T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

DERİN ÖĞRENME YÖNTEMİ İLE METİNSEL İFADELERDE DUYGU ANALİZİ

Nuray YILDIZ

Ağustos-2024
BATMAN

T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

DERİN ÖĞRENME YÖNTEMİ İLE METİNSEL İFADELERDE DUYGU ANALİZİ

Nuray YILDIZ

Danışman
Doç. Dr. Yılmaz KAYA

Ağustos-2024
BATMAN

TEZ KABUL VE ONAYI

Nuray Yıldız tarafından hazırlanan "Derin Öğrenme Yöntemi ile Metinsel İfadelerde Duygu Analizi" adlı tez çalışması 09/08/2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgi Teknolojileri Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Prof. Dr. Necmettin SEZGİN	**……..**
Danışman Doç. Dr. Yılmaz KAYA	……
Üye Doç. Dr. Melih KUNCAN	…..……
Yedek Üyeler	
Üye Dr. Öğr. Üyesi Abdülkerim ÖZTEKİN	……
Üye Doç. Dr. Funda KUNCAN	…….

Yukarıdaki sonucu onaylarım.

Dr. Öğr. Üyesi Ömer Murat ÖTER
Lisansüstü Eğitim Enstitüsü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Nuray YILDIZ

09 /08/2024

ÖZET

YÜKSEK LİSANS TEZİ

DERİN ÖĞRENME YÖNTEMİ İLE METİNSEL İFADELERDE DUYGU ANALİZİ

Nuray YILDIZ

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü
Bilgi Teknolojileri Anabilim Dalı

Danışman: Doç. Dr. Yılmaz KAYA

2024, 70 Sayfa

Jüri
Prof. Dr. Necmettin SEZGİN
Doç. Dr. Yılmaz KAYA
Doç. Dr. Melih KUNCAN
Dr. Öğr. Üyesi Abdülkerim ÖZTEKİN
Doç. Dr. Funda KUNCAN

Bu çalışmada, Türkçe metinlerde kullanıcı yorumlarının duygu analizini gerçekleştirmek amacıyla makine öğrenmesi ve derin öğrenme yöntemlerinin etkinliği incelenmiştir. Beyaz eşya ürünlerine ait kullanıcı yorumları, Trendyol ve Hepsiburada gibi popüler e-ticaret platformlarından toplanmış ve analiz edilmiştir.

Kelime bulutu yöntemi kullanılarak, yorumlardaki en sık geçen kelimeler görselleştirilmiştir. Olumlu yorumlarda "yüksek performans", "iyi fiyat", "memnun kaldım", "kurulumu kolay" gibi ifadeler öne çıkarken, olumsuz yorumlarda "gecikme", "kötü koku", "kalite kontrol eksikliği", "servis problemi" gibi ifadeler dikkat çekmiştir. Bu analizler, kullanıcıların genel memnuniyet düzeylerini ve karşılaştıkları sorunları görselleştirerek, işletmelere ürün ve hizmetlerini iyileştirmeleri konusunda içgörüler sunmaktadır.

Makine öğrenmesi yöntemleri arasında SVM, KNN ve Naive Bayes (NB) algoritmaları kullanılarak sınıflandırma işlemleri gerçekleştirilmiştir. %80-20 eğitim-test oranları ile NB modeli %94.02 başarı oranı ile en yüksek performansı sergilemiştir. SVM modeli %93.42, KNN modeli ise %86.75 başarı oranı elde etmiştir. NB modeli, kesinlik, hatırlama ve F1 ölçütlerinde de yüksek performans göstermiştir.

Derin öğrenme yöntemleri olarak LSTM ve 1D-CNN kullanılmış ve farklı eğitim-test oranları ile başarı oranları incelenmiştir. LSTM modeli %92.42, 1D-CNN modeli ise %91.42 başarı oranı ile iyi performanslar sergilemiştir.

Sonuç olarak, Naive Bayes modeli en yüksek başarı oranını elde ederek diğer modellere kıyasla en iyi performansı göstermiştir. Derin öğrenme yöntemlerinden LSTM ve 1D-CNN de yüksek başarı oranları ile dikkate değer performanslar sunmuştur. Bu çalışma, e-ticaret sektöründe kullanıcı yorumlarının duygu analizinde makine öğrenmesi ve derin öğrenme yöntemlerinin etkili olduğunu ortaya koymaktadır. Bu tür analizler, işletmelere müşteri geri bildirimlerini daha iyi değerlendirme ve müşteri memnuniyetini artırma konusunda önemli avantajlar sağlayabilir.

Anahtar Kelimeler: Kelime Çantası, Makine Öğrenmesi, Metin madenciliği, TF-IDF

ABSTRACT

MS THESIS

DEEP LEARNİNG FOR SENTİMENT ANALYSİS İN TEXTUAL EXPRESSİONS

Nuray YILDIZ

Batman University Graduate Education Institute

The Degree of Master of Science in Information Technologies

Advisor: Assoc. Prof. Yılmaz KAYA

2024, 70 Pages

Jury
Prof. Dr. Necmettin SEZGİN
Assoc. Prof. Dr. Yılmaz KAYA
Assoc. Prof. Dr. Melih KUNCAN
Dr. Lecturer Abdülkerim ÖZTEKİN
Assoc. Prof. Dr. Funda KUNCAN

In this study, the effectiveness of machine learning and deep learning methods for sentiment analysis in Turkish texts has been investigated. User comments on white goods products were collected from popular e-commerce platforms such as Trendyol and Hepsiburada and analyzed.

Using the word cloud method, the most frequently occurring words in the comments were visualized. Positive comments highlighted expressions like "high performance," "good price," "satisfied," and "easy installation," while negative comments included phrases such as "delay," "bad odor," "lack of quality control," and "service problem." These analyses visually represent the overall satisfaction levels of users and the problems they encounter, providing insights for businesses to improve their products and services.

Classification processes were carried out using machine learning methods such as SVM, KNN, and Naive Bayes (NB). With an 80-20 training-test ratio, the NB model exhibited the highest performance with a success rate of 94.02%. The SVM model achieved a success rate of 93.42%, while the KNN model achieved 86.75%. The NB model also demonstrated high performance in precision, recall, and F1-score metrics.

Deep learning methods such as LSTM and 1D-CNN were utilized, and success rates were examined with different training-test ratios. The LSTM model achieved a success rate of 92.42%, while the 1D-CNN model achieved 91.42%.

In conclusion, the Naive Bayes model outperformed other models with the highest success rate. Deep learning methods, specifically LSTM and 1D-CNN, also demonstrated notable performance with high success rates. This study illustrates the effectiveness of machine learning and deep learning methods for sentiment analysis in user comments within the e-commerce sector. Such analyses offer significant advantages for businesses in better evaluating customer feedback and enhancing customer satisfaction.

Keywords: Bag of Words, Machine Learning, Text Mining, TF-IDF

ÖNSÖZ

Bu tezin hazırlanması süreci, benim için oldukça heyecan verici ve öğretici bir yolculuk oldu. "Derin Öğrenme Yöntemi ile Metinsel İfadelerde Duygu Analizi" konusundaki bu çalışmamda bana destek olan ve her zaman yanımda olan kişilere teşekkür etmek istiyorum.

Öncelikle, bu çalışmanın her aşamasında bana rehberlik eden ve her türlü desteği sağlayan danışman hocam Doç. Dr. Yılmaz Kaya'ya en içten teşekkürlerimi sunarım. Kendisinin bilgi ve deneyimleri, bu çalışmanın şekillenmesinde büyük rol oynadı.

Bu süreçte her zaman yanımda olan sevgili eşim Deniz Yıldız'a da teşekkürlerimi iletmek istiyorum. Onun sabrı, anlayışı ve sürekli desteği olmadan bu tez asla tamamlanamazdı. Ayrıca, hayatımın en değerli varlıkları olan kızlarım Elif Mila ve Hiva'ya da teşekkür ederim. Onların neşesi ve sevgisi, bana her zaman güç verdi.

Nuray YILDIZ
BATMAN-2024

İÇİNDEKİLER

SİMGELER VE KISALTMALAR

Bu çalışmada yer alan bazı simgeler ve kısaltmaların açıklamaları aşağıda verilmiştir.

Simgeler	Açıklama
1D-CNN	1-Boyutlu Evrişimli Sinir Ağı, 1-Dimensional Convolutional Neural Network
YZ	Yapay Zeka, Artificial Intelligence
DA	Duygu Analizi, Sentiment Analysis
DDI	Doğal Dil İşleme, Natural Language Processing
FN	Yanlış Negatif , False Negative
FP	Yanlış Pozitif, False Positive
KNN	K-En Yakın Komşular, K-Nearest Neighbors
LSTM	Uzun Kısa Süreli Bellek, Long Short-Term Memory
NB	Naive Bayes
ReLU	Doğrultulmuş Doğrusal Birim, Rectified Linear Unit
RNN	Tekrarlayan Sinir Ağı, Recurrent Neural Network
SVM	Destek Vektör Makinesi, Support Vector Machine
TF-IDF Frequency	Terim Frekansı-Ters Belge Frekansı, Term Frequency-Inverse Document
TN	Gerçek Negatif , True Negative
TP	Gerçek Pozitif , True Positive
YSA	Yapay Sinir Ağı ,Artificial Neural Network

ŞEKİL DİZİNİ

ÇİZELGE LİSTESİ

1.GİRİŞ

Günümüzde internetin hızla gelişmesi, pek çok alanda köklü değişimlere sebep olmuştur. Özellikle ticaret dünyasında, bu değişimlerin etkileri oldukça belirgindir. İnternetin yaygınlaşması ve erişimin kolaylaşmasıyla birlikte, geleneksel mağaza alışverişlerine alternatif olarak e-ticaret siteleri önemli bir konum kazanmıştır. İnternet, E-ticaretin büyümesinde dünya çapında önemli bir rol oynamıştır. Artan internet kullanımı, bir ülkenin E-ticaret pazarının genişlemesine katkıda bulunmuştur. E-ticaret platformlarının yükselişi, ürünlerin ve hizmetin geliştirilmesi gerekliliğini beraberinde getirmiştir. E-ticaret siteleri, müşterilere geniş bir ürün yelpazesi sunarak alışverişi daha erişilebilir hale getirmiştir. E-ticaretin yükselişi, insanlara alışverişi fiziksel mağazalara gitmek zorunluluğu olmadan online olarak yapma imkanı sunmuştur (Aslan,2023).

E-ticaret siteleri, insanların alışveriş yapma alışkanlıklarını ve ticaret anlayışını kökünden değiştirmiştir. Artık insanlar, ürünleri incelemek, karşılaştırmak ve satın almak için fiziksel mağazalara gitmek zorunda değillerdir. Birkaç tıklama ile istedikleri ürünleri bulabilir, fiyatları karşılaştırabilir ve hatta dünyanın herhangi bir yerinden ürün satın alabilirler. Bu değişim, işletmeler için de büyük fırsatlar sunmaktadır. Küçük ve orta ölçekli işletmeler, geleneksel mağazalarda sahip olmadıkları geniş kitlelere ulaşma imkanı bulmuşlardır. E-ticaret siteleri, işletmelere düşük maliyetlerle geniş bir pazar alanı sunar ve böylece küresel ölçekte rekabet edebilmelerini sağlar. Günümüzde, tüketiciler sık sık satın aldıkları veya tükettikleri ürün ve hizmetler hakkında çeşitli platformlara olumlu veya olumsuz yorumlar yazmaktadırlar. Bu yorumlar, bir e-ticaret şirketi için online itibar yönetimi gibi konuları değerlendirirken, diğer kullanıcılar için önemli avantajlar sunabilir. Kullanıcılar, bir ürün veya hizmet satın alırken ilgili yorumları gözden geçirerek kendi kararlarını şekillendirme ve bu yorumları referans olarak kullanma eğilimindedirler. Yorumların sayısı her geçen gün artmakta ve dolayısıyla bu yorumların otomatik olarak analiz edilmesi gerekmektedir. Bu nedenle, günümüzde duygu analizi, bu tür yorumların analiz edilmesinde sıkça kullanılan bir yöntem olarak öne çıkmaktadır (Aytekin ve Bayram,2021). Ayrıca, internetin sunduğu veri analizi ve pazarlama araçları sayesinde, işletmeler müşterileriyle daha etkili bir şekilde etkileşime geçebilirler. Müşteri tercihlerini ve davranışlarını anlamak, kişiselleştirilmiş teklifler sunmak ve müşteri sadakatini artırmak için e-ticaret siteleri vazgeçilmez bir araç haline gelmiştir. Duygu analizi, internet üzerinde yaygın olarak kullanılan bir veri toplama ve analiz yöntemidir, özellikle web sitelerinden elde edilen

veriler üzerinde sıkça uygulanır. Veri madenciliği, günümüzde en aktif araştırma alanlarından biridir ve duygu analizi bu alanda önemli bir yer tutar, çünkü duygusal verilerin analizi giderek daha önemli hale gelmektedir. Bu analiz yöntemi, sadece bilgisayar bilimleri alanında değil, aynı zamanda sosyal bilimlerde de sıkça kullanılmaktadır, çünkü toplumsal konuların incelenmesinde ve iş dünyasında da değerli öngörü sunar. Özellikle tüketicilerin ürün ve hizmetlerle ilgili duygusal tepkilerini anlamak, işletmelere bu alanda stratejik avantajlar sağlar ve gerektiğinde ürün veya hizmetlerde revizyonlar yapma fırsatı sunar. Duygu Analizi, tüketicilerin beklentilerini belirlemede ve ürünlerin geliştirilmesinde etkili bir araç olarak kullanılır, çünkü müşteri geri bildirimlerinin duygusal boyutunu anlamak işletmelere rekabet avantajı sağlar (Varol M.ve Varol E., 2021).

Sonuç olarak, internetin evrimi, e-ticaret sitelerinin stratejik önemini artırmıştır. Hem tüketicilere hem de işletmelere sunduğu olanaklar ve sağladığı kolaylık, çağdaş ticaretin temel bir unsuru olarak kabul edilmektedir. Bununla birlikte, sektördeki hızlı değişim ve rekabet, işletmelerin sürekli olarak yenilenmesi ve büyümesi gerektiğini vurgulamaktadır.

Yapay zekâ, bilgisayarların veya bilgisayar destekli makinelerin insan benzeri nitelikleri, problemleri çözme yeteneği, anlama kapasitesi, anlam çıkarma yeteneği, genelleme yeteneği ve geçmiş deneyimlerden öğrenme gibi yüksek seviyede mantıksal işlemleri gerçekleştirme kabiliyeti olarak bilim dünyasında tanımlanmaktadır. Bu tanım, bilgisayar bilimi ve yapay zeka alanındaki araştırmacılar tarafından genellikle kabul edilen bir kavramdır ve yapay zekânın kapsamlı yönlerini ve yeteneklerini açıklamaktadır (Öztürk ve Şahin, 2018).

Doğal dil, insanların düşüncelerini ifade etmek ve iletişim kurmak için kullandığı karmaşık bir araçtır. Doğal dil işleme (DDİ) çalışmaları, metinleri anlamak için öncelikle sesli ya da yazılı metinden özellikler çıkararak ön işleme adımından geçer. Bu ön işleme adımından sonra, dilbilimsel açıdan önemli olan şekilbilim, sözdizim, anlambilim ve söylev işleme çalışmaları gerçekleştirilebilir. Ancak, bu çalışma alanları sözcük kökleri, sözcük bağlamları ve anlam açısından çeşitli zorluklarla karşılaşabilir. Bu zorlukları aşmak için geliştirilen dilbilgisine dayalı kural tabanlı DDİ çalışmaları, elle yapılan özelliklere dayanmaktadır. Ancak, elle yapılan özellikler zaman alır ve genellikle yetersiz kalabilir. Bu eksiklikleri gidermek için geliştirilen yapay zekâ yöntemlerinden biri olan derin öğrenme, yapay sinir ağı yapısı, güçlü donanımı ve büyük veri girdisi ile daha iyi sonuçlar elde edilmesini sağlamıştır. İnsanlar arasındaki

iletişim için kullanılan doğal dil, karmaşık bir araçtır ve bu dilin bilgisayarlar tarafından anlaşılması ve işlenmesi için doğal dil işleme (DDİ) çalışmaları önemli bir rol oynamaktadır. DDİ çalışmaları, öncelikle metinleri anlamak için sesli ya da yazılı metinden özellikler çıkararak bir ön işleme adımından geçmektedir. Bu adımdan sonra, şekilbilim, sözdizim, anlambilim ve söylev gibi dilbilimsel bileşenlerin işlenmesi gerçekleştirilebilir. Ancak, bu çalışma alanları bazı zorluklarla karşı karşıyadır, özellikle de sözcük kökleri, sözcük bağlamları ve anlam açısından. Bu zorlukları aşmak için geliştirilen kural tabanlı DDİ yaklaşımları, elle yapılan özelliklerden yararlanır ancak bu özelliklerin elde edilmesi zaman alıcı ve sınırlı olabilir. Bu eksiklikleri gidermek için derin öğrenme gibi yapay zekâ teknikleri kullanılabilir, bu teknikler daha iyi sonuçlar elde etmek için yapay sinir ağları yapısını, güçlü donanımı ve büyük veri girdisini kullanır. İnsanların iletişim aracı olarak kullandığı doğal dilin bilgisayarlar tarafından anlaşılması ve işlenmesi, doğal dil işleme (DDİ) adı verilen bir alanda araştırılmaktadır. DDİ çalışmaları, metin verilerini anlamak için öncelikle sesli veya yazılı metinlerden özellikler çıkararak bir ön işleme sürecinden geçer. Bu sürecin ardından, dilin şekilbilim, sözdizim, anlambilim ve söylev gibi farklı bileşenleri üzerinde çalışma gerçekleştirilir. Ancak, bu çalışma alanları bazı zorluklarla karşı karşıyadır, özellikle de kelime kökleri, kelime bağlamları ve anlam açısından. Bu zorlukları aşmak için geliştirilen kural tabanlı DDİ yaklaşımları, elle yapılan özelliklere dayanır ancak bu özelliklerin elde edilmesi zaman alabilir ve yetersiz olabilir. Bu eksiklikleri gidermek için kullanılan bir yapay zekâ yöntemi olan derin öğrenme, daha iyi sonuçlar elde etmek için yapay sinir ağı yapısını, güçlü donanımı ve büyük veri girdisini kullanır (Babüroğlu vd., 2019). Tüm bu avantajlar göz önüne alındığında, doğal dil işlemenin günümüzde ve gelecekte büyük bir öneme sahip olduğu açıktır.

Fikir madenciliği ve duygu analizi, insanların çeşitli sanal platformlarda ürünler, hizmetler, organizasyonlar, olaylar, siyasi düşünceler ve daha pek çok konu hakkında metinler aracılığıyla ifade ettikleri duyguları, fikirleri ve düşünceleri keşfetmeyi hedefler. Bu süreçte, insanların metinlerde gizli kalmış olan duygusal ifadelerini ve düşünsel derinliklerini ortaya çıkarmak amaçlanır. Duygu analizi, 2000'li yılların başından itibaren doğal dil işlemenin önemli bir alt dalı olarak kabul edilir ve yoğun bir şekilde araştırılan bir alan haline gelmiştir. Günümüzde duygu analizi çalışmaları, sadece doğal dil işlemenin sınırlarını aşmakla kalmaz, aynı zamanda veri madenciliği, web madenciliği ve metin madenciliği gibi disiplinlerle de sıkı bir ilişki içindedir. Bu çalışmalar, büyük veri setlerinden anlamlı öngörü çıkarabilmek için çeşitli

metotlar ve teknikler kullanır. Özellikle, duygu analizinin işaret ettiği duygu ifadeleri ve düşünceler, pazarlama stratejilerinden sosyal medya analizine kadar birçok alanda değerli bilgiler sağlayabilir. Dolayısıyla, fikir madenciliği ve duygu analizi, bilgi çağında önemli bir araştırma ve uygulama alanı olarak kabul edilir ve gün geçtikçe daha da önem kazanmaktadır (Özyurt ve Akçayol,2018). İnsanların düşüncelerini paylaşabilecekleri ve yorum yapabilecekleri birçok platformun ortaya çıkması, internetin gelişmesiyle birlikte gerçekleşmiştir. Sosyal medya uygulamaları, bloglar ve e-ticaret siteleri gibi her geçen gün yaygınlaşan yeni alanlar, bu büyümeyi desteklemektedir. Dolayısıyla, internet üzerindeki verilerin kullanımıyla ilgili çeşitli alanlar ortaya çıkmıştır. Bu verilerin işlenmesi ve analiz edilmesi için çeşitli teknikler ve yöntemler kullanılmaya başlanmıştır (Kumaş, 2021). Metinsel ifadelerden duygu analizi bu yöntemlerden birisidir. Duygu analizi, metnin ifade etmek istediği duygusal sınıfı belirlemeyi amaçlayan bir metin işleme süreci olarak temellendirilmiştir. Duygu analizi, duygusal kutupsallık olarak adlandırılan bir terimle verilen metinleri olumlu, olumsuz ve nötr olarak sınıflandırmayı hedefler. Zamanla, duygu analizi alanındaki çalışmalar, farklı duygu durumlarını belirlemeye yönelik analizlere de imkan tanımıştır. Genelde duygu durumlarını kodlamak için, veri kümesinde her bir metnin tek bir duygu ile etiketlenmesi veya birden fazla duygu ile etiketlenmesi gibi farklı yaklaşımlar bulunmaktadır (Seker, 2016).

Duygu Analizi(DA), metin madenciliğinin ve doğal dil işlemenin önemli bir parçasıdır. Duygu analizi kavramı, bir kişinin belirli bir konu, ürün, hizmet veya olay hakkında hissettiği duyguları ve bu konudaki tutumunu anlamak için kullanılır. Literatürde, duygu analizi farklı isimler altında da karşımıza çıkabilir; duygu durum analizi, fikir madenciliği, duygu sınıflandırması veya kanaat çıkarımı gibi. Ancak adlandırması ne olursa olsun, duygu analizinin temel amacı, bir metnin veya cümlenin içerdiği ifadelerin olumlu, olumsuz veya nötr olduğunu belirlemektir. Duygu analizi, genellikle metinlerde bulunan duygusal ifadeleri tanımlamak ve sınıflandırmak için kullanılan bir dizi teknik ve yöntemi içerir. Bu, insanların duygusal tepkilerini belirli bir ürün, hizmet, olay veya konu hakkında daha iyi anlamak ve bu bilgiyi işletmelerin karar alma süreçlerinde kullanmak için önemlidir. Örneğin, bir şirket yeni bir ürün piyasaya sürdüğünde, duygu analizi yoluyla tüketicilerin bu ürün hakkındaki düşüncelerini ve duygusal tepkilerini anlamak, pazarlama stratejilerini belirlemek ve ürünü iyileştirmek için kullanılabilir. Ayrıca, duygu analizi sosyal medya platformlarında da yaygın olarak kullanılır. Markalar, müşteri geri bildirimlerini izleyerek ve sosyal medya

etkileşimlerini analiz ederek marka itibarlarını değerlendirebilir ve yönetebilirler. Bu, kriz anlarında hızlı müdahale etmelerine ve müşteri memnuniyetini artırmak için gerekli adımları atmalarına olanak tanır. Kısacası, duygu analizi, metinlerdeki duygusal içerikleri anlamak ve sınıflandırmak için güçlü bir araçtır ve birçok alanda önemli bir rol oynar (Işık, 2019).

İşletmelerin uzun vadeli sürdürülebilirliği, büyük ölçüde müşteri ihtiyaçlarını karşılama yeteneklerine bağlıdır. Müşteri memnuniyetinin asıl amacı, bir işletmenin uzun vadeli varlığını sürdürmesinde kritik bir rol oynayan marka değerini artırmaktır. Bu nedenle, birçok işletme müşteri tercihlerini ve taleplerini anlamak için pazarlama araştırmalarına büyük yatırımlar yapmaktadır. Müşterilerin ürünleri nasıl algıladıklarını anlamak, etkili markalama ve konumlandırma stratejileri oluşturmak için hayati öneme sahiptir. Günümüz ticari ortamı, farklı aktörlerin dinamik ve karmaşık bir ortam oluşturduğu açıktır. Bu dinamik ortam, işletmeleri hızlı değişimlere ayak uydurmak ve doğru şekilde tepki vermek konusunda daha sistemik bir yaklaşım benimsemeye zorlamaktadır. Ancak, işletmelerin bu değişen ortama uyum sağlaması için kaynaklar ve zaman gerekmektedir. Ortak amaç ve çıkarlar doğrultusunda, işletmeler tamamlayıcı, sinerjik, etkileşimli ve hedefe yönelik eylemler geliştirmelidir. Bu bağlamda, rekabetçi araştırma olarak bilinen rakip analizi, bir firmanın ve sunduğu ürün veya hizmetlerin, benzer ürün veya hizmetler sunan diğer kuruluşlarla karşılaştırıldığında ne kadar rekabetçi olduğunu belirlemeye yönelik bir yöntemdir. Müşterilerin rakipler hakkındaki düşüncelerini ve motivasyonlarını anlamak, başarılı bir stratejik rekabet çalışmasının temel bir unsuru olarak kabul edilmektedir. Bu, işletmenin gerçek güçlü ve zayıf yönlerini belirlemeye yardımcı olacaktır. Bu tür bilgileri elde etmek için pazar araştırması yapılması gerekmektedir. Metin analitiği kullanarak duygu analizi yapmak, bu hedefe ulaşmada etkili bir yaklaşım sunmaktadır (Taherdoost ve Madanchian, 2023).

Duygu analizi, aynı zamanda fikir madenciliği olarak da bilinir ve doğal dil işleme (DDİ) tekniklerinden biridir. Bu teknik, bir metnin içerdiği duygusal tonu belirlemeye yönelik olarak kullanılır. Özellikle ürün ve hizmetlerle ilgili kullanıcı görüşlerini ve incelemelerini analiz ederek, bir organizasyon için harekete geçirilebilir bilgiler üretmek amacıyla yaygın bir şekilde uygulanır (Shaik vd., 2023).

Duygu analizi, bir bireyin, bir kavramın, bir konunun, bir olayın veya bir senaryonun temsil ettiği varlıkla ilgili insanların tutumlarının, görüşlerinin ve duygularının hesaplanarak analiz edilmesidir. Duygu analizi, bir sınıflandırma süreci olarak kabul edilebilir ve genellikle üç temel düzeyde gerçekleştirilir: belge düzeyi,

cümle düzeyi ve görünüm düzeyi. Belge düzeyindeki duygu analizinde amaç, bir görüş belgesini olumlu veya olumsuz bir duyguyla ilişkilendirerek kategorize etmektir. Bu süreçte, belge bütünü, tek bir temel bilgi birimi olarak ele alınır ve analiz edilir. Cümle düzeyindeki duygu analizi ise her cümlenin ifade ettiği duyguyu belirlemeye yöneliktir. Her cümle, içerdiği duyguya göre sınıflandırılır ve analiz edilir. Görünüm düzeyindeki duygu analizi, duyguları belirli varlıkların özelliklerine göre sınıflandırmak için duyarlılık analizi kullanır. Bu düzeyde, duyguların varlıkların farklı özelliklerine göre nasıl değiştiği incelenir ve analiz edilir (Geni vd., 2023).

Fikir madenciliği veya duygu analizi olarak adlandırılan süreç, belirli bir varlık veya konuyla ilgili duyguların, görüşlerin ve hislerin metin formatında ifade edilen davranışlarını analiz etmeyi içerir. Duygu analizinin başarısı, kelime, kelime öbeği, cümle ve belge düzeylerinde gerçekleştirilen ayrıntılı analizlere dayanır. Kelime düzeyinde yapılan analiz, bireylerin ürünler, hizmetler veya bunların nitelikleri hakkındaki düşüncelerini tespit etmeyi amaçlar. Örneğin, "Samsung M32'nin pil ömrü beni memnun etti" ifadesi, belirli bir ürünün belirli bir özelliği hakkında olumlu bir görüşü yansıtır. Cümle düzeyindeki duygu analizi, karmaşık ifadelerin içerdiği görüşlerin anlaşılmasına ve cümlenin genel duygusal tonunun belirlenmesine katkı sağlar. Belge düzeyindeki analiz ise, bir veya daha fazla cümlenin toplam görüşünü belirlemek için farklı yöntemlerin kullanılmasını içerir. Bu analizler, metnin genel duygusal durumunu ve içeriğini daha iyi anlamamıza yardımcı olur (Kahur ve Sharma, 2023).

Duygu analizi, bir metnin içerdiği duygusal tonu belirleyerek insanların tutumlarını, görüşlerini ve duygularını hesaplamalı olarak analiz etme sürecidir. Bu analiz, çeşitli alanlarda önemli bilgiler sağlayabilir; örneğin, karar verme süreçlerinde, sorunları çözme aşamalarında, kriz yönetimi süreçlerinde, yanlış anlamaları giderme çabalarında, istenen ürün ve hizmetleri sunma faaliyetlerinde, tüketicilerle etkileşimde bulunma süreçlerinde, ürün ve hizmet kalitesini artırma çabalarında, yeni pazarlama stratejileri geliştirme süreçlerinde ve satışları artırma çabalarında kullanılabilir. Ancak, duygu analizi yaparken karşılaşılabilecek bazı zorluklar ve sorunlar da mevcuttur. Örneğin, metnin öznel kısımlarını belirleme zorluğu yaşanabilir. Aynı kelimenin farklı bağlamlarda farklı duygusal anlamlara gelebilmesi bu süreci karmaşıklaştırabilir. Ayrıca, etki alanı bağımlılığı sorunuyla karşılaşılabilir; bir cümlenin farklı bağlamlarda farklı anlamlar taşıması durumu gibi. Alaycılığın tespiti de zorluklar arasında yer alır; bazı cümlelerde olumlu kelimeler kullanılarak aslında olumsuz bir anlam ifade

edilebilir. Bu gibi zorluklar, duygu analizi sürecini karmaşıklaştırabilir ve doğru sonuçlara ulaşmayı güçleştirebilir. Genel olarak, duygu analizi üç düzeyde gerçekleştirilir: cümle düzeyi, belge düzeyi ve unsur/özellik düzeyi. Cümle düzeyinde, her cümlenin içerdiği duygusal ton belirlenir. Belge düzeyinde, bir belgenin genel duygusal tonu tanımlanır ve olumlu veya olumsuz olarak sınıflandırılır. Unsur/özellik düzeyinde ise, belirli varlıkların veya özelliklerin içerdiği duygular incelenmeye çalışılır. Bu detaylı analizler, duygu analizi sürecinde daha kapsamlı ve doğru sonuçlar elde etmeyi sağlar (Kora ve Mohammed, 2023).

1.1.Amaç

Bu tez çalışmasının temel amacı, yapay zeka teknolojilerinin sağladığı avantajlardan yararlanarak makine öğrenimi yöntemlerini, Türkçe kullanıcı yorumlarından elde edilen veriler üzerinde duygu analizi yapmak için uygulamaktır. İlk aşamada, toplanan yorumlar olumlu ve olumsuz olarak etiketlenmiştir. Daha sonra, metinlerden Kelime Çantası (Bag Of Words) ve TF-IDF (Term Frequency-Inverse Document Frequency) gibi öznitelik grupları çıkarılmış ve bu özellikler kullanılarak LSTM(Uzun Kısa Süreli Bellek) ve 1D-CNN(1 Boyutlu Evrişimli Sinir Ağı) gibi derin öğrenme modelleriyle sınıflandırma işlemleri gerçekleştirilmiştir. Sınıflandırma yöntemleri ile elde edilen sonuçlar arasında karşılaştırma yapılmış ve hangi yöntemin daha başarılı sonuçlar verdiği ile ilgili bilgilere yer verilmiştir. Kelime Çantası, bir metinde geçen kelimelerin sıklığına dayalı olarak özniteliklerin çıkarılmasını sağlayan bir yöntemdir. Bu yöntem, metnin içeriği hakkında genel bir fikir verirken, kelime sırasını ve ilişkilerini göz ardı eder. Öte yandan, TF-IDF, bir kelimenin metinde ne kadar önemli olduğunu ölçen istatistiksel bir yöntemdir. Bu yöntem, nadir ve özgün kelimelerin daha fazla ağırlığa sahip olmasını sağlar, böylece metnin içeriği hakkında daha ayrıntılı bir fikir sunar. Bu çalışmanın bir diğer avantajı, hazır bir veri seti kullanılmamasıdır, bu da elde edilen sonuçların daha genel ve uygulanabilir olmasını sağlar. Ayrıca, Türkçe metinler üzerinde duygu analizi yaparken e-ticaret sitelerinden toplanan gerçek dünya verilerini kullanarak, literatürde az çalışılan derin öğrenme yöntemlerine odaklanarak önemli bir katkı sağlamayı amaçlamaktadır. Bu yaklaşım, alana yeni bir bakış açısı getirerek, derin öğrenme tekniklerinin duygu analizi gibi önemli uygulamalarda nasıl kullanılabileceğini göstermektedir.

Bu tez çalışması, e-ticaret sektöründe faaliyet gösteren firmaların müşteri yorumlarını değerlendirerek, bu yorumlardan duygu analizi yapılmasını sağlayacak bir model geliştirmeyi hedeflemektedir. Burada kullanılan yöntem derin öğrenme olarak isimlendirilmektedir. Derin öğrenme, büyük miktarda veriye dayalı olarak karmaşık desenleri ve ilişkileri otomatik olarak öğrenen yapay zeka tekniklerinden biridir.

Çalışmada, Trendyol ve Hepsiburada gibi bilinen ve popüler e-ticaret sitelerinden beyaz eşya ile ilgili toplanan kullanıcı yorumları üzerinde odaklanılmıştır. Bu sitelerin seçilme nedenleri arasında, bu platformlara erişimin kolaylığı, yorum çeşitliliğinin zenginliği ve bu sitelerin geniş bir kullanıcı tabanına sahip olması yer almaktadır. Bu yorumlar incelenerek, ürünlere ilişkin duygusal durumları (olumlu ve olumsuz yorumlar) içeren veri setleri oluşturulmuştur.

Tez çalışmasının önemli bir kısmı, bu veri setlerinin sınıflar arasındaki dağılımlarının dengeli olup olmadığını belirlemek ve doğru öznitelik seçimini yapmaktır. Daha sonra, bu faktörlerin, derin öğrenme algoritmalarının (örneğin, LSTM ve 1D-CNN gibi) performansını nasıl etkilediğini analiz etmek gerekmektedir. Bu analizler, daha etkili duygu analizi modellerinin tasarlanması ve e-ticaret firmalarının müşteri geri bildirimlerini daha verimli bir şekilde değerlendirmesine yardımcı olacaktır. Sonuç olarak, bu çalışma, e-ticaret sektöründe yapay zeka tabanlı duygu analizi alanında önemli bir katkı sağlamayı amaçlamaktadır.

1.2.Kapsam

Tez çalışmasının başlangıç aşamasında, metin madenciliği kavramı üzerine detaylı bilgiler sunulmuştur. Bu kapsamda, metin madenciliği alanındaki temel prensipler, yöntemler ve teknikler incelenmiştir. Daha sonra, çalışmanın odak noktası olan duygu analizi konusu ele alınmıştır. Duygu analizi yöntemlerine dair detaylı bir literatür taraması yapılmış ve bu alandaki önemli çalışmalara değinilmiştir.

Tez çalışmasının merkezinde yer alan uygulama çalışması, LSTM (Uzun Kısa Süreli Bellek) ve CNN-1 (1 Boyutlu Evrişimli Sinir Ağı) gibi derin öğrenme yöntemlerini kullanarak sonuç elde etmeyi hedeflemektedir. Bu yöntemlerin detayları incelenmiş ve uygulamada nasıl kullanılacakları açıklanmıştır. Ayrıca, modelin başarısını değerlendirmek için kullanılan ölçütler üzerinde durulmuştur.

Çalışmanın bir diğer bölümünde, daha önce gerçekleştirilen duygu analizi çalışmaları üzerine kapsamlı bir inceleme yapılmıştır. Bu inceleme, özellikle Türkçe

metinlerde duygu analizi alanında yapılan araştırmaların eksikliğine ve e-ticaret sitelerinden beyaz eşya yorumları ile elde edilen verilerle yapılan analizlerin sınırlı olduğuna işaret etmektedir. Bu noktada, tez çalışmasının bu boşluğu doldurmayı amaçladığı belirtilmiştir.

Sonuç olarak, yapay zeka ve makine öğrenmesi yöntemlerini kullanarak e-ticaret sitelerinden elde edilen veri setleri üzerinde Türkçe metinlerde duygu analizi yapılması, literatüre yeni bir yaklaşım sunmayı amaçlamaktadır. Bu çalışma, derin öğrenme yöntemlerinin duygu analizi alanında nasıl kullanılabileceğini göstererek, bu alanda yeni bir perspektif sunmaktadır.

2. KAYNAK ARAŞTIRMASI

Mevcut araştırmanın literatür taraması, derin öğrenme yöntemi ile metinsel ifadelerde duygu analizi konusunda ulusal ve uluslararası düzeyde gerçekleştirilen çalışmaları kapsamaktadır. Bu çalışmaların detaylı incelenmesi, mevcut araştırmanın savlarını ve iddialarını desteklemek üzere sağlam bir temel oluşturur. Literatürdeki bu çalışmaların analizi, mevcut araştırmanın alanındaki önemli gelişmeleri ve yöntemleri anlamamıza yardımcı olurken, araştırmanın yenilikçi katkılarını vurgulamak için de gereklidir. Bu bağlamda, derin öğrenme tekniklerinin, duygu analizinin ve diğer ilgili konuların nasıl kullanıldığına dair ayrıntılı bir anlayış sağlaması açısından önem arz etmektedir.

Kaur ve arkadaşları (2021), yaptıkları çalışmada R programlama dili kullanılarak Twitter verilerinin derinlemesine analizini içeren bir metot üzerinde çalışmışlardır. Twitter üzerindeki veriler, COVID-19, koronavirüs, ölümler, yeni vaka, iyileşen gibi belirli anahtar kelimeler kullanılarak toplanmıştır. Araştırma sürecinde, Hibrit Heterojen Destek Vektör Makinesi (H-SVM) adı verilen özgün bir algoritma geliştirilmiş ve bu algoritma Twitter verilerinin duygu analizini gerçekleştirmek üzere kullanılmıştır. Elde edilen sonuçlar, pozitif, negatif ve nötr duygulara ayrılmıştır, böylece duygusal eğilimler net bir şekilde belirlenmiştir. Ayrıca, önerilen algoritmanın performansı çeşitli kriterlerle değerlendirilmiştir; bu kriterler arasında hassasiyet, geri çağırma, F1 puanı ve doğruluk bulunmaktadır. Bu değerlendirme süreci, Tekrarlayan Sinir Ağı (RNN) ve Destek Vektör Makinesi (SVM) gibi yaygın kullanılan yöntemlerle yapılan analizlerle karşılaştırılarak gerçekleştirilmiştir. Bu kapsamlı inceleme, çalışmanın metodolojisini, bulgularını ve sonuçlarını daha derinlemesine anlamaya yardımcı olması düşünülmektedir. Twitter veri analizinin bireylerin ruh sağlığı üzerindeki etkisini belirlemek amacıyla duygu analizi gerçekleştirilmiştir. Duygu sınıflandırmasını gerçekleştirmek için Tekrarlayan Sinir Ağı (RNN) ve Destek Vektör Makinesi (SVM) gibi makine öğrenimi yöntemlerinden faydalanılmış ve elde edilen sonuçlar pozitif, negatif ve nötr duygu puanlarına göre kategorize edilmiştir (Kaur vd., 2021).

Bir diğer çalışmada duygu analizi ile ilgilenen Priyadarshini ve Cotton (2021) , yeni bir derin sinir ağı olan LSTM-CNN-grid arama tabanlı bir model önermektedir. Duygu analizi, akıllı şehir uygulamalarında kullanılabilecek doğal dil işlemenin en önemli araştırma alanlarından biridir. Geçmiş araştırmalar, yapay zeka ve diğer

geleneksel veya geleneksel olmayan yöntemler kullanılarak gerçekleştirilen duygu analizinin önemini vurgulamaktadır. Evrişimli sinir ağları (CNN), uzun kısa süreli bellek (LSTM), sinir ağları (NN), K-en yakın komşu (K-NN) gibi temel algoritmalar da çalışmada kullanılan çeşitli veri kümeleri üzerinde değerlendirilmiştir. Değerlendirme kriterleri arasında doğruluk, kesinlik, özgüllük, duyarlılık ve F-1 puanı bulunmaktadır. Çalışma sonuçları, önerilen modelin diğer temel algoritmalardan daha yüksek bir doğruluk oranına sahip olduğunu göstermektedir (%96 ve üzeri). Ayrıca, hiperparametre ayarının model performansını artırdığı gözlemlenmiştir. Gelecekte, bu tür hibrit makine öğrenimi tekniklerini araştırmak için diğer hiperparametre optimizasyon yöntemlerinin kullanılması planlanmaktadır. Ayrıca, ele alınan veri setinin çoğunlukla İngilizce kelimelerden oluştuğu göz önüne alındığında, farklı dillerdeki cümleleri içeren bir veri seti üzerinde duygu analizi yapılması ilginç olabilir. İngilizce dilinin çeşitliliği nedeniyle %100 doğruluk elde etmenin zor olduğu düşünüldüğünde, duygu analizinin insanlar ve botlar arasındaki farkı ayırt etmede önemli bir faktör olduğu yönündeki spekülasyonlar üzerinde durulması önemlidir (Priyadarshini ve Cotton, 2021).

Nandwani ve Verma (2021), duygu analizi ve duygu tespiti için kullanılan mevcut yöntemleri araştırmıştır. Buna göre Literatürde yapılan incelemeler, sözlük tabanlı yöntemlerin duygu analizi ve tespiti alanında etkili olduğunu ortaya koymaktadır. Sözlük tabanlı yaklaşımların esnek ve uygulanması kolay olduğu, derlem tabanlı yaklaşımların ise belirli bir alana özgü kurallar üzerine kurulduğu belirlenmiştir. Derlem tabanlı yöntemler genellikle daha doğru sonuçlar verirken, sözlük tabanlı yaklaşımlar genelleme yeteneğinden yoksun olabilir. Makine öğrenimi ve derin öğrenme algoritmalarının performansı, veri kümesinin boyutu ve ön işleme adımlarına bağlı olarak değişkenlik gösterir. Bazı durumlarda, makine öğrenimi modelleri metnin gizli özelliklerini başarıyla çıkaramayabilir. Büyük veri kümeleriyle çalışıldığında, derin öğrenme yöntemleri genellikle makine öğreniminden daha iyi sonuçlar verebilir. Özellikle, Tekrarlayan Sinir Ağları (RNN), uzun vadeli bağımlılıkları kapsayabilir ve etkili özellikler çıkarabilir. Bununla birlikte, dikkat mekanizmasına sahip RNN modelleri de oldukça başarılı olabilir. Sözlük tabanlı ve geleneksel makine öğrenimi yöntemlerinin sürekli olarak geliştiği ve daha iyi sonuçlar elde ettiği tespit edilmiştir. Ayrıca, ön işleme ve özellik çıkarma tekniklerinin, duygu analizi ve duygu tespiti yöntemlerinin başarısını önemli ölçüde etkilediği yapılan çalışmalarda görülmüştür (Nandwani ve Verma, 2021).

Başka bir çalışmada ise Rahman ve arkadaşları (2023) duyguların çok sınıflı sınıflandırılması için çeşitli ön işleme teknikleri ve makine öğrenmesi modelleri incelenmiştir. Duygu Analizi (SA), sıklıkla duygu tespiti olarak adlandırılan bir alanı içerir. Bu, metinlerden duyguların çıkarılması, tanımlanması veya karakterize edilmesi sürecidir. Genel olarak, bir metin belgesinin duygusu, genellikle olumlu ve olumsuz gibi ikili sınıflara ayrılır. Ancak, daha detaylı bir anlayış için duyguların ince taneli sınıflandırması tercih edilebilir. Bu yaklaşımın zorluğu ise ikili sınıflandırmaya göre daha yüksek olabilir. Öte yandan, çok sınıflı sınıflandırma durumunda performansın önemli ölçüde azalmasıyla karşılaşılabilir. Performansı artırmak için, çok katmanlı bir sınıflandırma modeli önerilmiştir. Uygulama bağlamında, sosyal medya metinlerine benzerlik gösteren film eleştirileri veri seti kullanılmıştır. Duygu sınıflandırma görevi için Karar Ağacı, Destek Vektör Makinesi ve Naïve Bayes gibi denetimli makine öğrenme modelleri kullanılmıştır. Tek katmanlı mimari modelleri, çok katmanlı modelle karşılaştırılmıştır. Çok katmanlı modelin sonuçları, tek katmanlı mimariye kıyasla hafif bir iyileşme göstermiştir. Ayrıca, çok katmanlı modellerin daha iyi hatırlanabilir olduğu ve bu durumun önerilen modelin daha fazla bağlam öğrenmesine olanak tanıdığı gözlemlenmiştir. Araştırmacıların, daha bağlamsal bilgilerle çok katmanlı modeller tasarlamak için önerilen modelin bazı eksikliklerini ele alması gerekebilir (Rahman vd., 2023).

Qorib ve arkadaşları (2023) yaptığı çalışmada, COVID-19 aşısı tereddütünü analiz etmek amacıyla üç duygu hesaplama yöntemini (Azure Machine Learning, VADER ve TextBlob) incemişlerdir. Beş farklı öğrenme algoritması (Random Forest, Logistics Regression, Decision Tree, LinearSVC ve Naïve Bayes) ve üç vektörizasyon yöntemi (Doc2Vec, CountVectorizer ve TF-IDF) kullanıldı. Kelime dağarcığının normalleştirilmesi için çömlekçi köklendirme, lemmatizasyon ve çömlekçi köklenme ile lemmatizasyon gibi üç farklı strateji uygulandı. Her normalleştirme stratejisi için 42 model tasarlandı, geliştirildi ve değerlendirildi. Araştırma sonuçları, COVID-19 aşısı tereddütünün zaman içinde azaldığını ve halkın aşılara karşı giderek olumlu ve umutlu hissettiğini göstermektedir. Ayrıca, çömlekçi köklendirme ve lemmatizasyonun birleştirilmesinin model performansını artırdığı tespit edildi. Son olarak, yapılan deneyin sonucunda TextBlob + TF-IDF + LinearSVC'nin doğruluk, kesinlik, hatırlama ve F1 puanı açısından kamuoyunun duygusal tepkilerini en iyi şekilde sınıflandırdığı görülmüştür. Bu, en iyi performansın, TF-IDF vektörleştirme ve LinearSVC sınıflandırma modeliyle TextBlob duyarlılık puanı kullanıldığında elde edildiği

anlamına gelir. Çalışmamız, TF-IDF vektörleştirme yöntemi ve TextBlob duyarlılık hesaplamasını kullanan LinearSVC'nin, önceki araştırmada kullanılan Gradient Boosting Machine model sınıflandırıcısından daha yüksek bir model doğruluğuna, %96.752'lik en yüksek değere ulaştığını göstermektedir. Ancak, CountVectorizer ve LinearSVC'nin birleştirilmesi, sınıflandırma modelinin doğruluğunun azalmasına neden olmaktadır. Bu durumun artan karmaşıklıktan kaynaklanabileceği düşünülmektedir (Qorib vd., 2023).

Haque ve arkadaşlarının (2023) önerdiği yöntemde, modelin maksimum doğruluğa ulaşmasını ve temel modellerle karşılaştırmalı bir analiz sunmayı amaçlamaktadır. Çok sınıflı duygu analizi (SA), bir metinde ifade edilen birden fazla görüşü tespit etmek için doğal dil işleme (NLP) ve metin madenciliği tekniklerini kullanır. Bengalce dilinde yapılan çalışmalar, üçlü sınıflandırma yöntemlerinin yetersiz performans sergilediğini göstermiştir. Bengalce metinlerin özellikleri, temel veri kümelerinin eksikliği ve ön işleme araçlarının kısıtlı kaynakları nedeniyle daha yüksek performans elde etmek zor olabilir. Derin öğrenme algoritmalarının dört duygu türünde daha üstün bir performans gösterdiğine dair kesin bir kanıt bulunmamaktadır. Bu nedenle, cinsel, dini, politik ve kabul edilebilir olarak etiketlenen Bengalce sosyal medya yorumlarını analiz etmek için bir CNN ve LSTM tabanlı denetimli derin öğrenme sınıflandırıcısı önerilmiştir. Bu çalışma, önerilen modelin maksimum doğruluğa ulaşmasını ve temel modellerle karşılaştırmalı bir analiz sunmayı amaçlamaktadır. Altı makine öğrenimi modeli, iki farklı özellik çıkarma tekniği kullanılarak temel modeller olarak kabul edilmektedir. Önerilen CNN ve LSTM tabanlı modelin performansı, 42.036 etiketli Bengalce sosyal medya yorumundan oluşan bir veri kümesinde %85,8 doğruluk ve 0,86 F1 puanı elde ederek çok sınıflı duygu analizi performansında büyük bir başarı elde edilmiştir. Önerilen modele ve en yüksek performanslı temel modele dayalı bir web uygulaması, sosyal medya yorumlarının gerçek duygusunu tespit etmek amacıyla geliştirilmiştir (Haque vd., 2023).

Khan ve Srivastava (2024) yaptıkları çalışmada, Twitter verilerindeki duyguları incelemek için farklı makine öğrenimi modelleri kullanmışlardır. Twitter duyarlılık analizi, sosyal medya platformu Twitter'da paylaşılan mesajlardaki genel duygusal içeriği inceleyen bir araştırma alanıdır. Bu analiz, makine öğrenimi ve doğal dil işleme tekniklerini kullanarak tweetleri otomatik olarak olumlu, olumsuz veya nötr olarak sınıflandırır. Bu teknikler, tek bir tweetten veya belirli bir konu veya olayla ilgili daha büyük bir veri setinden duygusal içerik çıkarmak için uygulanabilir. Bu duygusal

içeriklerin tanımlanması, makine öğrenimi modellerinin duygu analizi ve tahmininde avantaj sağlar. Twitter'da duyarlılık analizi yapmak için çeşitli yöntemler denenmiştir, bunlar arasında Vader, XGBoost ile CountVectorizer, XGBoost ile Gensim, Random Forest ile CountVectorizer, Random Forest ile Gensim, Tek Yönlü LSTM ve Çift Yönlü LSTM bulunmaktadır. Test edilen tüm bu makine öğrenimi modelleri arasında, Çift Yönlü LSTM'nin en üstün sonuçları ürettiği gözlemlenmiştir, %73 doğruluk oranı ile. Diğer modellerle karşılaştırıldığında, Çift Yönlü LSTM en yüksek performansı sergilemektedir. İlk olarak, Kaggle'da "fifa_world_cup_2022_tweets" adı altında bulunan veri seti yüklenecek. Bu veri seti, Katar'da düzenlenen FIFA Dünya Kupası'nın açılış gününe ait tweet'leri içermektedir. Daha sonra, veri seti önceden belirlenmiş görevler doğrultusunda ön işlemden geçirilecek, bu görevler arasında kullanıcı adlarının, URL'lerin ve stopword'lerin kaldırılması, lemmatizasyon ve tokenizasyon yer alacaktır. Ayrıca, etkinlik sırasında duygu dağılımı grafikleri, kelime bulutları ve zaman serisi duygu eğilimleri gibi görselleştirmeler oluşturularak içgörüler elde edilecektir. Ek olarak, pozitiflik, negatiflik ve tarafsızlık boyutlarını kapsayan bu modeller kullanılarak her bir tweet için duyarlılık puanları hesaplanacaktır. Son olarak, bu modeller arasında karşılaştırmalı bir çalışma yapılacaktır. Gelecekte, Sinir Ağı modeli potansiyel olarak diğer modellere kıyasla daha yüksek doğruluk sağlayabilir (Khan ve Srivastava, 2024).

Tan ve arkadaşları (2023) yayınladığı makalede, duygu analizi alanındaki en son gelişmelere genel bir bakış sunmakta olup, ön işleme teknikleri, özellik çıkarma yöntemleri, sınıflandırma teknikleri, yaygın olarak kullanılan veri kümeleri ve deneysel sonuçları incelemişlerdir. Ayrıca, duygu analizi veri kümelerinin ortaya çıkardığı zorlukları incelemekte ve gelecekteki araştırma olanaklarını tartışmışlardır. Duygu analizinin önemi göz önüne alındığında, bu makale, araştırmacılar ve uygulayıcılar için değerli bir kaynak olarak hizmet etmekte ve alandaki mevcut durumu hakkında kapsamlı bilgi sunmaktadır. Sunulan bilgiler, ilgili paydaşların en son duygu analizi gelişmeleri hakkında bilgi sahibi olmasına ve gelecekteki araştırmalara yön vermesine yardımcı olabilir. Duygu analizi, geniş bir uygulama yelpazesi nedeniyle doğal dil işlemede (NLP) kritik bir alandır. İlk duygu analizi çalışmalarında, metin verileri temizlenmiş, normalleştirilmiş ve frekans tabanlı özellikler kullanılarak temsil edilmiştir. Ancak, NLP'deki ilerlemelerle birlikte, araştırmacılar odaklarını derin öğrenme tekniklerine kaydırmışlardır. Derin öğrenme modelleri, önceden eğitilmiş kelime yerleştirmelerini kullanarak metin verilerini işlemekte ve daha karmaşık kalıpları tanımlamakta başarılı olmuştur. Bazı duyarlılık analizi çalışmaları, performansı

artırmak için birden fazla makine öğrenimi veya derin öğrenme modelini birleştirerek topluluk tabanlı bir yaklaşım benimsemektedir. Yaygın olarak kullanılan duygu analizi veri kümeleri arasında İnternet Film Veritabanı (IMDb), Twitter ABD Havayolu Duyarlılığı, Sentiment140 ve SemEval-2017 Görev 4 yer almaktadır. Ancak, mevcut yöntemler hala zayıf yapılandırılmış ve alaycı metinlere karşı savunmasızdır, bu da daha güvenilir dil modellerinin geliştirilmesi gerektiğini vurgular. Gelecekteki araştırmalar, duygusal yoğunluğu daha ince taneli sınıflarla ele alarak duygu analizi yöntemlerinin geliştirilmesine odaklanmalıdır. Ayrıca, bir konunun kutupsallık dağılımını hesaplayarak stratejik karar verme süreçlerinde kullanılacak yeni duygu ölçümleri üzerinde çalışılmalıdır (Tan vd., 2023).

Başka bir çalışmada Ahmed ve arkadaşları (2023), derin öğrenme modelleme tekniklerinin kapsamlı ve sistematik bir değerlendirmesini sunmaktadır. Bazı modeller, çeşitli yöntemlerle eğitilebilmekte ve kullanıldıkları alanlara bağlı olarak farklı verimlilik seviyeleri sergileyebilmektedir. Veri sınıflandırmasında hiyerarşik katmanların kullanılması ve eğitim sürecinde denetimin uygulanması, sağlam derin öğrenme modellerinin geliştirilmesi için kritik öneme sahiptir. Modellerin genellikle sağlam olmasına rağmen, mevcut tekniklerin hala eksikliklerinin bulunması eleştirilere neden olmaktadır. Büyük veri miktarının mevcut olduğu çeşitli alanlarda, derin öğrenme modellerinin eğitimi sırasında veri kalitesi bir zorluk olabilir. Ayrıca, derin öğrenme modellerinin eğitimi zaman ve maliyet açısından yüksek olabilir ve daha yüksek doğruluk için büyük miktarda doğru örneğe ihtiyaç duyulabilir, bu da günlük kullanımda veya hassas güvenlik sistemlerinde kullanımlarını kısıtlayabilir. Geliştirilen modeller ayrıca alanın spesifik ihtiyaçlarına göre şekillenebilir ve bu nedenle sınırlı uygulamalara sahip olabilir. Ayrıca, derin öğrenme modelleri, aldatma ve yanlış sınıflandırmaya karşı hassas olabilir, bu da bireylerin ve/veya kurumların sosyal ve finansal güvenliklerini tehdit edebilir. Çoğu modelin yerel minimumlara takılma eğilimi, çevrimiçi uygulamalar için uygun olmadığını göstermektedir. CNN'ler, RNN'ler, GAN'lar ve otomatik kodlayıcılar, birçok sektörde sıkça kullanılan derin öğrenme mimarileridir. Ancak, diğer mimarilerin DL kullanım alanlarında potansiyel uygulaması genellikle yeterince araştırılmamıştır. Bu çalışma, geleneksel DL mimarilerinin karşılaştığı zorlukların üstesinden gelebilecek hibrit gelişmiş DL modellerinin varlığını belirlemiştir. Ayrıca, üretken modellerin, daha az sayıda örneğe bağımlı oldukları için daha büyük esneklik sağladığı bulunmuştur. Gelecekteki ağlar, çarpık veya belirsiz girdi sorunlarını çözmeye yardımcı olabilecek bir dizi olası sonuç

üretme yeteneğine odaklanmalıdır. Parametrelerin, özellikle de hiperparametrelerin optimize edilmesi için yeni stratejilerin geliştirilmesi, daha fazla araştırmayı gerektiren bir olasılıktır. Kapsül ağları, katmanlar arasında bilgi yönlendirmenin gelişmiş bir yolunu sunarak gelecekte derin öğrenme modellerine egemen olabilir. Mevcut zorlukların üstesinden gelinirse, derin öğrenme modelleri yapay zeka alanında daha fazla inovasyona ve daha karmaşık sorunların çözümüne katkıda bulunabilir (Ahmed vd., 2023).

Bir diğer kaynak araştırmasına bakacak olursak Mitrović ve arkadaşları (2023), özellikle kısa metinlerde bir makine öğrenimi modelinin, insanların oluşturduğu metinler ile ChatGPT (Generative Pre-trained Transformer) tarafından oluşturulan metinler arasındaki farkı doğru bir şekilde belirleyip belirleyemeyeceğini inceleyen bir modelin eğitilmesini detaylandırmıştır. ChatGPT, çeşitli alanlardaki çeşitli türdeki sorulara dilbilgisi açısından kusursuz ve insanı andıran yanıtlar üretme yeteneğine sahiptir. Kullanıcıların ve uygulamaların sayısı artmaktadır ve bu durum, ne yazık ki, kullanım ve istismar arasında dengesiz bir artışa neden olmaktadır. Diğer yandan, ChatGPT tarafından üretilen metinlerle insanlar tarafından üretilen metinler arasındaki ayrımı öğrenmek için açıklanabilir bir yapay zeka çerçevesi kullanılmaktadır. Bu bağlamda amaç, modelin kararlarını analiz etmek ve belirli bir modelin veya özelliğin tanımlanıp tanımlanamayacağını belirlemektir. Bu çalışmaya göre, insanlar tarafından oluşturulan metinlerle ChatGPT tarafından oluşturulan metinleri karşılaştıran iki deneye odaklanmaktadır. Birinci deneyde, özel sorgulardan oluşturulan ChatGPT metinleri kullanılırken, ikinci deneyde, insanlar tarafından oluşturulan orijinal metinlerin yeniden ifade edilmesiyle oluşturulan metinler kullanılmaktadır. Bir dönüşümlü tabanlı modele ince ayar yapılır ve bunu tahminlerde bulunmak için kullanılır; bu tahminler daha sonra SHAP kullanılarak açıklanır. Model, şaşkınlık puanına dayalı bir yaklaşımla karşılaştırılır ve insan incelemeleri ile ChatGPT tarafından üretilen incelemeler arasındaki belirsizliğin, yeniden ifade edilen metinler kullanıldığında makine öğrenimi modeli için daha zor olduğu tespit edilir. Ancak yapılan çalışmada bu yaklaşım %79 doğruluk oranına sahiptir. Detaylı ifade edilebilirliği kullanarak, ChatGPT'nin sözel ifadelerin nazik , detaylara yer vermediğini, sözleri biraz abartarak güzelleştirdiğini, öznel olmayan ve büyük oranda duygulara yer vermediği gözlemlenmiştir (Mitrović vd., 2023).

Puh ve Bagić Babac (2023), ele aldıkları konu ile, turist yorumlarından duygu ve derecelendirmeyi öngörmek için makine öğrenmesi ve derin öğrenme modellerini

araştırma konusu yapmışlardır. Ayrıca yapılan çalışma, turist yorumlarında duygu analizi için makine öğrenmesi ve derin öğrenme modellerinin nasıl yorumlanabileceğine dair bir kılavuz sunmayı amaçlamaktadır. BiLSTM yönteminin hem duyarlılık hem de derecelendirme sınıflandırma görevlerinde üstün performans sergilediği bulunmuştur. Özellikle, BiLSTM modelindeki veriler, ilk olarak bağlamsal bilgi toplama görevi olan iki katmanlı BiLSTM'den geçirilmiş ve daha sonra sınıflandırmayı gerçekleştiren üç doğrusal katmana işlenmiştir. Modeller, incelemeleri kelime temsilini sağlamak için GloVe ile beşe ve ardından üçe sınıflandırmak üzere eğitime alındı. En sonunda, beş sınıf için en iyi performansı %72 doğruluk elde eden model gösterirken, üç sınıf için %89 seviyesinde olan model en iyisi olmuştur. Yöntemsel karşılaştırma amacıyla, Naïve Bayes ve SVM gibi diğer makine öğrenme modellerinin yanı sıra 1D evrişim ve LSTM gibi diğer derin öğrenme modelleri uygulamaya alınmıştır. Modelleme sonuçları, BiLSTM tabanlı derin öğrenme modelinin verilen iki görevde de en başarılı sonuçlara ulaştığı görülmüştür (Puh ve Bagić Babac, 2023).

Bir diğer çalışma ele alındığında Bayer ve arkadaşları (2023), hem uzun hem de kısa metinler için performansı artırmaya yönelik bir metin üretme yöntemi sunmuş ve değerlendirmişlerdir. Çeşitli makine öğrenimi örneklerinde yapılan araştırmalar, sınıflandırıcıların seçimi ve modellenmesinin yanı sıra eğitim verilerinin geliştirilmesinin de önemli olduğunu öne sürmüşlerdir. Bu nedenle, sınıflandırıcıların performansını artırmak için yapay olarak oluşturulan eğitim verileriyle ilgili veri büyütme yöntemleri geliştirilmiştir. Doğal Dil İşleme (NLP) alanında, metin dönüşümleri için evrensel kuralların oluşturulması zorlukları mevcuttur. Metin üretme yönteminin geliştirilmesiyle, kısa ve uzun metin görevlerinde önemli ilerlemeler kaydedilmiştir. Özellikle küçük veri analitiği bağlamında, yapılan çalışma %15,53 ve %3,56 oranında ek doğruluk kazanımları sağlamıştır, Bu da diğer veri artırma teknikleriyle karşılaştırıldığında dikkate değer bir iyileşme sağlamaktadır. Oluşturulan rejimlerin evrensel olarak uygulanabilir olmadığı düşünüldüğünde, gerçek dünya veri görevlerinde önemli gelişmeler kaydedilmiştir. Bu çalışma, farklı veri kümeleri üzerinde başarıyla uygulanabilen çıkarımlar ve kalıplar ele almaktadır (Bayer vd., 2023).

Anand ve arkadaşlarının (2023) önerdikleri bir çalışmada ise saldırgan dilin tespitinde derin öğrenme tekniklerine dayanan bir özellik seçimi ve sınıflandırma yaklaşımını ele almışlardır. Bu bağlamda, YouTube, Twitter ve Facebook gibi platformlardan elde edilen veri setleri, gürültü giderme, filtreleme ve dur sözcüklerinin

çıkarılması gibi ön işlemlerden geçirilmiş ve segmentlere ayrılmıştır. Daha sonra, bulanık mantık kurallarına dayalı olarak seçilen özellikler evrişimli sinir ağlarıyla çıkarılmıştır. Ardından, bu çıkarılan özellikler, topluluk tabanlı mimari kullanılarak sınıflandırılmıştır. Sosyal medya paylaşımları, saldırgan iletişimi yansıtabilmektedir. Bu tür içeriğin tespiti için hesaplamalı algoritmaların kullanılması, bu sorunla başa çıkmanın etkili bir yoludur. Bu araştırma, özellik seçimi ve sınıflandırmada kullanılan MOLD_DL (Çok Dilli Saldırgan Dil Tespiti) tekniklerini ve doğal dil işleme konularını ele almaktadır. Veri seti, YouTube, Twitter ve Facebook'tan toplanmış, gürültü giderme, filtreleme ve dur sözcüklerinin kaldırılması gibi ön işlemlerden geçirilmiş ve segmentlere ayrılmıştır. Bölümlenmiş veriler için özellik seçimi, bulanık tabanlı evrişimli sinir ağları (FCNN) kullanılarak gerçekleştirilmiştir. Daha sonra, seçilen özelliklerin çıkarılması ve sınıflandırılması, Bi-LSTM modeli ile Destek Vektör Makineleri (SVM) ve Naïve Bayes mimarisinin bir hibriti kullanılarak gerçekleştirilmiştir. Saldırgan dil tespitinin değerlendirilmesi, metnin duygularına göre otomatik olarak sınıflandırılarak gerçekleştirilmiştir. YouTube, Twitter ve Facebook veri kümeleri için yapılan deneysel analizler, karışıklık matrisiyle %98 doğruluk, %95 kesinlik, %90 geri çağırma, %92,5 F-1 puanı ve %45 RMSE elde edilmiştir(Anand vd., 2023).

Qi ve Shabrina (2023) makalesinde, İngiltere'nin büyük şehirlerindeki Twitter kullanıcılarından Kovid-19 ile ilgili verileri analiz eden ve bu süreci üç farklı aşamada gerçekleştiren bir model üzerinde çalışma yapmıştır. Başlangıçta, veri temizliği yapılır ve ardından tweetlerin duygu durumlarını belirlemek için denetimsiz sözlük tabanlı yaklaşımlar kullanılır. Daha sonra, Rastgele Orman, Multinomial Naïve Bayes ve SVC gibi denetimli makine öğrenimi modelleriyle eğitilmiş bir veri seti kullanılarak sınıflandırma işlemi yapılır. Sözlük tabanlı yaklaşımlar, Kovid-19 salgınına yönelik kamu algısının üç aşamalı değişimini gözler önüne serer. Şehirlerde, olumlu duyguların oranı artar, sonra azalırken olumsuz duyguların oranı farklı bir seyir izler. Ayrıca, ölüm ve teyit edilen vaka sayıları ile aşılama durumları incelenerek, artan teyit edilen vakaların ve azalan aşılama oranlarının olumsuz duyguların artmasına katkıda bulunabileceği öngörülebilir. Bu çıkarımı doğrulamak için daha fazla araştırma yapılması gereklidir. Denetimli makine öğrenimi sınıflandırıcıları için, her bir modelin hiper parametrelerini optimize etmek için Rastgele Arama yöntemi uygulanır. BOW ve TF-IDF özellik modelleri kullanılarak yapılan sınıflandırma sonuçları en iyi performansı sağlar ve %71'e kadar yüksek sınıflandırma doğruluğuna ulaşır. Word2Vec

gömme yöntemine dayalı sınıflandırıcılar, eğitim verilerinin yetersizliği nedeniyle düşük tahmin doğruluğuna sahiptir. Sonuç olarak, makine öğrenimi yaklaşımları ile duygu analizi yapmak, sözlüklerle sınırlı kalmadan metin özelliklerini doğru bir şekilde tespit edebilir. Bu makalenin İngiltere'deki Twitter kullanıcılarının Kovid-19 görüşlerini temsil ettiği ve sonuçların bu sınırlılık dikkate alınarak yorumlanması gerektiği unutulmamalıdır. Bu çalışma, çeşitli NLP yöntemlerinin kullanımı hakkında bilgiler sağlamıştır. Sözlük tabanlı yaklaşımlar için mevcut sözlükler, modern sosyal medya diline daha uygun hale getirilerek doğruluk artırılabilir. Daha ileri çalışmalar, duygu sınıflandırma sonuçlarını diğer faktörlerle birleştirerek hangi faktörlerin duygu değişikliklerine katkıda bulunduğunu analiz edebilir. Genel olarak, bu makale, BoW veya TF-IDF kullanarak SVC ile duygu analizi yapmanın, genel model doğruluğundan daha iyi performans gösteren farklı yöntemlerini göstermektedir (Qi ve Shabrina, 2023).

Bir başka araştırmada ise Huang ve arkadaşlarının (2023) amacı, mevcut e-ticaret platformlarında kullanılan duygu analizi tekniklerini ve gelecekteki gelişim yönlerini araştırmaktır., Son dönemlerde, doğal dil işlemenin pratik uygulamalarından biri olan duygu analizi, "fikir madenciliği" olarak da bilinir ve metinlerdeki gizli duygusal içerikleri tespit etmeyi amaçlar. Bu analiz, özellikle e-ticaret platformlarında, yorumlar ve incelemeler gibi değerli iş bilgisi içeren metinlerin büyük önem taşıdığı alanlarda yarar sağlar. Bu çalışmanın amacı, mevcut e-ticaret platformlarında kullanılan duygu analizi tekniklerini ve gelecekteki gelişim yönlerini araştırmaktır. Yapılan sistematik inceleme, mevcut literatürde bu konuya yönelik kapsamlı bir araştırma makalesinin bulunmadığını ortaya koymuştur. Bu çalışmanın bulguları, duygu analizi alanında çalışan araştırmacılara mevcut teknikler ve platformlar hakkında derinlemesine bir anlayış sunmanın yanı sıra gelecekteki araştırma yönelimlerine dair içgörüler sağlayabilir. Belirli anahtar kelimeler kullanılarak 271 makale incelenmiş ve bunlardan 54'ü deneysel çalışmalar olarak seçilmiştir. Bu makalelerin %48'i sadece makine öğrenimi tekniklerini kullanırken, %44'ü derin öğrenme teknikleri aracılığıyla duygu analizini ele almıştır ve %7'si her iki yaklaşımı da birleştirerek hibrit bir yöntem kullanmıştır. Ayrıca, yapılan inceleme Amazon ve Twitter'ın araştırmacılar arasında en popüler veri kaynakları olduğunu göstermiştir (Huang vd., 2023).

Savcı ve Das (2023)'a göre birbirinden farklı dillerde yazılan ve yazım hatası barındıran metinlerin başarısını neyin etkileyeceği ve hangi yönlerde daha başarılı olacağı, duygu analizini içeren bir çalışma yapılması elzem bir konu olmuştur ve buna uygun bir çalışma yürütmüşlerdir. Yaptıkları çalışmaya göre Türkçe, Arapça ve

İngilizce gibi çeşitli dillerden alınan e-ticaret verileri üzerinde duygu analizi yapılmış ve bu verilere dayanarak derin öğrenme ve makine öğrenme yöntemlerinin performansları karşılaştırmalı olarak değerlendirilmiştir. Verilerden olumlu, olumsuz ve nötr duygu sınıflarını belirlemek için önceden eğitilmiş çeşitli dil modelleri kullanılmış ve bu modellerin duygu analizi performansı üzerindeki etkisi incelenmiştir. Yapılan deneylerin sonuçları, önceden eğitilmiş dil modellerinin üç sınıf duygu analizinde başarılı sonuçlar verdiğini göstermiştir. Örneğin, Arapça veri setinde bert-base-multilingual-cased modeli %95,1 doğruluk oranıyla en yüksek performansı sergilemiştir. Benzer şekilde, İngilizce veri setinde distilroberta tabanlı model %91,4 doğruluk oranıyla en üst düzeyde başarı elde etmiştir. RNN gibi derin öğrenme modelleri İngilizce veri setinde %89,5 doğruluk oranıyla en iyi performansı gösterirken, Arapça veri setinde SVM gibi makine öğrenme yöntemleri %91,3 doğruluk oranıyla en üst düzeyde başarı sağlamıştır (Savcı ve Das, 2023).

Başka bir çalışmayı ele almak gerektiğinde Das ve arkadaşları (2023), yapılandırılmamış veriler üzerinde metin sınıflandırma için özellik ağırlıklandırma yöntemlerini ele alarak detaylandırmışlardır. Metin sınıflandırma, bir belgeyi çeşitli kategorilere ayırma işlemidir ve bu aralık, Doğal Dil İşleme (NLP) alanında temel bir öneme sahiptir. Terim Frekansı-Ters Doküman Frekansı (TF-IDF) ve NLP, metin sınıflandırmada sıkça kullanılan bilgi erişim yöntemleridir. Tavsiye edilen model, IMDB film incelemelerinden N-Gram'lar ve TF-IDF kullanarak duygu analizi için Amazon Alexa inceleme veri kümesini içerdi. Ardından, yöntemi doğrulamak için güncel sınıflandırıcılar kullanıldı; bunlar arasında Destek Vektör Makineleri (SVM), Lojistik Regresyon, Çok Terimli Naif Bayes (Çok Terimli NB), Rastgele Orman, Karar Ağacı ve k-en yakın komşular (KNN) bulunmaktadır. N-Gram'lar yerine TF-IDF özellikleriyle yapılan özellik çıkarımı, kritik bir doğruluk artışına yol gösterdi. TF-IDF, Rastgele Orman sınıflandırıcısında maksimum doğruluk (%93,81), kesinlik (%94,20), geri çağırma (%93,81) ve F1 skoru (%91,99) elde etti (Das vd., 2023).

Dang ve arkadaşlarına (2020) göre, derin öğrenme modellerinin ve ilgili tekniklerin sosyal medya verileri üzerinde duygu analizi yapmak için nasıl kullanıldığını ele almak gerekir. Bu kapsamda çalışma yapmışlardır ve bu amaçla giriş verilerini işlemek için kelime yerleştirme ve TF-IDF gibi yöntemlerden yararlanılmıştır. Duygu analizi için DNN, CNN ve RNN mimarileri incelenmiş ve bu modellerle kelime yerleştirme ve TF-IDF özellikleri birleştirilerek farklı konulardaki veri kümeleri üzerinde deneyler yapılmıştır. Ayrıca, alanda yapılan araştırmaları da tartışarak, derin

öğrenme modellerinin duygu analizi üzerindeki uygulamaları ve bu modellerin metin ön işleme teknikleriyle birleştirilmesi konusunda kapsamlı bir bakış sunulmuştur. Yapılan analizler sonucunda, duygu kutupluluğu analizinde en yaygın kullanılan modellerin DNN, CNN ve hibrit yaklaşımlar olduğu belirlenmiştir. Ancak, literatürdeki çalışmalarda yapılan karşılaştırmalı analizlerin eksikliği vurgulanmıştır. Ayrıca, çoğu makalenin hesaplama süresini göz ardı ederek sonuçlarını sunması güvenilirlik açısından sınırlılık oluşturmaktadır. Bu çalışmada yapılan deneyler, bu eksiklikleri gidermeye yönelik olarak tasarlanmıştır. Farklı veri kümeleri, özellik çıkarma teknikleri ve derin öğrenme modellerinin etkileri, özellikle duygu kutupluluğu analizi üzerinde odaklanılarak incelenmiştir. Sonuçlar, duygu analizi için derin öğrenme tekniklerinin kelime yerleştirme yöntemleriyle birleştirilmesinin TF-IDF kullanımından daha etkili olduğunu ortaya koymaktadır. Deneyler ayrıca, CNN'in diğer modellere göre daha iyi bir performans sergilediğini ve doğruluk ile işlemci kullanımı arasında denge sağladığını göstermiştir. RNN'nin çoğu veri setinde daha yüksek güvenilirlik sağladığı ancak hesaplama süresinin daha uzun olduğu belirlenmiştir. Çalışmadan elde edilen bir başka önemli sonuç ise algoritmaların etkililiğinin büyük ölçüde veri kümelerinin özelliklerine bağlı olduğudur, bu nedenle derin öğrenme yöntemlerinin daha geniş veri kümeleriyle test edilmesinin önemi vurgulanmıştır (Dang vd., 2020).

Bu çalışmada Gündüz (2023), Türkçe film incelemelerini içeren beyazperde.com sitesinden toplanan veriler üzerinde gerçekleştirilmiş olup, derin öğrenme modeli BERTurk'e dayalı bir metodolojiyi incelemiştir. Duygu analizi, metinlerdeki duygu, algı ve düşünceleri anlama sürecidir ve sıklıkla "görüş madenciliği" olarak adlandırılır. Bu çalışma, Türkçe film incelemelerini içeren beyazperde.com sitesinden toplanan veriler üzerinde gerçekleştirilmiş olup, derin öğrenme modeli BERTurk'e dayalı bir metodolojiyi incelemiştir. İlk deneyde, BERTurk modelinin dönüştürücü katmanından elde edilen derin temsiller, Destek Vektör Makineleri (DVM) modeliyle birleştirilerek sınıflandırma yapılmıştır. İkinci deneyde, BERTurk modeli ince ayarlanarak sınıflandırma işlemi gerçekleştirilmiştir. Son deneyde ise, ince ayarlı BERTurk modelinden elde edilen temsiller, DVM ile sınıflandırılmıştır. Deneyler sonucunda, en yüksek doğruluk oranının (%0.984) ince ayarlı BERTurk temsilleriyle elde edildiği gözlemlenmiştir. İnce ayar işlemi, doğruluk oranında %10'luk bir artış sağlarken, BERTurk'ten türetilen temsillerin DVM ile birleşimi kullanıldığında ise doğrulukta %5'lik bir artış gözlemlenmiştir (Gündüz, 2023).

Daşgın ve Adem' e göre (2022) günümüzde, erişimi kolay, esnek ve zamansal kısıtlamalardan az etkilenen şeyler daha fazla talep görmektedir. Bu çalışmada kurslara yapılan online yorumlar üzerinden derin öğrenme ve makine öğrenmesi yöntemleriyle bu kursların değerlendirilmesi ele alınmıştır. Son yıllarda, bireylerin mevcut olanakları ve teknolojik ilerlemelerin etkisiyle eğitim alanında ilgi alanlarında önemli değişiklikler yaşanmaktadır. İnsanlar artık kendi tercih ettikleri içeriklere, istedikleri zaman ve mekandan erişebilmek istemektedirler. Bu taleplerin bir sonucu olarak, Kitlesel Açık Çevrimiçi Ders (KAÇD) platformları gelişmeye başlamıştır. Bu platformlar, genellikle ücretsiz veya ücretli olarak birçok farklı konuda kursları içermektedir. Ancak, bir kursa kayıt olmadan önce, kullanıcıların yorumları ve kursun puanı gibi değerlendirmeler göz önünde bulundurulmaktadır. Ancak, tüm yorumları okuyarak bir kurs hakkında karar vermek zor olabilir. Bu nedenle, Udemy gibi bir KAÇD platformunda yer alan kurslara yapılan yorumlar incelenmiştir. Bu yorumlar üzerinden, klasik makine öğrenme ve derin öğrenme teknikleri kullanılarak kursların değerlendirmesi yapılmıştır. Klasik makine öğrenme algoritmaları arasında BayesNet, J48 ve OneR kullanılmış ve en yüksek doğruluk oranı %91.576 ile BayesNet algoritmasıyla elde edilmiştir. Veri seti üzerinde Random, GloVe ve Word2Vec gibi kelime gömme yöntemleri uygulanmıştır. Derin öğrenme modelleri arasında ise GRU ve CNN-LSTM hibrit mimarileri kullanılmış ve en yüksek doğruluk oranı GloVe kelime gömme yöntemi kullanıldığında %95.67 ile GRU mimarisiyle elde edilmiştir (Daşgın ve Adem, 2022).

3. MATERYAL VE METOT

3.1.Materyal

Bu tez çalışmasında, e-ticaret sitelerindeki beyaz eşya yorumlarını analiz etmek amacıyla özgün olarak oluşturulmuş bir veri seti kullanılmıştır. Araştırmanın amacı, kullanıcı yorumlarını olumlu ve olumsuz olarak sınıflandırmak ve bu sınıflandırma üzerinden duygu analizi yapmaktır. Yorumlar, olumlu ve olumsuz olmak üzere iki ana kategoriye ayrılmıştır; olumlu yorumlar "1", olumsuz yorumlar ise "0" olarak etiketlenmiştir. Toplamda 5022 yorumu içeren bu veri seti, 2508 olumlu ve 2514 olumsuz yorumdan oluşmaktadır.

Veri seti, Türkiye'nin önde gelen e-ticaret platformlarından Trendyol ve Hepsiburada'dan toplanmıştır. Yorumlar toplanırken, farklı beyaz eşya ürünleriyle ilgili yapılan değerlendirmeler dikkate alınmış ve en yüksek yıldız alan olumlu yorumlar ile en düşük yıldız alan olumsuz yorumlar seçilmiştir. Trendyol'daki yorumların genellikle kısa, tekrarlayan ve birbirini tekrar eden nitelikte olduğu gözlemlenirken, Hepsiburada'daki yorumların daha özenli, çeşitli ve gerçekçi olduğu tespit edilmiştir.

Toplanan yorumlar, ham veri olarak Excel formatında kaydedilmiş ve daha sonra Matlab ortamına aktarılmak üzere düzenlenmiştir. Bu aşamada, veri seti eğitim ve test olmak üzere ikiye ayrılmıştır. Farklı eğitim-test oranları denenmiş ve en başarılı sonuçlar %80 eğitim - %20 test oranında elde edilmiştir.

E-ticaret sitelerinde, kullanıcı yorumları sıklıkla konuşma dili ile yazılmakta ve anlam taşımayan Türkçe simgeler, ifadeler, şekiller, semboller ve hatta linkler içermektedir. Bu tür içerikler, duygu analizi açısından anlamsız olup, veri setinde gürültü yaratmaktadır. Bu nedenle, veri seti bu tür gürültülerden arındırılmış ve duygu analizi için daha uygun hale getirilmiştir.

Bu çalışma, e-ticaret sitelerindeki beyaz eşya yorumlarının duygusal analizine odaklanarak, kullanıcı yorumlarının daha derinlemesine anlaşılmasına ve değerlendirilmesine katkı sağlamayı amaçlamaktadır. Kullanıcı yorumlarının olumlu ve olumsuz olarak sınıflandırılması, e-ticaret sitelerinin müşteri memnuniyetini artırma stratejilerine ve ürün geliştirme süreçlerine önemli girdiler sağlayabilir. Ayrıca, bu tür analizler, tüketicilerin satın alma kararlarını etkileyen faktörlerin daha iyi anlaşılmasına yardımcı olabilir.

Sonuç olarak, bu çalışma, e-ticaret platformlarında yer alan kullanıcı yorumlarının sistematik bir şekilde analiz edilmesi ve sınıflandırılması yoluyla, beyaz eşya sektöründe müşteri geri bildirimlerinin daha etkin bir şekilde değerlendirilebileceğini göstermektedir. Bu da hem akademik araştırmalar hem de pratik uygulamalar açısından önemli bulgular sunmaktadır.

Çizelge 3.1. Olumlu yorumlar.

Değer	olumlu yorumlar
1	ürünü eşim için aldım gerçekten inanamadık çıkan havlara tozlara alın aldırın tavsiye
1	Kendime yaptığım en büyük iyilik oldu. Kesinlikle alınması gereken bir ürün hele ki ş
1	gerçekten çok memnun kaldım büyük rahatlık kurulumu da ücretsiz herkesin kullanı
1	1 yıl önce kendime 8 kg olan serisinden almıştım çok memnun kaldım şimdi yeni çıl
1	rün elime çok çabuk ve sorunsuz ulaştı ihtiyacı olan hiç düşünmeden alsın
1	Harika kendime yaptığım en güzel iyilik
1	Ne zamandır araştırıyordum. Bir türlü karar veremiyordum. Buradaki yorumlar sayes
1	kendim için yaptığım en iyi şey diyebilirim bir kurutma makinasında olması gereken
1	eşime aldım çok beğendi
1	Ürün gerçekten güzel.Kendime Arçelik aldım anneme bunu aldım .Bu daha güzel de
1	aklaşık 1 aydır kullanıyorum. baştan sonra her aşamasından çok memnumum. lüks
1	Evime ilk kez Altus aldım Öner'i üzerine ama şahane bir ürün servis hafta sonu olma
1	harika bir şey keşke daha önce alsaymışım

Çizelge 3.2. Olumsuz yorumlar.

değer	olumsuz yorumlar
0	Urun 15 gun sonra teslim oldu cok sorun yasadik kesinlikle tavsiye etmiyorum
0	Kötü puan vermemin sebebi ürünü aldıktan sonra yaklaşık 2 bin ₺ düşmesinden dolayı
0	Ayın 10 unda sipariş verdim.Manisa dan yola çıktı İstanbul a gelmek üzere.Bugün ayın
0	Hiç kimseye tavsiye etmiyorum koku yapıyor alalı 3 ay olmadı 1 ay bile düzgün kullana
0	Aldıktan sonra indirime girdi aynı durum başka ürün için diğer turuncu sitede yaşandı or
0	Sıfır ürünün üzerinde cizik var kalite kontrol yapılmadan gönderilmesi hiç hoş değil.
0	asla soğutmuyor.servisi aradım ama 3 mayısta gelecekler. kullanmadığım ürün için inşa
0	Teslim almadığım ürün teslim almış olarak işaretlendi. Kargo şirketi hala teslim etmedi.
0	Soğutması sıfır hiç güzel değil. Paranız çöp olur
0	Üründen hiç memnun değilim. Çok aşırı ses çıkartıyor ve ilk 2 bardakta istenilen soğukl
0	alalı 1,5 ay oldu aldığımızdan beri suyun tadında metalik bir tat var . tavsiye etmiyorum
0	tavye etmem soğutmuyor
0	Ürün elime 15 gün sonra ulaştı çok kötü kargo sistemi var alırken düşünün süreyi
0	su soğutması yeterli değil.
0	Baya baya Hafif bi ürün yani böyle ben sağlamım demiyor ...markasına güvenerek aldın
0	Ürün çok iyi, sağlam ulaştı. Fakat ulaşması çok uzun sürdü, kargoyu takip edebileceği

Veri madenciliği modellerinin çıktıları genellikle tahminlerden veya modellenmiş veri setlerinden oluşur. Kısacası, veri madenciliği, veri kümelerindeki

genellikle fark edilmeyen bilgilerin belirlenmesi ve bu bilgilere dayalı karar alma ve tahmin yapma çalışmaları için geliştirilmiş çeşitli tekniklerin kullanılması sürecidir. Bu tekniklerin uygulanmasıyla elde edilen modeller, veri madenciliğinin sonuçları olarak kabul edilebilir (Erden, 2021).

Karmaşık veri setlerinden anlamlı bilgilere erişmek için izlenmesi gereken adımlar şunlardır(Han vd., 2011).

1. Veri Seçimi: İncelemeye alınacak verilerin, veri tabanından veya başka bir kaynaktan alınması işlemidir.
2. Veri Entegrasyonu: Çeşitli kaynaklardan gelen verilerin birleştirilerek tek bir yapı haline getirilmesini ifade eder.
3. Veri Temizleme: Veri içindeki çelişkilerin ve gürültülerin temizlenerek düzeltilmesini sağlayan bir süreçtir.
4. Veri Dönüşümü: Verinin özetlenerek veya yeniden yapılandırılarak kullanılabilir hale getirilmesini ifade eder.
5. Veri Madenciliği: Akıllı yöntemlerin kullanılarak veri içindeki desenlerin ortaya çıkarılmasını sağlayan bir süreçtir.
6. Desen Değerlendirmesi: Önemli desenlerin, özel ölçümler kullanılarak tanımlanması ve analiz edilmesini içerir.
7. Bilgi Sunumu: Elde edilen bilgilerin, görselleştirme ve bilgi sunma teknikleriyle kullanıcılara aktarılmasını ifade eder.

Veri setimiz, belirtilen süreçleri göz önünde bulundurarak duygu analizi yöntemlerine uygun hale getirilmiştir.

3.2.Metot

3.2.1.Kelime torbası (bag of words)

Kelime Torbası modeli (bag of words - BOW), metinlerden özellik çıkarma yöntemlerinden biridir ve doğal dil işleme, metin madenciliği gibi alanlarda yaygın olarak kullanılır. Bu model, metinlerin yapısal özelliklerinden çok, içerdikleri kelimelerin varlığına ve frekansına odaklanarak metinlerin temsiliyetini sağlar. Modelin temel mantığı, bir dokümandaki her kelimenin, metnin anlamını taşıyan bağımsız bir birim olarak değerlendirilmesidir. Bu nedenle, kelimelerin metin içindeki dizilim sırası veya bağlamı dikkate alınmaz. Kelime çantası modeliyle, metin içerisindeki kelimelerin

frekanslarını veya bulunma durumlarını analiz ederek bir kelime haznesi oluşturulur. Bu hazne, modelin "kelime çantası" olarak adlandırılan bileşenidir. Kelime çantasındaki her bir kelime, metinlerin özellik vektöründe bir boyut oluşturur ve her metin, bu boyutlarda temsil edilir. Örneğin, "kitap" kelimesi kelime haznesinde bir boyut olarak yer alıyorsa, bir metindeki "kitap" kelimesinin frekansı, o metnin özellik vektöründeki karşılık gelen değeri belirler. Bu modelin en önemli avantajlarından biri, uygulamasının basit ve hesaplamalarının hızlı olmasıdır. Ayrıca, büyük metin veri kümelerinin işlenmesi ve analiz edilmesi için uygun bir yapıya sahiptir. Ancak, kelime çantası modelinin dezavantajları da vardır. Örneğin, kelimelerin dizilim sırasını ve bağlamı göz ardı ettiği için, metinlerin bağlamsal anlamını tam olarak yansıtamaz. Bu durum, aynı kelimelerin farklı bağlamlarda farklı anlamlar taşıdığı durumlarda modelin performansını sınırlayabilir. Sonuç olarak, kelime çantası modeli, metinlerin yüzeysel özelliklerini çıkararak hızlı ve basit bir şekilde analiz yapma olanağı sağlar. Metin sınıflandırma, bilgi erişimi ve belge kümeleme gibi görevlerde yaygın olarak kullanılır. Ancak, daha derin dil bilgisel analizler için ek yöntemlerle desteklenmesi gerekebilir (Görentaş, 2023).

Kelimelerin veri kümesi içindeki kullanım sıklığını tespit etmeye olanak tanır. Bu yaklaşım, veri kümesindeki belirli kelimelerin ne kadar yaygın olduğunu belirleyerek, veri kümesini en iyi temsil eden kelimeleri ortaya çıkarır. Böylelikle, metin analizi ve doğal dil işleme süreçlerinde, veri kümesinin genel özelliklerini ve içeriğini anlamak için hangi kelimelerin en önemli olduğunu saptama imkanı sunar. Kelimelerin frekanslarının hesaplanması, metnin temsiliyetini artırarak, metin madenciliği ve bilgi erişimi gibi alanlarda daha etkili sonuçlar elde edilmesini sağlar. Bu süreç, özellikle büyük ve karmaşık veri kümeleri üzerinde çalışırken, en belirgin ve anlamlı kelimeleri öne çıkarmaya yardımcı olur (Bayram,2023).

Kelime torbası (bag-of-words) yöntemi, bir metindeki kelimelerin veya kelime gruplarının varlığını veya tekrar sayısını dikkate alan bir metin analiz tekniğidir. Metinde kelimelerin sadece var olup olmadığına bakıldığında, bu yaklaşım ikili özellik seçimi (binary feature selection) olarak adlandırılırken, kelimelerin tekrar sayısı dikkate alındığında ise terim frekans ağırlığı (term frequency weighting) yöntemi olarak tanımlanır. Metin içerisindeki her bir kelime tek başına analiz ediliyorsa unigram, ikili gruplar halinde analiz ediliyorsa bigram ve n sayıda kelime grubu halinde analiz ediliyorsa n-gram modeli olarak adlandırılmaktadır. Bu yöntemi uygularken, öncelikle veri setindeki tüm benzersiz kelimelerden bir sözlük oluşturulmaktadır. Burada dikkat

edilmesi gereken nokta, metinlerin ön işleme aşamalarından geçirilmiş ve temizlenmiş olmasıdır. Aksi halde, örneğin büyük harflerle yazılmış bir kelimenin küçük harflerle yazılmış hali farklı olarak değerlendirilecek ve sözlükte iki ayrı giriş olarak yer alacaktır. Sözlük oluşturulduktan sonra, her bir kelime için ayrı bir sütun açılarak bir öznitelik matrisi oluşturulur. Bu matrisin satırları, veri setindeki her bir metin kaydındaki kelimelerin kaç kez tekrarlandığını gösteren sayılarla doldurulur. Öznitelik matrisinin oluşturulması, kelime torbası modelinin temel adımlarından biridir ve bu matris, metin madenciliği ve doğal dil işleme uygulamalarında önemli bir rol oynar. Matrisin satırları, metin belgelerini; sütunları ise sözlükteki kelimeleri temsil eder. Bu sayede, her metin belgesi, içindeki kelimelerin frekanslarına göre sayısal bir formata dönüştürülmüş olur. Bu sayısal temsil, makine öğrenmesi algoritmalarının metin verileri üzerinde çalışmasına olanak tanır (Tosun, 2021).

Kelime torbası (Bag of Words) yöntemi, doğal dil işleme ve bilgi çıkarımı alanlarında metinlerin sayısal temsili ve özelliklerinin belirlenmesi amacıyla kullanılan bir tekniktir. Bu yaklaşımda, bir metin Öğe ve Kayaalp'in (2021) yaptığı çalışmaya göre ve Çizelge 3.3'te gösterildiği gibi, içerdiği tüm kelimelerin bir kümesi olarak modellenir. Duygu analizi çalışmalarında, kelime torbası yöntemi, metinde duygusal ifadeleri belirlemede yararlı olabilecek kelimelerin listesini oluşturmayı amaçlar. Bu teknik, metinlerdeki duygusal içeriği tespit etmek ve analiz etmek için etkili bir yöntem olarak kabul edilir.

Çizelge 3.3. Bag of words yöntemi ile yorumların temsil edilmesi (Öğe ve Kayaalp,2021).

	1	2	3	4	5	6	7	8	9	10	11	Yorum
	This	movie	is	very	scary	and	long	not	slow	spooky	good	Uzunluğu
Yorum 1	1	1	1	1	1	1	1	0	0	0	0	7
Yorum 2	1	1	2	0	0	1	1	0	1	0	0	7
Yorum 3	1	1	1	0	0	0	1	0	0	1	1	6

3.2.2. Tf – ıdf yöntemler

TF-IDF (Term Frequency-Inverse Document Frequency), bir kelimenin belirli bir belgedeki önemini ölçmek için kullanılan bir yöntemdir. Bu yöntem, kelimenin o belge içerisindeki frekansı (TF) ile kelimenin yer aldığı belgelerin toplam belgeler içerisindeki oranının tersi (IDF) ile çarpılarak hesaplanır. Temelde, TF-IDF, belirli bir

kelimenin bir belgede ne kadar sık geçtiğini ve bu kelimenin genel veri setindeki nadirliğini dikkate alarak bir ağırlık değeri atar. Bu hesaplama, bir kelimenin belirli bir belge ile ne derece ilgili olduğunu belirlemeye yarar. Özellikle, az sayıda belgede sıkça rastlanan kelimeler, tüm belgeler genelinde yaygın olan kelimelere göre daha yüksek TF-IDF değerlerine sahip olma eğilimindedir. Bu, kelimenin özgüllüğünü ve bilgi değerini artırır. Özetle, TF-IDF, kelimenin belirli bir belgede ne kadar önemli olduğunu, bu kelimenin veri setindeki nadirliği ile birlikte değerlendirerek ortaya koyar. Bu yöntem, bilgi çıkarımı ve metin madenciliği gibi alanlarda yaygın olarak kullanılır ve belgeler arasındaki benzerliklerin belirlenmesinde de önemli bir rol oynar (Çelik ve Koç., 2021).

TF-IDF skoru aşağıdaki formül ile hesaplanır:

$$W_{i,j} = tf_{i,j} \times \log\left(N / dfi\right) \quad (3.1)$$

Burada,

- $tf_{i,j}$:i kelimesinin j belgesinde geçme sıklığını ifade eder.
- df_i :i kelimesini içeren belge sayısını gösterir.
- N, toplam belge sayısını belirtir.

TF-IDF ağırlıklandırma yöntemi, bir belgedeki kelimelerin önemini belirlemek için kullanılır ve bu önem, kelimenin frekansı ile ilişkilidir. Terim frekansı (TF) değeri, bir kelimenin bir veri seti içerisindeki tekrar sayısını ifade ederken, ters belge frekansı (IDF), kelimenin tüm belgeler içinde ne kadar nadir bulunduğunu ölçer. TF, kelimenin belirli bir belgede kaç kez geçtiğini hesaplarken, IDF, kelimenin genel belge koleksiyonunda ne kadar yaygın olduğunu değerlendirir. Eğer bir kelime, tüm eğitim belgeleri arasında sadece belirli bir belgede bulunuyorsa, bu durum o kelimenin o belge için ayırt edici ve belirleyici bir özellik olduğunu gösterir. Bu sayede TF-IDF, bir kelimenin bir belge içerisindeki önemini ve o belgenin diğer belgelerden farklılığını ortaya koyar (Göker ve Tekedere,2017).

Terim frekansı - Ters belge frekansı (TF-IDF), bir belgedeki kelimenin önem derecesini belirleyen bir ölçüttür. Bir kelimenin bir belge içinde daha sık görülmesi, o kelimenin TF-IDF değerinin yükselmesine neden olur (Kaynar vd., 2016).

Öğe ve Kayaalp'e göre (2021) TF-IDF, bir terimin bir metin içindeki geçiş sıklığını (Term Frequency, TF) ve bu terimin genel olarak ne kadar önemli olduğunu (Inverse Document Frequency, IDF) dikkate alan bir özellik çıkarım yöntemidir. Her kelimenin veya terimin kendine özgü TF ve IDF değerleri bulunur. Bir terimin TF ve IDF değerlerinin çarpımı, o terimin TFIDF ağırlığını belirler. Basitçe ifade etmek

gerekirse, bir terimin TFIDF puanı ne kadar yüksekse, o terim metinlerde o kadar nadir bulunur; bu puan ne kadar düşükse, terim o kadar yaygındır. Çizelge 3.4'te buna uygun bir örnek gösterilmiştir.

Çizelge 3.4.Örnek yorumlar için tf-ıdf puan hesaplanması (Öğe ve Kayaalp,2021).

Kelime	Yorum 1	Yorum 2	Yorum 3	TF(Yorum 1)	TF(Yorum 2)	TF(Yorum 3)	IDF	TF-IDF(Yorum 1)	TF-IDF(Yorum 2)	TF-IDF(Yorum 3)
This	1	1	1	1/7	1/8	1/6	0.00	0.00	0.00	0.00
movie	1	1	1	1/7	1/8	1/6	0.00	0.00	0.00	0.00
is	1	2	1	1/7	1/4	1/6	0.00	0.00	0.00	0.00
very	1	0	0	1/7	0	0	0.48	0.068	0.00	0.00
scary	1	1	0	1/7	1/8	0	0.18	0.025	0.022	0.00
and	1	1	1	1/7	1/8	1/6	0.00	0.00	0.00	0.00
long	1	0	0	1/7	0	0	0.48	0.068	0.00	0.00
not	0	1	0	0	1/8	0	0.48	0.00	0.060	0.00
slow	0	1	0	0	1/8	0	0.48	0.00	0.060	0.00
spooky	0	0	1	0	0	1/6	0.48	0.00	0.00	0.080
good	0	0	1	0	0	1/6	0.48	0.00	0.00	0.080

Kelime torbası (bag-of-words) yöntemi, metin analizi sırasında kelimelerin yalnızca bireysel kayıtlar içerisindeki frekanslarını dikkate alırken, TF-IDF yöntemi, veri setinin tamamını göz önünde bulundurarak hesaplamalar yapar. Özellik çıkarımı sürecinde temel amaç, algoritmalar için uygun giriş verilerini hazırlamaktır. Bu bağlamda, veri setinde çok sık tekrarlanan bir özelliğin bilgi değeri düşük olacağından, bu tür özelliklerin seçimi ve kullanımı minimize edilmelidir. TF-IDF yöntemi, kelimelerin genel önemini belirlerken, sıkça rastlanan kelimelerin etkisini azaltarak daha ayırt edici ve anlamlı özellikler elde edilmesini sağlar. Bu şekilde, algoritmaların performansı ve doğruluğu artırılır (Bonaccorso, 2018).

3.2.3. Sınıflandırma algoritmaları

3.2.3.1 K- en yakın komşu algoritması

K En Yakın Komşu (K Nearest Neighbour, k-NN) algoritması, ilk olarak 1950'lerin başında geliştirilmiştir. Bu yöntem, büyük eğitim setleri verildiğinde öğrenme sürecinde önemli ölçüde zaman kaybına yol açmaktadır. Bu nedenle, algoritma, yeterli bilgi işlem gücü kullanılabilir hale gelene kadar geniş çapta kullanılmamıştır(Han vd., 2011). K-En Yakın Komşu (k-NN) algoritması, öznitelik uzayında bulunan eğitim örneklerine dayanarak nesneleri sınıflandıran en basit örüntü tanıma yöntemlerinden biridir. Bu algoritma, verilen k değeri kadar en yakın komşunun sınıfına göre sınıflandırma yapar. k-NN algoritması, bir vektörün sınıflandırılmasını,

sınıfı bilinen vektörler kullanılarak gerçekleştirir. Test edilecek olan örnek, eğitim kümesindeki her bir örnek ile ayrı ayrı değerlendirilir. Örnekler arasındaki uzaklık genellikle Öklid uzaklık metriği ile hesaplanır. Öklid uzaklık formülü, n boyutlu iki nokta arasındaki uzaklığı hesaplar. Test edilecek örneğin sınıfını belirlemek için eğitim kümesindeki her bir örnek ile karşılaştırılır ve en yakın k adet örnek seçilir. Seçilen bu örneklerin hangi sınıfa ait olduğu incelenir ve test edilecek olan örnek bu sınıfa ait olarak sınıflandırılır. Bu şekilde, k-NN algoritması örnekler arasındaki benzerlikleri kullanarak sınıflandırma işlemi gerçekleştirir (Karakoyun ve Hacıbeyoğlu,2014).

Bu sınıflandırma algoritması, sınıflandırılacak olan imza örüntüsünün, daha önce sınıflandırılmış k tanesi benzer örüntüye bakılarak belirlenmesi prensibine dayanır. Bu yaklaşım, belirli sınıflara ait örneklerin olduğu bir örnek kümesinden yararlanarak, yeni bir örneğin hangi sınıfa ait olduğunu belirlemek için kullanılır. Bu yöntem, gözlemlerin sınıflandırılmasında doğruluk sağlayarak, önceden belirlenmiş sınıflar arasındaki benzerlikleri değerlendirir (Erten,2012).

K-En-Yakın-Komşu (k-NN) Algoritması, bir belgenin sınıflandırılması için kullanılan bir örüntü tanıma yöntemidir. Bu algoritma, sınıflandırmak istenen belgeye en yakın olan benzer belgelerin bulunmasına dayanır. En Yakın Komşu (k-NN) araması, bir belgenin benzerlik derecesini belirlemek için kullanılır; yani, bir belgenin kendisine en yakın olan diğer belgeleri bulmaya yarar. Yeni bir belgenin sınıflandırılması için, eğitim setinde bulunan belgeler kontrol edilir ve bu belgelerle yeni belge arasındaki benzerlik ölçülür. K-NN algoritması, bu benzerlik ölçüsüne dayanarak yeni belgenin sınıfını belirler. Yeni belge, kendisine en yakın olan k sayıdaki (önceden eğitilmiş) belgelerin sınıflarının çoğunluğuna göre sınıflandırılır. Burada k, dikkate alınacak en yakın komşu sayısını ifade eder ve sınıflandırma doğruluğunu etkileyen bir parametredir. Bu şekilde, k-NN algoritması, benzerlik esasına dayanarak belgelerin sınıflandırılmasını gerçekleştirir (Kaşıkçı ve Gökçen,2014).

K-En Yakın Komşu, bir verinin sınıflandırılması veya regresyonu için kullanılan bir algoritmadır, bu algoritma, bir verinin kendisine en yakın komşularının kimliğini kullanır. K, komşu sayısını belirten bir parametre olarak görev yapar ve doğru bir sınıflandırma ya da regresyon sonucu elde etmek için uygun bir K değerinin seçilmesi kritiktir. Bu nedenle, K değerinin doğru bir şekilde belirlenmesi, algoritmanın başarısını etkileyen önemli bir faktördür (Şahin vd., 2023).

Sınıflandırma sürecinde ana hedeflerden biri, nesnelerin özelliklerine göre sınıflandırılmasıdır. Bu amaca ulaşmak için farklı sınıflandırma algoritmaları

mevcuttur. Ancak, en yaygın kullanılanı olan K-En Yakın Komşu (KNN) Algoritması, bir özelliğin seçilerek diğer özelliklerle olan benzerliklerine dayalı olarak sınıflandırma yapar. Bu algoritma, nesneler arasındaki mesafelerin hesaplanmasında etkilidir. KNN Algoritması, yeni bir nesnenin sınıflandırılması için k değerine bakar. Bu değer genellikle tek sayı olacak şekilde seçilir, böylelikle sınıflandırma işleminde denge sağlanır. Yeni verinin mevcut veriyle karşılaştırılması sırasında, farklı metrikler kullanılabilir. Örneğin, Manhattan veya Öklid gibi uzaklık hesaplama yöntemleri, nesneler arasındaki benzerliklerin belirlenmesine yardımcı olur. Bu yöntemler, sınıflandırma sürecinde doğruluğu artırır ve algoritmanın etkinliğini güçlendirir (Çiçek ve Arslan,2020).

Hacıbeyoğlu ve arkadaşları(2023) ,K-NN komşu algoritması ile ilgili bir çalışma yapmışlardır. K-En Yakın Komşu algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılan bir denetimli makine öğrenmesi tekniğidir. KNN, özellikle veri kümelerinin küçük boyutlu olduğu ve sınıflar arasındaki ayrımın belirgin olduğu durumlarda etkili bir performans sergiler. Algoritma, yeni bir örneği sınıflandırmak için, bu örnek ile eğitim kümesindeki diğer örnekler arasındaki uzaklıkları hesaplar ve en yakın k komşunun sınıf değerlerini kullanarak sınıflandırma işlemini gerçekleştirir. Bu mesafe hesaplamaları için genellikle Öklid, Manhattan ve Minkowski uzaklık ölçüm formülleri tercih edilmektedir. Aşağıda formüller kısaca belirtilmiştir:

$$\left(\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\right) \tag{3.2}$$

$$\left(\sum_{i=1}^{n}|x_i - y_i|\right) \tag{3.3}$$

$$\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \tag{3.4}$$

Yukarıdaki denklemlerde X, sınıflandırılması gereken yeni veriyi, Y ise veri kümesindeki bir örneği ifade etmektedir. Hem X hem de Y, veri kümesindeki özellik sayısına karşılık gelen elemanları içeren vektörler olarak saklanmaktadır.

$$X=\{x1,x2,\ldots,xn\} \tag{3.5}$$

$$Y=\{y1,y2,\ldots yn\} \tag{3.6}$$

KNN algoritması, regresyon problemlerinde, önce uzaklık ölçütünü belirler ve ardından veri noktasına en yakın komşuları hesaplar. Bu komşuların hedef değerlerinin ortalamasını hesaplayarak tahminlerde bulunur. Bu süreçte, k sayısının optimum değerde seçilmesi kritik bir adımdır. Optimum k değerini belirlemek için, farklı k değerleri üzerinde testler gerçekleştirilir.

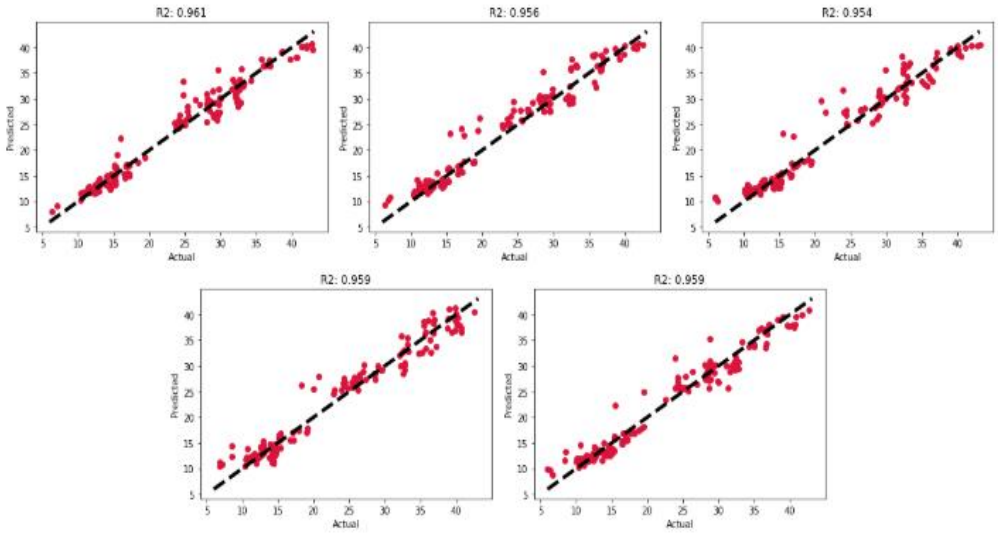
Regresyon analizi, bağımsız değişkenlerin belirli aralıklardaki değerlerine dayanarak bağımlı değişkenin tahmin edilmesini sağlar. Bu analiz, veri arasındaki ilişkiyi anlamak ve gelecekteki değerleri öngörmek için kullanılır. Regresyon analizinde çeşitli teknikler ve yaklaşımlar bulunmaktadır, bu yöntemler verinin özelliklerine ve amaçlarına bağlı olarak değişiklik gösterebilir.

Doğrusal regresyon, gözetimli öğrenme içinde temel bir araçtır ve genellikle basit ve etkili olarak kabul edilir. Bu yöntem, tek bir bağımsız değişkenin bulunduğu durumlarda

$y = ax + b$ denklemiyle ilişkiyi açıklamaya çalışır. Ancak günümüzde, makine öğrenmesi problemlerinde genellikle çok sayıda bağımsız değişken bulunmaktadır. Bu durumda, çok değişkenli doğrusal regresyon yöntemi, bağımsız değişkenler arasındaki ilişkiyi modellemeyi ve bu değişkenlerin bağımlı değişkene olan etkilerini anlamayı amaçlar. Çizelge 3.5'te sonuçların değişkenlik gösterdiği gözlenmektedir.

Çizelge 3.5. Isıtma yükü sınıf değeri için deneysel çalışma sonuçları(Hacıbeyoğlu vd.,2023).

Algoritma	MSE	MAPE	MAE	R-Kare
Lineer Regresyon	8,654	0,099	2,09	0,914
KNN Regresyon (k=3)	4,882	0,092	1,717	0,952
KNN Regresyon (k=5)	**4,301**	**0,084**	**1,529**	**0,958**
KNN Regresyon (k=7)	5,553	0,096	1,792	0,945
KNN Regresyon (k=9)	6,734	0,108	2,003	0,934
KNN Regresyon (k=11)	6,94	0,109	2,009	0,932



Şekil 3.1. Isıtma yükü sınıf değeri için k=5 değeri ile 5 çapraz doğrulama tahmin hatalarının şekilsel görünümü (Hacıbeyoğlu vd., 2023).

Şekil 3.1'de verilen sonuçlara bakıldığında, ısıtma yükü ve soğutma yükü tahminleri için en üstün performansın KNN algoritmasının k=5 değeri ile elde edildiği gözlemlenmektedir.

3.2.3.2. Karar destek makinası (support vector machine)

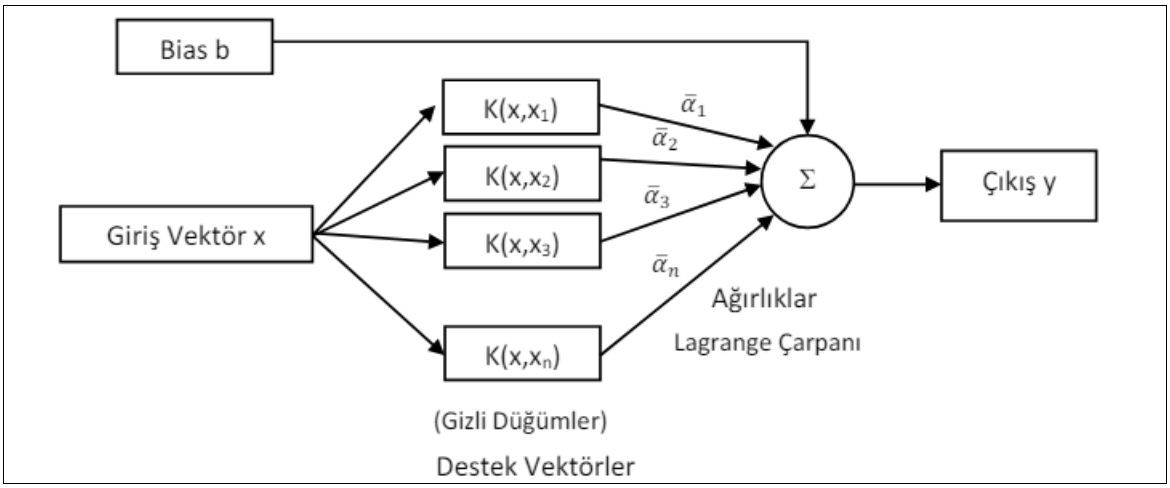
Destek Vektör Makineleri (SVM), veri sınıflandırma ve regresyonunda kullanılan güçlü bir algoritmadır. Bu algoritma, veri kümesini farklı sınıflara ayırmak için karar sınırlarını belirler. Karar sınırları, veri noktalarını en iyi şekilde ayırmak için en yakın iki nokta arasındaki eşit uzaklıkta çizilen vektörler tarafından belirlenir. Bu vektörler, destek vektörleri olarak adlandırılır ve modelin temelini oluşturur. Destek vektörleri, genellikle lineer, radyal veya sigmoid gibi çeşitli şekillerde oluşturulabilir. SVM modelleri, sınıflandırma ve regresyon problemleri için kullanılabilir. Bir dizi eğitim verisi sağlandığında, SVM modeli bu verilere dayanarak bir karar sınırı oluşturur. Bu karar sınırı, veri noktalarını sınıflandırmak ve yeni veri noktalarını tahmin etmek için kullanılır. SVM, veri noktalarını sınıflandırmak için doğru bir şekilde öğrenilmiş bir model sağladığı için genellikle başarılı sonuçlar verir (Cesur ve Efe,2023).

Destek Vektör Makineleri (SVM), istatistiksel öğrenme alanında önemli bir yer edinmiş ve 1960'lara dayanan köklere sahip bir algoritmadır. Vapnik ve Chervonenkis tarafından tanımlanmış olan SVM, birçok problemi başarılı bir şekilde çözmüş ve

özellikle sınıflandırma ve regresyon gibi alanlarda etkinliğini kanıtlamıştır. SVM'nin temel prensibi, verileri doğrusal olarak en iyi şekilde ayırmaktır. Ancak, doğrusal olarak ayrıştırılamayan veriler için SVM, dönüşüm teknikleri kullanarak farklı bir boyuta taşınma yeteneğine sahiptir. SVM'nin dikkate değer bir özelliği, öğrenme sürecinde görülmemiş verileri de sınıflandırabilmesidir. Bu genelleme yeteneği, SVM'nin diğer yöntemlerden ayrılmasını sağlar. Son yıllarda, SVM'nin popülerliği artmış ve yüz tanıma, veri madenciliği, biyoloji, gen analizi ve örüntü tanıma gibi çeşitli alanlarda kullanımı yaygınlaşmıştır. Ancak, SVM'nin etkili bir şekilde kullanılabilmesi için doğru özelliklerin seçilmesi gereklidir. Bu, doğru sonuçların elde edilmesi için kritik öneme sahiptir (Çiçek ve Arslan ,2020).

Destek Vektör Makineleri (SVM) algoritması, N-boyutlu bir alanda (N, özellik sayısı) veri noktalarını sınıflandıran bir hiperdüzlem bulmayı hedefler. İki sınıf arasında bir ayrım yapmak için kullanılabilecek birçok farklı hiperdüzlem vardır. Amacımız, maksimum marj (sınır) sağlayan, yani her iki sınıf arasındaki veri noktalarının en uzak mesafesini içeren bir düzlem bulmaktır. Bu maksimum marj, gelecekteki veri noktalarının daha güvenilir bir şekilde sınıflandırılabilmesi için bir güvence sağlar. Hiperdüzlemler, veri noktalarını sınıflandırmaya yardımcı olan karar sınırlarıdır. Hiperdüzlemin her iki tarafına düşen veri noktaları farklı sınıflarla ilişkilendirilir. Ayrıca, hiperdüzlemin boyutu özellik sayısına bağlıdır: Giriş özelliklerinin sayısı 2 ise, hiperdüzlem sadece bir çizgi oluşturur; 3 ise, iki boyutlu bir düzlem oluşturur. Ancak, özellik sayısı 3'ü aştığında, hiperdüzlemi hayal etmek zorlaşır (İlhan ve Sağaltıcı,2020).

Destek Vektör Makineleri (SVM), en uygun hiperdüzlemi bulmak için sınıflar arasındaki mesafeyi maksimize etmek amacıyla kullanılır. Şekil 3.2’te de gösterildiği gibi bu hiperdüzlem, verileri sınıflandırmak için kullanılan bir fonksiyondur. İki boyutta, çizgiler; üç boyutta, düzlemler; ve yüksek boyutlu uzaylarda ise hiperdüzlemler, öğeleri kategorize etmek için kullanılır. SVM tarafından belirlenen hiperdüzlem, iki sınıfı net bir şekilde ayırır ve en dış sınıftan gelen veri öğelerinden mümkün olduğunca uzak konumlanır. Destek vektörler, bu hiperdüzleme en yakın olan ve sınıflar arasındaki marjı belirleyen veri öğeleridir (Kına ve Biçek,2023).



Şekil 3.2.SVM genel işleyiş yapısı (Deka, 2014).

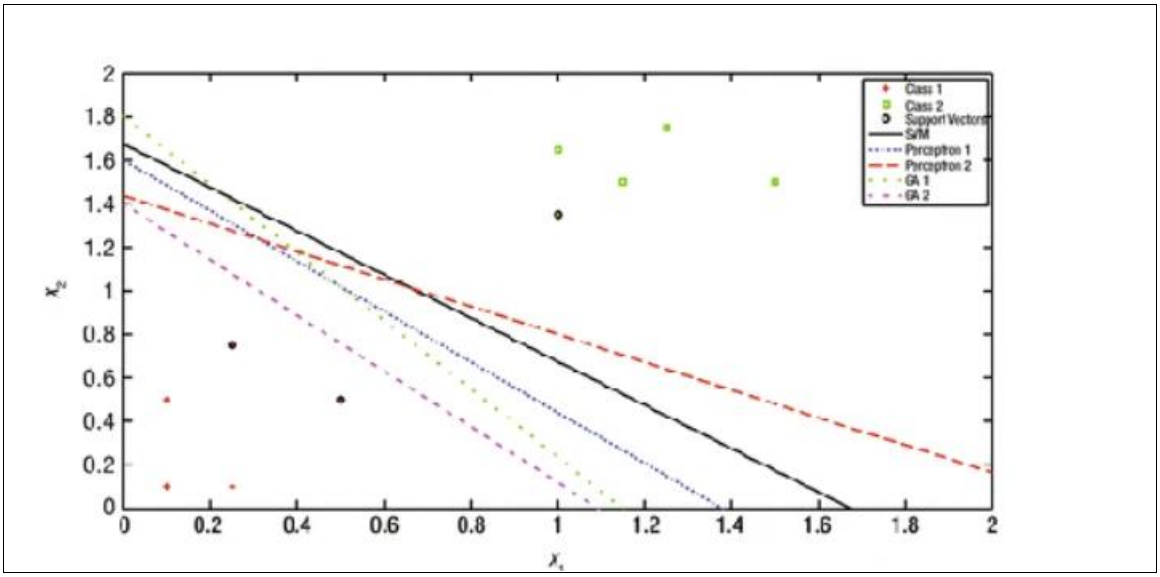
SVM' nin cazibesini oluşturan özelliklerden bazıları şunlarla özetlenebilir (Award vd., 2015):

SVM, seyreklik ilkesine dayanır. Diğer yöntemler gibi, modelin öğrenilmiş parametreleri eğitim aşamasında tüm veri setinin mevcut olmasını gerektirir, ancak model parametreleri belirlendikten sonra SVM, yalnızca destek vektörleri olarak adlandırılan bir alt kümeden gelecekteki tahminler için bağımsızdır. Destek vektörleri, hiperdüzlemlerin kenar boşluklarını tanımlar ve lagrange çarpanları aracılığıyla belirlenir. SVM'nin sınıflandırma karmaşıklığı, giriş alanının boyutundan ziyade destek vektörlerinin sayısına bağlıdır, bu da veri boyutu ve sınıf ayrılabilirliği tarafından belirlenen veri karmaşıklığına bağlı olarak değişir. Destek vektörlerinin sayısı, eğitim veri setinin yarısından fazla olamaz, ancak pratikte bu sınıra nadiren ulaşılır. Matematiksel olarak, SVM modeli, destek vektörlerinin ağırlıklı toplamını içerir, bu da test ve depolama gereksinimlerini azaltarak SVM'nin parametrik tekniklerle aynı avantajlara sahip olmasını sağlar.

SVM, çekirdek tabanlı bir yöntemdir. Diğer makine öğrenme tekniklerinde olduğu gibi, SVM optimumları analitik olarak bulunur. Makine öğrenimini bir dışbükey optimizasyon problemi olarak çözmeden önce, verileri daha yüksek boyutlu bir alana eşlemek için çekirdek hilesini kullanır. Gerçek veriler genellikle orijinal giriş uzayında doğrusal olarak ayrılabilir değildir, bu nedenle SVM, doğrusal olmayan ayırıcıları öğrenmek yerine önceden tanımlanmış çekirdek işlevlerini kullanarak verileri daha yüksek boyutlu bir alana eşler. Bu, doğrusal bir ayırıcı ile sınıfları doğru şekilde ayırmayı mümkün kılar. SVM optimizasyon aşamasında, haritalanan alanda yalnızca

doğrusal bir ayırt edici yüzeyin öğrenilmesi gerekir. Çekirdek fonksiyonunun seçimi ve ayarları, SVM'nin başarısı için kritiktir.

SVM, maksimum kenar boşluğu ayırıcısı olarak işlev görür. Diğer ayırt edici makine öğrenme tekniklerinden farklı olarak, SVM optimizasyon problemine bir kısıtlama getirir: hiperdüzlemin sınıflar arasında eşit ve maksimum uzaklıkta konumlandırılması gerekir. Bu, daha iyi genelleştirilebilecek bir hiperdüzlem bulmaya yönelik optimizasyon adımını zorlar. Eğitim popülasyonunun bir örneği üzerinde yapılan eğitim, henüz görülmemiş örnekler üzerindeki tahminleri etkileyebilir.



Şekil 3.3.SVM, algılayıcı ve GA hiperdüzlemleri için iki boyutlu, iki sınıflı çizim (Award vd., 2015).

Award ve arkadaşlarının (2015) oluşturduğu Şekil 3.3'te belirtildiği gibi iki boyutlu veri seti üzerinde SVM, algılayıcı ve GA sınıflandırıcılarının farklı hiperdüzlemlerini sergilemektedir. Dairelerle çevrili noktalar, destek vektörlerini gösterirken, her bir sınıflandırıcıya karşılık gelen hiperdüzlemler farklı renklerle vurgulanmıştır, böylece görselleştirme amacına uygun bir açıklık sağlanmıştır.

3.2.3.3. Naive bayes

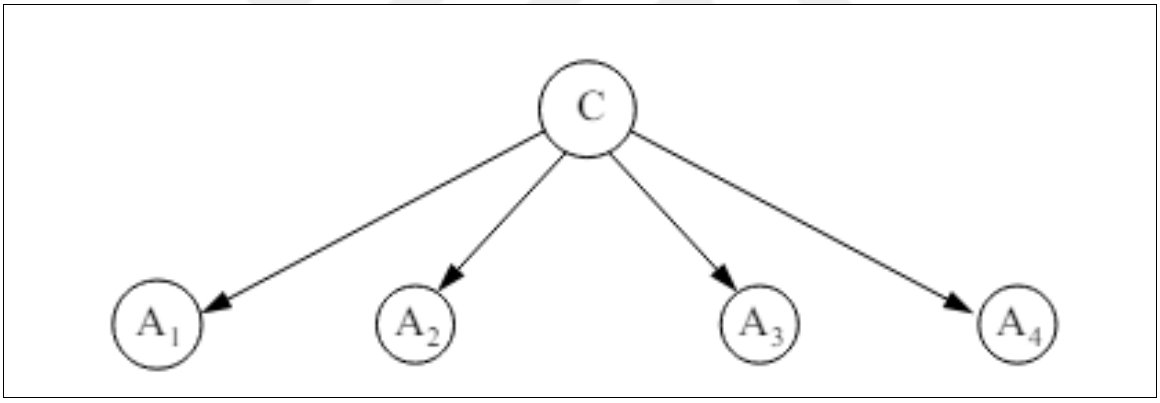
Naive Bayes sınıflandırma algoritması, Matematikçi Thomas Bayes'in ismini taşıyan ve veri sınıflandırma veya kategorize etme işlemlerinde kullanılan bir metottur. Bu algoritma, verilerin belli sınıflara veya kategorilere atanmasını sağlamak için olasılık temelli bir dizi hesaplama yöntemi kullanır. Naive Bayes sınıflandırmasında, öğrenme

için belirli bir sınıfa ait veriler sunulmalıdır. Bu veriler üzerinde yapılan olasılık hesaplamalarıyla, sisteme sunulan yeni test verilerinin hangi kategoriye ait olduğu tahmin edilmeye çalışılır. Öğretilmiş veri miktarının artması, test verilerinin doğru sınıflandırılma başarısını artırır. Naive Bayes sınıflandırma yöntemi, geniş bir uygulama alanına sahiptir; ancak, asıl önemli olan, veri tipi ve ilişkileri arasında nasıl bir bağlantı kurulduğudur. Bu algoritmanın matematiksel gösterimi aşağıdaki gibidir (Taşcı ve Şamlı,2020):

$$P(C_i \mid X) = (P(X \mid C_i) * P(Ci)) / P(X) \quad (3.7)$$

Burada:

- $P(C_i|X)$ X olayının meydana geldiği durumda $CiCi$ olayının gerçekleşme olasılığıdır.
- $P(X|C_i)$, C_i olayının gerçekleştiği durumda XX olayının gerçekleşme olasılığıdır.
- $P(Ci)$, C_i olayının öncül olasılığıdır.
- $P(X)$, X olayının gerçekleşme olasılığıdır.



Şekil 3.4. Naive bayes örneği (Zhang, 2004).

Naive Bayes, sınıf değişkeninin değeri verildiğinde tüm niteliklerin birbirinden bağımsız olduğu temel bir Bayes ağıdır. Bu duruma koşullu bağımsızlık denir ve gerçek dünya uygulamalarında çoğu zaman nadiren gerçekleştiği bilinmektedir. Saf Bayes'in bu kısıtlamalarını aşmak için basit bir yaklaşım, nitelikler arasındaki ilişkileri daha doğru bir şekilde temsil etmek için modeli genişletmektir. Genişletilmiş saf Bayes ağı veya kısaca Artırılmış Saf Bayes (ANB), sınıf düğümünün tüm öznitelik düğümlerine doğrudan bağlandığı ve öznitelik düğümleri arasında ilişkilerin bulunduğu bir genişletilmiş saf Bayes modelidir. Şekil de bir ANB örneği gösterilmektedir. Olasılık

açısından bakıldığında, bir ANB, G ile temsil edilen ortak bir olasılık dağılımını ifade eder (Zhang,2004).

Naive Bayes, her olası çıktı için bir olasılık atayarak tahminlerde bulunur. Bu olasılıklar, sonrasında bir tahminde birleştirilir. Kategorik problemlerde, sıfır kayıp durumunda, en uygun tahmin temel dağılımın modu olarak belirlenir. Ancak, sayısal problemlerde, optimal tahmin kayıp fonksiyonuna bağlı olarak ortalama veya medyan olabilir. Bu iki istatistik, temel dağılımın değişmesiyle neredeyse her zaman birlikte değişir, hatta küçük bir değişiklik bile bu durumu etkileyebilir. Bu nedenle, sayısal tahminlerde Naive Bayes, sınıflandırmaya odaklanırken yanlış olasılık tahminlerine karşı kullanılır. Bu çalışma, regresyon için Naive Bayes'in kullanımını açıklar ve yapay bir veri setinde bu yaklaşımın tercih edilen bir yöntem olduğunu gösterir. Daha sonra, çeşitli pratik öğrenme problemlerindeki performansını özetler. Standart veri kümeleri üzerindeki sınıflandırma doğruluğunun, regresyon bağlamında geçerli olmadığı sonucuna varılmıştır (Frank vd., 2000).

Naive Bayes, Bayesian çerçevesinin temel bir türüdür ve makine öğrenimi ile veri madenciliğinde kullanılan etkili ve verimli öğrenme algoritmalarına sahiptir. Bu model, istatistiksel yöntemlere dayanan bir sınıflandırma tekniği olan Bayesian ağ sınıflandırıcısını temel alır. Bayesian ağ sınıflandırıcısı, olayların öncül olasılıklarını ardıl olasılıklarla ustaca bağdaştırır. Naive Bayes algoritması, özellikle büyük veri kümeleriyle çalışırken ikili ve çoklu sınıf sınıflandırma problemlerini çözmek için uygundur. Metin madenciliği gibi alanlarda, kelimelerin ve sınıfların ortak olasılıklarını hesaplamak amacıyla yaygın olarak kullanılır. Bu algoritma, denetimli öğrenme kapsamında basit ve kullanıcı dostu bir sınıflandırma yöntemi olarak verilerin etiketlenmesi ve sınıflandırılması için tercih edilir. Bayesian teoremi kullanılarak, her bir kriterin sonuç üzerindeki etkileri olasılık değerleriyle hesaplanır ve böylece verilerin hangi sınıfa ait olduğu belirlenir (Gerdan vd., 2020).

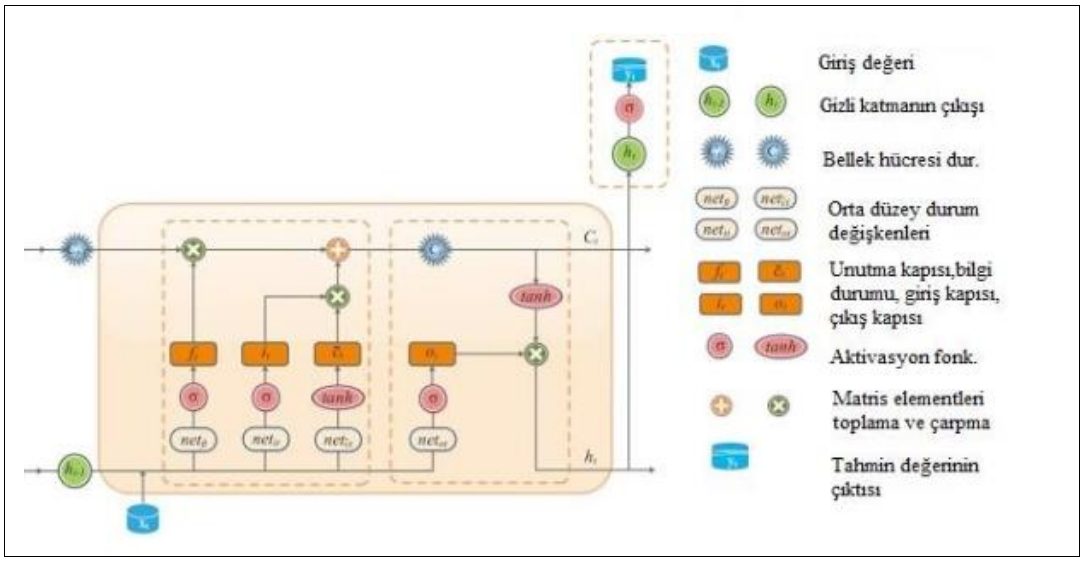
3.2.3.4. Lstm

LSTM (Long Short-Term Memory), derin öğrenmede kullanılan bir tür tekrarlayan sinir ağı (RNN) mimarisidir. RNN'lerin belirgin özelliği, girdiler arasında ilişki kurma ve bu ilişkileri hatırlama yetenekleridir. Bu mimariler, yenilenen yapıları ve sonuçları bir sonraki adıma aktarma mekanizmaları sayesinde, girdiler arasındaki bağıntıları öğrenir ve bu bağıntıları belleğinde saklar. LSTM, RNN'lerin yaşadığı

Kaybolan Gradyan İnişi (Vanishing Gradient Descent) sorununu çözmek için geliştirilmiştir. Bu sorun, büyük farklılıklar içeren girdilerin aktivasyon fonksiyonları sonrası çok küçük değerlere indirgenmesi ve bu küçük değerlerin zamanla sıfıra yaklaşması sonucu, modelin öğrenme kapasitesinin düşmesiyle ortaya çıkar. LSTM, bu problemi hücre durumu ve unutma, giriş ve çıkış kapıları gibi bileşenlerle çözerek, bilgilerin daha uzun süre hatırlanmasını sağlar. Bu yapı, LSTM modellerinin diğer RNN türlerine kıyasla daha iyi performans göstermesine olanak tanır(Öztürk ve Pashaei,2021).

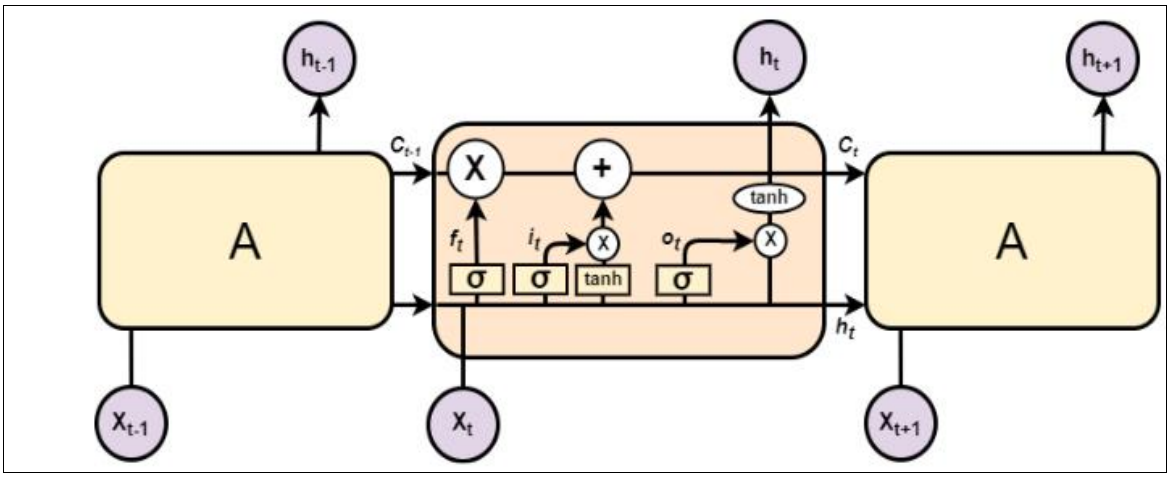
3.2.3.4.1.Lstm'nin yapısı

LSTM modelinin hafıza hücresi, üç adet doğrusal olmayan kapı ünitesinden oluşur: giriş kapısı, unutma kapısı ve çıkış kapısı (Wang F. vd., 2020). Bu yapı içinde, özellikle unutma kapısı ve çıkış aktivasyon fonksiyonu, LSTM'nin temel bileşenleri arasında yer alır ve önemlidir. Bu kapıların herhangi birinin eksik olması, LSTM'nin performansını önemli ölçüde etkileyebilir. Basit RNN yapıları, uzun vadeli bağımlılık sorunlarını ele almakta yetersiz kalırken, LSTM-RNN yapıları bu tür uzun vadeli bağımlılıkları öğrenme konusunda daha başarılıdır. Orijinal RNN'nin gizli katmanı, kısa vadeli girişlere karşı oldukça hassas bir h durumuna sahiptir. Buna karşılık, LSTM-RNN yapısı, hücre durumunu (c) ekleyerek uzun vadeli bağımlılığı muhafaza etmeyi amaçlar. Bu yeni eklenen hücre durumu, LSTM modelinde önemli bir bileşen olarak kabul edilir. Pratikte, her üç kapıdan da bilgi eklemek veya çıkarmak mümkündür(Avşar,2020). Aşağıdaki Şekil 3.5'te LSTM mimarisinin yapısı verilmiştir:

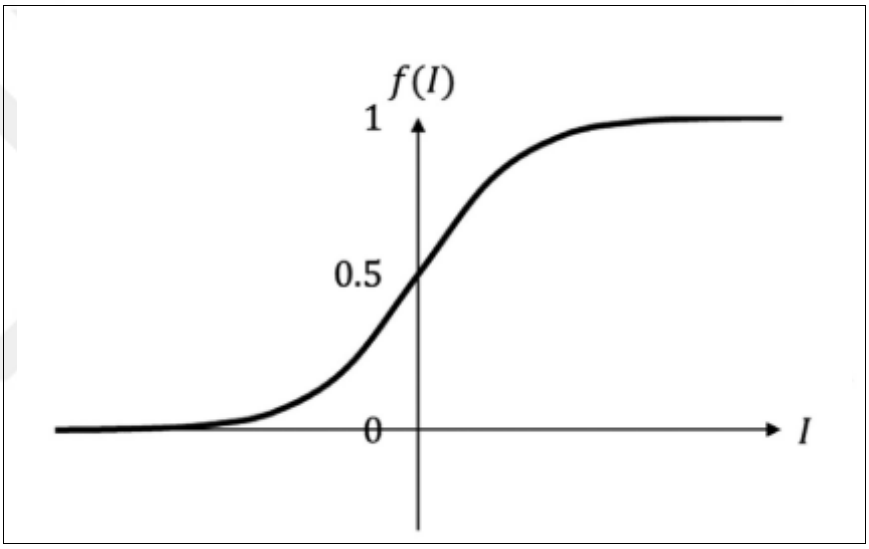


Şekil 3.5. LSTM ağ yapısının diyagramı: giriş, gizli ve çıkış katmanları (Avşar, 2020).

LSTM mimarisi, uzun dönem bağımlılıkları dikkate alarak metin sınıflandırması gibi görevlerde daha etkin bir performans sunabilir. Bu, LSTM'nin metin akışındaki geçmiş ve gelecek bilgileri göz önünde bulundurarak öznitelikleri uygun şekilde taşımasıyla gerçekleşir. Diğer yandan, tekrarlayan sinir ağları sadece bir tanjant kapısı bulundurur. Şekil 3.6'da bu durum detaylıca gösterilmiştir. LSTM mimarisinde, üç farklı kapı bulunmaktadır: unutma kapısı, giriş kapısı ve çıkış kapısı. İlk olarak, unutma kapısı, akışta hangi bilgilerin silineceğine karar verir. Bu karar, sigmoid fonksiyonu Şekil 3.7'de gösterildiği gibi kullanılarak alınır ve sonuçlar 0 ile 1 arasında ölçeklenir. Eğer sonuç 0'a yakınsa, o bilgi tamamen unutulur; eğer sonuç 1'e yakınsa, o bilgi bir sonraki katmana iletilir. Input kapısında, LSTM döngüsü içinde gelen bilgilerin saklanacağına karar verilir. Bu kapı, iki aşamadan oluşur. İlk olarak, sigmoid fonksiyonu kullanılarak belirlenen değerler, input geçidi tarafından tanımlanır. Daha sonra, tanjant fonksiyonu ile üretilen vektörle birleştirilerek, hücre durumu güncellenir. Son olarak, output kapısı oluşturulur. Bu aşamada, öncelikle sigmoid fonksiyonu uygulanır ve ardından tanh fonksiyonu ile filtreleme işlemi yapılır. Tüm bu adımlar tamamlandıktan sonra, hücre durumu, LSTM akışındaki bir sonraki hücreye aktarılacak şekilde hazır hale gelir (Kılınç vd., 2022).



Şekil 3.6. LSTM yapısı (Basiri vd., 2021; Hu vd., 2020).



Şekil 3.7. Sigmoid fonksiyonu (Matsubara vd., 2019).

Uzun Kısa Süreli Bellek (LSTM), derin öğrenme alanında önemli bir yer tutar ve uzun vadeli bağımlılıkları öğrenme kabiliyetiyle dikkat çeker. Bu algoritma, siber güvenlik gibi karmaşık alanlarda güçlü bir şekilde kullanılırken, metin tanıma, konuşma tanıma gibi daha basit uygulamalarda da etkin sonuçlar verir. LSTM'nin özelliği, girdi, çıktı ve unutma katmanlarından oluşan özel bir mimariye sahip olmasıdır. Bu mimari, LSTM'yi diğer özyinelemeli sinir ağı yapılarından ayıran bir özelliktir. LSTM'nin başlangıç aşaması, verilerin bellekte ne kadar süreyle saklanacağına karar verilmesidir. Örneğin, model için artık gerekli olmayan bir verinin, bellekten silinmesi için unutma mekanizması devreye girer. Zaman serilerinin LSTM modeline entegre edilmesi, bu serilerin denetimli öğrenme yaklaşımına dönüştürülmesiyle gerçekleşir. Sürgülü pencere yöntemi, zaman serilerinin denetimli öğrenme yaklaşımına dönüştürülmesinde

yaygın olarak kullanılan bir tekniktir. Bu yöntem, bir sonraki zaman adımındaki değeri tahmin etmek için önceki zaman adımlarının birleştirilmesini sağlar. Temel olarak, bu yöntem, gelecekteki bir değerin tahmininde, mevcut ve önceki zaman adımlarındaki verilerin kullanılmasına dayanır. Bu sayede, LSTM gibi karmaşık algoritmaların zaman serileriyle etkili bir şekilde çalışması sağlanır (Yusufoğlu vd., 2021).

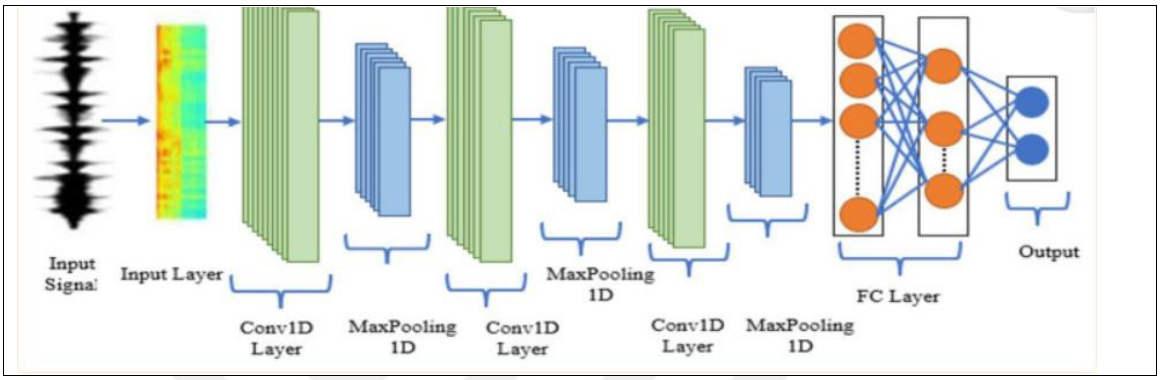
3.2.3.5.1d-cnn

1D-CNN modeli, Şekil 3.8'de ifade edildiği gibi zaman serisi verisini işlemek için özel olarak tasarlanmıştır ve ilk katmanı, bu veriyi modelin girişine uygun şekilde düzenler. Modelin ikinci katmanı, Conv1D katmanıdır. Conv1D katmanı, tek bir uzaysal veya zamansal boyut boyunca evrişim işlemi gerçekleştirir. Bu evrişim işlemi, girdi verisi üzerinde hareket eden bir evrişim çekirdeği (convolution kernel) kullanarak gerçekleştirilir ve bu sayede bir çıktı tensörü üretilir. Bu katman, verideki yerel özellikleri yakalayarak daha yüksek seviyeli temsiller oluşturmayı amaçlar. Modelin üçüncü katmanı, MaxPooling1D katmanıdır. MaxPooling1D, havuzlama işlemi gerçekleştirerek veriyi daha kompakt bir biçimde temsil eder. Bu katmanda, belirli bir havuzlama penceresi boyutundaki (pool_size) her bir segmentten maksimum değer seçilir. Havuzlama işlemi, modelin hesaplama yükünü azaltırken aynı zamanda konum değişimlerine karşı da belirli bir derecede dayanıklılık sağlar. Dördüncü katman olan Flatten katmanı, çok boyutlu veriyi tek boyutlu bir vektöre dönüştürür. Bu dönüşüm, verinin daha sonraki katmanlarda işlenebilmesi için gereklidir. Flatten katmanı, verinin boyutlarını düzleştirerek, onu Dense katmanlar için uygun hale getirir. Modelin beşinci ve altıncı katmanları ise Dense katmanlarıdır. Dense katmanlar, tam bağlantılı katmanlar olarak da bilinir ve her bir düğüm, önceki katmandaki tüm düğümlerle bağlantı kurar. İlk Dense katmanı, Flatten katmanından gelen tek boyutlu vektörü alır ve her bir girdiyi katmandaki her bir düğüme bağlar. Bu bağlantı, verinin daha yüksek seviyeli ve karmaşık özelliklerini öğrenmesini sağlar. İkinci Dense katmanı da benzer şekilde çalışır, ancak genellikle modelin nihai çıktısını üretmek için kullanılır. Bu katmanlarda, aktivasyon fonksiyonları kullanılarak verinin doğrusal olmayan dönüşümleri gerçekleştirilir ve böylece model, daha karmaşık veri yapıları ve ilişkilerini öğrenebilir. Bu yapı, zaman serisi verilerini işleyip anlamlı temsiller oluşturmak ve nihai olarak belirli bir görevi yerine getirebilmek için son derece etkili bir yöntem sunar. Modelin her bir katmanı, veriyi aşamalı olarak işleyip daha yüksek seviyeli

temsillere dönüştürerek, son katmanlarda istenilen çıktının elde edilmesini sağlar (Güler vd.,2023).Aşağıda formüller, 1D Relu, konvolüsyonel ve havuzlama katmanlarının çıktı uzunluklarını hesaplamak için kullanılır:

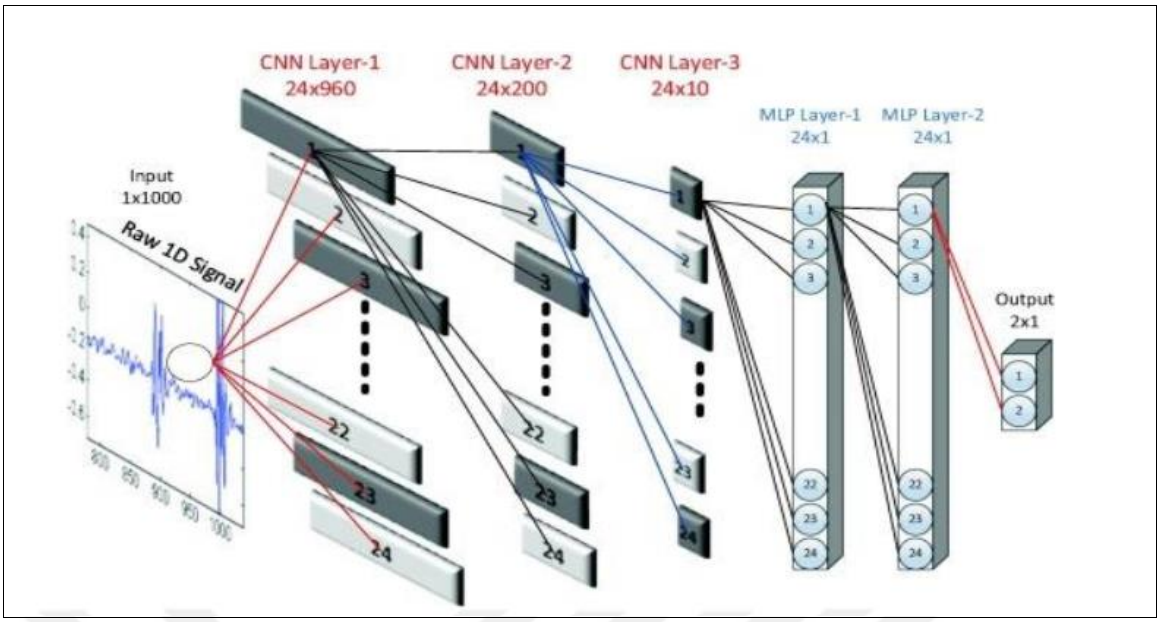
$$f(x)=\max(0,x) \tag{3.8}$$

$$\text{output\_shape}=(\text{input\_shape}-\text{pool\_size}+1)/\text{strides} \tag{3.9}$$



Şekil 3.8.1D CNN genel yapısı (Alsumaidaee vd., 2023).

Evrişimli Sinir Ağları (CNN) ile standart Yapay Sinir Ağları (YSA) arasındaki ana farklılık, evrişim katmanlarının varlığıdır. Bu katmanlar, girdi verilerinin önemli özelliklerini yakalayarak, bu özellikler arasındaki ilişkileri otomatik olarak çıkarabilme yeteneğine sahiptir. CNN'ler genellikle görüntü sınıflandırması gibi iki boyutlu veri analizlerinde kullanılır ve bu tür analizler için derin öğrenme mimarilerinde iki boyutlu evrişim katmanları (Conv2D) bulunur. Bununla birlikte, CNN'ler yalnızca iki boyutlu verilerle değil, aynı zamanda tek boyutlu ve üç boyutlu verilerle de çalışabilirler. Örneğin, zaman serisi verileri için tek boyutlu evrişim katmanları (Conv1D) ve üç boyutlu görüntüler için üç boyutlu evrişim katmanları (Conv3D) kullanılır. Böylece, veri türüne uygun evrişim katmanı seçilerek analiz yapılabilir ve bu da CNN'lerin çok yönlü kullanımını sağlar. Özellikle tek boyutlu veriler için, CNN'ler zaman serisi analizinde oldukça etkili bir yöntem sunar. Örneğin, referans evapotranspirasyon değeri tek boyutlu bir veri olduğundan, bu tür veriler için tek boyutlu evrişim katmanları kullanılır. Bu yaklaşım, verilerin özelliklerini etkin bir şekilde çıkararak daha doğru tahminlerin yapılmasını sağlar. CNN'lerin bu çok yönlü ve esnek yapısı, farklı veri türleriyle çalışmayı mümkün kılar ve derin öğrenme modellerinin performansını artırarak geniş bir uygulama alanı sunar (Saleh 2021).



Şekil 3.9.1D CNN'nin birbirini izleyen üç evrişim katmanı (Kiranyaz vd., 2021).

2D Konvolüsyonel Sinir Ağları (CNN'ler) genellikle iki boyutlu veriyle çalışmak için kullanılır, ancak 1D Konvolüsyonel Sinir Ağları (1D CNN'ler) olarak adlandırılan modifiye edilmiş bir versiyon, tek boyutlu veri setlerinde daha etkili olabilir. Araştırmalar, 1D CNN'lerin, etiketlenmiş verilerin sınırlı ve varyasyonların yüksek olduğu bazı uygulamalarda 2D CNN'lerden daha iyi performans gösterdiğini ortaya koymuştur (Kiranyaz ve diğerleri, 2021). 1D CNN'ler, düşük hesaplama gereksinimleri nedeniyle daha sığ mimariler kullanarak eğitim ve çıkarım süresini önemli ölçüde kısaltabilir. Bu özellikleri, onları gerçek zamanlı ve düşük maliyetli uygulamalar için ideal hale getirir. Şekil 3.9 'da gösterildiği gibi, 1D CNN'ler iki temel katman türünden oluşur: 1D konvolüsyonel katmanlar ve yoğun (dense) katmanlar. 1D konvolüsyonel katmanlar, aktivasyon fonksiyonları ve alt örnekleme (pooling) işlemlerini içerir. Yoğun katmanlar, genellikle Çok Katmanlı Algılayıcı (MLP) katmanları olarak da bilinir çünkü benzer şekilde çalışırlar. Giriş katmanı, ham tek boyutlu veriyi kabul eder ve pasif bir katman olarak işlev görürken, çıkış katmanı sınıfların sayısına eşit sayıda nörona sahiptir. 1D CNN'lerin konvolüsyonel katmanları, ham tek boyutlu veriyi işleyerek belirleyici özellikleri çıkarır ve bu özellikler daha sonra sınıflandırma için yoğun katmanlara iletilir. Bu yaklaşım, özellik çıkarma ve sınıflandırma işlemlerini birleştirerek sınıflandırma performansını optimize eder. 1D CNN'lerin yapılandırılmasında dikkate alınması gereken bazı hiperparametreler bulunur. Bunlar arasında gizli CNN ve MLP katmanlarının sayısı ve her katmandaki

nöron sayısı, her konvolüsyonel katmandaki filtre (kernel) boyutu, alt örnekleme faktörü ve havuzlama ile aktivasyon fonksiyonlarının seçimi yer alır. Şekil 10'daki örnek 1D CNN yapılandırmasında, üç gizli konvolüsyonel katman ve iki gizli yoğun katman bulunur. Bu yapılandırmada, alt örnekleme faktörü 4 olarak belirlenmiş ve tüm gizli konvolüsyonel katmanlarda filtre boyutu 41 olarak ayarlanmıştır (Kılıçkaya,2022).

3.2.4. Performans Ölçütleri

Karışıklık matrisi, verilerin sınıflandırılması sonucunda elde edilen çıktıların en az iki veya daha fazla kategoriye sahip olduğu durumlarda performansın ölçülmesinde yaygın olarak kullanılan bir araçtır. Bu matris, sınıflandırma modelinin her bir sınıf için doğru ve yanlış sınıflandırmalarını ayrıntılı bir şekilde görselleştirir ve bu sayede modelin performansını detaylı bir şekilde değerlendirmemize olanak tanır (Yarğı,2021). Bu çalışmada kullanılan veri seti, çeşitli makine öğrenimi algoritmalarıyla test edildikten sonra oluşturulan modellerin başarı performansını değerlendirmek için çeşitli istatistiksel ölçütler kullanılmaktadır. Model performansını karşılaştırırken, tahmin edilen değerler ile gerçek değerler arasında bir karşılaştırma yapılması gerekmektedir. Sınıflandırma problemleri için, bu karşılaştırma genellikle Karışıklık Matrisi (Confusion Matrix) olarak bilinen hata matrisine dayanır. Karışıklık matrisi, her bir sınıf için doğru ve yanlış sınıflandırmaları ayrıntılı olarak göstererek, modelin performansını değerlendirmeye yardımcı olur. Bu matris, sınıflandırma algoritmalarının güvenilirliğini ve etkinliğini anlamak için önemli bir araçtır (Kaba ve Kalkan,2022).

Çizelge 3.6. Karışıklık matrisi (Tanyılıdızı vd., 2018).

		Öngörülen Sınıf	
		EVET	HAYIR
Gerçek Sınıf	EVET	TP	FN
	HAYIR	FP	TN

Karışıklık Matrisi, sınıflandırma işlemlerinin sonuçlarını analiz etmek ve görselleştirmek için yaygın olarak kullanılan bir yöntemdir. Sınıflandırma performansını doğru bir şekilde değerlendirebilmek için doğruluk, duyarlılık ve özgünlük gibi metriklerin hesaplanması gerekmektedir. Bu metrikler, karışıklık matrisinden elde edilen verilere dayanarak hesaplanır. Çizelge 3.6 'da karışıklık matrisinin yapısı gösterilmiştir. Bu tabloda, satırlar gerçek sınıfları, sütunlar ise modelin tahmin ettiği sınıfları temsil eder. Karışıklık matrisindeki bu değerler, belirli formüller kullanılarak hesaplanır ve böylece modelin performansı değerlendirilir. Yöntemin doğruluğunu belirlemek için kullanılan önemli ölçütlerden biri doğruluk oranıdır. Doğruluk oranı, test veri setindeki tüm örnekler arasından doğru sınıflandırılanların oranını ifade eder. Bir başka kritik ölçüt ise duyarlılıktır. Duyarlılık, toplam pozitif örnekler içinde doğru sınıflandırılan pozitif örneklerin oranını gösterir. Özgünlük ise yanlış olarak sınıflandırılması gereken negatif örneklerin oranını belirtir. Bu üç ölçüt, modelin sınıflandırma başarımını detaylı bir şekilde anlamamıza yardımcı olur. Veri setinin eğitim ve test veri seti olarak hangi oranlarda bölüneceği, sınıflandırma performansını önemli ölçüde etkileyen bir faktördür. Eğitim verisi, modelin öğrenmesi için kullanılırken, test verisi modelin doğruluğunu değerlendirmek için kullanılır. Veri seti, belirli bir oranda bölünerek veya tamamen ayrı bir test veri seti kullanılarak doğruluk oranı belirlenebilir. Örneğin, veri seti %70 eğitim ve %30 test olarak bölünebilir. Ayrıca, çapraz doğrulama gibi teknikler kullanılarak veri setinin daha etkin bir şekilde değerlendirilmesi sağlanabilir. Bu yöntemler, modelin genel performansını ve geçerliliğini ölçmek açısından kritik öneme sahiptir. Doğruluk, duyarlılık ve özgünlük gibi metrikler, modelin ne kadar başarılı olduğunu anlamamıza yardımcı olur. Ayrıca, veri setinin eğitim ve test olarak nasıl bölündüğü de modelin performansı üzerinde büyük bir etkiye sahiptir. Bu nedenle, doğru veri bölme stratejileri ve değerlendirme metrikleri kullanılarak sınıflandırma modellerinin etkinliği artırılabilir (Tanyılıdızı vd., 2018). Aşağıda formülleri gösterilmiştir:

$$Doğruluk=(TP+TN)/(TP+FP+FN+TN) \quad (3.10)$$

$$Duyarlılık=TP/(TP+FP) \quad (3.11)$$

$$\text{Özgünlük= } TN/(TN+FP) \quad (3.8)$$

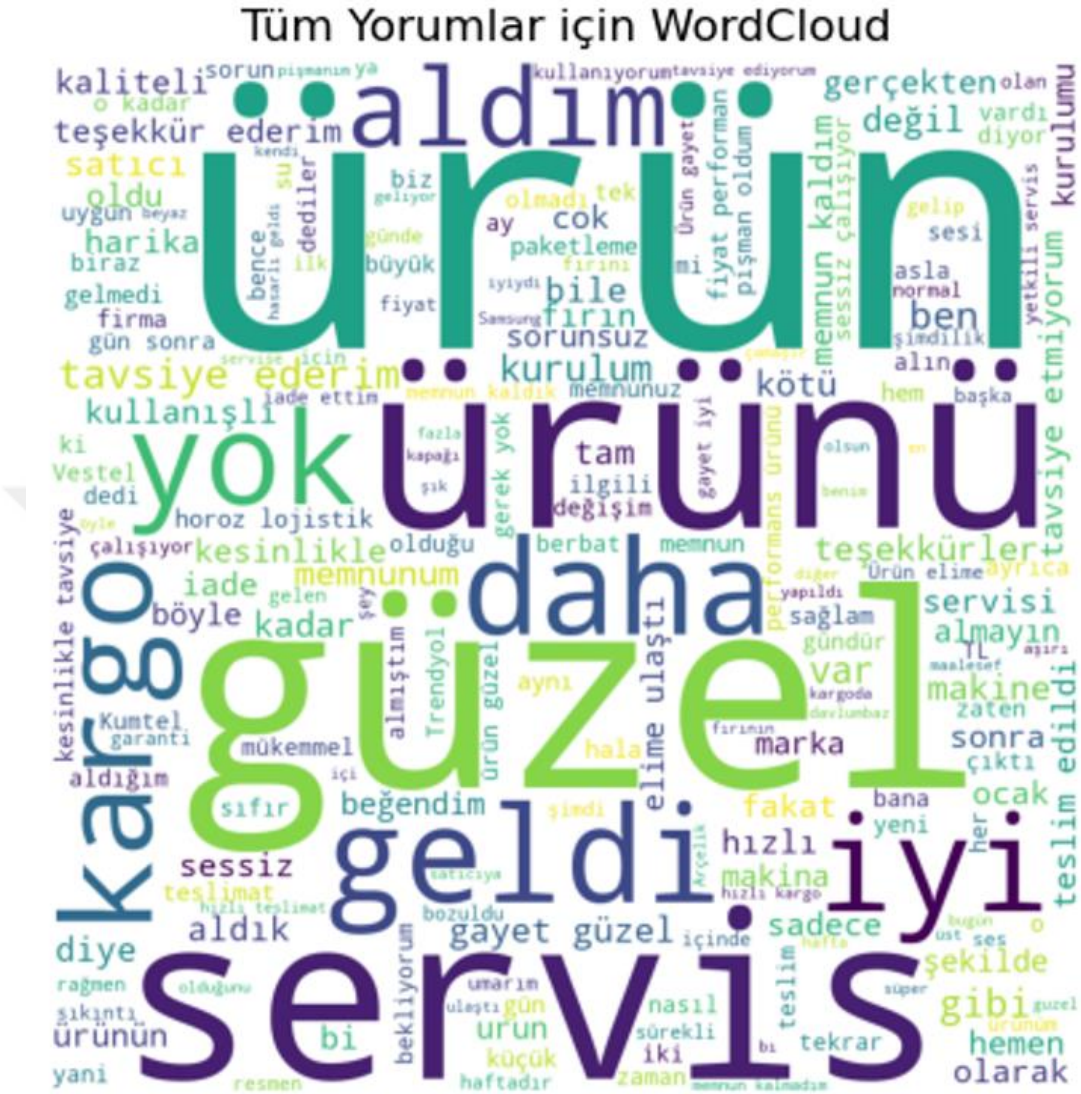
4.BULGULAR

Kelime bulutu, metin verilerindeki kelimelerin sıklığını görsel olarak temsil etmek için kullanılan bir tekniktir. Bu teknikte, metinde en sık geçen kelimeler daha büyük ve belirgin bir şekilde gösterilirken, daha az sık geçen kelimeler daha küçük ve soluk renklerde gösterilir. Kelime bulutları genellikle bir metnin anahtar kavramlarını ve temalarını hızlıca görselleştirmek için kullanılır. Kelime bulutları, çeşitli alanlarda kullanımı yaygın olan etkili bir görselleştirme aracıdır. Özellikle metin madenciliği, duygu analizi, müşteri geri bildirimi analizi gibi alanlarda metin verilerinin keşfedilmesi ve anlaşılmasında kullanılır. Ayrıca, sosyal medya analizi, pazar araştırmaları ve kullanıcı yorumlarının değerlendirilmesi gibi alanlarda da sıklıkla tercih edilir. Tüm yorumlara ait kelime bulutu Şekil 4.1'de verilmiştir. Ancak tüm yorumlara ait kelime bulutu verilmeden önce yorumlardan ["bir", "ve", "ile", "ancak", "ama", "bu", "için","da","çok","hiç","de","ne"] gibi gereksiz kelimeler elimine edilmiştir.

Olumlu kelimeler, kullanıcıların üründen memnuniyetini ve olumlu deneyimlerini ifade etmektedir. Örneğin, "yüksek performans", "iyi fiyat", "memnun kaldım", "kurulumu kolay" gibi ifadeler, kullanıcıların üründen genel olarak memnun kaldığını ve alışveriş deneyiminin olumlu olduğunu göstermektedir. Bu tür ifadeler, ürünün beklentileri karşıladığını ve hatta bazen aşıp, kullanıcıların beklentilerini aştığını vurgular. Ayrıca, olumlu yorumlarda ürünün kullanım kolaylığı, performansı ve fiyat-performans oranının öne çıktığı görülmektedir. Özellikle, "hızlı kargo", "kolay kurulum", "kaliteli malzeme", "yüksek performans" gibi ifadeler, kullanıcıların üründen elde ettikleri faydaları ve memnuniyetlerini belirtir. Olumlu yorumlara ait kelime bulutu Şekil 4.3'de gösterilmiştir.

Olumsuz kelimeler ise, kullanıcıların üründen memnun kalmadıkları veya beklenmedik sorunlarla karşılaştıklarını ifade eder. Örneğin, "gecikme", "kötü koku", "kalite kontrol eksikliği", "servis problemi" gibi ifadeler, kullanıcıların ürünle ilgili olumsuz deneyimlerini ve karşılaştıkları sorunları belirtir. Bu tür olumsuz yorumlar, genellikle ürünün beklenen performansı göstermemesi, teslimat sürelerinde yaşanan gecikmeler veya kalite sorunları gibi konulara odaklanır. Kullanıcılar, bu tür sorunlarla karşılaşmaları durumunda genellikle üründen memnun kalmazlar ve bu deneyimlerini diğer potansiyel alıcılarla paylaşarak onları bilgilendirmeye çalışırlar. Bu şekilde olumlu ve olumsuz kelimelerin yorumlanması, alışveriş yapmayı düşünen diğer kullanıcılara ürün hakkında daha kapsamlı bir görünüm sunar ve onlara karar

vermelerine yardımcı olabilir. Olumsuz yorumlara ait kelime bulutu Şekil 4.2'de gösterilmiştir.



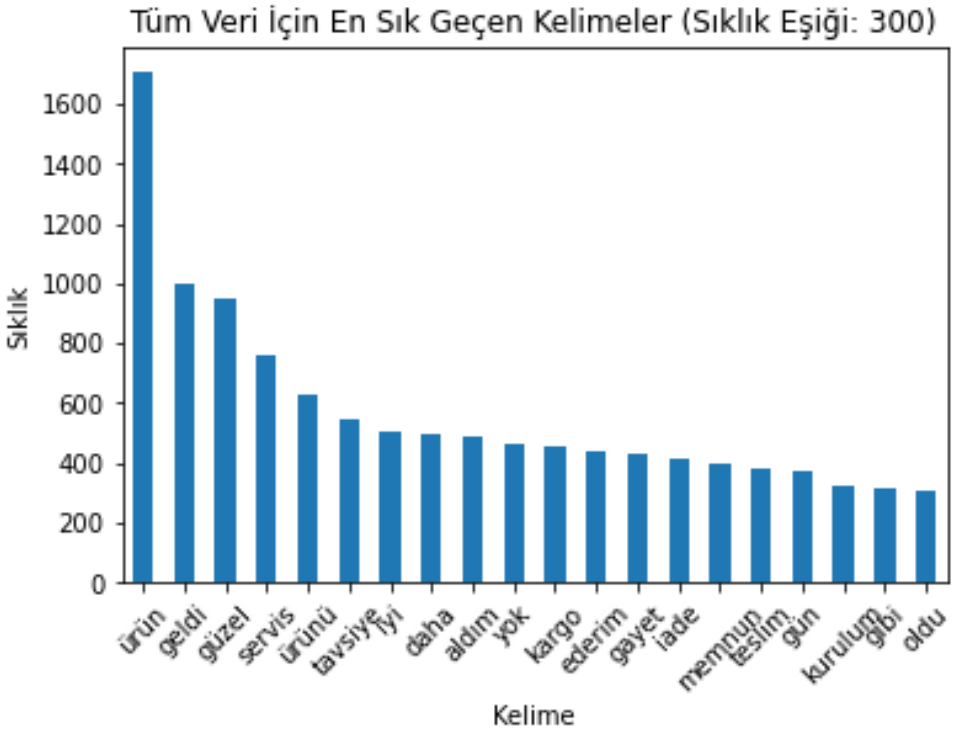
Şekil 4.1: Tüm yorumlara ait kelime bulutu

Şekil 4.2: Olumsuz yorumlara ait kelime bulutu.

Şekil 4.3: Olumlu yorumlara ait kelime bulutu.

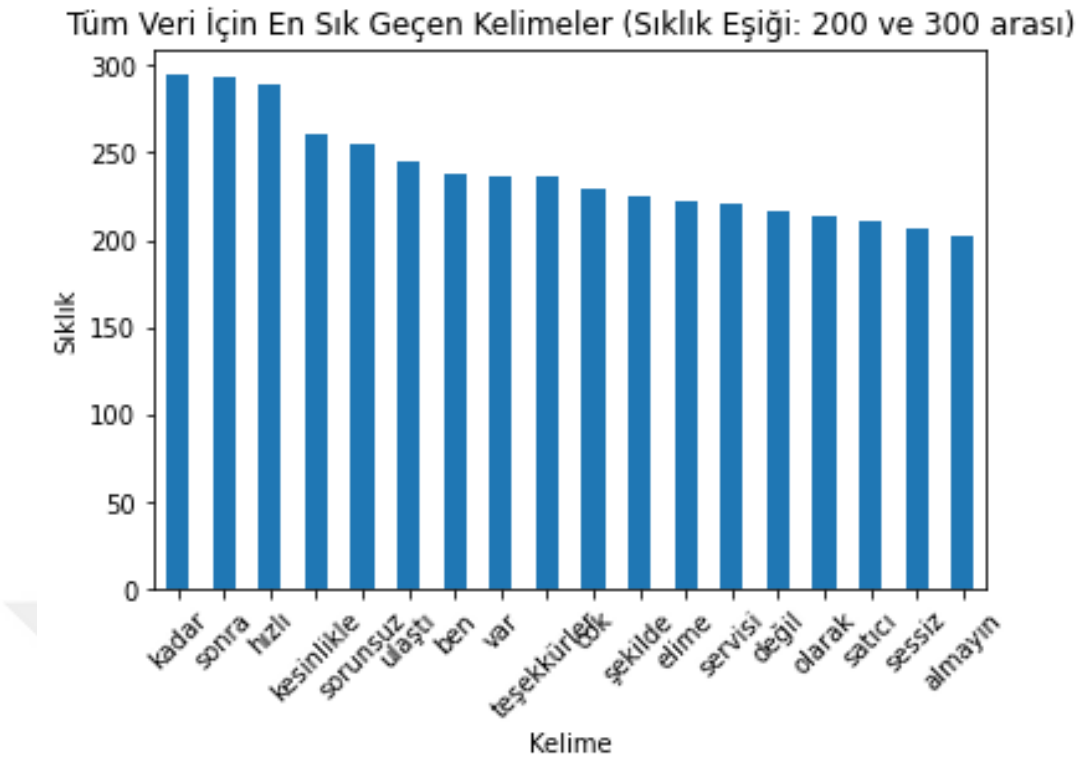
Metin madenciliğinde "bag of words" (kelime torbası) terimi, bir belgedeki metni analiz etmek için kullanılan bir modeldir. Bu modelde, metin belgesinde geçen her bir kelime bir torba içine atılır ve belgedeki kelime sıklıkları dikkate alınarak bir vektör oluşturulur. Bu vektör, belgenin içeriğini temsil eder. Kelime torbası modeli, metin sınıflandırma, duygu analizi, belge benzerliği gibi birçok metin madenciliği görevinde kullanılır. Örneğin, bir spam filtresi oluşturmak için gelen e-postaların içeriğini analiz etmek istediğinizi düşünelim. Her e-postadaki kelimeleri bir torbaya atar, daha sonra bu torbaları kullanarak e-postaları spam veya spam olmayan olarak sınıflandırabilirsiniz. Kelime torbası modeli, metinler arasındaki benzerlikleri veya farkları bulmak için de kullanılabilir. İki belgenin benzerliklerini ölçmek için, her iki belge için kelime torbası oluşturulur ve bu torbalar arasındaki benzerlik ölçülür. Bu modelin bir dezavantajı, kelime sırasını ve belge içindeki kelimenin konumunu dikkate almamasıdır. Ancak, basit ve etkili olması nedeniyle birçok uygulamada tercih edilir.

Aşağıdaki şekillerde kelime torbası elde ettikten sonra kelime sıklığı 300'üzerinde olan kelimeler ve 200-300 arasındaki kelimeler ve sıklık değerleri verilmiştir.



Şekil 4.4. Tüm veri seti içinde sıklık eşiği 300 üzerinde olan kelimeler.

Şekil 4.4'teki kelimelere bakıldığında tüm yorumlardan en çok geçen keliemeler ürün, geldi, güzel, servis, ürünü, tavsiye, iyi, daha, aldım, yok, kargo, ederim, gayet, iade, memnun, teslim, gibi ve oldu kelimeleridir. Yorumlarda en çok geçen kelime "ürün" olarak görülmektedir.

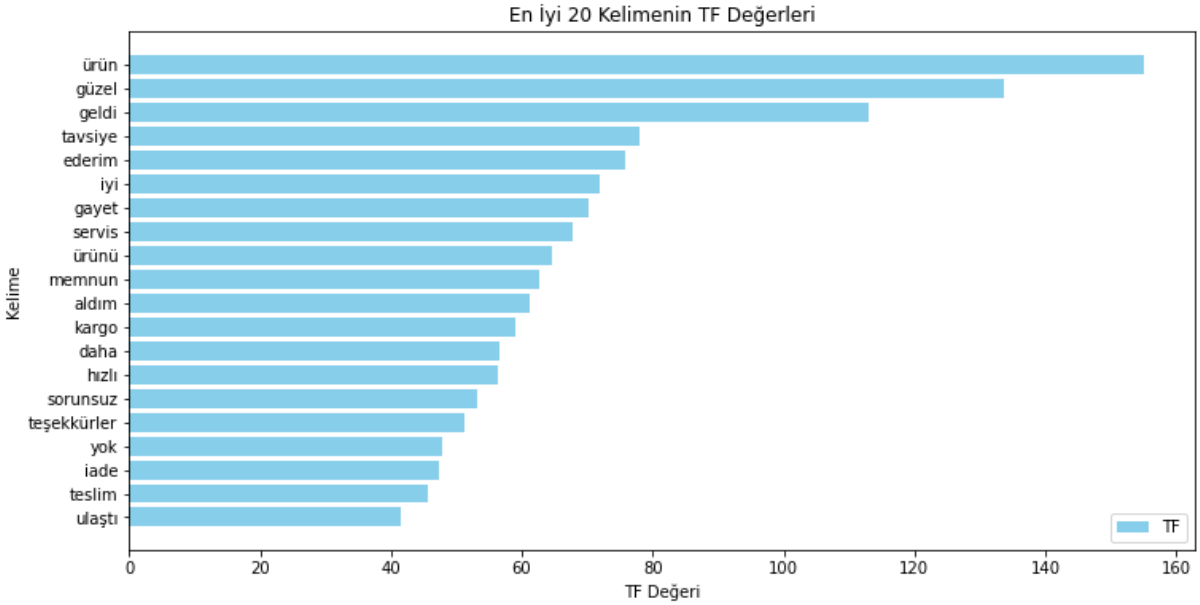


Şekil 4.5. Tüm veri seti içinde sıklık eşiği 200-300 arasında olan kelimeler.

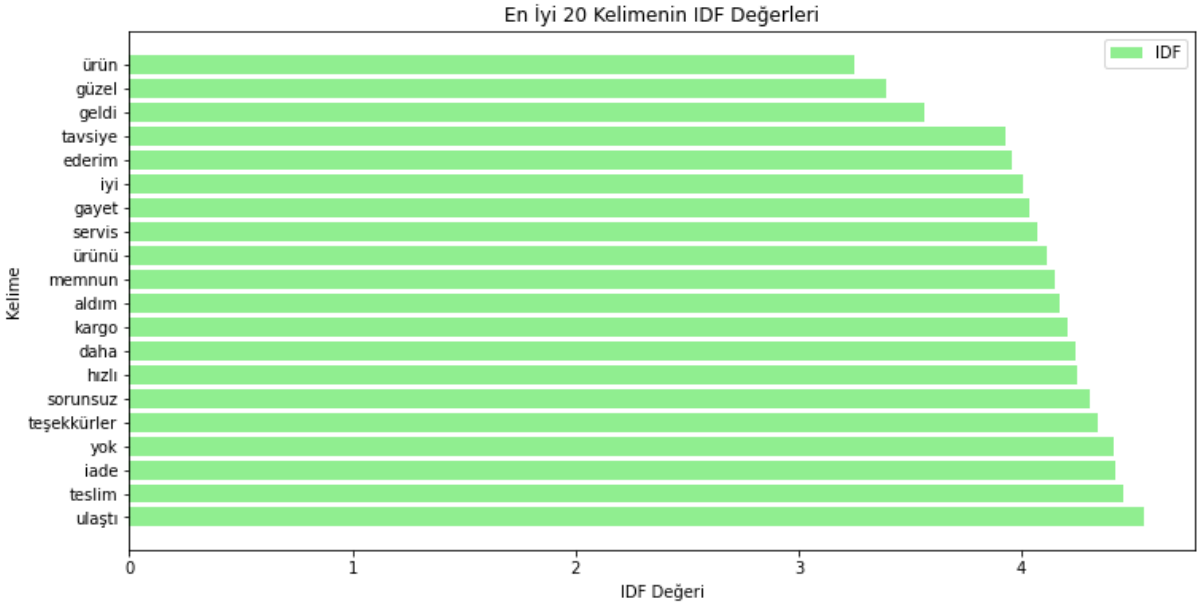
Şekil 4.5'e bakılğında sayıları 200-300 arasında olan keliemeler "kadar, sonra, hızlı, kesinlikle, sorunsuz, ulaştı, ben, var, teşekkürler, şekilde, elime, servisi, değil, olarak, satıcı, sessiz, almayın" kelimeler gözlenmiştir. Genel anlamda olumluluk bildiriren kelimelerdir. TF (Term Frequency - Terim Frekansı), bir belgedeki belirli bir terimin kaç kez geçtiğini gösterir. Bir terimin belgedeki sıklığı ne kadar yüksekse, TF değeri o kadar yüksek olur. TF değerinin amacı, bir terimin belgedeki önemini ölçmektir. Özellikle bir belgede sıkça geçen terimler genellikle önemli kavramları veya anahtar kelimeleri temsil eder. IDF (Inverse Document Frequency - Ters Belge Frekansı), belgeler koleksiyonundaki belirli bir terimin ne kadar yaygın veya nadir olduğunu ölçer. Yaygın terimlerin (genellikle "the", "and", "is" gibi) önemi düşükken, nadir terimlerin (özel isimler, uzman terimler gibi) önemi daha yüksektir. IDF, bir terimin genel önemini belirlemek için kullanılır.

Yeniden ifade edecek olursak: "Şekil 4.6 ve Şekil 4.7'de en çok gözlenen 20 kelimenin TF ve IDF değerleri verilmiştir. TF, bir terimin bir belgedeki sıklığını ölçerken IDF, terimin belgeler koleksiyonundaki yaygınlığını belirler. TF değeri, bir terimin belgedeki önemini yansıtırken, IDF değeri terimin genel önemini belirler. Bu

değerlerin bir araya gelmesi, bir terimin belgedeki önemi ile genel koleksiyondaki yaygınlığı arasındaki dengeyi sağlar."



Şekil 4.6. En çok gözlenen 20 kelimelere ait TF değerler.



Şekil 4.7. En çok gözlenen 20 kelimelere ait IDF değerler.

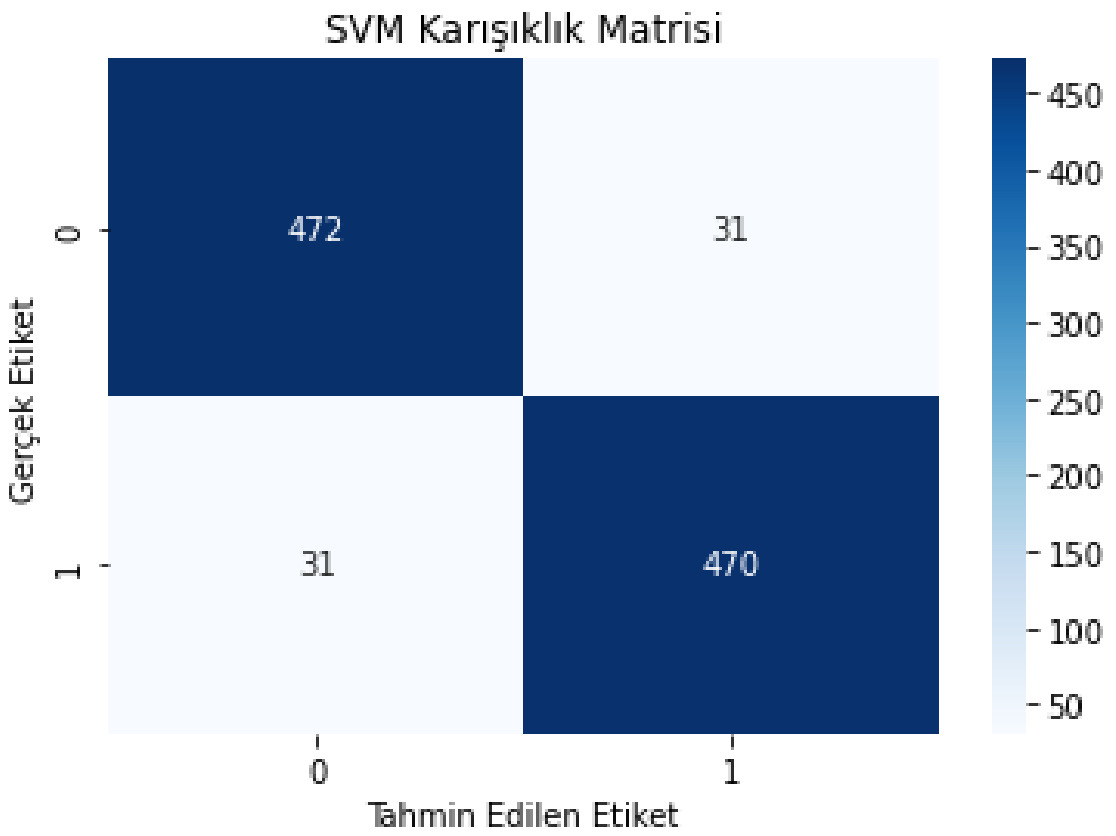
TF-IDF öznitelikler elde edildikten sonra klasik makine öğrenmesi yöntemler ile sınıflandırma işlemleri gerçekleştirilmiştir. SVM, Knn ve NB makine öğrenmesi yöntemler kullanılarak %80-20 eğitim-test oranları ile elde edilen başarı oranları Çizelge 3.7'de verilmiştir.

Çizelge 3.7. Makine öğrenmesi algoritmaları performans değerleri.

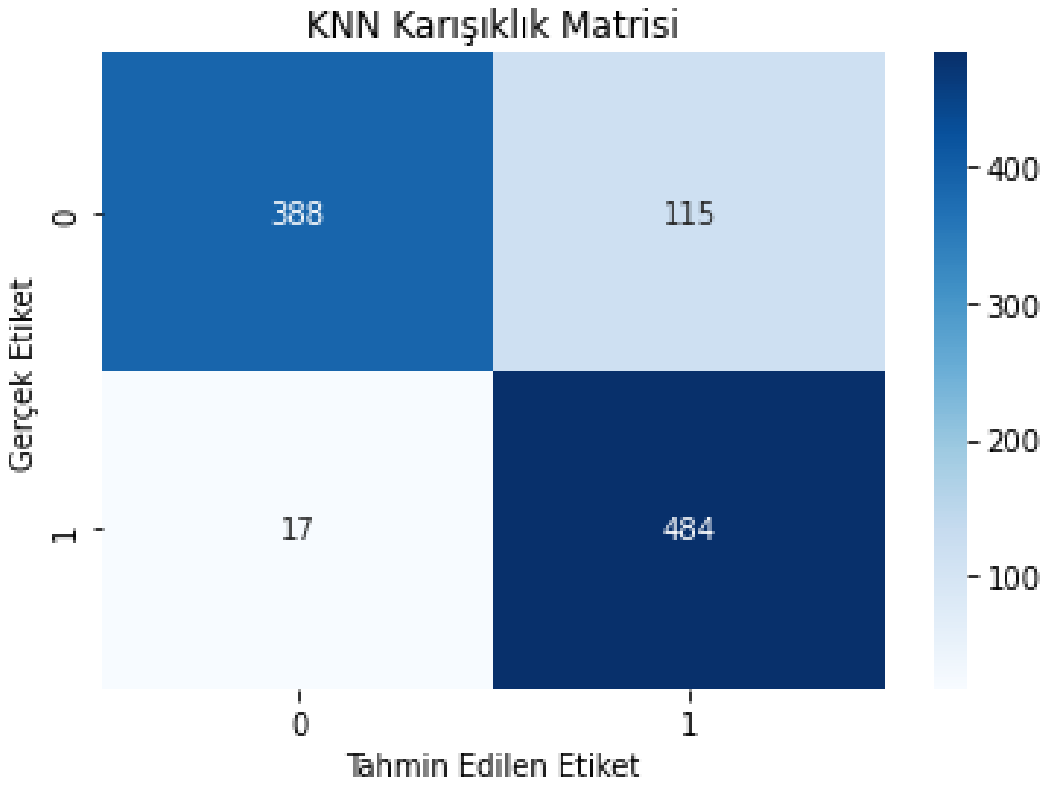
Model	**Başarı**	**Precision (kesinlik)**	**Recall (hatırlama)**		**F1 Ölçütü**
SVM	0.9342	0.9344	0.9342		0.9342
Knn	0.8675	0.8824	0.8675		0.8662
NB	0.9402	0.9403	0.9402		0.9402

Bu tablo, bir alışveriş sitesinde beyaz eşya ürünlerine yapılan yorumların duyarlılık ve kesinlik gibi ölçümlerle birlikte değerlendirilmesini sağlıyor. Her modelin başarısı ve performansı, doğruluk (başarı), kesinlik, hatırlama ve F1 ölçütü olarak ölçülüyor. SVM modeli, %93.42 başarı elde ederek oldukça etkileyici bir sonuç gösteriyor. Ayrıca, kesinlik, hatırlama ve F1 ölçütleri de oldukça yüksek. Bu, modelin hem olumlu hem de olumsuz yorumları doğru bir şekilde sınıflandırdığını gösteriyor. KNN modeli, %86.75 başarı oranıyla SVM'den biraz geride kalsa da hala kabul edilebilir bir performans sergiliyor. Ancak, kesinlik ve F1 ölçütü, SVM'ye kıyasla biraz düşük. Bu, modelin belirli durumlarda yanlış pozitif veya yanlış negatif sonuçlar üretebileceğini gösteriyor. NB (Naive Bayes) modeli ise %94.02 başarı oranıyla en yüksek başarıya sahip. Ayrıca, kesinlik, hatırlama ve F1 ölçütleri de oldukça yüksek. Bu, NB'nin diğer modellere kıyasla daha iyi bir performans sergilediğini gösteriyor.
Genel olarak, NB modeli en yüksek başarıyı gösteriyor gibi görünüyor. Ancak, SVM ve KNN modelleri de kabul edilebilir performans sergiliyor. Hangi modelin seçileceği, spesifik gereksinimler ve kullanım senaryolarına bağlı olabilir. Örneğin, hassas bir uygulama için kesinlik odaklı bir model tercih edilebilirken, genel doğruluk önemliyse başarı odaklı bir model tercih edilebilir.

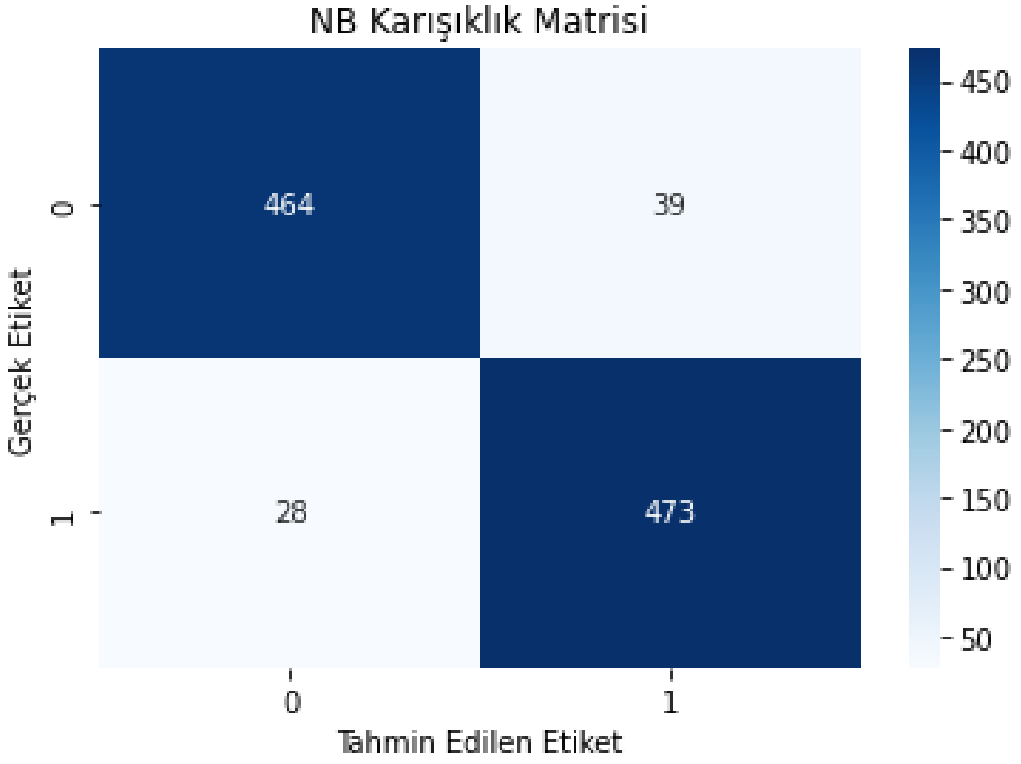
SVM, KNN ve NB için elde edilen karışıklık matrisleri Şekil 4.8, 4.9 ve 4.10'da verilmiştir. Karışıklık matrisleri üzerinden olumlu ve olumsuz yorumların başarı oranları gözlemlenebilir.



Şekil 4.8. SVM için karışıklık matrisi.



Şekil 4.9. KNN için karışıklık matrisi.



Şekil 4.10. NB için karışıklık matrisi.

NB makine öğrenmesi yöntemi KNN ve SVM modeline göre daha başarılı bulunmuştur. NB modelin farklı eğitim test oranlarına göre başarı oranları Çizelge 3.8'de verilmiştir.

Çizelge 3.8. NB modelin farklı eğitim-test oranları ile test edilmesi.

Eğitim -Test Oranı	Başarı	Precision (kesinlik)	Recall (hatırlama)	F1 Ölçütü
%50-50	0.9398	0.9406	0.9398	0.9398
%60-40	0.9362	0.9364	0.9362	0.9362
%70-30	0.9309	0.9310	0.9309	0.9309
%80-20	0.9402	0.9403	0.9402	0.9402

Bu tabloda, Naive Bayes (NB) modelinin farklı eğitim-test veri oranları için performansını gösteren sonuçlar bulunmaktadır. Her bir oran için başarı, kesinlik, hatırlama ve F1 ölçütü değerleri verilmiştir.

%50-50 eğitim-test oranında, NB modeli %93.98 başarı elde ederek oldukça yüksek bir performans sergilemiştir. Kesinlik, hatırlama ve F1 ölçütleri de oldukça yüksek seviyede tutulmuştur.

%60-40 eğitim-test oranında, başarı oranı %93.62'ye düşmüştür. Ancak, kesinlik, hatırlama ve F1 ölçütleri hala oldukça yüksek seviyelerdedir, modelin güvenilirliğini korur.

%70-30 eğitim-test oranında, başarı oranı biraz daha düşük bir seviyede (%93.09) ancak hala kabul edilebilir düzeydedir. Diğer performans ölçütleri de bu oranla uyumludur.

%80-20 eğitim-test oranında, NB modeli en yüksek başarı oranını (%94.02) elde etmiştir. Kesinlik, hatırlama ve F1 ölçütleri de diğer oranlarda olduğu gibi yüksek seviyelerde tutulmuştur.

Genel olarak, NB modelinin performansının eğitim-test oranıyla çok fazla değişmediği görülmektedir. Ancak, %80-20 oranında en yüksek başarı elde edilmiştir. Bu sonuçlar, modelin veri seti boyunca istikrarlı bir şekilde iyi performans gösterdiğini ve dengeli bir eğitim-test dağılımının etkili olduğunu göstermektedir.

Son olarak yorumların sınıflandırılmasında derin öğrenme metotları kullanılmıştır. LSTM ve 1D-CNN kullanılarak elde edilen başarı oranları 3.9'da verilmiştir.

Çizelge 3.9. Derin öğrenme başarı oranları.

Eğitim Test Oranı	LSTM	1D-CNN
%50-50	0.9081	0.8938
%60-40	0.9201	0.9021
%70-30	0.9209	0.9131
%80-20	0.9242	0.9142

Bu tabloda, derin öğrenme yöntemlerinden LSTM ve 1D-CNN'in farklı eğitim-test veri oranları için elde ettiği performans sonuçları verilmiştir.

%50-50 eğitim-test oranında, LSTM modeli %90.81 başarı elde ederken, 1D-CNN modelinin başarısı %89.38'dir. %60-40 eğitim-test oranında, LSTM modelinin başarısı %92.01 iken, 1D-CNN modelinin başarısı %90.21'dir. %70-30 eğitim-test oranında, LSTM modeli %92.09 başarı elde ederken, 1D-CNN modelinin başarısı %91.31'dir. %80-20 eğitim-test oranında, LSTM modelinin başarısı %92.42 iken, 1D-CNN modelinin başarısı %91.42'dir.

Genel olarak, LSTM modeli 1D-CNN modeline kıyasla daha yüksek başarı elde etmiştir. Ayrıca, her iki modelin de eğitim-test oranları arttıkça başarılarında artış gözlemlenmektedir. Bu sonuçlar, derin öğrenme yöntemlerinin belirli bir veri seti üzerinde eğitilirken ve test edilirken, veri seti boyunca tutarlı ve artan bir başarı gösterdiğini göstermektedir.

Naive Bayes (NB) modeli, en yüksek başarı oranını (%94.02) elde ederek diğer modellere kıyasla oldukça iyi bir performans sergiliyor. Ayrıca, kesinlik, hatırlama ve F1 ölçütleri de diğer modellere göre yüksek seviyede tutuluyor. Bu, NB'nin basit yapısı ve genel olarak iyi sonuçlar vermesiyle dikkat çekiyor. Destek Vektör Makineleri (SVM) modeli, %93.42 başarı oranı ile NB'den hemen sonra geliyor. Diğer performans ölçütleri de yüksek seviyelerde tutuluyor. SVM, genellikle karmaşık veri yapılarını iyi işleme kabiliyetiyle bilinir ve bu sonuçlar bu gücünü yansıtıyor. K-En Yakın Komşu (KNN) modeli, diğerlerine kıyasla biraz daha düşük bir performans sergiliyor (%86.75 başarı). Ancak yine de kabul edilebilir bir doğruluk seviyesine sahip. KNN, basit ve doğal bir yapıya sahiptir, ancak büyük veri setlerinde hesaplama yoğunluğu nedeniyle bazen pratikte kullanılabilirliği kısıtlanabilir. Derin öğrenme yöntemlerinden LSTM, %92.42 başarı oranı ile en yüksek performansı gösterirken, 1D-CNN modeli de iyi bir performans sergiliyor (%91.42 başarı). LSTM'in performansında eğitim-test oranları arttıkça belirgin bir iyileşme görülürken, 1D-CNN modelinde de benzer bir trend

gözlemleniyor. Genel olarak, NB ve SVM gibi geleneksel makine öğrenimi yöntemleri yüksek başarı oranları sağlarken, derin öğrenme modelleri de etkileyici performanslar sunuyor. Ancak her modelin avantajları ve dezavantajları olduğunu göz önünde bulundurarak, kullanılacak modelin spesifik gereksinimler ve veri seti özellikleri doğrultusunda seçilmesi önemlidir.

5.TARTIŞMA ve ÖNERİLER

Bu çalışma, e-ticaret sektöründe kullanıcı yorumlarının duygu analizini gerçekleştirmek üzere derin öğrenme yöntemlerinin etkinliğini değerlendirmektedir. Türkçe metinler üzerinde duygu analizi yapmak için derin öğrenme yöntemlerinin uygulanabilirliğini gösteren bu çalışma, Trendyol ve Hepsiburada gibi popüler e-ticaret platformlarından toplanan kullanıcı yorumlarına odaklanmıştır. Çalışmanın amacı, beyaz eşya ürünlerine ait kullanıcı yorumlarının analiz edilmesi ve bu yorumlardan elde edilen verilerin işletmelerin müşteri geri bildirimlerini daha etkin bir şekilde değerlendirmesine yardımcı olmasıdır.

Toplanan kullanıcı yorumları olumlu ve olumsuz olarak etiketlenmiş, bu yorumlardan öznitelikler çıkarılmıştır. Öznitelikler, Kelime Çantası (Bag of Words) ve TF-IDF (Term Frequency-Inverse Document Frequency) gibi yöntemlerle oluşturulmuştur. Veri setinin dengeli dağılımı sağlanmış ve yorumların doğru bir şekilde sınıflandırılabilmesi için gerekli ön işlemler yapılmıştır. Derin öğrenme modelleri, büyük veri setlerinden karmaşık desenleri öğrenme yeteneğine sahip olduklarından, metin madenciliği ve duygu analizi gibi karmaşık problemlerde başarılı sonuçlar elde etmiştir.

LSTM (Uzun Kısa Süreli Bellek) ve 1D-CNN (1 Boyutlu Evrişimli Sinir Ağı) modelleri kullanılarak, kullanıcılardan elde edilen yorumlar üzerinde duygu analizi gerçekleştirilmiştir. LSTM modeli, zaman serisi verilerini ve sıralı bilgileri işleyebilme kapasitesi sayesinde metinlerin bağlamsal ilişkilerini anlamada başarılı olmuştur. Özellikle, yorumlardaki duygu değişimlerini ve uzun vadeli bağımlılıkları tespit etmede LSTM'nin güçlü bir performans sergilediği gözlemlenmiştir. 1D-CNN modeli ise, özellikle metinlerdeki önemli öznitelikleri çıkarma ve bu özniteliklerden anlamlı sonuçlar elde etme konusunda etkili olmuştur. CNN modelleri, görüntü işlemelerinde yaygın olarak kullanılmalarına rağmen, tek boyutlu yapıları sayesinde metin verilerini de başarılı bir şekilde işleyebilmektedir. Bu çalışma, 1D-CNN'nin metin madenciliği alanında da rekabetçi performans sergileyebileceğini göstermiştir.

Elde edilen sonuçlar, derin öğrenme yöntemlerinin Türkçe metinlerde duygu analizinde yüksek doğruluk oranları ile çalıştığını ortaya koymuştur. Bu, literatürde az çalışılmış bir alan olan Türkçe metinlerde duygu analizi konusunda önemli bir katkı sağlamaktadır. Ayrıca, e-ticaret sitelerinden toplanan gerçek dünya verileri kullanılarak yapılan bu çalışma, pratik uygulamalarıyla literatürdeki boşluğu doldurmayı

hedeflemektedir. E-ticaret sektöründe, müşteri yorumlarının duygu analizi yapılması, işletmelerin müşteri memnuniyetini artırma, ürün ve hizmet kalitesini iyileştirme, pazarlama stratejilerini optimize etme gibi birçok alanda stratejik avantajlar sağlamaktadır. Bu tez çalışması, işletmelere müşteri geri bildirimlerini otomatik olarak analiz edebilecekleri etkili bir araç sunarak, rekabetçi avantaj elde etmelerine yardımcı olabilir.

Bu çalışma, Türkçe metinlerde kullanıcı yorumlarının duygu analizini gerçekleştirmek üzere derin öğrenme ve makine öğrenimi yöntemlerinin etkinliğini incelemektedir. Özellikle e-ticaret sektöründe, Trendyol ve Hepsiburada gibi popüler platformlardan toplanan beyaz eşya ürünlerine ait kullanıcı yorumları üzerinde odaklanılmıştır. Çalışma, kullanıcı yorumlarının duygu analizinin, işletmelerin müşteri geri bildirimlerini daha iyi değerlendirmelerine ve müşteri memnuniyetini artırmalarına yardımcı olabileceğini göstermeyi amaçlamaktadır. Kelime bulutu, metin verilerindeki kelimelerin sıklığını görsel olarak temsil eden etkili bir araçtır. Bu çalışmada, tüm yorumlara ait kelime bulutu oluşturulmuş ve gereksiz kelimeler elimine edilmiştir. Örneğin, olumlu yorumlarda "yüksek performans", "iyi fiyat", "memnun kaldım", "kurulumu kolay" gibi ifadeler öne çıkarken, olumsuz yorumlarda "gecikme", "kötü koku", "kalite kontrol eksikliği", "servis problemi" gibi ifadeler dikkat çekmektedir. Bu tür kelime bulutları, kullanıcıların genel memnuniyet düzeylerini ve karşılaştıkları sorunları görselleştirerek, işletmelere ürünlerini ve hizmetlerini iyileştirmeleri konusunda değerli içgörüler sunmaktadır.

Metin madenciliğinde sıkça kullanılan kelime torbası modeli, bir belgedeki kelimelerin sıklıklarına dayanarak metinleri temsil eder. Bu çalışmada, kelime torbası modeli kullanılarak, en sık geçen kelimeler belirlenmiş ve bu kelimeler üzerinde analizler gerçekleştirilmiştir. Tüm veri seti içinde en sık geçen kelimeler arasında "ürün", "geldi", "güzel", "servis", "ürünü", "tavsiye", "iyi", "daha", "aldım", "yok", "kargo", "ederim", "gayet", "iade", "memnun", "teslim", "gibi" ve "oldu" kelimeleri öne çıkmıştır. Kelime sıklıklarına dayanarak TF (Terim Frekansı) ve IDF (Ters Belge Frekansı) değerleri hesaplanmış ve en çok gözlenen 20 kelimenin TF ve IDF değerleri analiz edilmiştir.

Makine öğrenmesi yöntemleriyle yapılan sınıflandırma işlemlerinde SVM, KNN ve Naive Bayes (NB) algoritmaları kullanılmıştır. %80-20 eğitim-test oranları ile elde edilen sonuçlar, NB modelinin %94.02 başarı oranı ile en yüksek performansı sergilediğini göstermiştir. SVM modeli %93.42, KNN modeli ise %86.75 başarı oranı

elde etmiştir. NB modeli, kesinlik, hatırlama ve F1 ölçütlerinde de yüksek performans göstererek diğer modellere kıyasla daha başarılı bulunmuştur. NB modelinin farklı eğitim-test oranları ile performansı incelendiğinde, %80-20 oranında en yüksek başarı elde edilmiştir. Bu sonuçlar, modelin veri seti boyunca istikrarlı bir şekilde iyi performans gösterdiğini göstermektedir. Derin öğrenme yöntemleri olarak LSTM ve 1D-CNN kullanılmış ve farklı eğitim-test oranları ile elde edilen başarı oranları incelenmiştir. LSTM modeli, %92.42 başarı oranı ile en yüksek performansı göstermiştir. 1D-CNN modeli de %91.42 başarı oranı ile iyi bir performans sergilemiştir. Sonuç olarak, NB modeli en yüksek başarı oranını (%94.02) elde ederek diğer modellere kıyasla en iyi performansı sergilemiştir. Derin öğrenme yöntemlerinden LSTM ve 1D-CNN de yüksek başarı oranları ile dikkate değer performanslar göstermiştir. Bu çalışma, e-ticaret sektöründe kullanıcı yorumlarının duygu analizinde derin öğrenme ve makine öğrenmesi yöntemlerinin etkinliğini ortaya koymaktadır. Bu tür analizler, işletmelere müşteri geri bildirimlerini daha etkin bir şekilde değerlendirme ve müşteri memnuniyetini artırma konusunda önemli avantajlar sağlayabilir.

Gelecekte daha geniş veri setleri ve farklı model kombinasyonları ile bu alanda daha fazla araştırma yapılabilir. Örneğin, Transformer tabanlı modellerin (BERT, GPT gibi) duygu analizi üzerindeki etkisi incelenebilir. Ayrıca, daha karmaşık duygu durumlarını belirlemeye yönelik çok sınıflı sınıflandırma yaklaşımları da geliştirilebilir. Sonuç olarak, bu çalışma, e-ticaret sektöründe müşteri yorumlarının duygu analizini yaparak işletmelere değerli içgörüler sunmayı amaçlamaktadır. Derin öğrenme yöntemlerinin kullanılması, bu alandaki analizlerin doğruluğunu ve etkinliğini artırmakta, işletmelere müşteri memnuniyetini artırma ve rekabet avantajı sağlama konusunda önemli bir katkı sunmaktadır. Bu nedenle, duygu analizi alanındaki gelişmelerin ve derin öğrenme yöntemlerinin kullanımının, gelecekte daha da yaygınlaşması beklenmektedir.

Bu çalışmanın literatüre katkısı, Türkçe metinler üzerinde derin öğrenme modellerinin duygu analizi yapabilme potansiyelini ortaya koymasıdır. Geliştirilen modelin, e-ticaret sektöründe müşteri memnuniyetini artırma, ürün geliştirme ve pazarlama stratejilerini iyileştirme gibi konularda işletmelere önemli faydalar sağlayacağı düşünülmektedir. Elde edilen bulgular, derin öğrenme yöntemlerinin sadece İngilizce değil, Türkçe gibi farklı dillerde de etkili olabileceğini göstermektedir. Bu da, çok dilli duygu analizi çalışmalarının gelecekte daha fazla önem kazanacağına işaret etmektedir.

Çalışmanın sonuçları, işletmelere müşteri geri bildirimlerini daha etkin bir şekilde analiz etme ve değerlendirme olanağı sunarak, rekabet avantajı elde etmelerine yardımcı olabilecek niteliktedir. Ayrıca, bu tür modellerin geliştirilmesi ve uygulanması, e-ticaret sektöründe müşteri memnuniyetini artırma, ürün ve hizmet kalitesini iyileştirme ve pazarlama stratejilerini optimize etme gibi alanlarda önemli katkılar sağlayabilir. Bu nedenle, duygu analizi alanındaki gelişmelerin ve derin öğrenme yöntemlerinin kullanımının, gelecekte daha da yaygınlaşması beklenmektedir.

KAYNAKLAR

Ahmed, Sf, Alam, Msb, Hassan, M., Rozbu, Mr, Ishtiak, T., Rafa, N., ... & Gandomi, Ah (2023). Derin Öğrenme Modelleme Teknikleri: Mevcut İlerleme, Uygulamalar, Avantajlar Ve Zorluklar. *Yapay Zeka İncelemesi*, 56 (11), 13521-13617.

Alsumaidaee, Yam, Yaw, Ct, Koh, Sp, Tiong, Sk, Chen, Cp, Yusaf, T., ... Ve Raj, Aa (2023). Şalt Tesislerinde Korona Arızalarının 1d-Cnn, Lstm Ve 1d-Cnn-Lstm Yöntemleri Kullanılarak Tespiti. *Sensörler* (Basel, İsviçre), 23 (6).

Anand, M., Sahay, Kb, Ahmed, Ma, Sultan, D., Chandan, Rr Ve Singh, B. (2023). Özellik Seçimi Ve Topluluk Sınıflandırma Teknikleriyle Çevrimiçi Sosyal Ağlarda Saldırgan Dil Tespiti İçin Hesaplamada Derin Öğrenme Ve Doğal Dil İşleme. *Teorik Bilgisayar Bilimi*, 943, 203-218.And Techniques. Elsevier

Aslan, S. (2023). Doğal Dil İşleme Teknikleri Kullanarak E-Ticaret Kullanıcı İncelemelerinde Özellik Tabanlı Duygu Analizi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 35(2)*, 875-882.

Avşar, İ. İ. (2020). Kripto Paralar Ve Uluslararası Ticaret Üzerine Bir Araştırma Bibliyometrik, Lstm Ve Kümeleme Analizi.

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support Vector Machines For Classification. *Efficient Learning Machines: Theories, Concepts, And Applications For Engineers And System Designers*, 39-66.

Aytekin, Ç., & Bayram, M. A. (2021). Türkçe Metinler İçin Duygu Analizi Yaklaşımı İle İletişimde Bağlamdan Bağımsız Modellerin Geliştirilmesi Üzerine Bir Araştırma: Karma Veri Modeli Önerisi. *Yeni Medya Elektronik Dergisi*, 5(1), 12-25.

Babüroğlu, B., Tekerek, A., & Tekerek, M. (2019). Türkçe İçin Derin Öğrenme Tabanlı Doğal Dil İşleme Modeli Geliştirilmesi. *In 13th International Computer And Instructional Technology Symposium (Pp. 671-679).*

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). Abcdm: An Attention-Based Bidirectional Cnn-Rnn Deep Model For Sentiment Analysis. *Future Generation Computer Systems*, 115, 279–294. Https://Doi.Org/10.1016/J.Future.2020.08.005.

Bayer, M., Kaufhold, Ma, Buchhold, B., Keller, M., Dallmeyer, J. Ve Reuter, C. (2023). Doğal Dil İşlemede Veri Artırma: Uzun Ve Kısa Metin Sınıflandırıcılar İçin Yeni Bir Metin Üretme Yaklaşımı. *Uluslararası Makine Öğrenimi Ve Sibernetik Dergisi,14 (1*), 135-150.

Bayram, İ. (2023). Türkiye'de Kripto Para Farkındalığı Ve Tutumu: Duygu Analizi Ve İstatistiksel Analiz İle Bir Değerlendirme= Cryptocurrency Awareness And Attitudes İn Turkey: An Evaluation With Sentiment And Statistical Analysis (Master's Thesis, Sakarya Üniversitesi).

Bonaccorso, G., (2018). Machine Learning Algorithms: Popular Algorithms For Data Science And Machine Learning, *Packt Publishing*.

Cesur, E., & Efe, C. (2023). Kampüs İçi Kapalı Alanlarda Hava Kalitesinin Modellenmesi Ve Karar Destek Sistemi Geliştirilmesi. *Journal Of Intelligent Systems: Theory And Applications, 6(2),* 181-190.

Çelik, Ö., & Koç, B. C. (2021). Tf-Idf, Word2vec Ve Fasttext Vektör Model Yöntemleri İle Türkçe Haber Metinlerinin Sınıflandırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi, 23(67),* 121-127.

Çiçek, A., & Arslan, Y. (2020). Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması. *İleri Mühendislik Çalışmaları Ve Teknolojileri Dergisi, 1(1),* 13-19.

Dang, N. C., Moreno-García, M. N., & De La Prieta, F. (2020). Sentiment Analysis Based On Deep Learning: A Comparative Study. *Electronics,* 9(3), 483.

Das, M. Ve Alphonse, Pja (2023). Yapılandırılmamış Veri Seti Kullanılarak Tf-İdf Özellik Ağırlıklandırma Yöntemi Ve Analizi Üzerine Karşılaştırmalı Bir Çalışma. Arşiv Ön Baskı Arşiv:2308.04037.

Daşgın, R., & Kemal, A. D. E. M. (2022). Kitlesel Çevrimiçi Ders Platformlarında Kurslara Yapılan Yorumların Metin Madenciliği Kullanılarak Duygu Analizi. *International Journal Of Engineering Research And Development, 15(2),* 636-646.

Deka, P., C., 2014. Support Vector Machine Applications İn The Field Of Hydrology: A Review. *Applied Soft Computing,* 19, 372- 386.

Erden, C. 2021, Pyton İle Veri Madenciliği, 1. Baskı, İnkılap Kitabevi Yayın San. Tic. A.Ş., İstanbul, 1-4.

Erten, M. (2012). Sınıflandırma Yöntemleri Kullanılarak İmza Biyometriğine Dayalı Kişi Tanıma.

Frank, E., Trigg, L., Holmes, G. Ve Witten, Ih (2000). Regresyon İçin Naive Bayes. *Makine Öğrenimi,* 41, 5-25.

Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis Of Tweets Before The 2024 Elections İn Indonesia Using Indobert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (Jıtekı), 9(3),* 746-757.

Gerdan, D., Beyaz, A., & Vatandaş, M. (2020). Classification Of Apple Varieties: Comparison Of Ensemble Learning And Naive Bayes Algorithms İn H2o Framework. *Journal Of Agricultural Faculty Of Gaziosmanpaşa University (Jafag), 37(1),* 9-16.

Göker, H., & Tekedere, H. (2017). Fatih Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. *Bilişim Teknolojileri Dergisi, 10(3),* 291-299.

Görentaş, M. B. (2023). Yapay Zekâ Yöntemleriyle Uyuşmazlık Mahkemesi Kararlarının Tahmini.

Güler, E., Kakız, A. G. M. T., Gunay, F. B., Şanal, B., & Çavdar, T. (2023). Kapalı Mekân Ortamında 1d-Cnn Kullanarak Yapılan Doluluk Tespiti Sınıflandırması. *Karadeniz Fen Bilimleri Dergisi, 13(1),* 60-71.

Gündüz, H. (2023). Derin Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri Ve Destek Vektör Makineleri İle Türkçe Film Yorumları Üzerine Duygu Analizi. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi,* 26(2), 542-549.

Hacıbeyoglu, M., Çelik, M., & Çiçek, Ö. E. (2023). K En Yakın Komşu Algoritması İle Binalarda Enerji Verimliliği Tahmini. *Necmettin Erbakan Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 5(2),* 65-74.

Han, J., Kamber, M., Pei, J., (2011), Data Mining Concepts And Techniques Third Edition (Third Edition). *Morgan Kaufmann, Massachusetts*.

Han, J., Kamber, M., Ve Pei, J., 2011, Data Mining: Concepts And Techniques: Concepts

Haque, R., İslam, N., Tasneem, M. Ve Das, Ak (2023). Makine Öğrenimini Kullanarak Bengalce Sosyal Medya Yorumlarına İlişkin Çok Sınıflı Duygu Sınıflandırması. *Uluslararası Mühendislikte Bilişsel Hesaplama Dergisi, 4*, 21-35.

Hu, J., Wang, X., Zhang, Y., Zhang, D., Zhang, M., & Xue, J. (2020). Time Series Prediction Method Based On Variant Lstm Recurrent Neural Network. *Neural Processing Letters,* 52(2), 1485–1500. Https://Doi. Org/10.1007/S11063.020.10319-3

Huang, H., Asemi, A. Ve Mustafa, Mb (2023). E-Ticaret Platformlarında Duygu Analizi: Mevcut Tekniklerin Ve Gelecekteki Yönelimlerin Gözden Geçirilmesi. *Ieee Erişimi*.

Işık, N. (2019). Metin Madenciliği Yöntemleri İle E-Ticaret Markalarına Yönelik Sosyal Medya Yorumlarının Analizi (Doctoral Dissertation, Marmara Universitesi (Turkey)).

İlhan, N., & Sağaltıcı, D. (2020). Twitter'da Duygu Analizi. *Harran Üniversitesi Mühendislik Dergisi, 5(2),* 146-156.

Kaba, G., & Kalkan, S. B. (2022). Kardiyovasküler Hastalık Tahmininde Makine Öğrenmesi Sınıflandırma Algoritmalarının Karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 21(42),* 183-193.

Karakoyun, M., & Hacıbeyoğlu, M. (2014). Biyomedikal Veri Kümeleri İle Makine Öğrenmesi Siniflandirma Algoritmalarinin İstatistiksel Olarak Karşilaştirilmasi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi, 16(48),* 30-42.

Kaşıkçı, T., & Gökçen, H. (2014). Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7(1).

Kaur, G., & Sharma, A. (2023). A Deep Learning-Based Model Using Hybrid Feature Extraction Approach For Consumer Sentiment Analysis. *Journal Of Big Data, 10(1),* 5.

Kaur, H., Ahsaan, Su, Alankar, B. Ve Chang, V. (2021). Covıd-19 Tweetlerini Analiz Etmek İçin Önerilen Bir Duygu Analizi Derin Öğrenme Algoritması. *Bilgi Sistemleri Sınırları,* 1-13.

Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016, September). Makine Öğrenmesi Yöntemleri İle Duygu Analizi. *In International Artificial Intelligence And Data Processing Symposium (Idap'16)* (Vol. 17, No. 18, Pp. 17-18).

Khan, M. Ve Srivastava, A. (2024). Makine Öğrenimi Tekniklerini Kullanarak Twitter Verilerinin Duygu Analizi. *Uluslararası Mühendislik Ve Yönetim Araştırmaları Dergisi, 14 (1),* 196-203.

Kılınç, M., Aydın, C. & Tarhan, Ç. (2022). Kitle Fonlamasındaki Proje Metin İçeriklerinin Lstm İle Analizi. Journal Of Research İn Business, 7(1), E48-59.

Kına, E., & Biçek, E. (2023). Tweetlerin Duygu Analizi İçin Hibrit Bir Yaklaşım. *Doğu Fen Bilimleri Dergisi, 6(1),* 57-68.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., And Inman D. J. (2021) 1d Convolutional Neural Networks And Applications: A Survey, *Mechanical Systems And Signal Processing*, Vol. 151, P. 107398.

Kora, R., & Mohammed, A. (2023). An Enhanced Approach For Sentiment Analysis Based On Meta-Ensemble Deep Learning. *Social Network Analysis And Mining, 13(1),* 38.

Kumaş, E. (2021). Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması. E*skişehir Türk Dünyası Uygulama Ve Araştırma Merkezi Bilişim Dergisi, 2(2),* 1-5.

Matsubara, N., Teramoto, A., Saito, K., & Fujita, H. (2019). Generation Of Pseudo Chest X-Ray Images From Computed Tomographic Images By Nonlinear Transformation And Bone Enhancement. *Medical Imaging And Information Sciences,* 36(3), 141–146. Https://Doi.Org/10.11318/Mii.36.141.

Mitrović, S., Andreoletti, D. Ve Ayoub, O. (2023). Chatgpt Mi Yoksa İnsan Mı? Tespit Edin Ve Açıklayın. Chatgpt Tarafından Oluşturulan Kısa Metni Tespit Etmek İçin Makine Öğrenimi Modelinin Kararlarının Açıklanması. *Arşiv Ön Baskı Arşiv*:2301.13852 .

Nandwani, P. Ve Verma, R. (2021). Metinden Duygu Analizi Ve Duygu Tespiti Üzerine Bir İnceleme. *Sosyal Ağ Analizi Ve Madenciliği,* 11 (1), 81.

Öğe, Bc, & Kayaalp, F. (2021). Farklı Sınıflandırma Algoritmaları Ve Metin Temsil Yöntemlerinin Duygu Analizinde Performans Karşılaştırılması. *Düzce Üniversitesi Bilim Ve Teknoloji Dergisi, 9(6),* 406-416. Https://Doi.Org/10.29130/Dubited.1015320

Öztürk, K., & Şahin, M. E. (2018). Yapay Sinir Ağları Ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-İ Vekayi,* 6(2), 25-36.

Öztürk, Ö. F., & Pashaei, E. (2021). Konuşmalardaki Duygunun Evrişimsel Lstm Modeli İle Tespiti. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 12(4),* 581-589.

Özyurt, B., & Akcayol, Ma (2018). Fikir Madenciliği Ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. *Selçuk Üniversitesi Mühendislik, Bilim Ve Teknoloji Dergisi, 6(4),* 668-693. Https://Doi.Org/10.15317/Scitech.2018.160

Priyadarshini, I. Ve Cotton, C. (2021). Duyarlılık Analizi İçin Yeni Bir Lstm-Cnn-Grid Arama Tabanlı Derin Sinir Ağı. *Süper Bilgisayar Dergisi,* 77 (12), 13911-13932.

Puh, K. Ve Bagić Babac, M. (2023). Makine Öğrenimini Kullanarak Turist Yorumlarının Duyarlılığını Ve Derecelendirmesini Tahmin Etme. *Otelcilik Ve Turizm Anlayışları Dergisi,* 6 (3), 1188-1204.

Qi, Y., & Shabrina, Z. (2023). Sentiment Analysis Using Twitter Data: A Comparative Application Of Lexicon-And Machine-Learning-Based Approach. *Social Network Analysis And Mining, 13(1),* 31.

Qorib, M., Oladunni, T., Denis, M., Ososanya, E. Ve Cotae, P. (2023). Kovid-19 Aşısı Tereddütü: Covıd-19 Aşısı Twitter Veri Kümesinde Metin Madenciliği, Duygu Analizi Ve Makine Öğrenimi. *Uygulamalı Uzman Sistemler,* 212, 118715.

Rahman, H., Tariq, J., Masood, Ma, Subahi, Af, Khalaf, Oı Ve Alotaibi, Y. (2023). Denetimli Makine Öğrenimini Kullanarak Sosyal Medya Metinlerinin Çok Katmanlı Duyarlılık Analizi. Hesapla. Anne. Devam, 74, 5527-5543.

Saleh, A. A. A. R. E. (2021). Yarı-Nemli İklim Koşullarında Sürdürülebilir Su Yönetimi İçin Derin Öğrenme Kullanılarak Referans Bitki Su Tüketiminin Tahmin Edilmesi (Doctoral Dissertation, Bursa Uludag University (Turkey)).

Savci, P., & Das, B. (2023). Prediction Of The Customers' İnterests Using Sentiment Analysis İn E-Commerce Data For Comparison Of Arabic, English, And Turkish Languages. *Journal Of King Saud University-Computer And Information Sciences, 35(3),* 227-237.

Seker, S. E. (2016). Duygu Analizi (Sentimental Analysis). *Ybs Ansiklopedi,* 3(3), 21-36.

Sertaçkılıçkayahttps://Tez.Yok.Gov.Tr/Ulusaltezmerkezi/Tezgoster?Key=Rjzwh00omg4ina5sgvlggzahxowbfexf5j7zwdeyywprbr__Kmtxur50oyrftpd3 Https://Hdl.Handle.Net/20.500.14365/175

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment Analysis And Opinion Mining On Educational Data: A Survey. *Natural Language Processing Journal, 2,* 100003.

Şahin, F., Tulum, G., & Karaca, Ş. Anne Sağlığı Riski İçin Makine Öğrenmesi Modellerinin Performans Karşılaştırması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 14(4),* 547-553.

Taherdoost, H., & Madanchian, M. (2023). Artificial İntelligence And Sentiment Analysis: *A Review in Competitive Research. Computers*, 12(2), 37.

Tan, Kl, Lee, Cp Ve Lim, Km (2023). Duygu Analizine İlişkin Bir Anket: Yaklaşımlar, Veri Kümeleri Ve Gelecekteki Araştırmalar. *Uygulamalı Bilimler,* 13 (7), 4550.

Tanyılıdızı, E., Karabatak, M., Yıldırım, G., & Özpolat, Z. (2018). Siğil Tedavisinde Sınıflandırma Algoritmalarının Performans Analizi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 30(2),* 249-256.

Taşçı, M. E., & Şamlı, R. (2020). Veri Madenciliği İle Kalp Hastalığı Teşhisi. *Avrupa Bilim Ve Teknoloji Dergisi,* 88-95.

Tosun, İ. B. (2021). Makine Öğrenmesi İle Metin Sınıflandırma: Bakım Yönetim Sistemi Örneği (Master's Thesis, Sakarya Üniversitesi).

Varol, M. Ç., & Varol, E. Yenı Medyada Duygu Analızı Üzerıne Bır Değerlendırme: Bılgısayar Mühendıslığı Bılımlerı Doktora Tezlerı Incelemesı. *İletişim Araştırmaları Ve Film Çözümlemeleri,* 79.

Wang, F.-K., Cheng, X.-B., And Hsiao, K.-C. (2020). Stacked Long Short-Term Memory Model For Proton Exchange Membrane Fuel Cell Systems Degradation. *Journal Of Power Sources,* 448, 1-7. Doi:10.1016/J.Jpowsour.2019.227591.

Yarğı, V., & Postalcıoğlu, S. (2021). Eeg İşareti Kullanılarak Bağımlılığa Yatkınlığın Makine Öğrenmesi Teknikleri İle Analizi. *El-Cezeri, 8(1),* 142-154.

Yusufoğlu, H., Aydın, H., & Çetinkaya, A. (2021). Twitter Üzerindeki Finansal Tweetlerin Lstm Sinir Ağı Algoritması İle Duygu Analizi. *Veri Bilimi, 4(3),* 28-43.

Zhang, H. (2004). Saf Bayes'in Optimalliği. Aa, 1 (2), 3.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Nuray YILDIZ

Uyruğu : TC

EĞİTİM

Derece Lisans 2012

Üniversite : Ankara Gazi Üniversitesi Teknik Eğitim Fakültesi Bigisayar Sistemleri Öğretmenliği

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
11	Meb	Öğretmen
8	Meb	Alan Şefi
1	Meb	Uzman Öğretmen (Devam etmektedir)

UZMANLIK ALANI: Bilişim Teknolojileri Alan Öğretmeni

YABANCI DİLLER: İngilizce

YAYINLAR: İnternational Symposium on Architecture, Engineering and Design held on July 08-09,2024/Diyarbakır, Türkiye, in collaboration with Dicle Üniversity & İnstitute of Economic Development and Social Research with an oral presentation entitled SENTIMENT ANALYSIS FROM TEXTUAL EXPRESSIONS