

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ

**METİN MADENCİLİK YÖNTEMLERİLE ÜRÜN
YORUMLARININ OTOMATİK
DEĞERLENDİRİLMESİ**

DOKTORA TEZİ

Endüstri Yük. Müh. Kadriye ERGÜN

Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ
Tez Danışmanları : Doç. Dr. Cemalettin KUBAT
Yrd. Doç. Dr. Gültekin ÇAĞIL

Ekim 2012

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**METİN MADENCİLİĞİ YÖNTEMLERİ İLE ÜRÜN
YORUMLARININ OTOMATİK
DEĞERLENDİRİLMESİ**

DOKTORA TEZİ

Endüstri Yük. Müh. Kadriye ERGÜN

Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Bu tez **13/02/2012** tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.

Doç. Dr.
Cemalettin KUBAT
Jüri Başkanı

Yrd. Doç. Dr.
Bayram TOPAL
Üye

Prof. Dr.
Ramazan YAMAN
Üye

R. Yaman

Prof. Dr.
Nejat YUMUŞAK
Üye

Yrd. Doç. Dr.
Adem GÖLEÇ
Üye

A. Göleç

TE EKKÜR

Tezin en zor kısmı olan çalı ma konusunun tespitinden itibaren de erli bilgilerini benimle payla an ve beni yönlendiren danı man hocalarım Doç Dr. Cemalettin KUBAT ve Yrd. Doç.Dr. Gültekin ÇA IL'a te ekkürlerimi sunarım.

Çalı ma sürecinde de erli ele tiri ve önerileriyle katkı sa layan tez izleme jürisi hocalarım Prof. Dr. Nejat YUMU AK ve Yrd. Doç. Dr. Bayram TOPAL'a te ekkür eder, saygılarımı sunarım.

Tez kapsamında yazılım geli tirme ihtiyacını kar ılamak üzere yardımını hiç esirgemeyen Ar . Gör. Ra it CESUR'a te ekkürü bir borç bilirim.

Yine çalı mam boyunca moral desteklerini hiç azaltmadan devam ettiren her artta yanımda olan aileme te ekkür eder, saygılarımı sunarım.

Kadriye ERGÜN

28.12.2011

Ç İNDEK İLER

| | |
|-------------------------|------|
| TE EK KÜR..... | ii |
| Ç İNDEK İLER | iii |
| EK İLLER İLİSTESİ | v |
| TABLOLAR İLİSTESİ | vi |
| ÖZET..... | vii |
| SUMMARY..... | viii |

BÖLÜM 1.

| | |
|----------------------------------|----|
| G İRİ | 1 |
| 1.1. Tezin Amacı..... | 3 |
| 1.2. Çalışmanın Gerekçeleri..... | 4 |
| 1.3. Tezin Kapsamı..... | 10 |

BÖLÜM 2.

| | |
|---|----|
| METİN MADENCİLİK VE İLK İLDE İSİMLER..... | 13 |
| 2.1. Giriş | 13 |
| 2.2. Veri Madenciliği (Data Mining)..... | 14 |
| 2.2.1. Veri madenciliği yöntemleri..... | 16 |
| 2.2.1.1. Sınıflandırma (Classification) | 17 |
| 2.2.1.2. Kümeleme (Clustering) | 18 |
| 2.2.1.3. Birliktelik kuralları (Associaton rules) | 19 |
| 2.2.2. Veri madenciliğinden metin madenciliğine geçiş | 19 |
| 2.3. Metin Madenciliği (Text Mining)..... | 20 |
| 2.3.1. Bilgiye erişim (Information retrieval)..... | 24 |
| 2.3.2. Bilgi çıkarımı (Information extraction)..... | 28 |
| 2.3.3. Bilgiye erişim ve bilgi çıkarımının karşılaştırılması..... | 32 |

| | |
|--|-----------|
| 2.3.4. Metin madenciliğinin diğer uygulamaları | 34 |
| 2.4. Makine Öğrenmesi (Machine Learning) | 38 |
| 2.5. Doğal Dil İşleme..... | 41 |
| 2.6. Metin Madenciliğinin Adımları..... | 44 |
| 2.6.1. Metin koleksiyonu oluşturma | 45 |
| 2.6.2. Metin ön işleme..... | 45 |
| 2.6.3. Metin dönüştürme..... | 47 |
| 2.6.4. Özellik seçme..... | 50 |
| 2.6.5. Veri madenciliği..... | 51 |
| 2.6.6. Değerlendirme ve yorumlama..... | 51 |
| 2.7. Zekası (Business Intelligence)..... | 53 |
| 2.7.1. Zekası ve metin madenciliği ilişkisi..... | 56 |
| | |
| BÖLÜM 3. | |
| UYGULAMA: METİN MADENCİLİK YÖNTEMLERİLE ÜRÜN | |
| YORUMLARININ OTOMATİK DEĞERLENDİRİLMESİ | 62 |
| 3.1. Giriş | 62 |
| 3.2. Türkçe Dilinin Yapısı | 64 |
| 3.3. Özellik simleri ve Sıfatların Tespiti | 71 |
| 3.4. Açık Yapısı ve Derinlik Öncelikli Arama Algoritması | 77 |
| | |
| BÖLÜM 4. | |
| SONUÇLAR VE ÖNERİLER..... | 82 |
| KAYNAKLAR..... | 86 |
| ÖZGEÇMİŞ | 91 |

EKLER LİSTESİ

| | | |
|-----------|---|----|
| ekil 1.1. | Hane halkı bilişim teknolojileri kullanım oranları..... | 6 |
| ekil 1.2. | Kullanıcı yorumlarının elektronik ortamdaki yeri..... | 7 |
| ekil 2.1. | Veri madenciliği ile ilgili disiplinler..... | 15 |
| ekil 2.2. | Yorumların bilgiye dönüşüm süreci ve elde edilen bilginin paydaşları..... | 24 |
| ekil 2.3. | Metin madenciliğinin adımları..... | 44 |
| ekil 2.4. | Metin madenciliği adımları ve içerdikleri yaklaşımlar..... | 52 |
| ekil 2.5. | Veri madenciliği ve iş zekası..... | 56 |
| ekil 2.6. | İş zekasını çoklu kaynaklı ham metin bilgisinden çıkarmak için metin madenciliğini kullanma süreci..... | 59 |
| ekil 3.1. | Bilgi çıkarımı sistemi akı diyagramı..... | 62 |
| ekil 3.2. | Ürün yorumlarının elde edildiği kaynak..... | 65 |
| ekil 3.3. | Niteleme sonucu durumu | 77 |
| ekil 3.4. | Memnuniyet dereceleri sonucu durumu..... | 78 |
| ekil 3.5. | Nokia E-72'nin memnuniyet derecesi..... | 81 |

TABLolar L STES

| | | |
|-------------|--|----|
| Tablo 2.1. | Veri madencili i uygulama alanları..... | 15 |
| Tablo 2.2. | Veri madencili i ve metin madencili inin kar ıla tırılması..... | 22 |
| Tablo 2.3. | Metin madencili i metotlarının girdi ve çıktıları..... | 34 |
| Tablo 2.4. | Bilgi çıkarımı için kullanılan makine ö renmesi tekniklerinin sınıflandırılması..... | 40 |
| Tablo 2.5. | Metin madencili inin adımları..... | 45 |
| Tablo 3.1. | sim çekim ekleri..... | 63 |
| Tablo 3.2. | Örnek yorumlar..... | 66 |
| Tablo 3.3. | Cümlelerin morfolojik analizinin sonucu örne i..... | 68 |
| Tablo 3.4. | Ço ul eki, iyelik ekleri..... | 69 |
| Tablo 3.5. | ç çekim durum ekleri..... | 69 |
| Tablo 3.6. | Dı çekim durum ekleri, ilgi eki..... | 69 |
| Tablo 3.7. | Zaman ekleri..... | 70 |
| Tablo 3.8. | Ki i eklerine kar ılık gelen zaman ekleri..... | 70 |
| Tablo 3.9. | Eylem çekim ekleri, yapım ekleri..... | 71 |
| Tablo 3.10. | Ürün özellik isimleri ve e veya benzer kelimeler..... | 73 |
| Tablo 3.11. | Niteleme sıfatları puanları..... | 74 |
| Tablo 3.12. | A aç yapısında niteleme kuralları..... | 75 |
| Tablo 3.13. | Nitelenen-niteleyen ili ki puanları..... | 76 |
| Tablo 3.14. | Niteleme durumu..... | 77 |
| Tablo 3.15. | Örnek 1.'e ait 10 adet yorumun de erlendirilmesi | 79 |
| Tablo 3.16. | Örnek 1.'e ait 100 adet yorumun de erlendirilmesi | 79 |
| Tablo 3.17. | Örnek 2.'e ait yorumlar..... | 80 |
| Tablo 3.18. | Örnek 2.'ye ait yorumların be eni dereceleri..... | 81 |

ÖZET

Anahtar Kelimeler: Metin Madenciliği, Bilgi Çıkarımı, Doğal Dil İşleme, Derinlik Öncelikli Arama Algoritması

Tüketicilerin satın alma davranışlarını etkileyen en önemli unsurlardan birinin başta tüketicilerin görüşleri olduğu bilinmektedir. İnternet ortamında yapılan alı veri lerde ürün veya hizmetler hakkında görüş belirtmek ya da bu görüşlere ulaşmak kolaydır. Ancak bu avantaj aynı zamanda yorum sayısını arttırdığı için tüm yorumları incelemek zorla maktadır.

Bu çalışmada tüketiciye zaman kazandırmak için yorumlardan özet nitelikte bilgi çıkarılması konusu ele alınmıştır. Seçilen bir ürüne ait tüketici yorumları www.hepsiburada.com adı web sitesinden alınarak öncelikle morfolojik analiz işleminden geçirilmiştir. Ön işleme hazır hale getirilmiş kelimelerin analizi sonucunda sözcük türleri ve aldıkları ekler belirlenmiştir. Doğal dilde yazılmış bu metinlerden negatif veya pozitif anlamlar taşıyan bilginin çıkarılması için ürün özelliğini belirten kelimeler ve bunları niteleyen sıfatlar önceden tespit edilmiştir.

Yorumlardan otomatik olarak bilgi çıkarımını sağlayabilmek amacıyla doğal dilde yazılmış yorum cümleleri için Türkçe dilbilgisi kuralları ve cümle dizilişine göre ağaç yapısı oluşturulmuştur. Belirtilen özelliklerin tespit edilmesi için oluşturulan ağaç yapısı üzerinde Derinlik Öncelikli Arama algoritmasıyla arama yaparak sonuçları bir yazılım geliştirilmiştir. Yazılım, Zemberek doğal dil işleme kütüphanesinin de yazıldığı, geniş dokümantasyona sahip ve tam nesneye dayalı programlama dili olan Java'da gerçekleştirilmiştir. Yazılımın Java dilinde geliştirilmesi için NetBeans tercih edilmiştir. Buradan elde edilen sonuçlar SQL veritabanında saklanmaktadır. Bu veriler önceden tespit edilmiş ürün özelliklerine göre sorgulandığında tüketicilerin yaptığı yorumlardan memnuniyet derecesini gösteren sayısal hale dönüştürülmüş oranlar elde edilmiştir. Bu sistemden elde edilen bilgilerin tüketiciler kadar üreticiler için de fayda sağlayacağı üzerinde durulmuştur.

AUTOMATICALLY EVALUATING OF PRODUCT COMMENTS WITH TEXT MINING TECHNIQUES

SUMMARY

Keywords: Text Mining, Information Extraction, Natural Language Processing, Deep First Search Algorithm

The one of the important factors which affect consumers' purchasing behavior is known comments of another consumers. So purchases made in internet media make it easy to reach these views of them about products or services. However this advantage complicates to analyze all comments because this media brings about increasing the number of comments at the same time.

In this study, summary knowledge is aimed to extract from comments so as to save time for consumer. Consumer reviews of the selected product were primarily processed into morphological analysis by getting them on the web sites called www.hepsiburada.com. Word types and prefix or suffix was determined as a result of analysis of these words. Also, words and adjectives characterizing them were identified in order to extract knowledge indicating negative and positive meanings from these texts written in natural language.

A software system was developed to evaluate results with aim of determining the desired characteristics by creating tree structure. This system coded by using depth-first search algorithm. It was implemented in JAVA language being fully object-oriented programming language and having a wide range of documentation. Cause of selecting this language was that Zemberek being natural language processing library is built up by using it. NetBeans was chosen as this software developed in JAVA codes. Data being result of this process was stored in SQL database. When this data is queried according to the desired structure, numerical information designating the degree of consumers' satisfaction was obtained from comments of them about product characteristic.

BÖLÜM 1. G R

nsan dilinin temel niteli i ve en temel özelli i, sınırsız dü ünceler zincirini dile getirmek için sınırlı sayıda araçlar kullanmasıdır. “Galileo”, bunu ilk fark eden ki ilerden biridir ve “Dialogo” adlı eserinde, insanların altı üstü yirmi dört tanecik harfin kâ it üzerine farklı ekillerde yan yana dizilmesiyle en gizli dü üncelerini ifade etmelerini sa layan bir ileti im aracının ke fini hayranlıkla anlatmaktadır [1]. Toplumların dilleri farklı olsa da her dilin belirli sayıda olan harflerinin kombinasyonları kelimelere dönü ür. fade edilmek istenen dü ünceler, yazılı dilde kelimelerin ta ıdı ı anlamlara göre belirlenir. Bu nedenle seçti imiz kelimeler hayatımıza yön vermektedir.

nsano lu geli en teknoloji ile birlikte hemen her dü üncesini yazılı bir eilde elektronik ortamlara aktarabilmektedir. Bu aktarım sırasında, dü ünme eyleminin bir sonucu olarak fikirlerden olaylara, insanlardan, cansız varlıklara kadar her eyi, sınıflandırma ve beraberinde derecelendirme ihtiyacımız da kelimelerimize yansımaktadır. Ço u zaman farkında olmadan yaptı ımız bu analiz, bahsi geçen konu hakkında ba kalarının dü üncelerini veya tercihlerini etkileyebilir. Böyle bir etkinin mümkün olabilmesi ve anla ıldı ı ölçüde bilgi sa layabilmesi için yazılı ortamdaki dü üncelerin okunması gerekmektedir. Fakat bu i zaman alıcıdır ve artık zaman, günümüzün en de erli kaynaklarından biridir. çinde ya adı ımız bili im ça ı, elektronik ortamdaki mevcut ve toplanan veri miktarındaki hızlı artı la tanımlanmaktadır. Elektronik ortamda payla ılan bu dü ünceler de, bilgiye dönü meyi bekleyen çok sayıda verinin olu masına neden olmaktadır.

Günümüzün zaman kadar önemli ve en güçlü kaynağı ise bilgidir. En temel tanımıyla bilgi, yorumlanmış veri anlamına gelmektedir [2]. Bilgiyi elde etmek için eldeki mevcut verileri de erlendirmek yerine artık kendi verisini oluşturan yaklaşımlar tercih edilmeye başlanmıştır. Hızla gelişen bu teknolojik ortamda internet aynı zamanda çok önemli bir veri kaynağı olarak kabul edilmektedir. Bunun yanısıra internet, tüketicilerin bir ürün veya hizmet hakkındaki görüşlerini paylaşabilmeleri için de mükemmel bir ortam sayılmaktadır. Bu nedenle birçok kişi bir ürün veya hizmeti almadan önce o ürün veya hizmet hakkındaki ön araştırmalarını internette yer alan ürün bilgileri ve tüketici yorumlarını okuyarak yapmaktadır. Çünkü tüketici gözüyle bakıldığında satın alma davranışımızı etkileyen en önemli faktörlerden biri başkalarının o ürün hakkında ne düşündüğüdür.

Teknolojideki ilerlemeler neredeyse sınırsız imkanlar ortaya çıkarırken tüketiciler tarafından yapılan yorumların sayısı ve çeşitliliğinin de artmasına sebep olmaktadır. Ürün veya hizmet çeşitliliği gibi tüketici görüşlerinin de giderek arttığı günümüzde tercih yapmak için kısa zamanda ulaşılabilecekimiz bilgiye ihtiyaç duyarız. Bu nedenle tüketiciler kadar işletmeler için de önemli ve kefedilmemiş bilgilerle dolu olan ürün yorumlarının otomatik olarak de erlendirilmesi artık zorunlu hale gelmiştir.

Bu çalışmada, yapılan açıklamalar doğrultusunda başta tüketiciler olmak üzere firmalar için de fayda sağlamak amacıyla, doğrudan yazılmış ürün veya hizmet yorumlarından özet bilgi çıkaran bir sistem tasarlanmıştır.

1.1. Tezin Amacı

Bu tezde, özellikle internet üzerinden alı veri yapan bir tüketicinin karar vermesini kolayla tırmak için ürün veya hizmetler hakkında do al dilde yazılmı yorumlardan otomatik olarak de erlendirip kullanıcıya özet bilgi sunan bir sistem geli tirilmi tir.

Olu turulan sistemin çalı ma adımları a a ıda belirtildi i gibidir.

1. Herhangi bir ürüne ait yorumlar ve ürün özellikleri hakkındaki bilgiler www.hepsiburada.com adresinden JAVA'da yazılmı program parçası ile otomatik olarak alınır.
2. Türkçe do al dil i leme kütüphanesi olan Zemberek programı yardımıyla cümlelerin Morfolojik Analizi yapılır. Morfolojik Analiz sonucunda her bir kelimenin türü, ekleri, ek türleri ve kökleri tespit edilmi olur.
3. Herhangi bir konudaki fikrimizi ço unlukla sıfat-isim ve daha az kullanımla fiil-zarf ili kileri ile ifade ederiz. Bu nedenle cümlede niteleyen ve nitelenen kelimelerin ve aralarındaki ili ki derecesinin belirlenmesi gerekir. Bu ili kiler Türkçe'nin dilbilgisi yapısına ve cümle dizili ne göre olu turulmu a aç yapısı üzerinde Derinlik Öncelikli Arama algoritması kullanılarak geli tirilen yazılım yardımıyla otomatik olarak belirlenir.
4. Niteleyen(sıfat) kelimeler anlamlarına göre derecelendirilir. Anlamca olumlu kelimeler daha yüksek bir puanla derecelendirilirken olumsuzluk içeren kelimeler dü ük puanla temsil edilir. Bu ekilde niteleyen kelimeler artık rakamsal verilere dönü türülmü olur.
5. Seçilen ürün için özellik sözlü ü ve derecelendirilmi sıfat sözlü ü olu turulur. Türkçe'de kelimelerin pozitif veya negatif anlamlı olmaları kelime türüne veya köküne bakılarak anla lamaz ve bu konuda bir ayırım çalı ması yoktur. Bu nedenle sistemin gerçekte tirilebilmesi için en çok kullanılan niteleme sıfatları anlamlarına göre derecelendirilir. Bu derecelendirme için 1 ile 5 arasında puanlama yapılmı tir. Olu turulan sıfat sözlü ünden niteleme dereceleri alınır.

6. Tüm yorumlar incelendi inde ürünün hangi özelli ini hangi sıfatların temsil etti i belirlenmi oldu undan aynı özelli ine ait sıfatların olumluluk olumsuzluk derecelerinin ortalaması alınır. Bu sayede yorumlarda geçen her bir ürün özelli i için memnuniyet derecesine ula ılır. Bu dereceler oranlar halinde tablo ekinde gösterilir.

1.2. Çalışmanın Gereksinimleri

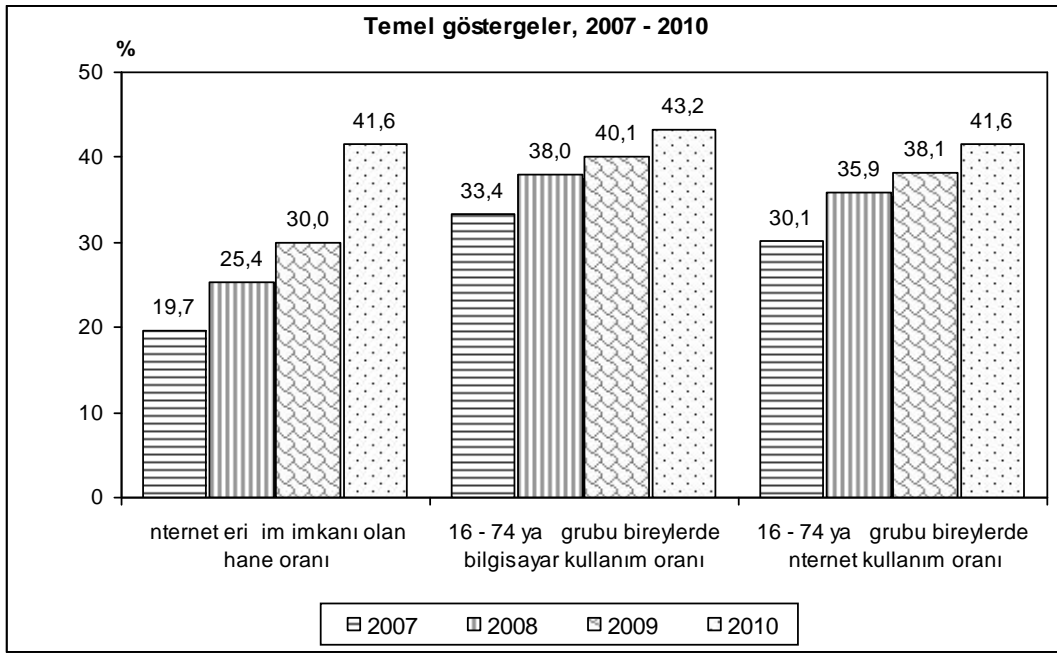
Kullanı amacına göre farklılık gösterse de internet günümüz dünyasının vazgeçilmez bir ö esi olmu durumdadır. Yapılan ara tırmalar neticesinde 2009 yılı Eylül ayı verilerine göre dünya nüfusunun %28.7'si internet kullanmaktadır ve internet, kullanımı her geçen gün artan kablosuz ve yeni nesil GSM araçları ile birlikte 800 milyon civarında cihaz üzerinden eri ilebilen bir ortam haline gelmi durumdadır [2]. Bili im teknolojilerindeki bu geli meler ve etkilerine ba lı olarak günümüz tüketim toplumunun seçimlerinde ve bu seçimlerin de erlendirilmesinde bazı de i imler gözlemlenmektedir. Bu de i imlerden en çok göze çarpanı alı veri alı kanlıklarımızın kazandı ı yeni boyuttur. Yukarıda verilen oranlardan da görüldü ü gibi elektronik ortamlardan faydalanma düzeyinin artması ve kullanımının yaygınla ması, alı veri tercihlerini, bazı ürün veya hizmetlerde internet ortamına ta ımı tır.

Bir internet kullanıcısı, elektronik ortamda potansiyel olarak mü teri ya da tüketici haline gelmektedir. Çünkü bu ortam bilgi edinmenin yanı sıra herhangi bir ürün veya hizmet ile ilgili çe itli satın alma seçeneklerini aya ımıza kadar getirmektedir. nternet üzerinden yapılan alı veri ler, yapısal olarak bakıldı ında elektronik ticaretin firmadan tüketiciye (Bussines to Consumer, B2C) gerçekleştirilen boyutudur. Elektronik ticaret ile tüketiciler, web siteleri üzerinden her türlü ürün veya hizmete eri ebilir, ürün ya da hizmet hakkında bilgi ve fiyat alabilir, rakip firmalarla kıyaslama yapabilir, elektronik ödeme, elektronik bankacılık ve

sigortacılık, danışmanlık gibi işlemleri yapabilirler. Elektronik ortamda alışveriş yapan tüketiciler, internette detaylı ürün bilgileri ve çok fazla çeşit seçeneği bulmanın avantajını yaşıyorlardır [3]. Davranışsal açıdan bakıldığında ise elektronik ticarete imkan sağlayan internet ortamı, fiziksel ortamın aksine tüketicinin tam anlamıyla özgür olduğu alandır. Sıradan mağazalarda karşılaşılan davranış normları bu ortamda geçerli değildir. Satın alma zorunluluğu olmadığından ürün ve hizmetlerin kod numaraları, gıda tüzükleri gibi bilgiler dikkate alınmaz. İnternet, alışveriş ortamını olabildiğince kişiselleştirmektedir. Ayrıca alışveriş için fiziksel bir çaba sarf edilmediğinden çok sayıda web sitesi kısa zamanda gezilebilir [3].

Tüm bu avantajlarına rağmen, müşteriler çoğu zaman internet üzerinden alışveriş yapmak konusunda çekingen davranmakta ve geleneksel alışveriş yöntemlerini tercih etmektedirler. Bunun en önemli sebebi ise güvenlidir. Bireylerin elektronik pazarlara ve pazarlama uygulamalarına yönelik güven tutumlarının analiz edildiği bir çalışmada, tüketicilerin elektronik pazarlarda bulunmaktan hoşlandıkları, fakat kendileri için risk içeren işlemleri yapmak konusunda istekli olmadıkları, ancak internetin geleceğini de olumlu gördükleri ve internetten alışverişin gittikçe yaygınlaşacağını düşündükleri saptanmıştır [3]. Güven sorununa rağmen internet üzerinden yapılan alışverişlerin yaygınlığı düşüncesini, Türkiye İstatistik Kurumundan alınan internet kullanımı ve amacına dair veriler de desteklemektedir.

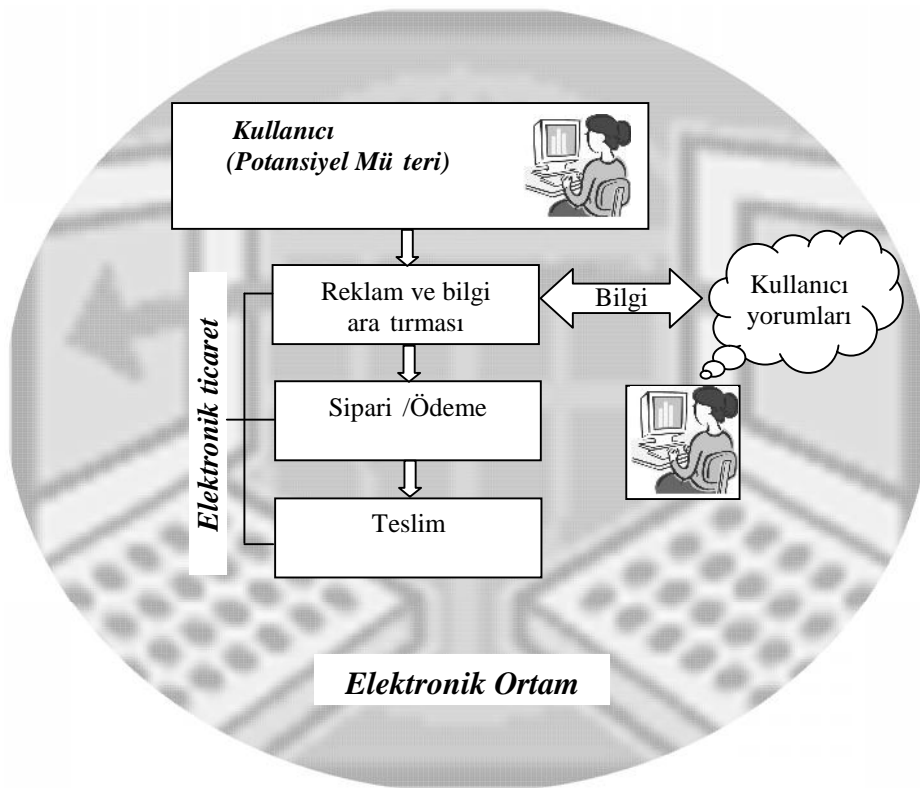
Şekil 1.1'de 2007-2010 yılları arasında bilişim teknolojilerinin kullanım oranları bilgisi yer almaktadır.



ekil 1.1. Hane halkı bilişim teknolojileri kullanım oranları[4]

Bu grafiğe göre 2010 yılı Nisan ayı içerisinde gerçekleştirilen hane halkı bilişim teknolojileri kullanım araştırması sonuçları, hanelerin %41.6'sının internet erişim imkanına sahip olduğunu göstermiştir. Bu oran 2009 yılının aynı ayında %30 olarak tespit edilmiştir. Buradan internet kullanıcılarının her geçen gün arttığı daha net bir şekilde görülmektedir. Yine aynı çalışmaya göre 2010 yılının son üç ayı içerisinde internet kullanan bireylerin, %55.7'si ürün ve hizmetler hakkında bilgi aramak için interneti kullanmıştır [4]. İnternet kullanan bireylerin kişisel kullanım amacıyla internet üzerinden ürün veya hizmet siparişini verme ya da satın alma oranı %15'tir. Aynı çalışmaya göre, son 12 ayda internet üzerinden mal veya hizmet siparişini veren ya da satın alan bireyler, %24.3 oranı ile en fazla giyim ve spor malzemelerini almışlardır. Bir önceki yıl ise aynı dönemde elektronik araçlar %32.2 ile ilk sırayı almıştır [4]. Bu oranlar satın alınan ürün veya hizmet farklılık gösterse de internet üzerinden yapılan alışverişlerin ihmal edilmeyecek kadar önemli olduğunu göstermektedir. Yine buradan yola çıkılarak bazı ürün veya hizmetlerde internet ortamında yapılan alışverişlerin, fiziksel ortamlarda yapılan alışverişlerden daha cazip hale geldiği söylenebilir.

nternet üzerinden yapılan alı veri lerin elektronik ticaret yönüyle bakıldı ında üç boyutu vardır. Birinci a ama, reklam ve bilgi ara tırması kapsayan satın alma öncesi a amadır. kinci a ama, sipari verme, satın alma ve ödemeyi kapsayan satın alma davranı ıdır. Üçüncü ve son a aması ise teslim a amasıdır [5,6]. Bu çalı mada amaç satın alma davranı ının tamamlanmasını sa lamak olmadı ından tüketici yorumlarının en çok yer aldı ı birinci a amadaki yorumların taraflar için faydalı bilgiye dönü üm süreci ile ilgilenilmi tir. Buradan hareketle ekil 1.2’de elektronik ortamda kullanıcı görü lerinin bilgiye dönü ümü ve elektronik ortam ile etkile imi görülmektedir.



ekil 1.2. Kullanıcı yorumlarının elektronik ortamdaki yeri

nternet kullanıcısını, davranı sal açıdan bakıldı ında bir tüketici gibi dü ünecek olursak, alı veri tercihi ve satın alma davranı ları u ekilde incelenebilir.

Bir tüketici karar alırken birçok iç ve dı faktörün etkisi altında kalmaktadır[6]. Bununla birlikte tüketici aynı zamanda davranı ları;

- kültürel etkiler,
- bireysel etkiler,
- grup etkileri

olmak üzere üç temel faktör üzerine kuruludur. Kültürel etkiler, davranı normları ve sosyalle meyi içerir. Bireysel etkilerde güdülenme, duygular, öğrenme ve hatırlama, tutum, algılama, rasyonel ve rasyonel olmayan düşünme, kişilik ve kişilik farklılıkları ve benlik özellikleri incelenir. Grup etkileri faktörü ise taklit etme ve öneri alma, aile, sosyal etkiler, etnik ve dinsel etkiler, sosyal sınıf, rol ve önderlerin etkisi gibi alt faktörlerden oluşmaktadır [5].

Tüketici davranışlarını etkileyen grup etkileri faktörlerinden de görüleceği gibi kişinin satın alma davranışını başka tüketicilerin görüşlerinden etkilenmektedir. İnternet üzerinden yapılan alıverişlerde geleneksel alıverişlerde olduğu gibi tüketiciyi yönlendiren en önemli faktörlerden biridir diğer tüketicilerin tutumlarıdır. Sanal ortamda bir ürünü fiziksel olarak inceleme imkânı olmadığından tüketici, ürün hakkında araştırmalarını ancak diğer tüketicilerin görüşlerini okuyarak tamamlayabilmektedir. Yapılan araştırmalar, insanların memnuniyetsizliklerini, memnuniyetlerinden daha fazla dile getirdiklerini göstermektedir. Bu durumda olumsuz görüşler satın alma tercihlerimize daha fazla etki etmektedir. Bu nedenle alacağımız ürün veya hizmetler hakkındaki tüm görüşleri ve varsa deneyimleri öğrenmek isteriz.

İnternet ortamında tüketici davranışını belirleyen diğer bir önemli etken de zaman tasarrufudur. Bu ortamda ürün veya hizmet ve firma bilgileri gibi bunlar hakkındaki tüketici görüşleri de daha kolay bir şekilde paylaşılmakta ve geniş kitlelere ulaşabilmektedir. Ancak bu avantajın bir sonucu da yapılan yorumların sayısının arttığı gerçektir. Web dünyasındaki bilgi ve belge sayısının her geçen gün üstel bir şekilde arttığını göz önüne alırsak yorum sayısının ve çeşitliliğinin de artması kaçınılmazdır. Bu artışa rağmen doğru kararı verebilmek için tüm görüş veya yorumları incelemek gerekebilir. Fakat bu durum başlangıçta söylediğimiz zaman tasarrufuna ters düşmektedir. Tüm yorumların gözden geçirilmesi ürün veya hizmet

çe itlili i ve yorum sayısının fazlalılı ı nedeniyle oldukça zaman alıcı bir i tir. Bu sebeple tüketicinin karar verebilmesini kolayla tırmak amacıyla yorumlardan otomatik olarak özet bilgi çıkarılması zorunlu hale gelmi tir. Çünkü teknoloji birçok konuda hayatımızı kolayla tırırken paralelinde çözülmesi gereken yeni problemlere de sebep olmaktadır. nternet ortamı buna en iyi örneklerden biridir. Bu teknoloji sayesinde bilgiye ula mak ve bilgiyi payla mak son derece kolayla mı ken, aynı zamanda bilgi kirlili i ya anmasına da yine internet imkanı neden olmaktadır. Teknolojideki her bir geli me yeni bir çözüm ve bununla beraber yeni bir problem olu turmaktadır. Her bir problem de teknolojinin geli mesini sa layan temel faktörlerden biridir. Bu yüzden teknoloji geli imi ve olu an problemler çift yönlü bir reaksiyon olarak dü ünülebilir. Burada ele alınan problem de teknolojik imkanların geli mesi nedeniyle ortaya çıkmı bir problem olup bir ürün veya hizmet hakkındaki tüm yorumların okunup de erlendirilmesinin uzun zaman alması ekinde tanımlanabilir.

Buraya kadar yapılan açıklamalar özetlenecek olursa,

- nternet üzerinden yapılan alı veri ler artmaktadır.
- Bu tür durumlarda alı veri gerçeikle mese bile insanlar ürün veya hizmetler hakkındaki dü üncelerini payla maktadırlar ve bu görü ler tercihlerimizi etkileyen en önemli faktörlerdir.
- Ürün veya hizmetler hakkındaki görü lerin tümünün incelenmesi uzun zaman alacak bir i tir ve özet nitelikte bilgi çıkarılması için bir sistemin geli tirilmesi gerekmektedir.

Bu çalı mada, yapılan açıklamalar do rultusunda ba ta tüketici olmak üzere firmalar için de fayda sa lamak amacıyla, ürün veya hizmet yorumlarından özet bilgi çıkaran bir sistem geli tirilmi tir.

Problemin çözümünde internet ortamındaki kullanıcı ya da tüketici yorumları veri olarak ele alınmıştır. Söz konusu veriler doğal dille yazılmış metinsel verilerdir. Bu nedenle elde edilen veriler, Doğal Dil İşleme yöntemlerinden geçirilerek analiz edilmeye hazırlanmıştır. Daha sonra Metin Madenciliği yöntemleri ile ürün veya hizmetler hakkında pozitif ya da negatif anlamlar taşıyan, özet nitelikte bilgiler çıkarılmıştır.

Doğru bilgiye otomatik bir değerlendirme sistemiyle ulaşıldığı takdirde zamandan tasarruf sağlanacağı gibi elde edilen bilgiler bir girdi olarak tüketiciler kadar firmalar için de fayda sağlayacaktır. Bu girdiler, karar vermeyi etkilediğinden geliştirilen sistem, aynı zamanda bir firmanın Zekası çalışmaları kapsamında yer alacak ve kurumsal zekanın gelişmesine katkı sağlayacaktır. Bu nedenle tezin 2. Bölümünde Metin Madenciliği ve Zekası arasındaki ilişkiye değinilmiştir.

1.3. Tezin Kapsamı

Tez beş bölüme ayrılmıştır. Giriş bölümünde tez konusu hakkında genel bilgiler verilmiş, tezin amacından bahsedilmiş ve çalışmanın gerekçeleri anlatılmıştır.

İnternet ortamında depolanan verilerin büyük çoğunluğu gerçekte yapısal olmayan belgelerden meydana gelmektedir. Bu verilerin incelenmesi için bazı tekniklere ihtiyaç duyulur. Bunlardan biri de Metin Madenciliği çalışmalarıdır. Metin Madenciliği çalışmalarının temelinde Veri Madenciliği vardır. Bu nedenle tezin ikinci bölümünde Veri Madenciliği konusuna özet bir şekilde değinilmiş, Metin Madenciliği hakkında ise detaylı bilgi verilmiştir.

Metin Madenciliği, işletme dokümanları, müşteri yorumları, web sayfaları ve XML dosyalarını içeren yapısal olmayan veri kaynaklarının herhangi bir çeşitten gelen

yeni, önceden bilinmeyen, potansiyel olarak kullanılabilir bilgiyi keşfetme sürecidir [6]. Metin Madenciliği metin derlemelerinden yeni bilgileri türetme işlemi olarak da tanımlanabilir. Elde edilen bilgiyle analizi yapılan metin kaynaklarında açık olarak görülmeyen ilişkiler, hipotezler veya etkileşimler olduğu anlaşılır [7].

Metin Madenciliği işlemleri, Veri Madenciliğine benzer olarak,

- Metin,
 - Metin Ön İşleme,
 - Özellik Üretimi,
 - Özellik Seçimi,
 - Veri Madenciliği (Örüntü Keşfi),
 - Yorum/Değerlendirme,
- adımlarından oluşmaktadır.

Zohar'a (2002) göre Metin Madenciliği metodları,

- Bilgiye Erişim (Information Retrieval),
- Bilgi Çıkarımı (Information Extraction),
- Web Madenciliği (Web Mining),
- Kümeleme (Clustering),

olmak üzere dört grupta toplanmaktadır[8].

Metin Madenciliği, Veri Madenciliğinin genel alanının bir parçası olarak düşünülmese de ayrı bir alan olarak kabul edilir. Bu nedenle çalışmanın devam eden kısmında Veri Madenciliği ve Metin Madenciliği arasındaki farklardan bahsedilmemiştir.

Veri Madenciliği çalışmalarını gerçekleştirmek için çeşitli yöntemlere ve bazı algoritmalara ihtiyaç vardır. Bu algoritmaların büyük bir bölümü Makine Öğrenmesi (Machine Learning) konusu adı altında yer almaktadır. Bu nedenle çalışmanın izleyen kısmında Makine Öğrenmesinden bahsedilmemiştir.

Tezin yine ikinci bölümünde devam eden kısımda Metin Madenciliği çalışmaları ve veri ön işlemede gerekli olan Doğal Dil İşleme çalışmaları bahsedilmiştir.

Metin Madenciliği çalışmaları sonucunda elde edilen bilgilerin firmalardaki Zekası çalışmaları katkısı iddia edildiğinden ikinci bölümün son kısmında Zekası hakkında bilgi verilmiş ve Metin Madenciliği arasındaki ilişki anlatılmıştır. Zekası en basit şekilde, verinin incelenmesi için fedilmesi ve dönüşümü ile bilginin elde edilmesi olarak tanımlanabilir [9].

Üçüncü bölümde ise uygulama amacıyla için gerekli olan metin ön işleme çalışmaları bahsedilmiştir. Bölümün devamında Türkçe'nin morfolojik analizi ve dilbilgisi kurallarının yardımıyla ürün yorumlarından özet bilgi çıkarılmasını sağlayan sistem anlatılmıştır. Dilbilgisi kuralları için ağ yapısı oluşturulmuş, niteleyen ve nitelenen kelimelerin tespiti için Derinlik Öncelikli Arama algoritması kullanılmıştır. Elde edilen çıktılarla birlikte sistemin performansı hakkında bilgi verilmiştir.

Dördüncü bölümde ise sonuç ve yapılması önerilen çalışmalara yer verilmiştir.

BÖLÜM 2. METİN MADENCİLİK VE BİLGİ KEŞİFLERİ

2.1. Giriş

Günümüzün en değerli iki kaynağı bilgi ve zamandır. Teknolojideki baş döndüren yenilikler, veri ve bilgiye kolayca ulaşılabilen ortamları ve bilgiyi elde etme imkanını arttırmıştır. Fakat her geçen gün artan veri ortamında istenilen bilgiye, istenilen zamanda ulaşmak artık kolay olmamaktadır. Kişilerin asla geri döndüremeyecekleri kaynakları olan zamanı etkin kullanmalarına ihtiyaçları vardır. Eldeki verinin yapısına ne olursa olsun doğru ve hızlı bir şekilde işlenmeli ve faydalı bilgiye dönüştürülmelidir. Bu nedenle özellikle yapılandırılmamış verilerden bilgi çıkarımı için bazı teknikler geliştirilmiştir. Bu tekniklerden biri de son zamanlarda önem kazanan yeni bir bilgi analiz etme yöntemi olan Metin Madenciliği konusudur. Metin Madenciliği, yapılandırılmamış çok büyük miktardaki verilerden yarı otomatik olarak bilgi çıkarımını imkânda kılan şekilde tanımlanabilir. Daha geniş bir ifade ile belirli bir formatta olmayan yazı tipindeki veriler içerisinde gizli olan nitelikli bilginin çıkarılması, düzensiz haldeki verinin formatlanması sürecini içermektedir [6]. Metin Madenciliği, Metin Veri Madenciliği (Text Data Mining) ve Metin Veritabanlarından Bilgi Keşfi (Knowledge Discovery from Textual Databases) olarak da adlandırılır [7].

Yukarıdaki isimlerden ve tanımlamalardan da anlaşılacağı gibi Metin Madenciliği konusunun çıkış noktası Veri Madenciliği çalışmaları olduğundan tezin bu bölümünde öncelikle Veri Madenciliği hakkında bilgi verilecektir.

2.2. Veri Madencili i (Data Mining)

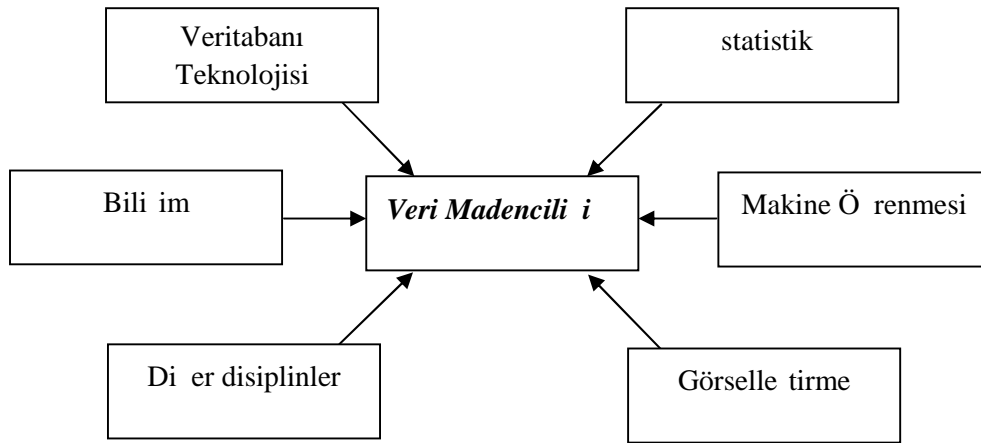
Geleneksel bilgiye ula ma yöntemleri, belge yı nları içinden ana konu ba lıklarına yönelik aramaları ba arılı bir ekilde kar ılayabilmektedir. Ancak, arama yapılan doküman sayısı artıkça, sorgu neticesinde gelen doküman sayısı binlerle ifade edilir hale gelebilmekte, bunun sonucu olarak da daha özel ihtiyaçları kar ılayabilecek çözümlere ihtiyaç duyulmaktadır. çinde ya adı ımız bili im ça nda elektronik ortamda mevcut verinin hızlı artı ı ve bilginin fazlala ması sebebiyle öncelikle, genelde Veri Tabanlarında Bilgi Ke fi olarak adlandırılan yeni bir paradigma ortaya çıkmı tır. Daha yaygın bir kullanımla bu alana Veri Madencili i denilmektedir [7,10]. Veri Madencili i en basit tanımıyla, büyük ölçekli veriler arasından de eri olan bir bilgiyi elde etme i idir [10].

Ba ka bir ifade ile Veri Madencili i yapısal veritabanlarında depolanmı verilerden geçerli, yeni, potansiyel olarak yararlı ve nihayetinde anla ılabilir örüntülerin tanımlanması i lemidir [11].

Di er bir tanımda ise veri madencili i, verilerin içerisindeki desenlerin, ili kilerin, de i imlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak ke fedilmesidir, denilmektedir [10,12].

Bu tanımlamalardan da anla ıldı ı üzere veriler arasındaki ili kileri ortaya koymak ve gerekti inde ileriye yönelik tahminlerde bulunmak veri madencili i çalı maları sayesinde mümkün olmaktadır. Bunun anlamı, veri madencili i bir kurumda üretilen tüm verilerin belirli yöntemler kullanarak var olan ya da gelecekte ortaya çıkabilecek gizli bilgiyi ortaya çıkarma süreci olarak de erlendirilmesidir. Bu açıdan bakıldı ında veri madencili i i inin kurumların Karar Destek Sistemleri için önemli bir yere sahip oldu u söylenebilir.

Veri madenciliği çalışmaları, sınıflandırma, ilişki kurma, kümeleme, regresyon, veri özetleme, derinliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Bu çalışmalar, aynı zamanda ekil 2.1’de gösterilen diğer teknolojileri de kapsamaktadır [10].



ekil 2.1. Veri madenciliği ile ilişkili diğer disiplinler [12]

Veri madenciliği özellikle e-ticaret, bilim, tıp, iş ve eğitim alanlarında geniş uygulama imkanı bulmuştur. Bunun yanı sıra, astronomi, biyoloji, finans, pazarlama, sigorta ve birçok alanda da uygulanmaktadır. Veri madenciliği teknikleri günümüzde özellikle işletmelerde de başarılı bir şekilde kullanılmaktadır. Bu uygulamaların başarıları Tablo 2.1’deki gibi özetlenmiştir.

Tablo 2.1. Veri madenciliği uygulama alanları [12]

| Bilim | Hayatı | Web | Devlet |
|--|---|--|--|
| - Astronomi -Biyoinformatik - laç kefi | -Reklam -CRM(Müşteri ilişkileri Yönetimi) ve Müşteri Modelleme -E-ticaret -Yatırım değerlendirme ve kar ılaştırma -Sa lık -Üretim -Spor/e lence -Telekom (telefon ve iletişim) -Hedef pazarlama | -Metin Madenciliği (haber grubu, e-mail, dokümanlar) -Web analizi -Arama motorları | -Terörle Mücadele -Kanun Yaptırımı -Vergi Kaçakçılarının Profiline Çıkarılması |

2.2.1. Veri madencili i yöntemleri

Veri madencili inde amaç çok büyük miktardaki ham veriden de erli bilginin çıkarılmasıdır. Bu amaç do rultusunda veri madencili inde çok sayıda yöntem ve algoritma uygulanmaktadır. Veri madencili inin ilk uygulama alanları, do ası gere i büyük veri kümeleridir. Veri madencili i çalı malarını verinin nerede oldu u, biçimi ve niteli i gibi etkenler de etkilemektedir. Özellikle ili kisel veritabanlarındaki verilerin analizi için kullanılan teknikler artık web sayfalarının yapısal diye tabir edilen belgelerin analizi gibi alanlarda kullanılmakta ve de i ik yakla ımların do masını sa lamaktadır [10,13,14].

Veri Madencili i, üç temel alana ayrılarak geli mi tir. Bunlardan ilki ve en önemlisi olan ve en eskiye dayanan klasik statistik bilimidir. Regresyon analizi, standart da ılım, standart sapma, diskriminant analizi ve güven aralıkları gibi verileri ve veriler arasındaki ili kiye inceleyen bu yöntemler klasik istatistik çalı malarıdır. Bu yöntemler ileri düzey istatistiksel analizin temelini olu turan yapı ta larıdır ve bugün kullanılmakta olan Veri Madencili i araç ve metotlarının esasını olu turur.

Veri Madencili inin gelişmesine neden olan di er bir alan ise Yapay Zeka'dır. Yapay Zeka, sezgisel yakla ımları temel alarak insan benzeri dü ünebilme prensibiyle istatistikten farklı metotlarla, istatistiksel problemlere yakla ır.

Di er ve son alan ise yine temellerini statistik ve Yapay Zeka'dan alan Makine Ö renmesi'dir. Makine Ö renmesi için Yapay Zeka'nın sezgisel yöntemleri ileri düzey istatistiksel yöntemlerle harmanlayıp evrimle tirerek geli ti i ileri düzey halidir, denilebilir[]. Makine Ö renmesi uygulandı ı bilgisayar sistemlerinde istatistiksel ve yapay zeka algoritmaları kullanılarak eldeki verinin de erlendirilmesine, bu verilerden sonuç çıkarılmasına ve bu sonuçlara bakılarak

kararlar alınmasına olanak sağlar. Bu konu ile ilgili daha detaylı bilgi ve Bilgi Çıkarımında kullanılan algoritma isimleri Bölüm 2.4'te yer almaktadır.

Yukarıda da bahsedildiği gibi birçok Veri Madenciliği yöntemi makine öğrenmesi, örüntü bulma ve istatistik disiplinlerinde denenmiş ve test edilmiş tekniklere dayanmaktadır. Veri Madenciliği yöntemleri ile ilgili olarak temel aldıkları farklı disiplin ve problem türüne göre çeşitli gruplandırmalar yapılabilir. Bunlardan biri de aşağıda verildiği gibidir [10,14].

- Sınıflandırma
- Kümeleme
- Birliktelik Kuralları

2.2.1.1. Sınıflandırma(Classification)

Sınıflandırma daha önceden belirlenmiş sınıflara veriyi yerleştirmek için kullanılacak fonksiyonun öğrenilmesi işlemi kapsar. Çıktılar önceden bilindiği için sınıflama veri kümesini denetimli olarak öğrenir. Tüm veriler kullanılarak bir etiketleme işlemi yapılmaz. Öğrenmenin amacı bir sınıflandırma modelinin oluşturulmasıdır. Öğrenme konusuna Bölüm 2.4'te yer alan Makine Öğrenmesi adı altında değinilmiştir.

Sınıflandırma başka bir deyişle, hangi sınıfa ait olduğunu bilinmeyen bir kayıt için sınıf belirleme sürecidir [10].

Verileri sınıflandırma yöntemlerine Karar Ağaçları örnek verilebilir. Karar ağaçları algoritmaları üç gruba ayrılmaktadır. Bunlar,

- Entropiye dayalı algoritmalar
- Sınıflandırma ve Regresyon Ağaçları(CART)
- Bellek tabanlı sınıflandırma Algoritmaları eklindedir.

Bu algoritmalar hakkında kısaca bilgi verecek olursak, bir sistemdeki belirsizliğin ölçüsüne “Entropi” denilir. Entropiye dayalı bölümlenmeyi kullanan algoritmalara örnek olarak ID3 ve onun gelişmiş biçimi olan C4.5 algoritmaları verilebilir.

Sınıflandırma ve Regresyon Ağaçları konusunda ise Twoing ve Gini algoritmalarından bahsedilebilir.

Bellek tabanlı sınıflandırma yöntemleri arasında k-En Yakın Komşu algoritması sayılabilir. Bu yöntem sınıfları belli olan bir örnek kümesindeki gözlemlerden yararlanarak örneğe katılacak yeni bir gözlemin hangi sınıfa olduğunu belirlemek amacıyla kullanılır [10,14].

Bunların dışında istatistiksel Sınıflandırma Modelleri arasında Bayes Ağları, Optimizasyona Dayalı sınıflandırma Modelleri arasında ise Destek Vektör Makinesi yöntemi örnek olarak verilebilir.

2.2.1.2. Kümeleme (Clustering)

Kümeleme işlemi verilen bir veri kümesini tanımlamak için sonlu sayıda sınıfa ya da kümeye bölmeyi kapsar. Birbirlerine benzeyen veri parçalarını ayırma işlemi olan kümeleme yöntemlerinin çoğu veri arasındaki uzaklığı kullanır. Örneğin, Öklid, Manhattan, Minkowski gibi uzaklık ölçütleri kümeleme işleminde alt ölçüt olarak

kullanılmaktadır. Kümelemedeki sınıflar birbirinden ayrı ve detaylı veya hiyerarşik yada örtü en niteliklerde olabilir [10,14].

Kümeleme yöntemleri temel anlamda hiyerarşik ve hiyerarşik olmayan yöntemler olmak üzere iki grup altında toplanır. Hiyerarşik kümeleme yöntemlerine en yakın komşu algoritması ve en uzak komşu algoritması örnek verilebilir. Hiyerarşik olmayan kümeleme yöntemleri arasında k-ortalama yöntemini sayılabilir [10,14,15].

2.2.1.3. Birliktelik kuralları (Association rules)

Veritabanı içinde yer alan kayıtların birbirleriyle olan ilişkilerini inceleyerek hangi olayların eş zamanlı birlikte gerçekleşebileceğini ortaya koymayan veri madenciliği yöntemlerine birliktelik kuralları denir [10,16].

Birliktelik kuralına ilişkin olarak geliştirilen bazı algoritmalar şunlardır; AIS (Agrawal vd., 1993), SETM (Houtsma ve Swami, 1995), Apriori (Agrawal ve Srikant, 1994), Partition (Savasere vd., 1995), RARM - Rapid Association Rule Mining (Das vd., 2001), CHARM (Zaki ve Hsiao, 2002). Bu algoritmalar içerisinde, ilk olanı AIS, en bilineni ise Apriori algoritmasıdır (Agrawal ve Srikant,1995) [17].

2.2.2. Veri madenciliğinden metin madenciliğine geçiş

Veri Madenciliğinin ilk uygulama alanları doğası gereği büyük veri kümeleridir. Veri Madenciliği çalışmalarına başlamak için öncelikle verinin niteliğini belirlemek gerekir. Veri yığınlarını niteliklerine göre üç grupta sınıflandırılır.

- Yapısal
- Yarı yapısal
- Yapısal olmayan

Yapısal veri, veritabanı ve veri ambarlarında tutulan ve SQL, OLAP gibi sorgulama yöntemleri ile sorgulanabilen veri türünü ifade eder. Yarı yapısal veriler ise metin, resim, grafik vs. olan belgelerdir. Belgelerin kim tarafından, hangi konuda ne zaman yazıldığı gibi bazı yapısal kısımları olmakla birlikte bir belgenin içeriğinin tam olarak anlaşılması ancak bir insan tarafından okunması ile ortaya çıkarılabilir. Yapısal olmayan veri ise ses ve görüntü gibi akan verilerdir [17].

İnternetlerdeki verilerin büyük çoğunluğu ise gerçekte yapısal olmayan belgelerde depolanmaktadır. Merrill Lynch ve Gartne tarafından yapılan son çalışmaya göre bütün internet verilerinin %85-90'ının yapısal olmayan formlarda depolandığı görülmüştür. Belgeler, pek çok bilgi içeriği belirli konulara özel olarak üretilen ve çoğunlukla birkaç kişi tarafından incelenmiş ve kişilere ait bilgisayarda muhafaza edilen verilerdir. Pek çok farklı formatta olabilir. Sade metin, Adobe Acrobat, MS Word, HTML, XML, vs. ve internet üzerinden http veya ftp protokolü vasıtasıyla erişilebilir durumda olabilirler. Bu tip verilerin büyük çoğunluğu sıklıkla deşifre edilebilir. Bu sınıftaki verileri analiz etmek çok değerli bilgilerin yakalanmasını sağlayabilir. Bu nedenle geleneksel durumda Metin Madenciliği kavramı ön plana çıkmaktadır [18,19].

2.3. Metin Madenciliği (Text Mining)

Günümüz teknolojisi her alanda çok sayıda ve çeşitli verinin oluşmasına neden olmaktadır ve bunlara kolay ulaşılmasını sağlamaktadır. Gittikçe büyüyen veri hacmi içinde önemli bilgiler kaybolup giderken, değerli bilgilere ulaşmak için özellikle yarı

yapısal veriler olan belgelerin içeriğinin belirlenmesi ve buna uygun sorgulanabilmesi ihtiyacı kendini hissettirmektedir [7]. Günlük hayatımıza baktığımızda gazeteler, makaleler, raporlar birer belgedirler ve veri niteliğindedirler. Geleneksel yöntemler bir yere kadar belgelerin onlardan faydalı ve gerekli bilgileri bulmaya yardımcı olsalar da asıl olarak gereken detay ve özel bilgilere bu yöntemler ile ulaşmak zordur. Bu nedenle bu tür verilerin incelenmesinde Metin Madenciliği teknikleri öne çıkmaktadır.

Veritabanları programlarının otomatik olarak işlenmesi için tasarlanmaktadır. Metin işleme ise insanların okuması için yazılır. Metinleri en azından insan kadar okuyabilen ve anlayabilen programlar yoktur [7,20]. Doğal dil işleme alanında gelinen en üstü noktalarla beraber beklenen gelecek için böyle bir program mevcut değildir. Birçok araştırmacı insanların yaptığı şekilde okuyan ve anlayan programları yazmamızdan önce aklın nasıl çalıştığını gösteren tam bir simülasyona ihtiyaç olacağını düşünmektedir. Bu konuda farklı disiplinlerden çalışmalara ihtiyaç vardır. Bunlardan birisi de Metin Madenciliği çalışmalarıdır.

Metin Madenciliği, işletme dokümanları, müşteri yorumları, web sayfaları ve XML dosyalarını içeren yapısal olmayan veri kaynaklarının herhangi bir çeşidinden gelen yeni, önceden bilinmeyen, potansiyel olarak kullanılabilir bilgiyi keşfetme sürecidir [21,22].

Başka bir deyişle Metin Madenciliği, derlemlerden yeni bilgileri türetme işlemi olarak da tanımlanabilir. Derlem (corpus-külliyyat), elektronik ortam üzerinde toplanmış yazılı ya da sözlü doğal dil seçkilerinin belli prensipler dahilinde bir araya getirilerek dilbilimsel ve yapısal bilgi ile geliştirilmiş halinden oluşan veritabanına denilmektedir. Metin Madenciliği çalışmalarından elde edilen bilgiyle, analiz edilmiş olan metin kaynaklarında açık olarak görülmeyen ilişkiler hipotezler veya eylemler olduğu anlaşılır [7,22].

Diğer bir tanımda ise Metin Madenciliği, işletme ortamında veya internet üzerindeki belgelerde bu belgeye benzer belgelerin olup olmadığını elle bir sınıflandırma gerekmeden benzerliği hesaplayabilmektir. Bu genelde otomatik olarak çıkarılan anahtar kelimelerin tekrarı sayesinde yapılır [23].

Metin Madenciliği, veri madenciliğinin genel alanının bir parçası olarak düşünülmese de genellikle veri madenciliğinden farklıdır. Ana farklılık, metin madenciliğinde örüntülerin olay tabanlı veritabanlarından daha çok, doğal dil metinlerinden çıkartılmasıdır. Tablo 2.2’de bu farklılıklar görülmektedir.

Tablo 2.2. Veri madenciliği ve metin madenciliğinin karşılaştırılması[19]

| Kriter | Veri Madenciliği | Metin Madenciliği |
|-----------------------------|----------------------------------|--|
| inceleme nesnesi | Sayısal ve kategorik veriler | Metinler |
| incelenen nesnelerin yapısı | Yapılandırılmış veriler | Yapılandırılmamış veriler |
| Amaç | Bilinmeyen ilişkileri belirlemek | Farkında olmadığımız anlamlara ulaşmak |
| Uygulama | Yaygın uygulama 1994’te başladı | Yaygın uygulama 2000’de başladı |

Yapılandırılmış bir formata dönüştürülen metinlerin geleneksel Veri Madenciliği teknikleriyle analizi süreci olan hem veri madenciliğinde hem de metin madenciliğinde gizli bilgilere bakılmakta ve genel yapay zeka, makine öğrenme ve istatistik algoritmaları kullanılmaktadır. Veri Madenciliğinde yapılandırılmış sayısal veri kullanılırken Metin Madenciliği yapılandırılmamış metinlerle ilgilidir. Veri Madenciliğinde, veri ambarlarında çıkartılmış, dönüştürülmüş ve yüklenmiş durumda bulunan verileri kullanırken metin madenciliği kesin olmayan verileri modellemeyle çalışmaktadır.

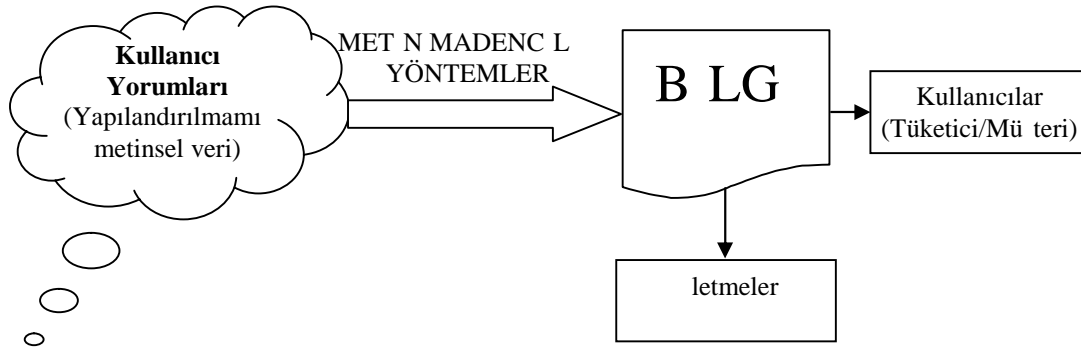
Metin Madenciliğinin ne yaptığına bakacak olursak en temel seviyede yapısal olmayan metin belgelerini sayısızla tırıp daha sonra Veri Madenciliği araç ve tekniklerini kullanarak onlardan anlamlı örüntüler çıkartmak olduğu görülür. Başka bir deyişle metin madenciliği, en genel haliyle doğal dilde yazılmış metinler içinden,

- aynı konudaki belgeleri bulur,
- birbiriyle ilişkili belgeleri bulur,
- ve bulunan belgeleri sıralar.

Metin Madenciliğinin çalışmaları sonucunda, metinsel verilerin büyük bir çoğunluğu oluşturan örneklerin, müşteri şikayet ve memnuniyet formları gibi kurumsal ilemlerden elde edilen metinlerden gelen anlamlı bilgiler, ürün geliştirme, hata izleme garanti süresi gibi konularda işletmeye girdi oluşturur [7,24].

Yapısal olmayan metinleri otomatik olarak işleme olarak da değerlendirilen Metin Madenciliği çalışmalarının kullanıldığı alanlardan biri, elektronik iletişim ve e-maillerdir. Metin Madenciliği yalnızca sınıflandırmaya ve gereksiz (junk) mailleri filtrelemeye yardımcı etmez aynı zamanda otomatik olarak cevap vermekte de kullanılır. Metin Madenciliği yargı, sağlık ve diğer endüstrilerde geleneksel olarak zengin belgeler ve sözleşmelerle elde edilen verilere de ulaşmayı sağlar [7,23].

Bu çalışmada internet kullanıcılarının herhangi bir ürün veya hizmet hakkında yaptıkları yorumlar yarı yapısal veriler olarak ele alınmaktadır. Çalışmanın sonunda elde edilecek bilginin paydaşları ve girdi oluşturan veri kaynağı ve bilgiye dönüşüm süreci ekil 2.2'deki gösterilmiştir.



ekil 2.2. Yorumların bilgiye dönüüm süreci ve elde edilen bilginin payda ları

Giderek artan belge yı nlarının faydalı bilgiye dönü türülmesini sa lamak için geli tirilen Metin Madencili i çalı maları, Bilgiye Eri im (Information Retrieval) ve Bilgi Çıkarımı (Information Extraction) olmak üzere iki alanda incelenmektedir.

Zohar'a (2002) göre Metin Madencili i metotları,

- Bilgiye Eri im (Information Retrieval),
- Bilgi Çıkarımı (Information Extraction),
- Web Madencili i (Web Mining),
- Kümeleme (Clustering),

olmak üzere dört grupta toplanmaktadır [8]. Bu tez çalı ması kapsamında yukarıda bahsedilen metotlardan sadece Bilgiye Eri im ve Bilgi Çıkarımı izleyen bölümde açıklanmı tır.

2.3.1. Bilgiye eri im (Information retrieval)

Bilgiye Eri im kavramı ilk kez Calvin Mooers tarafından 1948 yılında "Application of Random Codes to the Gathering of Statistical Information" ba lı ını ta ıyan yüksek lisans tezinde Information Retrieval terimi altında kullanılmı tır. Vickery,

Mooers'in kavramı İngilizce olarak getirdiği ilk tanımı şu şekilde aktarır. Bilginin bir depodan özelliklerine göre konusal olarak aranarak erişilmesi [25].

Bilgiye Erişim (BE), Metin Madenciliğinde ilk adım olarak nitelendirilmektedir. Bilgiye Erişimin amacı kullanıcıların bilgi ihtiyaçlarını karşılayacak olan belgeleri bulmasına yardımcı olmaktır.

Bilgiye Erişim, birçok konu alanına sahipliği nedeniyle geniş bir alana yayılmaktadır ve kullanıcıların belirli konulardaki belgeleri bulabilmesi gibi büyük bir topluluktan oluşan metni sunması için modeller geliştirilmiştir. Problem, kullanıcıya ne ile ilgilenmekte ve belirli bir konu kümesi hakkında belgeler nasıl sunulmalı ve tanımlanmalı gibidir [26].

Bilgiye Erişim, bilgi ihtiyacını karşılayan yapılandırılmamış materyalleri (genellikle dokümanlar) geniş bir koleksiyonun içerisinde bulmaktır. Eskiden bilgiye erişim sadece bazı meslek grupları tarafından özel amaçlar için kullanılmaktaydı. Fakat de i en günümüz dünyasında, milyonlarca insan e-mail ve web aramaları için kullanmaktadır. Böylelikle BE geleneksel veritabanı arama yöntemlerinin önüne geçmeye başlamıştır. BE bu tanımların dışında, yapılandırılmamış materyalleri içeren, temiz olmayan veriler ve anlamsız veriler ile ilgili problemleri de kapsar.

Birçok Bilgiye Erişim teknolojisinde kullanılan iki temel sunum yöntemleri vardır; vektör uzay modeli ve gizli anlambilimsel dizinleme (latent semantic indexing). Vektör uzayı modeli, belgeleri ve sorguları sunma maliyetini minimize edebilir. Belirli bir sorgunun kriterini, sırasıyla olası belgeleri ve özel sorguyu sunan iki vektör arasındaki öklit uzaklığını hesaplayarak karşılayan belgeleri etkin bir şekilde bulabilir. Gizli anlambilimsel dizinleme, özellikle çok anlamlılık ve çok anlamlılık gibi vektör uzayı modeline ait bazı kısıtlamaları dengelemek için geliştirilmiştir [25,26,27].

Bilgiye Erişim sistemlerinde kullanılan standart iki ölçüt vardır [27].

a) Doğruluk (Recall): Araştırmacı tarama yaptığı konularda bütün kaynaklara erişmek istemektedir. Bilgi sistemlerinde araştırmacının bu isteğinin karşılanma derecesi Doğruluk ile ifade edilir. Doğruluk, bir bilgi sisteminin sorgu ile ilgili olarak bulunduğu yayınların içindeki gerçekten sorgu ile ilgili olan yayınların sayısının veritabanında bulunan ilgili yayınların sayısına oranını gösterir [19, 26,27].

$$\text{Doğruluk} = \frac{\text{Veritabanı içinde dönen ilgili belge sayısı}}{\text{İlgili toplam doküman}} \quad (2.1)$$

b) Duyarlık (Precision): Araştırmacı istediği bilgileri çok fazla zaman harcamadan bulmak istemektedir. Zaman söz konusu olunca ilk akla gelen bilgi sisteminin tarama hızıdır. Ancak hızlı bir tarama sistemi araştırmacının amacı açısından yeterli değildir. Araştırmacının bilgi sisteminin kendisine sorgu ile ilgili olarak gösterdiği yayınlarda gerçekten ilgili olanları seçmesi gerekmektedir. Araştırmacının zamanının büyük bir kısmı da bu evrede harcanmaktadır. Araştırılan yayınları bulma süresini doğrudan etkileyen ve tarama sonuç listesinin iyiliğini gösteren bu özellik ise Duyarlık olarak adlandırılır. Duyarlık bir bilgi sisteminin sorgu ile ilgili olarak bulunduğu yayınların içindeki kullanıcının istediği yayınların sayısının bulunan yayınların sayısına oranıdır [19, 26,27].

$$\text{Duyarlık} = \frac{\text{Veritabanı içinde dönen ilgili belge sayısı}}{\text{geri döner doküman}} \quad (2.2)$$

Doğruluk ve Duyarlık ölçümlerinin her ikisini birden arttırmak bilgilerin tasnif edilmesi ile olur. Bu konudaki robotların Doğruluk(Recall) ve Precision(Duyarlık) oranları düşüktür. Kütüphanelerin ise yüksektir.

Bilgiye Erişim sistemlerinde doğruluk verme önemli bir rol oynar ve birçok farklı doğruluk verme modeli geliştirilmiştir. En yaygın olarak kullanılan model, yerel(local) ve genel(global) doğruluk verme yöntemlerinin bir arada kullanılmasıdır. Yerel doğruluk

vermede terim frekansı (term frequency) tf , genel a ırlık vermede ise idf , (inverse document frequency) ters doküman frekansı kullanılır [19, 26,27].

Terim Frekansı (tf), bir doküman içerisinde bir terimin tekrar sıklı ıdır Ters Doküman Frekansı(idf) bir terimin bütün doküman koleksiyonu içindeki önemidir ve a a ıda gösterildi i gibi Denklem 2.3 ile hesaplanır. Bunlara ba lı olarak terim a ırlı ı formülü Denklem 2.4'de görölmektedir. Denklem 2.4'deki N de eri, veri seti içindeki toplam belge sayısını göstermektedir.

$$idf_i = \frac{\log N}{df_i} \quad (2.3)$$

$$w_i = tf_i * idf_i \quad (2.4)$$

Bilgiye Eri im terimini Türkçe'de ilk kez 1971'de Aydın Köksal kullanmı tır. Köksal daha sonra bu kavramı Bili im Terimleri Sözlü ü'nde öyle tanımlanmı tır. Bir bilgiye eri im dizgesini (sistemini) kullanarak içerik bakımından ara tırılan konu ve kavramlarla ilgili olabilecek genellikle varlı ı bile bilenmeyen belgelerin izini bulmayı amaçlayan ara tırmadır. En eski Bilgiye Eri im çalı ması yine aynı ki i tarafından 1981 yılında gerçekte tirilmi ve 12 sorgu ile bilgisayar bilimindeki 570 belge kullanılmı tır [27].

Türkçe'de bu konudaki di er çalı malara bakılacak olursa, Solak ve Can (1994) 533 haber makalesine ait yı ını ve 71 sorguyu kullanmı lardır. Solak ve Can'ın çalı ması, kök bulma algoritması verilen bir kelimeyi sözlükte aramayı, kelimenin sonundan bir karakter silmeyi ve daha sonra yapısal analizi yapılandırmayı temel almaktadır [27].

Ekmeçio lu ve Willett (2000), 6289 boyutunda bir Türkçe haber belge yı ını ve 50 sorguyu kullanmı lardır. Sadece sorgu kelimelerini köklerine ayırmı lar ve

köklerine ayrılması ve ayrılmaması sorgu kelimelerini kullanarak kazanım etkililiğini karşılamaktadır [27].

Sever ve Bitirim (2003), çalışmalarında 2468 kanun belgesi ve 15 sorguyu temel alan sisteme ait uygulamayı tanımlamaktadır. İlk önce, yeni bir kök bulucunun daha üstün performansa sahip olduğunu ispatlamaktadır. Daha sonra, çekimli ve türetmeli kök bulucuların, kök bulamama durumuyla karşılaştırıldığında erişim duyarlılığı açısından % 25 civarında ilerleme sağladığını göstermektedir [27].

Pembe ve Say (2004) Türkçe Bilgiye Erişim problemini Türkçenin morfolojik, lexico-semantik ve sözdizimsel seviyelerindeki bilgileri kullanarak çalışmaktadır. Bazı sorgu zenginleştirme teknikleri ile kök bulucunun etkilerini tartışmaktadır. Deneylerinde, webden elde edilen farklı konuları ele alan 615 Türkçe belgeyi ve 5 uzun doğal dil sorgularını kullanmaktadır. Yedi farklı dizinleme ve kazanım kombinasyonunu kullanarak ve performans etkilerini ölçmektedir [27].

2.3.2. Bilgi çıkarımı (Information extraction)

Bilgi Çıkarımı konusu, genellikle bir metin üzerinde doğal dil işleme kullanılarak belirli kriterdeki bilgileri elde etmeyi hedefler. Bu işlem sırasında örneğin bir kalıba uygun olan verilerin çıkarılması istenebilir. Amaç çok miktardaki veriyi otomatik olarak işleyen bir yazılım üreterek insan müdahalesini en az seviyeye indirmektir. Bilginin çıkarılacağı ortam genellikle yazılı metinlerdir ancak bu metinlerin bulunacağı ortamlar değişebilir örneğin veri tabanları, internet üzerindeki dokümanlar veya taranmış metinler bu verinin kaynağı olabilir [28].

Bilgi Çıkarma yöntemleri metin içindeki unsurları varlıkları otomatik olarak çıkarır ve bunlar arasındaki ilişkileri ortaya koyar. Metin içindeki cümleler ve paragraflar içerdikleri önermelerle varlıklara ait bilgiler taşıır. Bilgi çıkarma teknikleri bu önermelere bağlı olarak belgeyi oluşturan varlıkları ve bu varlıklar arasındaki ilişkileri çıkarırlar [28,29].

Bilgi çıkarımı başka bir ifade ile geniş ölçekli bilgilerden özet çıkarılması olarak adlandırılabilir. Başka bir ifadeyle büyük veriyi onların içerisinden özet bilgiler elde edilmesidir. Anahtar kelimeler veya örnek dokümanlar gibi kullanıcı girişiyle bağlantılı olan bilgi ya da dokümanların bulunması bilgi çıkarımı örnekleridir. Bu çalışmalar sonucunda web sayfalarından bilgiler karıştırılarak bulunabilir, geniş ölçekli metinlerden özet bilgiler çıkarılabilir, sorgulara karşılık gelen ifadeler bulunabilir [30].

Bilgi çıkarım sürecinin en zor adımlarından birisi de veriyi belirli bir yapıya oturtmaktır. Örneğin internet üzerinde yayınlanan verilerin herhangi bir standart yapısı bulunmamakta, veriler dağınık halde istenildiği gibi yayınlanmaktadır.

Bilgi erişim yöntemlerine nazaran daha etkin sonuçlar elde edilmesini sağlayan bilgi çıkarma tekniklerinin avantajı belge içindeki içeriğin anlamını ön plana çıkaran terimlerin ve terimler arasındaki ilişkilerin bulunmasında yatar. Ancak bazen belgelerin incelenmesindeki amaç, daha önceden fark edilmemiş gerçeklerin ve ilişkilerin ortaya çıkarılmasıdır. Bu amaçla devreye bilgi keşif teknikleri girer. Bilgi keşif için kullanılan yöntemler metnin içeriklerini derler, birbiri ile entegre eder ve başka kaynaklardan elde edilen sonuçlarla birleştirilerek üst seviye bir anlam ve ilişki kümesi oluşturmaya çalışır. Özellikle konuya bağlı olarak terimler ve terimler arasındaki ilişkilerin üzerine de çalışılır ve konuya özel yapılar ve fonksiyonlara bağlı bir ilişki kümesi oluşturulur. Bu amaçla geliştirilen sistemlerin sadece belgeleri değil veritabanlarındaki verileri de kullanması gerekir [31].

Bilgi çıkarım i lemi, temelde anahtar kelime ve/veya benzerlik tabanlı çıkarımlara dayanmaktadır [31]. Anahtar kelime tabanlı bilgi çıkarımında, herhangi bir doküman ya da metinden bilgi çıkarılırken anahtar kelimelerden oluşan bir küme oluşturulur.

Benzerlik tabanlı çıkarım sistemleri ortak anahtar kelimeler kümesini temel alarak, benzer dokümanları bulmaktadır. Bu tür bir çıkarımın çıktısı, kelimelere yakınlığı ve birbirleriyle ilişkili derecelerini temel almaktadır. Günümüzde internet ve bilgi teknolojilerinin hızla gelişmesi ve insanların hayatında önemli bir yer tutması sebebiyle, bu ortamlardan bilgi çıkarımı önem kazanmıştır. Herhangi bir ürünün satış sitelerinden aranması ve karlı olarak değerlendirilmesinden, elektronik posta içeriklerinin yorumlanmasına kadar çeşitli uygulamalar internetten bilgi çıkarım i lemine örnek olarak düşünülebilir [28,32,33].

Bilgi Çıkarım sistemi sonuçlarının değerlendirilmesinde bilgi erişim sistemlerinde olduğu gibi duyarlılık ve doğruluk ölçütleri kullanılmaktadır. Fakat burada belgeler yerine, yapılan tahminler ölçüm değeri kenleri olarak kullanılmaktadır. Duyarlılık, sistemin doğru yaptığı tahminlerin tüm tahminlere bölümü ile hesaplanmaktadır. Doğruluk ise sistemin yaptığı doğru tahminlerin metinde bulunan bütün varlıkların sayısına bölünmesi ile elde edilmektedir [34,35].

Bilgi Çıkarımı konusunda literatürde birçok çalışmaya mevcuttur. Yapılan çalışmalar üzerinde çalışılan belgelerin metin özelliklerine göre farklılıklar göstermektedir. Eğer bilgi çıkarımı yapılan belgeler içinde herhangi bir yapısal özellik taşımayan düz yazıdan oluşuyorsa dilbilimi ile ilgili özellikler önem taşır. Öte yandan belgeler bir yapısal düzen taşıyorsa örneğin web sayfaları gibi o zaman çıkarılacak alanlar arasındaki noktalama işaretleri vb. özellikler önem kazanır. Web sayfaları gibi yapısal metinlerden bilgi çıkarımı için kullanılan algoritmalara örnekler aşağıda verilmiştir [35].

WHISK, bilgi çıkarım algoritması hem düz yazılarda hem de yapısal özellikler içeren belgelerde çalışabilir. Bu algoritma öğrenme kümesini kullanarak düzenli ifadeler (regular expression) tarzında kurallar öğrenir. Öğrenilen bu kurallar ile bilgi çıkarımı işlemi gerçekleştirilir [36].

SRV (Stochastic Real Valued) Algoritması Freitag tarafından önerilmiş bir bilgi çıkarım yöntemidir. Bu algoritma makine öğrenme tekniğinin bilgi çıkarım problemine bir uygulaması olarak görülebilir. Bu algoritma sonucunda da yine kurallar elde edilerek bilgi çıkarımı gerçekleştirilir. Web sayfaları üzerinde başarıyla ve makul sürede çalışabilen bir algoritmadır [36].

SRV algoritmasına ilave olarak literatürdeki iki farklı bilgi çıkarım yaklaşımı olarak RAPIER ve HMM tabanlı bilgi çıkarım yaklaşımı Sanner tarafından gerçekleştirilmiştir [36].

RAPIER bilgi çıkarım algoritması bir çeşit “detaydan-genele” (bottom-up) kural öğrenme algoritmasıdır. Algoritma önce bir ders sayfası için detaylı bir kural öğrenir. Daha sonra karılaşılan diğer bir etiketlenmiş veriler için bu kural mümkün olduğunca genelleştirmeye çalışır ki yeni görülen örnek de bu kural tarafından kabul edilsin. Eğer başka öğrenilen kural bu örneği kapsayacak şekilde genelleştiremiyorsa bu örnek için detaylı bir kural öğrenilir. Sonuç olarak öğrenme işlemi sonunda bir Bilgi Çıkarım kümesi elde edilmiş olur [36].

RAPIER algoritmasının öğrenildiği kurallar birçok çeşit bilgi üzerinde koşullar koyar. Bu bilgiler öyle sıralanabilir; kelimeler, kelimelerin cümle içinde kullanım şekilleri (isim, fiil, zamir, sıfat vs.), kelimelerin anlamsal sınıfları ve çıkarılacak bilgiyi çevreleyen kelimeler. RAPIER algoritması kelimelerin cümle içindeki kullanım şekillerine gereksinim duyduğu için literatürde kullanılan bir program olan Brill’in konuşma kısmı etiketleme (Part of Speech Tagger) programı kullanılabilir [36].

Sanner'in HMM bilgi çıkarımı programı Saklı Markov Modeli (Hidden Markov Model) tabanlı Bilgi Çıkarımı algoritmasını çalıştırmaktadır. Saklı Markov Modeli etiketlenmiş öğrenme kümesini kullanarak bir durum geçi (bağlantı) diyagramı öğrenir. Sanner HMM algoritmasının öğrendiği durum geçi diyagramı 6 farklı tip durum bilgisi içerir. BA LANGIÇ, B T , HEDEF, ÖN EK, SON EK, VE ARKA PLAN. Bu diyagram ilk başta BA LANGIÇ durumundadır. Durumlar arasındaki geçişler kelime ve ihtimal değerleri ile sağlanır. Bu kelime ve ihtimal değerleri öğrenme kümesinden öğrenilir. Bir test sayfası geldiğinde ise görülen kelimenin sistemin bulunduğu durumdan diğer bütün durumlara gitme ihtimallerine bakılarak ihtimal değeri en yüksek olan duruma sistem geçer bu kelime işlenmiş olur. Bilgi Çıkarımı işlemi ise HEDEF durum tipindeki durumlarda kullanılan kelimelerin belirlenmesiyle gerçekleştirilmiş olur [36].

Sanner HMM algoritmasında kullanılan model 3 ayrı HMM modelinin birlikte kullanılmasıyla oluşur: tekdüze model (uniform model), bağlam modeli (context model) ve tam geçişli model (full transition model). Bu modellerin farklı alanlar için farklı derecede önemleri vardır. Sanner HMM programı her bir alan için bu modellerin önemini ifade eden katsayı değerlerinin belirtilmesine konfigürasyon dosyaları sayesinde izin verir [36].

2.3.3. Bilgiye erişim ve bilgi çıkarımının gerçekleştirilmesi

Bilgi Çıkarımı, bilgi parçalarını çıkarmak için doğal dil işleme temel alan bir teknolojidir. Bu süreç girdi olarak metinleri alır ve çıktı olarak belirli bir formatta açık şekilde ifade edilebilecek veriler üretir. Bu veri kullanıcıların görüntü elde etmesi için doğrudan kullanılabilir veya daha sonra analiz etmek için veri tabanında veya elektronik tablolarda saklanabilir veya Google gibi internet arama motorlarında olduğu gibi Bilgiye Erişimi uygulamalarında dizinleme amaçlarını yerine getirmek için kullanılabilir [37].

Bilgi Çıkarımı, Bilgiye Eri imden oldukça farklıdır;

- Bilgiye Eri im sistemi uygun metinleri bulur ve bunları kullanıcıya sunar.
- Bilgi Çıkarımı uygulaması metinleri analiz eder ve sadece kullanıcıların ilgilendikleri metinlerden özel bilgi elde eder [37]

Bir örnekle açıklanmak istenirse, tarım ürünleri pazarlarını ilgilendiren ticari grup yapılarından bilgi bekleyen bir Bilgiye Eri im sistemi kullanıcısı uygun kelime listesini girecektir ve kar ılı nda olası e le meleri içeren belge kümesine (örne in gazete makaleleri) ula acaktır. Daha sonra kullanıcı belgeleri okuyacaktır ve bilgilerin içerisinde bir ayıklama i lemi gerçekle tirecektir. lem uygulandıktan sonra elektronik tablo halinde bilgi giri i yapılabilir ve bunlardan rapor veya sunu çizelgeleri olu turulabilir. Bunun tersine bir Bilgi Çıkarımı sistemi uygun irket ve grup adlarını do rudan ilgilendiren de erleri otomatik olarak elektronik tablo halinde sunacaktır [37].

Bilgiye Eri im sistemi ile kar ıla tırıldı nda Bilgi Çıkarımının avantajları ve dezavantajları vardır. Bilgi Çıkarım sistemleri i lem olu turma için daha zor ve daha bilgi duyarlı bir sistemdir. Özel alanlara ve senaryolara ba lı olan de i ken derecelerine göre çalı ır. Bilgi Çıkarımı, Bilgi Eri imden hesaplama açısından daha duyarlıdır. Bununla birlikte, geni çaptaki metin belgeleri ile ilgili uygulamalarda Bilgi Çıkarımı, Bilgiye Eri imden potansiyel olarak daha etkin çalı an bir sistemdir. Bunun nedeni insanların metin okuma sırasında kaybetti i zamanı önemli ölçüde azaltma potansiyeline sahip olmasından gelmektedir. Ayrıca, sonuçların farklı birkaç dilde sunulması gerekti i durumlarda, Bilgi Çıkarımının sabit biçimli, kesin sonuç veren yapısı Bilgiye Eri im tarafından bulunan birden çok dili içeren metinlerin yorumlanmasını gerektiren tam çeviri kolaylıklarını sa lama açısından kar ıla tırıldı nda nispeten anla ılır durumlar sa layacaktır [37].

Bu bilgiler ı ında Metin Madencili i metotlarının girdi ve çıktıları Tablo 2.3'te görüldü ü gibi özetlenebilir.

Tablo 2.3. Metin madencili i metotlarının girdi ve çıktıları [8]

| Bilgiye Eri im | Bilgi Çıkarımı | Web Madencili i | Kümeleme |
|--|---|---|--|
| Girdi: Metin Belgesi Kayna ı, Kullanıcı sorgusu (metin tabanlı) Çıktı: Sorgu ile ili kili olan sıralanmı belgeler kümesi | Girdi: Metinsel belgeler kayna ı yi tanımlanmı sınırlandırılmı sorgu Çıktı: li kili bilgi cümleleri li kili bilginin çıkarımı ve ili kili olmayan bilginin yok sayılması Önceden belirlenmi formatta çıktı ve ilgili bilgi linki. | Webteki özel bilginin çıkarımı ve metinsel belgelerin eri imi ve indekslenmesi | Benzer metin belgelerinin toplanması |

2.3.4. Metin madencili inin di er uygulamaları

1. Konu izleme: Kullanıcı profillerini kullanarak ve kullanıcı görü lerinden olu turulan belgelere ba lı olarak kullanıcı için ilginç olabilecek di er belgelerin tahmin edilmesidir [7].

2. Özetleme (Summarization): Okuyucuya zaman kazandırmak amacıyla belgenin aslını bozmadan metne hakim olarak özetlenmesi olarak tanımlanabilir. Ba ka bir deyi le otomatik metin özetleme bir bilgisayar programı aracılı ıyla istenilen metinlerin özetinin çıkarılmasıdır. Belge özetlemenin amacı bir belgenin amacını anlatan kısa bir özetinin otomatik olarak olu turulmasıdır. Etkin bir özetleme sistemi kullanıcıların arama sonucu olarak elde ettikleri belgelerin özetlerine bakarak tüm belgeyi inceleme zorunlulu u olmadan do ru belgeye ula ıp ula amadıklarını belirleyebilmeleridir. De i ik seviyelerde özetleme olu turmak mümkündür. Örne in

sadece anahtar kelimeleri içeren bir özet oluşturulabilir. Yada anahtar kelimeleri içeren cümlelerin seçilmesiyle bir özet oluşturulabilir. Daha ileri seviye özetleme teknikleri anahtar kelimeleri içeren cümleleri alıp anlamlı bir özet oluşturmak için yeniden düzenleyebilirler. Bunun için doğal dil işleme çalışmalarından yararlanırlar [7,38].

Bu konudaki ilk çalışma 1959 yılında Luhn adlı bir bilim adamı tarafından yapılmıştır. Luhn kelimelerin kullanım frekansından yararlanmıştır [38]. 1969 yılında Edmundson, 1989 yılında da Saltın konu ile ilgili çalışmalar yapmıştır. Edmundson'un yaptığı çalışmada kelime frekanslarına ek olarak,

- pucu veren ifadelerin (örneğin, sonuç olarak, özetle, bu makale gösteriyor ki, vs...)
 - Konu başlığı içeren kelimelerin
 - Cümlelerin bulunduğu yerin,
- bu özetleme işlemi sırasında incelendiği yeni özellikleri katmıştır.

Metin Özetleme çeşitlerinden bahsedilecek olunursa iki grupta toplanabilir.

- Cümle Seçerek Özetleme (Extract Summarization)
- Yorumlayarak Özetleme (Abstract Summarization)

Cümle Seçerek Özetlemede (Extract Summarization) özetlenecek metin önemli cümleler, istatistiksel metotlarla sezgisel çıkarımlarla veya bunların ikisinin kombinasyonu ile seçilerek bu cümlelerden oluşan bir özetleme yapılır. Bu özetlemede özeti oluşturulan cümleler içeriği akıllıca değerlendirilerek anlamlı cümleler değil, yazı içinden seçilmiş olan önemli cümlelerdir [38].

Yorumlayarak Özetlemede (Abstract Summarization) özetleme, özetlenecek metnin akıllıca yorumlanması ile yapılır. Bu özetlemede orijinal metindeki ifadeler akıllı bir şekilde kısaltılarak tekrar yazılmaya çalışılır [38].

3. Sınıflandırma: çinde önceden tanımlanmış konu kategorilerinin yer alacağı şekilde bir belgenin ana temalarının tanımlanmasıdır. Çerik bazlı belge yönetimi için belgelere ulaşımda esneklik amaçlanmaktadır. Metin sınıflama çalışması bu amaca ulaşmak için kullanılan bir adımdır ve konuyla ilgili yazılmış metinleri anlamlarına göre daha önceden belirlenmiş sınıflara ayırmaya çalışır. Günümüzde metin sınıflama kontrollü bir kelime haznesine bağlı olarak belgeleri indeksleme, belgeleri filtreleme, otomatik olarak metadata oluşturma web sayfalarını otomatik olarak hiyerarşik düzenlemeye tabi tutma gibi pratik olarak uygulanan pek çok alanda görmek mümkündür [7,38].

4. Kümeleme: Önceden belirlenmiş bir kategoriler kümesine sahip olmaksızın birbirine benzer belgelerin gruplandırılmasıdır.

Metin madenciliğindeki önemli noktalardan biri sınıflandırma ve kümeleme metodlarıdır. Karar ağaçları, makine öğrenmesi, istatistik gibi çeşitli teknikler bu nokta için kullanılmaktadır. Bunların içinden en önemlileri, karar ağaçları, yapay sinir ağları bulanık mantık, yaklaşımlı kümeler ve içerik öğrenmedir. Benzer belgelerin aranması da metin madenciliği uygulamasıdır ve benzerlik ölçümü ve sınıflandırma-kümeleme amaçlarını içerir. Yapay zeka tekniklerinin birçoğu bu amaçlarda kullanılır [39].

Temelde benzer belgelerin aranması için iki farklı yaklaşım vardır. Bunlardan ilki belgelerden anahtar kelimeler çıkarmaktır (anahtar kelime temeline dayanan yaklaşım). Sonrasında bu çıkarım nedeniyle benzerlik araştırılır. Diğer yaklaşım ise belgedeki tüm kelimeleri kullanma temeline dayanır (benzerlik temeline dayanan yaklaşım). Belgenin özellik vektörü belgedeki tüm kelimelerin yardımıyla belirlenir ve benzerlik araştırılır [39].

Bütün kelimeler ve belgeler arasındaki bu özellik, Vektör Uzay Modeli kullanılarak ortaya çıkarılır. Anahtar kelime karar verme işlemi hızlı yapılmasına rağmen bütün kelimelerden oluşan özellik vektörü daha kesin arama yapar. Belge kümeleme ve sınıflandırma yalnızca benzer belgelerin sınıflandırılmasında kullanılmaz aynı zamanda diğer metin madenciliği uygulamalarında da kullanılır. Bununla birlikte önceki çalışmalar temelde kümeleme ve sınıflandırma alanlarına odaklanmıştır. Kümeleme ve sınıflandırma için çeşitli metotlar kullanılmıştır [39].

5. Kavram bağlama-ekleme: Yayın olarak paylaşılan kavramları tanımlayarak ve böyle yaparak da geleneksel arama metotlarını kullanarak aradıklarını bulamayan kullanıcılara yardım ederek ilgili belgelerin bağlanmasıdır [7].

6. Soru cevaplama: Bilginin yönlendirdiği örüntü elemleriyle verilen bir soru için en iyi cevabın bulunmasıdır [7].

Soru cevaplama sistemleri bilgiye ulaşma ihtiyacı ile ortaya çıkmış olan ve genelde bilgisayar destekli yapılardır. Soru-cevap benzerliklerini karşılaştırarak ya da varolan kaynaklar üzerinde yapay zeka gibi insan türevi teknikler uygulanarak sorulara yeni cevaplar üretmeye çalışan sistemler geliştirilmiştir [39]

Metin Madenciliğinin temelini oluşturan Bilgiye Erişim ve Bilgi Çıkarımı alanlarında kullanılan yöntemler, Veri Madenciliğinde olduğu gibi Makine Öğrenmesi algoritmalarına dayanmaktadır. Bu nedenle devam eden kısımda Makine Öğrenmesi konusuna değinilmiştir.

2.4. Makine Öğrenmesi (Machine Learning)

Yapay Zekânın alt bir dalı olan Makine Öğrenmesi bilgisayarların “öğrenme” işlevini sağlayacak algoritma ve tekniklerin gelişimi ile ilgili bir çalışma alanıdır. Makine öğrenmesi insanlardaki geri beslemeli öğrenme yapısının bilgisayara taşınamı hali olarak görülebilir [40].

Makine Öğrenmesi, bilgisayarların algılayıcı verisi ya da veritabanları gibi veri türlerine dayalı öğrenimini olanaklı kılan algoritmaların tasarım ve geliştirme süreçlerini konu edinen bilim dalıdır. Makine öğrenmesinin odaklandığı konu bilgisayarlara karmaşık örüntüleri algılama ve veriye dayalı akılcı kararlar verebilme becerisi kazandırmaktır [40,41].

Makine öğrenmesi, Doğal Dil İşleme, Konu Ma ve El Yazısı Tanıma, Nesne Tanıma, Bilgisayar Oyunları, Robot Hareketleri, Arama Motorları ve Tıbbi Te his gibi birçok alanda kullanılır.

Makine Öğrenmesi üç önemli amaçtan oluşur.

- Dokümanların hazırlanması
- Öğrenme Metotlarının Uygulanması
- Öğrenmenin Performansının Değerlendirilmesi

Makine Öğrenmesinde öncelikle öğrenme yapılacak veri setinin uygulanacak öğrenme metoduna uygun bir şekilde hazırlanması gerekir. Öğrenme metodunda istatistiksel yöntemler kullanılır. Geliştirilen yeni metotlar da istatistiksel temelli metotlardır. Yeni bir metot bulunduktan sonra bu metodun performansı ölçülür ve diğer metotlarla karşılaştırılması yapılır [42].

Makine Ö renmesi kavramının geli mesi ile birlikte çok farklı metotlar ortaya çıkmı tır. Bu metotları kullanan algoritma tekniklerine göre sınıflandırılması öyledir [42]:

- Danı manlı Ö renme
- Danı mansız Ö renme
- Yarı Danı manlı Ö renme
- Ödüllü Ö renme
- Uyum Sa lam ile Ö renme
- Ö renme ile Ö renme

Makine Ö renmesi ve Bilgi Çıkarımı arasındaki ili kiye bakılacak olursa Bilgi Çıkarımı teknolojisi, son yirmi yıldır ilgili amaçları yerine getirmek için geli melerden yararlanmaktadır. Bu ilerlemelerin birço u Do al Dil lemeye deneysel yöntemlerin kullanımını temel alır. Do al Dil lemeye deneysel yakla ımlar tarafından ihtiyaç duyulan bilgi çe idi verildi inde Makine Ö renme teknikleri ço unlukla bunları kazanmak için kullanılmaktadır [43].

Tablo 2.4 bazı alan ba ımlı bilginin kazanımı için en son geli meleri yansıtan farklı tipteki Makine Ö renmesi yakla ımlarının bir sınıflandırılmasını göstermektedir. Bu tabloda bir ba langıç yakla ımına ait farklı versiyonlar ilk hücrede gruplandırılmı tır. Bu ko ul sütunu (KL) yakla ımlar tarafından ö renilen bilgi tipini ifade etmektedir. Üçüncü sütun (Paradigma) yakla ımı ö renme paradigmaları ile ili kilidir. Bu paradigmlar ya teklif niteli inde (örne in sıfır de erli mantık veya nitelik de erli mantık açısından ilgilenilen bir kavramın örneklerini sunmayı temel alır) veya ili kisel (örne in birinci dereceden mantık açısından ilgilenilen bir kavramın örneklerini sunmayı temel alır). Son sütun, ö renme için yakla ım tarafından kullanılan belge tiplerini göstermektedir [37].

Tablo 2.4. Bilgi çıkarımı için kullanılan makine öğrenmesi tekniklerinin sınıflandırılması [37]

| YAKLAŞIM | KL | PARADİGMA | Doküman türü | |
|---|----------|-----------------------------------|---|-----------------------------------|
| AutoSlog[Riloff,1993] AutoSlog-TS[Riloff,1996] [Harabagiu ve Malorano, 2000] PALKA[Kim ve Moldovan,1995] [Chat ve Biermann,1997] TIMES [Chat vd.,1999] [Basili vd.,2000] CRYSTAL [Soderland vd.,1995] [Soderland, 1997] WAVE [Aseltine,1999] ExDISCO [Yangarber,2000] [Yangarber,2003] ESSENCE [Catala,2003] DIPRE [Brin,1998] Snowball [Agichtein ve Gravano 2000] LIEP [Huffman,1995] WHISK[Soderland,1999] | kurallar | önermelerden öğrenme | düz metin | |
| | | | yarı yapısal belge | |
| | | | düz metin | |
| | | | yapılandırılmış belge/ yarı yapısal belge / düz metin | |
| EVIUS [Turmo ve Rodriguez,2002] SRV [Freitag 1998] RAPIER [Califf,1998] [Seymore vd.,1999] [Freitag ve McCallum,1999] [Freitag ve McCallum,2000] [McCallum vd. 2000] [Peshkin ve Pfeffer 2003] [Ray ve Craven 2001] [Skounakis vd. 2003] [Miller vd. 1998] [Miller vd. 2000] [Chieu ve Ng 2002] [Chieu ve Ng 2002] [Cox vd. 2005] [Kambhatla 2004] [Sun vd. 2003] Alice-SVM [Chieu vd. 2003] [Zelenko vd. 2003] [Zhao ve Grishman 2005] ELIE [Finn ve Kushmerick 2004] SNoW-IE [Roth ve Yih 2001] | | | ilişki öğrenme | yarı yapısal belge / düz metin |
| | | | | yarı yapısal belge / düz metin |
| | | | | yarı yapısal belge |
| | | düz metin | | |
| | | yarı yapısal belge / düz metin | | |
| | | yarı yapısal belge düz metin | | |
| | | düz metin | | |
| | | yarı yapısal belge | | |
| | | istatistiksel modeller | önermelerden öğrenme | yarı yapısal belge |
| | | | | düz metin |
| | | yarı yapısal belge / düz metin | | |
| | | yarı yapısal belge düz metin | | |
| | | düz metin | | |
| | | yarı yapısal belge | | |

Günümüzde çok büyük miktardaki metin belgelerini düzenlemek ve organize etmek ve bu verilerden yararlı bilgiler sağlamak için çeşitli çözüm yöntemleri sürekli aranmaktadır. Bu çözüm arayışları için temel olarak iki farklı noktaya odaklanılmaktadır. Bunlardan ilki ön işleme yapmaktır. Ön işleme yapmakla ilgili çeşitli

çalı malar vardır ve kök bulma (stemming) algoritmasını geli tirmek ve kelime listesinin durmasının belirlenmesi gibi konuları içerir [9]. Belgelerin sunumu ise di er inceleme konusudur ve ön i lemeyi de içine alır. Bu a amalar Do al Dil i leme çalı malarını gerektirdi inden izleyen kısımda bu konuya yer verilmi tir.

2.5. Do al Dil i leme

Metin Madencili inde veriden bilgi çıkarma yöntemlerinden biri olan do al dil i leme disiplini ile bilgi çıkarımında daha anlamlı sonuçlar elde edilmeye ba lanmı tır. Do al dil i leme ana i levi bir do al dili çözümlene, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını konu alan bir mühendislik alanıdır. Do al dil i leme çalı maları sayesinde insan-bilgisayar etkile iminin artırılması ba arılmı tır [44].

Metin Madencili inde, belgelerin analizi için, içeri inin anlamını ta ıyan kavramların tespit edilmesi gerekmektedir. Bu kavramlar kelimeler veya kelime grupları ile ifade edilir ve terimler olarak adlandırılır. Belge içindeki terimlerin çıkarılması ba lı ba na bir konudur ve do al dil i leme çalı maları kapsamında incelenen bir alandır. Do al dil i lemenin belge analizi sürecindeki en önemli faydası terimlerin yani kelimelerin ayrı tırılması, eklerinden arındırılarak anlamını kaybetmeyen en kısa biçimlerine dönü türülmesidir. Çünkü aynı anlam için kullanılan kelimeler dilbilgisi kuralları gere i farklı biçimlerde bulunabilir ve bu farklı kullanım biçimleri ortadan kaldırılmadı ı takdirde farklı anlam ta ıyan terimler gibi i leme alınarak, belgelerin gerçek anlamına ula ılmasını engelleyebilirler. Do al dil i leme çalı maları kapsamında yürütülen girişimler dört ana grup altında toplanabilir [45].

- Biçimbirimsel çözümlene (Morfolojik analiz)
- Sözdizimi çözümlene (Sentaktik çözümlene)

- Anlam çözümlemesi (Semantik çözümlemesi)
- Anlam karga asının giderilmesi

a) Morfolojik Analiz (Biçimbirim): Biçimbirim sözcüklerin yapısıyla ile ilgilenir. Türkçe için sözcüklerin türetilmesi ve ekler çok önem ta ır. Her dilde iki farklı ekilde sözcük olu turulabilir. Bunlardan biri çekim, di eri ise türetme yöntemidir. Çekim yoluyla sözcük olu turulurken bir sözcü ün farklı ekilleri kullanılır. Türetme ise var olan eski sözcüklere yapım ekleri eklenmesi yoluyla yeni sözcük olu turma yöntemidir [45,46,47].

b) Sözdizimsel Analiz (Sentaktik Analiz): Bilgisayarla do al dil modellemelerinde anlamsal analize geçmeden önce, kelimeler yı ınının geçerli bir cümle yapısı olu turup olu turmadı ı kontrol edilmelidir. Rasgele kelimelerin yan yana gelmesiyle geçerli bir cümle meydana gelmeyecektir. Geçerli bir cümle yapısı olu turulamadı ı zaman, buradan anlam çıkarılmasını beklemek yanlı olacaktır. “küçük ko mavi” kelimeler toplulu u anlamlı bir yapı olu turmamaktadır [45,46,47].

Sözdizimsel analiz, cümlenin yapısal bir tanımını olu turabilmek için morfolojik analizin sonuçlarını kullanır. Bu i lemi yapmanın amacı, ardı ardına gelen kelime yı ınlarının bu kelimeler yı ınının ifade etti i cümle birimlerini tanımlayan bir yapıya dönü türmektir. Cümle birimleri, kelimeler tamlamalar veya buna benzer cümle parçacıkları olabilir.

Türkçede cümleler en genel ekliyle özne, nesne ve yüklem unsurlarından olu ur. Cümle ile ifade edilmek anlam arttıkça, cümlelere yer tamlayıcısı, zaman tamlayıcısı gibi yeni unsurlar eklenecektir. Bunların yanı sıra, cümlelerin anlamlarını kuvvetlendirmek için edat, ba laç gibi unsurlar da cümlelere eklenebilir.

Dillerin sözdizimi açısından sınıflandırılması önem taşımaktadır. Çeşitli dillerde cümlenin temel ögeleri olan özne (Ö), nesne (N) ve yüklem (Y) düz cümledeki temel diziliminin Ö-N-Y, Ö-Y-N ve Y-Ö-N ekinde olduğu görülmüştür. Bunların yanı sıra, kullanılan diğer unsurlar, zarf tümleci (Z), dolaylı tümleş (D) olarak kabul edilebilir. Türkçenin anlamsal analizinden dolayı, önemsenen unsurlar cümlenin yüklemine yaklaştırılır [45,46,47].

c) Anlamsal Analiz (Semantik Analiz): Bir cümlenin ne demek istediğinin anlaşılması, diğer bir deyişle bir cümle ile ifade edilmek istenilen duygu veya düşüncenin ne olduğunun anlaşılması, anlamsal analiz yardımıyla yapılır.

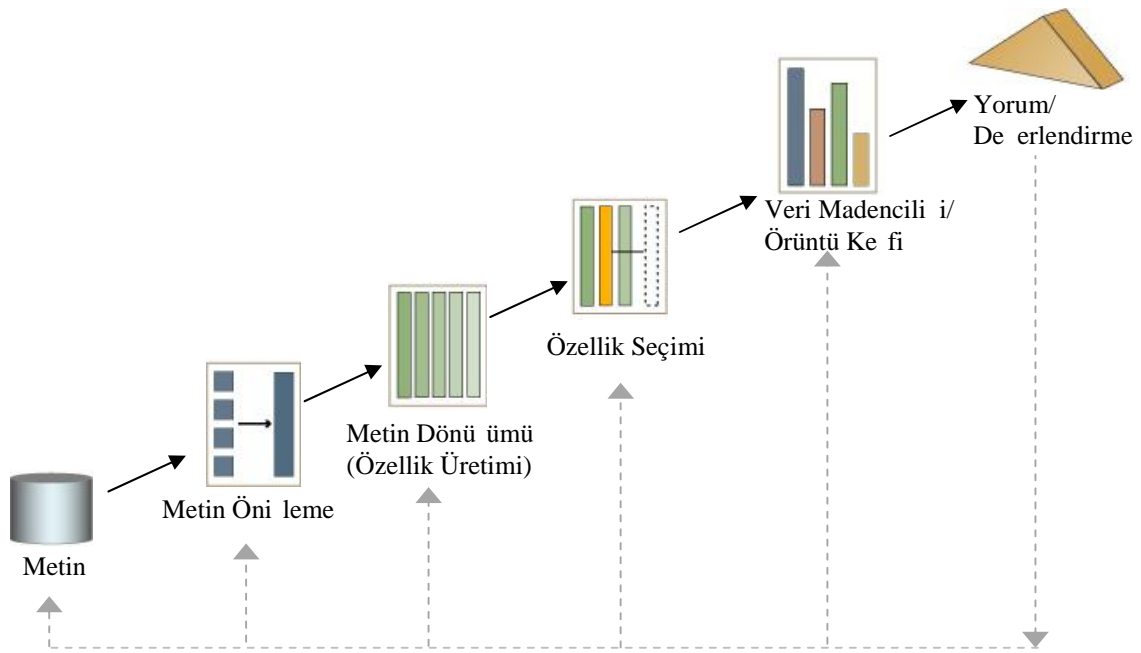
Kavram, bir nesnenin zihindeki tasarım biçimi; kelime ise bu biçimin göstergesidir. Gösterilmek istenilen, kavram; gösteren, kelime olarak adlandırılabilir. Kelimelerin belirttiği kavramların, çevredeki her türlü olayların, hareketlerin ve hayallerin zihindeki değerlendirili biçimine anlam denir. İnsanlar, doğada olup biten her şey için zihninde bir kavram oluşturur. Bu kavramların zihindeki değerlendirili biçimleri, anlamı meydana getirmektedir. Kelimeler, insanların zihinlerinde meydana getirdikleri canlanma ile anlam kazanırlar. Burada özetle, “kelimeler yansıttıkları kavramlara göre değerlendirilirler ve anlamlı duruma gelirler” denilebilir.

Anlamsal analiz yapılırken, öncelikli olarak kelimelerin tek tek veritabanından uygun nesnelere atılması gerekmektedir. Bu işlem, her zaman birebir eşleşme olmayabilir. Diğer bir deyişle, kelimelerin ifade ettikleri anlamlar her zaman bir tane olmayabilir. Ayrık kelimelerin bir cümledeki doğru anlamını bulma işlemine “kelime anlam berraklaştırılması” denir. Bu işlem, cümle içinde geçen bir kelimenin sözlükteki anlamlarının belirlenip bunlardan uygun olanının seçilmesidir. Cümle içinde geçen her bir kelime, diğer kelimelerin doğru anlamlarının ortaya çıkarılması için önem taşımaktadır [44,45,46,48].

Bu bilgiler 11'inde Metin Madencili adımlarını izleyen bölümdeki gibi açıklayabiliriz.

2.6. Metin Madenciliğin Adımları

Büyük miktardaki metinsel verilerden potansiyel olarak yararlı ve önceden bilinmeyen belirli bir önemi olan bilginin çıkarılması olarak nitelendirilen Metin Madenciliği ekilde görüldüğü gibi temelde altı adımdan oluşmaktadır. Metin Madenciliği ilelemleri, Veri Madenciliğine benzer olarak ekil 2.3'teki gibi özetlenebilir.



ekil 2.3. Metin madenciliğinin adımları [8]

Bu bilgilerden sonra Metin Madenciliği ilelemleri ve içerdikleri yöntemler Tablo 2.5'te görüldüğü gibi özetlenebilir. Bunlar hakkında detaylı bilgileri izleyen kısımda yer almaktadır.

Tablo 2.5. Metin madencili i adımları [8]

| Metin | Metin Öni leme | Metin Dönü ümü | Özellik Seçimi | Veri Madencili i/ Bilgi Ke fi | Yorum/ De erlendirme |
|-------|---|--|--|--|----------------------|
| | Söz dizimsel/ Semantik analiz Sözcük türü etiketleme Kelime anlamı belirginle tirme Ayrı tırma (parsing) | Kelime torbası, Kelimeler Kök bulma, Durdurma kelimeleri | Basit hesaplama statistik (boyut azaltma, ili kisiz özellikler) | Sınıflandırma (danı manlı) Kümeleme (Danı mansız) | Analiz Sonuçları |

2.6.1. Metin koleksiyonu olu turma

Igilenilen konularda bilgiye eri im sistemleri kullanılarak metin koleksiyonu olu turma sürecidir. Bu süreç, günümüzde genel olarak internet üzerinden, özellikle Google vb. arama motorları kullanılarak gerçekleştirilmektedir. Çevrim içi veritabanlarının yanı sıra veritabanlarında ya da ki isel bilgisayarlarda bulunan metin türü veriler ile olu turulan koleksiyonlar da metin madencili inde kullanılmaktadır [19,23].

2.6.2. Metin öni leme

Metni kelimelere ayırma, kelimelerin anlamsal de erlerini bulma (isim, sıfat, fiil, zarf, zamir vb.), kelimeleri köklerine ayırma ve gereksiz kelimeleri ayıklama, yazım kurallarına uygunlu unu tespit etmek ve var olan hataları düzeltmek gibi metin belgelerin yapıta ı olan kelimelerle ilgili i lemleri içeren süreçtir.

Metin madenciliğinin en büyük sorunu işleyeceği veri kümesinin yapısal olmamasıdır. Genellikle doğal dil kullanılarak yazılmış dokümanlar üzerinde çalışılan metin madenciliği alanında ön işleme aşaması veri temizlemenin yanında veriyi uygun formata getirme işlemini de gerçekleştirmektedir [35]. Belgeler için dizin oluşturmadan önce yapılacak ön işleme işlemleri şöyledir.

- Doküman dönüşüm

- a) Markup & Format Removal: Dokümanı oluşturan etiket ve özel formatların çıkarılması)
- b) Tokenization: Metin küçük harflere çevrilmesi ve noktalama işaretlerinin çıkarılması)

Sözcük türü etiketlemeye ait genel olarak iki yaklaşım bulunmaktadır. Kural tabanlı yaklaşımlar olası sözcük türü etiketlerini sınırlamak veya bu tip kelimelerin sözcük türü etiketini atamak için bir kelimenin kaynak bilgisini kullanan el yapımı dilbilimsel kurallar kümesini kullanır. Bu anlam çıkarma kuralları dönüşüm tabanlı öğrenme (transformed-based learning) yaklaşımını kullanarak da öğrenilebilir. İstatistiksel yaklaşımlar, Markov modeli, maksimum entropi modeli veya gizli bir Markov modelini içinde tanımlanabilir. Bu yöntemlerden ilk ikisinde metin kesin bir şekilde etiketlenir, diğer yöntem olan gizli markov zincirinde metin bulanık (belirsiz bir şekilde) etiketlenir. Bu etiketlere ait istatistiksel tahminler temel alınarak olası yorumlar ile birlikte seçim işlemi gerçekleştirir [35].

Sözdizimsel Analiz: Kelimelerin cümle içerisinde buldukları yerlere göre farklı anlamları vardır. Sözdizim kelimelerin cümle içindeki anlamlarını takip eder. Sözdizimsel analiz, sözdizimini (syntax) veya cümleyi oluşturan morfolojik öğelerin hiyerarşik kurallara uyumunu karşılaştırarak ölçümlemektir. Bu konu Bölüm 2.2.'de anlatılmıştır.

Semantik Analiz: Semantik sorgulama olarak da bilinen bu yöntem kullanıcıların sorgulamalarını konuma diline yakın ifadelerle oluşturmalarını sağlar. Bu yöntem sorgu ifadesini sözdizimsel ve anlamsal olarak inceler. Yani belgeler sadece içerdikleri kelimeler itibarıyla değil ayrıca daha üst seviye kelime grupları, kelimeler arasındaki ilişkiler bazında da indekslenir. Doğal Dil İşleme kapsamında yer alan bu konuda Bölüm 2.3'te yer verilmiştir.

2.6.3. Metin dönüştürme

Kelimelerin düzgün bir biçimde hecelerine ve eklerine ayrılmasından sonraki işlem, kelimelerin kökünün tespit edilmesidir. İngilizce için kullanılan Porter Stemmer Yöntemi gibi bir kök bulma algoritması kullanmak hızlı olması açısından önemli olsa da Türkçe gibi sondan ekli bir dilde başarı yüzdesi istenen düzeyde değildir ve özel durumları yakalayamamaktadır.

Türkçe alfabesi, Latin karakterlerden oluşur ve 8'i sesli 21'i sessiz olmak üzere 29 harften oluşur. Türkçe, metin akışı ve konumada kullanılan cümle yapısına göre özgür bileşen sırasına sahip bir dildir ve bu bileşenlerin sırası değişebilir ve değişebilir halde anlaşılır [49].

Türkçe, Fince ve Macarca'ya benzer eklemeli bir dildir. Bu gibi diller, farklı son ekler ile kelimeler veya kavramlar arasında sözdizimsel ilişkileri taşıyıcı ve karmaşık kelime yapılarına sahiptir. Türkçe kelimeler, bir köke bağlı olan çekimli ve türetmeli son ekler kullanılarak yapılandırılır.

Türkçe'de, son ek eklenmesi ile ortaya çıkan muhtemel kelime biçimi sayısı 11.313 gibi yüksek bir rakamdır [27].

ngilizce gibi, Türkçe'deki isimler bir cinsiyet ta imaz ve son ekler kelime tipine ba lı olarak de i tirilmez. Bununla birlikte, kelimelere son eklerin eklenmesinde bazı düzensizlikler vardır. Türkçe'de ön ek sadece sürekli kullanıma sahip sıfatların anlamını (ve nadiren zarflarla) arttırmak için kullanılır. “dolu” ve “dopdolu”, “tamam” ve “tastamam” gibi. Bunun gibi peki tirilmi zarflar, haber makaleleri haricinde hikaye anlatmak için daha uygundur. Kullanım olarak eskimi kelimelerde önek kullanımı (biperva-pervasız veya korkusuz anlamına gelir) veya “antisosyal” gibi batı dillerinden gelen ön ek kullanımına Türkçe'de seyrek olarak kar ıla ılır. Buna göre Türkçe için kök bulma yöntemleri hakkındaki uygulamalar a a ıda verilmi tir.

a) Türkçe için kök bulma yöntemleri: Kelimelerin belirtti i anlamları ke fedebilmek için öncelikle köklerini bulmak gereklidir. Dillerin yapısına göre kök bulma çalı maları farklılık göstermektedir. Buna göre Türkçe ve ngilizce dilleri için kök bulma algoritmalarından a a ıda örnekler verilmi tir [49].

- ngilizce için kök bulma: Porter Stemming Algorithm
 - <http://www.tartarus.org/~martin/PorterStemmer/>
- Türkçe için kök bulma: Zemberek projesi
 - <http://code.google.com/p/zemberek/>
 - <https://zemberek.dev.java.net/>

Zemberek, Türkçe ve di er Türkî diller için yazılımlı , biçimbirimsel çözümlene, yazım denetimi, sözcük üretme gibi temel Do al dil i lemlerini yapabilen açık kaynak kodlu Do al Dil i leme kütüphanesidir. Zemberek altyapısının destekledi i Do al Dil i leme i lemleri öyledir [49,50]:

- yazım denetimi
- biçimbirimsel çözümlene
- gövdeleme
- sözcük üretimi
- sözcük önerme

- sadece ASCII karakterle yazılmış sözcükleri Türkçe karakterli hale çevirme
- heceleme

Türkçe için sözlük kullanılarak gerçekleştirilen Zemberek projesine ek olarak şu yaklaşımlardan da bahsedilebilir.

Cebiro lu (2002), Türkçe kelimelerin aldıkları eklerin tümünün Sonlu Durum Makineleriyle ifade ederek “Sözlüksüz Köke Ulaşma Yöntemi” adındaki çalışmasında sözlük kullanmadan da kelime köküne ulaşılabilirliğini göstermiştir.

Can vd. (2008), Türkçe Bilgiye Erişim çalışmasını gerçekleştirebilmek için bu kapsamda Lemmatizer tabanlı kök bulmayı kullanmışlardır. Dil biliminde lemmatisation tek bir kelime olarak analiz edilebilecek farklı çekimli formdaki kelimelerin birlikte gruplanması lemdir ve kelime kökünden farklılık gösterir. Farklılık ise bir kök bulucu tek bir kelime üzerinde içerik bilgisi olmaksızın lemmayı yapar. Bununla birlikte sözcük türüne bağlı olarak farklı anlamlı kelimeler arasında ayırım yapamaz. Ancak kök bulucuların uygulanması daha kolaydır ve daha hızlı çalışır ve bazı uygulamalar için zayıf yanı önemli olmayabilir [27].

b) Snowball: Kök bulmak için tasarlanmış küçük bir karakter lemmadır. Snowball kullanılarak birçok dil için kök bulma algoritmaları geliştirilmiştir. Türkçe için snowball kullanılarak geliştirilen kök bulma algoritmaları Evren Kapusuz tarafından yürütülmektedir [49,51].

c) Kelime Türü: Bir kelimenin kökü bulunduğundan sonraki adım kelimenin türünün bulunmasıdır. Bu lemmayı Pos Tagging denir. Pos Tagging iki aşamadan oluşur. Birincisi eğitim(training) aşamasıdır. Bu fazda kelimelerin kökleri manuel olarak tanımlanmış algoritmalar kullanılarak Machine Learning sistemi vasıtasıyla öğrenilir. İkinci aşama ise tagging aşamasıdır. Bu aşamada, birinci adımda kullanılan algoritma, öğrenilen parametrelere göre yeniden öğrenilir ve kelimeler türlerine ayrılır.

Tagging i lemi için Karar A açları, Saklı Markov Modelleri, Destek Vektör Makineleri veya Dönü üm Temelli Etiketleme gibi yöntemler kullanılır. Bu tip bir geri beslemeli bir sistem kullanılarak İngilizce dilinde %96 gibi bir başarı yakalanabilmektedir.

d) Stopword i lemi: Tekrar eden ve tek başına anlam taşımayan kelimelere stopword kelimeleri denir. Bilgiye Erişimde bir stopword listesi, belgeleri bir dilerinden ayırt etme durumuna etkisi olmayan sıklıkla kullanılan kelimeleri içerir. Stopword kelimelerini azaltmak sorgu sürecinin verimini artırır. Bir stopword listesinin yapılandırması farklı ve bazen rastgele kararları içerir. Bilgiye Erişim literatüründe, verilen özel diller için farklı uzunluklarda stopword listelerini bulmak mümkündür.

e) Bag of Words: Bu amaçla gruplanan tüm dokümanlardaki tüm kelimelerin kullanım sıklıkları hesaplanır ve bir havuzda toplanır. Daha sonrasında ise bu kelimelerin derleri (Word Weighting) hesaplanır. Kelime deri, bir kelimenin belirli bir alan (sıklık, spor, politika, ...) ile ilgili bir metnin içinde bulunma sıklığı olarak açıklanabilir. Örneğin 10000 kelimelik spor kategorisindeki bir haberin içinde gol veya hakem kelimelerinin bulunma sıklıkları, aynı kelimelerin sıklık kategorisindeki bir haber içinde bulunma sıklığına göre kat be kat fazladır [51,52,53].

2.6.4. Özellik seçme

Metin madenciliği uygulamalarında her zaman gürültülü ve önemsiz bilgi içeren metin koleksiyonlarıyla uğraşma ihtiyacı bulunmaktadır. İlgili verilerin saptanması üzerine odaklanan özellik seçme, büyük miktarlardaki veriler üzerinde işlem yapılırken yükünü azaltmada yardımcı olmaktadır. Özellik seçme amacıyla, önlemden geçen metinlerdeki önemli kelimeleri (varlıkları) belirleme (isimler,

taamlamalar, bile ik kelimeler, kısıltmalar, sayılar, tarihler, para birimleri vb.) ve ili kili olmayan özelliklerin çıkarılması, sadece birkaç dokümanda gözlemlenen özelliklerin çıkarılması, birçok dokümanda gözlemlenen özellikleri azaltma vb. i lemleri yapılmaktadır [54,55].

2.6.5. Veri madencili i

Metin Madencili i adımları içinde yer alan Veri Madencili i hakkında bilgi Bölüm 2.3'de verilmi ti.

2.6.6. De erlendirme ve yorumlama

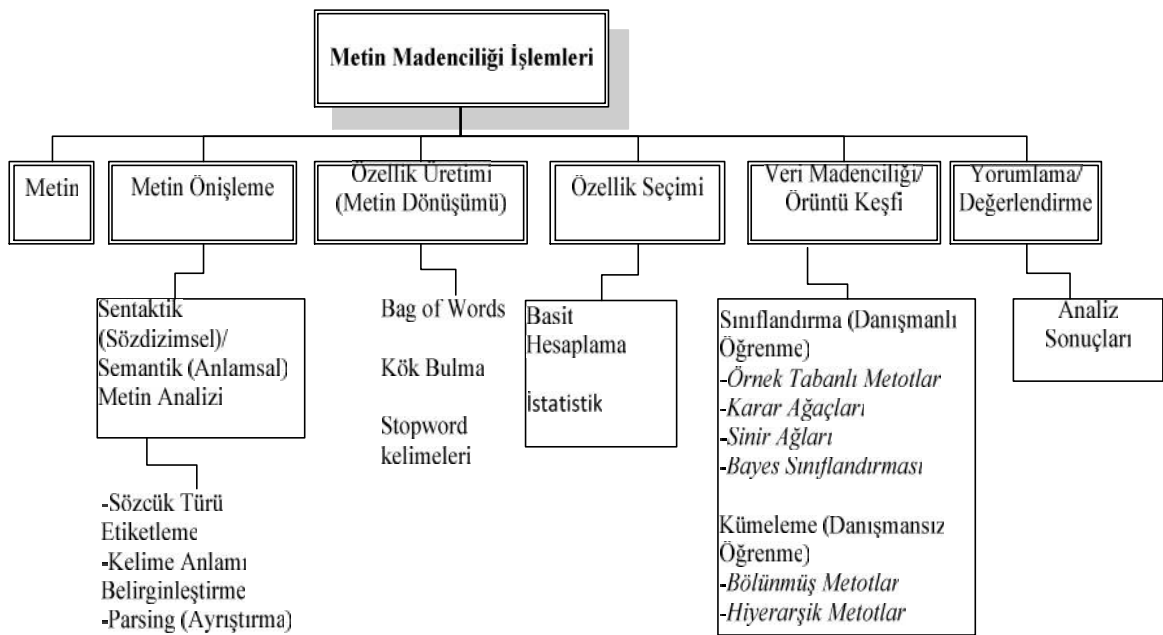
Veri madencili i yöntemleri ile verilerin analizinden elde edilen sonuçların de erlendirilip kullanıcıya uygun ve anla ılır bir eilde sunulması i lemidir.

Bu adımlar kapsamında Metin Madencili inde kullanılan yöntemlere ilave olarak Saraço lu ve di erlerinin (2008), çalı masında Metin Madencili inde kullanılan yakla ımların en önemlileri a ıda sıralanmı tır[7].

- Endüktif karar a açları (Apte, Damerau, & Weiss, 1998; Quinlan, 1986),
- Bayes (Sahami, Dumais, Heckerman, & Horvitz, 1998; Tzeras & Hartmann, 1993),
- Sinir a ları (Wiener, Pederson, & Weigend, 1995),
- K-en yakın kom uluk (Masand, Linoff, & Waltz, 1992; Tan, 2006; Weng & Lin, 2003)
- Küresel k-means (Dhillon & Guan, 2001),

- Destek Vektör Makinesi (Dumais, Platt, Heckerman, & Sahami, 1998; Joachims, 1997),
- Self Organizing Map (Klose et al., 2000; Yang & Lee, 2004; Yang & Lee, 2005)
- Bulanık Mantık (Miyamoto, 2001; Widyantoro & Yen, 2000).

Metin Madenciliği işlemleri ve içerdikleri yaklaşımlar ekil 2.4'teki gibi özetlenebilir.



ekil 2.4. Metin madenciliği adımları ve içerdikleri yaklaşımlar [8]

Metin Madenciliğinin veri madenciliğinde olduğu gibi kurumların karar destek sistemleri için önemli bir yere sahip olduğunu hatırlatmakta fayda vardır. Karar destek sistemleri farklı amaçlara hizmet eden birçok farklı uygulamanın genel adı olarak kullanılmaktadır. Bunlar raporlama araçları, veri ambarı, veri madenciliği, OLAP, Kurumsal Karne, Müteri İlişkileri Yönetimi gibi uygulamalardır. Günümüzde tüm bu uygulamalar, iş zekası verilen tek bir teknoloji ile çözümlenebilmektedir. Bu nedenle veri madenciliği uygulamaları İş Zekası çözümlerini de uzanmaktadır.

zleyen bölümde Zekası kavramı anlatılmış ve Metin Madenciliği arasındaki ilişki kiden bahsedilmiştir.

2.7. Zekası (Business Intelligence)

İşletmelerde farklı sistemler aracılığıyla hergün yüzlerce veri girişi yapılmaktadır. Sistemler arasındaki ilişkilerin giderek karmaşıklıklaştığı ortamda verilerin analiz edilmesi ve yararlı bilgiye ulaşılması, veri ambarının depolama alanı olarak ortaya çıkması, verilerin temizlenmesindeki gelişmeler, gelişen yazılım ve donanım kabiliyetleri ve web mimarisinin ortaya çıkması gibi etkilerin hepsi birden günümüzde Zekası olarak adlandırılan uygulamayı ön plana çıkarmaktadır [56].

Zekası sistemleri, endüstride yaygınlıkla kullanılmasına rağmen bu konudaki araştırmalar oldukça sınırlıdır. Zekası terimi, göreceli olarak yeni kabul edilirken bilgisayar temelli zekası sistemlerinin ortaya çıkışı yaklaşık 40 yıl kadar öncesine dayandırılmaktadır. Zekası terimi Karar Destek Sistemleri, Yönetim Bilgi Sistemleri ve Yönetim Bilişim Sistemleri kavramlarının yerini almıştır. Bu kavram ilk olarak 1989'da Howard Dresdner tarafından kullanılmıştır. Temelinde aslında Veri Madenciliği vardır. Ancak, Zekası, Veri Madenciliğini kendi içinde asimile ederek ufkunu genişletmeye başlamıştır. Çünkü Zekası uygulamaları veri olana kadar beklemek yerine, kendi verisini en hızlı biçimde oluşturma yollarını aramaktadır. Veri Madenciliği, Çevrimiçi Analitik İşleme (OLAP, Online Analytical Processing), sorgulama ve raporlama dahil olmak üzere birbiriyle ilgili çeşitli aktivitelerden oluşan bir disiplindir. Ancak başarı için bundan daha fazlasını bilmek gerekmektedir. Firmalar karar verme, maliyet düşürme, yeni fırsatları geliştirmek ve yeniden elden geçirilmesi gereken verimsiz süreçleri belirlemek için

Zekasını kullanırlar. Zekası kavramı veri analizinin yanında raporlama ihtiyaçlarını daha geniş bir açıdan bakarak sağlayan çözümler olarak düşünülebilir.

- Zekasının çe itli tanımları a a ıda yer almaktadır [57].
- Zekası, alınacak kararlara destek olmak üzere i dinamikleri üzerinde kesin ve belirgin bir anlayı geli tirmek için kullanılan metot ve süreçler bütünüdür. Di er bir ifade ile da ıtık veriler kullanılarak stratejik karar alma durumunda olan ki ilere bilgi üretilmesi sürecidir [57,58].
 - Zekası, veri analizi bakı açısı ile ki ilere kararlarını belirlemek ve varsayımlarda bulunmalarına yardımcı olabilmek amacıyla anlamlı bilgilerin toplanması sürecidir. Bilgi bakı açısıyla da verinin incelenmesi, ke fedilmesi ve dönü türülmesi ile bilginin elde edilmesidir [56].
 - Zekası bir organizasyonun ham verilerini analiz etmek için kullandı ı çe itli yazılım uygulamalarının tümünü kapsayan bir ifadedir [58].
 - Zekası sistemleri, veri elde etme, veri depolama ve rekabet üstünlü ü olan karma ık dahili bilgiyi analitik araçlar ile planlayıcılara ve karar vericilere sunan bilgi yönetimi süreçlerini bir araya toplayan sistemlerdir. Bu tanımda ifade edilmek istenen Zekası sistemlerinin karar vericilere yardımcı olacak bilgilerin do ru zamanda, do ru mekanda ve do ru ekilde karar vericilere ula tırılmasını sa lamakta oldu udur [59].
 - Zekası alanında uluslararası çözümler üreten Gartner irketi ara tırma bölümü ba kan yardımcısı olan Andreas Bitterer Zekasını u ekilde tanımlamaktadır. Zekası, kurulu ların etkinlik ve finansal fayda elde etmek amacıyla performansla ilgili gerekli kararları en iyi ekilde verebilmeleri ve ölçümleri yapabilmeleri, performansı en iyi ekilde yönetebilmeleri ve optimize edebilmeleri için bilginin kullanılmasıdır.
 - Bir ba ka tanım ise, Zekası kurumların karar verme ve yönetim kabiliyetlerinin artırılmasına yardımcı olan çok sayıda verinin kurumsal bilgiye dönü türülmesini ve böylelikle kurumların rekabet ortamında avantaj sa lamasına yönelik kavramlar, metotlar, süreçler ve yazılımlar bütünüdür.

Bu tanımlardan da anlaşıldığı üzere Zekası herhangi bir formattaki verinin yararlı bilgiye dönüştürülmesi süreci ile ilgilidir. Bu çalışmada söz konusu veriler ise doğal dille yazılmış metinsel veriler olan tüketici görüşleridir.

Zekası kavramının, Kurumsal Bilgi Sistemleri ve Karar Destek Sistemleri ile ayırım noktası ise Zekasının sadece yöneticilere yönelik olmayıp süreç içerisinde tüm paydaşların dahil edilmesini sağlamasıdır.

Zekasının yetenekleri aşağıdaki gibi sıralanabilir.

- a) Veri Yönetimi :Veri kazanımları, veri üretimi, veri dönüşümleri, veri tümleştirimi, mevcut verilerden bilgiye dönüştürülmesi, büyük çaplı verilerin yönetimi ve bakımı.
- b) Veri Analizi :OLAP (Online Analytical Processing, Çevrimiçi Analitik Süreç) ve OLTP (Online Transactional Processing) gibi veri işleme yetenekleri, veri sorgulama ve raporlama yetenekleri.
- c) Karar Destek :Etilim tahminleri, kararlı tırmalı analizler.
- d) Kurumsal Eniyileme :Modelleme yetenekleri [56,57].

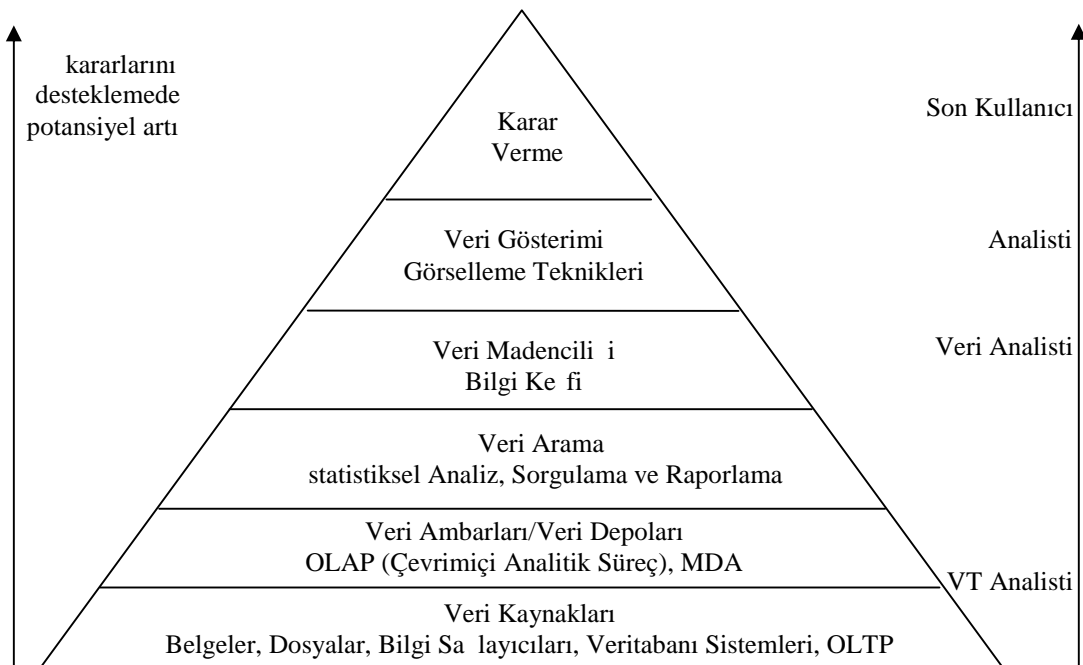
Zekası bahsedilen bu yetenekleri ile süreçlere destek veren Karar Destek Sistemleri ve Veri Madenciliği konuları üzerine, eniyilemeye yönelik modelleme yeteneklerini eklemektedir.

Zekası uygulamaları ayrıca Mü teri lileri Yönetiminde de kullanılmaktadır. Mü teri de erlerinin ve verilen de erin olu um sebeplerinin belirlenmesine ve böylelikle i letme açısından belirli bir e ik de eri üstünde de er ifade eden mü terilerden elde edilen faydanın arttırılmasına yönelik çalışmalara destek verecek araçları içermektedir.

2.7.1. Zekası ve metin madencili i ili kisi

Daha önce de bahsedildi i gibi Zekasının temelinde Veri Madencili i vardır. Metin Madencili i de, incelenen verilerin do al dilde yazılmı ve yapılandırılmamı veriler oldu u durumlardaki Veri Madencili i uygulaması olarak tanımlanabilir. Bu nedenle öncelikle Zekası çalı malarında “Veri Madencili i nerede yer almaktadır?”, sorusuna cevap vermek gerekmektedir.

Zekası terimi karar vermeyi destekleyen ve bilgi teknolojilerini temel alan bütün süreçler, teknikler ve araçlar için genel anlamda kullanılan bir ifadedir. Veri Madencili i, Zekasının önemli bir bile enidir. ekil 2.5’de farklı Zekası çözümlerinin taktik ve stratejik i kararları temelindeki potansiyel de erlerine göre mantıksal pozisyonlarını göstermektedir. Genel olarak yukarı çıkıldıkça karar vermeyi destekleyen bilginin de eri artmaktadır.



ekil 2.5. Veri madencili i ve i zekası [58]
(OLAP: Çevrimiçi Analitik Süreç, OLTP:Çevrimiçi Hareket leme , MDA:Çok Boyutlu Analiz, VT:Veritabanı)

Burada Zekası sisteminin en üst noktasında karar verme süreci görülmektedir. Bunun anlamı, Zekasının öncelikli hedefinin karar verme süreci olmasıdır. Zekası bütün araçları ve uygulamaları ile karar vermeyi destekler. A a ıdan yukarıya çıkıldıkça teknik kullanıcılardan teknik olmayanlara do ru gidilmektedir. Piramidin en altında Zekası sisteminin kaynakları ve elemanları yer almaktadır. Daha üst basamaklarda bu kaynaklardan alınan verilerin i lenip sunuldu u yapılar göze çarpmaktadır. Verilerin karar verici durumundaki son kullanıcılara aktarılması, karar verme sürecini ba latmaktadır.

Zekasının karar vermedeki etkileri sırasıyla a a ıdaki gibi sunulabilir.

- Zekasından sonra Karar Destek Sistemlerinde kontrol ve zamanlama yetene i artmı tır. Kontrolün artması bütün verilerin tek bir veri ambarında saklanması ve depolamanın merkezile mesi ile mümkün olmaktadır.
- Zamanlamadaki iyile me ise periyodik ve anlık raporlarla kullanıcıların do ru, güncel ve i e yarar bilgilere eri mesiyle sa lanmı tır.
- Artan kontrol ve zamanlama yetene i, karar vericilerin daha hızlı ve do ru kararlar almasına imkan tanımı tır.
- Karar sürecinde anlık durumun ortaya konmasının dı ında, veriler üzerinde çok boyutlu analizler gerçekleştirilerek gelece e yönelik tahminler de yapılabilir.
- Elde edilen bilgiler, Zekasının görsel gücü yüksek araçları ile karar vericilere sunulur.

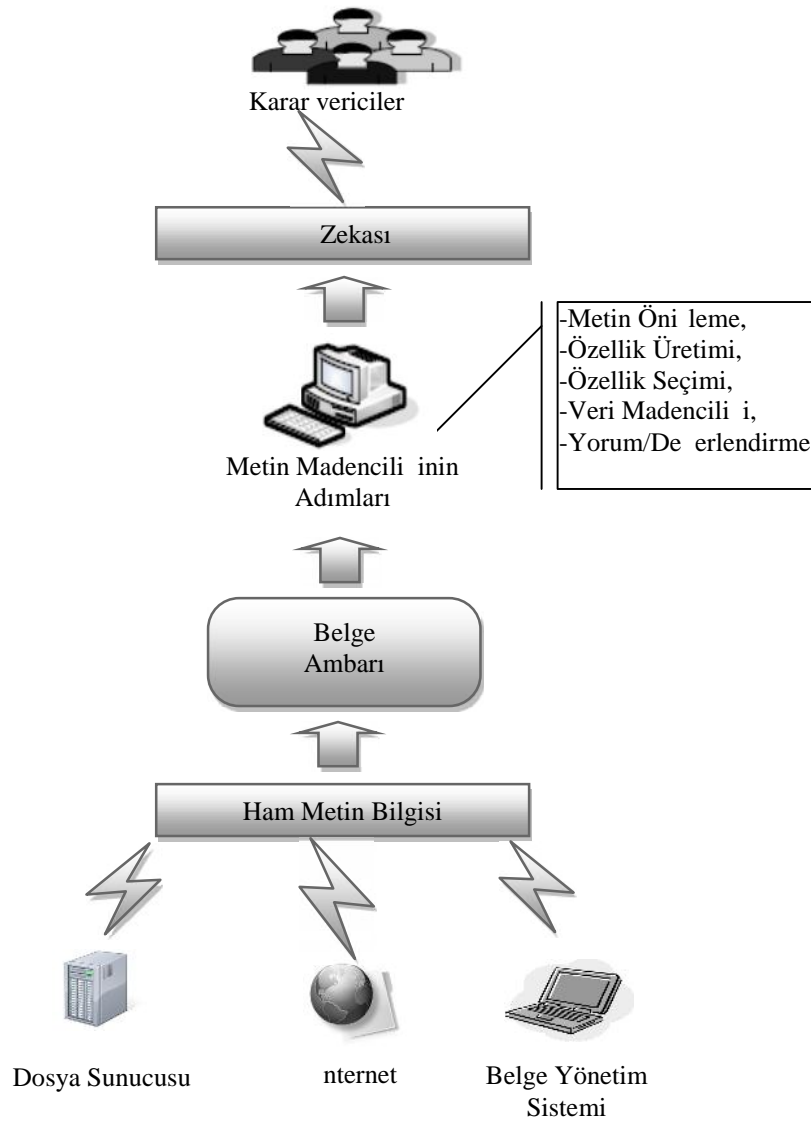
Zekası sistemleri kullanımı kolay ve anla ılır uygulamaları ile hem teknik hem de teknik olmayan kullanıcıların ihtiyaçlarına cevap verilebilecek yetenektedir.

Zekası sistemleri karar verme a amasında yukarıda sayılan birçok olumlu etkisine ra men yapılandırılmamı ve yarı yapılandırılmı veriler söz konusu oldu unda yetersiz kalmaktadır. Büyük i letmelerdeki veri, iyi bir ekilde organize edilmedi i

ve farklı formatlarda bulundu u için yapısı gere i genellikle karma ıktır. İletme yöneticileri, telefon, mail, online anket gibi kaynaklardan mü terilerin tepkilerini toplamaktadırlar. Veriler, perakende satılardan, pazarlama planları, kredi kartı kayıtları ve rakiplerin haberlerinden taranır ve gözden geçirilir. İletmelerin veri toplulu u böylece her geçen gün artar. Bugüne kadar, Veri Madencili i i sistemleri içindeki sayısal ve kısa karakter dizilerini iyi bir ekilde analiz etmeyi ba armı tır. Ancak, bu yapılandırılmı tipteki veriler, bilgi ve bili imi olu turmanın en kapsamlı formatı olan metin verilerini kapsamamaktadır. Bazı metinler proje durum bilgisi, pazarlama raporları, endüstri düzenlemelerinin detayları, rakiplerin rekabet stratejileri, patent uygulamalarında yeni teknolojilerin tanımları gibi i sistemleri için büyük öneme sahiptirler. Burada çok fazla bilgi bulunmaktadır. Bu nedenle, metin biçiminde büyük boyuttaki bilgi kayna ından faydalı bilgiyi çıkarıp bulmak için geli en bir teknoloji olan Metin Madencili i tekniklerine ba vurulmalıdır. Zekası kapsamında kullanılan Veri Madencili i uygulamaları do al dilleri temel alan bu tip verileri i leme ve bilgi olu turma konusunda bir etkiye sahip de ildir. Veri Madencili i, yeni teknoloji olan Metin Madencili ini geli tirmek için bir temel ve güçlü teknikler sa lamaktadır. İletmeler için en acil çözüm bekleyen konu metinlerden potansiyel i zekasını ortaya çıkarmaktır. Zekası kavramının kapsamı, i letme açısından rekabet avantajı sa lamak için, yeni ve güçlü bir araç olan Metin Madencili i yöntemleri ile geli tirilmelidir. Bu da i sistemlerinin büyük bölümünü içine alan metinsel verilerden bilgi ke finin bu kapsama dahil edilmesiyle olur.

Metin Madencili i bilgiyi yönetmek, ara tırmak ve analiz etmek için esnek yaklaşımlar önerebilir. Böylece metinsel malzemeleri de ele alma yetene i ile Veri Madencili i çalı malarını geli tirmektedir. Veri madencili i ve Metin Madencili i benzer konular olmasına ra men, bazı farklılıklara da sahiptirler.

ekil 2.6, Çoklu kaynak olan ham metin bilgisinden Zekasını çıkarmak için Metin Madencili i ve ili kili yöntemlerin kullanım sürecini tanımlamaktadır.



ekil 2.6. zekasını çoklu kaynaklı ham metin bilgisinden çıkarmak için metin madencili ini kullanma süreci[60]

Burada görülen belge ambarının önemi, belge ambarının sorgu ve analiz için ihtiyaçları kar ılamak amacıyla i lem metnini kolayca eri ilebilir ekilde yeniden yapılandırarak ham metinden gerekti i gibi bilgi üretmesini sa lamasıdır.

Veri Madencili inde devasa boyuttaki sayısal veriyi analiz etmek için veri ambarlarının deste ine benzer olarak, Metin Madencili inde belge ambarları Zekasını elde etmek ve karar verme i lemlerini desteklemek için etkin depolar sa lar.

Veri ambarları, ba langıçta farklı kaynaklardan gelen verinin üzerinde daha etkili ve daha kolay sorguların yapılmasını sa lamaktadır. Belge ambarları ise bu sistemde, e-mail, tam metin belgeleri, HTML dosyaları gibi do al dili temel alan yapılandırılmı veya yarı yapılandırılmı yapıdaki birçok kayna ı depolamak için tasarlanmaktadır. Depolanan metin üzerinde gerçekte tirilecek temel i lemler, özetleme, özellik çıkarımı, kümeleme, sınıflama ve konu takibidir. Burada metinsel bilginin gerçek yapısı ise, tam haldeki doküman, dokümanlardan otomatik olarak üretilmi özetler, dokümanların farklı dillerdeki çevirileri, yazarların isimleri, basım tarihleri ve anahtar kelimeler gibi dokümanlar hakkındaki metadatalar ile otomatik olarak çıkarılmı anahtar özellikler, benzer dokümanlar hakkında kümelenmi bilgi, tematik veya konusal indekslerden olu maktadır.

Zekası kapsamında yapılan Metin Madencili i çalı maları, ili kisel belge deposunun kullanımını ile bilgi yönetimi ve karar deste in etkinli ini ve verimlili ini arttırmayı sa lar. Bunun yanısıra, gelecekteki genel e ilimleri tahmin edebilir. Metin madencili i, devralmalar, pazar de i imleri veya zayıf çalı an performansı gibi önemli olaylar için Zekası sistemleri ile kullanıcının bilgiyi akıllıca kullanmasını ve uyarı olu turmasına yardımcı olabilir. Böylece kullanıcılar önleyici tedbirler alabilirler. Aynı zamanda, satı veya mü teri memnuniyetini veya çalı an moralini arttırmaya yardımcı olur [59,60].

Metin Madencili i çalı malarının i letme açısından bakıldı ındaki katkısı, i letmenin

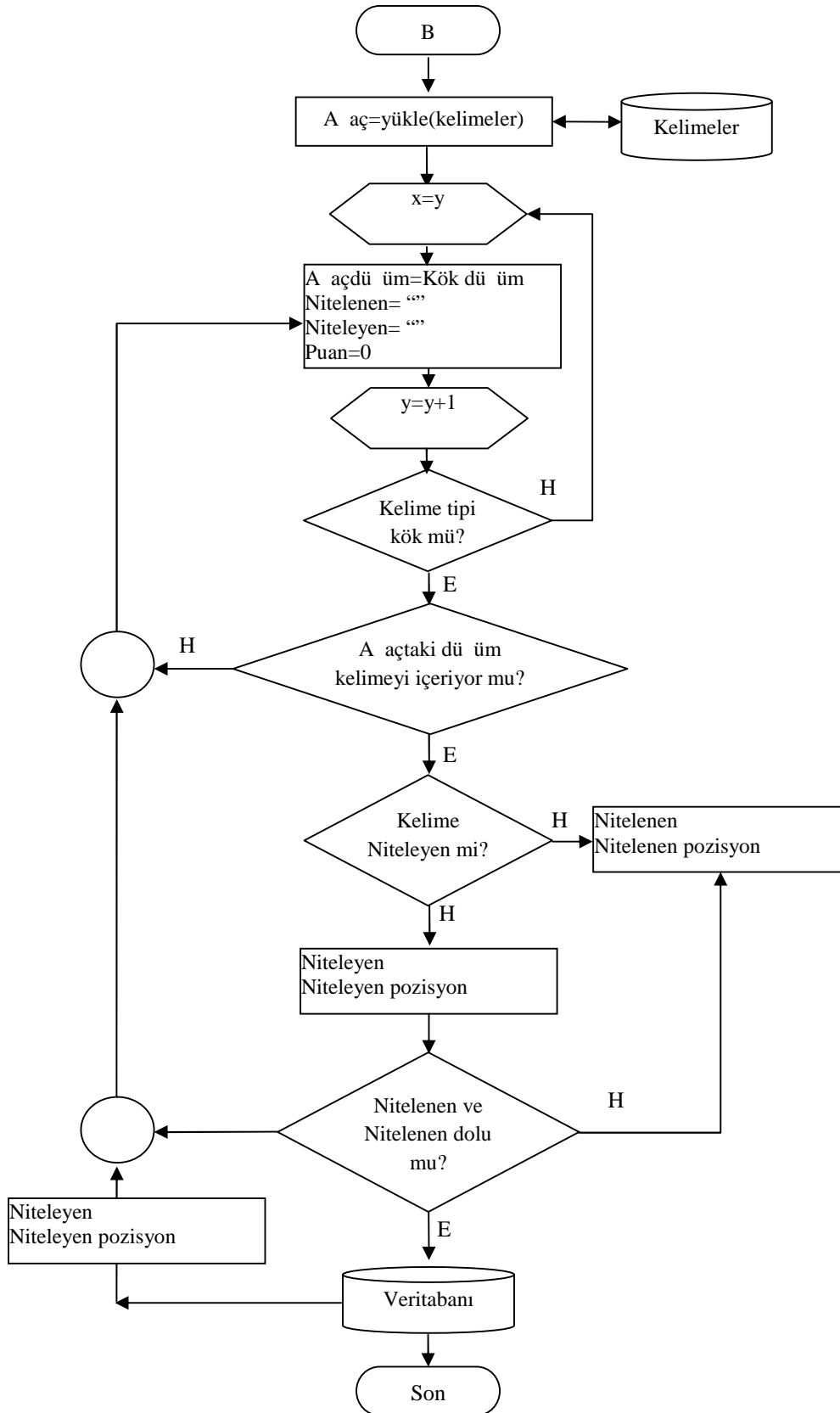
Zekası çalı maları kapsamında görülecektir. Bu kapsamda yapılan çalı malar, i letmeler için ürün geli tirme, ürün iyile tirme gibi konularda i letme için önemli bir rekabet üstünlü ü sa layacaktır. Fakat bu çalı mada elde edilen bilginin özellikle üzerinde durulacak payda ı i letmelerden ziyade tüketicilerdir.

BÖLÜM 3. METİN MADENCİLİK YÖNTEMLERİLE ÜRÜN YORUMLARININ OTOMATİK DEĞERLENDİRİLMESİ

3.1. Giriş

Tez kapsamında, doğal dildeki metinlerden bilgi çıkarımı için ağaç yapısı oluşturularak istenilen özelliklerin tespit edilmesi amacıyla Derinlik Öncelikli Arama algoritması uygulanmış ve sonuçların değerlendirilebilmesi yönelik olarak bir yazılım geliştirilmiştir. Yazılım, Zemberek doğal dil işleme kütüphanesinin yazıldığı, geniş dokümantasyona sahip ve tam nesneye dayalı programlama dili olan Java'da gerçekleştirilmiştir. Yazılımın Java dilinde geliştirilmesi için NetBeans tercih edilmiştir. Netbeans, çok geniş bir kullanıcı tabanı, büyüyen topluluğu, dünya çapında yüze yakın ortakları olan başarılı bir açık kaynak kod projesidir. NetBeans IDE'si programcıların yazma, derleme, hata bulma ve yüklemelerini sağlayan bir araç olarak geliştirme ortamıdır. Java ile yazılmış olmasına rağmen herhangi bir programlama dilini destekleyebilir [63]. Buradan elde edilen veriler SQL veritabanında saklanmaktadır. Bu veriler istenilen yapıya göre sorgulandıında ürün özellikleri hakkında bilgi elde edilmektedir. Tasarlanan sistemin akışması ekil 3.1'de gösterildiği gibidir.

Çalışma kapsamında incelenen ürün yorumları doğal dille yazıldığı için bu metinlerin dil kurallarına uygun bir şekilde düzenlenmesi gerekmektedir. Bu düzenlemenin yapılması için de kullanılan dile ait özellikler ve dilin yapısı bilinmelidir. zleyen kısımda Türkçe'nin özellikleri ve yapısı hakkında bilgi verilmiştir.



ekil 3.1. Bilgi çıkarımı sistemi akı diyagramı

3.2 Türkçe Dilinin Yapısı

Türkçe bitişken bir dildir. Her bir kelime bir kök ve bu kökün sonuna eklenmiş olan eklerden oluşur. Her ek kelimeye yeni bir anlam kazandırır. Kelimelerin aldığı ekler temelde yapım ve çekim ekleri olarak ikiye ayrılır. Yapım ekleri kelime köklerine bitişerek kelimenin hem anlamını hem de türünü belirler. Çekim ekleri ise kelimenin anlamını belirlemeden türünü belirler. Çekim ekleri ise kelimenin anlamını belirlemeden türünü belirler [61,62,64].

Çekim ekleri Tablo 3.1’de verilmiştir.

Tablo 3.1 Çekim ekleri

| Çokluk eki | Soru eki | Yelik ekleri | İlgili eki | Hal ekleri | Etilik eki |
|------------|-----------------------|---|------------|-------------------|--------------------|
| -ler, -lar | -mı, -mi, -mu, -mü | -m, -n, -ın, -in, -un, -ün, -ı, -i, -u, -ü, -sı, -si, -su, -sü, -mız, -miz, -muz, -müz, -nız, -niz, -nuz, -nüz, -ları | -ki | -i, -e, -de, -den | -ce, -çe, -ca, -ça |

Cümle içinde hangi ismin hangi sıfat tarafından nitelendiğini bulmak için kullanılan kurallar Tablo 3.12’de yer almaktadır.

Türkçe’de kelime türleri,

- isim (Ad),
- Sıfat,
- Zamir(Adıl),
- Zarf(Belirteç),
- Edat(İgeç),
- Bağlaç,
- Ünlem,
- Eylem(Fiil) olmak üzere 8 sınıfta incelenirler.

Bazı dilbilimciler kelime çe itlerini üç grupta toplamı lardır: sim (isim, sıfat, zamir), fiil ve edat (edat, ba laç ve ünlemler).

1. sim: Varlıkları ya da eylemleri bildiren, tanıtan kelimelerdir. Sokak, bulut, dü ünçe, Ayla, uyumak ve anlatmak gibi.
2. Sıfat: Varlıkları niteleyen ya da belirten kelimelerdir. Mavi, çalı kan, bu ve birinci gibi.
3. Zamir: simlerin yerini tutan kelimelerdir. Ben, sen, o, bu, u, herkes, kimse ve birisi gibi.
4. Fiil: , olu ve hareket bildiren kelimelerdir. Ko , vur, git ve oku gibi.
5. Zarf: Fiil, sıfat ve zarfların anlamlarını zaman, yer ve yön, durum, nicelik ve soru kavramlarıyla belirleyen ya da sınırlayan kelimelerdir. Dün, yarın, a a ı ve yukarı gibi.
6. Edat: Ba lı ba ına anlamları olmayan, ancak anlamlı kelimelerle ya da cümle içinde kullanıldı ı zaman anlamları sezilen kelimelerdir. çin, ile, e er ve kadar gibi.
7. Ba laç: Ba lı ba ına anlamları olmayan, ancak cümleleri ya da e görevli kelimeleri ya da kelime öbeklerini birbirine ba layan kelimelerdir. le, fakat ve ne..ne gibi.
8. Ünlem: Bir heyecanın etkisiyle a zımızdan çıkarak duygularımızı canlı biçimde anlatmaya yarayan kelimelerdir. Hey!, Ey!, Oh!, Of! ve Eyvah! gibi.

Evrendeki bütün canlı ve cansız varlıkları duygu dü ünçe durumları ve bütün bunların birbirleriyle ilgilerini kar ılayan sözcükler isimler olup ismi niteleyen kelimelere de sıfat denilmektedir. Sıfatlar, isimlerin niteliklerini, ne durumda olduklarını, ço u kez sayılarını, ölçülerini gösteren, soran ya da belirten sözcüklerdir. Bu özelliklerin ortaya çıkması için sıfatlar ismin önüne getirilir[64].

Sıfatlar, Niteleme Sıfatları ve Belirtme Sıfatları olmak üzere ikiye ayrılır. Niteleme sıfatları, varlıkların durumlarını biçimlerini renklerini kısaca nasıl olduklarını bildiren kelimelere verilen isimdir. Nasıl sorusunun kar ılı ı olan sıfatlardır. Bu

çalı mada bir ismin niteli i incelendi i için niteleme sıfatları üzerinde durulmu di erleri ihmal edilmi tir [64].

Di er tür sıfatlar ise belirtme sıfatlarıdır. Bu sıfatlar isimleri göstererek belirten sıfatlardır. Örne in bu, u, o, hangi, kaç, kaçınıcı, bazı, tüm gibi kelimeler belirtme sıfatlarıdır. Bu sıfatlar yorum de erlendirmede katkı sa lamadı ından belirlendikten sonra cümleden çıkarılmı tır.

Bu çalı manın de erlendirme a amasında isim ve sıfatların arasındaki ili kinin belirlenebilmesi için ismin çekim eki alıp almadı ına bakılmı tır. Çünkü Türkçe'de ba lanma durumu eklerle belirlenir. Kelimelerin aldı ı ekler ve ek türleri sonraki kısımda açıklanmı tır.

Tasarlanan sistemin çalı ma adımlarını özetleyen maddeler a a ıda belirtildi i gibidir.

1. Seçilen telefon ürününe ait 100 adet yorum ve ürün özellikleri hakkındaki bilgiler www.hepsiburada.com adresinden hazırlanan özel yazılım elde edilmi tir. ekil 3.2'de ürün yorumlarının elde edildi i kaynak görülmektedir. Bu yorumlar Excel Dosyasına Tablo 3.2'deki gibi kaydedilmi tir.



ekil 3.2. Ürün yorumlarının elde edildi i kaynak [65]

Tablo 3.2. Örnek yorumlar

| id | metin |
|----|---|
| 1 | BEN BU ÜRÜNÜ B R YIL ÖNCE 350 YTL ALDIM B R ÇOK ÖZELL K HEPS BU TELEFONUN Ç NE SI DIRILMI KULLANIMI GAYET BAS T EL N ZDE ÇOK İK B R GÖRÜNTÜSÜ VAR.FAZLA YORUMA GEREK YOK AMA UNU B L N BATARYASI FAZLA UZUN ÖMÜRLÜ DE L,B R SORUNUDA KULLANDI INIZ mpüç LERDEN HERHANG K ADET N TELEFON zsesi OLARAK AYARLIYAMIYORSUNUZ BU BEN M Ç N ÇOK ÖNEML B R ÖZELL KT AMA OLSUN Y NE DE GÜZEL KAMERASI PEK Y DE L AMA Y NE DE F YAT OLARAK DEAL B R ÜRÜN OLMU KAÇIRMAYIN |
| 2 | Bu fiyata bu kalite ancak mükemmel Tek sorun o da sorun sayılırsa arjı bir hafta gitse hiper mükemmel olacaktı Gün boyu yoğun kullanımla 2-3 gün arasında gidiyor. Ama tavsiye edilecek bir telefon. |
| 3 | ürün güzel tek sorunu resim ve kamera çekimlerinde görüntü kötü iyi de il yoksa fiyat ve telefon güzel birde mesaj sesi kötü yaaaa |
| 4 | Telefonu 2-3 hafta önce aldım güzel telefon her özelli i ucuza almak isteyenler için ideal fiyatı da güzel ve hkartını 1 gb almaya gerek yok 512 de yeter 80 arki var bende daha 80 tane daha alacak öyle gösteriyor, imdi telefonun +larına,-lerine geçelim: - erı: 1)Mesaj bildirim sesini kendiniz istedi i gibi ayarlayamıyorsunuz 4 tane var birini seçmek zorundasınız 2)Kamerası pek iyi de il ben de foto raf makinesi oldu u için pek dert etmedim,kamerasında zoom layarak çekerseniz telefonda pek belli olmuyor ama bilgisayara atınca çok bulanık oluyor o yüzden elinizden geldi i kadar zoomlamamaya çalı ın 3) arki dinlerken ba ka i lemler yapamıyorsunuz kapanıyor ancak radyo dinlerken yapabilirsiniz 4)Biri ça rı yapınca 2 saniye geçtikten sonra ses yükseliyor ilk ba ta az çıkıyor sonra birden artıyor birden +ları: 1)Kulaklıkta çok mükemmel bir sesi var çok ses çıkarıyor 2)30 tane radyo kanalı kaydedebilirsiniz ve radyosu çok iyi çekiyor 3) arj konusuna gelince çok abartıya kaçmamak suretiyle arki dinlerseniz günde 15-20 dk konu ursanız en az 2 gün gidiyor arj ve ben telefonu ilk aldı ım gün per embe ak amından pazartesi saat 17 ye kadar dayandı arji üstelik karı tırmama ra men,fazla kurcalamazsanız abartıya kaçmayacak ekilde arji iyi gidiyor 4)Kullanımı kolay 5)Hafif ve ık bir telefon Sonuç:Bu paraya alınacak en iyi telefon her telefonda mutlaka bir kusur vardır bunda da ufak tefek kusurlar olacak ama yakında alı ırsınız ve benim gibi Nokia dan geçenler biraz zorluk çekebilirler ama 3-4 günde alı ırsınız |
| 5 | bu fiyata böyle bir telefon kaçmaz derim. arkada larım ben almadan tavsiye etmemi lerdı ancak yine de aldım ve de çok memnunum. Almayı dü ünenlere tavsiye ederim |
| 6 | BU TELEFONUN KULLANICISIYIM ÇOK MEMNUNUM RADYO mpüç ÇALAR BLUETOOTH KAMERASIYLA TAM ANLAMIYLA TEKNOLOJ AYRICA ÇOK İK B R TAŞARIMA SAH P ÇOK ÖZELL K ARAMAYANLAR C N HAR KULADE B R ÜRÜN |
| 7 | Bu ürünü incelediniz ve birbirinden garip yorumların veya ba ka ürünlerin arasında kaldınız. Ben bu ürünü yaklaşık 5 ay önce hepsiburada.com dan aldım. Bu zamana kadar neler geldi ba ına telefonun ama bir kere bile tamire gitmedi. Suya bile dü tü. Hala çalı ıyor. Özelliklere gelelim. Öncelikle çok mükemmel bir organizatör toplantı randevu gibi hatırlatmalar için ve düzenlemeler için avuç içi bilgisayar gibi. Bu ürünün ayrıca sevdi im bir özelli i uçak modu. Çevrimdi ı oluyorsunuz ister otobüste ister uçakta mpüçleri tıkr tıkr dinliyorsunuz. Menüleri çok sa lam. Kesinlikle eksik bir menü yok. bana göre eksik iki özellik var. biri kızılötesi di eri ise foto raf makinasında fla yok. Batarya ilk kullanımda 2 gün gidiyor. uanda ö retmen hattı kullanıyorum ve ücretsiz görü me yapıyorum.. sürekli sürekli.. ve 2 güne bir arj ediyorum.. Oda yarı yolda kalmamak için.. E er batarya az gelir dersiniz bir batarya daha alın :) Msn olarak kullanıyorum, video çekiyorum. mpüç dinliyorum, radyosunu dinliyorum. Alacaksınız siyah rengini alın.. En ho renk o. Ve kesinlikle SAMSUNG dan a mayın.. Benden tavsiye.. |
| 8 | Ürünün sesi çok çıkmamakla beraber kamerası da pek iç açıcı de il ama onun dı ındaki özellikleri ve görünü ü güzel bu fiyata en iyisi. Bir dezavantajı da herkeste olması ama yine de bu kadar paranız varsa en iyisi alın. |

2. Geliştirilen yazılım yardımıyla Zemberek programı kullanılarak cümlelerin Morfolojik Analizi yapılır. Morfolojik Analiz sonucunda her bir kelimenin türü, ekleri, ek türleri ve kökleri tespit edilmiş olur.
3. Herhangi bir konudaki fikrimizi çoklukla sıfat-isim ve daha az bir eklede fiil-zarf ilikileri ile ifade ederiz. Bu nedenle cümlede niteleyen ve nitelenen kelimelerin ve aralarındaki iliki derecesinin belirlenmesi gerekir. Bu ilikiler Türkçe'nin dilbilgisi yapısına göre oluşturulmuş kurallar yardımıyla otomatik olarak belirlenir.
4. Metin Madenciliği yöntemleri kullanılarak niteleyen(sıfat) kelimeler anlamlarına göre derecelendirilir. Anlamca olumlu kelimeler daha yüksek bir puanla derecelendirilirken olumsuzluk içeren kelimeler düşük puanla temsil edilir. Bu eklede niteleyen kelimeler artık rakamsal verilere dönüştürülmüştür.
5. Oluşturulan sıfat sözlüğünden niteleme dereceleri alınır.
6. Tüm yorumlar incelendiğinde ürünün aynı özelliğe ait sıfatların olumluluk olumsuzluk derecelerinin ortalaması alınarak sonuç memnuniyet derecesine ulaşılmıştır. Bu bilgiler tablo ekinde sunulur.

Yukarıda da ifade edildiği gibi çalışmanın ilk aşaması ürün yorumlarının elde edilmesidir. Bu aşamadan sonra doğru bir değerlendirme yapabilmek için kelimelerin türleri, özellikleri tespit edilmelidir. Buna ilişkin açıklama aşağıda yer almaktadır.

3.3 Özellik Simgeleri ve Sıfatların Tespiti

Yazım hataları düzeltildikten sonra yorumlar ön işleme yöntemlerinden geçirilip işlenmeye hazır hale getirilmelidir. Metin ön işleme, Doğal Dil İşleme(NLP) yöntemlerinin kullanımını gerektirir. Bu aşamadan sonra üzerinde çalışılan dilin kurallarına göre elde edilecek çıktılar aşağıdaki şekilde olacaktır.

- Sözcüklerin kökleri ve türleri bulunur.

- Sözcüklerin aldıkları ekler ve türleri bulunur.

Bu amaçta elde edilen çıktıya ait örnek Tablo 3.3'teki gibidir. Elde edilen çıktıda;

- metin no, sisteme girilen cümle numarasını
- kelime no, o cümlede geçen kelime numaralarını göstermektedir. Cümlede geçen kelime ve aldığı eklerin türleri ayrı ayrı bulunmaktadır.

Tablo 3.3. Cümlelerin morfolojik analizinin sonucu örneği

| id | metin_no | kelime_no | kelime | incelenen | tip | tür |
|----|----------|-----------|--------|-----------|-------|----------------|
| 0 | 0 | 0 | BEN | ben | <kok> | ZAMIR |
| 1 | 0 | 1 | BU | bu | <kok> | ZAMIR |
| 2 | 0 | 2 | ÜRÜNÜ | ürün | <kok> | ISIM |
| 3 | 0 | 2 | ÜRÜNÜ | I | <ek> | ISIM_TAMLAMA_I |
| 4 | 0 | 3 | BİR | bir | <kok> | SAYI |
| 5 | 0 | 4 | YIL | yıl | <kok> | FİİL |

Türkçe'de temel cümle diziliş sırası, özne-tümleç-yüklem şeklindedir. Bu diziliş cümlelerin öğelerini göstermektedir. Fakat burada önemli olan kelime türleri ve birbirleriyle ilişkileridir. İsimlerin hangi sıfatlar tarafından nitelendirildiğini anlamak için isme getirilen çekim ekleri göz önüne alınmalıdır. Sıfat tamlamaları nitelik hakkında fikir vermektedir. Örneğin, “güzel ev.” Bu cümle sıfat ve isimden oluşmuş bir cümledir ve “ev”i nitelendirmektedir. “Evin kapısı mavi renklidir.” Bu cümlede ise “ev” ve “kapı” isim olmasına rağmen nitelenen isim sadece “kapı”dır. “Ev” ismi ek aldığı için sonra gelen kelimenin türüne ve ek alıp almadığına bakılarak hangi kelimenin nitelendirildiğine karar verilir. Sıfatlar ek almadığından sadece isme getirilen çekim ekleri ele alınmalıdır. Kelime türlerinin tespitinde ise Türkçe diline ilişkin çeşitli bilgi sistem problemlerinin çözülmesi için oluşturulmuş açık kodlu bağımsız bir kütüphane olan Zemberek programından faydalanılmıştır. Buna göre öncelikle kelimelerin aldıkları ek türleri ve Zemberek tarafından tespit edilmiştir. Aşağıdaki tablolarda Cebiroğlu(2002) tarafından yapılan çalışmada yer alan tanımlamalar temel alınarak oluşturulmuş ek türleri yer almaktadır [61].

Tablo 3.4. Ço ul eki, iyelik ekleri

| | |
|---------------|--------------------------|
| Ek Türü | Zemberek Kar ılı 1 |
| ÇO UL EK | ISIM_COGUL_LER |
| YEL K EKLER | |
| 1. TEK L AHIS | ISIM_SAHİPLİK_BEN_IM |
| 2. TEK L AHIS | ISIM_SAHİPLİK_SEN_IN |
| 3. TEK L AHIS | ISIM_SAHİPLİK_O_I |
| 1. ÇO UL AHIS | ISIM_SAHİPLİK_BİZ_IMİZ |
| 2. ÇO UL AHIS | ISIM_SAHİPLİK_SİZ_INİZ |
| 3. ÇO UL AHIS | ISIM_SAHİPLİK_ONLAR_LERİ |

Tablo 3.5. Ç çekim durum ekleri

| | | |
|-----------------------|--------------------|---|
| Ek Türü | Zemberek Kar ılı 1 | Ba lanma Durumu |
| DURUM EKLER | | |
| YALIN | ISIM_KOK | |
| -HAL YÜKLEME DURUMU | ISIM_BELIRTME_I | S M GEÇ L EYLEM |
| E-HAL YÖNELME DURUMU | ISIM_YONELME_E | S M GEÇ L EYLEM, S M EYLEMS , S M EDAT |
| DE-HAL BULUNMA DURUMU | ISIM_KALMA_DE | S M EYLEM, S MEYLEMS |
| DEN-HAL ÇIKMA DURUMU | ISIM_CIKMA_DEN | S M ÇEK ML EYLEM, S M F L MS , S M EDAT |
| LG DURUMU | ISIM_TAMLAMA_IN | S M S M, S M EDAT |

Tablo 3.6. Dı çekim durum ekleri, ilgi eki

| | | |
|----------|---------------------|--------------------------|
| Ek türü | Zemberek Kar ılı 1 | Anlam |
| -LE HAL | ISIM_BIRLIKTELIK_LE | B RL KTEL K (arkada -la) |
| -CE HAL | ISIM_GIBI_CE | GÖREL K (arkada -ça) |
| -L HAL | ISIM_BULUNMA_LI | SAH PL K (arkada -lı) |
| -S Z HAL | ISIM_YOKLUK_SIZ | YOKSUNLUK(arkada -sız) |
| LG EK | ISIM_BULUNMA_KI | SAH PL K |

Tablo 3.7. Zaman ekleri

| | |
|---------------------------|-------------------------------|
| Ek türü | Zemberek Kar ılı ı |
| ZAMAN EKLER | |
| MD K ZAMAN | FIIL_SIMDIKIZAMAN_IYOR |
| GEÇM (-D L) ZAMAN | FIIL_GECMISZAMAN_DI |
| DOLAYLI(-M L) GEÇM ZAMAN | FIIL_GECMISZAMAN_MIS |
| GELECEK ZAMAN | FIIL_GELECEKZAMAN_ECEK |
| GEN ZAMAN | FIIL_GENISZAMAN_IR |
| SÜRERL L K | FIIL_MASTAR_MEK+ISIM_KALMA_DE |
| EM R | FIIL_EMIR_O_SIN |
| GEREKL L K | FIIL_ZORUNLULUK_MELI |
| D LEK- ART | FIIL_SART_SE |
| STEK | FIIL_ISTEK_E |
| H KAYE B LE K ZAMAN | IMEK_HIKAYE_DI |
| R VAYET B LE K ZAMAN | IMEK_RIVAYET_MIS |
| ART B LE K ZAMAN | IMEK_SART_SE |

Tablo 3.8. Ki i eklerine kar ılık gelen zaman ekleri

| Ki i Ekleri | Tip1 (geçmi zaman, dilek, art) | Tip2(imdiki, gelecek, -di'li geçmi , geni , gereklilik, sürerlik) | Tip3 (istek) | Tip4 (emir) |
|-------------|--------------------------------|--|--------------|-------------|
| 1.TEK L | M | (Y)UM | YUM | - |
| 2.TEK L | N | SUN | SUN | - |
| 3.TEK L | - | - | - | SUN |
| 1.ÇO UL | K | (Y)UZ | LIM | - |
| 2.ÇO UL | NUZ | SUNUZ | SUNUZ | YUN, YUNUZ |
| 3.ÇO UL | LAR | LAR | LAR | SUNLAR |

Tablo 3.9. Eylem çekim ekleri, yapım ekleri

| Ek türü | Örnek | Zemberek Karılıklı |
|--|---------------------|------------------------|
| EYLEM ÇEKİM EKLERİ | | |
| KARMA İK(ZAMAN,TARZ) | MU | FIIL_DONUSUM_MIS |
| | (Y)ACAĞ | FIIL_DONUSUM_ECEK |
| | MAK | |
| | MAZLUK | |
| | MA | FIIL_DONUSUM_ME |
| | (Y)U | FIIL_DONUSUM_IS |
| | (Y)ACAĞ | |
| | DUK | FIIL_BELIRTME_DIK |
| | (Y)ASU | FIIL_DONUSUM_ESI |
| TASVİR (YETERLİK, VEDİLİK, SÜREKLİLİK, YAKINLIK) | (Y)ABİL(YETERLİLİK) | FIIL_YETENEK_EBİL |
| | (Y)ÜVER(VEDİLİK) | FIIL_TEZLİK_İVER |
| | (Y)AKAL(SÜREKLİLİK) | FIIL_SÜRERLİK_EKAL |
| | (Y)ADUR | FIIL_SÜRERLİK_EDUR |
| | (Y)AĞEL | FIIL_SÜRERLİK_EĞEL |
| | (Y)GÖR | FIIL_SÜRERLİK_EGOR |
| | (Y)AKOY | |
| | (Y)AYAZ | FIIL_YAKLASMA_AYAZ |
| ÇATI EKLER (ÖZNE-YÜKLEM,NESNE-YÜKLEM) | | |
| EDİLGENLİK | (U)L | FIIL_EDILGEN_IL |
| | (U)N | FIIL_EDILGENSESİL_N |
| DÖNÜLÜLÜK | (U)L | |
| | (U)N | |
| TELİK | (U) | FIIL_BERABERLİK_IS |
| NESNELERE GÖRE | | |
| ETTİRGENLİK | DUR | FIIL_ETTIRGEN_TIR |
| | (U)T | FIIL_ETTIRGEN_TEKRAR_T |
| OLDURGANLIK | DUR | |
| | (U)T | FIIL_OLDURGAN_T |
| YAPIM EKLERİ | | |
| SMDEN S M YAPMA | LUK | |
| | CU | |
| | CIK | |
| SMDEN EYLEM YAPMA | LA | ISIM_DONUSUM_LES |
| EYLEMDEN S M YAPMA | | |
| EYLEMDEN EYLEM YAPMA | | |

Tablo 3.9. Eylem çekim ekleri, yapım ekleri (devam)

| | | |
|-----------------------|----------------------|-----------------------|
| ISIM_DONUSUM_LE | | |
| FIILIMSILER | | |
| SIFAT FIIL (ORTAÇLAR) | an, -en, | FIIL_DONUSUM_EN |
| | ası, -esi, | FIIL_DONUSUM_ESI |
| | mez, -maz | FIIL_DONUSUM_MEZ |
| | ecek, -acak | FIIL_DONUSUM_ECEK |
| | mi , -mı , -mu , -mü | FIIL_DONUSUM_MIS |
| S M F L | ı ,i ,u ,ü | FIIL_DONUSUM_IS |
| | me,ma | FIIL_DONUSUM_ME |
| | mak,mek | FIIL_MASTAR_MEK |
| ZARFF L | ınca | |
| | dıkça | FIIL_DEVAMLILIK_DIKCE |
| | dı ında | |
| | ken | |
| | r... -mez | |
| | alı | |
| | erek | |
| | madan | |
| | meksizin | |
| | a...-a | |
| | ıp | FIIL_IMSI_IP |

Belirlenen bu ekler sayesinde niteleyen nitelenen kelimeleri kadar kim, ne, ne zaman sorularına da otomatik olarak cevap verilebilen sistemler geli tirilebilir.

Bu çalı mada seçilen ürün telefondur. Telefona ili kin ürün özellikleri ve bu özellik isimlerini içeren di er kelimeler mevcut yorumlardan manuel olarak tespit edilmi tir.

Bu çalı mada incelenen ürün telefondur. Telefona ait özellik tablosu ve bu özellikleri ifade eden e veya benzer kelimeler Tablo 3.10'da yer almaktadır.

Tablo 3.10. Ürün özellik isimleri ve e veya benzer kelimeler

| No | Özellik | E veya Benzer çeren Anlamlar | | | | | | |
|----|----------------------|------------------------------|---------|-----------|--------------|---------|------------|--------|
| 1 | batarya | arj | pil | | | | | |
| 2 | tasarım | dizayn | görünüm | ekran | dokunmatik | telefon | ürün | görünü |
| 3 | kullanım | kullanılabilirlik | telefon | tu takımı | renk | | | |
| 4 | ses | mp3 | müzik | kulaklık | | | | |
| 5 | kamera | foto raf | resim | çekim | zoom | video | çözünürlük | |
| 6 | fiyat | fiyat/performans | pahalı | ucuz | para | fiyat | | |
| 7 | boyut | ebat | a ırlık | ince | kalın | | | |
| 8 | Hafıza radyo tv | video | radyo | mobil TV | hafıza kartı | GB | Bayt | |
| 9 | ba lantı özellikleri | bluetooth | GPRS | GPS | internet | wifi | Navigasyon | 3G |
| 10 | i letim sistemi | symbian | V30 | yazılım | sürüm | | | |
| 11 | problem | sorun | eksi | | | | | |

Sıfatların anlamca pozitif yada negatiflikleri e it de ildir. Türkçe’de fiiller –ma ekiyle olumsuz hale getirebilirler. Fakat isim veya sıfatların anlamca olumlu olup olmadıkları aldıkları eke veya kelimenin köküne bakılarak belirlenemez. Bu nedenle anlamca olumlu ve olumsuz olan sıfatlar, bu sistemi olu turanlar tarafından belirlenmelidir. Örne in iyi ve kötü kelimeleri sıfattır. (Cümle içinde isim de olabilirler). “yi” anlamca olumlu iken “kötü” olumsuzdur. Bu yüzden sıfatlar derecelendirilmelidir. Bu nedenle niteleme sıfatlarının temsil ettikleri memnuniyet dereceleri 1 ve 5 arasında negatiften pozitive do ru artan bir likert ölçüsünde puanlanmı tır. Buna göre derecelendirilen sıfatlara ait örnek, Tablo 3.11’de verilmi tir. Bu tabloda sadece kesin yargı belirten sıfatlar derecelendirilmi tir. Aynı sıfatlar “de il” kelimesi ile birlikte kullanıldı nda de i en puanları da tabloda yer almaktadır.

Tablo 3.11. Niteleme sıfatları puanları

| sıfat | puan | de il |
|-----------|------|-------|
| acayip | 2 | 3 |
| acemi | 2 | 3 |
| acı | 1 | 4 |
| adil | 5 | 1 |
| aksak | 1 | 4 |
| ba arılı | 5 | 1 |
| ba arısız | 1 | 4 |
| basit | 3 | 2 |
| berbat | 1 | 4 |
| beter | 1 | 4 |

Bu çalı madaki en önemli sorun yapılan yorumlarda hangi ifadenin hangi özelli i niteledi ini bulmaktır. Bu i lemi otomatik olarak bilgisayara yaptırabilmek için bazı kurallara ihtiyacımızın oldu u açıktır. Bir dilde herhangi bir durum hakkındaki görü ü nasıl sorusuna verilen cevaplar belirlemektedir. Nasıl sorusunun cevabı isim-sıfat veya zarf-fiil ili kilerinden olu an bile enlerdir. Bu kelimeler arasındaki ili kiler, dilin yapısına göre kelime türü, kelime kökü ve kelimenin aldığı eklerden faydalanılarak olu turulan kurallar a aç yapısı haline getirilmi ve arama algoritması yardımıyla istenilen kalıplar elde edilmi tir.

3.4. A aç Yapısı ve Derinlik Öncelikli Arama Algoritması

Bu a amadan sonra olu turulan a aç yapısı üzerinde arama yapılabilmesi için niteleyen ve nitelenen kelimelerin tespit edildi i kurallar tablosu çıkarılmı tir. Tablo 3.12’de bu kurallar gözükmetedir. Bu sistemde yazım hataları sistemine benzer olarak niteleme çe itleri puanlandırılmı tir. Buna göre en yüksek puanda olan ili kideki kelimeler niteleyen ve nitelenen olarak belirlenmi tir.

- Niteleyen kelimeler →2,
- Nitelenen kelimeler →1 ile ifade edilmi tir.

Tablo 3.12. A aç yapısında niteleme kuralları

| agac | tip | tur | niteleme | puan |
|------|-----|--------------------|----------|------|
| 1 | kok | SIFAT | 2 | 1 |
| 1 | kok | ISIM | 1 | 0 |
| 2 | kok | ISIM | 1 | 1 |
| 2 | ek | ISIM_BELIRTME_I | 0 | 0 |
| 2 | kok | SIFAT | 2 | 1 |
| 2 | ek | ISIM_TANIMLAMA_DIR | 0 | 0 |
| 3 | kok | ISIM | 0 | 0 |
| 3 | ek | ISIM_TAMLAMA_IN | 0 | 0 |
| 3 | kok | ISIM | 1 | 1 |
| 3 | ek | ISIM_TAMLAMA_I | 0 | 0 |
| 3 | kok | SIFAT | 2 | 1 |
| 3 | kok | SIFAT | 2 | 1 |
| 4 | kok | ISIM | 0 | 0 |
| 4 | ek | ISIM_TAMLAMA_IN | 0 | 0 |
| 4 | kok | SIFAT | 2 | 1 |
| 4 | kok | SIFAT | 2 | 1 |
| 4 | kok | ISIM | 1 | 1 |
| 4 | ek | ISIM_BELIRTME_I | 0 | 0 |
| 5 | kok | ISIM | 1 | 1 |
| 5 | kok | BAGLAC | 0 | 1 |
| 5 | kok | ISIM | 1 | 1 |
| 5 | kok | SIFAT | 2 | 1 |
| 6 | kok | ISIM | 1 | 0 |
| 6 | ek | SIFAT | 2 | 0 |
| 6 | kok | ISIM_BULUNMA_LI | 0 | 1 |
| 6 | kok | ISIM | 1 | 1 |
| 6 | ek | ISIM_TANIMLAMA_DIR | 0 | 1 |
| 7 | kok | SIFAT | 2 | 1 |
| 7 | ek | ISIM_KUCULTME_CIK | 0 | 1 |
| 7 | kok | ISIM | 1 | 0 |
| 8 | kok | SIFAT | 2 | 1 |
| 8 | kok | SIFAT | 2 | 1 |
| 8 | kok | ISIM | 1 | 0 |
| 8 | ek | ISIM_COGUL_LER | 0 | 0 |

A aç yapısının olu turulmasından sonra bu yapı üzerinde kurallara uygun kalıpların belirlenebilmesi için bir yöntem geli tirilmelidir. Bu çalı mada arama algoritmalarından Derinlik Öncelikli Arama Algoritması kullanılmı tır.

Derinlik öncelikli arama (Depth-first search) algoritmasında en son açılan düüm takip edilerek yeni durumlar oluşturulur. Derinlik öncelikli aramada daima açılan en derin düümlerinden biri açılır. Potansiyel çözümünün çok derinlerde olmadığı durumlarda yaygın olarak kullanılmaktadır [66]. Bu çalışmada nitelenen kelime ile niteleyen kelime arasındaki kelime sayısının ortalama 5 kelime olduğu varsayılmıştır. Bu sınırlandırma olmadığı takdirde bir cümlede geçen niteleyen başka bir cümledeki kelimeyi nitelemektedir. Derinlik öncelikli aramada eğer amaçlanmayan düüme erişilmi ise veya açılacak düüm kalmamış ise açma işlemine daha üst seviyelerden devam edilir. Bu yapıya göre elde edilen örnek çıktısı Tablo 3.13’te verilmiştir.

Tablo 3.13. Nitelenen-niteleyen ilişkisi puanları

| | | | | |
|-----|-----|-------|----------|---|
| 164 | 163 | kötü | görüntü | 3 |
| 164 | 163 | kötü | görüntü | 0 |
| 164 | 163 | kötü | görüntü | 0 |
| 164 | 163 | kötü | görüntü | 0 |
| 164 | 163 | kötü | görüntü | 0 |
| 165 | 166 | iyi | de il | 2 |
| 165 | 166 | iyi | de il | 1 |
| 171 | 170 | güzel | telefon | 0 |
| 171 | 174 | güzel | mesaj | 0 |
| 177 | 175 | kötü | sesi | 3 |
| 177 | 175 | kötü | sesi | 1 |
| 177 | 178 | kötü | Telefonu | 1 |
| 177 | 178 | kötü | Telefonu | 1 |
| 185 | 180 | güzel | hafta | 1 |
| 185 | 186 | güzel | telefon | 1 |
| 185 | 186 | güzel | telefon | 1 |

Tablo 3.13’te son sütunda ilişki durumunun puanı yer almaktadır. En yüksek puanlı ilişki durumu tespit edildiğinde elde edilen çıktı Tablo 3.14’deki olacaktır. Burada dikkat edilen diğer husus ise sıfatın devamında “de il” ifadesinin yer alıp almadığıdır. Çünkü de il ifadesi belirtilen yargı tersine çevirmektedir. “iyi” ve “iyi de il” kelimeleri aynı sıfatı içermekte fakat anlam olarak farklılık göstermektedir. “iyi” kelimesi pozitif olarak değerlendirilirken “de il” ifadesi ile birlikte yer alması

anlamı negatif yapar. Bu nedenle “de il” ifadesi geli tirilen sistemde ayrıca tespit edilmi tir.

Tablo 3.14. Niteleme durumu

| id | niteleyenId | nitelenenId | niteleyen | nitelenen | puan | degil |
|-----|-------------|-------------|-----------|----------------------|------|-------|
| 185 | 164 | 163 | kötü | çekimlerinde görüntü | 3 | 0 |
| 186 | 177 | 175 | kötü | mesaj sesi | 3 | 0 |
| 189 | 891 | 889 | güzel | ve görünü ü | 3 | 0 |
| 200 | 185 | 186 | güzel | telefon | 1 | 0 |
| 201 | 203 | 204 | güzel | ve | 1 | 0 |
| 202 | 211 | 212 | yeter | arkı | 1 | 0 |
| 210 | 484 | 485 | iyi | telefon | 1 | 0 |
| 212 | 562 | 563 | çok | memnunum | 1 | 0 |

Belirlenen sıfat dereceleri geli tirilen sisteme dahil edilmesi için gerekli SQL kodları ekil 3.4’te ve memnuniyet dereceleri program çıktısı ise ekil 3.5’te yer almaktadır. gibi sonuç elde edilmektedir.

```

SELECT
    --sum(t1.[puan]) [sonuc]
    nitelenen,niteleyen
    ,t2.puan
FROM [ybs].[dbo].[niteleneSon][t1],[ybs].[dbo].[nitelene_sifatlari][t2] where t1.niteleyen like '%'+t2.sifat+'%' and
nitelenen like '%ses%' --group by t2.puan

```

| nitelenen | niteleyen | puan |
|------------|-----------|------|
| mesaj sesi | kötü | 4 |

ekil 3.4. Niteleme sonucu durumu

Örnek 1.

| ozellik | ortalama |
|----------|----------|
| batarya | 2,5 |
| kamera | 3 |
| kullanım | 0 |
| ses | 4 |
| tasarım | 0 |

| nitelenen | niteleyen | puan | ozellik |
|-----------------------|--------------|------|----------|
| BİLİN BATARYASI FAZLA | 5 | 5 | batarya |
| çekimlerinde görüntü | kötü | 4 | kamera |
| mesaj sesi | kötü | 4 | ses |
| şekilde şarjı | iyi | 0 | batarya |
| ve görünüşü | güzel | 0 | tasarım |
| telefon | güzel | 0 | kullanım |
| telefon | güzel | 0 | tasarım |
| 2)Kamerası | zorundasınız | 4 | kamera |
| telefon | iyi | 0 | kullanım |
| telefon | iyi | 0 | tasarım |
| KAMERASIYLA | TAM | 1 | kamera |
| telefon | güzel | 0 | kullanım |
| telefon | güzel | 0 | tasarım |
| ürün | güzel | 0 | tasarım |

ekil 3.5. Memnuniyet dereceleri sonucu durumu

ekil 3.5'te yukarıda belirten de erlerden "batarya" özelli i açıklanmak istenirse de erlendirilen yorumlar içinde seçilen ürünün bataryası hakkındaki memnuniyet oranının %50 oldu u anlamına gelmektedir.

Bu ürüne ait 10 adet yorumun de erlendirme sonucu Tablo 3.15'te, 100 adet yorumun de erlendirmesi ise Tablo 3.16'da görülmektedir.

Tablo 3.15. Örnek 1'e ait 10 adet yorumun değerlendirilmesi

| Ürün özellik ismi | Ortalama memnuniyet derecesi(%) |
|-------------------|---------------------------------|
| batarya | 50 |
| kamera | 60 |
| kullanım | 0 |
| ses | 80 |
| tasarım | 0 |

Tablo 3.16. Örnek 1'e ait 100 adet yorumun değerlendirilmesi

| Ürün özellik ismi | Ortalama memnuniyet derecesi(%) |
|-------------------|---------------------------------|
| batarya | 75 |
| boyut | 100 |
| fiyat | 0 |
| kamera | 48.57 |
| kullanım | 12.86 |
| ses | 60 |
| tasarım | 27.83 |

Örnek 2.

Seçilen 2. ürün Nokia E-72'dir. Ürüne ait 100 adet yorum aynı internet sitesinden aynı yöntemle elde edilmiştir. Bu yorumlar 3 tane 10'ar, 2 tane 50'er adet olarak gruplandırılmış ve gruplar arasındaki tutarlılık değerlendirilmiştir. Nokia E-72 ürününe ait yorumlardan örnekler Tablo 3.17'deki gibidir.

Tablo 3.17. Örnek 2.'ye ait yorumlar

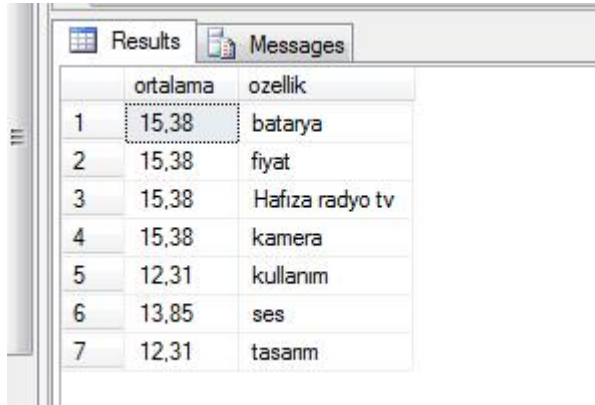
| |
|---|
| metin |
| İpari verdikten 1 gün sonra teslim aldım. TElefon her kes tarafından kullanılabilir ekilde yapılmı . ndirinde olması da güzel bir ey. Alacak olanlar acele etsin stok taki telefonlar bitmek üzere. |
| Ürünü 2 gün önce aldım.Batarya doluluk seviyesi ve çekim seviyesini gösteren bölüm baya 1 küçük,zil sesi seviyesi dü ük.(daha önce N95 kullanıyordum.)Telefona duvar ka ıdı yapmak isterseniz ekranı tam kaplamıyor(harici yüklemelerde)Telefonun kendi hafızasından duvar ka ıdı yaparsanız WWF adlı sitenin logosu ve e-posta adresi ekranda kalıyor.Yine de çok iyi bir telefon , ald ıma pi man olmadım.TAVS YE EDER M. |
| yaklaşık 6 aydır kullanıyorum. kullanımını rahat,ayrıca hafif taşıması da kolay :) çantamda cebimde kolayca ta ıyabiliyorum. video çekimi ve fotoğraf kalitesi de çok iyi. |
| Ürün bana biraz geç geldi. Hepsiburadadan birçok kez alı veri yaptım di erleri 2 gün içinde geliyordu. Onların hatırına bu kez bir ey demeyece im. Cihaz ayarlarından ba tıyan sorun ya adım. Bu cihaz için size tavsiyelerim olacak. Bir uygulama programı yükleyeceksiniz e er mutlaka dil seçene ini ingilizceye çevirin öyle proglarını kurun. Sonra tekrar dili türkçeye alırsınız.Cihaza alı mam biraz zaman aldı. Ama u anda çok memnunum. Ovi haritası mükemmel. Kaybolmak diye bir ey aklınıza gelmesin :) Anında konum payla ma uygulaması var. Resim çekip face de payla mak çok keyifli. |
| performans ve tasarım isteyenler kaçırmayın..süper bir tasarım.ve harika bir cep telefonu. indirimli fiyatı ise cabası.kaçırmayın. ben cep telefonunu çok be endim. |
| telefonu yakla ık 1 yıldır kullanıyorum.güzel telefon mesajla ma olayı süper kolay,gerçekten çok hızlı mesaj yazılabiliyor.ben telefonu elimden dü ürmem yinede bataryası 3 gün rahat gidiyor.ortalama kullanan bir insan için rahat 6 gün gider tam sarj olmu haliyle.ancak gelelim internetine ve ekranın yetersizli ine.evet e er ben internette çokça dola ırım diyosanız bu telefon size göre de il.ekran küçük oldu undan bir eyi bulmak ölüm oluyor.2. kötü yanı ise telefon birazcık büyük boyutlu videoları oynatamıyor.video oynatması o yüzden iyi de il.ayrıca içini 2-3 gb doldurduktan sonra telefon gerçekten yava lıyor. nokia artık kesmiyor arkada parayı tamamlayıp iphone 4 alma zamanı geldi |
| HEPS BURADAN S TES NS NDE ALDI M NOK A E 72 MODEL TELEFON ÇOK GÜZEL STED M K GÜN ÇER S NDE EL ME ULA TI DAHA ÖNCE NTERNET ALI VER LER NE GÜVEN M YOKTU AMA HEPS BURADANI TANIDIKTAN SONRA ÖN YARGIYI YIKTIM AMA SADECE Y NE SÖYLÜYORUM HEPS BURADAN S TES Ç N KONU UYORUM GÜVEN L R B R ALI VER S TES HERKESE TAVS YE EDER M |
| ürün nokianın tasarladı ı kusursuz cihazlardan biri i hayatında kullanacaklar için çok kullanı lı bir cihazdır. gerekli olan tüm dökümanlar içinde vardır . arj süreside oldukça iyidir 3 4 gün gider fazlakurcalamazsınız. |
| Dünki kaçırlımayacak fırsatla ne zamandır istedi im bu ürüne çok uygun fiyata sahip oldum, henüz elime ula madı ancak dün sipari vermeme birlikte kargoya verildi ini ö renmek gerçekten harika bir duygu. Göstermi oldu u özveriden ve sa ladı ı mü teri memnuniyetinden dolayı bence Hepsiburada.com en iyi alı veri sitelerinden birisi, tekrardan te ekkür ederim. Bundan sonra daha yakından takip edicem.Ürün için söylenecek bir ey yok harika bir Nokia modeli, aradı ımız bütün özelliklere sahip. |
| almayı dü ünen varsa kesinlikle kaçırmayın bu fiyata ba ka yerde bulamazsınız. ben hem kendime hem e ime aldım. arjı kullanımını ıklı ı internet hızı kamera kalitesi harika |

Aynı zamanda belirtilen internet sitesinde yer alan tüketicilerin yapılan yorumları be enme yüzdeleri Tablo 3.18'deki gibidir.

Tablo 3.18. Örnek 2.'ye ait yorumların be eni dereceleri

| Ürün özellik ismi | Be enme derecesi(%) |
|-------------------|---------------------|
| Mükemmel | 53 |
| Çok iyi | 35 |
| iyi | 8 |
| Fena De il | 2 |
| Çok kötü | 1 |

Geli tirilen yorum de erlendirme sistemiyle bu ürüne ait yorumlardan elde edilen sonuç memnuniyet durumu ekil 3.6'da gösterilmi tir.



| | ortalama | ozellik |
|---|----------|-----------------|
| 1 | 15,38 | batarya |
| 2 | 15,38 | fiyat |
| 3 | 15,38 | Hafıza radyo tv |
| 4 | 15,38 | kamera |
| 5 | 12,31 | kullanım |
| 6 | 13,85 | ses |
| 7 | 12,31 | tasarım |

ekil 3.6. Nokia E-72'nin memnuniyet derecesi

BÖLÜM 4. SONUÇ VE ÖNERİLER

İnternet kullanımının yaygınlaşması, tüketici olarak kullanıcıların satın alma davranışlarında bazı değişiklikleri beraberinde getirmektedir. İnternet ortamında yapılan alışverişler, kullanıcının daha fazla bilgiye daha kolay ulaşabilmesi, ürünler ve firmalar arasındaki kıyaslamaları daha kolay yapabilmesi ve coğrafi sınırların kalkması gibi nedenlerle fiziksel ortamdaki alışverişe göre tercih edilir hale gelmiştir.

Tüketicilerin satın alma davranışlarını etkileyen en önemli unsurlardan birinin başta tüketicilerin görüşleri olduğu bilinmektedir. Elektronik ortam sayesinde ürün ve firma bilgileri gibi tüketici görüşleri de geniş kitlelere ulaşmaktadır. Ancak bunun başta bir sonucu da ürünler hakkında yapılan yorumların sayısının artması gerçektir. Artan sayıdaki yorum içinde ürün ile ilgili memnuniyet derecesine ulaşmak oldukça zaman alıcıdır. Bu durum en kısa sürede istenilen bilgilere ulaşabilmek için otomatik olarak özet bilgi çıkarılmasını zorunlu hale getirmiştir.

Bu çalışmada internette yer alan ürün yorumlarından olumlu ve olumsuz görüşlerin sınıflandırılması sonucu kullanıcı ya da tüketiciye özet bilgi sunmayı hedefleyen bir yazılım geliştirilmiştir.

Tüketicilere zaman kazandırmak için yorumlardan özet nitelikte bilgi çıkarılması konusunun ele alındığı bu çalışmada seçilen bir ürüne ait tüketici yorumları www.hepsiburada.com adlı web sitesinden alınarak öncelikle morfolojik analizinden geçirilmiştir. Önemli kelimelere hazır hale getirilmiş kelimelerin analizi

sonucunda sözcük türleri ve aldıkları ekler belirlenmiştir. Doğal dilde yazılmış bu metinlerden negatif veya pozitif anlamlar taşıyan bilginin çıkarılması için ürün özelliğini belirten kelimeler ve bunları niteleyen sıfatlar önceden tespit edilmiştir. Sıfat anlamları 1 ile 5 puan arasında derecelendirilmiş ve “de il” kelimesinin bu dereceyi aksi yönde de ifade ettiği göz önüne alınmıştır.

Yorumlardan otomatik olarak bilgi çıkarımını sağlayabilmek amacıyla doğal dilde yazılmış yorum cümleleri için Türkçe dilbilgisi kuralları ve cümle dizilişine göre ağaç yapısı oluşturulmuştur. İstenilen özelliklerin tespit edilmesi için oluşturulan bu ağaç yapısı üzerinde Derinlik Öncelikli Arama algoritmasıyla arama yaparak sonuç ulaşılan bir yazılım geliştirilmiştir. Yazılım, Zemberek doğal dil işleme kütüphanesinin de yazıldığı, geniş dokümantasyona sahip ve tam nesneye dayalı programlama dili olan Java’da gerçekleştirilmiştir. Yazılımın Java dilinde geliştirilmesi için NetBeans tercih edilmiştir. Buradan elde edilen sonuçlar SQL veritabanında saklanmaktadır. Bu veriler önceden tespit edilmiş ürün özelliklerine göre sorgulandıktan sonra tüketicilerin yaptığı yorumlardan memnuniyet derecesini gösteren sayısal hale dönüştürülmüş oranlar elde edilmiştir. Burada sistem performansına en çok etki eden faktörün ağaç yapısındaki bulanma durumunun olduğu görülmüştür.

Geliştirilen bu sistemin Doğal Dil İşleme çalışmaları açısından kısıtlarına bakılacak olursa, Türkçe’de kelime anlamı ve türü cümlede geçtiği yere göre değiştiğinden bazı kelimelerin istenilen türünde olduğu görülmüştür. Bu sorunu gidermek için Kelime Anlamı Belirginleştirme çalışmaları yapılarak sisteme dahil edilmelidir.

İstenilen dil bilgisi kurallarının öğrenen bir yapı haline gelebilmesi için cep telefonlarında akıllı klavye olarak bilinen T9 klavye algoritmasının çalışmasından faydalanabilir.

Oluşturulan sistemdeki bir diller kısıtı, Türkçe’de sıfatların anlamca olumluluk olumsuzluk durumuna köke bakarak ulaşılamadığıdır. Bu nedenle dil bilimciler tarafından kelimelerin anlamca derecelerinin belirlenmesi ilerideki çalışmalar için de çok önemli ve faydalı olacaktır.

Yapılan çalışmada Doal Dilleme kütüphanesi olarak Zemberek programı kullanılmış fakat kelime türlerinin tespitinde yetersiz kaldığı görülmüştür. Bu nedenle kelime türlerini ve kelime anlamlarının derecelerini belirten yeni bir kütüphanenin oluşturulması çalışmadaki büyük bir kısıtı ortadan kaldıracaktır.

Geliştirilen sistem sonucunda çıkarılan bilginin faydaları iki boyutlu olacaktır. Tüketici veya internet kullanıcıları zamandan tasarruf ederken, üreticiler de müşteri önerileri ve şikayetlerine göre ürün veya hizmetlerin tasarımında değişiklik yapabilecek ya da sorunların giderilmesi için iyileştirme yollarına gideceklerdir. Ayrıca ürünün pazardaki yeri hakkında fikir sahibi olabileceklerdir. Bölüm 2.7’de Zeka’sı konusunda bahsedildiği gibi Metin Madenciliği çalışmalarından elde edilen bilgiler işletmelerin Karar Destek Sistemleri açısından büyük önem arz etmektedir.

Literatürde ve güncel hayatta uygulanan çeşitli iyileştirme teknikleri vardır. Bu tekniklerin öncelikli amacı hangi alanda iyileştirme yapılacağını tespit etmektir. Müşteri yorumlarının derecelenerek değerlendirilmesi bu soruya da cevap vermi olacaktır.

Tüketici yorumları ürün veya hizmetin tasarımı ya da sunulması hakkında da fikirler içerdiğinden buradan elde edilen bilgiler müşterilerin düşüncelerini tasarıma yansıtmak olarak nitelendirilebilir. Kalite Fonksiyon Göçerimi (Quality Function Deployment) çalışmalarına oldukça etkili veri sağlayacaktır. Çünkü bu çalışma aynı zamanda müşterinin sesine, müşteri firmaya ulaşmadan ulaşmayı da içine almaktadır.

Günümüzde işletmelerin gerçek patronları müşterilerdir değil yöneticilerden hareketle yine böyle bir sistemden elde edilecek bilgilerin Müşteri İlişkileri Yönetiminde büyük öneme sahip olacaktır.

Bu çalışmanın bir sonraki aşamasında nitelendirme kurallarını öğrenen zeki bir sistemin tasarlanması düşünülmektedir.

Bu sonuçların ürünler hakkında müşterilerin memnuniyeti ile ilgili yol gösterici bilgi vermesi görülmüştür.

KAYNAKLAR

- [1] CHOAMSKY, N., “Do a ve Dil Üzerine”, Sözcük Yayınları, 2010.
- [2] B LGEM Dergisi, Cilt:3, Sayı:5 Ocak-Nisan, 2011
- [3] TURAN, A.H., “Internet Alı veri i Tüketici Davranı ını Belirleyen Etmenler: Geli tirilmi Teknoloji Kabul Modeli (E-TAM) le Bir Model Önerisi”, Akademik Bili im, 2008.
- [4] Türkiye statistik Kurumu, 2011
- [5] KAYABA I, A., Elektronik(online) Alı veri te Lojistik Faaliyetlere Yönelik Mü teri ikayetlerinin Analizi ve Bir Alan Ara tırması, letme Ara tırmaları Dergisi, 2/2 (2010) 41-42, 2010.
- [6] AKAT, Ö., TA KIN Ç., ÖZDEM R A., “Uluslararası Alı veri Merkezi Tüketicilerinin Satın Alma Davranı ı: Bursa linde Bir Uygulama”, Uluda Üniversitesi Sosyal Bilimler Dergisi 2006/2, 2006.
- [7] DELEN, D., CROSSLAND, M.D., Seeding the survey and analysis of research literature with text mining, Expert Systems with Applications, 34 (2008) 1707–1720, 2008.
- [8] ZOHAR, E.Y., Introduction to Text Mining, Supercomputing, Automated Learning Group National Center for Supercomputing Applications, University of Illinois, 2002.
- [9] ERDEM R, Y.N., Kurumsal Zekası, Beykent Üniversitesi FBE, Yüksek Lisans Tezi, 2009.
- [10] ÖZKAN, Y., Veri Madencili i Yöntemleri, Papatya Yayıncılık, stanbul, 2008.
- [11] ALPAYDIN, E., Zeki Veri Madencili i: Ham Veriden Altın Bilgiye Ula ma Yöntemleri, Bili im 2000 E itim Semineri, 2000.
- [12] DEM R Z, A., Veri Madencili ine Giri , Veri Madencili i Ders Notları, <http://web.sakarya.edu.tr/~ademiriz/dms2006.htm>, 2006.
- [13] HAN, J., KAMBER M., Data Mining: Concept and Techniques, Morgan Kauffman Publishers, 2006.

- [14] S LAHTARO LU, G., Kavram ve Algoritmalarıyla Temel Veri Madencili i, Papatya Yayıncılık, stanbul, 2008.
- [15] BAYKASO LU, A., Veri Madencili i Ve Çimento Sektöründe Bir Uygulama , Akademik Bili im, Gaziantep, 2005
- [16] B LG N, T.T., Data Mining: Concept and Techniques, akademik.maltepe.edu.tr/~ttbilgin/DM/Week10.ppt, 2007.
- [17] ÇEL KYAY, E.K., Metin Madencili i Yöntemiyle Türkçede En sık Kullanılan ve Birbirini Takip Eden Harflerin Analizi ve Birliktelik Kuralları, Beykent Üniversitesi FBE, Yüksek Lisans Tezi, 2010.
- [18] ÖZYURT, Ö., “Türkçe Tabanlı Diyalog Sistemi ve nternet (Chat) Ortamlarından Bilgi Çıkarımı”, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2006.
- [19] P LAVCILAR, .F., Metin Madencili i le Metin Sınıflandırma, Yıldız Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2007.
- [20] SARAÇO LU, R., TÜTÜNCÜ, K., ALLAHVERD , N., A new approach on search for similar documents with multiple categories using fuzzy clustering, Expert Systems with Applications 34 (2008) 2545–2554, 2008.
- [21] MECCA, G., RAUNICH, S., PAPPALARDO, A., A new algorithm for clustering search results search results, Data & Knowledge Engineering, 62 (2007) 504–522, 2006.
- [22] WITTEN, I.H., Text Mining, Computer Science, University of Waikato, 2003.
- [23] O UZ, B., Metin Madencili i Teknikleri Kullanılarak Kulak Burun Bo az Hasta Bilgi Formlarının Analizi, Akdeniz Üniversitesi Sa lık Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2009.
- [24] MOHAMMAD, M.A., Text Mining: A Burgeoning Quality Improvement Tool, Msc. Thesis, METU, 2007.
- [25] TÜRKEE , M.K., Bilgi Eri iminde Tamlama Temelli Dizinleme, stanbul Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2007.
- [26] SEZER, E., Web Sayfaları için Anlamsal Eri im Sistemi, Hacettepe Üniversitesi FBE, Doktora Tezi, 2006.
- [27] CAN, F., KOÇBERBER, S., BALÇIK, E., KAYNAK, C., ÖÇALAN, H.C., VURSAVA , O.M., Information Retrieval on Turkish Texts, Journal of the American Society for Information Science and Technology, Vol.59, No.3, pp. 407-421, 2008.
- [28] DA , R., Web Kullanıcı Eri im Kütüklerinden Bilgi Çıkarımı, Fırat Üniversitesi FBE, Doktora Tezi, 2008.

- [29] KAISER, K., MIKSCH, S., Information Extraction A Survey, Vienna University of Technology Institute of Software Technology & Interactive Systems, 2005.
- [30] SODERLAND, S., Learning Information Extraction Rules for Semi-structured and Free Text, Machine Learning, Vol. 34, Nr. 1-3, p. 233-272, 1999.
- [31] KUSHMERICK, N., Wrapper Introduction for Information Extraction, University of Washington, Ph.D., 1997.
- [32] CUNNINGHAM, H., Information Extraction, Automatic, Encyclopedia of Language & Linguistic, Second Edition, V.5, pp. 665-677, 2006.
- [33] COWIE, J., WILKS, Y., Information Extraction, 1996
- [34] OFLAZER, K., Türkçe için Bir Sonlu Durumlu “Hafif” Doğal Dil Çözümleyicisi ve Bilgi Çıkarımı Uygulamasının Gerçekleştirilmesi, TÜB TAK PROJESİ, PROJE NO:199E027, 2002.
- [35] GÜVEN, A., Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciliği, Yıldız Teknik Üniversitesi FBE, Doktora Tezi, 2007.
- [36] ULUSOY, Ö., ALTINGÖVDE, S., ÖZCAN, R., Özelleştirilmiş Web Arama Motorlarının Otomatik Olarak Oluşturulması, TÜB TAK Projesi, Proje No:105E024, 2007.
- [37] TURMO, J., AGENO, A., CATALA, N., Adaptive Information Extraction, ACM Computing Surveys, Vol.38, No.2, Article 4, 2006.
- [38] TÜLEK, M., Türkçe için Metin Özetleme, İstanbul Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2007.
- [39] AMASYALI, F., Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulamaları. Yıldız Teknik Üniversitesi FBE, Doktora Tezi, 2008.
- [40] ORHAN, Z., ALTAN, Z., Makine Öğrenme Algoritmalarıyla Türkçe Sözlük Açıklama,
- [41] ÇİÇEK, S., Makine Öğrenmesi Teknikleri ile Tercüme Örneklerinden Öğrenilmesi, TÜB TAK Projesi, Proje No: EEEAG-244(197E011), 1999.
- [42] ÖZTEMEL, E., Yapay Sinir Ağları, Papatya Yayıncılık, İstanbul, 2006.
- [43] CHAU, M., CHEN H., A machine learning approach to web page filtering using content and structure analysis, Decision Support Systems, 44 (2008) 482–494, 2007.

- [44] DELBAĞ, A., Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi, İstanbul Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2008.
- [45] ÖZBİLGE, A., Türkçe Doğal Dili Anlamada İki-köşeli Ayrık Bilgiler Modeli ve Uygulaması, Sakarya Üniversitesi FBE, Yüksek Lisans Tezi, 2006.
- [46] NABİYEV, V.V., Yapay Zeka: İnsan-Bilgisayar Etkileşimi, Seçkin Yayıncılık, 3. Baskı, Ankara, 2010.
- [47] KESGİN, F., “Türkçe Metinler için Konu Belirleme Sistemi”, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, 2007.
- [48] SAY, B., Türkçe için Biçimbirimsel ve Sözdizimsel Olarak Üretilmiş Açık Yapılı Bir Derlem Oluşturma, TÜBİTAK EEEAG Projesi, 2003.
- [49] AYDIN, Ö., KILIÇASLAN, Y., Tümevarımlı Mantık Programlama ile Türkçe için Kelime Anlamı Belirginleştirme Uygulaması, Akademik Bilişim, Muğla, 2010.
- [50] ARIKAN, A., Bilgi Erişimde Dil Sorunları, İstanbul Üniversitesi SBE, Doktora Tezi, 2005.
- [51] ORHAN, Z., Türkçe Metinlerdeki anlam Belirsizliği Olan Sözcüklerin Bilgisayar Algoritmaları ile Anlam Belirginleştirilmesi, İstanbul Üniversitesi FBE, Doktora Tezi, 2006.
- [52] KARADENİZ, Z., Türkçe için Biçimbirimsel Belirsizlik Giderici, İstanbul Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2007.
- [53] ORHAN, Z., Türkçe Metinlerdeki Anlam Belirsizliği Olan Sözcüklerin Bilgisayar Algoritmaları ile Anlam Belirginleştirilmesi, İstanbul Üniversitesi FBE Doktora Tezi, 2006.
- [54] CEBEROĞLU, G., TANTU, A.C., ADALI E., ERENLER, Y., Sentetik Türkçe Sözcük Kökleri Üretimi, International XII. Turkish Symposium on Artificial Intelligence and Neural Networks - TAINN 2003, 2003.
- [55] ERYİĞİT, G., Türkçe'nin Bölümlük Ayrıştırılması, İstanbul Teknik Üniversitesi FBE, Doktora Tezi, 2006.
- [56] ATEŞ, H., Karar Vermede Zekasının Önemi: Tekstil Sektöründe Bir Araştırma, Dokuz Eylül Üniversitesi SBE, Yüksek Lisans Tezi, 2008.
- [57] ALADAĞ, M., Microsoft Visual Studio.Net & SQL Server 2008 ile Kurumsal Zekası Uygulamaları, Beykent Üniversitesi FBE, Yüksek Lisans Tezi, 2009.
- [58] GÜLEÇ, F.M., “Kurumsal Verilerin Yapay Zekâ Modelleri ile İşlenmesi için Modelleme Aracı Alt Yapı Tasarım Ve Gerçekleştirimi”, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2007.

- [59] ARSLAN, V., YILMAZ, G., “Karar Destek Sisteminin Kullanımı için Uygun Bir Model Geliştirilmesi”, HAVACILIK VE UZAY TEKNOLOJİLERİ DERGİSİ TEMMUZ 2010 CİLT 4 SAYI 4 (75-82), 2010.
- [60] GAO, L., CHANG, E., HAN, S., Powerful Tool to Expand Business Intelligence: Text Mining, World Academy of Science, Engineering and Technology 8,2005.
- [61] ÇEBROLU, G., Sözlüksüz Köke Ulaşma Yöntemi, Yüksek Lisans Tezi, .T.Ü. Fen Bilimleri Enstitüsü, 2002.
- [62] GÜL, M., Türkiye Türkçesinde Sözcük Türlerinin Sınıflandırılması, Mersin Üniversitesi SBE, 2008.
- [63] http://netbeans.org/index_tr.html, erişim tarihi: 20.12.2011.
- [64] YILMAZ, M.R., Türkiye Türkçesinde Sıfatlar (Tanım, Tasnif, Mukayese, Teklifler), Sakarya Üniversitesi SBE, Yüksek Lisans Tezi, 2010.
- [65] www.hepsiburada.com, erişim tarihi: 10.02.2012
- [66] BENZER, A. ., Yapay Zeka Uygulamalarında Kullanılan Arama Algoritmalarının Kıyaslanması, Gazi Üniversitesi Bilim Enstitüsü, Yüksek Lisans Tezi, 2007.

ÖZGEÇM

Kadriye Ergün, Çanakkale, Çan'da doğdu. İlk, orta ve lise eğitimini Çan'da tamamladı. 1993 yılında Çan Lisesi'nden mezun oldu. 1995 yılında başladığı Balıkesir Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği bölümünü 1999 yılında bitirdi. 2000 yılında Balıkesir Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği bölümünde ara tırma görevlisi olarak göreve başladı 2001 yılında başladığı Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Ana Bilim Dalı yüksek lisansını 2004 yılında tamamladı. 2006 yılından beri doktora çalışmasını tamamlamak üzere Sakarya Üniversitesi Endüstri Mühendisliği bölümünde ara tırma görevlisi olarak çalışmaktadır.