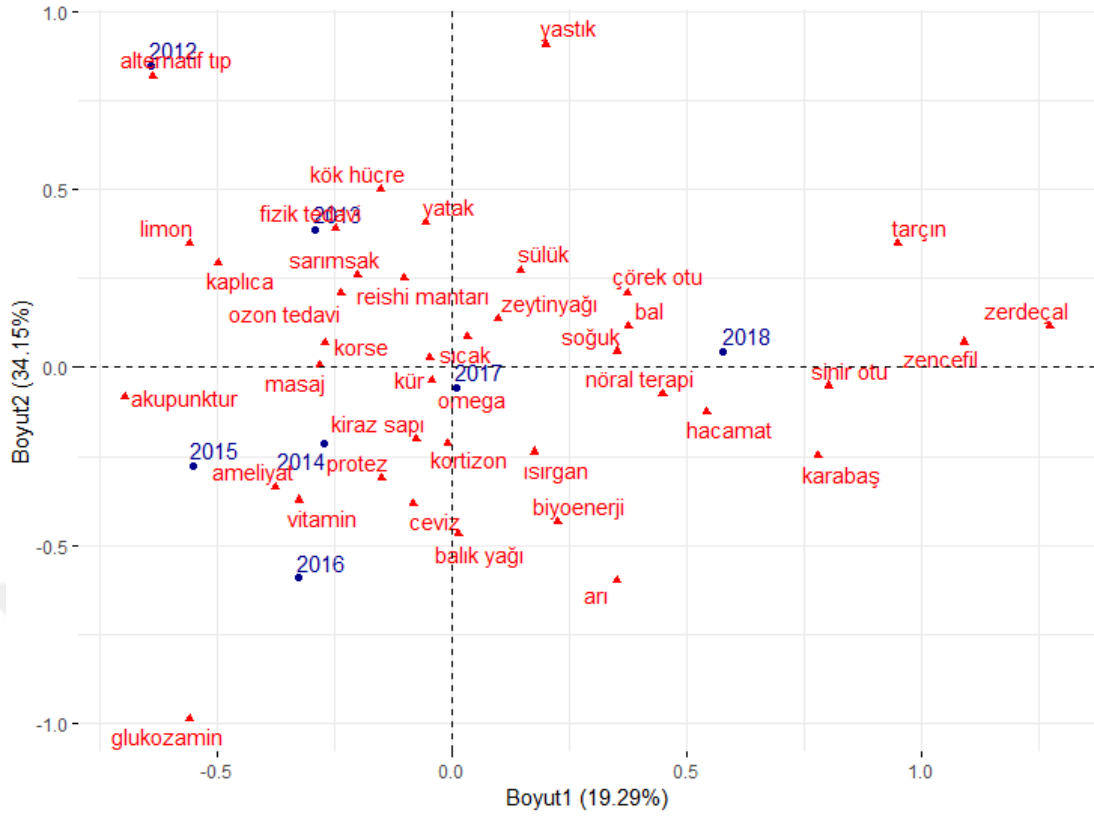
Tedavi kategorisini ait alternatif tedavi yöntemleri ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Alternatif tedavi yöntemleri ve yıllar arasında ilişki yoktur

H_1 : Alternatif tedavi yöntemleri ve yıllar arasında ilişki vardır

R programında 7×39 boyutunda bir matris oluşturulmuştur. 39, toplamda 5 ve üzeri sıklıkta geçen alternatif tedavi yöntemlerini göstermektedir. 7 ise toplam yılların sayısını göstermektedir. Test sonucunda $p < 0.05$ 'ten dolayı %95 güvenle H_0 hipotezi reddedilir. Sonuç olarak %95 güvenle alternatif tedavi yöntemleri ve yıllar arasında bir ilişki vardır.

Şekil 10.49'da basit uyum analizi sonucu verilmiştir. Boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %19.29'unu açıklarken, ikinci boyut ise %34.15'ini açıklamaktadır. Zerdeçal, zencefil, tarçın, sinir otu, karabaş, hacamat gibi yöntemlerin 2018 yılında daha yoğun görüldüğü gözükmektedir. Glukozamin 2016 yılında daha çok ön plandadır. 2014, 2015 ve 2016 yıllarında konuşulan tedavi yöntemleri benzerlik göstermektedir.



Şekil 10.49. Alternatif tedavi yöntemleri ve yıllar için basit uyum analizi sonucu

Bu boyutların alternatif tedavi yöntemlerini ve yılları hangi oranda açıkladığı Çizelge 10.33 ve Çizelge 10.34’te verilmiştir. Çizelge 10.33’te 2015 ve 2018 boyut 1, 2012 ve 2016 boyut 2, 2014 boyut 3, 2017 boyut 4 ve 2013 büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.33. Tedavi kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012	0.231	0.402	0.269	0.033	0.064
2013	0.169	0.293	0.240	0.001	0.297
2014	0.147	0.090	0.287	0.110	0.227
2015	0.398	0.100	0.010	0.029	0.069
2016	0.121	0.394	0.206	0.109	0.164
2017	0.000	0.013	0.045	0.858	0.000
2018	0.940	0.005	0.002	0.035	0.003

Çizelge 10.34’te ameliyat, hacamat, zerdeçal, akupunktur, limon, nöral terapi, korse, sinir otu, karabaş, zencefil ve tarçın yöntemleri boyut 1, vitamin, kiraz sapı, kök

hücre, balık yağı, alternatif tıp, biyoenerji, yastık ve glukozamin yöntemleri boyut 2, arı, çörek otu, kaplıca, omega, reishi mantarı, sıcak, soğuk ve zeytinyağı yöntemleri boyut 3, bal, fizik tedavi, ozon tedavi, protez, sarımsak ve sülük yöntemleri boyut 4, yatak, masaj, kür, kortizon, ısırğan ve ceviz yöntemleri ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.34. Boyutların alternatif tedavi yöntemlerini açıklama oranları

Alternatif tedaviler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
ameliyat	0.515	0.408	0.008	0.005	0.002
hacamat	0.757	0.039	0.121	0.031	0.037
yatak	0.008	0.416	0.015	0.004	0.484
fizik tedavi	0.148	0.368	0.011	0.438	0.010
sıcak	0.014	0.098	0.133	0.039	0.001
ısırğan	0.105	0.191	0.115	0.056	0.260
masaj	0.226	0.000	0.273	0.045	0.448
protez	0.027	0.114	0.140	0.515	0.044
kaplıca	0.273	0.095	0.453	0.151	0.025
vitamin	0.274	0.356	0.029	0.331	0.010
kiraz sapı	0.062	0.442	0.082	0.001	0.406
kür	0.081	0.028	0.016	0.262	0.519
zerdeçal	0.885	0.007	0.006	0.044	0.008
sülük	0.046	0.163	0.018	0.564	0.001
ozon tedavi	0.153	0.117	0.233	0.236	0.002
çörek otu	0.381	0.119	0.423	0.016	0.022
bal	0.331	0.032	0.030	0.472	0.014
akupunktur	0.606	0.008	0.055	0.172	0.158
limon	0.488	0.187	0.265	0.017	0.000
omega	0.005	0.004	0.811	0.000	0.136
kök hücre	0.043	0.474	0.176	0.000	0.299
kortizon	0.000	0.130	0.048	0.004	0.712
sarımsak	0.066	0.109	0.275	0.326	0.224
nöral terapi	0.487	0.013	0.095	0.179	0.197
balık yağı	0.000	0.484	0.009	0.053	0.128
arı	0.109	0.313	0.350	0.029	0.187

Çizelge 10.34. (devam)

Alternatif tedaviler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
alternatif tıp	0.232	0.383	0.200	0.000	0.184
zeytinyağı	0.034	0.066	0.663	0.000	0.010
korse	0.398	0.026	0.230	0.209	0.062
biyoenerji	0.112	0.418	0.076	0.029	0.001
yastık	0.036	0.742	0.001	0.141	0.050
reishi mantarı	0.016	0.098	0.377	0.191	0.112
soğuk	0.138	0.002	0.417	0.305	0.077
ceviz	0.006	0.136	0.050	0.036	0.326
sinir otu	0.817	0.003	0.014	0.022	0.133
karabaş	0.695	0.071	0.160	0.036	0.033
glukozamin	0.149	0.462	0.234	0.062	0.088
zencefil	0.952	0.004	0.017	0.000	0.010
tarçın	0.677	0.091	0.024	0.096	0.035

Tedavi kategorisine ait gönderi birliktelik analizi sonuçları

Toplam gönderi sayısı 707'dir. Bulunan tedavi yöntemlerinden 1 ve aşağısı yöntem barındıran gönderiler elenmiştir. Analizde toplam 193 gönderi kullanılmıştır. Kullanılan tedavi yöntemlerinin birliktelik kurallarını çıkarmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.01 ve minimum güven değeri 0.7 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 72 kural bulunmuştur. Bu kurallar Çizelge 10.35 ve Çizelge 10.36'da verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.35'te 32.1667 lift değeri ile en ilginç kural 'zencefil → sinir otu' kuralıdır. 'zencefil' ve 'sinir otu' kelimesi tüm gönderilerin %2.59'unda birlikte kullanılmıştır. Ayrıca 'zencefil' kelimesinin kullanıldığı gönderilerin %100'ünde 'sinir otu' kelimesi de kullanılmaktadır.

Lift değeri 24.1250 olan 'arı sütü→ arı' kuralında, 'arı sütü' ve 'arı' kelimesi tüm gönderilerin %1.04'ünde birlikte kullanılmıştır. Ayrıca 'arı sütü' kelimesinin kullanıldığı gönderilerin %100'ünde 'arı' kelimesi de kullanılmaktadır.

Lift değeri 16.0833 olan ‘maydanoz → limon’ kuralında, ‘maydanoz’ ve ‘limon’ kelimesi tüm gönderilerin %1.04’ünde birlikte kullanılmıştır. Ayrıca ‘maydanoz’ kelimesinin kullanıldığı gönderilerin %100’ünde ‘limon’ kelimesi de kullanılmıştır.

Lift değeri 14.8462 olan ‘tarçın → bal’ kuralında, ‘tarçın’ ve ‘bal’ kelimesi tüm gönderilerin %2.59’unda birlikte kullanılmıştır. Ayrıca ‘tarçın’ kelimesinin kullanıldığı gönderilerin %100’ünde ‘bal’ kelimesi de kullanılmıştır.

Çizelge 10.35. Tedavi kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[zencefil]	→ sinir otu	0.0259	1.0000	32.1667
[sinir otu]	→ zencefil	0.0259	0.8333	32.1667
[arı sütü]	→ arı	0.0104	1.0000	24.1250
[maydanoz]	→ limon	0.0104	1.0000	16.0833
[tarçın]	→ bal	0.0259	1.0000	14.8462
[sarımsak]	→ limon	0.0363	0.8750	14.0729
[kantaron yağı]	→ kür	0.0104	1.0000	12.8667
[zerdeçal]	→ bal	0.0259	0.7143	10.6044
[kiraz sapı]	→ ısırgan	0.0777	0.8824	8.9628
[ısırgan]	→ kiraz sapı	0.0777	0.7895	8.9628
[soğuk]	→ sıcak	0.0259	0.8333	8.4649
[kaplıca]	→ sıcak	0.0259	0.7143	7.2556
[protez]	→ ameliyat	0.0622	0.9231	7.1262

Çizelge 10.36’da üç ve üzeri birliktelik kurallarına göre Lift değeri 96.5 olan ‘balık yağı, kök hücre, omega → magnezyum’ kuralında ‘balık yağı, kök hücre, omega’ kelime grubu ve ‘magnezyum’ kelimesi tüm gönderilerin %1.04’ünde birlikte kullanılmıştır. Ayrıca ‘balık yağı, kök hücre, omega’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘magnezyum’ kelimesi de kullanılmaktadır.

Lift değeri 64.3333 olan ‘karahindiba, zerdeçal, hacamat, zencefil, sinir otu, bal, tarçın → kuru iğne’ kuralında ‘karahindiba, zerdeçal, hacamat, zencefil, sinir otu, bal, tarçın’ kelime grubu ve ‘kuru iğne’ kelimesi tüm gönderilerin %1.04’ünde birlikte kullanılmıştır. Ayrıca ‘karahindiba, zerdeçal, hacamat, zencefil, sinir otu, bal, tarçın’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘kuru iğne’ kelimesi de kullanılmaktadır.

Lift değeri 48.25 olan ‘zerdeçal, ceviz, zencefil, sinir otu, bal, tarçın → karahindiba’ kuralında ‘zerdeçal, ceviz, zencefil, sinir otu, bal, tarçın’ kelime grubu ve ‘karahindiba’ kelimesi tüm gönderilerin %1.04’ünde birlikte kullanılmıştır. Ayrıca ‘zerdeçal, ceviz, zencefil, sinir otu, bal, tarçın’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘karahindiba’ kelimesi de kullanılmaktadır.

Çizelge 10.36. Tedavi kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[balık yağı, kök hücre, omega, vitamin] →	magnezyum	0.0104	1.0000	96.5000
[karahindiba, zerdeçal, hacamat, zencefil, sinir otu, bal, tarçın] →	kuru iğne	0.0104	1.0000	64.3333
[zerdeçal, ceviz, zencefil, sinir otu, bal, tarçın] →	karahindiba	0.0104	1.0000	48.2500
[zerdeçal, hacamat, kuru iğne, zencefil, sinir otu, bal, tarçın] →	karahindiba	0.0104	1.0000	48.2500
[zerdeçal, zencefil, sinir otu, bal, tarçın] →	karahindiba	0.0207	1.0000	48.2500
[karahindiba, zerdeçal, ceviz, zencefil, sinir otu, bal] →	tarçın	0.0104	1.0000	38.6000
[karahindiba, zerdeçal, ceviz, sinir otu, bal, tarçın] →	zencefil	0.0104	1.0000	38.6000
[karahindiba, zerdeçal, hacamat, kuru iğne, zencefil, sinir otu, bal] →	tarçın	0.0104	1.0000	38.6000
[karahindiba, zerdeçal, hacamat, kuru iğne, sinir otu, bal, tarçın] →	zencefil	0.0104	1.0000	38.6000
[karahindiba, zerdeçal, zencefil, sinir otu, bal] →	tarçın	0.0207	1.0000	38.6000
[karahindiba, zerdeçal, sinir otu, bal, tarçın] →	zencefil	0.0207	1.0000	38.6000

Çizelge 10.36. (devam)

Gruplar	Sonuç	Destek	Güven	Lift
[karahindiba, zerdeçal, ceviz, zencefil, bal, tarçın]	→ sinir otu	0.0104	1.0000	32.1667
[karahindiba, zerdeçal, hacamat, kuru iğne, zencefil, bal, tarçın]	→ sinir otu	0.0104	1.0000	32.1667
[karahindiba, zerdeçal, zencefil, bal, tarçın]	→ sinir otu	0.0207	1.0000	32.1667
[kemik suyu, çörek otu]	→ balık yağı	0.0104	1.0000	32.1667
[balık yağı, çörek otu, vitamin]	→ omega	0.0104	1.0000	32.1667
[zeytinyağı, ceviz, omega, vitamin]	→ sinir otu	0.0104	1.0000	32.1667
[zeytinyağı, sinir otu, omega, vitamin]	→ ceviz	0.0104	1.0000	32.1667
[zeytinyağı, ceviz, sinir otu, vitamin]	→ omega	0.0104	1.0000	32.1667
[balık yağı, vitamin]	→ omega	0.0155	1.0000	32.1667
[çörek otu, kök hücre]	→ zeytinyağı	0.0104	1.0000	32.1667
[çörek otu, sıcak]	→ soğuk	0.0104	1.0000	32.1667
[çörek otu, omega, vitamin]	→ balık yağı	0.0104	1.0000	32.1667
[magnezyum, kök hücre, omega, vitamin]	→ balık yağı	0.0104	1.0000	32.1667
[magnezyum, balık yağı, kök hücre, vitamin]	→ omega	0.0104	1.0000	32.1667
[limon, sıcak]	→ zeytinyağı	0.0104	1.0000	32.1667
[ceviz, sinir otu, omega, vitamin]	→ zeytinyağı	0.0104	1.0000	32.1667
[ceviz, zencefil]	→ sinir otu	0.0155	1.0000	32.1667
[ceviz, sinir otu]	→ zencefil	0.0155	0.7500	28.9500
[karahindiba, ceviz, zencefil, sinir otu, bal, tarçın]	→ zerdeçal	0.0104	1.0000	27.5714
[karahindiba, hacamat, kuru iğne, zencefil, sinir otu, bal, tarçın]	→ zerdeçal	0.0104	1.0000	27.5714

Çizelge 10.36. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[karahindiba, zencefil, sinir otu, bal, tarçın]	→	zerdeçal	0.0207	1.0000	27.5714
[omega, vitamin]	→	balık yağı	0.0155	0.7500	24.1250
[kür, limon]	→	sarımsak	0.0207	1.0000	24.1250
[kemik suyu, balık yağı]	→	çörek otu	0.0104	1.0000	19.3000
[zeytinyağı, kök hücre]	→	çörek otu	0.0104	1.0000	19.3000
[zeytinyağı, çörek otu]	→	kök hücre	0.0104	1.0000	19.3000
[balık yağı, çörek otu, omega]	→	vitamin	0.0104	1.0000	19.3000
[zeytinyağı, ceviz, sinir otu, omega]	→	vitamin	0.0104	1.0000	19.3000
[balık yağı, omega]	→	vitamin	0.0155	1.0000	19.3000
[magnezyum, balık yağı, omega, vitamin]	→	kök hücre	0.0104	1.0000	19.3000
[magnezyum, balık yağı, kök hücre, omega]	→	vitamin	0.0104	1.0000	19.3000
[zeytinyağı, sarımsak]	→	limon	0.0104	1.0000	16.0833
[zeytinyağı, sıcak]	→	limon	0.0104	1.0000	16.0833
[kür, sarımsak]	→	limon	0.0207	1.0000	16.0833
[karahindiba, zerdeçal, ceviz, zencefil, sinir otu, tarçın]	→	bal	0.0104	1.0000	14.8462
[karahindiba, zerdeçal, hacamat, kuru iğne, zencefil, sinir otu, tarçın]	→	bal	0.0104	1.0000	14.8462
[karahindiba, zerdeçal, zencefil, sinir otu, tarçın]	→	bal	0.0207	1.0000	14.8462
[karahindiba, zerdeçal, kuru iğne, zencefil, sinir otu, bal, tarçın]	→	hacamat	0.0104	1.0000	12.8667
[ısrırgan, sıcak]	→	kiraz sapı	0.0104	1.0000	11.3529
[ısrırgan, çörek otu]	→	kiraz sapı	0.0155	1.0000	11.3529
[ısrırgan, karabaş]	→	kiraz sapı	0.0259	1.0000	11.3529
[kiraz sapı, sıcak]	→	ısrırgan	0.0104	1.0000	10.1579
[kiraz sapı, karabaş]	→	ısrırgan	0.0259	1.0000	10.1579

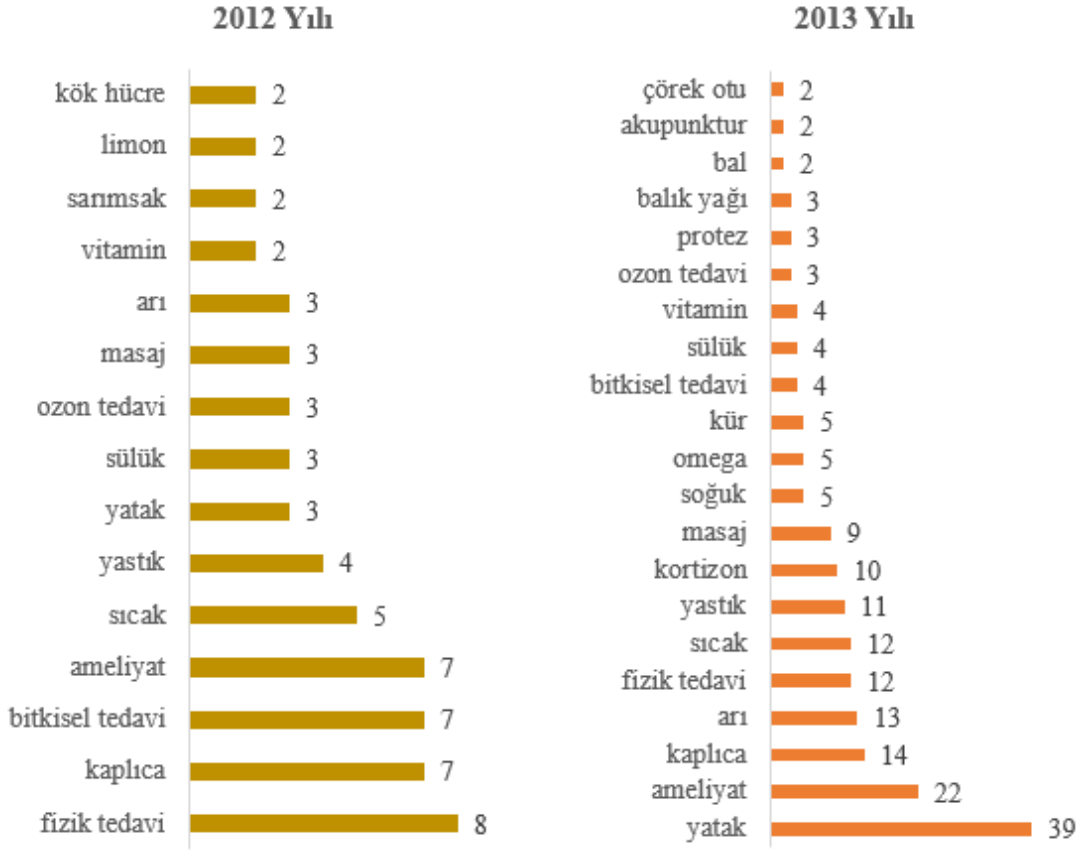
Çizelge 10.36. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[çörek otu, kiraz sapı]	→	ısırgan	0.0155	1.0000	10.1579
[soğuk, çörek otu]	→	sıcak	0.0104	1.0000	10.1579
[soğuk, kaplıca]	→	sıcak	0.0104	1.0000	10.1579
[kür, fizik tedavi]	→	sıcak	0.0104	1.0000	10.1579
[ısırgan, nöral terapi]	→	fizik tedavi	0.0104	1.0000	7.7200

10.10.2. Tedavi yorum sonuçları

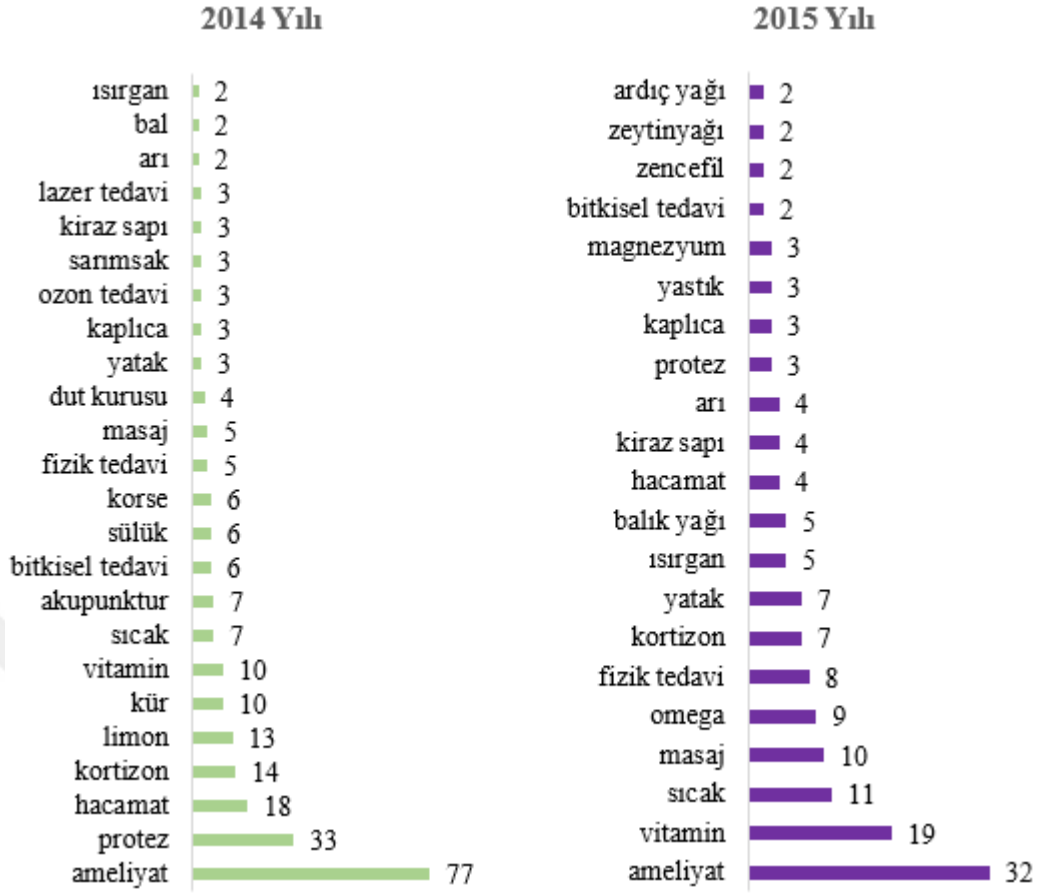
Toplam 13262 yorum içerisinde en çok konuşulan tedavi yöntemleri incelenmiştir. Yıllara göre kullanıcılar tarafından en çok yorum yapılan alternatif tedavi yöntemleri grafikler ile gösterilmiştir.

Şekil 10.50'ye göre 2012 yılında fizik tedavi, kaplıca, bitkisel tedavi ve ameliyat yöntemleri kullanıcılar tarafından en çok konuşulan yöntemlerdir. Ayrıca yatak ve ameliyat yöntemleri 2013 yılında yorumlar içerisinde sıkça kullanılmaktadır.



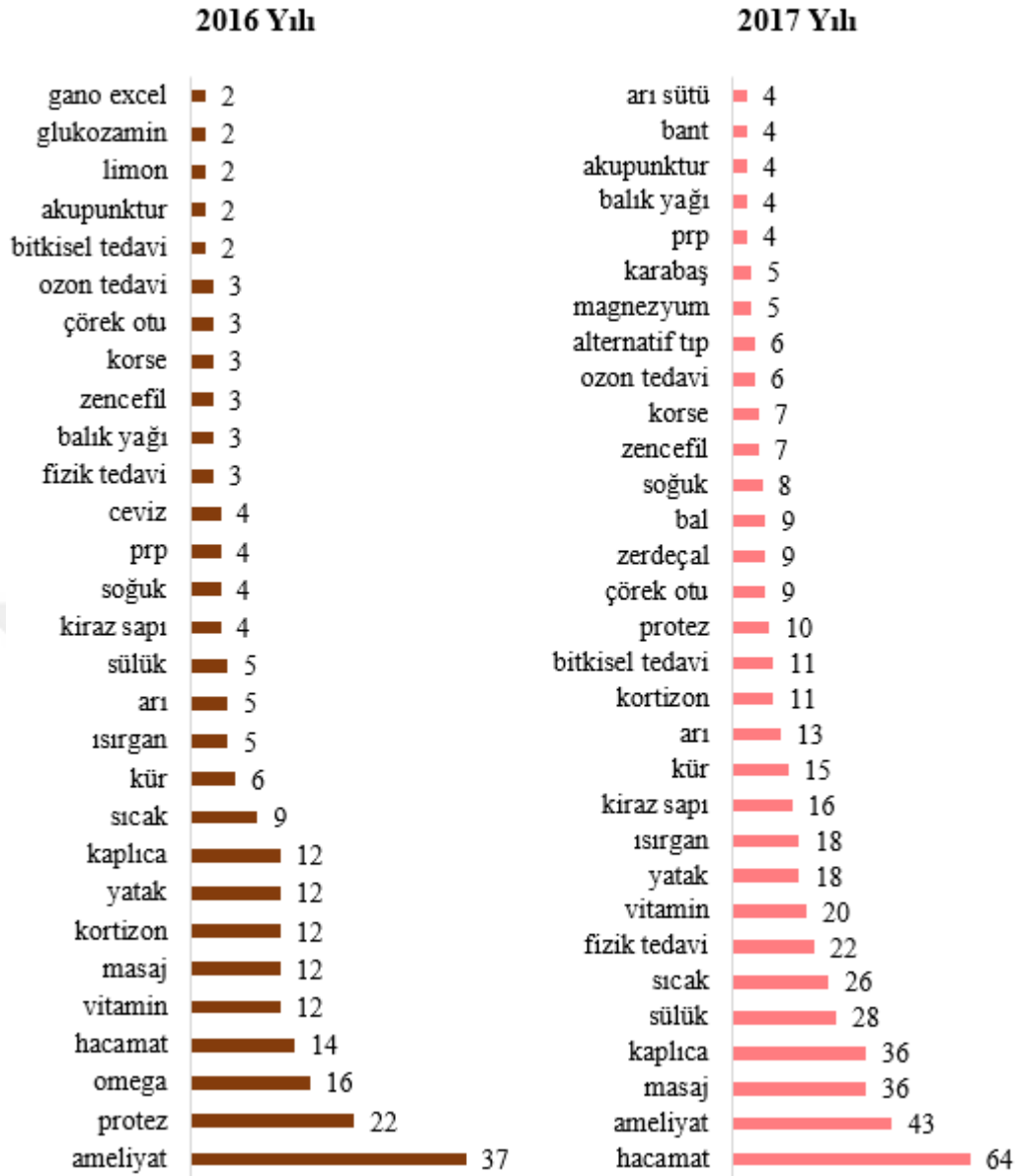
Şekil 10.50. 2012 ve 2013 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri

Şekil 10.51’de 2014 ve 2015 yıllarında kullanıcılar tarafından konuşulan alternatif tedavi yöntemleri grafiklerle gösterilmiştir. 2014 yılı içerisinde ameliyat ve protez gibi yöntemler sıkça dile getirilirken 2015 yılında ise ameliyat ve vitaminlerden sıkça bahsedilmektedir.



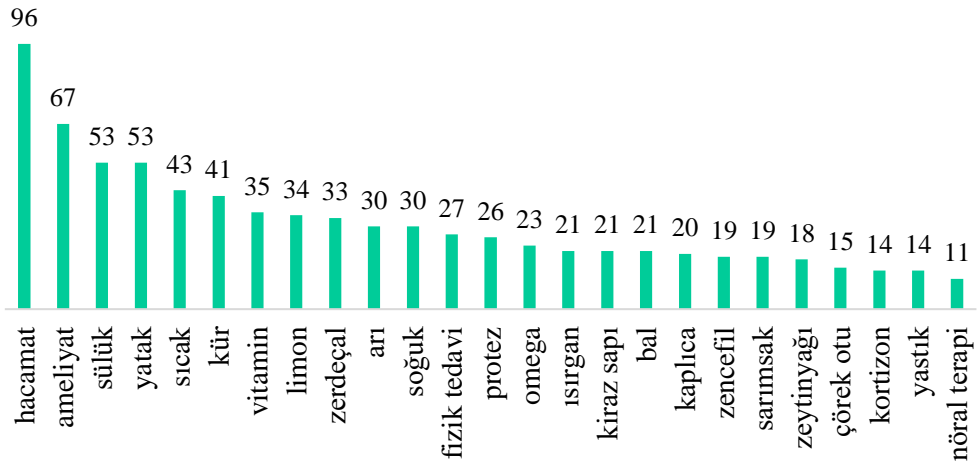
Şekil 10.51. 2014 ve 2015 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri

Şekil 10.52'ye göre 2016 yılında ameliyat ve protez, 2017 yılında ise hacamat ve ameliyat kullanıcılar tarafında sıkça konuşulan alternatif tedavi yöntemleridir.



Şekil 10.52. 2016 ve 2017 yıllarında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri

Şekil 10.53'te 2018 yılında kullanıcılar tarafından en çok konuşulan yöntemler ameliyat, protez ve sülük yöntemleridir.



Şekil 10.53. 2018 yılında kullanıcılar tarafından yorumlarda en çok konuşulan tedavi yöntemleri

Tedavi kategorisine ait yorumlarda geçen alternatif tedavi yöntemleri için birliktelik analizi sonuçları

Toplam yorum sayısı 13262'dir. Yorum verilerinde 2'den aşağı tedavi yöntemi barındıran yorumlar elenmiştir. Analizde toplam 559 yorum kullanılmıştır. Yorumlarda birlikte konuşulan ve kullanılan yöntemleri belirlemek amacıyla Apriori algoritması uygulanmıştır. Minimum destek değeri 0.01 ve minimum güven değeri 0.6 olarak tanımlanmıştır. Algoritma sonucunda toplam 17 kural bulunmuştur. Bulunan bu kurallar Çizelge 10.37 ve Çizelge 10.38'de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.37'de 24.0718 Lift değeri ile en ilginç kural 'yastık → yatak' kuralıdır. 'yastık' ve 'yatak' kelimesi tüm gönderilerin %1.61'inde birlikte kullanılmıştır. Ayrıca 'yastık' kelimesinin kullanıldığı gönderilerin %81.82'sinde 'yatak' kelimesi de kullanılmaktadır.

Lift değeri 23.2917 olan 'arı sütü → arı' kuralında, 'arı sütü' ve 'arı' kelimesi tüm gönderilerin %1.25'inde birlikte kullanılmıştır. Ayrıca 'arı sütü' kelimesinin kullanıldığı gönderilerin %100'ünde 'arı' kelimesi de kullanılmaktadır.

Lift değeri 14.4926 olan 'zencefil → zerdaçal' kuralında, 'zencefil' ve 'zerdaçal' kelimesi tüm gönderilerin %2.50'sinde birlikte kullanılmıştır. Ayrıca 'zencefil' kelimesinin kullanıldığı gönderilerin %70'inde 'zerdaçal' kelimesi de kullanılmıştır.

Lift değeri 13.5515 olan ‘magnezyum → vitamin’ kuralında, ‘magnezyum’ ve ‘vitamin’ kelimesi tüm gönderilerin %1.43’ünde birlikte kullanılmıştır. Ayrıca ‘magnezyum’ kelimesinin kullanıldığı gönderilerin %72.73’ünde ‘vitamin’ kelimesi de kullanılmıştır.

Çizelge 10.37. Tedavi yorum kategorisine ait alternatif tedavi yöntemleri için ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar		Sonuç	Destek	Güven	Lift
[yastık]	→	yatak	0.0161	0.8182	24.0718
[arı sütü]	→	arı	0.0125	1.0000	23.2917
[zencefil]	→	zerdeçal	0.0250	0.7000	14.4926
[magnezyum]	→	vitamin	0.0143	0.7273	13.5515
[kiraz sapı]	→	ısırgan	0.0519	0.7632	10.6651
[ısırgan]	→	kiraz sapı	0.0519	0.7250	10.6651
[sarımsak]	→	limon	0.0233	0.6500	10.0931
[karabaş]	→	kiraz sapı	0.0179	0.6667	9.8070
[protez]	→	ameliyat	0.0644	0.8000	8.4377
[ameliyat]	→	protez	0.0644	0.6792	8.4377
[sülük]	→	hacamat	0.0787	0.6875	5.8229
[hacamat]	→	sülük	0.0787	0.6667	5.8229

Çizelge 10.38’de üç ve üzeri birliktelik kurallarına göre Lift değeri 15.5278 olan ‘sarımsak, bal → limon’ kuralında ‘sarımsak, bal’ kelime grubu ve ‘limon’ kelimesi tüm gönderilerin %1.07’sinde birlikte kullanılmıştır. Ayrıca ‘sarımsak, bal’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘limon’ kelimesi de kullanılmaktadır.

Lift değeri 11.0329 olan ‘ısırgan, karabaş → kiraz sapı’ kuralında ‘ısırgan, karabaş’ kelime grubu ve ‘kiraz sapı’ kelimesi tüm gönderilerin %1.07’sinde birlikte kullanılmıştır. Ayrıca ‘ısırgan, karabaş’ kelimelerinin kullanıldığı gönderilerin %75’inde ‘kiraz sapı’ kelimesi de kullanılmaktadır.

Lift değeri 9.8070 olan ‘kür, ısırgan → kiraz sapı’ kuralında ‘kür, ısırgan’ kelime grubu ve ‘kiraz sapı’ kelimesi tüm gönderilerin %66.67’sinde birlikte kullanılmıştır. Ayrıca ‘kür, ısırgan’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘kiraz sapı’ kelimesi de kullanılmaktadır.

Çizelge 10.38. Tedavi yorum kategorisine ait alternatif tedavi yöntemleri için üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[sarımsak, bal]	→ limon	0.0107	1.0000	15.5278
[ısırgan, karabaş]	→ kiraz sapı	0.0107	0.7500	11.0329
[kür, ısırgan]	→ kiraz sapı	0.0107	0.6667	9.8070
[kür, kiraz sapı]	→ ısırgan	0.0107	0.6667	9.3167
[kiraz sapı, karabaş]	→ ısırgan	0.0107	0.6000	8.3850

Tedavi kategorisine ait yorum kelime sonuçları

Toplam 13262 yorum ön işlem aşamasından geçirilmiştir. Yorumları temsil eden 4433 adet kelime kökü bulunmuştur. Frekans değeri 10'dan aşağı olan kelimeler elenmiştir. Ayrıca belli başına anlam ifade etmeyen durak kelimeleri ve gereksiz kelimeler silinmiştir. Bu işlemler sonrasında yorumları temsil eden 281 kelime kökü belirlenmiştir. Yıllara göre en çok kullanılan 50 kelime Çizelge 10.39'da verilmiştir.

Çizelge 10.39. Tedavi kategorisine ait yıllara göre yorumlarda en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	tedavi, hastalık, doktor, ilaç, fayda, hasta, başla, tıp, egzersiz, hastane, besle, dikkat, spor, etki, hareket, fizik, uygula, tavsiye, yat, bel, tuz, yardım, bitkisel, namaz, kan, teşhis, değer, yöntem, ameliyat, vücut, kötü, ayak, normal, kaplıca, seans, deri, uzak, dert, yorum, moral, devlet, sağlık, boyun, sonuç, zarar, sistem, tanı, saygı, cihaz, sıcak
2013	tedavi, doktor, hastalık, yatak, ilaç, fayda, hasta, başla, yat, boyun, ameliyat, tavsiye, egzersiz, etki, bel, hareket, vücut, iğne, tıp, hastane, kötü, sırt, eczane, sağlık, sonuç, spor, tuz, yardım, uygula, iltihap, uyu, yağ, iyot, fizik, kaplıca, kemik, dikkat, ayak, zarar, sıcak, ürün, bol, bitkisel, teşhis, normal, yastık, büyük, omurga, sert, ortopedik
2014	doktor, tedavi, ameliyat, ilaç, hastalık, fayda, hasta, tavsiye, yardım, protez, kalça, etki, hastane, başla, iğne, kan, hareket, sağlık, uygula, mide, egzersiz, eklem, sonuç, tıp, bel, spor, muayene, yat, ayak, teşhis, hacamat, kafa, boyun, kortizon, kontrol, kahve, normal, öner, limon, taş, sırt, zarar, doktora, yöntem, test, kuru, kiraz, kötü, tahlil, destek
2015	tedavi, hastalık, doktor, ilaç, yardım, fayda, vitamin, ameliyat, hasta, etki, tavsiye, başla, sağlık, egzersiz, yağ, yat, iğne, spor, destek, fizik, sıcak, moral, göz, hastane, sonuç, masaj, uygula, kortizon, atak, uyku, hareket, ayak, doktora, kötü, bitkisel, vücut, ürün, bol, yatak, balık, kalça, teşhis, boyun, normal, yorum, değer, eksik, sistem, yüksek, besle

belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 19 kural bulunmuştur. Bu kurallar Çizelge 10.40 ve Çizelge 10.41’de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.40’da 113.8929 lift değeri ile en ilginç kural ‘kök → hücre’ kuralıdır. ‘kök’ ve ‘hücre’ kelimesi tüm gönderilerin %0.28’inde birlikte kullanılmıştır. Ayrıca ‘kök’ kelimesinin kullanıldığı gönderilerin %75’inde ‘hücre’ kelimesi de kullanılmaktadır.

Lift değeri 35.3912 olan ‘bağışık → sistem’ kuralında, ‘bağışık’ ve ‘sistem’ kelimesi tüm gönderilerin %0.99’unda birlikte kullanılmıştır. Ayrıca ‘bağışık’ kelimesinin kullanıldığı gönderilerin %72.41’inde ‘sistem’ kelimesi de kullanılmaktadır.

Lift değeri 21.9317 olan ‘yastık → yat’ kuralında, ‘yastık’ ve ‘yat’ kelimesi tüm gönderilerin %0.75’inde birlikte kullanılmıştır. Ayrıca ‘yastık’ kelimesinin kullanıldığı gönderilerin %72.73’ünde ‘yat’ kelimesi de kullanılmıştır.

Lift değeri 5.3484 olan ‘fizik → tedavi’ kuralında, ‘fizik’ ve ‘tedavi’ kelimesi tüm gönderilerin %1.98’inde birlikte kullanılmıştır. Ayrıca ‘fizik’ kelimesinin kullanıldığı gönderilerin %79.25’inde ‘tedavi’ kelimesi de kullanılmıştır.

Çizelge 10.40. Tahlil kategorisine ait yorumlar için ikili kelime kullanımlarını gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[kök] →	hücre	0.0028	0.7500	113.8929
[bağışık] →	sistem	0.0099	0.7241	35.3912
[yastık] →	yat	0.0075	0.7273	21.9317
[fizik] →	tedavi	0.0198	0.7925	5.3484
[ozon] →	tedavi	0.0056	0.7059	4.7641

Çizelge 10.41’de üç ve üzeri birliktelik kurallarına göre Lift değeri 69.7049 olan ‘çay, karabiber → kaşık’ kuralında ‘çay, karabiber’ kelime grubu ve ‘kaşık’ kelimesi tüm gönderilerin %0.21’inde birlikte kullanılmıştır. Ayrıca ‘çay, karabiber’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘kaşık’ kelimesi de kullanılmaktadır.

Lift değeri 59.0556 olan ‘kiraz, çay→ ısırgan’ kuralında ‘kiraz, çay’ kelime grubu ve ‘ısırgan’ kelimesi tüm gönderilerin %0.28’inde birlikte kullanılmıştır. Ayrıca ‘kiraz, çay’ kelimelerinin kullanıldığı gönderilerin %75’inde ‘ısırgan’ kelimesi de kullanılmaktadır.

Lift değeri 45.5571 olan ‘ısırgan, çay→ kiraz’ kuralında ‘ısırgan, çay’ kelime grubu ve ‘kiraz’ kelimesi tüm gönderilerin %0.28’inde birlikte kullanılmıştır. Ayrıca ‘ısırgan, çay’ kelimelerinin kullanıldığı gönderilerin %75’inde ‘kiraz’ kelimesi de kullanılmaktadır.

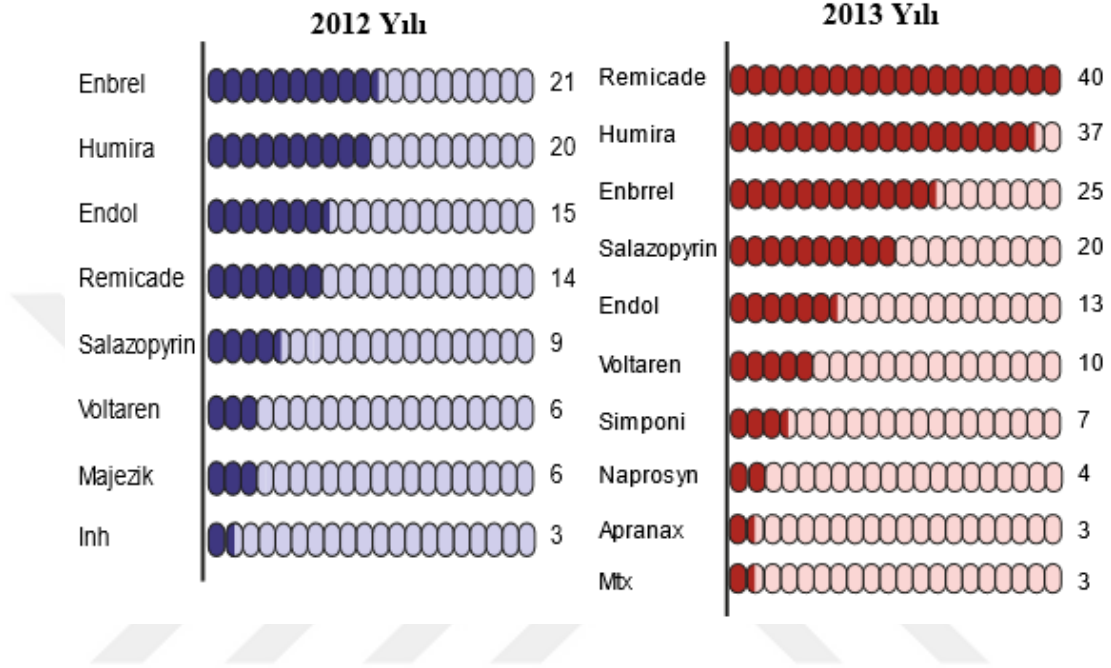
Çizelge 10.41. Tahlil kategorisine ait yorumlar için üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[çay, karabiber]	→ kaşık	0.0021	1.0000	69.7049
[kiraz, çay]	→ ısırgan	0.0028	0.7500	59.0556
[ısırgan, çay]	→ kiraz	0.0028	0.7500	45.5571
[hastalık, bağışık]	→ sistem	0.0033	0.7778	38.0128
[fayda, bağışık]	→ sistem	0.0021	0.7500	36.6552
[hasta, besle]	→ hastalık	0.0021	0.7500	6.8877
[bitkisel, uygula]	→ tedavi	0.0026	1.0000	6.7492
[fizik, ilaç]	→ tedavi	0.0021	1.0000	6.7492
[öner, fizik]	→ tedavi	0.0021	1.0000	6.7492
[tavsiye, fizik]	→ tedavi	0.0021	0.9000	6.0743
[fayda, fizik]	→ tedavi	0.0028	0.8571	5.7850
[hastane, fizik]	→ tedavi	0.0026	0.8462	5.7109
[fizik, kaplıca]	→ tedavi	0.0024	0.8333	5.6243
[doktor, fizik]	→ tedavi	0.0063	0.8182	5.5221

10.11. İlaç Kategorisine Ait Sonuçlar

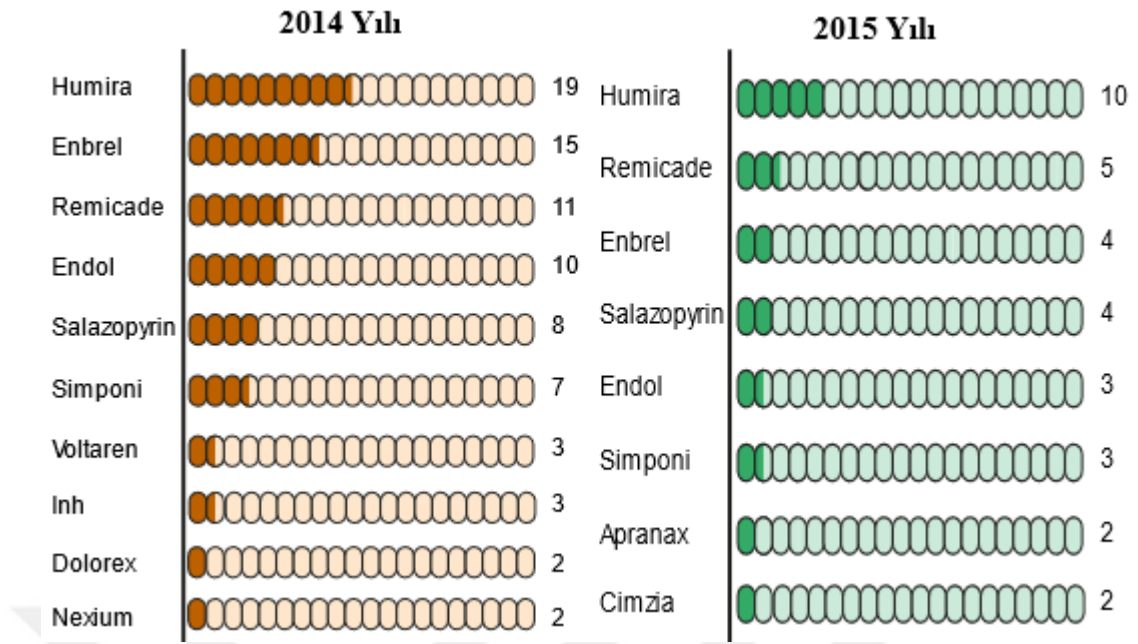
Toplam 1333 adet ilaç gönderi verileri analiz edilmiştir. Toplam kadın ve erkek kullanıcılar tarafından kullanılan 60 farklı ilaç ismi tespit edilmiştir. Bu ilaç isimleri gönderiler içerisinde geçip geçmediğine göre binary ağırlıklandırma yöntemi ile ağırlıklandırılmıştır. Kadın ve erkek kullanıcılar tarafından yıllara göre en çok kullanılan ilaç isimlerini göstermek amacıyla grafikler oluşturulmuştur.

Şekil 10.55’de erkek hastalar tarafından 2012 ve 2013 yıllarında gönderiler içerisinde en çok kullandıkları ilaçların isimleri grafikler ile verilmiştir. 2012 yılında kullanıcılar tarafından Enbrel, Humira ve Endol ilaç isimleri sıklıkla geçmektedir. 2013 yılında ise Remicade, Humira ve Enbrel ilaçlarının isimleri daha çok geçmektedir.



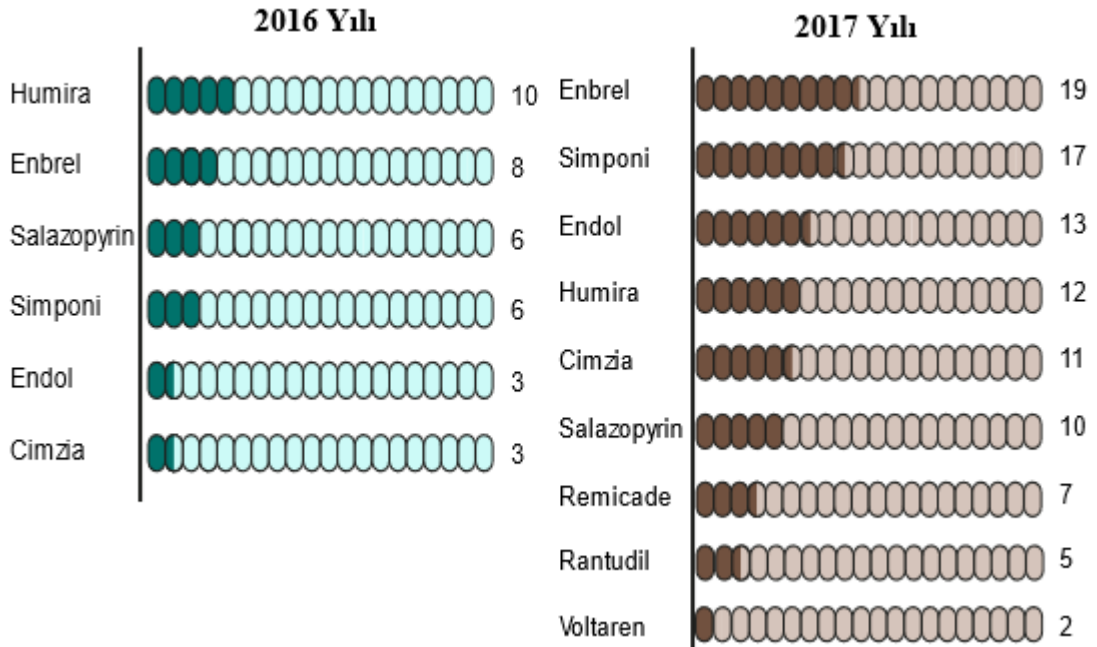
Şekil 10.55. 2012 ve 2013 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.56’ya göre 2014 yılından Humira, Enbrel ve Remicade, 2015 yılında ise Humira, Remicade ve Enbrel ilaçları erkek hastalar tarafından sıkça konuşulduğu görülmektedir.



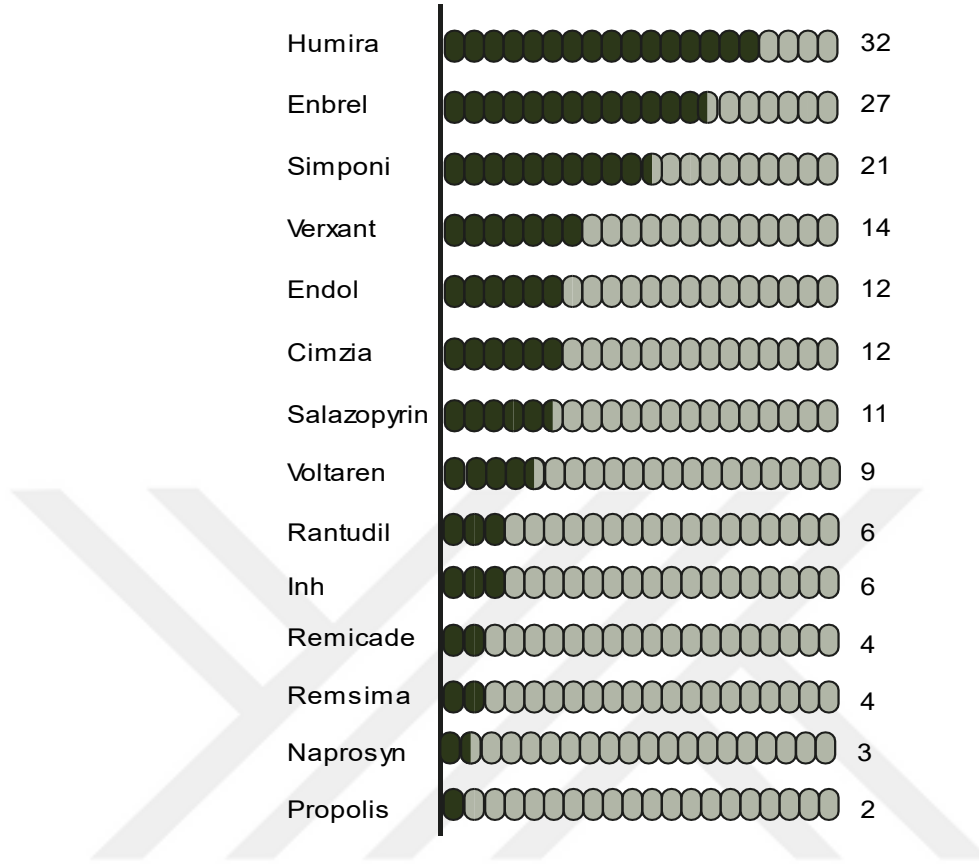
Şekil 10.56. 2014 ve 2015 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.57’de yer alan grafiklere göre 2016 yılında Humira, Enbrel ve Salazopyrin, 2017 yılında ise Enbrel, Simponi ve Endol ilaçlar isimleri erkek kullanıcılar tarafından sıkça kullanılmaktadır.



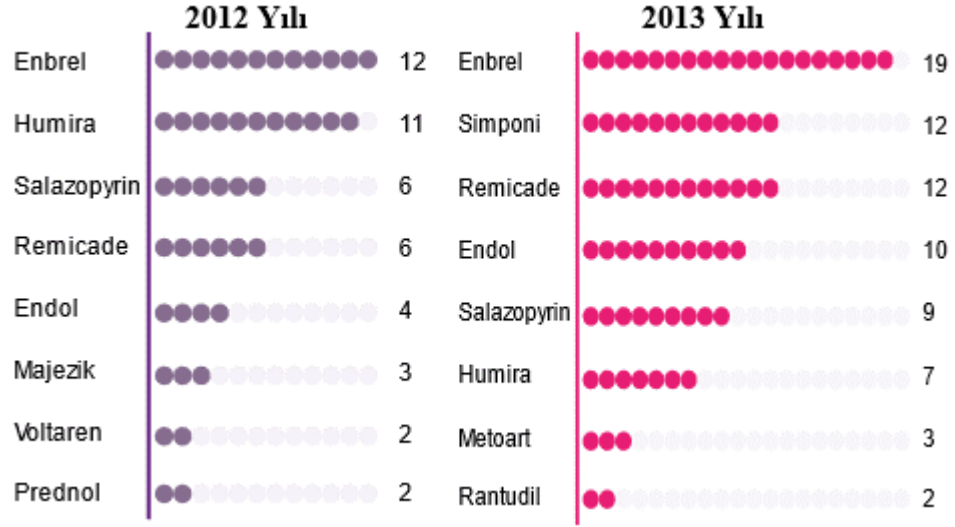
Şekil 10.57. 2016 ve 2017 yıllarında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.58'e göre erkek kullanıcılar tarafından Humira, Enbrel ve Simponi ilaçlar isimleri daha çok kullanılmaktadır.



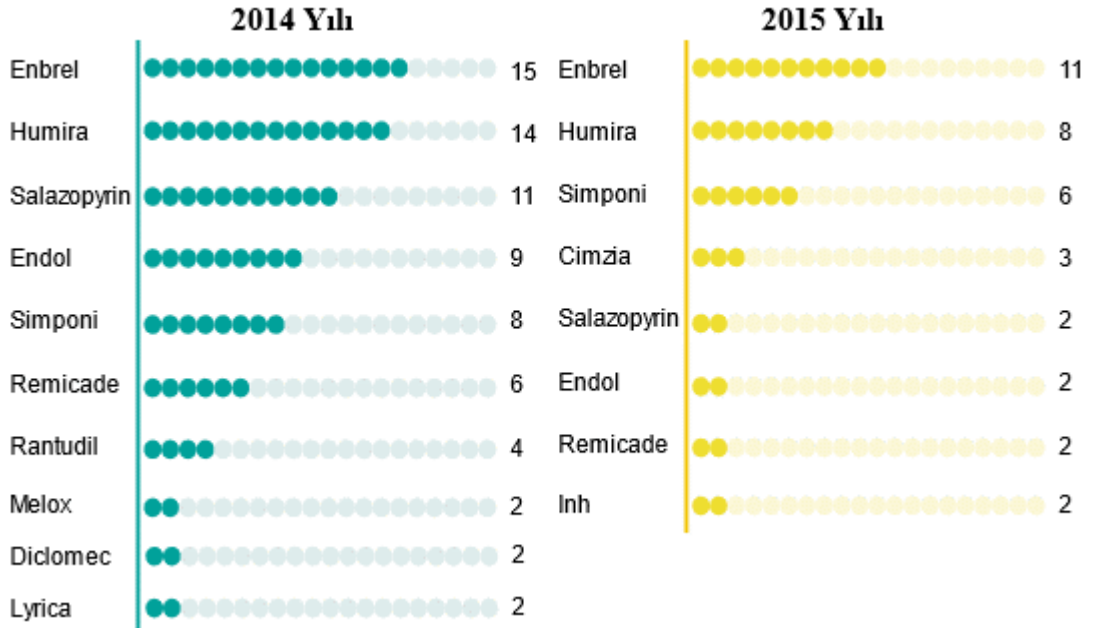
Şekil 10.58. 2018 yılında erkek kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.59'da kadın kullanıcılar tarafından 2012 ve 2013 yıllarında en çok kullanılan ilaç isimleri grafikler ile gösterilmiştir. 2012 yılında Enbrel, Humira ve Salazopyrin, 2013 yılında ise Enbrel, Simponi ve Remicade ilaçları kadın kullanıcılar tarafından sıklıkla konuşulmaktadır.



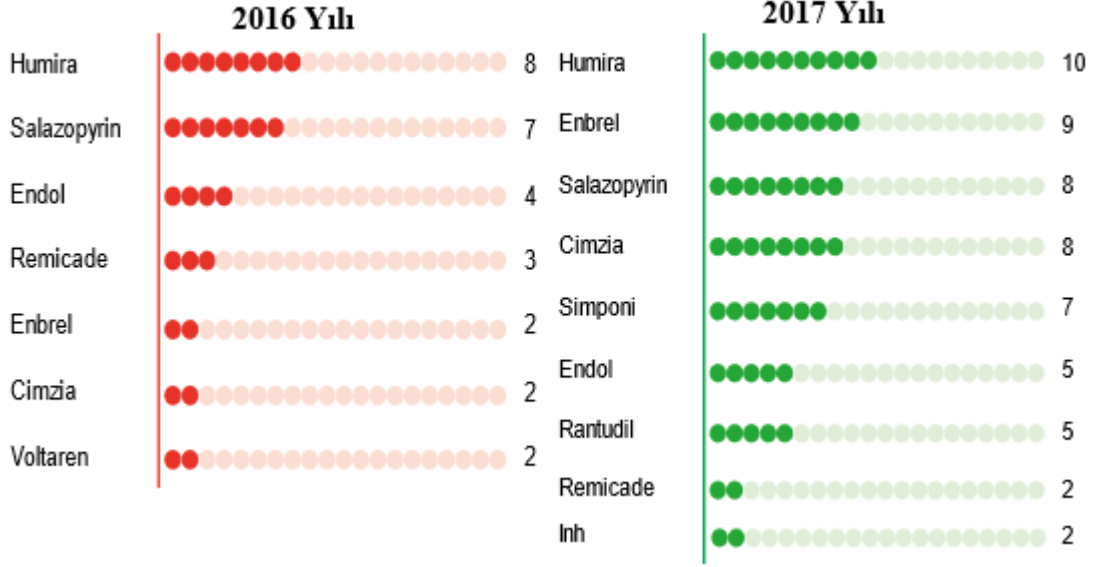
Şekil 10.59. 2012 ve 2013 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.60'a göre 2014 yılından kadın kullanıcılar tarafından Enbrel, Humira ve Salazopyrin, 2015 yılında Enbrel, Humira ve Simponi ilaçları hakkında daha çok konuşulmaktadır.



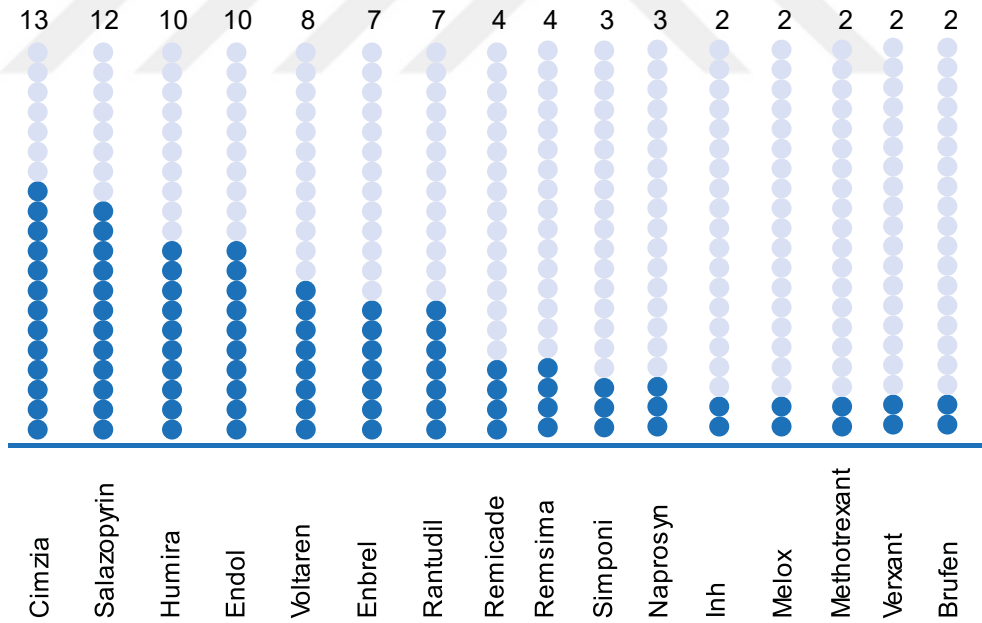
Şekil 10.60. 2014 ve 2015 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.61'e göre 2016 yılında Humira, Salazopyrin ve Endol, 2017 yılında Humira, Enrel ve Salazopyrin ilaçları isimleri kadın kullanıcılar tarafından sıklıkla kullanılmaktadır.



Şekil 10.61. 2016 ve 2017 yıllarında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.62'ye göre 2018 Cimzia ve Salazopyrin ilaçları sıklıkla kullanılmaktadır.



Şekil 10.62. 2018 yılında kadın kullanıcılar tarafından en çok kullanılan ilaç isimleri

Şekil 10.63'te kadın ve erkek kullanıcılar tarafından toplam olarak en çok kullanılan ilaçların isimleri verilmiştir. Toplam olarak hastalar tarafından Enbrel, Humira ve Salazopyrin ilaçları sıklıkla konuşulduğu görülmektedir.

Kullanılan ilaç isimleri için ilişkili örnekleme test sonuçları

Kullanıcılar tarafından konuşulan ilaçların yıllara göre farklılık gösterip göstermediğini test etmek amacıyla Friedman testi uygulanmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Kullanıcılar tarafından konuşulan ilaçlar yıllara göre farklılık göstermemektedir.

H_1 : Kullanıcılar tarafından konuşulan ilaçlar yıllara göre farklılık göstermektedir.

Test sonuçları Çizelge 10.42’de verilmiştir. Test sonucunda ($p < 0.05$) H_0 hipotezi reddedilir. Buna göre, kullanılan ilaçlar yıllara göre farklılık göstermektedir. 2013 yılı 2015 ve 2016’dan, 2014 yılı 2015 ve 2018’den, 2016 yılı 2018’den ve 2017 yılı 2018’den farklılık göstermektedir.

Çizelge 10.42. Kullanılan ilaçlara ait Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	60	2.60±6.91	0.00 (0.00-33.00) ^{abcd}	4.02		
2013	60	4.33±11.20	0.00 (0.00-52.00) ^{ad}	4.59		
2014	60	2.92±6.91	0.00 (0.00-33.00) ^{ac}	4.47		
2015	60	1.25±3.42	0.00 (0.00-18.00) ^b	3.08	68.93	<0.001
2016	60	1.30±3.32	0.00 (0.00-18.00) ^{bc}	3.28		
2017	60	2.67±6.68	0.00 (0.00-28.00) ^{bcd}	3.50		
2018	60	4.60±9.07	1.00 (0.00-42.00) ^a	5.06		

a-d; aynı harfe sahip yıllar arasında fark yoktur ($p > 0.05$)

Kullanılan ilaç isimleri için uyum analizi sonuçları

Kullanılan ilaçlar ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

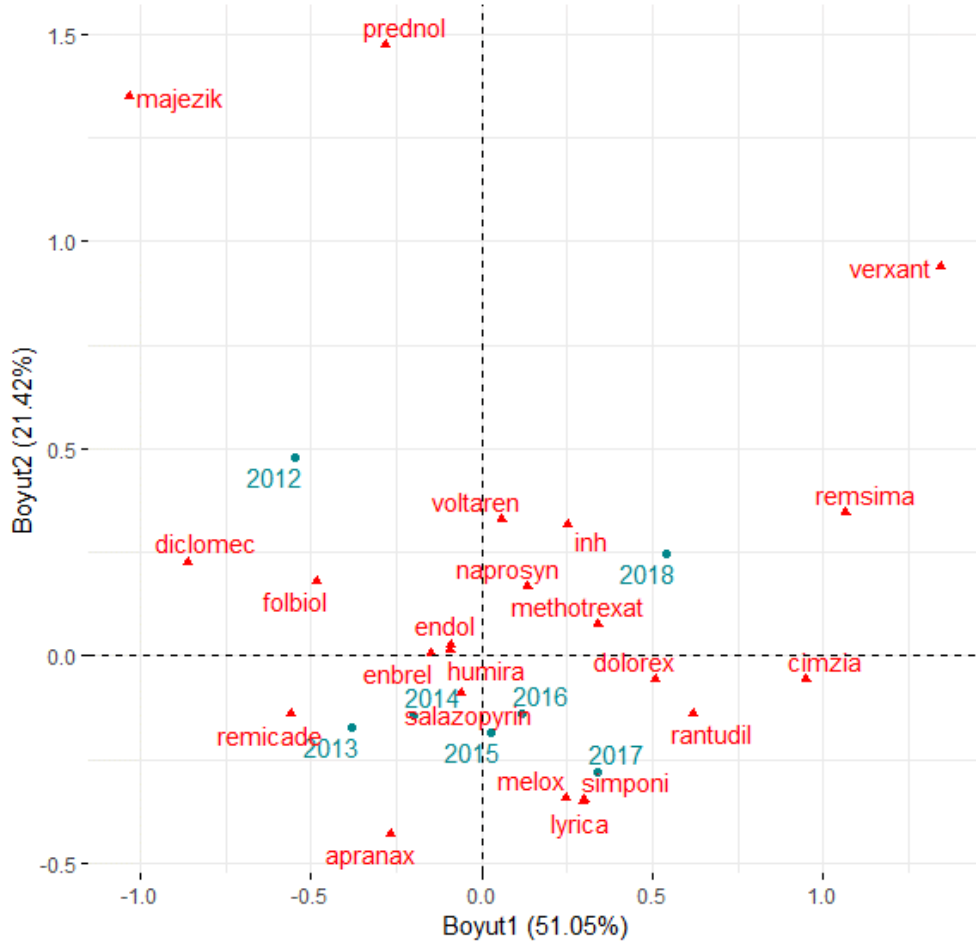
H_0 : Kullanılan ilaçlar ve yıllar arasında ilişki yoktur

H_1 : Kullanılan ilaçlar ve yıllar arasında ilişki vardır

R programında 7x22 boyutunda bir matris oluşturulmuştur. 22, toplamda 5 ve üzeri sıklıkta kullanılan ilaç isimlerini göstermektedir. 7 ise toplam yılların sayısını göstermektedir. Test sonucunda $p < 0.001$ ’den dolayı %99.9 güvenle H_0 hipotezi

reddedilir. Sonuç olarak %99.9 güvenle kullanılan ilaçlar ve yıllar arasında bir ilişki vardır.

Şekil 10.65'te basit uyum analizi sonucu verilmiştir. Boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam inertianın %51.05'ini açıklarken, ikinci boyut ise %21.42'sini açıklamaktadır.



Şekil 10.65. Kullanılan ilaçlar ve yıllar için basit uyum analizi sonucu

Prednol ve Majezik ilaçları 2012 yılında yoğunlaşırken, Verxant, Remsima ve Cimzia ilaçları 2018 yılında yoğunlaşmaktadır. Endol, Enbrel, Humira, Salazopyrin ilaçları yılların tamamında kullanılmaktadır.

Bu boyutların kullanılan ilaçları ve yılları hangi oranda açıkladığı Çizelge 10.43 ve Çizelge 10.44'te verilmiştir. Çizelge 10.43'te 2012, 2013, 2017 ve 2018 boyut 1, 2014 boyut 4, 2015 ve 2016 yılları büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.43. İlaç kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012	0.522	0.394	0.072	0.012	0.000
2013	0.576	0.118	0.269	0.023	0.011
2014	0.219	0.112	0.059	0.552	0.002
2015	0.003	0.140	0.007	0.131	0.637
2016	0.066	0.099	0.048	0.110	0.117
2017	0.368	0.254	0.240	0.085	0.053
2018	0.781	0.161	0.055	0.002	0.000

Çizelge 10.44'te enbrel, remicade, cimzia, rantudil, verxant, remsima, diclomec ve folbiol ilaç isimleri boyut 1, salazopyrin, simponi, voltaren majezik ve prednol boyut 2, endol ve naprosyn boyut 3, melox, methotrexat, lyrica ve dolorex boyut 4, humira, inh ve apranax ilaçları büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.44. Boyutların ilaçları açıklama oranları

İlaçlar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
humira	0.296	0.007	0.009	0.083	0.572
enbrel	0.677	0.000	0.148	0.066	0.003
salazopyrin	0.093	0.212	0.013	0.093	0.008
remicade	0.745	0.048	0.157	0.034	0.017
endol	0.268	0.026	0.398	0.028	0.280
simponi	0.415	0.557	0.007	0.004	0.000
cimzia	0.861	0.003	0.045	0.085	0.000
voltaren	0.019	0.629	0.333	0.010	0.009
rantudil	0.751	0.038	0.006	0.007	0.167
inh	0.165	0.266	0.117	0.074	0.351
verxant	0.546	0.268	0.164	0.011	0.003
naprosyn	0.034	0.053	0.611	0.036	0.002
remsima	0.862	0.090	0.030	0.003	0.014
majezik	0.313	0.536	0.126	0.020	0.004
melox	0.106	0.209	0.094	0.512	0.079
methotrexat	0.153	0.007	0.107	0.704	0.002
apranax	0.054	0.141	0.229	0.037	0.368

Çizelge 10.44. (devam)

İlaçlar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
diclomec	0.434	0.030	0.088	0.332	0.017
lyrica	0.076	0.108	0.172	0.603	0.001
prednol	0.033	0.893	0.046	0.026	0.003
dolorex	0.239	0.003	0.036	0.428	0.087
folbiol	0.472	0.064	0.277	0.039	0.117

İlaçlar için gönderi birliktelik analizi sonuçları

Toplam veri dosyası; 1333'dir. Bulunan anahtar kelimelerinden 2'den aşağı kelime barındıran yorumlar elenmiştir. Analizde toplam 268 gönderi kullanılmıştır. Birliktelik kurallarını çıkarmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.004 ve minimum güven değeri 0.6 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 25 kural bulunmuştur. Bu kurallar Çizelge 10.45 ve Çizelge 10.46'da verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.45'te 35.7333 Lift değeri ile en ilginç kural 'arveles → diclomec' kuralıdır. 'arveles' ve 'diclomec' kelimesi tüm gönderilerin %0.75'inde birlikte kullanılmıştır. Ayrıca 'arveles' kelimesinin kullanıldığı gönderilerin %66.67'sinde 'diclomec' kelimesi de kullanılmaktadır.

Lift değeri 4.7857 olan 'etol → endol' kuralında, 'etol' ve 'endol' kelimesi tüm gönderilerin %0.75'inde birlikte kullanılmıştır. Ayrıca 'etol' kelimesinin kullanıldığı gönderilerin %100'ünde 'endol' kelimesi de kullanılmaktadır.

Lift değeri 4.7018 olan 'arveles → voltaren' kuralında, 'arveles' ve 'voltaren' kelimesi tüm gönderilerin %0.75'inde birlikte kullanılmıştır. Ayrıca 'arveles' kelimesinin kullanıldığı gönderilerin %66.67'sinde 'voltaren' kelimesi de kullanılmıştır.

Lift değeri 2.7917 olan 'verxant → humira' kuralında, 'verxant' ve 'humira' kelimesi tüm gönderilerin %1.12'sinde birlikte kullanılmıştır. Ayrıca 'verxant' kelimesinin kullanıldığı gönderilerin %75'inde 'humira' kelimesi de kullanılmıştır.

Çizelge 10.45. Kullanılan ilaçlara ait ikili kullanımları gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[arveles]	→ diclomec	0.0075	0.6667	35.7333
[etol]	→ Endol	0.0075	1.0000	4.7857
[arveles]	→ Voltaren	0.0075	0.6667	4.7018
[verxant]	→ Humira	0.0112	0.7500	2.7917
[apranax]	→ humira	0.0075	0.6667	2.4815
[endol]	→ salazopyrin	0.1306	0.6250	2.1203
[rantudil]	→ salazopyrin	0.0448	0.6000	2.0354

Çizelge 10.46’da üç ve üzeri birliktelik kurallarına göre Lift değeri 44.6667 olan ‘voltaren, diclomec → majezik’ kuralında ‘voltaren, diclomec’ kelime grubu ve ‘majezik’ kelimesi tüm gönderilerin %0.75’inde birlikte kullanılmıştır. Ayrıca ‘voltaren, diclomec’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘majezik’ kelimesi de kullanılmaktadır.

Lift değeri 35.7333 olan ‘voltaren, majezik→ diclomec’ kuralında ‘voltaren, majezik’ kelime grubu ve ‘diclomec’ kelimesi tüm gönderilerin %0.75’inde birlikte kullanılmıştır. Ayrıca ‘voltaren, majezik’ kelimelerinin kullanıldığı gönderilerin %66.67’sinde ‘diclomec’ kelimesi de kullanılmaktadır.

Lift değeri 7.2432 olan ‘cimzia, humira, enbrel→ simponi’ kuralında ‘cimzia, humira, enbrel’ kelime grubu ve ‘simponi’ kelimesi tüm gönderilerin %0.75’inde birlikte kullanılmıştır. Ayrıca ‘cimzia, humira, enbrel’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘simponi’ kelimesi de kullanılmaktadır.

Çizelge 10.46. Kullanılan ilaçlara ait üç ve üzeri kullanımları gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[voltaren, diclomec]	→ majezik	0.0075	1.0000	44.6667
[voltaren, majezik]	→ diclomec	0.0075	0.6667	35.7333
[cimzia, humira, enbrel]	→ simponi	0.0075	1.0000	7.2432
[cimzia, enbrel, remicade]	→ simponi	0.0075	1.0000	7.2432
[cimzia, enbrel]	→ simponi	0.0112	1.0000	7.2432
[diclomec, majezik]	→ voltaren	0.0075	1.0000	7.0526
[salazopyrin, prednol]	→ endol	0.0075	1.0000	4.7857

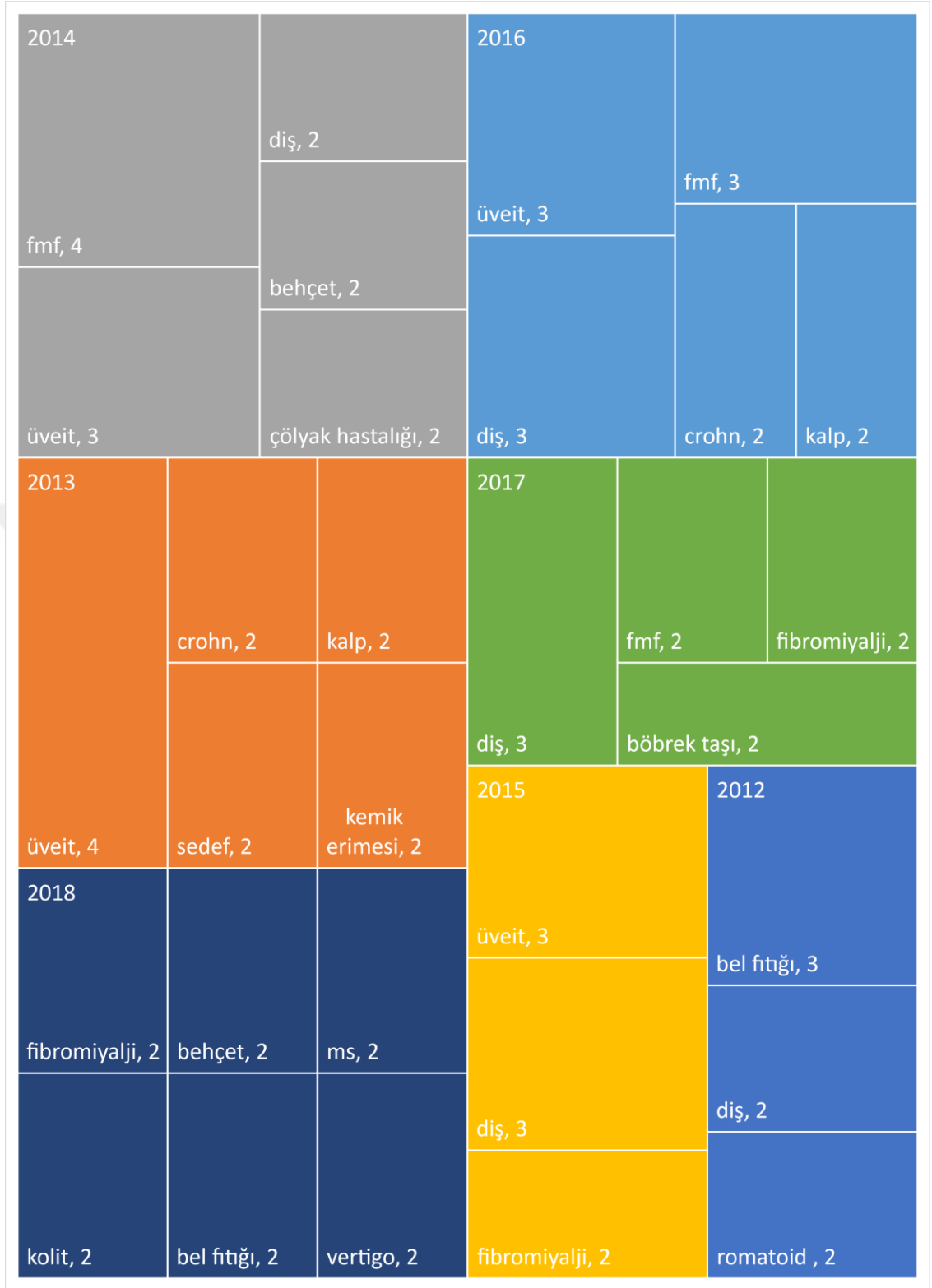
Çizelge 10.46. (devam)

Gruplar	Sonuç	Destek	Güven	Lift
[cimzia, enbrel, simponi]	→ remicade	0.0075	0.6667	4.3577
[humira, enbrel, simponi]	→ remicade	0.0149	0.6667	4.3577
[cimzia, humira, simponi]	→ enbrel	0.0075	1.0000	4.2540
[cimzia, simponi, remicade]	→ enbrel	0.0075	1.0000	4.2540
[humira, simponi, remicade]	→ enbrel	0.0149	0.8000	3.4032
[prednol, endol]	→ salazopyrin	0.0075	1.0000	3.3924
[simponi, remicade]	→ enbrel	0.0261	0.7778	3.3086
[cimzia, simponi]	→ enbrel	0.0112	0.6000	2.5524
[cimzia, enbrel, simponi]	→ humira	0.0075	0.6667	2.4815
[rantudil, endol]	→ salazopyrin	0.0112	0.6000	2.0354
[enbrel, endol]	→ salazopyrin	0.0112	0.6000	2.0354

10.12. Diğer Hastalıklar Kategorisine Ait Sonuçlar

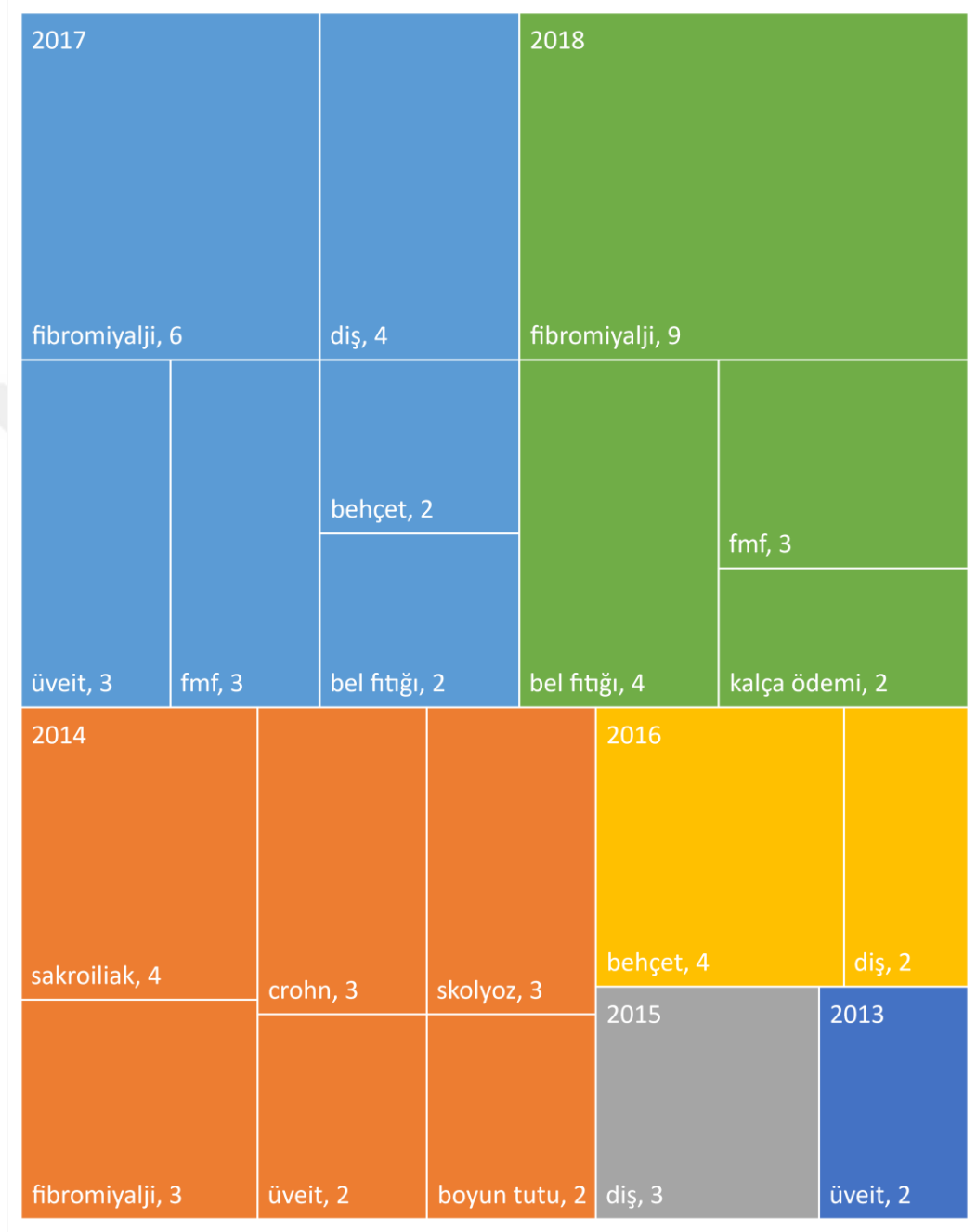
Diğer hastalıklar sınıfına ait toplam 404 adet gönderi analiz edilmiştir. Kullanıcılar toplam 86 farklı hastalıktan bahsedilmiştir. Kadın ve erkek kullanıcılar tarafından bu hastalıkların yıllara göre kullanımları ağaç haritası grafikleri ile gösterilmiştir. Ağaç haritalarında değişkenlerin her biri bir alan olarak temsil edilir ve değişkenin büyüklüğüne göre alanlar büyümektedir. Grafikler Şekil 10.66 ve Şekil 10.67’de verilmiştir.

Şekil 10.66’da erkekler tarafından 2012 yılında bel fitiği, 2013 yılında üveit, 2014 yılında fmf ve üveit, 2015 yılında üveit ve diş, 2016 yılında fmf, üveit ve diş, 2017 yılında diş, 2018 yılında ise fibromiyalji, Behçet, Ms, kolit, bel fitiği ve vertigo sıklıkla konuşulan hastalıklardır.



Şekil 10.66. Erkek kullanıcılar tarafından yıllara göre diğer hastalık isimlerinin kullanımı

Şekil 10.67’de kadın kullanıcılar tarafından 2013 yılında üveit, 2014 yılında sakroiliak, 2015 yılında diş, 2016 yılında Behçet, 2017 ve 2018 yıllarında fibromiyalji hastalıkları sıklıkla konuşulmaktadır.



Şekil 10.67. Kadın kullanıcılar tarafından yıllara göre diğer hastalık isimlerinin kullanımı

Konuşulan hastalık isimleri için ilişkili örneklem test sonuçları

Kullanıcılar tarafından konuşulan hastalıkların yıllara göre farklılık gösterip göstermediğini test etmek amacıyla Friedman testi uygulanmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Kullanıcılar tarafından konuşulan hastalıklar yıllara göre farklılık göstermemektedir.

H_1 : Kullanıcılar tarafından konuşulan hastalıklar yıllara göre farklılık göstermektedir.

Test sonuçları Çizelge 10.47’de verilmiştir. Test sonucunda ($p < 0.05$) H_0 hipotezi reddedilir. Buna göre, konuşulan hastalıklar yıllara göre farklılık göstermektedir. Bu farklılık 2012 yılı 2018’den, 2013 yılı 2018’den ve 2016 yılı 2018’den farklılık göstermektedir.

Çizelge 10.47. Kullanılan ilaçlara ait Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	86	0.40±0.95	0.00 (0.00-7.00) ^a	3.71		
2013	86	0.54±1.91	0.00 (0.00-16.00) ^a	3.57		
2014	86	0.84±2.78	0.00 (0.00-24.00) ^{ab}	3.97		
2015	86	0.74±2.91	0.00 (0.00-26.00) ^{ab}	3.91	40.72	<0.001
2016	86	0.69±2.38	0.00 (0.00-20.00) ^a	3.72		
2017	86	0.93±2.86	0.00 (0.00-24.00) ^{ab}	4.23		
2018	86	1.19±3.07	1.00 (0.00-26.00) ^b	4.90		

a-b; aynı harfe sahip yıllar arasında fark yoktur ($p > 0.05$)

Kullanılan hastalık isimleri için uyum analizi sonuçları

Konuşulan hastalıklar ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

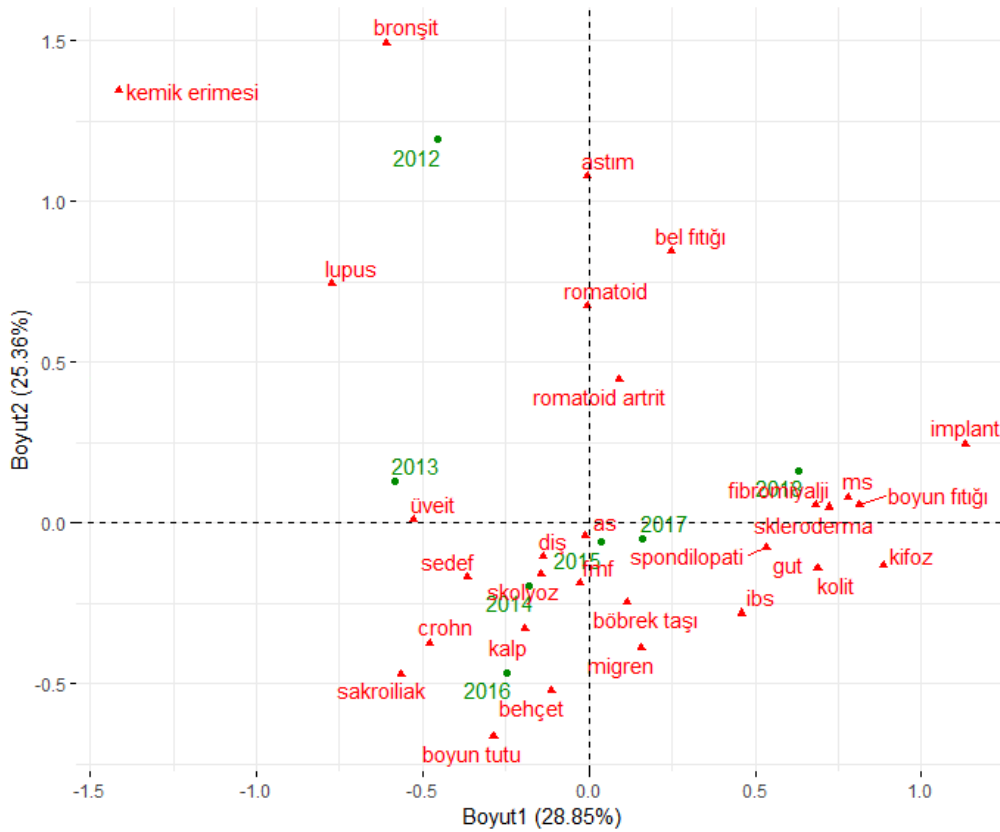
H_0 : Konuşulan hastalıklar ve yıllar arasında ilişki yoktur

H_1 : Konuşulan hastalıklar ve yıllar arasında ilişki vardır

R programında 7x30 boyutunda bir matris oluşturulmuştur. 30, toplamda 3 ve üzeri sıklıkta kullanılan hastalık isimlerini göstermektedir. 7 ise toplam yılların

sayısını göstermektedir. Test sonucunda $p>0.05$ 'ten dolayı %95 güvenle H_0 hipotezi reddedilmez. Sonuç olarak %95 güvenle konuşulan hastalıklar ve yıllar arasında bir ilişki yoktur.

Şekil 10.68'te basit uyum analizi sonucu verilmiştir. Boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %28.85'ini açıklarken, ikinci boyut ise %25.36'sını açıklamaktadır.



Şekil 10.68. Konuşulan hastalıklar ve yıllar için basit uyum analizi sonucu

Bu boyutların konuşulan hastalıkları ve yılları hangi oranda açıkladığı Çizelge 10.48 ve Çizelge 10.49'da verilmiştir. Çizelge 10.48'de 2013 ve 2018 boyut 1, 2012 ve 2016 boyut 2, 2014 ve 2017 yılları boyut 4, 2015 yılı büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.48. Diğer hastalıklar kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012	0.116	0.789	0.013	0.019	0.011
2013	0.496	0.024	0.261	0.045	0.007
2014	0.092	0.109	0.366	0.407	0.012
2015	0.004	0.011	0.039	0.081	0.858
2016	0.121	0.430	0.148	0.096	0.055
2017	0.093	0.009	0.208	0.301	0.136
2018	0.786	0.052	0.106	0.049	0.003

Çizelge 10.49'da fibromiyalji, üveit, crohn, ms, boyun fitiği, lupus, kolit, implant, gut, spondilopati, skleroderma ve kemik erimesi hastalıkları boyut 1, bel fitiği, behçet, romatoid, romatoid artrit, boyun tutu (boyun tutulumu), bronşit, migren ve astım boyut 2, kalp, sedef, böbrek taşı ve kifoza boyut 3, diş, sakroiliak, skolyoz ve ibs boyut 4, as ve fmf hastalıkları büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.49. Boyutların diğer hastalıkları açıklama oranları

Hastalıklar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
as	0.009	0.144	0.000	0.048	0.799
fibromiyalji	0.814	0.006	0.015	0.016	0.015
diş	0.055	0.033	0.177	0.699	0.023
üveit	0.896	0.000	0.025	0.010	0.036
fmf	0.003	0.176	0.175	0.004	0.618
bel fitiği	0.058	0.674	0.158	0.000	0.102
behçet	0.023	0.491	0.128	0.063	0.140
crohn	0.445	0.271	0.045	0.222	0.015
romatoid	0.000	0.861	0.056	0.049	0.034
kalp	0.058	0.169	0.594	0.003	0.176
sedef	0.168	0.036	0.436	0.063	0.000
romatoid artrit	0.025	0.599	0.004	0.004	0.365
sakroiliak	0.237	0.164	0.141	0.425	0.008
skolyoz	0.021	0.026	0.159	0.647	0.067
boyun tutu	0.104	0.563	0.124	0.011	0.012
ms	0.505	0.005	0.004	0.107	0.331

Çizelge 10.49. (devam)

Hastalıklar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
boyun fitiği	0.665	0.003	0.014	0.172	0.129
lupus	0.422	0.391	0.023	0.032	0.124
bronşit	0.094	0.565	0.169	0.020	0.086
ibs	0.178	0.067	0.001	0.356	0.345
migren	0.045	0.273	0.002	0.043	0.137
böbrek taşı	0.011	0.050	0.614	0.113	0.005
kolit	0.407	0.017	0.278	0.004	0.081
astım	0.000	0.556	0.054	0.109	0.058
kifoz	0.377	0.009	0.382	0.008	0.058
implant	0.629	0.029	0.089	0.022	0.207
gut	0.341	0.007	0.231	0.135	0.216
spondilopati	0.341	0.007	0.231	0.135	0.216
skleroderma	0.543	0.003	0.021	0.163	0.188
kemik erimesi	0.406	0.369	0.143	0.011	0.000

Konuşulan hastalıklar için birliktelik analizi sonuçları

Toplam veri dosyası, 404'tür. Bulunan anahtar kelimelerinden 2'den aşağı kelime barındıran yorumlar elenmiştir. Analizde toplam 116 gönderi kullanılmıştır. Birliktelik kurallarını çıkarmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.008 ve minimum güven değeri 0.7 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 81 kural bulunmuştur. Bu kurallar Çizelge 10.50 ve Çizelge 10.51'de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.50'de 58 Lift değeri ile en ilginç kural 'bursit → tenisçi dirseği' kuralıdır. 'bursit' ve 'tenisçi dirseği' kelimesi tüm gönderilerin %0.86'sında birlikte kullanılmıştır. Ayrıca 'bursit' kelimesinin kullanıldığı gönderilerin %100'ünde 'tenisçi dirseği' kelimesi de kullanılmaktadır.

Lift değeri 38.6667 olan 'kalp kacağı → kalp' kuralında, 'kalp kacağı' ve 'kalp' kelimesi tüm gönderilerin %0.86'sında birlikte kullanılmıştır. Ayrıca 'kalp kacağı' kelimesinin kullanıldığı gönderilerin %100'ünde 'kalp' kelimesi de kullanılmaktadır.

Lift değeri 38.6667 olan ‘koah → bronşit’ kuralında, ‘koah’ ve ‘bronşit’ kelimesi tüm gönderilerin %0.86’sında birlikte kullanılmıştır. Ayrıca ‘koah’ kelimesinin kullanıldığı gönderilerin %100’ünde ‘bronşit’ kelimesi de kullanılmıştır.

Lift değeri 29 olan ‘lyme → ms’ kuralında, ‘lyme’ ve ‘ms’ kelimesi tüm gönderilerin %0.86’sında birlikte kullanılmıştır. Ayrıca ‘lyme’ kelimesinin kullanıldığı gönderilerin %100’ünde ‘ms’ kelimesi de kullanılmıştır.

Çizelge 10.50. Konuşulan hastalıklara ilişkin ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[bursit]	→ tenisçi dirseği	0.0086	1.0000	58.0000
[kalp kaçağı]	→ kalp	0.0086	1.0000	38.6667
[koah]	→ bronşit	0.0086	1.0000	38.6667
[lyme]	→ ms	0.0086	1.0000	29.0000
[egzama]	→ sedef	0.0086	1.0000	19.3333
[beyaz leke]	→ as	0.0086	1.0000	1.2211
[plantar fasiit]	→ as	0.0086	1.0000	1.2211
[protein kaçağı]	→ as	0.0086	1.0000	1.2211
[amiloidoz]	→ as	0.0086	1.0000	1.2211
[brusella]	→ as	0.0086	1.0000	1.2211
[diyabet]	→ as	0.0086	1.0000	1.2211
[implant]	→ as	0.0086	1.0000	1.2211
[lenfoma]	→ as	0.0086	1.0000	1.2211
[pubis]	→ as	0.0086	1.0000	1.2211
[vertigo]	→ as	0.0086	1.0000	1.2211
[böbrek taşı]	→ as	0.0172	1.0000	1.2211
[gut]	→ as	0.0172	1.0000	1.2211
[astım]	→ as	0.0259	1.0000	1.2211
[kolit]	→ as	0.0259	1.0000	1.2211
[spondilopati]	→ as	0.0259	1.0000	1.2211
[crohn]	→ as	0.0690	1.0000	1.2211
[üveit]	→ as	0.1207	0.9333	1.1396
[fmf]	→ as	0.0862	0.9091	1.1100
[fibromiyalji]	→ as	0.1293	0.8824	1.0774

Çizelge 10.50. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[bel fitiği]	→	as	0.0517	0.7500	0.9158
[behçet]	→	as	0.0431	0.7143	0.8722

Çizelge 10.51’de üç ve üzeri birliktelik kurallarına göre Lift değeri 116 olan ‘fibromiyalji, bel fitiği, boyun fitiği → periferik spondilit’ kuralında ‘fibromiyalji, bel fitiği, boyun fitiği’ kelime grubu ve ‘periferik spondilit’ kelimesi tüm gönderilerin %0.86’sında birlikte kullanılmıştır. Ayrıca ‘fibromiyalji, bel fitiği, boyun fitiği’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘periferik spondilit’ kelimesi de kullanılmaktadır.

Lift değeri 116 olan ‘bel fitiği, kifoz, skolyoz → kalça ödemi’ kuralında ‘bel fitiği, kifoz, skolyoz’ kelime grubu ve ‘kalça ödemi’ kelimesi tüm gönderilerin %0.86’sında birlikte kullanılmıştır. Ayrıca ‘bel fitiği, kifoz, skolyoz’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘kalça ödemi’ kelimesi de kullanılmaktadır.

Lift değeri 116 olan ‘romatoid, kalp → kalp romatizması’ kuralında ‘romatoid, kalp’ kelime grubu ve ‘kalp romatizması’ kelimesi tüm gönderilerin %0.86’sında birlikte kullanılmıştır. Ayrıca ‘romatoid, kalp’ kelimelerinin kullanıldığı gönderilerin %100’ünde ‘kalp romatizması’ kelimesi de kullanılmaktadır.

Çizelge 10.51. İlaç gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar		Sonuç	Destek	Güven	Lift
[fibromiyalji, bel fitiği, boyun fitiği]	→	periferik spondilit	0.0086	1.0000	116.0000
[bel fitiği, kifoz, skolyoz]	→	kalça ödemi	0.0086	1.0000	116.0000
[romatoid, kalp]	→	kalp romatizması	0.0086	1.0000	116.0000
[romatoid, skleroderma]	→	lupus	0.0086	1.0000	116.0000
[romatoid, lupus]	→	skleroderma	0.0086	1.0000	116.0000
[gastrit, hiatal yetmezlik]	→	antral gastrit	0.0086	1.0000	116.0000
[gastrit, antral gastrit]	→	hiatal yetmezlik	0.0086	1.0000	116.0000
[as, kalp]	→	kalp kapakçığı	0.0086	1.0000	116.0000
[as, liken]	→	oral liken	0.0086	1.0000	116.0000

Çizelge 10.51. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[as, oral liken]	→	liken	0.0086	1.0000	116.0000
[as, spondiloartropati]	→	spondiloartrit	0.0086	1.0000	116.0000
[as, spondiloartrit]	→	spondiloartropati	0.0086	1.0000	116.0000
[antral gastrit, hiatal yetmezlik]	→	gastrit	0.0086	1.0000	58.0000
[fibromiyalji, as, skolyoz]	→	böbrek taşı	0.0086	1.0000	58.0000
[bel fitiği, kalça ödemi, skolyoz]	→	kifoz	0.0086	1.0000	58.0000
[as, hipertansiyon]	→	gut	0.0086	1.0000	58.0000
[fibromiyalji, bel fitiği, periferik spondilit]	→	boyun fitiği	0.0086	1.0000	38.6667
[fibromiyalji, as, bel fitiği]	→	astım	0.0086	1.0000	38.6667
[fibromiyalji, as, böbrek taşı]	→	skolyoz	0.0086	1.0000	38.6667
[bel fitiği, kifoz, kalça ödemi]	→	skolyoz	0.0086	1.0000	38.6667
[romatoid, kalp romatizması]	→	kalp	0.0086	1.0000	38.6667
[romatoid, spondilopati, as]	→	skolyoz	0.0086	1.0000	38.6667
[romatoid, as, skolyoz]	→	spondilopati	0.0086	1.0000	38.6667
[as, kalp kapakçığı]	→	kalp	0.0086	1.0000	38.6667
[as, bronşektazi]	→	astım	0.0086	1.0000	38.6667
[kalp romatizması, kalp]	→	romatoid	0.0086	1.0000	16.5714
[spondilopati, as, skolyoz]	→	romatoid	0.0086	1.0000	16.5714
[skleroderma, lupus]	→	romatoid	0.0086	1.0000	16.5714
[fibromiyalji, boyun fitiği, periferik spondilit]	→	bel fitiği	0.0086	1.0000	14.5000
[fibromiyalji, as, astım]	→	bel fitiği	0.0086	1.0000	14.5000
[boyun tutu, as]	→	crohn	0.0086	1.0000	14.5000
[kifoz, kalça ödemi, skolyoz]	→	bel fitiği	0.0086	1.0000	14.5000
[as, boyun fitiği]	→	bel fitiği	0.0086	1.0000	14.5000
[as, anemi]	→	fmf	0.0086	1.0000	10.5455
[bel fitiği, boyun fitiği, periferik spondilit]	→	fibromiyalji	0.0086	1.0000	6.8235

Çizelge 10.51. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[as, astım, bel fitiği]	→	fibromiyalji	0.0086	1.0000	6.8235
[as, böbrek taşı, skolyoz]	→	fibromiyalji	0.0086	1.0000	6.8235
[spondiloartrit, spondilartropati]	→	as	0.0086	1.0000	1.2211
[hipertansiyon, gut]	→	as	0.0086	1.0000	1.2211
[spondilopati, behçet]	→	as	0.0086	1.0000	1.2211
[fibromiyalji, fmf]	→	as	0.0086	1.0000	1.2211
[fibromiyalji, astım, bel fitiği]	→	as	0.0086	1.0000	1.2211
[fibromiyalji, böbrek taşı, skolyoz]	→	as	0.0086	1.0000	1.2211
[boyun tutu, crohn]	→	as	0.0086	1.0000	1.2211
[romatoid, spondilopati, skolyoz]	→	as	0.0086	1.0000	1.2211
[liken, oral liken]	→	as	0.0086	1.0000	1.2211
[anemi, fmf]	→	as	0.0086	1.0000	1.2211
[astım, bronşektazi]	→	as	0.0086	1.0000	1.2211
[astım, bronşit]	→	as	0.0086	1.0000	1.2211
[kolit, crohn]	→	as	0.0086	1.0000	1.2211
[sedef, crohn]	→	as	0.0086	1.0000	1.2211
[üveit, gut]	→	as	0.0086	1.0000	1.2211
[üveit, crohn]	→	as	0.0172	1.0000	1.2211
[kalp, kalp kapakçığı]	→	as	0.0086	1.0000	1.2211
[ms, crohn]	→	as	0.0086	1.0000	1.2211

10.13. Rapor Kategorisine ait Sonuçlar

Rapor kategorisine ait toplam 536 adet gönderiye veri ön işlem aşaması uygulandıktan sonra gönderileri temsil eden toplam 1659 adet kelime kökü bulunmuştur. Bu kelimeler terim frekansı ve binary ağırlıklandırma yöntemleri ile ağırlıklandırılmıştır. 2'den aşağı frekans değerine sahip olan kelimeler elenmiştir. Tek başına bir anlam ifade etmeyen gereksiz kelimeler de listeden çıkarılmıştır. Eleme sonucunda toplam 74 anahtar kelime belirlemiştir. Kullanıcılar tarafından rapor kategorisinde yıllara göre en çok kullanılan 50 kelime Çizelge 10.52'de verilmiştir.

Çizelge 10.52. Rapor kategorisine ait gönderiler için yıllara göre en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	rapor, emekli, hasta, asker, malulen, süre, engel, çalış, iş, heyet, gir, yaşa, muaf, memur, devlet, tecrübe, nöbet, kadro
2013	rapor, gün, emekli, engel, oran, sigorta, sağlık, kurul, çalış, indir, asker, vergi, kanun, hastane, hasta, başvur, yaş, ehliyet, yardım, sonuç, günü, yararlan, malulen, gir, belge, sosyal, prim, iş, memur, tedavi, maaş, aylık, işe, heyet, süre, yaşa, tecrübe, devlet, muaf, askeri, deęiş, otur, kadro, muayene, ücret, çık, başvuru, kart, süreç, ilaç
2014	rapor, engel, iş, kart, emekli, asker, sağlık, çalış, vergi, ehliyet, heyet, devlet, hasta, yardım, başvuru, ilaç, gün, hastane, sonuç, gir, yaşa, muayene, randevu, kurul, indir, tedavi, sevk, tayin, oran, günü, işe, muaf, ücret, çık, sürekli, kimlik, başvur, yararlan, malulen, belge, sosyal, kadro, tanı, kurula, muafiyet, maaş, süre, deęiş, otur, süreç
2015	ehliyet, rapor, sürücü, sağlık, muayene, asker, gün, hastane, hasta, indir, tedavi, engel, heyet, vergi, sonuç, emekli, tanı, süre, çalış, yaşa, sevk, tayin, oran, günü, muaf, çık, iş, başvuru, ilaç, işe, sürekli, muafiyet, kronik, polis, devlet, yardım, gir, başvur, yararlan, malulen, belge, deęiş, işlem, giriş, memur, evrak
2016	rapor, engel, emekli, ehliyet, sağlık, hastane, tedavi, iş, hasta, yaş, gün, indir, çalış, sevk, tayin, yardım, askeri, muayene, asker, heyet, vergi, yaşa, muaf, çık, ilaç, gir, yararlan, malulen, randevu, kurul, nöbet, günü, başvuru, işe, muafiyet, polis, memur, ücret, ertele
2017	engel, rapor, indir, hasta, asker, emekli, çalış, vergi, hastane, ücret, oran, heyet, yardım, gir, yaş, sonuç, maaş, muayene, muafiyet, aylık, ehliyet, iş, yaşa, memur, sağlık, muaf, gün, malulen, devlet, yararlan, evrak, günü, işe, başvur, belge, sevk, çık, kurul, nöbet, başvuru, polis, sosyal, tecrübe, tazminat, sürekli, giriş, otur, süreç, tedavi, ilaç
2018	rapor, engel, emekli, çalış, hasta, hastane, asker, indir, yardım, gün, vergi, ehliyet, sağlık, iş, başvur, oran, heyet, tedavi, muayene, malulen, süreç, gir, sonuç, yaşa, çık, kurul, tazminat, sevk, işlem, ertele, memur, muaf, yaş, maaş, devlet, yararlan, otur, süre, günü, başvuru, ayrıl, sigorta, bedel, nöbet, sürekli, giriş, prim, askeri, kurula, çürük

Belirlenen 74 anahtar kelimeyi frekans deęerlerine göre göstermek amacıyla kelime bulutu oluşturulmuştur. Şekil 10.69'daki kelime bulutuna göre 'rapor', 'emekli', 'asker', 'gün', 'ehliyet' gibi kelimelerin rapor kategorisine ait gönderilerde sıklıkla geçtięi görülmektedir.

Çizelge 10.53. Rapor kategorisini temsil eden kelimeler için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	74	0.42±1.03	0.00 (0.00-7.00) ^f	1.92		
2013	74	10.08±15.34	5.50 (0.00-89.00) ^e	5.10		
2014	74	3.70±5.22	2.00 (0.00-37.00) ^{bde}	4.05		
2015	74	3.66±7.06	1.00 (0.00-35.00) ^{cd}	3.33	238.26	<0.001
2016	74	1.93±3.53	1.00 (0.00-23.00) ^{cf}	2.63		
2017	74	7.68±12.66	3.00 (0.00-78.00) ^{be}	4.75		
2018	74	12.38±18.13	7.00 (0.00-132.00) ^a	6.22		

a-f; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Rapor kategorisine ait anahtar kelimeler için uyum analizi sonuçları

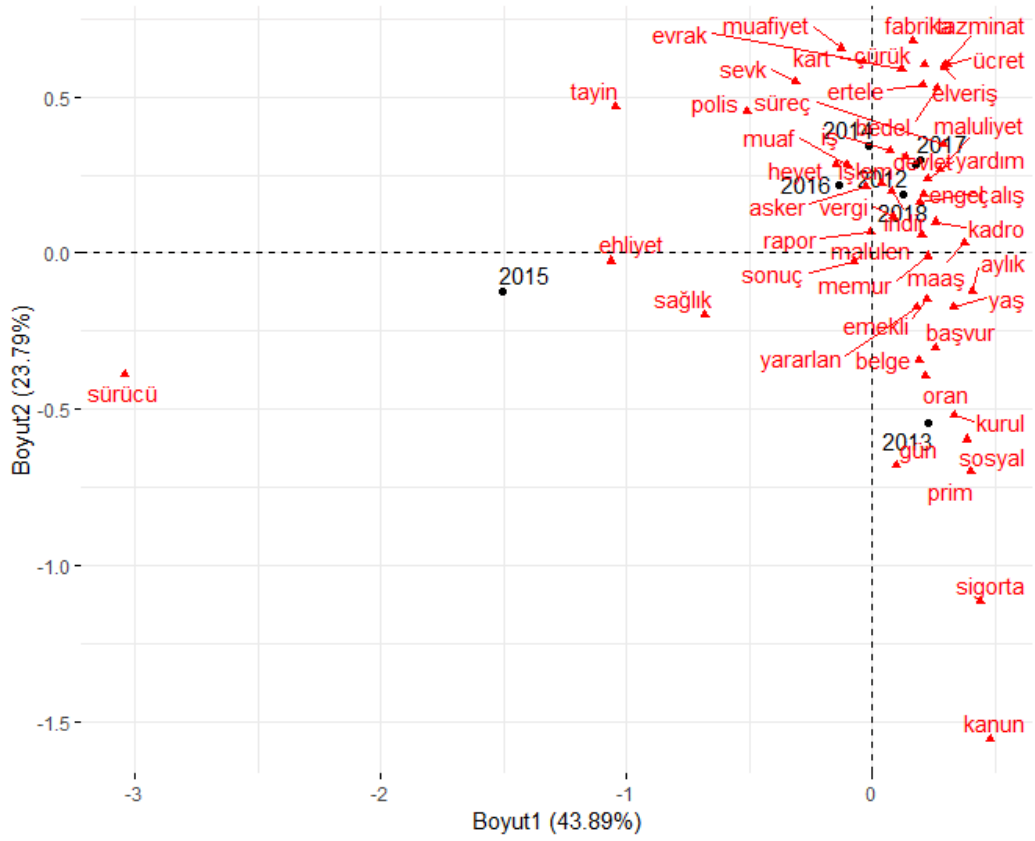
Rapor kategorisini temsil eden anahtar kelimeler ve yıllar arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Kullanılan anahtar kelimeler ve yıllar arasında ilişki yoktur

H_1 : Kullanılan anahtar kelimeler ve yıllar arasında ilişki vardır

R programında 7×48 boyutunda bir matris oluşturulmuştur. 48, toplamda 3 ve üzeri sıklıkta kullanılan anahtar kelimelerin sayısını göstermektedir. 7 ise toplam yılların sayısını göstermektedir. Test sonucunda $p>0.001$ 'ten dolayı %99.9 güvenle H_0 hipotezi reddedilir. Sonuç olarak %99.9 güvenle kullanılan kelimeler ve yıllar arasında bir ilişki vardır.

Şekil 10.70'te basit uyum analizi sonucu verilmiştir. Boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %43.89'ini açıklarken, ikinci boyut ise %23.79'unu açıklamaktadır.



Şekil 10.70. Rapor kategorisine ait kelimeler ve yıllar için basit uyum analizi sonucu

Bu boyutların kullanılan kelimeler ve yılları hangi oranda açıkladığı Çizelge 10.54 ve Çizelge 10.55'te verilmiştir. Çizelge 10.54'te 2015 boyut 1, 2013 boyut 2, 2012 ve 2017 boyut 3, 2014 ve 2018 boyut 4, 2016 yılı ise büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.54. Rapor kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012	0.014	0.037	0.078	0.005	0.027
2013	0.148	0.843	0.000	0.009	0.000
2014	0.000	0.182	0.294	0.478	0.042
2015	0.988	0.007	0.004	0.001	0.001
2016	0.032	0.082	0.022	0.065	0.771
2017	0.110	0.262	0.606	0.018	0.003
2018	0.096	0.211	0.200	0.485	0.004

Çizelge 10.55'te, çalış, ehliyet, emekli, malulen, polis, sağlık, sürücü, yararlan boyut 1, asker, başvuru, belge, gün, heyet, kanun, kurul, muafiyet, muaf, oran, prim,

sevk, sigorta, sosyal, yardım boyut 2, aylık, çürük, engel, evrak, fabrika, indir, iş, kadro, maaş, memur, rapor, ücret boyut 3, bedel, devlet, elveriş, ertele, işlem, kart, süreç, tayin, tazminat boyut 4, maluliyet, sonuç, vergi ve yaş kelimeleri büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.55. Boyutların rapor kategorisine ait kelimeleri açıklama oranları

Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
asker	0.009	0.527	0.070	0.052	0.163
aylık	0.287	0.027	0.659	0.000	0.027
başvur	0.264	0.363	0.080	0.170	0.102
bedel	0.030	0.121	0.199	0.626	0.011
belge	0.158	0.518	0.025	0.112	0.179
çalış	0.369	0.251	0.062	0.281	0.036
çürük	0.028	0.224	0.432	0.229	0.060
devlet	0.036	0.180	0.124	0.348	0.191
ehliyet	0.988	0.001	0.002	0.000	0.004
elveriş	0.064	0.266	0.030	0.606	0.022
emekli	0.506	0.227	0.040	0.014	0.116
engel	0.230	0.181	0.424	0.110	0.040
ertele	0.023	0.151	0.224	0.526	0.043
evrak	0.001	0.356	0.553	0.052	0.037
fabrika	0.016	0.262	0.567	0.000	0.122
gün	0.020	0.947	0.024	0.001	0.000
heyet	0.148	0.543	0.011	0.177	0.040
indir	0.028	0.174	0.741	0.029	0.020
iş	0.015	0.313	0.353	0.269	0.049
işlem	0.003	0.085	0.274	0.534	0.080
kadro	0.025	0.004	0.285	0.081	0.006
kanun	0.085	0.894	0.000	0.020	0.000
kart	0.004	0.098	0.313	0.426	0.139
kurul	0.247	0.592	0.091	0.036	0.004
maaş	0.400	0.003	0.513	0.003	0.081

Çizelge 10.55. (devam)

Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
malulen	0.127	0.010	0.053	0.028	0.100
maluliyet	0.227	0.209	0.182	0.033	0.304
memur	0.207	0.000	0.264	0.087	0.036
muaf	0.062	0.471	0.001	0.040	0.087
muafiyet	0.018	0.478	0.416	0.082	0.002
oran	0.178	0.580	0.146	0.006	0.089
polis	0.342	0.270	0.185	0.075	0.115
prim	0.223	0.674	0.012	0.078	0.009
rapor	0.001	0.102	0.641	0.040	0.135
sağlık	0.874	0.075	0.016	0.002	0.017
sevk	0.149	0.461	0.191	0.008	0.108
sigorta	0.133	0.863	0.001	0.000	0.002
sonuç	0.064	0.008	0.081	0.092	0.729
sosyal	0.195	0.473	0.013	0.261	0.055
süreç	0.102	0.149	0.168	0.509	0.050
sürücü	0.960	0.016	0.011	0.003	0.008
tayin	0.284	0.057	0.120	0.339	0.162
tazminat	0.071	0.296	0.013	0.585	0.024
ücret	0.091	0.351	0.431	0.115	0.009
vergi	0.143	0.276	0.136	0.009	0.304
yararlan	0.319	0.299	0.002	0.043	0.204
yardım	0.253	0.278	0.138	0.232	0.002
yaş	0.257	0.069	0.301	0.006	0.340

Rapor gönderileri için birliktelik analizi kuralı sonuçları

Gönderiler içerisinde 2'den aşağı anahtar kelime barındıran gönderiler elenmiştir. Analizde toplam 418 gönderi kullanılmıştır. Birliktelik kurallarını çıkarmak için Apriori algoritması kullanılmıştır. Minimum destek değeri 0.012 ve minimum güven değeri 0.7 olarak belirlenmiş ve model çalıştırılmıştır. Çalışma sonucunda toplam 96 kural bulunmuştur. Bu kurallar Çizelge 10.56 ve Çizelge 10.57'de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.56'da 7.3150 lift değeri ile en ilginç kural 'kurul → sağlık' kuralıdır. 'kurul' ve 'sağlık' kelimesi tüm gönderilerin %5.02'sinde birlikte kullanılmıştır. Ayrıca 'kurul' kelimesinin kullanıldığı gönderilerin %84'ünde 'sağlık' kelimesi de kullanılmaktadır.

Lift değeri 6.4718 olan 'sevk → hastane' kuralında, 'sevk' ve 'hastane' kelimesi tüm gönderilerin %4.07'sinde birlikte kullanılmıştır. Ayrıca 'sevk' kelimesinin kullanıldığı gönderilerin %94.44'ünde 'hastane' kelimesi de kullanılmaktadır.

Lift değeri 6,1926 olan 'ayrıl → çalış' kuralında, 'ayrıl' ve 'çalış' kelimesi tüm gönderilerin %1.91'inde birlikte kullanılmıştır. Ayrıca 'ayrıl' kelimesinin kullanıldığı gönderilerin %88.89'unda 'çalış' kelimesi de kullanılmıştır.

Lift değeri 5.5982 olan 'vergi → indir' kuralında, 'vergi' ve 'indir' kelimesi tüm gönderilerin %11.48'inde birlikte kullanılmıştır. Ayrıca 'vergi' kelimesinin kullanıldığı gönderilerin %85.71'inde 'indir' kelimesi de kullanılmıştır.

Çizelge 10.56. Rapor gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[kurul]	→ sağlık	0.0502	0.8400	7.3150
[sevk]	→ hastane	0.0407	0.9444	6.4718
[ayrıl]	→ çalış	0.0191	0.8889	6.1926
[vergi]	→ indir	0.1148	0.8571	5.5982
[indir]	→ vergi	0.1148	0.7500	5.5982
[askeri]	→ hastane	0.0191	0.8000	5.4820
[malulen]	→ emekli	0.0742	1.0000	4.4946
[maaş]	→ emekli	0.0383	0.7619	3.4245
[sosyal]	→ engel	0.0215	0.9000	3.1881
[kart]	→ engel	0.0191	0.7273	2.5763
[ücret]	→ rapor	0.0191	0.8889	1.7044
[otur]	→ rapor	0.0191	0.8889	1.7044
[süre]	→ rapor	0.0287	0.8571	1.6435
[sürekli]	→ rapor	0.0191	0.8000	1.5339
[oran]	→ rapor	0.0622	0.7879	1.5107
[tecrübe]	→ rapor	0.0215	0.7500	1.4381

Çizelge 10.56. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[kurul]	→	rapor	0.0431	0.7200	1.3806
[engel]	→	rapor	0.2010	0.7119	1.3650
[başvur]	→	rapor	0.0455	0.7037	1.3493
[iş]	→	rapor	0.0670	0.7000	1.3422

Çizelge 10.57’de üç ve üzeri birliktelik kurallarına göre Lift değeri 14.8622 olan ‘vergi, sağlık → kurul’ kuralında ‘vergi, sağlık’ kelime grubu ve ‘kurul’ kelimesi tüm gönderilerin %1.91’inde birlikte kullanılmıştır. Ayrıca ‘vergi, sağlık’ kelimelerinin kullanıldığı gönderilerin %88.89’unda ‘kurul’ kelimesi de kullanılmaktadır.

Lift değeri 13.3760 olan ‘indir, sağlık→ kurul’ kuralında ‘indir, sağlık’ kelime grubu ve ‘kurul’ kelimesi tüm gönderilerin %1.91’inde birlikte kullanılmıştır. Ayrıca ‘indir, sağlık’ kelimelerinin kullanıldığı gönderilerin %80’inde ‘kurul’ kelimesi de kullanılmaktadır.

Lift değeri 12.2613 olan ‘hastane, sağlık→ kurul’ kuralında ‘hastane, sağlık’ kelime grubu ve ‘kurul’ kelimesi tüm gönderilerin %2.63’ünde birlikte kullanılmıştır. Ayrıca ‘hastane, sağlık’ kelimelerinin kullanıldığı gönderilerin %73.33’ünde ‘kurul’ kelimesi de kullanılmaktadır.

Çizelge 10.57. Rapor gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralları

Gruplar		Sonuç	Destek	Güven	Lift
[vergi, sağlık]	→	kurul	0.0191	0.8889	14.8622
[indir, sağlık]	→	kurul	0.0191	0.8000	13.3760
[hastane, sağlık]	→	kurul	0.0263	0.7333	12.2613
[indir, kurul]	→	sağlık	0.0191	1.0000	8.7083
[kurul, vergi]	→	sağlık	0.0191	1.0000	8.7083
[kurul, sonuç]	→	sağlık	0.0191	1.0000	8.7083
[rapor, kurul, hastane]	→	sağlık	0.0215	1.0000	8.7083
[kurul, engel]	→	sağlık	0.0263	0.9167	7.9826
[rapor, kurul, engel]	→	sağlık	0.0239	0.9091	7.9167
[rapor, kurul]	→	sağlık	0.0383	0.8889	7.7407

Çizelge 10.57. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[emekli, indir, çalış]	→	vergi	0.0191	1.0000	7.4643
[indir, memur]	→	vergi	0.0191	1.0000	7.4643
[kurul, hastane]	→	sağlık	0.0263	0.8462	7.3686
[emekli, indir]	→	vergi	0.0478	0.9524	7.1088
[indir, hastane]	→	vergi	0.0263	0.9167	6.8423
[yardım, indir]	→	vergi	0.0239	0.9091	6.7857
[yararlan, indir]	→	vergi	0.0215	0.9000	6.7179
[rapor, emekli, indir]	→	vergi	0.0215	0.9000	6.7179
[hastane, vergi]	→	indir	0.0263	1.0000	6.5313
[emekli, çalış, vergi]	→	indir	0.0191	1.0000	6.5313
[emekli, vergi]	→	indir	0.0478	1.0000	6.5313
[vergi, memur]	→	indir	0.0191	1.0000	6.5313
[hasta, vergi]	→	indir	0.0215	1.0000	6.5313
[rapor, emekli, vergi]	→	indir	0.0215	1.0000	6.5313
[gün, vergi]	→	indir	0.0191	1.0000	6.5313
[indir, çalış]	→	vergi	0.0263	0.8462	6.3159
[rapor, sevk]	→	hastane	0.0215	0.9000	6.1672
[hasta, indir]	→	vergi	0.0215	0.8182	6.1071
[indir, sağlık]	→	vergi	0.0191	0.8000	5.9714
[yararlan, vergi]	→	indir	0.0215	0.9000	5.8781
[vergi, sağlık]	→	indir	0.0191	0.8889	5.8056
[rapor, indir]	→	vergi	0.0574	0.7500	5.5982
[rapor, engel, vergi]	→	çalış	0.0191	0.8000	5.5733
[yardım, vergi]	→	indir	0.0239	0.8333	5.4427
[rapor, vergi]	→	indir	0.0574	0.8276	5.4052
[engel, vergi]	→	indir	0.0215	0.8182	5.3438
[rapor, engel, vergi]	→	indir	0.0191	0.8000	5.2250
[çalış, vergi]	→	indir	0.0263	0.7333	4.7896
[indir, çalış, vergi]	→	emekli	0.0191	0.7273	3.2688
[rapor, çalış, vergi]	→	engel	0.0191	0.8889	3.1488
[indir, sağlık]	→	engel	0.0191	0.8000	2.8339
[rapor, muayene]	→	engel	0.0191	0.7273	2.5763

Çizelge 10.57. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[oran, sağlık]	→	engel	0.0191	0.7273	2.5763
[muayene, engel]	→	rapor	0.0191	1.0000	1.9174
[hastane, engel]	→	rapor	0.0287	1.0000	1.9174
[çalış, engel, vergi]	→	rapor	0.0191	1.0000	1.9174
[engel, sonuç]	→	rapor	0.0215	1.0000	1.9174
[yaşa, sağlık]	→	rapor	0.0215	1.0000	1.9174
[emekli, oran]	→	rapor	0.0263	0.9167	1.7576
[kurul, engel]	→	rapor	0.0263	0.9167	1.7576
[engel, vergi]	→	rapor	0.0239	0.9091	1.7431
[kurul, engel, sağlık]	→	rapor	0.0239	0.9091	1.7431
[indir, oran]	→	rapor	0.0215	0.9000	1.7257
[indir, engel, vergi]	→	rapor	0.0191	0.8889	1.7044
[oran, vergi]	→	rapor	0.0191	0.8889	1.7044
[hastane, sağlık]	→	rapor	0.0311	0.8667	1.6618
[indir, engel]	→	rapor	0.0287	0.8571	1.6435
[başvur, engel]	→	rapor	0.0239	0.8333	1.5979
[hasta, engel]	→	rapor	0.0359	0.8333	1.5979
[yardım, engel]	→	rapor	0.0335	0.8235	1.5791
[kurul, hastane, sağlık]	→	rapor	0.0215	0.8182	1.5688
[engel, gün]	→	rapor	0.0311	0.8125	1.5579
[engel, sağlık]	→	rapor	0.0407	0.8095	1.5522
[hasta, sağlık]	→	rapor	0.0287	0.8000	1.5339
[heyet, engel]	→	rapor	0.0191	0.8000	1.5339
[indir, sağlık]	→	rapor	0.0191	0.8000	1.5339
[oran, gün]	→	rapor	0.0191	0.8000	1.5339
[gün, sağlık]	→	rapor	0.0239	0.7692	1.4749
[kurul, sağlık]	→	rapor	0.0383	0.7619	1.4609
[hastane, sonuç]	→	rapor	0.0215	0.7500	1.4381
[indir, gün]	→	rapor	0.0215	0.7500	1.4381
[sonuç, sağlık]	→	rapor	0.0215	0.7500	1.4381
[ehliyet, sağlık]	→	rapor	0.0191	0.7273	1.3945
[hasta, iş]	→	rapor	0.0191	0.7273	1.3945
[hasta, gün]	→	rapor	0.0239	0.7143	1.3696
[oran, engel]	→	rapor	0.0239	0.7143	1.3696

10.14. Kişisel Deneyim ve Diyet Kategorisine ait Sonuçlar

Kişisel deneyim ve diyet gönderi sınıfına ait toplam 395 gönderi analiz edilmiştir. Metin ön işlem aşamasından sonra toplam 2718 tane kelime bulunmuştur. 5'ten aşağı sıklıkta geçen kelimeler elenmiştir. Durak kelimeler ve gereksiz kelimeler listeden çıkarılmıştır. İşlemler sonucunda toplam 242 tane özellik belirlenmiştir. Yıllara göre en çok kullanılan 50 kelime Çizelge 10.58'de verilmiştir.

Çizelge 10.58. Kişisel deneyim kategorisine ait gönderilerde en çok kullanılan 50 kelime

Yıl	Kelimeler
2012	hastalık, kullan, doktor, iğne, yap, ağrı, ilaç, yan, spor, hayat, tedavi, etki, sağlık, unut, kanser, verem, iyi, hasta, tavsiye, atak, yeni, fayda, egzersiz, hastane, gece, mutlu, sev, beyin, sistem, eski, ilaçla, fitik, bel, teşhis, vücut, yardım, süt, sürekli, dönem, sırt, fizik, tanı, sevgi, süreç, hücre, göz, eklem, yorum, rahatla, omurga
2013	ağrı, yap, kullan, ilaç, doktor, hastalık, iyi, tedavi, sigara, hasta, spor, bel, hareket, hayat, egzersiz, boyun, yeni, fayda, sırt, iltihap, teşhis, yan, etki, kalça, yastık, hastane, eski, yardım, baş, dil, uzun, özel, hisset, ayak, iğne, tavsiye, mutlu, ilaçla, sürekli, şeker, uygula, umut, boy, paket, bitki, masaj, sağlık, unut, dönem, tanı
2014	hastalık, ağrı, yap, ameliyat, kullan, iyi, hasta, yeni, ilaç, hayat, doktor, egzersiz, sonuç, tedavi, teşhis, etki, baş, mutlu, moral, vücut, besle, spor, bel, uzak, yemek, yüksek, dil, sağlık, test, tatlı, sigara, fayda, kalça, hastane, özel, iğne, tavsiye, önem, gece, kontrol, sistem, hareket, sırt, yan, eski, yardım, sürekli, şeker, unut, tanı
2015	hastalık, diyet, yap, ağrı, iyi, hasta, ilaç, kullan, iltihap, bağırsak, doktor, hayat, besle, yeni, kronik, sağlık, gıda, sev, yardım, yiyecek, uzun, sonuç, tedavi, teşhis, baş, dil, gece, şeker, dönem, yan, boyun, uygula, artrit, iltihabi, spor, fayda, hastane, özel, sistem, tanı, normal, süt, içecek, aşama, protein, kalça, önem, süreç, acı, vitamin
2016	yap, besle, ağrı, hastalık, iyi, ilaç, kullan, doktor, spor, vücut, hücre, bel, etki, tedavi, sağlık, yan, diyet, hasta, hayat, boyun, şeker, bağırsak, gıda, uzun, fayda, egzersiz, mutlu, sürekli, sonuç, kan, yağ, enerji, tavsiye, baş, dil, hastane, önem, eklem, unut, stres, hareket, doğal, çikolata, fizik, hormon, sistem, yemek, iğne, hisset, psikolojik
2017	ağrı, yap, hastalık, doktor, ilaç, kullan, spor, iyi, tedavi, hasta, hayat, iğne, besle, tavsiye, teşhis, bel, sağlık, baş, hareket, fizik, vücut, yeni, sevgi, iltihap, hisset, eklemek, etki, yan, önem, özel, boyun, fayda, değer, beyin, kalça, süt, diyet, şeker, egzersiz, kan, uzak, yardım, acı, yoğurt, uzun, sürekli, hastane, yemek, kilo, dönem
2018	ağrı, yap, hastalık, ilaç, doktor, spor, iyi, kullan, sigara, hayat, bel, hasta, besle, ayak, teşhis, sağlık, hastane, tedavi, baş, süt, fayda, dil, sonuç, sırt, etki, tavsiye, hareket, yeni, yan, yardım, acı, gece, uzun, kilo, vücut, beyin, kalça, diyet, dost, ameliyat, romatizma, sev, besin, hisset, önem, egzersiz, kan, uzak, kas, muayene

Çizelge 10.59. Kişisel deneyim ve diyet kategorisini temsil eden kelimeler için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	242	0.73 ± 1.81	0.00 (0.00 – 12.00) ^g	2.06		
2013	242	2.53 ± 4.91	1.00 (0.00 – 35.00) ^f	3.33		
2014	242	3.53 ± 5.87	2.00 (0.00 – 44.00) ^e	4.18		
2015	242	5.14 ± 8.66	3.00 (0.00 – 75.00) ^{da}	4.86	383.07	<0.001
2016	242	3.78 ± 6.28	2.00 (0.00 – 42.00) ^{ce}	4.06		
2017	242	6.10 ± 11.27	3.00 (0.00 – 102.00) ^{bd}	5.21		
2018	242	4.09 ± 7.98	2.00 (0.00 – 78.00) ^{ace}	4.31		

a-g; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Gönderileri temsil eden kelimeler için birliktelik analizi sonuçları

Toplam veri dosyası, 395’dir. 2’den aşağı anahtar kelimesi barındıran metin dosyaları elenmiştir. Analizde toplam 346 metin dosyası kullanılmıştır. Birliktelik kuralları için Apriori algoritması uygulanmıştır. Minimum destek değeri 0.07 ve minimum güven değeri 0.7 olarak belirlenmiştir. Toplam 65 tane kural bulunmuştur. Bu kurallar Çizelge 10.60 ve Çizelge 10.61’de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.60’da 2.3537 lift değeri ile en ilginç kural ‘egzersiz → kullan’ kuralıdır. ‘egzersiz’ ve ‘kullan’ kelimesi tüm gönderilerin %7.23’ünde birlikte kullanılmıştır. Ayrıca ‘egzersiz’ kelimesinin kullanıldığı gönderilerin %71.43’ünde ‘kullan’ kelimesi de kullanılmaktadır.

Lift değeri 2.3128 olan ‘uzun → iyi’ kuralında, ‘uzun’ ve ‘iyi’ kelimesi tüm gönderilerin %7.23’ünde birlikte kullanılmıştır. Ayrıca ‘uzun’ kelimesinin kullanıldığı gönderilerin %73.53’ünde ‘iyi’ kelimesi de kullanılmaktadır.

Lift değeri 2.1502 olan ‘sağlık → hastalık’ kuralında, ‘sağlık’ ve ‘hastalık’ kelimesi tüm gönderilerin %10.12’sinde birlikte kullanılmıştır. Ayrıca ‘sağlık’ kelimesinin kullanıldığı gönderilerin %79.55’inde ‘hastalık’ kelimesi de kullanılmıştır.

Lift değeri 2.1457 olan ‘egzersiz → yap’ kuralında, ‘egzersiz’ ve ‘yap’ kelimesi tüm gönderilerin %8.09’unda birlikte kullanılmıştır. Ayrıca ‘egzersiz’ kelimesinin kullanıldığı gönderilerin %80’inde ‘yap’ kelimesi de kullanılmıştır.

Çizelge 10.60. Kişisel deneyim ve diyet gönderi kategorisine ait ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift
[egzersiz] →	kullan	0.0723	0.7143	2.3537
[uzun] →	iyi	0.0723	0.7353	2.3128
[sağlık] →	hastalık	0.1012	0.7955	2.1502
[egzersiz] →	yap	0.0809	0.8000	2.1457
[teşhis] →	ağrı	0.1040	0.7826	1.9765
[bel] →	ağrı	0.1012	0.7609	1.9216
[spor] →	yap	0.1156	0.7143	1.9158

Çizelge 10.61’de üç ve üzeri birliktelik kurallarına göre Lift değeri 3.3491 olan ‘doktor, ağrı, kullan → ilaç’ kuralında ‘doktor, ağrı, kullan’ kelime grubu ve ‘ilaç’ kelimesi tüm gönderilerin %9.25’inde birlikte kullanılmıştır. Ayrıca ‘doktor, ağrı, kullan’ kelimelerinin kullanıldığı gönderilerin %84.21’inde ‘ilaç’ kelimesi de kullanılmaktadır.

Lift değeri 3.1334 olan ‘doktor, yap, kullan→ ilaç’ kuralında ‘doktor, yap, kullan’ kelime grubu ve ‘ilaç’ kelimesi tüm gönderilerin %7.51’inde birlikte kullanılmıştır. Ayrıca ‘doktor, yap, kullan’ kelimelerinin kullanıldığı gönderilerin %78.79’unda ‘ilaç’ kelimesi de kullanılmaktadır.

Lift değeri 3.1021 olan ‘doktor, kullan→ ilaç’ kuralında ‘doktor, kullan’ kelime grubu ve ‘ilaç’ kelimesi tüm gönderilerin %11.27’sinde birlikte kullanılmıştır. Ayrıca ‘doktor, kullan’ kelimelerinin kullanıldığı gönderilerin %78’inde ‘ilaç’ kelimesi de kullanılmaktadır.

Çizelge 10.61. Kişisel deneyim ve diyet gönderi kategorisine ait üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar		Sonuç	Destek	Güven	Lift
[doktor, ağrı, kullan]	→	ilaç	0.0925	0.8421	3.3491
[doktor, yap, kullan]	→	ilaç	0.0751	0.7879	3.1334
[doktor, kullan]	→	ilaç	0.1127	0.7800	3.1021
[ilaç, yap, kullan]	→	doktor	0.0751	0.7647	3.0766
[hastalık, ağrı, kullan]	→	ilaç	0.0838	0.7632	3.0351
[iyi, ağrı, kullan]	→	ilaç	0.0809	0.7568	3.0096
[ilaç, yap, ağrı]	→	doktor	0.0751	0.7429	2.9887
[ilaç, ağrı, kullan]	→	doktor	0.0925	0.7273	2.9260
[doktor, ağrı]	→	ilaç	0.1069	0.7255	2.8853
[iyi, ilaç, ağrı]	→	kullan	0.0809	0.8750	2.8833
[yap, ağrı, kullan]	→	ilaç	0.0838	0.7250	2.8833
[doktor, ilaç, yap]	→	kullan	0.0751	0.8667	2.8559
[doktor, ilaç, ağrı]	→	kullan	0.0925	0.8649	2.8499
[yap, ağrı, kullan]	→	doktor	0.0809	0.7000	2.8163
[doktor, yap, ağrı]	→	ilaç	0.0751	0.7027	2.7947
[doktor, ilaç]	→	kullan	0.1127	0.8478	2.7938
[ilaç, yap, ağrı]	→	kullan	0.0838	0.8286	2.7303
[ilaç, ağrı]	→	kullan	0.1272	0.8000	2.6362
[ilaç, hastalık, ağrı]	→	kullan	0.0838	0.7838	2.5828
[doktor, yap, ağrı]	→	kullan	0.0809	0.7568	2.4937
[spor, iyi]	→	yap	0.0723	0.9259	2.4835
[iyi, ilaç]	→	kullan	0.0867	0.7500	2.4714
[doktor, ağrı]	→	kullan	0.1098	0.7451	2.4553
[yap, sağlık]	→	hastalık	0.0723	0.8929	2.4135
[ilaç, hastalık]	→	kullan	0.1012	0.7292	2.4028
[ilaç, yap]	→	kullan	0.0983	0.7234	2.3838
[iyi, ilaç, kullan]	→	ağrı	0.0809	0.9333	2.3572
[ağrı, sağlık]	→	hastalık	0.0723	0.8621	2.3303
[hayat, yap]	→	iyi	0.0780	0.7105	2.2349
[doktor, ilaç, yap]	→	ağrı	0.0751	0.8667	2.1888
[ilaç, yap, kullan]	→	ağrı	0.0838	0.8529	2.1541

Çizelge 10.61. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[doktor, iyi]	→	yap	0.0809	0.8000	2.1457
[doktor, yap, kullan]	→	ağrı	0.0809	0.8485	2.1429
[tedavi, ilaç]	→	hastalık	0.0723	0.7813	2.1118
[ilaç, hastalık, kullan]	→	ağrı	0.0838	0.8286	2.0926
[doktor, ilaç, kullan]	→	ağrı	0.0925	0.8205	2.0722
[hayat, iyi]	→	yap	0.0780	0.7714	2.0691
[tedavi, kullan]	→	ağrı	0.0723	0.8065	2.0367
[doktor, ilaç]	→	ağrı	0.1069	0.8043	2.0314
[iyi, ilaç]	→	ağrı	0.0925	0.8000	2.0204
[doktor, ağrı, kullan]	→	yap	0.0809	0.7368	1.9763
[doktor, iyi]	→	ağrı	0.0780	0.7714	1.9483
[doktor, yap]	→	ağrı	0.1069	0.7708	1.9468
[ilaç, hastalık]	→	ağrı	0.1069	0.7708	1.9468
[iyi, kullan]	→	ağrı	0.1069	0.7708	1.9468
[doktor, ağrı]	→	yap	0.1069	0.7255	1.9459
[tedavi, ağrı]	→	hastalık	0.0780	0.7105	1.9206
[hayat, yap]	→	hastalık	0.0780	0.7105	1.9206
[doktor, kullan]	→	ağrı	0.1098	0.7600	1.9194
[hastalık, sağlık]	→	yap	0.0723	0.7143	1.9158
[yap, kullan]	→	ağrı	0.1156	0.7547	1.9061
[iyi, ilaç]	→	hastalık	0.0809	0.7000	1.8922
[doktor, ilaç, ağrı]	→	yap	0.0751	0.7027	1.8848
[ilaç, kullan]	→	ağrı	0.1272	0.7458	1.8835
[ilaç, yap]	→	ağrı	0.1012	0.7447	1.8807
[doktor, hastalık]	→	ağrı	0.0896	0.7381	1.8641
[hastalık, kullan]	→	ağrı	0.1098	0.7308	1.8456
[hastalık, sağlık]	→	ağrı	0.0723	0.7143	1.8040

10.15. Spor Kategorisine Ait Sonuçlar

Spor gönderi sınıfına ait toplam 855 gönderi analiz edilmiştir. 5'ten aşağı sıklıkta geçen kelimeler elenmiştir. Durak kelimeler ve gereksiz kelimeler listeden

çıkarılmıştır. İşlemler sonucunda toplam 255 tane özellik belirlenmiştir. Yıllara göre en çok kullanılan 50 kelime Çizelge 10.61’da verilmiştir.

Çizelge 10.62. Spor kategorisine ait gönderilerde yıllara göre en çok kullanılan 50 kelime

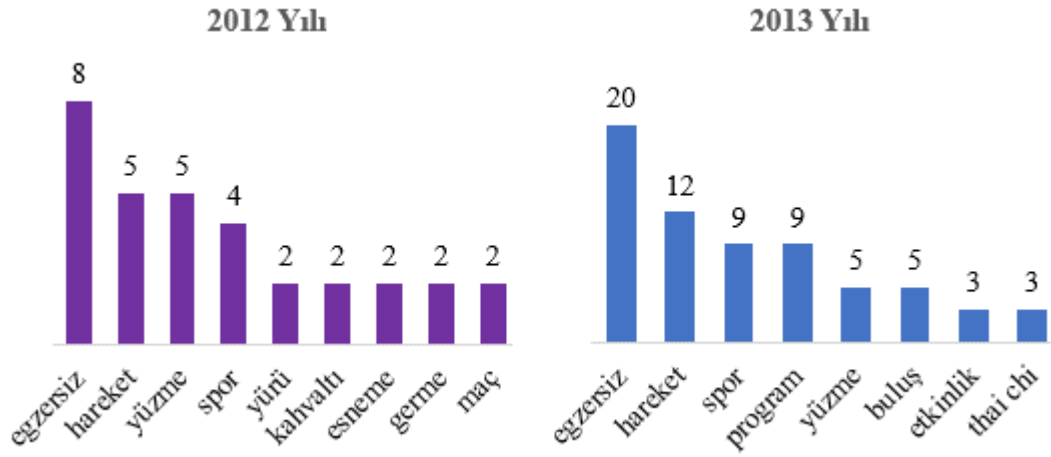
Yıl	Kelimeler
2012	yap, egzersiz, ağrı, hareket, ilaç, katıl, güzel, toplantı, yüz, başla, git, kullan, tedavi, fayda, kahvaltı, zaman, doktor, fazla, öğren, hastalık, çalış, sağlık, saat, yer, yürü, katılım, yad, uzun, yol, hasta, sabah, özel, fizik, teşhis, yatak, gir, yoga, acı, kötü, yaş, organize, yapı, nefes, önem, etki, kas, tavsiye, kişi, organizasyon, topla
2013	yap, egzersiz, program, toplantı, hareket, hastalık, katıl, ilaç, ağrı, git, hasta, saat, yüz, çalış, güzel, başla, kullan, paylaş, salon, tedavi, fazla, katılım, etki, hafta, buluş, etkin, soru, doktor, yer, yad, sabah, önem, erdem, burs, yaklaşık, zaman, yol, yapı, kişi, topla, emek, baş, yan, bel, yardım, omurga, mücadele, sor, bölüm, öğren
2014	yap, egzersiz, hastalık, ağrı, hafta, hareket, program, pazar, saat, yapı, yüz, başla, sabah, çalış, salon, ayak, fizik, bedeni, güzel, omurga, ücret, katıl, zaman, git, paylaş, tedavi, özel, şiş, ilaç, hasta, kullan, zor, kas, uyan, yavaş, toplantı, etki, yer, önem, baş, öğren, sağlık, büyük, yönelik, ölçü, boyun, göster, uzun, yatak, çek
2015	yap, egzersiz, başla, nefes, ağrı, çalış, hareket, kahvaltı, yer, hastalık, sabah, yüz, git, pazar, toplantı, sağlık, güzel, salon, zaman, hafta, hasta, göğüs, saat, ayak, katıl, paylaş, özel, buluş, program, fayda, eklem, sopa, şiş, ilaç, yatak, kişi, vücut, teknik, yapı, önem, baş, kaldır, etkin, destek, omuz, omurga, etki, yad, bel, kalça
2016	yap, egzersiz, ağrı, başla, hareket, yüz, hafta, çalış, yer, ilaç, toplantı, hasta, hastalık, güzel, kullan, zaman, salon, saat, yürü, sağlık, yapı, git, paylaş, buluş, vücut, zor, sonuç, doktor, etkin, kalk, gece, fayda, gerçek, havuz, yad, tedavi, büyük, sırt, çık, yardım, boyun, üst, esne, alet, kahvaltı, sabah, katıl, destek, kas, tavsiye
2017	yap, ağrı, egzersiz, hastalık, başla, hareket, yüz, zaman, sağlık, git, ilaç, kullan, yapı, yürü, paylaş, hasta, güzel, fayda, teşhis, kalk, tedavi, tavsiye, ağır, yardım, sabah, fizik, doktor, yan, uzun, uyu, bacak, bel, salon, saat, sağ, şifa, program, yaşam, moral, yer, boyun, görüş, özel, fazla, koş, eklem, kilo, önem, etkin, sırt
2018	yap, hasta, yüz, egzersiz, hastalık, başla, ağrı, teşhis, hareket, fayda, tecrübe, doktor, önem, buluş, toplantı, katılım, ilaç, paylaş, çek, havuz, sor, kış, sağlık, kullan, güzel, dernek, baş, katıl, gelir, mücadele, öğren, kambur, yad, kayıt, zaman, yürü, tedavi, yardım, sabah, bel, yaşam, yer, görüş, sırt, kişi, kalça, hafta, vücut, zor, otur

Anahtar kelimelerin tamamını göstermek amacıyla kelime bulutu oluşturulmuştur. Şekil 10.72’de ‘yüz’, ‘egzersiz’, ‘hareket’, ‘başla’, ‘kullan’ gibi kelimelerin gönderi içerisinde daha çok geçtiği görülmektedir.



Şekil 10.72. Spor kategorisine ait gönderiler için bulunan toplam 255 kelimeyi gösteren kelime bulutu

Spor kategorisine ait 855 gönderi analiz edilmiştir. Spor ve etkinlikle ilgili gönderilerde paylaşılan toplam 23 aktivite belirlenmiştir. Aktiviteler ikili ağırlıklandırma ile sayısallaştırılmıştır. Kadın ve erkek kullanıcılar tarafından yıllara göre kullanımları için grafikler oluşturulmuştur. Şekil 10.73'e göre 2012 yılında erkekler tarafından egzersiz, hareket ve yüzme, 2013 yılında ise egzersiz, hareket ve spor aktivite isimleri gönderilerde sıklıkla kullanılmaktadır.



Şekil 10.73. 2012 ve 2013 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri

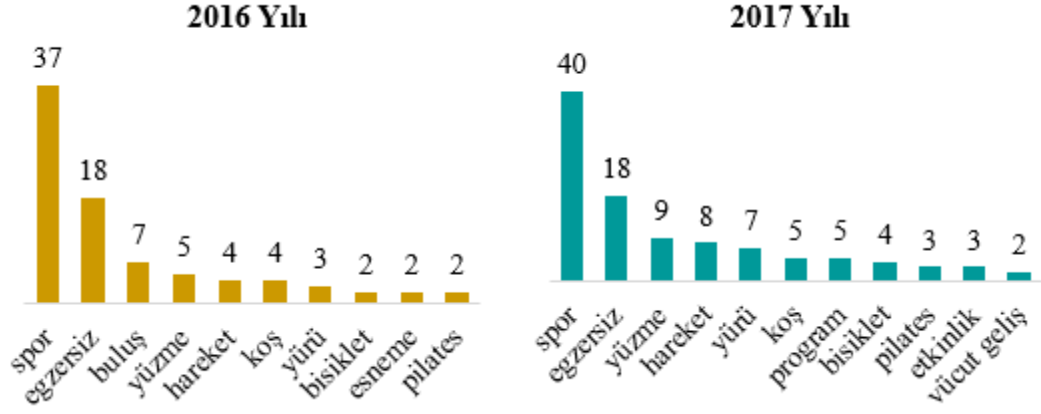
Şekil 10.74'te 2014 yılında spor, egzersiz ve hareket, 2015 yılında ise spor, egzersiz ve yüzme aktivite isimleri kullanıcılar tarafından sıklıkla kullanılmaktadır.



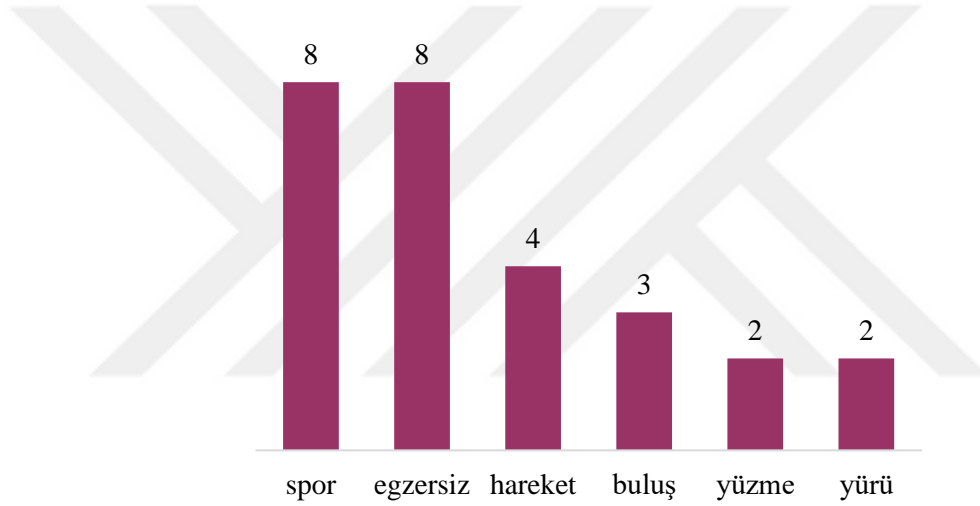
Şekil 10.74. 2014 ve 2015 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri

Şekil 10.75'e göre 2016 yılında spor, egzersiz, buluş, 2017 yılında spor, egzersiz ve yüzme aktivite isimleri erkek kullanıcıları tarafından sıklıkla kullanılmaktadır.

Şekil 10.76'ya göre 2018 yılında erkek kullanıcıları tarafından spor, egzersiz ve hareket kelimeleri sıklıkla geçmektedir.

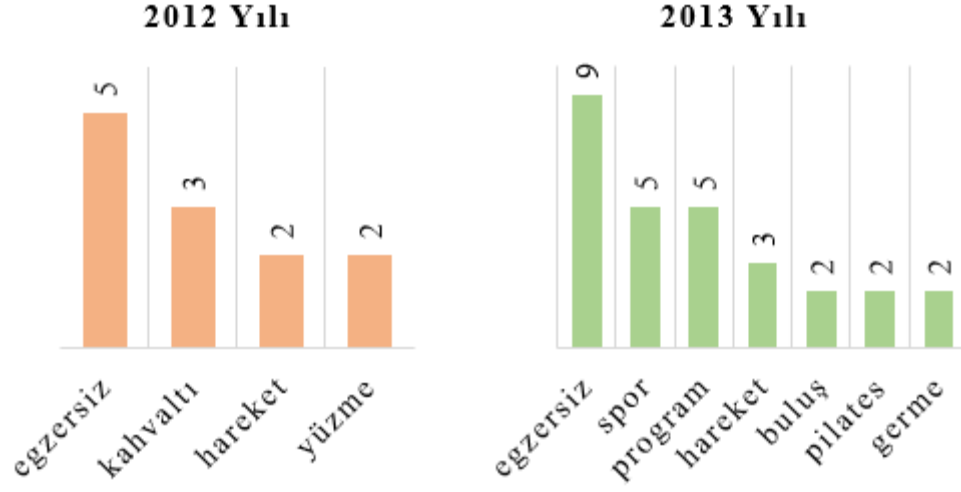


Şekil 10.75. 2016 ve 2017 yıllarında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri

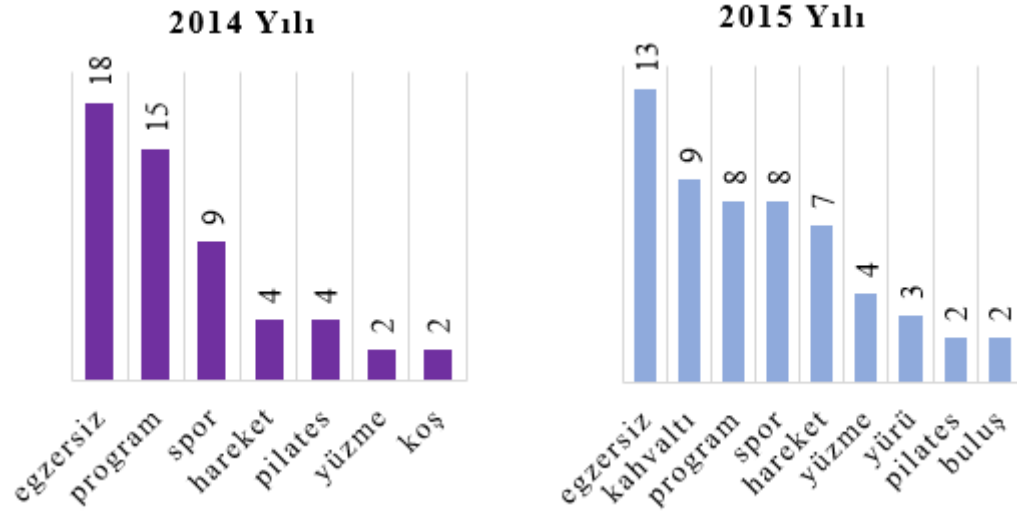


Şekil 10.76. 2018 yılında erkek kullanıcılar tarafından en çok kullanılan aktivite isimleri

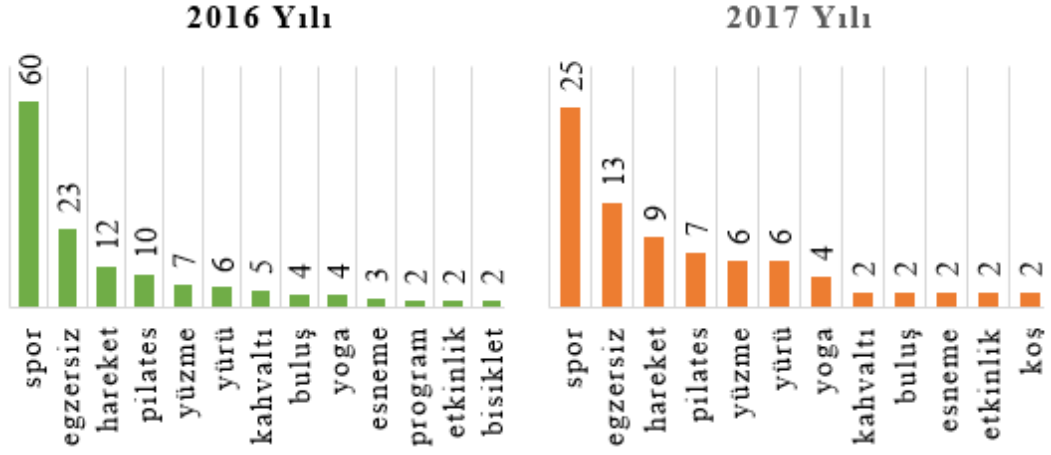
Şekil 10.77-Şekil 10.80 arası kadın kullanıcılar tarafından yıllar göre en çok kullanılan aktivite isimlerini göstermektedir. Kadın kullanıcılar tarafından 2012 yılında en çok egzersiz ve kahvaltı, 2013 yılında egzersiz ve spor, 2014 yılında egzersiz ve program, 2015 yılında egzersiz ve kahvaltı, 2016 ve 2017 yıllarında spor ve egzersiz, 2018 yılında yüzme ve spor gibi aktivitelerin en çok konuşulduğu görülmüştür.



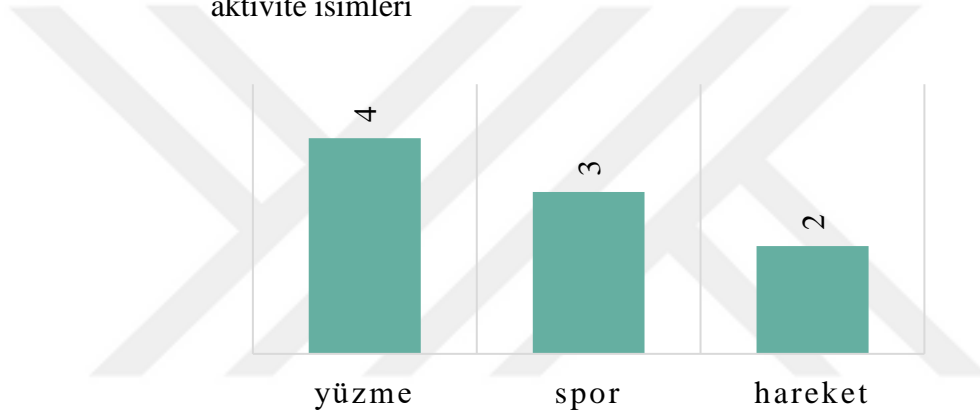
Şekil 10.77. 2012 ve 2013 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri



Şekil 10.78. 2014 ve 2015 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri



Şekil 10.79. 2016 ve 2017 yıllarında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri



Şekil 10.80. 2018 yılında kadın kullanıcılar tarafından en çok kullanılan aktivite isimleri

Spor gönderileri için ilişkili örneklem test sonuçları

Kullanıcılar tarafından kullanılan anahtar kelimelerin yıllara göre farklılık gösterip göstermediğini test etmek amacıyla Friedman testi uygulanmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Kullanıcılar tarafından kullanılan anahtar kelimeler yıllara göre farklılık göstermemektedir.

H_1 : Kullanıcılar tarafından kullanılan anahtar kelimeler yıllara göre farklılık göstermektedir.

Test sonuçları Çizelge 10.63'te verilmiştir. Test sonucunda ($p < 0.05$) H_0 hipotezi reddedilir. Buna göre, kullanılan kelimeler yıllara göre farklılık göstermektedir. Bu farklılık 2012 yılı 2013-2017'den, 2013 yılı 2015-2018'den, 2014 yılı 2015-2018'den, 2015 yılı 2018'den, 2016 yılı 2018'den ve 2017 yılı 2018'den farklılık göstermektedir.

Çizelge 10.63. Spor kategorisini temsil eden kelimeler için Friedman testi sonuçları

Yıllar	Anahtar Kelime Sayısı	Ortalama± Standart Sapma	Medyan (Minimum-Maksimum)	Ortalama Sıra	X^2	p
2012	255	1.52±3.08	1.00 (0.00-33.00) ^d	2.63		
2013	255	3.141±5.58	2.00 (0.00-53.00) ^c	3.95		
2014	255	3.44±5.75	2.00 (0.00-52.00) ^c	4.05		
2015	255	5.48±9.16	3.00 (0.00-103.00) ^b	5.05	448.83	<0.001
2016	255	4.76±9.18	3.00 (0.00-116.00) ^b	4.86		
2017	255	5.22±9.62	3.00 (0.00-111.00) ^b	5.04		
2018	255	1.23±2.43	0.00 (0.00-20.00) ^{ad}	2.40		

a-d; aynı harfe sahip yıllar arasında fark yoktur ($p>0.05$)

Spor kategorisine ait anahtar kelimeler için uyum analizi sonuçları

Spor kategorisinde ismi geçen aktivite isimleri ve yıllara göre cinsiyet grupları arasında ilişkinin olup olmadığı Pearson X^2 testi ile test edilmiştir. Sonuçlar için R programı kullanılmıştır. Bu amaç için test edilecek H_0 ve H_1 hipotezleri;

H_0 : Kullanılan aktivite isimleri ve yıllara göre cinsiyet grupları arasında ilişki yoktur

H_1 : Kullanılan aktivite isimleri ve yıllara göre cinsiyet grupları arasında ilişki vardır

R programında 14×23 boyutunda bir matris oluşturulmuştur. 23, gönderilerde ismi geçen aktiviteleri ifade eder. 14 ise yıllara göre cinsiyet gruplarının toplam sayısını göstermektedir. Test sonucunda $p<0.001$ 'den dolayı %99.9 güvenle H_0 hipotezi reddedilir. Sonuç olarak %99.9 güvenle kullanılan aktivite isimleri ve yıllara göre cinsiyet grupları arasında bir ilişki vardır.

Şekil 10.90'da basit uyum analizi sonucu verilmiştir. Boyut 1 ve boyut 2 olmak üzere iki boyut ile ifade edilmiştir. İlk boyut toplam değişimin %27.5'ini açıklarken, ikinci boyut ise %15.3'ünü açıklamaktadır. Taichi 2013 yılı erkek kullanıcılar için ayırıcı bir özelliktir ve en çok 2013 yılında kullanımı yoğunlaşmıştır. 'ip', 'program', 'kahvaltı' kelimeleri daha çok 2012, 2014 ve 2015 yılı kadın kullanıcıları tarafından yoğun olarak kullanılmıştır. 'futbol', 'maç' ve 'merkez' gibi kelimeler daha çok 2012 ve 2013 yılı erkek kullanıcılar tarafından yoğun kullanıldığı görülmektedir.

Çizelge 10.64. Spor kategorisine ait boyutların yılları açıklama oranları

Yıllar	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
2012E	0.001	0.058	0.408	0.378	0.070
2013E	0.382	0.354	0.103	0.068	0.060
2014E	0.090	0.056	0.000	0.175	0.021
2015E	0.002	0.328	0.189	0.244	0.213
2016E	0.305	0.010	0.141	0.014	0.126
2017E	0.145	0.106	0.208	0.030	0.195
2018E	0.037	0.097	0.053	0.155	0.000
2012K	0.000	0.335	0.216	0.022	0.204
2013K	0.325	0.000	0.001	0.004	0.028
2014K	0.620	0.001	0.319	0.007	0.004
2015K	0.331	0.393	0.001	0.099	0.058
2016K	0.575	0.004	0.052	0.009	0.148
2017K	0.457	0.002	0.005	0.000	0.276
2018K	0.078	0.014	0.064	0.023	0.013

Çizelge 10.65'te, egzersiz, program, salon, spor ve yürü kelimeleri boyut 1, kahvaltı, merkez ve thaichi kelimeleri boyut 2, futbol, germe, pilates ve yüzme boyut 3, bedel, bisiklet, buluş, esneme, etkinlik, havuz ve maç boyut 4, hareket, ip, koş, vücut geliş ve yoga kelimeleri büyük oranda boyut 5 tarafından açıklanmaktadır.

Çizelge 10.65. Boyutların spor kategorisine ait kelimeleri açıklama oranları

Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
bisiklet	0.129	0.086	0.077	0.239	0.093
buluş	0.001	0.002	0.191	0.417	0.065
egzersiz	0.477	0.050	0.000	0.001	0.064
esneme	0.305	0.009	0.015	0.462	0.002
etkinlik	0.015	0.000	0.057	0.603	0.089
futbol	0.001	0.237	0.259	0.002	0.006
germe	0.024	0.032	0.214	0.101	0.163
hareket	0.024	0.179	0.242	0.037	0.307
havuz	0.046	0.014	0.111	0.256	0.000
ip	0.171	0.149	0.001	0.145	0.235

Çizelge 10.65. (devam)

Kelimeler	Boyut 1	Boyut 2	Boyut 3	Boyut 4	Boyut 5
kahvaltı	0.011	0.702	0.197	0.010	0.026
koş	0.040	0.000	0.084	0.022	0.521
maç	0.019	0.071	0.136	0.382	0.196
merkez	0.059	0.408	0.283	0.065	0.000
pilates	0.074	0.009	0.222	0.022	0.208
program	0.888	0.000	0.085	0.000	0.001
salon	0.518	0.031	0.217	0.057	0.017
spor	0.699	0.007	0.199	0.036	0.002
thaichi	0.210	0.351	0.105	0.104	0.095
vücut geliş	0.052	0.063	0.113	0.031	0.557
yoga	0.152	0.187	0.001	0.003	0.364
yürü	0.368	0.040	0.002	0.050	0.024
yüzme	0.065	0.000	0.291	0.012	0.011

Spor kategorisine ait gönderileri temsil eden kelimeler için birliktelik analizi sonuçları

Toplam veri dosyası, 855'dir. 2'den aşağı anahtar kelimesi barındıran metin dosyaları elenmiştir. Analizde toplam 607 metin dosyası kullanılmıştır. Birliktelik kuralları için Apriori algoritması uygulanmıştır. Minimum destek değeri 0.02 ve minimum güven değeri 0.7 olarak belirlenmiştir. Toplam 72 tane kural bulunmuştur. Bu kurallar Çizelge 10.66 ve Çizelge 10.67'de verilmiştir. Çizelgelerde verilen kurallar Lift değerine göre büyükten küçüğe doğru sıralanmıştır.

Çizelge 10.66'da 8.8069 Lift değeri ile en ilginç kural 'kalın → sağlık' kuralıdır. 'kalın' ve 'sağlık' kelimesi tüm gönderilerin %2.14'ünde birlikte kullanılmıştır. Ayrıca 'kalın' kelimesinin kullanıldığı gönderilerin %81.25'inde 'sağlık' kelimesi de kullanılmaktadır.

Lift değeri 5.6044 olan 'mücadele → hastalık' kuralında, 'mücadele' ve 'hastalık' kelimesi tüm gönderilerin %2.14'ünde birlikte kullanılmıştır. Ayrıca 'mücadele' kelimesinin kullanıldığı gönderilerin %81.25'inde 'hastalık' kelimesi de kullanılmaktadır.

Lift değeri 3.3311 olan ‘şiş → egzersiz’ kuralında, ‘şiş’ ve ‘egzersiz’ kelimesi tüm gönderilerin %2.97’sinde birlikte kullanılmıştır. Ayrıca ‘şiş’ kelimesinin kullanıldığı gönderilerin %90’ında ‘egzersiz’ kelimesi de kullanılmıştır.

Lift değeri 2.8129 olan ‘etki → egzersiz’ kuralında, ‘etki’ ve ‘egzersiz’ kelimesi tüm gönderilerin %3.13’ünde birlikte kullanılmıştır. Ayrıca ‘etki’ kelimesinin kullanıldığı gönderilerin %76’sında ‘egzersiz’ kelimesi de kullanılmıştır.

Çizelge 10.66. Spor kategorisine ait gönderiler için ikili kelime kullanımlarını gösteren birliktelik kuralları

Gruplar	Sonuç	Destek	Güven	Lift	
[kalın]	→	sağlık	0.0214	0.8125	8.8069
[mücadele]	→	hastalık	0.0214	0.8125	5.6044
[şiş]	→	egzersiz	0.0297	0.9000	3.3311
[etki]	→	egzersiz	0.0313	0.7600	2.8129
[ücret]	→	egzersiz	0.0313	0.7308	2.7047
[program]	→	egzersiz	0.0527	0.7273	2.6918
[esne]	→	yap	0.0264	0.8889	1.9984
[yaşam]	→	yap	0.0247	0.8333	1.8735
[ayak]	→	yap	0.0313	0.8261	1.8572
[ağır]	→	yap	0.0214	0.8125	1.8266
[yaklaşık]	→	yap	0.0214	0.8125	1.8266
[boyun]	→	yap	0.0280	0.8095	1.8199
[gerçek]	→	yap	0.0280	0.8095	1.8199
[soru]	→	yap	0.0214	0.7647	1.7192
[fazla]	→	yap	0.0313	0.7600	1.7086
[çalış]	→	yap	0.0708	0.7544	1.6960
[otur]	→	yap	0.0247	0.7500	1.6861
[bel]	→	yap	0.0264	0.7273	1.6350
[vücut]	→	yap	0.0264	0.7273	1.6350
[zor]	→	yap	0.0346	0.7241	1.6280
[etki]	→	yap	0.0297	0.7200	1.6187
[uzun]	→	yap	0.0313	0.7037	1.5820
[ilaç]	→	yap	0.0626	0.7037	1.5820
[şiş]	→	yap	0.0231	0.7000	1.5737

Çizelge 10.67’de üç ve üzeri birliktelik kurallarına göre Lift değeri 14.6130 olan ‘sabah, program → pazar’ kuralında ‘sabah, program’ kelime grubu ve ‘pazar’ kelimesi tüm gönderilerin %2.14’ünde birlikte kullanılmıştır. Ayrıca ‘sabah, program’

kelimelerinin kullanıldığı gönderilerin %86.67'sinde 'pazar' kelimesi de kullanılmaktadır.

Lift değeri 10.3466 olan 'pazar, egzersiz→ program' kuralında 'pazar, egzersiz' kelime grubu ve 'program' kelimesi tüm gönderilerin %2.47'sinde birlikte kullanılmıştır. Ayrıca 'pazar, egzersiz' kelimelerinin kullanıldığı gönderilerin %75'inde 'program' kelimesi de kullanılmaktadır.

Lift değeri 10.1651 olan 'egzersiz, ücret→ program' kuralında 'egzersiz, ücret' kelime grubu ve 'program' kelimesi tüm gönderilerin %2.31'inde birlikte kullanılmıştır. Ayrıca 'egzersiz, ücret' kelimelerinin kullanıldığı gönderilerin %73.68'inde 'program' kelimesi de kullanılmaktadır.

Çizelge 10.67. Spor kategorisine ait gönderiler için üç ve üzeri kullanımları gösteren birliktelik kuralı

Gruplar	Sonuç	Destek	Güven	Lift
[sabah, program]	→ pazar	0.0214	0.8667	14.6130
[pazar, egzersiz]	→ program	0.0247	0.7500	10.3466
[egzersiz, ücret]	→ program	0.0231	0.7368	10.1651
[pazar, sabah]	→ program	0.0214	0.7222	9.9634
[pazar, program]	→ sabah	0.0214	0.8125	9.8638
[egzersiz, ücret]	→ salon	0.0231	0.7368	8.9453
[pazar, egzersiz]	→ sabah	0.0231	0.7000	8.4980
[program, hafta]	→ egzersiz	0.0214	1.0000	3.7012
[ücret, program]	→ egzersiz	0.0231	1.0000	3.7012
[pazar, program]	→ egzersiz	0.0247	0.9375	3.4699
[sabah, program]	→ egzersiz	0.0231	0.9333	3.4545
[ücret, salon]	→ egzersiz	0.0231	0.9333	3.4545
[sabah, salon]	→ egzersiz	0.0214	0.9286	3.4368
[program, salon]	→ egzersiz	0.0231	0.8750	3.2386
[yap, program]	→ egzersiz	0.0297	0.8182	3.0283
[pazar, sabah]	→ egzersiz	0.0231	0.7778	2.8787
[etki, yap]	→ egzersiz	0.0231	0.7778	2.8787
[sabah, yap]	→ egzersiz	0.0329	0.7143	2.6437
[çalış, salon]	→ yap	0.0214	1.0000	2.2481
[başla, hastalık, ağrı]	→ yap	0.0231	1.0000	2.2481
[başla, hastalık]	→ yap	0.0362	0.9565	2.1504

Çizelge 10.67. (devam)

Gruplar		Sonuç	Destek	Güven	Lift
[başla, çalış]	→	yap	0.0264	0.9412	2.1159
[çalış, ağrı]	→	yap	0.0247	0.9375	2.1076
[zor, ağrı]	→	yap	0.0214	0.9286	2.0876
[başla, sağlık]	→	yap	0.0264	0.8889	1.9984
[hareket, hastalık]	→	yap	0.0231	0.8750	1.9671
[zaman, hastalık]	→	yap	0.0231	0.8750	1.9671
[zaman, sağlık]	→	yap	0.0214	0.8667	1.9484
[hastalık, ağrı]	→	yap	0.0379	0.8519	1.9151
[egzersiz, ilaç]	→	yap	0.0264	0.8421	1.8932
[egzersiz, çalış]	→	yap	0.0247	0.8333	1.8735
[başla, zaman]	→	yap	0.0231	0.8235	1.8514
[başla, ilaç]	→	yap	0.0214	0.8125	1.8266
[ilaç, ağrı]	→	yap	0.0313	0.7917	1.7798
[hastalık, sağlık]	→	yap	0.0247	0.7895	1.7749
[başla, git]	→	yap	0.0231	0.7778	1.7486
[ilaç, kullan]	→	yap	0.0280	0.7727	1.7372
[başla, ağrı]	→	yap	0.0494	0.7692	1.7293
[fayda, başla]	→	yap	0.0214	0.7647	1.7192
[başla, hafta]	→	yap	0.0214	0.7647	1.7192
[ilaç, hastalık]	→	yap	0.0214	0.7647	1.7192
[egzersiz, ağrı]	→	yap	0.0478	0.7632	1.7157
[egzersiz, hareket]	→	yap	0.0362	0.7586	1.7055
[git, ağrı]	→	yap	0.0346	0.7500	1.6861
[egzersiz, saat]	→	yap	0.0231	0.7368	1.6565
[ağrı, sağlık]	→	yap	0.0231	0.7368	1.6565
[etki, egzersiz]	→	yap	0.0231	0.7368	1.6565
[başla, egzersiz]	→	yap	0.0297	0.7200	1.6187

11. TARTIŞMA VE SONUÇ

Bu çalışmada Ankilozan Spondilit hastalarının Facebook grubundaki paylaşımları 10 başlık altında kategorize edilerek toplanmış ve analiz edilmiştir. Bu çalışmada, yapılan literatür çalışmaları dikkate alındığında ağırlıklandırma yöntemi olarak en çok kullanılan ikili ağırlıklandırma, terim frekansı (TF) ve terim frekansı x ters doküman frekansı (TFxIDF) ağırlıklandırma yöntemleri gönderileri temsil eden terimleri ağırlıklandırmak için kullanılmıştır. Türkçe metinler için yapılan çalışmalarda kelime köklerini bulmak için kullanılan Zemberek kök bulma yöntemi bu çalışma için de kök bulma yöntemi olarak kullanılmıştır. Metinleri yapısal hale dönüştürmek, kelime köklerini belirlemek ve ağırlıklandırma yapmak için Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği bölümü Kemik Doğal Dil İşleme Grubu'nun geliştirdiği Java tabanlı Text2ArffV5 yazılımı kullanılmıştır.

Bölüm 2.7'de yapılan literatür çalışmalarına göre sınıflandırma işlemlerinde en çok kullanılan Naive Bayes, K-en yakın komşu, Karar ağaçları (J48) ve SMO algoritmaları bu çalışmada sınıflandırma işlemlerinde kullanılmıştır. En başarılı sonuç %53.333 oranında SMO yöntemi ile elde edilmiştir. Sınıflandırma algoritmalarının ikili ağırlıklandırma yönteminde daha başarılı sonuçlar verdiği tespit edilmiştir. Başarı performanslarının düşük olmasının en büyük sebebi Şekil 10.5'e göre sınıflar arası kullanılan ortak kelimelerin fazla olmasından kaynaklanmaktadır. İkinci bir neden ise hastalıkla ilişkili olan ve gruplar için ayırıcı özelliğe sahip olacak bazı terimlerin Zemberek kütüphanesinde yer almamasından kaynaklanmaktadır.

Şikayet, doktor, tahlil, tedavi, ilaç, diğer hastalıklar, rapor, kişisel deneyim ve diyet, spor kategorisi altında toplanan verilere ait bilgi çıkarımlarında bulunmuştur. Şikayet kategorisinden yıllara göre kadın ve erkek kullanıcılar tarafından en çok şikayet edilen bölgeler infografikler ile sunulmuştur. Bu bilgi çıkarımları doktorların hastalığın genel belirtisi ile bulunan bu şikayet bölgeleri arasındaki ilişkiyi incelemek ve araştırmak için çeşitli çalışmalar yapmasına olanak sağlayabilir. Doktor kategorisinde hastalar tarafından en çok doktor tavsiyesi sorulan iller haritalar şeklinde verilmiştir. Sağlık Bakanlığı tarafından en çok doktor tavsiyesi sorulan illere ait sonuçlar kullanılarak doktorların en çok arandığı Ankara, İstanbul, İzmir ve Bursa illerinde Romatoloji bölümlerinin ve doktorlarının sayısı artırılabilir. Ayrıca en çok

doktor tavsiyesi sorulan illerdeki romatoloji bölümlerinin yeterli olup olmadığı araştırılabilir ve o illerde ki popüler olan hastaneler tespit edilebilir. Kadın ve erkek kullanıcılar tarafından en çok konuşulan ilaçlar grafikler halinde gösterilmiştir. Bu sonuçlar sayesinde AS hastalığının tedavisinde en çok ve en az kullanılan ilaç isimleri belirlenmiştir. Bu sonuçlar ilaç firmaları, doktorlar ve hastalar tarafından farklı sağlık politikaları belirlenerek çeşitli amaçlar için kullanılabilir. En çok konuşulan alternatif tedavi yöntemlerine göre hastalığın tedavisinde kullanılan farklı tedavi yöntemlerinin neler olduğunu öğrenebilir. Diğer hastalıklar kategorisinde tespit edilen diğer hastalıklarının AS ile ilişkisini incelemek için doktorlar tarafından yeni çalışmaların ele alınmasına neden olabilir. Spor kategorisinde hastalar tarafından en çok yapılan aktivite isimleri verilmiştir. Hangi sporu yapacağı konusunda belirsiz kalan hastalar tarafından karar vermede etkili olabilir. Kısaca bu çıkarımlar doktorlar, ilaç firmaları ve hastalar tarafından çeşitli amaçlar için kullanılabilir.

Yıllara göre sınıfları temsil eden anahtar kelimelerden frekans değeri yüksek olan 50 kelime çizelgeler halinde sunulmuştur. Kategorileri temsil eden kelimelerin tamamının gösterimi için kelime bulutları oluşturulmuştur. Sınıfları temsil eden anahtar kelimelerin yıllara göre farklılık gösterip göstermediği test etmek amacıyla Friedman testi yapılmış ve test sonuçları çizelgeler halinde sunulmuştur. Test sonuçlarına göre kategorileri temsil eden anahtar kelimelerin yıllara göre farklılık gösterdiği tespit edilmiştir. Bu farklılığın en büyük nedeni yıllara göre kategorilere ait veri sayılarının farklı olmasından kaynaklanmaktadır. Örneğin, şikayet kategorisiyle ilgili 2012 yılında toplam 116 gönderi paylaşımı yapılmışken 2017 yılında toplam 256 gönderi paylaşılmıştır. Bu fark kelime ağırlık değerlerinde büyük farklılığa sebep olmaktadır. Anahtar kelimeler ile yıllar veya yıllara göre cinsiyet grupları arasında ilişkiyi ölçmek ve iki boyutlu uzayda göstermek için uyum analizi yapılmıştır. Uyum analizi ile ilgili sonuçlar iki boyutlu grafikler halinde verilmiştir.

Sınıflara ait ikili ve üzeri kelime kullanımlarını tespit etmek amacıyla birliktelik analizi yapılmıştır. Birliktelik analizi kullanılan akademik çalışmada sıklıkla Apriori algoritması uygulanmaktadır. Bu çalışmada da birliktelik analizi için Apriori algoritması kullanılmıştır. Algoritma minimum destek ve güven değerlerine göre çalışır. Literatür çalışmalarında sıklıkla destek değeri %10 ve güven değeri %80 olarak kullanılmaktadır. Algoritma için bu değerler girilerek çalıştırıldığında sınıflara ait çok az kural veya hiçbir kural elde edilememiştir. Bu çalışmada kategorilere ait ikili ve

üzeri kelime birliktelik kurallarını bulmak için farklı destek ve güven değerleri belirlenmiştir. Her sınıf için farklı değerler belirlenerek ideal kurallar elde edilmiştir. Çalışmada destek ve güven değerlerinin küçük olarak belirlenmesin amacı Lift değeri yüksek olan birliktelik kurallarını elde etmektir. Çünkü Lift değeri yüksek olan kurallarda elde edilen kelimelerin birlikte kullanımını daha anlamlıdır.

Ankilozan Spondilit Hastalığı hakkında birçok çalışma yapılmıştır. Ancak yapılan hiçbir çalışmada metin madenciliği yöntemleri kullanılmamıştır. Bu çalışma ile ilk defa Ankilozan Spondilit hastalarının sosyal medya gönderileri metin madenciliği yöntemleri ile analiz edilerek literatüre kazandırılmıştır. Ayrıca metin madenciliğinin Türkiye’de sağlık alanında uygulanmasına örnek bir çalışma olacaktır.

MM işlemlerinde, Türkçe Doğal Dil İşleme ilgili yapılan çalışmalar yetersiz kalmaktadır. Özellikle kelime köklerini bulmada kullanılan yöntemler ve Zemberek Kütüphanesi yetersiz kalmaktadır. Zemberek Kütüphanesindeki kelime kısıtlılığı sınıfları temsil edecek kelimelerin bulunmamasına sebep olmaktadır. Metin madenciliği ve Doğal Dil İşleme ile ilgili yapılacak çalışmalarda, çalışma hangi alan ile ilgili ise o alana ait kelime kütüphaneleri oluşturulmalıdır. Oluşturulan bu kelime kütüphaneleri ile daha geniş ve anlamlı sonuçlar elde edilebilir.

İleri ki çalışmalarda Sağlık Bilimleri ve Biyomedikal alanları için kelime kütüphaneleri geliştirilecektir. Ayrıca birliktelik analizi sonuçlarına göre anlamlı ikili kelime kullanımları için sağlık alanında kelime kütüphaneleri oluşturulacaktır.

KAYNAKLAR

- Açıkalin, B. ve Bayazıt, N. G. (2016). Türkçe Metinlerin Sınıflandırılmasında Önişlemenin Önemi. 24th Signal Processing and Communication Application Conference (SIU), 16 – 19 Mayıs, 2053-2056, Zonguldak, Türkiye.
- Aksu, K. 2015. Ankilozan spondilit tedavisinde sertolizumab pegol. *RAED Dergisi*, 7:1, 14-17.
- Akyel, N. ve Seçkin, K. 2012. K-en yakın komşuluk algoritmasının hile denetiminde kullanımı. *Muhasebe ve Vergi Uygulamaları Dergisi*, 5:1, 21-39.
- Alaeddinoğlu, M., Aydın, T. ve Dal, D. 2014. Birliktelik kuralları ile mekansal-zamansal veri madenciliği. *Fen Bilimleri Enstitüsü Dergisi*, 5:2, 191-212.
- Amanet, H. 2017. Türkçe sosyal medya metinlerinde duygu analizi. Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Ve Bilgisayar Bilimleri Anabilim Dalı, 95, Trabzon.
- Amasyalı, M. F., Balcı, S., Mete, E. ve Varlı, E. N. 2012. Türkçe metinlerin sınıflandırılmasında metin temsil yöntemlerinin performans karşılaştırılması. *EMO Bilimsel Dergi*, 2:4, 95-104.
- Amrit, C., Paauw, T., Aly, R. and Lavric, M. 2017. Identifying child abuse through text mining and machine learning. *Expert Systems with Applications*, 88, 402-418.
- Anonim (2019a). Facebook. <https://tr.wikipedia.org/wiki/Facebook> (Erişim Tarihi: 19.05.2019).
- Anonim (2019b). Twitter. <https://tr.wikipedia.org/wiki/Twitter> (Erişim Tarihi: 19.05.2019).
- Anonim (2019c). Youtube. <https://tr.wikipedia.org/wiki/YouTube> (Erişim Tarihi: 19.05.2019).
- Arslan, H., Kaynar, O. ve Yüksek, A. G. 2017. Kurumsal kolektif süreçler için e-posta iletilerinden görev keşfi ve gerçek zamanlı görev yönetim sisteminin geliştirilmesi. *Bilişim Teknolojileri Dergisi*, 10:4, 381-388.
- Aşlıyan, R. ve Günel, K. (2010). Metin İçerikli Türkçe Dokümanların Sınıflandırılması. 12. Akademik Bilişim Konferansı, 10-12 Şubat, 529-535, Muğla, Türkiye.
- Atasoy, Y. 2015. Veri madenciliği yöntemleri ile Ankilozan Apondilit hastalığında radyografik progresyona etkili faktörlerin analizi. Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü Enformatik Anabilim Dalı, 111, İstanbul.
- Ateş, Y. ve Karabatak, M. 2017. Nicel birliktelik kuralları için çoklu minimum destek değeri. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 29:2, 57-65.
- Ayata, D., Saraçlar, M. ve Özgür, A. (2017). Makine Öğrenmesi ve Kelime Vektör Temsili ile Türkçe Tweet Sentiment Analizi. 25th Signal Processing and Communications Applications Conference (SIU), 15-18 Mayıs, Antalya, Türkiye.

- Badal, V. D., Kundrotas, P. J. and Vakser, I. A. 2015. Text mining for protein docking. *PLOS Computational Biology*, 11:12, 1-21.
- Basaldella, M., Furrer, L., Tasso, C. and Rinaldi, F. 2017. Entity recognition in the biomedical domain using a hybrid approach. *Journal of Biomedical Semantics*, 8:51, 1-14.
- Bastem, K. ve Şeker, Ş. E. 2017. Veri madenciliği yöntemleri ile Twitter üzerinden MBTI kişilik tipi analizi. *Yönetim Bilişim Sistemleri Ansiklopedisi*, 4:2, 3-10.
- Başkaya, F. ve Aydın, İ. (2017). Haber Metinlerinin Farklı Metin Madenciliği Yöntemleriyle Sınıflandırılması. International Artificial Intelligence and Data Processing Symposium (IDAP 2017), 16-17 Eylül, IEEE, 1-5, Malatya, Türkiye.
- Bayer, H. 2011. Veri maddenciliğinde bir metin madenciliği uygulaması. Yüksek Lisans Tezi, Beykent Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 90, İstanbul.
- Bayrak, H. (2019). 2019 Türkiye İnternet Kullanım ve Sosyal Medya İstatistikleri. <https://dijilopedi.com/2019-turkiye-internet-kullanim-ve-sosyal-medya-istatistikleri/> (Erişim Tarihi: 07.02.2019).
- Bong, C. H., Kulathuramaiyer, N. and Ting Kiong, W. (2005). Categorical Term Descriptor: A Proposed Term Weighting Scheme for Feature Selection. 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005), 19-22 September, 313-316, Compiègne, France.
- Bülbül, H. 2015. Ankilozan Spondilit hastalarında miyokinlerin değerlendirilmesi. Tıpta Uzmanlık Tezi, İzmir Katip Çelebi Üniversitesi Atatürk Eğitim ve Araştırma Hastanesi İç Hastalıkları Kliniği, 47, İzmir.
- Büyüköztürk, Ş. 2005. Anket geliştirme. *Türk Eğitim Bilimleri Dergisi*, 3:2, 133-151.
- Calin, A., Porta, J., Fries, J. F. and Schurman, D. J. 1977. Clinical history as a screening test for Ankylosing Spondylitis. *The Journal of the American Medical Association*, 237:24, 2613-2614.
- Can, E. ve Amasyalı, M. F. (2016). Text2arff: Metin Temsil Kütüphanesi. 24th Signal Processing and Communication Application Conference (SIU), 16-19 Mayıs 2016, IEEE, 197-200, Zonguldak, Türkiye.
- Cangür, Ş., Sığırlı, D., Ediz, B., Ercan, İ. ve Kan, İ. 2005. Türkiye'deki özürlü grupların yapısının çoklu uyum analizi ile incelenmesi. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 31:3, 153-157.
- Canhası, E. 2007. Edebiyat eserlerinin web verilerine dayanarak sınıflandırılması. Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 93, Konya.
- Chen, Y.-L., Chang, C.-L. and Yeh, C.-S. 2017. Emotion classification of YouTube videos. *Decision Support Systems*, 101, 40-50.
- Christensen, K., Liland, K. H., Kvaal, K., Risvik, E., Biancolillo, A., Scholderer, J., Nørskov, S. and Næs, T. 2017. Mining online community data: The nature of ideas in online communities. *Food Quality and Preference*, 62, 246-256.

- Cingiz, M. Ö. ve Diri, B. (2012). Mikroblog Kullanıcılarının Kategorizasyonu. 20th Signal Processing and Communications Applications Conference (SIU), 18-20 Nisan, IEEE, 1-4, Muğla, Türkiye.
- Çakır, Ö. 2015. Ankilozan Spondilit'li hastalarda Tai Chi Chuan formunda planlanan egzersiz programının kassal uygunluk, denge ve fonksiyonel durum üzerine etkisi. Doktora Tezi, Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Fizik Tedavi ve Rehabilitasyon Anabilim Dalı, 71, İzmir.
- Çakmak, Z., Baş, M. ve Yıldırım, E. 2012. Gri ilişkisel analiz ve uyum analizi ile bir işletmede karşılaşılan üretim hatalarının incelenmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 17:1, 123-142.
- Çalış, K., Gazdağı, O. ve Yıldız, O. 2012. Reklam içerikli e-postaların metin madenciliği yöntemleri ile otomatik tespiti. *Bilişim Teknolojileri Dergisi*, 6:1, 1-7.
- Çeliksü, Z. 2017. Yabancı dizilerin altyazı ve Twitter yorumlarının metin madenciliği ile incelenmesi. Yüksek Lisans Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 46, İstanbul.
- Çelikyay, E. K. 2010. Metin madenciliği yöntemiyle Türkçe'de en sık kullanılan ve birbirini takip eden harflerin analizi ve birliktelik kuralları. Yüksek Lisans Tezi, Beykent Üniversitesi Fen Bilimleri Enstitüsü Matematik Bilgisayar Anabilim Dalı, 85, İstanbul.
- Çoban, E. 2016. Non Radyografik Aksiyel Spondiloartrit ile Ankilozan Spondilit hastalarının klinik , radyolojik özelliklerinin ve hastalık aktiviteleri arasındaki ilişkinin değerlendirilmesi. Tıpta Uzmanlık Tezi, İzmir Katip Çelebi Üniversitesi Atatürk Eğitim ve Araştırma Hastanesi İç Hastalıkları Kliniği, 45, İzmir.
- Çoban, Ö., Özyer, B. ve Özyer, G. T. (2015). Türkçe Twitter Mesajlarının Duygu Analizi. 23th Signal Processing and Communications Applications Conference (SIU), 16-19 Mayıs, IEEE, 2388-2391, Malatya, Türkiye.
- Doğan, B., Sarıkaya, U. ve Aktaş, A. 2017. R ile sosyal ağ madenciliği. *Marmara Fen Bilimleri Dergisi*, 29:3, 94-101.
- Dolgun, M. Ö., Özdemir, T. G. ve Oğuz, D. 2014. Veri madenciliğinde yapısal olmayan verinin analizi: Metin ve web madenciliği. *İstatistikçiler Dergisi*, 2, 48-58.
- Dondurucu, Z. B. ve Ulucay, A. P. 2015. Yeni medya ortamlarında nefret söylemi: Eşcinsellere yönelik nefret söylemi içeren videoların Youtube üzerinden incelenmesi. *International Journal of Social Sciences and Education Research*, 1:3, 875-902.
- Döven, S. 2013. Metin madenciliği ile dokümanlar arasındaki benzerliklerin bulunması. Yüksek Lisans Tezi, Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü Bilgi Teknolojileri, 92, İstanbul.
- Elbey, B. 2015. Ankilozan Spondilit'li hastalarda güncel tedavi yaklaşımları. *Dicle Tıp Dergisi*, 42:1, 123-127.
- Erben, Ş. E. ve Salı, J. B. 2015. Blogalanda blog yazarlarının etkileşim biçimleri ve görsel öğelerin kullanımı. *Global Media Journal: Turkish Edition*, 5:10, 176-207.

- Erdal, H. 2018. Yapay zeka teknikleri ve uzman sistemlerin karasal akıllı ulaşım sistemlerinin denetiminde kullanımı. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi*, 1:1, 32-39.
- Ergün, K., Kubat, C., Çağıl, G. ve Cesur, R. 2013. İnternet ortamındaki tüketici yorumlarından özet bilgi çıkarımı. *SAÜ Fen Bilimleri Dergisi*, 1, 33-40.
- Ergün, M. (2016). Eğitim Araştırmalarında Data Mining ve Text Mining Tekniklerinden Yararlanma. III. Uluslararası Dinamik, Araştırmacı ve Aktif Öğrenme Konferansı (İDEAL2016), 1-3 Eylül, Samsun, Türkiye.
- Erpolat, S. 2012. Otomobil yetkili servislerinde birliktelik kurallarının belirlenmesinde Apriori ve Fp-Growth algoritmalarının karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12:2, 137-146.
- Fluck, J. and Hofmann-Apitius, M. 2014. Review: Text mining for systems biology. *Drug Discovery Today*, 19, 140-144.
- Goh, Y. M. and Ubeynarayana, C. 2017. Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis & Prevention*, 108, 122-130.
- Göker, H. ve Tekedere, H. 2017. Fatih projesine yönelik görüşlerin metin madenciliği yöntemleri ile otomatik değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10:3, 291-299.
- Gönültaş, H. 2010. E-posta listelerinde metin kümeleme ve sosyal ağ analizi uyumu. Yüksek Lisans Tezi, Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği, 53, Kocaeli.
- Gün, Z. and Erdem, Z. Ç. 2014. Assessing the factors affecting the mathematics success via correspondence analysis method. *Adıyaman University Journal of Educational Sciences*, 4:2, 98-118.
- Gündüz, H. 2013. Borsa İstanbul (BIST) 100 endeksi yönünün ekonomi haberleri ile tahmin edilmesi. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 61, İstanbul.
- Gürbüz, F. ve Özyirmidokuz, E. K. 2018. Toplantı tutanaklarının analizi ile bir karar destek sistemi. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22:2, 257-268.
- Güven, A., Bozkurt, Ö. Ö. ve Kalıpsız, O. (2006). Gizli Anlambilimsel Dizinleme Yönteminin N-Gram Kelimelerle Geliştirilerek, İleri Düzey Doküman Kümelemesinde Kullanımı. Elektrik - Elektronik - Bilgisayar Mühendisliği Sempozyumu (ELECO 2006), 6-10 Aralık, Bursa, Türkiye.
- Güven, A., Bozkurt, Ö. Ö. ve Kalıpsız, O. (2007). Veri Madenciliğinin Geleceği. 9. Akademik Bilişim Konferansı, 31 Ocak - 2 Şubat, Kütahya, Türkiye.
- Hablemitoğlu, Ş. ve Yıldırım, F. 2012. Gençlerin gözünden sanal bir sosyal kapital olarak "Facebook". *Ankara Sağlık Bilimleri Dergisi*, 1:1, 1-20.
- Haltaş, A., Alkan, A. ve Karabulut, M. 2015. Metin sınıflandırmada sezgisel arama algoritmalarının performans analizi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 30:3, 417-427.
- Hark, C., Seyyarer, A., Uçkan, T. ve Karıcı, A. (2017). Doğal Dil İşleme Yaklaşımları ile Yapısal Olmayan Dökümanların Benzerliği. International Artificial

Intelligence and Data Processing Symposium (IDAP), 16-17 Eylül, IEEE, 1-6, Malatya, Türkiye.

- He, W., Tian, X., Tao, R., Zhang, W., Yan, G. and Akula, V. 2017. Application of social media analytics: a case of analyzing online hotel reviews. *Online Information Review*, 41:7, 921.
- İlhan, S., Duru, N., Karagöz, Ş. ve Sağır, M. (2008). Metin Madenciliği ile Soru Cevaplama Sistemi. Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, 26-30 Kasım, Bursa, Türkiye.
- Irmak, S., Köksal, C. D. ve Asilkan, Ö. 2012. Hastanelerin gelecekteki hasta yoğunluklarının veri madenciliği yöntemleri ile tahmin edilmesi. *Uluslararası Alanya İşletme Fakültesi Dergisi*, 4:1, 101-114.
- Jimeno Yepes, A., MacKinlay, A., Han, B. and Chen, Q. (2015). Identifying Diseases, Drugs, and Symptoms in Twitter MEDINFO 2015: EHealth-enabled Health, 19-23 August, Book of Abstracts, 643 - 647, Sao Paulo, Brazil.
- João, G., Paulo, R. and Duarte, T. 2016. A text mining-based review of cause-related marketing literature. *Journal of Business Ethics*:1, 111.
- Karaca, M. F. 2012. Metin madenciliği yöntemi ile haber sitelerindeki köşe yazılarının sınıflandırılması. Yüksek Lisans Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 99, Karabük.
- Karaca, M. F. ve Görgünoğlu, S. (2012). ColumnREADY: İnternet Gazeteleri Köşe Yazılarını Hazırlama Uygulama Yazılımı. 14. Akademik Bilişim Konferansı, 1-3 Şubat, 415-421, Uşak, Türkiye.
- Karaca, M. F., Günel, M. ve Taştan, A. A. (2015). Metin Sınıflandırmada Benzerlik Hesaplama Tekniklerinin Değerlendirilmesi. 17. Akademik Bilişim Konferansı, 4-6 Şubat, Eskişehir, Türkiye.
- Karataş, A. 2015. Ankilozan Spondilit'li hastalarda hastalık aktivitesinin uyku bozukluklarına etkisi. Tıpta Uzmanlık Tezi, Ankara Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı, 68, Ankara.
- Karystianis, G., Nevado, A. J., Kim, C. H., Dehghan, A., Keane, J. A. and Nenadic, G. 2018. Automatic mining of symptom severity from psychiatric evaluation notes. *International Journal of Methods in Psychiatric Research*, 27:1, e1602.
- Kasaba, E. ve Yıldıztepe, E. (2016). Destek Vektör Makinesi Yöntemi ile Bir Duygu Çözümlemesi. 18. Akademik Bilişim Konferansı, 3-5 Şubat, Aydın, Türkiye.
- Kaşıkcı, T. 2014. Metin madenciliği ile e-ticaret sitelerinin belirlenmesi. Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü Yönetim Bilişim Sistemleri, 140, Ankara.
- Kaşıkcı, T. ve Gökçen, H. 2014. Metin madenciliği ile e-ticaret sitelerinin belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7:1, 25-32.
- Kaya, M. İ. Y. 2010. Finansal haber makaleleri kullanılarak hisse senetleri fiyat değişimlerinin tahmin edilmesi. Yüksek Lisans Tezi, Yıldız Tenik Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı, 75, İstanbul.
- Kaya, Z. I. 2016. Mefv mutasyonlarının Ankilozan Spondilit hastalarındaki sıklığı, klinik ve radyolojik şiddetle ilişkisi. Uzmanlık Tezi, Eskişehir Osmangazi Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı, 59, Eskişehir.

- Kayıkçı, M. Y. ve Bozkurt, A. K. 2018. Dijital çağda Z ve Alpha kuşağı, robotlar ve turizmde yapay zeka uygulamaları. *Sosyal Bilimler Metinleri*, 2018:1, 54-64.
- Keleş, M. K. ve Özel, S. A. (2017). Uzaklık Ölçütleri ile Türkçe Metin Belgeleri Arasındaki Benzerliğin Belirlenmesi. International Conference on Computer Science and Engineering (UBMK), 5-8 October, IEEE, 316-321, Antalya, Turkey.
- Kılıç, E., Tavus, M. R. ve Karhan, Z. (2015). Çevrimiçi Haber Sitelerinden Çekilen Son Dakika Haberlerin Sınıflandırılması. 23rd Signal Processing & Communications Applications Conference (SIU), 16-19 Mayıs, IEEE, 363-366, Malatya, Türkiye.
- Kılınç, D., Borandağ, E., Yücelar, F., Tunalı, V., Şimşek, M. ve Özçift, A. 2016. KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, 28:3, 89-94.
- Kılınç, D., Bozyiğit, F., Yücelar, F., Akın, Ö. ve Borandağ, E. (2015). Metin Madenciliği Kullanarak Yazılım Kullanımına Dair Bulguların Elde Edilmesi. Ulusal Yazılım Mühendisliği Sempozyumu (UYMS), 9-11 Eylül, 257-265, İzmir, Türkiye.
- Kırmacı, B. ve Oğul, H. (2015). Şarkı Sözünden Yazar Tanıma. 23rd Signal Processing & Communications Applications Conference (SIU), 16-19 Mayıs, IEEE, 2489-2492, Malatya, Türkiye.
- Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S. and Gonzalez, G. H. 2016. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of Biomedical Informatics*, 62, 148-158.
- Kul, M. 2016. Ankilozan Spondilitli hastalarda sitokinler ve optik koherens tomografi bulgularının değerlendirilmesi. Tıpta Uzmanlık Tezi, Harran Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, 71, Şanlıurfa.
- Kuzucu, K. 2015. Müşteri memnuniyeti belirlemek için metin madenciliği tabanlı bir yazılım aracı. Yüksek Lisans Tezi, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 48, İstanbul.
- Lee, Y. J. and Park, J. Y. 2018. Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence. *Quality and Quantity*, 52:2, 653-667.
- Man, L., Sam-Yuan, S., Hwee-Boon, L. and Chew-Lim, T. (2005). A Comparative Study on Term Weighting Schemes for Text Categorization. International Joint Conference on Neural Networks (IJCNN), 31 July-4 August, 546-551, Quebec, Canada.
- Meral, M. ve Diri, B. (2014). Twitter Üzerinde Duygu Analizi. 22nd Signal Processing and Communications Applications Conference (SIU), 23-25 Nisan , IEEE, 690-693. Trabzon, Türkiye.
- Moghaddam, S. A. V. 2014. Etkin sınıflandırma için genetik algoritma tabanlı öznitelik alt küme seçimi. Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 111, Ankara.

- Mukherjee, S. and Bala, P. K. 2017. Gender classification of microblog text based on authorial style. *Information Systems and e-Business Management*, 15:1, 117-138.
- Murugesan, G., Abdulkadhar, S., Bhasuran, B. and Natarajan, J. 2017. BCC-NER: bidirectional, contextual clues named entity tagger for gene/protein mention recognition. *EURASIP Journal on Bioinformatics and Systems Biology*, 2017:1, 7.
- Mutlu, B. ve Bazarıcı, S. 2017. Marka işbirlikleri için yeni bir alan: YouTube içerik üreticileri ve kanal toplulukları üzerine netnografik bir araştırma. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*, 27, 28-45.
- Nanğır, M. 2013. Türk dili için çoklu sınıflandırıcı yöntemler ile duygu sınıflandırma. Yüksek Lisans Tezi, İstanbul Kültür Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 41, İstanbul.
- Nicandro, S., Benjamin, C. M. F. and Rashid, H. K. 2015. Mining known attack patterns from security-related events. *PeerJ Computer Science*, 1, 1-21.
- Nokhbeh Zaeem, R., Manoharan, M., Yang, Y. and Barber, K. S. 2017. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security*, 65, 50-63.
- Oğuz, B., Bİlge, U. ve Saka, O. (2007). Tıpta Metin Madenciliği. 4. Tıp Bilişimi Kongresi, 15-18 Kasım, 108-112, Antalya, Türkiye.
- Okuhara, T., Ishikawa, H., Okada, M., Kato, M. and Kiuchi, T. 2018. Contents of Japanese pro- and anti-HPV vaccination websites: A text mining analysis. *Patient Education and Counseling*, 101:3, 406-413.
- Onan, A. ve Korukoğlu, S. (2015). Görüş Madenciliğinde Sınıflandırıcı Topluluklar. 23rd Signal Processing & Communications Applications Conference (SIU), 16-19 Mayıs, IEEE, 212-215, Malatya, Türkiye.
- Özalp, N., Aykanat, S., Erdem, Z. ve Ayan, U. (2013). Yazar Tespiti için İyileştirilmiş Naïve Bayesian Algoritması. 21st Signal Processing and Communications Applications Conference (SIU), 24-26 Nisan, IEEE, 1057-1061, Haspolat, Türkiye.
- Özçakır, F. C. ve Çamurcu, A. Y. 2007. Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*:12, 21-37.
- Özen, S. 2017. Ankilozan Spondilit hastalarında subklinik kardiyak tutulum. Tıpta Uzmanlık Tezi, Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, 55, Ankara.
- Öztürk, G. ve Tanrısevdi, A. 2017. Uluslararası kruvaziyer ziyaretçilerine ait özelliklerin birliktelik kuralı modeli ile analizi. *Uluslararası İktisadi ve İdari Bilimler Dergisi*, 1, 131-148.
- Papanikolaou, N., Pavlopoulos, G. A., Theodosiou, T., Vizirianakis, L. S. and Lliopoulos, L. 2016. DrugQuest - a text mining workflow for drug association discovery. *BMC Bioinformatics*, 17:5, 333-415.
- Park, A., Conway, M. and Chen, A. T. 2018. Examining thematic similarity, difference, and membership in three online mental health communities from

- reddit: A text mining and visualization approach. *Computers in Human Behavior*, 78, 98-112.
- Pilavcılar, İ. F. 2007. Metin madenciliği ile metin sınıflandırma. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Mühendisliği Anabilim Dalı, 64, İstanbul.
- Salton, G., Wong, A., and Yang, C. S. 1975. A vector space model for automatic indexing. *Communications of the ACM*, 18:11, 613-620.
- Salur, M. U., Tokat, S. ve Aydılek, İ. B. (2017). Naive-Bayes Makine Öğrenmesi Algoritmasıyla Mahout Üzerinde Metin Sınıflandırma. Artificial Intelligence and Data Processing Symposium (IDAP), 16-17 Eylül, IEEE, 1-5, Malatya, Türkiye.
- Sancar, Y. 2016. Metin madenciliği kullanılarak talep tanıma ve yönlendirme sistemi. Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 47, Erzurum.
- Sarker, A. and Gonzalez, G. 2015. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of Biomedical Informatics*, 53, 196-207.
- Savaşan, S. 2011. Türkçe içeriklerden otomatik etiket bulutu oluşturma. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 69, İstanbul.
- Seçkin, K. 2011. Metin madenciliğinde kullanılan yöntemlerin karşılaştırılması: Siyasi parti liderlerinin grup genel toplantı konuşmaları ile bir uygulama. Yüksek Lisans Tezi, Sakarya Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 122, Sakarya.
- Silahtaroglu, G. ve Demircan, F. (2013). Çeviri yazılımlarında sözcüklerin bağlam içindeki anlamını algılamaya yönelik bir öneri. 14. Akademik Bilişim Konferansı, 5-7 Şubat, 23-25, Mersin, Türkiye.
- Sönmez, N. 2017. Çevrimiçi yorumların metin madenciliği ile analizi: İstanbul'daki alışveriş merkezleri üzerine bir çalışma. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 109, İstanbul.
- Şeker, Ş. E. (2010). Bilgi Getirimi ve Çıkarımı (Information Retrieval and Extraction). <http://bilgisayarkavramlari.sadievrenseker.com/2010/10/01/bilgi-getirimi-ve-cikarimi-information-retrieval-and-extraction/> (Erişim Tarihi: 17.06.2019).
- Şeker, Ş. E. 2015. Metin madenciliği (Text mining). *Yönetim Bilişim Sistemleri Ansiklopedisi*, 2:3, 30-32.
- Şen, Ü. S. 2005. Sanat eğitiminde bilimsel araştırma yöntemlerinin kullanılması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5:1, 343-360.
- Taha, S. M. 2011. Metin madenciliği ile doküman demetleme. Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü Bilgisayar Bilimleri Anabilim Dalı, 84, Ankara.
- Talebi, M. 2013. Facebook'da yorum madenciliği kullanarak kişilerin cinsiyet, yaş ve eğitim düzeylerinin tanımlanması. Yüksek Lisans Tezi, Karadeniz Teknik

Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 80, Trabzon.

- Tsai, M. F. and Wang, C. J. 2017. On the risk prediction and analysis of soft information in finance reports. *European Journal of Operational Research*, 257:1, 243-250.
- Tunalı, V. ve Bilgin, T. T. (2012). Türkçe Metinlerin Kümelenmesinde Farklı Kök Bulma Yöntemlerinin Etkisinin Araştırılması. Elektrik, Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO), 29 Kasım-1 Aralık, 598-602, Bursa, Türkiye.
- Türkmenoğlu, C. 2015. Türkçe metinlerde duygu analizi. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 47, İstanbul.
- Varol, M. 2011. Metin madenciliği yöntemlerini kullanarak Türkçe dökümanlarda tür ve yazar tanıma Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Elektronik Bilgisayar Eğitimi Anabilim Dalı 103, Isparta.
- Vural, Z. B. A. ve Bat, M. 2009. Siyasal seçim kampanyalarında yeni iletişim teknolojileri ve blog kullanımı: 2008 Amerika başkanlık seçimlerine yönelik karşılaştırmalı bir analiz. *Journal of Yaşar University*, 4:16, 2745-2778.
- Yeniçaktı, N. T. 2016. Hakla ilişkiler aracı olarak Instagram: Sosyal medya kullanan 50 şirket üzerine bir araştırma. *Selçuk Üniversitesi İletişim Fakültesi Akademik Dergisi*, 9:2, 92-115.
- Yigit, İ. O., Ateş, A. F., Guvercin, M., Ferhatosmanoğlu, H. ve Gedik, B. (2017). Çağrı Merkezi Metin Madenciliği Yaklaşımı. 25th Signal Processing and Communications Applications Conference (SIU), 15-18 Mayıs, Antalya, Türkiye.
- Yıldız, A. ve Demirci, M. (2017). Kurumsal E-posta Sınıflandırma Sistemi. (UBMK'17) 2nd International Conference on Computer Science and Engineering 5-8 Ekim, IEEE, 559-564, Antalya, Türkiye.
- Yıldız, O. 2017. Metin madenciliğinde anahtar kelime seçimi bir üniversite örneği. *Yönetim Bilişim Sistemleri Dergisi*, 2:3, 29-50.
- Yurt, E. A. 2015. Türkçe metinlerde duygu analizi. Yüksek Lisans Tezi, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 38, İstanbul.
- Zhou, J. and Fu, B.-q. 2018. The research on gene-disease association based on text-mining of PubMed. *BMC Bioinformatics*, 19:37, 1-8.

ÖZGEÇMİŞ

Adı ve Soyadı : Ertuğrul GÜMÜŞSU
Doğum Yeri : Ayvacık-SAMSUN
Doğum Tarihi : 03-01-1993
Yabancı Dili : İngilizce

Eğitim Durumu

Lise : Kavak Rıdvan Çelikel Fen Lisesi (2011)

Lisans : Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü (2016)

Çalıştığı Kurumlar

Ziraat Teknoloji AŞ – Stajyer (2015)

Yayınlar

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler:

Gümüşsu E. ve Murat N. (2019). Şehirlerarası yolcu taşımacılığı yapan firmalar hakkındaki şikayetlerin metin madenciliği yöntemleri ile incelenmesi. 2. Uluslararası İstatistik, Matematik ve Analitik Yöntemler Kongresi, İstanbul, 2019 (Özet Bildiri/Sözlü Sunum)

Gümüşsu E. Ve Murat N. (2019). Metin madenciliği yöntemleri ile otel müşteri yorumlarının sınıflandırılması. 2. Uluslararası İstatistik, Matematik ve Analitik Yöntemler Kongresi, İstanbul, 2019 (Özet Bildiri/Sözlü Sunum)