

Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi

Araştırma Makalesi/Research Article

Burcu KARAÖZ AKIN^{1*}, U. Tuğba GÜR SOY ŞİMŞEK²

¹ Kar Maden Ltd. Şti., Muğla, Türkiye

² Sayısal Yöntemler Anabilim Dalı, İşletme Fakültesi, İstanbul Üniversitesi, İstanbul, Türkiye

burcukaraoz@gmail.com, tugbasim@istabul.edu.tr

(Geliş/Received:09.10.2017; Kabul/Accepted:26.05.2018)

DOI: 10.17671/gazibtd.342419

Özet— Teknolojik ilerlemeler bilinen anlamdaki veri kavramını geliştirmekte ve çeşitli tipte, büyük boyutlarda veri setlerini ulaşılabilir kılmaktadır. Özellikle kişisel cihazların ve sosyal medya kanallarının yaygınlaşması ile bireyler yaşamları ve duyguları hakkında çeşitli bilgiler paylaşmaktadırlar. Bu da araştırmacılara anket gibi tekniklerle elde edilmesi güç boyutlarda veri sağlayan kaynaklar oluşmasını sağlamaktadır. Metin verileri de hem sosyal medya aracılığı ile en çok biriken veri çeşidinden biri olması hem de içerdiği duygular bakımından önem taşımaktadır. Kişilerin duygularını ifade etmek veya görüşlerini bildirmek amacıyla yazdıkları metinlerin analiz edilerek anlamlı bilgilere dönüştürülmesi ve değer yaratılması veri analitiğinin amaçlarından birisidir. Bu doğrultuda veri setinin analizi kadar çıktıların yorumlanarak değer yaratılması da büyük önem taşımaktadır. Bu amaçla çalışmada çıktıları yorumlayabilecek fakat yazılım altyapısı güçlü olmayan araştırmacıların da kolaylıkla faydalanabileceği bir model önerilmiştir. Böylece çıktıların değer bulacağı işletme, ekonomi ve sosyoloji gibi sosyal bilim alanlarından araştırmacıların gelişen teknoloji ve veri setlerini değerlendirmesi hedeflenmiştir. Yarı denetimli öğrenme ile pozitif ve negatif duygu sözlükleri genişletilmiş sözlük temelli duygu analizi tekniği kullanılan ve performansı yüksek, uygulama anlamında çeşitli düzeyde yazılım bilgisine sahip kişilere hitap eden bir duygu analizi modeli sunulmuştur.

Anahtar Kelimeler— duygu analizi, metin madenciliği, sosyal medya analitiği

Adaptive Learning Lexicon Based Sentiment Analysis Proposal

Abstract— Technological improvements cause to enlarge data sources, types and volume. Especially personnel devices and social media, which are valuable external resources, provides large datasets that is inaccessible to collect with survey or other traditional techniques. Mostly acquired type, Text data includes mysterious information about people's opinion or idea. Hence, gathering, analyzing the text data and interpreting the results creates significant value for the researchers from such as business, economy and sociology. Proposed algorithm assumes that each dataset needs specific opinion lexicons, so semi-supervised sentiment analysis algorithm is generated. Another feature of the proposed algorithm is simplicity. Without an expertise on programming, this algorithm can be applied for the social sciences researchers. Thus, the study introduces a model that makes sentiment analysis possible for lots of researcher who have little expertise on programming but can interpret the results and creates value.

Keywords— sentiment analysis, text mining, social media analytics

1. GİRİŞ (INTRODUCTION)

Teknolojinin gündelik hayatın önemli bir parçası haline gelmesi bilinen anlamdaki veri kavramının değişmesine neden olmuştur. Paylaşılan fotoğraflar, yazılan iletiler, konum bilgileri, ziyaret edilen web sayfaları gibi kişisel cihazlarla yapılan tüm aktiviteler kayıt altında tutulabilmektedir. Bu da veri boyutlarında büyümeye, yapısal olmayan ya da yarı yapısal formatta çeşitliliğe sahip veri tiplerinin ortaya çıkmasına, verinin oluşma ve analiz hızında artışa dolayısı ile değer yaratılmasına imkân

sunmakta yani büyük veri kavramını ortaya çıkarmaktadır. Sosyal paylaşım sitelerine yazılan iletiler, kullanıcı görüş veya şikâyetleri, bloglar, haber siteleri vb. birçok sosyal medya mecrasına her gün girilen milyonlarca ileti büyük hacimli veri setleri oluşturmaktadır. Bu durum birçok konuda sosyal medyadan elde edilen görüşlerin metin madenciliği teknikleri ile analizi edilebilmesi sayesinde, binlerce kişiye anket yapılma gereği duyulmadan önemli bulgulara ulaşmayı mümkün kılmaktadır. Metinlerden anlamlı bilgiler elde edilebilmesi için kullanılan yöntemlerden bazıları; konu modelleme, metin

sınıflandırma, duygu analizi, kelime bulutları veya birliktelik analizidir [1]. Konu modelleme, sosyal medya platformlarından alınan büyük ölçekli metin veri setinden en baskın konuları ayıklamayı hedeflemektedir. Duygu analizi, kişilerin yazdıkları metinlerden bahsettikleri konu hakkındaki duygularının olumlu mu, olumsuz mu yoksa nötr mü olduğunun anlaşılmasına çalışılmasıdır. Kelime bulutları sayesinde metin içerisinde geçen kelimeler, geçme sıklıklarına göre görselleştirilebilmektedir. Birliktelik analizi ile metinlerde hangi kelimelerin beraber kullanıldıklarını göstererek yorumlanmasını sağlamaktadır. Bu yöntemler sonuçları ile sağladıkları değerin yanı sıra veriyi yapısal hale dönüştürüp başka analitik çalışmalarda girdi olarak kullanılmasına da imkân vermektedir. Böylece müşterilerin yazdıkları iletiler işletme için daha kolay anlaşılır ve kullanılabilir hale gelebilmektedir. Bu analizler işletmelerin yapacağı yatırımların, pazarlama stratejilerinin, geliştireceği ürünlerin belirlenmesi aşamasında karar desteği sunmaktadır. Bu sayede müşterilerin sosyal medya da beğeni ile bahsettikleri ürünlerin piyasadan kaldırılması, belirli bir ürün hakkında sosyal medyadan yapılan şikâyetlerin göz ardı edilmesi, ürün ile ilgili geliştirilmesi/eklenmesi gereken özellikleri ileten müşteri görüşlerinin kaçırılması gibi sosyal medya verilerinin analiz edilmemesinden kaynaklı hatalar önenebilmektedir.

Metinlerden yazan kişinin hissettiklerine dair bilgi vermeyi amaçlayan duygu analiz tekniklerinin yoğun yazılım bilgisine ihtiyaç duyduğu algısı nedeniyle birçok sosyal bilimci, işletme araştırmacıları ilgili analizleri kullanmamaktadır. Metin verisinin analiz edilmesi büyük ölçüde doğal dil işleme alanının çalışma konusu olmasına karşın istatistiksel olarak anlamlı performans değerlerinin elde edilebildiği uygulaması daha basit olan algoritmalar, her alandan birçok kullanıcıya duygu analizi uygulama imkânı vermesi nedeniyle gereklidir. Öte yandan mevcut duygu sözlüklerine ek olarak veri setine has duygu ifadelerinin de sözlüklere dâhil edilmesi duygu analizinin çeşitli alanlarda yüksek performans ile sonuç vermesini sağlayabilmektedir. Bu amaçla çalışmada, Türkçe metinlerin yarı denetimli olarak bir kısmının etiketlenmesi ve bu etiketlenen metinlere göre duygu sözlüklerinin oluşturulduğu, R ve Excel-VBA gibi kolay ulaşılabilir ve kullanılabilen programlar ile geliştirildiği sözlük temelli duygu analizi algoritması sunulmaktadır.

2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Çalışmanın da veri kaynağı olan sosyal medya; kullanıcıların ürünler hakkında yorum ve şikâyetlerini paylaşabildikleri, politik görüşlerini ortaya koyabildikleri, kişisel görüş, beğeni ya da isteklerini belirtebildikleri çeşitli platformlar içermektedir. Üye sayısı ve günlük ortalama ileti sayısı bakımından ön planla olan sosyal medya platformlarından birisi Twitter'dır. Kişilerin 140 karakter ile görüşlerini paylaşabilmelerine imkân sağlayan Twitter araştırmacılar için de önemli bir veri kaynağı haline gelmektedir. Twitter'dan elde edilen kullanıcı iletilerine duygu analiz teknikleri uygulanarak kişilerin iletiyi yazarken hissettikleri duygunun belirlenmesini

amacıyla çalışmalar yapılmaktadır. Metinlerdeki duygunun ortaya çıkarılmasına çalışılan duygu analizi çalışmalarının bazıları sadece pozitif ve negatif duyguları belirlemeye çalışırken bazıları 0 ile 10 arasında bir skalaya (0-çok negatif, 10-çok pozitif) göre duyguları sınıflandırmışlardır [2,3]. Öte yandan hislerin ve bunların gücünün belirlenmeye çalışıldığı araştırmalar da yapılmaktadır [4]. Literatürde yer alan Twitter üzerinden derlenen verilerin çeşitli sektörlerdeki duygu analizi uygulamalarından bazıları aşağıda sunulmuştur. Bunun yanı sıra bu çalışmada da bir önerisi bulunan Türkçe veri setleri ile yapılmış duygu analizi çalışmalarına da yer verilmiştir.

Pazarlama sektöründe, müşterilerin firmalar hakkında görüşlerinin analizi amacıyla Twitter ve duygu analizi tekniklerinin beraber kullanıldığı çeşitli çalışmalar bulunmaktadır. He ve arkadaşları [5] bir sektörden belirledikleri üç firmanın Facebook ve Twitter üzerinden derledikleri veriler ile müşteri memnuniyetini belirlemeye çalışmışlardır. Bu amaçla takipçi, yorum ve paylaşım sayıları gibi sayısal değerlerin yanı sıra iletiler de metin madenciliği teknikleri kullanılarak değerlendirilmiştir. Mostafa [6], birçok firma için Twitter paylaşımlarında ilgili firma adlarının yanında en sık kullanılan kelimeleri belirlemeye çalışmıştır. Bu amaçla hava yolu, telekomünikasyon ve sağlık sektörü gibi çeşitli sektörlerden firmaları analiz etmiştir. Ghiassi [7] ise Twitter'dan bir marka hakkında müşteri yorumlarını derlemiş ve buna literatürdeki modellerden farklı olarak geliştirdikleri algoritmalar ile duygu analizi yapmışlardır.

Twitter verilerini finans alanında kullanan Zhang ve arkadaşları [8] günlük olarak tweetlerin korku mu umut mu içerdiklerini ölçüp bu değerlerin borsa göstergeleri ile korelasyonlarını analiz etmişlerdir. Buna göre ertesini gün için borsanın yönünü tahminleme imkânı olan bir model geliştirmişlerdir. Bollen ve arkadaşları [9] ise Twitter'dan derledikleri verileri ruh hali kategorilerine göre sınıflandırıp bunların DJIA kapanış fiyatlarının tahmininde kullanılmasına yönelik bir çalışma yapmışlardır. Günlük kapanış fiyatlarının artış azalış değişimlerinin tahmininde %86,7 oranında doğruluk elde etmişlerdir. Ardından Mittal ve Goel [10] de Twitter üzerinden halkın ruh hali skorunu belirleyip bunu borsadaki hareketin tahmininde kullanmıştır. Bu amaçla kullandıkları algoritma %75,56 doğruluk oranını sahiptir ve karşılaştırma yaptıkları Bollen ve arkadaşlarının [9] çalışmasından daha düşük bir oran elde etmişlerdir. Benzer şekilde Bouktif ve Awad [11] da Twitter'dan derledikleri veriler ile hisse senetleri kapanış fiyatları tahminlemeye çalışmışlardır. Geçmiş çalışmalar ile kıyaslayabilmek için yazarlar çalışmalarında karınca kolonisi optimizasyon algoritması kullanmışlardır. Porshnev ve arkadaşları [12] da borsa tahminlerinin gücünü arttırabilmek için Twitter mesajlarının duygu analizini kullandıkları çalışmalarında 755 milyon tweet kullanmışlardır. Destek vektör makinaları ve sinir ağları algoritmaları ile DJIA ve S&P500 hakkında tahmin yapmaya çalışmışlardır. Bunlara ek olarak Corea [13] yatırımcıların duygularının borsayı etkileyip etkilemediğini inceleyen çalışmada yatırımcıların

Twitter'a yazdıkları metinlere duygu analizi yapmıştır. Fakat tahminlemede duyguların değil atılan tweet sayılarının daha etkili olduğu sonucuna varmıştır.

Sinema filmlerinin gişe getirilerinin ya da izlenmelerinin tahmininde de Twitter ve duygu analizi beraber kullanılabilir. Asur ve Huberman [14] Twitter'dan belirli bir konu hakkında veri derlenip geliştirilen modelle piyasa için tahminlemeler yapılabileceğini öngörmüşlerdir. Bu amaçla geliştirdikleri doğrusal regresyon modeli ile Twitterdan sinema filmleri hakkında derledikleri veriler ile gişe kazancı değerlerinin arasındaki ilişkiyi ölçmüş ve gelecek tahmini yapmışlardır. Ayrıca tweetlere duygu analizi yaparak ilgili modeli geliştirmeye çalışmışlardır. Rui ve diğerleri [15], Twitter gibi sosyal medya kanalları aracılığı ile yapılan ağızdan ağıza iletişimin ürün satışları üzerinde etkiye sahip olduğunu ve bunun yönetilerek avantaja çevrilebileceğini, filmlerin gişe rakamları üzerindeki etkisi üzerinden ifade etmişlerdir. Kim ve diğerleri [16], Twitter ve Facebook sosyal medya sitelerinden filmler hakkında yapılan yorum sayılarını, haftalık eğilimlerini ve kaç salonda yayınlandığı gibi filmin yayınlanmasına ilişkin girdileri kullanarak gişe rakamını tahminlemeyi hedeflenmişlerdir.

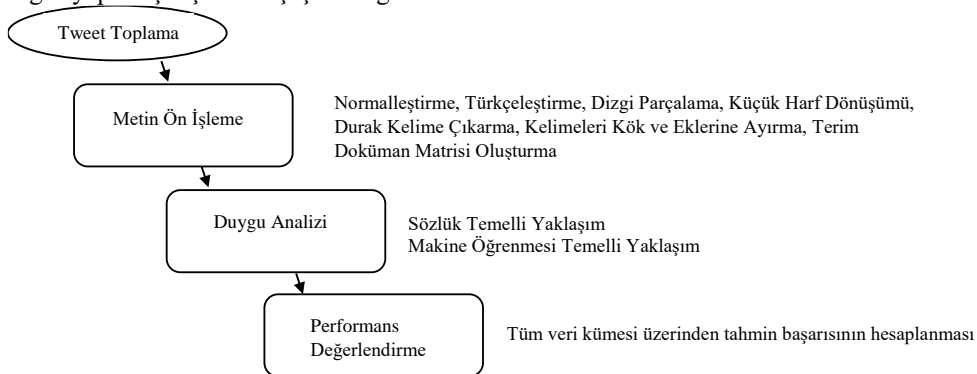
Sağlık alanında da sosyal medya analizinden yararlanılarak yapılan pek çok çalışma bulunmaktadır. Culotta [17] salgınların izlenmesi ve tahminlenmesi için Twitter'ı kullanmıştır. Twitter'dan grip ile ilgili mesajları derleyen yazar, çeşitli algoritmalar kullanarak bu mesajların "Hastalık Kontrol Merkezi" istatistik değerleri ile ilişkisini incelemiştir. Aramaki [18] 'de benzer bir amaçla grip ile ilgili tweetlerin pozitif ya da negatif olarak sınıflandırılmasında çeşitli makine öğrenmesi tekniklerinin performanslarını karşılaştırmıştır. Bodnar ve Salathe [19] ise grip ile ilişkili sayılabilecek en az bir kelime geçen tweetler üzerinde regresyon çalışması yapmıştır.

Twitter'dan Türkçe tweetlerin derlenmesi ile yapılan duygu analizi çalışmalarından bazıları şöyledir. Meral ve Diri [20] ile Çoban ve arkadaşları [21] Twitter'dan derledikleri Türkçe tweetler üzerinde duygu analizi çalışması yapmışlardır. Türkçe'nin sondan eklemeli bir dil olması nedeniyle duygu analizinde dikkat edilmesi gereken konulara vurgu yapan çalışmalar çeşitli algoritmaların

sınıflandırma performanslarını karşılaştırmışlardır. Eliaçık ve Erdoğan [22]'in çalışmasında geliştirdikleri yeni bir duygu analiz yöntemi ile finans topluluklarının Twitter'dan topladıkları verisini analiz etmişlerdir. Elde edilen duygu polarite değerleri ile Borsa İstanbul'un haftalık değerleri arasındaki korelasyonu incelenmiştir. Diri ve arkadaşları [23,24] seçilen Türkçe TV programları hakkında Twitter'dan veri derlemişlerdir. Bu metinlere çeşitli algoritmalar ile duygu analizi yaparak belirli bir skor atanmış ve bu skorlar ile gerçek reytingler karşılaştırılmıştır. Diri ve arkadaşlarının 2013 yılında yaptıkları çalışmada Twitter da paylaşılan mesajların 140 karakter ile sınırlı olmasından dolayı kişilerin kısaltma kullanarak yazdığı kelimelerin düzeltilmesinin gerekliliğine vurgu yapmışlardır [24]. Öte yandan diğer problemler olarak, tweet atan kesimin genellikle genç ve orta yaşlı olması, her program için aynı sayıda tweet atılmamış olması ve bazı sosyoekonomik sınıfların Twitter'da temsil edilmemesini belirtmişlerdir. Çalışmanın sonucunda sadece tweet sayılarının tek başına reyting belirlemede yeterli olmadığını bunun için duygu analizi sonuçlarının da kullanılarak geliştirilen bir reyting sıralaması yapıldığını ve bu sıralama ile gerçek reyting sıralamasının birbirine yakın olduğunu vurgulamışlardır. Akarsu ve Diri'nin 2016 yılında yaptıkları çalışmada ise Twitter'dan çekilen verilerin temizlenmesi ve düzenlenmesi için makine öğrenmesi algoritmalarının kullanılması üzerine yoğunlaşmışlardır [23]. Çeşitli algoritmaları çeşitli program türlerinin reytinglerinin tahminlemede kullanarak performanslarını karşılaştırmışlardır.

3. DUYGU ANALİZİ (SENTIMENT ANALYSIS)

Duygu analizi genel anlamıyla, kişilerin yazdıkları metinlerden görüşlerinin, tutumlarının, yaklaşımlarının veya duygularının belirlenmeye çalışılmasıdır [2,25]. Artan sosyal medya platformları ile kullanıcıların paylaştıkları subjektif metinler de artmıştır, bu da ilgili metinlerin analiz edilip anlam çıkarılması şeklinde bir araştırma alanı ortaya çıkarmıştır [26]. Duygu analizi literatürde, fikir madenciliği, duygu durum analizi, görüş madenciliği, duygu sınıflandırma gibi farklı isimlerle de yer almaktadır.



Şekil 1. Duygu analiz modelinin genel yapısı
(Stages of sentiment analysis)

Duygu analizi teknik yönleriyle mühendislik bilimlerinde, çıkarılan bilgilerin önemiyle de sosyal bilimlerde geniş çalışma alanı bulmaktadır. Mühendislik alanındaki çalışmalar genellikle doğal dil işleme ve daha iyi performanslı tahmin modelleri geliştirmeye yönelikken, sosyal bilimlerde duygu analizi çıktılarının işletmeler, toplum ve ekonomi için anlamlı bulgulara dönüştürülmesi hedeflenmektedir. Duygu analizi çalışmalarında ihtiyaç duyulan veriler, mikrobloglarda ifade edilen görüşler, ürün/film/otel/restoranlar hakkında yazılan yorumlar, Twitter ve Facebook gibi platformlarda paylaşılan düşünceler, haber makaleleri gibi çeşitli mecralardan elde edilebilmektedir. Sosyal medya platformlarından elde edilen büyük hacimli metin verileri duygu analizi sayesinde, toplumun mutluluk seviyesi, müşterilerin ürünler ve markalar hakkındaki görüşleri, halkın politikacılar veya hükümet hakkındaki görüşleri gibi çeşitli alanlardaki çalışmalarda kullanılan önemli değişkenlere dönüşmektedir [8,9,27]. Duygu analizinde kullanılacak metinler; doküman, cümle ve görüş seviyesinde olabildiğinden bir metnin tamamının ya da sadece bir cümlenin analizi mümkündür [28].

Kullanıcıların metinleri yazarken belirtmek istedikleri duyguların çıkarılmasını amaçlayan duygu analizinde, kullanılan teknikler genel hatlarıyla sözlük temelli ve makine öğrenmesi temelli olmak üzere iki ayrılmıştır [29]. Makine öğrenmesi temelli analiz tekniklerinde denetimli öğrenme modelleri kurularak yüksek performans sağlamaktadır. Fakat yüksek model kurma ve etiketleme maliyetleri ile karşı karşıya kalılabilmektedir. Sözlük temelli modeller ise uygulaması ve anlaşılması çok kolay olmasına rağmen daha düşük performans ve her yeni çalışmada uzun ön hazırlık süresi gibi dezavantajlara sahiptir [30]. Her iki tekniğin de uygulanabilmesi için öncesinde verilerin analize uygun hale getirildiği ön işleme sürecinden geçirilmesi gerekmektedir (Şekil 1). Metinlerdeki anlamsız kelimelerin ve sembollerin temizlenmesi metnin Türkçeleştirilmesi gibi çeşitli adımların izlenmesi ile veri seti uygulama için hazır hale getirilebilmektedir.

3.1 Metin Ön İşleme (Text Preprocessing)

Kullanıcıların sosyal medya platformlarında günlük konuşma dili ile ileti girmesi, Twitter iletilerinde 140 karakter sınırlaması olması gibi nedenler metinlerin analiz öncesi düzenlenmesini gerekli kılmaktadır. Metinlerin analize hazır hale getirilmesi için yapılan ön işlemler aşağıda açıklanmaktadır.

- Normalleştirme; “Çooooo güzel” örneğindeki gibi kelimelerin uzatılarak yazılması kelimelerin normalleştirilmesi ihtiyacını doğurmuştur. Öte yandan karakter sınırı nedeniyle izliyorum yerine izlyrm şeklinde yazılan kelimelerin de var olması analizin kalitesini düşürmemek adına önceden tespit edilip düzeltilmesi gereken unsurlardan birisidir.

- Türkçeleştirme; (“ç” , ”s” , ”ğ” , ”ı” ,”ö” , ”ü”) gibi Türkçe karakterlerin kullanılmaması “çok güzel” yerine “cok guzel” yazılmasına neden olmaktadır ve bunların Türkçe formata getirilmesi gerekmektedir. Böylece çok ve cok kelimesi iki ayrı kelime gibi algılanmayıp tek kelimeye dönüştürülmektedir.
- Dizgi parçalama: bir metinde yer alan tüm kelimelerin ayrı ayrı nesnelere haline getirilmesi işlemidir. Böylece “Havalar çok ısındı” cümlesi tek bir cümle olmaktan çıkıp “havalar”, “çok”, “ısındı” şeklinde üç ayrı özniteliğe dönüştürülmektedir.
- Küçük harf dönüşümü; tüm kelimelerin aynı formatta olmaları için yapılan işlemidir.
- Durak kelimelerin çıkarılması; duygu analizinde ihtiyaç duyulmayan, herhangi bir duygu anlamı içermeyen kelimelerin veri setinden temizlenmesi amacıyla yapılır. Bu sayede veri setindeki nesne sayısının azaltılması hedeflenir. Örneğin ben, sen, o gibi zamirler; ama, fakat, çünkü gibi bağlaçlar durak kelimeler arasında yer almaktadır.
- Kelimeleri kök ve eklerine ayırma; nesnelere boyutunun düşürülmesi ve fiillerdeki zaman ekleri, iyelik ekleri gibi duygu ifade etmeyen unsurların çıkarılması amacıyla yapılan işlemidir.
- Terim doküman matrisi oluşturma; her kelimenin metin ve tüm veri seti içerisinde kullanılma sıklığına göre terim frekans ağırlıklarının belirlenmesi ile oluşturulan matristir. Yani bir kelimenin tüm veri setindeki kelimelere oranla kullanılma sıklığı ile o kelimenin önemi belirlenebilmektedir. Bunun sağlanabilmesi için terim doküman matrisi oluşturulurken çeşitli normalleştirme teknikleri kullanılmaktadır [3], [31]. Aşağıda açıklanmış olan bu tekniklerde kullanılan formüllerde wf ilgili terimi/kelimeyi ifade etmektedir.

- Log frekansı: bulunan frekansların aşağıdaki log fonksiyonu ile normalleştirilmesidir (Eş.1). (wf-ham veri)

$$f(wf) = 1 + \log(wf) \quad \forall wf > 0 \quad (1)$$

- İkili frekans: terimin ilgili metinde kullanılıyorsa 1 kullanılmıyorsa 0 değerini alması ile yapılan normalleştirilmedir (Eş.2).

$$f(wf) = 1 \quad \forall wf > 0 \quad (2)$$

- Ters doküman frekansı (IDF): terimlerin metinler içerisindeki etkisinin belirlenmesi için veri setindeki toplam metin sayısının, kelimenin metinde geçme sayısına bölünmesinin logaritması ile elde edilen normalleştirilmedir. Frekans ile terimlerin ters doküman frekansları arasında ters orantı vardır (Eş.3).

$$IDF = \log \frac{\text{veri setindeki tüm metin sayısı}}{\text{kelimenin metinde geçme sayısı}} \quad (3)$$

- Terim frekansı- ters doküman matrisi (TF-IDF): log frekansı ve ters terim frekansı normalleştirme ile sentezlenmesi ile oluşturulmuş bir normalleştirme tekniğidir. Terimlerin geçtiği metin sayısının da önem kazandığı bir yöntemdir (Eş.4).

$$TD - IDF = \begin{cases} 0 & \text{if } wf = 0 \\ 1 + \log(wf) * IDF & \text{if } wf \geq 1 \end{cases} \quad (4)$$

Metin ön işleme adımları ile Türkçeleştirilmiş, kelimelerin normal formatlarında yazılı olduğu, veri setinde duygu ifade etmeyen kelimelerin yer almadığı metin veri seti oluşturulmuştur. Ardından her kelimenin ayrı bir nesne olarak düzenlendiği, en üst satırında kelimelerin, sol baş sütununda ise metin bilgilerinin yer aldığı, içerisinde de her kelimenin tüm veri seti içerisinde geçme sıklığının yer aldığı matrisin oluşturulması ile analize hazır veri seti oluşturulmuştur.

3.2 Duygu Analizi Teknikleri (Sentiment Analysis Techniques)

3.2.1. Sözlük Temelli Yaklaşımlar (Dictionary based approach)

Sözlük temelli duygu analizi yaklaşımlarında kelimeler ya da kelime gruplarının (n-grams) içerdikleri duyguya göre (pozitif/negatif) sınıflandırılması ile duygu sözlükleri oluşturulmaktadır. Bu sözlüklere göre metinlerin duygu durumları belirlenebilmektedir. Her metnin ne kadar pozitif ve negatif sözlük içerdiği incelenip, yüksek sayıda olan sınıfa atanması ile duygusu ortaya çıkarılmaktadır. Buna göre incelenecek olan bir metin daha çok pozitif terim içeriyorsa duygusu pozitif, daha çok negatif terim içeriyorsa negatif ikisi eşit ya da hiçbirini içermiyorsa nötr olarak sınıflandırılmaktadır.

Duygu sözlüklerinin oluşturulmasında manuel olarak duygu sözlüğü oluşturma, yazılı sözlüklerden yararlanarak duygu sözlüğü oluşturma ve bütüncü temelli duygu sözlüğü oluşturma gibi yöntemler kullanılmaktadır [32].

3.3 Makine Öğrenmesi Temelli Yaklaşımlar (Machine learning based approach)

Makine öğrenmesi temelli yaklaşımlar veri setindeki tüm kelimelerin ayrı ayrı birer öznitelik olarak değerlendirildiği kelime çantası (Bag of Words) tekniği ile sınıflandırma algoritmalarının kullanılmasıdır. Her kelimenin birer özellik olarak Karar Ağacı, Destek Vektör Makineleri, Yapay Sinir Ağları ve Naive Bayes gibi sınıflandırma algoritmalarında girdi olarak kullanılması ile metinlerin sınıflandırılarak duygularının tahmin edilmesi hedeflenmektedir.

Duyguların tahmin edilebilmesi amacıyla kurulan sınıflandırma modellerinde girdi olarak kullanılacak özniteliklerin belirlenmesi gerekmektedir. Bu amaçla metinler ön işlemeden geçirilerek işlevsiz / duygu ifade etmeyen kelimeler ve sembollerin çıkarılması sağlanır.

Temizlenmiş veri setindeki her kelimenin bir özniteliği temsil ettiği ve kelimelerin geçme sıklıklarının yer aldığı terim doküman matrisleri oluşturulur. Metin sayısı arttıkça terim doküman matrisinin boyutu da büyüyecektir. Bu nedenle Ki-kare bilgi kazanımı, terim frekansı ya da terim frekansı- ters terim frekansı değerleri kullanılarak temsil gücü yüksek özniteliklerin seçilmesi mümkündür [33]. Ardından denetimli öğrenme modeli için veri setinin bir kısmının etiketlenmesi ile eğitim seti oluşturulur ve sınıflandırma algoritmaları uygulanarak metinlerin duygu durumları belirlenebilir.

4. ÖNERİLEN DUYGU ANALİZ ALGORİTMASI (PROPOSED SENTIMENT ANALYSIS MODEL)

Kişiler, konuşurken ya da yazarken genel olarak iki şeyi amaçlarlar; gerçekleri belirtmek ya da görüşlerini bildirmek. Var olan herkesçe bilinen durumlar üzerine konuşulması ya da bu konular hakkında yazılması gerçekleri belirtmek amaçlıdır. Görüş bildirme amaçlı yapılan konuşmalar ya da yazılan iletiler ise kişiden kişiye değişen, kişinin o anki ruh halini, ilgili konu hakkındaki düşüncesini yansıtmaktadırlar. Duygu analizi görüş bildirilen bu iletilerin pozitif ya da negatif duygular ile ya da mutlu, üzgün, umutsuz vb. ruh halleriyle yazıldığı otomatik tespit edilmesini amaçlamaktadır.

Veri madenciliği ve metin madenciliği çalışmalarında sıklıkla tercih edilen R ve python gibi programlama dillerinde; yabancı diller için geliştirilmiş birçok duygu analizi kütüphanesi bulunmaktadır. Türkçe için geliştirilmiş böyle bir kütüphane ya da yazılım var olmamasına karşın bu alanda çok sayıda çalışma yapılmakta, çeşitli gruplarca yüksek performanslı doğal dil işleme ve duygu analizi modülleri geliştirilmeye çalışılmaktadır. Bu nedenle literatürdeki duygu analizi çalışmalarının çoğunluğu mühendislik temelli araştırmacılar tarafından yapılmaktadır.

Analizin gerektirdiği teknik alt yapıya karşın değer yaratılabilmesi için sonuçların yorumlanması ve mevcut durumun analizi ya da gelecek için plana dönüştürülmesi büyük önem taşımaktadır. Dolayısı ile sosyal bilimcilerin ya da işletmelerde çalışanların uygulama yapabileceği, sürecini ve sonuçlarını yorumlayabileceği duygu analizi algoritmaları gereklidir. Şekil 2'deki sözde kodda detaylıca açıklandığı gibi bu çalışmada sosyal bilimcilerin de duygu analizi tekniğini kullanabilmeleri için daha kolay erişilebilen platformlarda, göreceli olarak uygulaması daha basit bir duygu analizi algoritması sunulmaktadır. Bu amaçla, çalışmada Twitter'dan bir kanal ve o kanalda 8 aylık periyotta yayınlanan programlar hakkında yazılan iletiler toplanmıştır. Bu sayede binlerce kişiye anket yapılarak elde edilmesi çok güç olan anlık seyirci görüşleri sosyal medya sayesinde kısa bir sürede elde edilmiş ve bunlardan değer yaratılması hedeflenmiştir. Duygu analizi ile seyircilerin kanal ve program hakkında yazdıkları iletilerin olumlu, olumsuz ya da nötr olarak sınıflandırılan duygulardan hangisini içerdiği incelenmiştir. Böylece izleyicilerin ilgili kanalı ve programı izlerken olumlu mu olumsuz mu düşündükleri ortaya çıkarılabilecektir. Bu

bilgi ile kanal yöneticileri, program sorumluları gelecek için öngörülerde bulunabilecek, mevcut durumlarını analiz edip sorunları daha hızlı görüp çözme şansına sahip olabilecek veya gelecek için stratejiler belirleyebilecektir.

1. Başla
2. Tweet topla
3. Ön işlemeden geçir
4. İşlenmiş veri setindeki 2000 metni pozitif ve negatif olarak etiketle
5. Pozitif ve negatif etiketli veri setlerini ayır
 - 5.1. Pozitif veri setini için
 - 5.1.1. Terim frekans matrisini oluştur
 - 5.1.2. En sık kullanılan 1000 kelimeyi belirle
 - 5.1.3. Anlamli olan kelimeleri belirleyerek pozitif duygu sözlüğünü oluştur
 - 5.2. Negatif veri seti için
 - 5.2.1. Terim frekans matrisini oluştur
 - 5.2.2. En sık kullanılan 1000 kelimeyi belirle
 - 5.2.3. Anlamli olan kelimeleri belirleyerek negatif duygu sözlüğünü oluştur
6. Duygu analizinde kullanılacak pozitif ve negatif duygu sözlüklerini birleştir
7. Etiketlenmemiş veri setini aç
8. Her tweetin pozitif ve negatif sözlüklerinden içerdikleri kelimeleri belirle
9. Tweetlerin içerdikleri pozitif ve negatif kelime sayılarını karşılaştır
 - 9.1. Negatif kelime sayısı pozitif kelime sayısından büyük ise negatif duygu içeren tweet olarak etiketle
 - 9.2. Pozitif kelime sayısı negatif kelime sayısından büyük ise pozitif duygu içeren tweet olarak etiketle
 - 9.3. Pozitif kelime sayısı ile negatif kelime sayısı birbirine ya da sıfıra eşit ise nötr duygu içeren tweet olarak etiketle

10. Bitir

Şekil 2: Önerilen duygu analizi algoritmasının sözde kodu
(Pseudo code of proposed sentiment analysis algorithm)

Önerilen duygu analizi algoritmasının uygulamasında öncelikle geliştirilen yarı denetimli sözlük temelli duygu analiz algoritması ile 1.200.000 tweetin duygusu belirlenmeye çalışılmıştır. Sonrasında eğitim seti olarak

sözlük temelli yöntem ile belirlenen verilerin kullanıldığı bir makine öğrenmesi temelli duygu analizi modeli kurulmuş ve tahmin performansı incelenmiştir.

1. Adım

Sözlük temelli duygu analizi algoritmalarında kritik konulardan biri sözlük oluşturma aşamasıdır. Her veri seti kendine has jargonlara sahip olabilmektedir görüşünden yola çıkarak bu çalışmada yarı denetimli bir teknikle sözlük oluşturulmuştur. Bu amaçla 1000 adet tweet pozitif olarak 1000 adet tweet de negatif olarak etiketlenmiştir.

2. Adım

Sonrasında R programlama dili aracılığı ile veri seti metin ön işleme sürecine tabii tutulmuş; Türkçeleştirilme, küçük harfe çevirme, kelimeleri kök ve eklerine ayırma, dizgi parçalama durak kelimelerin çıkarılması ve terim frekans matrisi oluşturulmuştur.

3. Adım

Her iki veri seti için de ayrı ayrı yapılan bu işlemler sonucunda terim frekansları belirlenmiş ve;

- pozitif etiketli tweetlerde en sık geçen 1000 kelime
- negatif etiketli tweetlerde en sık geçen 1000 kelime

şeklinde iki farklı kelime seti elde edilmiştir. Böylece incelenen konunun kendine has pozitif ya da negatif duygu içeren sözcüklerinin kaçırılmaması sağlanmıştır.

4. Adım

Bu veri setleri içerisindeki anlamli kelimeler seçilerek “Pozitif Duygu Sözlüğü” ve “Negatif Duygu Sözlüğü” oluşturulmuştur.

5. Adım

Ardından Excel kullanılarak veri setindeki tüm iletilerin pozitif duygu sözlüğünden ve negatif duygu sözlüğünden kaçır kelime içerdikleri belirlenmiş ve sayıca yüksek olan duyguya etiketlenmiştir.

Duygu sözlüklerinin ve veri setinin küçük örnekleri üzerinden algoritmanın işleyişi ve adımları Tablo 1-4’de gösterilmektedir.

Tablo 1: Duygu sözlükleri örneği
(Lexicon examples)

Pozitif duygu sözlüğü	güzel, mükemmel, iyi
Negatif duygu sözlüğü	kötü, çirkin ...

Tablo 2: Pozitif duygu sözlüğü kontrolü ve skor belirleme-Adım 1
(Checking positive lexicon and defining the score- Step 1)

	güzel	mükemmel	iyi	...	Skor
Bu program çok iyi, sunucu da güzel	1	0	1		2
Ne kadar çirkin dizi	0	0	0		0
Bu akşam benim dizim var	0	0	0		0

Tablo 3: Negatif duygu sözlüğü kontrolü ve skor belirleme-Adım 2
(Checking negative lexicon and defining the score-Step 2)

	kötü	çirkin	...	Skor
Bu program çok iyi, sunucu da güzel	0	0		0
Ne kadar çirkin dizi	0	1		1
Bu akşam benim dizim var	0	0		0

Tablo 4: Duygu durumu belirleme için skor karşılaştırma- Adım 3
(Comparing the scores to assign the sentiment- Step 3)

	Poz. sözlük skoru	Neg.sözlük skoru	Sonuç
Bu program çok iyi, sunucu da güzel	2	0	Pozitif
Ne kadar çirkin dizi	0	1	Negatif
Bu akşam benim dizim var	0	0	Nötr

Sözlük temelli yaklaşım ile yapılan bu duygu analizi çalışmasının yarı denetimli olarak ifade edilmesinin nedeni sözlüklerin oluşturulurken etiketlenmiş veri setlerinden yararlanılmış olmasıdır.

1.200.000 tweetlik veri setinin önerilen yöntem ile duygu durumunun belirlenmesi işleminin ardından makine öğrenmesi temelli yöntemlerle algoritmanın performansı test edilmiştir. Bu amaçla önerilen yöntem ile etiketlenen veri seti çapraz değerlendirme yapılabilmesi ve daha küçük boyutlarda veri setleri ile çalışılması amacıyla 9 eşit parçaya ayrılmıştır. Modelin performansının ortaya çıkarılabilmesi amacıyla her bir veri seti %80'i eğitim, %20'si test olmak üzere ikiye ayrılmıştır.

Tablo 5: İkili karşılaştırma matrisi
(Confusion matrix)

		Gerçek	
		+	-
Tahmin	+	TP Doğru Pozitif	FP Yanlış Pozitif
	-	FN Yanlış Negatif	TN Doğru Negatif

Veri setlerine karar ağacı algoritması uygulanarak test verilerinin duygu sınıfları tahmin edilmeye çalışılmıştır. Tahmin performanslarının incelenebilmesi için Tablo 5'teki ikili matriste gösterildiği gibi veri setindeki gözlemlerin gerçek sınıf değerleri ile tahminlenen sınıf değerlerinin oranlanarak hesaplandığı doğruluk formülünden yararlanılmıştır (Eş. 5). Veri setlerine göre elde edilen doğruluk değerleri Tablo 6'da gösterilmektedir.

$$\text{Doğruluk} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN}) \quad (5)$$

Tablo 6: Duygu analizi algoritması performans değerleri
(Performance of sentiment analysis)

Veri seti	Doğruluk
1	%69
2	%66,8

Veri seti	Doğruluk
3	%67,4
4	%69,18
5	%67,73
6	%67,11
7	%70,18
8	%68,97
9	%67,22
Ortalama	%68,12

Literatürde yer alan yoğun programlama bilgisi kullanılan duygu analizi çalışmaları daha iyi performanslar ile sonuçlar üretmektedir ve birçok mühendislik temelli araştırmacı bu performansları arttırmaya yönelik algoritmalar üzerine çalışmaktadır. Fakat ilgili çalışmalar uygulama zorluğu nedeniyle her alandan araştırmacıların kullanabileceği yapıya sahip değillerdir. Bu da metinlerde gizli kalmış duygularının açığa çıkarılması ile yaratılacak değerlerden yoksun kalınmasına neden olmaktadır. Bu çalışma kapsamında geliştirilen model ise uygulama kolaylığı ön planda olan performans kalitesi yüksek bir duygu analizi algoritması önermektedir. Modelde sözlük oluşturma aşamasında R programlama dili, sözlük skoru belirleme ve karşılaştırma aşamalarında ise Excel ve VBA kullanılmıştır. Önceki bölümlerde de açıklandığı gibi R programlama dili günümüzde veri analitiği çalışmalarında sıklıkla kullanılan uygulama kolaylığı sunan bir yazılımdır. Excel ise Microsoft Office programlarından biri olarak milyonlarca bilgisayarda yer alan elektronik tabloları programdır. Bu da birçok araştırmacıya daha kolay kullanım imkânı sunmaktadır.

5. SONUÇLAR (RESULTS)

Teknolojik gelişmeler bilinen veri ve veri analizi kavramlarını hızla değiştirmektedir. Bu değişimin önemli bir ayağını metin verileri ve bu verilerden anlam çıkarılması adına yapılan analizler oluşturmaktadır. Müşterilerin ürünler hakkında yazdıkları geri bildirimler, otel veya restoran yorumları, mikrobloglarda kişilerin paylaştığı görüşler, birçok sosyal medya kanalında yazılan

yorumlar ve benzeri bir insan tarafından yazılmış metinlerin içerdiği duygunun araştırılabilir ve yorumlanabilir olması bu verilere değer katmaktadır. Bu verilerin analiz edilmesi ile elde edilen bilginin işletmelerce kullanılması ise işletmelere değer katmaktadır. Metin madenciliği olarak da adlandırılabilen metin verilerinin analizinde duygu analizi sıklıkla kullanılmaktadır. Duygu analizi sayesinde iletiyi yazan kişilerin hangi duygu ile ilgili metni yazdığı araştırılmaktadır. İletiler pozitif, negatif ya da nötr olarak sınıflandırılabilirdiği gibi mutlu, heyecanlı, üzgün şeklinde de sınıflandırılabilir. Böylece ürün ya da işletme hakkında yazılan bir iletiden müşterinin ilgili işletme veya ürünleri hakkında hangi tip düşünceye sahip olduğu anlaşılabilir.

Duygu analizi amacıyla çeşitli algoritmalar geliştirilmiştir. Bunlar çoğunlukla doğal dil işleme üzerine çalışan araştırmacıların ilgi alanı gibi görülse de aslında duygu analizi sonucu elde edilen bilgiler sosyal bilimler için büyük önem taşımaktadır. Fakat algoritmaların karmaşık ve kod bilgisi olmayan kişilerce kullanılması zor olmasından dolayı sosyal bilimcilerce çok uygulama yapılmadığı gözlenmiştir. Yazındaki çalışmaların çoğunluğu mühendislik ve bilişim alt yapısına sahip araştırmacıların, başarı performansını arttırmaya yönelik geliştirdikleri algoritmalarla oluşmaktadır. Çıktıların sosyal bilimlerde değerlendirilmesine yönelik çalışmalar sınırlı sayılarda kalmaktadır.

İngilizce dili için geliştirilen analiz edilmek istenen metin yazıldığında belli başarı oranları ile ilgili metnin duygusunu veren çeşitli kütüphane ve yazılımlar üretilmiştir. Fakat Türkçe dili için geliştirilmiş duygu analizi aracı olmadığından, bu çalışmada sözlük temelli bir algoritma önerilmiştir. Önerilen algoritma, her alandan araştırmacıya çalıştığı veri setine uygun duygu sözlüğünü oluşturup yoğun kodlama bilgisine ihtiyaç duymadan kolaylıkla duygu analizi çalışması yapılabilmektedir. Duygu analizi kapsamında geliştirilen duygu durumu belirleme algoritması sayesinde yaklaşık 1.200.000 tweet metninin hangi duyguyu barındırdığı istatistiksel olarak anlamlı bir doğruluk oranı (%68) ile belirlenebilmiştir.

Gelecek çalışmalarda metin ön işleme tekniklerini iyileştirip, temsil gücü daha yüksek duygu sözlükleri oluşturularak doğruluk oranının yükseltileceği çalışmalar yapılabilir. Duygu analizi çalışmaları sosyoloji, ekonomi, psikoloji ve işletme alanlarında önemli katkılar sağlayabileceğinden bu çalışmaların artırılması bunun için de performansı yüksek, uygulaması daha kolay analiz tekniklerinin geliştirilmesi gerekmektedir.

KAYNAKLAR (REFERENCES)

- [1] A. Kapucugil-İkiz, G. Özdağoğlu, "Text mining as a supporting process for VoC clarification", *Alphanumeric Journal*, 3(1), 25-40, 2015.
- [2] B. Pang, L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, In Proceedings of the 43rd annual meeting on association for computational linguistics", **Proceedings of the 43rd annual meeting on association for computational linguistics**, 115-124, Haziran 2005.
- [3] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, "Sentiment strength detection in short informal text", *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558, 2010.
- [4] A. Neviarouskaya, H. Prendinger, M. Ishizuka, "Textual affect sensing for sociable and expressive online communication", **International Conference on Affective Computing and Intelligent Interaction**, 218-229, 2007.
- [5] W. He, S. Zha, L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry", *International Journal of Information Management*, 33(3), 464-472, 2013.
- [6] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments", *Expert Systems with Applications*, 40(10), 4241-4251, 2013.
- [7] M. Ghiassi, J. Skinner, D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network" *Expert Systems with applications*, 40(16), 6266-6282, 2013.
- [8] X. Zhang, H. Fuehres, P. A. Gloor, "Predicting stock market indicators through twitter - I hope it is not as bad as I fear-" *Procedia-Social and Behavioral Sciences*, 26, 55-62, 2011.
- [9] J. Bollen, H. Mao, X. Zeng, "Twitter mood predicts the stock market", *Journal of Computational Science*, 2(1), 1-8, 2011.
- [10] A. Mittal, A. Goel, **Stock prediction using twitter sentiment analysis**, Stanford University, CS229, 2012.
- [11] S. Bouktif, M. A. Awad, "Ant colony based approach to predict stock market movement from mood collected on Twitter", **Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining**, 837-845, 2013.
- [12] A. Porshnev, I. Redkin, A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis", **In 2013 IEEE 13th International Conference on Data Mining Workshops**, 440-444, 2013.
- [13] F. Corea, "Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector", *Big Data Research*, 4, 70-74, 2016.

- [14] S. Asur, B. A. Huberman, "Predicting the future with social media", **Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01**, 492-499, 2010.
- [15] H. Rui, Y. Liu, A. Whinston, "Whose and what chatter matters? The effect of tweets on movie sales", *Decision Support Systems*, 55(4), 863-870, 2013.
- [16] T. Kim, J. Hong, P. Kang, "Box office forecasting using machine learning algorithms based on SNS data", *International Journal of Forecasting*, 31(2), 364-390, 2015.
- [17] A. Culotta, "Towards detecting influenza epidemics by analyzing Twitter messages", **In Proceedings of the first workshop on social media analytics - ACM**, 115-122, 2010.
- [18] E. Aramaki, S. Maskawa, M. Morita, "Twitter catches the flu: detecting influenza epidemics using Twitter", **In Proceedings of the conference on empirical methods in natural language processing (Association for Computational Linguistics)**, 1568-1576, 2011.
- [19] T. Bodnar, M. Salathé, "Validating Models For Disease Detection Using Twitter", **In Proceedings Of The 22nd International Conference on World Wide Web - ACM**, 699-702, 2013.
- [20] M. Meral, B. Diri, "Sentiment analysis on Twitter", **IEEE 22nd. International Conference On Signal Processing and Communications Applications**, 690-693, 2014.
- [21] Ö. Çoban, B. Özyer, G. T. Özyer, "Sentiment analysis for Turkish Twitter feeds", **IEEE 23th International Conference On Signal Processing and Communications Applications**, 2388-2391, 2015.
- [22] A. B. Eliaçık, N. Erdogan, "Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi", **UYMS**, 2015.
- [23] C. Akarsu, B. Diri, "Turkish TV rating prediction with Twitter", **IEEE 24th International Conference On Signal Processing and Communication Application**, 345-348, 2016.
- [24] D. Kayahan, A. Sergin, B. Diri, "Determination of TV programme ratings by twitter", **IEEE 21st International Conference On Signal Processing and Communications Applications**, 1-4, 2013.
- [25] S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger, "Social media analytics", *Wirtschaftsinformatik*, 56(2), 101-109, 2014.
- [26] R. Xia, C. Zong, S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification", *Information Sciences*, 181(6), 1138-1152, 2011.
- [27] M. N. Rahman, A. Esmailpour, J. Zhao, "Machine Learning with Big Data An Efficient Electricity Generation Forecasting System", *Big Data Research*, 5, 9-15, 2016.
- [28] B. Liu, "Sentiment analysis and opinion mining", *Synthesis lectures on human language technologies*, 5(1), 1-167, 2012.
- [29] W. Medhat, A. Hassan, H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, 5(4), 1093-1113, 2014.
- [30] C. Catal, M. Nangir, "A sentiment classification model based on multiple classifiers", *Applied Soft Computing*, 50, 135-141, 2017.
- [31] K. Çalış, O. Gazdağı, O. Yıldız, "Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti", *Bilişim Teknolojileri Dergisi*, 6(1), 1-7, 2013.
- [32] B. Liu, "Sentiment analysis and subjectivity", **In Handbook of Natural Language Processing**, Cilt 2, Chapman and Hall/CRC, 627-666, 2010.
- [33] L. Zhang, **Sentiment analysis on Twitter with stock price and significant keyword correlation**, Doktora Tezi, The University of Texas at Austin, Department of Computer Science, 2013.