

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

SOSYAL MEDYADA MARKALARIN İTİBARLAR
ANALİZİ

Bahar ŞARDAĞI

Danışman
Doç. Dr. Çiğdem TARHAN

İZMİR – 2020

YÜKSEK LİSANS
TEZ ONAY SAYFASI



YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Sosyal Medyada Markaların İtibarlar Analizi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

22.06.2020

Bahar ŞARDAĞI



ÖZET

Yüksek Lisans Tezi

Sosyal Medyada Markaların İtibarlar Analizi

Bahar ŞARDAĞI

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Yönetim Bilişim Sistemleri Anabilim Dalı

Yönetim Bilişim Sistemleri Programı

Günümüz dünyasında sosyal ağlar aktif bir şekilde kullanılmakta ve kullanıcılar bu platformlarda paylaşımlar yaparak görüşlerini ifade edebilmektedir. Bir marka hakkında yapılan paylaşımlar diğer kullanıcıların görüşlerini ve dolayısıyla ilgili markanın itibarını etkilemektedir. Belirtilen nedenlerden dolayı markalar sosyal medya platformlarında kendileri hakkında yapılan paylaşımları önemsemekte ve bu doğrultuda aksiyonlar almaktadır. Bu tez kapsamında sosyal medya paylaşımlarını kullanarak markaların itibarlarını analiz etmesini ve tüketiciler ile aralarındaki ilişkiyi yönetmesini sağlayan bir bilgi sistemi geliştirilmesi hedeflenmiştir. Tasarlanan sistem için markalar hakkında yapılan paylaşımları toplamak amacıyla çok katmanlı mimaride bir yazılım geliştirilmiştir. Toplanan içeriklerin olumlu veya olumsuz olarak sınıflandırılması sürecini insan gücüne kıyasla daha hızlı, masrafsız yapabilmek için makine öğrenmesi yöntemleri kullanılmıştır. Tasarlanan makine öğrenmesi modelleri arasından %90 F-ölçütü başarı skoruna sahip Naive Bayes modeli seçilmiştir. Son olarak seçilen makine öğrenmesi modeliyle otomatik olarak sınıflandırılan paylaşımları ve bilgilerini sunmak için bir web uygulaması olan Marka İtibar Yönetim Sistemi (MİYS) geliştirilmiştir. Bir markaya tanımlanan kullanıcı MİYS kullanarak ilgili marka hakkında yapılan paylaşımları, analizleri görüntüleyebilmektedir. Ayrıca tanımlanan kullanıcı MİYS üzerinden paylaşıma cevap verme, tekrar paylaşma, beğeni gibi bazı aksiyonlar alabilmektedir. İleriki

alıřmalarda sistemin birok sosyal medya platformuna entegre ve daha kapsamlı olacak řekilde geliřtirilmesi amalanmaktadır.

Anahtar Kelimeler: Makine ğrenmesi, Metin Madencilięi, Marka İtibarı, Sosyal Medya, Bilgi Sistemleri



ABSTRACT

Master's Thesis

Reputation Analysis of Brands in Social Media

Bahar ŞARDAĞI

Dokuz Eylül University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

Social networks are actively used nowadays and users can express their opinions by posting contents on these platforms. Social media posts about a brand affects the opinions of other users and therefore the reputation of the related brand. Thus, most of the brands follow posts about themselves on social media platforms and accordingly take actions. Within the scope of this thesis, it is aimed to develop an information system that enables brands to analyse their reputation in social media and manage their relationship with consumers in social networks. For the aimed system, software with multi-layer architecture has been developed in order to collect the posts about the brands. In the process of classification of the collected content as positive or negative, machine learning methods were applied, since using machine learning for the classification of contents is faster and inexpensive compared to human power. The Naive Bayes classification model with 90% F-criterion score was chosen among the machine learning models designed. Lastly, Brand Reputation Management System (BRMS), web application in MVC pattern, was developed to present the automatically classified posts and their analyses. The user, registered to the system for a brand, can view the posts and analysis about the related brand by using BRMS. Additionally, the registered user can take some actions such as responding, re-sharing, and like to a post via BRMS. In future work, it is aimed to integrate the current system (BRMS) to many social media platforms and make the system more comprehensive.

Keywords: Machine Learning, Text Mining, Brand Reputation, Social Media, Information Systems



SOSYAL MEDYADA MARKALARIN İTİBARLAR ANALİZİ

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
KISALTMALAR	x
ŞEKİLLER LİSTESİ	xi
TABLolar LİSTESİ	xii
FORMÜLLER LİSTESİ	xiii
GİRİŞ	1

BİRİNCİ BÖLÜM

LİTERATÜR ARAŞTIRMASI

1.1. MARKALAR İÇİN SOSYAL MEDYANIN ÖNEMİ	4
1.2. MAKİNE ÖĞRENMESİ VE METİN MADENCİLİĞİ	7

İKİNCİ BÖLÜM

MARKA AÇISINDAN İTİBAR VE SOSYAL MEDYA KAVRAMLARI

2.1. MARKA KAVRAMI	12
2.2. MARKA İTİBARI	14
2.3. SOSYAL MEDYA	17

ÜÇÜNCÜ BÖLÜM

MAKİNE ÖĞRENMESİ SÜRECİ

3.1. VERİLERİN TOPLANMASI	21
3.2. VERİLERİN ETİKETLENMESİ	24
3.3. VERİ ÖNİŞLEME	26
3.4. ÖZELLİK SEÇİMİ	29
3.5. MAKİNE ÖĞRENMESİ	31
3.6. MAKİNE ÖĞRENMESİ ALGORİTMALARININ DEĞERLENDİRİLMESİ	40

DÖRDÜNCÜ BÖLÜM

DENEYSEL ÇALIŞMA

4.1. SİSTEM GELİŞTİRME YAŞAM DÖNGÜSÜ	43
4.2. MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN UYGULANMASI	44
4.3. MARKA İTİBAR YÖNETİM SİSTEMİ (MİYS)	49
4.3.1. Analiz Ekranı	50
4.3.2. Tweet Arama Ekranı	51
4.3.3. Tweet Detay Bilgileri Ekranı	53
SONUÇ	56
KAYNAKÇA	58

KISALTMALAR

VTU	Veri Toplama Uygulaması
VEU	Veri Etiketleme Uygulaması
MİYS	Marka İtibar Yönetim Sistemi
TF	Terim Frekansı
TDF	Ters Doküman Frekansı
KNN	K-En Yakın Komşu (K-Nearest Neighbourhood)
YSA	Yapay Sinir Ağı
SVM	Destek Vektör Makineleri (Support Vector Machines)
DP	Doğru Pozitif
DN	Doğru Negatif
YP	Yanlış Pozitif
YN	Yanlış Negatif
SGYD	Sistem Geliştirme Yaşam Döngüsü

ŞEKİLLER LİSTESİ

Şekil 1: Sosyal Geri Bildirim Döngüsü	s.5
Şekil 2: Metin Madenciliği Süreci	s.20
Şekil 3: Veri Toplama Uygulaması Mimarisi	s.21
Şekil 4: Güncel Paylaşımları Toplama Metodunun Sözde (Pseudo) Kodu	s.22
Şekil 5: ER (varlık-ilişki) Diyagramı	s.23
Şekil 6: ER (varlık-ilişki) Diyagramı	s.25
Şekil 7: Veri Etiketleme Ekranı	s.26
Şekil 8: Makine Öğrenmesi Yöntemleri	s.32
Şekil 9: Denetimli Makine Öğrenmesi Süreci	s.33
Şekil 10: Karar Ağacı Örneği	s.35
Şekil 11: Rastgele Orman Süreci	s.36
Şekil 12: Yapay Sinir Ağları Modeli	s.37
Şekil 13: Yapay Sinir Ağ Modelleri Aktivasyon Fonksiyonları	s.38
Şekil 14: Destek Vektör Makineleri	s.40
Şekil 15: Sistem Geliştirme Yaşam Döngüsü	s.43
Şekil 16: Test Sonuçları	s.46
Şekil 17: İtibar Analiz Ekranı	s.51
Şekil 18: Tweet Arama Sayfası	s.52
Şekil 19: Kelime Bazlı Arama	s.53
Şekil 20: Tweet Detay Bilgileri Ekranı / Tweet Detay Bilgileri Sekmesi	s.54
Şekil 21: Tweet Detay Bilgileri Ekranı / Gönderen Bilgileri Sekmesi	s.55

TABLÖLAR LİSTESİ

Tablo 1: Üretici ve Tüketici Açısından Marka Kavramı	s.13
Tablo 2: Yazarlar Tarafından Yapılan İtibar Tanımları	s.15
Tablo 3: Geleneksel Medya ve Sosyal Medya Karşılaştırması	s.18
Tablo 4: Doküman Terim Matrisi	s.28
Tablo 5: TF-TDF Değer Tablosu	s.28
Tablo 6: Uzaklık Fonksiyonları ve Formülleri	s.34
Tablo 7: Ölçüt-Formül Tablosu	s.42
Tablo 8: Veri Kümesi Varyantları	s.45
Tablo 9: Denetimli Makine Öğr. Algoritmalarının Değerlendirme Sonuçları	s.48

FORMÜLLER LİSTESİ

Formül 1: Terim Doküman Frekansı	s.28
Formül 2: Ki-Kare	s.29
Formül 3: Bilgi Kazanımı	s.30
Formül 4: Entropi	s.35
Formül 5: Karar Ağacı Bilgi Kazanımı	s.35
Formül 6: Koşullu Olasılık	s.38
Formül 7: Naive Bayes Yaklaşımı	s.39
Formül 8: Günlük İtibar Puanı (GİP) Hesaplama	s.50



GİRİŞ

Markalar itibarlarını medya, basın bültenleri, web sitesi, çevrimiçi kaynaklar ve ağızdan ağıza aracılığıyla topladığı bilgilerle geliştirir. Bunlardan en önemlisi, bir şirketin itibarını toplam veya kısmi bilgi ve haberler yoluyla şekillendirmeye veya çarpıtmaya yardımcı olan medya, daha özel olarak sosyal medyadır (Kaul ve diğerleri, 2015).

Günümüzde sosyal medyanın her geçen gün kullanıcı sayısını artırması marka ve kurumların yönlerini bu platformlara çevirmesini beraberinde getirmiştir (Guräu, 2018). Kullanıcı dostu olarak nitelendirilen bu ortamlar kullanıcıların ve markaların içerik paylaşmasına izin verirken markalar açısından sadece ürün ya da hizmetlerinin tanıtıldığı yerler olarak kalmamaktadır. Bu ortamlar aynı zamanda markaların tüketicisiyle etkileşime girebildikleri, sorun çözebildikleri, hızlı geribildirim aldıkları bir ortam olmakla birlikte, tüketiciler açısından hem memnuniyetin hem de şikâyetlerin açıkça dile getirilebildiği, başkalarıyla paylaşılabilirdiği, fikir alışverişinde bulunabildikleri ortamlar haline gelmiştir. Sosyal medya bir markanın itibarının sadece ürün ve hizmetlerine yönelik reklamları ya da elde ettiğı başarıları ile değil; bu platformda insanların marka hakkındaki yorum, paylaşma ve beğeni gibi etkileşimlerle nasıl görüldüğünü değerlendirebilmesinde önemli katkılar sunmaktadır. Dolayısıyla bu platformda markalar açısından ürün ya da hizmetlerden çok tüketicilerin marka ya da kurum hakkındaki konuşmaları, değerlendirmeleri önemli hale gelmiştir.

Sosyal medyanın ortaya çıkmasından önce tüketiciler bir ürün ya da hizmet satın almak istediklerinde markanın ya da kurumun web sitesi, yayınlanan reklamları ve marka tarafından paylaşılan haberler gibi daha kontrollü gerçekleşen ortamlar aracılığıyla gerekli bilgileri elde etmekte ve sadece markanın sunduğı tek taraflı bilgilerle yetinmekteydi. Ancak sosyal medyanın ortaya çıkmasıyla birlikte tüketicilerin de bir ürün ya da hizmet alma sürecindeki arayışlarında davranış değişikliği meydana gelmeye başlamıştır. Artık bireyler satın alma davranışında bulunmak istediklerinde sadece kurum ya da markaların web sitelerindeki kontrollü bilgilerle yetinmemekte, bilgi arayış süreçlerine farklı platformları da dâhil etmektedir. Sosyal medya platformları tüketicilerin bilgi kaynaklarının çeşitlenmesi, markalar tarafından oluşturulan içeriklerin dışına çıkarak karar vermesi ve tercihte

bulunabilmesine imkân sağlamıştır. Çünkü tüketici artık karar verme sürecinde sadece marka ya da kurumun paylaştığı bilgilerle yetinmemekte aynı zamanda ürün ya da markaya yönelik sosyal medyada ne konuşulduğu, nasıl yorumların yapıldığı, beğeni ve şikâyetleri de dikkate almaktadır.

Her gün markalar ve firmalar hakkında sosyal ağlarda binlerce paylaşım yapılmaktadır. Bu paylaşımların insan gücüyle değerlendirilmesi masraflı ve zaman alan bir süreç olacaktır. Bu noktada tezin hipotezi, tüketici karar verme süreçlerinde ve marka algısının oluşmasında önemli etkisi olan sosyal medya paylaşımları üzerinden marka itibar analizinin bir bilişim sistemi ile yapılmasının insan gücüne kıyasla daha performanslı olacağıdır. Bu doğrultuda sosyal ağ paylaşımlarının otomatik olarak sınıflandırılması için metin madenciliği yöntemleri araştırılmış ve uygulanmıştır.

Metin madenciliği sürecinde ilk olarak markalar veya firmalar hakkında paylaşımları toplamak için çok katmanlı mimaride bir uygulama (Veri Toplama Uygulaması- VTU) geliştirilmiştir. Bu uygulama Twitter (Twitter, 2019) ara yüzünü kullanarak belirlenen markalar hakkında yapılan paylaşımları ilişkisel veri tabanına kayıt etmektedir. İkinci adımda denetimli makine öğrenmesi algoritmalarında kullanılmak üzere toplanan paylaşımların etiketlenmesi için bir kitle kaynak web uygulaması (Veri Etiketleme Uygulaması - VEU) geliştirilmiştir. Yeterli sayıda sosyal medya paylaşımı etiketlendikten sonra oluşturulan veri kümesinde çeşitli ön işleme adımlarıyla birlikte denetimli makine öğrenmesi modelleri Spyder (Python dili için yazılmış geliştirme ortamı) (Spyder, 2020) aracılığıyla kodlanıp deneyimlenmiştir. Yapılan testlerde %90 başarı oranına sahip olan Naive Bayes (Kibriya ve diğerleri, 2004) modeli sosyal medya paylaşımlarının otomatik olarak sınıflandırması için seçilmiştir.

Metin madenciliği sürecinden sonra markalar için otomatik olarak olumlu ya da olumsuz olarak sınıflandırılan içeriklerin analiz edilmesi ve yönetilebilmesi için bir web uygulaması (Marka İtibar Yönetim Sistemi- MİYS) geliştirilmiştir. Bu uygulamada markalar günlük olarak hesaplanan itibar skorlarını gözlemleyebilmekte, kendileri hakkında yapılan olumlu ve olumsuz paylaşımlar hakkında detaylı bilgileri (paylaşım zamanı, beğenilme sayısı, gönderen kullanıcı vb.) görüntüleyebilmektedir.

Ayrıca paylaşım zamanı veya içerdiği terime göre arama yapabilmekte ve paylaşımı yapan kullanıcılarla etkileşime girebilmektedir.

Sonuç olarak geliştirilen uygulamalar (VTU ve VEU) ve Python programlama dili geliştirilmiş sınıflandırma modeli ile markaların kendileri hakkında yapılan paylaşımların otomatik olarak sınıflandırıldığı, itibar analizini yapabildiği bir bilgi sistemi sunulmaktadır.

Tezin birinci bölümünde yapılan çalışma ile ilgili literatür araştırmasına yer verilmiştir. İkinci bölümde marka, marka itibarı ve sosyal medya kavramları hakkında genel bilgi verilmektedir. Sosyal medya paylaşımlarının otomatik olarak sınıflandırılması için uygulanan metin madenciliği adımları üçüncü bölümde açıklanmıştır. Dördüncü bölümde ise uygulanan metin madenciliği uygulamasının sonuçları yer almaktadır. Marka itibar yönetimi için geliştirilen web uygulaması MİYS dördüncü bölümde açıklanmaktadır. Son bölümde ise tez hakkında tartışmalar ve gelecek çalışmalar hakkında bilgiler verilmektedir.

BİRİNCİ BÖLÜM

LİTERATÜR ARAŞTIRMASI

1.1 MARKALAR İÇİN SOSYAL MEDYANIN ÖNEMİ

Web 2.0 ve sosyal medyanın ortaya çıkmasıyla birlikte tüketiciler markaya yönelik algı oluşumunda etkin rol almaya başlamışlardır. Bilgi ve deneyimlerini paylaşarak sosyal medya içeriğinde önemli katkıda bulundukları için karar verme süreçlerinde oldukça önemli yere sahiptirler.

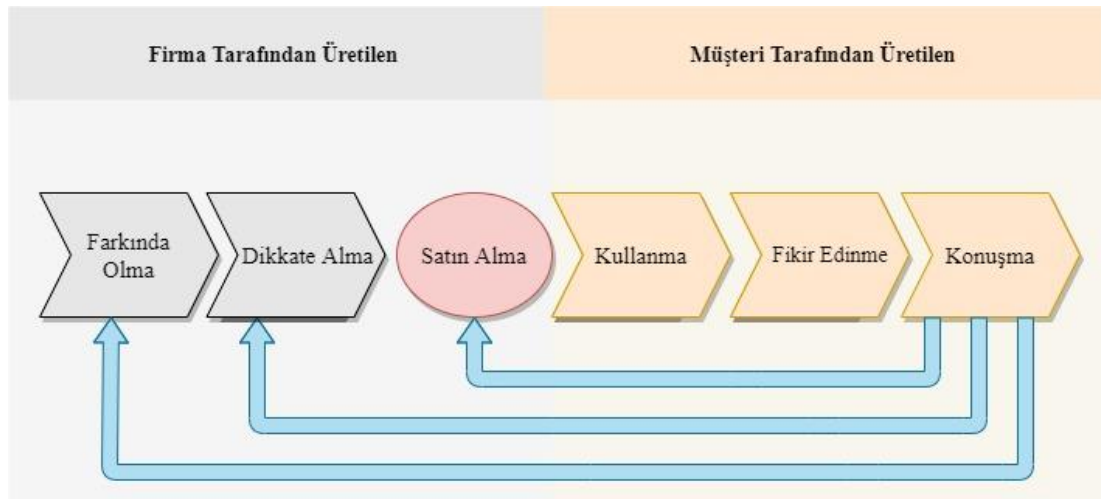
Tüketicinin markaya yönelik algısının oluşumunda sosyal medyayla birlikte bireysel müşterilerin yerini artık bir sayfa ya da ortak ağ içerisinde yer alan kolektif bir varlık olan müşteriler almaktadır. Bu nedenle bu topluluk karşısında markanın kendi itibarını, imajını oluşturması, koruması değişen tüketici yapısıyla giderek zorlaşmaktadır. Sosyal medyayı benimsemenin bir sonucu olarak, bazı yazarlar ve araştırmacılar, şirketlerin geleneksel iletişim modellerinden kaçınmaları ve tüketicilerin farkındalıklarını ve algılarını artırmak için sosyal medyadaki paydaşlarla açık, nitel ve güvenilir diyaloglar kurması gerektiği konusunda benzer görüşlere sahiptir (Sriram ve diğerleri, 2014).

Bernoff ve Schadler, Harvard Business Review 2010 yılındaki sayısında yayınladıkları yazısında sosyal medyayı bir kızgın tweetin bir markayı mahvedebileceği bir dünya olarak nitelendirerek markaların karşı karşıya kaldığı ve yönetmek zorunda oldukları durumu özetlemektedir (aktaran Kaul ve diğerleri, 2015:455). Benzer şekilde Berkshire Hathaway'in CEO'su ve en saygın işadamlarından biri olan Warren Buffett, "İtibar kazanmak 20 yıl, onu yıkmak ise beş dakika sürüyor. Bunu düşünürsen, işleri farklı yapacaksın." diyerek itibarın önemine vurgu yapmakta ve bu süreçte sosyal medyayı neden dikkate almamız gerektiğini aslında bize göstermektedir (Reputation Lessons from Warren Buffet, 2020). Bu nedenle artık kurumlar için bu ortamlarda yer alıp almamaktan ziyade bu ortamları tüketiciyle iletişim ve etkileşim kurmak, işbirliği yapmak, tüketiciyi dinlemek için kullanmak önemli hale gelmiştir. Özellikle her geçen gün artan kullanıcı sayısı kurumların yönetmek zorunda oldukları bu alanı daha da zorlaştırmaktadır.

Bunting ve Lipski 'nin (2001) sıkça atıfta bulunulan bir internet tahminine göre; radyo 50 milyonluk bir dinleyici kitlesine 38 yılda, televizyon 13 yılda ulaşırken internet hizmet sağlayıcıları sadece iki buçuk yılda ulaşmıştır. Bugün We are Social'ın (We Are Social, 2020) dünya çapında internet kullanımına yönelik olarak gerçekleştirdiği araştırma 2019 verilerine göre dünya üzerinde 4.3 milyar internet kullanıcısı varken 3.4 milyar insanın sosyal medya kullanıcısı olduğunu ortaya çıkarmıştır. Söz konusu sayılar kurumların karşı karşıya kaldığı kitleyi ve sosyal medyanın önemini ortaya koymaktadır.

Tüketici satın almak istediği hizmet veya ürün hakkındaki değerlendirmelere bu kadar kullanıcıya sahip bir platform olan sosyal medya aracılığı ile kolay bir şekilde ulaşabilmektedir. Satın alma tercih edildiği durumda ise tüketici kendi görüşlerine sahip olmaktadır. Bu görüşler sosyal ağlarda paylaşıldığı durumda gelecekteki tüketiciler tarafından satın alma kararlarında kullanılmak üzere yeni bir değerlendirmenin varlığından söz edilebilir. Bu döngü sürekli olarak Şekil 1'de görüldüğü gibi devam etmektedir. İyi bir ürün/hizmet üretmek kadar tüketici tarafından nasıl algılandığı, markaya yönelik düşünceler rekabetin yoğun olduğu günümüzde önemlidir.

Şekil 1: Sosyal Geri Bildirim Döngüsü



Kaynak: Fong ve Yazdanifard, 2014

2008 yılında yapılan bir araştırmada (Barnes ve diğerleri, 2008) sosyal medya ile müşteri memnuniyeti, marka itibarı ve sadakati arasındaki ilişki analiz edilmiştir. Sosyal medyanın yaygın şekilde kullanımının marka itibarını nasıl etkilediği incelenmiştir. Çalışma web tabanlı bir anket yoluyla gerçekleştirilmiştir. Anket kapsamında tüketicinin bir satın alma eylemi gerçekleştirmeden önce ürün veya hizmet hakkında bilgi edinmek için sosyal medyayı ne sıklıkla kullandığını araştırmaya yöneliktir. 320 aktif internet kullanıcısı ile yapılan ankette %70 den fazla kullanıcının satın alma sürecinde sosyal medyayı aktif kullandığı sonucuna varılmıştır.

Başka bir çalışmada (Öztamur ve Karakadılar, 2014) yazarlar küçük ve orta ölçekli işletmeler (Kobi) için sosyal medyanın rolünü analiz etmişlerdir. Rastgele seçilen dört firmaya ait sosyal medya hesapları belirli bir süre boyunca beğeni sayısı, güncelleme sıklığı, içeriğin zenginliği, dil kullanımı, kullanıcı ile etkileşimi gibi açılardan takip edilmiştir. Elde edilen sonuçlardan sosyal medya platformlarını bir pazarlama aracı olarak kullanmak isteyen işletmelerin zengin içerik oluşturmalarının önemli olduğu görülmüştür.

Uluslararası bir spor giyim markası için Twitter üzerinden paylaşılmış olan sosyal medya içeriklerinin analiz edildiği bir çalışmada (Krüger ve diğerleri, 2012) sosyal medyada marka iletişiminin önemi vurgulanmıştır. Çalışma kapsamında 60 gün içinde seçilen marka hakkında 289.513 twitter paylaşımı toplanmıştır. Rastgele seçilen 500 tweet ve kurumsal hesaplar tarafından paylaşılan tüm tweetler manuel olarak analiz edilmiştir. Analiz sosyal medyada iletişimin seçilen spor markası için oldukça önemli olduğunu göstermiştir. Aynı zamanda markanın ürünlerinin bilinirliğini arttırmak için Twitter’da iletişime stratejik olarak girdiği sonucuna varılmıştır.

Şirketlerle ilgili sahte hesaplar, yalan haberler ve içerikler, kişisel ve kurumsal skandallar, negatif yorumlar, rakip saldırıları, nefret siteleri vb. unsurlar şirket ya da markaya yönelik haberlerin hızla yayılması itibar konusunda sorunları beraberinde getirmektedir (Öngün,2014). Bu nedenle şirketler ya da markalar, sosyal medya konuşmalarını ele alma ve beklentileri yönetme süreçleri ve prosedürleri iyi planlanmamışsa veya daha geniş topluluklarla etkileşimleri stratejik olarak belirlenmemişse itibarları riski altındadır.

1.2 MAKİNE ÖĞRENMESİ VE METİN MADENCİLİĞİ

Bu tez kapsamında markaların sosyal medya itibarlarını değerlendirip analiz edebileceği bir bilgi sistemi sunmak amaçlanmıştır. Literatürde bu alanda yapılmış bir takım çalışmalar mevcuttur. Tespit edilen çalışmalar paylaşımların makine öğrenmesiyle otomatik olarak sınıflandırmasına yönelik olup, kullanıcılar tarafından bu sonuçların incelenebileceği bir sistem sunmamaktadır. Ayrıca incelenen çalışmalar genellikle İngilizce veri kümesi için yapılmıştır.

2015 yılında yapılan bir çalışmada (Vidya ve diğerleri,2015) mobil telefon kullanıcılarına ait Twitter paylaşımlarından duygu analizi yaparak marka itibarını bulmayı hedeflemiştir. Örnek model Haziran-Mart ayları arasında üç mobil firması hakkında toplanan 3000 Twitter paylaşımı üzerinden oluşturulmuştur. Veri temizleme ve veri dengeleme işlemlerinden sonra veri kümesi üzerinde Naïve Bayes, Destek Vektör Makineleri ve Karar Ağacı algoritmaları uygulanmış ve sırası ile %89.08, %89.33,%84.43 f-ölçütü skorları elde edilmiştir. En başarılı sonuç alınan Destek Vektör Makineleri çalışma kapsamında sınıflandırıcı model olarak seçilmiştir. Ayrıca elde edilen sonuçlardan markaya ait marka itibar puanı hesaplaması için bir ölçüm önerilmiştir. Bu ölçüm bu tez kapsamında MİYS içerisinde kullanılmıştır.

Mobil ödeme hizmetlerine yönelik kullanıcı görüşünü olumlu ve olumsuz olarak sınıflandırmak için yapılan başka bir çalışmada (Prihono ve Sari, 2019) makine öğrenmesi yöntemleri kullanılmıştır. 1 Mayıs - 31 Mayıs 2019 tarihleri arasında beş mobil ödeme firmasının her biri hakkında paylaşılan toplam 28.189 twitter paylaşımı üzerinden veri kümesi oluşturulmuştur. Birtakım veri ön işleme adımları uygulandıktan sonra veri kümesi üzerinde Destek Vektör Makineleri algoritması kullanılmıştır. Kullanılan algoritma için her bir marka özelinde doğruluk skoru ayrı ayrı hesaplanmıştır. Elde edilen doğruluk skorları %77.30 ile %80.82 arasında değişmektedir.

Birbirine rakip olan iki akıllı telefon markası için tüketici görüşlerini analiz etmek için sunulan başka bir çalışmada (Kim ve diğerleri, 2016) sözlük tabanlı duygu analizi ve makine öğrenmesi yöntemleri kullanılmıştır. İlgili markalar hakkında 24.206 Twitter paylaşımı toplanarak veri kümesi oluşturulmuştur. Veri ön işleme sürecinden sonra elde edilen veri kümesi üzerinde R programlama dilinde Destek

Vektör Makineleri algoritması kullanılmıştır. Çalışma sonucunda %74 tahminleme başarısı elde edilmiştir.

2016 yılında sosyal medya verilerini kullanarak belirli bir şirketin itibarını ölçen bir itibar yönetim sistemi önerilen başka bir çalışmada (Manaman ve diğerleri, 2016) makine öğrenmesi yöntemleri kullanılmıştır. Makine öğrenmesi yöntemlerini uygulamak üzere 4929 tane sosyal medya içeriğinden oluşan bir veri kümesi kullanılmıştır. Bu veri kümesinde her içerik pozitif, negatif, tarafsız ve alakasız olmak üzere etiketlenmiştir. Veri kümesi üzerinde Naïve Bayes, Yapay Sinir Ağları ve Destek Karar Vektörleri algoritmaları uygulanmıştır. Sırası ile %68.9, %67.4 ve %75 doğruluk skorları elde edilmiştir.

Markalara karşı tüketicilerin duygularını analiz etmeyi amaçlayan başka bir çalışmada (Mostafa,2013) yazar makine öğrenmesi yöntemlerine başvurmuştur. Tüketicilerin belirlenen markalar için Temmuz 2012 - Ağustos 2012 tarihleri arasında paylaştıkları sosyal medya içerikleri toplanarak veri kümesi oluşturulmuştur. Bu veri kümesi 3516 sosyal medya paylaşımını içermektedir. Veri kümesi üzerinde sözlük tabanlı makine öğrenmesi yöntemleri kullanılmıştır. Çalışma kapsamında değerlendirilen markalar için yapılan paylaşımların pozitif oranı sırasıyla % 54, %60, 28 bulunmuştur.

Cep telefonu, dizüstü bilgisayar vb. elektronik ürünlerle ilgili twitter paylaşımlarını makine öğrenimi yöntemlerini kullanarak analiz eden diğer bir çalışma 2013 yılında sunulmuştur (Neethu ve Rajasree, 2013). Nisan 2013 ve Mayıs 2013 tarihleri arasında 1200 tweet toplanarak veri kümesi oluşturulmuştur. Bu tweetler negatif ve pozitif içeriğe göre manuel olarak etiketlenmiştir. Veri kümesi üzerinde Naive Bayes, Destek Vektör Makineleri, Maksimum Entropi ve Topluluk algoritmaları uygulanmıştır. Naive Bayes algoritmasının diğer üç algoritmaya göre daha iyi hassasiyete sahip olduğu sonucuna varılmıştır. Ancak diğer üç algoritma kullanılarak %90 doğruluk skoru elde edilirken, Naive Bayes ile bu skor %89 olarak hesaplanmıştır.

İlaç ve kozmetik ürünlerle ilgili tüketici duygularının analiz edilmesine yönelik sunulan çalışmada (Isah ve diğerleri,2014) makine öğrenmesi ve metin madenciliği yöntemleri kullanılmıştır. Çalışma kapsamında seçilen anahtar kelimelerle ilgili (reçete, eczane, antibiyotik vb.) ilgili 11.431 twitter paylaşımı toplanmıştır.

Oluşturulan veri kümesinin %75 'i eğitim seti olmak üzere negatif ve pozitif olarak etiketlenmiştir. Veri kümesi üzerinde Naive Bayes algoritması uygulanmış ve %83 doğruluk skoru elde edilmiştir.

Makine öğrenmesi ve metin madenciliği yöntemlerinin kullanıldığı başka bir çalışmada (Mittal, ve Goel, 2012) borsa hareketleri tahmini hedeflenmiştir. Tahmin sonuçları ile mevcut durumu kıyaslayarak aralarındaki korelasyonu hesaplamayı hedeflemişlerdir. Çalışma kapsamında iki veri seti kullanılmıştır. Bunlardan biri seçilen hisse senedi endeksine ait belirli bir gün için açık, kapalı, yüksek ve düşük değerleri içermektedir. Diğerisi ise Haziran 2009 - Aralık 2009 tarihleri arasında toplanan Twitter verilerini içermektedir. Bu veriler üzerinde 4 farklı öğrenme algoritması kullanılmıştır: Doğrusal Regresyon, Lojistik Regresyon, Destek Vektör Makineleri, Bulanık Sinir Ağları. Bu algoritmalar arasından Bulanık Sinir ağları algoritmasından diğer algoritmalarından daha başarılı olarak %75.56 doğruluk skoru elde edilmiştir.

2018 yılında sunulan başka bir çalışmada (Hasan ve diğerleri, 2018) yazarlar makine öğrenme algoritmaları uygulayarak sosyal medya verileri üzerinden kullanıcıların bir konu hakkındaki görüşlerinin analizini yapmışlardır. Kullanıcılar tarafından paylaşılmış içerikleri içeren 6250 tweet seçilerek bir veri kümesi oluşturulmuştur. Bir takım veri ön işleme adımları uygulandıktan sonra veri kümesi üzerinde Naive Bayes ve Destek vektör makineleri algoritmaları uygulanmıştır. Naive Bayes algoritmasının uygulandığı veri kümesinde en yüksek skor olan %79 doğruluk skoru elde edilmiştir.

Veri madenciliği ve denetimli makine öğrenme algoritmalarının kullanıldığı bir çalışmada (Choi ve diğerleri, 2016) kötü hava koşullarının neden olduğu havayolu gecikmelerini tahminlemesi yapılmıştır. Çalışma kapsamında ABD iç uçuş verileri ve 2005 – 2015 hava durumu verileri kullanılmıştır. Oluşturulan veri kümeleri üzerinde Rastgele Orman, AdaBoost, K-En Yakın Komşu ve Karar Ağacı denetimli makine öğrenme algoritmaları uygulanmıştır. Sırası ile %80.36 , %71.43 , %35.71 ve %64.29 doğruluk skorları elde edilmiştir.

Bu alanda incelenen çalışmalar içerisinde bazı Türkçe veri kümesi üzerinde çalışmalar da mevcuttur. Gözlemlenen çalışmalar arasında en yüksek başarı oranı %85.64'tür.

Markalara ait itibar puanının tüketicilerin satın alma eylemi üzerindeki etkisinin değerlendirildiği bir çalışmada (Gok ve Fidan, 2019) veri madenciliği teknikleri kullanılmıştır. Çalışma kapsamında belirli bir marka hakkında 415 kişiye anket uygulanarak veri kümesi oluşturulmuştur. Bu veri kümesi üzerinde gerekli ön işleme adımları tamamlandıktan sonra Weka programı kullanılarak Karar Ağacı algoritması uygulanmıştır. Kullanılan algoritma ile araştırma yapılan marka üzerinde %85.64 doğruluk skoru elde edilmiştir.

2015 yılında sunulan başka bir çalışmada (Nalçakan ve diğerleri, 2015) yazarlar twitter platformunda bir marka hakkında paylaşılan içeriklerin pozitif, negatif ve nötr olduğu durumları analiz etmişlerdir. Çalışma kapsamında üç marka seçilerek her biri için Twitter üzerinden toplanan veriler ile ayrı veri kümesi oluşturulmuştur. Veri kümeleri üzerinde çeşitli makine öğrenmesi algoritmaları (Naive Bayes, Rastgele Orman, K-En Yakın Komşu vb.) uygulanmıştır. Üç marka için de en iyi sonucu Naive Bayes algoritmasından ortalama %70 doğruluk skoru ile elde edilmiştir.

Türkçe veri seti ile yapılan başka bir çalışmada (Albayrak ve diğerleri, 2017) yazarlar sosyal medyada belirli bir konu hakkında yapılan paylaşımlar üzerinde duygu analizi yapmışlardır. Seçilen konu hakkında yapılan paylaşımların pozitif, negatif ve nötr olarak sınıflandırılması amaçlanmıştır. 10-12 Ekim 2017 tarihleri arasında Twitter platformu üzerinden 12.739 sosyal medya paylaşımı toplanarak veri kümesi oluşturulmuştur. Elde edilen veri kümesi Python programlama dilinde Doğal Dil İşleme Kütüphanesi kullanılarak Türkçe duygu analizi sözlüğü olan SentiTurkNet veri kümesi ile karşılaştırılmış ve işlenmiştir. Çalışmada başarı değerlendirilmesi yapılmamış olup, paylaşımların %16 oranında pozitif, %5 oranında negatif ve %79 oranında ise nötr olduğu sonuçlarına varılmıştır.

2019 yılında sunulan bir çalışmada (Ayan ve diğerleri, 2019) firmaların Twitter paylaşımları incelenerek, firmalar ile paylaşımları arasındaki tutarlılık analiz edilmiştir. Ayrıca paylaşım yapılan konuları gruplayarak hangi konuların daha çok etkili olduğunu araştırmışlardır. Çalışma kapsamında seçilen belirli firmalar tarafından paylaşılan 3200 adet Twitter içeriği toplanarak veri kümesi oluşturulmuştur. Veri ön işleme adımlarının ardından ilk olarak veri kümesi üzerinde istatistik programları kullanılarak Uygunluk Analizi yapılmıştır. Analiz sonucunda firmaların özgün içerik oluşturarak sosyal medyada faaliyetlerini devam ettirmeye çalıştığına ulaşılmıştır.

Aynı zamanda veri seti paylaşım yapılan konulara göre gruplanarak firmaların en çok hangi konu bazında içerik oluşturduđu çalışmanın diğeri bir amacı olarak görölmektedir. Bu analiz ise firmaların özel teklif kategorisinde yaptığı paylaşımlarda daha çok kullanıcı etkileşimine ulaştığı görölmüştür.

Sonuç olarak bu tez kapsamında diğeri çalışmalardan farklı olarak (markalar hakkında yapılan sosyal medya paylaşımlarının otomatik olarak sınıflandırması) yanı sıra kullanıcıların bu sonuçları inceleyebileceği bu doğrultuda aksiyon alabileceği bir bilgi sistemi sunulmuştur. Ayrıca, incelen Türkçe veri kümesi üzerinde yapılan çalışmalara kıyasla oldukça başarılı sınıflandırma skoruna sahiptir.



İKİNCİ BÖLÜM

MARKA AÇISINDAN İTİBAR VE SOSYAL MEDYA KAVRAMLARI

Sosyal medya paylaşımları üzerinden marka itibarı analizi yapılan bu çalışmada marka, itibar ve sosyal medya kavramları çalışmanın etki alanını gözlemlemek adına önemlidir. Bu bölümde bu kavramlara ait çeşitli tanımlamalar ve yaklaşımlara yer verilmiştir.

2.1 MARKA KAVRAMI

Bir ürün ya da hizmeti diğer rakiplerinden farklılaştıran ayırt edilmesini sağlayan sembol ya da isimlerden oluşan marka kavramına yönelik çeşitli tanımlamalar mevcuttur.

Amerikan Pazarlama Birliği (AMA) (Amerikan Pazarlama Birliği, 2020) marka kavramını "bir satıcının malını veya hizmetini diğer satıcılardan farklı olarak tanımlayan bir ad, terim, tasarım, sembol veya başka bir özellik" şeklinde tanımlamaktadır. Kotler ve arkadaşları (1999) ise markayı bir satıcı veya satıcı grubunun mal veya hizmetlerini tanımlamak ve bunları rakiplerinden ayırmak için kullandıkları bir isim, terim, işaret, sembol, tasarım veya bunların bir kombinasyonu olarak ifade etmektedir. Bir başka tanıma göre ise marka, bir ürünün veya üreticinin ticari markası veya ayırt edici adıdır. Bir satıcının mal ve hizmetlerini tanımlamak için kullanılan bir isim, terim, işaret, sembol, tasarım veya herhangi bir kombinasyondur (Palumbo ve Herbig, 2000). Fill'e (2013) göre bir marka, şirketin paydaşların gözündeki konumunu gösterdiğini belirtirken; Keller (2013), bir markanın farkındalık yaratan ve pazarda itibarını artıran bir şey olduğunu belirtmektedir. Bir ürün çoğu durumda rakipler tarafından çoğaltılabilir ve üretilebilir, ancak bir marka başkaları tarafından kopyalanamayan münhasır ve nadirdir (aktaran Bodeklint ve diğerleri, 2017). Marka kavramının tüketiciler ve üreticiler açısından çeşitli yararlar sağlamaktadır. Bu yararlar Tablo 1'de gösterilmektedir.

Kotler'e (1999:571) göre, markalar, tüketiciyi ürün kalitesi hakkında bilgilendirir. Aynı markayı sürekli olarak satın alan alıcılar, ürünü her satın aldıklarında aynı kaliteyi elde edeceklerinin farkındadırlar. Ayrıca marka isimleri,

jenerik markalamanın aksine, istek ve ihtiyaçlarına uygun ürünleri bulma yeteneklerini artırarak tüketiciler için alışverişi kolaylaştırır. Son olarak, marka adları tüketicilerin dikkatini kendileri için faydalı olan yeni ürünlere çekmesini sağlar, çünkü marka ilk tanınma şeklidir. Marka şirketler için kendilerini rakiplerinden ve potansiyel rakiplerinden ayırmak açısından önemlidir. Markalar ayrıca şirketlerin değerlerini müşterilerin zihinlerine ve kalplerine iletmeleri için bir araçtır (Tresna ve Wijaya,2015). Marka bilinci oluşturma, ürünlere ve hizmetlere değer katar. Bu değer, markanın kullanımından kazanılan aşinalık, güvenilirlik ve risk azaltma ve markayı kullanan diğer kişilerin deneyimi gibi unsurlardan kaynaklanmaktadır (Palumbo ve Herbig, 2000).

Tablo 1: Üretici ve Tüketici Açısından Marka Kavramı

Marka Kavramı	
Üretici Bakış Açısından	Tüketici Bakış Açısından
• Marka imajı oluşturmaya yardımcı olur. • Fiyat karşılaştırmasını düşürür. • Marka ismi sayesinde üretici aynı ürünü üreten diğer üreticilerden farklı bir fiyat belirleme şansına sahip olabilir. • Ürün hattının büyümesine imkan sağlar. Ürün hattının büyümesini isteyen üretici için iyi bir marka ismi avantaj sağlar. • Üreticinin markası yoluyla pazarını belirli ölçüde korumasına yardımcı olur.	• Marka tüketici için hangi ürünün tatmin edici olduğunu karar vermede yardımcı olur. • Marka güven ve kalite hakkında mesajlar vererek tüketicilerin satın alma davranışına yardımcı olur. • Marka tüketicinin kendisi için yararlı sayılabilecek yeni ürünlere dikkat etmesini sağlayabilir.

Kaynak: Odabaşı ve Oyman ,2002

Günümüz tüketicisi, markanın işlevlerinden yararlanmanın yanında, marka ile kendisi arasında bir bağ kurmaktadır. Aralarında oluşan bu bağ tüketici için çok önemli bir konuma sahip olmakta hatta tüketici, kişiliğini adeta marka ile özdeşleştirmektedir (Eren ve Erge, 2012: 4457). Bir marka hem fiziksel hem de algısal bir varlıktır. Bir markanın fiziksel yönü bir süpermarket rafında veya bir hizmetin teslimatında bulunabilir. Bir markanın algısal yönü, psikolojik alanda - tüketicinin zihninde

bulunur. Başarılı bir markanın, tüketiciye özgü nitelikleri (yani kalite, zarafet, değer) gösteren tanınmış bir adı vardır. Marka kavramı bir tüketicinin, rakipleri benzer ürüne yönelik daha düşük fiyatlar teklif etse bile tekrar satın alma yeteneği olarak ifade edilir. Bu da ilerleyen dönemlerde marka sadakati kavramıyla açıklanabilir (Palumbo ve Herbig, 2000). Karmaşık bir yapı oluşturan marka kavramını her ne kadar kurumlar ürün, hizmet, renk, slogan, kimlik, tasarım vb. gibi unsurları bir araya getirerek ona bir kimlik oluşturmaya çalıştırsa da markanın algılanışını ya da imajını oluşturanlar müşterileri, tüketicileri ya da paydaşlarıdır. Dolayısıyla marka kendi kimliğini ne kadar yansıtsa da ancak paydaşlarının algısındaki yeri ile zihinlerde konumlanacaktır.

2.2 MARKA İTİBARI

İtibar kavramı Türk Dil Kurumu Sözlüğünde (TDK) saygınlık ve güvenilir olma durumu olarak ifade edilmektedir (Türk Dil Kurumu, 2020). İtibar, özellikle pazarlama ve işletme alanında hem akademide hem de uygulamada önemli ve ilgi gören bir kavram olarak karşımıza çıkmaktadır.

Özel ve Sert'e (2015:150) göre itibarın disiplinerarası bir tanımı yoktur çünkü insanlar tarafından oluşturulmuş sosyal bir yapıdır. İtibar, karaktere, kelimelere ve davranışlara dayalı olarak zaman içinde kazanılmaktadır. Literatürde itibar kavramına yönelik çeşitli tanımlar mevcuttur. Buna göre itibar, bir nesnenin - kurum veya şirket - dış veya iç hissedarlar tarafından nasıl algılandığıdır." (Mateši ve diğerleri, 2010). İtibar, bir şirketin ya da endüstrinin genel değerler dizisine dayanan güvenilirliği, itimada layık oluşu, sorumluluğu ve yeterliği konusunda çok sayıda kişisel ve kolektif yargıdan oluşmuş bir bütündür (Budak ve Budak, 2014). Yazarlar tarafından itibara yönelik olarak yapılan tanımlamalardan bazıları Tablo 2'de gösterilmiştir.

İtibar, bir şirketin çalışanlar, müşteriler, tedarikçiler, distribütörler, rakipler ve halk açısından genel değerlendirmelerini temsil eder (Fombrun ve Shanley, 1990). Kadıbeşgil, (2013:59) ise itibarın, bir şirketin yarattığı güvenin toplam pazar içindeki katkı payı ve şirketin elle tutulamayan değerlerinin taşıyıcısı olarak ifade ederek toplum tarafından beğenilen, takdir edilen bir şirket olmanın karşılığı olduğunu vurgulamaktadır.

Walsh ve Beatty (2007: 129) ise müşteriye temel olarak oluşturduğu itibar tanımında “bir markanın itibarını müşterinin, bir firmanın ürünlerine, hizmetlerine, iletişim faaliyetlerine, firma ve temsilcileri veya kurucuları ile etkileşimleri ve bilinen kurumlara verdiği tepkiler temelinde bir firmanın genel değerlendirme faaliyeti” olarak tanımlamaktadır.

Tablo 2: Yazarlar Tarafından Yapılan İtibar Tanımları

Yazar(lar)	İtibar tanımı
Fombrun ve Shanley, 1990:234	Markaların hedef kitleleri için toplumdaki statüsünü en üst düzeye çıkarabilmesi için temel özelliklerini ifade eden rekabetçi bir sürecin sonucudur.
Herbig and Milewicz, 1993: 18	İtibar, kurumun varlığı süresince gerçekleştirdiği önceki tüm işlemlerinin toplam bileşimidir, tarihsel bir kavramdır ve bir işletmenin eylemlerinin uzun süre boyunca tutarlı olmasını gerektirir.
Doney and Cannon, 1997: 37	Endüstrideki firma ve sektördeki kişilerin kurumun dürüst ve müşterileri ile ilgili olma derecesine yönelik inancının olmasıdır.
Weiss et al.,1999: 75	İmaj bir firmanın neyi temsil ettiğini ifade ederken; itibar, pazarın açısından ne kadar iyi yaptığını ifade eder.
Fombrun et al. 2000:243	İtibar, bir kurumun paydaşlarını temsil eden bir grup için değerli sonuçlar sağlama yeteneğinin toplu değerlendirmesidir.
Bromley, 2001:317	İtibar, kişi veya bir kurum hakkındaki görüşlerin, paydaş veya çıkar grupları arasında yayılmasıdır.
Wang et al., 2003: 76	Özünde itibar, bir firmanın geçmişteki eylemlerinin bir sonucudur.
Rose and Thomsen, 2004: 202	Tüm paydaşların bir kurum ya da markayı algılamasıyla özdeşdir, diğer bir deyişle firma hakkında paydaşların bildiklerini düşündüklerine dayanır ve bu yüzden bir marka ya da şirketin itibarı temelde insanların algılarını yansıtır.

Kaynak: Walsh ve Beatty, 2007

Tüketicilerin algıları, inançları ve önyargıları bir kurumun kuruluşunun söylediği ve yaptıkları kadar itibarı için önemlidir. Bir şirketin itibarı en iyi, şirketin eylemleri ve ifadeleri arasındaki etkileşimin ve “dış dünyanın” bu eylemlere ve ifadelere tepkisinin bir sonucu olarak anlaşılır (Bunting ve Lipski, 2000). Kurumsal

İtibar birtakım temel ilkelerden oluşmaktadır. Frost bu temel ilkeleri şu şekilde sıralamaktadır (Frost, 2006):

- Her işletmenin veya kuruluşun bir itibarı vardır.
- Bir itibarın iyi olması yeterli değildir. İtibarın açık olması daha önemlidir. Netlik, güç ve esnekliğe yol açar.
- İtibar, tüm paydaşlara uyan tek bir boyuttan ibaret değildir.
- Kurumsal itibarın temel bir değerler veya ilkeler kümesi olması gerekir.
- Bir gecede itibar oluşturulamaz (ancak bir görüntü oluşturulabilir).
- İtibar bir gecede yok edilebilir.
- İtibarı yönetmek, bilgi ve gerçeklerin analizine dayanan bir süreçtir. Asla sadece tahminlerle elde edilemez.

Güçlü bir kurumsal itibarın kuruma sağlayacağı yararlar aşağıdaki gibi sıralanabilir (Feldman ve diğerleri, 2014):

- Tüketicinin markanın ürün veya hizmet kalitesi konusundaki algısı satışların artmasına ve marka hakkında olumlu görüşlerin ağızdan ağıza (word-of-mouth) yayılmasına katkıda bulunur.
- Markanın bünyesinde nitelikli personel istihdam etmesi ve bunu koruması,
- Çalışanların moralindeki artışla birlikte verimliliği artırır.
- Kriz veya rekabetçi saldırıların etkisini azaltarak işletmenin değerini korumasına yardımcı olur,
- Pazara etki etmesinin yanı sıra aynı zamanda önemli işbirliklerini kolaylaştırır ve böylelikle uluslararası genişlemeye öncülük edip yardımcı olur,
- Güvenilirlik sebebiyle daha fazla sayıda yatırımcı çeker, piyasa değeri yükseltir ve kurum için riskleri azaltır.
- Kurumu rakiplerinden ayırır ve daha iyi pazar konumlandırması sağlar,
- Daha ucuz sermaye erişimine izin verir.

Dolayısıyla tüm bu zorluklar temelinde oluşturulan bir markanın varlığını sürdürülebilmesi ve koruması için her yönden gelebilecek tehlikelere karşı hazırlıklı olmasını beraberinde getirmektedir. Marka itibarı, oldukça rekabetçi küresel pazarlarda marka liderliği ölçeğindeki anahtar faktörlerden biridir. Bir markanın itibarı, o markayla ilgili bilinen tüm algıları içermektedir (Kıyat, 2017).

2.3 SOSYAL MEDYA

Sosyal medya, web 2.0'ın ideolojik ve teknolojik temelleri üzerine inşa edilen ve kullanıcı tarafından içeriğin oluşturulmasına ve deęiş tokuşuna izin veren internet tabanlı uygulama grubudur (Kaplan ve Haenlein, 2010). Bir başka tanıma göre, sosyal medya, kayıtlı kişisel profillerle oluşturulan bağlantıları paylaşan Facebook, Twitter ve LinkedIn gibi sosyal ağ hizmetlerini kapsayan bir kavramdır ve içerik üretme ve paylaşma işlevini geliştiren etkileşimli ortamlardan biridir (Seo ve dięerleri, 2020). Sosyal medya, sohbet ortamını kullanarak bilgi, bilgi ve görüş paylaşmak için çevrimiçi toplanan insanların toplulukları arasındaki etkinlikleri, uygulamaları ve davranışları ifade eder. Konuşma ortamı, içerik oluşturmayı ve kelimeler, resimler, videolar ve sesler biçiminde kolayca aktarmayı mümkün kılan Web tabanlı uygulamalardır.

Sosyal medya, internette kullanıcılarının deneyim alışverişinde bulunmalarına, içerik paylaşımlarına ve kendileri içerik oluşturmalarına olanak sağlayan ortamlardır. Sosyal medyadaki sosyal bileşen, bir şekilde birbirine baęlı olan insanlar tarafından yürütölen paylaşım, işbirliği, bilgilendirme ve iletişim gibi eylemlerle tanımlanmaktadır (Hurre ve Postatny,2015).

Safko'ya (Safko,2010) göre kavramdaki sosyal kısım, insanların dięer insanlarla baęlantı kurmak zorunda olduęu içgüdüsel ihtiyaçları ifade etmektedir. Bunu türümüzün başlamasından bu yana şu ya da bu şekilde yapmaktayız. Evimizde hissedebileceğimiz ve düşüncelerimizi, fikirlerimizi ve deneyimlerimizi paylaşabildiğimiz benzer düşönen insanlardan oluşın gruplara dâhil olmaya ihtiyacımız vardır. Kavramdaki medya kısmı ise “dięer insanlarla bu baęlantıları yaptığımız medyayı”, dolayısıyla “bu baęlantıları yapmak için kullandığımız teknolojileri” ifade etmektedir. Bir başka ifadeye göre "sosyal medya" teriminin anlamı onu oluşturan iki kelimeden türetilir. Medya genellikle reklamları ve fikirlerin veya bilgilerin yayınlar / kanallar aracılığıyla iletilmesini ifade ederken; sosyal kısmı, bir grup veya topluluk içindeki bireylerin etkileşimini ima eder (Sajid, 2016).

Sosyal medya; yüksek derecede ölçeklenebilir ve erişilebilir teknolojiler kullanmakta, bireysel olarak insanların dięer insanlarla iletişim kurmasını

kolaylaştırmakta ve etki sağlamaktadır (Akar, 2011). Bugün sosyal medyada insanlar fiziksel şartlar veya kan bağları nedeniyle değil, ilgi alanları, ortak zevkler, hayaller, korkular veya çıkarlar zemininde bir araya gelerek sosyal gruplar oluşturmaktadır (Varnalı, 2013).

Geleneksel medyaya bir alternatif olarak ele alınabilecek olan sosyal medya sahip olduğu birtakım özellikler nedeniyle farklılaşmaktadır. Geleneksel ve sosyal medya arasındaki bu fark aşağıdaki Tablo 3’de gösterilmektedir. Söz konusu bu farklılıklar aynı zamanda markalar açısından içinde bulundukları ortam ve karşılaşılacak sorunlar ve fırsatlar konusunda da bir yol göstericidir.

Tablo 3: Geleneksel Medya ve Sosyal Medya Karşılaştırması

Geleneksel medya	Sosyal medya
• Sabit, değiştirilemez. • Sınırlı ve gerçek zamanlı olmayan • Sınırlı, gecikmeli ölçüm • Arşive zayıf erişim • Sınırlı medya karmaşı • Bir kurulca yayımlanır. • Paylaşım desteklenmez. • Denetim	• Anlık güncellenebilir. • Sınırsız, gerçek zamanlı yorum • Anlık popülerlik ölçümü • Arşive erişilebilir. • Tüm medya karma hale getirilebilir. • Bireysel yayımcılar vardır. • Paylaşım ve katılım desteklenir. • Özgürlük

Kaynak: Akar, 2011:30

Safko (2010:4) kurumlar açısından sosyal medyanın avantajlı taraflarını ele aldığında bu ortamların insanların kurum ya da markanız hakkında konuşmalarını sağlayabilmek için önemli avantajları olduğunu belirtmektedir. Bu açıdan bakıldığında eğer kurum ya da marka olarak harika bir ürün veya hizmetiniz varsa daha fazla müşteri bulabilmek için, sektörel anlamda bir ödül kazandıysanız veya çalışmak için harika bir yer olarak biliniyorsa insanların sizinle ilgili konuşmasını istersiniz. Bu tip durumlarda insanların sizinle, işletmenizle veya markanızla ilgili konuşması doğal olarak kötü bir şey değildir. Bu nedenle bu konuşmalar üzerinde kontrole sahip olmayabilirsiniz, ancak bir sosyal medya stratejiniz varsa ve bazı temel

araç ve taktiklerle rahat edebiliyorsanız, sosyal medyayı kendi yararınıza kullanabilirsiniz.

Ancak tam tersinin olduğu durumların da olabileceği göz önünde bulundurulması gerekmektedir. Safko (2010:4) bu doğrultuda işletmeler için sosyal medyanın ilk üç kuralını şu şekilde ifade etmektedir:

- Sosyal medya tamamen konuşmaların etkinleştirilmesi ile ilgilidir.
- İleti dizilerini denetleyemezsiniz, kontrol edemezsiniz ancak onları etkileyebilirsiniz.
- Etki, Sosyal medyada etki; ekonomik açıdan daha uygulanabilir ilişkilerin oluşturulması için temeldir.

Tüm avantajları ve dezavantajlarıyla birlikte sosyal medya platformları marka ve kurumlar için çok önemli bir noktadadır. Bu platformlarda paylaşılan içerikleri yönetmek, analiz etmek ve paylaşım sahipleri ile iletişime geçmek markanın değerini arttırmaktadır. Bu tez kapsamında bu amaca yönelik araştırma ve çalışmalara yer verilmiştir.

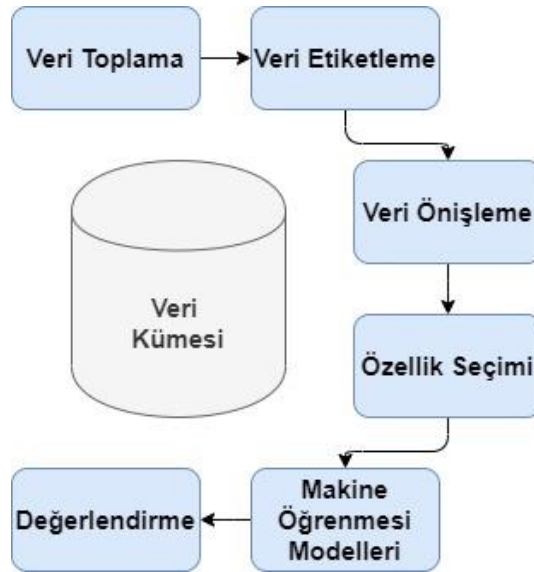
ÜÇÜNCÜ BÖLÜM

MAKİNE ÖĞRENMESİ SÜRECİ

Bu çalışma kapsamında, markaların sosyal medyadaki itibarlarının analizini yapmak amaçlanmıştır. Markalar veya kuruluşlar hakkında sosyal platformlarda yapılan paylaşımların olumlu yâda olumsuz olması platformdaki diğer kullanıcıların fikirlerini etkileyerek ilgili markaların itibarlarını belirlemektedir. Bu doğrultuda yapılan çalışmanın birincil hedefi markalar hakkında sosyal medyada yapılan paylaşımlarının toplanması ve değerlendirilmesidir. Toplanan içeriklerin değerlendirmesinde en büyük sorun ise bu işin insan gücüyle yapılmasının zor ve masraflı olmasıdır; çünkü sosyal platformlarda çok büyük sayılarda paylaşımlar yapılmaktadır. Belirtilen nedenlerden dolayı bu çalışma kapsamında süreci daha iyi yönetebilmek için metin madenciliği yöntemleri kullanarak toplanan içeriklerin sınıflandırılması hedeflenmiştir.

Önerilen çalışma veri toplama, veri ön işleme, özellik çıkarımı, denetimli makine öğrenmesi yöntemlerinin uygulanması ve uygulanan yöntemlerinin değerlendirmesi adımlarını içermektedir. Önerilen çalışmanın mimarisi Şekil 2'de sunulmaktadır.

Şekil 2: Metin Madenciliği Süreci

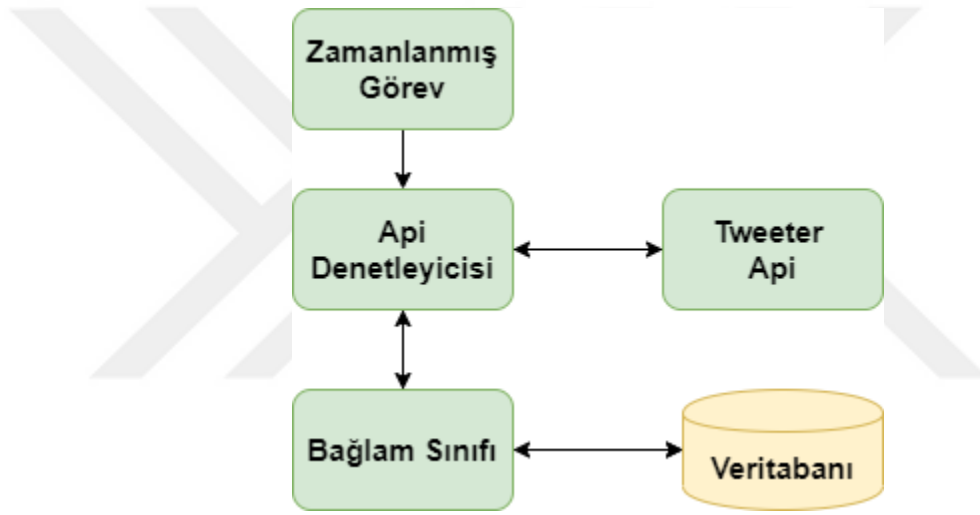


Kaynak: Yazar tarafından derlenmiştir.

3.1 VERİLERİN TOPLANMASI

Sosyal medyadaki markalar hakkında yapılan paylaşımların metin madenciliği teknikleriyle sınıflandırmak için bu çalışmada öncelikli olarak yeterli büyüklükte bir veri kümesinin toplanması planlanmıştır. Bu hedef doğrultusunda çok katmanlı tasarım mimarisinde bir yazılım geliştirilmiştir. İlgili yazılım Zamanlanmış Görev, Servis Denetleyicisi, Bağlam Sınıfı projelerinden oluşmak üzere üç katmanlı mimari yapısında geliştirilmiştir. İlgili yazılım mimarisinin şeması Şekil 3'de verilmektedir.

Şekil 3: Veri Toplama Uygulaması Mimarisi



Kaynak: Yazar tarafından derlenmiştir

Zamanlanmış Görev “.Net Core Framework” (.Net Core, 2020) kullanılarak geliştirilen bir konsol uygulamasıdır. Bu uygulama, Windows 10 (Windows 10, 2019) işletim sisteminde görev zamanlayıcısını kullanarak tanımlanan günler içerisinde 15 dakikada bir çalışacak şekilde ayarlanmıştır. **Zamanlanmış Görev** çalıştığında ise bulunduğu yazılım mimarisinin üst katmanında yer alan **Api Denetleyicisi** kütüphanesinden “GuncelPaylasimlariTopla” metodunu çağırılmaktadır.

Api Denetleyicisi Kütüphanesi resmi Twitter (Twitter, 2019) uygulama ara yüzünü kullanarak belirlenen markalar için güncel paylaşımların toplanması için “.Net Core Framework” kullanılarak geliştirilmiştir. Bu kütüphaneden örnek oluşturulduğunda verileri yasal olarak indirebilmek için kimlik bilgilerini Twitter

uygulama ara yüzüne gönderir. Kimlik bilgilerinin eşlenmesinden sonra yaratılan örnekten “GuncelPaylasimlariTopla” metodu çağırılarak belirlenen markalar için ilgili ara yüzden güncel paylaşımlar toplanmakta ve işlenmektedir. “GuncelPaylasimlariTopla” metodunun sözde kodu Şekil 4’de verilmektedir.

Şekil 4: Güncel Paylaşımları Toplama Metodunun Sözde (Pseudo) Kodu

```
GuncelPaylasimlariTopla()
{
  Markalar = Bağlam Sınıfı’ndan tanımlı markalar
  Markalar içerisindeki her marka için
    PaylasimlariAraKaydet (marka.İsmi)
    PaylasimlariAraKaydet (marka.SosyalMedyaHesapİsmi)
    PaylasimlariAraKaydet (marka.SosyalMedyaEtiketi)
  Döngü sonu
}

PaylasimlariAraKaydet(arama terimi)
{
  Paylasimlar=Twitter Api’den son 15 dakikada arama terimi içeren Türkçe paylaşımlar
  Paylasimlar içerisindeki her paylaşım için
    paylaşım işle ve bilgileriyle paylaşımModeli örneği yarat
    Bağlam Sınıfı’na paylaşımModeli örneğini gönder
  Döngü Sonu
}
```

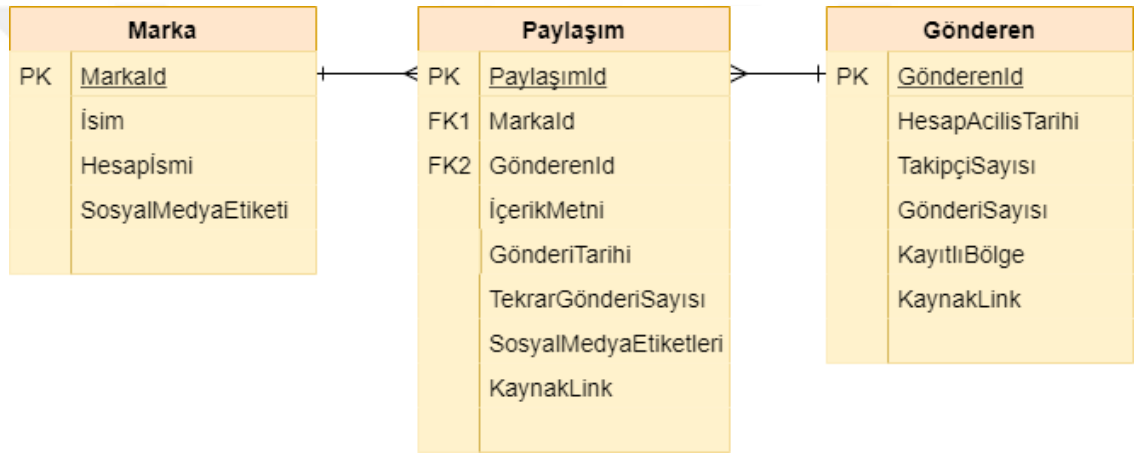
Kaynak: Yazar tarafından derlenmiştir.

“GuncelPaylasimlariTopla” metodu öncelikle **Bağlam Sınıfı**’ndan tanımlanmış marka modelinden nesneleri getirerek *Markalar* adlı listeye kaydetmektedir. Sonrasında bu metot *Markalar* listesinde bulunan her bir nesnenin ismi, sosyal medya hesap ismi ve sosyal medya etiketi için “*PaylasimlariAraKaydet*” metodunu çağırılmaktadır. “*PaylasimlariAraKaydet*” metodu ise kurumsal **Twitter Api** bağlantısını kullanarak topladığı son 15 dakikada *arama terimi* parametresini içeren Türkçe içerikleri *Paylasimlar* adlı listeye eklemektedir. İçerikleri toplama işleminden sonra *Paylasimlar* listesinden bulunan her nesneyi işleyerek bir model örneği haline getirip **Bağlam Sınıfı**’na göndermektedir.

Mimarinin son katmanı olan Bağlam Sınıfı Kütüphanesi “.Net Core Framework” ve “Entity Framework” kullanılarak geliştirilmiştir. Bağlam Sınıfı

kütüphanesini kullanarak tanımlanan model sınıfından üretilen örnekler ilişkisel veri tabanına kayıt edilebilmektedir. Ayrıca ilgili kütüphaneyi kullanarak istenen bilgiler ilişkisel veri tabanından model örnekleri olarak dönülebilmektedir. Uygulamanın ilişkisel veri tabanında yönetici tarafından tanımlanan markaların yansısı Twitter paylaşımları ve o paylaşımları yapan kullanıcıların bilgileri sırasıyla “Marka”, “Paylaşım” ve “Gönderen” varlıklarında saklanılmaktadır. İlgili varlıkların özellikleri ve aralarındaki ilişkiler Şekil 5’de verilen ER (varlık-iliski) diyagramında gösterilmektedir.

Şekil 5: ER (varlık-iliski) Diyagramı



Kaynak: Yazar tarafından derlenmiştir.

Marka varlığı ilgili kayıtlara dair yapılan paylaşımları arayabilmek için İsim, Hesap İsmi ve Sosyal Medya Etiketi özelliklerinden oluşmaktadır. Gönderen varlığı paylaşım yapan kullanıcıların bilgilerini saklamak için Hesap Açılış Tarihi, Takipçi Sayısı, Gönderi Sayısı, Kayıtlı Bölge ve Kaynak Link özelliklerinden oluşmaktadır. Son olarak Paylaşım varlığı toplanan içeriklerin özelliklerini saklamak üzere İçerik Metni, Gönderi Tarihi ve Tekrar Gönderi Sayısı, Etiketler ve Kaynak Link özelliklerinden oluşmaktadır.

Veri tabanında bir paylaşım sadece bir marka hakkında olabilirken, bir marka hakkında birden fazla paylaşım olabilecek şekilde tasarlanmıştır. Bunun sebebi ise veri tabanına kayıt edilen paylaşımların sadece bir marka hakkında olması şeklinde kısıtlanmasıdır. Ayrıca Gönderen ve Paylaşım arasındaki ilişki bir paylaşım sadece bir

kullanıcıya ait olabilirken, bir kullanıcı birden fazla paylaşıma sahip olabilecek şekilde tasarlanmıştır. Bu ilişkilere ait bilgiler Paylaşım varlığında var olan “MarkaId” ve “GönderenId” özellikleriyle saklanılmaktadır.

Sonuç olarak geliştirilen üç katmanlı mimariye sahip olan yazılım ile tanımlanan markalar için belirli günler içerisinde 15 dakikada bir güncel paylaşımlar düzenli ve efektif bir şekilde toplanmış olup veri tabanına kayıt edilmiştir.

3.2 VERİLERİN ETİKETLENMESİ

Verilerin toplanması adımından sonra, elde edilen sosyal ağ paylaşımlarının denetimli makine öğrenmesi modellerinde kullanılması için doğru ve efektif bir şekilde etiketlenmesi planlanmıştır.

Veri etiketleme işlemi ayrıntılara dikkat edilmesi gereken, tekrarlayan ve kullanım durumuna bağlı olarak değişiklik gösterebilen oldukça önemli bir süreçtir. Kullanılacak veri kümesinde yanlış etiketlemeler, üzerinde çalıştırılacak makine öğrenmesi modellerini yanlış yönlendirebileceği gibi değerlendirme sonuçlarında da hatalara neden olabilmektedir. Bu doğrultuda bu sürecin güvenli, daha hızlı ve sistematik yürütebilmesi için kitle kaynak kullanımı (crowdsourcing) modeli (Kruse ve Mukherjee,1998) oluşturulmuştur.

Bu çalışmada kitle kaynak modeli olarak .Net Core Framework yapısı kullanılarak bir web uygulaması geliştirilmiştir. Kitle kaynak kullanımı web uygulamasında, veri toplama uygulamasında kullanılan ilişkisel veri tabanındaki Paylaşım varlığına ek olarak “Etiket” ve “Etiketçi” varlıkları eklenmiştir. Web uygulamasının ilişkisel veri tabanının ER diyagramı Şekil 6’da gösterilmektedir.

Şekil 6: ER (varlık-ilişki) Diyagramı



Kaynak: Yazar tarafından derlenmiştir.

Veri tabanında bir etiket kaydı bir paylaşım için bir kullanıcının yaptığı veri etiketini ifade etmektedir. Bu web uygulamasına toplanan paylaşımları etiketlemesi için üç kullanıcı kayıt edilmiştir. Kullanıcılar birden fazla etiketleme işlemi yapabilmekte ve bir paylaşım farklı kullanıcılar tarafından yapılmış birden fazla etikete sahip olabilmektedir.

Sistemdeki kullanıcılara bir paylaşımın olumlu yâda olumsuz olma durumu örneklerle anlatılmış ve uygun olmayan paylaşımların filtrelenmesi konusunda bilgi verilmiştir. Kullanıcı kitlesel kaynak web uygulamasında oturum açtığında, veri tabanında var olan ve kendisinin daha önce etiketlemediği rastgele paylaşımlar Şekil 7'deki gibi gösterilmektedir.

Şekil 7: Veri Etiketleme Ekranı

Veri Etiketleme Web

Anasayfa

Hakkında

İletişim

Merhaba bahar.sardagi!

Çıkış

Etiketlenecek Paylaşımlar

Gönderen: [Redacted]

[Redacted] uzak durun [Redacted]. Sakın bilet almayın zarar ziyan kulliyen. [Redacted] 7 ay once ucuza al bileti 1 ay kala iptal etsinler. Su anda bilet fiyatları uçtu. Ee kim mesul 3 kisinin zararından?? Ben yaptım oldu ulkesi <https://t.co/bEMvcZJmFO>

☐ Olumlu

☐ Olumsuz

☐ UYGUN OLMAYAN VERİ

Kaydet

Kaynak: Yazar tarafından derlenmiştir.

Bu web uygulaması sayesinde, kullanıcı gösterilen paylaşımın içeriğine bakarak ilgili paylaşıma marka itibarı için olumlu yâda olumsuz olarak etiketleme işlemi yapabilmektedir. Ayrıca, kullanıcı anlamsız, belirsiz veya politik paylaşımları “Uygun Olmayan Veri” olarak etiketleyebilmektedir. Böylelikle muallak ve sakıncalı paylaşımlar veri kümesi dışında bırakılması amaçlanmıştır.

Veri etiketleme işleminin sonucunda yeterli sayıya ulaşıldığında, etiket tablosundan en az iki kullanıcı tarafından aynı etiketin yapıldığı paylaşımlar metin madenciliğinde kullanılmak üzere seçilmiştir. Böylelikle veri etiketleme süreci, geliştirilen kitle kaynak modelinde sistematik ve güvenli bir platformda kayıt altına alınarak yürütülmüştür. Ayrıca bir paylaşımın olumlu yâda olumsuz olması hakkındaki öznel yorumlar veya yönlendirmeler minimize edilmiştir.

3.3 VERİ ÖNİŞLEME

Metin madenciliğinde veri kümesinin etkin bir önışleme sürecinden geçmesi, makine öğrenmesi modellerinin başarısında oldukça önemli bir kıstastır. Ön işlemenin

temel amacı, metin dokümanlarındaki temel özellikleri veya anahtar terimleri elde etmek ve terim ile doküman arasındaki ilişki düzeyini artırmaktır. Bu doğrultuda önceki adımlarda oluşturulan veri kümesi (etiketleme sürecinde seçilmiş sosyal medya paylaşımları) üzerinde çeşitli metin temizleme teknikleri ve doküman terim matrisi yaklaşımları uygulanmıştır.

Metin temizlemede ilk adım olarak paylaşım metinleri için küçük harf dönüşümü yapılmıştır. Böylelikle makine öğrenmesi modellerinde büyük-küçük harf duyarlılığından oluşacak yanlış bilgi çıkarımın önüne geçilmesi amaçlanmıştır. İkinci adımda paylaşım metinleri içerisinde yer alan gereksiz sayılar kaldırılmıştır. Bunun sebebi ise paylaşım içerisinde yer alan sayıların bu tezdeki analiz çalışmasında çok alakalı ve ilişkili bir bilgi sunmamasından kaynaklanmaktadır. Son adım olarak gereksiz işaretlerin özellik olarak seçilmesine neden olabileceğinden dolayı, noktalama işaretleri ve semboller paylaşım içeriklerinden kaldırılmıştır.

Veri kümesinde yer alan temizlenmiş sosyal medya paylaşımları, makine öğrenmesi modellerinde kullanmak için doküman terim matrisi modeli kullanılmıştır. Doküman terim matrisi, bir metin doküman koleksiyonunda oluşan terimlerin sıklığını açıklayan matematiksel bir matristir (Torkkola,2004). Bu matriste, satırlar koleksiyondaki dokümanlara ve sütunlar koleksiyonda yer alan terimlere karşılık gelmektedir. Matristeki her bir hücre (doküman-terim ağırlığı) değerini belirlemek için çeşitli yaklaşımlar mevcuttur. Bu tez çalışmasında makine öğrenmesi modellerinde sıklıkla tercih edilen Terim Frekansı ve Terim Frekansı-Ters Doküman Frekansı yaklaşımlarının ikisi de uygulanmıştır.

Terim Frekansı (TF) yaklaşımında ağırlıklar bir terim ilgili dokümanda yer alma sayısı olarak belirlenmektedir (Xu ve Chen,2010). Böylelikle her bir doküman ağırlıklardan oluşan bir vektör olarak ifade edilebilmektedir. Örnek olarak “Çok çok keyifli bir uçuştı”, “Çok kötü bir uçuştı” ve “Güzel bir uçuştı.” sosyal paylaşımları bir koleksiyondaki dokümanlar olsun. Terim frekans yaklaşımını kullanarak ilgili paylaşımların doküman terim matrisi Tablo 4’de verilmektedir.

Tablo 4: Doküman Terim Matrisi

	Çok	Keyifli	Kötü	Bir	Uçuştı	Güzel
D1	2	1	0	1	1	0
D2	1	0	1	1	1	0
D3	0	0	0	1	1	1

Kaynak: Yazar tarafından derlenmiştir.

Terim Frekansı – Ters Doküman Frekansı (TF-TDF) TF ve Ters Doküman Frekansı (TDF) değerlerinin çarpılması ile hesaplanmaktadır (Wu ve diğerleri, 2008). N adet dokümandan oluşan bir terimin (TDF) değeri Formül 1’deki gibi hesaplanmaktadır.

Formül 1: Terim Doküman Frekansı

$$TDF = \log(N/DF)$$

Doküman Frekansı (DF) ise ilgili koleksiyonda t terimini içeren doküman sayısını ifade etmektedir. İlgili yaklaşım çoğu dokümanda sıklıkla geçebilecek “bir, bu, şu” gibi etkisiz kelimelerin ağırlıklarını düşürülmektedir. Örnek olarak Tablo 4’de verilen dokümanların TF-TDF değerleri Tablo 5’de verilmektedir.

Tablo 5: TF-TDF Değer Tablosu

	Çok	Keyifli	Kötü	Bir	Uçuştı	Güzel
D1	1.16	1.58	0	0	0	0
D2	0.58	0	1.58	0	0	0
D3	0	0	0	0	0	1.58

Kaynak: Yazar tarafından derlenmiştir.

Tablo 5’de görüldüğü üzere dokümanlarda sıklıkla geçen terimlerin ağırlıkları düşerken, nadir geçen terimlerin ağırlıkları artmaktadır.

Çalışmada oluşturulan veri kümesi üzerinde bahsedilen ön işleme teknikleri Scikit-learn (Scikit-Learn , 2019) kütüphanesi kullanılarak gerçekleştirilmiştir. Scikit-learn (sklearn olarak da bahsedilmektedir), Python programlama dili için açık kaynaklı

bir makine öğrenme kütüphanesidir. İlgili kütüphane çeşitli sınıflandırma, regresyon ve kümeleme algoritmalarına yansın metin madenciliği metotlarına da sahiptir.

3.4 ÖZELLİK SEÇİMİ

Ön işleme adımında elde edilen doküman terim matrisinin terim sayısı ile orantılı olarak çok büyük boyutlara ulaşması, makine öğrenmesi modellerinde karşılaşılan büyük zorluklardan bir tanesidir. Bir veri kümesinde bulunan on binlerce ve yüz binlerce terim (kelime), öğrenme süreci için yüksek bir hesaplama yüküne yol açmaktadır. Ayrıca, sınıflandırma aşamasında bazı alakasız ve gereksiz olabilecek terimler, sınıflandırma metotlarının tahminleme performansını düşürebilmektedir. Bu doğrultuda öğrenme sürecini hızlandırmak ve iyileştirmek için veri boyutunun azaltılması gerekli bir işlemdir. Metin sınıflandırma işlemlerinde veri boyutunun azaltılması genellikle özellik seçimi yaklaşımı ile yapılmaktadır.

Özellik seçimi, belirli bir veri kümesinde var olan özellik kümesinden, tanımlanmış özellik seçim (iyilik ölçütü) kriterine göre bir özellik alt kümesi oluşturma işlemidir. Bu tez çalışmasında özellik seçimi için Ki-Kare (Chi-Square) testi ve Bilgi Kazanımı (Information Gain) yaklaşımları uygulanmıştır.

Ki-Kare: iki değişkenin bağımsızlığını test etmek için istatistiksel çalışmalarda kullanılmaktadır (Zheng ve diğerleri,2004). Bu yaklaşımda, “herhangi bir iki değişken bağımsızdır” hipotezine dayanarak iki değişkenin birlikte beklenen gerçekleşme olasılığı B ilgili değişkenlerin bağımsız gerçekleşme olasılıklarının çarpılmasıyla elde edilmektedir. Gözlemlenen değer G ise veri kümesinde değişkenlerin birlikte gerçekleştiği sayıdır. Bu doğrultuda Ki-Kare skoru Formül 2’deki gibi hesaplanmaktadır.

Formül 2: Ki-Kare

$$\chi^2 = \sum \frac{(G_i - B_i)^2}{B_i}$$

İki değişken bağımsız olduğu durumda gözlemlenen sayı ve beklenen sayıya birbirine yakın olması durumunda daha küçük Ki-Kare skoru hesaplanmaktadır.

Bahsedilen durumdan dolayı, yüksek Ki-Kare skoru bağımsızlık hipotezinin yanlış olduğunu ifade etmektedir. Bir başka bir ifadeyle Ki-Kare değeri ne kadar yüksek ise iki değişken o derecede birbirine bağımlıdır. Özellik seçiminde bir özellik (terim) ve sınıfı (etiket) arasında Ki-Kare değeri ne kadar yüksekse seçilmesi o oranla başarılı sonuçlar vermesi beklenmektedir.

Örnek olarak bir veri kümesinde yarısı olumlu diğer yarısı olumsuz 3000 doküman olsun ve 300 doküman “rezalet” terimini içeriyor olsun. Bu veri kümesinde “rezalet” teriminin bir dokümanda geçme olasılığı $300/3000=0.1$ ’ken, bir dokümanın olumlu olması 0.5’dir. Bu durumda Ki-Kare hipotezine göre “rezalet” terimini içeren olumlu dokümanların beklenen sayısı $(0.5*0.1)*3000$ ’den (iki olayın bağımsız olduğu düşünüldüğünden) 150’dir. Eğer gözlemlenen değer beklenen değer 150’ye eşit olursa Ki-Kare skoru 0 olarak elde edilecektir. Bu da bir dokümanın “rezalet” terimini içermesinin dokümanın olumlu olup olmamasına bir etkisi olmadığını ifade etmektedir. Diğer yandan gözlemlenen değer beklenen değer 150’ye ne kadar uzak olursa Ki-Kare skoru o kadar yüksek olacaktır. Yüksek skor ise bir dokümanın rezalet terimini içermesinin ilgili dokümanın olumlu olup olmaması konusunda daha etkili olduğunu ifade etmektedir.

Bilgi Kazanımı: Makine öğrenimi çalışmalarında yaygın olarak kullanılan entropi tabanlı bir özellik değerlendirme yöntemidir (Roobaert ve diğerleri, 2006). Bilgi kazanımı hesaplaması göreceli olarak basit ve etkili olduğu için fikir madenciliği yâda metin sınıflandırma çalışmalarında özellik seçimi için sıklıkla tercih edilmektedir. Bir c sınıfı için t teriminin bilgi kazanımı Formül 3’deki gibi hesaplanmaktadır.

Formül 3: Bilgi Kazanımı

$$\begin{aligned}
 BG(t) = & - \sum_{i=1}^c P(c_i) \log P(c_i) \\
 & + P(t) \sum_{i=1}^c P(c_i|t) \log P(c_i|t) \\
 & + P(\bar{t}) \sum_{i=1}^c P(c_i|\bar{t}) \log P(c_i|\bar{t})
 \end{aligned}$$

C sınıfında i . kategori c_i olmak üzere, $P(c_i)$ ilgili kategorinin olasılığını ifade etmektedir. t teriminin dokümanlarda görülme olasılığı $P(t)$ olmak üzere, $P(c_i|t)$ t teriminin olduğu dokümanlarda i . kategorinin koşullu gerçekleşme olasılığıdır. t teriminin dokümanlarda görülmeme olasılığı $P(\bar{t})$ olmak üzere, $P(c_i|\bar{t})$ t teriminin olmadığı dokümanlarda i . kategorinin koşullu gerçekleşme olasılığıdır.

Bilgi Kazanımı yöntemi, eşik değerine dayanan filtre tabanlı bir özellik seçim tekniği olduğundan, genellikle bilgi kazancı değeri 0'dan büyük olan tüm özellikler öne çıkan özellikler olarak kabul edilmektedir.

3.5 MAKİNE ÖĞRENMESİ

Makine öğrenimi (ML), deneyim yoluyla otomatik olarak öğrenen ve gelişen bilgisayar algoritmaları üzerine bir çalışma alanıdır (Dietterich, 1997). Genel olarak makine öğrenmesi, yapay zekânın bir alt kümesi olarak görülmektedir. Makine öğrenmesi algoritmaları, karar almak veya tahminler yapmak için açıkça programlanmadan, "eğitim verisi" olarak bilinen örnek verilere dayalı bir matematiksel model oluşturmaktadır. Bu algoritma türleri geleneksel algoritmalar geliştirmenin zor veya mümkün olmadığı istenmeyen e-posta filtreleme gibi çeşitli alanlarda uygulanmaktadır.

Makine öğrenimi yaklaşımları öğrenme aşamasındaki gösterdikleri farklılıklara dayanarak üç ayrı kategoriye ayrılmaktadır (Sasakawa ve diğerleri, 2008). Bu kategoriler aşağıdaki gibidir.

- **Denetimli Öğrenme:** Öğrenme algoritmasına sunulan "denetmen" tarafından verilen örnek girdiler ve beklenen çıktılar doğrultusunda algoritmanın girdileri çıktılarla eşleyen genel bir kural öğrenmesidir.
- **Denetimsiz Öğrenme:** Öğrenme algoritmasına etiketlenmeden verilen veriler (beklenen çıktıları olmayan) sonucunda algoritmanın belirli kalıpları veya özellikleri keşfetmesidir.
- **Pekiştirmeli Öğrenme:** Geliştirilen program, belirli bir hedefi gerçekleştirmek için dinamik bir ortamla etkileşim (bir aracı sürmek veya bir rakibe karşı oyun oynamak gibi) halindeyken süreç içerisinde ödül ve ceza gibi geri dönüşlerle öğrenimin gerçekleştirilmesi durumudur.

Bahsedilen öğrenme yöntemleri ve genel olarak içerdikleri popüler algoritmalar şekil 8’de gösterilmiştir.

Denetimli öğrenme içerisinde yer alan regresyon ve sınıflandırma algoritmaları metin madenciliğinde sıklıkla kullanılmaktadır. Denetimli öğrenme algoritmalarının sıklıkla tercih edilme sebebi ise genellikle veri kümesinin kişiler tarafından etiketlenmiş dokümanlardan (metinlerden) oluşmasından kaynaklıdır.

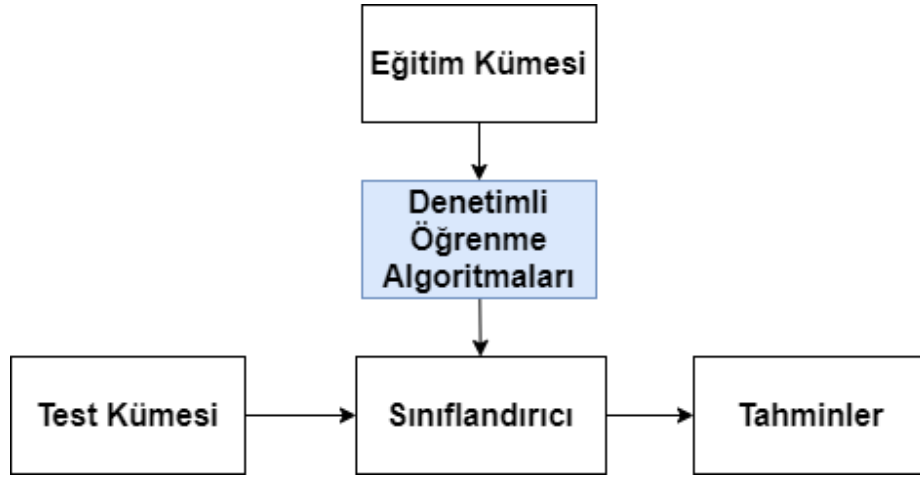
Şekil 8: Makine Öğrenmesi Yöntemleri



Kaynak: Yazar tarafından derlenmiştir.

Etiketleme işleminden sonra seçilmiş sosyal medya paylaşımları, ön işleme ve özellik seçimi adımlarından sonra oluşan veri kümesi denetimli öğrenme modeline uygun hale getirilmiştir. Sosyal medya paylaşımlarının otomatik olarak sınıflandırılması için uygulanan denetimli öğrenme süreci Şekil 9’da verilmektedir.

Şekil 9: Denetimli Makine Öğrenmesi Süreci



Kaynak: Yazar tarafından derlenmiştir.

Genel olarak bir denetimli öğrenme modelinde öğrenme süreci eğitim ve test olarak iki adımdan oluşmaktadır. Eğitim sürecinde veri kümesindeki bazı veriler eğitim kümesi olarak belirlenmektedir. Makine öğrenmesi algoritmaları eğitim kümesinden öğrenme gerçekleştirerek bir sınıflandırıcı model oluşturmaktadır. Test sürecinde ise oluşan sınıflandırma modelini kullanılarak eğitim kümesine alınmayan veriler (test kümesi) için tahmin yapılmaktadır. Test kümesindeki veriler için tahmin ve gerçek etiketlere bakarak makine öğrenmesi algoritmalarının başarısı hesaplanmaktadır. Süreç sonucunda ise genellikle doğru tahminleme yeteneği en yüksek sınıflandırıcı modeli oluşturan makine öğrenmesi algoritması tercih edilmektedir.

Bu tez çalışmasında öğrenme sürecinde uygulanan ve tahminle başarısı hesaplanan makine öğrenmesi algoritmaları aşağıda detaylı bir şekilde açıklanmıştır.

K-En Yakın Komşu: 1970'lerin başlarında istatistiksel tahmin ve örüntü tanımda kullanılmaya başlanmıştır. Basit bir mantığa sahip olan K-En Yakın Komşu (K-Nearest Neighbourhood - KNN) benzerlik metriğini kullanarak sınıflandırma işlemini yapmaktadır (Dudani, 1976).

Bu algorithma bir doküman için eğitim veri kümesinde ilgili dokümana en yakın olan K komşusuna bakılarak sınıflandırma yapılmaktadır. En yakın K komşu

içerisinde çoğunlukta olan sınıf hangisiyse ilgili doküman o sınıfa atanmaktadır. $K=1$ için vaka en yakın komşusunun sınıfına atanmaktadır.

K için en uygun değeri seçmek için öncelikli olarak verileri incelemek gerekmektedir. Genellikle büyük bir K değeri genel gürültüyü azalttığı için daha başarılı sonuçlar verebilmektedir, fakat bunun bir garantisi yoktur. İncelenen diğer çalışmalarda yer alan veri kümesi için en uygun K değeri 3 ile 10 arasında bulunmuştur.

Diğer yandan iki komşu yakınlığını, benzerliğini belirlemek için çeşitli fonksiyonlar mevcuttur. Popüler bazı uzaklık fonksiyonları ve n boyutlu x - y verileri için formülleri Tablo 6'da verilmektedir.

Tablo 6: Uzaklık Fonksiyonları ve Formülleri

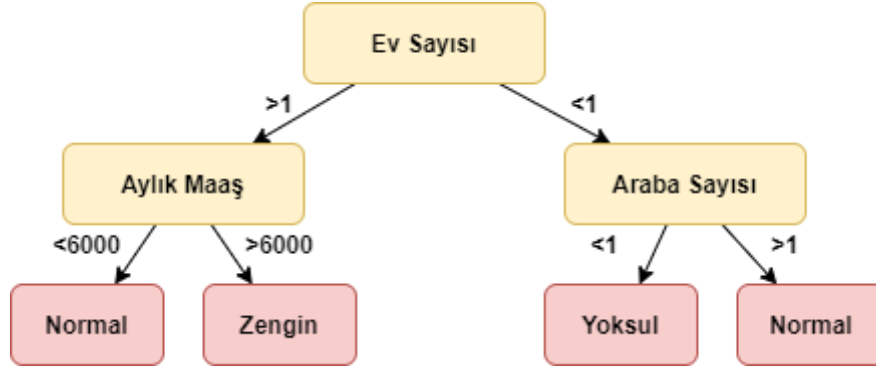
Fonksiyon	Formül
Euclidean	$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^n x_i - y_i $
Minkowski	$\sqrt[p]{\sum_{i=1}^n x_i - y_i ^p}$

Kaynak: Wettschereck, 1994

c4.5 Karar Ağacı: Ross Quinlan tarafından geliştirilen C4.5 algoritması eğitim veri kümesinden bir karar ağacı oluşturmak için kullanılmaktadır. Quinlan'ın önceki ID3 algoritmasının değişik bir sürümüdür (Ruggieri, 2002).

Bir karar ağacında her bir iç düğüm bir özelliği temsil etmekte, her dal bağlı olduğu özelliğin muhtemel değerlerini göstermekte ve her yaprak ise bir sınıf etiketini ifade etmektedir. Kökten yaprağa giden bütün yollar sınıflandırma kurallarını temsil etmektedir. Örnek olarak ev sayısı, aylık maaş, araba sayısı özellikleri ve maddi durum sınıf bilgisi olan veri kümesi için bir karar ağacı Şekil 10'da verilmektedir.

Şekil 10: Karar Ağacı Örneği



Kaynak: Yazar tarafından derlenmiştir.

Ev sayısı 1’den fazla ve aylık maaşı 7500’den yüksek olan yeni gelen bir örnek, Şekil 10’daki karar ağacına bakarak zengin olarak sınıflandırılacaktır.

C4.5 algoritması itératif olarak veri kümesini ve alt kümelerini sınıf bazında en etkin şekilde bölen özelliklerden başlayarak karar ağacı düğümleri seçilmektedir. Özellik seçimi işlemi bilgi kazanımı kistasına göre belirlenmektedir. Bilgi kazanımı bir özelliğin verileri sınıflandırmasındaki etkinliği ve entropide azalma olarak ifade edilmektedir. Bir X özelliği $p_1, p_2, \dots, p_k$ alt değerlerinde oluşsun, ilgili özelliğin entropisi Formül 4’deki gibi belirlenmektedir.

Formül 4: Entropi

$$Entropi = H(x) = - \sum_{i=1}^k p_i \log(p_i)$$

S örnek bir topluluk olsun, alt değerleri $p_1, p_2, \dots, p_k$ olan X özelliğinin bilgi kazanımı Formül 5’deki gibi hesaplanmaktadır. S_i , X özelliğinin p_i değerine sahip olanların alt kümesini temsil etmektedir.

Formül 5: Karar Ağacı Bilgi Kazanımı

$$Bilgi\ Kazanımı(S, X) = H(X) - \sum_{i=1}^k (|S_i|/|S|) Entropi(S_i)$$

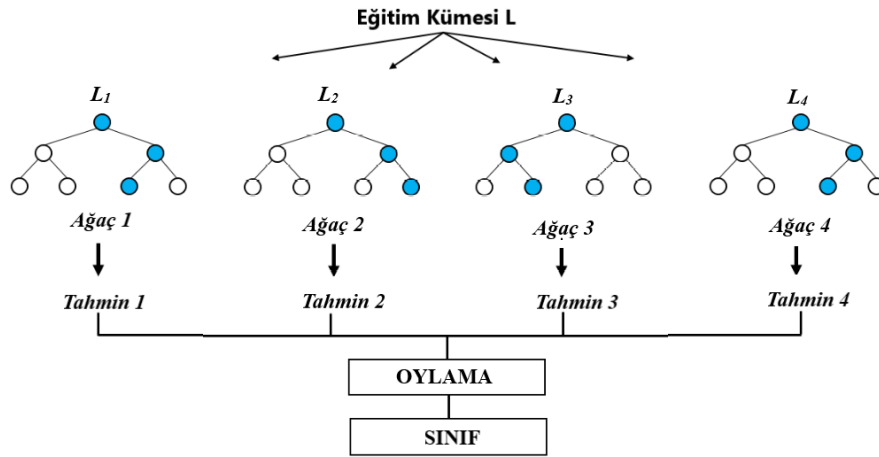
C4.5 algoritmasının başarısını artırmak için oluşturulan karar ağacında bazı ön budama veya budama yöntemleri kullanılmaktadır. Ön budama yöntemleri için ağaç

derinliđi, bir düğümdeki kayıt sayısı gibi durdurma kuralları uygulanabilmektedir. Bahsedilen budama yöntemleri kullanarak ağaç daha basit, yorumlanabilir hale getirilebilir ve aşırı öğrenmenin önüne geçilebilmektedir.

Rastgele Orman: Leo Breiman ve Adele Cutler tarafından önerilen karar ağaçlarından oluşan bir kolektif öğrenme yöntemidir (Pal, 2005). Algoritmanın ana hatları Şekil 11’de gösterilmektedir.

Rastgele orman algoritmasında ilk adımda p adet örnekten oluşan L eğitim kümesi için K tane yeni eğitim kümesi ($L_i, i=1...k$ için) yerine koyarak örneklenmektedir. $q < p$ olmak üzere, yeni oluşturulan eğitim kümeleri q adet örnek içermektedir. Ayrıca yeni eğitim kümeleri tekrar eden örnekler içerebilmektedir. İkinci adımda oluşturulan K tane eğitim kümesi için ayrı ayrı karar ağaçları eğitilerek rastgele orman oluşturulmaktadır. Eğitim aşamasından sonra yeni gelen bir veri için rastgele orman içerisindeki her bir karar ağacı sınıflandırma işleminde bulunup bu doğrultuda oy vermektedir. Rastgele ormada ilgili veri için en çok oyu alan sınıf sonuç olarak seçilmektedir.

Şekil 11: Rastgele Orman Süreci

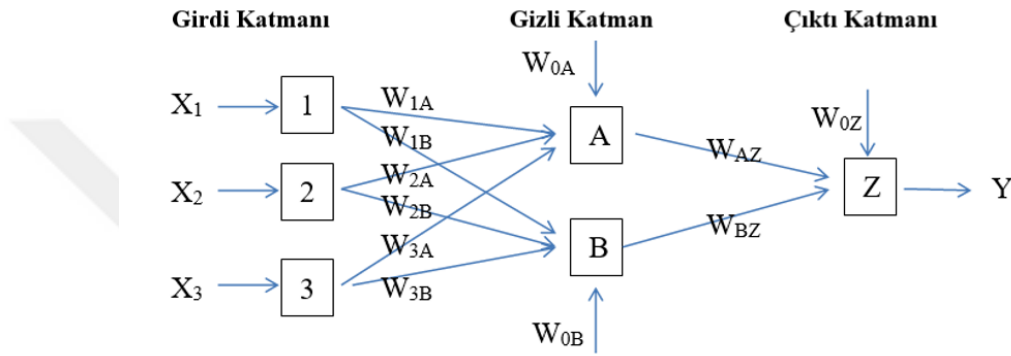


Kaynak: Yazar tarafından derlenmiştir.

Kolektif öğrenme yöntemi olan Rastgele Orman en başarılı denetimli öğrenme algoritmalarından bir tanesidir. Yapılan çalışmalarda oldukça güvenli tahminler yaptığı gözlemlenmiştir (Polikar, 2012).

Yapay Sinir Ağı (YSA): Biyolojik sinir ağlarının analogisinden esinlenerek tasarlanmış bir istatistiksel öğrenme yöntemidir (Psaltis ve diğerleri, 1988). YSA genel olarak birbirine bağlı mesaj gönderebilen "nöronlar" sistemi olarak ifade edilmektedir. Nöronlar arasında bağlantılar, deneyime dayalı olarak ayarlanabilen ve öğrenebilen sayısal ağırlıklara sahiptir. Örnek olarak $F(x_1, x_2, x_3) = Y$ için basit bir YSA modeli Şekil 12’de gösterilmektedir.

Şekil 12: Yapay Sinir Ağları Modeli



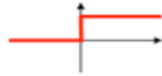
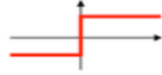
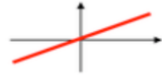
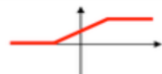
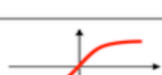
Kaynak: Yazar tarafından derlenmiştir.

Bir YSA modeli girdi, gizli ve çıktı olmak üzere üç katmandan oluşmaktadır. Şekil 12’deki modelde girdi katmanı 1, 2 ve 3 yapay nöronlarından oluşmaktadır. Bu yapay nöronlar tahminleme yapılmaya çalışılan fonksiyonun girdilerini temsil etmektedir. Şekil 12’deki modelde gizli katman A ve B nöronları çıktı katmanı ise Z nöronundan oluşmaktadır.

Katmanlardaki her bir nöron, bir bağlantı vasıtasıyla bir sonraki katmandaki nöronlarla bağlanmaktadır. Her bir bağlantı sinyal taşımaktadır ve taşıdığı sinyalin önemiyle orantılı olarak bir ağırlık değerine sahiptir. Bu doğrultuda her bir nöron (gizli katmanda bulunan yapay nöronlar hariç) kendisine gelen girdilere (sinyal) göre çıktı oluşturan bir aktivasyon fonksiyonuna sahiptir. Yapay sinir ağ modellerinde problem uzayına göre çeşitli aktivasyon fonksiyonları kullanılmaktadır. Gelen sinyaller $z = w_0 + x_1 * w_1 + x_2 * w_2 + \dots + x_n * w_n$ olmak üzere (girdiler ve ağırlıkların çarpımının toplamı) sıklıkla kullanılan aktivasyon fonksiyonları Şekil 13’de gösterilmektedir.

Şekil 13'deki girdi parametreleri x_1, x_2, x_3 için nöron bağlantıları ve aktivasyon fonksiyonu kullanılarak bir Y tahmini yapılmaktadır. Girdi parametreleri için gözlemlenen değer ve tahmin yapılan değer arasındaki fark hata olarak değerlendirilmektedir. Hesaplanan hata değeri için geriye yayılma (back propagation) algoritması kullanılarak bağlantı ağırlıkları güncellenmektedir.

Şekil 13: Yapay Sinir Ağ Modelleri Aktivasyon Fonksiyonları

Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer NN	

Kaynak: Aktivasyon Fonksiyonları,2020

Naive Bayes: Makine öğreniminde Bayes teoremindeki özellikler arasında naif bağımsızlık varsayımını uygulamaya dayanan basit olasılıklı sınıflandırma yöntemidir (Kibriya ve diğerleri, 2004). Sınıflandırılacak bir problem örneğinde $x = (x_1, \dots, x_n)$ bağımsız özelliklerin vektörü olsun, Naive Bayes her olası sınıf çıktısı için $p(C_k|x_1, \dots, x_n)$ olasılıklarını atayan koşullu olasılık modelidir. Bayes teoremini kullanarak koşullu olasılık Formül 6'daki gibi ayrıştırılmaktadır.

Formül 6: Koşullu Olasılık

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

Formülde yer olan olasılıklar takip eden listede açıklanmaktadır.

- $p(C_k|x)$: x özelliklerine sahip bir örneğin sınıfının k olma olasılığı (sonsal olasılık)
- $p(C_k)$: örneklemdeki herhangi bir örneğin sınıfının k olma olasılığı (öncül olasılık)
- $p(x|C_k)$: k sınıfına ait örneklerin x özelliklerine sahip olma olasılığı (olabilirlik)
- $p(x)$: örneklemdeki bir örneğin x özelliklerine sahip olma olasılığı (kanıt)

Naif bağımsızlık varsayımını ve Formül 7 kullanarak $p(C_k|x_1, \dots, x_n)$ hesaplanmaktadır.

Formül 7: Naive Bayes Yaklaşımı

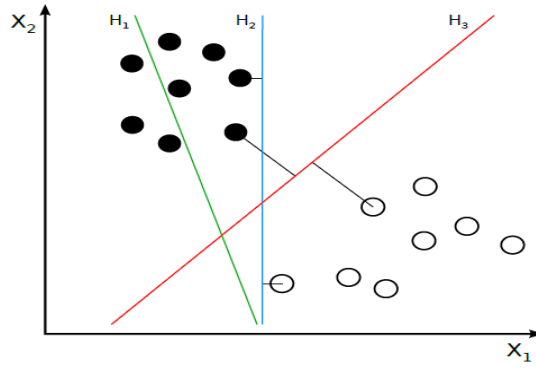
$$p(C_k|x_1, \dots, x_n) = p(C_k)p(x_1|C_k)p(x_2|C_k) \dots p(x_n|C_k)$$

Sonuç olarak verilen bir örnek tahmini için Naive Bayes yaklaşımında en yüksek sonsal olasılığa sahip sınıf seçilmektedir.

Naive Bayes yönteminin en büyük avantajı kolay uygulanabilir ve yüksek performanslı olmasıdır. Ayrıca, eğitim için çok büyük bir veri kümesine ihtiyaç duymamakta ve çoğu durumda başarılı sonuçlar vermektedir. Bahsedilen sebeplerden dolayı metin madenciliğinde Multinomial Naive Bayes modeli sıklıkla tercih edilmektedir.

Destek Vektör Makineleri: Makine öğreniminde, destek vektör makineleri (Support Vector Machines) (SVM) verileri analiz ederek örüntüleri tanımlayan ilişkili öğrenme algoritmalarına sahip denetimli öğrenme modelleridir. Sınıflandırma, regresyon veya diğer görevler için kullanılabilir (Suykens ve Vandewalle, 1999). Genel olarak, SVM yüksek veya sonsuz boyutlu bir uzayda bir hiper düzlem veya hiper düzlemler kümesi oluşturmaktadır. Bu yaklaşımda sezgisel olarak iyi bir ayrıştırma herhangi bir sınıfa ait örnekler arasında en büyük mesafeyi oluşturan bir hiper düzlem ile başarılmaktadır. Örnek olarak iki farklı sınıfa ait örnekler Şekil 14’de gösterilmektedir.

Şekil 14: Destek Vektör Makineleri



Kaynak: Yazar tarafından derlenmiştir.

Şekil 13’de H_1 sınıfları ayıramazken, H_2 ve H_3 sınıfları ayırabilmektedir. Fakat H_2 ayrıştırmayı küçük bir marjla sağlayabilmektedir. H_3 ise verilen örnek için maksimum mesafe (marj) ile ayrıştırma yapmaktadır.

Doğrusal olarak ayrıştırılamayan veriler için 1995 yılında Corinna Cortes ve Vladimir N. Vapnik farklı bir maksimum mesafe fikri önermişlerdir. İkili sınıflandırmayı doğrusal ayırabilecek herhangi bir hiper düzlem yoksa “Soft Margin” yöntemi örnekleri mümkün olduğunca en temiz bir şekilde ayrıştıran ve aynı zamanda ayrıştırma mesafesini maksimize eden bir hiper düzlem seçmektedir (Cortes ve Vapnik, 1995).

SVM yüksek boyutlu örneklem uzaylarında efektif bir şekilde çalışmaktadır. Ayrıca eğer boyut sayısı örnek sayısında fazla olduğu durumlarda da efektif sonuçlar verebilmektedir. En önemli nokta ise SVM eğer sınıflar arasında açık bir mesafe var ise oldukça başarılı sonuçlar vermektedir. Bir sosyal medya paylaşımın bir marka hakkında olumlu yâda olumsuz olması arasındaki fark yüksek olması beklendiğinden, SVM bu tez çalışması kapsamında uygulanmıştır.

3.6 MAKİNE ÖĞRENMESİ ALGORİTMALARININ DEĞERLENDİRİLMESİ

Verilen tüm problemlerde en iyi şekilde çalışacak bir sınıflandırma algoritması bulunmamaktadır, çünkü sınıflandırıcıların performansı büyük ölçüde veri kümesinin

özelliklerine göre değişkenlik göstermektedir (Narudin, 2016). Bu yüzden sınıflandırma algoritmalarının doğru bir şekilde değerlendirilmesi oldukça kritik bir öneme sahiptir.

Genel olarak denetimli öğrenme sürecinde veri kümesi eğitim ve test olarak ikiye ayrılmaktadır. Denetimli makine öğrenmesi algoritması eğitim veri kümesinden öğrenme sürecini tamamladıktan sonra test veri kümesi üzerinde değerlendirme yapılmaktadır. Eğitim ve test kümeleri ayrıştırılması verilerin özelliklerine, problemin amacına göre değişiklik göstermektedir. En basit yöntem olarak veri kümesi 2/3-1/3 ayırma kullanılmaktadır. Bu yöntemde eğitim kümesi veri setinin 2/3 oranında oluşturulmakta geri kalanlar ise test veri kümesine eklenmektedir. Genel olarak kullanılan ayrıştırma yöntemi ise k-Katlı Çapraz Geçerlemedir (Rodriguez ve diğerleri, 2009). Bu yöntemde L veri kümesi k parçaya ($i=1,2,...,k$ için L_k) ayrılmaktadır. Birinci adımda L_1 'i test için diğer k-1 parça eğitim için ayrılmaktadır. Bu işlem k adımda her L_k parçası test kümesi olacak şekilde yapılmaktadır. Sınıflandırma modelinin başarısı k test sonucunun ortalaması olarak değerlendirilmektedir.

Genel olarak bir sınıflandırma algoritmasının başarısının değerlendirmesinde dört temel veri yer almaktadır: doğru pozitif (DP) sayısı, doğru negatifler (DN), yanlış pozitifler (YP) ve yanlış negatifler (YN). İlgili terimlerin açıklamaları aşağıdaki gibidir.

- **Doğru Pozitif:** Test sonucu pozitif olarak etiketlenen gerçek pozitifler
- **Yanlış Pozitif:** Test sonucu pozitif olarak etiketlenen gerçek negatifler
- **Doğru Negatif:** Test sonucu negatif olarak etiketlenen gerçek negatifler
- **Yanlış Negatif:** Test sonucu negatif olarak etiketlenen gerçek pozitifler

Bu verileri kullanarak sınıflandırıcıların performanslarını ölçen birçok yöntem mevcuttur. Kesinlik (Precision), Duyarlılık (Recall) ve F-ölçütü (F-measure) ölçütleri bir sınıflandırma sisteminin kalitesini değerlendirmek için kullanılan popüler metriklerdir (Powers, 2011). İlgili ölçütlerinin formülleri Tablo 7'de verilmektedir.

Tablo 7: Ölçüt-Formül Tablosu

Ölçüt	Formül
Kesinlik	$DP / DP + YP$
Duyarlılık	$DP / DP + YN$
F-ölçütü	$(2 * Kesinlik * Duyarlılık) / (Kesinlik + Duyarlılık)$

Kaynak: Yazar tarafından derlenmiştir.

Test sonucunda doğru sınıflandırılmış pozitiflerin tüm pozitif olarak sınıflandırılan örneklere oranı kesinlik olarak tanımlanmamaktadır. Kesinlikte yanlış negatiflere dikkat edilmemektedir. Bu ölçütte önemli olan modelin testte pozitif olarak sınıflandırdığı örneklerin doğru olmasıdır.

Doğru sınıflandırılmış pozitiflerin tüm gerçek pozitiflere oranı ise duyarlılık olarak tanımlanmaktadır. Duyarlılıkta yanlış negatifler dikkate alınmaktadır. Bu ölçütte önemli olan gerçek pozitiflerin test sonucunda doğru bir şekilde tahmin edilmesidir.

Son olarak F-ölçütü kesinlik ve duyarlılığın harmonik ortalaması alınarak hesaplanmaktadır. Bu tez çalışmasında bahsedilen üç ölçüt de makine öğrenmesi algoritmalarının değerlendirilmesinde kullanılmıştır.

DÖRDÜNCÜ BÖLÜM

DENEYSEL ÇALIŞMA

Tez kapsamında markaların sosyal medyada kendileri hakkında yapılan olumlu veya olumsuz paylaşımları gözlemleyebilmesi ve yönetebilmesi için bir sistem geliştirilmesi hedeflenmiştir. Sistemi kısa süre içerisinde başarılı bir şekilde geliştirmek için Sistem Geliştirme Yaşam Döngüsü süreci uygulanmıştır. Bu bölümde ilk kısımda uygulanan Sistem Yaşam Döngüsü açıklanmıştır. İkinci kısımda ise markalar hakkında yapılan paylaşımların otomatik olarak sınıflandırılması için uygulanan makine öğrenmesi yöntemleri ve sonuçları verilmiştir. Son kısımda ise sistemin son kullanıcı ürünü olan Marka İtibar Yönetim Sistemi (MİYS) açıklanmıştır.

4.1 SİSTEM GELİŞTİRME YAŞAM DÖNGÜSÜ

Yüksek performanslı bilgi sistemleri geliştirmek için her adımın planlı bir şekilde uygulanması gerekmektedir. Sistem Geliştirme Yaşam Döngüsü (SGYD) mümkün olan en kısa sürede, en yüksek kalitede ve en düşük maliyetle bilgi sistemleri üretilmesini sağlayan bir süreçtir. Tez kapsamında geliştirilen sistem için uygulanan yaşam döngüsü Şekil 15’de verilmektedir.

Şekil 15: Sistem Geliştirme Yaşam Döngüsü



Kaynak: Tecim, 03.05.2020

Bu çalışmada çözülmesi gereken problem; sosyal ağlarda markaların itibarını etkileyen paylaşımların yönetilebilmesi olarak belirlenmiştir. Belirlenen problemin çözümünde ilk olarak sosyal medyadaki içeriklerin toplanması, içeriklerin otomatik olarak sınıflandırılması gerektiği tespit edilmiştir. Verilerin otomatik olarak sınıflandırılabilmesi için makine öğrenmesi yöntemleri araştırılmıştır. Gereksinimlerin belirlenmesi ve ihtiyaçların analizinden sonra iki adımdan oluşan bir sistem tasarlanmıştır. Sistemin birinci adımını paylaşımların toplanması ve toplanan paylaşımların makine öğrenmesi yöntemleriyle otomatik olarak sınıflandırılması olarak belirlenmiştir. İkinci adımda ise sınıflandırılan verilerin görselleştirilmesi, analiz edilebilmesi, yönetilebilmesi için bir web uygulaması geliştirilmesi hedeflenmiştir. Makine öğrenmesi sürecinde .Net ortamında geliştirilmiş Veri Toplama Uygulaması, Veri Etiketleme Uygulaması ve Python dili kullanarak denetimli sınıflandırma modelleri geliştirilmiştir. Sonraki süreçte .Net ortamında MVC mimarisinde bir web uygulaması olan Marka İtibar Yönetim Sistemi (MİYS) geliştirilmiştir. Denetimli sınıflandırma modelleri doğru tahmin edebilme başarısına göre test edilmiş olup, MİYS için kullanıcı etkileşim testleri yapılmıştır. Test sonuçlarına göre düzeltmeler yapıldıktan sonra sistem uygulamaya alınmıştır.

4.2 MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN UYGULANMASI

Bir marka itibar yönetim sistemi oluşturmak için belirlenen bir sosyal medya platformunda marka hakkındaki paylaşımların incelenmesi gereklidir. Sosyal medyada yer alan binlerce paylaşımın marka itibarı için pozitif yâda negatif olma durumunu insan gücüyle incelemek masraflı ve zaman alan bir süreç olacaktır. Bu tez çalışmasında metin madenciliği yöntemlerini kullanarak markalar hakkında paylaşılan sosyal medya içeriklerinin otomatik olarak sınıflandırılması amaçlanmıştır.

Öncelikle çalışma kapsamını özelleştirmek adına itibar analizi için ülkemizde faaliyet gösteren beş adet hava yolu firması seçilmiştir. Metodoloji bölümünde bahsedilen veri toplama uygulaması (VTU) kullanılarak 28/10/2019 ve 25/12/2019 tarihleri arasında ilgili firmalar hakkında yapılan 12,343 sosyal medya paylaşımı (Tweet) veri tabanına kayıt edilmiştir. Ayrıca paylaşımları yapan 10,171 Twitter kullanıcısının bilgileri de VTU aracılığıyla kayıt altına alınmıştır.

Belirlenen havayolu firmaları için yeterli sayıda içerik toplandıktan sonra veri kümesi oluşturmak için etiketleme işlemi Veri Etiketleme Uygulaması (VEU) aracılığıyla yapılmıştır. Uygulamaya kayıtlı olan üç kullanıcıya toplanan verilerin marka itibarı açısından nasıl sınıflandırılacağı (olumlu yâda olumsuz) ve hangi verilerin kapsam dışı bırakılması gerektiği detaylı bir şekilde açıklanmıştır. Veri etiketleme sürecinin sonunda bütün kullanıcılar tarafından aynı sınıfa atanan (hemfikir olunan) 3,533 adet içerik veri kümesine seçilmiştir. Seçilenlerin 1,187 tanesi olumlu, 2,346 tanesi olumsuz içeriklerdir.

Paylaşım metni ve sınıf (marka itibarı için pozitif yâda negatif olma durumu) özelliklerinden oluşan 3,533 seçilmiş veri için metodoloji kısmında bahsedilen ön işleme adımları uygulanmıştır. Ön işleme sonucunda 16,864 terimli Doküman Terim Matrisi (DTM) oluşmuştur. Matriste doküman-terim ağırlıkları belirlenirken terim frekansı ve Terim Frekansı – Ters Doküman Frekansı (TF-TDF) yaklaşımları ayrı ayrı uygulanmıştır.

16,864 özellikli (terim) DTM boyutunu indirgemek ve sınıflandırma işlemine pozitif katkı yapan özellikleri seçmek için iki ayrı yöntem uygulanmıştır. Birinci yöntem olan Ki-kare’de 15,502 özellik elenerek 1,362 özellik sınıflandırma işleminde kullanılmak üzere seçilmiştir. İkinci yöntem olan bilgi kazanımında 15,973 özellik elenerek 891 özellik sınıflandırma işleminde kullanılmak üzere seçilmiştir.

Ön işleme ve özellik seçimi aşamaları sonucunda 4 ayrı veri kümesi varyantı oluşmuştur. Bu varyantlar Tablo 8’de verilmektedir.

Tablo 8: Veri Kümesi Varyantları

Varyant	Terim Ağırlığı	Özellik Seçimi
Varyant ₁	Terim Frekansı	Ki-kare
Varyant ₂	Terim Frekansı	Bilgi Kazanımı
Varyant ₃	TF-TDF	Ki-kare
Varyant ₄	TF-TDF	Bilgi Kazanımı

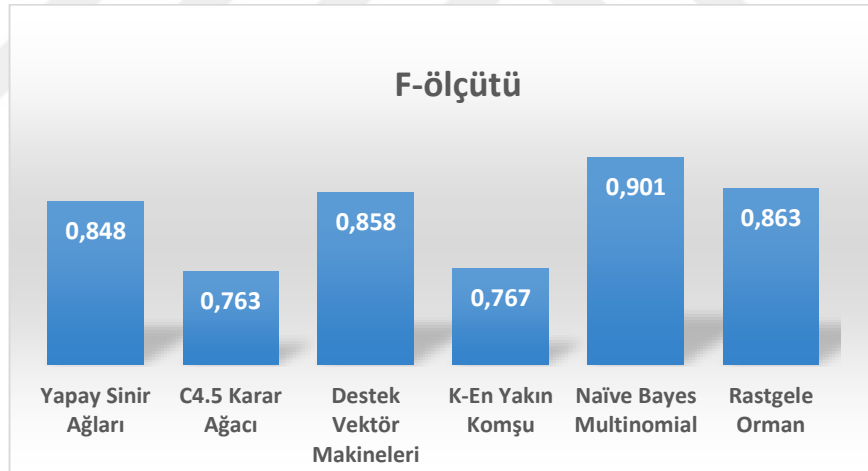
Kaynak: Yazar tarafından derlenmiştir.

Oluşan veri kümesi varyantları üzerinde altı farklı denetimli makine öğrenmesi algoritması uygulanmıştır. İlgili algoritmalar aşağıdaki gibidir.

- Naive Bayes
- Destek Vektör Makineleri
- Yapay Sinir Ağları
- C4.5 Karar Ağacı
- Rastgele Orman
- K-En yakın komşu

Eğitim kümesini ve test kümesini belirlemek için k-Katlı Çapraz Değerlendirme yöntemi kullanılmıştır. Makine öğrenmesi algoritmalarının test başarılarını ölçmek için Kesinlik, Duyarlılık ve F-ölçütü metrikleri hesaplanmıştır. Algoritmaların en başarılı olduğu eğitim kümesi varyantının test sonuçları Şekil 16’da verilmektedir.

Şekil 16: Test Sonuçları



Kaynak: Yazar tarafından derlenmiştir.

Oluşturulan dört veri kümesi varyantı üzerinde uygulanan altı farklı denetimli makine öğrenmesi algoritmalarının detaylı test değerlendirme sonuçları Tablo 9’da verilmektedir.

Sonuçlar gözlemlendiğinde ön işleme adımlarında TF-TDF ve Ki-Kare yöntemlerinin kullanıldığı veri kümesinde Naive Bayes algoritmasıyla oluşan sınıflandırma modeli %90 civarı F-ölçütü değeriyle dikkat çekmektedir. Rastgele

Orman, Destek Vektör Makineleri ve Yapay Sinir Ağları algoritmaları %84 üzeri F-ölçüt skorlarıyla başarılı sonuçlar vermiştir. Diğer yandan K-En Yakın Komşu ve C4.5 Karar Ağacı yöntemleri %80 altı F-ölçüt skoruyla diğer yöntemlere göre başarısız olmuşlardır.

Marka İtibar Yönetim Sisteminde kullanılmak üzere ön işleme adımlarında TF-TDF ve Ki-Kare yöntemlerinin kullanıldığı veri kümesinde oluşturulan Naive Bayes sınıflandırma modeli seçilmiştir. Bu sınıflandırma modelinin deneysel çalışma sonuçlarına dayanarak karşılaştığı bir sosyal medya içeriğini %90 oranında doğru sınıflandırması beklenmektedir.



Tablo 9: Denetimli Makine Öğrenmesi Algoritmalarının Değerlendirme Sonuçları

Varyant	Yöntem	Kesinlik	Duyarlılık	F-Ölçütü
Varyant ₁	Yapay Sinir Ağları	0,844	0,842	0,843
Varyant ₂	Yapay Sinir Ağları	0,848	0,849	0,848
Varyant ₃	Yapay Sinir Ağları	0,840	0,853	0,846
Varyant ₄	Yapay Sinir Ağları	0,842	0,847	0,844
Varyant1	C4.5 Karar Ağacı	0,765	0,759	0,762
Varyant2	C4.5 Karar Ağacı	0,765	0,759	0,762
Varyant3	C4.5 Karar Ağacı	0,766	0,760	0,763
Varyant4	C4.5 Karar Ağacı	0,766	0,760	0,763
Varyant1	Destek Vektör Makineleri	0,853	0,852	0,852
Varyant2	Destek Vektör Makineleri	0,858	0,858	0,858
Varyant3	Destek Vektör Makineleri	0,851	0,852	0,851
Varyant4	Destek Vektör Makineleri	0,861	0,856	0,858
Varyant1	K-En Yakın Komşu	0,776	0,759	0,767
Varyant2	K-En Yakın Komşu	0,780	0,742	0,761
Varyant3	K-En Yakın Komşu	0,760	0,731	0,745
Varyant4	K-En Yakın Komşu	0,774	0,719	0,745
Varyant1	Naïve Bayes	0,880	0,876	0,878
Varyant2	Naïve Bayes	0,860	0,859	0,859
Varyant3	Naïve Bayes	0,903	0,900	0,901
Varyant4	Naïve Bayes	0,880	0,878	0,879
Varyant1	Rastgele Orman	0,863	0,862	0,862
Varyant2	Rastgele Orman	0,861	0,859	0,860
Varyant3	Rastgele Orman	0,864	0,862	0,863
Varyant4	Rastgele Orman	0,862	0,860	0,861

Kaynak : Yazar tarafından derlenmiştir.

4.3 MARKA İTİBAR YÖNETİM SİSTEMİ (MİYS)

Yönetim Bilişim Sistemleri bir marka veya kuruluşa ait bilgilerin koordinasyonu, kontrolü, analizi ve görselleştirilmesi için kullanılan bilgi sistemleridir (Yurtsever ve Tecim,2016). Marka değerinin artmasında önemli yeri olan bu sistemler, yöneticilere kısa sürede etkili kararlar vermeleri için anlık ve uygun bilgiler sunmaktadır. Bu tez kapsamında yapılan çalışmada markaların sosyal medya üzerindeki itibarlarını ölçmeyi ve tüketici ile aralarındaki ilişkiyi yönetmeyi hedefleyen bir bilgi sistemi geliştirilmiştir.

Tez kapsamında geliştirilen bilgi sistemine Marka İtibar Yönetim Sistemi(MİYS) adı verilmiştir. MİYS .Net Core Framework kullanılarak MVC (Model-View-Controller) yazılım mimarisiyle geliştirilmiş bir web uygulamasıdır. .Net Core Framework Microsoft tarafından açık kaynaklı olarak kullanıma sunulmuş olup platform bağımsız uygulamalar geliştirmeye olanak sağlayan modern bir geliştirme ortamıdır. MVC(Model-View-Controller) ise uygulamanın iş mantığı ile kullanıcı ara yüzünü birbirinden ayrı olarak geliştirmeyi sağlayan bir yazılım mimarisidir (Leff ve Rayfield, 2001).

Sistemde kayıtlı marka hakkında yapılan sosyal medya paylaşımları tez kapsamında geliştirilen Veri Toplama Uygulaması(VTU) ile toplanmaktadır. Marka itibarının analizinde toplanan paylaşımların marka açısından bıraktığı izlenim önemli olduğu için sistemin ilk amacı paylaşımları hızlı ve doğru bir şekilde sınıflandırmaktır. Toplanan çok sayıda paylaşımların sınıflandırılması insan gücüyle zaman alıcı ve masraflı olabilmektedir. Bu aşamada herhangi bir insan gücüne gerek kalmadan toplanan verilerin otomatik olarak sınıflandırılması için tez kapsamında tasarlanan ve eğitilen %90 doğru sınıflandırma başarısına sahip olan Naive Bayes modeli kullanılmıştır. Böylelikle toplanan veriler 0.1 hata oranıyla çok hızlı ve masrafsız bir şekilde sınıflandırılabilir. Sistem elde edilen makine öğrenmesi sonuçlarını görselleştirerek kullanıcıya detaylı bilgiler sunmaktır. Ayrıca sistem marka hakkında olan sosyal medya içerikleriyle ilgili bazı aksiyonlara olanak tanımaktır.

Sisteme kayıtlı bir marka, kendisine tanımlanan yönetici hesabıyla kendi markasına ait analiz sonuçlarına erişebilmekte ve aksiyonlar alabilmektedir. Örnek kullanıcı etkileşimlerinin gösterilmesi için seçilen marka ise ülkemizde faaliyet

gösteren bir hava yolu (sistemde XHY olarak tanımlanmaktadır) firmasıdır. Bu doğrultuda uygulamanın içerdiği ekranlar takip eden kısımlarda açıklanmaktadır.

4.3.1 Analiz Ekranı

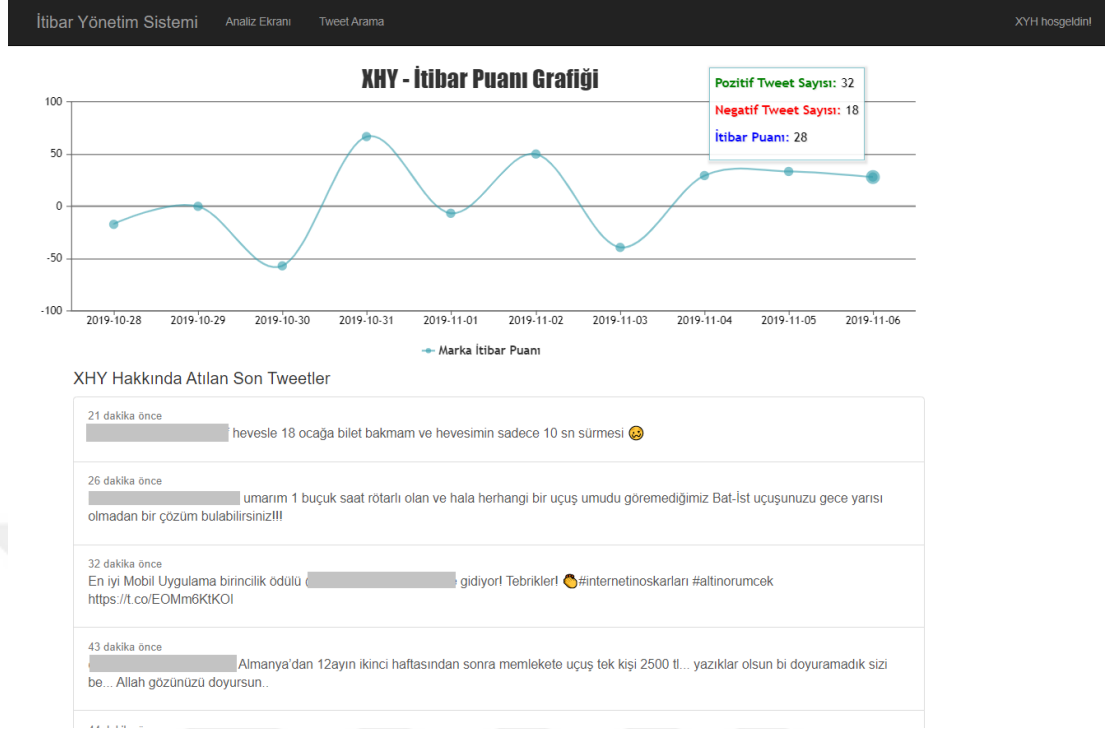
XHY (X Hava Yolu) firmasına ait kullanıcı sisteme giriş yaptıktan sonra marka hakkında paylaşılan sosyal medya içeriklerini analiz ekranında bulunan İtibar Puanı-Tarih grafiği yardımıyla inceleyebilmektedir. Günlük periyotlarla toplanan sosyal medya içerikleri makine öğrenmesi metotları ile pozitif ve negatif içerik olarak otomatik olarak sınıflandırılmış halde grafiğe yansıtılmıştır. Kullanıcı bu grafikte Negatif Tweet Sayısı, Pozitif Tweet Sayısı ve İtibar Puanı bilgilerine ulaşabilmektedir. Markanın itibar puanı bir gün içinde atılan pozitif paylaşım sayısının negatif paylaşım sayısından farkının toplam paylaşım sayısına oranıyla hesaplanmaktadır (Vidya ve diğerleri, 2015). Günlük İtibar Puanı (GİP) için hesaplama Formül 8’de gösterilmektedir.

Formül 8: Günlük İtibar Puanı (GİP) Hesaplama

$$GİP = \frac{(Pozitif Paylaşım Sayısı - Negatif Paylaşım Sayısı)}{(Pozitif Paylaşım Sayısı + Negatif Paylaşım Sayısı)} * 100$$

Aynı zamanda kullanıcı bu ekranda marka hakkında paylaşılan son sosyal medya içeriklerini görüntüleyebilmektedir. Şekil 17’de İtibar Analiz Ekranı gösterilmektedir.

Şekil 17: İtibar Analiz Ekranı



Kaynak: Yazar tarafından derlenmiştir.

4.3.2 Tweet Arama Ekranı

Analiz ekranında sayısal verilerle puan çizelgesini inceleyen kullanıcı belirli bir tarihte paylaşılan sosyal medya içeriklerini incelemek istediğinde “Tweet Arama” ekranını kullanabilmektedir. Örneğin kullanıcı Şekil 16’da bulunan XHY-İtibar Puan Çizelgesinde itibar puanının en yüksek veya en düşük hesaplandığı tarihlerde paylaşılan sosyal medya içeriklerini ayrıntılı incelemek isteyebilir. Bu tarz bir sistemde kullanıcıya sadece sayısal veriler sağlamak yeterli değildir. Kullanıcının marka hakkında paylaşılan içeriklerin hangi durumlarda pozitif ve negatif izlenim bıraktığı hakkında bilgi sahibi olması markanın gelecek dönemlerde alacağı aksiyonlar açısından oldukça önemlidir. Kullanıcı bu gibi durumlarda istediği periyottaki verilere Tweet Arama ekranından ulaşabilmektedir. Arama sonucunda kırmızı renkli satırlar marka hakkında negatif olan sosyal medya içeriklerini gösterirken, mavi renkli satırlar pozitif olan içerikleri göstermektedir. Ayrıca içeriklere ait Favorilenme ve Retweet sayısı gibi bilgiler sonuç tablosunda yer almaktadır. Sonuçlar (sosyal medya paylaşımları) Tarih, Tweet, Favorilenme Sayısı ve Retweet değerlerine göre

sıralanabilmektedir. Kullanıcının tarihe göre analiz yapması için geliştirilmiş olan Tweet Arama Sayfası Şekil 18’de görüldüğü gibidir.

Şekil 18: Tweet Arama Sayfası

İlbar Yönetim Sistemi

Analiz Ekranı

Tweet Arama

XYH hoşgeldin!

Tweet Arama

Seçili Tarih: 2019-11-03

Getir

Liste 10

ürün

Arama

Tarih	Tweet	Favorilenme Sayısı	Retweet Sayısı
22.29.13	TK 2984 SAW - Ankara uçuşundayız. 20.20 de havada olmamız gerekirken hala Bakan bekliyoruz. Milletvekilleri , vatandaşlar herkes bakan bekliyor. Bakanlarımız iyi bakmalıyız tabiki. Bu kafa ile nereye gideriz ?Zaman ayarsız Bakanlar, uçağa binince özür de dilemezler	34	0
08.42.06	Uçak içi eğlence sistemindeki filmlerde yer alan sevişme/öpmüş sahnelerini neden kesiyorsunuz? Yolcunun ne izleyeceğine siz mi karar vereceksiniz? Global bir hava yolunda böyle bir sansür kabul edilemez.	10	0
05.55.54	B787 Viyana'da dün akşam gerçekleştirdiği İstanbul-Viyana uçuşunu ilk kez TC-LLF tescilli B787 Dreamliner uçağı ile gerçekleştirdi. Viyana uçuşlarında dar gövde ve A330 tipi uçaklarını kullanmakta. @BoardingPassRO https://t.co/pGJ2saE4io	9	0
15.58.35	Hala havalimanın da uçak içinde bekletiliyoruz. Bu gidişle kayseriye gidemeyeceğiz (	3	0
08.27.47	3 Kasım 2019, Taşkent - İstanbul uçuşu Bunca kara propagandadan sonra uçaklarına bu suyu koymaması daha munasip olurdu. #hamidiyesu #hamidiyesukullan https://t.co/FQicnmf4zK	2	0
17.39.07	2 saat 45 dk rötar diyor simdi del otobüs tutun öyle gidelim Ankara İstanbul arasını Almanyadan mı geliyor bu uçak? #sikayetlar	2	0
00.05.44	Hizmet kaliteniz gün geçtikçe daha kötü oluyor.biletimi değiştirmek istiyorum sisteminizin çalışmadığını söylüyorsunuz.3 ve 5 yaşlarındaki çocukları ebeveyniden ayrı farklı koltuklara veriyorsunuz.star alianz üyesi olduğunuzdan eminmisiniz?	1	0
00.42.30	Trafik sebebiyle uçağı kacırdım.bir sonraki uçuş için bilet ayarlamaya çalışırken calcenter İbrahimBey'in inanılmaz özverisi,konusurken kesilen telefona ragmen yardım için anında tarafımıza geri dönüşü tamamen isini #çoksevmek ve %100 insanlığı ile ilgili. Benden tam puan	1	1

Kaynak: Yazar tarafından derlenmiştir.

Kullanıcı tarihe göre aradığı sosyal medya içeriklerinden belirli bir kelimeyi içeren verilere ulaşmak isteyebilir. Bu aşamada sağ üst köşede bulunan arama alanına arama yapmak istediği kelimeyi girerek sonuçlara ulaşabilir. Örnek olarak hava yolu firmaları için önemli yeri olan “Rötar” kelimesi için yapılan arama Şekil 19’da gösterilmiştir.

Şekil 19: Kelime Bazlı Arama

İtibar Yönetim Sistemi

Analiz Ekranı

Tweet Arama

XYH hoşgeldin!

Tweet Arama

Seçili Tarih: 2019-11-01

Getir

Liste: 10

ürün

Arama: rötar

Tarih	Tweet	Favorilenme Sayısı	Retweet Sayısı
00:23:32	Türkiye'nin en değerli markası [redacted] performansında ciddi bir düşüş var. Az önce havalimanındaydım bir çok uçuşta rötar var. Koltuktan para kazanma sevdasına 4 yaşında çocukla annesini ayrı ayrı koltuklara yerleştirecek kadar düşüncesiz bir kuruma dönüşmekte cabası !	1	0
20:03:23	Millet sizi rötar yapmıyor diye tercih ediyor, bu ne şimdi [redacted]	0	0

Sayfa: 1 / 1 (75 kayıt arasından filtrelendi.)

Önceki 1 Sonraki

Kaynak: Yazar tarafından derlenmiştir.

Bu sonuçlardan seçilen tarih için kaç kullanıcının rötar konusunda şikâyetini dile getirdiği gözlemlenebilmektedir.

4.3.3 Tweet Detay Bilgileri Ekranı

Kullanıcı Tweet Arama Ekranında görüntülediği sosyal medya içeriklerinin detaylı bilgilerine ilgili içeriğe tıklayarak Tweet Detay Bilgileri ekranından ulaşabilmektedir. Bu ekranda başlık kısmında seçilen içerik yer almaktadır. Kullanıcı bu içeriği markanın kurumsal sosyal medya hesabından cevaplama veya beğenme gibi işlemler yapabilmektedir. Cevaplama seçildiğinde kullanıcıyı ilgili içeriğin cevaplama linkine yönlendirmektedir. Beğenme seçildiğinde ise kullanıcının ilgili içeriği beğenme(favori) işlemi yapılmaktadır. Ayrıca bu ekranda detaylı bilgilerine erişilmek istenen içeriğe ait Tweet Bilgileri ve Gönderen Bilgileri olmak üzere iki ayrı sekme bulunmaktadır.

Tweet Bilgileri seçilen sosyal medya içeriğine ait gönderi tarihi, pozitif içerik durumu, favori sayısı, retweet sayısı, konu ve kaynak bilgilerine erişilmektedir.

“Pozitif İçerik” gönderinin makine öğrenmesi sonucunda içeriğinin sınıflandırma (pozitif veya negatif) bilgisidir. “Konu” bu içerikte hangi sosyal medya etiketlerini içerdiğini göstermektedir. “Kaynak” sosyal medya içeriğinin paylaşıldığı ilgili sosyal

medya linkini göstermektedir. Kullanıcı kaynak linke tıkladığında sosyal medya içeriğine direkt olarak erişebilmektedir. “Favorilenme Sayısı” ve “Retweet Sayısı” alanları paylaşılan içeriğin sosyal medyada ne kadar yayıldığına ölçülebilmesi adına önemli iki parametredir. Bu sayılar ne kadar yüksek olursa paylaşımın daha fazla kişiye ulaştığı öngörülebilir. Şekil 20’de Tweet Detay Bilgileri ve Tweet detay bilgilerini içeren görsel yer almaktadır.

Şekil 20: Tweet Detay Bilgileri Ekranı / Tweet Detay Bilgileri Sekmesi

İtibar Yönetim Sistemi

Analiz Ekranı

Tweet Arama

XYH hoşgeldin!

Tweet Detay Bilgileri

büyük bir rezilliği yaşadık. 40 kişi gidiş - dönüş bilet aldık. Gittik ama dönemedik. Asla olmayı hak etmiyorsunuz.Sorumlular hakkında gerekenin yapılmasını istiyorum.

★ Beğen

✍ Cevapla

Tweet Bilgileri

Gönderen Bilgileri

Gönderi Tarihi:

29.10.2019 13:51:39

Pozitif İçerik:

Hayır

Favori Sayısı:

6

Retweet Sayısı:

2

Konular:

Tekrar Gönderi:

Evet

Kaynak:

<https://twitter.com/sercin67/status/1189132685446434817>

Kaynak: Yazar tarafından derlenmiştir.

Gönderen Bilgileri sekmesinde sosyal medya içeriğini paylaşan kişiye ait bilgiler kullanıcı adı, kullanıcının sosyal medya hesabını oluşturduğu tarihi, takipçi sayısı, paylaşmış olduğu gönderi sayısı, kullanıcının kayıtlı olduğu bölge bilgisi ve kullanıcının sosyal medya hesabının linkini içeren kaynak bilgisi yer almaktadır. “Kullanıcının Takipçi Sayısı” paylaşımın retweet ve favorilenme olmaksızın minimum kaç kişiye ulaşabileceğini göstermektedir. Kullanıcının takipçi sayısı ne kadar fazla olursa marka hakkında paylaştığı içerik o kadar çok kişiye ulaşacaktır. Negatif olması durumunda marka için risk oluştururken, pozitif olması durumunda da marka açısından olumlu etki uyandırmaktadır. “Kaynak” sosyal medya içeriğinin paylaşan kişinin sosyal medya linkini göstermektedir. Kullanıcı kaynak linke

tıkladığında gönderen kullanıcı profiline direk olarak erişebilmektedir. Gönderen Bilgileri Sekmesi seçilmiş olan Tweet Detay Bilgileri ekranı Şekil 21’de verilmiştir.

Şekil 21: Tweet Detay Bilgileri Ekranı / Gönderen Bilgileri Sekmesi

İtibar Yönetim Sistemi

Analiz Ekranı

Tweet Arama

XYH hoşgeldin!

Tweet Detay Bilgileri

büyük bir rezillliği yaşadık. 40 kişi gidiş - dönüş bilet aldık. Gittik ama dönemedik. Asla A Star Alliance olmayı hak etmiyorsunuz.Sorumlular hakkında gerekenin yapılmasını istiyorum.

★ Beğen

✎ Cevapla

Tweet Bilgileri

Gönderen Bilgileri

Kullanıcı Adı:	sercin67
Hesap Tarihi:	20.12.2010 16:28:55
Takipçi Sayısı:	368
Gönderi Sayısı:	216
Kayıtlı Bölge:	Kdz.Ereğli
Kaynak:	sercin67

Kaynak: Yazar tarafından derlenmiştir.

SONUÇ

Sosyal medya ile birlikte internet üzerinden ulaşılabilen bilginin her geçen gün büyümesi, bu bilginin yönetimini ihtiyaç haline getirmiştir. Marka ve kurumlar kendi sosyal medya hesaplarını aktif olarak kullanarak kullanıcılar tarafından oluşturulan bu bilgiyi yönetmeye ve sorunlara çözüm bulmaya çalışmaktadır. Ancak bir sosyal ağ üzerinde belirli bir konu veya marka hakkında paylaşılan milyonlarca içerik olabilmektedir. Manuel olarak bu içeriklerin kontrolü, analiz edilmesi ve sonuca göre aksiyon alınması oldukça zor bir eylemdir.

Bu tez kapsamında sosyal medya kullanıcıları tarafından oluşturulan büyük miktarda veriyi otomatik olarak toplayan ve analiz eden bir sistemin markaların itibarlarına pozitif yönde etki edeceği düşünülmüştür. İlk aşamada marka hakkında sosyal medyada paylaşılan içerikleri belirli periyotlarla toplamak için bir uygulama(Veri Toplama Uygulaması) geliştirilmiştir. Toplanan verilerin denetimli makine öğrenmesi algoritmalarında kullanılmak üzere pozitif veya negatif olarak etiketlenmesi için Veri Etiketleme Uygulaması (VEU) geliştirilmiştir. Bu uygulama ile üç farklı kullanıcı kendilerine sunulan rastgele sosyal medya içeriklerini pozitif veya negatif olarak etiketlemişlerdir. En az iki kullanıcının aynı sınıflandırdığı içerikler makine öğrenmesi algoritmalarının uygulanacağı veri kümesine dâhil edilmiştir. Etiketleme sonrasında elde edilen veri kümesi üzerinde bir takım veri ön işleme adımları uygulanmıştır. Veri kümesi üzerinde Yapay Sinir Ağları, c4.5 Karar Ağacı, Destek Vektör Makineleri, K-En Yakın Komşu, Naive Bayes ve Rastgele Orman denetimli makine öğrenme algoritmaları uygulanmıştır. Uygulanan algoritmalar arasında %90 başarı oranına sahip Naive Bayes modeli sosyal medya içeriklerinin otomatik sınıflandırılması için seçilmiştir.

Çalışmanın ikinci aşamasında markaların kendileri hakkında sosyal ağlarda paylaşılan içeriklerin analizi ve denetimini sağlayan bir bilgi sistemi tasarlanmıştır. Bu sistemde Veri Toplama Uygulaması kullanılarak marka hakkında paylaşılan sosyal medya içerikleri günlük olarak toplanmaktadır. Toplanan bu içerikler geliştirilen makine öğrenmesi modeli ile otomatik olarak sınıflandırılmaktadır. Otomatik olarak pozitif veya negatif olarak sınıflandırılmış içerikler için gün bazlı bir marka itibar puanı hesaplanmaktadır. Sistemde markanın gün bazında aldığı pozitif ve negatif

içerik sayısı ve bu sayılara bağlı hesaplanan marka itibar puanı grafik olarak kullanıcı ile paylaşılmaktadır. Kullanıcı marka hakkında paylaşılan son sosyal medya içeriklerini görebilmektedir. Paylaşılan içeriğin kaç kişi tarafından tekrar paylaşıldığı veya beğenildiğini de inceleyebilmektedir. Ayrıca paylaştan kişi ile ilgili de detaylı bilgilere (takipçi sayısı, konum, hesap açılış tarihi) erişebilmektedir. Markanın paylaşımı yapan kişi ile kolay bir şekilde iletişime geçmesi için de sisteme çeşitli aksiyonlar (cevaplama, beğenme vb.) eklenmiştir.

Markaların günümüz rekabet ortamında varlıklarını sağlıklı bir şekilde sürdürebilmeleri için yöneticilerin anlık ve doğru bilgilere ihtiyaçları vardır. Bu bilgilere ulaşabilmek için Yönetim Bilişim Sistemleri kullanılarak insan gücü ve zaman tasarrufu yapılmaktadır. Tez kapsamında geliştirilen sistem markalar hakkında paylaşılan sosyal medya içeriklerini otomatik olarak analiz ederek yöneticilerin marka itibarına pozitif etki yaratmak amacı ile doğru ve hızlı karar almasına olanak sağlayan bir yönetim bilişim sistemidir.

İleriki çalışmalarda geliştirilen bilgi sisteminin Instagram, Facebook gibi bir çok sosyal medya platformuna entegre edilmesi amaçlanmaktadır. Böylelikle firmalar popüler sosyal ağlarda kendileri hakkında yapılan paylaşımlarına erişebilecek ve bütün sosyal ağlarını tek bir kanalda yönetebilecektir. Diğer yandan paylaşımların otomatik sınıflandırılması için seçilen makine öğrenmesi modelinde geliştirmeler yapılması amaçlanmaktadır. Bu doğrultuda birinci adım olarak daha büyük bir veri kümesi oluşturulması düşünülmektedir. İkinci adımda ise kelime temsili gibi gelişmiş metin madenciliği yöntemleri uygulanması düşünülmektedir. Sonuç olarak planlanan geliştirmelerle son kullanıcı için daha kapsamlı ve daha başarılı bir sistem oluşturulması hedeflenmektedir.

KAYNAKÇA

.Net Core Framework, 2019,

<https://www.microsoft.com/tr/download/details.aspx?id=17851> (01.10.2019)

Akar, E. (2011). Sosyal medya pazarlaması. Ankara: Elif yayınevi.

Aktivasyon Fonksiyonları,2020 <https://medium.com/datathings/dense-layers-explained-in-a-simple-way-62fe1db0ed75> (03.04.2020)

Albayrak, M., Topal, K. ve Altıntaş, V. (2017). Sosyal Medya Üzerinde Veri Analizi: Twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22, 1991-1998.

Amerikan Pazarlama Birliği, 2020, <https://www.ama.org/topics/branding/> (11.03.2020)

Ayan, B., Can, M. ve Gürsoy, U. T. (2019). Sosyal Medya Madenciliği İle Firmaların Twitter Verilerinin İncelenmesi, *International Journal of Economic and Administrative Studies*, (23).

Barnes, N. G., Cass, J., Getgood, S., Gillin, P. ve Goosieaux, F. (2008). Exploring The Link Between Customer Care And Brand Reputation in The Age of Social Media. S. f. NC Research (Ed.): Society for New Communication Research, 214.

Bernoff, J. ve Chadler, T. (2010). Empowered. Harvard Business Review, July–August, (s. 94–101).

Bodeklint, K., Unosson, W. ve Lindhe, A. (2017). Crisis Management and Brand Reputation: An Exploratory Study on Crisis Management And Its Effects on Brand Reputation.

Budak, G. ve Budak, G. (2014). İmaj Mühendisliği Vizyonundan Halkla İlişkiler. Nobel Akademik Yayıncılık

Bunting, M. ve Lipski, R. (2001). Drowned Out? Rethinking Corporate Reputation Management For The Internet. *Journal of Communication Management*, 5(2), 170-178.

Choi, S., Kim, Y. J., Briceno, S. ve Mavris, D. (2016, September). Prediction of weather-induced airline delays based on machine learning algorithms. *In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (s. 1-6).

Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

Dayanc Kiyat, G. B. (2017). The Relationship Between Reputation Quotient and Perceived Quality of Brands in Turkey's Mobile Communication Sector.

Dietterich, T. G. (1997). Machine-Learning Research. *AI magazine*, 18(4), 97-97.

Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *Transactions on Systems, Man, and Cybernetics*, (4), 325-327.

Eren, S. S. ve Erge, A. (2012), Marka Güveni, Marka Memnuniyeti ve Müşteri Değerinin Tüketicilerin Marka Sadakati Üzerine Etkisi. *Journal Of Yaşar University*, 7 (26), 4455 – 4482

Feldman, P. M., Bahamonde, R. A. ve Velasquez B. (2014). A New Approach For Measuring Corporate Reputation. *Revista de Administração de Empresas*, 54(1), 53-66.

Fombrun, C. ve Shanley, M. (1990). What's in a Name? Reputation Building And Corporate Strategy. *Academy of Management Journal*, 33, 233–258

Fong, N. H. ve Yazdanifard, R. (2014 January) The Impacts Of Social Media Marketing, Perceived Risks And Domain Specific Innovativeness On Online Consumer Behavior. *Global Journal of Commerce and Management Perspective*. 3. 22-26.

Frost, A. R. (2006). “Building and Sustaining a Dynamic Corporate Reputation”. *The IABC Handbook of Organizational Communication: A Guide to Internal*

Communication, Public Relations, Marketing, and Leadership kitabı içinde. (Ed: Tamara L. Gillis). John Wiley and Sons: San Francisco.

Gok, B. ve Fidan, U. (2019). Kurumsal İtibar Algısının Satın Alma Niyetine Etkisinin Karar Ağacı İle Haritalandırılması. *Business and Economics Research Journal*, 10(3).

Gurău, C. (2008). Integrated online marketing communication: implementation and management. *Journal of Communication Management*, 12(2), 169–184

Hasan, A., Moin, S., Karim, A. ve Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, 23(1), 11.

Hurtle, D. ve Postatny, J. (2015). *Social Media for Scientific Institutions: how to attract young academics by using social media as a marketing tool*. Springer.

Isah, H., Trundle, P. ve Neagu, D. (2014, September). Social Media Analysis For Product Safety Using Text Mining And Sentiment Analysis. *14th UK Workshop On Computational Intelligence (UKCI)* (s. 1-7).

Kadıbeşgil, S. (2013). İtibar Yönetimi. *Mediacat: İstanbul* 7. Baskı

Kaplan, Andreas M. ve Michael H. (2010). "Users of The World, Unite! The Challenges and Opportunities of Social Media". *Business Horizons* 53 (1): 59–68.

Kaul, A., Chaudhri, V. ve Cherian, D. (2015). Social Media: The New Mantra For Managing Reputation. *Vikalpa*, 40(4), 455-491.

Kibriya, A. M., Frank, E., Pfahringer, B. ve Holmes, G. (2004, December). Multinomial naive bayes for text categorization revisited. *In Australasian Joint Conference on Artificial Intelligence* (s. 488-499).

Kim, Y., Jeong J.R., Dwivedi R. Ve Zhang J. (2016), "Competitive Intelligence in Social Media Twitter: iPhone 6 vs. Galaxy S5", *Online Information Review*, Vol. 40 Iss 1.

Kotler, P., Armstrong, G., Saunders, J. ve Wong, V. (1999). Principles of Marketing, Second European Edition.

Kruse, H. ve Mukherjee, A. (1998, March). Preprocessing text to improve compression ratios. *In Proceedings DCC'98 Data Compression Conference* (Cat. No. 98TB100225) (s. 556).

Krüger, N., Stieglitz, S. ve Potthoff, T. (2012, December). Brand Communication In Twitter-A Case Study On Adidas. In PACIS (s. 161).

Leff, A. ve Rayfield, J. T. (2001, September). Web-application development using the model/view/controller design pattern. *In Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference* (s. 118-127).

Manaman, H. S., Jamali, S. ve AleAhmad, A. (2016). Online reputation measurement of companies based on user-generated content in online social networks. *Computers in Human Behavior*, 54, (s. 94-100).

Matešić, M., Vučković, K. ve Dovedan, Z. (2010, May). Should Academia Care About Online Reputation Management And Monitoring?. *In The 33rd International Convention MIPRO* (s. 852-857).

Mittal, A. ve Goel, A. (2012). Stock Prediction Using Twitter Sentiment Analysis. Stanford University, CS229 (<http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>), 15.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.

Nalçakan, Y., Bayramoğlu, Ş. S. ve Tuna, S. (2015). Sosyal Medya Verileri Üzerinde Yapay Öğrenme ile Duygu Analizi Çalışması. Technical Report.

Narudin, F. A., Feizollah, A., Anuar, N. B. ve Gani, A. (2016). Evaluation of machine learning classifiers for mobile malware detection. *Soft Computing*, 20(1), 343-357.

Neethu, M. S. ve Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. *In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (s. 1-5). IEEE.

Odabaşı, Y. ve Oyman, M.(2002). Pazarlama İletişimi Yönetimi, Media Cat Yayınları, 4.Basım,İstanbul

Öngün, T. F. (2014). Markalar için online itibar yönetimi. <https://www.trendweek.com/online-itibar-yonetimi/> Erişim Tar. 02.04.2020.

Özel, A. P. ve Sert, N. Y. (2015). Dijital Halkla İlişkiler: Kavram Ve Araçları. Derin Yayınları.

Öztamur, D. ve Karakadılar, İ. S. (2014). Exploring the role of social media for SMEs: as a new marketing strategy tool for the firm performance perspective. *Procedia-Social and Behavioral Sciences*, 150, 511-520.

Pal, M. (2005). Random Forest Classifier For Remote Sensing Classification. *International Journal of Remote Sensing*, 26(1), 217-222.

Palumbo, F. ve Herbig, P. (2000). The Multicultural Context Of Brand Loyalty. *European Journal of Innovation Management*.

Polikar, R. (2012). Ensemble Learning. In Ensemble Machine Learning (s. 1-34). Springer, Boston, MA

Powers, D. M. (2011). Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness And Correlation.

Prihono, O. F. ve Sari, P. K. (2019). Comparison Analysis Of Social Influence Marketing For Mobile Payment Using Support Vector Machine. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(4), 367-374.

Psaltis, D., Sideris, A. ve Yamamura, A. A. (1988). A Multilayered Neural Network Controller. *IEEE Control Systems Magazine*, 8(2), 17-21.

Reputation Lessons from Warren Buffet ,2020

<https://www.reputationdefender.com/blog/orm/reputation-lessons-warren-buffett> (2.04.2020)

- Rodriguez, J. D., Perez, A. ve Lozano, J. A. (2009). Sensitivity Analysis Of K-Fold Cross Validation In Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569-575
- Roobaert, D., Karakoulas, G. ve Chawla, N. V. (2006). Information Gain, Correlation And Support Vector Machines. in Feature Extraction (s. 463-470). Springer, Berlin, Heidelberg.
- Ruggieri, S. (2002). Efficient C4. 5 [classification algorithm]. *IEEE Transactions on Knowledge and Data Engineering*, 14(2), 438-444.
- Safko, L., 2010. The Social Media Bible : Tactics, Tools and Strategies for Business Success. 2nd edition ed. Hoboken, NJ: Wiley.
- Sajid, S. I. (2016). Social media and its role in marketing. *International Journal of Enterprise Computing and Business Systems*
- Salampasis, M., Paltoglou, G. ve Giachanou, A. (2014). Using Social Media for Continuous Monitoring and Mining of Consumer Behaviour. *International Journal of Electronic Business*, 11(1), 85-96.
- Sasakawa, T., Hu, J. ve Hirasawa, K. (2008). A Brainlike Learning System with Supervised, Unsupervised, and Reinforcement Learning. *Electrical Engineering in Japan*, 162(1), 32-39.
- Scikit-Learn ,2019, <https://scikit-learn.org/stable/> (15.10.2019)
- Seo, E.J.; Jin-Woo Park, J.W. ve Yu Jin Choi, Y.J.(2020). The Effect of Social Media Usage Characteristics on e-WOM, Trust, and Brand Equity: Focusing on Users of Airline Social Media. *Sustainability* (12).1691
- Spyder (2020) <https://www.spyder-ide.org/> (12.02.2020)
- Sriram, G. ve Kathiravan, C. (2014). Effect Of Social Media And Factors Influencing A Brand's Reputation. 366-378.
- Suykens, J. A. ve Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293-300.

Tecim, V. Sistem Geliştirme Yaşam Döngüsü. <http://debis.deu.edu.tr/userweb//vahap.tecim/dosyalar/sgyd.pdf>. (03.05.2020).

Torkkola, K. (2004). Discriminative Features for Text Document Classification. *Formal Pattern Analysis and Applications*, 6(4), 301-308.

Tresna, L. ve Wijaya, J. C. (2015). The impact of social media towards brand equity: An empirical study of mall X. *iBuss Management*, 3(2).

Türk Dil Kurumu, 2020, <https://sozluk.gov.tr/> (11.03.2020)

Twitter (2019) <https://twitter.com/> (20.11.2019)

Varnalı, K. (2013). Dijital Kabilelerin İzinde - Sosyal Medyada Netnografik Araştırmalar. İstanbul: MediaCat.

Vidya, N. A., Fanany, M. I. ve Budi, I. (2015). Twitter Sentiment To Analyze Net Brand Reputation of Mobile Phone Providers. *Procedia Computer Science*, 72, 519-526.

Walsh, G. ve Beatty, S. E. (2007). Customer-Based Corporate Reputation of a Service Firm: Scale Development And Validation, *Journal of the Academy Marketing Science*, 35, 127 – 143.

We Are Social (2020), <https://wearesocial.com/global-digital-report-2019> (03.03.2020).

Wettschereck, D. (1994). A Study of Distance-Based Machine Learning Algorithms.

Windows 10, 2019, <https://www.microsoft.com/tr-tr/windows/get-windows-10>, (10.09.2019)

Wu, H. C., Luk, R. W. P., Wong, K. F. ve Kwok, K. L. (2008). Interpreting Tf-Idf Term Weights As Making Relevance Decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3), 1-37.

Xu, Y. ve Chen, L. (2010). Term-Frequency Based Feature Selection Methods for Text Categorization. *In 2010 Fourth International Conference on Genetic and Evolutionary Computing* (s. 280-283). IEEE.

Yurtsever, M. ve Tecim, V. (2016). Mini Bilgisayarların Akıllı Kart Sistemleri İçin Tasarlanması ve Uygulanması: Üniversite Örneği. *Yönetim Bilişim Sistemleri Dergisi*. 2(1): 13-28.

Zheng, Z., Wu, X. ve Srihari, R. (2004). Feature selection for text categorization on imbalanced data. *ACM Sigkdd Explorations Newsletter*, 6(1), 80-89.

