



YÖNETİM BİLİŞİM SİSTEMLERİ DERGİSİ

<http://dergipark.gov.tr/ybs>

Yayın Geliş Tarihi: 07.09.2020
Yayına Kabul Tarihi: 12.11.2020
Online Yayın Tarihi: 27.12.2020

Cilt:6, Sayı:2, Yıl:2020, Sayfa:57-76
e-ISSN: 2630-550X

MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK SOSYAL MEDYADA MARKA İTİBAR ANALİZİ

Bahar ŞARDAĞI BOZYİĞİT ¹, Çiğdem TARHAN ¹

¹Yönetim Bilişim Sistemleri, İktisadi ve İdari Bilimler Fakültesi, Dokuz Eylül Üniversitesi, İzmir Türkiye

Özet

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 çalışma 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 paylaşım cevap verme, tekrar paylaşma, beğeni gibi bazı aksiyonlar alabilmektedir. İleriki v çalışmalarda sistemin birçok sosyal medya platformuna entegre ve daha kapsamlı olacak şekilde geliştirilmesi amaçlanmaktadır.

Anahtar Kelimeler: Makine Öğrenmesi, Metin Madenciliği, Marka İtibarı, Sosyal Medya, Bilgi Sistemleri

REPUTATION ANALYSIS OF BRAND IN SOCIAL MEDIA USING MACHINE LEARNING TECHNIQUES

Abstract

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 study, 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.

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 ettiği başarıları ile değil; bu platformda insanların marka hakkındaki yorum, paylaşma ve beğeni gibi etkileşimlerle nasıl görüldüğünü değerlendirebilmesinde önemli katkılar sunmaktadır (Şardağı ve Bayçu, 2018). 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 sunduğu 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. 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 olacaktır. 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.

LİTERATÜR TARAMASI

Bu çalışma 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 Naive 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 çalışma 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 önışleme 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 önışleme 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 Naive 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. Diğeri 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 diğeri algoritmalarından daha başarılı olarak %75.56 doğruluk skoru elde edilmiştir.

2018 yılında sunulan başka bir çalışmada (Hasan ve diğeri, 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.

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 diğeri, 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 diğeri, 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.

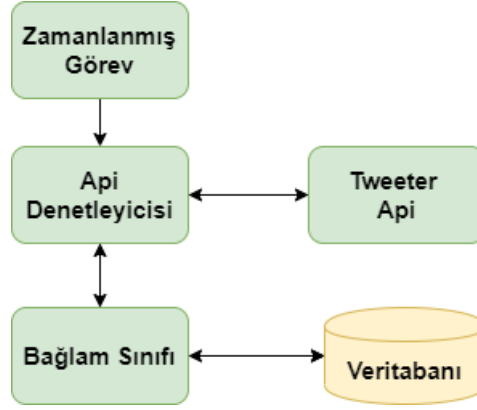
Sonuç olarak bu çalışma kapsamında diğeri çalışmalardan farklı olarak (markalar hakkında yapılan sosyal medya paylaşımlarının otomatik olarak sınıflandırması) yanı sıra kullanıcıların bu sonuçları inceleyebileceği bu doğrultuda aksiyon alabileceği bir bilgi sistemi sunulmuştur. Ayrıca, incelen Türkçe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

VERİ & YÖNTEM

Verilerin Toplanması: 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 mimarı yapısında geliştirilmiştir. İlgili yazılım mimarisinin şeması Şekil 1'de verilmektedir.

Şekil 1:Veri Toplama Uygulama Mimarisi



Zamanlanmış Görev “.Net Core Framework” (.Net Core, 2020) kullanılarak geliştirilen bir konsol uygulamasıdır. Bu uygulama, Windows 10 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 uygulama programlama 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 programlama ara yüzüne gönderir. Kimlik bilgilerinin eşlenmesinden sonra yaratılan örnekten belirlenen markalar için ilgili ara yüzden güncel paylaşımlar toplanmakta ve işlenmektedir.

Mimarinin son katmanı olan Bağlam Sınıfı Kütüphanesi “.Net Core Framework” ve “Entity Framework” kullanılarak geliştirilmiştir. Bağlam Sınıfı kütüphanesini kullanarak tanımlanan model sınıftan 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 yansira Twitter paylaşımları ve o paylaşımları yapan kullanıcıların bilgileri sırasıyla “Marka”, “Paylaşım” ve “Gönderen” varlıklarında saklanılmaktadır.

Marka varlığı ilgili kayıtlara dair yapılan paylaşımları arayabilmek için İsim, Hesap İsmi ve Sosyal Medya Etiketleri 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 paylaşımına 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.

Verilerin Etiketlenmesi: 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 yürütülebilmesi için kitle kaynak kullanımı 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.

Veri tabanında bir etiket kaydı bir paylaşım için bir kullanıcının yaptığı veri etiketini ifade etmektedir (Kılınç ve Teke, 2020). Bu web uygulamasına toplanan paylaşımları etiketleme 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 açıldığında, veri tabanında var olan ve kendisinin daha önce etiketlemediği rastgele paylaşımlar Şekil 2’deki gibi gösterilmektedir.

Şekil 2:Veri Etiketleme Uygulaması

Veri Etiketleme Web Anasayfa Hakkında İletişim Merhaba bahar.sardagi! Çıkış

Etiketlenecek Paylaşımlar

Gönderen: [Blurred Name]

[Blurred Name] uzak durun [Blurred Name]. Sakın bilet almayın zarar ziyan kulliyen. [Blurred Name] 7 ay önce 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

Bu web uygulaması sayesinde, kullanıcı gösterilen paylaşımın içeriğine bakarak ilgili paylaşımına 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.

Veri Önişleme: Metin madenciliğinde veri kümesinin etkin bir önişleme sürecinden geçmesi, makine öğrenmesi modellerinin başarısında oldukça önemli bir kıstastır. Ön işlemin 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 çıkarımının önüne geçilmesi amaçlanmıştır. İkinci adımda paylaşım metinleri içerisinde yer alan gereksiz sayılar kaldırılmıştır. Bunun sebebi ise paylaşım içerisinde yer alan sayıların bu çalışmadaki analizde 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 çalışmada makine öğrenmesi modellerinde sıklıkla tercih edilen Terim Frekansı ve Terim Frekansı-Ters Doküman Frekansı yaklaşımlarının ikisi de uygulanmıştır.

Özellik Seçimi: Ön işleme adımında elde edilen doküman terim matrisinin terim sayısıyla 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) kıstasına göre bir özellik alt kümesi oluşturma işlemidir. Bu çalışmada özellik seçimi için Ki-Kare (Chi-Square) testi ve Bilgi Kazanımı (Information Gain) yaklaşımları uygulanmıştır.

Makine Öğrenmesi: Makine öğrenimi (ML), deneyim yoluyla otomatik olarak öğrenen ve gelişen bilgisayar algoritmaları üzerine bir çalışma alanıdır (Dietterich, 1997; Aydın, 2018). 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 (Kılınç ve Aydın, 2019;).

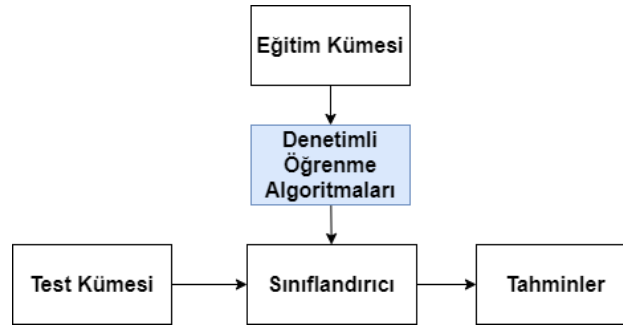
Makine öğrenimi yaklaşımları öğrenme aşamasındaki gösterdikleri farklılıklara dayanarak üç ayrı kategoriye ayrılmaktadır (Sasakawa ve diğerleri, 2008; Aydın, 2019). Bu kategoriler aşağıdaki gibidir.

- **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.

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.

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 3'te verilmektedir.

Şekil 3: Denetimli Makine Öğrenmesi Süreci



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 kullanarak 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 çalışmada öğrenme sürecinde uygulanan ve tahminleme başarısı hesaplanan denetimli makine öğrenmesi algoritmaları aşağıdaki gibidir.

Şekil 4: Uygulanan Makine Öğrenmesi Algoritmaları



Makine Öğrenmesi Algoritmalarının Değerlendirilmesi: 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 ($i=1,2,\dots,k$ 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 1'de verilmektedir.

Tablo 1: Ölçüt-Formül Tablosu

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

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 çalışmada bahsedilen üç ölçüt de makine öğrenmesi algoritmalarının değerlendirilmesinde kullanılmıştır.

DENEYSEL ÇALIŞMA

Makine Öğrenmesi Yöntemlerinin Uygulanması: 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 çalışmada 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 işlemine kullanılmak üzere seçilmiştir. İkinci yöntem olan bilgi kazanımında 15,973 özellik elenerek 891 özellik sınıflandırma işlemine kullanılmak üzere seçilmiştir.

Ön işleme ve özellik seçimi aşamaları sonucunda 4 ayrı veri kümesi varyantı oluşmuştur. Bu varyantlar Tablo 2'de verilmektedir.

Tablo 2: Veri Kümesi Varyantları

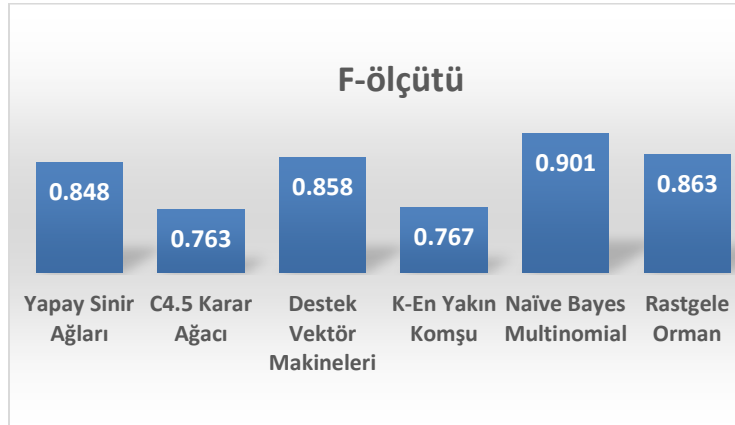
Varyant	Terim Ağırlığı	Özellik Seçimi
Varyant ₁	Terim Frekansı	Ki-kare
Varyant ₂	Terim Frekansı	Bilgi Kazanımı
Varyant ₃	TF-TDF	Ki-kare
Varyant ₄	TF-TDF	Bilgi Kazanımı

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 5'te verilmektedir.

Şekil 5: Test Sonuçları



Sonuçlar gözlemlendiğinde ön işleme adımlarında TF-TDF ve Ki-Kare yöntemlerinin kullanıldığı veri kümesinde Naive Bayes algoritmasıyla oluşan sınıflandırma modeli %90 civarı F-ölçüt 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 olmuştur.

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.

Marka İtibar Yönetim Sistemi: 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 çalışmada markaların sosyal medya üzerindeki itibarlarını ölçmeyi ve tüketici ile aralarındaki ilişkiyi yönetmeyi hedefleyen bir bilgi sistemi geliştirilmiştir.

Çalışma 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 çalışma 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 çalışma 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 sunmaktadır. Ayrıca sistem marka hakkında olan sosyal medya içerikleriyle ilgili bazı aksiyonlara olanak tanımaktadır.

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.

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

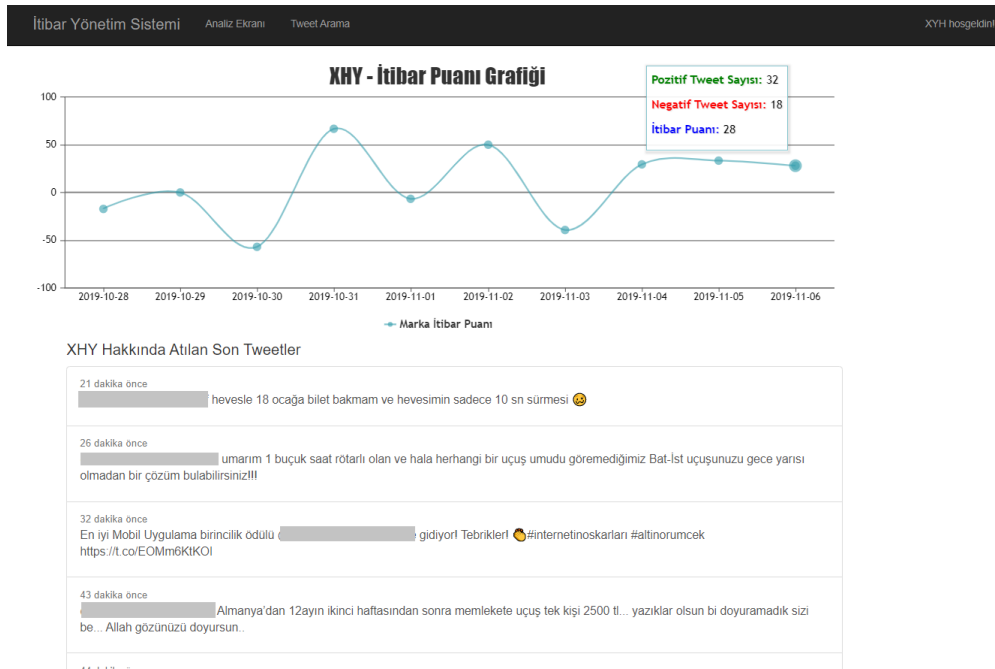
bilgilerine ulaşabilmektedir. Markanın itibar puanı bir gün içinde atılan pozitif paylaşım sayısının negatif paylaşım sayısından farkının toplam paylaşım sayısına oranıyla hesaplanmaktadır (Vidya ve diğerleri, 2015). Günlük İtibar Puanı (GİP) için hesaplama Formül 1’de gösterilmektedir.

Formül 1: 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 6’da İtibar Analiz Ekranı gösterilmektedir.

Şekil 6: İtibar Analiz Ekranı



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 6’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 7’de görüldüğü gibidir.

Şekil 7: Tweet Arama Sayfası

Tarih	Tweet	Favorilenme Sayısı	Retweet Sayısı
22.29.13	TK 2084 SAW - Ankara uçundaysız. 20.20 de havada olmamız gerekirken hala Bakan bekliyoruz Milletvekilleri , vatandaşlar herkes bakan bekliyor. Bakanlarımızı iyi bakmalıyız tabiki. Bu kafa ile nereye gideriz ?Zaman ayarsız Bakanlar, uçağa binince özur de dilemezler	34	0
08.42.06	Uçak içi eğlence sistemindeki filmlerde yer alan sevşme/öpişme 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 gerçekleştiği İstanbul-Viyana uçuşunu ilk kez TC-LLF tescilli B787 Dreamliner uçağı ile gerçekleştirdi. Viyana uçuşlarında dar gövde ve A330 tipi uçaklarını kullanmakta.	9	0
15.58.35	Hala havalimanında uçak içinde bekletiliyor. Bu gidile kaysenye gidemeyeceğiz (3	0
08.27.47	3 Kasım 2019, Taşkent - İstanbul uçuşu. Bunca kara propagandanadan sonra uçaklarına bu suyu koymaması daha mantıklı olurdu.	2	0
17.39.07	2 saat 45 dk rötar diyorsunuz del otobus tutan öyle gidelim Ankara İstanbul arası Almanyadan mı geliyor bu uçak? #şikayettar	2	0
00.05.44	Hizmet kaliteniz gün geçtikçe daha kötü oluyor bileti mi değiştirmek istiyorum sisteminizin çalışmadığını söylüyorsunuz, 3 ve 5 yaşlarındaki çocukları ebeveyninden ayrı farkı koltuklara veriyorsunuz,star alanız üyesi olduğunuzdan eminimsiniz?	1	0
00.42.30	Trafik sebebiyle uçağı kaçırdım bir sonraki uçuş için bilet ayarlamaya çalışırken calcenter IbrahimBey'in inanılmaz özveri konusurken kesilen telefona rağmen yardım için anında tarafımıza geri dönüşü tamamen isini #çoksevmek ve %100 insanlığı ile ilgili.Benden tam puan	1	1

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 8’de gösterilmiştir.

Şekil 8: Kelime Bazlı Arama

Tarih	Tweet	Favorilenme Sayısı	Retweet Sayısı
00:23:32	Türkiye'nin en değerli markası 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 şimdiki	0	0

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

Önceki 1 Sonraki

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

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 9’da Tweet Detay Bilgileri ve Tweet detay bilgilerini içeren görsel yer almaktadır.

Şekil 9:Tweet Detay Bilgileri Ekranı / Tweet Detay Bilgileri Sekmesi

İtibar Yönetim Sistemi Analiz Ekranı Tweet Arama XYH hoşgeldin!

Tweet Detay Bilgileri

büyük bir rezillliği yaşadık. 40 kişi gidış - dönüş bilet aldık. Gittik ama dönemedik. Asla [redacted] 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: [redacted]

Tekrar Gönderi: Evet

Kaynak: https://twitter.com/sercin67/status/1189132685446434817

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 10’da verilmiştir.

Şekil 10: 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 16:26:55
Takipçi Sayısı:	368
Gönderi Sayısı:	216
Kayıtlı Bölge:	Kdz.Ereğli
Kaynak:	sercin67

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 çalışma 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 paylaşan 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 kullanarak insan gücü ve zaman tasarrufu yapılmaktadır. Çalışma 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.

KAYNAKLAR

.Net Core Framework,2019, <https://www.microsoft.com/tr/download/details.aspx?id=17851> (01.10.2019)

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

Aydın, C. (2018). Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (14), 169-175. DOI: 10.31590/ejosat.458613

Aydın, C. (2019). Classification of the Fire Station Requirement with Using Machine Learning Algorithms. *I.J. Information Technology and Computer Science*, 1, 24-30. DOI: 10.5815/ijitcs.2019.01.03.

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

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

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

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

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

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

Kılınç, M. ve Aydın, C. (2019). Web Tabanlı İş Analitiğinin, İşletmelerdeki Kısa ve Orta Vadede Karar Verme Mekanizmasına Olan Etkisinin Araştırılması. *Yönetim Bilişim Sistemleri Dergisi*, Cilt:5, Sayı:1, Yıl:2019, Sayfa:64-85. ISSN: 2148-3752

Kılınç, M. ve Teke, İ. (2020). Veri Görselleştirmenin Bilgi Sistemlerinde Kullanımı: Web Tabanlı Mezun Bilgi Sistemi Örneği. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 18 (1); 96-109. ISSN: 1304-4796.

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

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

Leff, A., & 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* (pp. 118-127).

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

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., & Tuna, S. (2015). Sosyal Medya Verileri Üzerinde Yapay Öğrenme ile Duygu Analizi Çalışması. Technical Report.

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

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

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

Prihono, O. F., & 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.

Rodriguez, J. D., Perez, A., & 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

Sasakawa, T., Hu, J., & Hirasawa, K. (2008). A Brainlike Learning System with Supervised, Unsupervised, and Reinforcement Learning. *Electrical Engineering in Japan*, 162(1), 32-39.

Spyder (2020) <https://www.spyder-ide.org/> (12.02.2020)

Şardağı, E. ve Bayçu, S. (2018). İtibar Oluşumunda Medyada Görünürlük: Bankalar Üzerine Bir Araştırma. *Erciyes İletişim Dergisi*. 5. 108-122.

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

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

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

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.