

FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI



SAĞLIKLA İLGİLİ SOSYAL MEDYA FORUM VERİLERİNDE İÇERİĞİNİN
BİLGİLENDİRİCİ ÖZELLİKLERİ:
İLAÇ YAN ETKİLERİNİN ARAŞTIRMASI ÜZERİNE BİR ÇALIŞMA

YÜKSEK LİSANS TEZİ

ALKAN KAYA

Haziran 2018

SAĞLIKLA İLGİLİ SOSYAL MEDYA FORUM VERİLERİNDE İÇERİĞİNİN
BİLGİLENDİRİCİ ÖZELLİKLERİ:
İLAÇ YAN ETKİLERİNİN ARAŞTIRMASI ÜZERİNE BİR ÇALIŞMA

YÜKSEK LİSANS TEZİ

ALKAN KAYA

tarafından

OKAN ÜNİVERSİTESİ

Bilgisayar Mühendisliği Bölümüne

Yüksek Lisans

derecesi şartını sağlamak için sunulmuştur.

Onaylayan:

Chair

Name

Danışman

Doç. Dr. Pınar Yıldırım

Member's

Name

Üye

Doç. Dr. Rüyam Acar

Member's

Name

Üye

Doç. Dr. Erchan Aptoula

INFORMATIONAL CHARACTERISTICS OF HEALTH-RELATED SOCIAL MEDIA CONTENT: A STUDY ON THE INVESTIGATION OF DRUG SIDE EFFECTS

ABSTRACT

The ubiquity of Internet in Online Social Networks is creating new sources for healthcare information, characteristics on the content of pharmaceutical drug discussions we compared the side effects.

In this study, a research study has been performed by using the patient comments on the website www.webmd.com by utilizing data and text mining methods. The related comments have been analyzed by some methods including text parsing and some data such as patients' gender and age range, user type, information and duration of drug use have been systematized and the side effects of the drugs have been searched.

In this study, the data of the patient comments have been separated by text parsing methods and the frequencies of them have been calculated and the frequently seen side effects have been explored.

After analyzing the fields such as gender, age, user type, duration of use, the association rules (Apriori), classification rules (Knn), clustering rule (Kmeans) algorithms have been used and hidden relationships among these attributes have been investigated.

The results of this study will provide significant contributions to medical experts, researchers and pharmaceutical companies.

Keywords :Medical and health information; Content analysis; Social media; Data mining; Side effects; Apriori algorithm; Knn algorithm; Kmeans algorithm; Internet and Medical Informatics; Free Web Technologies; Mobile Software Development;

TEŐEKKÜR

Bu tez alıřmasın da her tŸrlŸ bilgi ve tecrŸbelerinden faydalandıđım ve hibir zaman yardımı esirgemeyen tez danıřmanım Sayın Do. Dr. Pınar Yıldırım'a ve Prof. Dr. B. Tevfik AkgŸn'e sonsuz teŐekkŸrlerimi sunarım

Ayrıca tez alıřmam boyunca teknik bilgi, fikir, zaman ve emeklerini esirgemeyen iř arkadařım; Emrah ŐimŐek'e ve Vural Akkaya'ya teŐekkŸrŸ bir bor bilirim

Son olarak tŸm hayatım boyunca hibir zaman desteklerini esirgemeyen ve her durumda yanımda olan aileme sonsuz teŐekkŸr ederim...

İÇİNDEKİLER

I. GİRİŞ.....	1
1.1. Tezin Amacı.....	3
1.2. Tezin Kapsamı	3
II. VERİ MADENCİLİĞİ VE ALGORİTMALAR.....	4
2.1. Veri Madenciliği	4
2.2. Veri Madenciliği Uygulama Süreci.....	6
2.2.1. Standart veri madenciliği süreçleri	6
2.2.2. Problemin tanımlanması	9
2.2.3. Verileri anlama	10
2.2.4. Verilerin hazırlanması.....	10
2.2.5. Modelin belirlenmesi	15
2.2.6. Modelin değerlendirilmesi	16
2.2.7. Modelin gösterimi.....	16
2.3. Veri Madenciliği Teknikleri	17
2.3.1. Birlikte kuralları	17
2.3.2. Sınıflandırma	18
2.3.3. Kümeleme	20
2.4. Veri Madenciliği Kullanım Alanları ve Yararları	22
2.5. Metin Madenciliği.....	23
2.5.1. Tarihçe	24
2.5.2. Metin madenciliği çalışma alanları.....	25
2.5.3. Metin sınıflandırma	25
III. VERİ MADENCİLİĞİ İÇİN SOSYAL MEDYA ÜZERİNDEN NESNE TABANLI BİR YAZILIM ARACI.....	26
3.1. Giriş.....	27
3.2. Yazılım Aracının Yapısı ve Tasarımı	27
IV. METİN MADENCİLİĞİ İŞLEMLERİ VE SONUÇLARI İÇİN WEB TABANLI BİR YAZILIM ÜZERİNDEN YORUMLARININ ANALİZİ	30
4.1. İçerik Oluşturma.....	31
4.2. Hasta Yorumlarının İşlem Süreçleri.....	32
4.3. Veri Hazırlama Süreci	33
4.3.1. Kullanıcı, durum ve değerlendirme verilerin temizlenmesi	33
4.3.2. Ham veri ve yorumlardan çıkan yan etkilerin birleştirilmesi.....	34

4.3.3. İlaç yan etki tespitlerinin dönüştürme işlemleri	35
4.3.4. İlaç yan etki tespitlerinin indirgeme işlemleri	36
4.4. Yan Etki İsim Listesinin Oluşturulması	37
4.5. Yorumlardan Like Metodu ile Yan Etki Çıkarım İşlemleri.....	39
4.6. Yorumlardan Soundex Metodu ile Yan Etki Çıkarım İşlemleri.....	41
4.7. Kullanıcı Bilgilerinin Nitelik Olarak Ayrılması.....	45
4.8. Analiz Edilen Verinin Weka Dosyasına Aktarılma İşlemi	46
4.9. İlaç Analiz.....	48
4.9.1. Yan etkilerde ilaç dağılımı	48
4.9.2. Veri istatistik	49
4.9.3. İlaçlardaki yan etki değerlerin frekansı.....	50
4.10. Grafik Gösterimler	52
4.10.1. İlacın veri durum grafikleri	53
4.10.2. İlacın yan etki frekans grafiği.....	55
4.11. Kullanılan Algoritmalar ve Modellerin Yapısı	56
4.11.1. Verilerin Weka ve web panel üzerinden Apiori algoritması ile analizi	57
4.11.2. Verilerin web panel üzerinden Knn algoritması ile analizi	62
4.11.3. Verilerin web panel üzerinden K-Means algoritması ile analizi	68
V. MOBİL UYGULAMA VE SONUÇLARI	74
5.1. Mobil Uygulama Alanı.....	74
5.2. Mobil Uygulama Sonuçları.....	75
VI. SONUÇ.....	80
KAYNAKÇA.....	82
EKLER.....	84
EK A - B - C.....	84
EK P1-P2...P12.....	107
EK U1-2-3	115
ÖZGEÇMİŞ.....	118

TABLO LİSTESİ

Tablo 2.1. Örnek Sınıflandırma Veri Seti.....	19
Tablo 2.2. Örnek Kümeleme Veri Seti	21
Tablo 2.3. Örnek Kümeleme Sonuç Görünümü	22
Tablo 3.1. Veri Seti Yapısı	29
Tablo 4.1. İşlenen Veri Yapısı	33
Tablo 4.2. Örnek Ham Veri Yapısı.....	34
Tablo 4.3. Örnek Yan Etki Yapısı	34
Tablo 4.4. Örnek Standardizasyon Parametreleri	35
Tablo 4.5. Örnek Yan Etki Frekans Tablosu	35
Tablo 4.6. Örnek İlaç Yan Etki Çıkarım Tablosu.....	36
Tablo 4.7. Örnek İndirgenmiş İlaç Yan Etki Çıkarım Tablosu.....	36
Tablo 4.8. Örnek Yan Etki Listesi	37
Tablo 4.9. Örnek K-Means Verisi.....	70
Tablo 4.10. Örnek K-Means Gözlem Değeri.....	71
Tablo 4.11. Örnek K-Means Gözlem Atama Tablosu	71
Tablo 4.12. Örnek K-Means Atama Tablosu.....	72
Tablo 4.13. Örnek K-Means Merkez Koordinant Tablosu	72
Tablo 4.14. Örnek K-Means Atama Tablosu.....	72
Tablo 4.15. Örnek K-Means Merkez Koordinant Tablosu	72
Tablo 4.16. Örnek K-Means Atama Tablosu.....	73
Tablo 4.17. Örnek K-Means Küme Sınıf Verisi	73

ŞEKİL LİSTESİ

Şekil 2.1. 6 SIGMA İşlem Basamakları.....	7
Şekil 2.2. SEMMA İşlem Basamakları.....	7
Şekil 2.3. CRISP-DM Sürecinin İşlem Basamakları	9
Şekil 2.4. Metin Madenciliği Çalışma Diyagramı	24
Şekil 3.1. Html Sayfası Okuyucu Programın Ekran Görüntüsü	28
Şekil 3.2. İlaç Kullanımı Hakkında Yapılmış Yorum Örneği	29
Şekil 4.1. Hasta Forumu Verilerinde İlaç Kullanımlarını Tanımlamak İçin Bir İş Akışı... 31	
Şekil 4.2. Uygulama İş Akışı	31
Şekil 4.3. Web Panel Yan Etki Çekim Robotu	38
Şekil 4.4. Mysql Yan Etki Tablosu.....	38
Şekil 4.5. Web Panel Yan Etki Çekim Like Robotu.....	39
Şekil 4.6. Örnek Like ile Yorumlardan Çıkan Yan Etki Mysql Tablosu.....	40
Şekil 4.7. Örnek Like ile Yorumlardan Çıkan Yan Etki Listesi	40
Şekil 4.8. Web Panel Yan Etki Çekim Soundex Robotu	43
Şekil 4.9. Örnek Soudex ile Yorumlardan Çıkan Yan Etki Mysql Tablosu	44
Şekil 4.10. Örnek Soundex ile Yorumlardan Çıkan Yan Etki Listesi	44
Şekil 4.11. Web Panel Üzerinden Kullanıcı Bilgilerini Ayırıştırma Robotu.....	45
Şekil 4.12. Kullanıcı Bilgilerinin Ayırıştırıldığı Mysql Tablosu	45
Şekil 4.13. Web Panel Üzerinden Kullanıcı Bilgilerini Ayırıştırma Robotu.....	46
Şekil 4.14. Web Panel Üzerinden Seçilen Örnek İlaçın Weka Dosyası	47
Şekil 4.15. Yan Etki Metot Mysql Tablosu	47
Şekil 4.16. Yan Etki Weka Çıkarım Mysql Tablosu	47
Şekil 4.17. Web Panel Üzerinden Yan Etkilerin İlaç Dağılım Görüntüsü.....	49
Şekil 4.18. Web Panel Üzerinden Veri İstatistik Görüntüsü	50
Şekil 4.19. Web Panel Üzerinden İlaç Frekans Filtre Görüntüsü	51
Şekil 4.20. Web Panel Üzerinden İlaç Frekans Görüntüsü.....	52

Şekil 4.21. Web Panel Üzerinden Grafik Seçimi.....	52
Şekil 4.22. Web Panel Üzerinden Yaş Cinsiyet Dağılım Grafiği.....	53
Şekil 4.23. Web Panel Üzerinden Cinsiyet Oran Grafiği	54
Şekil 4.24. Web Panel Üzerinden Kullanım Süresi Cinsiyet Dağılım Grafiği	54
Şekil 4.25. Web Panel Üzerinden Kullanım Süresi Yaş Dağılım Grafiği	55
Şekil 4.26. Web Panel Üzerinden İlaç Yan Etki Frekans Grafiği.....	56
Şekil 4.27. Web Panel Üzerinden Örnek Apriori Filtre Görünümü	61
Şekil 4.28. Web Panel Üzerinden Örnek Apriori Sonuç Görünümü	61
Şekil 4.29. Knn Uzaklık Formülleri.....	62
Şekil 4.30. Örnek Knn Grafik.....	63
Şekil 4.31. Örnek Knn İlişli Grafiği.....	63
Şekil 4.32. Örnek Knn Sınıflanmış Grafik.....	64
Şekil 4.33. Web Panel Üzerinden Örnek Knn Uygulaması	64
Şekil 4.34. Web Panel Üzerinden Örnek Uygulanan Knn Algoritması.....	67
Şekil 4.35. Web Panel Üzerinden Örnek K-Means Uygulaması	69
Şekil 4.36. Örnek Kümelenen Veri Görünümü	70
Şekil 5.1. Uygulamada Bulunan İlaç Yan Etki Listesi Görünümü	75
Şekil 5.2. Uygulamada Bulunan Örnek İlaç Frekansı.....	76
Şekil 5.3. Uygulamada Bulunan Örnek Grafik.....	77
Şekil 5.4. Uygulamada Bulunan Örnek Apriori Çıktısı.....	78
Şekil 5.5. Uygulamada Bulunan Örnek Knn Çıktısı.....	78
Şekil 5.6. Uygulamada Bulunan Örnek K-Means Çıktısı.....	79

I. GİRİŞ

Her geçen gün büyüyen bilişim sektörü alanında giderek artan verilerden elle tutulur kullanışlı bilgiler elde etmenin önemi artmış ve bilgiye ulaşma ve bilgiyi depolama sistemlerinin gücü çok büyük oranda artmıştır. Bu gelişmeler tüm sektörlerde faaliyet gösteren kurum ve kuruluşlar için yeni fırsatlar ortaya çıkmasına olanak sağlamakta olup veriyi mümkün olduğunca kullanıp bundan kazanç sağlama yoluna gitmek gerekmektedir.

Veri analizcileri tarafından yapısal olmayan büyük veri yığınlarından faydalanmak için bilgi çıkarım teknikleri kullanılarak faydalı bilgi ortaya çıkartmak için analiz yöntemleri ve kuralların uygulanması ile saklı kalmış bilgilerin keşfedilmesi ile anlamlı bilgilerin karar verme süreci içerisinde kullanılmasına kadar olan bütün süreci veri madenciliği kapsamaktadır.

Büyük veri tabanları içerisindeki verilerden anlamlı ve ilginç bilgilere ulaşmak için yeni algoritmalar ve metotlar geliştirilmiştir. Bu metotlara, veri madenciliği ve metin madenciliği algoritmaları örnek olarak verilebilir.

Sosyal medya da bulunan veriler genellikle yapısal olmayan verilerdir ve metin madenciliği içerisinde bu haliyle kullanılmazlar. Genellikle işletmelerde text, word, jpeg, pdf uzantılı dosyalarda veriler saklanmaktadır. Bunlarda yapısal olmayan verilere örnek

gösterilebilir. Öncelikle bu verilerin yapısal hale getirilmesi gerekir. Veriler yapısal hale getirildikten sonra üzerinde işlemler yapılabilir.

Bu çalışmada, ilaçların yan etkilerin kullanım sonrası gözlemlerinde değerlendirerek dünyada yaygın olarak görülen yanlış ilaç seçimi ve ilaç araştırmaların tedavisinde etkin olarak bilgi sağlayıcı olması öngörülmüştür.

İnternet sitesi üzerinden alınan yorumlar ile çeşitli yöntemler kullanılarak bir veritabanı oluşturulmuştur. Oluşturulan yapıda kural tabanlı metin ayrıştırma, eşleştirme ve çeşitli veri madenciliği yöntemleri ve araçları ile analiz edilmiş ve mobil ortama taşınmıştır. Çalışmada elde edilen sonuçlar ilgili tıp alanındaki uzmanlara da danışılarak değerlendirilmiştir.

İlaçların yan etkilerine ait çalışmalar, genellikle ilaç şirketleri tarafından yürütülmektedir. İnternet kullanımının yaygınlaşması ile ilaç kullanan kişiler çeşitli web siteleri aracılığı ile yaşadıkları süreçleri ve görülen yan etkileri paylaşmışlardır. Bu bilgiler ilaç üreten şirketler, doktorlar ve bilgi edinmek isteyenler için önemli bilgi kaynağı olmaktadır.

İlaç yorum siteleri ve tartışma forumların da yazılan yorumları klinik dil ve terminoloji kullanımına uygun olmakta ya da yanlış yazımlar ile ifade edilen paylaşılan verilerde anlamlı bilgiye ulaşmak için veri madenciliği çalışmalarının çok önemli olduğunu ortaya koymaktadır.

1.1 Tezin Amacı

Bu tez çalışmasının amacı, webMD web sitesi üzerinden ilaçlara yapılan yorumlarını bir yazılım aracı geliştirerek metin içerisinde yan etki tespit edecek bir yöntem geliştirmek ve bu bilgilere dayalı olarak ilgili kişilere, kullanıcılara analiz edilen verilere ait faydalı sonuçlar sunmaktır.

Bu çalışmada, webMD sosyal medya üzerindeki son 10 yıllık ilaç yorum kayıtları üzerinde çalışma yapılmıştır ve bu kayıtlar ile sistem eğitilerek kelimelerden yapısal veri elde edilmiştir. Visual Studio.NET platformu üzerinde C# programlama dili ile bir Windows Form uygulaması geliştirilmiştir.

1.2 Tezin Kapsamı

Tez çalışmasında kullanılan veri, yapısal türde olmayan veridir ve bu veriyi kullanabilmek için yapısal veri şekline çevirmek gerekir. Metin madenciliği sürecinde bu çevrim yapılmaktadır.

Tezin ilk bölümünde giriş yapılmakta ve tezin amacı açıklanmaktadır. İkinci bölümde, Metin madenciliği konusu ve metin madenciliği adımları anlatılmaktadır. Üçüncü bölümde, Tıp ve sağlık hizmetlerinde veri madenciliği anlatılmaktadır. Dördüncü bölümünde, Veri madenciliği için sosyal medya üzerinden nesne tabanlı bir yazılım aracı anlatılmaktadır. Beşinci bölümde, Metin madenciliği işlemleri ve sonuçları için web tabanlı bir yazılım üzerinden yorumların analizi anlatılmaktadır. Altıncı bölümde, Mobil uygulama ve gösterilmektedir. Tezin sonuç kısmıyla tez bitirilmektedir.

II. VERİ MADENCİLİĞİ VE ALGORİTMALAR

2.1. Veri Madenciliği

Genel olarak, Büyük miktarlarda verinin elde edilmesi ile bu verinin kullanışlı ve anlamlı bilgiye dönüştürülme işlemi denilebilir.

Veri madenciliği kısaca, büyük miktarda veriden anlamlı bilginin (knowledge) elde edilmesidir (extraction & mining). Knowledge mining de denilebilir. Birçok araştırmacı bu yüzden Veri Madenciliği yerine Knowledge Discovery from Data ifadesini kullanmaktadır.

Veri madenciliği konusunda bahsi geçen geniş verideki geniş kelimesi, tek bir iş istasyonunun sahip olduğu bütünsel ve hücrel belleğine sığamayacak kadar büyük veri kümelerini ifade etmektedir. Yüksek hacimli veri ise, tek bir iş istasyonundaki ya da bir grup iş istasyonundaki disklere sığamayacak kadar fazla veri anlamındadır. Dağıtık veri ise, farklı coğrafi konumlarda bulunan birbirinden bağımsız salt verileri anlatır[1].

Veri madenciliği, veri bütününde gizli kalmış, önceden bilinmeyen ve potansiyel olarak kullanışlı olan anlamlı bilginin çıkarımıdır[2].

Veri analiz yöntemlerinde doğal dil yöntemleri kullanılmaktadır yapılmış çalışmalara baktığımızda sosyal medya da günden güne artan veriler araştırmalara çok önemli kaynak sağlamaktadır. Örneğin araştırılan ilaçların ilişki yapısı internet sitelerinden metin içerisinde bilgi çıkarımı yapılarak veri içerisinde gizli kalmış bilgileri veri madenciliği teknikleri ile ilaç özellik çıkarımları sağlanmıştır. İlaç kullanım özellikleri sınıflandırılmış verilerin durumu incelenmiş olup bu sayede doğru ilaç kullanımı ve yanlış ilaç kullanımı aynı tablolarda skor değerleri referans verilerek anlamlı bilgi sunumu haline gelmiş ve önceden bilinmeyen potansiyel olarak kullanışlı olan anlamlı bilgiler kazanılmıştır.

Veri madenciliği için yapılmış araştırma metotlarında fikir madenciliği ilgiyi üzerine çekmektedir. Örneğin İlaç forum siteleri üzerinden alınan bilgiler içerisinde metinsel özellikler sağlık için yeni teknikler geliştirmek tıbbi bilgilerin içeriği hakkında en önemli alanı ortaya koymuştur. Bu durum da web' de hasta bilgileri içerisinde öznitelik alınarak bilgilendirici analiz yapılmakta olup bu gibi tekniklerin sonuçlarına bakıldığında yararlı sonuçlardan çıktığı bilinmektedir. Bilimsel özellikler laboratuvar ortamında karşılık gelen bilgi sayısı ile kullanım sonrası gelen bilgiden çok az ve yetersiz olması bu gibi tekniklerin çok önemli olduğunu ortaya çıktığını göstermiştir.

2.2. Veri Madenciliği Uygulama Süreci

Veri Madenciliği uygulanırken, standart veri madenciliği süreçleri, problemin tanımlanması, verileri anlama, verilerin hazırlanması, modelin belirlenmesi ve modelin dağılımı gibi birçok süreçten geçmektedir. Bu süreçlerin detayları aşağıdaki gibidir.

2.2.1. Standart veri madenciliği süreçleri

Altı Sigma (Six Sigma), SEMMA (Sample, Explore, Modify, Model, Assess) ve CRISP-DM (Cross- Industry Standard Process for Data Mining) metodolojileri en yaygın olarak kullanılan veri madenciliği süreçleridir.

2.2.1.1. Altı Sigma (Six Sigma)

Altı Sigma'nın tedarik zincirindeki sorunları ve zayıf yönleri için işletmelerde kalite iyileştirme strateji süreçlerini tanımlanması amacıyla (**define**), ölçme (**measure**), analiz (**analyze**), iyileştirme (**improve**) ve kontrol (**control**) kolay ve etkili istatistik yöntemlerinin kullanıldığı bir iş yönetim stratejisi olarak tanımlanmaktadır[3]. Bu çalışmada Altı Sigma değerlendirme aşamalarında dikkate alınmış ve işlem durumları veri analiz süreçlerinde kullanılmıştır. Şekil 2.1.' de 6 SIGMA'nın süreci aşama aşama olarak gösterilmiştir



Şekil2.1. 6 SIGMA Sürecinin İşlem Basamakları

2.2.1.2. SEMMA (sample, explore, modify, model, assess)

SEMMA, SAS şirketi tarafından yaratılan ve veri madenciliğinde yürütülen süreçlerin özünü gösteren, örnekle (Sample), araştır (explore), değiştir (modify), modelle (model), değerlendir (assess) kelimelerinin baş harflerinden oluşmaktadır[4]. Bahsedilen SEMMA süreçlerinin aşamaları aşağıdaki gibidir.



Şekil 2.2. SEMMA İşlem Basamakları

Örnekle (Sample), modelleme için belirlenen yapıdan, veri seçme ve veri örnekleme ile başlar. Örneklenen veri seti verimli kullanılabilir kadar küçük ama içerdiği bilgi açısından yeterince büyük olmalıdır. Bu aşama aynı zamanda veri bölümlenmeyle de ilgilidir.

Araştır (Explore), bu aşama beklenmeyen eğilimleri arayarak veriler üzerindeki ilişkileri keşfeder, yeni bakış açısı ve fikirler kazandıran süreci kapsar.

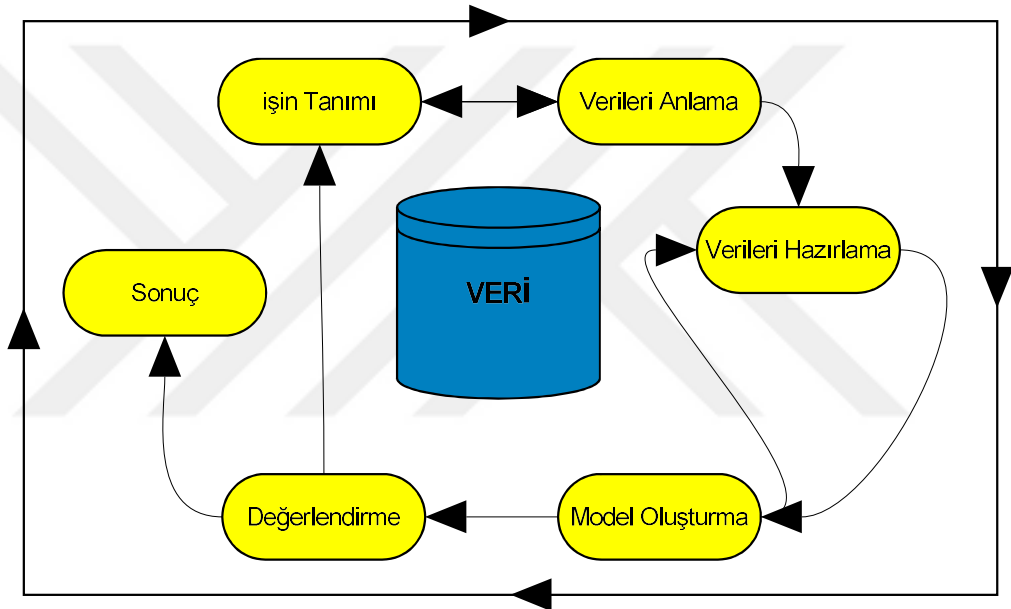
Değiştir (Modify), veri modelleme süreci için hazırlık yaparak, veri oluşturma, veri değiştirme ve veri seçimi sürecini kapsar.

Modelle (Model), aşamasında sonuçları alabilmek için hazırlanan veri seti üzerinde çeşitli veri madenciliği tekniklerinin uygulandığı süreçtir.

Değerlendir (Asses), son aşamada modelleme süreci sonuçlarının değerlendirilmesini, oluşturulan modellerin güvenilirlik ve kullanılabilirlik durumunu gösterir[5].

2.2.1.3. CRISP-DM (cross- industry standard process for data mining)

CRISP-DM modeli, bir veri madenciliği projesini altı aşamaya ayırır ve farklı aşamalar arasında ileri geri gitmeye ihtiyaç duyulmasına izin verir. Yaklaşımın ismi CRISP-DM (CRoss-Industry Standard Process for data mining). Orijinal üyeleri Daimler-Benz, SPSS ve NCR olan bir consorsiyum tarafından geliştirilmiştir. CRISP-DM modeli, bir veri bilimi projesine eşit derecede iyi uygulanır.



Şekil 2.3. CRISP-DM Sürecinin İşlem Basamakları

2.2.2. Problemin tanımlanması

Veri madenciliğinde, en önemli unsur problemin açık ve net bir şekilde ortaya konularak uygun çözüm yöntemlerinin geliştirilmesidir. Çözüm yöntemlerinin doğru olması, problem analizlerinin anlaşılır olmasını sağlamak ve sınıflandırılması iyi oluşturulmasına bağlıdır, başarılı sonuçlar alındığında faydaları elde edilecek başarısız sonuçlarda dezavantajları, işletmeye olan maliyeti nasıl olacağına yer verilmelidir.

Bu tez çalışmasında tanımlanan problem, her geçen gün artan ilaç kullanımların nasıl etkilerini olduğunu nasıl sonuçlar verdiğini mevcut kullanım yapan hasta veya bakıcıların görüşlerini internet de ki sosyal paylaşım platformları üzerinden alarak analiz ederek kullanıcı üzerindeki yan etkileri ortaya koymaya çalışılmaktadır.

2.2.3. Verileri anlama

Günümüzde verinin ve bilginin bir güç olarak kabul edilmesiyle birlikte metin madenciliği kavramı önem kazanmıştır. İnsanların uzun bir yazıdan alması gereken fikirler, metnin içerisinden seçilen anahtar kavramlardan çok daha kolay bir şekilde anlaşılabilir.

Başlangıç verilerinin oluşturulması, verileri tanımaya ve anlamaya yönelik analizlerin yapıldığı veri madenciliği sürecidir. Ayrıca veri kalitesi sorunları da bu aşamada değerlendirilir. Veri madenciliği uzmanları verileri anlama aşamasında çok sık bilgi alışverişinde bulunarak daha anlamlı sonuçlar elde edilir.

2.2.4. Verileri hazırlama

Veri hazırlama aşaması, ham veriden başlayarak son veriye kadar yapılması gereken bütün düzenlemeleri içermektedir. Veriler hazırlanırken verileri ön işlemeden geçirme sürecinde hangi değişkenleri ve hangi kayıtların analizde kullanılacağına karar verilir. Genellikle veriler birden çok farklı kaynaklardan toplanarak eksik veya gürültülü verilerin olup olmadığı kontrol edilen veri madenciliği süreçlerindedir.

Veri temizlemenin hedefi; girilen veri içerisinde aynı formatta olmayan veya yanlış girilmiş verilerin kaldırılmasıdır. Eksik veri alanları, o alanın ortalaması alınarak da doldurulabilir ya da bu eksik veriler tamamen kaldırılır. Gürültülü veriler ise veri girişinin hatalı olması durumunda ortaya çıkmaktadır. Veri hazırlama dört aşamada gerçekleştirilir.

- Veri Temizleme,
- Veri Birleştirme,
- Veri Dönüştürme,
- Veri İndirgeme

2.2.4.1. Veri temizleme

Veri setinden anlamlı bilgilerin elde edilmesi için veri madenciliği uygulama sürecinin ilk basamağı olan veri temizleme işlemi yapılır. Örnek olarak eksik, tutarsız veya uygun olmayan veriler ile karşılaşılabilir.

Verileri amacına uygun hale getirebilmek için veri doldurma, sapan verileri azaltma, verilerin tutarsızlığını düzeltme gibi metotlar ile veriler temizlenir. Verilerde bulunan gürültünün temizlenmesi ve eksik verilerin yerine yenilerinin belirlenerek konulması gerekebilir bunun için eksik değer bulunduran kayıt ya da kayıt kümeleri veri setinden de atılabilir.

2.2.4.2. Veri birleřtirme

İřlenebilir veriler elde etmek için farklı veri tabanlarından ya da veri kaynaklarından elde edilen verilerin birlikte deęerlendirilmesi için farklı yapıda ki verilerin ortak bir yapıya dnřtrlmesi bu ařamada yapılır.

Veri birleřtirme için veri ambarı yapılabilir rneęin 2010-2017 yılları için e ticaret satıř tutarlarından oluřan bir veri kmesinin bulunduęunu varsayalım bu yıllar için yıllık satıř tutarları tek bir tabloda toplandıęında veri birleřtirmesi yapılmıř olur. Sonu olarak elde edilen veri kmesinin hacmi daha kktr fakat yapılacak analiz için bir bilgi kaybı sz konusu deęildir. Veri ambarları ise ok deęiřkenli birleřtirilmiř bilginin saklandıęı veri tabanlarıdır. Veri ambarları zet bilgiye herhangi bir hesaplama yapmadan hızlı bir biimde eriřilmesini olanak verir.

Veri birleřtirme yapılırken birok durumun nemle dikkatli olması gerekir. rneęin arařtırmacı iki farklı veritabanında Tablo_No ile Tablo_İndeks alanlarının aynı olduęundan emin olması için yapısını bilmesi gerekir. Buna veriye iliřkin veri anlamında st veri (metadata) denir[6].

2.2.4.3. Veri dönüştürme

Bazı veri madenciliği araçları metin değerleri ile işlem yapabilir; fakat bazı araçlar da tüm alanların sayısal olmasını gerektirebilir. Sıralı değerlerde matematiksel işlemler ve/veya karşılaştırmalar (küyüktür, küçüktür gibi) yapabilmek için bu değerleri sayısal verilere dönüştürmek gerekiyor. Örneğin üniversite sene sonu ağırlıklı puanı (GPA) Denklem 2-1...4 ' de olduğu gibi sayısal verilere dönüştürülür:

$$A+ \rightarrow 4.0 \quad (2-1)$$

$$A- \rightarrow 3.7 \quad (2-2)$$

$$B+ \rightarrow 3.3 \quad (2-3)$$

$$B \rightarrow 3.0 \quad (2-4)$$

Kategorik puanlar sayısal değerlere dönüştürüldüğünden matematiksel işlemler yapılabilir bu sayede veri madenciliği için uygun formlara dönüştürülürler.

Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden birisidir. Veri normalleştirme min-max bu tekniklerinden birisidir[7]. Min-max normalleştirme ile orijinal veriler yeni veri aralığına doğrusal dönüşüm ile dönüştürülürler. Bu veri genellikle 0-1 Aralığında olur.

2.2.4.4. Veri indirgeme

Veri madenciliği uygulamalarında kullanılan verinin çıkacak sonuca etki etmeyeceğine inanılıyorsa veri sayısı ya da değişkenlerin sayısı azaltılabilir.

Veri indirgeme yöntemleri;

- Veri Sıkıştırma (Data Compression)
- Boyut indirgeme (Dimension Reduction)
- Veri birleştirme (Data Aggregation)
- Kesikleştirme (Cutting)

Veri Sıkıştırma: Orijinal verileri temsil edecek indirgenmiş veya sıkıştırılmış veriler elde etmeye dayalı bir yöntemdir. Veri şifreleme veya veri dönüştürme veri sıkıştırmada uygulanan yöntemlerdir.

Boyut indirgeme: Oldukça popüler bir yöntem olduğu bilinmektedir. Karar ağaçları ile ele alınan çıktı değişkenini en iyi temsil edecek değişken kümesi belirlenmesidir.

Veri birleştirme: Bu yöntem bir önceki alt başlık olan veri dönüştürmede kullanılan bir yöntem olarak anlatılmaktadır. Benzer şekilde veri indirgemedeki veri birleştirme yöntemi kullanılmaktadır.

Kesikleştirme: Bazı veri madenciliği algoritmaları sadece kategorik değişkenlere göre işlem yapabilmektedirler. Bu tür algoritmalarının kullanılması gerektiğinde sürekli verilerin kesikli verilere dönüştürülmesi ile elde edilen kategorik değerler, orijinal veri değerlerinin yerine kullanılmaktadırlar.

2.2.5. Modelin belirlenmesi

Belirtilen probleme uygun modelin belirlenmesi, çok fazla sayıda modelin kurup test edilmesi ile olur. Bu nedenle veri hazırlama ve modelin belirlenmesi adımları, en etkin model bulununcaya kadar devam eder. Modelin belirlenmesi, denetimli ve denetimsiz öğrenme olarak ikiye ayrılır[8].

Denetimli öğrenmede sınıflar belirlenir. Bu belirlenen sınıflara göre kriterlere ayrılır. Her sınıf için çeşitli örneklemeler yapılır. Sistem bu örneklerden yola çıkarak her bir sınıfa ait özellikleri ve bu özellikleri ifade eden nitelik cümleleri ile tanımlar. Örnekler yeteri kadar tanımlandığında nitelik cümleleri yeni gelen örneklere uygulanır ve yeni örneklerin hangi sınıfa ait olduğu model yardımıyla belirlenir. Burada kurulan modelin doğruluğunu test etmek için elimizdeki örnek verilerin bir kısmı test için ayrılabilir. Model kurulduktan sonra test verilerinden alınan sonuçlara bakılır. Bu şekilde model üzerinde doğrulama testi yapılabilir.

Denetimsiz öğrenmede, denetimli öğrenmenin aksine daha önceden bilinen sınıflandırmalar ve etiketlemeler yoktur. Örnekler gözlemlenir, çıktı veya sonuca bağlı değişken bulunmaz. Bu örneklerin arasındaki benzerlikler tespit etmeye çalışılır. Bu modelde benzer örüntü kurallarına sahip veriler gruplanabilir[9].

2.2.6. Modelin deęerlendirilmesi

Modelin deęerlendirilmesi iin burada belirlenen modelin alıřmanın amacına tam anlamıyla gerekleřtirmesi veya ilk sure de belirlenen problemin tam olarak analiz edilmesi gereklidir. Bu sebeple modelin deęerlendirme sureci ok nemlidir. Gerekirse tm adımlar en bařtan bařlanarak gzden geermeli ya da deęiřiklik zerinde durulmalıdır. Bu sure sonunda veri madencilięinde elde edilen ıktıları kullanılıp kullanılmayacaęıyla ilgili karar alınır[10].

2.2.7. Modelin gsterimi

Burada nceki adımlarda gerekleřtirilen iřlemler artık son kullanıcının kullanacaęı ve anlayacaęı řekilde kullanıma sunulur. Amacına uygun yorumlanan sistemlerde ncelik sıralaması yapılır ve kullanıcının ncelikli hangi iřlem yapacaęını sistem model daęılımı yapılırken kurgulanmıř olur bu sayede daha kullanıřlı ulařabilir hızlı ve anlaşılır iřlemler ortaya ıkarılmaktadır.

Model sistemleri etkileyecek dzeyde kurguladıęında destekleyici aralar ile veri tabanlarından bilgi keřfi yapacaęımız bir sure sz konusu olmaktadır. Modelin daęılımı sistemin bu ařamasında surecin en nemli kısmı oluřturmaktadır.

2.3. Veri Madenciliği Teknikleri

Veri madenciliği teknikleri işletmelerin zaman içerisinde sahip olduğu büyük boyutlardaki verilerden alışlagelen istatistik yöntemlerle elde edilemeyen veya elde edilmesi güç olan bilgileri elde etmek için bir çeşit verileri işleme ve çözümlene olanak sağlamaktadır. Veri madenciliği yöntemleri temelde 3 aşamada gerçekleştirilir.

- Birliktelik kuralları
- Sınıflandırma
- Kümeleme

2.3.1. Birliktelik Kuralları

Bir çok firma, büyük miktarda veriyi saklarlar. Örneğin süpermarketler daha sonra müşteri ilişki yönetimi, envanter yönetimi veya promosyonların belirlenmesi için kasalardaki satın alma işlemlerini, fiş ve fişin içeriği bazında tutarlar. Bu veriye genel olarak market sepeti verisi adı verilir.

Birliktelik analizi çalışması, yukarıdaki örnekteki gibi veriler içinde ilgi çekici ilişkilerin saptanması amaçlar. Çalışma sonunda saptanan ilgi çekici ilişkilere birliktelik kuralları ismi verilir. Örneğin aşağıdaki gibi bir kural yukarıdaki veriden çıkarılabilir.

{Çocuk Bezi} => {Bira}

Bu kural çocuk bezi alanların, bira alma olasılığının yüksek olduğunu söylemektedir. Süper market yöneticileri bu tür verilerden yararlanarak çapraz satışa ilişkin promosyonlar yapmaktadırlar.

İstatistiksel bir bakış açısı ile birliktelik analizi için bir dizi değişken arasındaki korelasyonun yarı otomatik tespiti diyebiliriz.

Birliktelik analizi market sepeti verisi dışında, tıbbi tanı konulması, bioinformatik , yerbilimleri ve web madenciliği gibi alanlarda da kullanılmaktadır. Örneğin yer bilimcileri, birliktelik analizi kullanarak, atmosferdeki değişimler ile gezegenin yüzeyindeki değişimlerin birliktelik ilişkileri üzerine çalışmaktadırlar.

Birliktelik analizinde 2 önemli konu vardır. Bunlardan birincisi; büyük verileri işlemenin aldığı zaman ve ihtiyaç duyduğu donanım anlamında işlemsel olarak çok maliyetli olması. İkincisi ise, analiz sonucunda bulunan birliktelik örüntülerinin bazılarının şansa bağlı ve gerçeği yansıtmayan örüntüler olabilmesi. Birliktelik analizine ilişkin yapılan çalışmalar, bu iki konu üzerine yoğunlaşır.

2.3.2. Sınıflandırma

Sınıflandırma kavramı, basitçe bir veri kümesi (data set) üzerinde tanımlı olan çeşitli sınıflar arasında veriyi dağıtmaktır. Sınıflandırma algoritmaları, verilen eğitim kümesinden bu dağılım şeklini öğrenirler ve daha sonra sınıfının belirli olmadığı test verileri geldiğinde doğru şekilde sınıflandırmaya çalışırlar. Örneğin Tablo 2.1.' de öğrencilerden toplanan bir veri kümemiz bulunsun ve alt satırlarında ortalamaları olsun.

Tablo 2.1. Örnek Sınıflandırma Veri Seti

Sıra No	Yaş	Boy	Kilo	Cinsiyet
1	19	181	75	Erkek
2	20	183	90	Erkek
3	20	175	70	Erkek
4	21	179	80	Erkek
5	19	162	50	Kız
6	21	171	57	Kız
7	22	169	55	Kız
1	20,67	167,33	54,00	Kız Ortalama
2	20,00	179,50	78,75	Erkek Ortalama

Bir sınıflandırıcı yöntemimiz, Tablo 2.1.' de veri setine bakarak verilen yaş boy ve kilo değerlerine göre bir öğrencinin cinsiyetini öğrenecek olsun. Yani yukarıdaki veri kümesini bir eğitim kümesi olarak kullanacağız. Ardından gelen yeni bir kayıt için, yaş, boy ve kilo değerleri verildiğinde, sınıflandırıcımız cinsiyetini otomatik olarak tahmin edecek. Çok sayıdaki sınıflandırma algoritmalarından basit birini seçelim. Diyelim ki sınıflandırma algoritmamız, verilen etiketteki değerlerin ortalamasını alacak ve bu ortalama değer, öğrendiği değer olacak. Ardından gelen test değerleri için bulmuş olduğu ortalamaya uzaklığına bakacak ve kime yakınsa o etiketten kabul edecek.

Diyelim ki yeni gelen ve test edilmesini istediğimiz değer de, yaş 21, boy 165, kilo 60 değerlerinde bir kişi olsun. Şimdi Knn algoritmamız öğrendiği değerlere göre bu yeni gelen kişinin cinsiyetini tahmin etmeye çalışacak. Basitçe her değere olan mesafeyi hesaplayacak (burada da çok farklı mesafe hesaplama algoritmaları olmasına karşılık biz yine amacımız temel kavramlar olduğu için konuyu basit tutarak Öklit mesafesi (euclidean distance) kullanılmıştır. Mesafeler sıralanır ve en minimum uzaklığa bağlı olarak en yakın

komşular bulunur. Algoritmamızın Erkek tanımından öğrendiği değerler ile yeni gelen kişinin mesafesini Denklem 2.5.' de gibi hesaplanmıştır:

$$\sqrt{((20-21)^2+(179.5-165)^2+(78.75-60)^2)}=23.72 \quad (2-5)$$

Benzer şekilde kızlar için öğrendiğimiz değere olan mesafesini Denklem 2.6.' da gibidir.

$$\sqrt{((20.66-21)^2+(167.33-165)^2+(54-60)^2)}=6.44 \quad (2-6)$$

Buna göre algoritmamızın verdiği değer, erkeklere olan mesafesinin 23.72 olduğu ve kızlara olan mesafesinin 6.44 olduğudur. Yani Öklit uzaklığı daha yakın olmasından dolayı algoritmamız yeni gelen kişiyi kız sınıfından tanımlamıştır.

2.3.3. Kümeleme

Kümeleme algoritmaları verilerin kendi aralarındaki benzerliklerin göz önüne alınarak gruplandırılması işlemidir. Kümeleme yöntemlerinin çoğu veri arasındaki uzaklıkları kullanır. Nesnelere birbirine olan yakınlık ve uzaklıklarına göre kategorilere ayrılır. Öncelikle eldeki veriler incelenir. Birbirine benzeyen nesnelere bir kümeye, benzemeyenler ise başka bir kümeye ayrılır. Böylece ilk olarak kümelemeye hangi nesnelere dâhil edileceği belirlenir.

Kümeleme algoritmalarının amacı birbirleri ile benzer özellikleri olan verileri alt kümelere ayırmaktır. Ele alınan verilerin metinlerden oluştuğu durumlarda kümeleme yöntemlerini uygulamak için öncelikle dokümanların sayısal verilere çevrilmesi gerekir. Her bir dokümanın özellik matrisleri oluşturularak kümeleme algoritmalarında uygulanabilir hale getirilir. Veri boyutu çok büyük olduğunda, algoritmaların çalışma zamanları oldukça

uzun olabilir. Bu yüzden doğru seçimler yaparak algoritmaların daha kısa sürede en iyi çözümü vermesi beklenir.

Tablo 2.2. Örnek Kümeleme Veri Seti

Gözlemler	Değişken 1	Değişken 2	Küme Üyeliği
X1	4	2	C1
X2	6	4	C1
X3	5	1	C2
X4	10	6	C1
X5	11	8	C2

Tablo 2.2.' de kümelerin sayısına başlangıçta $k=2$ biçimde karar veriyoruz. Başlangıçta tesâdüfi olarak $C1 = \{X1,X2,X4\}$, $C2 = \{X3,X5\}$ olarak iki küme belirliyoruz. Belirtilen iki kümenin merkezi Denklem 2.7 - 8.' de ki hesaplanır.

$$M1 \{(4+6+10)/3, (2+4+6)/3\} = \{6.67,4.0\} \quad (2-7)$$

$$M2 \{(5+11)/2, (1+8)/2\} = \{8.00,4.5\} \quad (2-8)$$

$M1$ ve $M2$ merkezlerinden olan uzaklık minimum olması istendiğinden Denklem 2.9 -10.' da ki gibi Öklit uzaklık formülü kullanılarak $X1$ gözlem için hesaplanır.

$$d(M1,X1) = \sqrt{(6.67 - 4)^2 + (4 - 2)^2} = 3,33 \quad (2-9)$$

$$d(M2,X1) = \sqrt{(8 - 4)^2 + (4,5 - 2)^2} = 4,72 \quad (2-10)$$

Tüm gözlem noktalarının merkeze uzaklığı hesaplanmış ve bu işlemlerden anlaşılıyor ki $X1$ gözlem değerinin $M1$ ve $M2$ merkezlerine uzaklıkları $d(M1,X1) < d(M2,X1)$ oldundan

M1,X1 daha yakın görünüyor. Benzer şekilde diğer gözlem noktalarına uyguladığımızda küme üyelikleri Tablo 2.3.' de ki tablo elde edilir.

Tablo 2.3. Örnek Kümeleme Sonuç Görünümü

Gözlemler	M1 den uzaklık	M2 den uzaklık	Küme Üyeliği
X1	$d(M1,X1) = 3,33$	$d(M2,X1) = 4,72$	C1
X2	$d(M1,X2) = 0,67$	$d(M2,X2) = 2,06$	C1
X3	$d(M1,X3) = 3,43$	$d(M2,X3) = 4,61$	C1
X4	$d(M1,X4) = 3,89$	$d(M2,X4) = 2,50$	C2
X5	$d(M1,X5) = 5,90$	$d(M2,X5) = 4,61$	C2

Bu durumda yeni kümeler $C1=\{X1, X2, X3\}$, $C2=\{X4, X5\}$ olarak belirlenecektir.

2.4. Veri Madenciliği Kullanım Alanları ve Yararları

Veri madenciliğinin Kullanım alanları oldukça geniştir. Farklı bilim dallarında ve sektörlerde uygulama alanları, analiz edilen verinin yapısı ve boyutuna göre farklılaşmak olup bu zamanlarda karar alma süreci içinde olduğumuz birçok alanda veri madenciliği çalışmaları sürekli olarak çalışmaktadır.

Veri madenciliği çalışma alanları bilimsel araştırma amacıyla yapılan çalışmalara kaynak sağlamaktadır. Bu çalışmada yapılmak istenen analizin araçları veri madenciliği tarafından sağlanmış olup kendisine özgü bilimsel sonuçlar ortaya koymuştur.

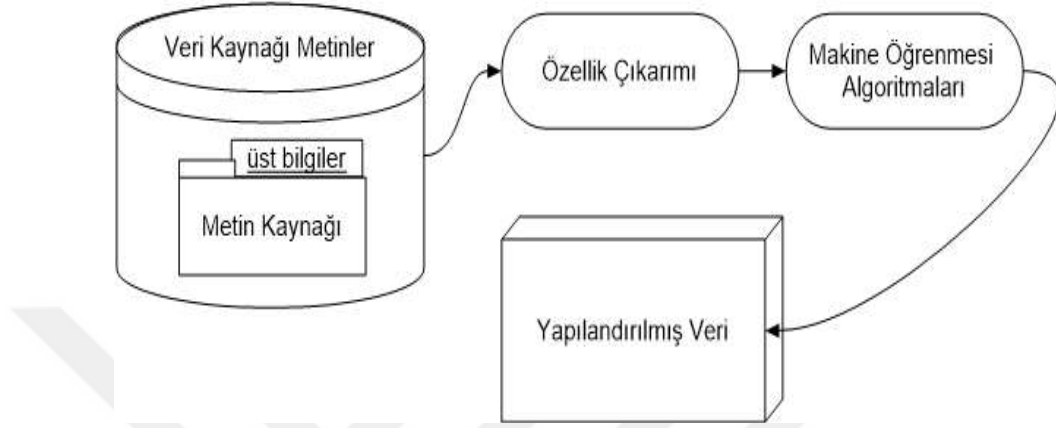
2.5. Metin Madenciliği

Metin madenciliği çalışmaları metni veri deposu olarak ele alır. Diğer bir tanımla metin üzerinden anlamlı veri çıkarmayı amaçlar. Çalışma mantığı şema olarak Şekil 2.4.' de belirtilmiştir. Metin madenciliği, metinlerin sınıflandırılması, bölütlenmesi (clustering), metinlerden kavram çıkarılması (concept/entity extraction), metinler için sınıf taneciklerinin üretilmesi (production of class labels), metinlerde görüş analizi yapılması (sentimental analysis), metin özetlerinin çıkarılması (document summarization) ve metnin özü ile ilgili ilişki modellemesi (entity relationship modelling) gibi çalışmaları hedefler[11].

Yukarıdaki hedeflere ulaşılması için metin madenciliği çalışmaları kapsamında enformasyon getirmesi (information retrieval), hece analizi (lexical analysis), kelime frekans dağılımı (Word frequency distribution), örüntü tanıma (pattern recognition), etiketleme (tagging), enformasyon çıkarımı (information extraction), veri madenciliği (data mining) ve hatta görselleştirme (visualization) gibi yöntemleri kullanmaktadır.

Metin madenciliği çalışmaları, metin kaynaklı literatür deki diğer bir çalışma alanı olan doğal dil işleme (natural language processing, NLP) çalışmaları ile çoğunlukla beraber yürütülmektedir. Doğal dil işleme çalışmaları en çok yapay zeka altındaki dil bilim bilgisine dayalı çalışmaları ele almaktadır. Metin madenciliği çalışmaları ise daha çok istatistiksel olarak metin üzerinden sonuçlara erişmeyi hedefler. Metin madenciliği

çalışmaları sırasında çoğu zaman doğal dil işleme kullanılarak özellik çıkarımı da yapılmaktadır[12].



Şekil 2.4. Metin Madenciliği Çalışma Diyagramı

2.5.1. Tarihçe

Analizcilerin metin madenciliği çalışmaları 1980' lerin ortalarında başladığı bilinmektedir, fakat bilim ve teknolojik gelişmelerin ışığında 2000' li yıllardan sonra önem kazanmıştır. Metin madenciliğinin tanımlar arası bir çalışma alanı olup bilgi üretimi, makine öğrenmesi, veri madenciliği, istatistik ve işlemsel dilbilim kavramlarının çalışmaktadır. Günümüzde bilginin büyük kısmı metin olarak tutulmaktadır[13]. Bu yüzden metin madenciliğinin de bilgi çıkarımı zor ve modeller geliştirilerek sağlıklı bilgi çıkarılmaları gereklidir. Ayrıca çok dilli veri madenciliği gibi farklı dillerden aynı özellikleri taşıyan değerlerin çıkarılması son derece önemlidir.

2.5.2. Metin madenciliği çalışma alanları

Metin madenciliği sıklıkla aşağıdaki problemlerle ilgilenir. Bazı durumlarda, metnin içerisinde özel bazı bilgilerin metin madenciliğine konu olması mümkündür[14]. Çeşitli amaçlarla metnin içerisinde sınıflı bilgilerin çıkarılması istenebilir. Örneğin bu tez çalışmamda, alınan ilaç yorumları içerisindeki yan etkileri çıkararak sıralamak (event ordering) üzerine çalışmış, İngilizce de ki kelime yapılarını, durumu belirten kelime içerikleri ve bütün bu kelime kökleri arasındaki olası ilişkileri gösteren özel bir metod kurgulanmıştır.

Metin madenciliği çalışmalarında duygu analizi yöntemi kullanılmaktadır. Metinler içinde geçmekte olan duygusal ifadeleri çıkarmayı amaçlar. En çok kullanılan duygusal kutupsallıktır. Ayrıca bu işleme göre bir konu içinde geçen yazımların olumsuz veya olumlu olmasına göre de iki sınıfa ayrılmayı hedeflemiştir. Bu durum metin içinde geçen kelimelerin duygu araştırma konusundaki çıkarımlara faydalı sonuçlar ortaya koymaktadır.

2.5.3. Metin sınıflandırma

Metin sınıflandırmada öncelikli olarak her bir kaydın önceden belirlenmiş bir sınıfta olması amaçlanmaktadır. Sınıf olmak için her bir kaydın belirli ortak özellikleri olması gerekir. Ortak özelliklere sahip olan bu verilerin hangi özellikleriyle bu sınıfa girdiğini tanımlayan algoritma, sınıflandırma algoritmasıdır. Sınıflandırma algoritması denetimli öğrenme kategorisine giren bir öğrenme biçimidir.

III. VERİ MADENCİLİĞİ İÇİN SOSYAL MEDYA

ÜZERİNDEN NESNE TABANLI BİR YAZILIM

ARACI

Sosyal ağlarda bulunan verilerin analiz çalışmaları yaklaşımlara göre iki ayrı durumda araştırılabilir. Bağlantı tabanlı ve yapısal analizler. Kullanılması amaçlanan bu analiz yöntemlerinin sosyal ağlarda bulunan bağlantıları işleyerek bir veri seti oluşturmak istenmiştir. İşlem düğümleri arasında yorumlanan, değerlendirme veya kullanıcı bilgileri sınıflara ayrılarak çıkarılması hedeflenmektedir. Ağdaki hareketliliğe bağlı olarak kullanılan sabit analiz veya hareketli (dynamic) analiz yöntemi seçilebilir. Örneğin tezlerin referansları takip edildiği bir ağda sabit analiz yapılırken, Sosyal ağda verinin sürekli arttığı bir ağda hareketli analiz yöntemine ihtiyaç olmaktadır.

Bu çalışmada, sosyal medya da duygu analizi çalışması doğal dil işleme yapılmıştır. Sosyal medya aracı olarak ilaç alanında ziyaretçiler tarafından en yaygın kullanılan sağlık web sitelerinden biri olan webMD sitesi seçilmiştir. Çünkü bu site, popüler, erişim kolaylığı, 22 yılı aşkın faaliyet geçmişi ve ayda ortalama 150 milyondan fazla ziyaretçisi vardır ve büyük miktarda ilaç yorumu içermektedir. WebMD forumu kullanıcılar tarafından hasta bilgileri barındırılan saygın bir sağlık sitesidir. Avantajlarının yanında dezavantajları da vardır. Örneğin ilaçların yorum ve değerlendirme listesinin tek sayfada olmaması birden fazla sayfayı ziyaret etmeyi gerektirdiğinden dolayı sayfa sayfa internet sitesinde gezinme problemini ortaya çıkarmaktadır.

3.1. Giriş

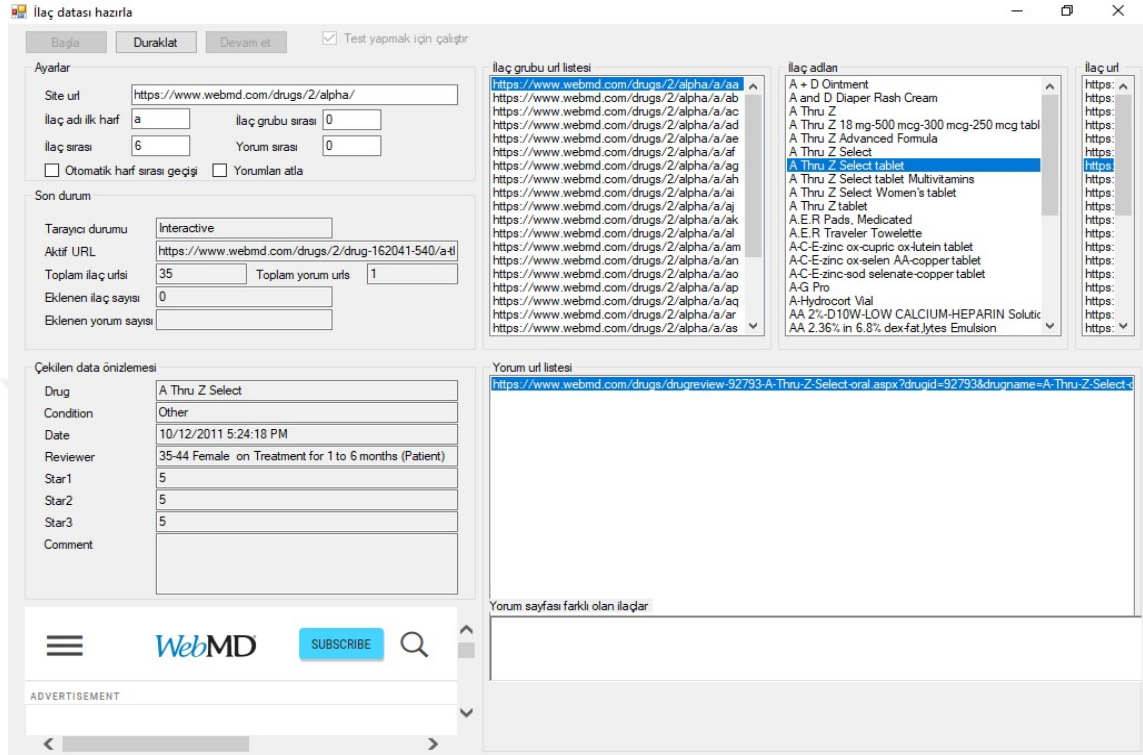
Hasta tedavilerin tıbbi satış sürecinden önce klinik çalışmalarda etkinlik değerlendirmesi yapılmaktadır ve bu çalışmalar genellikle ilaç şirketleri tarafından yapılır ama laboratuvar ortamında kısa sürede etkin sonuçlar almak çok maliyetli ve zordur. İnternet kullanımının yaygınlaşması ile ilaç kullanan kişiler sosyal medya forum siteleri üzerinden günden güne kullanım bilgilerine dayalı yorumlarını paylaşmakta olup bu durum veri madenciliği araştırma konusuna olan ilginin artmasına neden olmuştur. Bu çalışmada ilaç kullanan kişilerin webMD üzerinden yapmış olduğu yorumlar da veri madenciliği ve biomedikal metin madenciliği araçlarını kullanarak ilaç yan etkilerinin kullanımını konusunda veri analizi yapmayı hedeflemiştir.

Hastaların kullandıkları ilaçlar hakkında kullanım süresi, yaş aralığı, kullanıcı tipi, cinsiyet ve kullanım nedeni gibi bilgilerin paylaşıldığı, 22 yılı aşkın süredir faaliyet gösteren ve Amerika da bir kuruluş olan www.webmd.com web sitesinde ki ilaçlar hakkın da yapılan yorumlar analiz edilerek elde edilen sonuçlar hastalara, doktorlara yol gösterecektir.

3.2. Yazılım Aracının Yapısı ve Tasarımı

Sosyal medya da yaygın olan webMD internet sitesinde yapısal olarak denetimsiz halde bulunan ilaçlar hakkında yapılan yorumlar, C#.Net dilinde geliştirilen ve Şekil 3.1.' de ara yüzü bulunan "ilaç verisi hazırla" uygulaması ile verilerin okunması sağlanmıştır. Ayrıca Ek-A' da program da kullanılan html veri okuma ve parçalanmasını içeren metotlar, Ek-B'

de kullanım bilgilerinin niteliklerinin ayrıştırılmasına ait kodlar, Ek-C' de ise bu işlemlerin kurgulandığı veri tabanı modeli gösterilmiştir.



Şekil 3.1. Html Sayfası Okuyucu Programın Ekran Görüntüsü

Bu uygulamanın içerisinde bulunan web tarayıcı nesnesi ile ilgili internet sayfası html formatta okunarak Tablo 3.1.' de belirtilen yapıda toplamda dokuz adet nitelik olacak şekilde kural tabanlı metin parçalama yöntemleri uygulanarak veri tabanına kayıt işlemi yapılmış ve toplam 221.359 adet kayıt bulunan veri seti oluşturulmuştur.

WebMD' de ilaç kullanımı örnekleri Şekil 3.2.' de belirtilmiş olup bir hastanın depresyon tedavisinde ek ilaçlar tedavisinde Abilify oral kullanıldığında kendisi için kötü sonuçlar doğurduğunu ve kullanım sonrası yaşadığı yan etkiyi 2 yıl sonra bile yaşamak da olduğunu bu ilgili format da göstermiştir.

Tablo 3.1. Veri Seti Yapısı

Nitelik Adı	Nitelik Açıklaması	Nitelik Tipi	Örnek Veri
Comment ID	Kayıt No	integer	1
Drug Name	İlaç Adı	text	Abilify oral
Evaluation Date	Değerlendirme Tarihi	date	10/15/2017 3:13:44 PM
Condition	Kullanım Nedeni	text	Additional Medications to Treat Depression
Reviewer	Kullanıcı Bilgileri	text	65-74 Female on Treatment for 1 to 6 months (Patient)
Effectiveness	Etki Puanı	integer	2
Ease of Use	Kullanım Kolaylığı Puanı	integer	3
Satisfaction	Memnuniyet Puanı	integer	1
Comment	Yorum	text	Terrible made me clench my teeth and rock back & forth. Two yrs later I still do it. Terrible side effects for me. not recommended by me.

Hastalık ve tedavi bilgilerini içeren forumlar ile hastalar ve sağlık hizmeti sağlayıcıları arasında paylaşım kaynak halindedir. Onun içinde temel demografiyi doldurduktan sonra hasta forumu değerlendirmelerine izin verilir. Bu durum sistematik verilerde bilgi yürütmek için çok arzu edilen bir çalışma alanı oluşturmaktadır.

Condition: Additional Medications to Treat Depression 10/15/2017 3:13:44 PM

Reviewer: 65-74 Female on Treatment for 1 to 6 months (Patient)

Effectiveness ★ ★ ☆ ☆ ☆

Ease of Use ★ ★ ★ ☆ ☆

Satisfaction ★ ☆ ☆ ☆ ☆

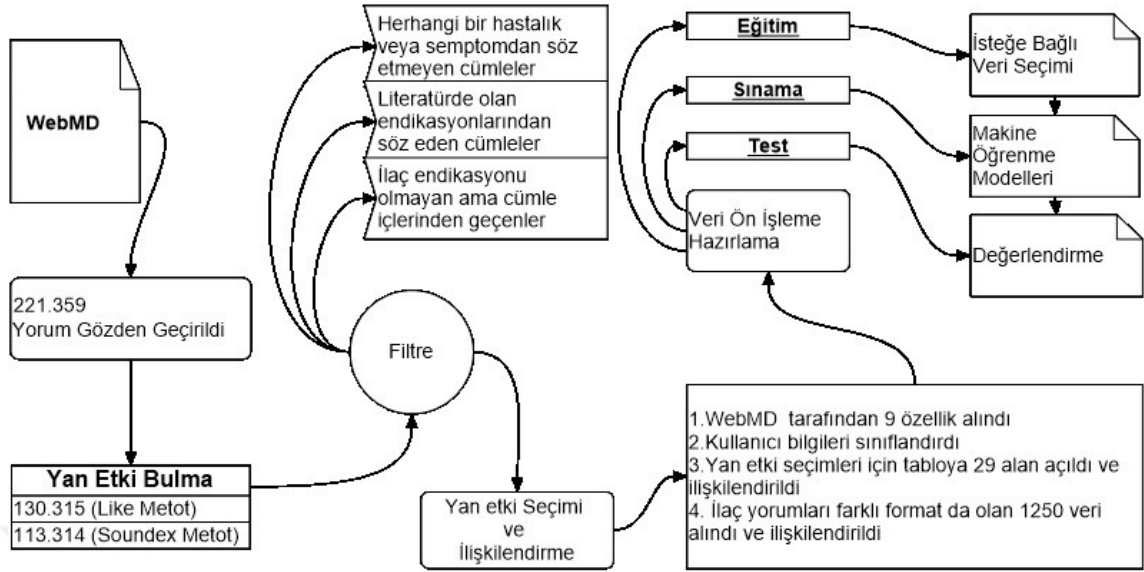
Comment:
 Terrible made me clench my teeth and rock back & forth. Two yrs later I still do it.
 Terrible side effects for me. not recommended by me.

Şekil 3.2. İlaç Kullanımı Hakkında Yapılmış Yorum Örneği.

IV. METİN MADENCİLİĞİ İŞLEMLERİ VE SONUÇLARI İÇİN WEB TABANLI BİR YAZILIM ÜZERİNDEN YORUMLARININ ANALİZİ

Makine öğrenimi için istenen veri formatı son derece önemlidir. Veri setinin sınıflandırılmış ve standardı belli olması kullanılan algoritmalar için elverişli bir zemin kazandırmaktadır ama her zaman formatı ve sınıfı belli veri elde etmek zordur.

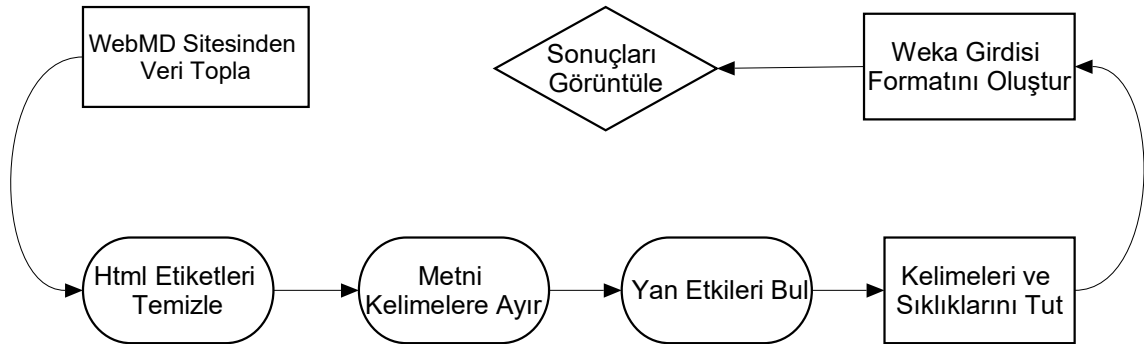
Bu çalışmada, istatistiksel veriler elde etmek için standart veri madenciliği araçları kullanılarak ilaç yorumlarından tespit edilen yan etki çıkarımlarını ve tekrar durumlarını ortaya çıkarılması hedeflenmiştir. Bu aşamada veri toplanma sürecinden izlenen yol WebMD tarafından yayınlanan incelemeleri toplayarak başlanılmış, hastalar ve sağlık hizmeti sağlayıcıları tarafından tedavi bilgilerini içeren yorum verilerini alınmıştır. Veri filtreleme, anlamlandırma, özellik yapımı ve seçimi, veri ön işleme, makine öğrenimi için model eğitimi ve değerlendirmesi için Şekil 4.1.' de aşamaları tanımlı olan iş akışı izlenmiştir. Temel demografik bilgiler doldurduktan sonra hasta forumu cinsiyet ve yaş grubu, kullanıcılar dâhil olmak üzere alınan bilgiler ilaçların etkinlik açısından değerlendirmelerine izin verir, kullanım kolaylığı ve genel memnuniyet ilaç deneyimleri hakkında verilmiş puanlar analiz sürecine dâhil edilme olanağı vermiştir. Birçok sosyal medya platformundan üstün olan webMD' den 1 Ocak 2008 ile 9 Eylül 2017 tarihleri arasında 221.329 adet kullanıcı yorumu alınmış ve her yorum metni doğal dil işleme için metin madenciliği analizi ile değerlendirilmiştir.



Şekil 4.1. Hasta Forumu Verilerinde İlaç Kullanımlarını Tanımlamak İçin Bir İş Akışı

4.1. İçerik Oluşturma

Toplanan 221.359 adet ilaç yorumun içerikleri veri olarak kullanılmıştır. Metin içerikleri doğal dil işleme süreçlerinden geçirilerek kelime köklerinden ayrılmış her bir kelimenin karşılığı değerlendirmeye alınmıştır Şekil 4.2.’ de belirtildiği gibidir. Bu projede ilaç yan etkilerinin yorumlar içerisinde tespit edilmesi hedeflemiştir. Kullanılan ilaç, durum ve demografik bilgiler alınarak veri madenciliği araçları ile analiz de kullanılmıştır.



Şekil 4.2. Uygulama İş Akışı

4.2. Hasta Yorumlarının İşlem Süreçleri

Hastaların kullandıkları ilaçlar hakkında, kullanım nedeni, görülen yan etkiler, yaş aralığı, kullanım süresi ve cinsiyet gibi bilgilerin ilaçlarda bilinen ama ortaya çıkmayan gizli özelliklerini ortaya çıkmasında çok önemli veri kaynağı olmuştur. Süreçleri aşağıda listelenmiştir;

- **Veri Toplama (Data Collection):** İlaç yan etkilerini tespit etmek için WebMD sitesinden hasta yorumları toplanılmıştır.
- **Veri Kümesi Kullanımı (Dataset Usages) :** İnternet ortamından alınan düzensiz veriler sınıflandırılarak yapısal hale getirilmiştir.
- **Veri Filtreleme (Data Filtering) :** Metinlerde geçen her kelime işleme alınmış ve anlam olarak uygun olup olmadığı kategorik olarak alınmıştır.
- **Özellik Seçimi (Feature Selection) :** Metinlerde geçen ifadelerin ilaç yan etkisi içermesi durumunda adet ve tekrar sayısı kayıt altına alındı.
- **Veri Ön İşleme (Data Preprocessing) :** Yapılandırılmış ve sınıflandırılmış veri setleri ilişkilendirilmiş bütünlüğü korunmuş ilgili format da analize sunulması için hazır hale getirilmiştir.
- **Makine Öğrenim Modelleri (Machine Learning Models) :** Veri setine uygun kullanılan makine öğrenim algoritmaları veri bütünlüğü içinde gizli kalmış bilgileri ortaya çıkarmış ve analizde değerlendirmeye sunulmuştur.
- **Değerlendirme (Evaluation) :** Hasta yorumlarından çıkan yan etki analiz sonuçları hasta bilgisine, ilaç şirketlerine ve ilgi duyulan araştırmalara kaynak olarak değerlendirilmiştir.

4.3. Veri Hazırlama Süreci

Veri hazırlama sürecinde veri temizleme, birleştirme, dönüştürme, indirgeme ve bu teknikler ile ilişkili çeşitli algoritmalar kullanılarak aşağıdaki işlemler yapılmıştır.

4.3.1. Kullanıcı, durum ve değerlendirme verilerin temizlenmesi

WebMD üzerinden alınan ilaç yorumu tanımlı standartlarda yer alacak bilgilere erişmek için temizleme yöntemi kullanılarak Tablo 4.1.' de olacak şekilde kural tabanlı veri temizleme işlemi yapılmıştır. Durum için 221.358 adet kayıt gerçekleştirilmiştir, kullanıcı için 217.669 adet kayıt gerçekleştirilmiştir, Değerlendirmeler için 219.303 adet kayıt gerçekleştirilmiştir[15].

Tablo 4.1. İşlenen Veri Yapısı

Veri Türü	Siteden Alınan Veri	İşlenen Veri
Condition	Condition: Other	Other
Reviewer	Reviewer: LTCProf, 65-74 Female on Treatment for 6 months to less than 1 year (Patient)	LTCProf, 65-74 Female on Treatment for 6 months to less than 1 year (Patient)
Effectiveness	Current Rating: 5	5
Ease of Use	Current Rating: 4	4
Satisfaction	Current Rating: 4	4

4.3.2. Ham veri ve yorumlardan çıkan yan etkilerin birleştirilmesi

İlaç yorumlarından analiz verisi elde edebilmek için tanımlı standartlar da yer alacak bilgilere erişmek için birleştirme yöntemi kullanılarak Tablo 4.2.' de WebMD den alınarak sınıflandırılan veriyi Tablo 4.3.' de ki yorumlardan çıkarılan ilaç yan etkileri ile kural tabanlı veri birleştirme işlemi yapılmıştır. Bu durum için 119.993 adet kayıt için gerçekleştirilmiştir[15].

Tablo 4.2. Örnek Ham Veri Yapısı

ID	İlaç Adı	Durum	Cinsiyet	Kullanım Süresi	Yaş Aralığı	Kullanıcı Tipi
1	Jakafi	Myelofibrosis		6 months to less than 1 year	35-44	Patient
2	Jakafi	Myelofibrosis		6 months to less than 1 year	75 or over	Patient
3	Jakafi	Myelofibrosis	Female	less than 1 month	75 or over	Caregiver
4	Jakafi	Myelofibrosis	Male	6 months to less than 1 year	55-64	Patient
5	Jakafi	Myelofibrosis	Female		55-64	Patient
6	Jakafi	Myelofibrosis	Female	1 to 6 months	45-54	Patient

Tablo 4.3. Örnek Yan Etki Yapısı

ID	İlaç Adı	Yan Etki 1	Yan Etki 2	Yan Etki 3	Yan Etki 4
1	Jakafi	Myelofibrosis	Pain		
2	Jakafi	Haemoglobin	Myelofibrosis		
3	Jakafi	Anemic	Haemoglobin	Myelofibrosis	Pain
4	Jakafi	Dizziness	Myelofibrosis	Night sweats	
5	Jakafi	Infection			
6	Jakafi	Myelofibrosis			

4.3.3. İlaç yan etki tespitlerinin dönüştürme işlemleri

Yazılan yorumları daha anlaşılır ve standart bir yapıya getirmek amacı ile Tablo 4.4.' de ki gibi bulunan yan etki niteliğindeki verilerde dönüştürme işlemi yapılmıştır. Sonuç olarak Tablo 4.4.' de Giriş Yan Etki verisi dönüşüm sonrası Dönüştürülmüş Yan Etkiler verisi olacak şekilde kullanıma alınmış ve standardizasyon sağlanmıştır. Sedrax web panel de tüm dönüşüm işlemleri yorum analiz (Soundex) menüsünde yan etki adetleri seçim yapıldığın listeleniyor. Tablo 4.5.' de yan etki frekans hesaplamaları gösterilmiştir[15].

Tablo 4.4. Örnek Standardizasyon Parametreleri

Giriş Parametresi	Dönüştürülmüş Yan Etkiler
pain-	Pain
)pain	Pain
PAIN"	Pain
paine	Pain
paain	Pain
ppain	Pain
pa8in	Pain
paihn	Pain
pain*	Pain
,pain	Pain

Tablo 4.5. Örnek Yan Etki Frekans Tablosu

Yan Etki	Frekans (Adet)
pain.	9982
pain,	5597
pain!	310
pain)	84
pain;	82
pain?	80
Pain	58

4.3.4. İlaç yan etki tespitlerinin indirgeme işlemleri

İlaç yorumlarından çıkan yan etki verilerinde yüksek sayıda tekrar olduğu tespit edilmiştir.

Tablo 4.6.' da ki gibi tekrar eden örnek kayıtlarda veri indirgeme işlemi yapılmıştır.

İndirgeme işleminden sonra tekrar eden kayıtlar için (Yan Etki Adet) alanında frekans hesaplaması yapılmış ve frekansı en yüksek olan kayıt bâz alınarak tekrar eden 91.276 kayıt verileri güncellenmiştir. Tablo 4.7.' de frekans hesaplamaları gösterilmiştir[15].

Tablo 4.6. Örnek İlaç Yan Etki Çıkarım Tablosu

Yorum ID	İlaç Adı	Yan Etki Adı
31	Yasmin 28	Anxiety
31	Yasmin 28	Anxiety
31	Yasmin 28	Anxiety
31	Yasmin 28	Anxiety
32	Yasmin 28	Bladder pain
33	Yasmin 28	Fear
34	Yasmin 28	Nightmare
35	Yasmin 28	Pain
36	Yasmin 28	Panic attack
37	Yasmin 28	Rage
37	Yasmin 28	Rage

Tablo 4.7. Örnek İndirgenmiş İlaç Yan Etki Çıkarım Tablosu

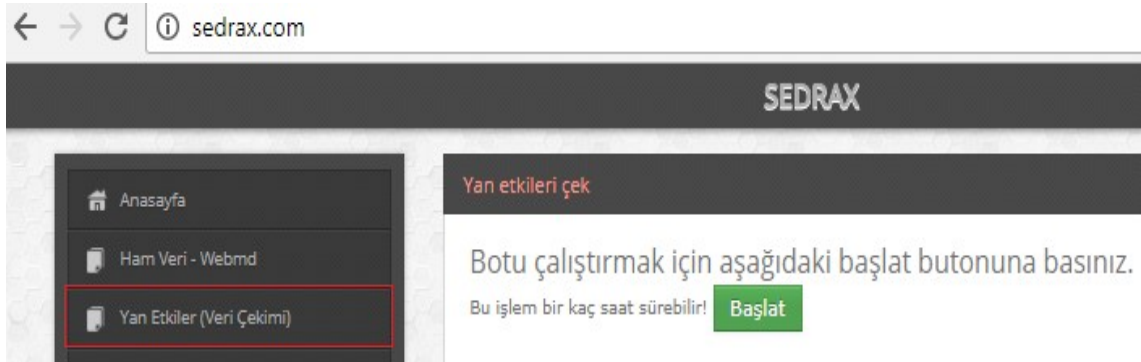
Yorum ID	İlaç Adı	Yan Etki Adı	Yan Etki Adet
31	Yasmin 28	Anxiety	4
37	Yasmin 28	Rage	2
32	Yasmin 28	Bladder pain	1
33	Yasmin 28	Fear	1
34	Yasmin 28	Nightmare	1
35	Yasmin 28	Pain	1
36	Yasmin 28	Panic attack	1

4.4. Yan Etki İsim Listesinin Oluşturulması

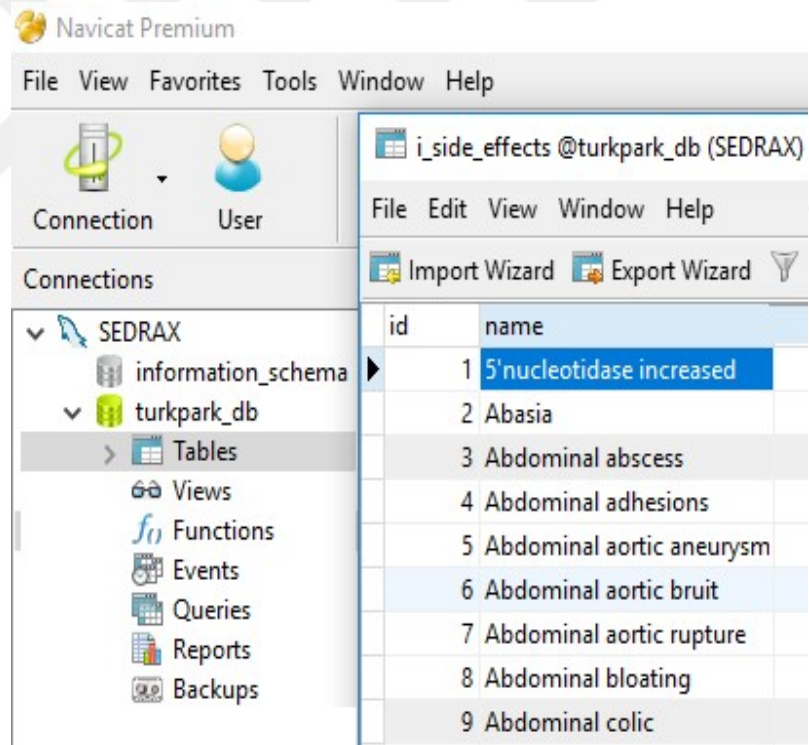
Yan etki isimlerinin listesi, Avrupa Moleküler Biyoloji Laboratuvarının bünyesinde çalışmalarını yürüten Sider4 (Side Effect Resource – Yan Etki Kaynakları) araştırma grubunun yayınlanmış olduğu web sitesinden alınmıştır[16]. İlaç yan etkileri araştırma grubu, Yan etkiler veritabanının oluşturulmasına yönelik çalışmalar yapan bir araştırma grubudur. Yan etkileri çekmek için web panel üzerinde Yan Etkiler (Veri Çekimi) menüsünde Şekil 4.3.’ de ki gibi Php Curl kütüphanesi ile hazırlanmış sayfa kaynağını alacak fonksiyon işlem butonu ile yan etki listesi alınır. Sider4 web sitesinden çekildikten sonra yan etki ayrıştırma işlemi yapılmış olup Mysql veritabanı aracılığıyla Şekil 4.4.’ de ki gibi i_side_effects tablosuna kayıt edilmiştir[15]. Tablo 4.8.’ de örnek yan etki listesi gösterilmiştir.

Tablo 4.8. Örnek Yan Etki Listesi

ID	Yan Etki Adı
1	Abdominal distension (Karın şişliği)
2	Abscess (Apse)
3	Acute abdomen (Akut karın)
4	Felt faint (Baygınlık hissetmek)
5	Aggression (Saldırganlık)
6	Agitation (Çalkalama)
7	Pain (Ağrı)
8	Infection Enfeksiyon
9	Nausea (Mide bulantısı)
10	Headache (Baş ağrısı)
11	Fatigue (Yorgunluk)
12	Dizziness (Baş dönmesi)
13	Depression (Depresyon)
14	Anxiety (Kaygı)
15	Insomnia (Uykusuzluk hastalığı)



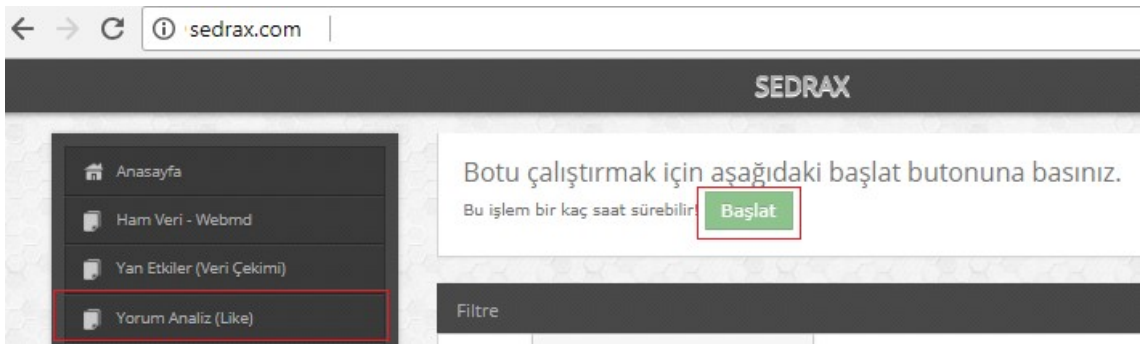
Şekil 4.3. Web Panel Yan Etki Çekim Robotu



Şekil 4.4. Mysql Yan Etki Tablosu

4.5. Yorumlardan Like Metodu ile Yan Etki Çıkarım İşlemleri

İlaç yorumlarından yan etki bilgisini çıkarmak için metin içinde kelime anlamı olarak benzeyenleri tespit etmek bulmak gerekir. Bu çalışmada, yan etki çıkarımları sağlamak için Like metodu kullanılmıştır. Yan etkileri bulmak için web panel üzerinde Yorum Analiz (Like) menüsünde Şekil 4.5.' de ki gibi Php yazılım dilinde ile hazırlanmış wep uygulaması Like fonksiyonu işlem butonu ile yan etkileri yorumlardan tespit edilmesi sağlanmıştır[15]. Sql tarafından da kullanımı içeren bu metot kelimenin başında, sonunda ve içinde olması durumunda benzerliği bulabilmektedir. Like komutu ilaç yorumlarında geçen kelimelerin benzerliğini vermektedir. Bu sayede Sider4 üzerinden indirilen ilaç yan etkileri sırasıyla kelime benzerliği web panel üzerinde yazılan Sedrax programı sayesinde karşılaştırarak çıkarmakta ve Şekil 4.6.' da ki gibi Mysql veritabanına i_comment_relations tablosuna yorum id, ilaç id, yan etki id, yan etki toplam, bilgileri ile kayıt altına almıştır. Toplamda 221.359 yorum bu işlemde geçirilmiş olup 130.315 yorumda yan etki tespit edilmiş ve 308.657 adet yan etki bulunmuştur.



Şekil 4.5. Web Panel Yan Etki Çekim Like Robotu

i_comment_relations @turkpark_db (SEDRAx) - Table

i_comment_relations @t... x

File Edit View Window Help

Import Wizard Export Wizard Filter Wizard Grid

id	comment_id	drug_id	se_id	se_count
1	18	4	2	1
2	20	4	87	2
3	20	4	4765	2
4	21	4	2468	1
5	21	4	4765	1
6	21	4	6445	1
7	23	4	455	1
8	23	4	660	1
9	23	4	4134	4

Şekil 4.6. Örnek Like ile Yorumlardan Çıkan Yan Etki Mysql Tablosu

İlaç yorumlarından Like metodu ile çıkan yan etkiler kullanıcıların bilgisine sunulması için web panel üzerinde listesi yapılmış ve Şekil 4.7.' de gösterilmiştir. Kullanıcı bu listede sıralama yapabilir, istediği ilacın yan etki bilgisini filtreleyebilir ve kaç adet yan etki olduğunu görebilir[15].

Filtre

İlaç Seçiniz Ekle

İlaç	ID	Sil
Jakafi	21	Sil

Listele

Soundex İle Yan Etki Çıkarımı

Show 10 entries Search:

ID	İlaç Adı	Yanetki Adı	Adet
860	Jakafi	Myelofibrosis	6

Şekil 4.7. Örnek Like ile Yorumlardan Çıkan Yan Etki Listesi

4.6. Yorumlardan Soundex Metodu ile Yan Etki Çıkarım İşlemleri

Soundex algoritması, İngilizce kelimelerin söyleniş biçimlerine göre hazırlanmış bir fonetik algoritmadır. Bu algoritmanın kurgulanmasının genel amacı; telaffuzları benzeyen kelimelerin kurgulanmış algoritması ile aynı karakter katarına çevrilmeleri sonucunda eşleştirme yapması ve benzeyen kelimelerin yazımında değişiklik olsa bile ortaya çıkarılmasına olanak vermesidir. Ayrıca Soundex algoritması, fonetik algoritmalarının en yaygın ve en çok bilinendir. Algoritma yan etki değerinin ilaç yorumlarına yanlış girilen yan etkiler ile karşılaştırma yaparak bu yazım hatalarına karşılık gelen bilgileri almak için modellenmiştir. Bu çalışmada ilaç yorum içeriği kelime köklerinden ayrılarak Soundex algoritmasının işleminden geçirilmesi sonucunda yan etkiler ile benzerliği ortaya çıkarılmasını sağlayarak yorumlayıcının ilaç yan etki türünü yanlış yazımında analiz sürecine dahil edilmesi sağlanmıştır.

Tarihsel süreç de Soundex algoritması, Robert Russell ve Margaret Odell tarafından modellenmiştir, U.S. Patenti 1,261,167 ve U.S. Patenti 1,435,663 ile patentlenmiştir. Soundex kodlaması 1960'larda ortaya çıkarılmak için 'Communications of the ACM', 'Journal of the ACM of the Association for Computing Machinery' gibi bazı dergilere konu olması sebebiyle ünlenmiş ve özellikle Donald Knuth, 'The Art of Computer Programming' yazmış olduğu makalede nasıl bir yapıda olduğu anlatılmaktadır[17].

Soundex algoritma kodu ilki bir harf, diğer üçü ise rakamlardan oluşan dört karakterli bir koddur; harf olarak kelimenin ilk harfi, sayı olarak da geri kalan harflerin belirli bir kurala göre numaralandırılmasıyla ortaya çıkan sonuç alınır. Benzer telaffuzlu sözcükler ise aynı

numaralarla kodlanır; örneğin, benzer telaffuzlu B, F, P ve V harfleri 1 ile kodlanır. Sesli harfler kodlama işlemini etkiler, ancak bu sesli harf kelimenin başında bulunmuyorsa ortaya çıkan sonucu asla doğrudan etkilemez.

Tam algoritma aşağıdaki gibidir:

Karakter katarının ilk harfini kod başında bulundurur.

Eğer ilk harf "a, e, h, i, o, u, w, y" harflerinden herhangi biri değilse, bu harfleri metinden çıkarılır.

Sırasıyla tüm harflere aşağıdaki numaralandırmayı kelime grubuna uygular;

- b, f, p, v = 1
- c, g, j, k, q, s, x, z = 2
- d, t = 3
- l = 4
- m, n = 5
- r = 6

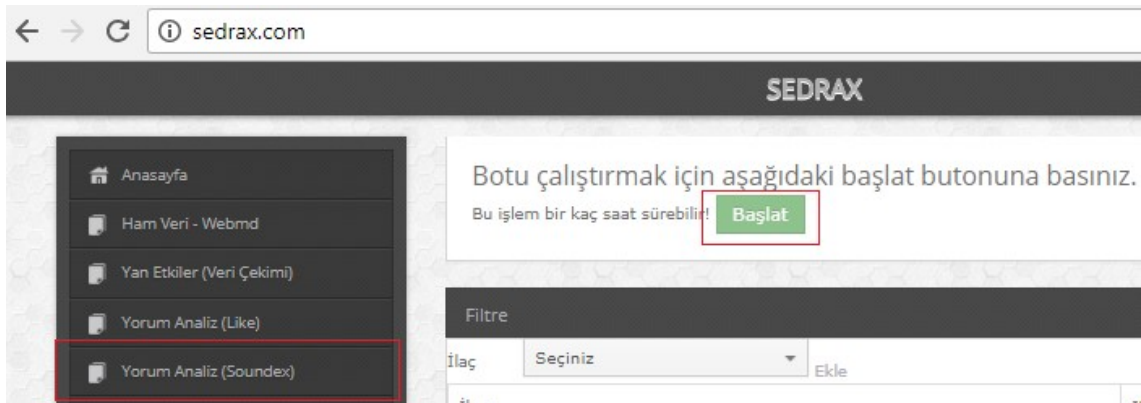
Eğer numaralandırmada aynı numarayı almış iki ya da daha fazla harf yan yanaysa (ilk işlemde önce) ya da bu harflerin arasında h veya w harfi varsa aynı olanları atlar.

İlk dört karakteri sonuç olarak gönderir; eğer sonuç dört karakterden az çıkıyorsa, dört karakter tamamlanması için sonuna sıfırlar eklenir. (örn. A22 → A220).

Algoritma kullanılırken; örneğin "Rubin" in kodu R150 iken, "Robert" ve "Rupert" isimleri aynı kodu döndürür; R163[17].

İlaç yorumlarından yan etki bilgisini çıkarmak için metin içinde kelimeler ile Sider4 üzerinden indirilen yan etki benzeyenleri tespit etmek bulmak gerekir.

Bu projede yan etki çıkarımları sağlamak için Soundex metodu da kullanılmıştır. Yan etkileri bulmak için web panel üzerinde Yorum Analiz (Soundex) menüsünde Şekil 4.8.' de ki gibi Php yazılım dilinde ile hazırlanmış wep uygulaması Soundex fonksiyonu işlem butonu ile yan etkileri yorumlardan tespit edilmesi sağlanmıştır[15]. İlaç yorumlarında geçen kelimelerin Soundex kod yapısı web panel üzerinde yazılan Sedrax programı sayesinde dönüştürülmüş ve yan etki benzerliğini bulmak için Sider4 üzerinden indirilen ilaç yan etkileri sırasıyla Soundex kod bloğuna ile karşılaştırılmıştır. Aynı kod benzerliği olan kelimeler Şekil 4.9.' da ki gibi Mysql veritabanına i_comment_relations_soundex tablosuna yorum id, ilaç id, yan etki id ve yan etkiye benzeyen kelime, bilgileri ile kayıt altına almıştır. Toplamda 221.359 yorum bu işlemde geçirilmiş olup 113.314 yorumda yan etki tespit edilmiş ve 254.449 adet yan etki bu metot ile de bulunmuştur. Çıkan yan etkiler kullanıcıların bilgisine sunulması için web panel üzerinde listesi yapılmıştır[15]. Şekil 4.10.' da gösterildiği gibidir.



Şekil 4.8. Web Panel Yan Etki Çekim Soundex Robotu

Untitled @turkpark_db (SEDRIX) - Query *

File Edit Format View Window Help

Run Stop Explain Export Wizard New Load

Query Builder Query Editor

```

1 SELECT comment_id,se_id,word,SOUNDEX(word)
2 FROM `i_comment_relations_soundex`;
3

```

Message	Result1	Profile	Status	
	comment_id	se_id	word	SOUNDEX(word)
	18	2	Abasia	A120
	20	87	acne	A250
	20	87	acne	A250
	20	4765	pain	P500
	23	455	anxiety,	A523
	23	4134	migraines	M2652

Şekil 4.9. Örnek Soudex ile Yorumlardan Çıkan Yan Etki Mysql Tablosu

SEDRIX

Botu çalıştırmak için aşağıdaki başlat butonuna basınız.
Bu işlem bir kaç saat sürebilir! [Başlat](#)

Filtre

İlaç Seçiniz [Ekle](#)

İlaç	ID	Sil

[Listele](#)

Soundex İle Yan Etki Çıkarımı

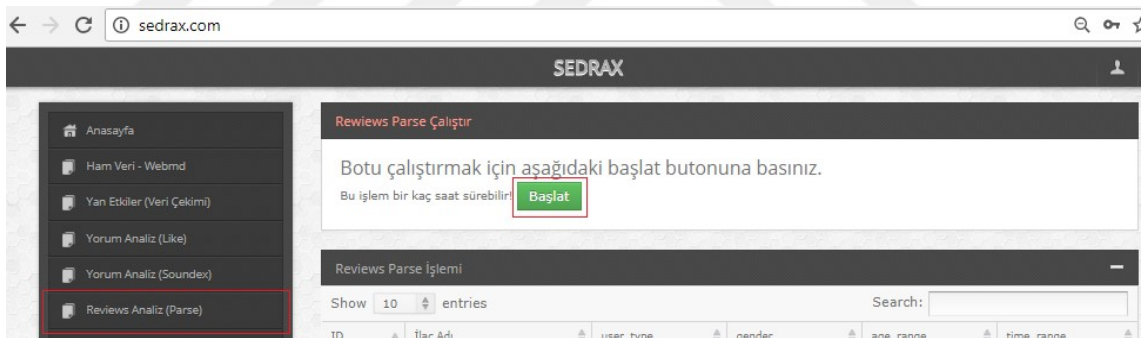
Show 10 entries Search:

ID	İlaç Adı	Yanetki Adı	Adet
27496	hydrocodone-acetaminophen	Pain	2709
27499	hydrocodone-acetaminophen	Surgery	299

Şekil 4.10. Örnek Soundex ile Yorumlardan Çıkan Yan Etki Listesi

4.7. Kullanıcı Bilgilerinin Nitelik Olarak Ayrılması

İlaç yorumların WebMD sitesinden aldıktan sonra kullanıcı bilgilerinin birleşik olarak geldiği görülmüş analiz aşamasında kullanmak için gelen bilgilerin sınıflandırılması gerektiğini için metin üzerinde string katar işlemesi yapılmış ve ayrıştırılmıştır. Ayrıştırmak için web panel üzerinde Reviews Analiz (Parse) menüsünde Şekil 4.11.' de ki gibi Php yazılım dilinde ile hazırlanmış web uygulaması Parse fonksiyonu işlem butonu ile kullanıcı bilgilerinin ayrıştırma işlemi yapılmıştır[15]. Örnek: `<p class="reviewerInfo">Reviewer: anom, 45-54 Male on Treatment for 1 to less than 2 years (Patient) </p>` html tag ile gelen kullanıcı bilgisi ayrıştırılarak sınıflandırılmış olup bu işlem 217.669 yorumda gerçekleşmiştir. Şekil 4.12.' de örnek gösterilmiş ayrıca Ek-B' de program klası gösterildi.



Şekil 4.11. Web Panel Üzerinden Kullanıcı Bilgilerini Ayrıştırma Robotu

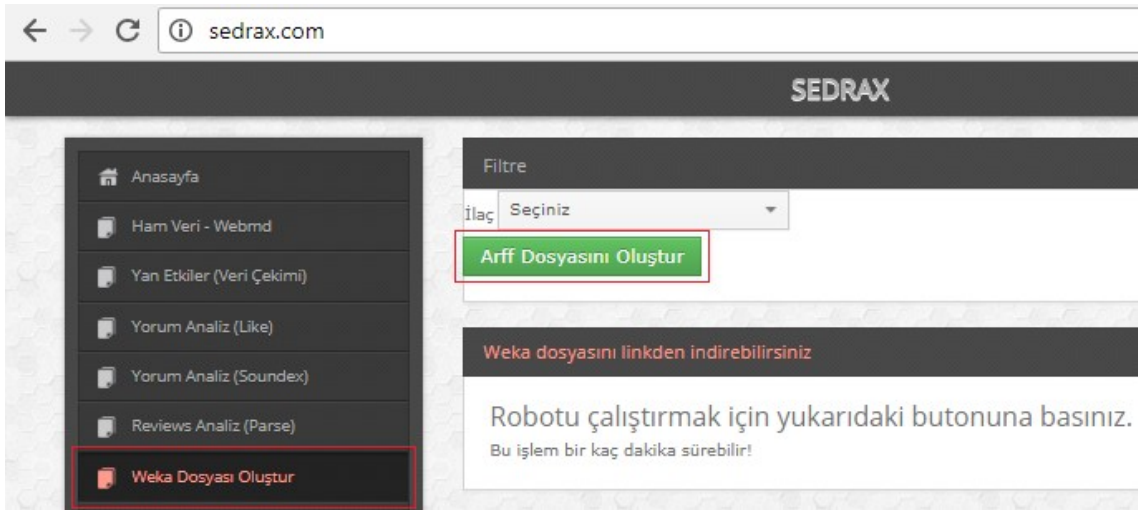
reviewer	screen_name	age_range	gender	time_range	user_type
ANOM, 45-54 Male on Treatment for 1 to less than 2 years (Patient)	ANOM,	45-54	Male	1 to less than 2 years	Patient

Şekil 4.12. Kullanıcı Bilgilerinin Ayrıştırıldığı Mysql Tablosu

4.8. Analiz Edilen Verinin Weka Dosyasına Aktarılma İşlemi

WEKA (Waikato Environment for Knowledge Analysis), Yeni Zelanda Waikato Üniversitesi'nde geliştirilen bir veri madenciliği ve makine öğrenmesi yazılımı olup Java programlama dili ile geliştirilmiştir. WEKA, birçok sınıflandırma tekniğini içermektedir. WEKA' da, Ön işleme, Sınıflama, Kümeleme, Birliktelik Kuralları, Nitelik Seçme ve Görselleştirme işlemleri gerçekleştirilebilir.

Bu çalışmada, ilaç yorumlarından tespit edilen yan etkileri kullanıcı demografik bilgilerine ekleyip analizde kullanılması için ön işlemden geçirilip Weka'nın kabul ettiği dosya tiplerinden Arff dosyalarının okunması sağlanmıştır. Şekil 4.14.'de Arff dosyasının oluşturulması gösterilmiştir. Weka'ya analiz verisini aktarmak için web panel üzerinde Weka Dosyası Oluştur menüsünde Şekil 4.13.'de ki gibi ilaç seçimi yapıldığında Şekil 4.15. ve Şekil 4.16. Mysql tablolarını kullanılarak Php yazma fonksiyonu Arff içeriğine uygun dosya oluşturmaktadır[15].



Şekil 4.13. Web Panel Üzerinden Kullanıcı Bilgilerini Ayrıştırma Robotu


```

← → ↻ ⓘ sedrax.com
@relation 'Jadenu'
@attribute condition {'Too Much Iron in the Blood for a Long Period of Time'}
@attribute effectiveness {'1E','2E','3E','4E','5E',''}
@attribute easeofuse {'1K','2K','3K','4K','5K',''}
@attribute satisfaction {'1M','2M','3M','4M','5M',''}
@attribute gender {'Female','Male',''}
@attribute time_range {'less than 1 month','1 to 6 months','6 months to less than 1 year','1 to less than 2 years','2 to less than 5 years','5 to less than 10 years','10 years or more',''}
@attribute age_range {'0-2','3-6','7-12','13-18','19-24','25-34','35-44','45-54','55-64','65-74','75 or over',''}
@attribute user_type {'Patient','Caregiver','consumer',''}
@attribute se_1 {'Rash','Anaemia',''}
@attribute se_2 {'Chronic iron overload',''}
@attribute se_3 {'Iron overload',''}
@attribute 'Class' {'Jadenu'}

@data
'Too Much Iron in the Blood for a Long Period of Time','3E','2K','1M','Female','less than 1 month','55-64','Patient','Rash','','','Jadenu'
'Too Much Iron in the Blood for a Long Period of Time','5E','5K','1M','','','1 to 6 months','65-74','Patient','Anaemia','Chronic iron overload','Iron overload','Jadenu'

```

Şekil 4.14. Web Panel Üzerinden Seçilen Örnek İlaçın Weka Dosyası

Untitled @turkpark_db (SEDRAx) - Query *

i_side_effects @turkpark... x Untitled @turkpark_db (... x

File Edit Format View Window Help

Run Stop Explain Export Wizard New Load Save S

Query Builder Query Editor

```
1 SELECT id, `name`,accept_method FROM `i_side_effects`;
```

Message Result1 Profile Status

id	name	accept_method
1	5'nucleotidase increased	lk
2	Abasia	sd

Şekil 4.15. Yan Etki Metot Mysql Tablosu

i_comment_rel @turkpark_db (SEDRAx) - Table

File Edit View Window Help

Import Wizard Export Wizard Filter Wizard Grid View Form View

id	comment_id	drug_id	se_id	se_count	accept_method
1	18	4	2	1	sd
2	20	4	87	2	sd
3	20	4	4765	1	sd
4	23	4	455	1	sd
5	23	4	4134	4	sd
6	29	4	87	1	sd

Şekil 4.16. Yan Etki Weka Çıkarım Mysql Tablosu

4.9. İlaç Analiz

İlaç yan etkilerinin araştırılmasında, çeşitli analiz yöntemleri kullanılmakta ve bunlar arasında frekans analiz yöntemleri tercih edilmektedir.

İlaç analizlerindeki hedef endikasyonların önemli ilişki bilgilerin sunulmasıdır. Örneğin; sürekli ya da bazen kullanılan ilaçların yan etkilerinin durumu ortaya konulduğunda kullanıcı bu duruma uygun hareket edebilir. Analizlerde ilaçların yan etki durumlarını bilmek son derecede önemlidir. İlaç kullanımına bağlı olarak hastaların belirli bir yan etki olayın görülme riskini artırdığı durumlarda, önlem alınacak bilgiye sahip olması tedavi için önem arz etmektedir.

Bu çalışmada, literatür taramalarında yayımlanan raporlar hakkında bir süredir süregelen ilaç kullanımına bağlı endişeleri gün yüzüne çıkarılmak için çalışılmıştır. Hâlihazırda ulaşılabılır olmayan ilaç kullanım bilgileri veri madenciliği araçları ile analiz edilmiş yan etki çıkarımları sağlanmış ve frekans değerleri kullanıcı bilgisine web ve mobil ortamda sunulmuştur.

4.9.1. Yan etkilerde ilaç dağılımı

Sider4 üzerinden alınan ve literatür de geçen ilaç yan etkileri WebMD internet sitesinden alınan hasta yorumları ile eşleştirilmiş çıkan sonuçlar ilaç dağılımı yapılarak veri tabanına kaydedilmiştir. Yan etki ilaç dağılım listesini web panel üzerinde Yan Etkiler – İlaç Dağılımı menüsünde Şekil 4.17.' de ki gibi gösterilmiştir[15].

ID	Yan Etki Adı	İlaç Sayısı
4765	Pain	2469
4356	Nausea	1332
2959	Headache	1104
3416	Infection	1035
2000	Dizziness	1029
6277	Surgery	975
6289	Swelling	968
5537	Rash	944
1936	Diarrhoea	915
1862	Depression	870

Şekil 4.17. Web Panel Üzerinden Yan Etkilerin İlaç Dağılım Görüntüsü

4.9.2. Veri istatistik

Neredeyse tüm çalışma alanlarında istatistiksel veri analizi büyük öneme sahiptir. Hemen her alanda söz konusu çalışmaların bilimsel birer araştırmaya dönüştürülebilmesi ancak ve ancak istatistiksel yaklaşımları kullanmakla mümkündür. WebMD internet sitesi üzerinden alınan ilaç yorum bilgileri veri durumu hakkında web panel üzerinde genel kullanılan verinin durumu hakkında istatistiksel bilgi sunulmuştur. Cinsiyet dağılımı, yaş dağılımı, kullanım süresi sınıflarına göre değerlendirilmiş olup en iyi ilaçlar veri sınıfından gelen etki, kullanım kolaylığı ve memnuniyet puanlarına “5” olma durumuna göre frekansı alınmıştır. En iyi tıbbi durumlar veri sınıfın da ki ilgili duruma istinaden yapılan yorumun en yüksek verilen puan durumlarına göre frekansı alınmış, en büyük olumsuz olaylar ilaç yorumlarından çıkan yan etkilerine verilen en düşük puanlarına göre frekansı alınmıştır ve Şekil 4.18.’ de gösterilmiştir[15].

Data Frekansı	
Cinsiyet Dağılımı	145402 (% 65,69) Female 59893 (% 27,06) Male 16064 (% 7,26) Bilinmeyen
Yaş Dağılımı	45-54 (48804) 55-64 (44549) 35-44 (33495) 25-34 (30971) 65-74 (23669) 19-24 (15532) Bilinmeyen (9890) 75 or over (8454) 13-18 (4239) 7-12 (1058) 3-6 (501) 0-2 (197)
Kullanım Süresi	less than 1 month (63976) 1 to 6 months (53019) 2 to less than 5 years (26567) 1 to less than 2 years (20943) 6 months to less than 1 year (20202) 5 to less than 10 years (13979) Bilinmeyen (13716) 10 years or more (8957)
En iyi ilaçlar Etki - Memnuniyet - Kullanım Kolaylığı	phentermine HCL (1010) hydrocodone-acetaminophen (919) lisinopril (606) lisinopril Solution (605) Adipex-P (563) Lexapro (528) trazodone HCL (454)

Şekil 4.18. Web Panel Üzerinden Veri İstatistik Görüntüsü

4.9.3. İlaçlardaki yan etki değerlerin frekansı

İlaç yorumlarından alınan yan etki bilgileri, yorumlayıcının ilaç kullanımına bağlı olarak yaşadığı yan etkisi de olabilir bu ilacı kullanmadan önce böyle bir yan etki durumunu yaşıyor da olabilir. Doğal dil işleme bilgi çıkarım metotları böyle bir ayırım için yetersiz kaldığından bu çalışmada çıkarılan yan etkiyi kullanıma bağlı olarak değerlendirmeye alınmıştır.

Bazı ilaçları birlikte almak kötü bir reaksiyona neden olabilir. Buna etkileşim denir. Örneğin, bir ilaç başka ilaçlar için sorun yaratabilecek yan etkilerin ortaya çıkmasına neden olabilir. Bu sebep göz önüne alınarak yapılan çalışmanın ilaç kullanımına bağlı olarak nasıl yan etki değişimi olduğunu ortaya konulmuştur. Sedrax web yazılımı

üzerinden kullanıcı seçtiği ilacı analiz için filtre alanına ekleyebilir birlikte kullandığı ilacın yan etki durumu için birden fazla seçim yapabilir ve kullanıcı duruna uygun veri filtresi uygulayıp sonuçları listeleyebilir. Şekil 4.19.' da ki gibidir.

İlaç yorumlarından çıkan yan etkiler Sedrax web panel üzerinde İlaç Frekansı menüsünde demografik bilgiler ve kullanım bilgileri ile sorgulama yaptığında ilaç yan etki bilgileri frekans hesaplamaları yapılmıştır. Şekil 4.20.' da gösterildiği gibidir[15].

The screenshot shows the SEDRAX web panel interface. The browser address bar displays 'sedrax.com'. The main content area is titled 'SED RAX' and features a 'Filtre' (Filter) section. The filter section contains several dropdown menus for selecting criteria: 'İlaç' (Drug), 'Yan Etki' (Side Effect), 'Kullanım Süresi' (Usage Duration), 'Yaş Aralığı' (Age Range), 'Cinsiyet' (Gender), 'Tip' (Type), 'Durum' (Status), 'Etki' (Effect), 'Kullanım Kolaylığı' (Ease of Use), and 'Memnuniyet' (Satisfaction). The 'İlaç' dropdown is currently set to 'Seçiniz' (Select). Below the filter section, a table displays the results of the filter. The table has four columns: 'İlaç' (Drug), 'Durum' (Status), 'Yorum Sayısı' (Number of Comments), and 'Sil' (Delete). The table contains two rows of data:

İlaç	Durum	Yorum Sayısı	Sil
Yasmin 28	Birth Control	168	Sil
Jakafi	Myelofibrosis	19	Sil

Şekil 4.19. Web Panel Üzerinden İlaç Frekans Filtre Görüntüsü

Yanetki Frekansı		
Yanetki	İlaç (Yorum Sayısı)	Tekrar Sayısı
+ Anxiety	1 (13)	18
+ Acne	1 (17)	18
+ Mood swings	1 (17)	18
+ Migraine	1 (11)	17
- Pain	2 (13)	15
	Yasmin 28 (11)	13
	Jakafi (2)	2

Şekil 4.20. Web Panel Üzerinden İlaç Frekans Görüntüsü

4.10. Grafik Gösterimler

Veri görselleştirmenin en önemli amaçlarından biri normalde klasik formatta sunulan karmaşık veriler yerine projemizde uyguladığımız kolay algılanabilir grafik ara yüzler ile rahatça anlaşılır hale getirmektir. Şekil 4.21.' de gösterilen grafikler kullanılmıştır[18][15].

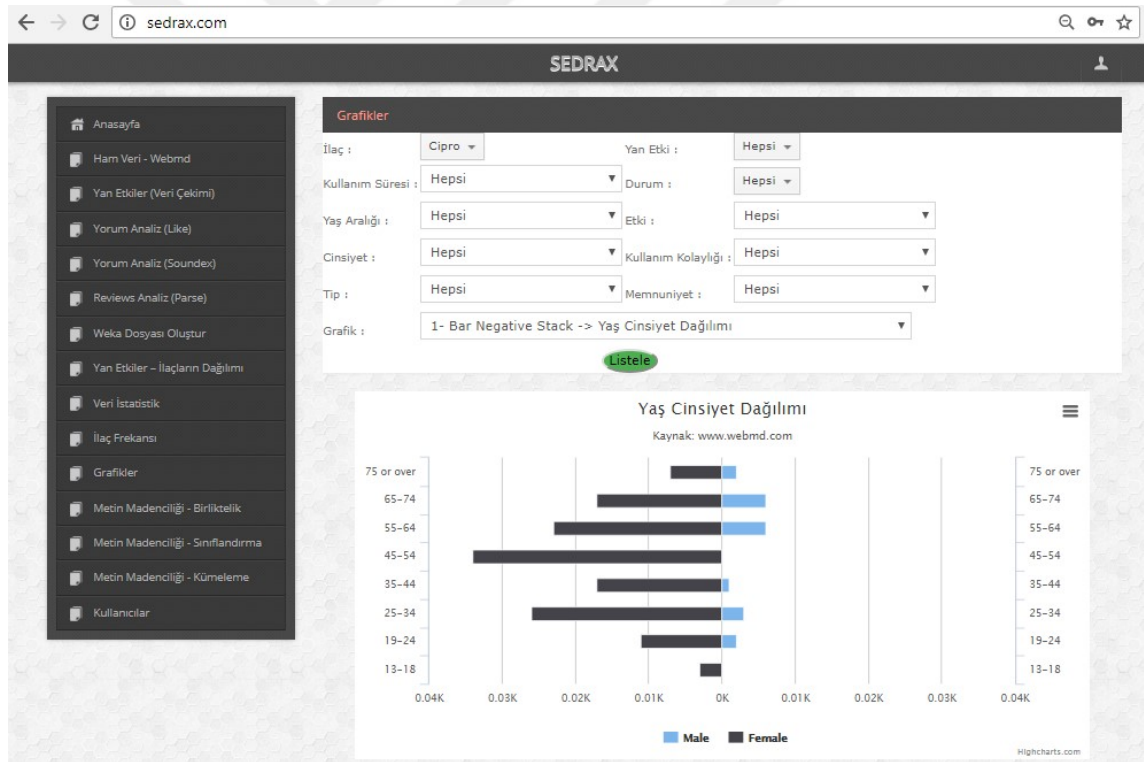


Şekil 4.21. Web Panel Üzerinden Grafik Seçimi

4.10.1. İlacın veri durum grafikleri

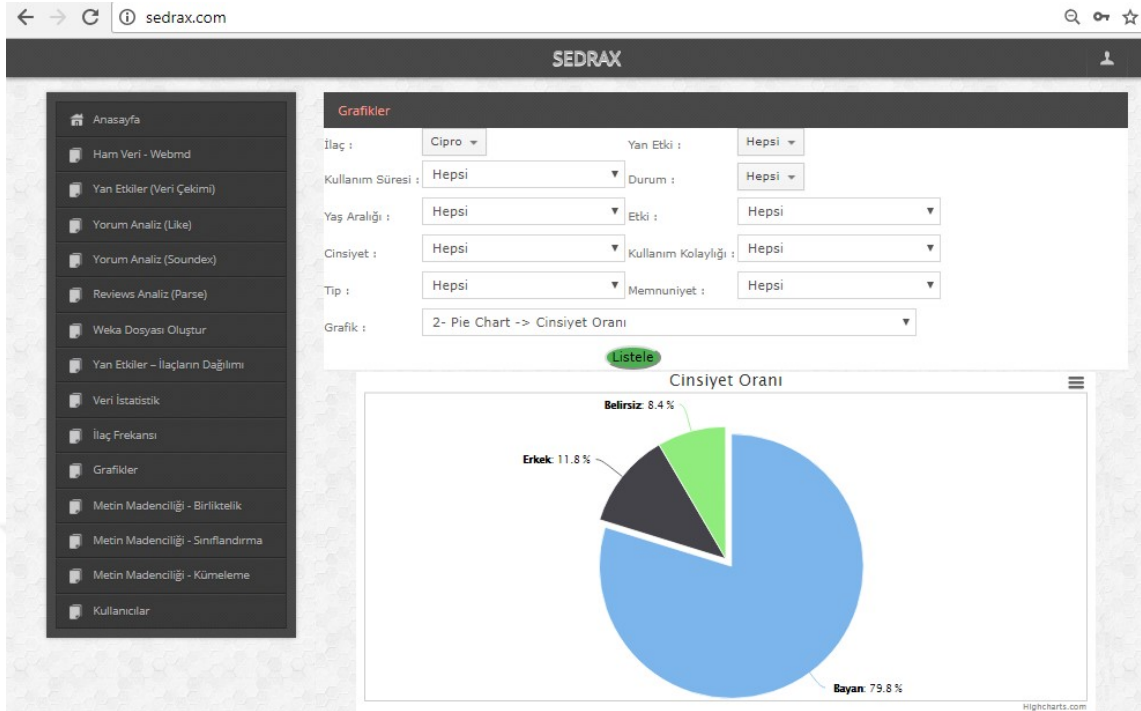
Veri dağılımımızın yapısını daha iyi tanımlamak için grafiklerden destek almak gerekir. İlaç kullanım bilgilerini analiz etmek için özetlenen bilgiler istenirse grafik olarak ta sunulması daha anlaşılır olmaktadır. Burada önemli bir husus veri tipimize göre grafik seçilmesidir. Veri tipimize nazaran seçmemiz gereken grafik türleri şöyledir;

Yaş cinsiyet dağılımı : Bar Negative Stack formatında verinin sorgulama ve filtreleme alanına dayalı olarak grafiksel gösterimi Şekil 4.22.' de ki gibidir[15].



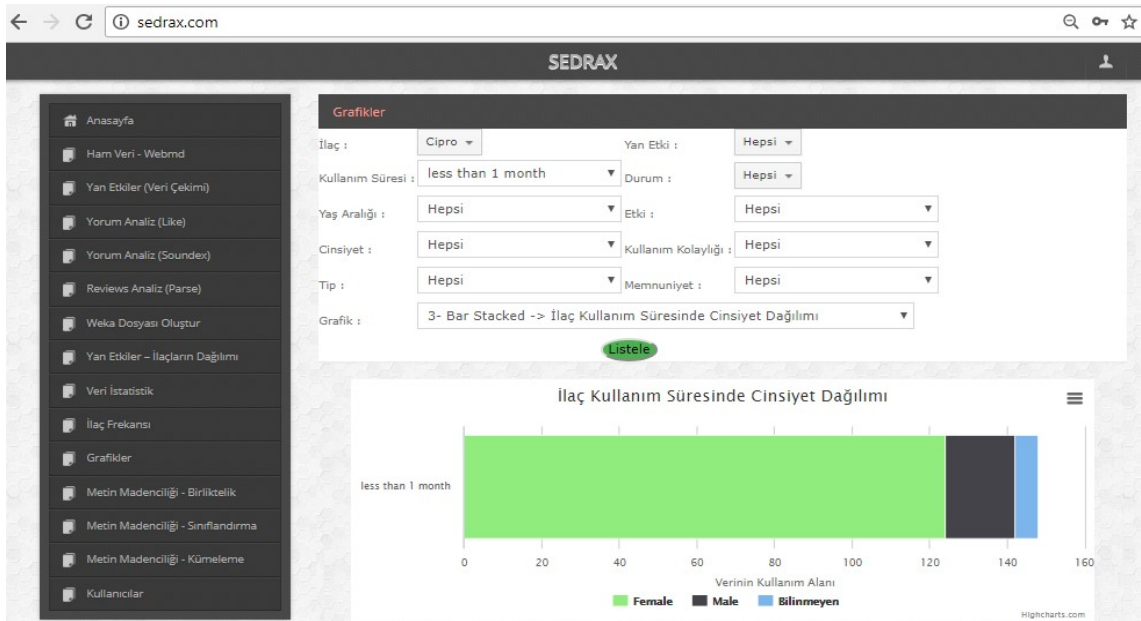
Şekil 4.22. Web Panel Üzerinden Yaş Cinsiyet Dağılım Grafiği

Cinsiyet oranı : Pie Chart formatında verinin sorgulama ve filtreleme alanına dayalı olarak grafiksel gösterimi Şekil 4.23.' de ki gibidir[15].



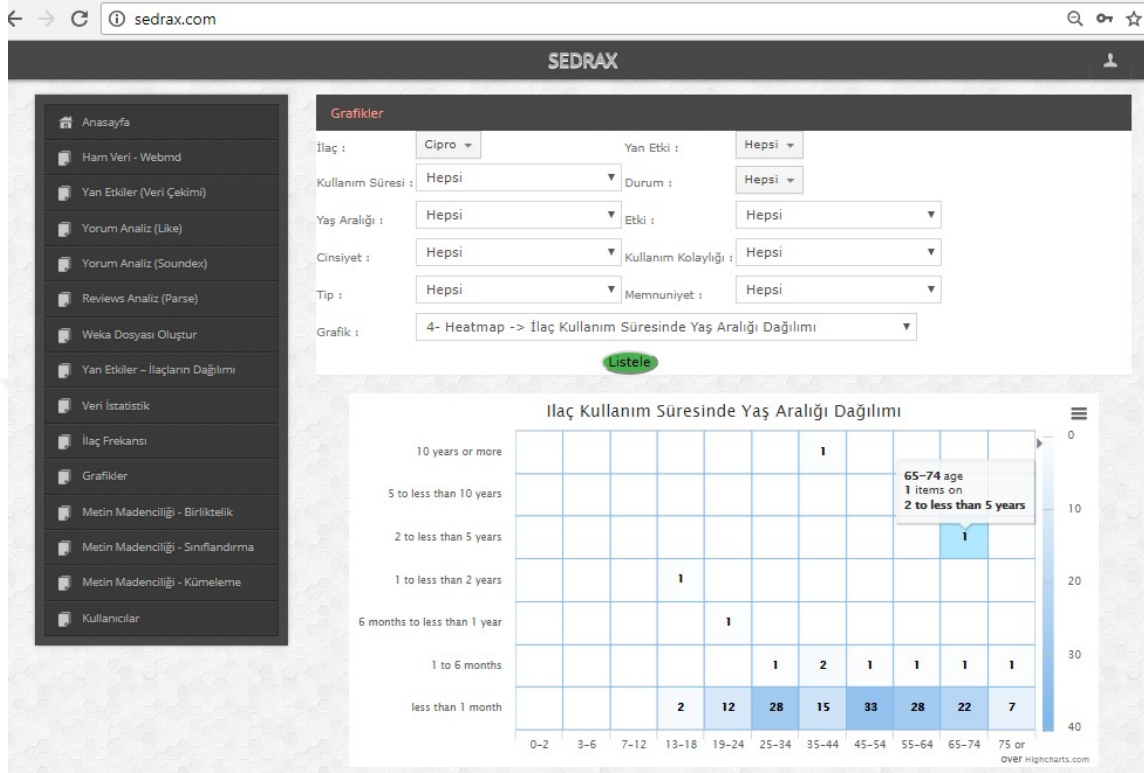
Şekil 4.23. Web Panel Üzerinden Cinsiyet Oran Grafiği

İlaç kullanım süresinde cinsiyet dağılımı : Bar Stacked formatında verinin sorgulama ve filtreleme alanına dayalı olarak grafiksel gösterimi Şekil 4.24.' de ki gibidir[15].



Şekil 4.24. Web Panel Üzerinden Kullanım Süresi Cinsiyet Dağılım Grafiği

İlaç kullanım süresinde yaş dağılımı : Heatmap formatında verinin sorgulama ve filtreleme alanına dayalı olarak grafiksel gösterimi Şekil 4.25.' de ki gibidir[15].



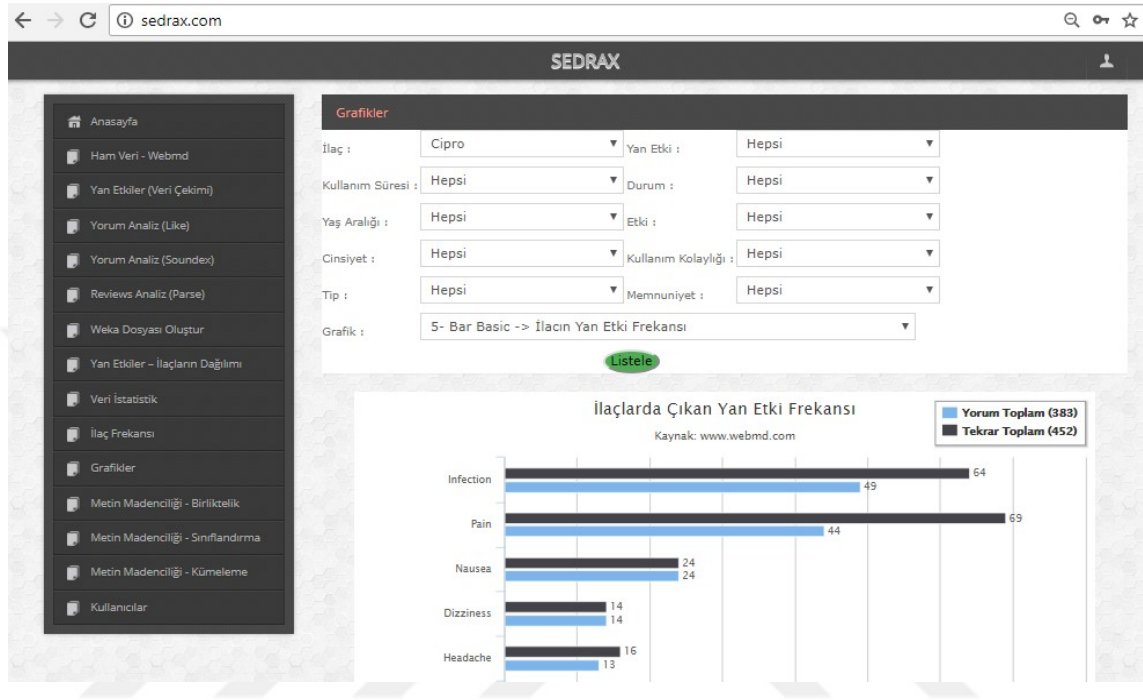
Şekil 4.25. Web Panel Üzerinden Kullanım Süresi Yaş Dağılım Grafiği

4.10.2 İlacın yan etki frekans grafiği

İlaçların yan etkilerini ortaya çıkarabileceği ciddi komplikasyonlar sağlığımızı bozabilir bu neden ile bilgilendirici özellikleri grafiksel olarak kullanıcıya yaşanmış tecrübeleri analiz verisi olarak frekans hesaplamasını sunmak son derece önemlidir.

İlaç yorumlarından çıkan yan etkiler Sedrax web panel üzerinde Grafikler menüsünde demografik bilgiler ve kullanım bilgileri ile sorgulama yaptığında ilaç yan etki bilgileri

frekans hesaplamaları grafik olarak (bar basic) formatında kullanıcı bilgisine sunulması sağlanmıştır. Şekil 4.26.' da gösterildiği gibidir[15].



Şekil 4.26. Web Panel Üzerinden İlaç Yan Etki Frekans Grafiği

4.11. Kullanılan Algoritmalar ve Modellerin Yapısı

Bu çalışmada, Apriori, Knn ve K-Means algoritmaları kullanılmıştır. Web sitesine yapılan ilaç yorumlarının içerisinde, klinik dil ve terminolojiye ait yan etki kelimeler ile yorumlayıcı bilgileri arasında birliktelik kuralları çerçevesinde anlamlı ilişkileri, sınıfları ve örüntüleri saptamaya çalışılmıştır.

4.11.1. Verilerin Weka ve web panel üzerinden Apriori algoritması ile analizi

Veri madenciliğinde kullanılan veriler arasındaki ilişkiyi çıkarmak için geliştirilmiş algoritmalarından bir tanesidir. Algoritmanın adı, daha önceki kullanımlara bağlı olduğu dolay, latince, önce anlamında olan “prior” kelimesinden türemiştir. Yüzeysel olarak bakıldığında birliktelik kuralı çıkarımlarında kullanılmak da olan bir algoritmadır. Algoritmanın amacı, veri seti bağlantılarını tespit etmektir.

Algoritma model olarak, aşağıdan yukarıya modeli kullanmaktadır ve her seferinde tek bir değeri araştırmak da ve bu değer diğerleri ile arasında ki ilginç ilişkileri keşfetmek için çalışmaktadır. Ayrıca algoritma tüm değerler için çalıştığından dolayı, arama algoritmasına benzerliği bilinmektedir. Kural tabanlı bir makine öğrenme yöntemidir.

Apriori konusu üzerindeki ilk yayını yapan kişiler Agrawal ve Srikat, tarafınan algoritmanın müsvette kodu (pseudo code) aşağıdaki şekilde verilmiştir.

```

L1 = {sık tekrarlanan öğeleri içeren küme}
for(k=2; L k-1 != 0 ; k++){
    Ck = apriori-üret (L k-1);
    forall (işlemler t ÎD){
        Ct ⊂ (Ck,t);
        forall adaylar c ∈ Ct{
            c.sayaç++;
        }
    }
}

```

$$L_k = \{ c \in C_k \mid c.\text{sayaç} = \text{enküçük} \}$$

$$\}$$

$$\text{Sonuç} = \bigcup_k L_k;$$

Algoritmayı anlatmak gerekirse, işlemdeki L , veri tabanı veya veri setinde bulunan öğeleri temsil etmektedir. Apriori üret fonksiyonundan üretilen her bir C_k kümesi için, t bir işlem (transaction) olmak üzere, üretilen bu C_k kümesinin t işlemini tutan sayacı, her münasebette bir arttırıyoruz. Ardından sıradaki L değerini alarak algoritmamız çalışmaya devam ediyor. Sonuç olarak da üretilen bu L değerlerinin birleşim kümesini döndürüyoruz.

Yukarıdaki algoritmada belirsiz olan apriori-üret fonksiyonunu ise aşağıdaki şekilde verebiliriz:

C_k 'ya ekle

select p.eleman1, p.eleman2, ... p.eleman k-1, q.eleman k-1

for(L k-1 p'den L k-1 q'ya kadar)

where p.eleman1, p.eleman2, ... p.eleman k-2 = q.eleman k-2 , p.eleman k-1 < q.eleman k-1

forall elemanlar $c \in C_k \{$

 forall c 'nin (k-1)-alt kümesi s için{

 if($s \notin L_{k-1}$){

c 'yi C_k 'dan sil

 }

 }

}

Yukarıdaki algoritma, basitçe bütün adayların üretilerek elemanların aday kümelerine konulmasını hedeflemiştir[19].

Bu kısımda, ilaç yorumlarına ait veriler üzerinden oluşturulan veri seti; ilaç kullanım bilgileri, kullanıcı demografik bilgiler ve yorumlardan çıkan yan etki bilgileri birliktelik kuralına göre analiz edilecektir.

Çalışmada ilk olarak veriseti, Php programlama dilinde geliştirmiş olan Apriori kütüphanesine uygun program örnek kodlar derlenmiştir[20]. Apriori kodlarını sedrax web panel üzerinde kullanılması sağlanmış olup birliktelikte kullanılan ölçütler aşağıdaki gibi sistemde tanımlanmıştır[15].

Destek (Support) değeri: öge kümesinin veri kümesinde ne sıklıkta görüldüğünün bir göstergesi olup Şekil 4.28.' de apriori kuralında kullanılmıştır.

- $\text{support}(X \rightarrow Y) = \text{support}(XUY)$, range: [0,1]

Güven (Confidence) değeri: Kuralın ne sıklıkla doğru olduğunun bir göstergesidir. X değer grubunun Y değer grubunu da alma olasılığını ortaya koyar. Aşağıdaki gibi formüle edilmekte olup Şekil 4.28.' de apriori kuralında kullanılmıştır.

- $\text{confidence}(X \rightarrow Y) = \text{support}(X \rightarrow Y) / \text{support}(X)$, range: [0,1]

Yükseltme (Lift) değeri: Hem kuralın güvenini hem de genel veri kümesini dikkate almaktadır. X ve Y'nin oransal olarak bağımsız olması halde, ne kadar birlikte kullanıldığını hesaplamaktadır. Aşağıdaki gibi formüle edilmekte olup Şekil 4.28.' de apriori kuralında kullanılmıştır.

- $\text{lift}(X \rightarrow Y) = \text{confidence}(X \rightarrow Y) / \text{support}(Y)$, range: [0,∞]

Kaldıraç (Leverage) değeri: X ve Y' nin birlikte bulunmaları ile birlikte bulunmama arasındaki farkı ölçmektedir ve X ile Y değerleri istatistiksel olarak bağımlı ortaya çıkarmaya çalışmaktadır. Aşağıdaki gibi formüle edilmekte olup Şekil 4.28.' de apriori kuralında kullanılmıştır.

- $\text{leverage}(X \rightarrow Y) = \text{support}(X \rightarrow Y) - \text{support}(X) \times \text{support}(Y)$, range: $[-1, 1]$

Kanaat (Conviction) değeri: itimat ölçümünün alternatifi olarak düşünülebilir. Kanaat ölçümü hesaplanırken, X değerini, Y değerine olmadan görülme olasılıkları ortaya çıkarır; şayet X ve Y frekans değerleri bağımlı ise, X'in Y'den bağımsız olarak görülme oranları hesaplayarak bulunur. Bu işlemde kanaat değeri ilgi değerine anımsatmaktadır. Fakat ilgi değeri, yönü olan bir ölçüm işlemi olup ardışık bilgileri değerlendirmede kullanılmaktadır. Aşağıdaki gibi formüle edilmekte olup Şekil 4.28.' de apriori kuralında kullanılmıştır.

- $\text{conviction}(X \rightarrow Y) = 1 - \text{support}(Y) / 1 - \text{confidence}(X \rightarrow Y)$, range: $[0, \infty]$

Yazmış olduğum web yazılımda (Sedrax) örnek kütüphane ile oluşturulan Apriori uygulaması entegre edilmiş ve ölçüm kuralları oluşturulmuş olup panel üzerinden çalıştırılmıştır[15]. Şekil 4.27. 'de örnek Apriori kullanımı gösterilmiş ve Şekil 4.28. ' de sonuçları görüldüğü gibi modellenmiştir.

Oluşturulan veri setinde WEKA3.8 yazılımı kullanılmıştır. Web panel Sedrax üzerinden Weka Arff dosyası oluşturulması sağlanmış bu sayede birliktelik kurallarını öğrenmek için veri seti hazır duruma getirilmiş ve sonuçları değerlendirmeye açılmıştır[15][21].

The screenshot shows the SEDRAX web interface. On the left is a navigation menu with options like 'Anasayfa', 'Ham Veri - Webmd', 'Yan Etkiler (Veri Çekimi)', 'Yorum Analiz (Like)', 'Yorum Analiz (Soundex)', 'Reviews Analiz (Parse)', 'Weka Dosyası Oluştur', 'Yan Etkiler - İlaçların Dağılımı', 'Veri İstatistik', 'İlaç Frekansı', 'Grafikler', 'Metin Madenciliği - Birliklik', 'Metin Madenciliği - Sınıflandırma', 'Metin Madenciliği - Kümeleme', and 'Kullanıcılar'. The 'Metin Madenciliği - Birliklik' option is highlighted.

The main content area is titled 'Metin Madenciliği - Birliklik (Apriori)'. It contains a configuration form with the following fields:

- İlaç : Jakafi
- Kullanım Süresi : Hepsi
- Yaş Aralığı : Hepsi
- Cinsiyet : Hepsi
- Tip : Hepsi
- NumRules : 10
- MaxScan : 2
- MinSup : 1
- MinConf : 90
- Durum : Hepsi
- Etki : Hepsi
- Kullanım Kolaylığı : Hepsi
- Memnuniyet : Hepsi

Below the form is a 'Listele' button. The output section is titled 'Association Rules' and shows the following rules:

```

Time: 0.01 second(s)
*****
Jakafi 7 ==> Myelofibrosis 7 = supp:(1,00) conf:(1,00) lift:(0,54) lev:(-0,86) conv:(0,00)
Patient 6 ==> Myelofibrosis 6 = supp:(0,86) conf:(1,00) lift:(0,54) lev:(-0,73) conv:(0,00)
Patient 6 ==> Jakafi 6 = supp:(0,86) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
SK 4 ==> Myelofibrosis 4 = supp:(0,57) conf:(1,00) lift:(0,54) lev:(-0,49) conv:(0,00)
SK 4 ==> Jakafi 4 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
SE 4 ==> Myelofibrosis 4 = supp:(0,57) conf:(1,00) lift:(0,54) lev:(-0,49) conv:(0,00)
SE 4 ==> Jakafi 4 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
SM 3 ==> SK 3 = supp:(0,43) conf:(1,00) lift:(1,75) lev:(0,18) conv:(0,00)
SM 3 ==> SE 3 = supp:(0,43) conf:(1,00) lift:(1,75) lev:(0,18) conv:(0,00)
SM 3 ==> Myelofibrosis 3 = supp:(0,43) conf:(1,00) lift:(0,54) lev:(-0,37) conv:(0,00)

```

Şekil 4.27. Web Panel Üzerinden Örnek Apriori Filtre Görünümü

The screenshot shows the 'Association Rules' output for a different filter configuration. The title is 'Association Rules' and the time is 'Time: 0.08 second(s)'. The output is as follows:

```

*****
Birth Control 76 ==> Yasmin 28 76 = supp:(1,00) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
Yasmin 28 76 ==> Birth Control 76 = supp:(1,00) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
Patient 73 ==> Birth Control 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
Patient 73 ==> Yasmin 28 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
Female 73 ==> Birth Control 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
Female 73 ==> Yasmin 28 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
5E 43 ==> Birth Control 43 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
5E 43 ==> Patient 43 = supp:(0,57) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
5E 43 ==> Yasmin 28 43 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
5E 43 ==> Female 43 = supp:(0,57) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
5K 41 ==> Yasmin 28 41 = supp:(0,54) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
5K 41 ==> Female 41 = supp:(0,54) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
5K 41 ==> Patient 41 = supp:(0,54) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
5K 41 ==> Birth Control 41 = supp:(0,54) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
25-34 34 ==> Birth Control 34 = supp:(0,45) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
25-34 34 ==> Patient 34 = supp:(0,45) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)

```

Şekil 4.28. Web Panel Üzerinden Örnek Apriori Sonuç Görünümü

4.11.2. Verilerin web panel üzerinden Knn algoritması ile analizi

Bu algoritmanın adı K komşu algoritmasıdır. Çalışma mantığı nesnelerin bir biri arasında yakınlık ilişkilerine göre sınıflandırma işlemi yapar. Doğrusal ayrıştırma yöntemi ile kordinat düzleminde çalışır. Uzaklık bulunurken euclidean, manhattan, minkowski gibi farklı uzaklık formülleri ile hesaplanabilir. Şekil 4.29.' da formülleri gösterilmiştir.

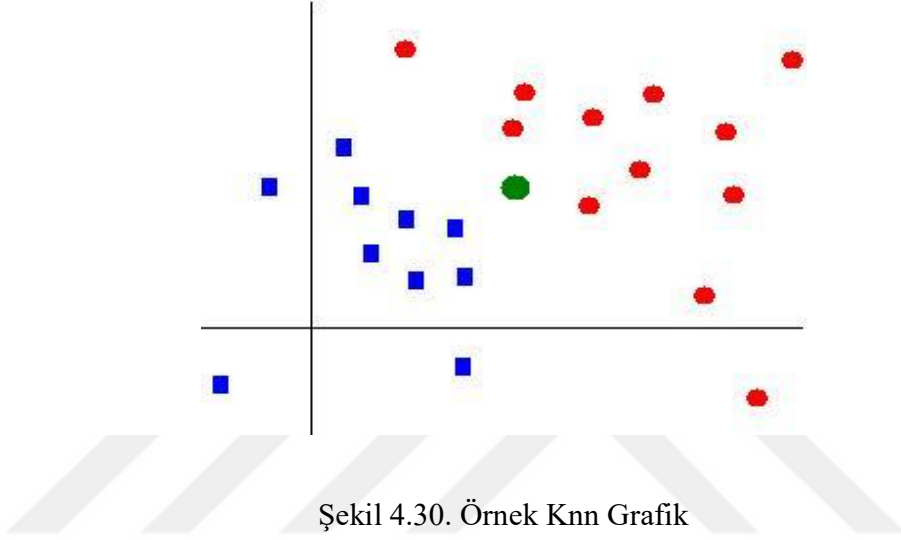
Distance functions

Euclidean	$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^k x_i - y_i $
Minkowski	$\left(\sum_{i=1}^k (x_i - y_i)^q \right)^{1/q}$

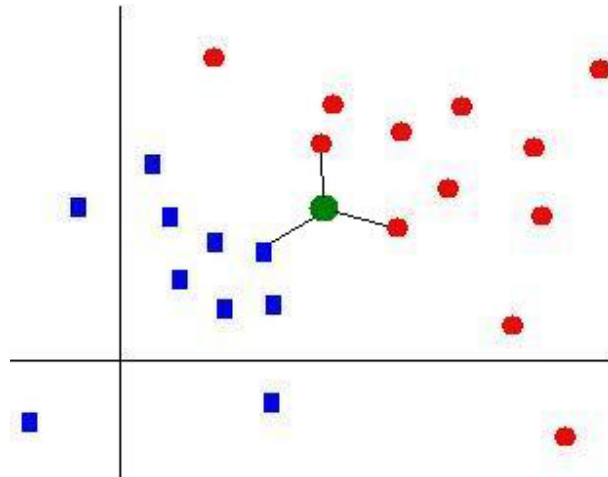
Şekil4.29. Knn Uzaklık Formülleri

Sınıflandırma (classification) algoritmasına göre sınıflandırma işlemi yapılırken ortaya çıkarılan özelliklerden, sınıflandırılmak için kullanılmak istenen yeni nesnenin önceden tanımlanan nesnelere k adedine yakınlığı ölçülmesidir. Diyelimki k = 2 için yeni bir nesne sınıflandırılmak isteniyor ve bu olayda daha önce sınıflandırılmış nesnelere en yakın 2 tanesi işleme alınır. Bu nesnelere bakılır ve hangi sınıfa dahil edilmişse, yeni nesne de o sınıfa eklenir. Mesafe ölçülmesini için genelde öklit uzaklık kullanılır.

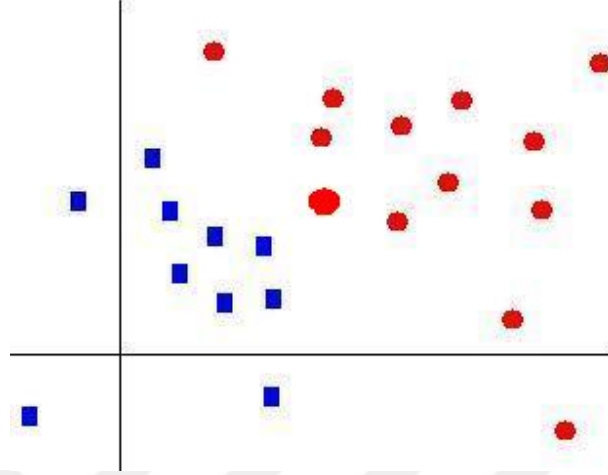
Örneğin Şekil 4.30.' da görünen veri seti özelliklerine bakıldığında 2 boyutlu koordinat sistemi üzerine yerleştirilmiş olan noktaları bakalım. Bu örnekteki noktaların birbiri ile ayrılması durumunda doğrusal ayrıştırma problemi ortaya çıkmaktadır ve knn algoritması metoduna göre çözülebilir. Şekil 4.30.' da yeşil noktalı yeni bir üyenin geldiğini düşünelim.



Şekilde ki yeni gelen noktanın en yakın olduğu 3 kümeyi tespit edelim. Şekil 4.31.' de ki gibi görünür.



En yakın olan 3 noktanın iki adedi kırmızı üyeler olduğuna göre bu durumda yeni gelen noktanın Şekil 4.32.' de olduğu gibi kırmızı olarak sınıflandırabiliriz.



Şekil 4.32. Örnek Knn Sınıflanmış Grafik

Bu projede ilaç kullanıma bağlı olarak oluşturulan veri seti üzerinden Knn algoritması kullanılmış olup Şekil 4.33.' de gösterildiği gibi modellenmiştir.

SEDRAx

Metin Madenciliği - Sınıflandırma (Knn)

İlaç : Yasmin 28
 Kullanım Süresi : 1 to 6 months
 Yaş Aralığı : 7-12
 Cinsiyet : Female
 Tip : Patient
 K : 1

Durum : Birth Control
 Etki : 4
 Kullanım Kolaylığı : 5
 Memnuniyet : 5

1-76

Listele

Crying

Time	Age	Gender	Type	Effectiveness	Easeofuse	Satisfaction	Condition	Side Effect	Distance	Normalization Distance
1 to 6 months	19-24	Female	Patient	4	5	3	Birth Control	Crying	2	0,40
less than 1 month	19-24	Female	Patient	4	5	5	Birth Control	Fatigue	2	0,43

Şekil 4.33. Web Panel Üzerinden Örnek Knn Uygulaması

K-En Yakın Komşu algoritması en yakın komşuya bakar sınıflandırmayı buradan gelen bilgiye göre yapar. Çalışmamızda, Şekil 4.33.' de görüldüğü gibi filtre alanından seçilen ilaç adı veri seti üzerinden analiz için Knn verisini oluşturmaktadır. Knn kütüphanesine uygun Php yazılım dilinde yazılmış örnek yazılım kodu web panele uyarlanmıştır[22]. Buradaki amaç ilaç kullanımına bağlı olarak panel üzerindeki kullanıcı profiline veri seti üzerindeki komşuluk ilişkisi çıkarmak ve olası kullanıcıya en yakın ilaç yan etki bilgisini vermektir. Kullanıcı seçimin uzaklık değerlerini tespit etmek ilaç aşağıdaki işlemler yapılmıştır.

- Projemizde K sınıflandırılmamış örneği durumlardadır ve panel üzerinden filtre alanına oluşturulan sınıf ile seçilen ilaç verisi üzerinden örnek Knn algoritması çalıştırılmıştır.
- Kullanıcı panel filtresi üzerinden seçmiş olduğu parametreler veri seti üzerinden eşitlik “var - yok“ karşılaştırması yapar ve parametre sayısı kadar karşılığı olmayan değerler için uzaklık değerini bir arttırır. Bu değerler en yüksek uzaklık değer frekansına sahip olacak şekilde tabloya yansıtılır. Uzaklık değeri Şekil 4.34.' de örnek distance değeri olarak hesaplanmıştır.
- K en yakın komşu değeridir. $K = 1$ 'i seçmek en yakın komşu değeri ile aynı olacaktır, çünkü sadece en yakın 1'e bakar. $K = N$ (burada N, eğitim örneklerinin sayısıdır ve seçilen ilaç yorumu kadardır) sınıflamayı tüm ilaç yorum örneklerin uzaklık frekansı hesaplaması yaparak değerlendirir.
- Veri seti üzerinden normal uzaklık değerleri aynı olması durumunda K değeri seçimi karar verilmesi için normalizasyon uzaklık değeri eklenmiştir.

En Düşük – En Yüksek Normalleştirilmesi (Min - Max Normalisation) : Yaygın olarak kullanılan bu metot da, veri grubu içerisindeki en yüksek ve en düşük değerleri ele alarak işlem yapar. İşlemdeki tüm veriler, bu mantıkla işleme alınarak normalleştirilir. Bu işlemdeki amaç en düşük değerini 0 ve en yüksek değerini 1 olacak yapıda normalleştirilmesini sağlamak ve işlemdeki bütün verileri bu 0 - 1 aralığında olmasını sağlar. Denklem 5.1.' de kullanılan formül gösterilmiştir.

$$MinMax = \frac{(V - Min\ marks)}{Max\ marks - Min\ marks} (newMax - newMin) + newMin \quad (5-1)$$

Örneğin; V1=8, V2=10, V3=20 değerler olsun ve min-max normalizasyon işlemini Denklem 5-1' de ki formül ile uyarlanmıştır. Denklem 5-2,5-3,5-4 sonuçları gösterilmiştir.

$$MinMax = \frac{(8 - 8)}{20 - 8} * (1 - 0) + 0$$

$$MinMax = \frac{(0)}{12} * 1$$

$$MinMax = 0 \quad (5-2)$$

$$MinMax = \frac{(10 - 8)}{20 - 8} * (1 - 0) + 0$$

$$MinMax = \frac{(2)}{12} * 1$$

$$MinMax = 0.16 \quad (5-3)$$

$$MinMax = \frac{(20 - 8)}{20 - 8} * (1 - 0) + 0$$

$$MinMax = \frac{(12)}{12} * 1$$

$$MinMax = 1 \quad (5-4)$$

Normalleştirilmiş değerler $V1=0$, $V2=0.16$, $V3=1$ olmaktadır.

- Web panel üzerinde uygulanan Knn sorgulama içerisinde normal uzaklık değeri eşit olması durumunda ilaç kullanım süresi ve yaş aralığı parametreleri kullanılarak veri seti üzerinde min-max normalizasyonu yapılmıştır. Öklit uzaklıkları değerleri Denklem 5-5' de gösterilen formül ile işleme alınmış olup Şekil 4.34.' de olduğu gibi normalizasyonlu uzaklık(Normalization Distance) değerleri projeye uygulanmıştır. K değeri aynı olması durumunda en yakın komşu değerleri bu yöntem tespit edilmiştir[15].

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

(5-5)

Metin Madenciliği - Sınıflandırma (Knn)

İlaç : Cipro
 Kullanım Süresi : less than 1 month
 Yaş Aralığı : 19-24
 Cinsiyet : Male
 Tip : Patient
 K : 1

Durum : Birth Control
 Etki : 4
 Kullanım Kolaylığı : 5
 Memnuniyet : 5

1-140

Listele

Infection, Pain, Sensitisation, Burning sensation

Time	Age	Gender	Type	Effectiveness	Easeofuse	Satisfaction	Condition	Side Effect	Distance	Normalization Distance
less than 1 month	19-24	Male	Patient	4	5	1	Bacterial Urinary Tract Infection	Infection, Pain, Sensitisation, Burning sensation	2	0,00
less than 1 month	55-64	Male	Patient	4	5	5	Bacterial Urinary Tract Infection	Infection	2	0,57
less than 1 month	35-44	Male	Patient	4	5	1	Bacterial Urinary Tract Infection	Disability, Infection, Pain	3	0,29

Şekil 4.34. Web Panel Üzerinden Örnek Uygulanan Knn Algoritması

4.11.3. Verilerin panel üzerinden K-Means algoritması ile analizi

Veri madenciliğinde kümeleme yöntemlerinden K-Means algoritması çalışma yöntemi x adet veri objesinden ortaya çıkan verilerin, giriş değişkeni olarak tanımlanan y adet kümeye ayırmaktadır. Amaç, gerçekleştirilen kümeleme işlemi sürecinde ortaya çıkan kümelerin, kümede olan benzerliklerinin en üst derecede olmasını sağlamak ve küme içi benzerliklerinin minimum dereceye gelmesini hedeflemektedir. Kümelerin yakınlığı, merkez olarak kabul edilen bir nokta ile kümedeki diğer noktalar arasındaki uzaklıkların ortalama değeri ele alarak ölçülür ve hesaplanan değer kabul edilir[23].

K-Means en çok kullanılan gözetimsiz öğrenme metotlarından birisidir. Atama veya yönlendirme yapılırken öncelik tüm verilerin sadece bir kümeye ait olabilmesine olanak verir. Bu sebeple, çok net bir kümeleme algoritması olduğu bilinmektedir. Atama işlemi merkez noktanın kümeyi temsil etmesini sağlayan bir işlem durumudur. Küme içeriğindeki nesnelere aynı oranda çıkma eğilimi vardır ve küresel kümeleri bulmaya çalışır[24].

Bu projede ilaç kullanıma bağlı olarak yorumlanmış işlenmiş, sınıflandırılmış veri seti üzerinden K-Means algoritması kullanılmış olup Şekil 4.35.' de gösterildiği gibi modellenmiştir.

K-Means algoritması ile ilaç kullanım bilgileri ve yan etki çıkarımları analiz edilerek yazılan web sitesi(Sedrax) yazılımı aracılığı ile analiz edilmiş olup sağlık alanında ilaç kullanımına dikkat çekecek sonuçlar yorumlanmıştır[15].

Metin Madenciliği - Kümeleme (Kmeans)

İlaç : Yasmin 28 Yan Etki : Seçiniz

Kullanım Süresi : Seçiniz Durum : Seçiniz

Yaş Aralığı : Seçiniz Etki : Seçiniz

Cinsiyet : Seçiniz Kullanım Kolaylığı : Seçiniz

Tip : Seçiniz Memnuniyet : Seçiniz

X : Kullanım Süresi Y : Yaş Aralığı

K (Grup Adedi) : 3

Grup	Adet	Oran
0	24	32,43 %
1	23	31,08 %
2	27	36,49 %

Cluster 0: Birth Control,Female,25-34,less than 1 month,Patient,1E,1K,1M
Cluster 1: Birth Control,Female,13-18,1 to less than 2 years,Patient,5E,5K,5M
Cluster 2: Birth Control,Female,25-34,10 years or more,Patient,5E,5K,5M

Şekil 4.35. Web Panel Üzerinden Örnek K-Means Uygulaması

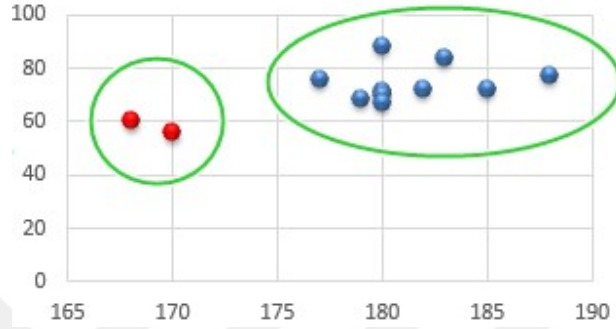
İlaç yorum veri seti alınarak K-Means bölümlenmeli kümeleme algoritmalarıyla ilaç yan etkilerine göre kümelere ayrılması hedeflenmiştir. Uygulama K-Means kütüphanesine uygun php yazılımı ile yazılmış kod kullanılmıştır[25]. Web ara yüzünde geliştirilerek internet ortamına “sedrax.com” adlı web sitesinden ve android mobil ortamında kullanıcılara sunulmuştur[15].

Elde edilen sonuçlar yorumlanırsa, K-Means algoritması oluşturulan veri setindeki ilaç seçimine göre ve x ve y olmak üzere 2 adet veri sınıf parametresi üzerinden k adet gruba (cluster - küme) ayırması için tasarlanmıştır.

- Seçilen ilaç yorum verisi işlem görür ve k grubu kadar K-Means işlem yapar.
- Gruplara ayrılmış veri seti ilaç kullanımların yan etki gruplarını da çıkarır.

Kullanılan algoritma daha iyi anlaşılması için aşağıdaki örneğe göz atmak gereklidir.

Örnek K-Means senaryosu: Boy ve kilo bilgisine sahibiz. Bu iki değişkeni kullanarak nesnelere yükseklik ve ağırlık bilgilerine göre gruplandırılması yapılmıştır. Şekil 4.36.' da grafiksel olarak gruplaması sağlanmıştır. K-Means algoritması kullanılarak gruplama işlemini bu doğrultu da tanımlanması beklenir.



Şekil 4.36. Örnek Kümelen Veri Görünümü

1. Adım: Bu veri kümesinde, kümelene için (Yükseklik ve Ağırlık) dikkate alınmaktadır. Tablo 5.9.' da gösterilen veri seti kullanılmıştır.

Tablo 4.9. Örnek K-Means Verisi

Yükseklik	Ağırlık
185	72
170	56
168	60
179	68
182	72
188	77
180	71
180	70
183	84
180	88
180	67
177	76

2. Adım: Küme merkezini Bu örnekte, K değeri 2 olarak kabul edilir. Kümelenmiş merkezler ilk 2 gözlemle başlatılır. Tablo 4.10.' da gösterilmiştir.

Tablo 4.10. Örnek K-Means Gözlem Değeri

Küme	İlk Merkez	
	Yükseklik	Ağırlık
K 1	185	72
K 2	170	56

3. Adım: Öklit uzaklığı Denklem 5.6.' da ki formül ile hesaplanır. Bir gözlem başlangıç kümesi belirlenir ve merkezi değer 1 ve değer 2 arasındaki Öklit mesafesine göre hesaplanır. Atamalar minimum Öklit mesafesine göre yapılır. Tablo 4.11.' de her bir küme Öklit uzaklığına göre hesaplanmıştır. Gözlem noktası veri setindeki ilk değer olduğundan koordinat merkez noktası değişmemiştir.

$$\text{Euclidean Distance} = \sqrt{(X_H - H_1)^2 + (X_W - W_1)^2}$$

(5-6)

Tablo 4.11. Örnek K-Means Gözlem Atama Tablosu

Öklit Mesafe 1	Öklit Mesafe 2	Küme Atama
$(185-185)^2 + (72-72)^2 = 0$	$(185-170)^2 + (72-56)^2 = 21,93$	1
$(170-185)^2 + (56-72)^2 = 21,93$	$(170-170)^2 + (56-56)^2 = 0$	2

4. Adım: Bir sonraki gözlem noktası olan Yükseklik 168, Ağırlık 60 değerleri ile küme merkez gözlem noktası Tablo 4.12.' de ki gibi gözlem atama noktası belirlenir ve sonraki yeni merkez küme koordinatı Tablo 4.13.' de ki gibi hesaplanır.

Tablo 4.12. Örnek K-Means Atama Tablosu

Öklit Mesafe 1	Öklit Mesafe 2	Küme Atama
$(168-185)^2 + (60-72)^2$ = 20.808	$(168-185)^2 + (60-72)^2$ = 4.472	2

Tablo 4.13. Örnek K-Means Merkez Koordinat Tablosu

Küme	Güncellenmiş Merkez Veri Sırası(168 , 60)	
	Yükseklik	Ağırlık
K 1	185	72
K 2	$(170 + 168) / 2 = 169$	$(56 + 60) / 2 = 58$

5. Adım: Bir sonraki gözlem noktası için Öklit hesaplanır. Tablo 4.14.' de gösterilmiş ve ataması yapılmıştır.Yapılan atama neticesinden merkez koordinat güncellenmiş olup Tablo 4.15.' de gösterilmiştir.

Tablo 4.14. Örnek K-Means Atama Tablosu

Öklit Mesafe 1	Öklit Mesafe 2	Küme Atama
7,21	14,14	1

Tablo 4.15. Örnek K-Means Merkez Koordinat Tablosu

Küme	Güncellenmiş Merkez Veri Sırası(179,68)	
	Yükseklik	Ağırlık
K 1	$(179 + 185) / 2 = 182$	$(68 + 72) / 2 = 70$
K 2	169	58

Son Adım: Tüm gözlemler aynı format da işlemden geçmiş olup merkez koordinat noktası Tablo 4.16.' da gösterilmiş olup Tablo 4.17.' de veri setine küme sınıfı atanmıştır[26].

Tablo 4.16. Örnek K-Means Atama Tablosu

Küme	Son Merkez	
	Yükseklik	Ağırlık
K 1	182	72
K 2	169	58

Tablo 4.17. Örnek K-Means Küme Sınıf Verisi

Yükseklik	Ağırlık	Atanan Küme
185	72	1
170	56	2
168	60	2
179	68	1
182	72	1
188	77	1
180	71	1
180	70	1
183	84	1
180	88	1
180	67	1
177	76	1

Örnek çalışmada olduğu gibi bu çalışmada ilaç veri seti 2 parametreye göre K-Means algoritmasıyla çalıştırılmış ve elde edilen analiz sonuçları karşılaştırılmalı olarak incelenerek web ve mobil ortamda sunulmuş ve ilaç yan etki kullanımına yönelik bilgiler sunması sağlanmıştır.

V. MOBİL UYGULAMA VE SONUÇLARI

Mobil uygulamada 2008 - 2017 yılları arasında gerçekleşen yaklaşık 221.359 adet ilaç yorum kaydı báz alınarak yapılmıştır. Uygulama android stüdyo ile Java programlama dili kullanılarak geliştirilmiş olup Mysql veritabanı ile haberleşmesi sağlanması için web API (Application Programming Interface) kullanılmıştır. Uygulama android ortamında yazılmış ve google play üzerinde yayına alınmıştır[27]. Kullanım için anroid kullanıcıların EK U2' de görülen mobil uygulama kayıt formuna bilgilerini girmeleri gereklidir. EK U1' de kayıt bilgilerine göre giriş yapılabilir. Ana sayfada kullanılan özellikler şunlardır; Yan Etki Listesi, İlaç Frekansı, Grafikler, Metin Madenciliği Birliktelik, Metin Madenciliği Sınıflandırma, Metin Madenciliği Kümeleme ve Veri İstatistik olarak tasarlanmıştır EK U3' de gösterilmiştir.

5.1. Mobil Uygulama Alanı

Teknoloji geliştikçe mobil uygulamaları geliştirme çabaları da artmaktadır. Mobil uygulamaların yaygın olarak kullanımı kullanıcıların ulaşabilirlik düzeyini arttırmaktadır. Bu çalışmada analiz edilen yorumlardan ilaç yan etkileri düzeyi, sıklığı gibi kriterlere göre sınıflandırılarak bilgi alabilirler. İlaç alanında araştırma yapmak isteyen kullanıcıların Sedrax[27] uygulamasından arama veya filtre özelliklerini kullanarak kolayca yan etki analiz sonuçlarına erişebilir durumunda olup bunun için uyumlu telefonlar, tabletler ve bilgisayarlar üzerinden android uygulama mağazalarından ücretsiz olarak indirebilirler.

5.2. Mobil Uygulama Sonuçları

Uygulama da işlenmiş, sınıflandırılmış ve denetimli veri eklenerek sistem analizine sunulmuştur. Uygulama içerisinde ilişkilendirilmiş test edilen veriler veri tabanı üzerinde mobil ortama taşınmış olup sağlık ve tıp alanında bilgi sağlayacak kaynak haline getirilmiştir. Bu uygulama da Şekil 5.1.' de WebMD üzerinden çekilen ilaç yorumlarından tespit edilen yan etkilerin hangi ilaç da ne kadar görüldüğü bulunabilir. Şekil 5.2.' de Kullanıcı seçtiği ilacın yan etki bilgisini adet bâzlı alır ve kaç yorumda görüldüğünü, kaç defa tekrar ettiği bilgisini alabilir. Şekil 5.3.' de grafiksel olarak ilaç yan etki oranlarını bilgisini alır. Şekil 5.4.' de veri madenciliği algoritmalarından apriori' nin kullanımını yapabilir. Şekil 5.5.' de Knn algoritmasının sınıflı veriler üzerindeki durumuna bakarak analiz yapabilir. Şekil 5.6.' da K-Means algoritmasının veriler üzerindeki grupsal dağılımına bakarak ilaç yan etkisinin grubunu öğrenebilir.



ID	Yan Etki Adı	İlaç Sayısı
4765	Pain	2469
4356	Nausea	1332
2959	Headache	1104
3416	Infection	1035
2000	Dizziness	1029
6277	Surgery	975
6289	Swelling	968
5537	Rash	944
1936	Diarrhoea	915
1862	Depression	870

Şekil 5.1. Uygulamada Bulunan İlaç Yan Etki Listesi Görünümü

Sedrax

Filtre

İlaç: Seçiniz

Ekle

Yan Etki: Hepsi

Kullanım Süresi: Hepsi

Yaş Aralığı: Hepsi

Cinsiyet: Hepsi

Tip: Hepsi

Durum: Hepsi

Etki: Hepsi

Kullanım Kolaylığı: Hepsi

Memnuniyet: Hepsi

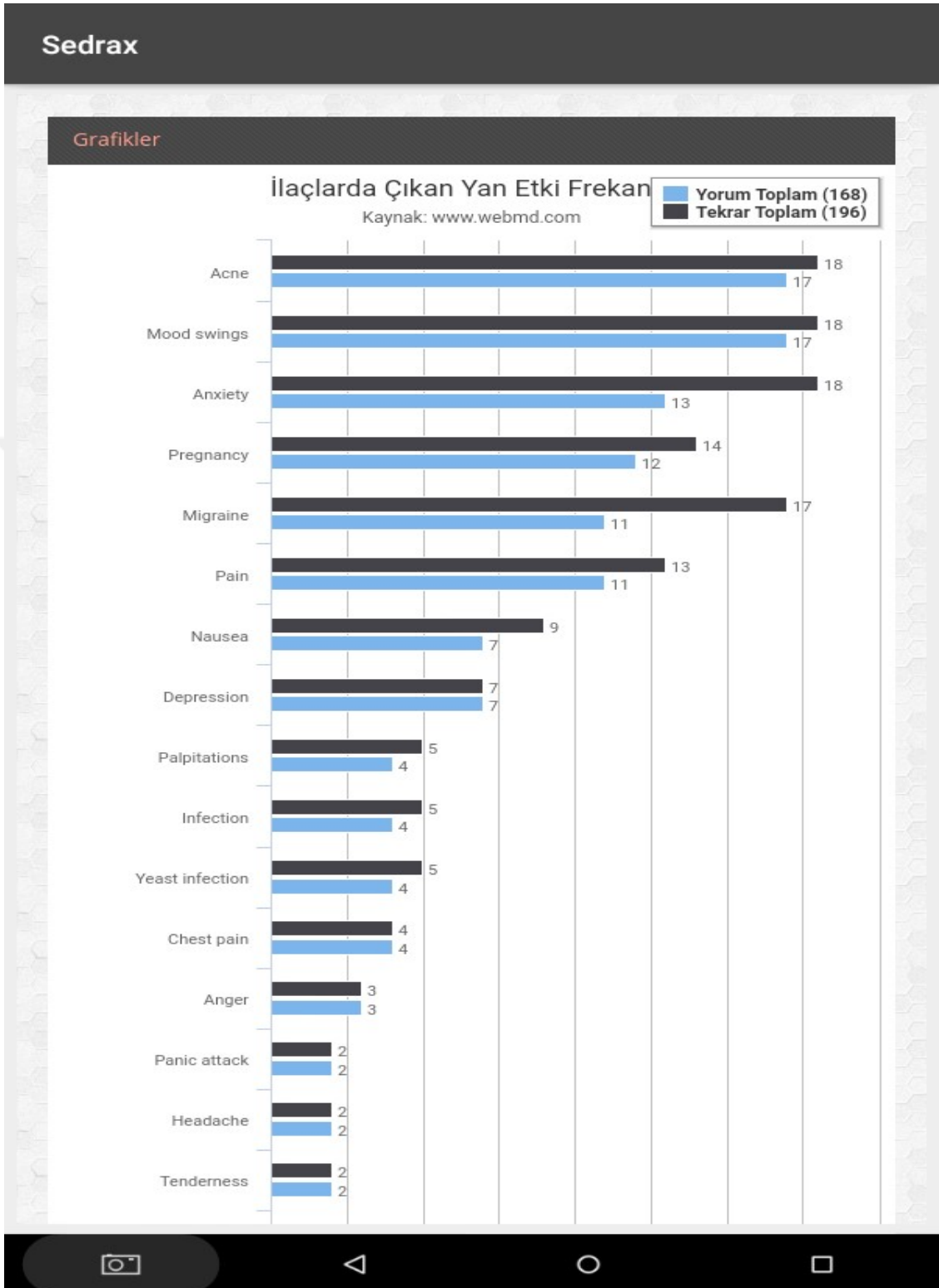
İlaç	Durum	Yorum Sayısı
Yasmin 28	Birth Control	168
YAZ	Birth Control	626
Jakafi	Myelofibrosis	19
Jadenu	Too Much Iron in the Blood for a l	4

Listele

Yanetki Frekansı

Yanetki	İlaç (Yorum Sayısı)	Tekrar Sayısı
+ Acne	2 (93)	113
+ Mood swings	2 (92)	106
+ Migraine	2 (40)	62
+ Pain	3 (46)	60
+ ...	2 (46)	50

Şekil 5.2. Uygulamada Bulunan Örnek İlaç Frekansı



Şekil 5.3. Uygulamada Bulunan Örnek Grafik

Metin Madenciliği - Birliklilik (Apriori)

Metin Madenciliği - Birliklilik (Apriori)

Association Rules

Time: 0.04 second(s)

Birth Control 76 ==> Yasmin 28 76 = supp:(1,00) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 Yasmin 28 76 ==> Birth Control 76 = supp:(1,00) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 Patient 73 ==> Birth Control 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 Patient 73 ==> Yasmin 28 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 Female 73 ==> Birth Control 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 Female 73 ==> Yasmin 28 73 = supp:(0,96) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 5E 43 ==> Birth Control 43 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 5E 43 ==> Patient 43 = supp:(0,57) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 5E 43 ==> Female 43 = supp:(0,57) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 5E 43 ==> Yasmin 28 43 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 5K 41 ==> Birth Control 41 = supp:(0,54) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 5K 41 ==> Female 41 = supp:(0,54) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 5K 41 ==> Patient 41 = supp:(0,54) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 5K 41 ==> Yasmin 28 41 = supp:(0,54) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 25-34 34 ==> Birth Control 34 = supp:(0,45) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)
 25-34 34 ==> Patient 34 = supp:(0,45) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 25-34 34 ==> Female 34 = supp:(0,45) conf:(1,00) lift:(1,04) lev:(0,02) conv:(0,00)
 25-34 34 ==> Yasmin 28 34 = supp:(0,45) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)

Şekil 5.4. Uygulamada Bulunan Örnek Apriori Çıktısı

Metin Madenciliği - Sınıflandırma (Knn)

Metin Madenciliği - Sınıflandırma (Knn)

Numbness, Pain

Time	Age	Gender	Type	Effectiveness	Easeofuse	Satisfaction	Condition	Side Effect	Distance	Normalization Distance
6 months to less than 1 year	19-24	Female	Patient	4	3	1	Birth Control	Numbness, Pain	3	0,00
6 months to less than 1 year	19-24		Patient	4	4	1	Birth Control	Anxiety, Mood swings	3	0,00
6 months to less than 1 year	19-24	Female	Patient	4	5	2	Birth Control	Mood swings	3	0,00

Şekil 5.5. Uygulamada Bulunan Örnek Knn Çıktısı



Şekil 5.6. Uygulamada Bulunan Örnek K-Means Algoritması Sonuçları

VI. SONUÇ

Bu çalışma kapsamında, “www.webmd.com” adresinde bulunan 2008 - 2017 tarihleri arasında yazılmış ilaç yorumlarına ait yapılan, (toplam 221.359 adet kayıt) C#.Net dilinde yazılan web sayfası okuyucu (html okuyucu) program kullanılarak elde edilen veriler Mysql veritabanı sisteminde bir araya getirilerek veri kümesi oluşturulmuştur. Çok geniş bir çalışma alanı olan veri madenciliğinin birliktelik, sınıflandırma ve kümeleme algoritmaları üzerinde analiz edilmiştir.

Çalışmada, birliktelik kurallarından Apriori, sınıflandırma algoritmasında Knn ve kümeleme algoritmasında ise K-Means algoritması kullanılmıştır.

Çalışmamız da, ilaç yorumların sınıflara ayrılması sağlanmış ve daha sonra elde edilen bilgiler çerçevesinde cinsiyet, ilaç kullanım süresi, yaş aralığı, kullanıcı tipi ve kullanım nedenleri ile hangi yan etkilerin kullanımları birlikte görüldüğü tespit edilmeye çalışılmıştır. Veri madenciliği araçlarından Apriori, Knn ve K-Means algoritmaları sayesinde analiz edilmiş olup web ortamında yazılan “sedrax.com” internet sitesinden ve android mobil mağazasında “sedrax” uygulamasıyla beraber kullanıma sunulmuştur.

Bu çalışmanın zayıf yönleri, ilaç yorumlarından çıkarılan yan etkilerin hastalık öncesi yaşanan durumlardan kaynaklı olarak da yorumlarda geçme durumu olmasından dolayı tam olarak ilaç yan etkisi olarak yorumlanmasını zorlaştırmaktadır. WebMD sitesinde tüm ilaçlara yorum yapılmış olmaması nedeniyle veri setinde karşılığı olmayan cevaplar

karşımıza çıkmak da olup dengeli bir dağılım bulunmamaktadır. Ayrıca analiz edilen verilerin, internet üzerinde bulunan web sitesinden elde edilmiş olması ve günümüz teknolojisi ile bu verilerin gerçek kişiler ve gerçek etkileri tamamen yansıttığı onaylanamadığı için verilerin doğruluğundan şüphe edilmesi çalışmanın eksik yönlerinden sayılabilir.

Sağlık, bireyler ve ülkeler için güvence altına alınması gereken en büyük değerdir. Bu önem çerçevesinde yapılan çalışma yapısı, kurulan sistem modelin tek bir yapı üzerinde toplanıp tamamen otomatik bir hale getirerek sürekli olarak yenilenen ilaç yan etki sonuçlarını görebilme olanağı vermiş olup yeni markalaşan ilaçlar içinde sonuçlarını görebileceğimiz bir çalışma başlatılabilir. Bunlara ek olarak günümüz internet teknolojilerinin gelişmesi büyük oranda veri kaynakların çıkarmış olup kurulan model çıkarımları geliştirilebilir ve ilaç yorum içerikleri daha yararlı bilgi çıkarımları sağlayacak ilişkisel kelime darağacına uygun cevaplar alınabilir. Örneğin bir web sitesinde yorum yapan kişilerin yorumda ki metin bağlantıları iyi kurgulanmış modeller ile sonuçları anlamlandırılabilir.

Sonuç olarak bu çalışmadan çıkan sonuçlar, ilaç kullanımdan önce bilgi almak isteyenlere, doktorlara, ilaç alanında araştırma yapan Ar-Ge firmalarına, ilaç üretimi yapan şirketlere büyük bir katkı sağlayacağı, önemli bir bilgi kaynağı olarak görülebilir. Yapılan çalışma betik yazılım dili ile anlamlandırılmış olup görselleştirilmiştir. Ek-P1-P2...P12' de web panel görüntüleri gösterilmiştir.

KAYNAKÇA

1. Viki-zero, özgür ansiklopedi “*Veri Madenciliği*”
“<http://www.wikizero.net/index.php?q=aHR0cHM6Ly90ci53aWtpcGVkaWEub3JnL3dpa2kvVmVyaV9tYWRlbnNpbGnEn2k>”, 30 Mayıs 2018
2. C.Clifton, B.Thuraisingham, “*Emerging Standards for Data Mining*,” Computer Standards & Interfaces, syf.187-193, 2001.
3. Y.S.Türkan, E.Manisalı, M. F. Çelikkol, *Evaluation of critical success factors effect on six sigma project success in Turkey’s manufacturing sector*, Journal of Engineering and Natural Sciences, syf.105-117, 2009.
4. A.S.Koyuncugil, *Veri Madenciliği Ders Notları: Yönetim Bilişim Sistemleri*, <http://www.koyuncugil.org/files/ders/bolum6.pdf>. Ankara. 27 Aralık 2010
5. Tahminciler, Erhan. “*Erythromcin İlacının Yan Etkilerinin Araştırılması Üzerine Veri Madenciliği Çalışması*”, syf. 9-10, 2014.
6. Bilekdemir, Gökhan, “*Veri Madenciliği Tekniklerini Kullanarak Üretim Süresi Tahmini Ve Bir Uygulama*”, <http://acikerisim.deu.edu.tr/xmlui/handle/12345/10890>, s.12, 2010
7. R.J.Roiger, M.W.Geatz, *Data Mining A Tutorial: Based Primer*, Wesley, USA, syf.350, 2003.
8. Hand, D. Mannila, H. Smyth, F. “*Principles of Data Mining*”. England: Cambridge Ma and London. The MIT Press: A. Bradford Book, 2011
9. Oğuzlar, Ayşe. “*Temel Metin Madenciliği*”. Bursa: Dora Yayıncılık Ltd. Şti. 2011
10. Dolgun, M. Özgür. “*Büyük alışveriş merkezleri için veri madenciliği uygulamaları*”. Yayınlanmış Yüksek Lisans Tezi. Ankara: Hacettepe Üniversitesi, 2006
11. Sadi. Evren. Şeker, Khaled. Al-Naami, “*Sentimental Analysis on Turkish Blogs via Ensemble Classifier, (DMIN'13)*”. Proceedings of the 2013 International Conference on Data Mining, ISBN=1-60132-239-9 , pp. 10-16, 2013
12. Sadi. Evren. Şeker, “*Metin Madenciliği*”. MISSozluk. 5 Mart 2016 tarihinde kaynağından arşivlendi. Erişim tarihi: 15 Haziran 2014.
13. Justin. Langseth, “*Unstructured Data and the 80 Percent Rule*”. <https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>, 1 Ağustos 2008.
14. Şadi Evren ŞEKER, “*Turkish Query Engine on Library Ontology*”, IKE12, Internet Knowledge Engineering, 2012, ISBN:1-60132-222-4, Pages:26-33.
15. Alkan.KAYA, “*SEDRAx—Side Effect from Drugs*”, <http://www.sedrax.com/>, 2018.
16. Yan Etkiler Araştırma Grubu, *EMBL: Avrupa Moleküler Biyoloji Laboratuvarı Sider4*, <http://sideeffects.embl.de/>, 2018.
17. Soundex, “*fonetik algoritma*”, <http://www.wikizero.net/index.php?q=aHR0cHM6Ly90ci53aWtpcGVkaWEub3JnL3dpa2kvU291bmRleA>. 2018.
18. HighCharts, “*Dünyanın En Popüler JavaScript Grafik Motoru*”, <https://www.highcharts.com/download>, 2017

19. E.Şeker, “*Bilgisayar Kavramları*”,
<http://www.bilgisayarkavramlari.com/2011/09/07/apriori-algoritmasi/>, 12 Haziran 2018
20. God D. Apo, “*Apriori Örnek Çalışma*”, <https://github.com/filip179/apriori/>, 01 Mayıs 2018
21. WEKA, *Weka 3.8: “Data Mining Software in Java”* “
<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>, 01 Mayıs 2018
22. james. hamilton, “*Knn Örnek Çalışma*”, https://old.jameshamilton.eu/datamining/k-nearest_neighbour.php?source/, 01 Mayıs 2018
23. P.Berkhin, *Survey of Clustering Data Mining Techniques*, San Jose, California, USA, Accrue Software Inc., 2002.
24. M.İşık, A.Y.Çamurcu, *K-means, K-medoids ve Bulanık C-means Algoritma Performans Tespiti*, 2007.
25. Christ Prateek. Prasanna Xaxa, “*K-Means Örnek Yazılım Kodu*”,
https://github.com/cppxaxa/kmeans_cluster_php_library/blob/master/kmeans.php/, 01 Mayıs 2018
26. Christ Prateek. Prasanna Xaxa, “*K-Means Örnek Demo*”,
<https://www.spider.nitt.edu/~cppxaxa/kmeans/> 01 Mayıs 2018
27. Alkan. KAYA, “*Android Mobil Uygulama*”,
<https://play.google.com/store/apps/details?id=com.sedrax.app.sedrax>, 2018.

EKLER

EK A

HTML VERİ OKUYUCU PROGRAMIN KODLARI

```
using System;
using System.Collections.Generic;
using System.Collections.Specialized;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.IO;
using System.Linq;
using System.Net;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
namespace IlacDatasiHazirla
{
    public partial class Form1 : Form
    {
        String domain = "https://www.webmd.com";
        int ilacSayisi;
        int commentCount;
        Boolean isLoad, isLoad2, isLoad3;
        String tmp;
        HtmlDocument subDocument;
        String strComment;
```

```
StreamWriter sw;
String drugName;
int step = 1;
int webCount = 0;
int drugCount = 0;
int sendCommentCount = 0;
String strUrl;
String currentMethod;
Boolean loadVar1, loadVar2, loadVar3;
Boolean isOnlyComment = false;
String strCondition, strDate, strReviewer, strStar1, strStar2, strStar3;

public Form1()
{
    InitializeComponent();
}

public void urlListesiHazirla()
{
    String baseUrl = txtSiteUrl.Text + txtDrugFirstChar.Text;
    int i;
    String url;
    if (txtDrugFirstChar.Text == "0")
    {
        lstDrugGroupUrl.Items.Add(baseUrl);
    }
    else
    {
        for (i = 97; i < 97 + 26; i++)
        {
```

```
        url = baseUrl + "/" + txtDrugFirstChar.Text + ((char)i).ToString();
        lstDrugGroupUrl.Items.Add(url);
    }
}
}

public void eksikUrlListesiHazirla()
{
    lstDrugName.Items.Add("alpha lipoic acid");
    lstDrugName.Items.Add("Apple Cider Vinegar Plus");
    lstDrugName.Items.Add("biotin tablet");
    lstDrugName.Items.Add("black cohosh");
    lstDrugName.Items.Add("caffeine");
    lstDrugName.Items.Add("Cinnamon capsule");
    lstDrugName.Items.Add("Co Q-10");
    lstDrugName.Items.Add("DHEA tablet");
    lstDrugName.Items.Add("echinacea capsule");
    lstDrugName.Items.Add("Fish Oil 60 mg-90 mg-500 mg capsule");
    lstDrugName.Items.Add("flaxseed oil capsule");
    lstDrugName.Items.Add("folic acid");
    lstDrugName.Items.Add("garlic");
    lstDrugName.Items.Add("ginger");
    lstDrugName.Items.Add("ginseng");
    lstDrugName.Items.Add("glucosamine sulfate 500 mg tablet");
    lstDrugName.Items.Add("GOLDEN SEAL root");
    lstDrugName.Items.Add("Green Tea");
    lstDrugName.Items.Add("hoodia gordonii extract capsule");
    lstDrugName.Items.Add("Iron (ferrous sulfate) 325 mg (65 iron) tablet");
    lstDrugName.Items.Add("Kava capsule");
    lstDrugName.Items.Add("lysine");
```

```
lstDrugName.Items.Add("magnesium");
lstDrugName.Items.Add("MEGA taurine");
lstDrugName.Items.Add("melatonin");
lstDrugName.Items.Add("milk thistle capsule");
lstDrugName.Items.Add("MSM capsule");
lstDrugName.Items.Add("niacin tablet");
lstDrugName.Items.Add("potassium");
lstDrugName.Items.Add("saw palmetto 80 mg capsule");
lstDrugName.Items.Add("selenium tablet");
lstDrugName.Items.Add("St. John's wort");
lstDrugName.Items.Add("Vitamin B-6");
lstDrugName.Items.Add("vitamin E 100 unit capsule");
lstDrugName.Items.Add("zinc 50 mg tablet");
lstDrugName.Items.Add("5-HTP");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-767-ALPHA-LIPOIC+ACID.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-816-APPLE+CIDER+VINEGAR.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-313-BIOTIN.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-857-BLACK+COHOSH.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-979-CAFFEINE.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-1002-CASSIA+CINNAMON.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-938-COENZYME+Q-10.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-331-DHEA.aspx?drugid");

lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-
supplements/ingredientreview-981-ECHINACEA.aspx?drugid");
```


IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-993-FISH+OIL.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-990-FLAXSEED+OIL.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-1017-FOLIC+ACID.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-300-GARLIC.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-961-GINGER.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-967-AMERICAN+GINSENG.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-807-GLUCOSAMINE+SULFATE.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-943-GOLDENSEAL.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-960-GREEN+TEA.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-1079-HOODIA.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-912-IRON.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-872-KAVA.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-237-LYSINE.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-998-MAGNESIUM.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-1024-TAURINE.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-940-MELATONIN.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-138-MILK+THISTLE.aspx?drugid");

IstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-522-MSM+METHYLSULFONYLMETHANE.aspx?drugid");

```

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-924-NIACIN+AND+NIACINAMIDE+VITAMIN+B3.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-851-POTASSIUM.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-971-SAW+PALMETTO.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-1003-SELENIUM.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-329-ST+JOHN%27S+WORT.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-934-PYRIDOXINE+VITAMIN+B6.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-954-VITAMIN+E.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-982-ZINC.aspx?drugid");

        lstDrugUrl.Items.Add("https://www.webmd.com/vitamins-supplements/ingredientreview-794-5-HTP.aspx?drugid");

        lstDrugName.SelectedIndex = -1;

        lstDrugUrl.SelectedIndex = -1;

        isOnlyComment = true;
    }

private void btnStart_Click(object sender, EventArgs e)
{
    if (txtDrugFirstChar.Text == "")
    {
        MessageBox.Show("Harf girilmedi!");
        return;
    }

    chkTest.Enabled = false;

    txtDrugCount.Text = "0";

```

```
if (txtDrugFirstChar.Text == "-1")
{
    eksikUrlListesiHazirla();
    callUrl2();
}
else
{
    urlListesiHazirla();
    callUrl1();
}
btnStart.Enabled = false;
button1.Enabled = true;
}

private void callUrl1()
{
    if (int.Parse(txtDrugGroupOrder.Text) > 0 && loadVar1 == false)
    {
        lstDrugGroupUrl.SelectedIndex = int.Parse(txtDrugGroupOrder.Text);
        loadVar1 = true;
    }
    else
    {
        if (lstDrugGroupUrl.Items.Count > 0 && lstDrugGroupUrl.SelectedIndex <
lstDrugGroupUrl.Items.Count - 1)
        {
            if (lstDrugGroupUrl.SelectedIndex > 0)
            {
                if (chkAuto.Checked)
                {
```

```
        lstDrugGroupUrl.SelectedIndex++;
    }

    else
    {
        MessageBox.Show("Data çekme işlemi duraklatıldı!");
        timer1.Stop();
        return;
    }
}
else
{
    lstDrugGroupUrl.SelectedIndex++;
}
}
else
{
    MessageBox.Show("Data çekme işlemi bitti");
    timer1.Stop();
    return;
}
}

txtDrugGroupOrder.Text = lstDrugGroupUrl.SelectedIndex.ToString();
webBrowser.Navigate(lstDrugGroupUrl.SelectedItem.ToString());
txtCurUrl.Text = lstDrugGroupUrl.SelectedItem.ToString();
currentMethod = "parseHtml1";
timer1.Start();
}

private void parseHtml1()
```

```

{
    String s;
    lstDrugName.Items.Clear();
    lstDrugUrl.Items.Clear();
    var links = webBrowser.Document.GetElementsByTagName("a");
    foreach (HtmlElement link in links)
    {
        s = link.GetAttribute("data-metrics-link");
        if (link.GetAttribute("data-metrics-link") == "link")
        {
            lstDrugUrl.Items.Add(link.GetAttribute("href"));
            lstDrugName.Items.Add(link.OuterText);
        }
    }
    txtTotalDrugCount.Text = lstDrugUrl.Items.Count.ToString();
    currentMethod = "callUrl2";
    timer1.Start();
}

```

//ilaç detay sayfasını getirir

private void callUrl2()

```

{
    if (int.Parse(txtDrugOrder.Text) > 0 && loadVar2 == false)
    {
        lstDrugUrl.SelectedIndex = int.Parse(txtDrugOrder.Text);
        lstDrugName.SelectedIndex = lstDrugUrl.SelectedIndex;
        loadVar2 = true;
    }
    else
    {

```

```

        if (lstDrugUrl.Items.Count > 0 && lstDrugUrl.SelectedIndex <
lstDrugUrl.Items.Count - 1)
        {
            lstDrugUrl.SelectedIndex++;
            lstDrugName.SelectedIndex = lstDrugUrl.SelectedIndex;
        }
        else
        {
            currentMethod = "callUrl1";
            timer1.Start();
            return;
        }
    }
    txtDrugOrder.Text = lstDrugUrl.SelectedIndex.ToString();
    webBrowser.Navigate(lstDrugUrl.SelectedItem.ToString());
    txtCurUrl.Text = lstDrugUrl.SelectedItem.ToString();
    drugName = lstDrugName.SelectedItem.ToString();
    if (!chkTest.Checked)
        sendDrugDB();
    currentMethod = "parseHtml2";
    timer1.Start();
}

```

```

private void parseHtml2()
{
    Boolean isComment = false;
    lstCommentUrl.Items.Clear();
    var links = webBrowser.Document.GetElementsByTagName("a");
    foreach (HtmlElement link in links)
    {

```

```
if (link.GetAttribute("data-metrics-module") == "dr-seereviews")
{
    isComment = true;
    strUrl = link.GetAttribute("href");
    tmp = link.OuterText;
    tmp = tmp.Replace("Read Reviews (", "");
    tmp = tmp.Replace(")", "");
    if (tmp != "")
        commentCount = int.Parse(tmp);
    else
        commentCount = 0;
    break;
}
}

if (!isComment)
{
    links = webBrowser.Document.GetElementsByTagName("span");
    foreach (HtmlElement link in links)
    {
        if (link.GetAttribute("className") == "totalreviews")
        {
            isComment = true;
            strUrl = lstDrugUrl.SelectedItem.ToString();
            tmp = link.OuterText;
            tmp = tmp.Replace(" Total User Reviews", "");
            if (tmp != "")
                commentCount = int.Parse(tmp);
            else
                commentCount = 0;
        }
    }
}
```

```
        break;
    }
}
}
if (isComment)
{
    txtCurUrl.Text = strUrl.ToString();
    if (commentCount > 0)
    {
        for (int i = 0; i <= (commentCount / 5); i++)
        {
            lstCommentUrl.Items.Add(strUrl + "&pageIndex=" + i +
                "&sortBy=3&conditionFilter=-500");
        }
    }
    currentMethod = "callUrl3";
    timer1.Start();

    txtTotalCommentUrlCount.Text = lstCommentUrl.Items.Count.ToString();
}
else
{
    txtNoComment.AppendText(lstDrugName.SelectedItem.ToString() + "\n");
    currentMethod = "callUrl2";
    timer1.Start();
}
}

//ilaç yorum sayfasını getirir
private void callUrl3()
{
```



```
if (chkSkipComment.Checked)
{
    currentMethod = "callUrl2";
    timer1.Start();
    return;
}

if (int.Parse(txtCommentOrder.Text) > 0 && loadVar3 == false)
{
    lstCommentUrl.SelectedIndex = int.Parse(txtCommentOrder.Text);
    loadVar3 = true;
}
else
{
    if (lstCommentUrl.Items.Count > 0 && lstCommentUrl.SelectedIndex <
lstCommentUrl.Items.Count - 1)
    {
        lstCommentUrl.SelectedIndex++;
    }
    else
    {
        currentMethod = "callUrl2";
        timer1.Start();
        return;
    }
}

IlacDatasiHazirla.Memory.FlushMemory();
System.Threading.Thread.Sleep(1000);
txtCommentOrder.Text = lstCommentUrl.SelectedIndex.ToString();
webBrowser.Navigate(lstCommentUrl.SelectedItem.ToString());
```

```
txtCurUrl.Text = lstCommentUrl.SelectedItem.ToString();
currentMethod = "parseHtml3";
timer1.Start();
}
private void parseHtml3()
{
    var links = webBrowser.Document.GetElementsByTagName("div");
    foreach (HtmlElement link in links)
    {
        if (link.GetAttribute("className") == "userPost")
        {
            parseComment(link.OuterHtml);
            if (!chkTest.Checked)
                sendCommentDB();
        }
    }
    currentMethod = "callUrl3";
    timer1.Start();
}
private void timer1_Tick(object sender, EventArgs e)
{
    txtWebBrowserState.Text = webBrowser.ReadyState.ToString();
    if (webBrowser.ReadyState == WebBrowserReadyState.Complete || webCount == 70)
    {
        timer1.Stop();
        webCount = 0;
        switch (currentMethod)
        {
            case "parseHtml1":
                parseHtml1();
```

```
        break;
    case "parseHtml2":
        parseHtml2();
        break;
    case "parseHtml3":
        parseHtml3();
        break;
    case "callUrl1":
        callUrl1();
        break;
    case "callUrl2":
        callUrl2();
        break;
    case "callUrl3":
        callUrl3();
        break;
    }
} else if (webBrowser.ReadyState == WebBrowserReadyState.Interactive)
{
    webCount++;
}
if (webCount == 100)
    webBrowser.Refresh();
}
public void sendCommentDB()
{
    var request = (HttpWebRequest)WebRequest.Create("http://www.turk-
park.xyz/api/comment");
    var postData = "drug_name="+drugName;
    postData += "&comment=" + strComment;
```

```

postData += "&condition=" + strCondition;
postData += "&date=" + strDate;
postData += "&reviewer=" + strReviewer;
postData += "&star1=" + strStar1;
postData += "&star2=" + strStar2;
postData += "&star3=" + strStar3;
postData += "&source=" + lstCommentUrl.SelectedItem.ToString();
var data = Encoding.ASCII.GetBytes(postData);
request.Method = "POST";
request.ContentType = "application/x-www-form-urlencoded";
request.ContentLength = data.Length;
using (var stream = request.GetRequestStream())
{
    stream.Write(data, 0, data.Length);
}
try
{
    var response = (HttpWebResponse)request.GetResponse();
    var responseString = new
StreamReader(response.GetResponseStream()).ReadToEnd();
}
catch(Exception e)
{
    MessageBox.Show(e.Message);
}
sendCommentCount++;
txtCommentCount.Text = sendCommentCount.ToString();
}
public void sendDrugDB()
{

```

```
var request = (HttpWebRequest)WebRequest.Create("http://www.turk-
park.xyz/api/drug");
var postData = "drug_name=" + drugName;
var data = Encoding.ASCII.GetBytes(postData);
request.Method = "POST";
request.ContentType = "application/x-www-form-urlencoded";
request.ContentLength = data.Length;
using (var stream = request.GetRequestStream())
{
    stream.Write(data, 0, data.Length);
}
try
{
    var response = (HttpWebResponse)request.GetResponse();
    var responseString = new
StreamReader(response.GetResponseStream()).ReadToEnd();
}
catch (Exception e)
{
    MessageBox.Show(e.Message);
}
txtDrugCount.Text = (int.Parse(txtDrugCount.Text) + 1).ToString();
}
public void parseComment(String commentHtml)
{
    string[] s1 = commentHtml.Split('>');
    if (isOnlyComment) // sadece yorumları çekeceğimiz data ise
    {
        Console.WriteLine(s1[4]);
        Console.WriteLine(s1[7]);
    }
}
```

```

Console.WriteLine(s1[10]);
Console.WriteLine(s1[19]);
Console.WriteLine(s1[27]);
Console.WriteLine(s1[35]);
Console.WriteLine(s1[43]);
strCondition = s1[4].Replace("</span", "");
strDate = s1[7].Replace("</div", "");
strReviewer = s1[10].Replace("Reviewer: ", "").Replace("</p", "");
strStar1 = s1[19].Replace("\n\t\t\tCurrent Rating: ", "").Replace("</span", "");
strStar2 = s1[27].Replace("\n\t\t\tCurrent Rating: ", "").Replace("</span", "");
strStar3 = s1[35].Replace("\n\t\t\tCurrent Rating: ", "").Replace("</span", "");
if (s1[46] == "Hide Full Comment</a")
    strComment = s1[43].Replace("<br", "").Replace("Comment:</strong", "");
else
    strComment = s1[48].Replace("<br", "").Replace("Comment:</strong", "");
}
else
{
    Console.WriteLine(s1[3]); //Condition "\n\t\t\t\tCondition: Other</div"
    Console.WriteLine(s1[5]); //date "8/30/2016 3:26:53 AM</div"
    Console.WriteLine(s1[8]); //revier "Reviewer: </p"
    Console.WriteLine(s1[15]); //1. yıldız "\n\t\t\t\tCurrent Rating: 2</span"
    Console.WriteLine(s1[23]); //2. yıldız "\n\t\t\t\tCurrent Rating: 3</span"
    Console.WriteLine(s1[31]); //3. yıldız "\n\t\t\t\tCurrent Rating: 3</span"
    Console.WriteLine(s1[37]); //comment
    strCondition = s1[3].Replace("\n\t\t\t\t\tCondition: ", "").Replace("</div", "");
    strDate = s1[5].Replace("</div", "");
    strReviewer = s1[8].Replace("Reviewer: ", "").Replace("</p", "");
    strStar1 = s1[15].Replace("\n\t\t\t\tCurrent Rating: ", "").Replace("</span", "");
    strStar2 = s1[23].Replace("\n\t\t\t\tCurrent Rating: ", "").Replace("</span", "");

```

```

strStar3 = s1[31].Replace("\n\t\t\tCurrent Rating: ", "").Replace("</span", "");
if (s1[46] == "Hide Full Comment</a")
    strComment = s1[44].Replace("<br", "").Replace("Comment:</strong", "");
else
    strComment = s1[46].Replace("<br", "").Replace("Comment:</strong", "");
}
drugName = lstDrugName.SelectedItem.ToString();
txtODrugName.Text = drugName;
txtOCondition.Text = strCondition;
txtODate.Text = strDate;
txtOReviewer.Text = strReviewer;
txtOStar1.Text = strStar1;
txtOStar2.Text = strStar2;
txtOStar3.Text = strStar3;
txtOComment.Text = strComment;
}
private void button1_Click(object sender, EventArgs e)
{
    timer1.Stop();
    button1.Enabled = false;
    button2.Enabled = true;
}
private void button2_Click(object sender, EventArgs e)
{
    timer1.Start();
    button1.Enabled = true;
    button2.Enabled = false;
}
}
}

```

EK B

İLAC KULLANICI BİLGİLERİN NİTELİK OLARAK AYRILMA İŞLEMLERİNİN KODLARI

```
public function validate($data){
    $err_msg = "";
    //todo: buraya kurallar eklenecek
    if (!$data['drug_name'])
        $err_msg .= 'İlaç adı boş olamaz!';
    if ($err_msg){
        CdrLib::addAlert("error",$err_msg);
        return false;
    }

    return true;
}

public function parse_all_reviewers(){
    global $db;

    $commentModel = new Comment();
    $commentList = $this->getCommentList('Y', 5000, true);

    foreach($commentList as $c){
        $data = $this->parse_reviewer($c['reviewer']);

        if ($data)
            $commentModel->editComment($c['id'], $data);
    }
}
```



```
}

```

```
public function parse_reviewer($str){

```

```
    $data = array();

```

```
    $arrAgeRanges = array('0-2','3-6','7-12','13-18','19-24','25-34','35-44','45-54','55-64','65-74','75');

```

```
    $arrTimeRanges = array('less than 1 month','1 to 6 months','6 months to less than 1 year','1 to less than 2 years','2 to less than 5 years','5 to less than 10 years','10 years or more');

```

```
    //cinsiyet

```

```
    $index = strpos($str, " Male ");

```

```
    if ($index !== false)

```

```
        $gender = "Male";

```

```
    $index = strpos($str, " Female ");

```

```
    if ($index !== false)

```

```
        $gender = "Female";

```

```
    $str = str_replace($gender,"",$str);

```

```
    //yaş

```

```
    foreach($arrAgeRanges as $age){

```

```
        if ($age == "75")

```

```
            $age = "75 or over";

```

```
        $index = strpos($str, $age." ");

```

```
        if ($index !== false)

```

```
        {

```

```
            $age_range = $age;

```

```
            $str = str_replace($age.' ','',$str);

```

```
            break;

```

```
        }

```

```
    }

```

```
    //kullanıcı türü

```

```

$index = strpos($str, "(Patient)");
if ($index !== false)
    $user_type = "Patient";
$index = strpos($str, "(Caregiver)");
if ($index !== false)
    $user_type = "Caregiver";
$str = str_replace("($user_type.)", "", $str);
//zaman aralığı
foreach($arrTimeRanges as $time){
    $strTime = "on Treatment for ".$time;
    $index = strpos($str, $strTime);
    if ($index !== false)
    {
        $time_range = $time;
        $str = str_replace($strTime, "", $str);
        break;
    }
}
$str = str_replace(' ', "", $str);
$str = trim($str);
$screen_name = $str;
$data = array(
    'user_type' => $user_type,
    'gender' => $gender,
    'age_range' => $age_range,
    'time_range' => $time_range,
    'screen_name' => $screen_name
);
return $data;
}

```

EK C

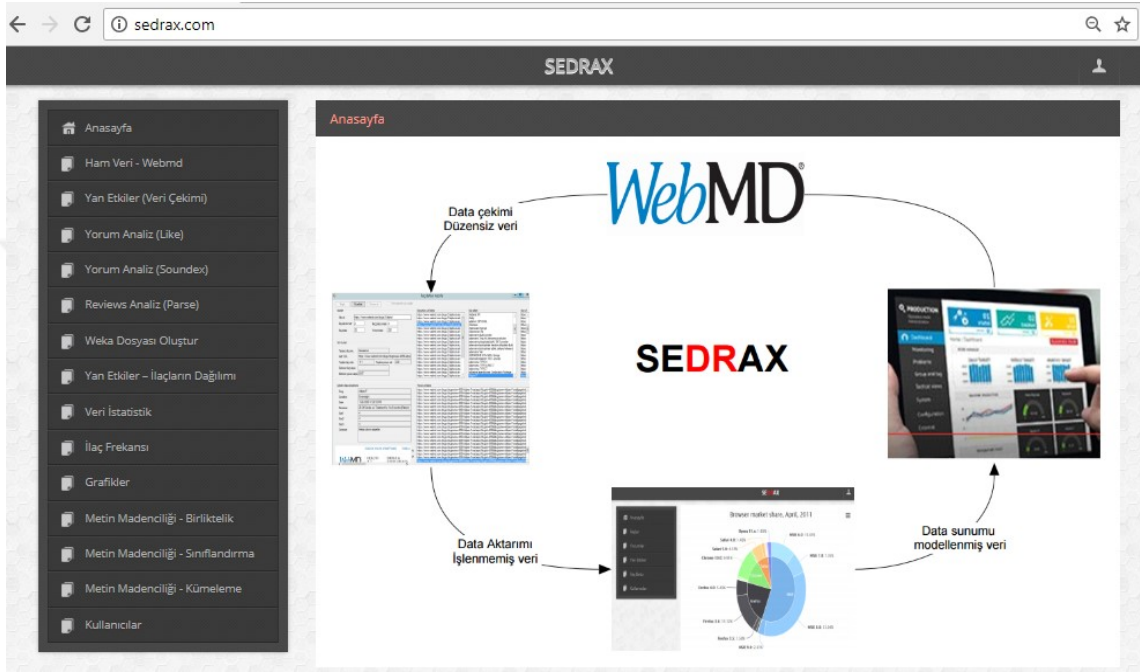
OLUŞTURULAN VERİ SETİ YAPISI

<input type="checkbox"/> i_comments <input type="checkbox"/> id <input type="checkbox"/> drug_id <input type="checkbox"/> drug_name <input type="checkbox"/> comment <input type="checkbox"/> comment_tr <input type="checkbox"/> condition <input type="checkbox"/> date <input type="checkbox"/> reviewer <input type="checkbox"/> effectiveness <input type="checkbox"/> easeofuse <input type="checkbox"/> satisfaction <input type="checkbox"/> user_type <input type="checkbox"/> gender <input type="checkbox"/> age_range <input type="checkbox"/> time_range <input type="checkbox"/> screen_name <input type="checkbox"/> like_count <input type="checkbox"/> is_read <input type="checkbox"/> is_reviewer_parse <input type="checkbox"/> is_stanford_nlp <input type="checkbox"/> soundex_count <input type="checkbox"/> source <input type="checkbox"/> status <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_age <input type="checkbox"/> id <input type="checkbox"/> age_range <input type="checkbox"/> age_range_partner <input type="checkbox"/> age_range_partner_genel <input type="checkbox"/> i_side_effects <input type="checkbox"/> id <input type="checkbox"/> name <input type="checkbox"/> name_tr <input type="checkbox"/> partner <input type="checkbox"/> drug_count <input type="checkbox"/> accept_method <input type="checkbox"/> status <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_time_use <input type="checkbox"/> id <input type="checkbox"/> time_range <input type="checkbox"/> time_range_partner <input type="checkbox"/> i_druqs <input type="checkbox"/> id <input type="checkbox"/> name <input type="checkbox"/> status <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_gender <input type="checkbox"/> id <input type="checkbox"/> gender <input type="checkbox"/> gender_partner <input type="checkbox"/> i_filters <input type="checkbox"/> id <input type="checkbox"/> type <input type="checkbox"/> name <input type="checkbox"/> ord <input type="checkbox"/> status <input type="checkbox"/> created_at <input type="checkbox"/> updated_at
	<input type="checkbox"/> i_comment_relations_soundex <input type="checkbox"/> id <input type="checkbox"/> comment_id <input type="checkbox"/> drug_id <input type="checkbox"/> se_id <input type="checkbox"/> word <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_comment_relations <input type="checkbox"/> id <input type="checkbox"/> comment_id <input type="checkbox"/> drug_id <input type="checkbox"/> se_id <input type="checkbox"/> se_count <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	
<input type="checkbox"/> i_users <input type="checkbox"/> id <input type="checkbox"/> role <input type="checkbox"/> name <input type="checkbox"/> email <input type="checkbox"/> username <input type="checkbox"/> password <input type="checkbox"/> status <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_comment_rel <input type="checkbox"/> id <input type="checkbox"/> comment_id <input type="checkbox"/> drug_id <input type="checkbox"/> se_id <input type="checkbox"/> se_count <input type="checkbox"/> accept_method <input type="checkbox"/> created_at <input type="checkbox"/> updated_at	<input type="checkbox"/> i_knn_sonuc <input type="checkbox"/> qkey <input type="checkbox"/> id <input type="checkbox"/> code <input type="checkbox"/> se <input type="checkbox"/> isSelected <input type="checkbox"/> distance <input type="checkbox"/> distance2	<input type="checkbox"/> apriori_sonuc <input type="checkbox"/> qkey <input type="checkbox"/> description <input type="checkbox"/> ord1 <input type="checkbox"/> ord2 <input type="checkbox"/> ord3 <input type="checkbox"/> i_user_type <input type="checkbox"/> id <input type="checkbox"/> user_type <input type="checkbox"/> user_type_partner
	<input type="checkbox"/> i_commmnet_condition <input type="checkbox"/> id <input type="checkbox"/> name		

EK P1

OLUŞTURULAN WEB PANEL

ANASAYFA



EK P2

WEB PANEL HAM VERİ LİSTESİ

ID	İlaç Adı	Yorum	condition	effectiveness	easeofuse	satisfaction
18	Yasmin 28	this my Abasia first time using contraceptve, on my first 2 wk, i have bigmentation on the legs itching on the growin area	Birth Control	1	1	1
19	Yasmin 28	I recently ordered yasmin since I know it works like wonder for birth control. I ordered it from buybirthcontrolpills dot com.	Birth Control	4	5	5
20	Yasmin 28	I was put on Yasmin for a combination of birth control, acne and extremely painful, heavy and unpredictable periods. It is the only type of contraceptive pill I've been on so I can't compare it to others but what I've experienced with Yasmin has been great. My acne has almost gone and I no longer have any pain whatsoever when on my period, which is a huge change from before, and although it is still quite heavy it isn't as bad as before and I know exactly when it is going to start and end. I can even skip a period when I'm on holiday! Really satisfied with Yasmin and hopefully will continue to use it for many more years.	Birth Control	5	5	5
21	Yasmin 28	I went on Yasmin because I had incredibly painful cramps, and I no longer have any cramps what so ever when on Yasmin. It has worked great for me, the only thing that I have noticed is that I'm loosing more hair than what normally falls out when brushing it, etc. I believe this was caused by the pill because I am young	Birth Control	5	5	4

EK P3

WEB PANEL YORUMLARDAN YAN ETKİ BULMA “LIKE”

BOT Çalıştır

Botu çalıştırmak için aşağıdaki başlat butonuna basınız.
Bu işlem bir kaç saat sürebilir! [Başlat](#)

Filtre

İlaç: Ekle

İlaç	ID	Sil
Yasmin 28	4	Sil

[Listele](#)

Like İle Yan Etki Çıkarımı

Show: entries Search:

ID	İlaç Adı	Yanetki Adı	Adet
1	Yasmin 28	Abasia	1
2	Yasmin 28	Acne	2
3	Yasmin 28	Pain	2

EK P4

WEB PANEL YORUMLARDAN YAN ETKİ BULMA “SOUNDEX”

BOT Çalıştır

Botu çalıştırmak için aşağıdaki başlat butonuna basınız.
Bu işlem bir kaç saat sürebilir! [Başlat](#)

Filtre

İlaç: Ekle

İlaç	ID	Sil
Yasmin 28	4	Sil

[Listele](#)

Soundex İle Yan Etki Çıkarımı

Show: entries Search:

ID	İlaç Adı	Yanetki Adı	Adet
1	Yasmin 28	Abasia	1
2	Yasmin 28	Acne	18
4	Yasmin 28	Pain	13
5	Yasmin 28	Anxiety	18

EK P5

WEB PANEL KULLANICI BİLGİLERİNİN AYRIŞTIRILMASI

SED RAX

Reviews Parse Çalıştır

Botu çalıştırmak için aşağıdaki başlat butonuna basınız.
Bu işlem bir kaç saat sürebilir! [Başlat](#)

Reviews Parse İşlemi

Show 10 entries Search:

ID	İlaç Adı	user_type	gender	age_range	time_range
22291	hydrocodone-acetaminophen	Patient	Female	65-74	1 to 6 months
22292	hydrocodone-acetaminophen	Patient	Female	35-44	less than 1 month
22293	hydrocodone-acetaminophen	Patient	Male	65-74	less than 1 month
22294	hydrocodone-acetaminophen				
22296	hydrocodone-acetaminophen	Patient		35-44	less than 1 month
22300	hydrocodone-acetaminophen	Patient	Female	65-74	5 to less than 10 years
22301	hydrocodone-acetaminophen	Patient	Male	55-64	1 to less than 2 years
22302	hydrocodone-acetaminophen	Patient		45-54	less than 1 month
22303	hydrocodone-acetaminophen	Patient		65-74	2 to less than 5 years
22304	hydrocodone-acetaminophen	Patient	Female	13-18	less than 1 month

EK P6

WEB PANEL İLAÇLARDA ÇIKAN YAN ETKİ LİSTESİ

SED RAX

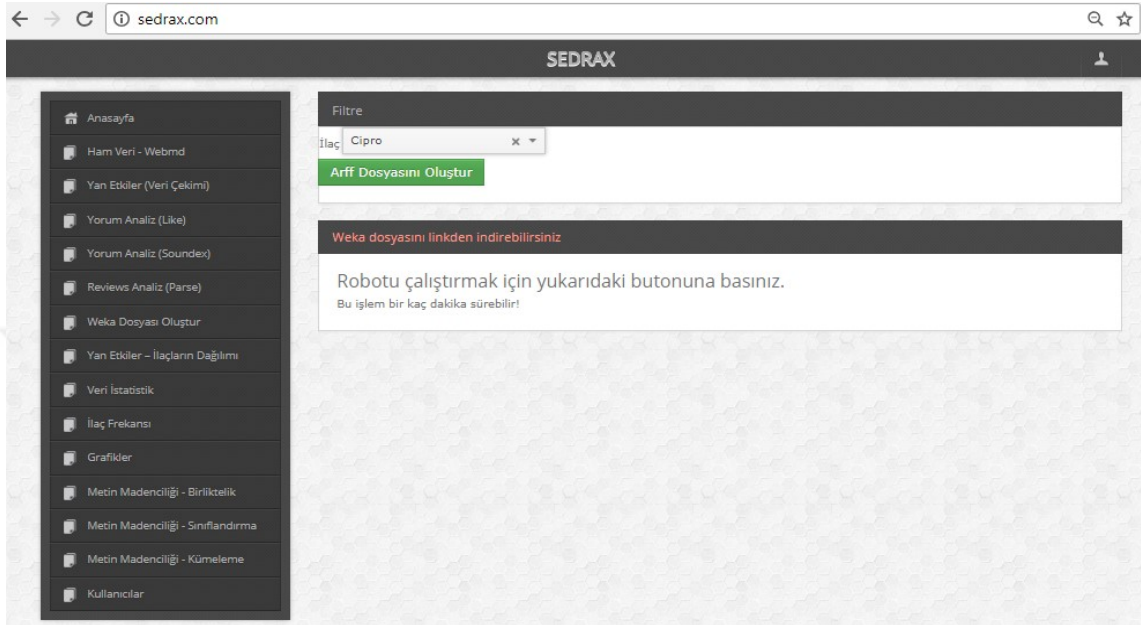
Yan Etki Listesi

Show 25 entries Search:

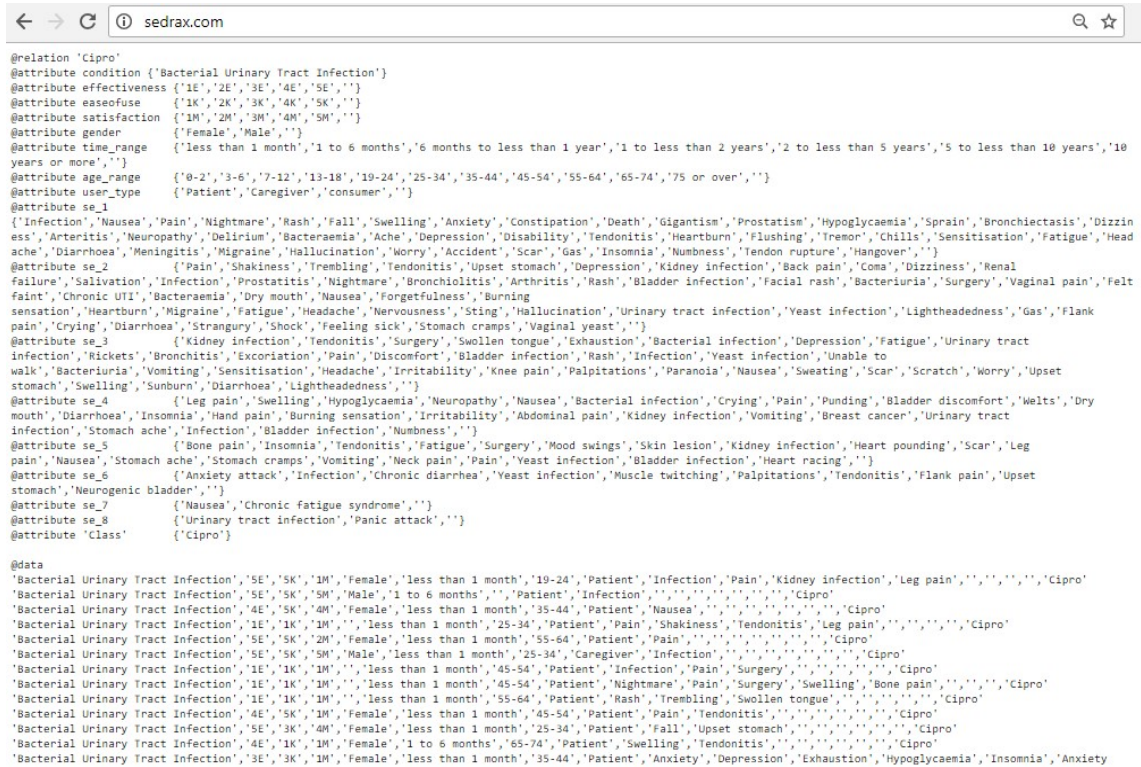
ID	Yan Etki Adı	İlaç Sayısı
4765	Pain	2469
4356	Nausea	1332
2959	Headache	1104
3416	Infection	1035
2000	Dizziness	1029
6277	Surgery	975
6289	Swelling	968
5537	Rash	944
1936	Diarrhoea	915
1862	Depression	870
2491	Fatigue	839
4134	Migraine	835
455	Anxiety	820
2843	Gout	816
6956	Vomiting	787

EK P7

WEB PANEL İLAÇLARIN WEKA DOSYASINA AKTARILMASI



WEB PANEL WEKA DOSYASININ İÇERİK GÖRÜNTÜSÜ



EK P8

WEB PANEL YAN ETKİ LİSTESİ

SED RAX

Yan etkileri çek

Botu çalıştırmak için aşağıdaki başlat butonuna basınız.
Bu işlem bir kaç saat sürebilir! [Başlat](#)

Yan Etki Ham Veri

Show 10 entries Search:

ID	Yan Etki Adı
1	5'nucleotidase increased
2	Abasia
3	Abdominal abscess
4	Abdominal adhesions
5	Abdominal aortic aneurysm
6	Abdominal aortic bruit
7	Abdominal aortic rupture
8	Abdominal bloating
9	Abdominal colic
10	Abdominal cramps

Showing 1 to 10 of 7,058 entries

First Previous 1 2 3 4 5 Next Last

EK P9

WEB PANEL VERİ İSTATİSTİK LİSTESİ

SED RAX

Data Frekansı

Cinsiyet Dağılımı	145402 (% 65,69) Female 59893 (% 27,06) Male 16064 (% 7,26) Bilinmeyen
Yaş Dağılımı	45-54 (48804) 55-64 (44549) 35-44 (33495) 25-34 (30971) 65-74 (23669) 19-24 (15532) Bilinmeyen (9890) 75 or over (8454) 13-18 (4239) 7-12 (1058) 3-6 (501) 0-2 (197)
Kullanım Süresi	less than 1 month (63976) 1 to 6 months (53019) 2 to less than 5 years (26567) 1 to less than 2 years (20943) 6 months to less than 1 year (20202) 5 to less than 10 years (13979) Bilinmeyen (13716) 10 years or more (8957)
En iyi ilaçlar Etki - Memnuniyet - Kullanım Kolaylığı	phentermine HCL (1010) hydrocodone-acetaminophen (919)

EK P10

WEB PANEL İLAÇ FREKANS LİSTESİ

The screenshot shows the SEDRAX web panel interface. The browser address bar displays 'sedrax.com'. The main content area is titled 'SED RAX' and features a sidebar on the left with various navigation options. The main area is divided into a filter section and a table of drug frequencies.

Filter Section:

- İlaç: Seçiniz
- Yan Etki: Hepsi
- Kullanım Süresi: Hepsi
- Yaş Aralığı: Hepsi
- Cinsiyet: Hepsi
- Tip: Hepsi
- Ekle
- Durum: Hepsi
- Etki: Hepsi
- Kullanım Kolaylığı: Hepsi
- Memnuniyet: Hepsi

Table of Drug Frequencies:

İlaç	Durum	Yorum Sayısı	Sil
Jadenu	Too Much Iron in the Blood fo	4	Sil

Yanetki Frekansı Table:

Yanetki	İlaç (Yorum Sayısı)	Tekrar Sayısı
- Anaemia	1 (1)	2
	Jadenu (1)	2
+ Chronic iron overload	1 (1)	1
+ Iron overload	1 (1)	1

EK P11

WEB PANEL İLAÇ FREKANS İÇERİK GÖRÜNTÜSÜ

The screenshot shows the SEDRAX web panel interface displaying the content view of a drug frequency entry. The browser address bar displays 'tr/comment/all/19/343'. The main content area is titled 'SED RAX' and features a search bar and a table of drug frequencies.

Search Bar:

Search:

Table of Drug Frequencies:

ID	İlaç Adı	Yorum
577	Jadenu	I have Sideroblastic Anemia combined with Chronic Iron Overload and was prescribed Jadenu in gut issues, extreme tiredness and muscle weakness caused by skyrocketing liver functions. My t

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

EK P12

WEB PANEL VERİ MADENCİLİĞİ ALGORİTMASI APRIORİ

← → ↻ sedrax.com

SEDRAx

- 🏠 Anasayfa
- 📄 Ham Veri - Webmd
- 📄 Yan Etkiler (Veri Çekimi)
- 📄 Yorum Analiz (Like)
- 📄 Yorum Analiz (Soundex)
- 📄 Reviews Analiz (Parse)
- 📄 Weka Dosyası Oluştur
- 📄 Yan Etkiler - İlaçların Dağılımı
- 📄 Veri İstatistik
- 📄 İlaç Frekansı
- 📄 Grafikler
- 📄 Metin Madenciliği - Birliklilik
- 📄 Metin Madenciliği - Sınıflandırma
- 📄 Metin Madenciliği - Kümeleme
- 📄 Kullanıcılar

Metin Madenciliği - Birliklilik (Apriori)

İlaç :

Kullanım Süresi : Durum :

Yaş Aralığı : Etki :

Cinsiyet : Kullanım Kolaylığı :

Tip : Memnuniyet :

NumRules : //Listelenecek verinin adetinin belirlir

MaxScan : //Maksimum gerçekleştirilen döngü sayısını verir(değer/100)

MinSup : //Minimum destek oranı verir(değer/100)

MinConf : //Minimum güven oranı verir(değer/100)

Listele

Association Rules

Time: 0.01 second(s)

Jakafi 7 ==> Myelofibrosis 7 = supp:(1,00) conf:(1,00) lift:(0,54) lev:(-0,86) conv:(0,00)

Patient 6 ==> Myelofibrosis 6 = supp:(0,86) conf:(1,00) lift:(0,54) lev:(-0,73) conv:(0,00)

Patient 6 ==> Jakafi 6 = supp:(0,86) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)

SK 4 ==> Myelofibrosis 4 = supp:(0,57) conf:(1,00) lift:(0,54) lev:(-0,49) conv:(0,00)

SK 4 ==> Jakafi 4 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)

SE 4 ==> Myelofibrosis 4 = supp:(0,57) conf:(1,00) lift:(0,54) lev:(-0,49) conv:(0,00)

SE 4 ==> Jakafi 4 = supp:(0,57) conf:(1,00) lift:(1,00) lev:(0,00) conv:(0,00)

SM 3 ==> SK 3 = supp:(0,43) conf:(1,00) lift:(1,75) lev:(0,18) conv:(0,00)

SM 3 ==> SE 3 = supp:(0,43) conf:(1,00) lift:(1,75) lev:(0,18) conv:(0,00)

SM 3 ==> Myelofibrosis 3 = supp:(0,43) conf:(1,00) lift:(0,54) lev:(-0,37) conv:(0,00)

WEB PANEL VERİ MADENCİLİĞİ ALGORİTMASI

KNN

← → ↻ sedrax.com

SEDRAx

- 🏠 Anasayfa
- 📄 Ham Veri - Webmd
- 📄 Yan Etkiler (Veri Çekimi)
- 📄 Yorum Analiz (Like)
- 📄 Yorum Analiz (Soundex)
- 📄 Reviews Analiz (Parse)
- 📄 Weka Dosyası Oluştur
- 📄 Yan Etkiler - İlaçların Dağılımı
- 📄 Veri İstatistik
- 📄 İlaç Frekansı
- 📄 Grafikler
- 📄 Metin Madenciliği - Birliklilik
- 📄 Metin Madenciliği - Sınıflandırma
- 📄 Metin Madenciliği - Kümeleme
- 📄 Kullanıcılar

Metin Madenciliği - Sınıflandırma (Knn)

İlaç :

Kullanım Süresi : Durum :

Yaş Aralığı : Etki :

Cinsiyet : Kullanım Kolaylığı :

Tip : Memnuniyet :

K : 1-76

Listele

Mood swings

Time	Age	Gender	Type	Effectiveness	Easeofuse	Satisfaction	Condition	Side Effect	Distance	Normalization Distance
less than 1 month	19-24	Female	Patient	5	5	3	Birth Control	Mood swings	0	0,00
2 to less than 5 years	19-24	Female	Patient	5	5	3	Birth Control	Migraine	1	0,67
less than 1 month	19-24	Female	Patient	4	5	5	Birth Control	Fabigue	2	0,00
1 to 6 months	19-24	Female	Patient	5	5	2	Birth Control	Acne	2	0,17
1 to 6 months	19-24	Female	Patient	5	4	3	Birth Control	Pregnancy	2	0,17

WEB PANEL VERİ MADENCİLİĞİ ALGORİTMASI KMEANS

← → ↻ sedrax.com

SEDRAX

Metin Madenciliği - Kümeleme (Kmeans)

İlaç : Yan Etki :

Kullanım Süresi : Durum :

Yaş Aralığı : Etki :

Cinsiyet : Kullanım Kolaylığı :

Tip : Memnuniyet :

X : Y :

K (Grup Adedi) :

Clustered Instances

Grup	Adet	Oran
0	4	66,67 %
1	1	16,67 %
2	1	16,67 %

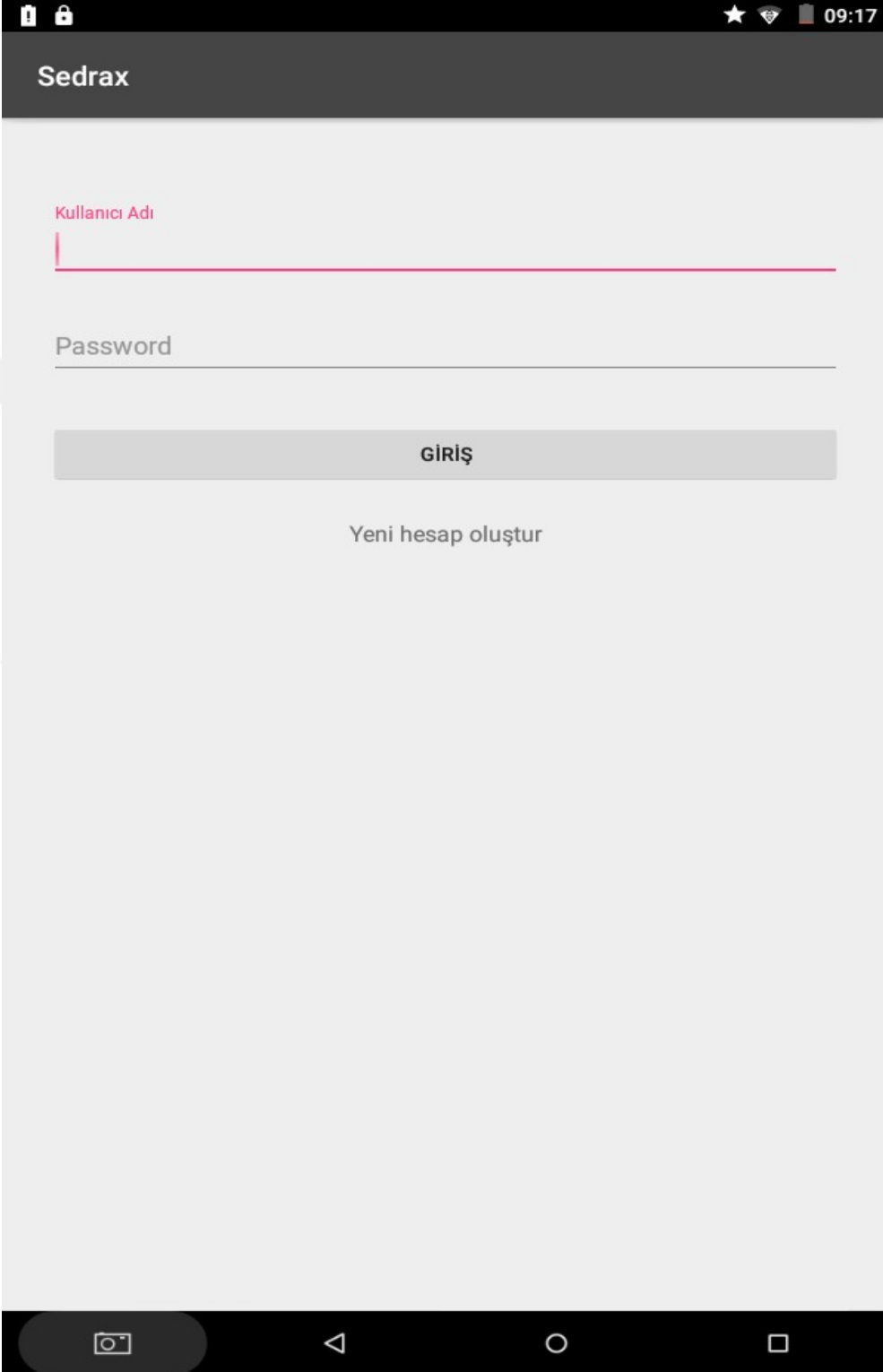
Cluster 0: Myelofibrosis,,35-44,6 months to less than 1 year,Patient,4E,4K,4M
Cluster 1: Myelofibrosis,,75 or over,6 months to less than 1 year,Patient,4E,4K,4M
Cluster 2: Myelofibrosis,Female,75 or over,less than 1 month,Caregiver,5E,5K,5M

Grup	ID	condition	gender	age_range	time_range	user_type	eff_st	eas_st	sat_st	Yan Etki
0	578	Myelofibrosis		35-44	6 months to less than 1 year	Patient	4	4	4	Myelofibrosis,Pain

Anasayfa
Ham Veri - Webmd
Yan Etkiler (Veri Çekimi)
Yorum Analiz (Like)
Yorum Analiz (Soundex)
Reviews Analiz (Parse)
Weka Dosyası Oluştur
Yan Etkiler - İlaçların Dağılımı
Veri İstatistik
İlaç Frekansı
Grafikler
Metin Madenciliği - Birliklilik
Metin Madenciliği - Sınıflandırma
Metin Madenciliği - Kümeleme
Kullanıcılar

EK U1

MOBİL UYGULAMAYA GİRİŞ



The image shows a mobile application login screen for 'Sedrax'. The screen has a dark grey header with the app name 'Sedrax' in white. Below the header, there are two input fields: 'Kullanıcı Adı' (Username) and 'Password'. The 'Kullanıcı Adı' field has a red underline, and the 'Password' field has a grey underline. Below the input fields is a grey button labeled 'GİRİŞ' (Login). Underneath the button is a link that says 'Yeni hesap oluştur' (Create new account). The screen also shows a status bar at the top with a lock icon, a star icon, a Wi-Fi icon, and the time '09:17'. At the bottom, there is a navigation bar with icons for home, back, and app drawer.

Sedrax

Kullanıcı Adı

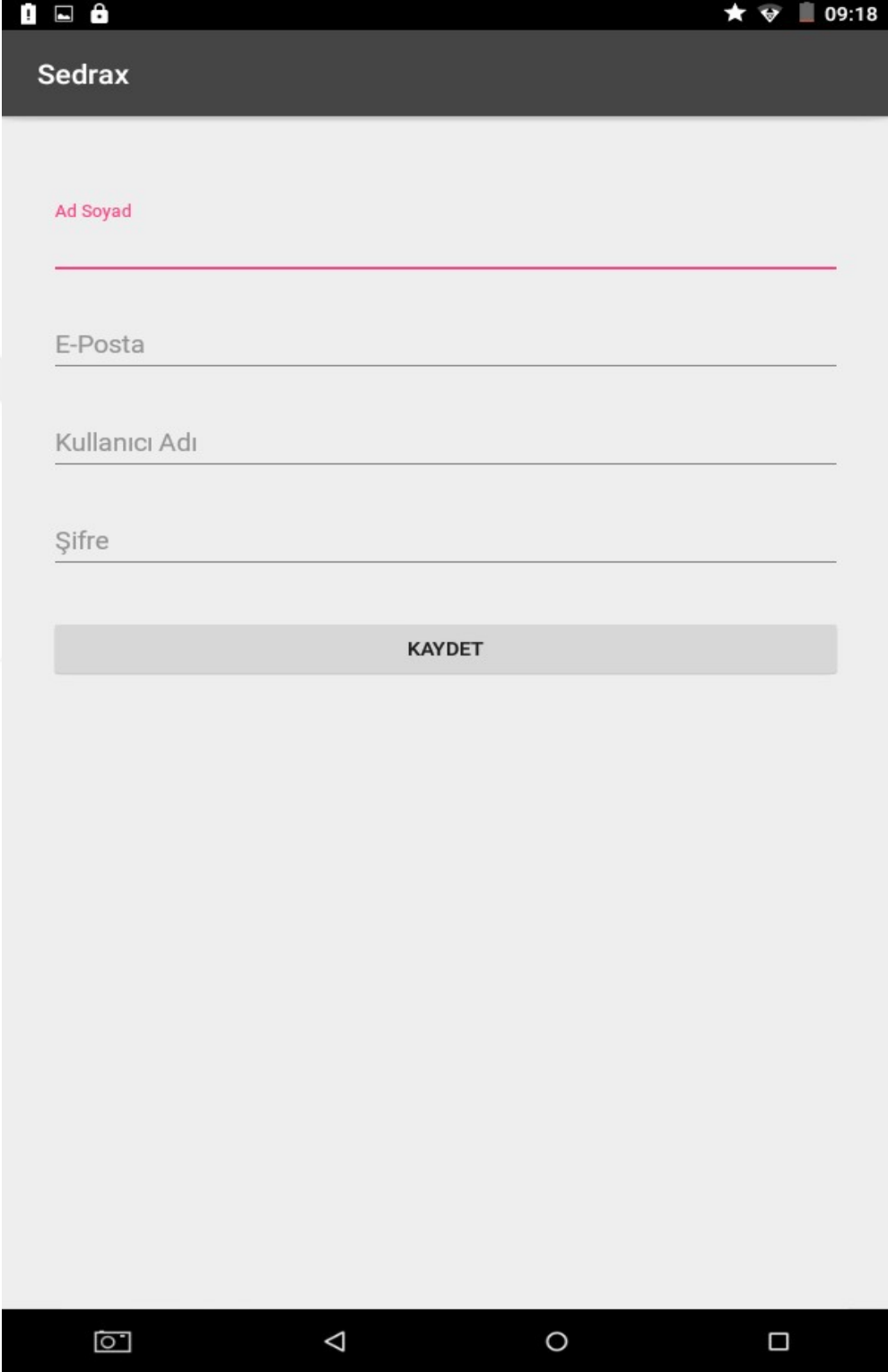
Password

GİRİŞ

Yeni hesap oluştur

EK U2

MOBİL UYGULAMAYA KAYIT OLMA



The image shows a mobile application registration screen. At the top, there is a dark header bar with the text 'Sedrax' on the left and system icons (star, Wi-Fi, battery, and time 09:18) on the right. Below the header, the form consists of four input fields: 'Ad Soyad' (Name and Surname) with a pink underline, 'E-Posta' (Email) with a grey underline, 'Kullanıcı Adı' (Username) with a grey underline, and 'Şifre' (Password) with a grey underline. A grey button labeled 'KAYDET' (Save) is positioned below the password field. At the bottom of the screen, there is a black navigation bar with three icons: a camera, a back arrow, and a square.

EK U3

MOBİL UYGULAMA ANASAYFASI



ÖZGEÇMİŞ

Ad Soyad: Alkan Kaya,

Alkan Kaya 1988' de Kahramanmaraş Elbistan'da doğdu. İlkokulu Yukarıpınar'da, Ortaokulu Necatibey' de, liseyi Mehmetçik Lisesi'nde okudu. 2005' de Adıyaman Üniversitesi Gölbaşı Meslek Yüksek Okulunun da Bilgisayar Teknolojisi ve Programcılığı bölümünde ön lisans öğrenimini 2007 yılında tamamladı. 2009'da Anadolu Üniversitesi İşletme bölümünde başladığı lisans öğrenimini 2012 yılında tamamladı. Okan Üniversitesi Bilgisayar Mühendisliği programında tezli yüksek lisan öğrenimine 2014 yılında başladı. Halen RMG Petrol Ürünleri Dağıtım Sanayi A.Ş. firmasında otomasyon ve bilgi işlem müdürü olarak çalışmalarını sürdürmektedir.

FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI



SAĞLIKLA İLGİLİ SOSYAL MEDYA FORUM VERİLERİNDE İÇERİĞİNİN
BİLGİLENDİRİCİ ÖZELLİKLERİ:
İLAÇ YAN ETKİLERİNİN ARAŞTIRMASI ÜZERİNE BİR ÇALIŞMA

YÜKSEK LİSANS TEZİ

ALKAN KAYA

tarafından

YÜKSEK LİSANS

derecesi şartını sağlamak için hazırlanmıştır.

Haziran 2018

Program: Bilgisayar Mühendisliği Yüksek Lisans