

Makine Öğrenmesi ve Derin Öğrenmeye Dayalı Duygu Analizinde Metin Temsil Yöntemlerinin Sınıflandırma Başarımına Etkisinin İncelenmesi

Aleyna KAHRAMAN¹, Durmuş Özkan ŞAHİN^{2*}, Dilara BIYIKLI³,
Feyzanur AYTEKİN⁴, Hasan Basri DARGA⁵

Öz

Bu çalışmada, iki farklı veri seti üzerinde çeşitli metin temsil yöntemleri kullanılarak ikili ve üçlü sınıflandırma işlemleri gerçekleştirilmiştir. Metin temsil yöntemleri olarak TF-IDF, GloVe, Word2Vec, FastText ve Bag of Words kullanılmıştır. Makine öğrenimi algoritmalarından Naive Bayes, Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Orman, Yapay Sinir Ağı, En Yakın Komşu Algoritması, Karar Ağacı, XGBoost ve LightGBM uygulanmıştır. Derin öğrenme algoritmaları olarak ise Evrişimli Sinir Ağı, Tekrarlayan Sinir Ağı ve Uzun Kısa Süreli Bellek kullanılmıştır. Elde edilen sonuçlarla, kullanılan metin temsil yöntemleri ve algoritmaların performansları karşılaştırılmıştır. Amazon veri setinde, makine öğrenimi yöntemleri arasında en yüksek doğruluk oranı LightGBM algoritması, derin öğrenme yöntemleri arasında ise TF-IDF ve FastText kullanan LSTM algoritması tarafından elde edilmiştir. IMDb veri setinde, makine öğrenimi yöntemleri arasında en yüksek doğruluk oranı Lojistik Regresyon algoritması, derin öğrenme yöntemleri arasında ise FastText kullanan LSTM algoritması tarafından elde edilmiştir.

Anahtar Kelimeler: Makine öğrenimi, Derin öğrenme, Duygu analizi, Metin temsil yöntemleri, Doğal dil işleme, Metin sınıflandırma.

Examining the Effect of Text Representation Methods on Classification Performance in Sentiment Analysis Based on Machine Learning and Deep Learning

Abstract

In this study, binary and multi-class classification were performed on two different datasets using various text representation methods in conjunction with multiple machine learning and deep learning algorithms. The text representation methods employed include TF-IDF, GloVe, Word2Vec, FastText, and Bag of Words. The machine learning algorithms applied were Naive Bayes, Logistic Regression, Support Vector Machines, Random Forest, Artificial Neural Networks, K-Nearest Neighbors, Decision Trees, XGBoost, and LightGBM. For deep learning algorithms, Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory were utilized. The performance of the text representation methods and algorithms was compared based on the results obtained. In the Amazon dataset, the highest accuracy among the machine learning methods was achieved by the LightGBM algorithm. Among the deep learning methods, the LSTM algorithm using TF-IDF and FastText achieved the highest accuracy. In the IMDb dataset, the Logistic Regression algorithm provided the highest accuracy among the machine learning methods, while the LSTM algorithm using FastText achieved the highest accuracy among the deep learning methods.

Keywords: Machine learning, Deep learning, Sentiment analysis, Text representation methods, Natural language processing, Text classification.

^{1,2,3,4,5}Ondokuz Mayıs Üniversitesi, Bilgisayar Mühendisliği Bölümü, Samsun, Türkiye, aleyna.kahraman@bil.omu.edu.tr
durmus.sahin@bil.omu.edu.tr dilara.biyikli@bil.omu.edu.tr feyzanur.aytekin@bil.omu.edu.tr basri.darga@bil.omu.edu.tr

*Sorumlu Yazar/Corresponding Author

1. Giriş

Son yıllarda internete erişim imkânlarının artması ve sosyal medya platformlarının çoğalmasıyla birlikte büyük miktarda işlenmemiş veri ortaya çıkmıştır. Kullanıcıların bu platformlarda paylaştığı veriler, veri madenciliği ve analitik yöntemler kullanılarak anlamlı bilgi çıkarımı amacıyla analiz edilmektedir. Bu bağlamda öne çıkan çalışma alanlarından biri de duygu analizidir. Duygu analizi, metinlerin olumlu, olumsuz veya nötr olarak sınıflandırılmasını sağlayarak insanların duygusal durumlarını anlamaya yönelik önemli bir araştırma alanıdır.

Duygu analizi, yaklaşık 20 yıl önce ortaya çıkan bir alan olarak önemli gelişmeler kaydetmiş ve ticari uygulamalarda geniş bir kullanım alanına sahip olmuştur. Ancak, (Poria ve ark., 2020) tarafından yapılan çalışmada vurgulandığı gibi, bu alanın olgunlaştığı algısına rağmen hâlâ keşfedilmemiş ve yeterince araştırılmamış yönler bulunmaktadır. Duygu analizi genellikle basit bir olumlu-olumsuz sınıflandırma görevi olarak algılansa da çalışma, bu görevin aslında karmaşık ve çok yönlü olduğunu, insan güdülerini, niyetleri ve bağlamsal nüansları içerdiğini vurgulamıştır. Duygu analizi çalışmalarında duyarlılık analizi, sarkazm analizi ve çok modlu duyarlılık analizi gibi alt alanlardaki çalışmaların önemine işaret edilirken, bu çalışmaların duygu analizi alanında hâlâ çözülmesi gereken sorunlara ışık tuttuğu belirtilmiştir (Poria ve ark., 2020). Bu nedenle çalışma, duygu analizi alanındaki keşfedilmemiş potansiyeli vurgulayarak bu alanda daha fazla araştırma yapma çağrısında bulunmaktadır.

Duygu analizi, özellikle pazarlama, müşteri hizmetleri ve halkla ilişkiler gibi çeşitli alanlarda yaygın olarak kullanılmaktadır. Örneğin, şirketler ürün ve hizmetleri hakkında müşteri geri bildirimlerini analiz ederek, müşteri memnuniyetini artırmaya yönelik stratejiler geliştirebilir. Siyasi partiler ve kampanya yöneticileri, seçmenlerin duygusal tepkilerini ölçerek seçim stratejilerini şekillendirebilir. Aynı şekilde, medya ve eğlence sektörü de izleyici tepkilerini analiz ederek içerik üretiminde bulunabilir.

Makine öğrenimi ve derin öğrenme tekniklerindeki gelişmeler, duygu analizinin doğruluğunu ve verimliliğini önemli ölçüde artırmaktadır. Bu teknolojiler sayesinde, büyük veri kümelerinden karmaşık desenlerin çıkarılması mümkün hale gelmiştir.

1.1. Literatür Araştırması

(Rawat ve ark., 2022) yapmış oldukları çalışmada, aşıyla ilgili atılmış tweetleri kullanarak kamuoyunun aşıyla ilgili duyarlılığını duygu analizi yöntemiyle değerlendirmektedirler. Aşıyla ilgili tweetlerin veri seti Kaggle platformundan alınmıştır. Özel karakterler, hashtag'ler, noktalama işaretleri, bağlantılar ve satır sonları gibi metnin ilgisiz kısımlarını kaldırmak için Spacy kütüphanesi

kullanılmıştır. İlgisiz tüm kelime veya sembollerin kaldırılması için Nexttext kütüphanesi kullanılmıştır. TF-IDF yöntemini kullanarak işlenmiş tweetler vektörlere dönüştürülmüştür. Sınıflandırma işlemi için Bernoulli Naïve Bayes (BNB), Destek Vektör Sınıflandırıcı (SVC) ve Lojistik Regresyon (LR) modelleri kullanılmıştır. Veri setinde 78.771 adet olumlu tweet, 23.993 adet olumsuz tweet bulunmuş ve 125.443 tweet'in çoğunun nötr olduğu tespit edilmiştir. En yüksek doğruluk oranı %97 ile SVC modeli kullanılarak elde edilmiştir. BNB modelinin doğruluğu %81, LR modelinin doğruluğu ise %95 olarak raporlanmıştır.

(Demir ve Bilgin, 2023) tarafından yapılan çalışmada, Türkçe haber metinlerinden elde edilen veriler üzerinde BERT tabanlı dil modelleri ve makine öğrenmesi algoritmaları kullanılarak duygu analizi gerçekleştirilmiştir. BERT tabanlı dil modelleri olarak ALBERT, DistilBERT ve RoBERTa, makine öğrenme algoritmaları olarak ise Naive Bayes, Destek Vektör Makinesi ve Rastgele Orman yöntemleri kullanılmıştır. Veri seti 5000 adet olumlu ve olumsuz cümle içermektedir ve verilerin %90'ı eğitim, %10'u test için kullanılmıştır. Çalışmaların sonuçları incelendiğinde, dil modelleri ile yapılan çalışmaların doğruluk değerlerinin, makine öğrenmesi algoritmalarına göre daha yüksek değerlere ulaştığı görülmüştür. DistilBERT, RoBERTa ve ALBERT dil modellerinin başarımları sırasıyla %80, %80 ve %77 olarak kaydedilmiştir. Makine öğrenmesi algoritmalarının sıralaması ise Naive Bayes, Destek Vektör Makinesi ve Rastgele Orman olup, elde edilen sonuçlar sırasıyla %71, %68 ve %68 olarak tespit edilmiştir.

(Öge ve Kayaalp, 2021) tarafından yapılan çalışmada, IMDB internet sayfasında yer alan film yorumlarından oluşan bir veri seti kullanılmıştır. Veri setinde 25.000 adet pozitif ve 25.000 adet negatif olarak etiketlenmiş yorum bulunmaktadır. Python programlama dilindeki kütüphaneler kullanılarak Naive Bayes, Lojistik Regresyon, Support Vector Machine, Decision Tree, Random Forest ve Artificial Neural Network sınıflandırma algoritmalarının duygu analizi kapsamında performans karşılaştırması yapılmıştır. Bag of Words, TF-IDF, Word2Vec ve FastText metin temsil yöntemleri kullanılarak temizlenmiş veri seti sayısal olarak temsil edilmiştir. Veri setinin eğitimi ve test edilmesi aşamasında $k=5$ olacak şekilde k -fold cross validation tekniği kullanılmıştır. 6 farklı sınıflandırma yöntemi için elde edilen sonuçlar doğruluk, kesinlik, geri çağırma ve F1 score hesaplanarak ayrıntılı bir karşılaştırma yapılmış ve sonuçlar kaydedilmiştir. Sonuçlar karşılaştırıldığında en iyi performans, 0.87 doğruluk, 0.87 kesinlik, 0.88 geri çağırma değerleriyle 9.7 saniyede Word2Vec ile Lojistik Regresyon yöntemi kullanılarak elde edilmiştir.

(Rumelli ve ark., 2019) tarafından Türkçe metinler için bir duygu analizi modeli geliştirilmiştir. Bu model geliştirilirken, sözlük tabanlı yaklaşımlar ile makine öğrenmesi algoritmaları birlikte kullanılmıştır. Duygu analizinin en temel yöntemi olarak, cümlenin içindeki her kelimenin sözlükteki puan değerine bakarak toplama dayalı bir model ile hesaplama yapılmıştır. Makine öğrenmesi modelleri, metinlerin polarite puanlarını baz alarak doğru sonuçlar bulan duygu analizi yapmak için

eğitilmiştir. Projede NB, RF, SVM ve k-NN algoritmaları olmak üzere dört farklı makine öğrenmesi yöntemi kullanılmıştır. NB ve SVM için R dilindeki 'e1071' kütüphanesi kullanılmıştır. SVM algoritması için maliyet çekirdeği ve gama değerlerini değiştirme olanağı bu paket tarafından sağlanmıştır. Projede radyal, lineer ve polinom çekirdekler değiştirilerek model hayata geçirilmiş ve RF için "randomForest" isimli paket kullanılarak 'tree' parametresi 100, 425 ve 750 olarak değiştirilerek model eğitilmiştir. kNN algoritması için 'class' kütüphanesi kullanılmış ve k değeri sadece tek sayılar kontrol edilerek 3'ten 71'e çıkarılmıştır. Son olarak doğrulama setindeki performansı yüksek olduğundan k değeri 33 olarak belirlenmiştir. Çapraz doğrulamanın, makine öğrenimi yöntemlerinin değerlendirilmesi üzerinde önemli bir etkisi olduğu belirtilmiştir. Bu nedenle, 100 defa çapraz doğrulama yapılmıştır. Veri setinin %80'i eğitim, %20'si test verisi olarak bölünmüştür. Tüm modellerin ortalama doğruluğunun 0.73 civarında değiştiği görülmüştür. Bu da modellerin doğrulukları arasında çok büyük bir fark olmadığını göstermektedir. NB ve kNN 1 dakikadan az sürede işlemi tamamlarken, SVM ve RF'nin daha uzun süreye ihtiyaç duyduğu gözlemlenmiştir.

(Poornima ve Priya, 2020) tarafından yapılan çalışmada, Twitter verileri kullanılarak cümle sınıflandırmasında SVM, Multinomial Naive Bayes ve Lojistik Regresyon algoritmalarının performans karşılaştırması yapılmıştır. Twitter'dan alınan ham tweetler genellikle gürültülü bir veri kümesine neden olmaktadır. Çeşitli yöntemler kullanılarak veri seti temizlenmiştir. Bigramlar kullanılarak veri kümesinden özellikler çıkarılmıştır. Bu özellikler, sınıflandırmada olumsuzluğu modellemek için etkili bir yol sağlamıştır. Lojistik regresyonun doğruluğunun %86.23, destek vektör makinesinin doğruluğunun %85.69 ve Multinomial Naive Bayes'in doğruluğunun %83.54 olduğu gözlemlenmiştir.

(Sharma ve ark., 2023) tarafından yapılmış olan çalışmada, çok terimli Naive Bayes, Destek Vektör Makineleri ve Lojistik Regresyon'dan yararlanarak duygu analizi çalışmalarında kullanılan tüm makine öğrenimi yaklaşımları için araştırma yapılmış, karşılaştırma analizi yapılmış ve performans değerlendirme verileri sunulmuştur. CNN ve LSTM gibi derin öğrenme modellerinin, fikir madenciliği ve duygu sınıflandırma görevlerini gerçekleştirmek için kullanıldığında diğer modellerden daha iyi performans gösterdiği görülmüştür. Makine öğrenimi modeli olan Rastgele Orman (RF) ve derin öğrenme modeli olan RNN'in performansının benzer olabileceği, ancak sınıflandırıcıların veri kümelerine göre değişen eğilimlerinin olduğu tespit edilmiştir. Sonuçlara göre en iyi doğruluk oranı %93.5 ile CNN kullanılarak elde edilmiştir. Aynı veri kümesinde yapılan çalışmada, LSTM'nin %77.75 doğruluğa sahip olduğu gözlemlenmiştir.

(Joshi ve ark., 2023) tarafından yapılan çalışmada, duygusal analiz ve kategorizasyon için Twitter verileri kullanılmıştır. Duygu analizinin uygun veri görselleştirmesi ve ön işleminin gerekli olduğu belirtilmiştir. Makine öğrenimi algoritması kullanılarak olumlu veya olumsuz duyguların

analizi yapılmıştır. Naive Bayes sınıflandırma algoritması kullanılmıştır. Verilerin kesinlik değerlerinin RF yönteminde %75, karar ağacı yönteminde %75.36, eğitimde %84.93 ve testte %78.58 olduğu sonucuna ulaşılmıştır.

(He ve ark., 2023) tarafından yapılan çalışmada, çevrimiçi ürün deneyimi elde etmeyi amaçlayan ve metin analiz tekniklerini makine öğrenimi algoritmalarıyla birleştiren bir füzyon duyarlılığı analizi yöntemi geliştirilmiştir. Yöntem temel olarak üç adımdan oluşmaktadır. İlk olarak, duygu sözlüğünün duygusal bilgiye duyarlılığından esinlenerek, sözlük duygu özellikleri çıkarılmıştır. Daha sonra, incelemelerin duyarlılık kutuplarını belirlemek için SVM algoritması kullanılmıştır. Buna dayanarak, duyarlılık konuları LDA modeli aracılığıyla incelemelerden çıkarılmıştır. Ayrıca, duygusal bilgilerin atlanmasını önlemek için sözlük anlamsal benzerliğe dayalı olarak genişletilmiştir. Tüketicilerin Amazon'daki çevrimiçi kitap okuma deneyimlerinin araştırılması, yöntemin uygulanabilirliğini ve geçerliliğini doğrulamıştır. Sonuçlar, yöntemin incelemelerin duygusal eğilimlerini doğru bir şekilde belirlediğini ve incelemelerden okuma deneyimlerini etkileyen unsurları yakaladığını göstermektedir. Genel olarak, araştırmanın çevrimiçi ürün deneyimini araştırmak ve müşteri taleplerini takip etmek için etkili bir yol sağladığı, gelecekteki ürün iyileştirme ve pazarlama stratejisi optimizasyonunu güçlü bir şekilde desteklediği görülmüştür.

(Dhariwal ve ark., 2023) tarafından yapılan çalışmada, sesli duygu analizi araçlarından transkripsiyon hizmetlerine ve doğal dil işleme tekniklerine kadar geniş bir yelpazede kullanılan hesaplamalı bir yaklaşım olan "çatallama ve karıştırma" yönteminin önerilmesi amaçlanmıştır. Metin duyarlılığı analizi sözlüklerini, AssemblyAI ve VADER gibi, ses verilerinden duyguları çıkarmak ve kategorilere ayırmak için kullanarak gerçek zamanlı duyarlılık analizi modeli oluşturmak hedeflenmiştir. Bu analizin sonuçları, tüm Hindistan Radyosu'nun 'Akashwani' tarafından yapılan radyo yayınlarında ifade edilen duygulardaki kalıpları ve eğilimleri ortaya koyarak haber servislerinin duygu analizi konusundaki yöntemlerin geliştirilmesi amaçlanmıştır. Metin duyarlılığı analizi ve sesli duyarlılık analizi arasındaki avantaj ve dezavantajları değerlendiren araştırma, bu iki yöntemin birleşimiyle daha kapsamlı ve etkili bir sonuç elde edilebileceğini ortaya koymuştur. Ayrıca, bu çalışmanın gelecekte afet yönetimi gibi gerçek dünya sorunlarına çözüm bulmak ve radyo sinyallerinin kullanıldığı iletişimdeki rolünü güçlendirmek adına önemli bir araştırma kaynağı olabileceği vurgulanmıştır. Ancak, analiz modellerini çalıştıran bağımsız bir sistem geliştirmenin zorlukları, modelin eğitildiği veri setinin kalitesi ve çeşitliliği, afet yönetimi durumlarında kullanılacak olan NVV veri setlerinin ihtiyaçları gibi konular da ele alınması gerektiği sonucuna varılmıştır.

(Li ve ark., 2020) tarafından yapılan çalışmada, canlı yorumlama platformu Danmaku üzerine odaklanılmıştır. Danmaku, kullanıcıların videoları izlerken gerçek zamanlı olarak yorum yapabildikleri bir sistemdir. Ancak, mevcut duygu sınıflandırma yöntemleri Danmaku verileri için

uygun değildir. Bu sorunu çözmek için makalede bir duygu sözlüğü oluşturulmuş ve Naive Bayes tabanlı yeni bir yöntem önerilmiştir. Tencent Video'dan seçilen Danmaku verileri üzerinde yapılan testler, önerilen SDD-NB modelinin diğerlerine göre %14.8 daha yüksek doğruluk oranına sahip olduğunu göstermiştir.

(Fu ve ark., 2018) tarafından yapılan çalışmada, duygu analizi için kullanılan LSTM'lerin başarılı bir yöntem olarak vurgulandığı ancak yalnızca kelime gömmelerinin yetersiz olduğu belirtilmiştir. Bu eksikliği gidermek amacıyla LE-LSTM adlı bir sözlük ile geliştirilmiş LSTM modeli önerilmiştir. GloVe kullanılarak yapılan deneylerde, LE-LSTM'nin standart LSTM modeline kıyasla %2 daha yüksek doğruluk elde ettiği gözlemlenmiştir. Ayrıca, duygu kelimesinin dil bilimsel rolünü modellemek için kayıp fonksiyonuna düzenleyici eklenmesi ve farklı boyutlardaki CNN mimarisi de metin kategorizasyonu için önerilmiştir. Gerçekleştirilen deneylerde, LE-LSTM, Bi-LSTM ve Tree-LSTM modellerinin tüm veri kümelerinde standart LSTM'den daha iyi performans sergilediği gösterilmiştir. Bu modellerin test edildiği beş veri kümesi (IMDB, Yelp2013, NB4000, MR ve Book4000) üzerinde elde edilen sonuçlar tarafından, özellikle LE-LSTM'nin diğer modellere göre daha etkili olduğunu ortaya konmuştur. Bu önerilen model, kelime gömülerinin anlamsal bilgi taşıma sorununu çözmek ve duygu analizi performansını artırmak için etkili bir yöntem sunmaktadır.

(Wang ve ark., 2021) tarafından yapılan çalışmada, geleneksel kelime gösterimine duygu bilgisi ekleyememe sorununu çözmek amacıyla Word2Vec ve GloVe temelli RGWE yöntemi önerilmektedir. RGWE yöntemiyle duyarlılık, konum ve POS özelliklerini ekleyerek duygu analizinin doğruluğunun artırılması hedeflenmektedir. Yapılan deneylerde, RGWE'nin SSWE ve Seninfo+TF-IDF'ye göre daha iyi bir performans sergilediği görülmüştür. İkili sınıflandırma, üçlü sınıflandırma ve CNN için sırasıyla %89.86, %69.1 ve %48.85 ortalama F1 değerleri elde edilmiştir. Bu başarı, SemEval, SST1, SST2, IMDB, Amazon ve Yelp gibi altı klasik veri kümesi üzerinde gözlemlenmiştir. RGWE, duygusal içerikli metin analizi için güçlü bir kelime gömme teknolojisi olarak öne çıkmaktadır.

(Zhang ve Zheng, 2016) tarafından yapılan çalışmada, Çince metinler üzerinde duygu analizi için TF-IDF ile ağırlıklı özellik hesaplaması kullanılmıştır. SVM ve ELM (Extreme Learning Machine) çekirdekli modeli, eğitim örnek kümesini öğrenerek test kümesinde duygu sınıflandırma problemini çözmek için kullanılmıştır. Çalışmada, ELM'nin yüksek verimliliği ve uygulanabilirliği ile duygu analizi alanında önemli bir çözüm sunduğu vurgulanmıştır. Metin dosyaları kullanılarak oluşturulan veri setinde, gürültü temizleme, segmentasyon ve durdurma kelimesi filtreleme gibi adımlar gerçekleştirilmiştir. TF-IDF kullanılarak belgeler arasındaki fark belirlenmiş ve MATLAB ortamında simülasyonlar yapılmıştır. Deney sonuçlarıyla, çekirdekli ELM'nin Çince metinlerde duygu analizi uygulamalarında SVM'ye kıyasla daha etkili olduğu gösterilmiştir. Hem doğruluk oranları hem de eğitim-test süreleri açısından çekirdekli ELM, avantajlı bir performans sergilemiştir.

Bu çalışma ile Çince metinlerde duygu analizinde yeni yaklaşımların ve çözümlerin geliştirilmesine katkı sağlanmıştır.

(Xu ve Song, 2023) yapmış oldukları çalışmada, metin duygu sınıflandırması alanında geleneksel ve derin öğrenme modellerini aşan yeni bir yaklaşım sunmuştur. Çalışmada, BERT ile birleştirilmiş iki hibrit sinir ağı modeli tanıtılmış ve bu modeller Chnsenticrop ve Weibo veri kümeleri üzerinde eğitilmiştir. Veriler, temizlenmiş ve SST-2 formatına dönüştürülmüş olup, 80:20 oranında eğitim-test setlerine ayrılmıştır. Kelime tabanlı metinler, 512 boyutlu Word2Vec vektörleri kullanılarak temsil edilmiştir. Çeşitli yöntemler, örneğin CNN, LSTM, GRU, Bert+BiLSTM ve Bert+CNN karşılaştırılmış ve en etkili yaklaşım belirlenmiştir. Bert+CNN modelinin, diğerlerine göre daha yüksek doğruluk, geri çağırma ve F1 skorlarına sahip olduğu gösterilmiştir. LSTM'nin en yüksek geri çağırma oranını, GRU'nun ise en yüksek F1 puanını elde ettiği belirtilmiştir. Analiz sonuçları, metin duygu sınıflandırmasında Bert+CNN modelinin etkili bir performans sergilediğini ortaya koymaktadır.

(Rizk ve Asal, 2021) tarafından yapılan çalışmada, çeşitli makine öğrenimi, sinir ağları ve kelime gömme modellerini değerlendiren kapsamlı bir çalışmanın özeti sunulmuştur. Araştırmada, BERT ile LSTM-çift yönlü kodlama modelinin en yüksek test doğruluğuna ulaştığını ve kelime gövdeleri ile derin öğrenme mimarilerinin geleneksel modellere göre üstün performans sergilediğini gösterilmiştir. Ayrıca, Word2Vec'in farklı modellerle birleştirilmesini inceleyen çalışmada, bu hibrit yaklaşımın %89.2'lik bir doğruluk oranıyla diğer modellere göre daha etkili olduğu ortaya konmuştur. DistilBERT adlı önceden eğitilmiş bir dil modelinin tanıtıldığı ve orijinal BERT'e kıyasla %40 daha küçük, %97 oranında bilgi koruyabilen ve %60 daha hızlı olduğu kanıtlanmıştır. Araştırmada, IMDB film incelemeleri üzerinde yapılan denemelerde, Transformatörden Çift Yönlü Kodlayıcı Temsilleri (BERT) gömülü Uzun Kısa Süreli Bellek (LSTM) modelinin %93'lük doğrulukla en iyi performansı sergilediği sonucuna ulaşılmıştır. Bu bağlamda, DistilBERT'in daha küçük boyutta ve hızlı çalıştığı vurgulanmıştır.

(Hammad ve Anwar, 2019) yapmış oldukları çalışmada, Sindhi dilinde duygu analizi gerçekleştirmek için denetimli makine öğrenimi tekniklerinden faydalanmışlardır. Sindhi tweetlerini içeren bir veri kümesi, önceden tanımlanmış sözlüksel veri tabanı ve Google Translate API yardımıyla etiketlenmiştir. Veri kümesi, temizlenmiş ve tokenizasyon işleminden geçirilmiştir. Duygu analizi için SVM, NB, DT ve k-NN algoritmaları kullanılmıştır. Elde edilen sonuçlar, Sindhi tweet veri setinde DT ve k-NN algoritmalarının iyi bir performans sergilediğini, SVM'nin bunları takip ettiğini göstermektedir. Özellikle SVM'nin, doğruluk, kesinlik, geri çağırma ve F1 puanı ölçütlerinde yüksek bir performans sergilediği belirtilmiştir. Veri kümesi boyutları arttıkça doğruluk değerlerinin arttığı, ancak kesinlik ve geri çağırma değerlerinin değişkenlik gösterdiği gözlemlenmiştir. Bu bağlamda, Sindhi dilinde duygu analizi için denetimli makine öğrenimi

yöntemlerini değerlendiren bu araştırmada, özellikle SVM'nin yüksek performans gösterdiğini vurgulanmıştır.

(Pandya ve ark., 2021) yapmış oldukları çalışmada, makine öğrenimi ve derin öğrenme modelleri kullanarak Twitter duygu analizi yapmayı hedeflemişlerdir. 1.6 milyon tweet, bu çalışmanın veri kümesi olarak kullanılmış ve tweetler olumlu veya olumsuz duygulara göre sınıflandırılmıştır. Bu amaçla Naive Bayes ve Destek Vektör Makineleri gibi geleneksel makine öğrenimi modelleri kullanılmıştır. Makine öğrenimi ve derin öğrenme modellerini eğitmek için Sentiment140 veri kümesi kullanılmış olup, bu küme 800.000 negatif ve 800.000 pozitif tweet içermektedir. Veriler, ön işleme tabi tutulmuş; tweetlerden kullanıcı adı, bağlantı, noktalama işaretleri, rakamlar ve stopwords gibi ifadeler kaldırılmış ve tüm harfler küçük harfe dönüştürülmüştür. Daha sonra bu veriler köklerine ayrıştırılmıştır. Elde edilen veri kümesi görselleştirilerek negatif ve pozitif kelime bulutu oluşturulmuştur. Veri kümesi kelimelerden oluştuğu için, vektöre dönüştürmek amacıyla Word2Vec modeli kullanılmıştır. Veri kümesi %80 eğitim ve %20 test olarak iki ayrı kümeye dönüştürülmüştür; toplam tweet sayısı üzerinden 1.28 milyon tweet eğitim için, 320.000 tweet test için kullanılmıştır. Model eğitimlerinde Naive Bayes (NB), Destek Vektör Makineleri (SVM), RNN, LSTM ve Bi-LSTM kullanılmıştır. Sonuçlar doğruluk, F1 skoru, geri çağırma skoru ve hassasiyete göre karşılaştırılmıştır. Destek Vektör Makineleri %76.53 doğrulukla daha iyi performans sergilemiştir. Ancak, derin öğrenme modelleri makine öğrenimi modellerine göre daha iyi performans gösterdiği belirtilmiştir.

(Singh ve ark., 2022) tarafından yapılan çalışmada, TF-IDF ve makine öğrenimi algoritmaları kullanılmıştır. Veri kümesi olarak ABD Havayolları duyarlılık tweetleri kullanılmıştır. Tweetler olumlu, olumsuz ve nötr olmak üzere üç kategoride sınıflandırılmıştır. Veriler önce ön işleme tabi tutulmuş; gereksiz kelimeler ve noktalama işaretleri gibi karakterler silinmiş ve büyük harfler küçük harflere dönüştürülmüştür. Geriye kalan metin daha küçük parçalara bölünerek tokenizasyon işlemi uygulanmıştır. Daha sonra sözcüklerin köküne indirgemek için stemming işlemi yapılmıştır. Kalan verilere TF-IDF yöntemi uygulanarak kelimelerin uygulanma sıklığı elde edilmiştir. Veri kümesi, %80 eğitim ve %20 test olarak ayrılmıştır. Bu çalışmada RF, GB, XGBoost ve SVM olmak üzere dört sınıflandırma tekniği doğruluk, hassasiyet, hatırlama ve F1 skoru ölçütlerine göre karşılaştırılmıştır. SVM'nin %83.74 oranı ile en yüksek doğruluğu elde ettiği görülmüştür. Aynı zamanda, %84 hassasiyet ve %87.85 F1 skoru ile diğer tekniklerden daha iyi sonuç verdiği tespit edilmiştir.

(Kim, 2021) yapmış olduğu çalışmada, müşteri yorumlarının inovasyon yöneticileri için önemli bir kaynak olduğunu vurgulayarak, duygu analizinin bu yorumlardan elde edilen istatistiksel verilerle mümkün olduğu bir perspektife odaklanmıştır. Metin madenciliği yöntemlerini kullanarak müşteri duyarlılığını değerlendirmek amacıyla çeşitli ölçütler önerilmiş ve Linguistic Inquiry and Word

Count (LIWC) gibi metin madenciliği paketleri üzerindeki alternatifler incelenmiştir. Duyarlılık analizi için pozitif ve negatif eğilimli kelimelerin sayılarına dayanan bir yaklaşım benimsenmiş ve bu yöntemlerin sınırlamaları tartışılmıştır. Çalışmada, duyarlılık puanları ile çevrimiçi inceleme derecelendirmeleri arasındaki dağılım farkını göstererek özellikle çevrimiçi derecelendirmelerin öznel olduğunu vurgulanmıştır. Ayrıca, pazarlamacıların müşteri görüşlerini değerlendirmek ve karşılaştırmak için duyarlılık puanlarını kullanabileceği alternatif yollar sunulmuştur. Sonuç olarak, metin madenciliği ve duygu analizinin pazarlamacılara müşteri geri bildirimlerini etkili bir şekilde değerlendirme ve inovasyon performanslarını ölçme konusunda yardımcı olabileceği öne sürülmüştür.

(Chandra ve Jana, 2020) yapmış oldukları çalışmada, duygu analizi konusundaki araştırmalara ve kullanılan yöntemlere odaklanılarak insanların görüşlerini, duygularını ve tutumlarını belirli konularla ilişkilendiren bir alana dikkat çekilmiştir. Makine öğrenimi ve derin öğrenme tekniklerinin dil işleme istatistiksel yöntemleriyle birleştirilmesi ile duygu analizinin gelişimine önemli katkılar sağlanmıştır. Sosyal medyanın etkisi ve kullanıcıların duygularını dijital medya aracılığıyla ifade etmeleriyle, özellikle Twitter gibi platformlarda zengin bir veri kaynağı oluşturulmuştur. Duygu analizi için kullanılan çeşitli yöntemler ve teknikler, duyarlılık sözlüklerinin algoritmik oluşturulması, duyarlılık endeksi oluşturma ve önem değerlendirmesi gibi aşamaları kapsamaktadır. Çalışmada, bu yöntemlerin yanı sıra derin öğrenme modelleri ve makine öğrenimi sınıflandırıcıları kullanarak gerçekleştirilen duygu analizi için önerilen yaklaşımlar detaylı bir şekilde açıklanmıştır. Derin öğrenme modelleriyle yapılan deneylerde, çeşitli modellerin performansları karşılaştırılmış ve LSTM-CNN'nin en iyi sınıflandırma sonuçlarını verdiği sonucuna varılmıştır. Deneyler ve sonuçların analizi bölümünde, Twitter API ve çeşitli konulardaki veri kümeleri kullanarak gerçekleştirilen deneylerin sonuçlarına odaklanılmıştır. Makine öğrenimi modellerinin doğruluk yüzdesinin %81 ile %90 arasında kaldığı belirtilirken, derin öğrenme modellerinin %85 ile %97 aralığında daha iyi sonuçlar elde ettiği vurgulanmıştır. Sonuç ve gelecek kapsamı bölümünde, derin öğrenme modellerinin daha iyi performans gösterdiği ve gelecekte bu alanda yapılacak çalışmaların bu modellerin tasarımı ve uygulanması üzerine odaklanabileceği sonucuna varılmıştır. Sosyal medya verilerinin analizinin, işletmelere bilgilendirici iç görüler ve gelir sağlama potansiyeli sunduğu belirtilmiştir.

(Saraswat ve ark., 2023) yapmış oldukları çalışmada, ses dosyalarının makine öğrenimi kullanılarak duygu analizi yapılması ve ses verilerinin metinsel sınıflandırma ile analiz edilmesi hedeflenmiştir. Duygu analizi, genellikle metin madenciliği teknikleri ile işlenen verilere odaklanmıştır. Ancak, ses verilerinin duygu analizi alanındaki araştırmalar hala erken aşamadır. Ses verilerinin analizinde kullanılan temel özellikler arasında ses frekansı, konuşma öznitelikleri ve zamansal öznitelikler bulunmaktadır. Projenin temelini oluşturan sistem mimarisi, ses dosyalarından

metin transkriptleri oluşturmayı içermektedir. Bu transkriptler, daha sonra çeşitli metin madenciliği araçları kullanılarak işlenir. Metin verileri analize hazırlanırken önce ses dosyalarının metne dönüştürülmesi, ardından ön işleme adımları ile belirli bir özellik kümesinin çıkarılması gerçekleştirilir. Son olarak, denetimli öğrenme teknikleri kullanılarak metin transkriptleri kategorize edilir. Naïve Bayes, Lojistik Regresyon ve Sayım Vektörleştirici gibi çeşitli sınıflandırma algoritmaları kullanılarak duygu analizi gerçekleştirilir. Naïve Bayes sınıflandırıcısı kullanılarak elde edilen sonuçlar, özellikle pozitif ve nötr duyguların tanınmasında yüksek doğruluk göstermektedir. Duygu analizörü, çeşitli duyguları tanıma doğruluğuyla şu sonuçları vermektedir: Neşe %94, Üzüntü %93, Öfke %85, Korku %74, İğrenme %70, Utanç %56, Sürpriz %52. Duygu analizörünün başarı oranının, güçlü duygusal değerlere sahip kelimelerin konuşmada belirgin bir şekilde kullanılmasına bağlı olduğu gözlemlenmiştir. Bu kelimeler, tahmin skorunu ve doğruluğunu etkilemede önemli bir rol oynamaktadır. Model, 34.000'den fazla veri noktası üzerinde eğitildiği için %86'lık bir doğruluk sağlamaktadır, bu da diğer ses duygu analizörlerine kıyasla (metinsel sınıflandırma tabanlı ve harici donanım cihazına ihtiyaç duymadan) oldukça yüksektir. Daha büyük veri setleri ve gelişmiş veri temizleme teknikleri kullanılarak gelecekteki çalışmaların, duygu analizi alanında daha kesin ve güvenilir sonuçlara ulaşabileceği belirtilmiştir.

(Can ve Alataş, 2017) çalışmasında, sosyal medyanın ve internet üzerindeki sosyal ağların giderek daha fazla önem kazanmasıyla birlikte popülerleşen duygu analizi konusu genel hatlarıyla ele alınmıştır. Bu kapsamda, duygu analizi üzerine yapılan araştırmalar incelenmiş, kullanılan algoritmalar ve uygulamalar detaylandırılmıştır. Güncel bir değerlendirme yapabilmek adına son yıllarda gerçekleştirilen çalışmalar analiz edilmiş ve belirli kategorilere ayrılarak sunulmuştur. Bu araştırmalar, duygu analizi tekniklerini gerçek dünya problemlerine uygulayarak alana önemli katkılar sağlamıştır. İncelemeler sonucunda, duygu sınıflandırması ve özellik seçimi algoritmalarının geliştirilmesinin halen önemli bir araştırma alanı olduğu görülmüştür. Ayrıca, duygu sınıflandırma problemlerinin çözümünde en yaygın kullanılan makine öğrenme algoritmalarının Naive Bayes ve Destek Vektör Makineleri olduğu tespit edilmiştir. Yapılan analizler, İngilizce dilinde birçok çalışma bulunduğunu ve dolayısıyla geniş bir kaynak havuzunun mevcut olduğunu ortaya koymuştur. Ancak, İngilizce dışındaki dillerde, özellikle de Türkçe alanında, kaynak ve araştırma eksikliği göze çarpmaktadır. İngilizceden sonra en yaygın kullanılan veri kaynağı olarak WordNet öne çıkmaktadır. Son dönemde duygu analizi çalışmalarında mikro-bloglar, bloglar ve forumlar gibi yeni bilgi kaynaklarının kullanılmaya başlandığı görülmektedir. Bu platformlar, insanların güncel konular hakkındaki düşüncelerini ve duygularını anlamak açısından büyük bir öneme sahiptir. Daha derinlemesine analizler için ise sosyal ağ siteleri ve mikro-bloglar kritik veri kaynakları olarak değerlendirilmektedir. Bunun yanı sıra, algoritmaların test edilmesi amacıyla kullanılabilecek çeşitli veri setleri de bulunmaktadır. Genel olarak bakıldığında, duygu analizi alanında hâlâ tartışmaya açık

pek çok konu olduğu ve özellikle Türkçe üzerine yapılan çalışmaların yetersiz kaldığı sonucuna ulaşılmıştır.

(Satrya ve ark., 2022) tarafından yapılan çalışmada, Twitter üzerindeki kripto para birimleri ile ilgili olumlu ve olumsuz eğilimlerin incelenerek, toplumun bu dijital varlıklara yönelik tutumlarının anlaşılması amaçlanmıştır. Duygu analizi veya fikir madenciliği yöntemleri kullanılarak Twitter'daki veriler sınıflandırılmış ve olumlu ve olumsuz eğilimler ortaya çıkarılmaya çalışılmıştır. Bu analiz, özellikle Bitcoin'e odaklanarak kripto para birimlerinin genel algısını belirleyebilir ve yatırımcılar ile işletmeler için önemli bir referans olabilir. Araştırmacılar, daha önceki çalışmaların incelenmesine dayanarak bu çalışmada SVM algoritmasını kullanmışlardır. Literatürde SVM'nin duygu analizi konusunda yüksek doğruluk sağladığı belirtilmiştir. Elde edilen Twitter verileri SVM kullanarak sınıflandırılarak olumlu ve olumsuz eğilimlere dair sonuçlar elde edilmeye çalışılmıştır. Veri toplama işlemi, Twint kütüphanesi kullanılarak gerçekleştirilmiş ve Twitter API'sine gerek duyulmamıştır. Elde edilen veriler öncelikle temizlenmiş ve duygu analizi için uygun hale getirilmiştir. Veri kümesinin dengesizliği, alt örnekleme yöntemi kullanılarak ele alınmış ve analiz için hazırlanan veriler SVM algoritmasıyla işlenmiştir. Sonuçlar, kullanılan yöntemin yüksek doğruluk değerleri elde ettiğini göstermiştir. 80:20 oranında eğitim-test bölme durumunda en yüksek doğruluk %94.64 olarak bulunmuştur. Hassasiyet, özgüllük ve F1 skoru gibi değerler de incelendiğinde, SVM algoritmasının kripto para duygu analizi konusunda etkili bir araç olduğu görülmüştür. Bu sayısal değerler, analizin güvenilirliğini ve gerçek dünya uygulamalarındaki potansiyel etkisini daha fazla vurgulamaktadır.

(Sarkar ve Bhowmick, 2017) tarafından yapılan çalışmada Bengalce tweetler üzerinde duygu analizi gerçekleştirilmiştir. Çalışmada, Multinomial Naive Bayes ve SVM gibi sınıflandırıcılar kullanarak duygu kutupluluğu tespiti yapılmıştır. N-gram Tokenizer ve SentiWordnet gibi yöntemler kullanılarak veriler işlenmiştir. Veriler, olumluluk, olumsuzluk ve tarafsızlık olmak üzere üç duyarlılık puanına göre sınıflandırılmıştır. Öncelikle veriler ön işleme tabi tutulmuş, ardından etiketli vektörlerle bir sınıflandırıcı eğitilmiş ve bu model tweetler üzerinde test edilmiştir. Kullanılan Bengalce veri kümesinde eğitim veri kümesi olarak 356 negatif, 368 pozitif ve 276 nötr tweet, test veri kümesi olarak ise 500 etiketlenmemiş tweet olmak üzere toplamda 1500 tweet bulunmaktadır. Deney sonuçlarına göre, Naive Bayes multinomial unigram ve bigram kullanılarak SentiWordnet ile %44.20 doğrulukla en iyi sonuçlar elde edilmiştir. SVM tarafından unigram ve SentiWordnet özellikleri kullanılarak %45 doğrulukla en iyi performans gösterilmiştir. Ayrıca, önerilen SVM tabanlı sistem, SAIL 2015 yarışmasında kullanılan en iyi tweet duyarlılık tespit sistemleriyle karşılaştırıldığında daha iyi performans sergilediği belirtilmiştir.

(Jayakody ve Kumara, 2021) tarafından yapılan çalışmada, Twitter'daki ürün incelemelerinin kullanılarak duygu analizi üzerinde çalışıldığı belirtilmiştir. Veri olarak 3548 açıklamalı tweet kullanılmıştır. İki farklı vektörleştirme yöntemi kullanılarak üç denetimli makine öğrenimi modeli

karşılaştırılmıştır. Veriler; SVM, Lojistik Regresyon, K-Nearest Neighbor algoritmaları kullanılarak sınıflandırılmıştır. Vektörleştirici olarak TF-IDF ve Sayım Vektörleştirici kullanılmıştır. Veriler olumlu-olumsuz duygulara göre ayrıştırılmıştır. Tweetler pozitif olanlar 1, negatif olanlar ise 0 şeklinde etiketlenmiştir. Daha sonra boş tweetler kaldırılarak kalan tweetler üstünden istenmeyen içerikler silinmiştir. 2 kelimedenden daha az tweetlerde silinerek kalan metinler küçük harfe dönüştürülüp veri hazırlama aşaması tamamlanmıştır. Eldeki veri kümesi eğitim ve test olmak üzere iki veri kümesine ayrılmıştır. Eğitim veri kümesinin oranı %70, test veri kümesinin oranı ise %30 olarak bildirilmiştir. Makine Öğrenimi algoritmaları kullanılarak sonuçlar elde edilmiştir. Elde edilen sonuçlara bakıldığında Sayım Vektörleştirici kullanılarak yapılan çalışmada Lojistik Regresyonun diğer algoritmalara göre daha iyi sonuç verdiği gözlemlenmiştir. TF-IDF kullanılarak yapılan çalışmada ise SVM algoritmasının %87.41 ile diğer algoritmalara kıyasla daha iyi sonuç verdiği gözlemlenmiştir. Diğer bir çalışma olarak veri kümesinin bölünmesinin sonuca etkisini belirlemek için eğitim ve test olmak üzere her ikisi de %50 oranla bölünmüş ve çalışma uygulanmıştır. Sayım vektörleştirici kullanıldığında Lojistik Regresyon, TF-IDF kullanıldığında SVM algoritması daha iyi sonuç vermiştir. En doğru sonuçları ise %87.2 doğruluk oranı ile Sayım Vektörleştirici kullanılan Lojistik Regresyon algoritması tarafından verildiği görülmüştür. Genel sonucun veri bölme işleminden etkilenmediği görülmektedir.

(Bhasin ve Das, 2021) yapmış oldukları çalışmada, Apache Hadoop ile makine öğrenimi teknikleri incelenmiştir. Çalışmada ilk olarak hashtag temelinde tweetler toplanmış daha sonra bir ön işlemden geçirilmiştir. Özellik çıkarımı yapılmıştır. Makine öğrenimi modellerinden LR, NB ve SVM kullanılmıştır. Hadoop büyük verileri işlemek için kullanılan açık kaynaklı bir çerçeve olarak tanımlanmıştır. Map-Reduce algoritması bir Hadoop modülü olarak üç aşamadan oluşmaktadır; Eşleyici aşaması, Karıştırma aşaması ve İndirgeyici aşaması. Eşleyici aşaması tweetlerin girdi olarak alınıp işlenir. Karıştırma aşaması karıştırma, kopyalama ve sıralama olarak üç aşamaya ayrılmaktadır. İndirgeyici aşama ise tweetlerin pozitif ve negatif olarak belirlendiği aşamadır. Hadoop Dağıtılmış Dosya Sistemi, bir dosya sisteminin bilgisayarlar üzerinde çalışmasını sağlayan sistemdir. Apache Spark Map Reduce'un kısıtlamalarını kaldırmak için tasarlanan toplu reduce seçimi olarak tanımlanmaktadır. Hızlı sorgu yapmak için kullanılmaktadır. Çalışmanın sonucunda veri kümesinin boyutu çok büyük olduğunda Naive Bayes makine öğrenimi algoritmasının diğer algoritmalara göre daha sağlam olduğu belirtilmiştir.

(Jagadeesan ve ark., 2022) tarafından yapılan çalışmada, Twitter duygu analizinin gerçekleştirilmesi için makine öğrenimi yöntemi kullanılmıştır. Çalışmanın amacı, Twitter'daki ırkçı, kadın düşmanı, dine ve mültecilere karşı yapılan nefret söylemlerinin makine öğrenimi kullanılarak tespit edilip tanımlanmasıdır. Metodoloji dört adımdan oluşmaktadır. İlk adım olarak, veri koleksiyonları toplanmıştır. Daha sonra, bu koleksiyonlar ön işleme tabi tutulmuş ve istenmeyen

karakterler metinden kaldırılmıştır. Tokenizasyon uygulanmış, lemmatizasyon ve stemming işlemleri uygulanarak bir kelimenin bütün alternatifleri bir kategoride sınıflandırılıp kelimenin köküne yoğunlaştırılmıştır. Bu iki işlem Doğal Dil İşleme yaklaşımıdır. Üçüncü adım olarak, özellik çıkarma uygulanmıştır. Son adımda ise duygu sınıflandırması yapılmıştır. Bu adımda, Rastgele Orman Sınıflandırıcı, Lojistik Regresyon, Karar Ağacı Sınıflandırıcı, Destek Vektör Makinesi ve XGBoost Sınıflandırıcı olmak üzere beş algoritma ile deneyler yapılmıştır. Hata matrisi oluşturularak doğruluk oranları incelenmiştir. Sonuçlara bakıldığında, doğruluk oranı %95 ile Rastgele Orman ve Destek Vektör Makinesi algoritmaları en iyi sonucu verdiği gözlemlenmiştir.

(Aytuğ, 2018) tarafından yapılan çalışmada, Twitter üzerinde duygu analizi için psikodilbilimsel bir yaklaşım benimsenmiş ve çeşitli metodolojik adımlar takip edilmiştir. Veri seti, İngilizce tweet'lerden oluşmakta olup olumlu, olumsuz ve nötr duygular içermektedir. Toplamda 12.600 tweet içeren veri seti, dengeli bir şekilde etiketlenmiş 4200 negatif, 4200 pozitif ve 4200 nötr tweet içermektedir. Veri Seti Koleksiyonu aşamasında, Twitter Streaming API ve Twitter4J kütüphanesi kullanılarak tweet'ler toplanmıştır. Veri Ön İşleme aşamasında, tweet'lerin düzensiz ve gayri resmi doğası dikkate alınarak, baş harfler, gereksiz tekrarlar ve yanlış kullanılan harfler gibi belirli sorunlar ortadan kaldırılmıştır. Özellik Mühendisliği aşamasında, duygu analizi için psikodilbilimsel özellikler çıkarmak amacıyla LIWC (Dil bilimsel Sorgulama ve Kelime Sayısı) kullanılmıştır. Bu çerçevede, dilin duygusal, bilişsel ve yapısal yönlerini tanımlamak için kullanılan bir metin analizi uygulamasıdır. Sınıflandırma Algoritmaları aşamasında, çeşitli özellik mühendisliği şemalarının tahmin performansını değerlendirmek için Naïve Bayes, Destek Vektör Makineleri, k-En Yakın Komşu algoritması ve Lojistik Regresyon kullanılmıştır. En yüksek tahmin performansı Naïve Bayes algoritması ile sağlanmıştır. Ensemble Öğrenme Yöntemleri aşamasında, topluluk öğrenme algoritmaları olan Bagging, AdaBoost ve Rastgele Altuzay yöntemleri kullanılmıştır. Bagging, farklı eğitim setleri üzerinde eğitilmiş zayıf öğrenme algoritmalarını birleştirerek daha yüksek tahmin performansı elde etmeyi amaçlayan bir yöntemdir. AdaBoost ile sınıflandırılması zor olan veri noktalarına odaklanarak sağlam bir sınıflandırma modeli oluşturulması amaçlanmaktadır. Rastgele Altuzay yöntemi, rastgele seçilen özellik altuzayları üzerinde eğitilen birden fazla sınıflandırıcıyı birleştiren bir topluluk öğrenme algoritmasıdır. Deney analizi ve sonuçlar aşamasında, 10 kat çapraz doğrulama yöntemi kullanılarak çeşitli özellik kümeleri ve sınıflandırıcılarla yapılan deneyler değerlendirilmiştir. Psikodilbilimsel özelliklerin ve topluluk öğrenme yöntemlerinin duygu analizi performansını arttırdığı gözlemlenmiştir. Özellikle, dilbilimsel süreçler, psikolojik süreçler ve kişisel kaygıların birleştirildiği topluluk özellik kümeleri, Naïve Bayes sınıflandırıcısı ile birlikte kullanıldığında %89.10'luk bir sınıflandırma doğruluğu elde edilmiştir. F-skoru sonuçları da sınıflandırma doğrulukları ile paralellik göstermiştir. Bu çalışmanın sonuçlarından, tweet verilerinde

duygu analizi için psikodilbilimsel özelliklerin etkili bir şekilde kullanılabileceği ve topluluk öğrenme yöntemlerinin sınıflandırma algoritmalarının performansını arttırabileceği elde edilmiştir.

(Kemaloğlu ve ark., 2021) yapmış oldukları çalışmada, Facebook, Twitter, Instagram, YouTube ve Rss verilerini kullanarak duygu analizi üzerinde çalışmaktadır. LSTM modeli, özellikle Twitter verileriyle eğitilen modeller arasında en yüksek doğruluk oranını elde etmiştir. Multinomial Naive Bayes ise %86.06 başarı oranıyla dikkat çekmektedir. Rastgele Orman algoritması, kelime tabanlı yöntemlerde %89.5 başarı sağlamıştır. Ancak özellikle TF-IDF kullanılarak geliştirilen LSTM modeli %83.3 başarı oranı vermiştir. Bu çalışmada, 5 farklı sosyal medya platformundan API'ler kullanılarak veri toplanmış ve veri seti %30 test, %70 eğitim verisi olarak ayrılmıştır. Python tabanlı bir yazılım kullanılarak geliştirilen sistemde, numpy, pandas, keras ve scikit-learn gibi kütüphaneler kullanılmıştır. Analiz sonuçlarına göre, en yüksek başarı oranı %84.46 ile LSTM modelinden elde edilmiş ve bu nedenle geliştirilen sistemle birlikte kullanılmak üzere LSTM modeli tercih edilmiştir.

(Hassan ve ark., 2021) yapmış oldukları çalışmada, tweetlerin duyarlılık analizini kullanarak COVID-19 salgınına tespit etmeye yönelik bir araştırma gerçekleştirilmiştir. Bu bağlamda Twitter veri kaynağı olarak seçilmiştir ve COVID-19 ile ilgili anahtar kelimeler içeren birçok tweet toplanmıştır. Toplanan bu veriyi daha sonra filtrelemek ve iki etikete ayırmak amacıyla farklı ön işleme teknikleri uygulanmıştır. Bu tekniklerden biri geçerli veriyi temsil ederken diğeri geçersiz veriyi ifade etmektedir. Ön işlemeden sonra ait olduğu verinin bir dizi kategorisini tanımlamak için sınıflandırma algoritmaları kullanılmıştır. Sınıflandırma için Karar Ağacı Algoritmasını, Naïve Bayes Algoritmasını ve Destek Vektör Makinesi Algoritmasını kullanılmıştır. Deneylerimiz için en uygun yolu belirlemek amacıyla dört farklı k-kat değeri kullanılmıştır. Bu adımı, araştırmanın doğruluk testlerini uyguladığı performans değerlendirme adımı izlemiştir. Deneylerden, SVM'nin verilerin doğruluk açısından sınıflandırılmasına en uygun olduğu gözlemlenmiştir. Bu araştırma, COVID-19 hastası olan ve herhangi bir nedenle test yapılmayan hastaları tespit etmek için bu tür bir yaklaşımın kullanılabileceğini göstermektedir.

(Liu ve ark., 2022) tarafından yapılan çalışma, Çince üzerinden yürütülmüş ve duygular 7 ana kategoride sınıflandırılmıştır. Bu ana kategoriler, iki ana duygu kategorisinde birleştirilmiştir. “Mutlu” ve “iyi” ifadeleri olumlu duygu; “üzgün”, “kızgın”, “korkmuş” ve “şaşkın” olumsuz duygu olarak sınıflandırılmıştır. Olumlu duyguların sıfır puandan büyük, olumsuz duyguların sıfır puandan küçük ve nötr duyguların ise sıfır puan olarak değerlendirilmesi sağlanmıştır. Bu çalışmada bazı kurallar uygulanmıştır. İlk kural olarak, olumsuz bir kelimenin bir duygudan sonra gelmesi durumunda cümlelerin duygusunun tam tersine çevrilmesi ve ona göre puanlandırma yapılması sağlanmıştır. İkinci kural olarak, karşıt bağlaç kullanıldığında cümlelerin olumsuza çevrilmesi ve puan ağırlığının eksi olması sağlanmıştır. Üçüncü kural ise derece zarflarına farklı ağırlıklar verilmesi üzerinedir. Bu kuralların birçok olumlu ve olumsuz yanlarının olduğu belirtilmiştir. Bu çalışmada,

duyarlılık sözlüğü ve TF-IDF yöntemini birleştiren yeni bir yöntem tasarlanmıştır. Tüm duygu kelimelerinin ağırlıkları E olarak değerlendirilmiş ve bir metnin toplam ağırlığı 1 olarak sınırlandırılmıştır. Daha sonra, ağırlığın ikiye bölünmesi ve tüm duygu kelimelerine a, nötr kelimelere ise 1-a ifadesinin TF-IDF değerleriyle orantılı olarak atanması sağlanmıştır. Bu verilerle bir denklem elde edilmiştir. Duygu analizi için Yu Hua'nın "Çiseleyen Yağmurda Çılgınlıklar" adlı romanı 16 bölüm şeklinde ayrıştırılarak veri olarak kullanılmıştır. Sonuçlar incelendiğinde, toplamda 2079 cümlelerin analiz edildiği gözlemlenmiştir. Metindeki verilerin duygu dağılımı incelendiğinde, romanın daha çok olumsuz duygular içerdiği tespit edilmiştir. Değerlendirme olarak kesinlik, geri çağırma ve f1 puanı hesaplanmıştır. Duygu sözlüğüne dayalı kurallar, TF-IDF ve bu çalışmada önerilen yöntemlerin doğruluk oranı karşılaştırıldığında en yüksek oranı önerilen yöntemin sağladığı belirtilmiştir. Geri çağırma için yöntemler karşılaştırıldığında, pozitif duygularda en yüksek oranı duygu sözlüğüne dayalı kurallar, nötr ve negatif duygularda ise önerilen yöntemin sağladığı gözlemlenmiştir. F1 skoru karşılaştırıldığında, üç durumda da en yüksek oranı önerilen yöntemin verdiği tespit edilmiştir. Sonuç olarak, önerilen yöntemin en iyi performansı gösterdiği belirtilmiştir.

(Başarslan ve Kayaalp, 2023) yapmış oldukları çalışmada, TripAdvisor ve Rotten Tomatoes gibi web sitelerindeki açık kullanıcı geri bildirimlerine duygu analizi uygulanmıştır. Word2Vec ve TF-IDF yöntemleri kullanılarak dört makine öğrenme ve iki topluluk öğrenme yöntemiyle sonuçlar elde edilmiştir. Topluluk öğrenme, özellikle çoğunluk oylaması, tekli öğrenme yöntemlerinden daha etkili sonuçlar gösterdiği belirtilmiştir. Eğitim-test ayrımları %80-%20 ve %70-%30 olarak iki şekilde yapılmıştır. Metin ön işleme adımları arasında yorum temizleme, özel karakterleri çıkarma, küçük harfe dönüştürme ve stemming yer almaktadır. LR, SVM, NB ve KNN sınıflandırıcıları kullanılarak gerçekleştirilen deneylerde, scikit-learn kütüphanesi Python ile uygulanmıştır. Bu çalışmada, topluluk öğrenme yöntemlerinin duyarlılık sınıflandırma sürecine önemli katkı sağladığı gösterilmiştir.

(Karayığit ve ark., 2018) tarafından yapılan çalışmanın amacı, öğrenciler tarafından toplanan Türkçe tweetler aracılığıyla sistem hakkındaki düşünceleri öğrenmek ve duygu analizi yapmaktır. Türkçe dilinde duygu analizi için makine öğrenmesi yaklaşımlarının daha iyi performans gerçekleştirdiği yapılan araştırmalarla ortaya konulmuştur. Bu nedenle bu çalışmada makine öğrenimi tabanlı yöntemler tercih edilmiştir. Twitter aracılığıyla konu ile ilgili hashtag kullanılarak 63699 adet tweet toplanmıştır. Veri ön işleme dört adıma ayrılmıştır: temizleme, normalizasyon, tokenizasyon ve durak kelimeleri kaldırma. Temizleme kısmında gereksiz kelimeler, bağlantılar veri kümesinden çıkarılmıştır. Yabancı dil kullanılan tweetler "Language Detection API" ile tespit edilip aynı şekilde veri kümesinden çıkarılmıştır. Geriye 4652 adet tweet kalmıştır. Normalleştirme kısmında ise noktalama işaretleri kaldırılmış ve tüm harfler küçük harfe dönüştürülmüştür. Tokenizasyon kısmında kalan veriyi anlamlı parçalara bölme işlemi yapılmıştır. Son adımda durak

kelimeler çıkarılarak kalan metin daha anlamlı hale getirilmiştir. Duygu analizinde üç ayrı duygu kullanılmıştır; olumlu, olumsuz ve nötr. Vektör uzayı modeli olarak ise Bag of Words ve TF-IDF kullanılmıştır. Duygu sınıflandırması için bu çalışmada Destek Vektör Makinesi, Lojistik Regresyon, K-En Yakın Komşu ve Yapay Sinir Ağları kullanılmıştır. Veri seti 1075 negatif, 3345 nötr ve 232 pozitif veriden oluşmaktadır. Veri setinde açık bir şekilde görülen dengesizlik vardır. Bu sebeple karşılaştırma için F1 skoru kullanılmıştır. Sonuçlar incelendiğinde Lojistik Regresyon sınıflandırıcının TF-IDF ve BoW vektör uzaylarını kullanarak diğer makine öğrenimi yöntemlerine göre daha iyi sonuç verdiği görülmüştür. Detaylı incelemede SVM, ANN ve kNN yöntemlerinin BoW vektör uzayı modeli kullanılan çalışmada, LR yönteminin ise TF-IDF modeli kullanan çalışmada çok daha iyi sonuç verdiği gözlemlenmiştir.

(Aydın ve ark., 2021) yapmış oldukları çalışmada, duygu analizi ve fikir madenciliği arasındaki ilişkiyi açıklayarak bunlar üzerinde bir inceleme yapmak amaçlanmıştır. Ayrıca bu çalışmada Türkçe'nin diğer dillerden daha farklı olmasına ve bir kelime ile bir cümle kurulabilmesine dikkat çekilmiş, duygu analizi yaparken başka bir dilin modellerinin kullanılamayacağı fakat duygu analizinde kullanılan yöntemlerin kullanılabileceği vurgulanmıştır. Duygu analizi ve fikir madenciliği incelendiğinde fikir madenciliğinin akademik alanda, duygu analizinin ise endüstride kullanıldığı görülmektedir. Farklılıklar olsa dahi iki alanında birlikte çalıştığı gözlemlenmiştir. Sonuçlar kısmında Türkçe duygu analizi yapmak için veri kümesi bulunmaması gibi zorluklar yaşandığına, bu zorlukların sosyal medyadan veri elde edilmesi ile aşılabileceğine değinilmiştir. Aynı zamanda bu konu ile ilgili yapılan projelerin yetersiz olması gereken bütçenin sağlanması ile aşılabileceği belirtilmiştir. Eğitim setinin kategori olarak düzgün ayarlanmasının analiz üzerinde doğrudan etkili olduğu söylenmiştir.

(Salur ve Aydın, 2022) tarafından yapılmış olan çalışmada, Türkçe tweetler için yeni bir duygu sınıflandırma modeli tasarlanmıştır. Veri ön işlemede silinen karakterlerden, emojiler gibi sözcüklerden özellik çıkarma ve diğer özelliklerle birleştirilmesine dayalı bir çalışma yapılmıştır. Veri kümesinde 6887 adet pozitif ve 7890 adet negatif tweet bulunmaktadır. Tweetler anlamlı veya anlamsız birçok karakter, kelime içermektedir. Bu kelimelerin daha anlamlı olabilmesi için veri ön işleme yapılmıştır. Veri ön işlemede URL bilgisi, fazla boşluklar, noktalama işaretleri, kullanıcı adı, durak kelimeler, rakamlar ve tek karakterler silinmiş karakter tekrarı içerdiği için anlamsız olan kelimeler düzeltilmiştir. Ayrıca daha önce yapılan çalışmalarda Türkçe tweetlerin analizinde kelime gövdeleme işlemi, duygu analizin performansına katkı sağlamadığı için bu çalışmada kullanılmadığına değinilmiştir. Kalan veriler FastText kullanılarak metin temsili oluşturulmuş ve bu veri kümesinden LSTM, BiLSTM ve GRU kullanılarak derin özellik çıkarımı yapılmıştır. Veri ön işleme kısmında silinen içeriklerin aslında duygu belirttiği ve Twitter'a özgü ifadeler olduğu görülmüştür. Bu içeriklerden 12 adet elle özellik çıkarımı yapılmıştır. Bu özellikler, sözdizimsel

olarak; toplam karakter sayısı, büyük harf sayısı, küçük harf sayısı, rakam sayısı, noktalama işareti sayısı ve URL sayısı olarak toplam 6 tane, Twitter'a özgü olarak; Bahsetme (mention) sayısı ve Hashtag sayısı olarak toplam 2 tane ve anlamsal olarak; pozitif emoji sayısı, negatif emoji sayısı, pozitif ünlem sayısı, negatif ünlem sayısı olarak toplam 4 tane özellikten oluşmaktadır. Elle çıkarılan özellikler LSTM, BiLSTM ve GRU kullanılarak derin özellikler çıkarılmıştır. Daha sonra FastText kullanılarak yapılan özellik çıkarımı ile elle yapılan özellik çıkarımı birleştirilerek duygu sınıflandırma katmanına iletilmiştir. Bu çalışmada veriler eğitim için %90 ve test işlemi için %10 olarak ayrılmıştır. Modelin başarısını değerlendirmek için karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisinden elde edilen değerlerle doğruluk, kesinlik, duyarlılık ve F1 skoru hesaplanmıştır. Derin özellik çıkarma modelinde doğruluk oranları incelendiğinde BiLSTM'in daha yüksek doğruluk verdiği gözlemlenmiştir. Genel sonuçlarda en yüksek doğruluk oranı %75.71 oran ile elle özellik çıkarma ve BiLSTM'in birleşiminden oluşan modelde gözlemlenmiştir. Çıkan sonuçlar incelendiğinde en yüksek sonuçların özellik birleştirme modellerinde olduğu görülmektedir. Bu çalışma daha önce yapılan çalışmalar ile karşılaştırıldığında önerilen modelin daha iyi performans gösterdiği gözlemlenmiştir.

(Tan ve ark., 2023) tarafından yapılmış olan çalışmada, duygu analizi sırasında alaycılık tespitini artırmaya yönelik bir çerçeve sunulmuştur. Alaycı ifadelerin duygu sınıflandırmasını olumsuz etkilediği belirtilmiş ve bu sorunu çözmek adına çoklu görev öğrenimini içeren bir derin sinir ağı kullanılmıştır. Çerçeve, duygu sınıflandırması ve alaycılık sınıflandırması görevlerini aynı anda gerçekleştirmek üzere Çift Yönlü LSTM ağını eğitmeyi içermektedir. Çalışma, bu yöntemin diğer modellere göre daha iyi performans sergilediğini göstermektedir. Veri kümesi olarak Kaggle'dan elde edilen duyarlılık ve alaycılık veri kümeleri kullanılmıştır. Her iki veri kümesi için aynı doğal dil işleme teknikleri ve ön işleme yöntemleri uygulanmıştır. Çerçeve, özellikle paylaşılan bir katman kullanarak aşırı uyum riskini azaltmayı amaçlayarak duygu analizi performansının artırılması hedeflenmiştir. Önerilen çerçeve, duyarlılık sınıflandırması için %91, alaycılık sınıflandırması için %92 F1 puanları elde edilmiştir. Bağımsız Bi-LSTM modeline kıyasla, çerçeve duyarlılık ve alaycılık sınıflandırmalarında sırasıyla %3 ve %1 iyileşme sağlanmıştır. Çerçeve, çoklu görev öğrenimi kullanarak eğitildiğinden hem eğitim süresini hem de hesaplama gücünü daha verimli bir şekilde kullandığı sonucuna varılmıştır.

(Tsai ve Huang, 2021) yapmış oldukları çalışmada, evcil hayvanların duyguları yapay zekâ sistemleri vasıtasıyla analiz edilmiştir. İnsan duygu analizi çerçevesini evcil hayvanlar için uyarlayarak görüntü ve ses özelliklerini birleştirerek duygu analizi yapılmıştır. Evcil hayvanların yüz kaslarının sınırlamaları nedeniyle duruşları, ses özellikleri ise spektrogram kullanılarak değerlendirilmiştir. Nesne etiketlerini tespit etmek ve tanımlamak için Mask R-CNN kullanılarak derin öğrenme teknolojileriyle evcil hayvanların duygu analizi gerçekleştirilmiştir. Kullanılan

teknolojiler arasında Logitech Webcam C925 kamera, gömülü sistemlerde çekirdek bilgi işlem platformu, Python programlama dili ve Tensorflow derin öğrenme kütüphanesi bulunmaktadır. Ortalama yürütme süresi 1.32 dakika olan önerilen yöntemin doğruluğu %85.71'dir. Bu yöntemin, temassız bir şekilde evcil hayvan duygu tanımlamasında başarılı bir doğruluk oranına sahip olduğu ve sahibine bilgi sağladığı vurgulanmıştır.

(Chen ve ark., 2020) tarafından yapılan çalışmada, askeri alandaki duygu analizi için özel olarak geliştirilen MILSentic adlı duygu sözlüğünün performansı değerlendirilmiştir. Yapılan deneylerde, mevcut duygu sözlükleri ile MILSentic'in kombinasyonları kullanılarak gerçekleştirilen analizlerde, MILSentic'in eklenmesinin kutupluluk tahmininde %1,5 ve %1,4 daha yüksek doğruluk ve F1-ölçümü sağladığı belirlenmiştir. Ayrıca, çalışma, LSTM ve Bi-LSTM modelleri kullanılarak gerçekleştirilen parametre kalibrasyonunu içermektedir. Bu bağlamda, ağ katmanları ve aktivasyon fonksiyonları üzerinde yapılan ayarlamalarda, doğruluk ve F1-ölçümü değerlerindeki değişiklikler gözlemlenmiştir. Bi-LSTM modelinin, özellikle Tanh aktivasyon fonksiyonu ve belirli bir ağ katmanı sayısı ile birlikte kullanıldığında en iyi performansı sergilediği belirtilmiştir. Bu çalışmada, askeri duygu analizi için özel veri kaynaklarının etkili bir şekilde kullanılmasının, duygu sözlüklerinin geliştirilmesinin ve derin öğrenme modellerinin parametre ayarlarının performans artışına nasıl katkı sağlayabileceğini detaylı bir şekilde ele alınmıştır.

Bu çalışmalar, duygu analizi alanında çeşitli teknikler ve metodolojilerin geliştirilmesine önemli katkılar sağlamaktadır. Yapay zekâ ve makine öğrenimi tekniklerinin kullanımıyla, metinlerden duygusal içeriklerin daha doğru ve derinlemesine analiz edilmesi mümkün olmuştur. Özellikle, dil ve kültürel farklılıkları dikkate alan çalışmalar tarafından duygu analizinin evrensel bir uygulama alanı olduğunu gösterilerek bu alandaki bilimsel ilerleme desteklenmiştir. Yeni duygu sözlükleri ve büyük veri setleri oluşturma çabaları, analizlerin doğruluğunu ve kapsamını artırarak endüstriyel ve akademik kullanımlarda yeni olanaklar sunmaktadır.

1.2. Motivasyon ve Katkı

Sosyal medya platformlarının yaygınlaşması ve kullanıcıların bu platformlarda yoğun şekilde veri üretmesi, duygu analizine olan ilgiyi artırmıştır. Duygu analizi, sosyal medya verilerinden anlamlı bilgi çıkarımı sağlayarak çeşitli alanlarda önemli uygulamalara olanak tanımaktadır. Bu çalışmanın motivasyonu, makine öğrenimi ve derin öğrenme tekniklerini kullanarak duygu analizinde yüksek doğruluk oranları elde etmek ve farklı algoritmalar ile veri temsili yöntemlerini karşılaştırarak en etkili yaklaşımları belirlemektir.

Var olan çalışmalar genellikle sınırlı sayıda veri kümesi ve algoritma kullanarak gerçekleştirilmiştir. Bu çalışmada ise daha kapsamlı bir yaklaşım benimsenmiş, farklı veri kümeleri

ve çeşitli algoritmalar kullanılarak duygu analizinin performansı değerlendirilmiştir. Bu sayede, duygu analizinde hangi yöntemlerin daha etkili olduğunu belirlemek ve bu alandaki literatüre katkıda bulunmak amaçlanmıştır. Bu çalışmanın ana katkıları şöyle özetlenebilir:

- **Kapsamlı Literatür İncelemesi:** Çalışma kapsamında 39 farklı makale incelenmiş ve mevcut yöntemler ile yaklaşımlar hakkında derinlemesine bilgi sunulmuştur. Bu literatür taraması, çalışmanın temellerini güçlendirerek literatürdeki mevcut boşlukları ve geliştirme alanlarını belirlemeye yardımcı olmuştur.
- **Çeşitli Veri Kümelerinin Kullanımı:** İki farklı veri kümesi kullanılarak, farklı veri kaynaklarından elde edilen sonuçlar karşılaştırılmıştır. Bu yaklaşım, duygu analizi modellerinin genellenebilirliğini ve farklı veri kümelerindeki performansını değerlendirme imkânı sunmaktadır.
- **Çok Sınıflı Karşılaştırmalar:** Hem iki sınıflı hem de üç sınıflı sınıflandırma yapılmış, böylece farklı sınıflandırma düzeylerinde elde edilen sonuçlar incelenmiştir. Bu, duygu analizinin farklı sınıflandırma senaryolarındaki performansını karşılaştırma imkânı sunmaktadır.
- **Çeşitli Metin Temsil Yöntemleri:** Çalışmada, beş farklı metin temsil yöntemi kullanılmış ve metinlerin farklı şekillerde temsil edilmesinin duygu analizine etkisi araştırılmıştır. Bu yöntemlerin karşılaştırılması, hangi temsil yöntemlerinin daha etkili olduğunu belirlemeye yardımcı olmaktadır.
- **Geniş Algoritma Yelpazesi:** Dokuz makine öğrenimi algoritması ve üç derin öğrenme algoritması kullanılarak sınıflandırma yapılmıştır. Bu geniş algoritma yelpazesi, farklı algoritmaların duygu analizi performansını karşılaştırma ve en etkili olanları belirleme imkânı sunmaktadır.

Bu çalışma, duygu analizi alanında mevcut literatürde yer alan çalışmalardan birkaç önemli noktada ayrılmaktadır. İlk olarak, çoğu çalışma belirli bir metin temsil yöntemi ve sınıflandırma algoritmasına odaklanırken, bu çalışmada beş farklı metin temsil yöntemi (TF-IDF, GloVe, Word2Vec, FastText ve Bag of Words) ile dokuz makine öğrenimi ve üç derin öğrenme algoritması detaylı bir şekilde karşılaştırılmıştır. Bu kapsamlı yaklaşım, farklı veri kümelerinde en etkili yöntemin belirlenmesine olanak tanımaktadır. İkinci olarak, literatürde çoğu duygu analizi çalışması tek bir veri kümesi üzerinde gerçekleştirilmektedir. Bu çalışmada ise iki farklı veri kümesi (Amazon ve IMDb) üzerinde detaylı analizler yapılarak yöntemlerin genel uygulanabilirliği test edilmiştir. Üçüncü olarak, literatürde yaygın olarak ikili sınıflandırma yöntemleri kullanılmasına rağmen, bu çalışmada hem ikili hem de üçlü sınıflandırma senaryoları incelenerek farklı sınıflandırma düzeylerinde yöntemlerin başarımı karşılaştırılmıştır. Son olarak, elde edilen sonuçlar, literatürdeki benzer çalışmalarla kıyaslanarak hangi yöntemlerin daha başarılı olduğu gösterilmiştir.

1.3. Organizasyon

Çalışmanın geri kalan bölümleri şu şekilde organize edilmiştir: Bölüm 2'de çalışmanın altyapısını oluşturan programlama dili, kullanılan kütüphaneler ve veri kümeleri açıklanacaktır. Bölüm 3'te, yapılan sınıflandırma çalışmalarının sonuçlarına odaklanılacak ve elde edilen sonuçlar ayrıntılı bir şekilde incelenerek analiz edilecektir. Bölüm 4'te ise çalışmanın genel değerlendirmesi yapılacak, elde edilen sonuçlar üzerinden çalışmanın başarısı ve etkinliği değerlendirilecektir.

2. Deneysel Ayarlamalar

Bu bölümde, duygu analizi çalışmasının altyapısını oluşturan tekniklere değinilecektir.

2.1. Programlama Dili ve Kütüphaneler

Bu çalışmada Python programlama dili ve çeşitli Python kütüphanelerinden yararlanılmıştır. Veri analizi ve işlemesi için Pandas, bilimsel hesaplamalar için NumPy, makine öğrenimi algoritmaları için Scikit-Learn, derin öğrenme modelleri için Tensorflow, metin madenciliği için Gensim bu kütüphanelerden bazılarıdır.

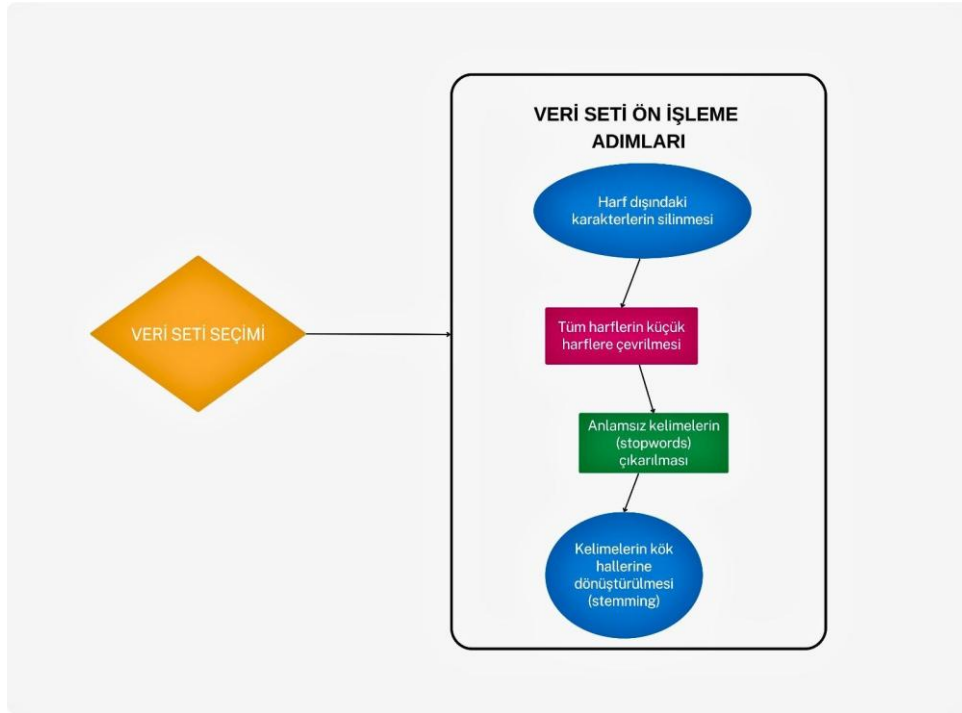
2.2. Kullanılan Veri Kümeleri

Çalışmada iki farklı veri kümesi kullanılmıştır. Bu veri kümeleri Kaggle platformundan alınmıştır. İlk veri kümesi Amazon sitesindeki müzik ürünlerine aittir ve olumlu, olumsuz ve nötr duyguları içeren toplam 10.000 veriden oluşmaktadır. Diğer veri kümesi ise IMDb film yorumlarından oluşmakta olup, olumlu ve olumsuz duyguları içeren ilk 20.000 veriden meydana gelmektedir.

Her iki veri setini analiz etmek için verilerin belirli bir oranı eğitim ve test setleri olarak ayrılmıştır. Çalışmada, model performansının tutarlı bir şekilde değerlendirilebilmesi için veri kümelerinin %80'i eğitim, %20'si ise test seti olarak ayrılmıştır. Rastgelelik tohumu olarak 42 değeri belirlenmiş ve bu parametrenin, verilerin her çalışmada aynı şekilde ayrılmasını sağlayarak sonuçların tekrarlanabilirliğini garanti ettiği görülmüştür. Bu yöntem, model performansının değerlendirilmesi için tutarlı ve güvenilir bir temel oluşturmuştur.

2.3. Veri Ön İşleme

Veri ön işleme aşaması, metin verilerinin analiz için uygun hale getirilmesini sağlar. Bu çalışmada, Şekil 1’de gösterilen adımlar izlenerek boş satırların temizlenmesi, her bir metin ögesinin küçük harfe dönüştürülmesi, sayıların ve noktalama işaretlerinin kaldırılması, durak kelimelerin (stopwords) çıkarılması ve kelimelerin kök hallerinin bulunması gibi işlemler uygulanmıştır. Bu adımlar, metin verilerinin analiz edilebilir ve tutarlı bir formata getirilmesine katkıda bulunarak daha doğru sonuçlar elde edilmesini sağlamaktadır.



Şekil 1. Veri ön işleme adımları

2.4. Performans Metrikleri

Bu çalışmada, duygu analizi modellerinin performansını değerlendirmek amacıyla Doğruluk, Hassasiyet, Duyarlılık ve F1-Skor metrikleri kullanılmıştır. Eşitliklerde yer alan TP gerçek pozitifleri, TN gerçek negatifleri, FP yanlış pozitifleri ve FN yanlış negatifleri ifade etmektedir. Bu metrikler, modelin doğru ve yanlış tahminlerini detaylı bir şekilde analiz ederek, duygu analizi performansının kapsamlı bir değerlendirmesini sağlamaktadır. Çalışmada kullanılan algoritmaların her biri için bu metrikler hesaplanmış ve sonuçlar karşılaştırılarak en etkili yöntemlerin belirlenmesi amaçlanmıştır.

Doğruluk, modelin doğru tahmin ettiği örneklerin toplam örnek sayısına oranıdır. Eşitlik 1’de doğruluk metriğinin matematiksel gösterimi verilmiştir.

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Hassasiyet, pozitif sınıf olarak tahmin edilen örnekler arasından gerçekten pozitif olanların oranını göstermektedir. Yanlış pozitifleri minimize etmek için önemlidir. Eşitlik 2’de hassasiyet metriğinin matematiksel gösterimi verilmiştir.

$$Hassasiyet = \frac{TP}{TP + FP} \quad (2)$$

Duyarlılık, gerçek pozitif örneklerin model tarafından doğru tahmin edilme oranını göstermektedir. Eşitlik 3’te duyarlılık metriğinin matematiksel gösterimi verilmiştir.

$$Duyarlılık = \frac{TP}{TP + FN} \quad (3)$$

F1-Skor, hassasiyet ve duyarlılığın harmonik ortalamasıdır ve sınıflandırma modelinin genel performansını değerlendirir. Özellikle dengesiz veri setlerinde faydalıdır. Eşitlik 4’te F-1 skor metriğinin matematiksel gösterimi verilmiştir.

$$F - 1 \text{ Skor} = 2 \times \frac{Hassasiyet \times Duyarlilik}{Hassasiyet + Duyarlilik} \quad (4)$$

3. Yöntem

Bu bölümde, makine öğrenmesi temelli duygu analizi sisteminin altyapısında kullanılan metin temsil yöntemleri ve sınıflandırma algoritmaları detaylı olarak açıklanmaktadır.

3.1. Metin Temsil Yöntemleri

Metin temsil yöntemleri, ham metni makine öğrenimi ve derin öğrenme algoritmalarının işleyebileceği sayısal vektörlere dönüştürmek amacıyla uygulanmaktadır. Bu çalışmada, TF-IDF, Bag of Words, GloVe, FastText ve Word2Vec gibi yöntemler kullanılmıştır. Her bir yöntem, farklı prensiplere dayanarak kendine özgü avantajlar sunmaktadır.

TF-IDF, belgelerdeki terimlerin önemini sayısal olarak ölçen bir tekniktir. Terim frekansı (TF), belirli bir kelimenin belgede kaç kez geçtiğini ifade ederken, ters belge frekansı (IDF), kelimenin tüm belgeler arasındaki yaygınlığını ölçerek sık rastlanan kelimelerin ağırlığını düşürür (Patil ve ark., 2023). Bir kelimenin TF ve IDF değerlerinin çarpımı, o kelimenin TF-IDF ağırlığını belirler. Bag of Words yöntemi, metni kelime dizilimini dikkate almadan, her kelimenin kaç kez geçtiğini sayarak sayısal verilere dönüştürür (Patil ve ark., 2023). Bu yaklaşım, duygu analizi gibi uygulamalarda, belirli duyguları ifade eden kelimelerin tespitini kolaylaştırır. Word2Vec, büyük veri kümelerinden sözcük yerleştirmelerini öğrenmek için geliştirilen, sinir ağı tabanlı bir modeldir (Patil ve ark., 2023).

Bu model, her kelimeyi yüksek boyutlu bir vektör uzayında temsil ederek, kelimeler arasındaki anlamsal benzerlikleri yakalamayı amaçlar. Word2Vec, kelimelerin bağlamını dikkate alarak, metin içindeki ilişkileri öğrenir ve her kelimeyi, benzer anlamlara sahip diğer kelimelere yakın konumlandıran bir vektörle ifade eder (Öge ve Kayaalp, 2021). FastText, Word2Vec'in geliştirilmiş bir versiyonu olarak, kelimeleri alt birimlerine ayırarak analiz eder (Patil ve ark., 2023). Bu sayede dilde nadir görülen veya hiç görülmemiş kelimeler için bile anlamlı vektörler oluşturabilir. GloVe, kelimeler arasındaki ortak görülme istatistiklerini kullanarak vektör temsilleri oluşturan bir modeldir (Patil ve ark., 2023). Büyük metin verilerindeki kelime ilişkilerini yakalayarak, anlamsal ve sözdizimsel benzerlikleri matematiksel olarak modelleyebilir.

3.2. Sınıflandırma Algoritmaları

Bu çalışmada, sosyal medya verileri üzerinde duygu analizi gerçekleştirmek amacıyla çeşitli makine öğrenimi ve derin öğrenme algoritmalarından faydalanılmıştır. Makine öğrenimi algoritmaları kapsamında, Scikit-learn kütüphanesinden K-Neighbors Classifier, Random Forest Classifier, Lojistik Regresyon, Decision Tree Classifier, Bernoulli Naive Bayes, Support Vector Machine (SVM) ve Multi-Layer Perceptron (MLP) Classifier algoritmaları uygulanmıştır. Ayrıca, LightGBM ve XGBoost gibi gelişmiş gradient boosting algoritmalarından yararlanılarak metin verilerinin sınıflandırılması sağlanmıştır. Derin öğrenme algoritmaları ise TensorFlow ve Keras kütüphaneleri kullanılarak gerçekleştirilmiş olup, LSTM, SimpleRNN ve Conv1D katmanları ile modeller oluşturulmuştur. Bu geniş algoritma yelpazesi, duygu analizi performansını optimize etmek ve farklı yöntemlerin etkinliğini karşılaştırmak amacıyla tercih edilmiştir.

4. Sonuçlar ve Tartışma

Bu bölümde sonuçlar tablolar halinde verilerek detaylandırılmıştır. Tablo 1 ve Tablo 2, Amazon veri seti üzerindeki derin öğrenme ve makine öğrenmesi yöntemlerinin performanslarını karşılaştırmaktadır. Tablo 3 ve Tablo 4 ise IMDb veri seti üzerinde gerçekleştirilen analizlerde derin öğrenme ve makine öğrenmesi yöntemlerinin sonuçlarını sunmaktadır.

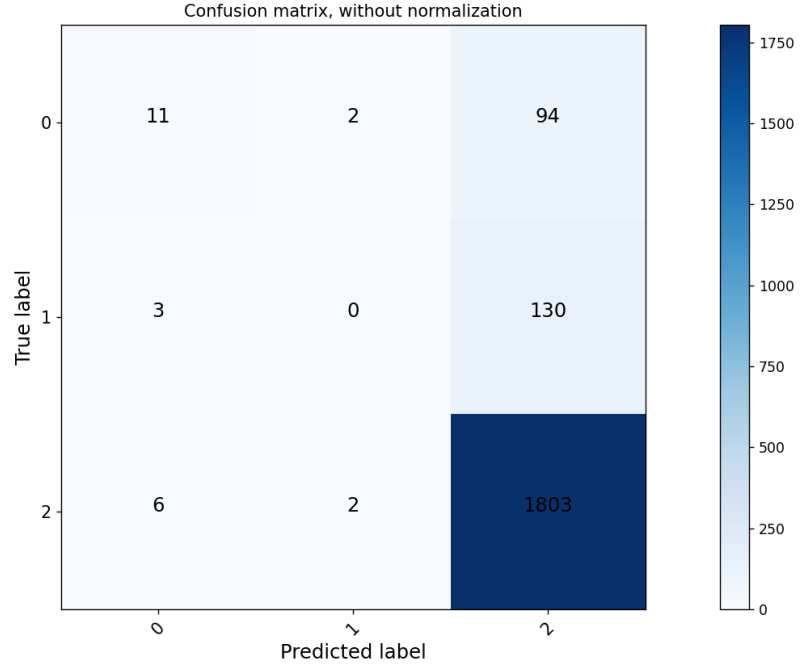
Tablo 1, Amazon veri setinde gerçekleştirilen derin öğrenme çalışmalarının sonuçlarını detaylı bir şekilde sunmaktadır. Bu çalışmanın sonuçlarına göre, derin öğrenme modellerinin doğruluk oranları %88.00 ile %88.44 arasında değişmektedir. En yüksek doğruluk oranı %88.44 ile FastText temsil yöntemi ve LSTM algoritması kombinasyonu tarafından elde edilmiştir. Genel olarak, TF-IDF, GloVe, ve FastText temsil yöntemleri, CNN, RNN, ve LSTM algoritmaları ile birlikte kullanıldığında yüksek doğruluk oranları sergilemiştir. Özellikle CNN ve LSTM algoritmaları, çoğu temsil

yöntemiyle yüksek performans göstermiştir. Bununla birlikte, LSTM algoritması, özellikle FastText ve TF-IDF ile birlikte kullanıldığında, genellikle yüksek doğruluk oranları sağlamıştır.

Tablo 1. Amazon veri seti derin öğrenme sonuçları.

Metin Temsil Yöntemi	Sınıflandırma Algoritması	Doğruluk(%)	Hassasiyet	Duyarlılık	F1-Score
TF-IDF	CNN	88.29	0.90	0.88	0.83
	RNN	88.00	0.83	0.88	0.83
	LSTM	88.29	0.90	0.88	0.83
WORD2VEC	CNN	88.10	0.79	0.88	0.83
	RNN	88.15	0.78	0.88	0.83
	LSTM	88.00	0.81	0.88	0.84
GLOVE	CNN	88.29	0.90	0.88	0.83
	RNN	88.29	0.90	0.88	0.83
	LSTM	88.24	0.80	0.78	0.83
FASTTEXT	CNN	88.29	0.90	0.88	0.83
	RNN	88.29	0.87	0.88	0.83
	LSTM	88.44	0.90	0.88	0.83
BINARY	CNN	88.05	0.90	0.88	0.83
	RNN	88.29	0.90	0.88	0.83
	LSTM	88.29	0.78	0.88	0.83

Şekil 2’de verilen karışıklık matrisi, Amazon veri seti ile duygu analizi çalışmasında kullanılan FastText + LSTM modellerinin performansını göstermektedir. Model, pozitif duygu sınıfını (2) çok iyi tahmin etmektedir; 1803 doğru tahmin ile bu sınıfta başarılı olmuştur. Ancak, negatif (0) ve nötr (1) duygu sınıflarında performansı düşüktür; negatif duyguları sadece 11 doğru tahminle, nötr duyguları ise hiç doğru tahmin edememiştir. Genel olarak, model pozitif duyguları tespit etmekte başarılı, ancak negatif ve nötr duyguları ayırt etmede yetersizdir.



Şekil 2. LSTM karşılık matrisi - Amazon veri seti

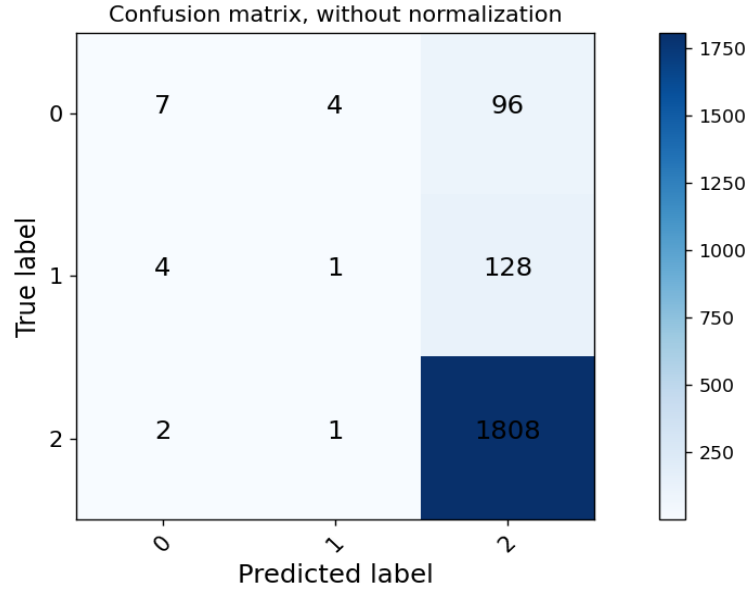
Tablo 2. Amazon veri seti makine öğrenimi sonuçları.

Metin Temsil Yöntemi	Sınıflandırma Algoritması	Doğruluk(%)	Hassasiyet	Duyarlılık	F1-Score
TF-IDF	BNB	84.05	0.79	0.84	0.81
	LR	88.34	0.82	0.88	0.83
	RF	88.24	0.78	0.88	0.83
	DT	83.32	0.82	0.83	0.83
	SVM	88.29	0.84	0.88	0.83
	ANN	87.03	0.84	0.87	0.85
	KNN	87.51	0.79	0.88	0.83
	XGBOOST	88.20	0.83	0.88	0.84
	LightGBM	88.29	0.84	0.88	0.84
WORD2VEC	BNB	88.10	0.81	0.88	0.83
	LR	88.05	0.81	0.88	0.83
	RF	88.34	0.87	0.88	0.83
	DT	80.30	0.82	0.80	0.81
	SVM	88.29	0.84	0.88	0.83
	ANN	83.71	0.83	0.84	0.83
	KNN	86.44	0.81	0.86	0.83
	XGBOOST	88.54	0.84	0.89	0.84
	LightGBM	88.39	0.83	0.88	0.84
GLOVE	BNB	84.73	0.79	0.85	0.82
	LR	87.95	0.80	0.88	0.83
	RF	88.20	0.83	0.88	0.83
	DT	77.96	0.80	0.78	0.79
	SVM	88.29	0.90	0.88	0.83

	ANN	87.56	0.81	0.88	0.83
	KNN	87.03	0.80	0.87	0.83
	XGBOOST	88.10	0.83	0.88	0.83
	LightGBM	88.20	0.83	0.88	0.83
FASTTEXT	BNB	82.35	0.80	0.82	0.81
	LR	88.05	0.82	0.88	0.83
	RF	88.29	0.85	0.88	0.83
	DT	77.86	0.80	0.78	0.79
	SVM	88.29	0.90	0.88	0.83
	ANN	87.46	0.81	0.87	0.83
	KNN	86.15	0.80	0.86	0.83
	XGBOOST	88.20	0.80	0.88	0.83
	LightGBM	88.10	0.78	0.88	0.83
BINARY	BNB	83.95	0.79	0.84	0.81
	LR	86.78	0.84	0.87	0.85
	RF	88.29	0.78	0.88	0.83
	DT	83.08	0.82	0.84	0.83
	SVM	88.29	0.78	0.88	0.83
	ANN	86.25	0.83	0.86	0.84
	KNN	88.29	0.85	0.88	0.83
	XGBOOST	88.34	0.83	0.88	0.84
	LightGBM	88.34	0.84	0.88	0.85

Tablo 2, Amazon veri setinde yapılan makine öğrenimi çalışmalarının sonuçlarını kapsamlı bir şekilde sunmaktadır. Makine öğrenimi modellerinin doğruluk oranları %77.86 ile %88.54 arasında değişmektedir. En yüksek doğruluk oranı %88.54 ile Word2Vec temsil yöntemi ve XGBOOST algoritması kombinasyonu tarafından elde edilmiştir. Genel olarak, XGBOOST, LightGBM, ve SVM algoritmaları, farklı metin temsil yöntemleriyle kullanıldığında yüksek doğruluk oranları sağlamıştır.

Şekil 3'te verilen karışıklık matrisi, Amazon veri seti ile yapılan duygu analizi çalışmasında Word2Vec + XGBoost yapısının performansını göstermektedir. Model, pozitif duyguları (2) tespit etmekte oldukça başarılı olup 1808 doğru tahmin yapmıştır, ancak negatif (0) ve nötr (1) duyguları ayırt etmede yetersiz kalmaktadır. Negatif duyguları sadece 7 doğru tahminle, nötr duyguları ise sadece 1 doğru tahminle sınıflandırmıştır. Bu durum, modelin negatif ve nötr duyguları pozitif olarak sınıflandırma eğiliminde olduğunu göstermektedir.



Şekil 3. XGBoost karışıklık matrisi - Amazon veri seti

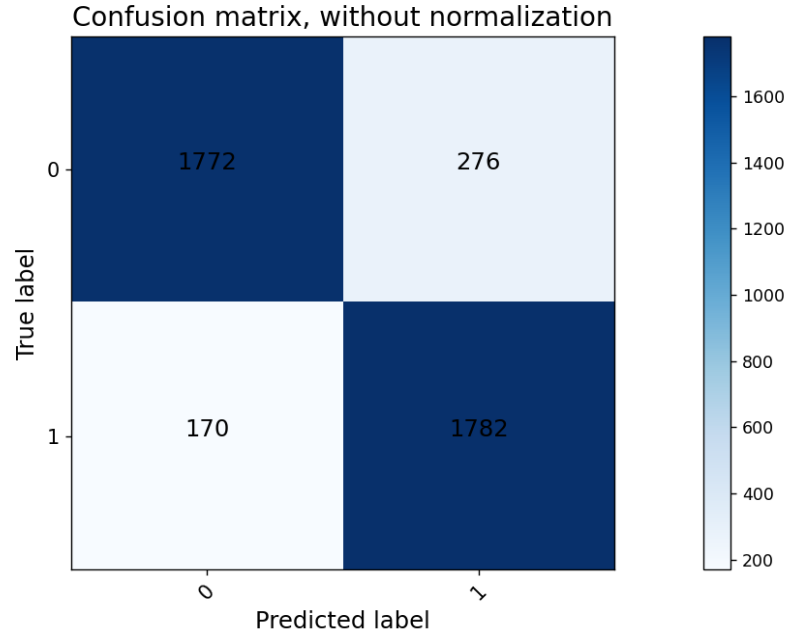
Tablo 3, IMDb veri setinde gerçekleştirilen makine öğrenimi çalışmalarının derinlemesine incelenmesini sunmaktadır. Çalışma %51.25 ile %88.85 arasında doğruluk oranları göstermiştir. En yüksek doğruluk oranı %88.85 ile TF-IDF ve SVM kombinasyonu tarafından elde edilmiştir. Genel olarak, TF-IDF ve Binary metin temsil yöntemleri, SVM ve LR algoritmaları ile yüksek doğruluk oranları sağlamıştır. TF-IDF ve SVM en iyi performansı verirken, GloVe temsil yöntemi genellikle daha düşük doğruluk oranları göstermiştir.

Tablo 3. IMDb veri seti makine öğrenimi sonuçları.

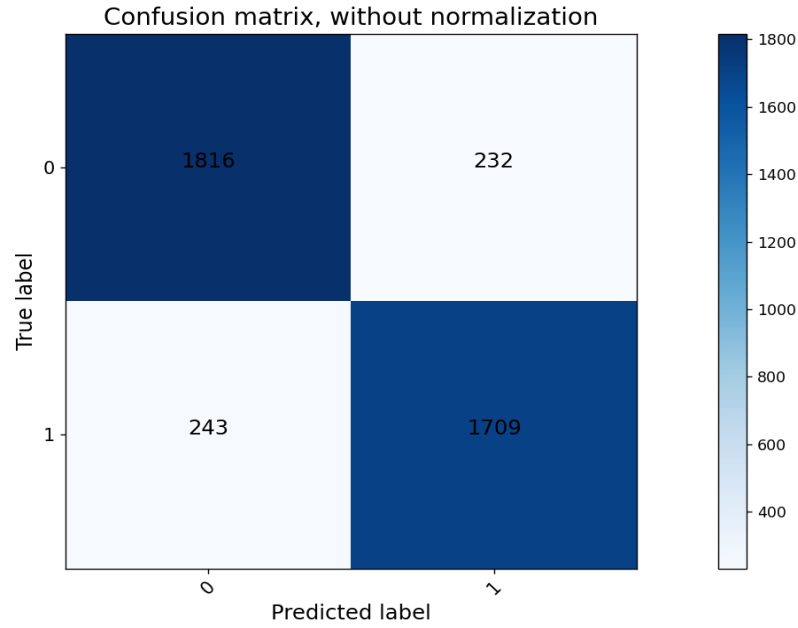
Metin Yöntemi	Temsil	Sınıflandırma Algoritması	Doğruluk(%)	Hassasiyet	Duyarlılık	F1-Score
TF-IDF		BNB	85.65	0.86	0.86	0.86
		LR	88.37	0.88	0.88	0.88
		RF	84.05	0.84	0.84	0.84
		DT	70.17	0.70	0.70	0.70
		SVM	88.85	0.89	0.89	0.89
		ANN	87.45	0.87	0.87	0.87
		KNN	74.52	0.75	0.75	0.74
		XGBOOST	83.52	0.84	0.84	0.84
		LightGBM	84.70	0.85	0.85	0.85
WORD2VEC		BNB	51.25	0.59	0.51	0.35
		LR	86.65	0.87	0.87	0.87
		RF	83.02	0.83	0.83	0.83
		DT	71.97	0.72	0.72	0.72
		SVM	86.87	0.87	0.87	0.87

	ANN	84.47	0.85	0.85	0.85
	KNN	80.55	0.81	0.81	0.81
	XGBOOST	85.20	0.85	0.85	0.85
	LightGBM	84.77	0.85	0.85	0.85
GLOVE	BNB	67.00	0.67	0.67	0.67
	LR	75.75	0.76	0.76	0.76
	RF	73.15	0.73	0.73	0.73
	DT	61.65	0.62	0.62	0.62
	SVM	75.05	0.75	0.75	0.75
	ANN	72.60	0.73	0.73	0.73
	KNN	66.12	0.66	0.66	0.66
	XGBOOST	74.02	0.74	0.74	0.74
	LightGBM	73.77	0.74	0.74	0.74
FASTTEXT	BNB	71.75	0.72	0.72	0.72
	LR	85.07	0.85	0.85	0.85
	RF	79.85	0.80	0.80	0.80
	DT	69.35	0.69	0.69	0.69
	SVM	85.22	0.85	0.85	0.85
	ANN	81.70	0.82	0.82	0.82
	KNN	77.05	0.77	0.77	0.77
	XGBOOST	82.82	0.83	0.83	0.83
	LightGBM	82.67	0.83	0.83	0.83
BINARY	BNB	85.72	0.86	0.86	0.86
	LR	87.12	0.87	0.87	0.87
	RF	84.22	0.84	0.84	0.84
	DT	70.70	0.71	0.71	0.71
	SVM	88.07	0.88	0.88	0.88
	ANN	86.15	0.86	0.86	0.86
	KNN	59.87	0.62	0.60	0.57
	XGBOOST	84.10	0.84	0.84	0.84
	LightGBM	85.07	0.85	0.85	0.85

Şekil 4’te verilen karışıklık matrisi, IMDb veri seti ile yapılan duygu analizi çalışmasında TF-IDF + SVM algoritmasının performansını göstermektedir. Model, pozitif (1) ve negatif (0) duyguları oldukça başarılı bir şekilde ayırt edebilmektedir. Negatif duygular için 1772 doğru tahmin ve 276 yanlış tahmin yapılmıştır. Pozitif duygular içinse 1782 doğru tahmin ve 170 yanlış tahmin yapılmıştır. Bu sonuçlar, modelin genellikle yüksek doğrulukla çalıştığını ve her iki duygu sınıfını da iyi ayırt edebildiğini göstermektedir.



Şekil 4. SVM karışıklık matrisi - IMDB veri seti



Şekil 5. LSTM karışıklık matrisi - IMDB veri seti

Tablo 4, IMDB film incelemeleri veri seti üzerinde gerçekleştirilen derin öğrenme çalışmalarının sonuçlarını detaylandırmaktadır. Çalışma %48.00 ile %88.12 arasında değişen doğruluk oranları göstermiştir. En yüksek doğruluk oranı %88.12 ile FastText temsil yöntemi ve LSTM algoritması kombinasyonu tarafından elde edilmiştir. Genel olarak, TF-IDF ve FastText temsil yöntemleri CNN ve LSTM algoritmaları ile yüksek doğruluk oranları sağlamıştır. Glove ve Binary temsil yöntemleri ise genellikle daha düşük doğruluk oranları sergilemiş, özellikle RNN algoritmaları ile düşük performans göstermiştir.

Tablo 4. IMDb veri seti derin öğrenme sonuçları.

Metin Temsil Yöntemi	Sınıflandırma Algoritması	Doğruluk(%)	Hassasiyet	Duyarlılık	F1-Score
TF-IDF	CNN	87.24	0.88	0.87	0.87
	RNN	51.80	0.75	0.49	0.32
	LSTM	51.20	0.75	0.51	0.35
WORD2VEC	CNN	84.77	0.85	0.85	0.85
	RNN	76.67	0.80	0.77	0.76
	LSTM	85.37	0.85	0.85	0.85
GLOVE	CNN	75.02	0.77	0.75	0.75
	RNN	69.57	0.71	0.70	0.69
	LSTM	80.54	0.82	0.81	0.80
FASTTEXT	CNN	85.07	0.86	0.85	0.85
	RNN	73.65	0.79	0.74	0.73
	LSTM	88.12	0.88	0.88	0.88
BINARY	CNN	78.50	0.79	0.79	0.78
	RNN	48.00	0.49	0.48	0.33
	LSTM	52.00	0.75	0.52	0.36

Şekil 5’te verilen karışıklık matrisi, IMDb veri seti ile yapılan duygu analizi çalışmasında FastText + LSTM algoritmasının performansını detaylı olarak vermektedir. Model, negatif duyguları (0) 1816 doğru tahmin ve 232 yanlış tahminle, pozitif duyguları (1) ise 1709 doğru tahmin ve 243 yanlış tahminle sınıflandırmıştır. Bu sonuçlar, modelin duygusal sınıflandırmada oldukça başarılı olduğunu, ancak 232 ve 243 yanlış tahminle bazı hatalar yapabildiğini göstermektedir.

4.1. Önceki Çalışmalar ile Kıyaslama

Tablo 5’te, literatür çalışmasında incelenen çalışmaların sonuçları ile bu çalışmadaki veri setlerinden elde edilen en yüksek doğruluk oranı sonuçları karşılaştırılmaktadır. Bu tablo, farklı metin temsil yöntemleri ve algoritmaların performansını değerlendirmek için önemli bir referans sağlar. Genel olarak, en çok kullanılan metin temsil yöntemleri arasında TF-IDF, Word2vec ve Glove bulunmaktadır. Bu yöntemler, farklı veri setlerinde çeşitli makine öğrenimi ve derin öğrenme algoritmaları ile birleştirilerek test edilmiştir. Doğruluk oranları, veri seti ve kullanılan yöntemlere bağlı olarak geniş bir yelpazede değişmektedir.

Çalışmada, Amazon veri seti için Word2vec temsili ve XGBoost algoritması kullanılarak %88.54 doğruluk oranına, IMDb veri seti için ise TF-IDF temsili ve SVM algoritması kullanılarak

%88.85 doğruluk oranına ulaşılmıştır. Bu sonuçlar, literatürdeki benzer çalışmalara kıyasla yüksek doğruluk oranlarına işaret etmektedir ve kullanılan metin temsil yöntemleri ile algoritmaların etkinliğini göstermektedir. Genel olarak, çalışmamızdaki sonuçlar, farklı metin temsil yöntemleri ve algoritmaların performansını değerlendirmek için önemli bir referans sağlar ve duygu analizi alanında seçilen yöntemlerin ve algoritmaların güçlü performansını ortaya koymaktadır.

Tablo 5 göz önüne alındığında, literatürde yer alan benzer araştırmalarla bu çalışma karşılaştırıldığında belirli tutarlılıklar ve farklılıklar görülmektedir. Örneğin, (Öge ve Kayaalp, 2021) tarafından yapılan çalışmada IMDB veri seti kullanılarak çeşitli makine öğrenimi algoritmalarının duygu analizi performansları incelenmiş ve Lojistik Regresyon algoritmasının en yüksek doğruluk oranına ulaştığı raporlanmıştır. Bu çalışmada da IMDB veri seti üzerinde Lojistik Regresyon algoritması %88 doğruluk oranı ile en yüksek başarılarından birini göstermiştir, bu sonuç literatürdeki bulgularla uyumludur. Benzer şekilde, (Fu ve ark., 2018) tarafından yapılan çalışmada, metin duygu analizi için LSTM modellerinin kelime gömme yöntemleri ile birleştirildiğinde yüksek performans sergilediği ortaya konulmuştur. Çalışmamızda da FastText temsili ile LSTM algoritması, IMDB veri setinde %88.12 doğruluk oranı ile en iyi sonucu vermiştir, bu da LSTM'in uzun vadeli bağıntıları yakalama konusunda literatürde belirtilen avantajlarını desteklemektedir. Bununla birlikte, (Rumelli ve ark., 2019) tarafından yapılan çalışmada, Türkçe metinler için klasik makine öğrenimi algoritmalarının belirli durumlarda derin öğrenme yöntemlerine kıyasla daha verimli çalıştığı vurgulanmıştır. Amazon veri setinde elde edilen sonuçlar da LightGBM algoritmasının %88.39 doğruluk oranı ile en iyi makine öğrenimi modeli olduğunu göstermekte ve bu bulgu literatürdeki benzer çalışmalarla örtüşmektedir. Bu karşılaştırmalar, çalışmada elde edilen bulguların mevcut literatürdeki yöntemlerle büyük ölçüde uyumlu olduğunu ve duygu analizi alanındaki genel eğilimleri desteklediğini göstermektedir.

Çalışmada elde edilen yüksek doğruluk oranlarının başlıca sebepleri arasında kullanılan metin temsil yöntemlerinin ve sınıflandırma algoritmalarının dikkatli bir şekilde seçilmesi ve farklı kombinasyonlarla test edilmesi yer almaktadır. Öncelikle, TF-IDF ve FastText gibi metin temsil yöntemleri, kelime sıklığı ve kelimeler arası ilişkiyi daha iyi modelleyerek metinlerin daha anlamlı bir şekilde temsil edilmesini sağlamıştır. Ayrıca, LightGBM ve XGBoost gibi gelişmiş gradient boosting algoritmaları, büyük veri kümeleri üzerinde daha yüksek doğruluk oranları elde edilmesini mümkün kılmıştır. Bu algoritmalar, veri içindeki önemli özellikleri belirleme ve sınıflandırma kararlarını optimize etme konusunda daha etkili çalışmaktadır. Bunun yanı sıra, LSTM modeli, metinlerdeki uzun vadeli bağıntıları başarılı bir şekilde yakalayarak derin öğrenme tabanlı duygu analizinde diğer yöntemlere kıyasla daha iyi sonuç vermiştir. Literatürde yapılan çalışmaların çoğunda, tek bir metin temsil yöntemi veya sınıflandırma algoritması kullanılırken, bu çalışmada birden fazla yöntem detaylı olarak karşılaştırılmış ve en iyi performans gösteren kombinasyon

belirlenmiştir. Ayrıca, iki farklı veri kümesi üzerinde yapılan deneyler sayesinde yöntemin genel uygulanabilirliği test edilmiş ve yüksek doğruluk oranlarının veri setlerine bağlı olmadığı gösterilmiştir. Bu faktörler bir araya geldiğinde, çalışmada elde edilen sonuçların literatürdeki diğer çalışmalara kıyasla genel olarak daha başarılı olmasını sağlamıştır.

Tablo 5. Var olan çalışmalar ile kıyaslama.

Çalışmalar	Metin Temsil Yöntemi	Algoritmalar	Doğruluk %	Veri Kümesi
(Rawat ve ark., 2022)	TF-IDF	BNB SVC LR	81 97 95	Twitter
(Demir ve Bilgin, 2023)	-	DistilBERT RoBERT AlBERT NB SVC RF	80 80 77 71 68 68	Haber Metinleri
(Öge ve Kayaalp, 2021)	-	LR	87	IMDb
(Fu ve ark., 2018)	Glove	WALE-LSTM LSTM	96 80.8	Kitap İncelemesi
(Rumelli ve ark., 2019)	Bigram	k-NN NB SVM RF	73.8 73.2 73.4 73.3	Hepsiburada Kullanıcı Yorumları
(Wang ve ark., 2021)	RGWE	CNN Bİ-LSTM	90.1 91.0	IMDb
(Hassan ve ark., 2021)	N-Grams	SVM k-NN DT NB	81.8 58.1 63.7 51.6	Twitter
(Başarslan ve Kayaalp, 2023)	Word2vec	SVM LR k-NN NB	87.8 81.9 80.6 81.6	Film Yorumları
(Poornima ve Priya, 2020)	TF-IDF	LR MNB SVM	86.23 83.54 85.69	Twitter

(Zhang ve Zheng, 2016)	TF-IDF	ELM kernels	88.74	Çince Metinler
		SVM	88.54	
(Sharma ve ark., 2023)	-	CNN	93.5	Çince Ürün
		LSTM	77.5	İncelemeleri
(Xu ve Song, 2023)	Word2vec	CNN	97.9	Chnsenticorp
		LSTM	97.68	veri seti
		GRU	97.89	
		Bert+BiLSTM	97.37	
		Bert + CNN	98.34	
(Aydın ve ark., 2021)	TF-IDF	SVM	73.50	Twitter
		LR	75.84	
		k-NN	62.03	
		ANN	70.39	
(Salur ve Aydın, 2022)	Fasttext	LSTM	74.33	Twitter
		Bİ-LSTM	74.68	
		GRU	74.32	
		BERT	75.34	
(Singh ve ark., 2022)	TF-IDF	RF	77.90	Twitter
		SVM	83.76	
		GB	80.46	
		XGBoost	81.55	
(Saraswat ve ark., 2023)	Bag-of-Words	ANN	84	Ses Dosyası
(Pandya ve ark., 2021)	Word2vec	SVM	76.53	Twitter
		RNN	78.15	
		LSTM	80.21	
		Bi-LSTM	81.33	
(Kemaloğlu ve ark., 2021)	Bag-of-Words	MLR	80.91	Twitter
		RF	83.93	
(Hammad ve Anwar, 2019)	N-gram	SVM	66.00	Twitter
		NB	62.30	
		DT	68.20	
		K-NN	67.40	
(Jayakody ve Kumara, 2021)	TF-IDF	SVM	86.47	Twitter
		LR	87.2	
		K-NN	83.87	

<i>BU ÇALIŞMA</i>	<i>Word2vec</i>	<i>XGBoost</i>	<i>88.54</i>	<i>AMAZON</i>
<i>BU ÇALIŞMA</i>	<i>TF-IDF</i>	<i>SVM</i>	<i>88.85</i>	<i>IMDb</i>

5. Genel Değerlendirme ve Sonuçlar

Bu çalışmada, Amazon ve IMDb veri setleri kullanılarak duygu analizi gerçekleştirilmiş ve çeşitli metin temsilleri ile sınıflandırma algoritmalarının performansları karşılaştırılmıştır. Amazon veri setinde üç sınıflı (olumlu, olumsuz, tarafsız), IMDb veri setinde ise iki sınıflı (olumlu, olumsuz) sınıflandırma yapılmıştır. Elde edilen sonuçlar, derin öğrenme ve makine öğrenmesi yöntemlerinin farklı sınıflandırma problemlerindeki performansını ortaya koymaktadır.

Amazon veri seti üzerinde yapılan analizlerde, derin öğrenme yöntemleri arasında en yüksek doğruluk oranı %88.30 ile TF-IDF ve FastText kullanan LSTM algoritması tarafından elde edilmiştir. Bu sonuç, LSTM'in metinlerdeki uzun süreli bağıntıları yakalamadaki üstün başarısını ortaya koymaktadır. CNN ve RNN modelleri de benzer doğruluk oranlarına ulaşmış, ancak LSTM biraz daha yüksek performans sergilemiştir. Öte yandan, makine öğrenmesi yöntemleri arasında LightGBM algoritması %88.39 doğruluk oranı ile en yüksek performansı gösterirken, Lojistik Regresyon algoritması %88.34 doğruluk oranı ile dikkat çekici bir performans sergilemiştir. Elde edilen bulgular, makine öğrenmesi algoritmalarının da yüksek doğruluk oranlarına ulaşabildiğini ve bazı durumlarda derin öğrenme yöntemlerine kıyasla daha hızlı ve verimli olduğunu ortaya koymaktadır.

IMDb veri seti üzerinde gerçekleştirilen analizlerde, derin öğrenme yöntemleri arasında en yüksek doğruluk oranı %87.84 ile FastText kullanan LSTM algoritması tarafından elde edilmiştir; bu sonuç LSTM'in metinlerdeki uzun süreli bağıntıları başarılı bir şekilde yakaladığını göstermektedir. Ayrıca, Word2Vec ile LSTM %85.37 doğruluk oranı ile yüksek bir performans sergilemiştir. Makine öğrenmesi yöntemleri arasında ise Lojistik Regresyon %88.85 doğruluk oranı ile en yüksek performansı gösterirken, Support Vector Machine algoritması %88.07 doğruluk oranı ile dikkat çekici bir başarı elde etmiştir. Bu sonuçlar, iki sınıflı duygu analizinde makine öğrenmesi algoritmalarının oldukça etkili olduğunu ortaya koymaktadır.

Sonuç olarak, bu çalışmada Amazon ve IMDb veri setlerinde duygu analizi gerçekleştirilmiş ve derin öğrenme ile makine öğrenmesi yöntemleri karşılaştırılmıştır. Amazon veri setinde en yüksek doğruluk oranı LightGBM algoritması ile elde edilirken (%88.39), IMDb veri setinde Lojistik

Regresyon algoritması %88.85 ile en yüksek doğruluk oranına ulaşmıştır. Derin öğrenme modelleri, özellikle büyük ve karmaşık veri setlerinde daha avantajlı olmasına rağmen, bazı durumlarda makine öğrenmesi modelleri daha hızlı ve verimli sonuçlar sunabilmektedir. İki sınıflı ve üç sınıflı sınıflandırma problemleri, model performansı ve sınıflandırma zorluğu açısından farklılıklar göstermektedir. Bu bulgular, duygu analizi projelerinde kullanılacak yöntemi belirlerken dikkate alınması gereken önemli faktörleri ortaya koymaktadır.

Teşekkür

Bu çalışmada açık kaynaklı iki veri kümesi kullanılmıştır. İlk veri kümesine <https://www.kaggle.com/datasets/eswarchandt/amazon-music-reviews> bağlantından erişilebilir. İkinci veri kümesine ise <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews> bağlantından erişilebilir. Bu veri kümelerini hazırlayıp kullanıma sunan kuruluşlara teşekkür ederiz.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Aydın, Z. E., Ozturk, Z. K., & Cıcek, Z. I. E. (2021). Turkish sentiment analysis for open and distance education systems. *Turkish Online Journal of Distance Education*, 22(3), 124-138.
- Başarslan, M. S., & Kayaalp, F. (2023). Sentiment analysis with ensemble and machine learning methods in multi-domain datasets. *Turkish Journal of Engineering*, 7(2), 141-148.
- Bhasin, A., & Das, S. (2021, May). Twitter sentiment analysis using Machine Learning and Hadoop: A comparative study. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)* (pp. 267-272). IEEE.
- Can, U., & Alatas, B. (2017). Duygu analizi ve fikir madenciliği algoritmalarının incelenmesi. *International Journal of Pure and Applied Sciences*, 3(1), 75-111.

- Chandra, Y., & Jana, A. (2020, March). Sentiment analysis using machine learning and deep learning. In *2020 7th international conference on computing for sustainable global development (INDIACom)* (pp. 1-4). IEEE.
- Chen, L. C., Lee, C. M., & Chen, M. Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, 24(11), 8187-8197.
- Dhariwal, N., Akunuri, S. C., & Banu, K. S. (2023). Audio and Text Sentiment Analysis of Radio Broadcasts. *IEEE Access*, 11, 126900-126916.
- Demir, E., & Bilgin, M. (2023, September). Sentiment Analysis from Turkish News Texts with BERT-Based Language Models and Machine Learning Algorithms. In *2023 8th International Conference on Computer Science and Engineering (UBMK)* (pp. 01-04). IEEE.
- Fu, X., Yang, J., Li, J., Fang, M., & Wang, H. (2018). Lexicon-enhanced LSTM with attention for general sentiment analysis. *IEEE Access*, 6, 71884-71891.
- Hammad, M., & Anwar, H. (2019, November). Sentiment analysis of sindhi tweets dataset using supervised machine learning techniques. In *2019 22nd International Multitopic Conference (INMIC)* (pp. 1-6). IEEE.
- Hassan, S. M., Khan, J., Khan, M. A., Khan, M. S., Ahmad, I., & Khan, M. (2021). Detecting COVID-19 Pandemic Using Sentiment Analysis of Tweets. *Artificial Intelligence Theory and Applications*, 1(2), 39-47.
- He, H., Zhou, G., & Zhao, S. (2022). Exploring E-commerce product experience based on fusion sentiment analysis method. *IEEE Access*, 10, 110248-110260.
- Jagadeesan, M., Saravanan, T. M., Selvaraj, P. A., Ali, U. A., Arunsivaraj, J., & Balasubramanian, S. (2022, December). Twitter Sentiment Analysis with Machine Learning. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 681-686). IEEE.
- Jayakody, J. P. U. S. D., & Kumara, B. T. G. S. (2021, December). Sentiment analysis on product reviews on twitter using Machine Learning Approaches. In *2021 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 1056-1061). IEEE.
- Joshi, V., Patel, S., Agarwal, R., & Arora, H. (2023, March). Sentiments Analysis using Machine Learning Algorithms. In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1425-1429). IEEE.
- Karayığit, H., Acı, Ç., & Akdağlı, A. (2018). A Review of Turkish Sentiment Analysis and Opinion Mining. *Balkan Journal of Electrical and Computer Engineering*, 6(2), 94-98.
- Kim, R. Y. (2021). Using online reviews for customer sentiment analysis. *IEEE Engineering Management Review*, 49(4), 162-168.
- Kemaloğlu, N., Küçükşille, E., & Özgünsür, M. (2021). Turkish sentiment analysis on social media. *Sakarya University Journal of Science*, 25(3), 629-638.
- Li, Z., Li, R., & Jin, G. (2020). Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. *IEEE Access*, 8, 75073-75084.
- Liu, H., Chen, X., & Liu, X. (2022). A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis. *IEEE Access*, 10, 32280-32289.
- Onan, A. (2018). Sentiment analysis on Twitter based on ensemble of psychological and linguistic feature sets. *Balkan Journal of Electrical and Computer Engineering*, 6(2), 69-77.
- Öge, B. C., & Kayaalp, F. (2021). Farklı Sınıflandırma Algoritmaları ve Metin Temsil Yöntemlerinin Duygu Analizinde Performans Karşılaştırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(6), 406-416.
- Pandya, V., Somthankar, A., Shrivastava, S. S., & Patil, M. (2021, December). Twitter sentiment analysis using machine learning and deep learning techniques. In *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)* (pp. 1-5). IEEE.
- Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A survey of text representation and embedding techniques in nlp. *IEEE Access*, 11, 36120-36146.
- Poornima, A., & Priya, K. S. (2020, March). A comparative sentiment analysis of sentence embedding using machine learning techniques. In *2020 6th international conference on advanced computing and communication systems (ICACCS)* (pp. 493-496). IEEE.
- Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research. *IEEE transactions on affective computing*, 14(1), 108-132.
- Rawat, A., Maheshwari, H., Khanduja, M., Kumar, R., Memoria, M., & Kumar, S. (2022, May). Sentiment analysis of Covid19 vaccines tweets using NLP and machine learning classifiers. In *2022 International*

- Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* (pp. 225-230). IEEE.
- Rizk, Y. E., & Asal, W. M. (2021, October). Sentiment analysis using machine learning and deep learning models on movies reviews. In *2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 129-132). IEEE.
- Rumelli, M., Akkuş, D., Kart, Ö., & Isik, Z. (2019, October). Sentiment analysis in Turkish text with machine learning algorithms. In *2019 Innovations in intelligent systems and applications conference (ASYU)* (pp. 1-5). IEEE.
- Salur, M. U., & Aydın, İ. (2022). Türkçe tweetler için derin özellik çıkarımı tabanlı yeni bir duygu sınıflandırma modeli. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(1), 1-13.
- Satrya, R. N., Pratiwi, O. N., Fa'rifah, R. Y., & Abawajy, J. (2022, November). Cryptocurrency sentiment analysis on the twitter platform using support vector machine (svm) algorithm. In *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)* (pp. 01-05). IEEE.
- Sarkar, K., & Bhowmick, M. (2017, December). Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines. In *2017 IEEE Calcutta Conference (CALCON)* (pp. 31-36). IEEE.
- Saraswat, S., Bhardwaj, S., Vashistha, S., & Kumar, R. (2023, March). Sentiment Analysis of Audio Files Using Machine Learning and Textual Classification of Audio Data. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-5). IEEE.
- Sharma, S., Pandey, A., Kumar, V., Ohdar, D., Pillai, A. R., & Mahajan, M. (2023, May). Recent Trends in Sentiment Analysis using Different Machine Learning based Models: A Short Review. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 474-481). IEEE.
- Singh, S., Kumar, K., & Kumar, B. (2022, May). Sentiment analysis of Twitter data using TF-IDF and machine learning techniques. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* (pp. 252-255). IEEE.
- Tan, Y. Y., Chow, C. O., Kanesan, J., Chuah, J. H., & Lim, Y. (2023). Sentiment analysis and sarcasm detection using deep multi-task learning. *Wireless personal communications*, 129(3), 2213-2237.
- Tsai, M. F., & Huang, J. Y. (2021). Sentiment analysis of pets using deep learning technologies in artificial intelligence of things system. *Soft Computing*, 25(21), 13741-13752.
- Wang, Y., Huang, G., Li, J., Li, H., Zhou, Y., & Jiang, H. (2021). Refined global word embeddings based on sentiment concept for sentiment analysis. *IEEE Access*, 9, 37075-37085.
- Xu, L., & Song, Y. (2023, April). Comparison of text sentiment analysis based on traditional machine learning and deep learning methods. In *2023 4th International Conference on Computer Engineering and Application (ICCEA)* (pp. 692-695). IEEE.
- Zhang, X., & Zheng, X. (2016, July). Comparison of text sentiment analysis based on machine learning. In *2016 15th international symposium on parallel and distributed computing (ISPDC)* (pp. 230-233). IEEE.