

Türkçe Metinlerde Sözlük Tabanlı Yaklaşımla Duygu Analizi

Sentiment Analysis Using Dictionary Based Approach in Turkish Texts

Önder DEMİR¹  Abdoulaye Issa BABAN CHAWAI²  Buket DOĞAN¹ 

¹Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, 34722, Göztepe Kampüsü, Kadıköy, İstanbul

²Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, 34722, Göztepe Kampüsü, Kadıköy, İstanbul

Öz

Metin madenciliği yöntemleri, yapılandırılmamış metinlerden otomatik olarak faydalı bilgilerin çıkarılmasını kolaylaştırmak için önem kazanmaktadır. Bu amaçla kullanılan metin madenciliği yöntemleri ile büyük miktardaki metin verileri, kısa zamanda ve yüksek performans ile analiz edilebilmektedir. Bu çalışmanın temel amacı, kelime sıklığı, bilgi çıkarma, sınıflandırma, veri işleme ve çıkarma gibi araçlarla Türkçe metinlerinin işlenmesi ve görselleştirilmesidir. Metin madenciliği alanlarından biri olan duygu analizi veya fikir madenciliği, görüşler, tutumlar ve duygular gibi öznel bilgilerin algılanmasını otomatikleştirmek için kullanılmaktadır. Bu tezde sunulan çalışmada, sosyal medya paylaşımları, roman, müşteri yorumları gibi Türkçe metinlerin analizine ihtiyaç duyulduğu durumlarda kullanılabilir yeni bir araç önerilmektedir. Geliştirilen uygulama esas olarak duygu analizi yaparken “kelime düzeyinde” çalışmaktadır. Buna ek olarak, duygu analizinde kullanılan sözlük sıfatlar, isimler, fiiller, zarflar, ön ekler, son ekler türünde kelimeler içermektedir. Kullanılan birinci sözlük 5.000 kelime, ikinci ve üçüncü sözlük yaklaşık 25.000 kelime içermektedir. Geliştirilen iki yöntemle üç farklı veri seti üzerinde deneyler yapılmıştır. Birinci yöntem bu veri setlerinde sırasıyla %77,14, %72,78 ve %74,17 doğruluk oranına ulaşmıştır. İkinci yöntem ise aynı veri setleri üzerinde sırasıyla %82,85, %74,92 ve %77,50 doğruluk oranına erişmiştir. Yapılan deneysel çalışmaların sonuçları incelendiğinde sistemin hatalı karar verdiği cümlelerde hatalı yazılan kelimeler, kinayeli yorumlar, art niyetli yazılmış yorumlar, bazı deyimlerin ve kelimelerin birden fazla anlamının olması, karşılaştırma içeren cümlelerin bir taraf için olumlu duygu taşıırken diğer taraf için olumsuz duygu taşıması gibi durumların hatalı tespit işleminde etkin olduğu gözlemlenmiştir.

Anahtar Kelimeler: Metin madenciliği, duygu analizi, metin sınıflandırması, makine öğrenmesi

Abstract

Text mining methods are important to facilitate the extraction of useful information from unstructured texts automatically. Text mining methods used for this purpose can analyze large amounts of text data in a short time and with high performance. The main purpose of this study is to process and visualize Turkish texts by means of word frequency, information extraction, classification, data processing and extraction. Emotional analysis or idea mining, one of the text mining fields, is used to automate the perception of subjective information such as opinions, attitudes and emotions. In the study presented in this thesis, a new tool that can be used in situations where analysis of Turkish texts such as social media shares, novels and customer comments is needed is proposed. The developed application works mainly at the word level while doing emotion analysis.

The developed sentiment analysis software system's preprocessing stage is very important. In the preprocessing phase, the text is cleared and made ready for use to reveal its properties. At this stage, words that are shorter than three letters are removed on the text, uppercase words are converted to lowercase letters, numbers are removed, punctuation marks are removed, extra spaces between words are removed, stop words are removed, stemming / lemming is performed. The document term is then converted to a document term matrix. At the end of these operations, a converted text is obtained. The purpose of this step is to format the text so that it can be processed by the next layer. In this way, many text review tasks are made easier, more effective and can be performed with commonly used tools. With the regular text obtained after the preprocessing phase, it is possible to obtain the features of the text, summarize, visualize and integrate it into effective workflows using natural language processing. Regular text data has a specific structure: each variable creates a column, each observation a row, each observational unit a table. The final component of data preparation is a table representing the frequency of terms used throughout the document. At this stage, a structured representation of data such as the document term matrix is provided first. Each row represents

a document, and each column shows the terms that appear. To analyze the frequencies of these words in the corpus, a probability status table of each word is created. The content of this table is the repeat value of the word. This table is called the Document Term Matrix (DTM). DTM table shows how many times a word is used in a text. The distribution of words in the text is used for text analysis and classification. In this study Bayes classification is used, because it is the most suitable method to classify with probability values obtained from the frequency values of words in the text using DTM tables.

The dictionary used in emotion analysis includes adjectives, nouns, verbs, adverbs, prefixes, suffixes. The first dictionary used contains 5,000 words, the second and third dictionary contains approximately 25,000 words. Experiments were performed on three different data sets with two methods developed. The first data set contains 35 positive and 35 negative comments, obtained from a website where a feature film is interpreted. The second dataset consists of reader comments on a book e-commerce site. This dataset contains 91309 comments. The third dataset contains 10,000 positive and 10,000 negative texts. The first method reached 77.14%, 72.78% and 74.17% accuracy in these data sets, respectively. The second method reached 82.85%, 74.92% and 77.50% accuracy rate on the same data sets, respectively. When the results of the experimental studies are examined, when the sentences that the system makes a wrong decision are examined, the words that are written incorrectly, allusive comments, maliciously written comments, having more than one meaning of some idioms and words, the comparison containing sentences have positive emotion for one side and negative emotion for the other side. was found to be effective in the detection process.

The developed system also visualizes the types of emotions in the text examined. There is no restriction on the length of the text that the developed system will examine. However, as the number of text or words to be analyzed increased, the processing time increased. In order to reduce the processing time, a preprocessing procedure was applied to the texts and the words and punctuation marks that do not have a feeling were removed from the vocabulary list to be examined. However, it was observed that this preprocessing did not solve all the difficulties of emotion analysis. Factors such as misspelled words, allusive interpretations, the existence of more than one meaning of some idioms and words, and cultural differences in expressions were observed as misclassified sentences in the test studies.

Keywords: Text mining, sentiment analysis, text classification, machine learning

GİRİŞ

Günümüzde, internet hem kişisel hem de profesyonel bilgi alışverişi için vazgeçilmez bir araçtır. 2019 yılında metin yazmak veya okumak için internet platformunu kullanan 4,39 milyar aktif çevrimiçi kullanıcı bulunmaktadır. Bu

rakam Ocak 2018'e göre 366 milyon (%9) artış göstermektedir [1].

İnternet; kullanıcılara, bilgi aramada ve erişiminde büyük ölçüde yardımcı olmaktadır. İnternet ortamı; bilgi almak, otel rezervasyonu yapmak, ürün satın almadan önce diğer kullanıcıların incelemelerini veya yorumları kontrol etmek, bir sinema filmi ile ilgili yorumları okumak gibi amaçlarla ürünler ile ilgili bilgi sahip olmak için kullanılan ilk araçtır.

2019 yılında 82,4 milyon nüfusa sahip Türkiye'nin nüfusunün %72'sini oluşturan 59,36 milyon internet kullanıcısı vardır [2]. Türkiye'nin yalnızca bir aylık e-ticaret kullanıcı etkinliklerine göre; internet kullanıcılarının %52'si satın almak üzere çevrimiçi(online) ürün veya hizmet aramakta ve bir perakende e-ticaret mağazasını ziyaret etmektedir. Bu kullanıcıların %43'ü de aradığı ürünle ilgili satın almayı gerçekleştirmektedir [3].

İnternet üzerinde dolaşan verilerin hızlı artışıyla metin madenciliği kullanımı da artmıştır. İnternet kullanıcılarının ürettikleri verinin %80'i yapılandırılmamış verilerden oluştuğu bilinmektedir[4]. Yapılandırılmamış veriler, yasal belgeler, tıbbi kayıtlar, web sayfaları, görüntüler, gibi veri türlerinden oluşur. Metinsel veriler kitaplar, gazeteler, makaleler, bloglar, sosyal ağlar gibi birçok farklı biçimde bulunabilmektedir. Metin madenciliği, bu yapısal olmayan (unstructured) veri kaynaklarından verilerin analizini yaparak, yapılandırılmış (structured) veri elde etmeyi amaçlamaktadır.

Metin madenciliği, doğal dil halindeki metinlerden anlamlı bilgileri toplamaya çalışan bir alandır. Belirli amaçlar için yararlı olan bilgilerin elde edilmesi için metnin analiz edilmesi süreci olarak tanımlanabilir[5]. Elde edilen bilgi ile, analiz edilen metin kaynaklarında açık olarak görülmeyen ilişkiler, hipotezler ve eğilimler tespit edilir. Metin madenciliği, karmaşık metinlerin daha kolay bir şekilde anlaşılmasını sağlar.

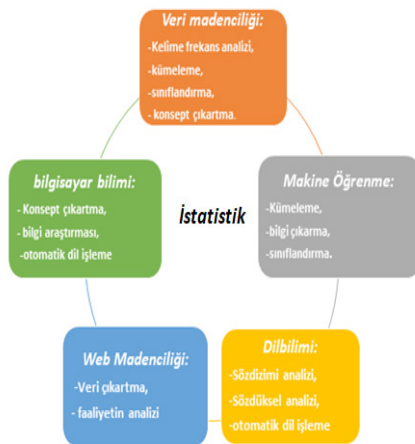
İnternet tüm dünyada, halka açık yorumların, görüşlerin ve duyguların paylaşımı için en yaygın kullanılan iletişim platformudur. Birçok internet kullanıcısı sahip olduğu ürünler, ziyaret ettiği yerler, okuduğu kitaplar veya günlük olaylar hakkında görüşlerini, düşüncelerini ve duygularını internet siteleri veya sosyal ağlar aracılığıyla paylaşmaktadır.

Duygu analizi, insanların sanal ortamlarda ürünler, hizmetler, organizasyonlar, olaylar, siyasi düşünceler, bazı konulardaki tutumları, duruşları gibi konular hakkında görüşlerini ifade ettikleri metinler içinde saklı olan duygu, fikir ve düşünceleri ortaya çıkarmayı, genel duyarlılık kutuplarını belirlemeyi amaçlamaktadır [6]. Metin madenciliğinin bir alt bölümü olan ve fikir madenciliği olarak da bilinen duygu analizi, doğal dil işleme, metin analizi ve hesaplamalı

dilbilimciliği gibi bilimlerle yakın ilişki içindedir. Duygu analizi bu bilim dallarındaki yaklaşımları da kullanarak, kaynak materyallerdeki öznel bilgiyi tanımlamakta ve ortaya çıkarmaktadır [7,8]. Duygu analizi, pazarlamadan müşteri hizmetlerine kadar çeşitli uygulamalarda yaygın olarak uygulanmaktadır. Metin madenciliğinde sunulan yöntemleri ve duygu analizinde kullanılan teknikleri aracıyla metnin ana duygusu tespit edilir. Duygu analizi metin madenciliğinde de önemli bir yere sahiptir. Metin içerisindeki anahtar kelimelere bakılarak metnin konusu belirlendikten sonra metnin “duygu haritası” çıkarılır. Bu işlem metnin okuyucuya aktarmak istediği duygunun daha kolay anlaşılmasını sağlar. Duygu analizi, genellikle duygu polaritesi ve duygu skoru olarak adlandırılan iki kavrama bağlıdır. Duygu polaritesi veya kategorisi pozitif veya negatif bir ikili değerdir [9].

Bu çalışmanın amacı metin madenciliği yöntemleriyle “Duygu Analizi” işlemini gerçekleştirmektir. Gerçekleştirilen çalışmada, Türkçe metin verileri ile metin madenciliği uygulamaları geliştirilerek, bu alanda yer alan sınırlı çalışmalara katkıda bulunmak amaçlanmıştır. Bu amaçla İngilizce ve Türkçe sözcükler içeren Afinn, Bing, Nrc ve Sentitürk sözlükleri bir araya getirilip, değerlendirilerek duygu analizindeki kutupsallıkların belirlenmesi hedeflenmiştir. Sistemin görevi, kendisine sunulan metni oluşturan sözcükleri analiz ederek, o metnin içerdiği en baskın duyguya göre veya polariteye göre sınıflandırma gerçekleştirmektir.

Metin madenciliği; Şekil 1’de görülen dilbilimi, bilgisayar bilimi, veri madenciliği, makine öğrenmesi teknikleri, web madenciliği olarak belirtilen beş alan ile yakın ilişki içerisinde [10].



Şekil 1. Metin madenciliği ile ilgili Alanlar

Dilbilim; insanın dil kullanımının bütün boyutlarını araştıran, kural koymayı değil, dilin doğasını, işleyişini doğru betimlemeyi hedefleyen, dil olgusunu bütün yönleriyle anlamaya çalışan bir alandır. Dilbilim ile ilgili temel sorular; dil bilgisi, dil becerileri, dilin üretilmesinde ve anlaşılmasında dil bilgisinin nasıl edinildiği ve konuşlandırıldığıdır. Dilbilimciler, insanların dili nasıl kullandıklarını merak eder ve matematiksel modeller kullanarak modelleştirirler [11]. Bilgisayar bilimde amaç, metnin analiz edilmesi için gereken mühendislik problemlerine ilişkin en uygun çözümleri oluşturmaktır. NLP (Natural Language Processing) çeşitli metin işleme ve yapay zeka yöntemlerini kullanarak büyük miktarda mevcut bilgiyi metin biçiminde keşfetmektir [12]. İstatistik bilimi, bilgisayar bilimlerinde uygulanan bir çok algoritma da olduğu gibi metin madenciliği algoritmalarının da merkezinde yer almaktadır. Veri madenciliği ve makine öğrenme tekniklerinde amaç az bilginin mevcut olduğu bir alanı keşfetmek veya gelecekteki gözlemleri doğru bir şekilde tahmin etmektir. Veri madenciliği, veri analiz ve karar verme sürecini ifade eder. Makine öğrenmesi, gelecekteki verilerin gözlemlenenlerden nasıl tahmin edileceğini bilmeyi amaçlamaktadır[13].

Web Madenciliği, internet kullanıcılarının davranış analizini içermektedir. Web madenciliği üç çeşit analiz içerir. İlk analiz, sitelerin yapısını web sitelerinin arasındaki bağlantı sayısına göre araştırmaktır (bağlantılar). İkinci analiz, kullanıcıların tarihçelerini anlamak için web sitelerinin kullanılmasıdır[145]. Böylece, bağlantıların sayısı ve web sitesinde nasıl gezinildiği incelenir. Üçüncü analiz ise web sitesi temalarında yer alan kelimelerin tanımlanmasına karşılık gelen içerik analizidir. İçerik, veri kaynağına bağlı olarak çeşitli şekillerde gelir (sosyal ağlar, forumlar, bloglar, siteler).

Metin madenciliği işlemlerinde metin sınıflandırma önemli bir alandır. Bu işlemin yapılabilmesi için yukarıda belirtilen beş alan içerisinden makine öğrenmesi ve bilgi arama öncelikli alanlar olacaktır. Duygu analizi yaparken bilgi arama (Information Retrieval), metin sınıflandırma (Text Categorization), makine öğrenmesi (Machine Learning) veya metin arama önemli etki alanları olmuştur.

Dehkharghani gerçekleştirdiği çalışmada, SENTITURKNET isimli bir Türkçe kelimelerin duygu kutuplarını belirten bir sözlük oluşturmuştur. SENTITURKNET sözlüğü English WordNet, SentiWordNet ve SenticNet sözlüklerinden yararlanılarak oluşturulmuştur. Kelimenin İngilizcedeki karşılığının doğal anlamını bulmak amacıyla English WordNet Gloss, kelimenin olumlu, olumsuz, tarafsız gibi kutupsal anlamı için

SentiWordNet ve duygusal anlamı için senticNet sözlükleri kullanılmıştır. Kelimelere olumlu, olumsuz ve tarafsızlık olarak toplamları 1 eden puanlar verilmiştir. Örneğin good kelimesinin olumlu puanı P:0,75, Olumsuz puanı N: 0, Tarafsız puanı O: 0,25'dir. Duygu analizi gerçekleştirebilmek için kelime düzeyinde inceleme yeterli değildir. Bu amaçla cümle, doküman bazlı inceleme ve sınıflandırma yapılmıştır. Çalışmada lojistik regresyon, yapay sinir ağları, destek vektör makineleri algoritmaları kullanılmıştır. Geliştirilen uygulama WEKA yazılımı kullanılarak Twitter üzerinden elde edilen veri seti üzerine uygulanmıştır [15].

Al-Bndi ve Salman'ın çalışmasında bir mikroblog sitesi olan twitter'dan elde edilen veriler üzerinde duygu analizi yapılmıştır. Bir öznitelik çıkarıcı ile kelimelerin frekansı bulunmuştur. Sonra öznitelikler kullanılarak tweet cümleleri sınıflandırılmıştır. Şirkete, kişiye, ürüne, filme, yerlere, etkinliğe gibi farklı farklı alana ait twitter data seti ele alınmıştır: 1000, 5000, 10000, 25000, 50000 ve 131.042 tweet cümlesinden oluşan veri setleri ile sırasıyla unigram için %70,19; %74,37; %74,93; %76,32; %76,32; %77,15 , bigram için ise %69,08; %75,20; %75,14; %74,65; %76,60; %76,88 doğruluk oranı elde edilmiştir. Kelime sayısının azlığı sınıflandırma işlemini zorlaştırdığından duygu belirten , J , L gibi simgeler de değerlendirilmiştir [17].

Yengi'nin çalışmasının amacı ise kullanıcı puanlama temelli öneri sistemlerinin, kullanıcı puanları yerine duygu analizinden elde edilen değerler ile gerçekleşmesidir. Bu çalışmada büyük veri analiz yöntemleri ile öneri sistemlerinin kişiselleştirilmesinin önemi ortaya konmuş ve bu sistemlerin temel yöntemlerine sunduğu katkı raporlanarak sonuçları analiz edilmiştir. Analiz sonuçlarına göre, verinin hacim olarak büyümesi ve kişiselleştirme adımları başarımlarını arttırmıştır. Aynı zamanda verinin hacminin büyümesinin hata değerine olan olumsuz etkisi de görülmektedir. Çalışmada sözlük tabanlı ve doğal dil işleme tabanlı duygu analizi kullanılmıştır. Geliştirilen yöntem Amazon ve Movie Lens verisetleri üzerine uygulanmıştır. Ayrıca Mahaout makine öğrenmesi ve doğal dil işleme kütüphaneleri kullanılmıştır. Her cümlelerin içerdiği duygu tespit edilip toplanarak tüm metnin duygusu analiz edilmektedir. Duygu değeri çok kötü (-2), kötü (-1), nötr (0), iyi (1), çok iyi (2) olarak beş farklı şekilde temsil edilmektedir [17].

Nassar'ın gerçekleştirdiği tez çalışmasında, duyguları sınıflandırmak için sıklıkla kullanılan makine öğrenme yöntemlerini araştırarak, Arapça belgeler üzerinde duygu analizinin performans sonuçlarının iyileştirilmesi amaçlanmıştır. Aynı zamanda ML ile modelleme sürecinde girdilerin temsilinde yeni vektör biçimleri oluşturulmuştur. İkinci olarak maksimum, ortalama ve çıkarma gibi toplama işlevlerine

dayalı farklı formüller kullanarak terim ve belge düzeylerinde sözcük tabanlı yaklaşımın uygulanması amaçlanmıştır. Arapça'da sözcüğe dayalı duygu analizinin terim ve belge seviyesi ile birlikte kullanılabilceğini göstermek için kullanılmıştır. Destek vektör makinesi, karar ağacı ve yapay sinir ağı gibi farklı makine öğrenmesi yöntemleri hem tek başlarına ve hem de kurullarla birlikte uygulanmıştır. Bu çalışmada elde edilen en başarılı sonuçlar yapay sinir ağı ile sınıflandırma gerçekleştirildiğinde elde edilmiştir. Genel olarak makine öğrenmesi yaklaşımı sonuçları sözlük kullanımını yaklaşımına göre daha iyi sonuçlar vermiştir. Çalışmada Rushdi-Saleh tarafından oluşturulan OCA veri seti kullanılmıştır. Bu veri seti yarısı olumlu, yarısı olumsuz 500 film yorumundan oluşmaktadır [18].

Türkmenoğlu tarafından gerçekleştirilen çalışmada ise, hem İngilizce hem de Türkçe için yapılan çalışmalarda kullanılan makine öğrenmesi ve sözlük tabanlı metotlara yeni özellikler eklenerek iki farklı veri seti üzerinde değerlendirme yapılmıştır. Daha önce İngilizce ve Türkçe metinler için çalışılmış metotlardan MÖ ve sözlük tabanlı DA olmak üzere iki ayrı metot Türkçe metinler için uygulanmıştır. Bu metotlar kısa ve uzun metinler olmak üzere iki farklı Türkçe veri kümesine uygulanıp başarımları ölçülmüştür. Kelimeler asıl anlamlarını köklerinde barındırdığından gereksiz ekler atılıp, asıl anlamı içeren kelime köklerine ulaşılmıştır. Varlık/yokluk (-lı,-li,-sız,-siz) eklerini ve olumsuzluk bildiren ekleri (-me,-ma) kaldırılmamış, özel bir işaret konularak muhafaza edilmiştir. Sözlük tabanlı yöntem için her kelimesi taşıdığı duygu yönelimine göre puanlandırılmış bir sözlük kullanılarak bir metnin duygu yönelimini bulmaya çalışılmıştır. Kullanılan sözlüğü oluşturmak için Sentistrength sözlüğü Türkçeye çevrilmiştir. Veri seti Twitter yorumları veri kümesi ve film yorumlarından oluşmaktadır. Twitter veri kümesine uygulandığında sözlük tabanlı metot ile %75,2, makine öğrenmesi tabanlı metot ile ise, Karar Destek Makineleri (KDM) sınıflandırıcısı kullanılarak, %85 başarı elde edilmiştir. Film yorumları veri kümesine uygulandığında ise sözlük tabanlı metot ile %79,5, MÖ tabanlı metot ile KDM sınıflandırıcısı kullanılarak %89 başarı elde edilmiştir [19].

MATERYAL ve YÖNTEM

Bu çalışmada Afinn, Bing ve NRC olmak üzere üç farklı sözlük kullanılmıştır. Bu sözlüklerin tamamı çalışma kapsamında Türkçeye çevrilmiştir.

Afinn sözlüğü 2009 ve 2011 yılları arasında Finn Årup Nielsen tarafından - 5 (negatif) ve +5 (pozitif) arasında bir tamsayı ile ve elle derecelendirilen İngilizce terimler

listesidir. Negatif değerler olumsuz duyguları, pozitif değerler ise olumlu duyguları temsil eder [20].

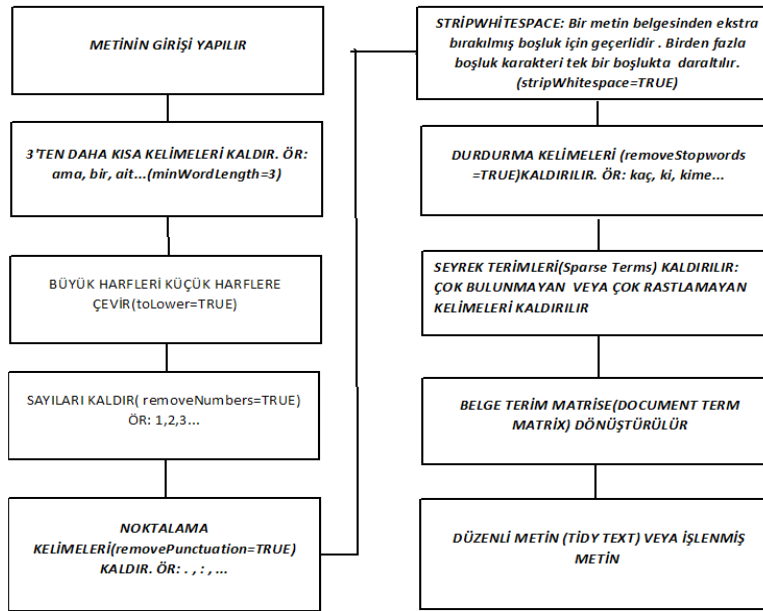
BİNG: Bing Liu ve arkadaşları tarafından oluşturulmuştur, kelimeleri ikili bir biçimde pozitif ve negatif kategoriler olarak sınıflandırır [21].

NRC: Saif Mohammad and Peter Turney tarafından oluşturulmuştur. Kelimeleri pozitif, negatif, öfke, beklenti, tiksinti, korku, sevinç, üzüntü, şaşkınlık ve güven kategorilerine sınıflandırır [22].

SentiTurkNet: Üç tane farklı sözlükten yararlanarak oluşturulmuştur. İngilizce “wordNet” içinde yaklaşık 117.000 cümle vardır. Bu sözlük bir küme olarak eşanlı terimleri gruplandırır. “sentiWordNet”, İngilizcedeki duyarluk analizi görevlerini desteklemek amacıyla oluşturulmuştur. Negatif, pozitif, nötr olarak her kelimenin 3 farklı kutup puanını verir.

2.1 Geliştirilen sistemin yapısı

Geliştirilen duygu analizi yazılım sistemi üç aşamalı çalışarak analiz işlemini gerçekleştirmektedir. Önışleme aşamasının dikkatli biçimde gerçekleştirilmesi sistemin başarısı için çok önemlidir. Çünkü eğer bu aşama doğru bir biçimde tamamlanırsa, takip eden diğer kısımlar da başarılı bir şekilde tamamlanabilecektir. Bu aşamada metnin temizlenmesi ve özelliklerinin ortaya çıkarılması için kullanıma hazır hale getirmesi sağlanmaktadır. Şekil 2’de önışleme aşamasının işlem adımları görülmektedir. Metin üzerine sırasıyla üç harften daha kısa kelimeler kaldırılır, büyük harflerden oluşan kelimeler küçük harflere çevrilir, rakamlar kaldırılır, noktalama işaretleri kaldırılır, kelimeler arasında fazladan bırakılmış boşluklar kaldırılır, durdurma kelimeleri kaldırılır, stemming/lemming yapılır. Ardından belge terim matrisine (document term matrix) dönüştürülür.



Şekil 2 Önışleme aşamasının işlem adımları

Bu işlemlerin sonunda dönüştürülmüş bir metin elde edilmektedir. Bu aşamanın amacı metinleri bir sonraki katman tarafından işlenebilecekleri şekilde biçimlendirmektir. Bu sayede birçok metin incelemesi görevi daha kolay, etkili ve yaygın kullanılan araçlarla gerçekleştirilebilir hale getirilir. Önışleme aşaması sonrasında elde edilen düzenli metin ile, metnin özneliklerinin elde edilmesi, özetlenmesi, görselleştirilmesi ve doğal dil işleme kullanarak etkin iş akışlarına entegre edilmesine mümkün olmaktadır. Düzenli metin

verileri belirli bir yapıya sahiptir: Her değişken bir sütun, her gözlem bir satır, her gözlemsel birim bir tablo oluşturur.

Veri hazırlamanın son bileşeni tüm belgede kullanılan terimlerin sıklığını temsil eden bir tablodur. Bu aşamada öncelikle belge terim matrisi gibi verilerin yapılandırılmış bir gösteriminin oluşturulması sağlanır. Her satır bir dokümanı temsil eder ve her sütun görünen terimleri gösterir. Şekil 3’te örnek bir korpus için kelimelerin dağılımı görülmektedir.

	Global occ.	Global %
bir	236	4.6283585
ve	114	2.2357325
çok	111	2.1768974
kitap	101	1.9807805
bu	76	1.4904883
ama	57	1.1178662
da	56	1.0982546
kadar	45	0.8825260
için	44	0.8629143
de	42	0.8236909
kitabı	42	0.8236909
gibi	38	0.7452442
ben	27	0.5295156

Şekil 3. Korpustaki kelimelerinin dağılımı

Bu kelimelerin korpustaki frekanslarını analiz etmek için her kelimenin bir olasılık durum tablosu oluşturulur. Bu tablonun içeriği kelimenin tekrar değeridir. Bu tablo, Belge Terim Matrisi (BTM) olarak adlandırılır. BTM tablosu, bir kelimenin bir metinde kaç kez kullanıldığını göstermektedir. Şekil 4'te örnek bir BTM tablosu görülmektedir.

Docs	ama	bir	bu	çok	da	de	için	kadar	kitap	ve
X100	0	4	2	1	0	0	0	2	5	0
X116	0	5	0	1	1	0	1	0	1	3
X123	0	4	1	1	1	5	0	0	1	2
X26	3	8	1	4	3	3	1	0	1	4
X32	2	5	3	0	1	0	0	0	0	0
X40	0	2	1	0	0	0	0	1	1	2
X66	0	4	0	0	0	1	0	2	2	1
X71	0	4	2	1	1	0	0	0	0	2
X87	0	4	2	6	0	0	1	0	1	5
X92	0	5	0	1	2	1	0	0	0	1

Şekil 4. Örnek bir BTM tablosu

Metin analizi ve sınıflandırması yapmak için metindeki kelimelerin oluşumu ve dağılımı kullanılır. Sınıflandırma algoritması olarak, metin içerisinde kelimelerin tekrar değerlerinden elde edilen olasılık değerleri ve sözlükteki duygu temsili ile sınıflandırma yapmaya en uygun yöntem olduğu için Bayes sınıflandırıcı seçilmiştir. Bu çalışmada Bayes sınıflandırıcı skorları belirlemek için üç farklı yaklaşım uygulanmıştır.

Birinci yöntem de metnin sınıflandırması öfke, nefret, korku, mutluluk, üzüntü, sürpriz olarak adlandırılan altı duyguya göre yapılmaktadır. Düzenli metin oluşturmak için metin ilk olarak matris yapısına dönüştürülür. Oluşturulan matristeki her kelime teker teker sözlükteki kelimelerle karşılaştırılır. Sözlükte bulunan kelimelerin indeksi tutulduktan sonra, her farklı duygu için kaç adet kelime var olduğunu sayılır. Her kelime için Bayes algoritması aracılığıyla duygu

skoru hesaplanır. Hangi duygunun skoru büyükse, metin o skorun kategorisine ait olacak şekilde sınıflandırılır. Şekil 5'te bu yöntem ile değerlendirilen bir metne ait çıktı görüntüsü görülmektedir.

```

Console Terminal
~/Metinmadenciligi/ >
[1] "KELİME: ağırılık KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: bağı KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: becerikli KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bekle KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bıkmak KAT: nefret SKOR: 4.95582705760126"
[1] "KELİME: brava KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bu KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bunalımlı KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: büyü KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: çekememezlik KAT: öfke SKOR: 5.97888576490112"
[1] "KELİME: çekicilik KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: te KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: in KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: in KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: yazık KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: zor KAT: üzüntü SKOR: 6.38687931936265"
OFKE NEFRET KORKU MUTLULUK ÜZÜNTÜ
[1] "7.81480342948936" "7.81480342948936" "7.81480342948936" "48.4050372901726" "39.749200
0263026"
SÜRPRİZ EN_UYGUN_DUYGU
[1] "3.31499375915909" "mutluluk"
> |

```

Şekil 5. Duygu sınıflandırma yönteminin ekran çıktısı

İkinci yöntemde sınıflandırma kutupluluğa (Olumlu veya Olumsuz) ve kategoriye (zayıf veya Güçlü) göre yapılır. Polarite sınıflandırma için önce ilk yöntemde olduğu gibi metin matrisine dönüştürülür. Özellik sözlüğündeki kelimelerle matristeki her kelime karşılaştırılır. Bulunan kelimelerin indeksi, polaritesi ve kategorisi tutulur. Eğer polaritesi güçlü (strong) ise skoruna güçlü değeri verilir aksi halde zayıf(weak) değeri verilir. Her kelime için hem kategorisine hem de polaritesine göre duygu skoru hesaplanır. Hangi kelimenin kategori skoru büyükse, metni o skorun kategorisine göre sınıflandırılır. Şekil 6'da bu yöntem ile değerlendirilen bir metne ait çıktı görüntüsü görülmektedir.

```

Console Terminal
~/Metinmadenciligi/ >
[1] "WORD: ağır CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: ağır CAT: negative POL: weaksubj SCORE: 9.5887680769643"
[1] "WORD: bağı CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: başlamak CAT: negative POL: strongsubj SCORE: 9.17492042674492"
[1] "WORD: becerikli CAT: positive POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: bekle CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: bıkmak CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: bunalımlı CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: büyü KAT: negative POL: weaksubj SCORE: 9.5887680769643"
[1] "WORD: çekememezlik CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: çekicilik CAT: negative POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: te CAT: negative POL: strongsubj SCORE: 9.17492042674492"
[1] "WORD: in CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: in CAT: negative POL: weaksubj SCORE: 9.5887680769643"
[1] "WORD: dur CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: fiyatl CAT: negative POL: weaksubj SCORE: 9.5887680769643"
[1] "WORD: karakter CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: meor CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: var CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: vermek CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: yazık CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: yok CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: zor CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "POS: 0.924955939796596 NEG: 0.51109955884509 RATIO: 1.80973730810232"
POS NEG POS/NEG BEST_FIT
[1] "0.924955939796596" "0.51109955884509" "1.80973730810232" "positive"
> |

```

Şekil 6. Polarite sınıflandırma yönteminin ekran çıktısı

Veri hazırlama aşamasından sonra tutulan kelimeler DTM tablosuna dahil edilir. Daha sonra, bu kelimelerin her birinin sıklığı hesaplanır ve azalan düzende sıralanırlar. Tablo 1'de örnek bir kelime frekans tablosu görülmektedir.

Tablo 1. Örnek Frekans Tablosu

Kelime	Frekans
1 otel	239

2	kitap	109
3	gün	88
4	değil	87
5	yemek	80
6	aynı	65
7	kötü	64
8	güzel	60
9	oda	60

Metin verilerini görselleştirmenin en kolay ve en yaygın yolu kelime bulutlarıdır. Amaç, her bir kelimeyi görüntülemektir. Bu gösterimi kelimenin metin içerisindeki sıklığı o kelime için kullanılan fontun büyüklüğünü etkilemektedir. En sık görülen kelime, en büyük yazı tipiyle gösterilir. Bu kelimeler, metinde ifade edilen ana düşüncüyü veya konuyu yansıtmaktadır.

Şekil 7’de sıklıkları verilen kelimelerden oluşan kelime bulutu görülmektedir.

	word	freq	V3
<u>kitap</u>	kitap	100	2.82
<u>kitabı</u>	kitabı	43	1.21
<u>değil</u>	değil	26	0.73
<u>okudum</u>	okudum	26	0.73
<u>kitabın</u>	kitabın	23	0.65
<u>hayal</u>	hayal	20	0.56
<u>fazla</u>	fazla	17	0.48
<u>okudugum</u>	okudugum	16	0.45
<u>sıkıcı</u>	sıkıcı	16	0.45
<u>begenmedim</u>	begenmedim	15	0.42
<u>güzel</u>	güzel	15	0.42
<u>merak</u>	merak	14	0.40
<u>oldugu</u>	oldugu	14	0.40
<u>vazarın</u>	vazarın	14	0.40
<u>üzerin</u>	üzerin	13	0.37
<u>okurken</u>	okurken	12	0.34
<u>oldugunu</u>	oldugunu	12	0.34
<u>olması</u>	olması	12	0.34
<u>acıkkası</u>	acıkkası	11	0.31
<u>akıcı</u>	akıcı	11	0.31
<u>kitapta</u>	kitapta	11	0.31
<u>tavsiy</u>	tavsiy	11	0.31
<u>yazar</u>	yazar	11	0.31
<u>yer</u>	yer	11	0.31
<u>aldım</u>	aldım	10	0.28

Şekil 7. Kelime sıklıkları ve kelime bulutları

Şekil 7’deki kelime bulutu incelendiğinde “kitap”, “kitabı” “okudum” ve “değil” frekans tablosundaki mevcut verileri doğrulayan 4 kelime olduğu görülmektedir. Diğer önemli kelimeler ise “begenmedim”, “hayal”, “sıkıcı”, “merak” ve “güzel” olup kitabın karakteristiğine odaklandığını göstermektedir. Buradan da kitapla ilgili yorumların çoğunun olumsuz olduğunu anlaşılabilir.

BULGULAR VE TARTIŞMALAR

Geliştirilen sistemle farklı veri setleri kullanılarak deneyler yapılmıştır. İlk veri setinde bir sinema filmlerinin yorumlandığı bir web sitesinden elde edilen 35 pozitif, 35 negatif

yorum yer almaktadır [23]. İkinci veriseti bir kitap e-ticaret sitesinde yer alan okuyucu yorumlarından oluşmaktadır. Bu veriseti 91309 yorum içermektedir [24]. Üçüncü verisetinde ise 10.000 olumlu 10.000 olumsuz metin yer almaktadır [25].

Veriseti 1 üzerinde yapılan deneylerde yöntem 1 kullanılarak 35 negatif yorumun 27’si negatif polariteli olduğu sistem tarafından tespit edilmiştir. Sekiz adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 31 tanesi negatif 4 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür. Yöntem 1’in toplam doğruluk oranı %77,14 iken yöntem 2’nin toplam doğruluk oranı %82,85 olmuştur. Deneylerle ilgili veriler Tablo 2’de sunulmuştur.

Tablo 2. Veriseti 1 ile elde edilen sonuçlar

Veriseti 1	Yöntem 1	Doğruluk Oranı	Yöntem 2	Doğruluk Oranı
35 Negatif Metin	27 Negatif ve 8 Pozitif	%77,14	31 Negatif ve 4 pozitif	%88,57
35 Pozitif Metin	27 Pozitif ve 8 Negatif	%77,14	27 Pozitif ve 8 Negatif	%77,14

İkinci verisetinde 700 adet negatif, 700 adet pozitif cümle yer almaktadır. Bu veriseti üzerinde yapılan deneylerde yöntem 1 kullanılarak 700 negatif yorumun 491’i negatif polariteli olduğu sistem tarafından tespit edilmiştir. 210 adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 525 tanesi negatif 175 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür.

Pozitif yorum cümleleri üzerinde yapılan deneylerde yöntem 1 700 cümlelerin 528’ini pozitif, yöntem 2 ise 524 tanesini pozitif olarak değerlendirmiştir. Yöntem 1’in toplam doğruluk oranı %72,78,14 iken yöntem 2’nin toplam doğruluk oranı %74,92 olmuştur. Deneylerle ilgili veriler tablo 3’te sunulmuştur.

Tablo 3. Veriseti 2 ile elde edilen sonuçlar

Veriseti2	Yöntem 1	Doğruluk Oranı	Yöntem 2	Doğruluk Oranı
700 Negatif Metin	491 Negatif ve 209 Pozitif	%70,14	525 Negatif ve 175 pozitif	%75
700 Pozitif Metin	528 Pozitif ve 172 Negatif	%75,42	524 Pozitif ve 176 Negatif	%74,85

10.329 negatif cümle ve 10.833 pozitif cümleleri içermektedir. Bu veri seti daha iyi değerlendirebilmek için 20 tane pozitif data seti ve 20 tane negatif alt veri setine bölünerek deneyler gerçekleştirilmiştir.

Bu veriseti üzerinde yapılan deneylerde yöntem 1 kullanılarak 10.329 negatif yorumun 7.423'si negatif polariteli olduğu sistem tarafından tespit edilmiştir. 2.906 adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 7.773 tanesi negatif 2.556 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür.

pozitif yorum cümleleri üzerinde yapılan deneylerde yöntem 1 10.833 cümlelerin 8.273 tanesini pozitif olarak nitelendirmiştir. 2.560 cümle ise negatif olarak nitelendirilmiştir. yöntem 2 10.833 cümlelerin 8.628 tanesini pozitif olarak nitelendirmiştir. 2.205 cümle ise negatif olarak nitelendirilmiştir

Yöntem 1'in toplam doğruluk oranı %74.11 iken yöntem 2'nin toplam doğruluk oranı %77,445 olmuştur. Deneylerle ilgili veriler tablo 4'te sunulmuştur.

Tablo 4. Veriseti 3 ile elde edilen sonuçlar

Veriseti 3	Yöntem1	Doğruluk Oranı	Yöntem2	Doğruluk Oranı
10.833 Cümle	Pozitif: 8.273 Negatif: 2.560	%76,36	Pozitif: 8.628 Negatif: 2.205	%79,64
10.329 negatif cümle	Negatif: 7.423 Pozitif: 2.906	%71,86	Negatif: 7.773 Pozitif: 2.556	%75,25

SONUÇLAR

Bu çalışmada metin madenciliği ile bir duygu analizi sistemi geliştirilmiştir. Çeşitli veri setleri üzerinde yapılan testlerle sistemin performansı ölçülmüştür. Geliştirilen sistem ayrıca incelediği metindeki duygu türlerini görselleştirmektedir. Geliştirilen sistemin inceleyeceği metin uzunluğu ile ilgili bir kısıtı bulunmamaktadır. Fakat analiz edilecek metin veya kelime sayısı arttıkça, işleme süresinin arttığı görülmüştür. İşlem süresini azaltmak için metinler üzerinde bir önışleme prosedürü uygulanmış ve bir duygu taşımayan kelimeler ve noktalama işaretlerinin incelenecek kelime listesinden çıkarılması sağlanmıştır. Ancak bu önışlemenin de duygu analizinin tüm zorluklarını gideremediği gözlenmiştir. Hatalı yazılan kelimeler, kinayeli yorumlar, art niyetli yazılmış yorumlar, bazı deyimlerin ve kelimelerin birden fazla anlamının olması, karşılaştırma içeren cümlelerin bir taraf için olumlu duygu taşıırken diğer taraf için olumsuz duygu taşıması, yorum yazarının kültür farklılığı gibi durumlar duygu analizini zorlaştırmaktadır. Yapılan test çalışmalarında sistemin hatalı karar verdiği cümleler incelendiğinde bu gibi durumlarla karşılaşılmıştır.

Bu alandaki gelecek çalışmalarda negatif anlam içeren cümlelerin analizi üzerinde daha fazla çalışılmalıdır. Türkçe'de olumsuzluk anlamı veren bir çok kelime vardır. Bu

kelimeler tek başına olumlu görülen bir cümleyi olumsuz hale getirebilmektedirler. Bu da pek çok duygu analizi sisteminin başarısını etkilemektedir. Fiillerin belirttiği duygular üzerine daha çok çalışma yapılmalıdır. Sıfat ve zarflar üzerine yoğunluklu yapılan çalışmalar çok yüksek başarıya erişmemektedir. Türkçe için geliştirilen sözlük tabanlı çalışmaların başarısının artması için mecaz anlamları da içeren sözlüklerin üretilmesi de gereklidir.

KAYNAKLAR

- [1] Simon Kemp, report author , Digital around the world in 2019, the essential line data you need to understand global mobile, internet, and social media use P7. <https://wearesocial.com/global-digital-report-2019>.
- [2] Dr. Cebrail Taşkın, MBA Chief Technology Officer (CTO) & Chief Digital Officer, YENİ DEĞER: BÜYÜK VERİ, 2017 https://www.linkedin.com/pulse/yeni-de%C4%9Fer-b%C3%BCy%C3%BCK-veri-dr-cebrail-ta%C5%9Fk%C4%B1n-mba?articleId=632.862.5124677341184#comments-632.862.5124677341184&trk=public_profile_post
- [3] Simon Kemp, report author, Digital around the world in 2019, the essential line data you need to understand global mobile, internet, and social media use, we are social 2019 report for Turkey <https://wearesocial.com/global-digital-report-2019>
- [4] www.marketingturkiye.com.tr, haberler, 2017 rakamlar, Türkiye'de e-ticaret istatistikleri <https://www.marketingturkiye.com.tr/haberler/rakamlarla-2017-turkiye-e-ticaret-istatistikleri/>
- [5] Text mining, Ian H. Witten Computer Science, University of Waikato, Hamilton, New Zealand.
- [6] Akcayol, M & Özyurt, Barış. (2018). Fikir Madenciliği Ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. Selcuk University Journal of Engineering ,Science and Technology. 6. 668-693. 10.15317/Scitech.2018.160.
- [7] Cothrel J, Williams RL. On-line communities: helping them form and grow. J Knowl Manag. 1999 Mar; 3(1): 54-60
- [8] Hsiung RC. Suggested principles of Professional ethics for the online provision of mental health services. Stud Health Technol Inform. 2001;84(Pt 2): 1296-300
- [9] Liu, B., Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
- [10] Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text mining: techniques, applications and issues. International Journal of Advanced Computer Science and Applications, 7(11), 414-418.
- [11] Stubbs M. Text and corpus analysis: Computer-assisted studies of language and culture. Blackwell Oxford; 1996.
- [12] Roche E, Schabes Y. Finite-state Language Processing. 1997. (MIT Press).
- [13] <https://www.quora.com/What-is-sentiment-analysis-and-real-life-examples>, Wonderflow BV Panamalaan 6h

- [14] O’Dea B, Wan S, Batterham PJ, Callear AL, Paris C, Christensen H. Detecting suicidality on Twitter. *Internet Interv* 2015 May;2(2): 183-8.
- [15] Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), 667-685.
- [16] Al-Bnd, A., & Salman, M. (2015). Sentiment analysis and opinion mining via microblogging in social media like: twitter.
- [17] Yengi, Y., & Omurca, S. İ. (2016). Distributed Recommender Systems with Sentiment Analysis. *Avrupa Bilim ve Teknoloji Dergisi*, 4(7).
- [18] Nassar, A. (2017). Multilevel Sentiment Analysis In Arabic. Hacettepe University/ Graduate School of Science and Engineering/ Department of Computer Engineering.
- [19] Türkmenoğlu, C. (2015). Türkçe Metinlerde Duygu Analizi (Doctoral dissertation, Fen Bilimleri Enstitüsü). İstanbul Teknik Üniversitesi/ Fen Bilimleri Enstitüsü/ Bilgisayar Mühendisliği Anabilim Dalı.
- [20] Nielsen F (2011) A new anew: evaluation of a word list for sentiment analysis in microblogs. In The ESQ2011 Workshop on “Making Sense of Microposts”.
- [21] Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.
- [22] Mohammad, S. M., & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.
- [23] “Türkçe Metinlerin Siniflandırılmasında Metin Temsil Yöntemlerinin Performans Karşılaştırılması”, M.Fatih AMASYALI, Sümeyra BALCI, Esra Nur VARLI, Emrah METE ,*EMO Bilimsel Dergi*, Cilt 2, Sayı 4, 95-104, 2012.
- [24] “Türkçe Sıfat Sözlüğü ile Duygu Analizi”, İslam Mayda, Mehmet Korkmaz, Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU), 4-6 Ekim 2018, Adana.
- [25] Naderalvojud, B., Ucan, A., & Sezer, E. A. (2018), “HUMIR at IEST-2018: Lexicon-Sensitive and Left-Right Context-Sensitive BiLSTM for Implicit Emotion Recognition”, In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 182-188)