



İnternet ortamındaki tüketici yorumlarından özet bilgi çıkarımı

Kadriye Ergün^{1*}, Cemalettin Kubat², Gültekin Çağlı², Raşit Cesur²

^{1*}Balıkesir Üniversitesi, Mühendislik Mimarlık Fakültesi, Endüstri Mühendisliği, Balıkesir

²Sakarya Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, Sakarya

13.09.2012 Geliş/Received, 27.11.2012 Kabul/Accepted

ÖZET

Bu çalışmada, metin madenciliği tekniklerinden yararlanılarak bir ürün hakkındaki yorumları otomatik olarak değerlendirip özetlemek için geliştirilmiş bir sistem tanıtılmıştır. İncelenen veriler doğal dilde yazılmış metinler olduğundan öncelikle morfolojik analiz işleminden geçirilmiştir. Metinler içerisinde ürün özelliğini gösteren kelimeler ve pozitif-negatif anlamlı sıfatlar belirlenmiştir. Niteleyen ve nitelenen kelimelerin tespiti için ise Türkçe dilbilgisi kurallarına göre ağaç yapısı oluşturulmuştur. Ağaç yapısı üzerinde Derinlik Öncelikli Arama algoritması kullanan bir yazılım geliştirilmiştir. Yazılımın çalıştırılması sonucunda elde edilen veriler SQL veritabanında depolanmaktadır. Bu verilerden ürünün herhangi bir özelliğine göre sorgulama yapıldığında, o özellik hakkında memnuniyet derecesini ifade eden sayısal bilgiler elde edilerek yorumlanmıştır.

Anahtar Kelimeler: Metin madenciliği, Bilgi çıkarımı, Doğal dil işleme

Abstract information extraction from consumer's comments on internet media

ABSTRACT

In this study, a system developed to summarize by automatically evaluating comments about product with using text mining techniques will be described. The data has been primarily went through morphological analysis process, because they are texts written in natural language. Words and adjectives meaning positive or negative are determined. They show product features in texts. The tree structure is established according to Turkish grammar rules as subordinate and modified words are designated. The software which uses the depth-first search algorithm on the tree structure is developed. Data from result of software is stored in the SQL database. When any inquiry is made from these data depending on any property of product, numerical information which indicates the degree of satisfaction about this property is obtained. .

Keywords: Text Mining, Information Extraction, Natural Language Processing

* Sorumlu Yazar / Corresponding Author

1. GİRİŞ

Teknolojinin gelişmesi tüketicilerin davranışlarında bazı değişiklikler meydana getirmiştir. Bunlardan birisi de alışveriş alışkanlıklarının kazandığı yeni boyuttur. Gelişen teknoloji, tüketicilere internet ortamında da alışveriş imkanı sağladığından bazı ürün veya hizmetlerin satın alma işlemleri sanal ortamda daha çok yapılar hale gelmiştir. Çünkü bu tür bir alışverişlerde tüketicinin daha fazla bilgiye, daha kolay ulaşabilmesi, ürün ve firmalar arasındaki karşılaştırmaları daha kolay yapabilmesi ve coğrafi sınırların kalkması gibi avantajlar bulunmaktadır [1]. İnternet kullanıcılarının aynı zamanda potansiyel bir müşteri ya da tüketici olduğu varsayıldığında satın alma davranışlarını etkileyen faktörler arasında fiziksel alışverişlerde olduğu gibi başkalarının görüşleri de yer almaktadır.

Tüketicilerin satın alma davranışlarını etkileyen en önemli unsurlardan birinin başka tüketicilerin görüşleri olduğu bilinmektedir. Ürün ve firma bilgileri gibi tüketici görüşleri de elektronik ortam sayesinde, satın alma işlemi gerçekleşsin ya da gerçekleşmesin geniş kitlelere ulaşmaktadır. Ancak bunun başka bir sonucu da ürün veya hizmetler hakkında yapılan yorumların sayısının arttığı gerçeğidir. Artan sayıdaki yorumlar içinde ürün ile ilgili memnuniyet derecesine ulaşmak oldukça zaman alıcı ve zordur. Bu durum, en kısa sürede istenilen bilgilere ulaşabilmek için otomatik olarak özet bilgi çıkarılmasını zorunlu hale getirmiştir.

Bu çalışmada, internet ortamında yer alan ürün yorumlarından olumlu ve olumsuz görüşlerin metin madenciliği yöntemleri ile analiz edilmesi sonucu tüketiciye özet bilgi sunulması amacıyla geliştirilmiş bir sistem anlatılmıştır.

Çalışmanın devamında metin madenciliği hakkında bilgi verilmiş olup geliştirilen sistem açıklanmaktadır. Geliştirilen sistemin beklenen faydaları ve gelecek çalışmalar sonuç bölümünde yer almaktadır.

2. METİN MADENCİLİĞİ

Veriler nitelikleri bakımından yapısal, yarı yapısal ve yapısal olmayan olmak üzere üç grupta toplanmaktadır. Yarı yapısal veriler, metin, resim, grafik vs. olan belgelerden oluşmaktadır. Günlük hayata bakıldığında gazeteler, makaleler, raporlar birer belgedirler. Veri niteliği taşıyan bu belgeler çok önemli bilgiler içermelerine rağmen çoğu zaman bu bilgilere ulaşamaz. Bu tür verilerin incelenmesinde metin madenciliği teknikleri öne çıkmaktadır [2].

Metin madenciliği, belirli bir formatta olmayan yazı tipindeki veriler içerisinde gizli olan nitelikli bilginin çıkarılması, düzensiz haldeki verinin formatlanması sürecini içermektedir [3].

Metin madenciliğinde veri kaynakları, işletme dokümanları, müşteri yorumları, web sayfaları ve xml dosyaları gibi yarı yapısal veya yapısal olmayan verilerden oluşmaktadır. Buradan elde edilen bilgilerden, analizi yapılmış metin kaynaklarında açık olarak görülmeyen ilişkiler, hipotezler veya eğilimler olduğu anlaşılır [4,5].

Metin madenciliği metotları dört grupta toplanabilir [6].

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

Bu yöntemler kısaca açıklanmak istenirse, Bilgiye Erişim, metin madenciliğinde ilk adım olarak nitelendirilmektedir. Bilgiye erişimin amacı kullanıcıların bilgi ihtiyaçlarını karşılayacak olan belgeleri bulmasına yardımcı olmaktır [7]. Bilgi Çıkarımı, genellikle bir metin üzerinde doğal dil işleme kullanılarak belirli kriterdeki bilgileri elde etmeyi hedefler [8]. Web madenciliği, access log dosyaları, kullanıcı kayıt bilgileri, oturum ve hareket bilgileri, site yapısı ve içeriği gibi çeşitli yapıdaki web sayfaları dokümanlarını ve kayıt bilgilerini incelemek, bunlardaki kalıpları keşfetmek için veri madenciliği tekniklerinin kullanılması olarak tanımlanabilir [9]. Kümeleme ise önceden belirlenmiş bir kategoriler kümesine sahip olmaksızın birbirine benzer belgelerin gruplandırılması işlemi olarak tanımlanabilir [10].

Bu metotlara ait girdi kaynakları ve elde edilen çıktılar Tablo 1.'de gösterilmiştir.

Tablo 1. Metin madenciliği metotlarının girdi kaynakları ve çıktıları [6]

Bilgiye Erişim	Bilgi Çıkarımı	Web Madenciliği	Kümeleme
<i>Girdi:</i> Metin Belgesi Kaynağı, Kullanıcı sorgusu (metin tabanlı)	<i>Girdi:</i> Metinsel belgeler kaynağı İyi tanımlanmış sınırlandırılmış sorgu <i>Çıktı:</i> İlişkili bilgi cümleleri İlişkili bilginin çıkarımı ve ilişkili olmayan bilginin yok sayılması Önceden belirlenmiş formatta çıktı ve ilgili bilgi linki.	Webteki özel bilginin çıkarımı metinsel belgelerin erişimi ve indekslenmesi	Benzer metin belgelerinin toplanması

Metin madenciliği işlemleri, veri madenciliğine benzer olarak,

- Metin,
- Metin Önışleme,
- Özellik Üretimi,
- Özellik Seçimi,
- Veri Madenciliği (Örüntü Keşfi),
- Yorum/Değerlendirme,

adımlarından oluşmaktadır [6]. Bu adımların içerdikleri işlemler Tablo 2.'de gösterilmiştir [6].

Tablo 2. Metin madenciliği işlemleri

Metin Önışleme	Metin Dönüşümü	Özellik Seçimi	Veri Madenciliği/ Bilgi Keşfi	Yorum/ Değerlendirme
Söz dizimsel/ Semantik analiz	Kelime torbası,	Basit hesaplam a	Sınıflandırma (Danışmanlı)	Analiz Sonuçları
Sözcük türü etiketleme	Kelimeler Kök bulma,	İstatistik (boyut azaltma, ilişkisiz özellikler)	Kümeleme (Danışmansız)	
Kelime anlamı belirginleştirme	Etkisiz kelimeler			
Ayrıştırma (parsing)				

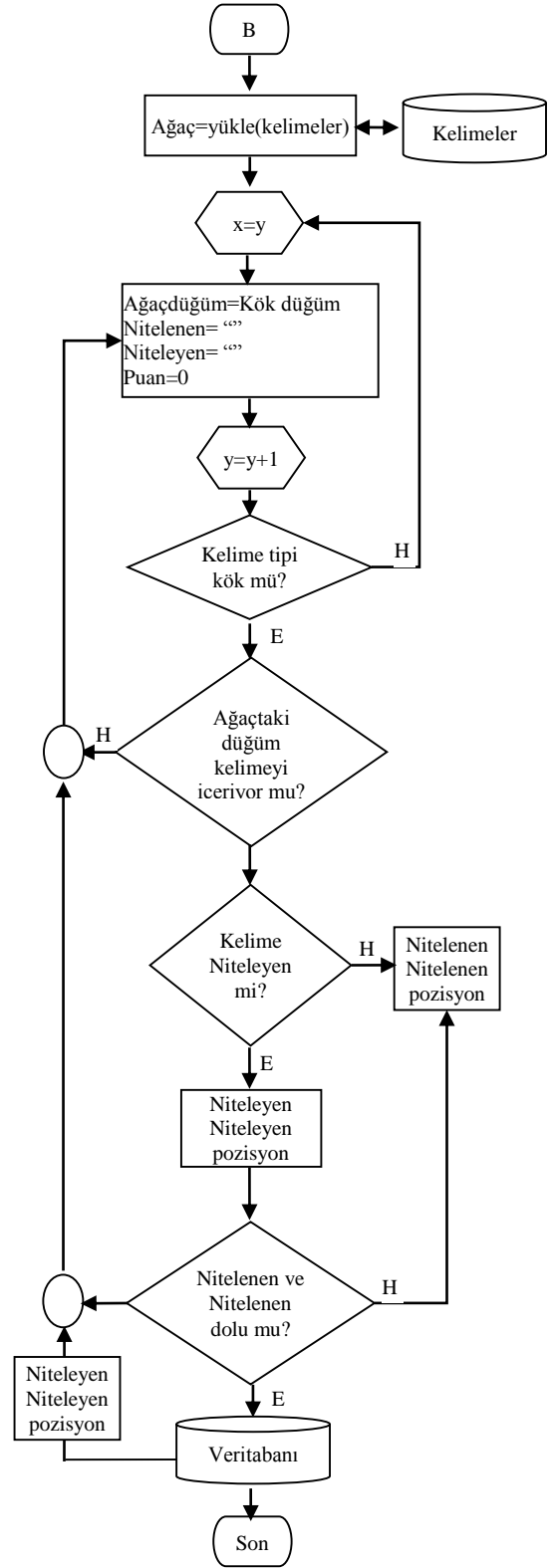
Türkçe metinler için metin madenciliğinin tüm aşamalarını tek başına uygulayabilen bir program henüz geliştirilmemiştir. Buna karşın metin madenciliği kapsamında Türkçe'de yapılan bazı projeler şunlardır. Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü Kemik Doğal Dil İşleme Grubu tarafından özet çıkarma, heceleme gibi metin madenciliği çalışmaları yapılmaktadır. Kelimelerin anlamları ve aralarındaki ilişkilere odaklanmış büyük ve uzun vadeli Wordnet projesi ise Sabancı Üniversitesi tarafından yürütülmektedir [11]. Ayrıca Türkçe'de tümcenin sözdizimsel ve biçimbirimsel olarak işaretlenmiş olduğu Türkçe Ağaç Yapılı Derlem çalışması ODTÜ-Sabancı işbirliğince yapılmıştır [12].

Bu bilgiler kapsamında geliştirilen sistem izleyen bölümde anlatılmıştır.

3. UYGULAMA

Geliştirilen yazılımda niteleme kuralları için ağaç yapısı oluşturulmuş, istenilen ürün özelliklerinin tespitinde Derinlik Öncelikli Arama algoritması uygulanarak elde edilen sonuçlar değerlendirilmiştir. Yazılım, Zemberek doğal dil işleme kütüphanesinin de yazıldığı Java'da gerçekleştirilmiştir. Yazılımın Java dilinde geliştirilmesi için NetBeans tercih edilmiştir. Buradan elde edilen veriler SQL veritabanında saklanmaktadır. Bu veriler istenilen yapıya göre sorgulandığında ürün özellikleri hakkında sayısal olarak ifade edilen memnuniyet

dereceleri elde edilmektedir. Tasarlanan sistemin akış şeması Şekil 1.'de gösterildiği gibidir.



Şekil 1. Bilgi çıkarımı sistemi akış şeması

Bu akışa göre geliştirilen sistemin aşamaları aşağıdaki gibi özetlenebilir.

Seçilen telefon ürününe ait 100 adet yorum ve ürün özellikleri hakkındaki bilgiler www.hepsiburada.com adresinden hazırlanan özel bir yazılım ile elde edilmiştir.

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.

Sıfat-isim ilişkileri, Türkçe'nin dilbilgisi kurallarının ağaç yapısına dönüştürülmesi sonucu ve derinlik öncelikli arama algoritması yardımıyla otomatik olarak belirlenir.

Niteleyen (sıfat) kelimeler anlamlarına göre derecelendirilir. Anlamca olumlu kelimeler daha yüksek

bir puanla derecelendirilirken olumsuzluk içeren kelimeler düşük puanla temsil edilir. Bu şekilde niteleyen kelimeler artık rakamsal verilere dönüştürülmüş olur.

Oluşturulan sıfat sözlüğünden niteleme dereceleri alınır. Tüm yorumlar incelendiğinde ürünün aynı özelliğine ait sıfatların olumluluk-olumsuzluk derecelerinin ortalaması alınarak sayısal bir memnuniyet derecesine ulaşılır. Bu bilgiler tablo şeklinde sunulur.

Bu çalışmada incelenen ürün telefondur. Telefona ilişkin ürün özellik isimleri ve bu özellik isimlerini içeren diğer kelimeler mevcut yorumlardan manuel olarak tespit edilmiştir. Telefona ait ürün özelliği tablosu ve bu özellikleri ifade eden eş veya benzer kelimeler Tablo 3.'de yer almaktadır.

Tablo 3. Ürün Özellik İsimleri ve Eş veya Benzer Kelimeler

özellik	eş-benzer içeren anlamlar						
<i>batarya</i>	şarj	pil					
<i>tasarım</i>	dizayn	görünüm	ekran	dokunmatik	telefon	ürün	görünüş
<i>kullanım</i>	kullanılabilirlik	telefon	tuş takımı	renk			
<i>ses</i>	mp3	müzik	kulaklık				
<i>kamera</i>	fotoğraf	resim	çekim	zoom	video	Çözünürlük	
<i>fiyat</i>	fiyat/performans	pahalı	ucuz	para	fiyat		
<i>boyut</i>	ebat	ağırlık	ince	kalın			
<i>Hafıza radyo tv</i>	video	radyo	mobil TV	hafıza kartı	GB	Bayt	
<i>bağlantı özellikleri</i>	bluetooth	GPRS	GPS	internet	wifi	Navigasyon	3G
<i>işletim sistemi</i>	symbian	V30	yazılım	sürüm			
<i>problem</i>	sorun	eksi					

Bu aşamaların açıklaması ve geliştirilen sisteme ait kısıtlar ve kabuller izleyen bölümlerde yer almaktadır.

3.1. Uygulamaya Ait Kabuller

Metin madenciliğinin ilk aşaması olan ön işleme sürecinin uygulanabilmesi için kullanılan dilin özelliklerinin bilinmesi gerekmektedir. Buna göre, 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 değiştirirler. Çekim ekleri ise kelimenin anlamını değiştirmemekle birlikte, türünü değiştirebilmektedir.

Türkçe'de kelime türleri isim, sıfat, zamir, zarf, edat, bağlaç, ünlem ve eylem olmak üzere 8 sınıfta incelenirler. Bu çalışmada değerlendirme için isim ve sıfatlar kullanıldığından sadece ürün özellik isimlerinin çekim eki alıp almadığına bakılmıştır.

Geliştirilen sistemde öncelikli sorun, kelime türlerinin tespitinden sonra hangi kelimenin ürünün hangi özelliğini nitelediğini bulmaktır. Bir ürünün bir özelliği hakkında olumlu ya da olumsuz düşüncüyü cümle içinde yer alan sıfatlar belirtmektedir. Sıfatlar, niteleme sıfatları ve belirtme sıfatları olmak üzere ikiye ayrılır. Niteleme sıfatları, varlıkların durumlarını biçimlerini renklerini kısaca nasıl olduklarını bildiren kelimelere verilen isimdir. Kısaca nasıl sorusunun karşılığı olan sıfatlardır. Bu çalışmada bir ismin niteliği incelendiği için niteleme sıfatları ile ilgilenilmiştir.

Türkçe’de sıfatların anlamca pozitif ya da negatiflikleri eşit değildir. Dilimizde 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, yazarlar tarafından belirlenmelidir. Örneğin “iyi” ve “kötü” kelimeleri sıfattır. (Cümle içinde isim de olabilirler). “İyi” anlamca olumlu iken “kötü” olumsuzdur. Bu yüzden sıfatlar derecelendirilmelidir. Bu nedenle niteleme sıfatlarının temsil ettikleri memnuniyet dereceleri 1 ve 5 arasında negatiften pozitive doğru artan bir skala ile puanlanmıştır. Türk Dil Kurumunca hazırlanan sözlükler kaynak alınıp en çok kullanıldığı düşünülerek derecelendirilen 500 adet sıfat ile ilgili örnek puanlama, Tablo 4.’te verilmiştir. Bu tabloda sadece kesin yargı belirttiği düşünülen sıfatlar derecelendirilmiştir. Bir sıfat “değil” kelimesi ile birlikte kullanıldığında anlamda değişim olduğundan değişen puanı da aynı tabloda yer alarak sisteme dahil edilmiştir.

Tablo 4. Niteleme Sıfatları Puanları

sıfat	puan	değil
acayip	2	3
acemi	2	3
acı	1	4
adil	5	0
aksak	1	4
başarılı	5	0
başarısız	1	4
basit	3	2
berbat	1	4
beter	1	4

Yazılımdaki diğer bir kabul ise niteleme ilişkilerinin belirlenmesinde görülmektedir. Buna göre 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.

3.2. Ürün Özellik İsimleri ve Niteleme İlişkilerinin Belirlenmesi

Yorumlar ön işleme yöntemlerinden geçirilip işlenmeye hazır hale getirilmelidir. Metin ön işleme, Doğal Dil İşleme yöntemlerinin kullanımını gerektirir. Bu aşamada Türkçe diline ilişkin çeşitli bilgi işlem problemlerinin çözümlenmesi için oluşturulmuş açık kaynak kodlu bağımsız bir kütüphane olan **Zemberek** programından faydalanılmıştır. Bu adımdan sonra elde edilen çıktılar şu şekilde olmaktadır.

- Sözcüklerin kökleri ve türleri bulunur.
- Sözcüklerin aldıkları ekler ve türleri bulunur.

Bu çıktıya ait örnek Tablo 5.’teki gibidir.

Tablo 5. Cümlelerin morfolojik analizinin sonucu örneği

id	metin_no	kelime_no	kelime	incelenen	tip	tur
0	0	0	BEN	ben	<kok>	ZAMIR
1	0	1	BU	bu	<kok>	ZAMIR
2	0	2	ÜRÜN Ü	ürün	<kok>	ISIM
3	0	2	ÜRÜN Ü	I	<ek>	ISIM_TAM LAMA_I
4	0	3	BİR	bir	<kok>	SAYI
5	0	4	YIL	yıl	<kok>	FIIL

Bu adımdan sonra isimlerin hangi sıfatlar tarafından nitelendiğini belirlemek gerekir. Bunu belirleyebilmek için isme getirilen çekim ekleri göz önüne alınmıştır. Bu çalışmada Cebiroğlu(2002) tarafından yapılan yüksek lisans tezinde Sonlu Durum Makineleri ile ifade edilen ek türleri ve tanımlamalar temel alınmış ve bunlara Zemberek programında karşılık gelen tanımlamaları belirlenmiştir [13].

3.3. Ağaç Yapısı ve Arama Algoritması

Bilgi çıkarımı için geliştirilen yazılımda en önemli kısıt, daha önce de belirtildiği gibi yapılan yorumlarda hangi ifadenin hangi özelliği nitelediğini bulmaktır. Bu işlemi otomatik olarak gerçekleştirebilmek için bazı kurallara ihtiyacımızın olduğu açıktır. Bir dilde herhangi bir durum hakkındaki görüşü nasıl sorusuna verilen cevaplar belirlemektedir. Nasıl sorusunun cevabı *isim-sıfat* veya *zarf-fiil* ilişkilerinden oluşan bileşenlerdir. Bu kelimeler arasındaki ilişkiler, dilin yapısına göre kelime türü, kelime kökü ve kelimenin aldığı eklerden faydalanarak oluşturulan kuralların ağaç yapısı haline getirilmesi ve arama algoritması yardımıyla elde edilmiştir. Ağaç yapısındaki niteleme kuralları Tablo 6.’da görülmektedir. Bu yapıda aynı sıfatın başka bir cümledeki ismi nitelemesini engellemek için niteleme çeşitleri puanlandırılmıştır. Buna göre en yüksek puanda olan ilişkideki kelimeler niteleyen ve nitelenen olarak belirlenmiştir. *Niteleyen kelimeler 2, Nitelenen kelimeler 1* ile ifade edilmiştir.

Tablo 6. Ağaç Yapısında Niteleme Kuralları ve Puanları

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

Ağaç yapısının oluşturulmasından sonra bu yapı üzerinde kurallara uygun kalıpların belirlenebilmesi için Derinlik Öncelikli Arama Algoritması kullanılmıştır. Derinlik öncelikli arama (Depth-first search) algoritmasında en son açılan düğüm takip edilerek yeni durumlar oluşturulur. Derinlik öncelikli aramada daima ağacın en derin düğümlerinden biri açılır. Potansiyel çözümünün çok derinlerde olmadığı durumlarda yaygın olarak kullanılmaktadır. Derinlik öncelikli aramada eğer amaçlanmayan düğüme erişilmiş ise veya açılacak düğüm kalmamış ise açma işlemine daha sığ seviyelerden devam edilir [14]. Bu yapıya göre elde edilen örnek çıktı Tablo 7.'de verilmiştir.

Tablo 7. Niteleme-Niteleyen İlişki 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	değil	2
165	166	iyi	değil	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 7.'de son sütunda ilişki durumunun puanı yer almaktadır. En yüksek puanlı ilişki durumu tespit edildiğinde sonuç çıktı Tablo 8.'deki gibi olacaktır. Burada dikkate alınan diğer husus ise sıfatın devamında “değil” ifadesinin yer alıp almadığıdır. Çünkü değil ifadesi belirtilen yargı tersine çevirmektedir. “iyi” ve “iyi değil” kelimeleri aynı sıfatı içermekte fakat anlam olarak farklılık göstermektedir. “iyi” kelimesi pozitif olarak değerlendirilirken “değil” ifadesi ile birlikte yer alması anlamı negatif yapar. Bu nedenle yukarıda da belirtildiği gibi “değil” ifadesi geliştirilen sistemde ayrıca tespit edilmiştir.

Tablo 8. Niteleme Durumu

Niteleyen Id	Nitelenen Id	niteleyen	nitelenen	puan	degil
164	163	kötü	çekimlerinde görüntü	3	0
177	175	kötü	mesaj sesi	3	0
891	889	güzel	ve görünüşü	3	0
185	186	güzel	telefon	1	0
203	204	güzel	ve	1	0
211	212	yeter	şarkı	1	0
484	485	iyi	telefon	1	0
562	563	çok	memnunum	1	0

Bu adımların sonucunda Şekil 2.'deki gibi bir özet bilgi tablosu elde edilmektedir. Burada belirtilen değerlerden “batarya” özelliği açıklanmak istenirse değerlendirilen

yorumlar içinde ürünün bataryası hakkındaki memnuniyet oranının %50 olduğu anlaşılmaktadır.

N79		Ara	
ozellik ortalama			
batarya	2,5		
kamera	3		
kullanım	0		
ses	4		
tasarım	0		
nitelenen	niteleyen	puan	ozellik
BİLİN BATARYASI	FAZLA	5	batarya
çekimlerinde görüntü	kötü	4	kamera
mesaj sesi	kötü	4	ses
şekilde şarjı	iyi	0	batarya
ve görüntüsü	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

Şekil 2. Memnuniyet derecesi sonucu durumu

Aynı ürüne ait 10 adet yorumun değerlendirme sonucu Tablo 9.'da, 100 adet yorumun değerlendirmesi ise Tablo 10.'da görülmektedir.

Tablo 9. 10 adet yorumun değerlendirilmesi

Ürün özellik ismi	Ort. Memnuniyet derecesi(%)
batarya	50
kamera	60
kullanım	0
ses	80
tasarım	0

Tablo 10. 100 adet yorumun değerlendirilmesi

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

4. SONUÇ VE GELECEK ÇALIŞMALAR

Günümüzde internet işlenmeyi bekleyen verileri içeren en büyük kaynak haline gelmiştir. Verilerin bilgiye dönüştürülmesi aşamasında yapılarındaki farklılıklar nedeniyle geleneksel veri madenciliği yöntemleri yetersiz kaldığından metin madenciliği tekniklerine

ihtiyaç duyulmuştur. Metin madenciliği teknikleri kullanılarak yapılan bu çalışmada, internet ortamında yer alan doğal dilde yazılmış metinsel verilerden otomatik olarak bilgi çıkarılmasını sağlayan bir sistem geliştirilmiştir. Söz konusu metinsel veriler, www.hepsiburada.com adresli web sitesinden alınmıştır. Seçilen ürüne ait 100 adet tüketici yorumu geliştirilen sistem yardımıyla değerlendirilerek ürün özellikleri hakkında olumlu ve olumsuz görüşlerin oranı belirlenmiş, sonuçlar tablo halinde sunulmuştur. Bu yapıyı gerçekleştirebilmek için anlamca olumlu yada olumsuz kelimeler yazarlar tarafından derecelendirilmiştir. Yorumların sayısı arttıkça ürünün farklı özelliklerinden de bahsedildiği görülmüş ve değerlendirme sonucunu değiştirdiği tespit edilmiştir. Elde edilen sonuçların ürünler hakkında müşterilerin memnuniyeti ile ilgili yol gösterici bilgiler verdiği görülmüştür.

Bu değerlendirmeler sonucunda çıkarılan bilginin faydaları iki boyutlu olacaktır. Tüketici veya internet kullanıcıları karar verebilmek için zamandan tasarruf ederken, üreticiler de müşteri öneri ve şikayetlerine göre ürün veya hizmetlerin tasarımında değişiklik yapabilecek ya da sorunların giderilmesi için iyileştirme yollarına gidebileceklerdir. Ayrıca ürünün pazardaki yeri hakkında fikir sahibi olabileceklerdir. Bu nedenle gelecek çalışmalarda bu sistemin ürün geliştirme veya ürün iyileştirme süreci sistemlerine dahil edilmesi planlanmaktadır.

KAYNAKLAR

- [1] TURAN A.H., "İnternet Alışverişi Tüketici Davranışını Belirleyen Etmenler: Geliştirilmiş Teknoloji Kabul Modeli (E-TAM) İle Bir Model Önerisi", Akademik Bilişim, 2008.
- [2] GÜVEN, A., Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciliği, Yıldız Teknik Üniversitesi FBE, Doktora Tezi, 2007.
- [3] ÖZYURT Ö., "Türkçe Tabanlı Diyalog Sistemi ve İnternet (Chat) Ortamlarından Bilgi Çıkarımı", Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2006.
- [4] SARAÇOĞLU R., TÜTÜNCÜ K., ALLAHVERDİ N., A new approach on search for similar documents with multiple categories using fuzzy clustering, Expert Systems with Applications 34 (2008) 2545–2554, 2008.
- [5] 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.
- [6] ZOHAR, E.Y., Introduction to Text Mining, Supercomputing 2002, Automated Learning Group National Center for Supercomputing Applications, University of Illinois, 2002.

- [7] TÜRKEEŞ, M.K., Bilgi Erişiminde Tamlama Temelli Dizinleme, İstanbul Teknik Üniversitesi FBE, Yüksek Lisans Tezi, 2007.
- [8] DAŞ, R., Web Kullanıcı Erişim Kütüklerinden Bilgi Çıkarımı, Fırat Üniversitesi FBE, Doktora Tezi, 2008.
- [9] BAYKAL A., ÇOŞKUN C., Web Madenciliği Teknikleri, Akademik Bilişim, 2009.
- [10] ÖZKAN, Y., Veri Madenciliği Yöntemleri, Papatya Yayıncılık, İstanbul, 2008.
- [11] OĞUZLAR, A., Temel Metin Madenciliği, Dora Yayınları, Bursa, 2011.
- [12] <http://www.ii.metu.edu.tr/tr/category/tags/turkce-derlem>
- [13] CEBİROĞLU, G., Sözlüksüz Köke Ulaşma Yöntemi, Yüksek Lisans Tezi, İ.T.Ü. Fen Bilimleri Enstitüsü, 2002.
- [14] BENZER, A.İ., Yapay Zeka Uygulamalarında Kullanılan Arama Algoritmalarının Kıyaslanması, Gazi Üniversitesi Bilişim Enstitüsü, Yüksek Lisans Tezi, 2007.