

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ TEZLİ YÜKSEK
LİSANS PROGRAMI**

**TÜRKÇE E-TİCARET ÜRÜN YORUMLARININ
SINIFLANDIRILMASI**

HAZIRLAYAN

BURCU MELİS TOPRAK

YÜKSEK LİSANS TEZİ

ANKARA – 2023

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ TEZLİ YÜKSEK
LİSANS PROGRAMI**

**TÜRKÇE E-TİCARET ÜRÜN YORUMLARININ
SINIFLANDIRILMASI**

HAZIRLAYAN

BURCU MELİS TOPRAK

YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

DOÇ. DR. SELDA GÜNEY

ANKARA – 2023

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Elektrik-Elektronik Mühendisliği Anabilim Dalı Elektrik-Elektronik Mühendisliği Tezli Yüksek Lisans Programı çerçevesinde Burcu Melis Toprak tarafından hazırlanan bu çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 12/09/2023

Tez Adı: Türkçe E-Ticaret Ürün Yorumlarının Sınıflandırılması

Tez Jüri Üyeleri (Unvanı, Adı- Soyadı, Kurumu)

İmza

Prof. Dr. Hamit Erdem, Başkent Üniversitesi

.....

Dr. Öğr. Üyesi Koray Açıcı, Ankara Üniversitesi

.....

Doç. Dr. Selda Güney, Başkent Üniversitesi

.....

ONAY

Prof. Dr. Faruk Elaldı

Fen Bilimleri Enstitüsü Müdürü

Tarih : / /

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

Tarih: 25/09/2023

Öğrencinin Adı, Soyadı : Burcu Melis Toprak

Öğrencinin Numarası : 22010627

Anabilim Dalı : Elektrik-Elektronik Mühendisliği

Programı : Tezli Yüksek Lisans

Danışmanın Unvanı/Adı, Soyadı : Doç. Dr. Selda Güney

Tez Başlığı : Türkçe E-Ticaret Ürün Yorumlarının Sınıflandırılması

Yukarıda başlığı belirtilen Yüksek Lisans tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 85 sayfalık kısmına ilişkin, 25/09/2023 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %10'dur.

Uygulanan filtrelemeler:

1. Kaynakça hariç
2. Alıntılar hariç
3. Beş (5) kelimeden daha az örtüşme içeren metin kısımları hariç

“Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını” inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası :

ONAY

Tarih: ... / ... / 20...

Öğrenci Danışmanı

Doç. Dr. Selda Güney

TEŞEKKÜR

Hem yüksek lisans eğitimim hem de bu tez çalışması süresince her konuda tüm samimiyetiyle destek olan ve yol gösteren değerli tez danışman hocam Doç. Dr. Selda Güney'e en içten teşekkürlerimi sunarım.

Eğitim hayatım boyunca her zaman yanımda olan motivasyonlarını ve desteklerini eksik etmeyen aileme teşekkür ederim.



ÖZET

Burcu Melis TOPRAK

TÜRKÇE E-TİCARET ÜRÜN YORUMLARININ SINIFLANDIRILMASI

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı

2023

Günümüzde e-ticaret ürün incelemeleri, çevrim içi alışverişte oldukça önemli bir rol oynamaktadır. Teknolojinin hayatımızdaki önemi ve çevrim içi alışverişe olan yoğun ilgi nedeniyle ürün yorumları ürünü satın alma aşamasında oldukça önemlidir. Ürüne verilen puanlar ve yazılan yorumlar arasında zaman zaman uyumsuzluk yaşanmaktadır. Bu nedenle yazılan yorumların metin sınıflandırma kullanılarak gruplandırılması ile ürün hakkında daha objektif değerlendirilme sağlanacağı düşünülmektedir. Metin sınıflandırmada oldukça kullanışlı ve etkili olan denetimli ve denetimsiz makine öğrenimi algoritmalarının yanı sıra derin öğrenme algoritmaları da oldukça popülerdir ve başarı oranları yüksektir. Bu tez çalışmasının amacı, Türkçe metin sınıflandırması için farklı makine öğrenmesi yöntemlerinin başarılarını incelemektir. Kullanılan veri setinde, çevrim içi alışveriş sitelerinde bir ürün altına yapılan yorumlar toplanmış ve yorumların olumlu, olumsuz ya da nötr olmasına göre sınıf etiketleri verilerek veri seti oluşturulmuştur. Toplam 15170 yorumun yer aldığı veri setinde 6799 olumlu, 6978 olumsuz ve 1393 tarafsız yorum bulunmaktadır. Sınıflandırma aşamasında, bu tez çalışmasında sınıflandırıcı olarak Evrişimsel Sinir Ağı (ESA) önerilmiştir. Ayrıca önerilen yöntem, Karar Ağaçları, Lineer Diskriminant Analiz, Uzun Kısa Süreli Bellek, İkinci Dereceden Diskriminant Analiz, Çift Yönlü Uzun Kısa Süreli Bellek, Verimli Lineer Destek Vektör Makineleri, Geçitli Tekrarlayan Birim, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel ve Destek Vektör Makineleri ile karşılaştırılmıştır. En yüksek başarı Evrişimsel Sinir Ağları kullanıldığında %90,77 doğruluk ile elde edilmiştir.

ANAHTAR KELİMELER: Destek Vektör Makineleri, Evrişimsel Sinir Ağları, Metin Sınıflandırma, Türkçe Metin Sınıflandırma, Uzun Kısa Süreli Bellek

ABSTRACT

Burcu Melis TOPRAK

CLASSIFICATION OF TURKISH E-COMMERCE PRODUCT REVIEWS

Başkent University Institute of Science and Engineering

Department of Electrical and Electronics Engineering

2023

Nowadays, e-commerce product reviews play a very important role in online shopping. With the importance of technology in our lives and the intense interest in online shopping, the classification of these comments with text classification quite important. In addition to supervised and unsupervised machine learning algorithms, which are very useful and effective in text classification, deep learning algorithms are also very popular and have high success rates. The aim of the study is to provide a brief overview of machine learning methods for text classification. In the data set used, the comments under a product on online shopping sites were collected and a dataset was created by giving class labels according to whether the comments were positive, negative or neutral. There are 6799 positive, 6978 negative and 1393 neutral comments in the data set, which includes a total of 15170 comments. In the classification phase, Convolutional Neural Network (CNN) is proposed as a classifier. Also the proposed method is compared with Decision Trees, Linear Discriminant Analysis, Long Short Term Memory, Quadric Discriminant Analysis, Bidirectional Long Short Term Memory, Efficient Linear Support Vector Machines, Gated Recurrent Unit, Efficient Logistic Regression, Naive Bayes, K-Nearest Neighbor , Ensemble Models, Artificial Neural Networks, Kernel and Support Vector Machines. The highest success was obtained with 90.77% accuracy when using Convolutional Neural Networks.

KEYWORDS: Convolutional Neural Networks, Long Short Term Memory, Support Vector Machines, Turkish Text Classification, Text Classification

ÖNSÖZ

Günümüzde küresel anlamda e-ticaret üzerinden yapılan alışverişlerin popülerliği giderek artmaktadır. Bu sebeple e-ticaret sayfalarına yapılan yatırımlar, e-ticaret uygulamalarına yapılan geliştirmeler de oldukça önem arz etmektedir. Bu tez çalışmasında önem dikkate alınarak kullanıcılara kolaylık sağlayabileceği, ürün değerlendirmelerinin doğruluğunu arttıracakı düşünülen ürün yorumları sınıflandırması yapılmıştır. Bu sınıflandırma süresince çeşitli metin ön işleme tekniklerinin uygulanmasının yanı sıra öznitelik çıkarımı ve sınıflandırma için de çeşitli algoritmalar kullanılmıştır. Yapılan bütün deneyler analiz edilmiş ve en yüksek başarı kriterlerine sahip algoritma bu sınıflandırma işlemi için önerilmiştir. Tezin yürütülmesi ve geliştirilmesi konusunda çalışmaya olan önemli katkılarından dolayı Başkent Üniversitesi Doç. Dr. Selda Güney' e teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR.....	i
ÖZET.....	ii
ABSTRACT	iii
ÖNSÖZ	iv
İÇİNDEKİLER	v
TABLolar LİSTESİ.....	vii
ŞEKİLLER LİSTESİ.....	viii
SİMGELER VE KISALTMALAR	xii
1. GİRİŞ.....	1
1.1. Çalışmanın Amacı.....	4
1.2. Tezin Organizasyonu	5
2. E-TİCARET VE MAKİNE ÖĞRENMESİ	6
2.1. E-Ticaretin Avantajları	6
2.2. E-Ticaretin Dezavantajları	7
2.3. Küresel E-Ticaret.....	7
2.4. Türkiye’de E-Ticaret	8
2.5. Makine Öğrenmesi Yöntemlerinin Kullanıcıya Yönelik Faydaları	8
2.6. Makine Öğrenmesi Yöntemlerinin E-Ticaret Platformlarına Yönelik Faydalar.....	9
3. LİTERATÜRDEKİ ÇALIŞMALAR.....	11
4. KULLANILAN VERİ SETİ	16
5. KULLANILAN SINIFLANDIRMA ALGORİTMALARI.....	18
5.1. Makine Öğrenmesi.....	18
5.1.1. Denetimli makine öğrenmesi	19
5.1.2. Denetimsiz makine öğrenmesi	19
5.1.3. Yarı denetimli makine öğrenmesi	20
5.1.4. Pekiştirmeli makine öğrenmesi	20
5.2. Klasik Makine Öğrenmesi Algoritmaları	20
5.3. Derin Öğrenme Algoritmaları	22
5.3.1. Uzun kısa süreli bellek (UKSB).....	23
5.3.2. Çift yönlü uzun kısa süreli bellek (Çift yönlü UKSB)	25

5.3.3. Evrişimsel sinir ağıları (ESA).....	25
5.3.4. Geçitli tekrarlayan birim (GTB).....	27
5.4. Başarı Kriterleri.....	28
6. UYGULANAN YÖNTEMLER.....	30
6.1. Metin Verisinin Ön İşlemesi	30
6.2. Öznitelik Çıkarımı	32
6.3. Sınıflandırma	32
6.3.1. DVM sınıflandırma algoritmasının uygulanması	32
6.3.2. UKSB sınıflandırma algoritmasının uygulanması.....	33
6.3.3. Çift yönlü UKSB sınıflandırma algoritmasının uygulanması	40
6.3.4. ESA sınıflandırma algoritmasının uygulanması.....	47
6.3.5. GTB sınıflandırma algoritmasının uygulanması	54
7. SONUÇ VE ÖNERİLER.....	63
KAYNAKLAR	65

TABLÖLAR LİSTESİ

	Sayfa
Tablo 1.1. Uygulanan yöntemler	2
Tablo 3.1. Literatür çalışmaları	13
Tablo 4.1. Veri setine ait örneklendirme	16
Tablo 6.1. Tokenizasyon örneklendirmesi	31
Tablo 6.2. 30 epok için UKSB deney sonuçları	34
Tablo 6.3. UKSB öznitelik çıkarımı kullanıldığında deney sonuçları	37
Tablo 6.4. 30 epok için Çift Yönlü UKSB deney sonuçları	40
Tablo 6.5. Çift Yönlü UKSB öznitelik çıkarımı kullanıldığında deney sonuçları	44
Tablo 6.6. 30 epok için ESA gözlem sonuçları	47
Tablo 6.7. ESA öznitelik çıkarımı kullanıldığında deney sonuçları	52
Tablo 6.8. 30 epok için GTB deney sonuçları	55
Tablo 6.9. GTB öznitelik çıkarımı kullanıldığında deney sonuçları	58
Tablo 6.10. Kullanılan sınıflandırma algoritmalarına göre elde edilen başarı değerleri	61
Tablo 6.11. Aynı veri setine sahip çalışmalarla karşılaştırma	62

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. Küresel anlamda e-ticaretin yıllara göre satış hacmi.....	7
Şekil 2.2. E-ticaretin Türkiye'nin ekonomik kalkınma vizyonuna desteği.....	8
Şekil 4.1. Kullanılan veri setindeki kelime dağılımları	17
Şekil 5.1. Yapay zeka algoritmalarının alt dalları	18
Şekil 5.2. Makine öğrenmesi türleri.....	19
Şekil 5.3. Karar Ağacı yapısı [15]	21
Şekil 5.4. DVM çalışma prensibi [17]	22
Şekil 5.5. YSA mimarisi [19]	23
Şekil 5.6. UKSB mimarisi [22].....	24
Şekil 5.7. Çift Yönlü UKSB mimarisi [25]	25
Şekil 5.8. ESA mimarisi [27].....	26
Şekil 5.9. Evrişim katman süreci [28].....	26
Şekil 5.10. İki farklı yöntem kullanılarak yapılan havuzlama işlemi [28].....	27
Şekil 5.11. GTB mimarisi [29]	27
Şekil 6.1. Metin sınıflandırma süreci blok diyagramı.....	30
Şekil 6.2. Kelime çantası modeli özellikleri.....	32
Şekil 6.3. DVM karışıklık matrisi (çapraz doğrulama yüzdesi=20)	33
Şekil 6.4. UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	34
Şekil 6.5. UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	35
Şekil 6.6. UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5).....	35
Şekil 6.7. UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5).....	36

Şekil 6.8. Kelime çantası modeli uygulandığında UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)	36
Şekil 6.9. Kelime çantası modeli uygulandığında UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	37
Şekil 6.10. UKSB öznitelik çıkarımında Yapay Sinir Ağları (Dar) karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)	39
Şekil 6.11. UKSB öznitelik çıkarımında Yapay Sinir Ağları (Dar) ROC eğrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	39
Şekil 6.12. Çift Yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30).....	41
Şekil 6.13. Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)	42
Şekil 6.14. Çift yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)	42
Şekil 6.15. Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)	43
Şekil 6.16. Kelime çantası modeli uygulandığında Çift yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)	43
Şekil 6.17. Kelime çantası modeli uygulandığında Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)	44
Şekil 6.18. Çift Yönlü UKSB öznitelik çıkarımında Yapay Sinir Ağları (Orta) karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)	46
Şekil 6.19. Çift Yönlü UKSB öznitelik çıkarımında Yapay Sinir Ağları (Orta) ROC eğrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)	46
Şekil 6.20. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=30).....	48
Şekil 6.21. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=30).....	49

Şekil 6.22. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30).....	49
Şekil 6.23. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30).....	49
Şekil 6.24. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5).....	50
Şekil 6.25. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5).....	50
Şekil 6.26. Kelime çantası modeli uygulandığında ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)	51
Şekil 6.27. Kelime çantası modeli uygulandığında ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5).....	51
Şekil 6.28. ESA öznitelik çıkarımında Verimli Lojistik Regresyon karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)	53
Şekil 6.29. ESA öznitelik çıkarımında Verimli Lojistik Regresyon ROC eğrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5).....	54
Şekil 6.30. GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	55
Şekil 6.31. GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30).....	56
Şekil 6.32. GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5).....	56
Şekil 6.33. GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5).....	57
Şekil 6.34. Kelime çantası modeli kullanıldığında GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)	57
Şekil 6.35. Kelime çantası modeli kullanıldığında GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)	58

Şekil 6.36. GTB öznitelik çıkarımında Verimli Lineer DVM karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5).....	60
Şekil 6.37. GTB öznitelik çıkarımında Verimli Lineer DVM ROC eğrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)	60



SİMGELER VE KISALTMALAR

ADB	Adaptive Data Boosting (Uyarlanabilir Veri Yükseltme)
Çift Yönlü UKSB	Çift Yönlü Uzun Kısa Süreli Bellek
DN	Doğru Negatif
DP	Doğru Pozitif
DVM	Destek Vektör Makineleri
ESA	Evrişimsel Sinir Ağları
GB	Gradient Boosting (Gradyan Yükseltme)
GTB	Geçitli Tekrarlayan Birim
K-EYK	K En Yakın Komşu
LDA	Lineer Diskriminant Analiz
LR	Lojistik Regresyon
MLP	Çok Katmanlı Algılayıcı (Multilayer Perceptron)
NB	Naif Bayes
RO	Rastgele Orman
SMO	Sıralı Minimum Optimizasyon
TSA	Tekrarlayan Sinir Ağları
UKSB	Uzun Kısa Süreli Bellek
XGBoost	Aşırı Gradyan Yükseltme (Extreme Gradient Boosting)
YSA	Yapay Sinir Ağları
YN	Yanlış Negatif
YP	Yanlış Pozitif

1. GİRİŞ

Çevirim içi alışveriş platformlarının zaman ve enerji tasarrufu yapılmasına olanak sağlaması, her yaş grubuna hitap etmesi ve oldukça geniş ürün çeşitliliğine sahip olması kullanıcı kitlesini giderek artırmaktadır. Ancak çevirim içi alışveriş yaparken ürünlerin gözle görülememesi ürün değerlendirmesini zorlaştırmaktadır. Bu da e-ticaretin önündeki en büyük engel olarak gösterilebilir. Tüketicilerin çevirim içi alışveriş yaparken en büyük referansları ürün yorumları olmuştur. Bu sebeple ürün yorumlarının ürünlere ait duygulara göre sınıflandırılması ürün seçiminde ve değerlendirmesinde önemli ölçüde kolaylık sağlamaktadır. Kullanıcılar ürüne ait yorumlarının sınıflandırılması sonucunda elde edilen sonuçları dikkate alarak hareket ettiğinde daha az bir eforla daha doğru seçimler yapabilir.

Metin sınıflandırma e-ticaret ürün yorumları üzerinden duygu analizi yapılmasının yanı sıra film yorumlarında duygu analizi, sosyal medyada yazılan yorumlar üzerinden hakaret tespiti ve şikayet sayfalarına yazılan yorumların güvenilebilirliğinin derecesi gibi birçok alanda da kullanılabilir.

Literatürdeki çalışmalar da göz önüne alındığında metin sınıflandırmaya olan ilginin ve kullanım alanlarının giderek arttığı görülebilir. Burada en büyük eksiklik ise Türkçe metinlere ait sınıflandırmanın literatürdeki diğer çalışmalara oranla oldukça az sayıda olmasıdır.

Bu sebeple bu çalışmada Türkçe metin sınıflandırılmasına yönelik literatürdeki eksiklik, e-ticaretin günümüzdeki gelişimi ve tüketicilerin ürün yorumlarındaki analiz ihtiyaçları göz önüne alınarak Türkçe e-ticaret ürün yorumlarının sınıflandırılması yapılmıştır. Çalışmada çeşitli sınıflandırma algoritmaları kullanılarak ürün yorumlarının en yüksek doğruluk yüzdesi elde edilecek şekilde sınıflandırılması ve literatürdeki eksikliklerin olabildiğince önüne geçilebilmesi amaçlanmıştır.

En uygun yöntemin belirlenmesi aşamasında öznitelik çıkarımı için kelime çantası modeli, Uzun Kısa Süreli Bellek, Çift Yönlü Uzun Kısa Süreli Bellek, Evrişimsel Sinir Ağları ve Geçitli Tekrarlayan Birim algoritmaları uygulanmıştır. Sınıflandırma yapılırken ise Karar Ağaçları, Lineer Diskriminant Analiz, Uzun Kısa Süreli Bellek, İkinci Dereceden Diskriminant Analiz, Çift Yönlü Uzun Kısa Süreli Bellek, Verimli Lineer Destek Vektör Makineleri, Geçitli Tekrarlayan Birim, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel, Destek Vektör Makineleri ve Evrişimsel Sinir Ağları algoritmaları kullanılarak çeşitli testler yapılmıştır. Uygulanmış olan

yöntemler Tablo 1.1’de gösterilmiştir. Yapılan testler sonucunda ürün yorumlarının sınıflandırılmasında en yüksek doğruluk oranını veren yöntem olarak hem öznitelik çıkarımında hem de sınıflandırma aşamasında ESA algoritmasının kullanıldığı yöntem önerilmiştir.

Tablo 1.1. Uygulanan yöntemler

Metin Verisinin Önışlemesi	Öznitelik Çıkarımı	Sınıflandırma
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Kelime Çantası Modeli	Destek Vektör Makineleri
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Kelime Çantası Modeli	Uzun Kısa Süreli Bellek
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Kelime Çantası Modeli	Çift Yönlü Uzun Kısa Süreli Bellek
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Kelime Çantası Modeli	Evrişimsel Sinir Ağları
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü	Kelime Çantası Modeli	Geçitli Tekrarlayan Birim

Noktalama İşaretleri ve Boşlukların Kaldırılması		
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Uzun Kısa Süreli Bellek	Karar Ağaçları, Destek Vektör Makineleri, Lineer Diskriminant Analiz, Uzun Kısa Süreli Bellek, İkinci Dereceden Diskriminant Analiz, Verimli Lineer Destek Vektör Makineleri, , Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Çift Yönlü Uzun Kısa Süreli Bellek	Karar Ağaçları, Destek Vektör Makineleri, Lineer Diskriminant Analiz, Çift Yönlü Uzun Kısa Süreli Bellek, İkinci Dereceden Diskriminant Analiz, Verimli Lineer Destek Vektör Makineleri, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Evrişimsel Sinir Ağları	Karar Ağaçları, Destek Vektör Makineleri, Lineer Diskriminant Analiz, Evrişimsel Sinir Ağları, İkinci Dereceden Diskriminant Analiz, Verimli Lineer Destek

		Vektör Makineleri, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel
• Tokenizasyon • Büyük/Küçük Harf Dönüşümü • Noktalama İşaretleri ve Boşlukların Kaldırılması	Geçitli Tekrarlayan Birim	Karar Ağaçları, Destek Vektör Makineleri, Lineer Diskriminant Analiz, Çift Yönlü Geçitli Tekrarlayan Birim, İkinci Dereceden Diskriminant Analiz, Verimli Lineer Destek Vektör Makineleri, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Kernel

1.1. Çalışmanın Amacı

Ürün yorumlarına bakılarak ürün hakkında doğru değerlendirmeyi yapabilmek e-ticaretteki en önemli noktadır. Ürünler potansiyel müşteriler tarafından ne kadar doğru değerlendirilerek satın alınırsa müşteri memnuniyeti o kadar artacak ve satıcının ve/veya e-ticaret uygulamasının tercih edilebilirliği o kadar artacaktır.

Kullanıcılar ürün yorumlarının üst sıralarda görünmesi amacı ile olumsuz yaptıkları yorumlara yüksek puanlar verebilmektedir. Bu durum ürünler incelenirken tüketiciler tarafından en çok karşılaşılan problemlerden biridir. Bu sebeple, bu çalışma makine öğrenmesi ve derin öğrenme algoritmalarını kullanarak metin sınıflandırma işlemi ile ürüne ait yorumlar üzerinden bir duygu analizinde bulunmayı amaçlamaktadır. Bu çalışma kullanılarak ürünlere ait daha doğru değerlendirmeler yapılabilecektir.

1.2. Tezin Organizasyonu

Elde edilen ürün yorumlarının makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak sınıflandırılması sonucu ürün yorumlarının duygu analizi ile başarılı bir şekilde sınıflandırılmasını amaçlayan bu tez çalışması, aşağıdaki gibi dokuz farklı bölümden oluşacak şekilde kategorize edilmiştir.

Çalışmanın birinci bölümünde tez çalışmasının amacı doğrultusunda yapılan çalışmalardan bahsedilmiş, tezin organizasyonu sunulmuştur.

Tezin ikinci bölümünde e-ticaret ve e-ticaret uygulamaları hakkında genel bilgiler verilmiş olup çevirim içi olarak alışveriş imkânı sunan bu uygulamaların hayatımızdaki yerinden bahsedilmiştir. Ayrıca ülkemizde ve küresel boyutta e-ticaretin öneminden ve güncel durumundan bahsedilmiştir. Bunun yanı sıra yine aynı bölümde makine öğrenmesi ve derin öğrenme algoritmalarının e-ticaret üzerindeki etkisinden ve öneminden bahsedilmiştir.

Çalışmanın üçüncü bölümünde farklı alanlarda yapılan çalışmalar da dahil olmak üzere metin sınıflandırma ile yapılan duygu analizlerine ait literatürdeki diğer çalışmalar ve bu çalışmalarda kullanılan yöntemlere değinilmiştir.

Çalışmanın dördüncü bölümünde bu tez çalışmasında kullanılmış olan veri setinin içeriğinden bahsedilmiştir.

Çalışmanın beşinci bölümünde ise metin verisinin sınıflandırma uygulanabilir hale gelmesi amacıyla yapılan veri ön işleme süreci, bu tez çalışması sürecinde kullanılmış olan sınıflandırma algoritmaları çalışma prensipleri detaylandırılarak anlatılmıştır.

Çalışmanın altıncı bölümde yapılmış olan çeşitli deneyler elde edilen sonuçlar sunulmuştur.

Çalışmanın son bölümünde çalışmaya ait sonuçlar değerlendirilmiş ve gelecek çalışmalara ışık tutması amacıyla önerilere yer verilmiştir.

2. E-TİCARET VE MAKİNE ÖĞRENMESİ

Kısaca e-ticaret olarak tanımlanan elektronik ticaret, ürün ve hizmetlerin tüketiciler ile internet ortamında buluşturulduğu ticaret türüdür. E-ticarete ürünlerin reklamı, satışı, ödemesi, kullanıcıya ulaştırılması ve kullanıcının yaşadığı problemlere yönelik destek ve yönlendirmenin sağlanması da internet üzerinden gerçekleştirilmektedir.

2.1. E-Ticaretin Avantajları

E-ticaretin avantajları aşağıdaki şekilde sıralanabilir.

- 7 gün 24 saat hizmet verilebilir olması.
- E-ticaret siteleri üzerinden sanal bir mağaza açmanın maliyeti fiziksel bir mağaza açmanın maliyetine kıyasla oldukça azdır.
- Tüketici için istediği ürüne erişebilmek e-ticaret platformlarında fiziksel mağazalara kıyasla daha kolay, daha kısa sürede ve daha az maliyetle erişilebilir durumdadır.
- Tüketiciler aynı ürünlerin farklı satıcılar üzerinden daha uygun fiyatlı muadillerini kolaylıkla görebilir ve ürün seçimini buna göre yapabilir.
- İnternet reklamcılığı ve e-ticaret kampanya yönetimi sayesinde işletmeler kampanyalarını çok daha hızlı bir şekilde duyurup satışlarını arttırabilmektedir.
- Çevirim içi yapılan her türlü tıklanma, satış, beğenme gibi işlemlerin internet ortamında kayıtlı tutulması sayesinde kullanıcıların arz ve taleplerine yönelik her türlü analiz daha kolay yapılabilir.
- E-ticaret platformları markasını küreselleştirmek isteyen işletmelere bu isteklerini gerçekleştirebilmeleri için daha olnaklı bir ortam hazırlar.
- Tüketiciler almak istedikleri ürünlerin yorumlarını okuyarak ilk izlenimden bir süre kullanım sonrası izlenime kadar ürünü satın alan diğer tüketicilerin düşüncelerine kolaylıkla ulaşabilir.

2.2. E-Ticaretin Dezavantajları

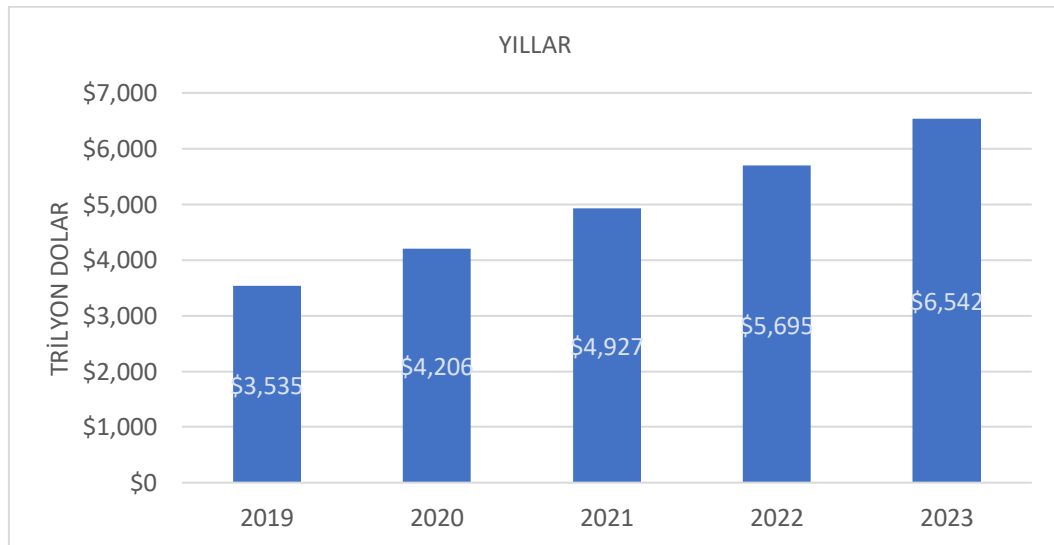
E-ticaretin avantajlarının oldukça fazla olduğu gibi tüketicileri e-ticaretten uzaklaştıran, e-ticaret kullanımını azaltan bazı dezavantajlar da bulunmaktadır. Bu dezavantajlar aşağıdaki şekilde sıralanabilir.

- Tüketicinin ürünü canlı olarak görme, inceleme ve deneme imkânı yoktur. Bu sebeple de iade oranları fiziksel mağazalardan yapılan satışlara göre e-ticaret platformlarında daha fazladır.
- Potansiyel müşteriler e-ticaret ortamlarında kredi kartı, adres, ad ve soy ad bilgilerini paylaşmaktan çekinebilir.
- Ürün satın alındıktan sonra beklenmesi gereken bir teslimat süresi vardır.

2.3. Küresel E-Ticaret

Tarihsel olarak çevirim içi alışveriş sistemi hayatımıza 1990’larda girmiştir. Hala lider konumda olan Amazon ve eBay firmaları 1995 yılında ortaya çıkmıştır. 20 yıldan fazla bir süredir aktif olarak hayatımızda olan e-ticaretin küresel satış hacmi 2022 yılının sonunda yapılan analizlere göre 5.6 trilyona ulaşmıştır [1].

Özellikle gelişmekte olan ülkelerde yer alan orta gelire ait nüfusun ve bu nüfusa ait harcama eğilimlerinin hızla artmasıyla beraber e-ticarete olan ilginin de hızla arttığı görülmektedir.



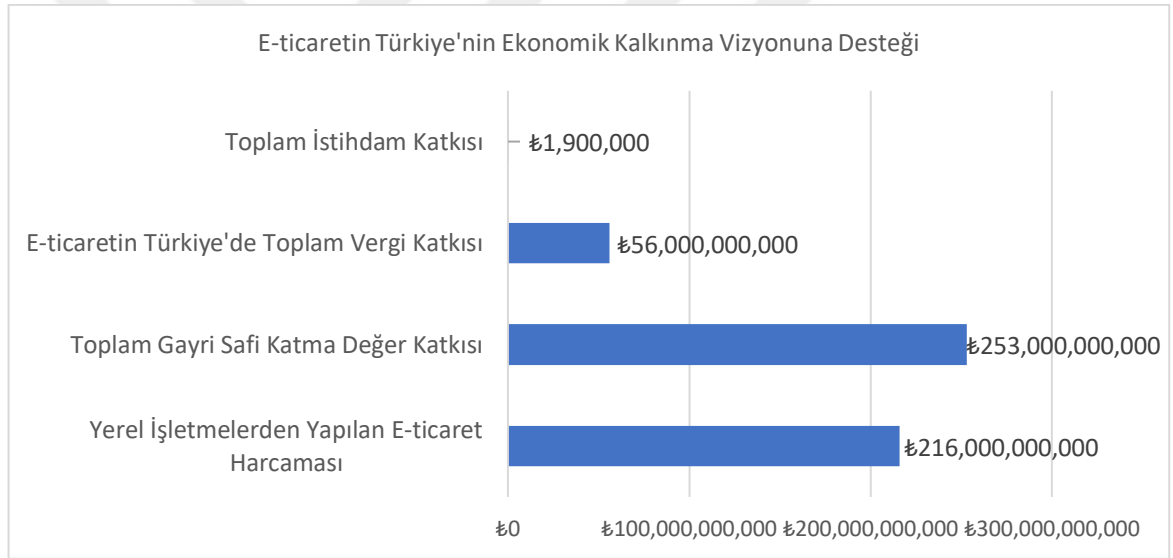
Şekil 2.1. Küresel anlamda e-ticaretin yıllara göre satış hacmi

2.4. Türkiye’de E-Ticaret

Türkiye’de de küresel piyasada olduğu gibi e-ticaret platformlarına yönelen ilgi katlanarak artmaktadır. Bu ilginin artmasına bağlı olarak da e-ticaretten elde edilen gelirlerin ülke ekonomisine olan katkısı da önemli bir yer oluşturmaktadır.

2020 TÜBİSAD (Türkiye Bilişim Sanayicileri Derneği) verilerine göre Covid-19 salgınının da büyük ölçüde etkili olmasıyla beraber her 5 kişiden 4’ü e ticareti deneyimlemiştir ve buna bağlı olarak e-ticaretin kullanımının artmasıyla GSYH (Gayri Safi Yurtiçi Hasıla)’ye oranının %4,5’a ulaşmış ve Türkiye olgun pazarlar arasına girmiştir [2].

E-ticaretin gelişmesinin Türkiye ekonomisinin kalkınmasında da önemli bir rol oynadığı söylenebilir.



Şekil 2.2. E-ticaretin Türkiye’nin ekonomik kalkınma vizyonuna desteği

2.5. Makine Öğrenmesi Yöntemlerinin Kullanıcıya Yönelik Faydaları

Hayatımızın her alanında olduğu gibi e-ticarete de makine öğrenmesi ve derin öğrenme algoritmaları oldukça etkili bir rol oynamaktadır. Sınıflandırma amacıyla kullanılan bu algoritmalar hem kullanıcı hem de e-ticaret platformları için birçok açıdan fayda sağlamaktadır.

E-Ticaret platformlarında yapılan ürün yorumları kullanıcılara satın alınan ürünün; kalite, fiyat/performans beklentisini karşılama ve ürünün gerçekten belirtilen özelliklere

sahip olması gibi değerlendirmelerin yapılabilmesi konusunda oldukça kolaylık sağlamaktadır.

Ürünlerin satın alınırken verilen puanlara göre değerlendirilmesi ise bazı durumlarda ürünün yanlış değerlendirilmesine sebep olmaktadır. Bu durumla ilgili örnek olarak, kullanıcının olumsuz yorumunun diğer yorumlar arasında üst sırada görülmesi veya yüksek puanıyla dikkat çekmesi amacıyla yüksek puan verilerek yazılan olumsuz yorumlar verilebilir.

Ürünlerin sadece puanlarına bakılarak yanlış değerlendirilmesi sebep olan bir diğer durum ise yorumu yapan bazı kullanıcıların olumlu veya olumsuz olan düşüncelerini her zaman puanlama sisteminde doğru bir şekilde karşılıklılandırmamasıdır.

Bu durumlara aşağıdaki yorumlar örnek verilebilir.

- Yorumun üstte görünmesi için 5 puan verdim, ürün hiç beklediğim gibi çıkmadı kumaşı çok kalitesiz ve lekeli geldi. (*****)
- Ürün kesinlikle görseldeki ile aynı değil ama satıcının ilgili olması ve hızlı bir şekilde ürünü kargolamasından dolayı 4 puan verdim. (****)
- Arkadaşlar yorum dikkat çeksin diye 5 puan veriyorum kesinlikle almayın çok kalitesiz verdiğiniz paraya değmiyor. (*****)
- Ürün çok güzel ama bedenler çok küçük bluzu kafamdan geçiremedim (**)
- Renkleri fotoğraftakinden daha soluk ama kaliteli beğendim. (***)

Bu sebeple ürünlerin en doğru şekilde değerlendirilebilmesi için ürün yorumlarının incelenmesi şarttır. Bu incelemede de makine öğrenmesi ve derin öğrenme algoritmaları kullanarak yorumların olumlu veya olumsuz olma durumlarına göre sınıflandırılabilmesi kullanıcılar için ürün analizlerini oldukça kolaylaştıracaktır.

2.6. Makine Öğrenmesi Yöntemlerinin E-Ticaret Platformlarına Yönelik Faydalar

E-Ticaret platformları da yapılan ürün yorumlarını sınıflandırılması ile:

- Arama yapılan ürünlere ait sonuçlarda en çok olumlu yoruma sahip olan ürünün en başta gösterilmesini sağlayabilir. Mevcut yapılarda da

kullanılan bu filtrelemenin sınıflandırma algoritmaları kullanılarak iyileştirilmesi sağlanabilir.

- Aynı ürünün satışını yapan fakat ürüne ait yüklenen görseller ile ürünün gerçekliği arasındaki bağlantıya, kargolama hızı müşteri ile iletişimi gibi birçok değerlendirmeye göre olumlu/olumsuz yorum oranları farklılık gösteren satıcıların da puanlandırılması konusunda ürünlerin sınıflandırılması sonucunda elde edilen veriler kullanılabilir.



3. LİTERATÜRDEKİ ÇALIŞMALAR

Literatürde metin sınıflandırma üzerine oldukça fazla çalışma mevcuttur. Bu çalışmaların büyük bir kısmı ise duygu analizi amacıyla yapılan sınıflandırmalardır. Duygu analizi yapabilmek için seçilen başlıca metinler ise haber metinleri, sosyal medya ağlarında yazılan yorumlar ve e-ticaret platformlarında yer alan ürün yorumlarıdır. Çeşitli metin verileri, veri ön işleme metotları ve sınıflandırma algoritmaları kullanılarak yapılmış birçok çalışma bulunmaktadır.

Bu çalışmalardan birisi Hatice Nizam ve Saliha Sıla Akın tarafından 2017 yılında gerçekleştirilen, makine öğrenmesi temelli yaklaşımlardan biri olan denetimli makine öğrenmesi yöntemleri ile duygu analizi çalışmasıdır [4]. Bu çalışmada, veri seti olarak gıda sektöründeki farklı üreticilerin farklı ürünleri ile ilgili (twitter'da) paylaşımlarından elde edilen 2 farklı veri seti kullanılmıştır. Veri setlerinin her ikisi de pozitif, negatif ve nötr olmak üzere 3 sınıfa ayrılmıştır. A veri setinde pozitif sınıfta 1113, negatif sınıfta 277 ve nötr sınıfta 610 veri olmak üzere toplam 2000 veri bulunmaktadır. B veri setinde ise pozitif sınıfta 257, negatif sınıfta 277 ve nötr sınıfta 290 veri olmak üzere toplam 824 veri bulunmaktadır. Çalışmada sınıflandırma için Naif Bayes (NB), Rastgele Orman (RO), Sıralı Minimum Optimizasyon (SMO) Sınıflandırması, Karar Ağacı-J48 Sınıflandırması, 1- En Yakın Komşular (K-EYK) - IB1 Sınıflandırma Algoritmaları kullanılmıştır. Sonuçlar incelendiğinde SMO Algoritması uygulandığında %72,33 doğruluk oranı ile en yüksek başarımın B veri setinde elde edildiği görülmüştür.

Fatih Samet Çetin ve Gülşen Eryiğit tarafından 2017 yılında yapılan çalışmada, Türkçe restoran yorumları veri seti kullanılarak duygu analizi yapılmıştır. Çalışma sürecinde kullanılan veri setinde toplamda 1415 adet cümle bulunmaktadır. Veriler pozitif, negatif ve nötr olmak üzere 3 sınıfa ayrılmıştır. Duygu sınıfını belirlemek için cümle analizi sonucunda hedef terimin komşu sözcüklerinden öznitelik seçimine dayalı doğrusal bir sınıflandırma yöntemi olan DVM kullanılmış ve en başarılı sonuç %76,1 doğruluk değeri elde edilmiştir [5].

2020 yılında Özer Çelik ve Gürkan Kaplan tarafından yapılan başka bir çalışmada yeniden örnekleme teknikleri kullanılarak SMS verisi üzerinde bir metin sınıflandırma çalışması yapılmıştır [6]. Bu çalışmada, çok sayıda kullanıcının telefonlarına gelen 4203 adet SMS verisine (Reklam SMS: 2374, Normal SMS: 1277, Spam SMS: 552) çeşitli makine öğrenme teknikleri ile metin sınıflandırması yapılmıştır. Sınıflandırma sonucunda SMS

verileri reklam, normal ve spam olmak üzere 3 farklı şekilde etiketlenmiştir. Çalışma süresince Lojistik Regresyon (LR), K-En Yakın Komşu, Destek Vektör Makineleri, Naif Bayes, Karar Ağaçları, Rastgele Orman, Uyarlanabilir Veri Yükseltme (ADB), Gradyan Yükseltme (GB), Aşırı Gradyan Yükseltme (XGBoost) ve Yapay Sinir Ağı (YSA) algoritmaları kullanılmış olup sınıflandırma sonucunda en yüksek doğruluk yüzdesini veren algoritma %80,1 ile Lojistik Regresyon algoritması olmuştur.

Özer Çelik ve Burak Can Koç'un 2021 yılında yaptıkları çalışmada 6 farklı haber sitesinden derlenerek hazırlanan bilim-teknoloji, eğitim, kültür-sanat, sağlık, siyaset ve spor kategorilerinde her bir kategoride 2000 metin olmak üzere toplam 12000 adet Türkçe haber metni bulunan bir veri seti üzerinde çalışmışlardır. Bu çalışmada TF-IDF, Word2Vec ve FastText vektör modellerinin sınıflandırma başarı oranları karşılaştırılması yapılmış ve FastText ile elde edilen vektör modelinin %95,75 ile en yüksek başarı oranını verdiği görülmüştür [7].

Türkçe ürün yorumları verisi ile duygu analizi yapılmış olan bir çalışma ise 2021 yılında Buğra Polat tarafından yapılmıştır [8]. Bu çalışmada kullanılan veri seti "Hepsiburada" adlı e-ticaret platformuna ait ürün yorumları ile oluşturulmuştur. Veri setinde toplamda 272,216 tane yorum bulunmaktadır. Kullanılan sınıflandırmada yorumlar olumlu ve olumsuz olmak üzere 2 farklı sınıfa ayrılmıştır. Sınıflandırmanın performansını arttırmak amacıyla FastText tarafından CBOW ve Skipgram yaklaşımıyla oluşturulan önceden eğitilmiş kelime gömme modelleri kullanılmıştır. Sınıflandırma algoritması olarak Evrişimsel Sinir Ağları (ESA), Uzun Kısa Süreli Bellek (UKSB) ve Geçitli Tekrarlayan Birim (GTB) algoritmaları kullanılmış olup en yüksek başarıyı %93.73 doğruluk yüzdesi ile GTB algoritması vermiştir.

2021 yılında yapılmış olan bir başka çalışmada ise Emrah Aydemir, Metin Işık ve Türker Tuncer tarafından Türkçe haber metinlerinin sınıflandırılması yapılmıştır [9]. Bu çalışmada kullanılan veri setinde her bir haber metni ortalama 1586 sözcükten oluşan 2248 adet haber metni toplanmıştır. Bu haber metinlerinden en uzun haberler kültür-sanat kategorisine ait iken en kısa haberler ise magazin kategorisinde olmuştur. Sınıflandırma sonucunda haberler yaşam, dünya, ekonomi, kültür-sanat, magazin, otomobil, spor ve teknoloji olacak şekilde 8 farklı sınıfa ayrılmıştır. Çalışmada Çok Terimli Naif Bayes ve Rastgele Orman algoritmaları kullanılmış olup en yüksek başarılı sınıflandırma %99,86 ile Rastgele Orman algoritması ile elde edilmiştir.

Metin sınıflandırma amacıyla 2021 yılında Ali Aycan Kolukısa tarafından yapılmış olan başka bir çalışmada ise hem Türkçe hem de İngilizce olan iki farklı veri seti ile deneyler

yapılmıştır [10]. İngilizce veri seti <https://www.kaggle.com/rmisra/news-category-dataset> adresinde önceden hazırlanmış olarak mevcut olan İngilizce haber içerik, başlık ve sınıf etiketleri gibi bilgilerin bulunduğu verilerden eleyerek alınan verilerdir. Veri setinden dengeli bir şekilde veri seçimi yapılarak veri seti 4 kategoriden 5'er cümle olacak şekilde 20 adet cümleye indirilmiştir. Türkçe veri setinde ise Türkçe makalelerin özet kısımlarından alarak hazırlanmış olan ve 3 farklı gruba ait 16'şar cümleden oluşan toplamda 48 adet cümle verisi kullanılmıştır. Sınıflandırma aşamasında yapay sinir ağı yapılarından olan UKSB, Çok Katmanlı Algılayıcı (MLP) ve Tekrarlayan Sinir Ağı (TSA) sınıflandırma algoritmaları kullanılmış olup Türkçe veri seti için en yüksek doğruluk yüzdesini %62,5 ile MLP algoritması verirken, İngilizce veri seti için en yüksek doğruluk yüzdesini %87,5 ile yine MLP algoritması vermiştir.

2021 yılında Hasibe Büşra Doğru, Sahra Tilki, Akhtar Jamil ve Alaa Ali Hameed tarafından haber metinleri kullanılarak yapılmış olan başka bir çalışmada ise yine hem Türkçe hem de İngilizce veri setleri kullanılarak 2 farklı deney yapılmıştır [11]. Türkçe haber metinleriyle yapılmış olan deneyde 3600 adet haber metninden oluşan TTC-3600 veri seti, İngilizce haber metinleriyle yapılmış olan deneyde ise 2225 adet haber metninden oluşan BBC-NEWS veri seti kullanılmıştır. Türkçe veri setinde 6 sınıf olacak şekilde, İngilizce veri setinde ise 5 sınıf olacak şekilde sınıflandırma yapılmıştır. En yüksek sınıflandırma doğruluğunu her iki veri seti için de ESA algoritması vermiş olup Türkçe veri setinde %94,17'lik, İngilizce veri setinde ise %96,41'lik doğru yüzdesi elde edilmiştir.

Bu çalışmadaki veri seti ile aynı veri setinin kullanıldığı, Barış Ulusoy tarafından 2022 yılında yapılmış olan çalışmada TF-IDF yöntemi ile birlikte Naif Bayes sınıflandırma algoritması kullanılarak %85,49'luk bir doğruluk değeri elde edilmiştir [12].

Bu çalışmadaki veri seti ile aynı veri setinin kullanıldığı Şeyma Sarıgil tarafından 2021 yılında yapılan çalışmada ise GTB algoritması kullanılarak %78,92'lik bir doğruluk değeri elde edilmiştir [13].

Tablo 3.1. Literatür çalışmaları

Yazarlar	Çalışma Yılı	Kullanılan Veri Seti	Metin Dili	Veri Sayısı	Sınıf Sayısı	Kullanılan Yöntem	Doğruluk Değeri
Hatice Nizam,	2017	Twitter verileri	Türkçe	2000	3	SMO algoritması	%66,40

Saliha Sıla Akın [4]		Twitter verileri	Türkçe	824	3	SMO algoritması	%72,33
Fatih Samet Çetin, Gülşen Eryiğit [5]	2017	Restoran yorumları verileri	Türkçe	1415	3	DVM algoritması	%76,1
Özer Çelik, Gürkan Kaplan [6]	2020	SMS verileri	Türkçe	4203	3	Lojistik Regresyon	%80,1
Özer Çelik, Burak Can Koç [7]	2021	Haber verileri	Türkçe	12000	6	FastText yöntemi (sınıflandırma algoritması paylaşılmamış)	%95,75
Buğra Polat [8]	2021	Anonymous 2018e veri seti	Türkçe	272216	2	Geçitli Tekrarlayan Birim algoritması	%93,73
Emrah Aydemir, Murat Işık, Türker Tuncer [9]	2021	Haber verileri	Türkçe	2248	8	Rastgele Orman algoritması	%99,86
Ali Aycan Kolukısa [10]	2021	News Category Dataset (https://www.kaggle.com/rmisra/news-category-dataset)	İngilizce	20	4	Çok Katmanlı Algılayıcı	%87,5
		Türkçe makale verileri	Türkçe	48	3	Çok Katmanlı Algılayıcı	%62,5

Hasibe Büşra Doğru, Sahra Tilki, Akhtar Jamil, Alaa Ali Hameed [11]	2021	TTC-3600 veri seti	Türkçe	3600	6	Evrişimsel Sinir Ağları algortiması	%94,17
		BBC-News veri seti	İngilizce	2225	5	Evrişimsel Sinir Ağları algortiması	%96,41
Barış Ulusoy [12]	2022	e- ticaret_urun _yorumlari. csv	Türkçe	15170	3	Naif Bayes	%85,49
Şeyma Sarıgil [13]	2021	e- ticaret_urun _yorumlari. csv	Türkçe	15170	3	Geçitli Tekrarlayan Birim algoritması	%78,92
Bu Çalışma	2023	e- ticaret_urun _yorumlari. csv	Türkçe	15170	3	Evrişimsel Sinir Ağları algoritması	%90,77

4. KULLANILAN VERİ SETİ

Çalışma süresince kullanılan veri seti, veri bilimi için oldukça sık kullanılan KAGGLE veri bilimi platformundan alınmıştır. Kullanılan veri seti birden fazla e-ticaret platformundan çeşitli ürünlerin altındaki yorumlar çekilerek hazırlanmıştır. Toplamda 15170 yorumun yer aldığı veri setinde 6799 olumlu, 6978 olumsuz ve 1393 nötr yorum yer almaktadır. Veri setinde belirtilen olumlu yorumlar kullanıcılar memnun olduğu, olumsuz yorumlar kullanıcıların memnun kalmadığı, nötr yorumlar ise kullanıcılar ne tam anlamıyla memnun ne de tam anlamıyla memnun kalmadığı durumlarda yazdıkları yorumlar olarak da düşünülebilir.

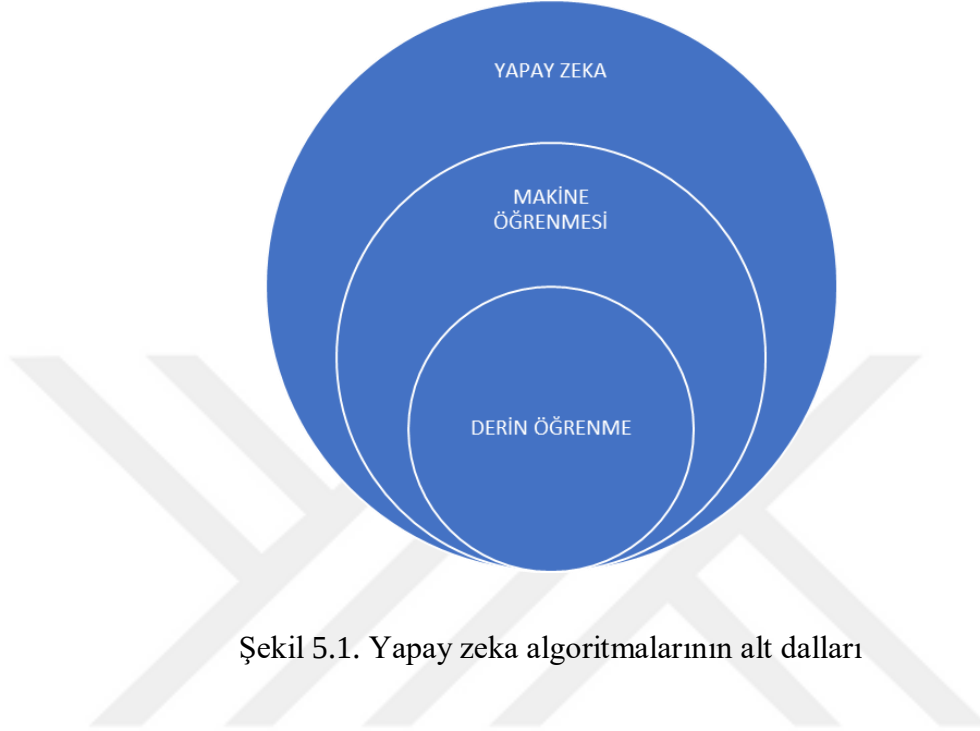
Veri setinin uygulanan sınıflandırma yapısındaki girdileri yazılan yorumlardır. Metin verisindeki olumlu, olumsuz ve nötr olma durumu ise sınıflandırmanın çıktılarıdır. Veri setinin örneklendirilmesi amacıyla Tablo 4.1’de verilerin bir parçası paylaşılmıştır.

Tablo 4.1. Veri setine ait örneklendirme

evet anlatıldığı gibi	1
Daha öncede aynısını almıştım çok güzel ve kaliteli bir ürün.	1
35 TL ye berber makinesi powertec almıştım. 300 liraya aldığım dünya markası yerine onu kullandım. powertec yine şaşırtmadı harika bir ürün. genelde sakal kesiyorum çok iyi kesiyor. hiç yoruma felan bakmayın daha alın gitsin.	1
Ürün sadece aldığım gün güzel bir şekilde kesti. sonrasında saatlerce uğraşıp sadece sakalımı toplayabildim. Allah kahretsin başka bı makine alacam ama bu sefer powertec mi ? ASLA	0
urun gelmiyor 10 gundur bekliyoruz magaza iletisim sifir siparis tamamlandi diyor urun gelmiyor dolandırıldik	0
berbat ötesi paketleme hasar görmüş kap ve çizik olan gözlük baya baya kötü	0
Görüntü kalitesi bir tık düşük ama fena değil	2
Boyutu ufak sağlamlık olarak iş görür kilolu iseniz tavsiye etmem ama 55-60 kilolarında bir insan gayet rahat kullanır	2

5. KULLANILAN SINIFLANDIRMA ALGORİTMALARI

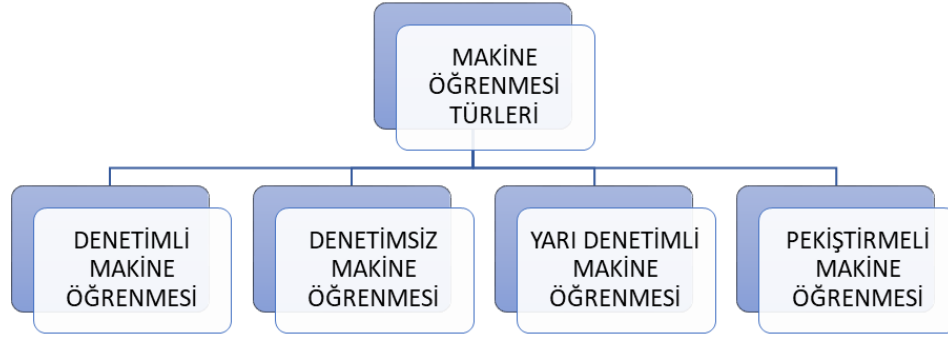
Sınıflandırma algoritmaları makine öğrenmesi altında farklı türlere göre incelenebilir.



5.1. Makine Öğrenmesi

Makine öğrenmesi, insanların öğrenme yapısından yola çıkılarak oluşturulan büyük verilerin işlenmesinde ve veriler üzerinden çıkarım yapılmasında kullanılan modeller olarak tanımlanabilir. Yapay zekanın bir alt kümesi olarak konumlandırılabilen makine öğrenmesi yöntemleri ile sınıflandırma, kümeleme ve regresyon (eğri uydurma) yapılarak geçmişteki veri referans alınıp oluşturulan modele göre yeni veriler için bir karar mekanizması çalışır.

Makine öğrenmesi yöntemleri denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme yöntemleri olarak Şekil 5.2’de de görüldüğü gibi 4 ana başlık altında incelenebilir.



Şekil 5.2. Makine öğrenmesi türleri

5.1.1. Denetimli makine öğrenmesi

Denetimli öğrenme yöntemlerinde her bir verinin hangi kategoriye dahil olduğunu gösteren etiketleri vardır. Etiketlenmiş veriler algoritma içerisinde tahmin/karar verme adımında alınır. Sınıflandırma ve regresyon (eğri uydurma) denetimli makine öğrenmesine örnek verilebilir.

Denetimli öğrenmede, makine örnek olarak öğretilir. Operatör, makine öğrenimi algoritmasına istenen girdileri ve çıktıları içeren bilinen bir veri kümesi sağlar ve algoritmanın bu girdilere ve çıktılara nasıl ulaşılacağını belirlemek için bir yöntem bulması gerekir. Operatör sorunun doğru cevaplarını bilirken, algoritma verilerdeki kalıpları tanımlar, gözlemlerden öğrenir ve tahminler yapar. Algoritma tahminlerde bulunur ve operatör tarafından düzeltilir ve bu süreç, algoritma yüksek düzeyde doğruluk/performans elde edene kadar devam eder.

5.1.2. Denetimsiz makine öğrenmesi

Denetimsiz makine öğrenme yöntemlerinde denetimli yöntemlerin aksine verilerin kategorize edildikleri etiketleri yoktur. Denetimsiz makine öğrenmesinde amaç oluşturulan algoritmanın veriler arasında bağlantı kurup gruplara ayırmaktır. Kümeleme işlemi denetimsiz makine öğrenmesine örnek verilebilir.

Burada makine öğrenimi algoritması, kalıpları belirlemek için verileri inceler. Talimat verecek bir cevap anahtarı veya insan operatör yoktur. Bunun yerine makine, mevcut verileri analiz ederek korelasyonları ve ilişkileri belirler. Denetimsiz bir öğrenme sürecinde, büyük

veri kümelerini yorumlamak ve bu verileri buna göre adreslemek için makine öğrenimi algoritmasına bırakılır. Algoritma, yapısını açıklamak için bu verileri bir şekilde düzenlemeye çalışır. Bu, verileri kümeler halinde gruplandırmak veya daha düzenli görünecek şekilde düzenlemek anlamına gelebilir.

5.1.3. Yarı denetimli makine öğrenmesi

Yarı denetimli makine öğrenmesinde, denetimli makine öğrenmesinden farklı olarak etiketlenmiş verilerin yanında etiketlenmemiş veriler de kullanılır. Burada anlamlı olarak etiketlenmiş veriler algoritmanın verileri anlayabilmesini sağlarken etiketsiz veriler algoritma için tamamen yeni verilerdir. Bu şekilde algoritma etiketli verilere dayanarak etiketsiz verileri etiketlemeyi öğrenir.

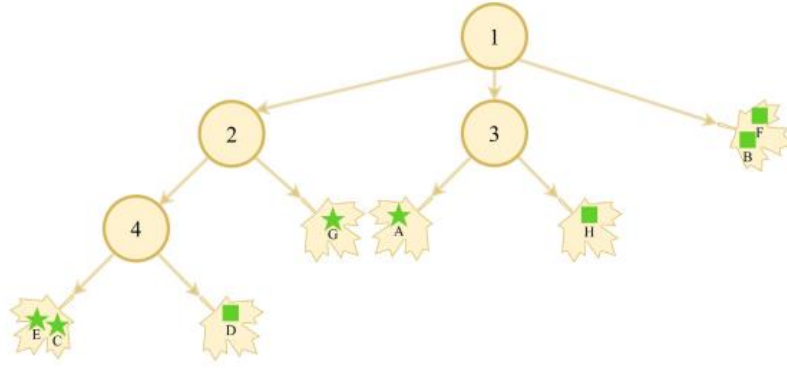
5.1.4. Pekiştirmeli makine öğrenmesi

Pekiştirmeli öğrenme, amaca yönelik ne yapılması gerektiğini öğrenen bir makine öğrenmesi yaklaşımıdır. Bu yaklaşımda makine öğrenmesi algoritması kural tanımlayarak farklı olasılıkları keşfetmeye çalışır. Öğretici sistemin ürettiği doğru veya yanlış sonuçlarına göre bir ödül-ceza sistemi uygulanır ve oluşturulan sistem geçmiş deneyimlerden öğrenerek mümkün olan en iyi sonucu elde etmek için yaklaşımını duruma göre uyarlamaya başlar.

5.2. Klasik Makine Öğrenmesi Algoritmaları

Bu tez çalışması boyunca klasik makine öğrenmesi yöntemlerinden Karar Ağaçları, Lineer Diskriminant, İkinci Dereceden Diskriminant, Verimli Lineer DVM, Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller ve Destek Vektör Makineleri kullanılmıştır.

Karar ağacı yapısı Şekil 5.3’de görüleceği gibi yaprak ve düğüm noktalarından oluşan bir ağaç yapısına benzer yapıdadır. Ağaç yapısındaki her düğüm bir özneliği, her dal bir test sonucunu, her yaprak ise test sonuçlarında oluşan örnekleri temsil eder.



Şekil 5.3. Karar Ağacı yapısı [15]

Verileri doğrusal olarak birbirlerinden en iyi şekilde ayırmayı amaçlayan Lineer Diskriminant Analiz (LDA) veri kümeleri içerisindeki varyansı minimize, sınıflar arasındaki ayrımı maksimize ederek boyut azaltması yapan bir yöntemdir.

İkinci Dereceden Diskriminant Analiz, LDA'nın sık kullanılan varyantlarından biridir. Çalışma prensibi olarak aynı LDA gibi sınıflar arası ayrımı maksimum sınıf içi verilerin varyansını minimum tutmayı amaçlar.

Lojistik regresyon, lineer olarak ayrılabilir verilerde kullanılan istatistiksel bir makine öğrenmesi yöntemidir.

Naif Bayes algoritması Bayes teoremine bağlı, olasılıksal bir sınıflandırma algoritmasıdır. [15] Bu algortmada öğretilen veriler göz önüne alınarak yeni veriler tahmin edilirken her bir öznelik çiftinin diğerlerinden olasılıksal olarak bağımsız olduğu varsayımına dayanılır.

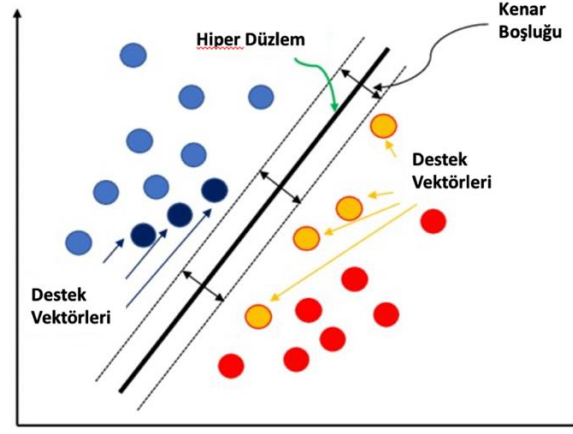
K-EYK algoritmasında veriler uzaklık hesabına dayanan bir benzerliğe göre sınıflandırılır. K değişkeni en yakın komşu sayısını ifade eder ve veri setindeki başarısına göre değişken ayarlanabilir.

Birleşik modeller sınıflandırıcıların performanslarını arttırmak amacıyla belirli sınıflandırıcılarla belirli metodların kombinasyonu olarak tanımlanabilir. Birleşik modellerden bazıları Güçlendirilmiş Ağaçlar, Torbalanmış Ağaçlar, Alt Uzay Diskriminant ve Alt Uzay K-EYK olarak örneklendirilebilir.

En basit tanımıyla Karar Ağaçlarına AdaBoost metodunun uygulanmasıyla güçlendirilmiştir ağaçlar, torbalama metodunun uygulanmasıyla ise torbalanmış ağaçlar birleşik modelleri oluşturulmuştur.

Diskriminant Analiz sınıflandırıcısına Alt Uzay metodunun uygulanması ile Alt Uzay Diskriminant birleşik modeli, K-EYK sınıflandırıcısına Alt Uzay metodunun uygulanması ile ise Alt Uzay K-EYK birleşik modeli oluşturulmuştur.

1990'lu yılların başında, Vladimir Vapnik ve arkadaşları tarafından AT T Bell Laboratuvarlarında geliştirilen Destek Vektör Makineleri sınıflandırma ve regresyon problemlerinde oldukça sık kullanılan makine öğrenmesi yöntemlerinden biridir [14]. Destek Vektör Makineleri çok boyutlu verileri sınıflara ayıran bir hiper düzlem elde ederek genelleştirilecek veriler için uygun bir sınıflandırıcı olmayı amaçlar. Lineer olarak ayrılabilen sınıfların belirlenmesinde sıklıkla kullanılan DVM modeli, kernel fonksiyonları sayesinde lineer ayrıştırılamayan girdi uzayını daha yüksek boyutlu lineer ayrılabilir uzaya taşıyarak lineer olmayan verilerin sınıflandırılmasında da başarıyla kullanılmaktadır. [16]



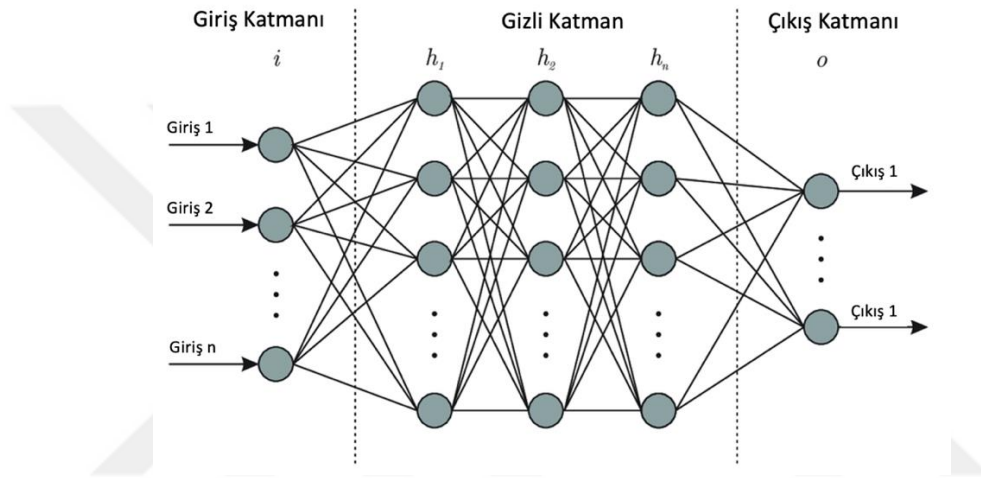
Şekil 5.4. DVM çalışma prensibi [17]

Makine öğrenmesinde Kernel teknikleri, doğrusal sınıflandırıcıların verileri daha yüksek boyutlu bir uzaya eşleyerek doğrusal olmayan verilerde kullanımını sağlar.

5.3. Derin Öğrenme Algoritmaları

Makine öğrenme algoritmalarının alt başlığı olan derin öğrenme insan beyninin yapısından esinlenilerek oluşturulan yapay sinir ağları üzerine kurgulanmış bir makine öğrenmesi yaklaşımı olarak tanımlanabilir. Yapay sinir ağlarının çoklu girdi ve çıktıya sahip katmanlardan oluşmakta olup, her bir katman bir öncekinin çıktısını bir girdi olarak kabul eder.

Yapay Sinir Ağları, insan beyninden ve insan beyninin öğrenme sürecinden esinlenerek oluşturulmuş yapılardır. Yapay Sinir ağları, Şekil 5.5’de de görülebileceği üzere 3 katmana ayrılır. Giriş katmanı, veri girişi ile sorumlu katmandır. Bu katmanda girdiler genellikle aktivasyon fonksiyonları tarafından üretilen sınır değerler içerisinde normalleştirilir. Gizli katman, sistemle ilişkili örüntüleri çıkarmakla sorumlu nöronlardan oluşur. Son olarak çıkış katmanı da nöronlardan oluşur. Bu nedenle önceki katmanlarda nöronlar tarafından gerçekleştirilen işlemde kaynaklanan nihai ağ çıktılarından üretilmesinden ve sunulmasından sorumludur. [18]



Şekil 5.5. YSA mimarisi [19]

Derin öğrenme algoritmaları makine öğrenme algoritmalarına kıyasla çok daha fazla verinin bulunduğu daha karmaşık problemlerin çözümleri için uygundur.

Bu tez çalışması kapsamında yapılan deneyler süresince veri setine derin öğrenme algoritmalarından UKSB, Çift Yönlü UKSB, ESA ve GTB sınıflandırma algoritmaları uygulanmıştır. Sınıflandırma sonucunda elde edilen sonuçlar başarı kriterlerine göre değerlendirilmiştir. Karşılaştırma başarı kriteri doğruluk değeri olarak belirlenmiştir.

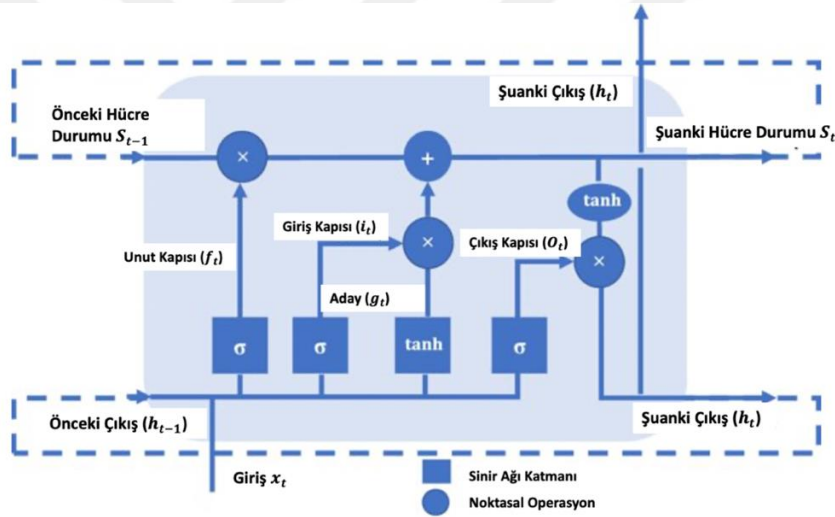
5.3.1. Uzun kısa süreli bellek (UKSB)

1997 yılında Jürgen Schmidhuber ve Sepp Hochreiter tarafından ilk kez sunulan Uzun Kısa Süreli Bellek bir Tekrarlayan Sinir Ağı modeli olarak bilinmektedir. Ses ve görüntü işlemenin yanı sıra doğal dil işlemede de oldukça sık kullanılan bir tür tekrarlayan sinir ağı modelidir. Karmaşık verileri işleme yeteneği nedeniyle, konuşma tanıma, makine çevirisi gibi birçok farklı uygulama UKSB'den yararlanmış olup son zamanlarda fiyat tahmini, hava

durumu tahmini gibi birçok problemle ilgili yapılan çalışmalarda da UKSB tahmin yeteneğini büyük ölçüde kanıtlamıştır. [20]

Geleneksel TSA yapısından farklı olarak UKSB mimarisinde uzun süreli bellek ile TSA sinir ağlarının eğitilmesinde yaşanan problemler bütünüyle giderilmiştir. [21] Geleneksel TSA yapısı tek bir tanh katmanı içerirken, UKSB mimarisinde iletişim halinde olan 4 farklı katman vardır. UKSB mimarisinde içerdği bu 4 farklı katman sayesinde anlamlı bilgileri hücreler boyunca taşıyan bir iletişim hattı oluşturulur. İletişim hattı boyunca bilgilerin gerekli veya gereksiz olduğunu belirleyen kapılardır. (Şekil 5.6) Her bir kapı Sigmoid aktivasyon fonksiyonu kullanarak gerekli bilgileri 1, gereksiz bilgileri 0 olarak belirler ve iletimini sağlar.

UKSB algoritmasının dezavantajı olarak tasarım ve programlamadaki karmaşıklığı ve doğru sonuçlar üretmek için büyük miktarda verilerin gerekli olması örnek gösterilebilir.



Şekil 5.6. UKSB mimarisi [22]

$$S_t = f_t * S_{t-1} + i_t * g_t$$

$$H_t = O_t * \tan h(S_t)$$

$$X_t = \text{Giriş Vektörü}$$

$$f_t = \text{Unut Kapısı}$$

$$g_t = \text{Aday Hafıza}$$

$$H_{t-1} = \text{Önceki Hücre Çıktısı}$$

$$S_{t-1} = \text{Önceki Hücre Durumu/Hafızası}$$

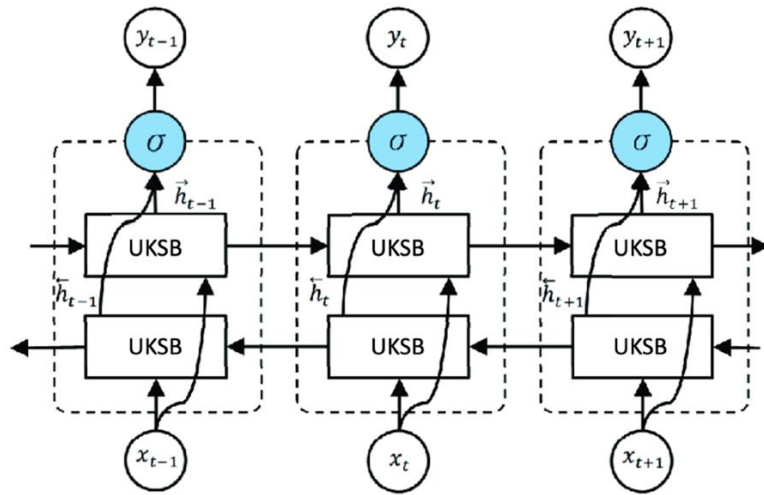
$$H_t = \text{Şuanki Hücre Çıktısı}$$

$$S_t = \text{Şuanki Hücre Durumu/Hafızası}$$

5.3.2. Çift yönlü uzun kısa süreli bellek (Çift yönlü UKSB)

Özellikle doğal dil işlemede kullanılan Çift Yönlü UKSB, TSA sinir ağı yapılarından biridir. Çift Yönlü UKSB mimarisinin yalnızca önceki durum bilgisini yakalayabilen ancak gelecekteki durum bilgisini yakalayamayan UKSB hücresinin eksikliğini gidermek için ortaya atıldığı söylenebilir. [23] Standart UKSB yapısından farklı olarak, Şekil 5.7’de de görülebileceği gibi girdi her iki yönde de akar ve ağ yapısı her iki taraftan gelen bilgileri kullanma yeteneğine sahiptir. [24]

Bu sayede ağı her adımda girdi dizisinin hem ileri hem de geri yönünden öğrenmeyi güçlendirmesini sağlar. Bu esneklik sayesinde ağ, bir sonraki adım dizisini tahmin etmek için hem önceki hem de gelecekteki bağlam bilgilerine bakabilir.

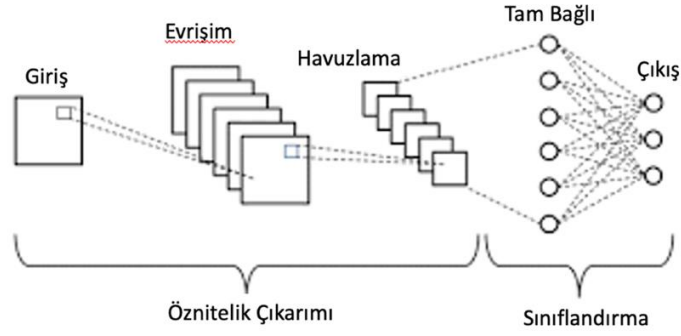


Şekil 5.7. Çift Yönlü UKSB mimarisi [25]

5.3.3. Evrimsel sinir ağları (ESA)

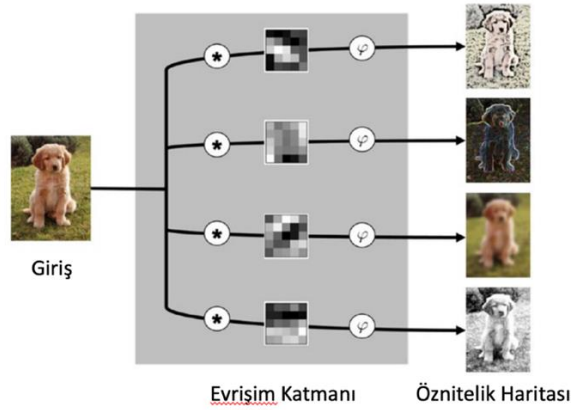
ESA, hayvan görsel korteksinin organizasyonundan ilham alan, özniteliklerin uzamsal hiyerarşilerini düşükten yükseğe doğru otomatik ve uyarlanabilir bir şekilde öğrenmek için tasarlanmış olup görüntü ve metin işlemede oldukça sık kullanılan bir yöntemdir. [26]

ESA yapısı Şekil 5.8’de gösterildiği gibi giriş, evrişim, havuzlama, tam bağlı ve çıkış katmanlarından oluşur. Evrişim ve havuzlama katmanlarında öznitelik çıkarımı işlemi gerçekleştirilirken, tam bağlı katmanda sınıflandırma işlemi gerçekleştirilir.



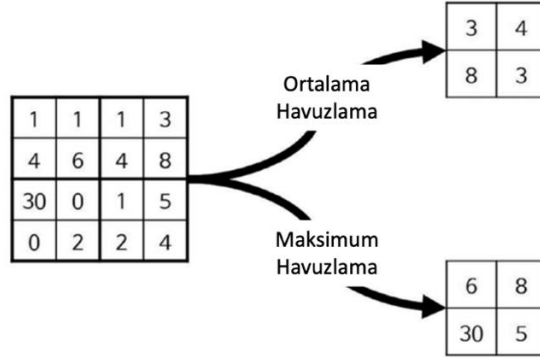
Şekil 5.8. ESA mimarisi [27]

Evrişim katmana gelen veriler öznitelik matrislerine dönüştürülür. Bu katmanda diğer sinir ağı katmanlarından farklı olarak ağırlıklar yerine görüntüleri dönüştüren filtreler içerir. evrişim katmanı, Şekil 5.9’da gösterildiği gibi evrişim filtreleriyle aynı sayıda öznitelik haritası üretir. [27]



Şekil 5.9. Evrişim katman süreci [28]

Havuzlama katmanı ise öznitelik haritalarının boyutunu küçültmek veya azaltmaktan sorumludur. Bu şekilde ağ yapısında işlenmesi gereken parametlerin sayısını azalttığı için işlemler daha hızlı gerçekleştirilmiş olur.

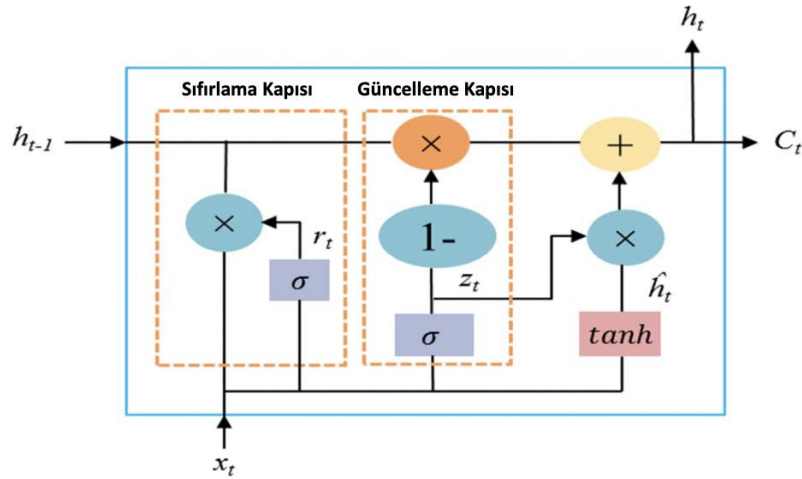


Şekil 5.10. İki farklı yöntem kullanılarak yapılan havuzlama işlemi [28]

5.3.4. Geçitli tekrarlayan birim (GTB)

TSA yapısının bir türü olan GTB, 2014 yılında Cho ve arkadaşları tarafından önerilen bir modeldir. TSA yapılarında her bir katmandaki ağırlıklar geriye yayılım ile iletilirken gradient değerleri geriye doğru ilerledikçe katlanarak küçülerek kaybolmaktadır. Bu problem kaybolan gradyan problemi olarak adlandırılır. GTB modelinde ise bu kaybolan gradyan sorununun üstesinden gelinmiştir. [29]

GTB yapısının mimarisi UKSB yapısına oldukça benzerdir fakat GTB mimarisinde UKSB mimarisinden farklı olarak bir çıkış kapısı yoktur. GTB mimarisi Şekil 5.11’de verildiği gibi “Sıfırlama” ve “Güncelleme” olmak üzere 2 kapıdan oluşmaktadır.



Şekil 5.11. GTB mimarisi [29]

GTB mimarisinde yer alan Güncelleme kapısı UKSB mimarisindeki Giriş ve Unutma kapıları ile oldukça benzer bir çalışma prensibine sahiptir. Bu kapıda hangi bilgilerin tutulup, hangi bilgilerin atılacağına karar verilir.

Sıfırlama kapısında ise geçmiş bilgilerin ne kadarının tutulup ne kadarının atılacağına karar verilir.

5.4. Başarı Kriterleri

Bu tez çalışmasında sınıflandırma sonuçları; kesinlik, duyarlılık, F1-skor, kappa ve doğruluk başarı kriterlerine göre elde edilen başarılar göz önüne alınarak değerlendirilmiştir.

Kesinlik değeri modelin tahmin ettiği doğru pozitif örnek sayısının ve toplam pozitif örnek sayısına oranı olarak tanımlanabilir. Kesinlik değeri çok sınıflı veri setlerinde (5.1) ve (5.2)'de gösterildiği şekilde hesaplanabilir. [30]

$$Kesinlik_i = \frac{DP_i}{DP_i + YP_i} \quad (5.1)$$

$$Kesinlik = \frac{\sum_{i=0}^{Sınıf Sayısı} Kesinlik_i}{Sınıf Sayısı} \quad (5.2)$$

Duyarlılık değeri ise gerçek pozitif örnek sayısının, gerçek pozitif ve yanlış negatif örneklerin toplamına oranıyla hesaplanır. Duyarlılık metriğinin hesaplaması da (5.3) ve (5.4)'de gösterildiği şekilde hesaplanabilir. Modelin tahminin kaç tane pozitif veriyi kaçırdığını ortaya koyar. [31]

$$Duyarlılık_i = \frac{DP_i}{DP_i + YN_i} \quad (5.3)$$

$$Duyarlılık = \frac{\sum_{i=0}^{Sınıf Sayısı} Duyarlılık_i}{Sınıf Sayısı} \quad (5.4)$$

Makine öğrenmesinde en sık kullanılan başarı kriterlerinden biri olan F1-skor değeri ise (5.5)'de gösterildiği üzere kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanır.

$$F1 - Skor = \frac{2 \times Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (5.5)$$

Bir diğer başarı kriteri olan kappa değeri en basit tanımıyla gerçek ve tahmin edilen değerler arasındaki uyumayı -1 ve 1 aralığında gösterir. Kappa değerinin 1'e yakınlığı

sınıflandırmanın başarısı ile doğru orantılıdır. [31] Kappa değeri (5.6) ve (5.7)'de gösterildiği şekilde hesaplanabilir.

$$Temel \ Çizgi = \sum \frac{(YP+DN) \times (DN+YN)}{(DN+YN+DP+YP)^2} \quad (5.6)$$

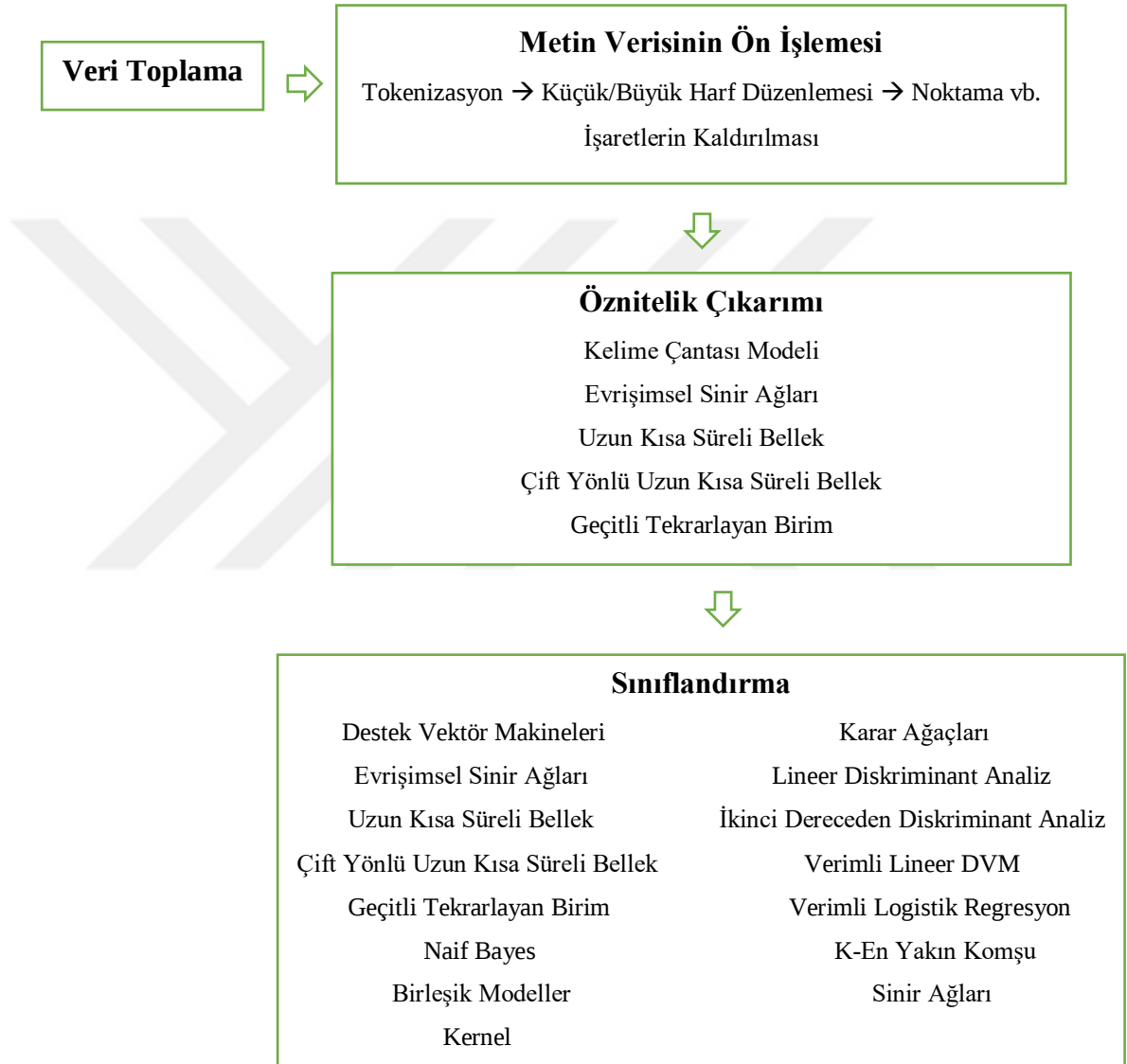
$$Kappa = \frac{Doğruluk - Temel \ Çizgi}{1 - Temel \ Çizgi} \quad (5.7)$$

Makine öğrenmesinde en çok kullanılan değer olan doğruluk değeri doğru tahmin edilen değerlerin tüm tahmin edilen değerlere oranı olarak tanımlanabilir ve (5.8)'de gösterildiği şekilde hesaplanabilir.

$$Doğruluk = \frac{\sum DP + \sum DN}{Tüm \ Gözlemlerin \ Sayısı} \quad (5.8)$$

6. UYGULANAN YÖNTEMLER

Metin sınıflandırma işlemi bu tez çalışması boyunca 3 farklı aşamada uygulanmıştır. Metin sınıflandırma işlemi Şekil 6.1’de görüldüğü üzere metin verileri elde edildikten sonra metin verisinin ön işlemesi ile başlar, öznitelik çıkarımı ve sınıflandırma işlemi ile devam eder.



Şekil 6.1. Metin sınıflandırma süreci blok diyagramı

6.1. Metin Verisinin Ön İşlemesi

Bu bölümde kullanılacak olan veri seti belirlendikten sonra sınıflandırma işlemine geçilmeden önce yapılmış olan normalleştirme, ön işleme sürecinden bahsedilmiştir.

İnternet ortamından temin edilen tüm veri setlerinde olduğu gibi e-ticaret sayfalarında ürün yorumları toplanarak elde edilmiş olan bu veri setinde de, metin verilerinin normalleştirilmesi hem daha anlamlı bir sınıflandırma yapılabilmesi hem de sınıflandırma başarısının arttırılabilmesi açısından oldukça önem taşımaktadır. Metin verisinin daha anlamlı bir şekilde sınıflandırılabilmesi amacıyla oldukça sık kullanılan bazı ön işleme yöntemleri vardır. Bunlar aşağıdaki gibi sıralanabilir.

- Tokenizasyon
- Tüm harflerin küçük harflere dönüştürülmesi
- Noktalama, vurgu ve diğer aksan işaretlerinin kaldırılması

Ön işlem süreci sınıflandırmada girdi olarak kullanılacak olan metin verilerin tokenizasyonu ile başlar. Tokenizasyonu kısaca bir ifadeyi, cümleyi, paragrafı veya tüm metin belgesini tek tek kelimeler veya terimler gibi daha küçük birimlere bölmektir. Bu küçük birimlerin her birine token adı verilir. Tokenizasyon sayesinde metin verileri daha küçük ve işlenebilir parçalara ayrılmış olur.

Tablo 6.1. Tokenizasyon örneklendirmesi

3 token	evet anlatıldığı gibi
15	daha öncede almıştım bu cihazdan ense ve sakal tüketmek için on numara sıfıra yakın alıyor
7	denedim ürün çok güzel herkese tavsiye ederi
4	ürün hala elime geçmedi
6	tavsiye edebileceğim çok güzel bir makina

*Veri setinin yola çıkılarak örneklendirme amacıyla yapılmış olan tokenizasyonu

Tokenlaştırılmış metin verilerinin sınıflandırma işleminde bir girdi olarak kullanılabilmesi amacıyla wordEncoding fonksiyonu ile metin verileri kategorik vektör dizilerine dönüştürülür. Ardından belgeleri tam sayı dizilerine dönüştürmek için doc2sequence fonksiyonu kullanılır.

Tüm bu ön işlemlerin ardından öznitelik çıkarımı veya sınıflandırma algoritmaları uygulanır.

6.2. Öznitelik Çıkarımı

Çalışma süresince tüm sınıflandırma algoritmalarından önce veri setinden öznitelik çıkarılması amacıyla öncelikle tüm sınıflandırma modellerine kelime çantası modeli uygulanmıştır. Sonrasında ise öznitelik çıkarımı için ESA, UKSB, Çift Yönlü UKSB ve GTB algoritmaları kullanılmıştır.

Kelime çantası; öznitelik seçimi, metin ve görüntülerin sınıflandırması gibi amaçlara yönelik kullanılan ve özellikle doğal dil işlemede oldukça sık rastladığımız bir modeldir. Doğal dil işlemede kullanılan kelime çantası modelinde, metin verileri kelime torbası olarak temsil edilir [32]. Bu modelin oluşturulmasında önemli olan kelimelerin sıklığı olduğundan sadece tekrarlayan kelimeler dikkate alınır. Tekrarlama sıklığı bu çalışmada 2 olarak belirlenmiştir.

```
bag =  
bagOfWords with properties:  
  Counts: [12136x17183 double]  
  Vocabulary: ["evet" "anlatıldığı" "gibi" "daha" "öncede" "almıştım" "bu" "cihazdan" "ense" "ve" "sakal" "tüketmek" "için" ... ]  
  NumWords: 17183  
  NumDocuments: 12136
```

Şekil 6.2. Kelime çantası modeli özellikleri

Derin ağlardan çıkarılan öznitelikler ise tüm derin ağ modellerinin tam bağlı katmanından elde edilmiştir. UKSB, Çift yönlü UKSB, ESA ve GTB modelleri oluşturulduktan sonra model eğitimi sağlanmıştır. Sonrasında oluşturulan modelin tam bağlı katmanından öznitelikler çekilmiştir.

6.3. Sınıflandırma

Sınıflandırma aşaması ön işleme metotları ve öznitelik çıkarımları uygulandıktan sonra uygulanmış olup sonuçlar paylaşılmıştır.

6.3.1. DVM sınıflandırma algoritmasının uygulanması

Ön işleme ve kelime çantası modeli ile öznitelik çıkarımı süreçlerinden sonra veri seti sınıflandırılmaya hazır hale getirilmiş ve DVM algoritması kullanılarak sınıflandırma sürecine başlanmıştır. Sınıflandırma süresince çapraz doğrulama yüzdesi değiştirilerek en yüksek doğruluğu veren değer aranmıştır. Çapraz doğrulama yüzdesi %10 olarak seçildiğinde elde edilen doğruluk değeri %87,41, çapraz doğrulama yüzdesi %20 seçildiğinde doğruluk değeri %87,57, çapraz doğrulama yüzdesi %30 seçildiğinde ise elde edilen doğruluk değeri %87,28 olmuştur. Sınıflandırmanın sonucunda DVM için elde edilen

en yüksek doğruluk değeri %87,57 olarak elde edilmiştir. En yüksek doğruluğa sahip karışıklık matrisi Şekil 6.3’de verilmiştir.

Şekil 6.3, DVM karışıklık matrisi (çapraz doğrulama yüzdesi=20) olarak adlandırılan bir tabloyu göstermektedir. Tablo, 'Doğru Sınıf' (0, 1, 2) ve 'Tahmin Edilen Sınıf' (0, 1, 2) başlıklarıyla yapılandırılmıştır. Tablo içeriği aşağıdaki gibidir:

Doğru Sınıf \ Tahmin Edilen Sınıf	0	1	2
0	1226	36	49
1	93	1268	67
2	77	55	163

Şekil 6.3. DVM karışıklık matrisi (çapraz doğrulama yüzdesi=20)

6.3.2. UKSB sınıflandırma algoritmasının uygulanması

Veri setinin ön işleme tamamlandıktan sonra UKSB yapısının ağ mimarisi oluşturulmaya başlanır. UKSB ağ mimarisi oluşturulurken ilk önce, dizi verilerini ağı sokmak için bir giriş katmanı eklenmiştir ve giriş boyutunu 1 olarak ayarlanmıştır. Ardından bir kelime gömme katmanı ve kelime kodlamasıyla aynı sayıda kelimeler eklenmiştir. Ardından, UKSB katmanı eklenip gizli birimlerin sayısı belirlenmiştir. UKSB katmanını bir diziden etikete sınıflandırma probleminde kullanmak için, çıktı modu 'son' olarak ayarlanmıştır. Son olarak sınıf sayısı kadar büyüklüğe sahip tamamen bağlantılı bir katman, bir softmax katmanı ve bir sınıflandırma katmanı eklenmiştir. Verilerin eğitimi sürecinde en uygunlayıcı olarak Adam yöntemi kullanılmış ve veriler her epokta karıştırılmıştır. Ardından oluşturulan ağ yapısı işleme sokulmuştur.

Veri seti eğitim ve test olarak ayrılırken en ideal çapraz doğrulama yüzdesinin belirlenmesi amacıyla denemeler yapılmıştır. Bunun yanı sıra deney süresince çapraz doğrulama yüzdesinde, kelime gömme katmanı boyutunda ve gizli birimlerin sayısında değişiklikler yapılarak değişimler gözlemlenmiştir. Gözlem sonuçlarına ait elde edilen değerler Tablo 6.2’de verilmiştir.

Sınıflandırma sonucunda UKSB yöntemi kullanılarak elde edilen en yüksek doğruluk değeri %86,35 olmuştur. Sınıflandırma performansını temsilen karışıklık matrisi Şekil 6.4’de, UKSB eğitim süreci Şekil 6.5’de verilmiştir.

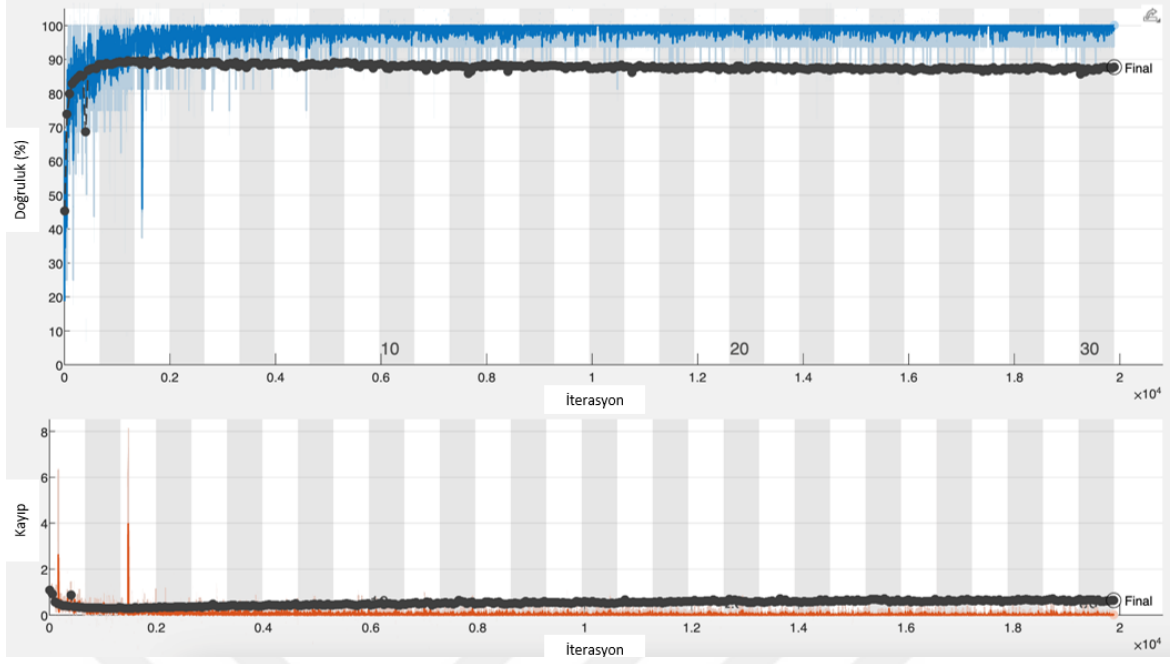
Tablo 6.2. 30 epok için UKSB deney sonuçları

Çapraz Doğrulama Yüzdesi	Kelime Gömme Katmanı Boyutu	Gizli Birimlerin Sayısı	Doğruluk (%)
%10	50	100	86,16
%20	50	100	86,35
%30	50	100	87,06
%30	100	100	86,40
%30	200	100	86,66
%30	50	50	87,45
%30	50	200	87,70

0	1867	106	120
1	72	1901	67
2	80	115	223
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.4. UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)



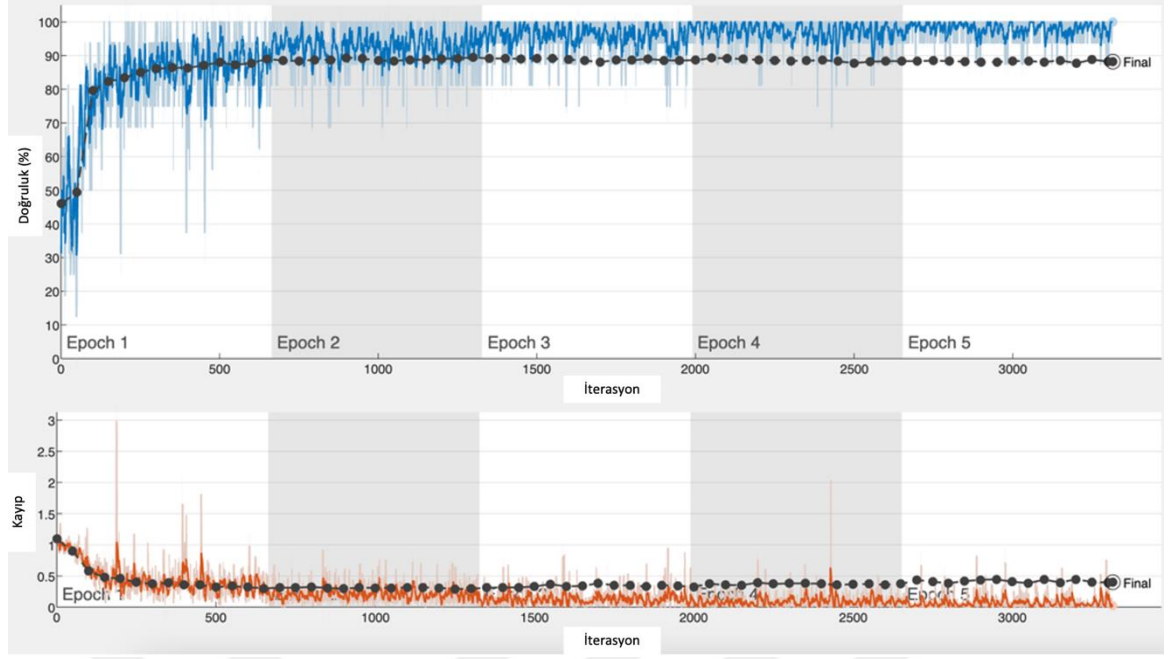
Şekil 6.5. UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)

Şekil 6.5’de verilen eğitim süreci gözlemlendiğinde yaklaşık olarak 5.epoktan itibaren başarı değerinin düştüğü ve kaybın arttığı gözlemlenmiştir. Bu sebeple 30.epok olarak çalıştırılan eğitim 5. Epok olarak değiştirilmiştir. Elde edilen yeni karışıklık matrisi Şekil 6.6’da, gözlemlenen yeni eğitim süreci ise Şekil 6.7’de paylaşılmıştır. Epok değeri 5 olarak seçildiğinde doğruluk değeri %88,24 olarak elde edilmiştir.

	0	1	2
0	1968	53	73
1	107	1841	92
2	112	98	207
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.6. UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

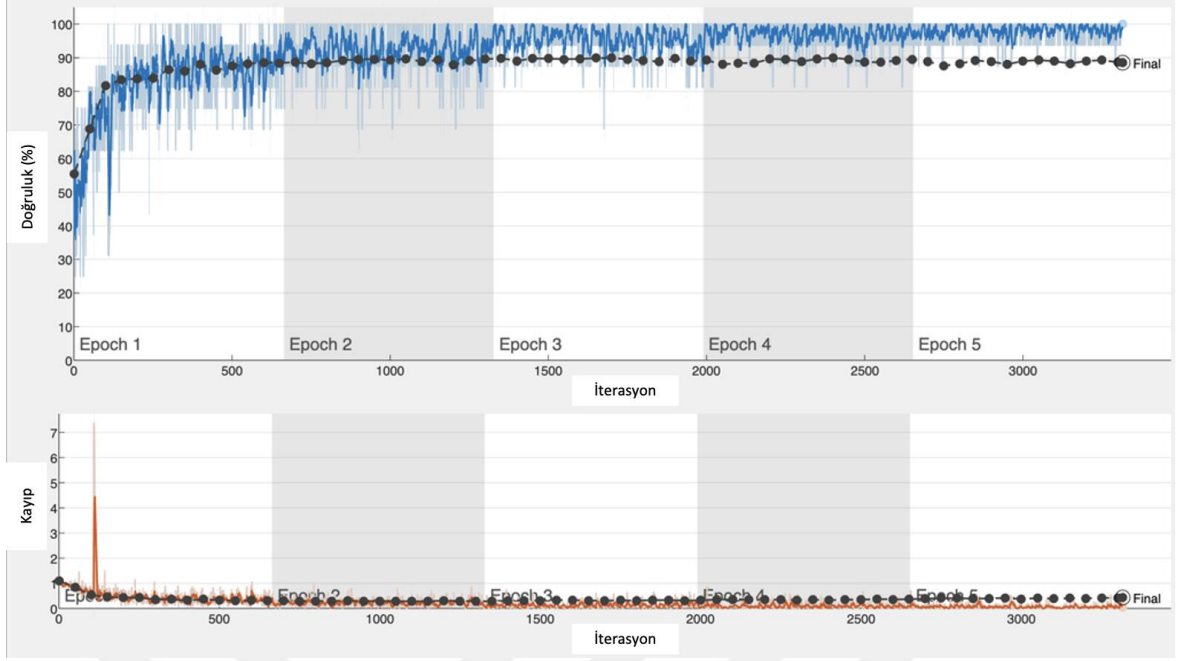


Şekil 6.7. UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

En yüksek başarıyı veren parametreler ve epok değeri için elde edilen verilerin ardından öznitelik çıkarımı amacıyla kelime çantası modeli uygulanmış olup karışıklık matrisi Şekil 6.8'de, eğitim süreci ise Şekil 6.9'de paylaşılmıştır. Kelime çantası modeli uygulandığından yeni doğruluk değeri %88,51 olarak elde edilmiş olup, doğruluğun arttığı gözlemlenmiştir.

0	1902	103	89
1	51	1898	90
2	104	86	228
	0	1	2
	Tahmin Edilen Sınıf		

Şekil 6.8. Kelime çantası modeli uygulandığında UKSB karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)



Şekil 6.9. Kelime çantası modeli uygulandığında UKSB eğitim süreci (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)

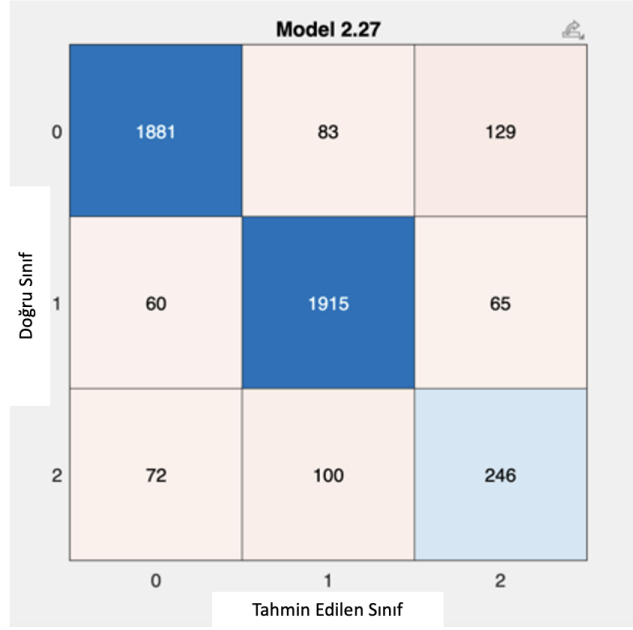
Son aşama olarak öznitelik çıkarımı için UKSB algoritmasının kendisi kullanılmıştır. Bu aşamada veri seti UKSB ile eğitilmiş olup, eğitim sonunda tam bağlı katmandan öznitelikler çıkarılmıştır. Çıkarılan öznitelikler klasik makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırılmıştır. Elde edilen sonuçlar Tablo 6.3’de paylaşılmıştır. UKSB algoritmasının kullanıldığı tüm deneylere ait sonuçlar analiz edildiğinde en yüksek başarımın Verimli Logistik Regresyon ve Sinir Ağaçları (Dar) uygulandığında %88,82 doğruluk değeri ile elde edildiği gözlemlenmiştir. En yüksek doğruluğu veren her iki algoritma kıyaslandığında ise Yapay Sinir Ağları (Dar) algoritmasının kappa başarı değerinin daha yüksek olması sebebiyle en yüksek başarıyı veren algoritma Yapay Sinir Ağları (Dar) olarak değerlendirilmiştir.

Tablo 6.3. UKSB öznitelik çıkarımı kullanıldığında deney sonuçları

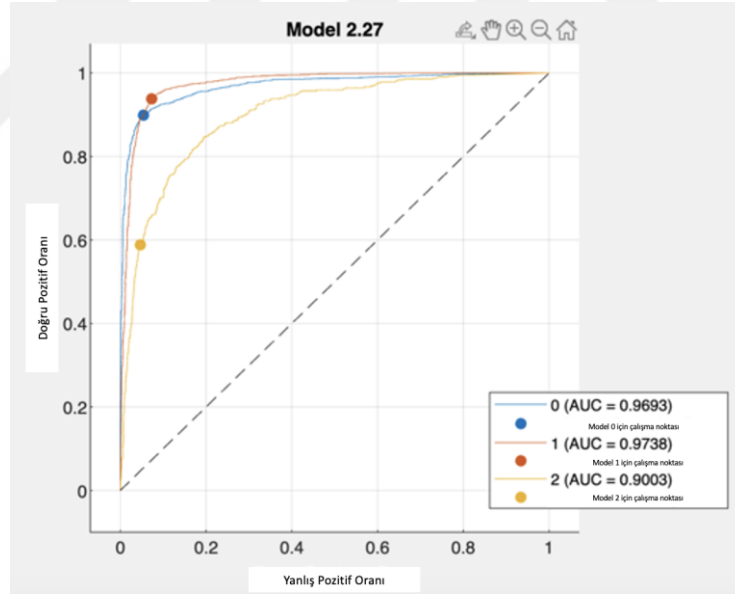
Kullanılan Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Karar Ağaçları (Orta Ağaç)	0,80	0,80	0,80	0,804	88,62

Lineer Diskriminant Analiz	0,81	0,78	0,79	0,786	87,34
İkinci Dereceden Lineer Diskriminant Analiz	0,81	0,79	0,80	0,800	88,31
Verimli Logistik Regresyon	0,81	0,80	0,80	0,807	88,82
Verimli Lineer DVM	0,81	0,80	0,80	0,806	88,70
Naif Bayes (Gaussian)	0,82	0,78	0,80	0,789	87,52
Destek Vektör Makineleri (Lineer DVM)	0,81	0,80	0,80	0,805	88,66
K-En Yakın Komşu (Kaba)	0,81	0,80	0,80	0,807	88,75
Birleşik Modeller (Güçlendirilmiş Ağaçlar)	0,81	0,80	0,80	0,805	88,64
Yapay Sinir Ağları (Dar)	0,81	0,80	0,80	0,808	88,82
Kernel (DVM Kernel)	0,81	0,80	0,80	0,806	88,75

En yüksek doğruluk değerinin elde edildiği sonuçlara ait karışıklık matrisi ve ROC eğrisi Şekil 6.10'da ve Şekil 6.11'de verilmiştir.



Şekil 6.10. UKSB öznitelik çıkarımında Yapay Sinir Ağları (Dar) karışıklık matrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)



Şekil 6.11. UKSB öznitelik çıkarımında Yapay Sinir Ağları (Dar) ROC eğrisi (kelime gömme katman boyutu=50, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)

UKSB algoritmasının hem sınıflandırma hem de öznitelik çıkarımında kullanıldığı tüm deneylere ait sonuçlar incelendiğinde en yüksek başarımın özniteliklerin UKSB' nin tam bağlı katmanından çekildiği sonrasında ise Yapay Sinir Ağları (Dar) algoritması ile sınıflandırıldığında %88,82 doğruluk değeri ile elde edildiği gözlemlenmiştir.

Elde edilen ROC eğrisi incelendiğinde en yüksek başarının 1 etiketli sınıfta yani pozitif yorumların sınıflandırılmasında elde edildiği görülmüştür. En çok hatanın ise veri setindeki en az veri sayısına sahip 2 etiketli yani nötr yorumlarda yapıldığı gözlemlenmiştir. Bu durumun önüne geçmek amacıyla veri setindeki nötr yorumlarının sayısının artırılması hedeflenmektedir.

Yapay sinir ağları çok sınıflı ve yüksek veri sayısına sahip yorumlarda başarılı olması sebebiyle sınıflandırılma aşamasında klasik makine öğrenmesi yöntemlerinin önüne geçerek diğer algoritmalara kıyasla daha yüksek bir başarı değerleri vermiştir.

6.3.3. Çift yönlü UKSB sınıflandırma algoritmasının uygulanması

Veri setinin ön işleme tamamlandıktan sonra Çift Yönlü UKSB yapısının ağ mimarisi oluşturulmaya başlanır. Çift Yönlü UKSB ağ mimarisi oluşturulurken ilk önce, dizi verilerini ağa sokmak için bir giriş katmanı eklenmiştir ve giriş boyutunu 1 olarak ayarlanmıştır. Ardından bir kelime gömme katmanı ve kelime kodlamasıyla aynı sayıda kelimeler eklenmiştir. Ardından, Çift Yönlü UKSB katmanı eklenip gizli birimlerin sayısı belirlenmiştir. Çift Yönlü UKSB katmanını bir diziden etikete sınıflandırma probleminde kullanmak için, çıktı modu 'son' olarak ayarlanmıştır. Son olarak sınıf sayısı kadar büyüklüğe sahip tamamen bağlantılı bir katman, bir softmax katmanı ve bir sınıflandırma katmanı eklenmiştir. Verilerin eğitimi sürecinde en uygunlayıcı olarak Adam yöntemi kullanılmış ve veriler her epokta karıştırılmıştır. Ardından oluşturulan ağ yapısı işleme sokulmuştur.

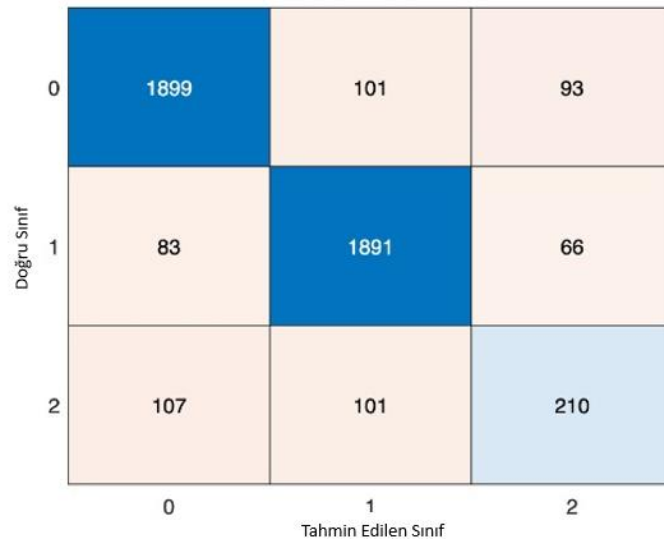
Veri seti eğitim ve test olarak ayrılırken en ideal çapraz doğrulama yüzdesinin belirlenmesi amacıyla denemeler yapılmıştır. Bunun yanı sıra deney süresince çapraz doğrulama yüzdesinde, kelime gömme katmanı boyutunda ve gizli birimlerin sayısında değişiklikler yapılarak değişimler gözlemlenmiştir. Gözlem sonuçlarına ait elde edilen değerler Tablo 6.4’de verilmiştir.

Sınıflandırma sonucunda Çift Yönlü UKSB yöntemi kullanılarak elde edilen en yüksek doğruluk değeri %87,89 olmuştur. Sınıflandırma performansını temsilen karışıklık matrisi Şekil 6.12’de, Çift Yönlü UKSB eğitim süreci Şekil 6.13’de verilmiştir.

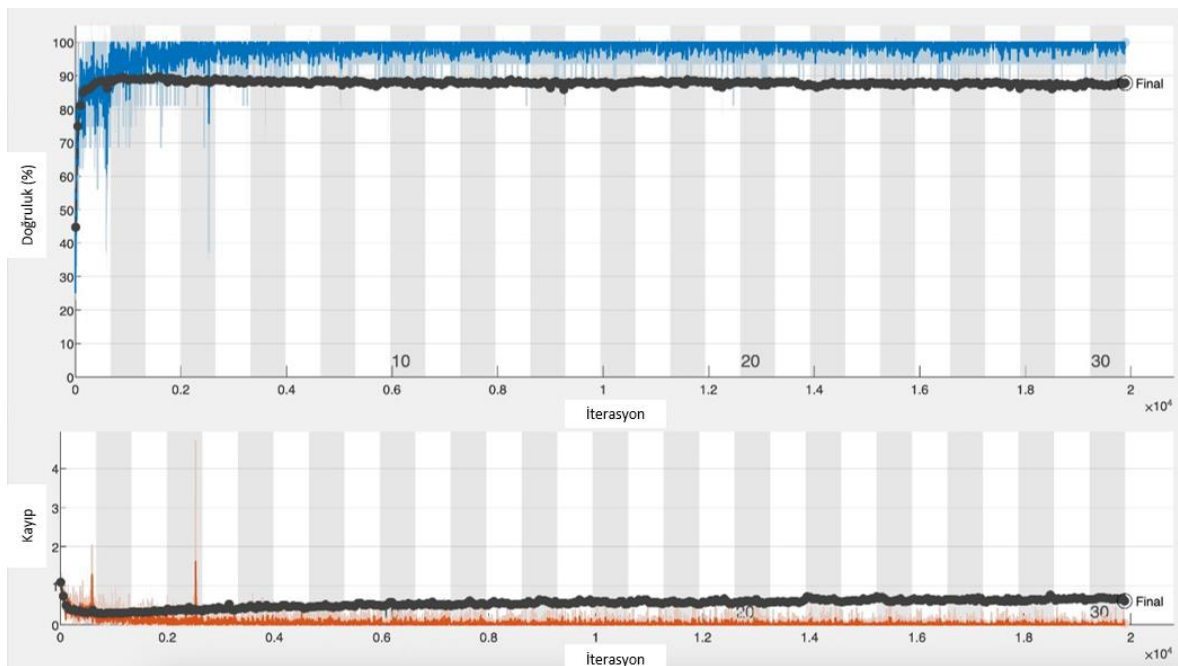
Tablo 6.4. 30 epok için Çift Yönlü UKSB deney sonuçları

Çapraz Doğrulama Yüzdesi	Kelime Gömme Katmanı Boyutu	Gizli Birimlerin Sayısı	Doğrululuk (%)

%10	50	100	85,89
%20	50	100	86,88
%30	50	100	87,28
%30	100	100	87,70
%30	200	100	87,89
%30	200	50	86,16
%30	200	200	86,11



Şekil 6.12. Çift Yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)

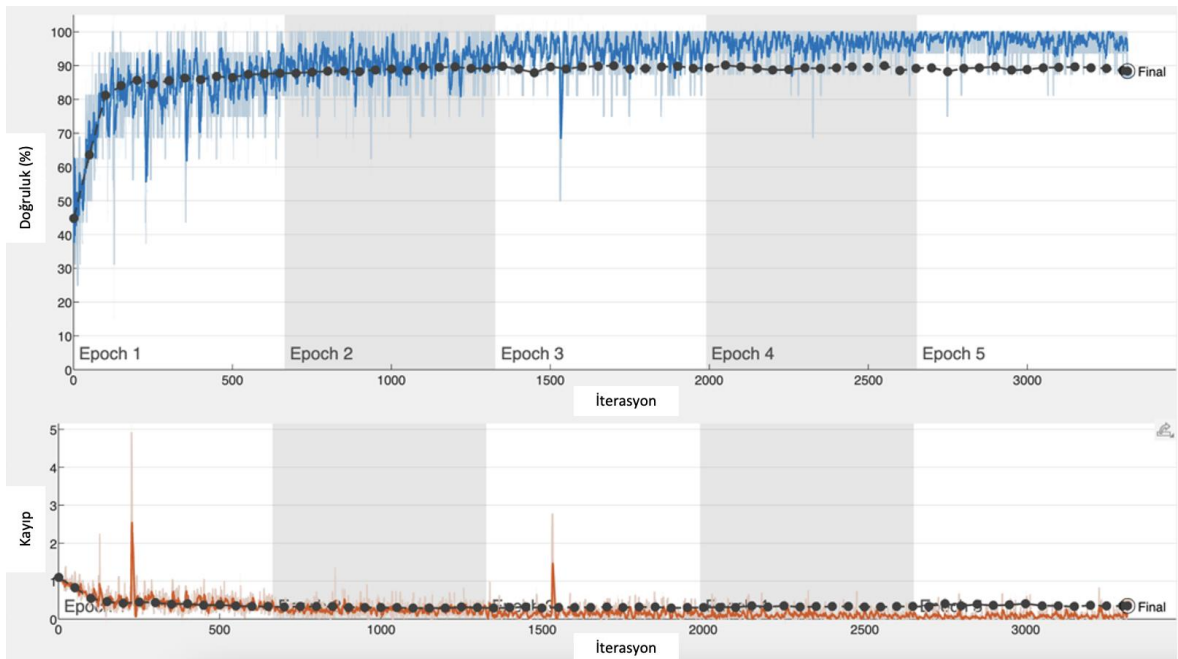


Şekil 6.13. Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)

Şekil 6.13’de verilen eğitim süreci gözlemlendiğinde yaklaşık olarak 5.epoktan itibaren başarı değerinin düştüğü ve kaybın arttığı gözlemlenmiştir. Bu sebeple 30.epok olarak çalıştırılan eğitim 5. Epok olarak değiştirilmiştir. Elde edilen yeni karışıklık matrisi Şekil 6.14’de, gözlemlenen yeni eğitim süreci ise Şekil 6.15’de paylaşılmıştır. Epok değeri 5 olarak seçildiğinde doğruluk değeri %88,42 olarak elde edilmiştir.

Doğru Sınıf	Tahmin Edilen Sınıf		
	0	1	2
0	1950	47	97
1	66	1814	159
2	86	72	260

Şekil 6.14. Çift yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)

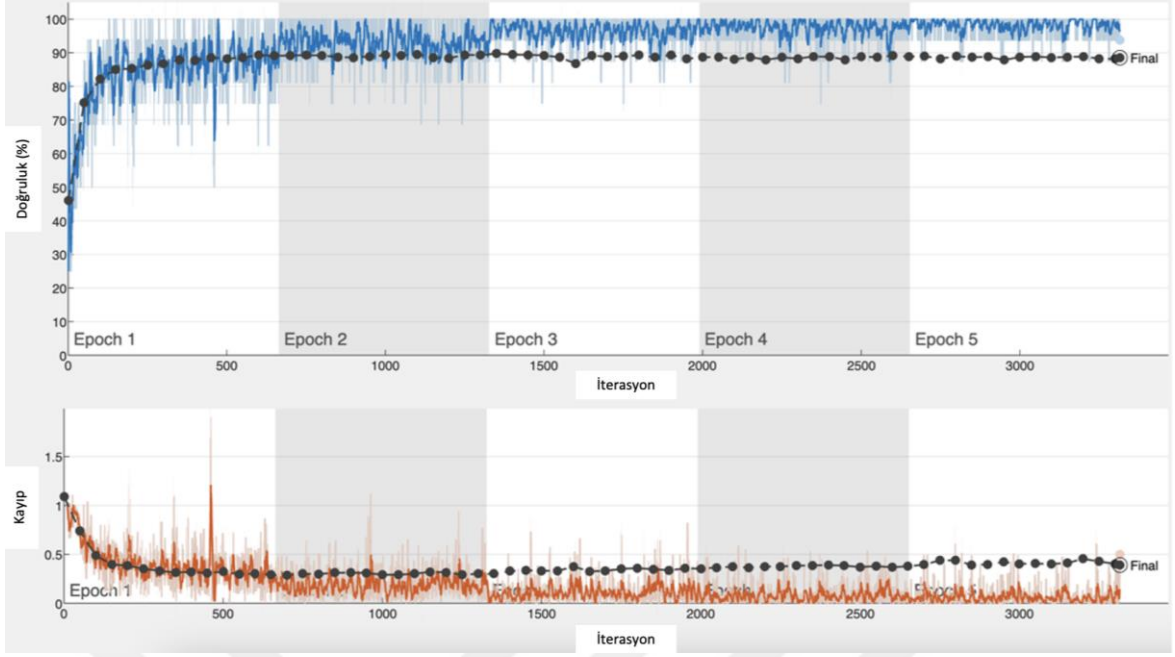


Şekil 6.15. Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)

En yüksek başarıyı veren parametreler ve epok değeri için elde edilen verilerin ardından öznitelik çıkarımı amacıyla kelime çantası modeli uygulanmış olup karışıklık matrisi Şekil 6.16’da, eğitim süreci ise Şekil 6.17’de paylaşılmıştır. Kelime çantası modeli uygulandığından yeni doğruluk değeri %88,57 olarak elde edilmiş olup, doğruluğun arttığı gözlemlenmiştir.

0	1916	77	101
1	66	1891	82
2	90	104	224
	0	1	2
	Tahmin Edilen Sınıf		

Şekil 6.16. Kelime çantası modeli uygulandığında Çift yönlü UKSB karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)



Şekil 6.17. Kelime çantası modeli uygulandığında Çift Yönlü UKSB eğitim süreci (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=5)

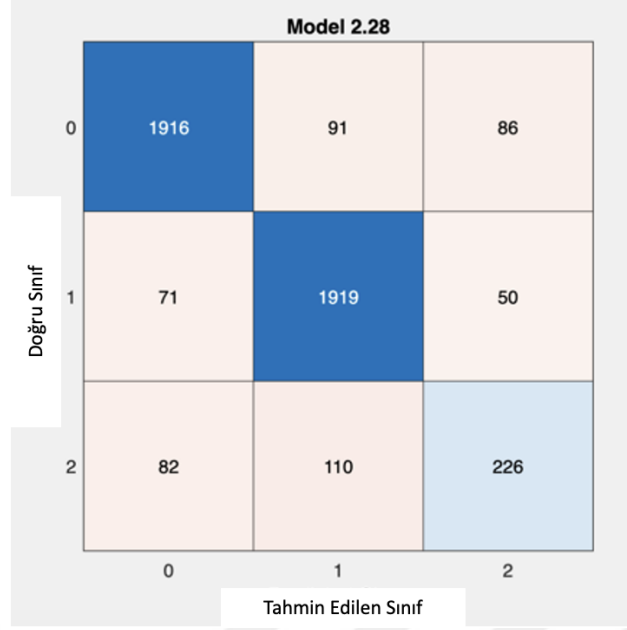
Son aşama olarak öznitelik çıkarımı için Çift Yönlü UKSB algoritmasının kendisi kullanılmıştır. Bu aşamada veri seti Çift Yönlü UKSB ile eğitilmiş olup, eğitim sonunda tam bağlı katmandan öznitelikler çıkarılmıştır. Çıkarılan öznitelikler klasik makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırılmıştır. Elde edilen sonuçlar Tablo 6.5’de paylaşılmıştır. Çift Yönlü UKSB algoritmasının kullanıldığı tüm deneylere ait sonuçlar analiz edildiğinde en yüksek başarının Yapay Sinir Ağları (Orta) uygulandığında %89,23 doğruluk değeri ile elde edildiği gözlemlenmiştir.

Tablo 6.5. Çift Yönlü UKSB öznitelik çıkarımı kullanıldığında deney sonuçları

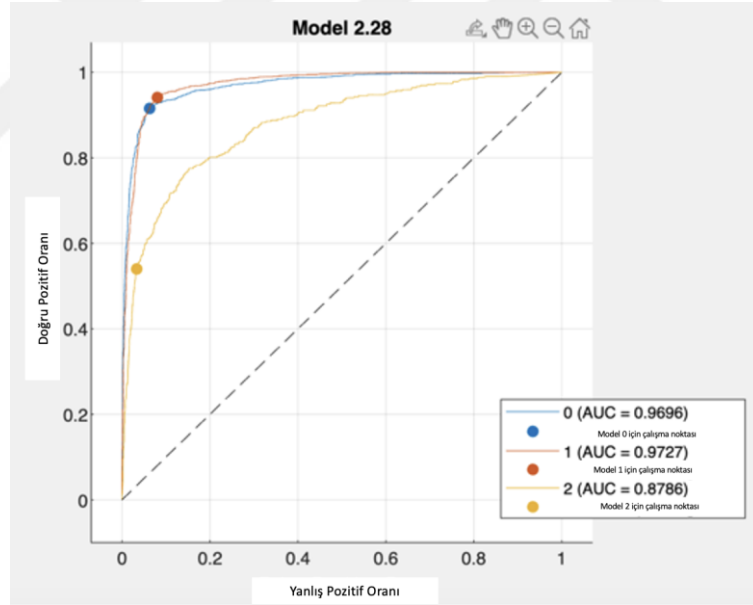
Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Karar Ağaçları (Kaba Ağaç)	0,79	0,82	0,80	0,807	88,95
Lineer Diskriminant Analiz	0,81	0,78	0,79	0,794	87,92
İkinci Dereceden	0,80	0,80	0,80	0,804	88,62

Lineer Diskriminant Analiz					
Verimli Logistik Regresyon	0,79	0,81	0,80	0,807	88,90
Verimli Lineer DVM	0,80	0,82	0,81	0,809	89,01
Naif Bayes (Gaussian)	0,81	0,79	0,80	0,798	88,18
Destek Vektör Makineleri (İnce Gaussian DVM)	0,80	0,81	0,80	0,813	89,21
K-En Yakın Komşu (Kübik)	0,80	0,82	0,81	0,812	89,17
Birleşik Modeller (Güçlendirilmiş Ağaçlar)	0,79	0,82	0,80	0,810	89,06
Yapay Sinir Ağları (Orta)	0,80	0,82	0,81	0,813	89,23
Kernel (Lojistik Regresyon Kernel)	0,79	0,82	0,80	0,810	89,08

En yüksek doğruluk değerinin elde edildiği sonuçlara ait karışıklık matrisi ve ROC eğrisi Şekil 6.18’de ve Şekil 6.19’da verilmiştir.



Şekil 6.18. Çift Yönlü UKSB öznitelik çıkarımında Yapay Sinir Ağları (Orta) karışıklık matrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)



Şekil 6.19. Çift Yönlü UKSB öznitelik çıkarımında Yapay Sinir Ağları (Orta) ROC eğrisi (kelime gömme katman boyutu=200, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)

Çift Yönlü UKSB algoritmasının hem sınıflandırma hem de öznitelik çıkarımında kullanıldığı tüm deneylere ait sonuçlar incelendiğinde en yüksek başarımın özniteliklerin Çift Yönlü UKSB' nin tam bağlı katmanından çekildiği sonrasında ise Yapay Sinir Ağları (Orta) algoritması ile sınıflandırıldığında %89,23 doğruluk değeri ile elde edildiği gözlemlenmiştir.

Elde edilen ROC eğrisi incelendiğinde en yüksek başarımın 1 etiketli sınıfta yani pozitif yorumların sınıflandırılmasında elde edildiği görülmüştür. En çok hatanın ise veri setindeki en az veri sayısına sahip 2 etiketli yani nötr yorumlarda yapıldığı gözlemlenmiştir. Bu durumun önüne geçmek amacıyla veri setindeki nötr yorumlarının sayısının artırılması hedeflenmektedir.

Yapay sinir ağları çok sınıflı ve yüksek veri sayısına sahip yorumlarda başarılı olması sebebiyle sınıflandırılma aşamasında klasik makine öğrenmesi yöntemlerinin önüne geçerek daha yüksek başarı değerleri vermiştir.

6.3.4. ESA sınıflandırma algoritmasının uygulanması

Veri setinin ön işleme tamamlandıktan sonra ESA yapısının ağ mimarisi oluşturulmaya başlanır. ESA ağ mimarisi oluşturulurken ilk önce, dizi verilerini ağı sokmak için bir giriş katmanı eklenmiştir ve giriş boyutunu 1 olarak ayarlanmıştır. Ardından bir kelime gömme katmanı ve kelime kodlamasıyla aynı sayıda sözcükler eklenmiştir. Ardından n-gram uzunlukları 2, 3, 4 ve 5 için, bir evrişim katmanı, bir grup normalleştirme katmanı, bir ReLU katmanı, bir bırakma katmanı ve bir havuzlama katmanı içeren katman blokları oluşturulmuştur. Ardından bir birleştirme katmanı ve tam bağlı katman eklenmiştir. Son olarak çıktıları sınıflandırmak için; çıktı boyutu K olan bir tamamen bağlı katman, bir softmax katmanı ve K'nın sınıf sayısı olduğu bir sınıflandırma katmanı eklenmiştir.

Veri seti eğitim ve test olarak ayrılırken en ideal çapraz doğrulama yüzdesinin belirlenmesi amacıyla denemeler yapılmıştır. Bunun yanı sıra deney süresince çapraz doğrulama yüzdesinde, kelime gömme katmanı boyutunda ve gizli birimlerin sayısında değişiklikler yapılarak değişimler gözlemlenmiştir. Gözlem sonuçlarına ait elde edilen değerler Tablo 6.6'da verilmiştir.

Sınıflandırma sonucunda ESA yöntemi kullanılarak %89,45 doğruluk değeri elde edilmiştir. Sınıflandırma performansını temsilen karışıklık matrisi Şekil 6.20'de, ESA eğitim süreci Şekil 6.21'de verilmiştir.

Tablo 6.6. 30 epok için ESA gözlem sonuçları

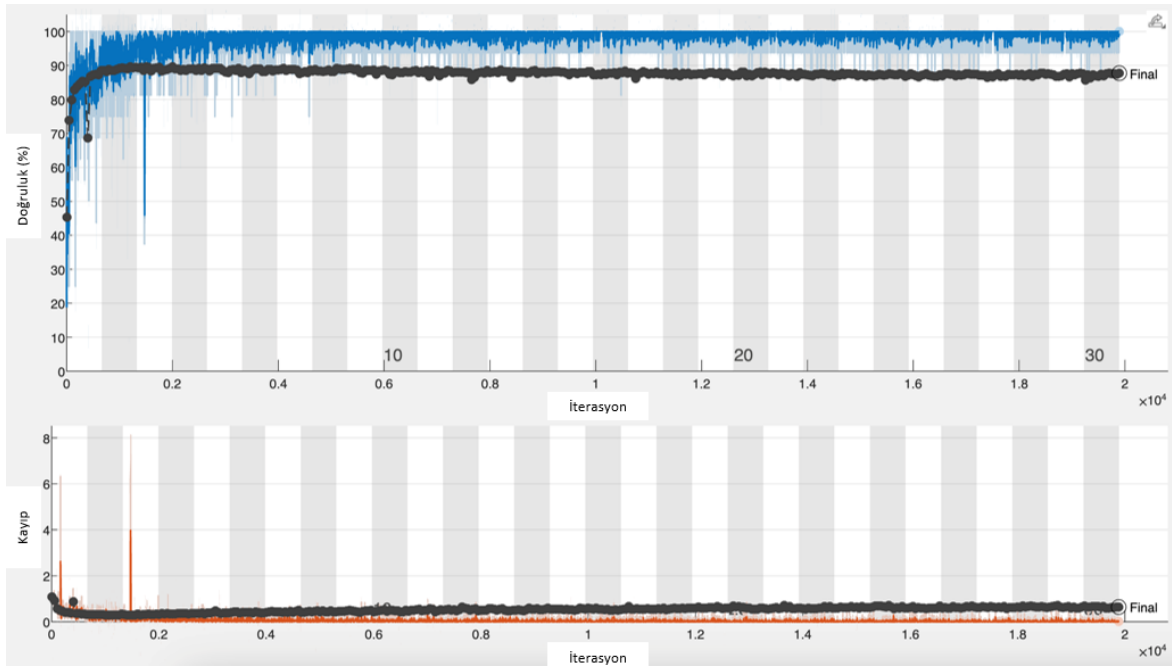
Çapraz Doğrulama Yüzdesi	Kelime Gömme Katmanı Boyutu	Filtre Sayısı	Doğruluk (%)
%10	200	100	89,12

%20	200	100	88,79
%30	200	100	88,64
%10	100	100	89,45
%10	300	100	88,66
%10	100	200	88,53
%10	100	50	87,34

0	636	34	27
1	13	651	16
2	33	37	70
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.20. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=30)

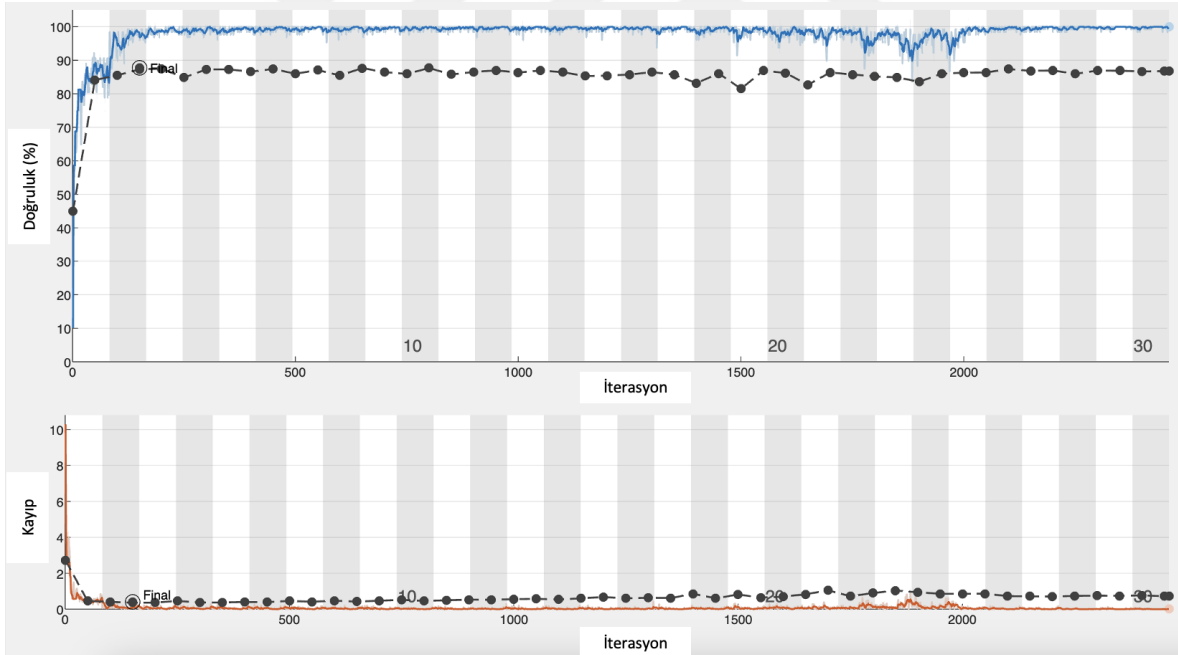


Şekil 6.21. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=30)

0	1961	71	62
1	122	1872	46
2	147	110	160
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.22. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)



Şekil 6.23. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=30, epok=30)

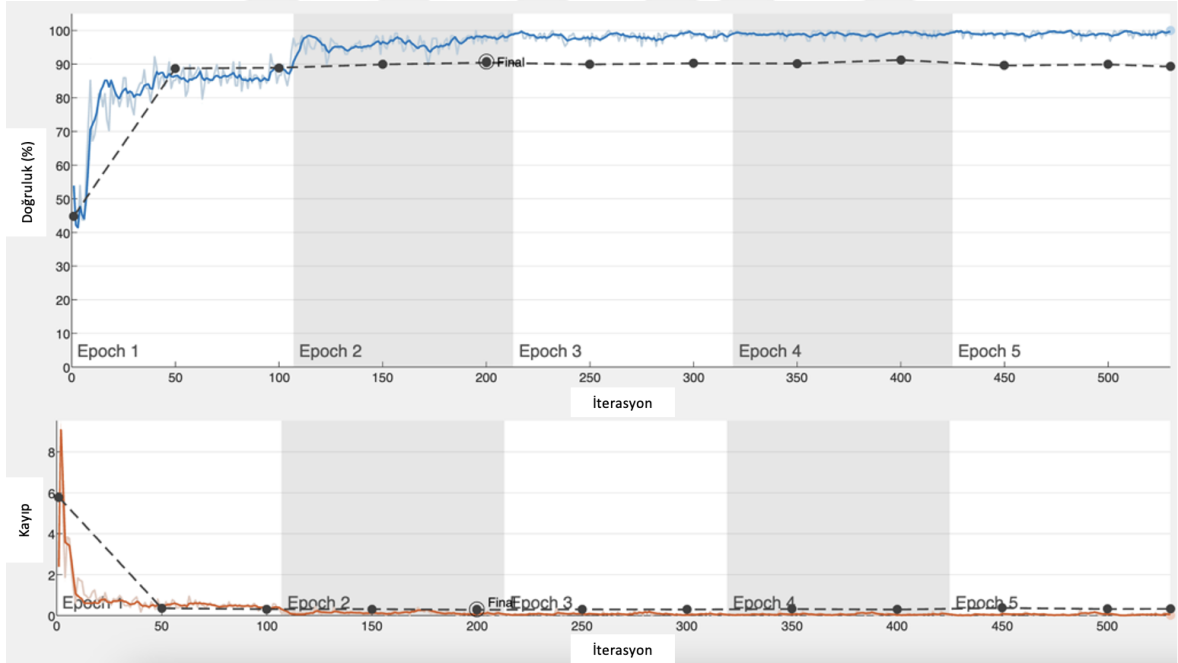
Şekil 6.21’de verilen eğitim süreci gözlemlendiğinde yaklaşık olarak 5.epoktan itibaren başarı değerinin düştüğü ve kaybın arttığı gözlemlenmiştir. Bu sebeple 30.epok olarak çalıştırılan eğitim 5. Epok olarak değiştirilmiştir. Elde edilen yeni karışıklık matrisi

Şekil 6.22’de, gözlemlenen yeni eğitim süreci ise Şekil 6.23’de paylaşılmıştır. Epok değeri 5 olarak seçildiğinde doğruluk değeri %90,77 olarak elde edilmiştir.

0	646	17	34
1	9	644	27
2	28	25	87
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.24. ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)



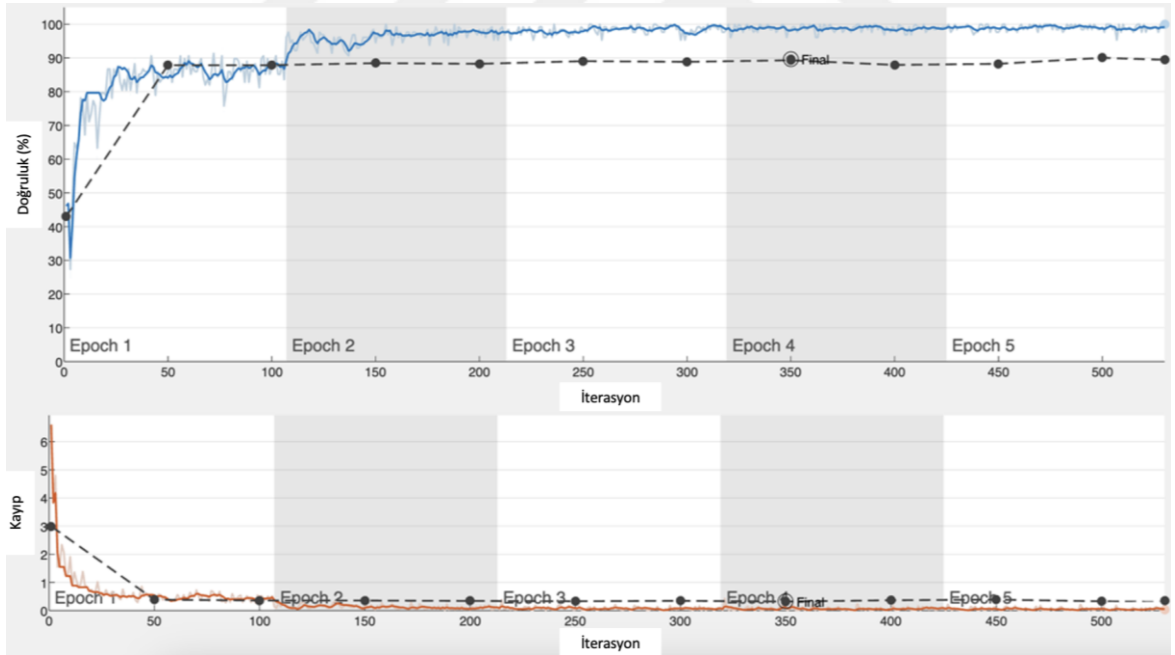
Şekil 6.25. ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)

En yüksek başarıyı veren parametreler ve epok değeri için elde edilen verilerin ardından öznitelik çıkarımı amacıyla kelime çantası modeli uygulanmış olup karışıklık matrisi Şekil 6.24’de, eğitim süreci ise Şekil 6.25’de paylaşılmıştır. Kelime çantası modeli

uygulandığından yeni doğruluk değeri %89,65 olarak elde edilmiş olup, doğruluğun azaldığı gözlemlenmiştir. Bu sebeple kelime çantası modelinin bu sınıflandırma algoritması için başarıyı artırıcı nitelikte bir algoritma olmadığı görülmektedir.

Doğru Sınıf	0	1	2
0	644	21	32
1	24	627	29
2	21	30	89
	0	1	2
	Tahmin Edilen Sınıf		

Şekil 6.26. Kelime çantası modeli uygulandığında ESA karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)



Şekil 6.27. Kelime çantası modeli uygulandığında ESA eğitim süreci (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)

Son aşama olarak öznitelik çıkarımı için ESA algoritmasının kendisi kullanılmıştır. Bu aşamada veri seti ESA ile eğitilmiş olup, eğitim sonunda tam bağlı katmandan öznitelikler çıkarılmıştır. Çıkarılan öznitelikler klasik makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırılmıştır. Elde edilen sonuçlar Tablo 6.7'de paylaşılmıştır. ESA

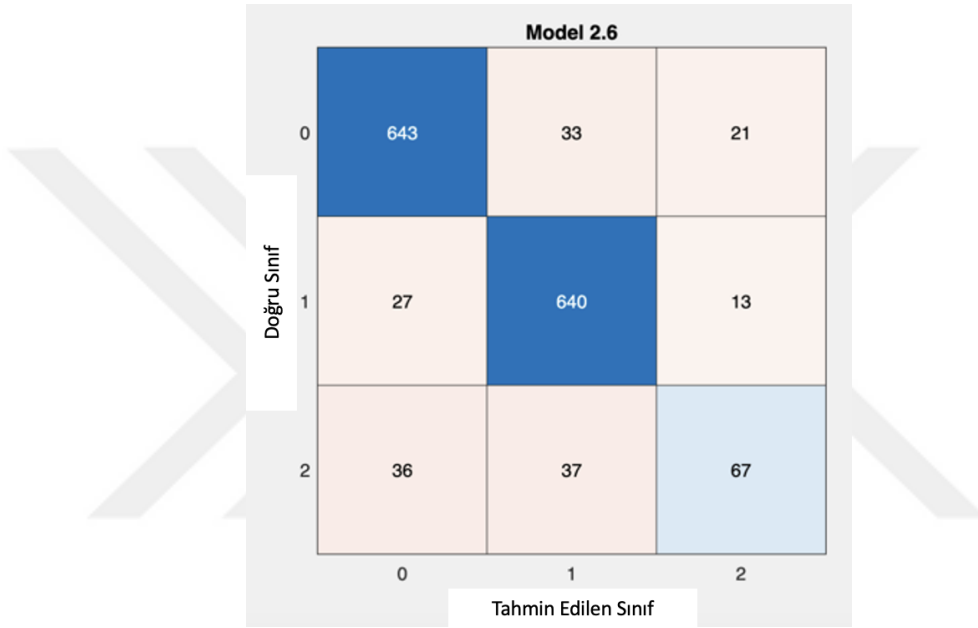
algoritmasının kullanıldığı tüm deneylere ait sonuçlar analiz edildiğinde en yüksek başarımın Verimli Lojistik Regresyon uygulandığında %88,99 doğruluk değeri ile elde edildiği gözlemlenmiştir.

Tablo 6.7. ESA öznitelik çıkarımı kullanıldığında deney sonuçları

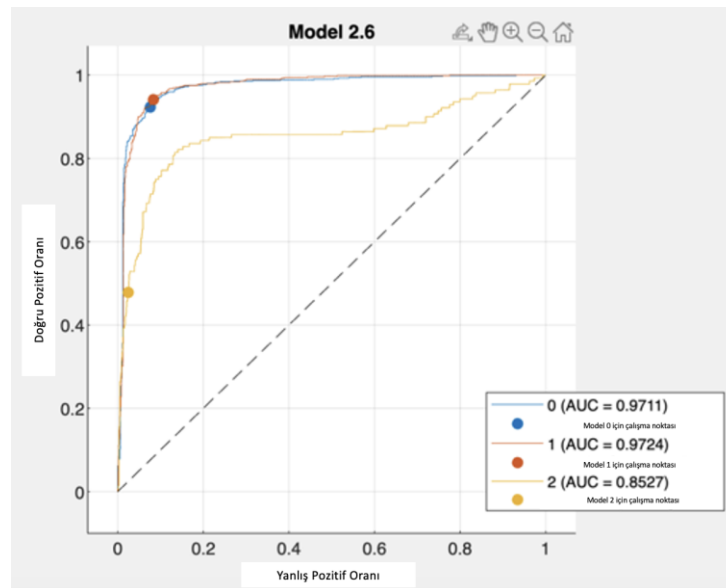
Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Karar Ağaçları (Kaba Ağaç)	0,76	0,82	0,79	0,793	88,27
Lineer Diskriminant Analiz	0,78	0,82	0,80	0,801	88,66
İkinci Dereceden Lineer Diskriminant Analiz	0,79	0,81	0,80	0,804	88,73
Verimli Logistik Regresyon	0,78	0,82	0,80	0,807	88,99
Verimli Lineer DVM	0,78	0,82	0,80	0,804	88,79
Naif Bayes (Kernel)	0,78	0,81	0,79	0,802	88,60
Destek Vektör Makineleri (Orta Gaussian)	0,79	0,82	0,80	0,805	88,86
K-En Yakın Komşu (Kübik)	0,79	0,82	0,80	0,805	88,86
Birleşik Modeller (Güçlendirilmiş Ağaçlar)	0,78	0,81	0,79	0,803	88,73

Yapay Sinir Ağları (Orta)	0,79	0,81	0,80	0,804	88,73
Kernel (DVM Kernel)	0,64	0,84	0,73	0,745	86,03

En yüksek doğruluk değerinin elde edildiği sonuçlara ait karışıklık matrisi ve ROC eğrisi Şekil 6.26’da ve Şekil 6.27’de verilmiştir.



Şekil 6.28. ESA öznitelik çıkarımında Verimli Lojistik Regresyon karışıklık matrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)



Şekil 6.29. ESA öznitelik çıkarımında Verimli Lojistik Regresyon ROC eğrisi (kelime gömme katman boyutu=100, gizli birimlerin sayısı=100, çapraz doğrulama yüzdesi=10, epok=5)

ESA algoritmasının hem sınıflandırma hem de öznitelik çıkarımında kullanıldığı tüm deneylere ait sonuçlar incelendiğinde en yüksek başarının %10 çapraz doğrulama yüzdesi, 100 kelime gömme katman boyutu, 100 gizli birim ve 5 epok süren eğitim değeri seçildiğinde ve hem öznitelik çıkarımında hem de sınıflandırmada ESA kullanıldığında %90,77 doğruluk değeri ile elde edildiği görülmüştür.

Elde edilen ROC eğrisi incelendiğinde en yüksek başarının 1 etiketli sınıfta yani pozitif yorumların sınıflandırılmasında elde edildiği görülmüştür. En çok hatanın ise veri setindeki en az veri sayısına sahip 2 etiketli yani nötr yorumlarda yapıldığı gözlemlenmiştir. Bu durumun önüne geçmek amacıyla veri setindeki nötr yorumlarının sayısının artırılması hedeflenmektedir.

ESA algoritması ile çapraz doğrulama yüzdesi %10 seçilerek veri seti eğitildikten sonra test edilecek veri sayısının diğer algoritmalarla kıyasla daha az olması sebebiyle Verimli Lojistik Regresyon algoritması diğer algoritmalarla kıyasla daha yüksek bir başarı değeri sağlamıştır.

6.3.5. GTB sınıflandırma algoritmasının uygulanması

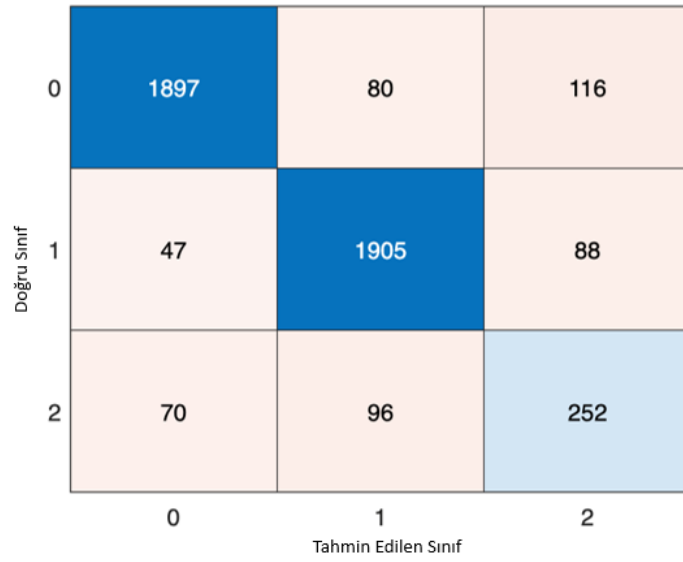
Veri setinin ön işlemi tamamlandıktan sonra GTB yapısının ağ mimarisi oluşturulmaya başlanır. GTB ağ mimarisi oluşturulurken ilk önce, dizi verilerini ağa sokmak için bir giriş katmanı eklenmiştir ve giriş boyutunu 1 olarak ayarlanmıştır. Ardından bir sözcük gömme katmanı ve sözcük kodlamasıyla aynı sayıda sözcükler eklenmiştir. Ardından GTB katmanı ve tam bağlı katman eklenmiştir. Son olarak çıktıları sınıflandırmak için; çıktı boyutu K olan bir tamamen bağlı katman, bir softmax katmanı ve K'nın sınıf sayısı olduğu bir sınıflandırma katmanı eklenmiştir.

Veri seti eğitim ve test olarak ayrılırken en ideal çapraz doğrulama yüzdesinin belirlenmesi amacıyla denemeler yapılmıştır. Bunun yanı sıra deney süresince çapraz doğrulama yüzdesinde, kelime gömme katmanı boyutunda ve gizli birimlerin sayısında değişiklikler yapılarak değişimler gözlemlenmiştir. Gözlem sonuçlarına ait elde edilen değerler Tablo 6.8'de verilmiştir.

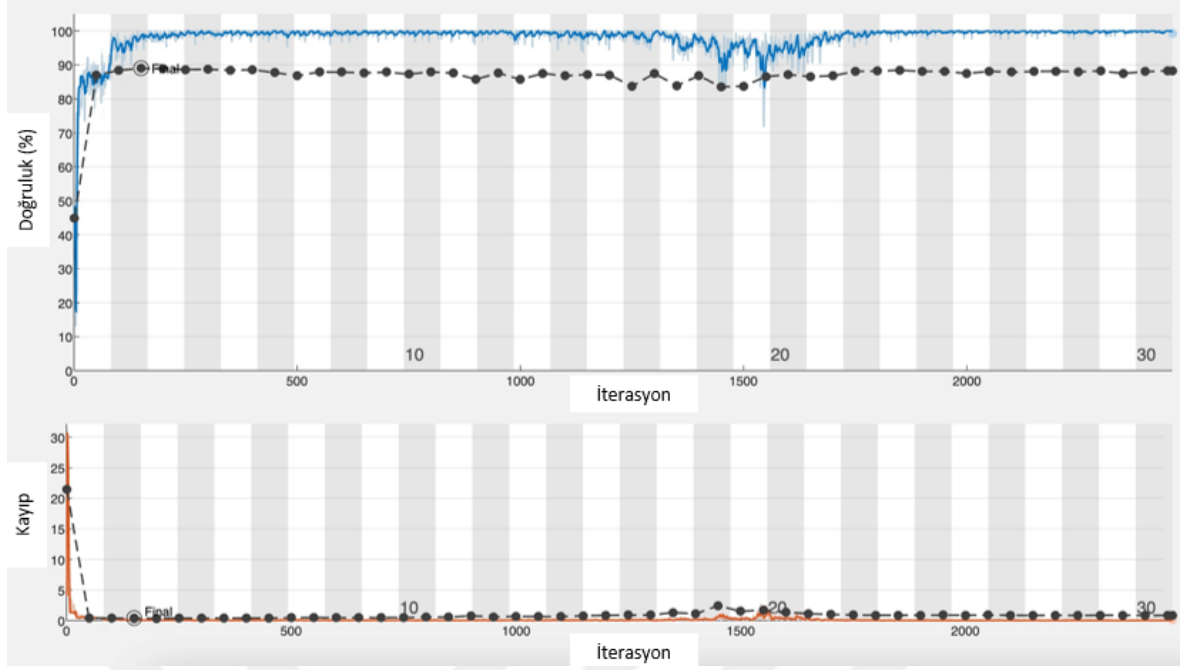
Sınıflandırma sonucunca GTB yöntemi kullanılarak %89,08 doğruluk değeri elde edilmiştir. Sınıflandırma performansını temsilen karışıklık matrisi Şekil 6.28’de, GTB eğitim süreci şekil 6.29’da verilmiştir.

Tablo 6.8. 30 epok için GTB deney sonuçları

Çapraz Doğrulama Yüzdesi	Kelime Gömme Katmanı Boyutu	Gizli Birimlerin Sayısı	Doğruluk (%)
%10	100	200	87,80
%20	100	200	87,61
%30	100	200	88,60
%30	200	200	87,85
%30	300	200	89,08
%30	300	50	89,04
%30	300	100	88,53



Şekil 6.30. GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)



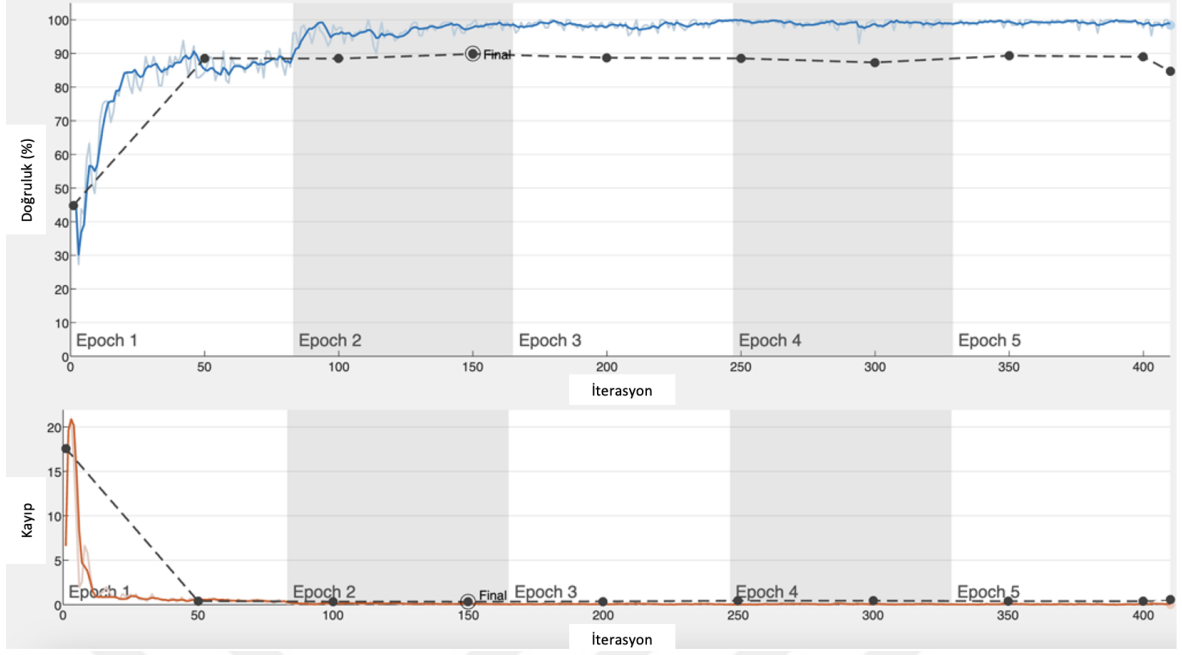
Şekil 6.31. GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=30)

Şekil 6.29’da verilen eğitim süreci gözlemlendiğinde yaklaşık olarak 5.epoktan itibaren başarı değerinin düştüğü ve kaybın arttığı gözlemlenmiştir. Bu sebeple 30.epok olarak çalıştırılan eğitim 5. Epok olarak değiştirilmiştir. Elde edilen yeni karışıklık matrisi Şekil 6.30’da, gözlemlenen yeni eğitim süreci ise Şekil 6.31’de paylaşılmıştır. Epok değeri 5 olarak seçildiğinde doğruluk değeri %89,96 olarak elde edilmiştir.

0	1940	56	97
1	56	1893	91
2	72	85	261
	0	1	2

Tahmin Edilen Sınıf

Şekil 6.32. GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

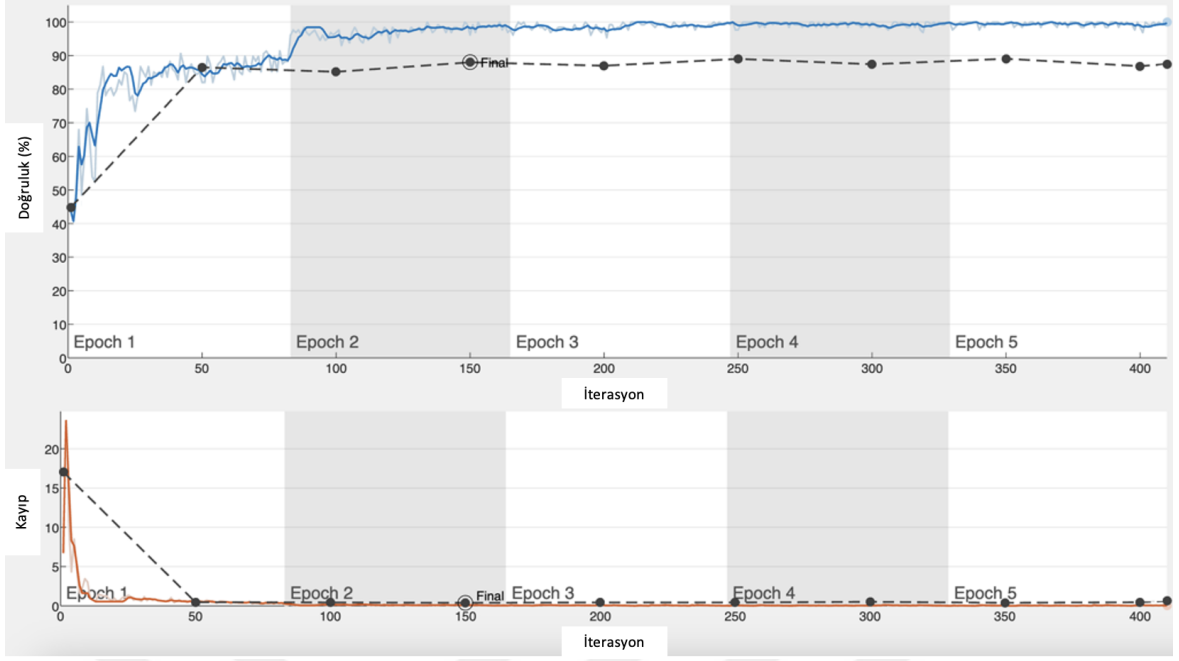


Şekil 6.33. GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

En yüksek başarıyı veren parametreler ve epok değeri için elde edilen verilerin ardından öznitelik çıkarımı amacıyla kelime çantası modeli uygulanmış olup karışıklık matrisi Şekil 6.32'de, eğitim süreci ise Şekil 6.33'de paylaşılmıştır. Kelime çantası modeli uygulandığından yeni doğruluk değeri %88,09 olarak elde edilmiş olup, doğruluğun azaldığı gözlemlenmiştir. Bu sebeple kelime çantası modelinin bu sınıflandırma algoritması için başarıyı artırıcı nitelikte bir algoritma olmadığı görülmektedir.

0	1856	85	152
1	35	1882	123
2	53	94	271
	0	1	2
	Tahmin Edilen Sınıf		

Şekil 6.34. Kelime çantası modeli kullanıldığında GTB karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)



Şekil 6.35. Kelime çantası modeli kullanıldığında GTB eğitim süreci (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

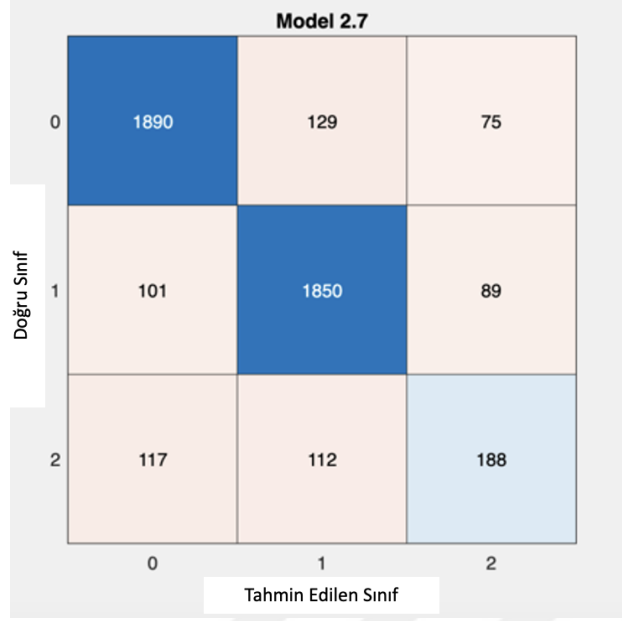
Son aşama olarak öznitelik çıkarımı için GTB algoritmasının kendisi kullanılmıştır. Bu aşamada veri seti GTB ile eğitilmiş olup, eğitim sonunda tam bağlı katmandan öznitelikler çıkarılmıştır. Çıkarılan öznitelikler uygulaması kullanılarak klasik makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırılmıştır. Elde edilen sonuçlar Tablo 6.9’da paylaşılmıştır. GTB algoritmasının kullanıldığı tüm deneylere ait sonuçlar analiz edildiğinde en yüksek başarımın Verimli Lineer DVM uygulandığında %86,31 doğruluk değeri ile elde edildiği gözlemlenmiştir.

Tablo 6.9. GTB öznitelik çıkarımı kullanıldığında deney sonuçları

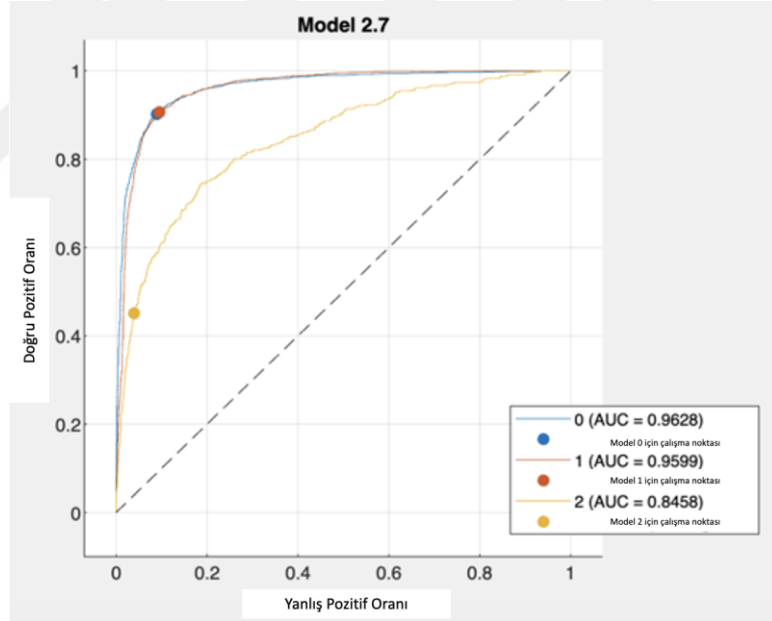
Kullanılan Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Karar Ağaçları (Kaba Ağaç)	0,74	0,76	0,75	0,755	86,00
Lineer Diskriminant Analiz	0,76	0,76	0,76	0,744	85,10

İkinci Dereceden Lineer Diskriminant Analiz	0,76	0,77	0,76	0,757	85,96
Verimli Logistik Regresyon	0,75	0,77	0,76	0,761	86,29
Verimli Lineer DVM	0,75	0,77	0,76	0,761	86,31
Naif Bayes (Kernel)	0,77	0,75	0,76	0,750	85,43
Destek Vektör Makinelere (Kaba Gaussian)	0,75	0,77	0,76	0,760	86,22
K-En Yakın Komşu (Orta)	0,76	0,77	0,76	0,761	86,20
Birleşik Modeller (Güçlendirilmiş Ağaçlar)	0,75	0,76	0,75	0,756	85,98
Yapay Sinir Ağları (Dar)	0,75	0,77	0,76	0,759	86,16
Kernel (DVM)	0,75	0,77	0,76	0,759	86,16

En yüksek doğruluk değerinin elde edildiği sonuçlara ait karışıklık matrisi ve ROC eğrisi Şekil 6.34’de ve Şekil 6.35’de verilmiştir.



Şekil 6.36. GTB öznitelik çıkarımında Verimli Lineer DVM karışıklık matrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)



Şekil 6.37. GTB öznitelik çıkarımında Verimli Lineer DVM ROC eğrisi (kelime gömme katman boyutu=300, gizli birimlerin sayısı=200, çapraz doğrulama yüzdesi=30, epok=5)

GTB algoritmasının hem sınıflandırma hem de öznitelik çıkarımında kullanıldığı tüm deneylere ait sonuçlar incelendiğinde en yüksek başarımın %30 çapraz doğrulama yüzdesi, 300 kelime gömme katman boyutu, 200 gizli birim ve 5 epok süren eğitim değeri seçildiğinde ve hem öznitelik çıkarımında hem de sınıflandırmada GTB kullanıldığında %89,96 doğruluk değeri ile elde edildiği görülmüştür.

Elde edilen ROC eğrisi incelendiğinde en yüksek başarımın 0 etiketli sınıfta yani negatif yorumların sınıflandırılmasında elde edildiği görülmüştür. En çok hatanın ise veri setindeki en az veri sayısına sahip 2 etiketli yani nötr yorumlarda yapıldığı gözlemlenmiştir. Bu durumun önüne geçmek amacıyla veri setindeki nötr yorumlarının sayısının artırılması hedeflenmektedir.

Destek Vektör Makineleri çok sınıflı ve doğrusal olarak ayrıştırılabilen verilerde oldukça başarılı bir yöntemdir. Bu çalışmada kullanılan veri setinde de problemin doğrusal olarak ayrıştırılabilmesi sebebiyle bu yöntemde Destek Vektör Makineleri diğer algoritmalara kıyasla daha yüksek başarı değerleri vermiştir.

Tez çalışması süresince çeşitli öznitelik çıkarma yöntemleri, çeşitli sınıflandırıcıların parametreleri değiştirilerek en yüksek başarıyı elde etme hedefiyle veri seti üzerinde test edilmiştir. Tablo 6.10'da gösterildiği üzere yapılmış olan tüm bu testler sonucunda en yüksek doğruluk değeri hem öznitelik seçiminde hem de sınıflandırılma ESA'nın kullanıldığında %90,77 doğruluk değeri ile elde edilmiştir.

ESA yapısı; TSA modelinden türeyen UKSB, Çift Yönlü UKSB, GTB ile karşılaştırıldığında hem öznitelik çıkarımı hem de sınıflandırma adımında daha yüksek başarılar vermiştir. ESA yapısı örüntüleri uzay boyutunda algılayabilirken, TSA yapısından türeyen diğer algoritmalar örüntüleri zamansal boyutta görebilirler. Bu sebeple UKSB, Çift Yönlü UKSB ve GTB algoritmaları sıraları verilerde daha başarılı olurken, ESA algoritması görüntü ve metin işlemede çok daha başarılıdır. Buna ek olarak öznitelik çıkarımı ve sınıflandırma aşamasında ESA algoritması kullanıldığında evrişim katmanında oluşturulan öznitelik matrislerinin havuzlama katmanında boyutlarının azaltılmasıyla işlenecek parametreler azalmış ve diğer deneylere kıyasla en hızlı sınıflandırma süresi elde edilmiştir.

Tablo 6.10. Kullanılan sınıflandırma algoritmalarına göre elde edilen başarı değerleri

Öznitelik Çıkarma Yöntemi	Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Kelime Çantası Modeli	Destek Vektör Makineleri (DVM)	0,79	0,80	0,79	0,786	87,57
Uzun Kısa Süreli Bellek (UKSB)	Yapay Sinir Ağları (Dar)	0,81	0,80	0,80	0,808	88,8

Öznitelik Çıkarma Yöntemi	Sınıflandırma Algoritması	Kesinlik	Duyarlılık	F1-Skor	Kappa	Doğruluk (%)
Çift Yönlü Uzun Kısa Süreli Bellek (Çift Yönlü UKSB)	Yapay Sinir Ağları (Orta)	0,80	0,82	0,81	0,813	89,23
Evrişimsel Sinir Ağları (ESA)	Evrişimsel Sinir Ağları (ESA)	0,83	0,83	0,83	0,841	90,77
Geçitli Tekrarlayan Birim (GTB)	Geçitli Tekrarlayan Birim (GTB)	0,83	0,82	0,82	0,827	89,96

Yapılan tüm sınıflandırma çalışmalarında olduğu gibi bu çalışmada da kullanılan veri setinin dengeli bir veri seti olup olmaması, veri setindeki toplam veri sayısı, toplanan verilerin gerçekliğe uygun olması, sınıf sayısı gibi kriterler sınıflandırma performansını oldukça etkilemektedir. Bu çalışmada kullanılan veri setinin birden fazla platformdan toplanarak elde edilmesi, literatürdeki çalışmalar ile kıyaslandığında yeterli veri sayısına sahip olması ve bu verilerin veri seti içerisinde dengeli dağılması sebebiyle oldukça gerçekçi bir veri seti olduğu söylenebilir. Bu sebeple sınıflandırma sonucunda elde edilen başarıların da gerçekliğe yakınlığının yüksek olduğu söylenebilir.

Tablo 6.11. Aynı veri setine sahip çalışmalarla karşılaştırma

Çalışma	Doğruluk Değeri (%)
Barış Ulusoy [12]	85,49
Şeyma Sarıgil [13]	78,92
Bu çalışma	90,77

Tablo 6.11’de görüldüğü üzere bu çalışmada kullanılan veri seti ile aynı veri setinin kullanıldığı Barış Ulusoy [12] ve Şeyma Sarıgil [13] tarafından yapılan çalışmalarda elde edilen başarının üzerine çıkmıştır.

7. SONUÇ VE ÖNERİLER

Günümüzde e-ticarete olan ilginin giderek artması ile beraber e-ticaret platformlarının sayısı ve bu platformlarda bulunan ürün çeşitliliği giderek artmaktadır. Bu platformda tüketici için hangi satıcıdan hangi ürünün satın alınmasına karar verilmesi, satıcı için ise müşteri memnuniyetinin artırılması oldukça büyük önem taşımaktadır. Bu amaçla ürünlere ait yorumların ifade ettikleri duygulara göre sınıflandırılarak duygu analizlerinin yapılması oldukça önemlidir. Bu çalışmada da ürün yorumlarının olumlu, olumsuz ve nötr olarak sınıflandırılması hedeflenmiştir. Bu hedef doğrultusunda klasik makine öğrenmesi ve derin öğrenme modelleri kullanılmıştır.

Yapılan deneyler sonrasında elde edilen sonuçlar başarı kriterlerine göre değerlendirilmiştir. Metin verisinin sınıflandırmaya uygun hale gelmesi amacıyla normalleştirilmesi ve ön işleme uygulanması sınıflandırmanın başarısı açısından oldukça önemlidir. Yapılan deneyler sonucunda “E-Ticaret Ürün Yorumları” veri setine Kelime Çantası modeli, Uzun Kısa Süreli Bellek, Çift Yönlü Uzun Kısa Süreli Bellek, Evrişimsel Sinir Ağları, Geçitli Tekrarlayan Birim ile öznitelik çıkarımı uygulanmasının yanı sıra; Karar Ağaçları, Evrişimsel Sinir Ağları, Lineer Diskriminant Analiz, Uzun Kısa Süreli Bellek, İkinci Dereceden Diskriminant Analiz, Çift Yönlü Uzun Kısa Süreli Bellek, Verimli Lineer DVM, Geçitli Tekrarlayan Birim , Verimli Logistik Regresyon, Naif Bayes, K-En Yakın Komşu, Birleşik Modeller, Yapay Sinir Ağları, Destek Vektör Makineleri, Uzun Kısa Süreli Bellek, Çift Yönlü Uzun Kısa Süreli Bellek, Evrişimsel Sinir Ağları ve Geçitli Tekrarlayan Birim algoritmaları ile sınıflandırma yapılmıştır. En yüksek doğruluk değerini veren algoritmanın %90,77 ile Evrişimsel Sinir Ağları algoritması olduğu görülmüş olup bu tez çalışmasında hem öznitelik çıkarımında hem de sınıflandırmada ESA algoritmasının kullanıldığı bu model önerilmektedir.

Bu çalışmada kullanılan veri seti ile elde edilen doğruluk değerinin aynı veri setinin kullanıldığı diğer çalışmalara kıyasla daha yüksek bir doğruluk değeri olması çalışmanın değerini vurgulamaktadır.

Yapılan tüm sınıflandırma çalışmalarında olduğu gibi bu çalışmada da kullanılan veri setinin dengeli bir veri seti olup olmaması, veri setindeki toplam veri sayısı, toplanan verilerin gerçekliğe uygun olması, sınıf sayısı gibi kriterler sınıflandırma performansını oldukça etkilemektedir. Bu sebeple literatürdeki çalışmalar ile kıyaslandığında oldukça yüksek veri sayısına sahip olması ve veri seti elde edilirken çeşitli e-ticaret platformlarının

kullanılmış olması çalışmanın gerçekliğe yakınlığını arttırmıştır. Bu sebeple sınıflandırma sonucunda elde edilen başarıların da gerçekliğe yakınlığının yüksek olduğu söylenebilir.

Gelecek çalışmalarda farklı metin ön işleme yöntemlerinin ve farklı sınıflandırma algoritmalarının denenmesi ile sınıflandırma sonucunda daha yüksek başarılar elde edilebilir. Buna ek olarak, yapılan her ürün yorumunda yorumun olumlu veya olumsuz olmasıyla doğru orantılı olarak puanlandırılmadığı göz önüne alınarak ürün yorumlarının gerçek puanlarının bu çalışmadan elde edilen olumlu olumsuz yorumlar doğrultusunda ağırlıklandırılarak ürünün gerçek puan değerinin hesaplanmasında kullanılabileceği düşünülmektedir. Son olarak elde edilen sonuçlar değerlendirildiğinde uygulanan sınıflandırma algoritmalarının doğruluk değerleri arasındaki farkın istatistiksel açıdan ne kadar anlamlı olduğu test edilebilir.



KAYNAKLAR

- [1] Cumali Çiçek, “Türkiye’de ve küreselde e-ticaret trendleri ” in *E-Ticaret sitesi açma rehberi*, 2022. [Online]. Available: books.google.com.tr/books?id=8xR_EAAAQBAJ&printsec=frontcover&hl=tr#v=onepage&q&f=false. Accessed: 9 September 2023.
- [2] *Ekonominin dönüştürücü gücü: e-ticaret etki analizi*, 2021. [Online]. Available: tubisad.org.tr/tr/images/pdf/tubisad-e-commerce_impact_assessment-launchpresentation.pdf. Accessed: 9 September 2023.
- [3] İ. Kavaklı, *E-Ticaret rehberi*, 2018. [Online]. Available: books.google.com.tr/books/about/E_T%C4%B0CARET_REHBER%C4%B0.html?id=HPC3DwAAQBAJ&redir_esc=y, Accessed: 9 September 2023.
- [4] H. Nizam, S. S. Akın, “Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması”, *inetconf19*, November 2020. [Online]. Available: inet-tr.org.tr/inetconf19/bildiri/10.pdf, Accessed: 9 September 2023.
- [5] F. S. Çetin, G. Eryiğit, “Türkçe hedef tabanlı duygu analizi için alt görevlerin incelenmesi – hedef terim, hedef kategori ve duygu sınıfı belirleme”, *Journal of Information Technologies*, vol. 11, issue 1, pp. 43-56, January 2018.
- [6] Ö. Çelik, G. Kaplan, “Yeniden Örnekleme teknikleri kullanarak SMS verisi üzerinde metin sınıflandırma çalışması” , *Erciyes University Journal of Institute Of Science and Technology*, vol. 36, issue 3, December 2020.
- [7] Ö. Çelik, B. C. Koç, “TF-IDF, Word2vec ve Fasttext vektör model yöntemleri ile Türkçe haber metinlerinin sınıflandırılması”, *Dokuz Eylül University Faculty of Engineering Journal of Science and Engineering*, vol. 23, issue 67, pp. 121-127, January 2021.
- [8] B. POLAT, “Türkçe ürün yorumları verisi ile duygu analizi”, M.S. Thesis, Dept. Computer Eng., Ankara Üniversitesi, Ankara, Turkey, June 2021. [Online].

Available:

dspace.ankara.edu.tr/xmlui/bitstream/handle/20.500.12575/73816/10401060.pdf?sequence=1&isAllowed=y, Accessed: 9 September 2023.

- [9] E. Aydemir, M. Işık, T. Tuncer, “Türkçe haber metinlerinin Çok Terimli Naive Bayes algoritması kullanılarak sınıflandırılması”, *Fırat University Journal of Engineering Science*, vol. 33, issue 2, pp. 519-526, 2021, doi: 10.35234/fumbd.871986.
- [10] A. A. Kolukısa, “Yapay Sinir Ağları ile metin sınıflandırması - Weka Derin Öğrenme Eklentileri Üzerine İnceleme”, August 2021, doi: 10.13140/RG.2.2.12331.77606.
- [11] H. B. Dogru, S. Tilki, A. Jamil, A. A. Hameed, “Deep Learning-based classification of News Texts Using Doc2Vec Model”, *International 2021 1st Conference on Artificial Intelligence and Data Analytics (CAIDA)*, 2021.
- [12] B. Ulusoy, Kaggle, 2022. [Online]. Available: github.com/barissulusoy/kaggle/blob/837e5f6b564d27fc05e51c85c05c9cc3faf0d46d/r-n-yorumlar-duygu-analizi.ipynb, Accessed: 9 September 2023.
- [13] Ş. Sarıgil, Github, 2021. [Online]. Available: kaggle.com/code/seymasa/trendyol-comment, Accessed: 9 September 2023.
- [14] D. Boswell, Introduction to Support Vector Machines, 2002. [Online]. Available: pzs.dstu.dp.ua/DataMining/svm/bibl/IntroToSVM.pdf, Accessed: 9 September 2023.
- [15] F. Maleki, K. Ovens, K. Najafian, B. F. MEng, C. Reinhold, R. Forghani, “Overview of Machine Learning part 1: fundamentals and classic approaches”, *Neuroimaging Clinics of North America*, vol. 30, issue 4, November 2020, doi: 10.1016/j.nic.2020.08.007.
- [16] O. Kaynar, M. Yıldız, Y. Görmez, A. Albayrak, “Sentiment analysis with machine learning techniques”, *International Artificial Intelligence and Data Processing Symposium (IDAP'16)*, pp 17-18, September 2016.
- [17] O. Manjrekar, M. Duduković, “Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique”, *Chemical Engineering Science*, vol. 2, May 2019, doi: 10.1016/j.cesx.2019.100023.

- [18] I. N. Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, S. F. R. Alves, “Artificial Neural Networks: a practical course”, *Springer*, 2017, doi: 10.1007/978-3-319-43162-8.
- [19] F. Bre, J. M. Gimenez, V. D. Fachinotti,” Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks”, *Energy and Buildings*, vol. 158, pp. 1429-1441, January 2018, doi: 10.1016/j.enbuild.2017.11.045.
- [20] S. Arifin, A. K. Wijaya, R. Nariswari, A. Yudistira, Suwarno, Faisal, D. Wihardini, “Long Short-Term Memory (LSTM): Trends and future research potential”, *International Journal of Emerging Technology and Advanced Engineering*, vol. 13, pp. 2250-2459, May 2023, doi: 10.46338/ijetae0523_04.
- [21] A. A. SÜZEN, Z. YILDIZ, T. YILMAZ, “LSTM tabanlı derin sinir ağı ile ayak taban basınç verilerinden VKİ durumlarının sınıflandırılması”, *BEU Journal of Science*, vol. 8, issue 4, pp 1392-1398, 2019.
- [22] L. Mu, F. Zheng, R. Tao, Z. Quingzhou, Z. Kapelan, “Hourly and daily urban water demand predictions using a Long Short-Term Memory based model”, *American Society of Civil Engineers*, vol. 146, issue 9, June 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001276.
- [23] B. Bohara, R. I. Fernandez, V. Gollapudi, X. Li, “Short-Term aggregated residential load forecasting using BiLSTM and CNN-BiLSTM”, *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*, 2022, doi: 10.1109/3ICT56508.2022.9990696.
- [24] Hongyu Jia, Yaoyu Yang, Jintang An, Rui Fu, “A ship trajectory prediction model based on Attention-BiLSTM Optimized by the whale optimization algorithm”, *Applied Sciences*, vol. 13, pp. 4907, 2023, doi: 10.3390/app13084907.
- [25] Y. Li, L. N. Harfiya, K. Purwandari, Y. D. Lin, “Real-time cuffless continuous blood pressure estimation using deep learning model”, *Sensors 2020*, vol. 20, pp. 5606, 2020, doi: doi.org/10.3390/s20195606.
- [26] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, “Convolutional Neural Networks:

- an overview and application in radiology” *Insights Imaging*, vol. 9, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.
- [27] Van Hiep Phung, Eun Joo Rhee, “A high-accuracy model average ensemble of Convolutional Neural Networks for classification of cloud image patches on small datasets”, *Applied Sciences*, vol. 9, pp. 4500, October 2019, doi: 10.3390/app9214500.
- [28] P. Kim ,”Convolutional Neural Network”, MATLAB deep learning, pp. 121–147, 2017, doi: 10.1007/978-1-4842-2845-6_6.
- [29] S.Mohsen, “Recognition of human activity using GRU deep learning algorithm”, *Multimedia Tools and Applications*, May 2023, doi: 10.1007/s11042-023-15571-y.
- [30] R.Katarya, A. Raturi, A. Mehndiratta, A. Thapper, “Impact of machine learning techniques in precision agriculture”, *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, pp. 1-6, February 2020, doi: 10.1109/ICETCE48199.2020.9091741.
- [31] P. Czodrowski, “Count on kappa”, *Journal of Computer-Aided Molecular Design*, vol. 28, pp. 1049–1055, July 2014, doi: 10.1007/s10822-014-9759-6.
- [32] W. A. Qader, M. M. Ameen, B. I. Ahmed, “An overview of Bag of Words;importance, implementation, applications, and challenges”, *Fifth International Engineering Conference on Developments in Civil & Computer Engineering Applications*, June 2019, doi: 10.1109/IEC47844.2019.8950616.