



Farklı dillerdeki metinler üzerinde yapılan duygu analizi çalışmalarının incelenmesi

Analysis of sentiment analysis studies on texts in different languages

Seda Kılıçer^{1,2*}, Ruya Samli²¹ Beykent Üniversitesi, Bilgisayar Mühendisliği Bölümü, sedakilicer@beykent.edu.tr

ORCID: 0000-0002-3464-6219

² İstanbul Üniversitesi Cerrahpaşa, Bilgisayar Mühendisliği Bölümü, ruyasamli@istanbul.edu.tr

ORCID: 0000-0002-8723-1228

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 13 Temmuz 2022
Revizyon 24 Ağustos 2022
Kabul 15 Eylül 2022
Online 30 Eylül 2022

Anahtar Kelimeler:

Duygu analizi,
Sözlük tabanlı yaklaşım,
Makine öğrenmesi yaklaşımı,
Hibrit yaklaşım,
Sınıflandırma algoritmaları.

ÖZ

Teknolojinin gelişmesiyle birlikte veri miktarı hızlı bir şekilde artmaktadır, sürekli artan bu verilerden faydalı bilgilerin bulunması, ihtiyaçlarımız ile ilgili çözümlerin manuel bir şekilde veriler arasında tespit edilebilmesi pek mümkün olmayacaktır. Bunun için veriler işlenmediği sürece anlamlı analizlerin yapılarak faydalı bilgilerin elde edilmesi çok zor olacaktır. Bu sebeple bu büyük verilerin sistemler aracılığıyla hızlı bir şekilde bilgiye dönüştürülmesi ve insanlar tarafından kullanılabilir hale dönüştürülebilmesi hedeflenmektedir.

Duygu analizi, bir metnin olumlu, olumsuz veya nötr gibi duygulardan hangisine daha yakın olduğunu sonucunun farklı yöntemler kullanılarak elde edilmesidir. Bu çalışmada son on yılda (2013-2022) yapılan duygu analizi çalışmaları ele alınmış, değerlendirilmiş ve sınıflandırılmıştır. Bu amaçla sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve hibrit yaklaşım yöntemleri kullanılarak duygu analizi yapılan 68 çalışma incelenmiştir. Çalışmalar ilk olarak duygu analizinde kullanılan yöntemlere göre, sonrasında verilerin elde edilme şekillerine ve incelenen metnin diline göre sınıflandırılmıştır. Bunlar dışında duygu analizinde kullanılan yöntemlere göre uygulanan adımlar detaylı olarak incelenmiştir. Duygu analizi çalışmalarında elde edilen başarı oranlarının artırılabilmesi için incelenen çalışmalar tablolar aracılığıyla karşılaştırılmıştır. Her yöntem kendi içerisinde değerlendirilerek daha başarılı sonuçlar elde edilebilmesi için incelenen kriterler ve elde edilen sonuçlar arasındaki bağlantılar araştırılmıştır.

ARTICLE INFO

Article history:

Received 13 July 2022
Received in revised form 24 August 2022
Accepted 15 September 2022
Available online 30 September 2022

Keywords:

Sentiment analysis,
Dictionary-based approach,
Machine learning approach,
Hybrid approach,
Classification algorithms.

ABSTRACT

With the development of technology, the amount of data is increasing rapidly, it will not be possible to find useful information from this ever-increasing data and to determine the solutions related to our needs manually among the data. As long as the data is not processed for this purpose, it will be very difficult to obtain useful information by making meaningful analyzes. For this reason, it is aimed to transform this big data into information quickly through systems and to make it usable by people.

Sentiment analysis is the use of different methods to obtain the result of which of the emotions a text is closer to, such as positive, negative or neutral. In this study, sentiment analysis studies conducted in the last ten years (2013-2022) were discussed, evaluated and classified. For this purpose, 68 studies in which sentiment analysis was performed using dictionary-based approach, machine learning approach and hybrid approach methods were examined. Studies were first classified according to the methods used in sentiment analysis, and then according to the way the data were obtained and the language of the text examined. Apart from these, the steps applied according to the methods used in sentiment analysis were examined in detail. In order to increase the success rates obtained in sentiment analysis studies, the studies examined were compared through tables. Each method is evaluated within itself and the connections between the criteria examined and the results obtained are investigated in order to obtain more successful results.

Doi: 10.24012/dumf.1143479

* Sorumlu Yazar

Giriş

İnsanlar çoğu zaman diğer insanların fikirlerine ihtiyaç duymaktadır. Bir konu hakkında tercih yapmaları gerektiği zamanlarda, bir ürün veya mekan hakkında karar vermeleri gerektiği durumlarda diğer insanların deneyimleri merak edilmektedir. Bunun için insanların duygu ve düşüncelerini paylaşacakları çeşitli imkanlar verilmiştir. Farklı kategorilere göre farklı platformlarda çok sayıda yorum bulunabilir. Ancak bu yorumların tek tek okunması ve bir sonuca varılması çok zor olacağı için bu işlemin sistemler tarafından hızlıca yapılması beklenmektedir. Birçok yorumun hızlıca değerlendirilebilmesi için metinlerin analiz edilerek bir sonuç elde edilebilmesi için metinler üzerinde duygu analizi yapılmaktadır.

Duygu analizi işlemi, sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve hibrit yaklaşım olmak üzere 3 farklı şekilde yapılabilir. Sözlük tabanlı yaklaşım genel olarak veri miktarının az olduğu durumlarda tercih edilen bir yöntemdir. Çünkü incelenecek kelimeler ve her kelimenin ağırlığını içeren bir sözlük kullanılarak olumlu, olumsuz ve nötr şeklinde sonuçların elde edilmesi hedeflenmektedir. Makine öğrenmesi yaklaşımı, metinler üzerinde çeşitli sınıflandırma algoritmaları kullanılarak duygu analizi yapılmasıdır. Hibrit yaklaşım, sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımını birlikte kullanılarak metin üzerinde duygu analizi yapılmasıdır. Ancak hibrit yaklaşım olarak değerlendirilen bazı çalışmalarda sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımları ayrı ayrı uygulanarak elde edilen sonuçlar karşılaştırılmaktadır. Burada incelenen çalışmalarda her iki şekildeki çalışmalardan da bahsedilmiştir.

Metinler üzerinde yapılan duygu analizi çalışmaları incelendiğinde verilerin farklı yöntemlerle elde edildiği, farklı dille yazılan metinlerin kullanıldığı ve sonuçların elde edilebilmesi için farklı yaklaşımların kullanıldığı gözlenmiştir. Bu nedenle incelenen çalışmalar bu kriterler dikkate alınarak incelenmiştir.

2. Sözlük tabanlı yaklaşım

Metinler üzerinde yapılan duygu analizi çalışmaları detaylı olarak incelenirken ilk olarak sözlük tabanlı yaklaşım kullanılarak yürütülen çalışmalar incelenmiştir. Burada 2016-2022 yılları arasında yapılmış toplam 13 çalışma (7'si Türkçe, 6'sı farklı diller) ele alınmıştır.

Sözlük tabanlı yaklaşım kullanılarak yapılan duygu analizi çalışmalarında ilk olarak Twitter sosyal medya sitesinden elde edilen Türkçe metinler üzerine yapılan çalışmalar incelenmiştir. Twitter sosyal medya sitesi insanların sınırlı sayıda harf ve karakter kullanarak duygu ve düşüncelerini paylaştıkları bir paylaşım sitesidir. Bu platformda insanlar etiket kelimeler kullanarak aynı konu hakkında toplu olarak duygu ve düşüncelerini herkese açık bir şekilde paylaşabilmektedir. Böylece bir konu hakkında farklı insanlar tarafından belirtilen farklı görüşler ile ilgili bir duygu veya düşünce kümesi oluşturulmaktadır. Bu etiketler sayesinde Twitter sosyal medya sitesi üzerinden bir konu hakkında hızlı bir şekilde çok fazla veri elde edilebilmektedir. Ayrıca Twitter sosyal medya sitesinden Twitter uygulama programlama arayüzü (Application Programming Interface -API) kullanılarak kolay ve hızlı bir şekilde veri elde edilebilmektedir. Bunun için Twitter sosyal

medya sitesi duygu analizi çalışmalarında sıklıkla tercih edilmiştir. Bu çalışmalarda kullanılan veri kümeleri, sözlükler ve çalışma sonucunda elde edilen sonuçlar aşağıda detaylı olarak anlatılmıştır.

2016 yılında yapılan [1] referanslı çalışmada kullanılan veriler Twitter ortamından elde edilmiştir. Bu çalışmada elle etiketleme yapılarak oluşturulan 500, 1.200 ve 5.100 adet tweet içeren üç farklı veri seti oluşturulmuştur. Türk Dil Kurumu'na ait verilerden faydalanılarak elde edilen sözlük kullanılarak duygu analizi yapılmış ve elde edilen sonuçların kıyaslanabilmesi için F-skoru tercih edilmiştir. Çalışma sonucunda sözlük tabanlı yöntem ile en yüksek F-skor değerleri sırasıyla %70, %71 ve %72 olarak elde edilmiştir. 2019 yılında yapılan [2] referanslı çalışmada Twitter üzerinden Aziz SANCAR ile ilgili 300 tweet, Beşiktaş ile ilgili 364 tweet ve Arçelik ile ilgili 537 tweetten oluşan 3 farklı veri setleri üzerinde sözlük tabanlı yaklaşım kullanılarak duygu analizi yapılmış ve en yüksek doğruluk oranı Aziz SANCAR konusu ile ilgili veri seti ile %88,2 olarak elde edilmiştir. 2019 yılında yapılan [3] referanslı çalışmada hava durumu ve çok izlenen Survivor programı ile ilgili Twitter verileri üzerinde inceleme yapılmıştır. Duygu kütüphanesi oluşturularak veriler üzerinde duygu analizi yapılmış ve beklenen değerler ile elde edilen sonuçların uyumlu olduğu gözlenmiştir. 2021 yılında yapılan [4] referanslı çalışmada kullanılan veriler sözlük tabanlı yaklaşımla ve insanlar tarafından sınıflandırılarak sistemin performansı karşılaştırılmıştır. Sistem ile insanların sınıflandırmasını karşılaştırabilmek için Twitter üzerinden elde edilen veriler kullanılmış ve sözlük olarak da SentiTurkNet sözlüğü kullanılmıştır. Duygu ifadesi insanlar tarafından net bir şekilde anlaşılan cümlelerde başarılı olduğu ancak negatif ve yoruma açık olan cümlelerin duygu analizinde yeterince başarılı sonuçlar elde edilemediği gözlenmiştir.

Sözlük tabanlı yaklaşım ile hazır veri setleri veya farklı web sitelerinden elde edilen Türkçe metinler üzerinde yapılan duygu analizi çalışmaları aşağıda incelenmiştir.

2017 yılında yapılan [5] referanslı çalışmada film yorumlarından oluşan 14.949 cümle WordNet duygu sözlüğü kullanılarak incelenmiş ve film yorumları veri setinde en yüksek doğruluk % 55 olarak elde edilmiştir. 2019 yılında yapılan [6] referanslı çalışmada dengeli ve farklı sayılarda sinema film yorumları,e-ticaret sitesindeki kitap yorumları ve metinden oluşan 3 farklı veri seti kullanılmıştır. Afinn, Bing, Nrc, SentiTurk sözlükleri kullanılarak kutupluluğa (Olumlu veya Olumsuz) ve kategoriye (Zayıf veya Güçlü) göre duygu analizi sonucunda sırasıyla %82,85, %74,92 ve %77,50 doğruluk oranları elde edilmiştir. 2021 yılında yapılan [7] referanslı çalışmada Yükseköğretim kurumundaki Türkçe tezler taranarak elde edilen veriler manuel olarak oluşturulan duygu sözlüğü kullanılarak incelenmiştir. Duygu sözcükleri el ile etiketlendikten sonra verilerin %10'u 10 kişi tarafından tekrar etiketlenmiş ve %95 oranında başarılı olduğu gözlenmiştir.

Sözlük tabanlı yaklaşım kullanılarak Türkçe olmayan metinler üzerinde de duygu analizi çalışmaları yapılmıştır. Aşağıda incelenen çalışmalarda kullanılan veriler Twitter sosyal medya sitesinden veya farklı web sitelerden elde edilmiştir.

Twitter verileri kullanılarak yapılmış olan duygu analizi çalışmalarından biri 2018 yılında [8] referanslı çalışmada

Türkçe ve İngilizce tweetler toplanmıştır. İngilizce tweetlerin duygu analizi için Rsentiment içerisindeki duygu sözlüğü kullanılırken Türkçe tweetlerin analiz edilebilmesi için yeni bir duygu sözlüğü geliştirilmiş ve geliştirilen Türkçe sözlük ile manuel değerlendirmenin tutarlı olduğu sonucu elde edilmiştir. 2019 yılında yapılan [9] referanslı çalışmada Sentiment.140 isimli sitedeki Twitter'dan elde edilen veriler kullanılmıştır. Bu çalışmada SentiWordNet sözlüğü kullanılmış ve en yüksek doğruluk %74,80 olarak elde edilmiştir. 2022 yılında yapılan [10] referanslı çalışmada Twitter sosyal medya sitesinden elde edilen 110.883 tweet kullanılmış ve Bing sözlüğü kullanılarak duygu analizi sonucunda salgın ile ilgili paylaşılan metinlerin olumlu bir yaklaşıma sahip olduğu sonucu elde edilmiştir. 2022 yılında yapılan [11] referanslı çalışmada Twitter sosyal medya sitesinden elde edilen Sinovac ve BionTech aşılı ile ilgili 21.676 tweet incelenmiştir. Burada NRC (National Research Council of Canada - Kanada Ulusal Araştırma Konseyi) sözlüğü kullanılmıştır. Kelime düzeyinde yapılan analiz sonucunda Biontech marka aşısı ile ilgili pozitif yorumların daha çok olduğu ve her iki marka için de tarafsız yorumların benzer oranlarda olduğu gözlenmiş, cümle düzeyinde yapılan analiz sonucunda ise Sinovac marka aşısı ile ilgili daha olumlu paylaşımlar yapıldığı sonucu elde edilmiştir. Yine kullanıcıların markalara göre duygularının analizinde de Sinovac marka aşısı daha çok güven duyulduğu sonucu elde edilmiştir. Twitter dışında farklı yollardan elde edilen veriler kullanılarak sözlük tabanlı yaklaşımla duygu analizi çalışmaları incelendiğinde 2017 yılında yapılan [12] referanslı çalışmada seyahat platformu olan Ly.com'dan elde edilen 30.000 yorum incelenmiştir. Çince kelimelerde oluşan Affective Lexicon Ontology sözlüğü kullanılmış ve duygu analizi sonucu turistik yerlerin sıralaması ile sitedeki sıralamaların tutarlı olduğu sonucu elde edilmiştir 2018 yılında [13] referanslı çalışmada Amazon ürün incelemelerinden ve SemEval-2007'deki haber başlıklarını içeren veriler kullanılmıştır. WKWSCI (Wee Kim Wee School of Communication and Information-Wee Kim Wee İletişim ve Bilgi Okulu) duygu sözlüğü tanıtılmıştır. Daha sonra bu sözlük ile elde edilen sonuçlar ve Hu & Liu Fikir Sözlüğü, MPQA (Multi-perspective Question Answering Çok Perspektifli Soru Cevaplama) Öznellik Sözlüğü, Genel Sorgulayıcı, NRC(National Research Council Canada - Ulusal Araştırma Konseyi Kanada) sözlük duyarlılığı sözlüğü, SO-CAL (Semantic Orientation Calculator - Anlamsal Yönelim Hesaplayıcısı) sözlükleri ile elde edilen sonuçlar karşılaştırılmıştır. Ürün incelemelerinin belge düzeyinde duyarlılık ile duygu analizinde en yüksek doğruluk oranı Hu & Liu sözlüğü ile %75 olarak elde edilirken haber başlıklarından oluşan verilerde duygu analizinde en yüksek doğruluk oranı WKWSCI sözlüğü ile %69 olarak elde edilmiştir. Sözlük tabanlı yaklaşım kullanılarak incelenmiş çalışmalarda kullanılan metnin dili, kullanılan veri seti, kullanılan sözlük gibi kriterlerin elde edilen sonuçlar üzerindeki etkisi merak edilmiştir. Bu amaçla Tablo 1'de sözlük tabanlı yaklaşım kullanılarak yapılan duygu analizi çalışmalarında incelenen dil, veri seti, kullanılan sözlük ve bu çalışmalarda elde edilen sonuçlar arasındaki ilişkiler incelenmiştir.

Tablo 1. Sözlük tabanlı yaklaşımda farklı kriterlerin karşılaştırılması.

| Referans | Dil | Veri Seti | Sözlükler | Sonuçlar |
|----------|------------------|---|--------------------------------------|---|
| [1] | Türkçe | Twitter: VS1, VS2, VS3 veri setleri | Yeni sözlük | En yüksek F-skoru: • VS3 veri seti ile %72 olarak elde edilmiştir. |
| [2] | Türkçe | Twitter: •Aziz Sancar •Beşiktaş •Arçelik veri setleri | Yeni sözlük | En yüksek doğruluk oranı: • Aziz Sancar veri seti ile %88,2 olarak elde edilmiştir. |
| [3] | Türkçe | Twitter: •Mevsimleri n etkisi için genel tweetler •Survivor isimli program ile ilgili tweetler | Türkçe duygu kütüphanesi | Duygu analizi sonrasında karşılaşılan sonuçların beklenen değerlerle uyumlu olduğu gözlemlenmiştir. |
| [4] | Türkçe | 50 cümle | SentiTurk Net | •Yorumla açık cümleler için başarılı sonuçlar elde edilmiştir. •Negatif cümle analizi konusunda ise uygulamanın yeteriz kaldığı gözlenmiştir. |
| [5] | Türkçe | Olumlu ve olumsuz olarak etiketlenmiş film yorumlarınd an oluşan 14.949 cümle | Türkçeye çevrilen WordNet | •En yüksek doğruluk oranı: %55 olarak elde edilmiştir. |
| [6] | Türkçe | •Filmler ile ilgili yorumlar: 35 pozitif, 35 negatif, •Ticaret verileri ile ilgili: 700 pozitif 700 negatif, •Metinlerden elde edilen veriler: 10.000 olumlu 10.000 olumsuz | Farklı sayıda kelime içeren 3 sözlük | • Olumlu/Olumsuz, Zayıf/Güçlü özelliklerine göre incelendiğinde en yüksek belirtilen sırayla doğruluk oranları: %82,85, %74,92 ve %77,50 olarak elde edilmiştir. |
| [7] | Türkçe | Yüksek Öğrenim Kurumu web sitesinden indirilen 113 tez çalışması | Yeni marka imaj sözlüğü | Rastgele seçilen %10'luk bir kısmın 10 kişiye tekrar etiketlenilmesi sonucunda, %95 tutarlı bir sözlük oluşturulmuştur. |
| [8] | Türkçe/İngilizce | Twitter: •Türkçe 97.850 tweet •İngilizce 250.857 tweet | •Rsentiment •Yeni Türkçe sözlük | •Türkçe tweetler genel olarak olumlu, olumsuz nötr arasında eşit dağıldığı sonucu elde edilmiştir. •İngilizce tweetler nötr ve olumsuz arasında dağıldığı sonucu elde edilmiştir. |
| [9] | Türkçe Olmayan | Sentiment.140 veri | •SentiWordNet | •SentiWordNet sözlüğü |

| | | | | |
|------|----------------|--|---|--|
| | | setinden elde edilen veriler | •Domain Focused sözlüğü | kullanılarak %68,20 doğruluk elde edilmiştir. •Hibrit sözlük kullanılarak %74,80'lik bir hassasiyet elde edilmiştir. Salgın ile ilgili yapılan paylaşımların olumlu yaklaşıma sahip olduğu gözlenmiştir. |
| [10] | Türkçe Olmayan | 110.883 Twitter verisi | Bing sözlüğü | Salgın ile ilgili yapılan paylaşımların olumlu yaklaşıma sahip olduğu gözlenmiştir. |
| [11] | Türkçe Olmayan | Twitter: Sinovac ve BionTech aşılı ile ilgili 21.676 tweet | NRC sözlüğü | •Kelime düzeyinde yapılan analiz sonucunda Biontech marka aşılı ile ilgili pozitif yorumların daha çok olduğu gözlenmiştir. •Cümle düzeyinde yapılan analiz sonucunda Sinovac marka aşılı ile ilgili daha olumlu paylaşımlar yapıldığı sonucu elde edilmiştir. |
| [12] | Türkçe Olmayan | Çinli seyahat rezervasyon platformu olan Ly.com'dan alınan 130.000 yorum | Çince Affective Lexicon Ontology | Hesaplama yöntemi kullanılarak en iyi on turistik yer puanlanmış ve sıralama sonuçlarının Ly.com'un sıralamasıyla oldukça tutarlı olduğu gözlenmiştir. |
| [13] | Türkçe Olmayan | Amazon ürün incelemesi veri seti SemEval-2007 veri seti | •WKWSC I •Hu & Liu •MPQA •Özellik Sözlüğü •Genel Sorgulayıcı •NRC sözlük duyarlılığı sözlüğü, •SO-CAL | •Ürün veri seti ile en yüksek doğruluk oranı Hu & Liu sözlüğü ile %75 olarak elde edilmiştir. •SemEval veri seti ile en yüksek doğruluk oranı WKWSC sözlüğü ile %69 olarak elde edilmiştir. |

3. Makine öğrenmesi yaklaşımı

Metinler üzerinde yapılan duygu analizi çalışmaları detaylı olarak incelemeye devam edilirken bu kısımda makine öğrenmesi yaklaşımı kullanılarak yürütülen çalışmalar incelenmiştir. Makine öğrenmesi yaklaşımı kullanılarak 2013-2022 yılları arasında yapılmış toplam 41 çalışma (15'i Türkçe, 26'sı farklı diller) ele alınmıştır.

Makine öğrenmesi yaklaşımı kullanılarak yapılan duygu analizi çalışmalarında ilk olarak Twitter sosyal medya sitesinden elde edilen Türkçe metinler üzerine yapılan çalışmalar incelenmiştir. Bu çalışmalarda kullanılan veri kümeleri, veriler üzerinde yapılan işlemler, makine öğrenmesi yaklaşımı algoritmaları ve çalışma sonucunda elde edilen sonuçlar aşağıda detaylı olarak anlatılmıştır.

2013 yılında yapılan [14] referanslı çalışmada Telekom sektörüne ait iki farklı şirketin Twitter gönderilerinden iki

veri seti oluşturulmuştur. Metinleri temsil etmek için Karakter N-gram ve Kelime kökleri yöntemleri ve terim ağırlıklandırma için eğitici ve geleneksel yöntemler kullanılmıştır. Naive Bayes (NB), Sıralı Minimum Optimizasyon (Sequential Minimal Optimization-SMO), Karar Ağacı (Decision Tree-DT (J48)), Rastgele Orman (Random Forest- RF) ve 1-En Yakın Komşu (1-Nearest Neighbors-IB1) algoritmaları kullanılmıştır. Her iki veri seti ile de en yüksek başarı Karakter N-gram ve RF eğitici terim ağırlıklandırma yöntemi kullanılarak SMO algoritmasıyla %65 olarak elde edilmiştir. 2014 yılında yapılan [15] referanslı çalışmada gıda sektöründeki farklı firmaların ürünleri ile ilgili Twitter gönderilerinden dengeli ve dengesiz iki farklı veri seti oluşturulmuştur. NB, RF, SMO, DT (J48) ve IB1 algoritmaları kullanılmıştır. En yüksek doğruluk dengeli veri seti ve SMO algoritması kullanılarak %72,33 olarak elde edilmiştir. 2015 yılında yapılan [16] referanslı çalışmada Samsung, Apple, LG firmaları ile ilgili Twitter gönderilerinden 3 farklı veri seti oluşturulmuştur. NB, RF, DT (J48), Destek Vektör Makinesi (LibSVM), K-En Yakın Komşu (K-Nearest Neighbours-KNN) algoritmaları kullanılmıştır. En yüksek doğruluk oranları Apple veri setinde NB algoritması ile %52,9, Samsung veri setinde NB algoritması ile %56,53 ve LG veri setinde NB algoritması ile %65,63 olarak elde edilmiştir. 2015 yılına ait olarak incelenen bir diğer çalışma olan [17] referanslı çalışmada Twitter gönderilerinden elde edilen verilerde özniteliklerin çıkarılması için Kelime torbası (Bag of Words-BoW) ve N-gram yöntemleri kullanılmıştır. KNN, NB, Destek Vektör Makinesi (Support Vector Machine-SVM) ve Çok Terimli Naive Bayes (Multinomial Naive Bayes-MNB) algoritmaları kullanılmıştır.

En yüksek sınıflandırma başarısı N-gram kullanılarak KNN ve SVM algoritmaları ile %65 olarak elde edilmiştir. 2016 yılında yapılan [18] referanslı çalışmada Twitter'dan elde edilen veriler kullanılmış ve bu veriler konu bilgisine göre etiketlenmiştir. Özniteliklerin çıkarılmasında BoW ve N-gram yöntemleri kullanılmıştır. SVM, NB, MNB ve KNN algoritmaları kullanılmıştır. Çalışma sonucunda en yüksek başarı BoW ile MNB algoritması kullanılarak %92,50 olarak bulunmuştur.

2017 yılında yapılan [19] referanslı çalışmada Telekomünikasyon sektöründeki bir firmaya ait Twitter gönderilerinden elde edilen veriler kullanılmıştır. Niteliklerin bulunmasında BoW yöntemi kullanılarak Kikare, Bilgi Kazanımı, Sorgu Genişletme Sıralaması ve Karınca Kolonisi Optimizasyonu (KKO) gibi farklı nitelik seçim yöntemleri karşılaştırılmıştır. Maksimum Entropi (ME) kullanılmıştır ve en yüksek F-skoru KKO kullanılarak %78 olarak elde edilmiştir. 2017 yılında incelenen bir diğer çalışma olan [20] referanslı çalışmada Twitter gönderilerinden elde edilen veriler ve NB, SVM, Lojistik Regresyon ((Logistic Regression-LR) algoritmaları kullanılmıştır. En yüksek doğru sınıflandırma oranı öznitelik temsili olarak 1-gram+2-gram kullanılarak NB ve LR algoritmaları ile %77 olarak bulunmuştur.

2020 yılında yapılan [21] referanslı çalışmada Twitter gönderilerinden elde edilen veriler daha yüksek başarı oranı elde edilebilmesi için 5 farklı konu başlığına göre olumlu ve olumsuz şeklinde kurallı eğitim setleri oluşturularak analiz edilmiş ve LR algoritması kullanılmıştır. Performans ölçütü olarak Alıcı İşletim karakteristiği (Receiver Operating

Characteristic-ROC) eğrisinin altındaki alan (Area Under Curve -AUC) değeri kullanılmıştır ve en iyi sonuç sokağa çıkma yasağı konu başlığı için %98 olarak elde edilmiştir. 2021 yılında yapılan [22] referanslı çalışmada Twitter gönderilerinden elde edilen veriler Pandas isimli kütüphane kullanılarak pozitif ve negatif olarak etiketlenmiştir. NB, KNN, SVM, LR ve DT algoritmaları kullanılmıştır. Performans ölçütü olarak F-skoru kullanılmıştır ve en iyi sonuç SVM algoritması ile %73 olarak elde edilmiştir.

Makine öğrenmesi yaklaşımı ile hazır veri setleri veya farklı web sitelerinden elde edilen Türkçe metinler üzerinde yapılan duygu analizi çalışmaları aşağıda incelenmiştir.

2017 yılında yapılan [23] referanslı çalışmada elektronik cihazlar ile ilgili müşteri yorumlarını içeren bir e-ticaret sitesinden elde edilen yorumlar kullanılmıştır. RF algoritması kullanılmıştır ve %84,23 doğruluk elde edilmiştir. 2019 yılında yapılmış olan [24] referanslı çalışmada bir e-ticaret sitesindeki üç farklı cep telefonu markası ile ilgili yapılmış yorumlar kullanılmıştır. Veri etiketleme işlemi için kullanıcıların verdikleri yıldız bilgisi kullanılmıştır. NB, SVM, KNN algoritmaları kullanılmıştır. En yüksek doğruluk oranı SVM algoritması ile pozitif etiketliler için %77,69 ve negatif etiketliler için %74,19 olarak elde edilmiştir. 2019 yılında yapılan [25] referanslı çalışmada eğitim modeli için Türkçe duygu veri seti (The Turkish emotion data set-TREMO) hazır veri setindeki veriler kullanılmıştır. Yapay Sinir Ağları (Artificial Neural Network-ANN), SVM, RF ve KNN algoritmaları kullanılmıştır. Karışıklık matrisi oluşturulmuştur ve en yüksek doğruluk ANN algoritması kullanılarak mutluluk etiketi için %89 olarak bulunmuştur. 2019 yılında incelenmiş olan [26] referanslı çalışmada Facebook sitesinden şirketlerle ilgili alınan veriler kullanılmıştır. ANN, DT, SVM, NB, LR, KNN ve Extreme Gradient Boosting (XGBoost) algoritmaları kullanılmıştır. Bütün algoritmalar ile yaklaşık %70 oranında doğruluk elde edilmiştir ancak en yüksek doğruluk LR algoritması ile %74,13 olarak elde edilmiştir. 2020 yılında yapılmış olan [27] referanslı çalışmada çevrimiçi kitap sitesindeki kullanıcı yorumlarını içeren dengeli bir veri seti kullanılmıştır.

Çok Katmanlı Algılayıcı (Multi-Layer Perceptron-MLP), NB, SVM ve LR algoritmaları kullanılmıştır. En yüksek sınıflandırma başarısı MLP algoritması ile %89 olarak elde edilmiştir. 2021 yılında yapılan [28] referanslı çalışmada Yemek Sepeti sitesinden elde edilen veriler kullanılmıştır. ANN, SVM, KNN, NB ve Doğrusal Regresyon algoritmaları kullanılmıştır. En yüksek doğruluk oranı ANN algoritması kullanılarak %86 olarak elde edilmiştir.

Makine öğrenmesi yaklaşım kullanılarak Türkçe olmayan metinler üzerinde duygu analizi çalışmaları yapılmıştır. Aşağıda incelenen çalışmalarda kullanılan veriler Twitter sosyal medya sitesinden veya farklı web sitelerden elde edilmiştir.

Twitter verileri kullanılarak yapılmış olan duygu analizi çalışmalarından biri 2019 yılında yapılan [29] referanslı çalışmada Twitter gönderilerinden elde edilen veriler pozitif ve negatif olarak etiketlenerek kullanılmıştır. NB, KNN, SVM algoritmaları kullanılmıştır. En yüksek doğruluk NB algoritması ile %80,90 olarak elde edilmiştir. 2019 yılında yapılan [30] referanslı çalışmada Twitter gönderilerinden elde edilen veriler incelenmiştir. Toplanan veriler 5 gönüllü

kişi tarafından pozitif ve negatif olarak etiketlenmiştir. Fonksiyon olarak Doğrusal Ridge Regresyonu ve NB algoritmaları, başarı ölçütü olarak F-skor kullanılmıştır ve en yüksek F-skor Ridge Regresyonu ile %96,9 olarak elde edilmiştir. 2019 yılında yapılan [31] referanslı çalışmada Twitter gönderilerinden ve Kaggle Twitter ABD havayoluna ait verilerden oluşan 2 farklı veri seti kullanılmıştır. NB, RF, SVM, DT, KNN algoritmaları uygulanan çalışmada, en yüksek doğruluk oranları SVM algoritması ile Twitter veri seti için %83,67 olarak ve havayolu veri seti için %69 olarak elde edilmiştir.

2019 yılında yapılan [32] referanslı çalışmada Sentiment.140 Twitter gönderileri kullanılmıştır. SVM, RF, K-Ortalama Kümeleme algoritmaları kullanılmıştır. En yüksek doğruluk oranı K-Ortalama Kümeleme algoritması kullanılarak %74 olarak elde edilmiştir. 2020 yılında yapılan [33] referanslı çalışmada Kaggle Twitter ABD havayoluna ait veriler pozitif, negatif ve nötr olarak ayrılarak kullanılmıştır. ANN, DT ve SVM algoritmaları kullanılmıştır. En yüksek doğruluk oranı ANN algoritması ile %75,99 olarak elde edilmiştir. 2020 yılında [34] referanslı çalışmada Twitter gönderilerinden elde edilen veriler ve SVM, NB, LR, RF algoritmaları kullanılmıştır. En yüksek doğruluk NB algoritması ile yaklaşık %76 olarak elde edilmiştir. 2020 yılında yapılan [35] referanslı çalışmada farklı telefon modeller ile ilgili Twitter gönderilerinden elde edilen veriler ve NB, SVM, DT, RF algoritmaları kullanılmıştır. En yüksek doğruluk SVM algoritması ile %79,90 olarak elde edilmiştir. 2020 yılında yapılan [36] çalışmasında Twitter gönderilerinden elde edilen ve pozitif, negatif ve nötr olarak gruplandırılmış veriler kullanılmıştır. NB, SVM, ME algoritmaları kullanılmıştır. En yüksek doğruluk NB algoritması ile %86 olarak elde edilmiştir. 2020 yılındaki bir diğer incelenen çalışma olan [37] referanslı çalışmada Haryana Meclisi anketleri, The Sky is Pink film yorumları, United Nations General Assembly (UNGA), Howdy Modi konuları ile ilgili Twitter gönderilerinden elde edilen farklı veri setleri kullanılmıştır. NB, SVM, RF, Uzun Kısa Vadeli Bellek Ağları (Long Short-Term Memory-LSTM) algoritmaları kullanılmıştır. En yüksek doğruluk oranları Haryana Meclisi anketleri NB algoritması ile %96, The Sky is Pink SVM ile %98,29, UNGA NB algoritması ile %98, Howdy Modi RF algoritması ile %98,24 olarak elde edilmiştir. 2021 yılındaki [38] referanslı çalışmada Kaggle Twitter ABD havayoluna ait veriler kullanılmıştır. Terim ağırlıklandırma için Terim Frekansı-Ters Belge Frekansı (Term Frequency-Inverse Document Frequency-TF-IDF) ve BoW yöntemleri kullanılmıştır. SVM, MNB, RF, LR algoritmaları kullanılmıştır. En yüksek doğruluk oranı BoW ağırlıklandırma yöntemi ve SVM ve LR algoritmaları ile %77 olarak elde edilmiştir. 2021 yılında yapılan [39] referanslı çalışmada Twitter gönderilerinden elde edilen veriler manuel olarak pozitif ve negatif olarak etiketlenerek kullanılmıştır. SVM, LSTM ve Bernoulli Naive Bayes (BernoulliNB) algoritmaları karşılaştırılmıştır. En yüksek doğruluk oranı LSTM algoritması ile %79 olarak elde edilmiştir. 2021 yılında yapılan [40] referanslı çalışmada Twitter gönderilerinden elde edilen veriler kullanılmıştır. Oluşturulan modelin eğitimi için Sentiment.140 veri seti ve NB, LR, SVM algoritmaları kullanılmıştır. Algoritmaların performansını değerlendirmek için alıcı çalışma

özelliklerinin altındaki alan (Area Under the Receiver Operating Characteristics-AUROC) ve doğruluk oranları kullanılmıştır. En yüksek doğruluk oranı Doğrusal Destek Vektör Makinesi (Linear Support Vector Machine-LinearSVM) algoritması ile %83,71 olarak elde edilmiş ancak AUROC değeri göz önüne alındığında LR algoritması ile %82,47'lik doğruluk oranı elde edilmiştir. 2021 yılında yapılan [41] referanslı çalışmada havayolu ile ilgili Twitter gönderileri kullanılmıştır. DT, SVM, KNN, SMO algoritmaları ve KNN+SVM, SMO+DT (SMODT) sınıflandırıcıları kullanılmıştır. En yüksek doğruluk SMO+DT ile %89,47 olarak elde edilmiştir.

Twitter dışında farklı yollardan elde edilen veriler kullanılarak makine öğrenmesi yaklaşımıyla duygu analizi çalışmaları incelendiğinde 2016 yılında yapılan [42] referanslı çalışmada internet film veri tabanı (Internet Movie DatabaseIMDB) sitesindeki film yorumlarından elde edilen pozitif ve negatif olarak dengeli bir veri seti kullanılmıştır. NB, MLP, SVM ve Merkez Tabanlı Sınıflayıcı algoritmaları karşılaştırılmıştır. Bu çalışma sonucunda eğitim veri setinde ANN algoritması ile %89 doğruluk elde edilirken test veri setinde MLP ve SVM algoritmaları ile %75 doğruluk elde edilmiştir. Tüm veriler ile en yüksek doğruluk oranı MLP algoritması ile %86 olarak elde edilmiştir. 2018 yılında [43] referanslı çalışmada UC (University of California) Irvine makine öğrenmesi deposundaki veriler kullanılmıştır. LR, DT (C4.5) ve NB algoritmaları kullanılmıştır.

En yüksek doğruluk LR algoritması ile %79 olarak elde edilmiştir. 2018 yılında yapılan [44] referanslı çalışmada IMDB sitesindeki İngilizce yazılmış film yorumları ve Beyazperde.com sitesindeki Türkçe yazılmış film yorumlarından elde edilen farklı boyutlardaki ancak pozitif ve negatif olarak dengeli olan 2 farklı veri seti kullanılmıştır. Terim ağırlıklandırılmasında Terim Frekansı (TF), TF-IDF, 1-gram, 2-gram, 1-2-gram ve Word2vec yöntemleri kullanılmıştır. NB, LR, DT, KNN, SVM, ANN algoritmaları kullanılmıştır. Bu çalışma sonucunda 1-gram yönteminin diğer N-gram yöntemlerine göre daha başarılı olduğu ve TF-IDF yönteminin de TF yöntemine göre daha iyi sonuçlar verdiği gözlenmiştir. İngilizce film yorumları üzerine yapılan incelemede en yüksek doğruluk ANN algoritması ile %88 olarak elde edilmiş iken Türkçe film yorumları üzerine yapılan incelemede en yüksek doğruluk LR algoritması ile %84 olarak elde edilmiştir. 2019 yılında yapılan [45] referanslı çalışmada TripAdvisor web sitesindeki çevrimiçi müşteri yorumlarından elde edilen veriler kullanılmıştır. TextBlob ve NB algoritması kullanılmıştır. En yüksek doğruluk NB algoritması ile %72 olarak elde edilmiştir. 2019 yılında yapılan [46] referanslı çalışmada Yelp veri setindeki veriler kullanılmıştır. NB, MNB, LR, BernoulliNB, Doğrusal Destek Vektör Kümeleme (Linear Support Vector Clustering-Linear SVC) algoritmaları karşılaştırılmıştır. En yüksek doğruluk oranı NB algoritması ile %79 olarak elde edilmiştir.

2019 yılındaki [47] referanslı çalışmada ise Twitter gönderilerinden ve farklı sitelerden elde edilen veri setleri kullanılmıştır. Bunlar Twitter sosyal medya sitesinden oluşan Arapça Duygu tweetleri Veri Kümesi (Arabic Sentiment tweets Dataset-ASTD), Stanford Twitter duyarlılığı külliyesi (Stanford), Twitter ABD Havayolu Duyarlılığı Veri Kümesi (ABD Havayolu), Qaym4, TripAdvisor web sitelerinden elde edilen restoran

yorumlarından oluşturulan veri kümesi (RES), Uber Ride İncelemeleri Veri Kümesi (Uber) veri setleridir. DT, KNN, LR, MLP algoritmaları kullanılmıştır. En yüksek doğruluk oranları RES veri setinde MLP algoritması ile %81, ASTD veri setinde LR algoritması ile %68, Stanford veri setinde LR algoritması ile %79, Uber veri setinde BernoulliNB ve KNN algoritmaları ile %84, ABD havayolu veri setinde BernoulliNB algoritması ile %76 olarak elde edilmiştir. 2019 yılında yapılan [48] referanslı çalışmada ise Tripadvisor.com web sitesinden elde edilen oteller ile ilgili yorumlar ve değerlendirme puanları kullanılmıştır ve elde edilen veriler değerlendirme puanlarına göre olumlu ve olumsuz şeklinde ayrılmıştır. LR, KNN, NB ve SVM algoritmaları karşılaştırılmıştır. En yüksek doğruluk oranı LR algoritması ile %92 olarak elde edilmiştir. 2020 yılında [49] referanslı çalışmada uzaktan eğitim öğrencileri için hazırlanan çevrimiçi materyaller ile ilgili yapılan geri bildirimlerden elde edilen veriler kullanılmıştır. DT, MLP, SVM, XGBoost, KNN, Çok Terimli Lojistik Regresyon ve Gaussian NB algoritmaları kullanılmıştır. En yüksek doğruluk oranı LR algoritması ile %77 olarak elde edilmiştir. 2021 yılında yapılan [50] referanslı çalışmada olumlu ve olumsuz yorum sayıları dengeli olan Amazon, Imdb, Yelp sitelerinden elde edilen veriler ve bu veri setlerinin birleştirilmesi ile oluşturulan yeni bir veri seti kullanılmıştır. SVM, KNN, NB, DT, RF algoritmaları kullanılmıştır. En yüksek doğruluk oranları SVM algoritması ile Amazon veri seti kullanılarak %79, Imdb veri seti kullanılarak %80, Yelp veri seti kullanılarak %79 ve birleştirilmiş veri seti kullanılarak %85 olarak elde edilmiştir. 2021 yılında yapılan [51] referanslı çalışmada Kaggle Amazon ürün yorumları verileri kullanılmıştır. NB, SVM, DT, LR algoritmaları kullanılmıştır. En yüksek doğruluk DT algoritması ile %94 olarak elde edilmiştir. 2022 yılında yapılan [52] referanslı çalışmada Apple, IBM ve NIKE şirketleri ile ilgili Yahoo Finance API'sinden elde edilen veriler kullanılmıştır. Doğrusal Regresyon, DT, SVM, LSTM, Radyal Temel Fonksiyon ve Tekrarlayan Sinir Ağı (Recurrent Neural Network-RNN) algoritmaları kullanılmıştır. Bu çalışmada elde edilen başarılar mutlak hata yüzlerine göre kıyaslanmıştır ve en düşük hata oranı LSTM algoritmasıyla %2,81 olarak bulunmuştur. Apple, IBM ve NIKE için stok tahmini yapılarak gerçek değerler ile tahmini değerlerin birbirine çok yakın olduğu sonucu elde edilmiştir. 2022 yılında yapılan [53] referanslı çalışmada IMBD sitesinden elde edilen olumlu ve olumsuz olarak işaretlenmiş veriler kullanılmıştır. NB, SVM, RF, MLP algoritmaları kullanılmıştır. En yüksek doğruluk MLP algoritması ile %88 olarak elde edilmiştir. 2022 yılında yapılan [54] referanslı çalışmada FoodPanda isimli web sitesinden Bangla dilinde elde edilen veriler kullanılmıştır. MNB, DT, LR, RF, Linear SVM algoritmaları kullanılmıştır. Terim ağırlıklandırma için sayma vektörü ve TF-IDF yöntemleri kullanılmıştır. Bu çalışmada derin öğrenme için RNN, LSTM, Kapılı tekrarlayan birim (Gated Recurrent Unit-GRU) algoritmaları kullanılmıştır. En yüksek doğruluk oranı TF-IDF yöntemi ve LR algoritması ile %75 olarak elde edilmiş iken derin öğrenme yöntemlerinden Word2Sec yöntemi ve LSTM kullanılarak %90 olarak elde edilmiştir.

Makine öğrenmesi yaklaşımı kullanılarak geçmişten günümüze doğru incelenen çalışmalarda kullanılan veri seti, duygu analizi yapılan metnin dili, veri ön işlemede kullanılan

yöntemler, makine öğrenmesi yaklaşımında kullanılan algoritmalar gibi farklı kriterler ile elde edilen başarı oranları arasında nasıl bir bağlantının olduğu incelenmek istenmiştir. Bu sebeple farklı kriterler ve elde edilen sonuçlar kullanılarak oluşturulan tablolar ile bu kriterlerin sonuçlar üzerindeki etkileri tespit edilmeye çalışılmıştır.

Makine öğrenmesi yaklaşımı kullanılarak incelenmiş çalışmalarda kullanılan metnin dili, kullanılan veri seti, veriler üzerinde yapılan ön işlem adımları, kullanılan programlar ve çalışmalarda kullanılan makine öğrenmesi algoritmaların elde edilen başarıya olan etkisi incelenmek istenmiştir. Bu sebeple Tablo 2’ de makine öğrenmesi yaklaşımı kullanılarak yapılan duygu analizi çalışmalarındaki incelenen dil, veri seti, ön işlem adımları, kullanılan programlar ve ilgili çalışmalarda kullanılan algoritmalar ve çalışma sonucunda kullanılan algoritmalarından en başarılı algoritma ile elde edilen sonuçlar listelenerek belirtilen kriterler ile elde edilen sonuçlar arasındaki ilişkiler incelenmiştir.

Tablo 2. Makine öğrenmesi yaklaşımında farklı kriterlerin karşılaştırılması.

| Referans | Dil | Veri Seti | Ön İşlem Adımları | Kullanılan Program | Algoritmalar | Sonuçlar |
|----------|--------|---|--|--------------------|---|---|
| [14] | Türkçe | Twitter: Pozitif, negatif ve nötr olarak etiketlenmiş 2 farklı veri seti | Kelime kökleri Karakter gramları TF IDF RF,F1,F2... | Weka | •SVM •DT •RF •IB1 •NB | En yüksek başarı SVM ile; •A veri kümesi ile %65, •B veri kümesinde %65,3 olarak elde edilmiştir. |
| [15] | Türkçe | Twitter: Pozitif, negatif ve nötr olarak etiketlenmiş dengeli (824 veri) dengesiz (2.000 veri) iki veri seti | N-Gram TF | Weka | •NB •DT •IB1 •SMO •RF | En yüksek doğruluk oranı, SMO ile; •Dengeli veri seti ile %72,33 iken •Dengesiz veri seti ile %66,40 olarak elde edilmiştir. |
| [16] | Türkçe | Twitter: Samsung, Apple LG markaları ile ilgili 500’er tweet | Yazım Düzeltici (Python) | Weka | •NB •DT •KNN •RF •Destek Vektör Makinesi (LibSVM) | En yüksek doğruluk NB ile; •Apple için %53,30, •Samsung için %56,53 ve •LG için %65,63 olarak elde edilmiştir. |
| [17] | Türkçe | Twitter: Türkçe mesajı | Vektör Uzay Modeli BoW N-gram TF TF-IDF Boolean, CfsSubset algoritması Vektör uzay modeli Binary, TF | Weka | •SVM •NB •MNB •KNN | En yüksek doğru sınıflandırma yüzdesi: N-gram yöntemi ve MNB ile %66 olarak elde edilmiştir. |
| [18] | Türkçe | Twitter: 14.777 adet mesaj | TF-IDF BoW, N-gram Korelasyon Tabanlı Öznitelik Seçme BoW | | •NB •MNB •KNN •SVM •ME | En yüksek doğru sınıflandırma yüzdesi: •MNB ile %66 iken, •Konuya göre etiketleme işlemi yapıldıktan sonra %92,50 olarak elde edilmiştir. |
| [19] | Türkçe | Twitter: Telekomunikasyon ile ilgili 2.043 tweet | KKO Ki-kare Bilgi Kazanımı Sorgu Genişletme Sıralaması | Weka | • ME | En yüksek F-skoru: KKO öznitelik kullanılarak ME ile %78 olarak elde edilmiştir. |
| [20] | Türkçe | Twitter: Pozitif ve negatif dengeli 10.600 tweet | N-gram | Weka | •NB •LR •SVM | En yüksek doğru sınıflandırma: NB ve LR ile %77 olarak elde edilmiştir. |
| [21] | Türkçe | Twitter: 2.000.000 tweet | BoW TF-IDF | Python Ms Sql | • LR | En iyi AUC değeri: •Rastgele oluşturulan eğitim setlerine göre Sokağa Çıkma Yasağı ile ilgili olarak %82, |

| | | | | | | |
|------|-------------------|--|--|-----------------------|--|--|
| [22] | Türkçe | Twitter: Pozitif ve negatif dengeli 32.000 tweet | TF-IDF | Python | <ul style="list-style-type: none"> • NB • KNN • SVM • LR • DT | <ul style="list-style-type: none"> •Kurallı eğitim gruplarına göre maske için %98 olarak elde edilmiştir. <p>En yüksek F-skoru: SVM ile %73 olarak elde edilmiştir.</p> |
| [23] | Türkçe | E-ticaret sitesinden eğitim (9.100), Test (3.100) veri 3 farklı cep telefonu | Word2vec | | <ul style="list-style-type: none"> • RF | En yüksek doğruluk: RF ile %84,23 olarak elde edilmiştir. |
| [24] | Türkçe | markasına ait pozitif ve negatif dengeli 1000 yorum | | R | <ul style="list-style-type: none"> •NB •KNN SVM | En yüksek doğruluk: •Pozitif tahminlemede SVM ile %74 •Negatif tahminlemede NB ile %83 olarak elde edilmiştir. |
| [25] | Türkçe | TREMO veri seti: 25.989 veri | Vektör uzay model Karşılıklı bilgi yöntemi TF-IDF | | <ul style="list-style-type: none"> •ANN •RF •SVM •KNN •ANN •SMO •LR •XGBoost •DT •NB •KNN | En yüksek doğruluk: %82-%89 arasında bulunmuştur. |
| [26] | Türkçe | Facebook sitesinden: 8770 yorum | Sayma Vektörü | Weka | <ul style="list-style-type: none"> •MLP •NB •SVM •LR •ANN •SV •KNN •NB •Doğrusal Regresyon •NB •SVM •KNN | En yüksek doğruluk: LR ile %74,13 olarak elde edilmiştir. |
| [27] | Türkçe | Kitap yorumu: 700 pozitif, 700 negatif yorumdan oluşan dengeli veri kümesi | TF-IDF | Phyton Rapid Miner | <ul style="list-style-type: none"> •MLP •NB •SVM •LR | En yüksek sınıflandırma başarısı: MLP ile %89 bulunmuştur. |
| [28] | Türkçe | Yemek Sepeti: olumlu ve olumsuz dengeli 676 bin veri | Word2vec | Python | <ul style="list-style-type: none"> •ANN •SV •KNN •NB •Doğrusal Regresyon •NB •SVM •KNN | En yüksek doğruluk: ANN ile %86 olarak elde edilmiştir. |
| [29] | Türkçe Olmayan | Twitter: 443 tweet | Tokenizasyon | Python Rapid Miner | <ul style="list-style-type: none"> •NB •SVM •KNN | En yüksek doğruluk: NB ile %80,90 olarak elde edilmiştir. |
| [30] | Türkçe Olmayan | Twitter: Pozitif veya negatif olarak etiketlenmiş 162.000 tweet | TF-IDF | MYSQL Python | <ul style="list-style-type: none"> • NB •Doğrusal Ridge Regresyon | En yüksek F-skoru: Ridge Regresyonu ile %96,9 olarak elde edilmiştir. |
| [31] | Türkçe Olmayan | Twitter: 89.743 tweetten oluşan ve 14.641 tweetten oluşan ABD havayolu veri seti Sentiment.140'da n: eğitim (13.418), test (2.368) Kaggle'dan | TF-IDF | | <ul style="list-style-type: none"> •NB •RF •SVM •DT •KNN | En yüksek doğruluk: •Twitter veri setinde SVM ile %83,67, •Havayolu veri setinde SVM ile %69 olarak elde edilmiştir. |
| [32] | Türkçe Olmayan | Twitter: ABD havayolu veri setinden eğitim için 10.248 test için 4.392 tane 1.600.000 sınıflandırılmış tweet | N-gram TF-IDF | | <ul style="list-style-type: none"> •SVM •RF •K-Ortalama Kümeleme | En yüksek doğruluk: K-ortalama Kümeleme ile %74 bulunmuştur. |
| [33] | Türkçe Olmayan | Twitter: ABD havayolu veri setinden eğitim için 10.248 test için 4.392 tane 1.600.000 sınıflandırılmış tweet | N-gram | Python | <ul style="list-style-type: none"> •ANN •DT •SVM | En yüksek doğruluk: ANN ile %75,99 olarak elde edilmiştir. |
| [34] | Türkçe Olmayan | Twitter: Google Samsung, Mi Apple ve diğer cep telefonlarla ilgili eğitim için 6.000 test için 2.000 tweet | BoW | | <ul style="list-style-type: none"> •SVM •NB •LR •RF | En yüksek doğruluk: NB ile yaklaşık %76 olarak bulunmuştur. |
| [35] | Türkçe Olmayan | Twitter: Google Samsung, Mi Apple ve diğer cep telefonlarla ilgili eğitim için 6.000 test için 2.000 tweet | N-gram | | <ul style="list-style-type: none"> •SVM •NB •RF •DT | En yüksek doğruluk: SVM ile %79,90 olarak elde edilmiştir. |
| [36] | Türkçe Olmayan | Twitter: 1.000 tweet | | Python | <ul style="list-style-type: none"> •NB •SVM •ME | En yüksek doğruluk: NB ile %86 olarak elde edilmiştir. |

| | | | | | | |
|------|------------------|---|--|---------------|---|--|
| [37] | Türkçe Olmayan | Twitter: Haryana Meclisi anketleri ile 16000 tweet, The Sky is Pink, filmi ile ilgili yorumlar, UNGA, yaklaşık 5.000 tweet, Howdy Modi hakkında tweetler Kaggle'dan | Tokenization N-gram Kelime Bölütleme (POS) | | •NB •SVM •RF •LSTM | En yüksek doğruluk: •Haryana Meclisi anketleri NB ile %96, •The Sky is Pink SVM ile %98,29, •UNGA NB ile %98, •Howdy Modi RF ile %98,24 olarak elde edilmiştir. |
| [38] | Türkçe Olmayan | Twitter: ABD Havayolu verilerinden 14.640 kayıt | BoW TF-IDF | | •SVM •MNB •RF •LR | En yüksek doğruluk: SVM ve LR ile %77 olarak elde edilmiştir. |
| [39] | Türkçe Olmayan | Twitter: 4.035 tweet | Sayma Vektörü TF-IDF | NLTK | •LSTM •BernoulliNB •NB •SVM | En yüksek doğruluk: LSTM ile %79 olarak elde edilmiştir. |
| [40] | Türkçe Olmayan | Kaggle: Pozitif, negatif dengeli 1,6 milyon veri | | | •NB •Linear SVM •LR •SVM •KNN | En yüksek doğruluk: •DDVS ile %83,71 ve •LR ile %82,47 olarak elde edilmiştir. |
| [41] | Türkçe Olmayan | Twitter: Havayolu ile ilgili 1000 tweet | | Python | •KNN+SVM •SMO+DT •SVM •SMO •DT | En yüksek doğruluk: SMO+DT ile %89,47 olarak elde edilmiştir. |
| [42] | Türkçe Olmayan | Internet Movie Database (IMDb) sitesindeki pozitif ve negatif eşit 2.000 film yorumu | Vektör Uzay Modeli Sayma Vektörü TF-IDF | Python Matlab | •NB •MLP •SVM •Merkez Tabanlı Sınıflayıcı | En yüksek doğruluk: •Eğitim veri setinde ANN ile %89,73 •Test veri seti için SVM ve ANN ile %75 olarak elde edilmiştir. |
| [43] | Türkçe Olmayan | UCI makine öğrenme havuzundan İngilizce için IMDB'deki 50.000 yorum, | Word2vec BoW N-gram TF TF-TDF | R | •LR •NB •DT •NB •LR •DT | En yüksek doğruluk: LR algoritması ile %79 olarak elde edilmiştir. |
| [44] | Türkçe/İngilizce | Türkçe için Beyazperde.com'dan 10.662 yorum | | Python | •ANN •KNN •SVM | En yüksek doğruluk: •İngilizce film yorumları ANN ile %88,50 •Türkçe film yorumları LR ile %84,02 olarak elde edilmiştir. |
| [45] | Türkçe Olmayan | TripAdvisor: 337 veri | | WebHarvy WEKA | •Text Blob •NB •BernoulliNB | En yüksek doğruluk: NB ile %72,06 olarak elde edilmiştir. |
| [46] | Türkçe Olmayan | Restoranlarla ilgili YELP veri seti | POS | Python | •NB •MNB •LR •Linear SVC | En yüksek doğruluk: NB ile %79,12 olarak elde edilmiştir. |
| [47] | Türkçe Olmayan | Arapça ve İngilizce dillerinde veri setleri | Vektör Uzay Modeli BoW TF-IDF | Python | •LR •Perceptron •MLP •BernoulliNB •DT •MNB •KNN | Doğruluk oranları: •RES veri setinde MLP ile %81 •ASTD veri setinde LR ile %68 •Stanford veri setinde LR ile %79 •Uber sveri setinde BernoulliNB ve KNN ile %84 •US havayolu veri setinde BernoulliNB ile %76 olarak elde edilmiştir. |
| [48] | Türkçe Olmayan | Tripadvisor.com: 708 otel yorumu | Vektör Oluşturma: TF | Python | •LR •KNN •SVM •NB •XGBoost •Gaussian NB •Çok Terimli Lojistik | En yüksek doğruluk: LR ile %92'lik olarak elde edilmiştir. |
| [49] | Türkçe Olmayan | Çevrimiçi materyallerle ilgili 6.059 geri bildirim | Üçlü likert yöntemi | Python | Regression •DT •MLP •SVM •KNN | En yüksek doğruluk: LR algoritması ile %77,5 olarak elde edilmiştir. |
| [50] | Türkçe Olmayan | Amazon İmdb Yelp Anonimden olumlu ve olumsuz dengeli 3.000 veri | TF-IDF | Python | •SVM •KNN •RF •NB •DT | En yüksek doğruluk: SVM ile, •Amazon veri seti ile %79, •İmdb veri seti ile %80, •Yelp veri seti ile %79, •Birleştirilmiş veri setiyle %85 olarak elde edilmiştir. |

| | | | | | | |
|------|----------------|---|---|--------|--|--|
| [51] | Türkçe Olmayan | Kaggle: 28.332 satır ürün yorumu | TF-IDF | Phyton | <ul style="list-style-type: none"> •NB •SVM •LR •DT •Doğrusal Regresyon | En yüksek doğruluk: DT ile %94 olarak elde edilmiştir. |
| [52] | Türkçe Olmayan | Twitter: Apple IBM NIKE şirketleri ile ilgili tweetlerden | | | <ul style="list-style-type: none"> •RNN •SVM •LSTM •DT •NB | En düşük hata oranı: LSTM algoritması ile 2.81 olarak elde edilmiştir. |
| [53] | Türkçe Olmayan | IMBD sitesinden elde edilen yorumlar | TF-IDF | | <ul style="list-style-type: none"> •SVM •RF •MLP •RF •MNB | En yüksek doğruluk: MLP ile %88 olarak elde edilmiştir. |
| [54] | Türkçe Olmayan | Web sitesinden elde edilen 520 pozitif 520 negatif yorum | Sayma Vektörü TF-IDF N-gram Derin öğrenme Kelime dizisi | | <ul style="list-style-type: none"> •DT •LR •RNN •LSTM •GRU •Linear SVM | En yüksek doğruluk: <ul style="list-style-type: none"> •Makine öğrenmesi LR ile %75, •Derin öğrenme LSTM ile %90 olarak elde edilmiştir. |

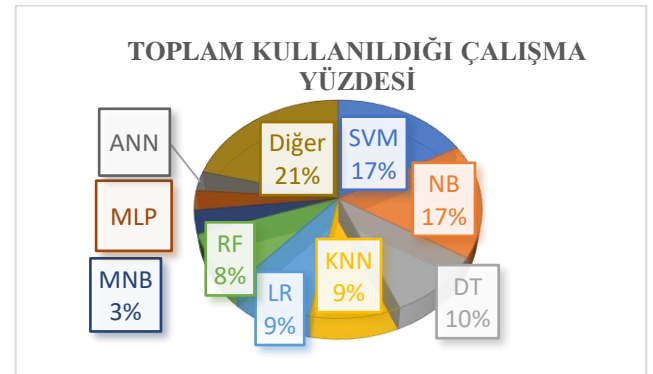
Makine öğrenmesi yaklaşımı kullanılarak incelenen 41 çalışmada kullanılan algoritmalar çoğunlukla aynı olmasına rağmen her çalışmada farklı sayıda algoritma kullanılarak toplamda 172 tane algoritmadan bahsedilmiştir. Tablo 3'te makine öğrenmesi yaklaşımında algoritmaların kullanıldığı çalışma sayıları gösterilmiştir.

Tablo 3. Makine öğrenmesi yaklaşımında algoritmaların kullanıldığı çalışma sayıları.

| Kullanılan Algoritma | Toplam Kullanıldığı Çalışma Sayısı |
|---------------------------------|------------------------------------|
| IB1 | 2 |
| ANN | 5 |
| BernoulliNB | 3 |
| MLP | 5 |
| Çok Terimli Lojistik Regresyon | 1 |
| Destek Vektör Makinesi (LibSVM) | 1 |
| Linear SVM | 2 |
| Doğrusal Regresyon | 2 |
| Linear SVC | 1 |
| XGBoost | 2 |
| Gaussian NB | 1 |
| GRU | 1 |
| DT | 17 |
| KNN | 16 |
| KNN+SVM | 1 |
| K-Ortalama Kümeleme | 1 |
| Doğrusal Ridge Regresyon | 1 |
| LR | 15 |
| ME | 3 |
| Merkez Tabanlı Sınıflayıcı | 1 |
| MNB | 6 |
| NB | 29 |
| RF | 14 |
| Regression | 1 |
| SMO | 3 |
| SVM | 28 |
| RNN | 2 |
| LSTM | 4 |
| Diğer | 4 |

Tablo 3'teki veriler incelendiğinde sırasıyla en çok NB, SVM algoritmaların kullanıldığı gözlenmiştir. Bu algoritmalar incelendiğinde eğitici makine öğrenmesi algoritmaları oldukları gözlenmiştir. Ayrıca bu algoritmaların özellikleri incelendiklerinde metin türündeki veriler ile uyumlu çalıştığı ve yüksek başarı elde edildiği gözlenmiştir.

Bu sebeplerden metinler üzerinde makine öğrenmesi yaklaşımı ile duygu analizi sırasında NB ve SVM algoritmalarının sıklıkla tercih edildiği düşünülmektedir. Şekil 1'de makine öğrenmesi yaklaşımında kullanılan algoritmaların kullanıldığı çalışmalarda oranları görülmektedir.



Şekil 1. Algoritmaların toplam kullanılma oranları.

4. Hibrit yaklaşımı

Metinler üzerinde yapılan duygu analizi çalışmaları detaylı olarak incelemeye devam edilirken bu kısımda hibrit yaklaşım kullanılarak yürütülen çalışmalar incelenmiştir. Hibrit yaklaşım kullanılarak 2013-2022 yılları arasında yapılmış toplam 14 çalışma (5'i Türkçe 9'u farklı diller) ele alınmıştır.

Hibrit yaklaşım kullanılarak yapılan duygu analizi çalışmalarında ilk olarak Twitter sosyal medya sitesinden elde edilen Türkçe metinler üzerine yapılan çalışmalar incelenmiştir. Bu çalışmalarda kullanılan veri kümeleri, sözlük yaklaşım için kullanılan sözlük, makine öğrenmesi yaklaşımı için kullanılan algoritmalar ve çalışma sonucunda elde edilen sonuçlar aşağıda detaylı olarak anlatılmıştır.

2015 yılında yapılan [55] referanslı çalışmada film yorumlarından ve Twitter sosyal medya sitesinden elde edilen metinlerde duygu analizi yapılmıştır. Bu çalışmada sözlük tabanlı duygu analizi yöntemi ve makine öğrenmesi yaklaşımı yöntemleri ayrı ayrı uygulanarak elde edilen sonuçlar karşılaştırılmıştır. Sözlük tabanlı yöntem için

Türkçe'ye çevrilen Sentistrength sözlüğü kullanılmıştır. Makine öğrenmesi yaklaşımı için de SVM, NB, DT algoritmaları karşılaştırılmıştır. Twitter veri kümesi üzerinde her iki yöntem uygulanarak duygu analizi yapıldığında sözlük tabanlı yöntem kullanıldığında %75,2 iken makine öğrenmesi yaklaşımı yöntemi kullanıldığında SVM algoritması ile %85 başarı elde edilmiştir. Film yorumlarından elde edilen veri kümesinde ise sözlük tabanlı yöntem kullanıldığında %79,5 iken makine öğrenmesi yaklaşımı yöntemi kullanıldığında SVM algoritması ile %89 başarı elde edilmiştir. 2018 yılında yapılan [56] referanslı çalışmada 2018 seçimlerinde Cumhurbaşkanlığı adayı ve parti seçimi tahmini için Twitter verileri kullanılmıştır. Bu çalışmada sözlük tabanlı yaklaşım için duygu skoru hesaplanmıştır ve makine öğrenmesi yaklaşımı için CART, RF ve NB algoritmaları kullanılmıştır. Cumhurbaşkanı adayları ile ilgili verilerin analizi sonucunda en yüksek duygu skoruna sahip aday Muharrem İnce ve duygu skoru %50 olarak elde edilmiştir. Makine öğrenmesi yaklaşımına göre en yüksek doğruluk Selahattin Demirtaş için RF algoritması ile %84 olarak bulunmuştur. Siyasal Partiler ile ilgili çalışma sonucunda en yüksek duygu skoru ortalaması %32 ile Ak parti iken makine öğrenmesi yaklaşımına göre en yüksek doğruluk HDP için RF algoritması ile %86 olarak elde edilmiştir. 2021 yılında yapılan [57] referanslı çalışmada kullanılan veriler Twitter üzerinden elde edilmiştir. Sözlük tabanlı ve Bert tabanlı makine öğrenmesi kullanılmıştır. Sözlük tabanlı duygu analizi için SentiTurkNet sözlüğü kullanılmıştır. Sözlük tabanlı makine öğrenmesi yaklaşımı modeli ile duyarlılık 1 olarak elde edilmiştir.

Hibrit yaklaşım ile hazır veri setleri veya farklı web sitelerinden elde edilen Türkçe metinler üzerinde yapılan duygu analizi çalışmaları aşağıda incelenmiştir.

2013 yılında yapılan [58] referanslı çalışmada duygu analizinde sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımı kullanılarak elde edilen başarı oranları karşılaştırılmıştır. Bu çalışmada kullanılan veriler BeyazPerde.com adlı internet sitesindeki film yorumlarından elde edilmiştir. Sözlük tabanlı yaklaşım için Türkçe'ye çevrilen SentiWordNet sözlüğü kullanılmıştır ve makine öğrenmesi yaklaşımı için SVM, NB, KNN, DT (C4.5) algoritmaları kullanılmıştır. En yüksek F-skoru sözlük tabanlı yaklaşımla %59 ve makine öğrenmesi yaklaşımıyla SVM algoritması ile %82 olarak elde edilmiştir. 2019 yılında yapılan [59] referanslı çalışmada film yorumları ile ilgili veriler beyazperde.com sitesinden, otel yorumları ile ilgili veriler otelpuan.com sitesinden ve tweetlerden elde edilen 3 farklı veri seti kullanılmıştır. Bu çalışmada sözlük tabanlı yaklaşım için genişletilmiş SentiTurkNet sözlüğü kullanılmıştır ve makine öğrenmesi yaklaşımı için NB, SVM ve DT(J48) algoritmaları kullanılmıştır. Hibrit yaklaşım kullanılarak en yüksek doğruluk oranları film veri seti ile NB algoritması ile %88,93 otel veri seti ile SVM algoritması ile %91,96 ve Twitter verileri ile NB algoritması ile %83,37 olarak elde edilmiştir.

Hibrit yaklaşım kullanılarak diğer dillerde olan metinler üzerinde de duygu analizi çalışmaları yapılmıştır. Aşağıda incelenen çalışmalarda kullanılan veriler Twitter sosyal medya sitesinden veya farklı web sitelerden elde edilmiştir. Twitter verileri kullanılarak yapılmış duygu analizi çalışmalarından 2018 yılında yapılan [60] referanslı

çalışmada halka açık olan SemEval 2016 veri seti kullanılmıştır. Bu çalışmada duyarlılık puanları oluşturulmuştur ve SVM algoritması kullanılmıştır. Bu çalışma sonucunda SVM algoritması ile duygu puan vektörü dikkate alınmadan %79,6 olarak doğruluk elde edilirken duygu puan vektörü dikkate alınarak %81 doğruluk elde edilmiştir. 2019 yılında yapılan [61] referanslı çalışmada kullanılan veriler Yunanca tweetlerden elde edilmiştir. Bu çalışmada hazır sözlükler kullanılmıştır. RF, DT ve XGBoost algoritmaları karşılaştırılmıştır. Bu çalışma sonucunda en yüksek doğruluk oranı RF algoritması ile %80 olarak elde edilmiştir. 2022 yılında yapılan [62] referanslı çalışmada sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve hibrit yaklaşım uygulanarak çalışmalar sonucunda elde edilen sonuçlar karşılaştırılmıştır. Bu çalışmada Twitter'daki verilerden oluşan 3 farklı hazır veri seti incelenmiştir. Sözlük tabanlı yaklaşım için SentiWordNet, TextBlob, VADER, SentiStrength, Hu ve Liu, MPQA, KWWSO sözlükleri karşılaştırılmıştır. Makine öğrenmesi için SVM, NB, LR algoritmaları karşılaştırılmıştır. En yüksek doğruluk oranları sözlük tabanlı yaklaşım ile birleştirilmiş veri seti kullanılarak VADER sözlüğü ile %57, makine öğrenmesi yaklaşımı kullanılarak LR algoritması ile %70,2 olarak elde edilmiştir. Hibrit yaklaşım sonucunda elde edilen sonuçlar karşılaştırıldığında en yüksek F-skoru birleştirilmiş veri seti ile TextBlob ve LR algoritması kullanılarak %75,3 olarak elde edilmiştir.

Twitter dışında farklı yollardan elde edilen veriler kullanılarak hibrit yaklaşımla duygu analizi çalışmaları incelendiğinde 2018 yılında [63] referanslı çalışmada kullanılan veriler BookMyShow, IMDB, Rotten Tomatoes, Netflix gibi kaynaklardan manuel olarak elde edilmiştir. SentiWordNet sözlüğü ve makine öğrenmesi yaklaşımı için NB ve Linear SVM algoritmaları kullanılmıştır. En yüksek doğruluk NB algoritması ile %89 olarak elde edilmiştir. 2019 yılındaki [64] referanslı çalışmada LG G7 telefonunun kutudan çıkarılması videosunun yorumlarından elde edilen veriler kullanılmıştır. Sözlük tabanlı yaklaşım için VADER sözlüğü kullanılmıştır ve makine öğrenmesi yaklaşımı için NB algoritması kullanılmıştır. Bu çalışma sonunda doğruluk oranı %79,78 olarak elde edilmiştir. 2020 yılında yapılan [65] referanslı çalışmada veriler Jingdong platformundaki tablet bilgisayarın inceleme verilerinden elde edilmiştir. Manuel etiketli veri setine derin öğrenme modeli ve SVM algoritması uygulanmıştır. Çalışma sonunda en yüksek doğruluk oranı derin öğrenmeye dayalı duygu sınıflandırma yöntemiyle %87 olarak elde edilmiştir. 2020 yılında yapılmış bir diğer çalışma olan [66] referanslı çalışmada cep telefonu incelemeleri ile ilgili Amazon Kaggle veri seti kullanılmıştır. Sözlük tabanlı yöntem için VADER kullanılmış. LR, Hybrid Bag-Boost ve SVM algoritmaları karşılaştırılmış, en yüksek doğruluk Hybrid Bag-Boost algoritması ile %94 olarak elde edilmiştir. 2021 yılında yapılan [67] referanslı çalışmada restoran ile ilgili veriler Kaggle'dan elde edilmiştir. Sözlük tabanlı yaklaşım için VADER sözlüğü kullanılmıştır. LR, Destek Vektör Kümeleme (Support Vector Clustering-SVC), Linear SVC, MNB, XGBoost algoritmaları kullanılmıştır. En yüksek doğruluk XGBoost algoritması ile %67,5 olarak elde edilmiştir. 2022 yılında yapılan [68] referanslı çalışmada Sentiment.140 veri setinde bulunan 400.000 veri incelenmiştir. Sözlük tabanlı yaklaşım için Jaccard benzerlik

indeksi ve öznellik sözlüğü kullanılmıştır. KNN, Tamamlayıcı Naive Bayes (Complement Naive Bayes-CNB), RF, MLP ve SVM algoritmaları kullanılmıştır. Bu çalışma sonucunda sözlük tabanlı yaklaşım kullanılarak %59,5 doğruluk elde edilmiştir. Makine öğrenmesi yaklaşımı ile en yüksek doğruluk oranları eğitim verisi ile RF algoritması ile %98,8 ve test verisi ile SVM algoritması ile %76,7 olarak elde edilmiştir.

Hibrit yaklaşım kullanılarak incelenen çalışmalarda kullanılan veri seti, duygu analizi yapılan metnin dili, veri ön işleme kullanılan yöntemler, sözlük tabanlı yaklaşım için kullanılan sözlük, makine öğrenmesi yaklaşımında kullanılan algoritmalar gibi farklı kriterler ile elde edilen başarı oranları arasında nasıl bir bağlantının olduğu incelenmek istenmiştir.

Bu sebeple farklı kriterler ve elde edilen sonuçlar kullanılarak oluşturulan tablolar ile bu kriterlerin sonuçlar üzerindeki etkileri tespit edilmeye çalışılmıştır.

Hibrit yaklaşımı kullanılarak yapılan duygu analizi çalışmalarında incelenen metnin dilinin, veri setinin büyüklüğünün, elde edilen verilere uygulanan ön işlem adımlarının, kullanılan sözlüklerin ve kullanılan makine öğrenmesi algoritmaların elde edilen başarıya olan etkisi incelenmek istenmiştir. Bu sebeple Tablo 4'te hibrit yaklaşımında farklı kriterler ile elde edilen sonuçlar arasındaki ilişkiler incelenmiştir.

Tablo 4. Hibrit yaklaşımında farklı kriterlerin karşılaştırılması.

| Referans | Dil | Veri Seti | Ön İşleme Adımları | Kullanılan Sözlükler | Kullanılan Algoritmalar | Sonuçlar |
|----------|----------------|--|-------------------------------------|--|--------------------------------|--|
| [55] | Türkçe | Twitter: 4.324 tweet Beyazperde.com: 20.244 yorum | N-gram POS BoW | Sentistrength | SVM, NB, DT | En yüksek F-skoru: Sözlük Tabanlı Yöntem: Twitter için %75,2, yorumlar için %79,5 Makine Öğrenmesi: Twitter için %85 (SVM), yorumlar için %89 (SVM ve NB) En yüksek doğruluk: |
| [56] | Türkçe | Adaylar ile ilgili 95.478 tweet, partiler ile ilgili 134.878 tweet | Terim doküman matrisi oluşturulmuş. | Manuel duygu skoru hesaplanmış sözlük | CART, RF, NB | Duygu skorlarının aritmetik ortalaması: Adaylar: Muharrem İnce için %50 Partiler: Ak Parti %32 Makine Öğrenmesi: Adaylar: Selahattin Demirtaş için %84 (RF) Partiler: HDP %86 (RF) Sözlük tabanlı yaklaşım: Duyarlılık: 1 |
| [57] | Türkçe | 4.575 Twitter verisi | | SentiTurkNet | Bert | Bert tabanlı makine öğrenmesi: Doğruluk %99 |
| [58] | Türkçe | BeyazPerde.com: 2.305 yorum | N-gram TF-IDF, A-TF. | SentiWordNet | SVM, KNN, DT, NB | En yüksek F-skoru: Türkçe görüş sözlüğü: %59, Makine öğrenmesi: %82 (SVM) |
| [59] | Türkçe | Beyazperde.com, otelpuan.com'dan ve Kemik doğal dil işleme (NLP) sitesindeki veriler | | Genişletilmiş SentiTurkNet | NB, SVM, DT(J48) | En yüksek doğruluk: Film yorumları için: %88,93 (NB) Otel yorumları için: %91,96 (SVM) Kemik NLP verileri için: %83,37 (NB) |
| [60] | Türkçe Olmayan | SemEval 2016 Twitter veri seti | N-gram TF, TF-IDF, Binary | Duyarlılık puanları belirlenmiş sözlük | SVM | En yüksek doğruluk: duygu puan vektörü ve SVM ile %81 |
| [61] | Türkçe Olmayan | 1.640 Yunanca tweet | N-gram | Yunanca sözlük | RF,DT, XGBoost | En yüksek doğruluk RF ile %80 |
| [62] | Türkçe Olmayan | Weather Sentiment: 1.000 tweet DecarboNet: 600 tweet Kaggle: 396 tweet | BoW, TF-IDF | SentiWordNet, TextBlob,VADER, SentiStrength, Hu ve Liu, MPQA,WKWSC | SVM, NB, LR | En yüksek doğruluk: Birleştirilmiş veri seti : Sözlük tabanlı yaklaşım: VADER ile %57 Makine öğrenmesi: LR ile %70,2 Hibrit yaklaşım: TextBlob ve LR ile %75,3 |
| [63] | Türkçe Olmayan | BookMyShow,IMDB Rotten Tomatoes, Netflix sitelerinden film yorumları | TF-IDF | SentiWordNet | NB, Linear SVM | En yüksek doğruluk NB ile %89 |
| [64] | Türkçe Olmayan | Youtube'daki LG G7 cep telefonu ile ilgili 6.248 yorum | Sayma Vektörü | VADER | NB | En yüksek doğruluk %79,78 |
| [65] | Türkçe Olmayan | Jingdongi: 14.831 yorum | | Duygu sözlüğü | SVM, Evrişimli Sinir Ağları | En yüksek doğruluk: duygu sözlüğü ve Evrişimli Sinir Ağları ile %87 |
| [66] | Türkçe Olmayan | Amazon'daki cep telefonları yorumları | N-gram | VADER | SVM, LR, Hybrid Bag-Boost | En yüksek doğruluk: Hybrid Bag-Booster ile %94 |

| | | | | | | |
|------|----------------|---|----------------------------|---|-----------------------------------|---|
| [67] | Türkçe Olmayan | Kaggle: 1.000 yorum | BoW, TF-IDF, Sayma Vektörü | VADER | LR, MNB, SVC, Linear SVC, XGBoost | En yüksek doğruluk: XG Boost ile %67,5 |
| [68] | Türkçe Olmayan | Sentiment.140 veri setindeki 400.000 veri | Bow, Sayma Vektörü | Jaccard benzerlik indeksi ve Öznellik sözlüğü | KNN, CNB, RF, MLP, SVM | En yüksek doğruluk: Sözlük tabanlı yaklaşım: %59,5 eğitim verisi: RF ile %98,8 test verisi: SVM ile %76,7 |

Hibrit yaklaşımı kullanılarak incelenen 14 çalışmada kullanılan sınıflandırma algoritmalar çoğunlukla aynı olmasına rağmen her çalışmada farklı sayıda algoritma kullanılarak toplamda 39 tane algoritmadan bahsedilmiştir. Tablo 5'te hibrit yaklaşımda algoritmaların kullanıldığı çalışma sayıları gösterilmiştir.

Tablo 5. Hibrit yaklaşımda algoritmaların kullanıldığı çalışma sayıları

| Kullanılan Algoritma | Toplam Kullanıldığı Çalışma Sayısı |
|--------------------------------------|------------------------------------|
| Bert | 1 |
| CART | 1 |
| CNB | 1 |
| DT | 4 |
| Evrişimli Sinir Ağları | 1 |
| Hybrid Bag-Boost | 1 |
| KNN | 2 |
| Linear SVC | 1 |
| Linear SVM | 1 |
| LR | 3 |
| MLP | 1 |
| NB | 7 |
| RF | 3 |
| SVC | 1 |
| SVM | 8 |
| Topluluk Öğrenme (Ensemble Learning) | 1 |
| XGBoost | 2 |

Tablo 5'teki veriler incelendiğinde sırasıyla en çok SVM, NB algoritmaların kullanıldığı gözlenmiştir. Bu algoritmalar incelendiğinde eğitici makine öğrenmesi algoritmaları oldukları gözlenmiştir. Ayrıca bu algoritmaların özellikleri incelendiğinde metin türündeki veriler ile uyumlu çalıştığı ve yüksek başarı elde edildiği gözlenmiştir. Bu sebeplerden metinler üzerinde Hibrit yaklaşım ile duygu analizi sırasında NB ve SVM algoritmalarının sıklıkla tercih edildiği düşünülmektedir.

5. Sonuç

2013-2022 yılları arasında yapılan metin üzerinden duygu analizi elde edilmesine yönelik çeşitli çalışmalar incelenmiştir. Metin üzerinden duygu analizinin elde edilmesi işleminde sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve hibrit yaklaşım olmak üzere 3 farklı yöntemin kullanıldığı gözlenmiştir. Duygu analizi işleminde kullanılan metinlerin Türkçe veya diğer dillerde olduğuna dikkat edilerek incelemeler yapılmıştır. Duygu analizi sırasında uygulanan her yöntemin özellikleri belirlenerek incelenen çalışmalar bu özellikler dikkate her çalışma uygulandığı yöntem dikkate alınarak karşılaştırılmıştır. Daha sonra ise yöntemler özelliklerine ve elde edilen sonuçlara göre karşılaştırılmıştır.

Sözlük tabanlı yaklaşım kullanılarak yapılan duygu analizi çalışmalarında kullanılan veri setlerinin elde edilen başarıya etkisi incelendiğinde, kullanılan veriler detaylı olarak incelendiğinde veri setindeki veri miktarları ile başarı oranları arasında bir bağlantı olduğu gözlenmemiştir. Metnin dilinin elde edilen sonuçlara etkisi incelendiğinde, Türkçe ve Türkçe olmayan metinlerde yapılan duygu analizi çalışmalarında elde edilen başarılar karşılaştırıldığında metnin dili ile elde edilen sonuçlar arasında bir bağlantı olduğu gözlenmemiştir. Sözlük tabanlı yaklaşım kullanılarak yapılan duygu analizi çalışmalarında kullanılan duygu sözlüğünün elde edilen başarıya etkisi incelenmiştir. Türkçe metinler için yeni Türkçe duygu sözlüğü oluşturularak yapılan duygu analizlerindeki başarı oranının diğer diller için oluşturulan duygu sözlüklerinin Türkçeye çevrilmesi ile yapılan duygu analizlerindeki başarı oranlarından yüksek olduğu gözlenmiştir. Ayrıca yeni Türkçe duygu sözlüğü kullanılarak yapılan duygu analizlerindeki sonuçların diğer dillerdeki metinler üzerinde yapılan duygu analizi sonuçlarından daha başarılı olduğu gözlenmiştir. Bu durum Türkçe dil yapısı incelenerek oluşturulacak daha kapsamlı duygu sözlüklerine ihtiyaç olduğunu göstermiştir.

Makine öğrenmesi yaklaşımı kullanılarak yapılan duygu analizi çalışmalarında farklı miktardaki veri setleri kullanılarak elde edilen sonuçlar karşılaştırıldığında veri miktarları ile elde edilen sonuçlar arasında bir bağlantı gözlenmemiştir. Metnin dilinin elde edilen başarıya etkisi incelendiğinde, metin dili Türkçe olan çalışmalarda başarı oranlarının diğer dillerde olan metinlerdeki başarı oranlarına göre daha düşük olduğu gözlenmiştir. Metin dili Türkçe olan çalışmalarda veri sayısının diğer dillerde olan çalışmalarda veri sayısına göre fazla olmasına rağmen elde edilen başarı oranının daha düşük olduğu gözlenmiştir. Bu durum makine öğrenmesi yaklaşımı kullanılarak duygu analizi çalışmalarında incelenen dilin sonuçlar üzerinde etkili olduğunu göstermiştir. Makine öğrenmesi yaklaşımı kullanılarak incelenen çalışmalarda Türkçe metinlerde konu ile ilgili etiketleme veya kurallı eğitim verileri kullanılarak yapılan çalışmalarda başarı oranlarının daha yüksek olduğu gözlenmiştir. Elde edilen verilere uygulanan ön işlem adımlarının ve kullanılan programların elde edilen başarıya olan etkisi incelendiğinde, birçok çalışmada ön işlem adımlarının benzer olduğu görülmektedir. Bu sebeple bu adımlar ile elde edilen başarı oranları arasındaki bağlantı hakkında tam bir yorum yapılamamasına rağmen birçok çalışmada bu adımların tercih edilmesinde elde edilen başarı üzerinde etkili olduğu düşünülmektedir. Aynı şekilde bu çalışmalarda benzer programlar kullanıldığı gözlenmiştir. Makine öğrenmesi yaklaşımı kullanılarak elde edilen sonuçlar incelendiğinde en başarılı sonuçların

SVM, NB, LR algoritmaları kullanılarak elde edildikleri gözlenmiştir. Ancak aynı çalışmada bütün algoritmalar uygulanmadığı için yine de net bir şekilde bu algoritmaların başarı oranının yüksek olduğunu söylenememektedir.

Hibrit yaklaşım kullanılarak duygu analizi yapılan çalışmalar incelendiğinde bazı çalışmalarda sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımı ayrı ayrı uygulanarak elde edilen sonuçlar karşılaştırılırken diğer çalışmalarda ise her iki yöntem birlikte uygulanmıştır. Bu çalışmalar sonucunda elde edilen sonuçlar farklı kriterlere göre karşılaştırılmıştır. İlk olarak farklı miktardaki veriler kullanılarak elde edilen sonuçlar karşılaştırıldığında veri miktarı ile elde edilen başarı oranları arasında bir bağlantı olduğu gözlenmemiştir. Hibrit yaklaşım olarak sözlük tabanlı yaklaşımın, makine öğrenmesi yaklaşımın ve her iki yöntemin birlikte uygulandığı çalışmalar incelendiğinde sözlük tabanlı yaklaşım ile elde edilen başarı oranlarının diğer iki yöntem ile elde edilen başarı oranlarından daha düşük olduğu gözlenmiştir. Makine öğrenmesi yaklaşımı kullanılarak elde edilen sonuçların sözlük tabanlı yaklaşım kullanılarak elde edilen sonuçlardan daha başarılı olmasına rağmen hibrit yaklaşım kullanılarak elde edilen sonuçlardan daha düşük olduğu gözlenmiştir. Duygu analizi yapılacak olan metnin dili incelendiğinde her iki yöntemin birlikte uygulandığı hibrit yaklaşım kullanılan çalışmalarda incelenen metnin dilinin elde edilen sonuçlar üzerinde etkili olduğu düşünülmektedir. Çünkü Türkçe metinlerde ve diğer dillerdeki metinler üzerinde de hibrit yaklaşım kullanılarak elde edilen sonuçların yakın olduğu gözlenmiştir. Hibrit yaklaşım kullanılarak yapılan duygu analizi çalışmalarında makine öğrenmesi yaklaşımı uygulanırken kullanılan algoritmaların dağılımları incelendiğinde en başarılı sonuçların SVM, NB, RF algoritmaları kullanılarak elde edildikleri gözlenmiştir. Sözlük tabanlı, makine öğrenmesi ve hibrit yaklaşım kullanılarak yapılan duygu analizi çalışmaları incelendiğinde makine öğrenmesi yaklaşımı kullanılarak elde edilen sonuçların sözlük tabanlı yaklaşım kullanılarak elde edilen sonuçlardan daha iyi olduğu gözlenmiştir. Hibrit yaklaşım kullanılarak yapılan duygu analizi çalışmalarında ise diğer iki yaklaşıma göre daha yüksek doğruluk oranları elde edildiği gözlenmiştir. Literatür taraması sırasında hibrit yaklaşım kullanılarak özellikle Türkçe metinlerde olmakla birlikte diğer dillerdeki metinlerde de çok fazla çalışma bulunamamıştır. Bu sebeple hibrit yaklaşım ile ilgili daha iyi yorum yapılabilmesi için daha fazla çalışmaya ihtiyaç duyulmaktadır.

Kaynaklar

- [1] E. Sercan Akgül, C. Ertano ve B. Diri, "Twitter Verileri İle Duygu Analizi", *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, c. 22, sayı. 2, ss. 106-110, 2016, DOI: 10.5505/pajes.2015.37268
- [2] G. Yurtalan, M. Koyuncu ve Ç. Turhan, "A. Polarity Calculation Approach For Lexicon-Based Turkish Sentiment Analysis", *Turkish Journal Of Electrical Engineering & Computer Sciences*, c. 27, sayı. 2, ss 1325-1339, 2019, DOI:10.3906/elk-1803-92
- [3] S. Yıldırım, Y. B. Salman ve S. Ayvaz, "Türkçe Duygu Kütüphanesi Geliştirme: Sosyal Medya Verileriyle Duygu Analizi Çalışması", *Avrupa Bilim Ve Teknoloji Dergisi*, sayı.16, ss. 51-60, 2019, DOI: 10.31590/ejosat.537085
- [4] İ. N. Yoldaş, "Türkçe Metinlerde Duygu Analizi: Sözlük Tabanlı Yaklaşım Ve İnsanların Tepkilerinin Karşılaştırılması", *Eskişehir Türk Dünyası Uygulama Ve Araştırma Merkezi Bilişim Dergisi*, c. 2, sayı.1, ss1-6, 2021.
- [5] M. Önal, "Sözlüksel Benzeşim Yöntemi İle Android Tabanlı Duygu Analizi Uygulaması", Yüksek Lisans Tezi, Mekatronik Mühendisliği Anabilim Dalı, Fırat Üniversitesi, Elazığ, Türkiye, 2017.
- [6] Ö. Demir, A. I. Baban Chawai ve B. Doğan, "Türkçe Metinlerde Sözlük Tabanlı Yaklaşımla Duygu Analizi Ve Görselleştirme", *International Periodical Of Recent Technologies In Applied Engineering*, c.1, sayı. 2, ss. 58-66, 2019, DOI: 10.35333/porta.2019.98
- [7] E. Özdemir Akcan, "Marka İmajı Üzerine Türkçe Duygu Sözlüğü Geliştirme Çalışması", *Acta Infologica*, c. 5, sayı. 2, ss. 415-433, 2021, DOI: 10.26650/acin.908724
- [8] N. Öztürk ve S. Ayvaz, "Sentiment Analysis On Twitter: A Text Mining Approach To The Syrian Refugee Crisis", *Telematics And Informatics*, c. 35, sayı. 1, ss. 136-147, 2018.
- [9] A. John, A. John ve R. Sheik, "Context Deployed Sentiment Analysis Using Hybrid Lexicon", *In 1st International Conference On Innovations In Information And Communication Technology (ICIICT)*, Chennai, India, 25-26 Nisan, 2019.
- [10] N. Tuna, A. Sebatlı Sağlam ve F. Çavdur, "Covid-19 Salgını İle İlgili Paylaşımlar Üzerinde Veri Analizi", *Bilişim Teknolojileri Dergisi*, c.15, sayı.1, ss. 13-23, 2022, DOI: 10.17671/gazibtd.928990.
- [11] E. Özdemir Akcan ve N. Sütütemiz, "Analysis of Brand Perceptions of Covid-19 Vaccines By Sentiment Analysis on Social Media", *Pamukkale University Journal of Social Sciences Institute*, sayı. 49, ss.145-162, 2022, DOI: 10.30794/pausbed.994700
- [12] Y. Ding, B. Li, Y. Zhao ve C. Cheng, "Scoring Tourist Attractions Based On Sentiment Lexicon", *In 2nd Advanced Information Technology, Electronic And Automation Control Conference (IAEAC)*, Chongqing, Çin, 25-26 Mart, 2017.
- [13] C.S. Khoo ve S. B. Johnkhan, "Lexicon-Based Sentiment Analysis: Comparative Evaluation Of Six Sentiment", *Journal Of Information Science*, c.44 sayı.4, ss. 491-511, 2018, DOI: 10.1177/0165551517703514
- [14] M. Çetin ve M. F. Amasyalı, "Eğitici ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi", *In 21st Signal Processing and*

- Communications Applications Conference (SIU)*, Haspolat, Türkiye, 24-26 Nisan, 2013.
- [15] H. Nizam ve S.S. Akın, “Sosyal Medyada Makine Öğrenmesi İle Duygu Analizinde Dengeli Ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması”, *XIX. Türkiye’de İnternet Konferansı*, Bornova, İzmir, Türkiye, 27-29 Kasım, 2014.
- [16] Nalçakan, Y. Bayramoğlu, Ş. S. Tuna, S. Sosyal Medya Verileri Üzerinde Yapay Öğrenme İle Duygu Analizi Çalışması. Technical Report, 2015.
- [17] Ö. Çoban, B. Özyet ve G. Tümüklü Özyer, “Türkçe Twitter Mesajlarının Duygu Analizi”, *In 23rd Signal Processing and Communications Applications Conference (SIU)*, Malatya, Türkiye, 16-19 Mayıs, 2015.
- [18] Ö. Çoban, “Metin Sınıflandırma Teknikleri İle Türkçe Twitter Duygu Analizi”, Yüksek Lisans Tezi, Bilgisayar Mühendisliği Anabilim Dalı, Atatürk Üniversitesi, Erzurum, Türkiye, 2018.
- [19] T. Parlar, E. Sara., S.A. Özel, “Türkçe Twitter Verilerinde Duygu Analizi için Nitelik Seçim Yöntemlerinin Karşılaştırılması”, *In 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya, Türkiye. 15-18 Mayıs, 2017.
- [20] O. Aytuğ, “Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi”, *Yönetim Bilişim Sistemleri Dergisi*, c.3, sayı.2, ss. 1-14, 2017.
- [21] G. Sarıman ve E. Mutaş, “Covid-19 Sürecinde Twitter Mesajlarının Duygu Analizi”, *Euroasia Journal of Mathematics Engineering Natural and Medical Sciences*, c.7, sayı. 10, ss. 137-148, 2020, DOI: 10.38065/euroasiaorg.149
- [22] E. Kumaş, “Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması”, *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, c.2, sayı.2, ss. 1-5, 2021.
- [23] N. Pervan, H.Yalın Keleş, “Sentiment Analysis Using A Random Forest Classifier On Turkish Web Comments”, *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, c. 59, sayı. 2, ss. 69-79, 2017, DOI:10.1501/commua1-2_0000000105
- [24] E. Karamanlı, “Makine Öğrenmesi Algoritmaları Kullanarak, Metin Madenciliği Ve Duygu Analizi İle Müşteri Deneyiminin Geliştirilmesi”, Yüksek Lisans Tezi, İşletme Anabilim Dalı, İstanbul Üniversitesi, İstanbul, Türkiye, 2019.
- [25] M. A. Toçoğlu, A. Çelikten ve A. Alpkoçak, “Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması”, *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, c. 21, sayı. 63, ss. 719-725, 2019, DOI: 0.21205/deufmd.2019216303
- [26] Ö. Çelik ve A. Aslan, “Gender Prediction From Social Media Comments With Artificial Intelligence”, *Sakarya University Journal of Science*, c.23, sayı.6, ss.1256-1264, 2019, DOI: [10.16984/sofenbilder.559452](https://doi.org/10.16984/sofenbilder.559452)
- [27] S. Tuzcu, “Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi İle Sınıflandırılması”, *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, c.1, sayı.2, ss.1-5, 2020.
- [28] Ö. Aktaş, B. Coskuner ve İ. Soner, “Turkish Sentiment Analysis Using Machine Learning Methods: Application on Online Food Order Site Reviews”, *Journal of Artificial Intelligence and Data Science*, c.1, sayı.1 ss.1-10, 2021.
- [29] M. Wongkar ve A. Angdresy, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter”, *In Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Endonezya, 16-17 Ekim, 2019.
- [30] B. Ayan, B. Kuyumcu ve B. Ceylan, “Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi İle Tespiti”, *Gazi University Journal of Science Part C: Design and Technology*, c.7, sayı. 2 , ss. 495-502, 2019, DOI: 10.29109/gujsc.561806
- [31] P.B. Sunitha, S. Joseph ve P.V. Akhil, “A study on the performance of supervised algorithms for classification in sentiment analysis”, *In IEEE Region 10 Conference (TENCO)*, Kochi, Hindistan, 17-20 Ekim, 2019.
- [32] S. Al-Hadhrani, N. Al-Fassam, H. Benhidour, “Sentiment Analysis Of English Tweets: A Comparative Study Of Supervised And Unsupervised Approaches”, *In 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyad, Suudi Arabistan, 1-3 Mayıs, 2019.
- [33] G.R. Kumar, K.V. Sheshanna, G.Anjan Babu, “Sentiment Analysis For Airline Tweets Utilizing Machine Learning Techniques”, *In International Conference on Mobile Computing and Sustainable Informatics EAI/Springer Innovations in Communication and Computing*, Bölüm 75, ss. 791-799, Springer Nature Switzerland AG 2021.
- [34] K. Jayamalini, M. Ponnaivaikko ve J. Kothandan, “A Comparative Analysis Of Various Machine Learning Based Social Media Sentiment Analysis And Opinion Mining Approaches”, *Advances in Mathematics: Scientific Journal*, c.9, sayı.11, ss. 10195-10209, 2020, DOI: <https://doi.org/10.37418/amsj.9.12.1>
- [35] S.G. Kanakaraddi, A.K. Chikaraddi, K.C. Gull ve P.S.Hiremath, “Comparison Study Of Sentiment Analysis Of Tweets Using Various Machine Learning Algorithms”, *In International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, Hindistan, 26-28 Şubat, 2020.
- [36] L. Madloi ve R. Patel, “Twitter Sentiments Analysis Using Machine Learning Methods”,

- International Conference for Emerging Technology (INCET), Belçika, Hindistan, 5-7 Haziran, 2020.*
- [37] S. Zahoor ve R. Rohilla, "Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study", *In International Conference on Advances in Computing, Communication & Materials (ICACCM)*, Dehradun, Hindistan, 21-22 Ağustos, 2020.
- [38] T.H.K. Tusar ve T. Islam, "A Comparative Study Of Sentiment Analysis Using NLP And Different Machine Learning Techniques On US Airline Twitter Data", *In International Conference on Electronics, Communications and Information Technology (ICECIT)*, Dakka, Bangladeş, 14-16 Eylül, 2021.
- [39] M. Tripathi, "Sentiment Analysis of Nepali COVID19 Tweets Using NB SVM and LSTM". *Journal of Artificial Intelligence*, c. 3, sayı. 3, ss. 151-168, 2021, DOI: <https://doi.org/10.36548/jaicn.2021.3.001>
- [40] N. Yadav, O. Kudale, A. Rao, S. Gupta ve A. Shitole, "Twitter Sentiment Analysis Using Supervised Machine Learning". In *Intelligent Data Communication Technologies and Internet of Things*, Singapur, 2021, ss.631-642.
- [41] A. Naresh ve P.V. Krishna, An "Efficient Approach For Sentiment Analysis Using Machine Learning Algorithm", *Evolutionary Intelligence*, c.14, sayı.2, ss. 725-731. 2021, DOI: <https://doi.org/10.1007/s12065-020-00429-1>
- [42] O. Kaynar, Y. Görmez, M. Yıldız ve A. Albayrak, "Makine öğrenmesi yöntemleri ile Duygu Analiz", *In International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Türkiye, 17-18 Eylül, 2016.
- [43] Y. M. Kızılkaya ve A. Oğuzlar, "Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili İle Kıyaslanması", *Karadeniz Uluslararası Bilimsel Dergi*, c.37, sayı. 37, ss. 90-98, 2018, DOI: 10.17498/kdeniz.405746.
- [44] A. Albayrak, "Duygu Analizinde Farklı Vektör Temsil Yöntemleri Ve Sınıflayıcıların Karşılaştırılması", Yüksek Lisans Tezi, Yönetim Bilişim Sistemleri Anabilim Dalı, Sivas Cumhuriyet Üniversitesi, Sivas, Türkiye, 2018.
- [45] R. A. Laksono, K.R. Sungkono, R. Sarno ve C. S. Wahyuni, "Sentiment Analysis Of Restaurant Customer Reviews On Tripadvisor Using Naïve Bayes", *In 12th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Endonezya, 18 Temmuz, 2019.
- [46] S. Hemalatha ve R. Ramathmika, "Sentiment Analysis Of Yelp Reviews By Machine Learning", *In International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, Hindistan, 15-17 Mayıs, 2019.
- [47] M. K. Elhadad, K.F. Li ve F. Gebali, "Sentiment Analysis of Arabic and English Tweets". *In Workshops of the International Conference on Advanced Information Networking and Applications*, Cham, İsviçre, 2019, ss.334-348.
- [48] H. E. İnan, "Otel Yorumlarının Sınıflandırılmasında Makine Öğrenmesi Algoritmalarının Karşılaştırılması ve Eskişehir Otellerine Yönelik Yorumların Duygu Analizi", *20. Ulusal ve 4. Uluslararası Turizm Kongresi*, Eskişehir, Türkiye, 16-19 Ekim, 2019.
- [49] U. Ö. Osmanoglu, O. N. Atak, K. Çağlar, H. Kayhan ve T.C. Can, "Sentiment Analysis For Distance Education Course Materials: A Machine Learning Approach", *Journal of Educational Technology and Online Learning*, c. 3, sayı.1, ss. 31-48, 2020, DOI: 10.31681/jetol.663733
- [50] E. Çelik, D. Dal ve T. Aydın, "Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması", *Avrupa Bilim ve Teknoloji Dergisi*, sayı. 27, ss. 880-889, 2021, DOI: 10.31590/ejosat.905259.
- [51] Ö.F. Göçgün ve A. Onan, "Amazon Ürün Değerlendirmeleri Üzerinde Derin Öğrenme/Makine Öğrenmesi Tabanlı Duygu Analizi Yapılması", *Avrupa Bilim ve Teknoloji Dergisi*, sayı. 24, ss. 445-448, 2021, DOI: 10.31590/ejosat.902674
- [52] B. Shaikh, A. Iyer, M. Koneti ve S. Iyengar, Stock Price Prediction with Sentimental Analysis, *In 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, Hindistan, 20-22 Ocak, 2022.
- [53] Y. Sikhi, S.A. Devi, S. K. Jasti ve M. S. Ram, "Sentimental Analysis Through Speech and Text for IMDB Dataset", *Proceedings of the Fourth International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, Hindistan, 20-22 Ocak, 2022.
- [54] M.I.H. Junaid, F. Hossain, U.S. Upal, A. Tameem, A. Kashim ve A. Fahmin, "Bangla Food Review Sentimental Analysis using Machine Learning", *In 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, ABD, 26-29 Ocak, 2022.
- [55] C. Türkmenoğlu, "Türkçe Metinlerde Duygu Analizi", Yüksek Lisans Tezi, Bilgisayar Mühendisliği Anabilim Dalı, İstanbul Teknik Üniversitesi, İstanbul, Türkiye, 2015.
- [56] Y. M. Kızılkaya, "Duygu Analizi Ve Sosyal Medya Alanında Uygulama" Doktora Tezi, İstatistik Anabilim Dalı, Uludağ Üniversitesi, Bursa, Türkiye, 2018.
- [57] E. Karaahmetoğlu, S. Ersöz, O. Karaahmetoğlu, "Sosyal Ağ Tabanlı Verilerden Faydalanarak Korona Virüs Konulu Duygu Analizi Çalışması", *Ergonomics*, c.4, sayı.1, ss. 47-54, 2021, DOI: 10.33439/ergonomi.824333

- [58] B.İ. Sevindi, “Türkçe Metinlerde Denetimli Ve Sözlük Tabanlı Duygu Analizi Yaklaşımlarının Karşılaştırılması,” Yüksek Lisans Tezi, Bilgisayar Mühendisliği, Gazi Üniversitesi, Ankara, Türkiye, 2013.
- [59] B. Erşahin, Ö. Aktaş, D. Kilinc ve M. Erşahin, “A hybrid sentiment analysis method for Turkish”, *Turkish Journal of Electrical Engineering & Computer Sciences*, c. 27, sayı. 3, ss. 1780-1793, 2019, DOI: 10.3906/elk-1808-189
- [60] S. Naz, A. Sharan ve N. Malik, “Sentiment Classification On Twitter Data Using Support Vector Machine”, *In IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Santiago, Şili, 3-6 Aralık, 2018.
- [61] D. Belevesslis, C. Tjortjjs, D. Psaradelis ve D. Nikoglou, “A Hybrid Method For Sentiment Analysis Of Election Related Tweets”, *In 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*. Pire, Yunanistan, 20-22 Eylül, 2019.
- [62] N. M. Sham ve A. Mohamed, “Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches”, *Sustainability*, c.14, sayı.8, ss. 4723-4751, 2022, DOI: 10.3390/su14084723
- [63] R. Bandana, “Sentiment Analysis Of Movie Reviews Using Heterogeneous Features”, *In 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, Kalkuta, Hindistan, 4-5 Mayıs, 2018.
- [64] V. D. Chaithra, “Hybrid Approach: Naive Bayes And Sentiment VADER For Analyzing Sentiment Of Mobile Unboxing Video Comments”, *International Journal of Electrical and Computer Engineering (IJECE)*, c.9, sayı.5, ss. 4452-4459, 2019, DOI: 1011591/IJECE.v9i5
- [65] M. Zhang, “E-Commerce Comment Sentiment Classification Based On Deep Learning”, *In IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, Çengdu, Çin, 10-13 Nisan, 2020.
- [66] S. K. Pathuri, N. Anbazhagan ve G. Prakash, “Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost Algorithm”, *In 7th International Conference on Smart Structures and Systems (ICSSS)*, Chennai, Hindistan, 23-24 Temmuz, 2020.
- [67] K. Yadav ve M. Patel, “Sentiment Analysis on Restaurant Review using Hybrid Approach”, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, c.9 sayı.5 ss. 1999-2006, 2021, DOI: <https://doi.org/10.22214/ijraset.2021.34737>
- [68] D. Dankhara, “A Review of Sentiment Analysis of Tweets”, 2022, https://www.researchgate.net/publication/360076165_A_Review_of_Sentiment_Analysis_of_Tweets#fullTextFileContent