Mühendislik Bilimleri ve Tasarım Dergisi
6(2), 324 – 333, 2018
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.384791

Araştırma Makalesi Research Article

METİN MADENCİLİĞİ TEKNİKLERİ İLE SOSYAL AĞLARDA BİLGİ KEŞFİ

Asım Sinan YÜKSEL [1*], Fatma Gülşah TAN[2]

[1]Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Isparta, Türkiye
[2]Manisa Celal Bayar Üniversitesi, Kırkağaç Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Manisa, Türkiye

Anahtar Kelimeler

Metin Madenciliği,
Duygu Analizi,
Sosyal Ağlar,
Doğal Dil İşleme,
Karar Destek Sistemleri

Anahtar Kelimeler

Sosyal ağların popülerleşmesi ile birlikte sosyal medya ortamlarındaki veri akışı artmıştır. Verinin artması, yapısal olmayan veri miktarını arttırmış, dolayısıyla veriden anlam çıkarabilme yeteneği olan, otomatik olarak duygu analizi yapan karar destek sistemlerinin önemi artmıştır. Bu çalışmada sosyal ağlardan daha yararlı bilgilere ulaşmak, kişilerin bulundukları çevrede gitmeyi planladıkları mekânlar hakkındaki görüşlerini otomatik olarak yorumlamak amacıyla, metin madenciliği ve duygu analizi teknikleri uygulanarak bir karar destek sistemi geliştirilmiştir. Olumlu ve olumsuz görüşler gerçek zamanlı olarak belirlenmekte ve duygu analizleri otomatik bir şekilde yapılmaktadır. Metin madenciliği tekniklerinin uygulanması için gerekli olan veri setleri insanların dünya üzerinde herhangi bir şehirde yemek, içmek, alışveriş yapmak veya ziyaret etmek için mekânları bulmalarını sağlayan Foursquare sosyal paylaşım platformunda bulunan yorumlardan oluşturulmuştur. İngilizce için kullanılan duygu analizi yöntemlerinin Türkçede göstereceği tepkiler incelenmiş ve Türkçeye özel yeni bir yöntem geliştirilerek duygu analizi metotları ile birleştirilmiştir. İkili sınıflandırmada %81,97, üçlü sınıflandırmada ise %84,49 başarı oranı elde edilmiştir.

KNOWLEDGE DISCOVERY IN SOCIAL NETWORKS USING TEXT MINING TECHNIQUES

Keywords

Text Mining,
Sentiment Analysis,
Social Networks,
Natural Language Processing,
Decision Support Systems

Keywords

With the popularization of the social networks, data flow in social media environments has increased. Increase in the number of non-structural data has increased the importance of decision-support systems that performs automatic sentiment analysis and have the ability to derive meaning from data. In this study, a decision support system was developed by applying text mining and sentiment analysis techniques in order to reach more useful information from social networks and automatically interpret the opinions of the people about the places they planned to visit. Positive and negative opinions are determined in real time and sentiment analysis is done automatically. The data set that is needed to implement the text mining techniques was collected from comments on the Foursquare social sharing platform, which allows people to find places to eat, drink, shop or visit in any city in the world. The reactions of the sentiment analysis methods used in English to the Turkish were examined and a new method specific to Turkish was developed and combined with sentiment analysis methods. Success rates of 81,97% and 84.49% were achieved in binary and ternary classifications respectively.

Alıntı / Cite

Yüksel, A.S., Tan, F.G., (2018). Metin Madenciliği Teknikleri ile Sosyal Ağlarda Bilgi Keşfi, *Journal of Engineering Sciences and Design*, 6(2), 324-333.

Yazar Kimliği / Author ID (ORCID Number)

A. S. Yüksel, 0000-0003-1986-5269
F. G. Tan, 0000-0002-2748-0396

[*] İlgili yazar / Corresponding author: asimyuksel@sdu.edu.tr, +90-246-211-1380

Başvuru Tarihi / Submission Date	27.01.2018
Revizyon Tarihi / Revision Date	02.04.2018
Kabul Tarihi / Accepted Date	29.05.2018
Yayım Tarihi / Published Date	23.06.2018

1. Giriş

Sosyal ağların kullanımının artmasıyla birlikte yararlı ve net bilgi elde etmenin önemi de artmıştır. Bu hızlı artış yapısal olmayan verilerin miktarında da artışa neden olmuştur. Veriler işlenmediği sürece veri tabanlarında anlamsız bir yığın olarak depolanmaktadır (Dolgun, 2006).

Türkçe metinlerde metinsel verilerle çalışmak ve anlamlı sonuçlar elde etmek oldukça zordur. Metin madenciliğini veri madenciliğinden ayıran fark da burada ortaya çıkmaktadır. Metin madenciliğinde veri seti belli olan, düzgün oluşturulmuş veri tabanlarından ziyade, doğal dil işleme kullanılarak elde edilen düzgün metinler kullanılır. Bu sebeple metin madenciliği yöntemleri doğal dil işleme teknikleri ile birlikte kullanılarak karmaşık yapıların bulunduğu ifadeler anlamlı bir şekilde sınıflandırılmalıdır.

Doğal dil işleme çalışmaları çoğunlukla dil bilim bilgisiyle bütünleşmiş yapay zekâ çalışmalarını kapsamaktadır. Doğal dil işleme çalışmalarında, bir yazının anlaşılması ve bunlardan anlam çıkarma, belirtilmek istenen düşüncenin bulunması, bir konuşmaya ya da sorulan sorulara yanıt verilmesi hedeflenmektedir (Adalı, 2012).

Sosyal medyanın çok sık kullanılması, üzerinde en çok veri barındıran platform olmasına neden olmuş ve metin madenciliği çalışmaları yapmak kaçınılmaz hale gelmiştir. Bu çalışma kapsamında Foursquare sosyal paylaşım platformu üzerinde yapılan yorumlar veri kaynağı olarak kullanılmış ve bu platformdan elde edilen veriler üzerinde metin madenciliği teknikleri uygulanarak mekânlar hakkında belirtilmek istenilen düşüncelerin ortaya çıkarılması hedeflenmiştir. Foursquare platformu kişilerin gittikleri mekânlar hakkında şikâyetlerini ve memnuniyetlerini paylaşabildikleri bir ortamdır. Bu şekilde mekân sahiplerinin eksikliklerini gidermeleri ve hizmet kalitelerini arttırmaları yönünde adım atmaları amaçlanmıştır.

Karar destek sistemlerinin olmadığı durumlarda insanların duygu ve düşüncelerinin değerlendirilmesi çok uzun zaman alabilmektedir. Bir sosyal medya kullanıcısının bir mekân hakkında yüzlerce yorumu okuması ve buna göre karar vermesi zahmetli bir süreçtir. Bu kapsamda, önce doğal dil işleme teknikleri kullanılarak kullanıcı yorumlarında geçen ve yanlış yazılan kelimeler düzenlenmiş, daha sonra da düzenlenen veriler kullanılarak metin madenciliği teknikleri ile hızlı ve verimli bir şekilde duygu analizi yapmayı sağlayan karar destek sistemi geliştirilmiştir. Geliştirilen karar destek sistemi ile bu tarz problemlerin ortadan kaldırılması, duygu analizlerinin daha hızlı ve verimli yapılması, olumlu ve olumsuz görüşlerin otomatik olarak belirlenmesi amaçlanmıştır.

Yapılan çalışma ile var olan doğal dil işleme ve metin madenciliği tekniklerinin geliştirilen Sosyal Bilgi Keşfi Algoritması (SBKA) ile iyileştirilmesi ve kullanılan yöntemlere katkı sağlanması amaçlanmıştır.

2. Bilimsel Yazın Taraması

Delibaş (2008), yapmış olduğu çalışmada, doğal dil işleme teknikleri ile Türkçenin biçimsel yapısının çözümlemesinin yapılması üzerinde durmuştur. Çalışmanın amacı, eldeki Türkçe metin üzerindeki yazım yanlışlarının bulunması, bu yanlışların ayıklanması ve düzeltilmesidir. Yapılan bu çalışmanın daha önceden yapılan çalışmalara göre başarı oranının yüksek olduğunu belirtmiştir.

Pang vd. (2008), yapmış oldukları çalışmada, metin sınıflandırma teknikleri kullanmışlardır. Sonuç olarak belgenin yalnızca öznel bölgelerine uygulanan Makine Öğrenmesi yöntemini önermişlerdir. Bu teknik duygu analizi için uyarlanmış ve duygu sınıflandırması yapılmıştır. En başarılı sonuç Destek Vektör Makineleri (DVM) sınıflandırma algoritması ile alınmıştır.

Eroğul (2009), yapmış olduğu çalışmada, veri kaynağı olarak Türkçe film yorumlarını kullanmıştır. Bu veri seti üzerinde DVM ve doğal dil işleme tekniklerini kullanarak duygu analizi yapmıştır. Yaptığı ikili sınıflamada (Olumlu, Olumsuz) %85 başarı elde etmiştir.

Go vd. (2009), yapmış oldukları çalışmada, veri kaynağı olarak Twitter'dan elde ettikleri 800.000 olumlu, 800.000 olumsuz yorumu kullanmışlardır. Makine Öğrenmesi algoritmalarından Sade Bayes, Destek Vektör Makineleri, Maksimum Entropi yöntemini kullanarak bi-gram, unigram ile verileri birleştirerek analiz etmişler ve duygu analizi yapmayı amaçlamışlardır. En iyi sonuç Maksimum Entropi unigram ile bi-gramın birlikte kullanılması sonucu %83 doğruluk payıyla elde edilmiştir.

Agarwal vd. (2011), yapmış oldukları çalışmada, veri kaynağı olarak Twitter sosyal medya aracını kullanmışlar ve buradan elde ettikleri verileri etiketlemişlerdir. Bu veriler üzerinde ikili (Olumlu, Olumsuz) ve üçlü (Olumlu, Olumsuz, Nötr) sınıflama yapmışlardır. Buna göre ikili sınıflamada Senti-Feature ve Unigram, üçlü sınıflamalarda Tree Kernel, Senti-Feature en iyi sonucu vermiştir.

Adalı (2012), yapmış olduğu çalışmada, insanlık tarihinde önemli bir yeri olan dilin yapısını ve kurallarını incelemiş, doğal dil işleme kapsamında yer alan temel konuları, doğal dil işleme uygulamalarını ve dillere ilişkin özellikleri tanıtmıştır.

Vural vd. (2013), yapmış oldukları çalışmada, Makine Öğrenmesi ve sözlük tabanlı yöntemleri kullanarak duygu analizi yapmayı amaçlamışlardır. Sözlük tabanlı uygulamalarda ilk çalışmalar İngilizce dili için yapılmıştır. İngilizce dili için yapılmış olan "Dictionary of Affect in Language (DAL)" in farklı dillere nasıl çevrileceği incelenmiş ve Türkçe için ilk sözlük tabanlı uygulama olan SentiStrength'i geliştirmişlerdir.

Yıldırım vd. (2015), Türkçe dili için sosyal medya metinleri üzerinde doğal dil işleme çalışmaları yapmışlardır.

Twitter üzerinden toplanan 12790 yorumu el ile etiketlemişlerdir (3541 olumlu, 4249 negatif ve 5000 Nötr). 3'lü gruplandırma ile yapılan duygu analizi çalışmalarında Türkçe metinler üzerinde %78,83 başarı elde etmişlerdir.

Dumanlı (2017), yapmış olduğu çalışmada, veri seti olarak Twitter'dan haber ajansları tarafından gönderilen 81174 adet yorum kullanmıştır. Kullanıcıların etkileşim seviyesini belirlemek amacıyla tweetlerin, retweet ve beğeni sayıları açısından analizini gerçekleştirmiştir. Analizde haber tweetlerinin 48 saat içerisindeki etkileşim oranları, günün belli zaman dilimlerinde gönderilen tweetlerin kullanıcılar tarafından ne kadar retweet edildiği ve beğenildiği, retweet ve beğeni sayılarının toplam takipçi sayılarına oranları hesaplanarak kullanıcıların etkileşim düzeyleri belirlenmiştir.

Tablo 1'de bu çalışma kapsamında incelenen, sosyal medyayı veri kaynağı olarak kullanan, doğal dil işleme veya sınıflandırma algoritmaları kullanılarak duygu analizi yapan diğer çalışmalar ve özetleri verilmiştir.

Tablo 1. İncelenen diğer makale özetleri

Çalışmayı Yapan	Yıl	Sınıflandırma	Algoritma	Veri Kaynağı
Piryani vd.	2017	İkili Sınıflandırma	Naive Bayes	Twitter
Liu vd.	2017	İkili Sınıflandırma	PROMETH EE II	Otomobil Ürün Yorumları
Vilares vd.	2017	İkili Sınıflandırma	SO-CAL, SentiStrength	Film Yorumları
Schumaker vd.	2016	İkili Sınıflandırma	Central Support Model	Twitter
Williams vd.	2016	İkili Sınıflandırma	Naive Bayes, Destek Vektör Makineleri	British National Corpus, Learn English Today
Ptaszynski vd.	2013	İkili Sınıflandırma	Semantik	Fast Food Yorumları
Balahur vd.	2012	İkili Sınıflandırma	Destek Vektör Makineleri, K nearest neighbor, Naive Bayes	Media Analysis Şirketi
Cao vd.	2011	İkili Sınıflandırma	Zayıf ve Yarı Denetimli Sınıflandırma	Film Yorumları, IMDB, Amazon

Yapılan analizler sonucunda İngilizce dilinde birçok çalışma ve kaynak olduğu, İngilizce dışındaki dillerde yapılan çalışmaların ise yetersiz bulunduğu görülmüştür.

3. Materyal ve Yöntem

Bu çalışmanın kapsamı Foursquare'den toplanan veri seti üzerinde doğal dil işleme teknikleri kullanılarak yanlış yazılan kelimelerin tespitini ve düzeltilmesini, düzeltilen cümlelerden öznitelik seçme yöntemleri ile Türkçe metinlerdeki gereksiz kelimelerin tespitini ve ayrıştırılmasını, sınıflandırma algoritmalarını ve metin madenciliği tekniklerini kullanarak cümleler üzerinde duygu analizini içermektedir.

3.1. Veri Kaynağının Toplanması ve Doğal Dil İşleme Aşaması

Sosyal ağlarda metin madenciliği tekniklerini uygulayabilmek için gerekli olan veri seti Foursquare sosyal medya platformunda bulunan mekânlara ait yorumlardan elde edilmiştir. Toplanan bu yorumlar Foursquare API'si kullanılırken "query=food" ile sınırlandırılmış, bu şekilde türü yemek-içecek olan mekânlara ait yorumlar toplanmıştır. Bu çalışmada kullanılan tüm tablolar Microsoft SQL Server Management Studio üzerinde oluşturulmuş ve veriler tablolarda depolanmıştır.

Geliştirilen algoritmanın ve duygu analizi sonucunun doğruluk oranını test etmek için Türkçe yorumlar toplanmış ve veri tabanına kaydedilmiştir. 128 mekâna ait toplam 7086 yorum hem geliştirilen algoritmanın hem de diğer sınıflandırma ve duygu analizi algoritmalarının test aşamasında kullanılmıştır.

Doğal dil işleme, bilgisayar yardımıyla dilin işlenmesi üzerinde durur. Doğal dil işlemenin ana görevi doğal bir dili çözümleyecek, yorumlayacak bilgisayar sistemlerini tasarlamak ve gerçekleştirmektir (Rich, 1991).

Bu alanda yapılmış çalışmaların çoğu geçerli bir dil olan İngilizce için geliştirilmiştir. Kullanılan kural ve algoritmaların birebir Türkçeye çevrilmesi Türkçe dilinin biçimsel farklılığından dolayı pek mümkün değildir. Bu aşamada doğal dil işleme teknikleri ile yazım hatalarının bulunması ve bu hataların en yüksek başarı ile düzenlenmesi amaçlanmıştır.

Sınıflandırma algoritmalarının ve duygu analizi çalışmalarının doğru sonuç verebilmesi için veri setinin temiz ve doğru yazılmış metinlerden oluşması gerekmektedir. Tablo 2'de gösterildiği gibi sosyal medyadan toplanan yorumların dil bilgisi kurallarına uygun olmadığı görülmüştür. Bu tarz sorunlar duygu analizi yapmayı zorlaştırmaktadır.

Tablo 2. Foursquare'de bulunan mekânlara ait yorumlar

Mekan Id	Yorum
539061b3498e21833a537777	Dnynz. Cheese California
4d7278e427ddb60cac57ce1b	Davukları möthiş
502568cee4b0273849bbd37c	Corvasi cook güzel sezar salatasida fena değil elinize saglik
4d511362c5ff6ea83b71a007	Enfesssssss

Bu çalışma kapsamında doğal dil işleme çalışmaları yapılırken Türkçe doğal dil işleme kütüphanesi olan Zemberek ve ITU Turkish NLP Web Service API'si kullanılmıştır.

3.1.1. Zemberek Kütüphanesi Kullanımı

Zemberek açık kaynak kodlu Türkçe doğal dil işleme kütüphanesidir (Akın vd., 2007). İçerisinde kullanılan dile ait kelime ve bunların köklerini barındırabildiği için yanlış yazılan kelimeleri düzenlemekte ve en doğru kelimeyi önermektedir. Tablo 3'te, Tablo 1'de verilen verilerin Zemberek kütüphanesi kullanılarak düzenlenmiş hali verilmiştir.

Tablo 3. Zemberek kütüphanesi kullanılarak düzenlenen yorumlar

Mekan Id	Yorum
539061b3498e21833a537777	Dnynz. Cheese California
4d7278e427ddb60cac57ce1b	Tavukları müthiş
502568cee4b0273849bbd37c	Corvasi cook güzel mezar salatasıda fena değil elinize sağlık
4d511362c5ff6ea83b71a007	Enfesssssss

Zemberek'in kelime köklerine ayrıştırma başarısı diğer yöntemlere göre yüksek olsa da kelimelerde sadece +-1 harf hatasını düzeltebilmesi yönüyle sınıflandırma aşamasında geliştirdiğimiz yönteme ek olarak ITU Turkish NLP Web Service kullanılmıştır. Şekil 1'de geliştirilen uygulamada Zemberek kütüphanesi kullanılarak yanlış yazılan cümleler ile nasıl sonuç alındığını gösteren bir örnek verilmiştir.



Şekil 1. Zemberek kütüphanesi kullanarak cümle düzenleme

3.1.2. ITU Turkish NLP Web Service Kullanımı

ITU Turkish NLP Web Service, İstanbul Teknik Üniversitesinde Natural Language Processing grubu tarafından geliştirilen Türk doğal dil işleme araçları ve uygulama programlama arayüzlerini (API) içermektedir (Eryiğit, 2014).

Tablo 4'te, Tablo 1'de verilen verilerin ITU Turkish NLP Web Service kullanılarak düzenlenmiş hali verilmiştir.

Tablo 4. ITU Turkish NLP Web Service kütüphanesi kullanılarak düzenlenen yorumlar

Mekan Id	Yorum
539061b3498e21833a537777	Deneyiniz. Cihetse California
4d7278e427ddb60cac57ce1b	Tavukları müthiş

502568cee4b0273849bbd37c	Çorbaya çok güzel Sezar sağlatasıda fena değil elinize sağlık
4d511362c5ff6ea83b71a007	Enfes

ITU Turkish NLP Web Service Zemberek kütüphanesi gibi kelime bazlı değil de cümle bazlı çalıştığından dolayı ve araştırmacılara doğal dil işleme araçlarını ön işleme, söz dizim, morfoloji ve varlık gibi katmanlarla da sunduğu için daha kullanışlıdır. Bu çalışma kapsamında ITU Turkish NLP Web Service'nin Normalization metodu kullanılarak yanlış yazılan kelimelerin tespiti yapılmıştır. Şekil 2'de geliştirilen uygulamada ITU Turkish NLP Web Service kullanılarak yanlış yazılan cümleler ile nasıl sonuç alındığını gösteren örnek verilmiştir.



Şekil 2. ITU Turkish NLP Web Service kullanarak cümle düzenleme

3.2. Öznitelik Seçme ve Duygu Analizi

Metin madenciliği alanında kullanılan öznitelik seçme ve ön işleme aşaması gereksiz kelimelerin temizlenmesinin yanında bunları uygun formata getirme işlemini de gerçekleştirmektedir (Feldman ve Sanger, 2007).

Öznitelik seçme, Foursquare üzerinden elde edilen verilerden en değerli ve belirleyici değişkenlerin ve gereksiz kelimelerin tespiti için kullanılmaktadır. Ön işlemenin temizleme aşamasında Türkçede sık kullanılan kelimeler, noktalama işaretleri, zarflar, fiiller, yani tek başına anlam ifade etmeyen kelimeler elimizdeki veri setinde aranır. Bu şekilde Foursquare' den elde edilen yorumlarda sıkça geçen ve yorumlamak için duygu analizine katkı sağlamayan kelimeler elenir.

Duygu analizi insanların duygularını, görüşlerini değerlendiren, bir konu hakkındaki davranışlarını yazılı dil üzerinden analiz eden çalışma alanıdır (Liu , 2012). İnternetin hayatımızı sanallaştırdığı bir dönemde insanlar duygu ve düşüncelerini artık sosyal ağlarda belirtmektedirler. Artan veri akışından dolayı duygu analizi, doğal dil işleme ve metin madenciliği çalışma alanlarının en aktif çalışma konularından biri olmuştur.

Duygu analizi ile anlam vurgusunun tespiti için Foursquare'den elde edilen yorumların doğal dil işleme çalışmaları ile yapısal veriler haline getirilmesi gerekir. Bu şekilde duygu analizi, hazırlanmış yapısal veri seti kullanılarak geliştirilen algoritmalar ve makine öğrenmesi yöntemleri ile gerçekleştirilir. Foursquare' den elde edilen yorumlarda duygu analizi yapmak ve duygu içeren kelimeleri tespit etmek için Tablo 5'te örnek verilen önceden oluşturulan kelime listesi kullanılmıştır. Veri setindeki cümlelerin içerdiği kelimeler eğer bu liste içerisinde varsa kelimelerin değerleri toplanıp ortalaması alınarak cümlenin içerdiği duygu tespit edilmektedir.

Tablo 5. Duygu analizinde kullanılan örnek kelime listesi

Kelime	Pozitif/Negatif
muhteşem	1
lezzetli	1
lezzetsiz	0

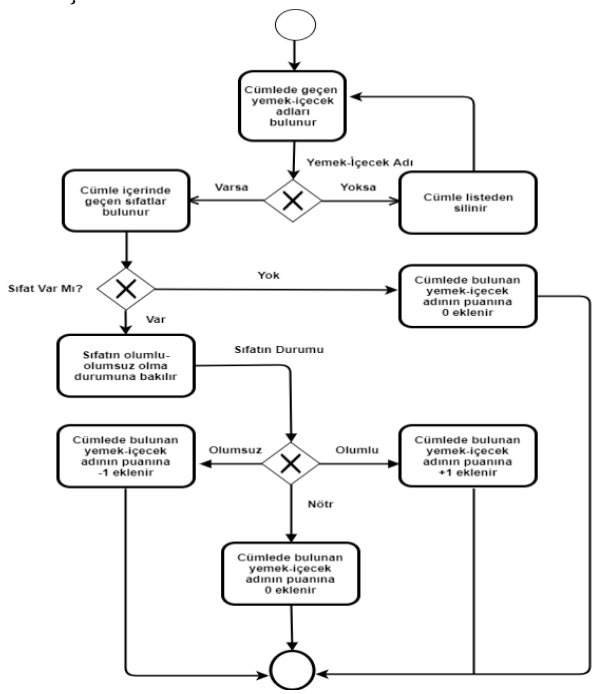
Sözlük tabanlı yaklaşım olarak adlandırılan bu yöntem Türkçe metinler üzerinde yapıldığı için %100 başarı elde edilememiştir. Çünkü cümle içerisinde geçen duygu analizi içeren veri listesi haricindeki kelimeler cümlenin anlamını değiştirebilmektedir. Örneğin "Mantı lezzetli değildi." cümlesini değerlendirdiğimizde cümle olumsuzdur. İçerisinde geçen sıfatlara bakıldığında "lezzetli" kelimesi cümleye pozitif anlam katarken "değildi" kelimesi cümlenin anlamını değiştirmektedir. Bu tarz cümleler duygu analizi çalışmalarında doğruluk oranını düşüreceğinden bu tarz problemlerin tespit edilip algoritmanın bunlara göre düzenlenmesi gerekmektedir.

Duygu analizi yapılırken sadece cümlenin olumlu-olumsuz olma durumu değil o mekânda en çok hangi yemek veya içeceğin beğenildiği de bulunmaya çalışılmıştır. Bunun için Tablo 6'da örnek verildiği gibi yemek-içecek adlarının tutulduğu veri listesi oluşturulmuştur.

Tablo 6. Duygu analizinde kullanılan örnek yemek-içecek listesi

Kelime
Abugannuc
Bomonti Kebap
Sifonda Kahve
Dağ Kekikli Kremalı Tavuk

Şekil 3'te duygu analizi çalışmaları yapılırken puanlamanın nasıl olduğunu gösteren akış diyagramı verilmiştir.

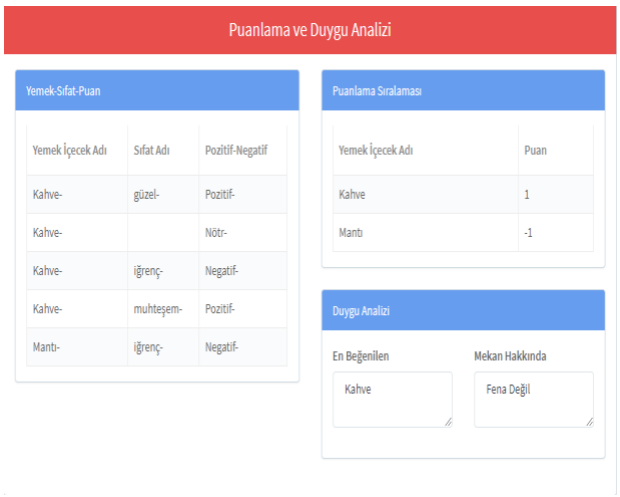


Şekil 3. Duygu analizi ve puanlama akış diyagramı

Puanlandırma aşamasında eğer cümle içerinde yemek-içecek geçiyorsa ve herhangi bir duygu analizi içeren kelime kullanılmamışsa bu cümlenin puanı 0 olarak hesaplanır. Olumlu cümleler için +1, olumsuz cümleler için -1, nötr cümleler için 0 puanlaması yapılır. Yemek-içecek adına göre sınıflandırma yapılarak en çok hangi yemek-içeceğin geçtiği ve bunun beğenilip beğenilmediği bulunur. Şekil 4'te ve Şekil 5'te geliştirilen uygulamada, yorumlar üzerinde doğal dil işleme, duygu analizi, metin madenciliği teknikleri ile sınıflandırma çalışmaları uygulandığında elde edilen sonuç gösterilmiştir.



Şekil 4. Yorum üzerindeki doğal dil işleme sonucu



Şekil 5. Yorum üzerindeki duygu analizi ve puanlama sonucu

3.3. Geliştirilen Algoritmanın Başarı Oranının Değerlendirilmesi

Testler 128 farklı mekân ve bunlara ait 7086 adet yorum üzerinde yapılmıştır. Bu testler;

- ITU Turkish NLP Web Service ile SBKA'nın test edilmesini
- Zemberek kütüphanesi kullanılarak SBKA'nın test edilmesini
- Weka'da bulunan Naive Bayes algoritması kullanılarak yapılan sınıflandırma ile test edilmesini
- Verilerin tek tek elle işlenerek ve Naive Bayes algoritması uygulanarak test edilmesini
- İngilizce cümleler üzerinde duygu analizi yapan API kullanılarak test edilmesini içermektedir.

Elle tek tek işleme aşamasında veri gereksiz kelimelerden temizlenmiş, yanlış yazılan kelimeler düzenlenmiş, cümle içerisindeki anlamına göre yorum cümlelere ayrılmıştır.
İngilizce cümleler ile üçlü (Olumlu, Olumsuz, Nötr)(Tablo 7), Türkçe cümleler ile ikili (Olumlu, Olumsuz)(Tablo 8) ve üçlü (Olumlu, Olumsuz, Nötr)(Tablo 9) sınıflandırma yapılmış ve aynı zamanda Türkçe yorumlarda cümle içerisinde geçen yemek-içecek adları etiketlenerek geliştirilen algoritmanın başarı oranı test edilmiştir.

Tablo 7. İngilizce cümlelerde üçlü sınıflandırılan örnek veri listesi

Mekan Id	Yorum	Olumsuz/ Olumsuz
42fa9500f964a520e2261fe3	The cornbread isn't too great.	0
5632c07e498eaa6c566acefc	Must try the fried chicken!	1
48b5de88f964a520b7511fe3	Lasagna isn't great.	0

Google Translate API'si (Google Inc., 2018) kullanılarak önce Türkçe yorumlar İngilizceye çevrilmiştir. Daha sonra İngilizce metinler üzerinde duygu analizi yapan Text Analyis API (Aylien Inc., 2018) kullanılarak cümlenin olumlu, olumsuz, nötr olma durumları bulunmuştur. Bu şekilde elle tek tek etiketlenip 3'lü sınıflandırılan veri ile birlikte geliştirilen algoritmanın başarısı test edilmiştir.

Tablo 8. Türkçe cümlelerde ikili sınıflandırılan örnek veri listesi

Yorum	Olumsuz/ Olumsuz	Beğenilen Yemek-İçecek
Mantısı ve yemekleri harika.	1	Mantı
Dağ kekikli kremalı tavuk süper..	1	Dağ Kekikli Kremalı Tavuk
ama işletmecilik çok zayıf .	0	NULL

Tablo 9. Türkçe cümlelerde üçlü sınıflandırılan örnek veri listesi

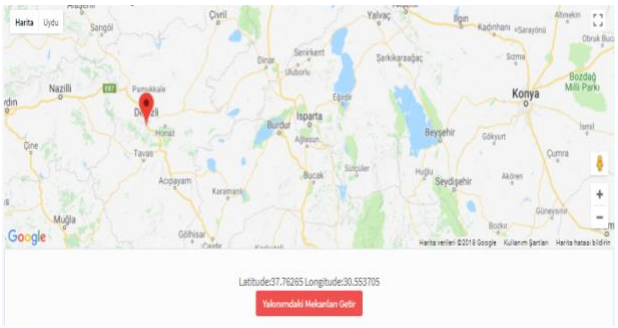
Yorum	Olumsuz/ Olumsuz/ Nötr	Beğenilen Yemek-İçecek
arnavut ciğeri fena değil	2	arnavut ciğeri
Sandviçler ve kurabiyeler bir harika bence :)	1	sandviç, kurabiye
şakşuka ve peynir kötü	0	şakşuka, peynir

WEKA arayüzü üzerinde bulunan Sade Bayes algoritması kullanılarak veri tabanına kaydedilen işlenmiş yorumlar (ikili, üçlü) sınıflandırılmış ve geliştirilen algoritma ile kıyaslanmıştır.

Etiketleme işleminde negatif cümleler için 0, pozitif cümleler için 1, nötr cümleler için 2 kullanılmıştır.

3.4. Anlık Çekilen Veriler Üzerinde Metin Madenciliği Tekniklerinin Uygulanması

Bu çalışma kapsamında metin madenciliği tekniklerini uygulamak ve her aşamada Foursquare'ye istek göndermemek için veri tabanına kaydedilen mekânlar ve bunlara ait yorumlar kullanılmıştır. SBKA metin madenciliği, doğal dil işleme ve duygu analizi çalışmalarını birleştirerek oluşturduğumuz algoritmadır. Visual Studio platformu ve Asp.Net teknolojisi kullanılarak web tabanlı bir uygulama geliştirilmiş (Yuksel vd., 2018) ve geliştirilen algoritmanın gerçek zamanlı performansını test etmek için anlık çekilen veriler üzerinde SBKA ile denemeler yapılmıştır. Şekil 6'da gösterildiği gibi veriler çekilirken harita üzerinde GPS bilgileri kullanılarak seçilen koordinatlara göre Foursquare'den o bölgede bulunan mekân listesi alınmaktadır. Şekil 7'de ise seçilen koordinatlar yakınında bulunan mekânlar listelenmekte ve yorumla tuşuna basıldığında seçilen mekânın yorumları üzerinde SBKA uygulanarak elde edilen sonuçlar kullanıcıya gösterilmektedir.



Şekil 6. GPS bilgilerinin seçilip Foursquare'ye gönderilmesi

Venue ID	Venue Name	Operations
55e5e43b498e5dfaf8708ffd	The Piano Cafe & Performance Hall	Yorumla
5a53442bf62e092a22595a40	Panino Fabs Pizza	Yorumla
570ce401498e16f83f4e4177	WİSH Fast&Food	Yorumla
5a7ca66e69e77b4a8adb2e0d	Buongiorno Bistro & Cafe	Yorumla
5633aa79498e64092a3927c5	Dürüm Bar	Yorumla
4f5a2d60e4b083219d268424	Minta	Yorumla
5662eaf1498e16a3d5f76782	LEZZET DURAĞI	Yorumla
59c03b61db1d8147ca06caf3	Gecekondu	Yorumla

Şekil 7. GPS bilgilerine göre mekânların listelenmesi

Şekil 8'de anlık verilerle çalışırken uygulanan algoritmanın akış diyagramı verilmiştir.



Şekil 8. Yakınımdaki mekânlar akış diyagramı

4. Araştırma Bulguları

4.1. Deneysel Sonuçlar

Literatürde kullanılan metin madenciliği tekniklerinin yanında geliştirdiğimiz algoritma Foursquare'den çekilen yorumlara uygulanmış ve Türkçe dili yüksek başarı oranları elde edildiği görülmüştür. Doğal dil işleme çalışmalarında ITU Turkish NLP Web Service'nin Zemberek kütüphanesine göre uygulanabilirliğinin ve kullanılabilirliğinin daha kolay olduğu görülmüştür. Yapılan web tabanlı uygulama (Yuksel vd., 2018) ile hem duygu analizine etki edecek kelime listesi hem de yemek-içecek listesi genişletilebilir. Yapılan çalışma mekânlar hakkında olduğu için bu liste bu çalışmaya özel olarak oluşturulmuştur. Farklı konularda veri seti oluşturularak geliştirilen algoritmaya uygulanması ve duygu analizi çalışmalarının yapılması mümkündür.

SBKA ile yapılan sınıflandırma çalışmasının başarısını kıyaslamak için elle tek tek etiketlenip ikili ve üçlü sınıflandırılan 7086 adet Türkçe yorum kullanılmıştır. İkili sınıflandırmaya ek olarak Makine Öğrenmesi yöntemlerinden biri olan Sade Bayes algoritması ile testler gerçekleştirilmiştir.

Tablo 10. İkili sınıflandırma test sonuçları

Algoritma	Olumlu	Olumsuz	Başarı Oranı %
El ile etiketleme	6028	1058	-
SBKA	5204	1882	81,97
Sade Bayes	4519	2567	73,8

Üçlü sınıflandırma için de yorumlar İngilizce diline çevrilmiş ve İngilizce metinlerde üçlü sınıflandırma yapan API (Aylien Inc., 2018) kullanılmıştır.

Tablo 11. Üçlü sınıflandırma test sonuçları

Algoritma	Olumlu	Olumsuz	Nötr	Başarı Oranı %
El ile etiketleme	5594	996	526	-
SBKA	5204	851	1031	84,49
Text Analysis API	3308	1592	2186	59.38
Sade Bayes	4680	760	1646	78

İkili sınıflandırmada geliştirilen SBKA ile %81,97, Sade Bayes algoritması kullanılarak yapılan sınıflandırmada %73,8 başarı oranı elde edilmiştir. Üçlü sınıflandırmada SBKA ile %84,49, İngilizce cümleler üzerinde sınıflandırma yapan API kullanılarak yapılan sınıflandırmada %59,38 başarı oranı elde edilmiştir. SBKA'nın üçlü sınıflandırmada daha yüksek başarı oranı elde ettiği görülmüştür.

Geliştirilen yöntemin Türkçeye kolay bir şekilde uygulanabileceği gösterilmiştir. SBKA ile Türkçe dili için yapılan diğer yöntemler kıyaslandığında yüksek bir başarı oranı elde edildiği görülmüştür.

Sınıflandırma çalışmalarıyla birlikte mekânda en çok hangi yemek veya içeceğin beğenildiği bulunmaya çalışılmıştır. Buna göre ITU Turkish NLP Web Service kullanılarak SBKA ile 128 mekân için 85 mekân, Zemberek kütüphanesi kullanılarak SBKA ile 107 mekân, WEKA kullanılarak yapılan sınıflandırma ile 78 mekân için beğenilen yemek-içecek doğru bulunmuştur. Mekân sayısı ve buna bağlı olarak yorum sayılarının artması doğruluk oranını artıracaktır.

5. Sonuç ve Tartışma

Bu çalışma kapsamında Foursquare sosyal medya platformu üzerinde bulunan mekânlar hakkında yapılan yorumlar üzerinde çalışılmıştır ve duygu analizi ile de yapılan yorumların olumlu ya da olumsuz olma durumu analiz edilmiştir. Bu şekilde duygu analizi, mekânlar için mekânın sosyal medya üzerinde nasıl bir havaya sahip olduğunu tespit etmek, kriz anında eksiklere hızlı bir şekilde müdahale edebilmek, ürünlerin başarı oranlarını tespit etmek ve rakiplere karşı durumunu kontrol edebilmek için önemli bir analiz çalışmasıdır.

Bir mekân hakkında yapılan yorumlar tüm sosyal medyaya yayılır. Kullanıcı tarafından bu verilerin takibi bir süre sonra imkânsız hale gelir. Bu yüzden hem kullanıcıların hem de mekân sahiplerinin internet ortamında yapılan yorumları otomatik ve doğru bir şekilde analiz edebilecek karar destek sistemlerine ihtiyaçları vardır. Karar destek sistemleri, karar verme sürecinin hızlandırılması, verimliliğin arttırılmasında kullanılacak raporlamaların kolayca yapılması için çok önemlidir. Bu doğrultuda, yazılan binlerce yorumun kullanıcı tarafından tek tek okunmasına gerek kalmadan kullanımı kolay, hızlı cevap verebilen bir karar destek sistemi geliştirilmiştir. Geliştirilen karar destek sistemi ile mekân bazlı olarak yorumlar olumlu, olumsuz, nötr şekilde gruplandırılarak kategorilere ayrılır. Aynı zamanda oluşturulan yemek-içecek veri seti ile mekânın beğenilip beğenilmediğinin yanında en çok hangi yemek-içeceğin beğenildiği bulunarak kullanıcıya gösterilir.

Çalışmanın devam ettirilmesi durumunda geliştirilen algoritmanın farklı alanlarda da başarılı sonuçlar verecek şekilde, alan bağımsız çalışması sağlanabilir. Bunun yanında Türkçe dilinin yapısından dolayı duygu analizi yapılırken cümleleri ayrıştırmak için kullanılan yöntem basit dil bilgisi kuralları geliştirilerek değiştirilebilir. Sosyal ağlara özel olarak etiketlenmiş geniş kapsamlı veri listesi oluşturularak analiz için farklı yöntemler geliştirilebilir. SBKA'nın uygulanması, test edilmesi ve sonuçların gösterilmesi için web tabanlı bir uygulaması geliştirilmiştir. Buna ek olarak geliştirilen yöntem mobil cihazlarda çalışacak bir uygulamaya dönüştürülebilir.

Teşekkür

4853-YL1-17 numaralı proje ile desteklenen bu çalışmada Süleyman Demirel Üniversitesi Bilimsel Araştırma Projeleri Yönetim Birimi'ne teşekkür ederiz.

Conflict of Interest / Çıkar Çatışması

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir.
No conflict of interest was declared by the authors.

Kaynaklar

Adalı, E., 2012. Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (2).

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R., 2011. Sentiment Analysis of Twitter Data. LSM '11 Proceedings of the Workshop on Languages in Social Media, 30-38.

Akın, A.A., Akın, M.D. (2007). Zemberek, An Open Source NLP Framework for Turkic Languages. Structure, 10, 1-5.

Aylien, Inc., 2018. Text Analysis API. Çevrimiçi (Erişim, 3 Nisan 2018): https://developer.aylien.com/text-api-demo?text=&language=en&tab=classify-taxonomy

Balahur, A., Hermida J.M., Montoyo, A., 2012. Detecting Implicit Expressions of Emotion in Text: A Comparative Analysis. Decision Support Systems, 53 (4), 742-753.

Cao, Q., Duan, W., Gan, Q., 2011. Exploring Determinants of Voting for the ''helpfulness'' of Online User Reviews: A Text Mining Approach. Decision Support Systems, 50 (2), 511-521.

Delibaş, A., 2008. Doğal Dil İşleme İle Türkçe Yazım Hatalarının Denetlenmesi. Yayınlanmış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, Türkiye.

Dolgun, M., 2006. Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları. Yayınlanmış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, Türkiye.

Dumanlı Kürkçü, D., 2017. Haber Tweetleri Üzerinden Kullanıcı Etkileşiminin Analizi, Erciyes İletişim Dergisi, 5 (1), 314-324.

Eroğul, U., 2009. Sentiment Analysis in Turkish. Published Master's Thesis, Middle East Technical University, Ankara, Turkey.

Eryiğit, G., 2014. ITU Turkish Natural Language Processing Web Service, 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014).

Feldman, R., Sanger, J., 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press, New York USA.

Google, Inc., 2018. Google Translate Uygulama Programlama Arayüzü. Çevrimiçi (Erişim, 3 Nisan 2018): http://www.google.com/translate_t?hl=en&ie=UTF8&text={0}&langpair={1}.

Go, A., Bhayani, R., Huang, L., 2009, Twitter Sentiment Classification Using Distant Supervision, CS224N Project Report, Stanford, 1 (12).

Liu , B., 2012. Sentiment Analysis and Opinion Mining. California: Morgan & Claypool Publishers, San Rafael California.

Liu, Y., Bi, J.W, Fan, Z.P., 2017. Ranking Products through Online Reviews: A Method Based on Sentiment Analysis Technique and Intuitionistic Fuzzy Set Theory. Information Fusion, 36, 149-161.

Pang, B., Lee, L., 2008. Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, 2 (1-2), 1-135.

Piryani, R., Madhavi, D., Singh, V.K., 2017. Analytical Mapping of Opinion Mining and Sentiment Analysis Research during 2000-2015, Information Processing and Management, 53 (1), 122-150.

Ptaszynski, M., Dokoshi H., Oyama, S., Rzepka, R., Kurihara, M., Araki, K., Momouchi, Y., 2013. Affect Analysis in Context of Characters in Narratives. Expert Systems with Applications, 40 (1), 168–76.

Rich, E., Knight K., 1990. Artificial Intelligence 2nd. Columbus: McGraw-Hill Higher Education, Columbus USA.

Schumaker, R.P., Jarmoszko, A.T., Labedz, C.S., 2016. Predicting wins and spread in the Premier League Using a Sentiment Analysis of Twitter, Decision Support Systems, 88, 76-84.

Vural, A.G, Cambazoglu, B.B., Senkul, P., Tokgoz, Z.O., 2013. A Framework for Sentiment Analysis in Turkish: Application to Polarity Detection of Movie Reviews in Turkish. Gelenbe E., Lent R. (Edt.), Computer and Information Sciences III, içinde (s. 437-445). London: Springer.

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., Spasic, I., 2015. The Role of Idioms in Sentiment Analysis. Expert Systems with Applications 42 (21),7375–7385.

Yıldırım, E., Çetin, F., Eryiğit, G., Temel, T., 2015. The Impact of NLP on Turkish Sentiment Analysis, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 7 (1), 43-51.

Yuksel, A.S, Tan F.G. 2018, Metin Madenciliği Teknikleri Ve Uygulaması. Çevrimiçi (Erişim, 2 Nisan 2018): http://ileriteknoloji.net