

Yapay Zekâ Çağında Duygu Analizi: Büyük Dil Modellerinin Yükselişi ve Klasik Yaklaşımlarla Karşılaştırılması

*Makale Bilgisi / Article Info

Alındı/Received: 15.05.2024

Kabul/Accepted: 05.08.2024

Yayımlandı/Published: xx.xx.xxxx

Sentiment Analysis in the Age of Artificial Intelligence: The Rise of Large Language Models and Comparison with Classical Approaches

Muhammed Abdulhamid KARABIYIK^{1*}, Asım Sinan YÜKSEL², Fatma Gülşah TAN³

¹ Niğde Ömer Halisdemir Üniversitesi, Bor Meslek Yüksekokulu, Bilgisayar Programcılığı Bölümü, Niğde, Türkiye

² Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Isparta, Türkiye

³ Süleyman Demirel Üniversitesi, Bilgi İşlem Daire Başkanlığı, Isparta, Türkiye

© Ayfon Kocatepe Üniversitesi

Öz

Duygu analizi, günümüzde hem bireylerin hem de şirketlerin karar alma süreçlerinde önemli bir rol oynayan, farklı veri kaynaklarından elde edilen bilgileri inceleyerek insan duygularını anlamayı sağlayan bir yöntemdir. Duygu analizi klasik makine öğrenmesi yöntemleriyle gerçekleştirildiğinde, metinlerin duygusal içeriğini anlamak için genellikle belirli özelliklerin elle seçilmesini ve öznelik mühendisliği gerektirir. Bu yöntemler, duygusal ifadelerin karmaşıklığını ve çok katmanlı yapısını tam olarak yakalayamamakta ve genellikle belirli bağlamlarda başarılı olmaktadır. Ancak, büyük dil modelleri, derin öğrenme prensiplerine dayanarak, karmaşık dil yapılarını daha etkili bir şekilde öğrenebilirler. Bu modeller, büyük ölçekteki metin verilerini işleyerek genel dil bilgisiyle donanmışlardır. Dolayısıyla, duygu analizi gibi görevlerde kullanıldıklarında, öznelik mühendisliği gerektirmeden karmaşık duygusal ifadeleri daha doğru bir şekilde çözebilirler. Çalışmamızda, üç farklı veri seti kullanılarak büyük dil modelleri ve klasik yöntemlerin duygu analizindeki performansları karşılaştırılmıştır. Sonuçlar, büyük dil modellerinin klasik yöntemlere kıyasla daha yüksek doğruluk oranları elde ettiğini, duygu analizi alanında gelecekte daha fazla kullanılacağını ve bu alandaki çalışmalara önemli katkılar sağlayacağını göstermektedir.

Anahtar Kelimeler Duygu Analizi; Büyük Dil Modelleri; Doğal Dil İşleme; İstem Mühendisliği, Makine Öğrenme.

Abstract

Sentiment analysis is a method that plays an important role in the decision-making processes of both individuals and companies today and enables understanding human emotions by examining information obtained from different data sources. When sentiment analysis is performed with traditional machine learning methods, it often requires hand-selection of specific features and feature engineering to understand the sentiment of texts. These methods cannot fully capture the complexity and multilayered nature of emotional expressions and are often successful in certain contexts. However, large language models can learn complex language structures more effectively, based on deep learning principles. These models are equipped with general language knowledge by processing large-scale text data. Therefore, when used in tasks such as sentiment analysis, they can more accurately decode complex emotional expressions without requiring feature engineering. In our study, the performances of large language models and classical methods were compared using three different data sets. The results show that large language models achieve higher accuracy rates compared to classical methods, will be applied more in the field of sentiment analysis in the future, and will make significant contributions to studies in this field.

Keywords Sentiment Analysis; Large Language Models; Natural Language Processing; Prompt Engineering; Machine Learning.

1. Giriş

Yapay zekâ, insan zekasını taklit etmeyi ve hatta aşmayı amaçlayan, bilgisayar bilimlerindeki en karmaşık ve çok yönlü konulardan biridir. Ancak, bu hedefe ulaşma süreci, bilgi işleme gücü ve algoritmik karmaşıklık açısından bir dizi zorlukla karşı karşıya kalmıştır. Yapay zekâ teorisi, bu zorlukları aşmak için çeşitli yaklaşımlar geliştirmiştir. Ancak geçmiş dönemlerdeki teknolojinin verimlilik, bellek ve işlem gücü açısından yetersizliği bu alandaki pratik uygulamaların önünde bir engel olarak durmuştur. Günümüzde bilgisayar teknolojisinin hızla ilerlemesiyle birlikte, bu durum değişmeye başlamış ve yapay zekânın

günlük hayata entegrasyonu giderek artmıştır. Özellikle yapay zekânın alt dalı olan derin öğrenme tekniklerinin gelişimiyle, yapay zekâ tabanlı çözümler daha yüksek hız ve doğruluk performanslarını elde etmişlerdir. Derin öğrenme yaklaşımları, klasik yapay zekâ yöntemleriyle çözülemeyen problemlerin aşılmasına yardımcı olmalarıyla dikkat çekmekte, zaman ve fiyat maliyetlerinin, sistem hata oranlarının düşürülmesine imkân sağlamaktadırlar (Khakurel vd. 2018).

Yapay zekâ teknikleri, görüntü, ses, sinyal ve metin gibi verilere dayalı problemlerin çözümünde hayati bir rol oynamaktadır. İnternetin yaygınlaşmasıyla birlikte, sosyal

medya platformları, bloglar, haber siteleri ve diğer çevrimiçi kaynaklar aracılığıyla özellikle metin tabanlı verilerin miktarı hızla artmaktadır. Bu artış son yıllarda doğal dil işleme çalışmalarına büyük bir ivme kazandırmıştır (Jouis ve Shafei 2018). Araştırmacılar, metin verilerini analiz etmek, anlamak ve kullanmak için yeni yöntemler geliştirmekte ve bu alandaki potansiyeli keşfetmektedirler. Bu bağlamda, doğal dil işleme alanında önemli bir araştırma ve geliştirme potansiyeline sahip temel konulardan birisi de duygu analizi olmuştur. Duygu analizi, kullanıcı etkileşimli sistemlerde hedef kitlelerin tepkilerinin tespiti açısından önemli bir konudur (Yue vd. 2019). Duyguların başarılı bir şekilde tespit edilmesi, ticari ve sosyal konularda kurumların operasyonlarını faydalı bir şekilde geliştirmelerini sağlamaktadır. Faydalarının yanında belirli zorlukları da beraberinde getirmektedir. Bu zorluklardan ilki, insan duygularının geniş çeşitliliği ve metinlerdeki duygusal ifadelerin karmaşıklığı ve belirsizliğidir. Dil ve kültürel farklılıklar da bir başka zorluktur; çünkü duygusal ifadeler, dil ve kültürel bağlamlardan etkilenmekte ve bu da yapay zekâ modellerinin genelleme yeteneği kazanmasını zorlaştırmaktadır. Ayrıca, metinlerde sıklıkla kullanılan ironi ve mizah gibi dil işlevleri, duygu analizi modelleri için zorluklar oluşturmaktadır. Son olarak, etiketli veri toplama süreci öznel olabilmekte ve insanlar arasında duygusal ifadelerin yorumlanması farklılık gösterebilmektedir. Bu da model eğitimi için ek bir zorluk yaratmaktadır. Bu zorluklar, metin tabanlı duygu analizi alanında çalışan araştırmacılar ve uygulayıcılar için önemli sorunlar oluşturmakta ve yeni yöntemlerin geliştirilmesini ve mevcut modellerin iyileştirilmesini gerekli kılmaktadır (Khurana vd. 2023).

Büyük dil modelleri (BDM), son yıllarda yapay zekâ alanında büyük bir gelişme göstermiştir. Çok büyük metin ve kod veri kümeleri üzerinde eğitilmiş olan bu modeller, metin üretme, makine çevirisi, farklı türde içerikler yazma ve sorulan sorulara cevap verme gibi çeşitli görevlerde insan benzeri performans sergilemektedirler. Duygu analizinde, klasik gözetimli sınıflandırma yöntemlerinden faydalanılmaktadır (Hasan vd. 2018). Ancak, BDM'lerin gelişimiyle birlikte bu alan yeni bir boyut kazanmıştır. Temelinde derin öğrenme tekniklerinden faydalanılması, büyük miktarda verilerle eğitilmiş olmaları doğal dilin yapısını kolaylıkla öğrenmelerini sağlamıştır.

Bu çalışmada klasik yöntemler ve BDM'lerin duygu analizi problemindeki başarıları karşılaştırılmıştır. Yapılan karşılaştırmalarda üç farklı veri seti kullanılmıştır. Bu veri setleri, hava yolu şirketlerine, finansal gelişmelere ve filmlere yapılan yorumları kapsamaktadır. Klasik yöntemlerde deneyler farklı sayıda eğitim verisi kullanılarak gerçekleştirilmiştir. Deneyin her adımında eğitim seti genişletilerek doğruluk oranları yeniden hesaplanmıştır. Eğitim setinin her adımda

genişletilmesindeki amaç ise klasik yöntemlerin daha fazla veriyle eğitilerek BDM'lerin performanslarını yakalayıp yakalamadıklarını gözlemlemektir. BDM deneylerinde ise Sıfır Atış İstem (Zero Shot Prompting) yöntemi kullanılarak ön eğitilmiş modellerin doğruluk oranı hesaplanmıştır. Sıfır atış istem, sadece BDM'de yer alan bilgiyi kullandığı için ek bir eğitim maliyeti oluşturmamıştır.

2. Materyal ve Metot

Duygu analizi için seçilen klasik makine öğrenmesi teknikleriyle BDM'leri karşılaştırma amacınız, duygu analizi alanında mevcut en etkili ve güncel yöntemlerin belirlenmesidir. Klasik yöntemlerden, duygu analizi gibi doğal dil işleme görevlerinde sıklıkla kullanılan ve kapsamlı bir literatür desteğine sahip olan Rastgele Orman (RO), Lojistik Regresyon (LR), Destek Vektör Makineleri (DVM), Gauss Sade Bayes (GSB), K - En Yakın Komşu (K-EYK), AdaBoost ve Yapay Sinir Ağları (YSA) seçilmiştir.

BDM'lerde, son yıllarda yapay zekâ alanında devrim niteliğindeki ilerlemelerin ürünleri olan BERT, GPT 3.5 Turbo ve Gemini modelleri seçilmiştir. Bu modeller, derin öğrenme prensiplerine dayanarak geniş bir dil anlama yeteneğiyle bilinmektedirler. Özellikle BERT, Transformer mimarisi sayesinde doğal dil işleme alanında çığır açmış ve birçok dil işleme görevinde liderlik etmiştir. GPT 3.5 Turbo ise OpenAI tarafından geliştirilen ve daha önceki versiyonların performansını büyük ölçüde aşan son derece büyük bir dil modelidir. Gemini ise daha yeni bir model olmasına rağmen, özellikle genel anlamda daha dengeli bir performans sağlamak amacıyla tasarlanmıştır.

BDM'ler ve klasik yöntemlerin performansları, üç farklı veri seti kullanılarak karşılaştırılmıştır. Veri setleri dengeli bir şekilde oluşturulmuştur; dolayısıyla, performans değerlendirmesi için yalnızca doğruluk oranı kullanılmıştır. Modellerin performansını daha adil bir şekilde karşılaştırabilmek için test veri seti sabit tutulmuştur.

2.1 Veri Setleri

Performans testlerinde kullanılmak üzere 3 farklı veri seti hazırlanmıştır. Bu veri setleri hava yolu şirketlerine, finansal gelişmelere ve filmlere yapılan yorumlar olarak 3 ana kategoriden oluşmaktadır. Veri setleri eğitim ve test olarak iki parçaya ayrılırken, test veri setleri sabit boyutta tutulmuştur. Bunun sebebi, farklı büyüklüklerdeki eğitim setleriyle geliştirilen modellerin aynı veri seti üzerindeki başarılarının kıyaslanabilmesidir. Testler, eğitim setlerindeki veriler sınıf başına 50 veri artırılarak yapılmıştır. Hava yolları yorumlarını içeren veri setinde pozitif, negatif ve nötr olmak üzere 3 sınıf mevcuttur (Rane ve Kumar 2018). Başlangıç olarak 150 eğitim ve 150 test verisi olmak üzere modeller eğitilmiştir. Testler,

eğitim veri seti boyutu her adımda 150 veri artırılarak tekrarlanmıştır. Toplam 46 adet test yapılmış ve son modelin eğitimi 7050 eğitim ve 150 test verisiyle gerçekleştirilmiştir.

Finansal gelişmelere yapılan yorumları içeren veri setinde pozitif, negatif ve nötr olmak üzere 3 sınıf mevcuttur (Malo vd. 2014). Başlangıç olarak 150 eğitim ve 150 test veri seti olmak üzere modeller eğitilmiştir. Eğitim veri seti boyutu her adımda 150 veri artırılarak testler tekrarlanmıştır. Toplam 16 adet test yapılmış ve son modelin eğitimi 2400 eğitim ve 150 test verisiyle gerçekleştirilmiştir. Film yorumlarını içeren veri setinde pozitif ve negatif olmak üzere 2 sınıf mevcuttur (Maas vd. 2011). Başlangıç olarak 100 eğitim ve 100 test verisi olmak üzere modeller eğitilmiştir. Eğitim seti boyutu her adımda 100 veri artırılarak testler tekrarlanmıştır. Toplam 30 adet test yapılmış ve son modelin eğitimi 3000 eğitim ve 100 test verisiyle gerçekleştirilmiştir. Veri setleriyle ilgili özet bilgiler Çizelge 1’de gösterilmiştir.

Çizelge 1. Kullanılan veri setleri temel bilgileri

Veri Seti	Eğitim seti		Test Seti Boyutu
	Başlangıç Boyutu	Eğitim seti Bitiş Boyutu	
Finans	150	2400	150
Hava Yolları	150	7050	150
Film	100	3000	100

2.2 Metin Ön İşleme

Doğal dil işlemede, makine öğrenmesi yöntemlerinin kullanılmasında yapısal zorluklar ortaya çıkmaktadır. Bu sebeple metinler bazı ön işlem aşamalarından geçirilmektedir. Çalışmamızda kullanılan ön işleme adımları Şekil 1’de gösterilmiştir.

2.3 Klasik Makine Öğrenmesi Yöntemleri

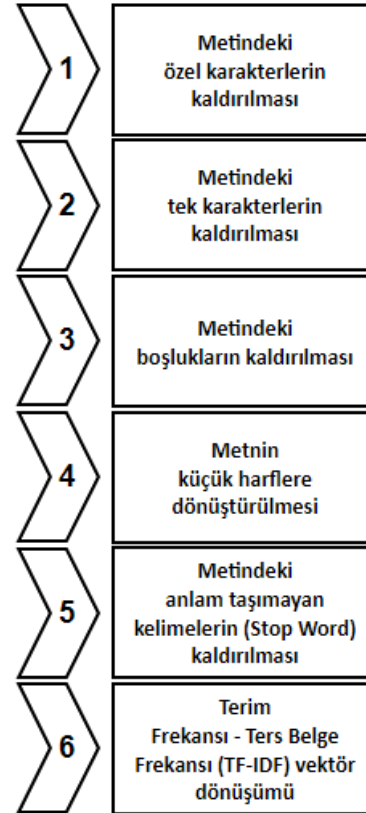
Klasik makine öğrenmesi yöntemleri ile metin sınıflandırma, veri toplama, metin ön işleme ve özellik çıkarma adımlarını içerir. Bu adımların ardından bir model eğitilir ve değerlendirilir.

Eğitilen model, yeni metin verilerini belirli kategorilerde sınıflandırmak için kullanılır. Duygu analizinde ise metnin hedef duygulara göre etiketlenmesi gerekir. Örneğin, ifadenin duygusu insan kararıyla pozitif, negatif ve nötr olarak etiketlenir. Etiketlenen ifade, metin ön işleme adımlarından geçirilerek sınıflandırma modelinin eğitimi yapılır. Modelin eğitimi tamamlandıktan sonra elde edilen model duygu analizinin tespitinde kullanılır.

2.3.1 Rastgele Orman

Rastgele Orman algoritması, temel olarak karar ağaçları üzerine kurulmuş bir yöntemdir. Alt kümelere ayrılmış

birden fazla karar ağacının sonuçlarını değerlendirir ve alt küme sonuçlarının yoğunluğuna göre nihai sonucu elde eder. Kullanılan bu dağıtık yapı modeli ezberden (aşırı uyum) koruyarak modelin daha etkili sonuçlar elde etmesini sağlar. Ancak alt karar ağaçlarının sayısı arttıkça oluşan karmaşıklık kullanım zorlukları oluşturmaktadır (Kılınçarslan vd. 2020). Şekil 2’de RO yöntemine ait çalışma prensibi gösterilmiştir. Deneysel olarak kullanılan RO yönteminde toplam alt ağaç sayısı 200 olarak kullanılmıştır.



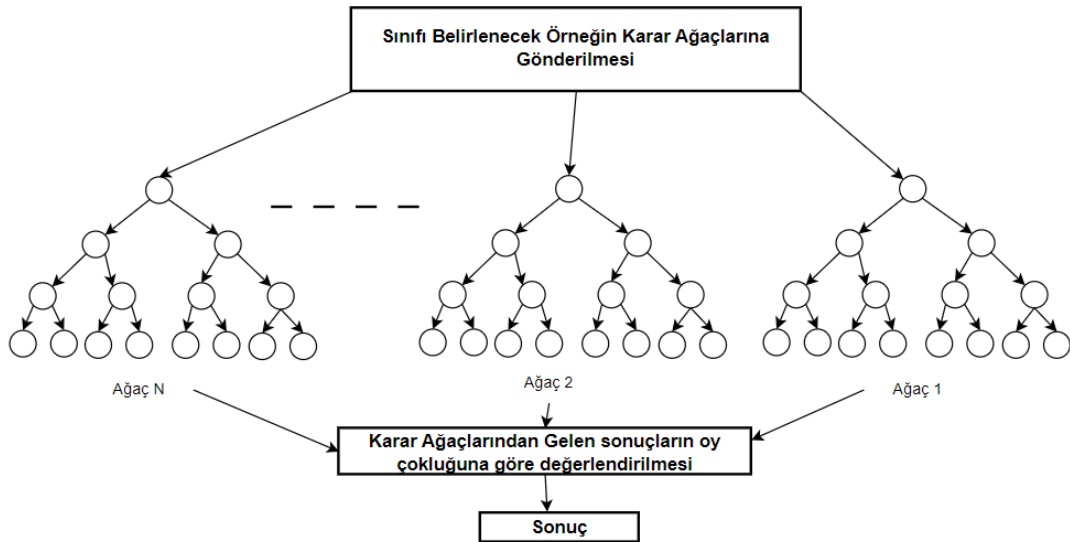
Şekil 1. Metin ön işleme adımları.

2.3.2 Lojistik Regresyon

Lojistik Regresyon algoritması, denetimli öğrenme tabanlı bir sınıflandırma yöntemidir. Genellikle iki sınıf içeren problemler için kullanılır. Ancak çok sınıflı problemlerde kullanılmak üzere geliştirilmiş çoklu LR yöntemleri de bulunmaktadır. Çoklu LR algoritması, hedef verinin hangi sınıfa ait olduğunu olasılıksal bir sonuçla tahmin eder. Hedef verinin en yüksek olasılığa sahip sınıfa ait olduğunu varsayar (Asan vd. 2023). LR olasılık formülü Denklem 1’de gösterilmiştir.

$$P(x) = \frac{1}{1 + e^{-(x-\mu)/s}} \quad (1)$$

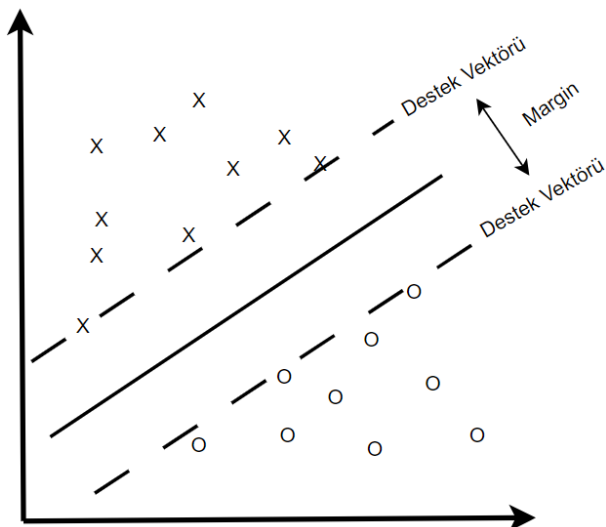
Deneysel olarak kullanılan LR modelinin parametreleri düzenli hale getirme gücünün tersi olan ve C ile sembolize eden parametre değeri 1.0’dır. Optimizasyon problemi için kullanılan çözücü algoritma ise lbfgs’dır.



Şekil 2. Rastgele Orman Yöntemi Çalışma Prensipleri.

2.3.3 Destek Vektör Makineleri

Destek Vektör Makineleri algoritması, denetimli öğrenme yöntemlerinden biridir ve hem regresyon hem de sınıflandırma problemleri için kullanılabilir. DVM, verileri analitik bir düzlemdeki dağılımlarına göre sınıflandırır ve bu sınıfları ayıran en uygun karar sınırlarını belirlemeye odaklanır. Bu sınırlar, veri noktalarının en yakın olduğu destek vektörlerine dayanır. DVM, veri noktalarını sınıflara ayırmak için bu destek vektörlerini kullanarak çalışır. DVM'nin birincil amacı, veri noktalarını farklı sınıflara en iyi şekilde ayıran bir karar sınırı belirlemek ve böylece etiketlenmemiş veri noktalarının doğru sınıflandırılmasını sağlamaktır (Yücesoy 2020). DVM sınıflar arası vektör yerleşimi Şekil 3'te gösterilmiştir.



Şekil 3. Destek Vektör Makineleri vektör yerleşimi ve sınıflar arası ayırım.

Deneylerde kullanılan LR modelinin parametreleri düzenli hale getirme gücünün tersi olan ve C ile sembolize eden parametre değeri 1.0'dır. Sınıfların ayrılması için kullanılan Kernel yöntemiyle lineerdir.

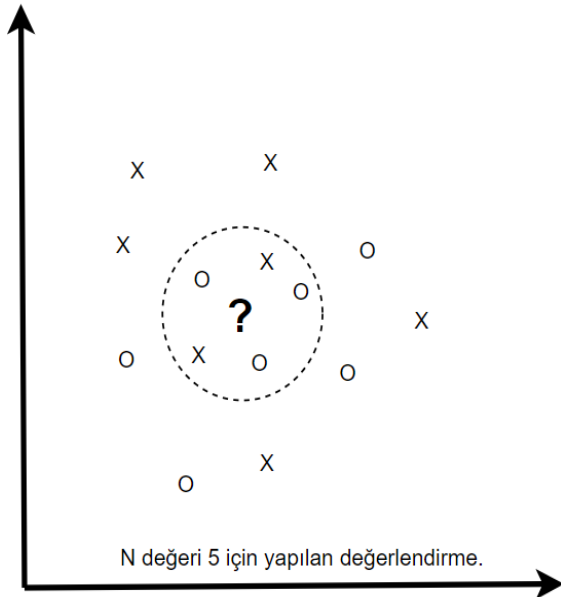
2.3.4 Gauss Sade Bayes

Sade Bayes sınıflandırıcı algoritması, verilerin sınıflar arası dağılım oranını hesaplamak için Bayes teoreminin formülünü kullanır. Bu algoritma, verilerin belirli niteliklerle birlikte hangi sınıfa ait olduğunu tahmin etmek için kullanılır. GSB algoritması, Sade Bayes algoritmasının bir alt türüdür ve benzer bir çalışma prensibine sahiptir. Bu algoritma, eğitim setindeki her sınıf için özelliklerin normalleştirilmiş verileri üzerinden ortalama değerlerini hesaplar. Sınıflandırılacak yeni bir veri noktasının, hesaplanan ortalama sınıf değerlerinden hangisine daha yakınsa o sınıfa ait olduğu varsayılır. Bu yöntem, veri noktalarının özelliklerine göre sınıflandırma yaparken sınıflar arasındaki önemli farklılıkları dikkate alır ve bu da genellikle etkili bir sınıflandırma sonucuna yol açar (Bezek Güre 2024). Denklem 2'de GSB formülü gösterilmiştir.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (2)$$

2.3.5 K En Yakın Komşu

K-EYK, denetimli öğrenme yöntemini kullanan bir sınıflandırma algoritmasıdır. Sınıfı tahmin edilecek verinin kendisine en yakın sınıf örneklerine olan mesafesini öklid veya manhattan uzaklık formüllerini kullanarak hesaplar. K adet komşunun ölçülen mesafelerinden en yakın komşuluğa sahip sınıf tahmin edilecek verinin sınıfı olduğu varsayılır (Gül ve Kalyoncu 2020). K-EYK yöntemi için mesafe ve komşu sayıları Şekil 4'te gösterilmiştir. Deneylerde kullanılan K-EYK yöntemi için hesaplanması gereken en yakın komşu sayısı 5 olarak belirlenmiştir.



Şekil 4. K En Yakın Komşu Yöntemi Çalışma Prensibi.

Zayıf Öğrenici 1		
x	x	x
x	y	y
y	y	y

Zayıf Öğrenici 2		
x	x	x
x	y	y
y	y	y

Güçlü Öğrenici 2		
x	x	x
x	y	y
y	y	y

Zayıf Öğrenici 3		
x	x	x
x	y	y
y	y	y

Şekil 5. AdaBoost Yöntemi Çalışma Prensibi.

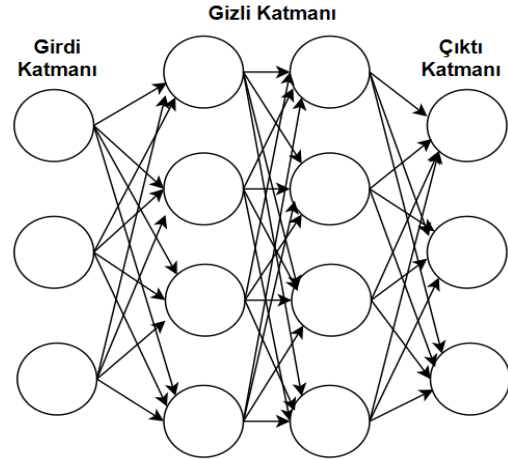
2.3.6 AdaBoost

Adaboost algoritması, belirli sayıda oluşturulan zayıf öğrencilerden alınan sonuçlar üzerine kurulu bir sınıflandırıcıdır. Zayıf öğrencilerin sonuçlarında iteratif olarak kat sayıları değiştirilir ve daha başarılı sonuçlar elde edilir. Belirli bir iterasyon sayısından sonra nihai sonuca ulaşılmaktadır. Adaboost algoritması, ezbere (aşırı uyum) eğilimli bir algoritmadır (Sağbaş 2023). Çalışma prensibi Şekil 5'te gösterilmiştir.

Deneylerde kullanılan AdaBoost yöntemi için kullanılan parametreler güçlendirme algoritması için SAMME ve her iterasyon için rastgele durum sayısı (random state) 42 olarak kullanılmıştır.

2.3.7 Yapay Sinir Ağları

YSA, denetimli bir öğrenme yöntemi kullanmaktadır. Belirli sayıda iterasyonla kendini eğiterek sonuç üretir. 3 katmandan oluşmaktadır. Bu katmanlar giriş katmanı, gizli katman ve çıkış katmanıdır. Giriş katmanı, girdi olarak kullanılan verinin modellendiği nöronlardan oluşmaktadır. Gizli katman sayısı problem türüne göre değişebilen ağırlık değişimlerinin yapıldığı nöronların bulunduğu bölümdür. Çıkış katmanı ise hedef sınıfların benzerlik oranlarını veren katmandır (Subası ve Erdem 2021). YSA katmanları Şekil 6'da gösterilmiştir.



Şekil 6. Yapay Sinir Ağı katmanları.

Deneylerde kullanılan YSA için aktivasyon fonksiyonu relu'dur. Kullanılan gizli katmanın boyutu 10'a 10 olarak tasarlanmıştır.

2.4 Büyük Dil Modelleri

BDM'ler, genellikle derin öğrenme tekniklerine dayalı büyük ölçekli YSA'lardır. Bu modeller, milyonlarca veya milyarlarca metin örneğiyle eğitilirler. Eğitim sürecinde, geniş metin veri setlerini kullanarak dilin yapısını, bağlamsal ilişkileri ve kelime kullanımını öğrenirler. Metinlerin anlamını anlama, cümle oluşturma, çeviri gibi çeşitli doğal dil işleme görevlerinde oldukça etkilidirler. Ancak BDM'ler doğal bir dilin esnek yapısından dolayı her zaman doğru bilgileri verememektedir. Bu problemin üstesinden gelebilmek ve modelin istenilen çıktıları daha etkili bir şekilde üretmesini sağlamak için istem mühendisliği tekniklerinden faydalanılmaktadır. Örneğin, bir metin sınıflandırma görevi için özel bir giriş veya talimat oluşturarak, modelin belirli bir metni daha doğru bir şekilde sınıflandırması sağlanabilmektedir. Bu çalışmada, istem mühendisliği tekniği olarak sıfır atış

istemi kullanılmıştır (Bulat ve Tzimiropoulos 2024). Sıfır atış yöntemi, bir dil modeline daha önce öğretilmemiş bir görevi veya konuyu tanımlamadan önce, bu görevi veya konuyu doğru bir şekilde anlamasını ve ona ilişkin bilgi üretmesini sağlayan bir yöntemdir. Model, belirli bir talimat veya giriş olmadan, genel dil anlayışı ve eğitimiyle, yeni bir görevi çözmek veya bilgi sağlamak için kullanılabilir. Sonuç olarak, modelin esnekliği ve genel dil anlama yeteneği kabiliyeti gelişmiş olur.

2.4.1 BERT

BERT, BDM'lerin ilk örneklerinden birisidir. Temel olarak metin üretmek için geliştirilen model farklı problemlerin çözümünde de kullanılmaktadır. BERT, yapısı gereği farklı doğal dil işleme problemlerinde kullanılacağı zaman probleme özel bir şekilde yapılandırılmalıdır. Bu çalışmada kullanılan BERT modeli duygu analizi yapan ön eğitilmiş bir yapıya sahiptir (Sousa vd. 2019).

2.4.2 GPT 3.5 Turbo

GPT 3.5 Turbo, 14 Kasım 2023 yılında OpenAI firması tarafından yayınlanan BDM'dir. Bünyesinde 500 milyardan fazla parametre bulundurmaktadır. Farklı alanlarda kullanılabilen dil modeli özellikle doğal dil işleme problemlerinde oldukça başarılıdır. GPT 3.5 Turbo'nun, son bilgi güncellemesi Ocak 2022 yılında gerçekleşmiştir (Shahin vd. 2024).

2.4.2 Gemini

Gemini, 2024 yılında Google AI tarafından piyasaya sürülmüştür. Google AI tarafından daha önce piyasaya sürülen BARD ve LaMDA dil modellerinin birleşiminden oluşmasından dolayı çoklu dil modeli olarak da tanımlanmaktadır. Çoklu model özelliğinden dolayı Gemini metin, görüntü, ses ve video gibi farklı türdeki giriş değerlerini de kullanabilmektedir (Khurdula vd. 2024).

3. Bulgular ve Tartışma

Yapılan deneylerde ilk olarak hava yolu veri seti kullanılarak her yöntem için 47 adet model eğitilmiştir. Modeller pozitif, negatif ve nötr olmak üzere üç sınıflı tahminler üretmektedir. Başlangıç olarak 150 adet veri olan bir eğitim seti kullanılmıştır.

Finansal yorumlar veri seti kullanılarak her yöntem için 16 adet model eğitilmiştir. Modeller pozitif, negatif ve nötr olmak üzere üç sınıflı tahminler üretmektedir. Başlangıç olarak 150 adet veri olan bir eğitim seti kullanılmıştır.

Film yorumları veri seti kullanılarak her yöntem için 30 adet model eğitilmiştir. Modeller pozitif ve negatif olmak üzere iki sınıflı tahminler üretmektedir. Başlangıç olarak 100 adet veri olan bir eğitim seti kullanılmıştır.

Her model eğitimi için sınıf başı 50 adet veri eklenerek eğitim seti artırılmış ve her model bu yeni veri setiyle

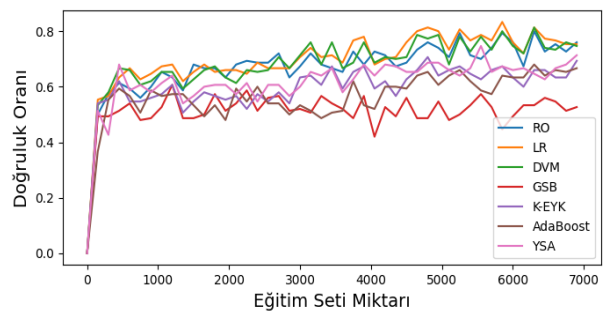
yeniden eğitilmiştir. Modellerdeki sınıf sayılarının farklı seçilmesindeki temel sebep sınıf sayılarının modeller üzerindeki etkisini ölçmektir. Oluşturulan modellerin aynı veri seti üzerinde test edilebilmesi için test seti sabit tutulmuştur. Model eğitimlerinde kullanılan test setlerindeki veri miktarı, finans ve hava yolları yorumları test setleri için 150 adet film yorumları test seti içinse 100 adet veriye sahiptir.

Modellerin performanslarının değerlendirilmesinde doğruluk oranı kullanılmıştır. Veri setlerindeki sınıflar arası veri miktarının eşit olarak yapılandırılması performans değerlendirilmesinde doğruluk oranını verimli hale getirmektedir. Doğruluk oranı karmaşıklık matrisi tabanlı bir değerlendirme yapmaktadır. Karmaşıklık matrisi, modelin Doğru pozitif (TP), Doğru negatif (TN), Yanlış pozitif (FP) ve Yanlış negatif (FN) cevaplarıyla oluşturulmaktadır. Denklem 3'te doğruluk oranı formülü verilmiştir.

$$\text{Doğruluk oranı} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

3.1 Klasik Yöntem Sonuçları

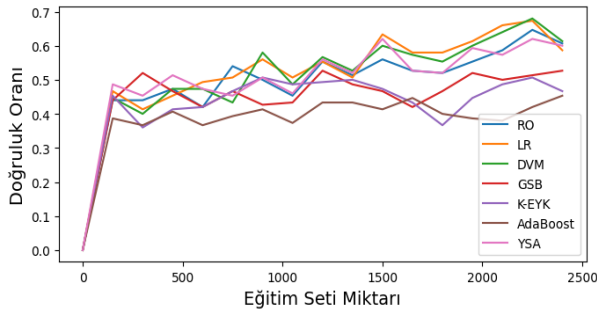
Klasik yöntemlerle gerçekleştirilen deneyin ilk adımında Hava yolu şirketlerinin hizmetlerine yapılan yorumlar için 150 eğitim seti ve 150 test seti kullanılarak modeller eğitilmiştir. Bu eğitimin sonunda AdaBoost algoritması 0.36 doğruluk oranıyla en düşük başarıyı elde ederken, 0.55 doğruluk oranıyla en yüksek başarıyı LR algoritması elde etmiştir. En yüksek doğruluk oranına ise 5850 eğitim ve 150 test seti içeren adımda ulaşılmıştır. Bu adımda, 0.83 doğruluk oranıyla en yüksek performansı LR yöntemi elde etmiştir. Şekil 7' de deneyin her bir adımı için doğruluk oranları gösterilmiştir.



Şekil 7. Hava yolu şirketlerinin hizmetlerine yapılan yorumlar için yapılan üç sınıflı duygu analizi sonuçları.

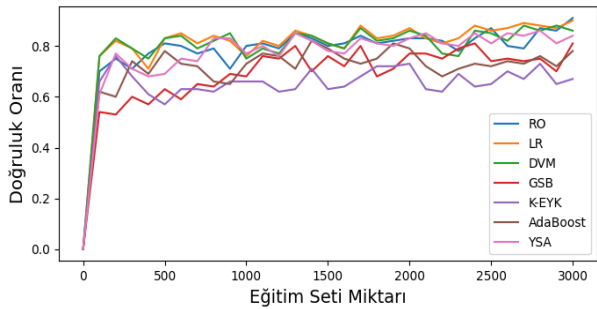
Deneyin diğer bölümünde finansal gelişmelere yapılan yorumlar için 150 eğitim seti ve 150 test seti kullanılarak modeller eğitilmiştir. Bu eğitimin sonunda AdaBoost algoritması 0.38 doğruluk oranıyla en düşük başarıyı yakalarken, 0.48 doğruluk oranıyla en yüksek başarıyı YSA algoritması elde etmiştir. En yüksek doğruluk oranına ise 2250 eğitim ve 150 test seti kullanılan adımda ulaşılmıştır. Bu adımda 0.68 doğruluk oranıyla en yüksek performansı

DVM yöntemi elde etmiştir. Şekil 8’ de deneyin her bir adımı için doğruluk oranları gösterilmiştir.



Şekil 8. Finansal gelişmelere yapılan yorumlar için üç sınıflı duygu analizi sonuçları.

Deneyin son bölümünde filmlere yapılan yorumlar için 100 eğitim seti ve 100 test seti kullanılarak modeller eğitilmiştir. Bu eğitimin sonunda YSA algoritması 0.61 doğruluk oranıyla en düşük başarıyı yakalarken, 0.76 doğruluk oranıyla en yüksek başarıyı LR ve DVM algoritmaları elde etmiştir. En yüksek doğruluk oranına ise 3000 eğitim ve 100 test seti kullanılan adımda ulaşılmıştır. Bu adımda 0.91 doğruluk oranıyla en yüksek performansı RO yöntemi göstermiştir. Şekil 9’ da deneyin her bir adımı için doğruluk oranları gösterilmiştir.



Şekil 9. Filmlere yapılan yorumlar için iki sınıflı duygu analizi sonuçları.

3.2 Büyük Dil Modelleri Sonuçları

BDM’lerle yapılan deneylerde klasik yöntemlerde kullanılan test seti üzerinde deneyler yapılmıştır. Sıfır atış tekniği kullanılarak yapılan duygu analizinde Gemini modelinin tüm veri setlerinde en yüksek performansı gösterdiği görülmüştür. BERT modeli performans açısından en düşük başarıyı sergilemiştir. Çizelge 2’de modellerin veri setlerine göre doğruluk oranları gösterilmiştir.

Çizelge 2. BDM’lerin üç farklı veri setindeki duygu analizi sonuçları.

Model	Hava yolu şirketlerine yapılan yorumlar	Finansal gelişmelere yapılan yorumlar	Film yorumları
BERT	0.6733	0.5266	0.69
GPT 3.5 Turbo	0.76	0.5933	0.96
Gemini	0.76	0.7733	0.98

3.2 Karşılaştırmalı Model Sonuçları

BDM ve klasik yöntemler kendilerine has gereksinimlere sahiptirler. BDM’ler geliştirilirken yüksek boyutlu verilerden faydalanırlar ve üretilen model birçok farklı problemde kullanılabilir. Klasik yöntemlerde modeller probleme uygun veri setleriyle eğitilir ve dil yapısı çoğunlukla dikkate alınmaz. Yapılan deneyler sonucunda klasik yöntemler bazı veri setlerinde ne kadar çok veri kullanılsa da BDM başarısını yakalayamamıştır. Başarıları eşit olan durumlardaysa ise probleme uygun yüksek bir veri ihtiyacı olmaktadır. Çizelge 3’te veri miktarına göre en başarılı sonuçları veren BDM’ler ve klasik makine öğrenmesi modellerinin karşılaştırılması gösterilmiştir.

Çizelge 3. Veri miktarına göre BDM ve Klasik model başarı karşılaştırılması.

BDM Modeli	Klasik Model	Veri Seti	Veri Sayısı
BERT	LR	Hava Yolları	1050
GPT 3.5 Turbo	DVM	Hava Yolları	3150
GEMINI	DVM	Hava Yolları	3150
BERT	RO, LR, DVM	Film	100
GPT 3.5 Turbo	-	Film	-
GEMINI	-	Film	-
BERT	LR, DVM	Finansal	900
GPT 3.5 Turbo	LR,DVM,YSA	Finansal	1500
GEMINI	-	Finansal	-

Çizelge 3’te BDM Modeli alanında deneylerde kullanılan BDM’ler gösterilirken Veri seti kısmında deneyde kullanılan veri setini göstermektedir. Veri sayısı kısmında ise klasik yöntemlerde kullanılan modellerin başarısının BDM modelini geçtiği noktada kaç adet eğitim verisi kullanıldığını göstermektedir. Model kısmındaysa BDM sonuçlarını geçebilen klasik yöntem modeli varsa listelenmektedir.

Hava yolları yorumları veri seti için karşılaştırmalı sonuçlar incelendiğinde LR yöntemi 1050 veriyle eğitildiğinde BERT modelinin başarısına ulaşırken DVM yöntemi 3150 veriyle eğitildiğinde GPT 3.5 Turbo ve Gemini modellerinin başarısına ulaşmaktadır.

Finansal yorumlar veri seti için karşılaştırmalı sonuçlar incelendiğinde LR ve DVM yöntemleri 900 veriyle eğitildiğinde BERT modelinin başarısına ulaşırken LR, DVM ve YSA yöntemleri 1500 veriyle eğitildiğinde GPT 3.5 Turbo modelinin başarısına ulaşmaktadır. Klasik yöntemler çalışmada kullanılan tüm eğitim seti verisi kullanılmasına rağmen Gemini modelinin başarısına ulaşamamaktadır.

Film yorumları veri seti için karşılaştırmalı sonuçlar incelendiğinde RO, LR ve DVM yöntemleri 100 veriyle eğitildiğinde BERT modelinin başarısına ulaşırken hiçbir klasik yöntem GPT 3.5 Turbo ve Gemini başarısına ulaşamamıştır.

BDM ve klasik yöntemler beraber değerlendirildiğindeyse modeller Finans yorumları veri setinde ortak bir başarısızlığa sahiptir. Bu başarısızlığın temel nedeni olarak finansal terimlerin klasik duygu analizi yapısından farklı olmasıdır. Film yorumları veri setinde olan yüksek başarı oranıysa veri setindeki yorum metinlerinin detaylı olması ve veri setinin iki sınıfa sahip olması olarak gözlemlenmiştir.

4. Sonuçlar

BDM'ler sıfır atış, birkaç atış istem yöntemleri ile genelleme yapma ve yeni görevleri yerine getirme yeteneği kazanabilmektedirler. Yalnızca dil modelleme hedefleri üzerine eğitilmiş olmalarına rağmen, bu modeller, bir soruyu yanıtlamak veya özetleme yapmak gibi eğitim almadıkları yeni görevlerde nispeten iyi performans gösterebilirler.

Bu çalışmada, BERT, GPT, Gemini gibi BDM'lere sıfır atış yöntemi uygulandığında, herhangi bir eğitim süreci olmadan duygu analizi görevini ne ölçüde yerine getirebildiği incelenmiştir. BDM'lerin metinlerdeki duyguları negatif, pozitif ya da nötr olarak sınıflandırabildikleri ve performanslarının da klasik denetimli makine öğrenmesi yöntemlerine rakip olduğu deneylerle kanıtlanmıştır.

Makine öğrenmesi yöntemlerinde kullanılan eğitim örneklerin sayısı arttıkça performansları mümkün olan iyi sabit değere ulaştığı ve BDM'lerin ise en iyi algoritmanın performansını geçtiği gözlemlenmiştir. Sonuçlarımız, BDM'lerin bağlam içi örnekler verilmesi dahi duygu analizi gibi karmaşık görevleri yerine getirebildiklerini ortaya koymuştur. Gelecek çalışma olarak, BDM'lerin duygu analizi konusundaki hassasiyetlerini artırmak için farklı istem mühendisliği yaklaşımları denenebilir. Ayrıca, BDM'lerin duygusal bağlamları daha iyi anlayabilmesi için belirli metin türleri veya kültürel farklılıklar üzerinde daha fazla çalışma yapılabilir. Bu şekilde, BDM'lerin duygu analizi alanında daha etkili ve genel geçer sonuçlar elde etmesi sağlanabilir.

Etik Standartlar Bildirgesi

Yazarlar tüm etik standartlara uyduklarını beyan ederler.

Yazarlık Katkı Beyanı

Yazar 1: Kaynaklar, Deney, Yazma, Fikir Sahibi

Yazar 2: Kaynaklar, Araştırma, Yazma

Yazar 3: Araştırma, Deney, Yazma

Çıkar Çatışması Beyanı

Yazarların bu makalenin içeriğiyle ilgili olarak beyan edecekleri hiçbir çıkar çatışması yoktur.

Verilerin Kullanılabilirliği

Yazarlar, bu çalışmanın bulgularını destekleyen ana verilerin makale içerisinde mevcut olduğunu beyan ederler.

5. Kaynaklar

- Asan, M.E., Taşkın, H., Alemdar, M., ve Capoglu, R., 2023, Tiroit kanseri hastalık tanısında lojistik regresyon kullanımı. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*.
<https://doi.org/10.17341/gazimmfd.1253193>
- Bezek Güre, Ö., 2024, Classification of Liver Disorders Diagnosis using Naïve Bayes Method. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, **13**, 153–160.
<https://doi.org/10.17798/bitlisfen.1361016>
- Bulat, A. ve Tzimiropoulos, G., 2024, Language-Aware Soft Prompting: Text-to-Text Optimization for Few- ve Zero-Shot Adaptation of V & L Models. *International Journal of Computer Vision*, **132**, 1108–1125.
<https://doi.org/10.1007/s11263-023-01904-9>
- Gül, E. ve Kalyoncu, M., 2020, Ağır Vasıta Hava Kompresörü Piston Segmanı Aşınması Durumlarında K-En Yakın Komşu Algoritmasının Sınıflandırma Performansının İncelenmesi. *European Journal of Science ve Technology*.
<https://doi.org/10.31590/ejosat.802958>
- Hasan, A., Moin, S., Karim, A., ve Shamshirbve, S., 2018, Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical ve Computational Applications*, **23**, 11.
<https://doi.org/10.3390/mca23010011>
- Jouis, C. ve Shafei, B., 2018, Big textual data. In *Proceedings of the 10th International Conference on Management of Digital EcoSystems* ((New York, NY, USA: ACM)), pp. 7–12.
- Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., ve Zhang, W., 2018, The Rise of Artificial Intelligence under the Lens of Sustainability. *Technologies*, **6**, 100.
<https://doi.org/10.3390/technologies6040100>
- Khurana, D., Koli, A., Khatte, K., ve Singh, S., 2023, Natural language processing: state of the art, current trends ve challenges. *Multimedia Tools ve Applications*, **82**, 3713–3744.
<https://doi.org/10.1007/s11042-022-13428-4>
- Khurdula, H.V., Pagutharivu, A., ve Soung Yoo, J., 2024, The Future of Feelings: Leveraging Bi-LSTM, BERT with Attention, Palm V2 & Gemini Pro for Advanced Text-Based Emotion Detection. In *SoutheastCon 2024* ((IEEE)), pp. 275–278.
- Kılınçarslan, Ş., Şimşek Türker, Y., ve İNCE, M., 2020, Prediction of heat-treated cedar wood swelling ve shrinkage values with artificial neural networks ve rveom forest algorithm. *Mühendislik Bilimleri ve Tasarım Dergisi*, **8**, 200–205.
<https://doi.org/10.21923/jesd.825442>
- Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., ve Potts, C., 2011, Learning Word Vectors for

- Sentiment Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* ((Portlve, Oregon, USA: Association for Computational Linguistics)), pp. 142–150.
- Malo, P., Sinha, A., Korhonen, P., Wallenius, J., ve Takala, P., 2014, Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science ve Technology*, **65**, 782–796.
<https://doi.org/10.1002/asi.23062>
- Rane, A. ve Kumar, A., 2018, Sentiment Classification System of Twitter Data for US Airline Service Analysis. In *2018 IEEE 42nd Annual Computer Software ve Applications Conference (COMPSAC)* (IEEE), 769–773.
- Sağbaşı, E.A., 2023, MFCC Öznitelikleri ve Adaboost Topluluk Öğrenme Yöntemi Kullanılarak Uyku Seslerinin Sınıflandırılması. *Computer Science*. Cilt: IDAP-2023 : International Artificial Intelligence and Data Processing Symposium Sayı: IDAP-2023, 31 - 36
<https://doi.org/10.53070/bbd.1347221>
- Shahin, M., Chen, F.F., Hosseinzadeh, A., Maghanaki, M., ve Eghbalian, A., 2024, A novel approach to voice of customer extraction using GPT-3.5 Turbo: linking advanced NLP ve Lean Six Sigma 4.0. *The International Journal of Advanced Manufacturing Technology*, **131**, 3615–3630.
<https://doi.org/10.1007/s00170-024-13167-w>
- Sousa, M.G., Sakiyama, K., Rodrigues, L. de S., Moraes, P.H., Fernvees, E.R., ve Matsubara, E.T., 2019, BERT for Stock Market Sentiment Analysis. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* ((IEEE)), pp. 1597–1601.
- Subaşı, A. ve Erdem, K., 2021, Hibrit nanoakışkanların özgül ısılarının yapay sinir ağları ile tahmin edilmesi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, **37**, 377–388.
<https://doi.org/10.17341/gazimmfd.880340>
- Yücesoy, E., 2020, Konuşmacının Yaş ve Cinsiyetine Göre Sınıflandırılmasında DVM Çekirdeğinin Etkisi. *El-Cezeri Fen ve Mühendislik Dergisi*.
<https://doi.org/10.31202/ecjse.707179>
- Yue, L., Chen, W., Li, X., Zuo, W., ve Yin, M., 2019, A survey of sentiment analysis in social media. *Knowledge ve Information Systems*, **60**, 617–663.
<https://doi.org/10.1007/s10115-018-1236-4>