T.C.
KAYSERİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
HESAPLAMALI BİLİMLER VE MÜHENDİSLİK
ANABİLİM DALI

YENİ TÜRKÇE DUYGUSAL VERİ SETİ ÜZERİNDE KONUŞMADAN DUYGU TANIMA

Hazırlayan
Emel ÇOLAKOĞLU

Danışman
Dr. Öğr. Üyesi Serhat HIZLISOY

Yüksek Lisans Tezi

Ağustos 2023, KAYSERİ

T.C.
KAYSERİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
HESAPLAMALI BİLİMLER VE MÜHENDİSLİK
ANABİLİM DALI

YENİ TÜRKÇE DUYGUSAL VERİ SETİ ÜZERİNDE KONUŞMADAN DUYGU TANIMA

(Yüksek Lisans Tezi)

Hazırlayan
Emel ÇOLAKOĞLU

Danışman
Dr. Öğr. Üyesi Serhat HIZLISOY

Ağustos 2023, KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Bu tezin kendi çalışmam olduğunu, planlanmasından yazımına kadar hiçbir aşamasında etik dışı davranışımın olmadığını, tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları kaynaklar listesine aldığımı, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

Emel ÇOLAKOĞLU

TEZ İNTİHAL RAPORU TESLİM FORMU

T.C.
KAYSERİ ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü Müdürlüğü

Öğrenci No : 4034620006
Anabilim Dalı : Hesaplamalı Bilimler ve Mühendislik ABD
Program Adı : Tezli Yüksek Lisans
Tez Başlığı : Yeni Türkçe Duygusal Veriseti Üzerinde Konuşmadan Duygu Tanıma

Yukarıda bilgileri verilen tez çalışmasının a) Giriş, b) Ana bölümler ve c) Sonuç kısımlarından oluşan (Kapak, Önsöz, Özet, İçindekiler ve Kaynakça hariç) toplam 69 sayfalık kısmına ilişkin 06/09/2023 tarihinde ***Turnitin*** intihal programından aşağıda belirtilen filtreleme uygulanarak alınmış olan özgünlük raporuna göre tezin benzerlik oranı**: %9'**dur.

Uygulanan filtrelemeler:

1- Giriş dahil
2- Ana Bölümler dahil
3- Sonuç dahil
4- Alıntılar dahil
5- Kapak hariç
6- Önsöz ve Teşekkür hariç
7- İçindekiler hariç
8- Kaynakça hariç
9- Özet hariç
10- Yedi (7) kelimeden daha az örtüşme içeren metin kısımları hariç

Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Önerisi Hazırlama, Tez Yazma ve Teslim Yönergesini inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini, aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim. 07/09/2023

Danışman: Dr. Öğr. Üyesi Serhat HIZLISOY Öğrenci: Emel ÇOLAKOĞLU

YÖNERGEYE UYGUNLUK

"Yeni Türkçe Duygusal Veriseti Üzerinde Konuşmadan Duygu Tanıma" adlı Yüksek Lisans Tezi, Kayseri Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi'ne uygun olarak hazırlanmıştır.

Hazırlayan	**Danışman**
Emel ÇOLAKOĞLU	Dr. Öğr. Üyesi Serhat HIZLISOY

Hesaplamalı Bilimler ve Mühendislik ABD Başkanı

Prof. Dr. Ercan KARAKÖSE

KABUL VE ONAY TUTANAĞI

Dr. Öğr. Üyesi Serhat HIZLISOY danışmanlığında **Emel ÇOLAKOĞLU** tarafından hazırlanan "**Yeni Türkçe Duygusal Veriseti Üzerinde Konuşmadan Duygu Tanıma**" adlı bu çalışma jürimiz tarafından Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü **Hesaplamalı Bilimler ve Mühendislik** Anabilim Dalında **Yüksek Lisans Tezi** olarak kabul edilmiştir.

25 /08 / 2023

JÜRİ:

Danışman	: Dr. Öğr. Üyesi Serhat HIZLISOY	……
Üye	: Dr. Öğr. Üyesi Ömer Faruk SARITAŞ	……
Üye	: Dr. Öğr. Üyesi Fatma Özge ÖZKÖK	……...……

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun ……...… tarih ve ……..…… sayılı kararı ile onaylanmıştır.

….. /…… / ..…

Doç. Dr. Ahmet TERZİ
Enstitü Müdürü

TEŞEKKÜR

Yüksek lisans çalışmamda bana yardımlarını ve desteğini eksik etmeyerek tez çalışmamı başarı ile tamamlamamı sağlayan danışman hocam Sayın Dr. Öğr. Üyesi Serhat Hızlısoy'a; Tezin gelişmesine yönlendirici görüş ve önerileriyle yardımcı olan ve manevi desteklerini esirgemeyen hocam Sayın Dr. Öğr. Üyesi Recep Sinan Arslan'a; Bu süreçteki sabır ve destekleri için sevgili eşim Ömer Faruk Çolakoğlu'na ve biricik çocuklarım Fatih Mehmet ve Yağmur'a; Tüm içtenliğimle teşekkür ederim.

Emel ÇOLAKOĞLU, Kayseri 2023

YENİ TÜRKÇE DUYGUSAL VERİ SETİ ÜZERİNDE KONUŞMADAN DUYGU TANIMA

Emel ÇOLAKOĞLU

Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü

Yüksek Lisans Tezi, Ağustos 2023

Danışman: Dr. Öğr. Üyesi Serhat HIZLISOY

ÖZET

İnsanlar arasındaki iletişimin en doğal ve hızlı yöntemi konuşmaktır. Konuşma sinyalleri konuşmanın içeriği haricinde konuşan kişinin ruh hali, yaşı, cinsiyeti, fizyolojisi ve lehçesi gibi birçok veriyi de barındırmaktadır. Duygu ise gerçekleşen bir olayın, kişilerin ya da varlıkların insanın iç dünyasında oluşturduğu etkidir. Duygular mutluluk, heyecan, öfke, sakinlik gibi gruplara ayrılabilir. Bireyler duygularını hem beden dili yoluyla hem de sesli iletişim ile birbirlerine aktarırlar. Ancak bazı durumlarda beden dilini yorumlamak ya da görebilmek mümkün olmamaktadır. Böyle durumlarda sesten duygunun çıkarımı daha fazla önem kazanmaktadır. Bu alanda yapılan ilk çalışmalarda genelde görüntü sinyallerinden duygu analizi yapılması üzerinde durulmuştur. Son yıllarda ise ses sinyali üzerinden duygu analizi çalışmaları gelişme göstermeye başlamıştır. Literatür incelendiğinde ise konuşmadan duygu tanıma alanında birçok çalışma olduğu görülmektedir. Yapılan çalışmalarda hem makine öğrenim hem de derin öğrenme algoritmaları kullanılmakta ve oldukça başarılı tanıma oranlarına ulaşılmıştır.

Bu tez kapsamında ise önceki çalışmalardan farklı olarak verisetini kendimiz oluşturduk. Bunun için Türk yapımı dizi ve filmler incelendi. Veriseti içindeki tüm kayıtlar 5 saniyeden oluşmakta ve 4 duygu içermektedir. Burada tezin önceki çalışmalardan en önemli farkı konuşmadan duygu tanıma için oluşturulmuş; yüksek veri kalitesi ve etiketlendirmesine sahip olan hazır veri setlerinin kullanılmamasıdır. Ayrıca Türkçe dili üzerinde yayın yapan çok az sayıda çalışma bulunmaktadır. Diğer bir noktada konuşmacı bağımsız bir model oluşturulmuş olmasıdır. Bu durumda aslında model performansını olumsuz etkileyebilecek bir durumdur. İncelenen çalışmalarda düşük tanıma oranları nedeniyle genelde tercih edilmemiştir.

Özniteliklerin çıkarılmasında OpenSMILE uygulaması kullanılmıştır. Ön işleme sürecinde sıralama, standardizasyon, resample, discretize işlemleri uygulanmıştır. Öznitelik seçme sürecinde farklı algoritmalar ve uygulamalar incelenmiş. Performansa

ek bir katkı gözlemlenmediği için süreç içinde bu kısım kullanılmamıştır. Sınıflandırma sürecinde ise hem makine öğrenim hem de derin öğrenme algoritmaları kullanılmıştır. Makine öğrenim tarafında en başarılı tanıma oranı Lojistik Regresyon (%94,25) ile elde edilmiştir. Derin öğrenme tarafında ise DNN modeli (%92,57) ile yüksek tanıma oranına ulaşılmıştır.

Anahtar Kelimeler: Makine Öğrenmesi, OpenSMILE, Derin Öğrenme, Konuşmadan Duygu Tanıma

SPEECH EMOTION RECOGNITION IN A NEW TURKISH EMOTIONAL SPEECH DATASET

Emel ÇOLAKOĞLU

Kayseri University, Institute of Graduate Education

Master Thesis, July 2023

Supervisor: Assistant Professor Serhat HIZLISOY

ABSTRACT

The most natural and fast method of communication between people is talking. Apart from the content of speech, speech signals also contain many data such as the mood, age, gender, physiology and dialect of the person speaking. Emotion is the effect of an event, people or beings in the inner world of a person. Emotions can be divided into groups such as happiness, excitement, anger, calmness. Individuals convey their emotions to each other both through body language and through voice communication. However, in some cases, it is not possible to interpret or see body language. In such cases, the extraction of emotion from the sound becomes more important. In the first studies in this field, it was generally focused on the analysis of sentiment from image signals. In recent years, sentiment analysis studies over sound signal have started to develop. When the literature is examined, it is seen that there are many studies in the field of speech to emotion recognition. In the studies carried out, both machine learning and deep learning algorithms are used and highly successful recognition rates have been achieved.

Within the scope of this thesis, unlike previous studies, we created the dataset ourselves. For this purpose, Turkish series and films were examined. All records in the dataset consist of 5 seconds and contain 4 emotions. Here, the most important difference of the thesis from previous studies was created for emotion recognition without speaking; not to use ready-made data sets with high data quality and labeling. In addition, there are very few studies published on the Turkish language. Another point is that an independent model of the speaker has been created. In this case, it is actually a condition that can adversely affect the performance of the model. In the studies examined, it was generally not preferred due to low recognition rates.

The OpenSMILE implementation was used to extract attributes. In the pre-processing process, sorting, standardization, resample, discretize processes were applied. During the attribute selection process, different algorithms and implementations were examined.

Since no additional contribution to performance was observed, this part was not used in the process. In the classification process, both machine learning and deep learning algorithms were used. On the machine learning side, the most successful recognition rate was obtained by Logistic Regression (94.25%). On the deep learning side, a high recognition rate was achieved with the DNN model (92.57%).

Keywords: Machine Learning, OpenSMILE, Deep Learning, Speech Emotion Recognition, emo_large

İÇİNDEKİLER

YENİ TÜRKÇE DUYGUSAL VERİ SETİ ÜZERİNDE KONUŞMADAN DUYGU TANIMA

BİLİMSEL ETİĞE UYGUNLUK ..... **i**
TEZ İNTİHAL RAPORU TESLİM FORMU ..... **ii**
YÖNERGEYE UYGUNLUK ..... **iii**
KABUL VE ONAY TUTANAĞI ..... **iv**
TEŞEKKÜR ..... **v**
ÖZET ..... **vi**
ABSTRACT ..... **viii**
SİMGELER ve KISALTMALAR LİSTESİ ..... **xiii**
TABLOLAR LİSTESİ ..... **xvi**
ŞEKİLLER LİSTESİ ..... **xvii**
GİRİŞ **1**

BİRİNCİ BÖLÜM

TEMEL KAVRAMLAR

1.1. Konuşma 4
1.2. Duygu 4
1.3. Duygu Sınıflandırma 5
1.4. Ses Sinyali 6
1.5. Konuşmadan Duygu Tanıma 7
1.6. Konuşmadan Duygu Tanımada İzlenen Aşamalar 8
1.6.1. Veriseti Seçimi 8
1.6.2. Öznitelik Çıkarma 9
1.6.3. Ön İşleme Süreci 9
1.6.4. Öznitelik Seçme 10
1.6.5. Sınıflandırma 10
1.6.5.1. Makine Öğrenmesi Nedir? 11

1.6.5.2. Makine Öğrenim Sınıflandırıcı Türleri 11
1.6.5.3. Derin Öğrenme Nedir? 12
1.6.5.4. Model Performans Değerlendirme Ölçütleri 13

İKİNCİ BÖLÜM
LİTERATÜR TARAMASI

2.1. Sık Kullanılan Veri Setleri 15
2.2. Yöntem ve Modelleme Süreci 16

ÜÇÜNCÜ BÖLÜM
METODOLOJI

3.1. Sistem Mimarisi 23
3.2. Veriseti 26
3.3. Özniteliklerin Çıkarılması 27
3.3.1. Prozodik Öznitelikler 28
3.3.1.1. Perde 28
3.3.1.2. Enerji 28
3.3.1.3. Süre 28
3.3.2. Spektral Öznitelikler 29
3.3.2.1. MFCC 29
3.3.2.2. LPCC (Doğrusal Öngörü Kepstral Katsayıları) 30
3.3.2.3. Formantlar 30
3.3.3. Öznitelik Çıkarılmasında Kullanılan Uygulamalar 31
3.3.3.1. OpenSMILE 31
3.3.3.2. MIRtoolbox 35
3.3.3.3. JAudio 36
3.4. Ön İşleme Süreci 36
3.4.1. Sıralama 36
3.4.2. Resample 37
3.4.3. Standardizasyon 37

3.4.4. Discretize....38
3.5. Özniteliklerin Seçilmesi....38
3.6. Sınıflandırma....42
3.6.1. Makine Öğrenmesi....42
3.6.1.1. Destek Vektör Makineleri (DVM)....42
3.6.1.2. Lojistik Regresyon (LR)....43
3.6.1.3. K-en Yakın Komşu (KNN)....44
3.6.1.4. Karar Ağacı (KA)....45
3.6.1.5. Naive Bayes (NB)....45
3.6.1.6. Lineer Diskiriminant Analizi (LDA)....46
3.6.1.7. AdaBoosts (ADA)....46
3.6.1.7. Gradyan Artırma (GB)....46
3.6.1.8. Ekstra Ağaçlar (EA)....47
3.6.1.9. XGBoosts (XGB)....47
3.6.1.10. Rastgele Orman (RO)....47
3.6.2. Derin Öğrenme....48
3.6.2.1. Derin Sinir Ağları (DNN)....48
3.6.2.2. Evrişimli Sinir Ağları (CNN)....49
3.6.2.3. Tekrarlayan Sinir Ağları (RNN)....50
3.6.2.4. Uzun Kısa Süreli Bellek (LSTM)....51

DÖRDÜNCÜ BÖLÜM

DENEYSEL SONUÇLAR

4.1. Analiz Sonuçları ve Karşılaştırmalar....52
4.2. Tartışma....65
4.3. Sonuç....68

KAYNAKÇA....69
ÖZ GEÇMİŞ....77
TEZ İNTİHAL RAPORU....78

SİMGELER ve KISALTMALAR LİSTESİ

Simgeler

∑ **:** Sigma

Π **:** Pi

Kısaltmalar

AE **:** Otomatik Kodlayıcı (Autoencoder)

BLSTM **:** Çift Yönlü Uzun Kısa Dönem Bellek (Bidirectional Long Short-Term Memory)

ZC **:** Sıfır Geçişi (Zero Crossing)

CNN **:** Evrişimli Sinir Ağı (Convolutional Neural Network)

ÇKA **:** Çoklu Katman Algılayıcı

DNN **:** Derin Sinir Ağı (Deep Neural Network)

DVM **:** Destek Vektör Makineleri (Support Vector Machine)

EML **:** Ensemble Makine Öğrenimi

GTCC **:** Gammaton Kepstral Katsayıları (Gammatone Cepstral Coefficients)

HNR **:** Harmonik-Gürültü Oranı (Harmonic-to-Noise Ratio)

HSF **:** Yüksek Düzeyli Özellik (High-Level Feature)

IDS **:** İleri Doğru Seçim

KELM **:** Çekirdek Tabanlı Aşırı Öğrenme Makinesi (Kernel Extreme Learning Machine)

WA **:** Ağırlıklı Ortalama (Weighted Average)

KNN **:** K-En Yakın Komşuluk (K-Nearest Neighbor)

LFPC **:** Düşük Frekanslı Algısal Kepstral Katsayıları (Low-Frequency Perceptual Cepstral Coefficients)

LLD **:** Düşük Düzeyli Açıklayıcı (Low-Level Descriptor)

LPCC **:** Doğrusal Tahmin Kepstral Katsayıları (Linear Predictive Cepstral Coefficients)

ADA **:** Aykırı Değer Algılama

LSTM **:** Uzun Kısa Dönem Bellek (Long short-term memory)

MEDC **:** Ortalama Öklidyen Mesafe Sınıflandırıcısı (Mean Euclidean Distance Classifier)

ReLU **:** Doğrultulmuş Doğrusal Birim (Rectified Linear Unit)

MFCC **:** Mel-Frekans Kepstral Katsayıları (Mel-Frequency Cepstral Coefficients)

MFMC **:** Çok Çerçeve ve Çok Kanal (Multi-Frame and Multi-Channel)

MGDCC **:** Değiştirilmiş Grup Gecikmesi Kepstral Katsayıları (Modified Group Delay Cepstral Coefficients)

ML **:** Makine Öğrenimi (Machine Learning)

PCA **:** Temel Bileşenler Analizi (Principal Component Analysis)

RNN **:** Tekrarlayan Sinir Ağı (Reccurrent Neural Network)

SFFS **:** Sıralı Kayan İleri Seçim (Sequential Floating Forward Selection)

SMO **:** Sıralı Minimal Optimizasyon (Sequential Minimal Optimization)

SNN **:** İğnecikli Sinir Ağları (Spiking Neural Network)

TBA **:** Temel Bileşen Analizi

TCN **:** Zamansal Evrişimli Ağ (Temporal Convolutional Network)

TEO **:** Teager Enerji Operatörü (Teager Energy Operator)

UA **:** Ağırlıksız doğruluk (Unweighted accuracy)

ZCR **:** Sıfır Geçiş Oranı (Zero-Crossing Rate)

DL **:** Derin Öğrenme (Deep Learn)

LR **:** Lojistik Regresyon

LDA **:** Lineer Diskriminant Analizi

KA **:** Karar Ağacı

NB **:** Naive Bayes

EA **:** Ekstra Ağaç

RO **:** Rastgele Orman

GNB : Gauss NB

ADA **:** AdaBoosts

GA **:** Gradyan Artırma

XGB **:** XGBoosts

YSA **:** Yapay Sinir Ağları

TABLOLAR LİSTESİ

Tablo 1.1. *Karışıklık Matrisi* 13
Tablo 2.1. *EmoDB'de Duygu Sayıları* 15
Tablo 3.1. *391 Kayıtlık Veri Setinde Duygu Dağılımı* 26
Tablo 3.2. *862 Kayıtlık Veri Setinde Duygu Dağılımı* 27
Tablo 3.3. *1099 Kayıtlık Veri Setinde Duygu Dağılımı* 27
Tablo 3.4. *emo_large Öznitelikler* 32
Tablo 3.5. *64 Düşük Seviyeli Tanımlayıcı (IS13_Compare.Conf)* 33
Tablo 3.6. *Interspeech 2013 Öznitelik Seti Fonksiyonları* 33
Tablo 3.7. *60 Düşük Seviyeli Tanımlayıcı (IS11_speaker_state.conf)* 34
Tablo 3.8. *The InterSpeech 2011 Veri Setinde Kullanılan Fonksiyonlar* 34
Tablo 3.9. *Sınıflandırıcılarda Varsayılan Değerden Farkı Atanan Parametreler* 48
Tablo 4.1. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V1 Veriseti ve MFCC)* 52
Tablo 4.2. *Hiper Parametreler (DVM Sınıflandırıcısı İçin)* 53
Tablo 4.3. *K-kat Çapraz Doğrulama Parametreleri* 54
Tablo 4.4. *Sonuçlar ve Tüm Algoritmaların Karşılaştırılması (10 kat çapraz Doğrulama ile Karşılaştırma)* 55
Tablo 4.5. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V1 Veriseti)* 57
Tablo 4.6. *Lojistik Regresyon Parametreleri* 58
Tablo 4.7. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V2 Veriseti)* 60
Tablo 4.8. *DVM Parametreleri* 61
Tablo 4.9. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V3 Veriseti)* 63
Tablo 4.10. *ML ve DL Algoritmaları Sonuçların Karşılaştırılması* 65
Tablo 4.11. *Önceki Çalışmalar ile Karşılaştırmalar* 66

ŞEKİLLER LİSTESİ

Şekil 1.1. *Plutchik'in Duygu Çarkıfeleği* 6
Şekil 3.1. *İlk Modelin Mimari Yapısı* 23
Şekil 3.2. *İkinci Modelin Mimari Yapısı* 24
Şekil 3.3. *MFCC İşlem Adımları* 29
Şekil 3.4. *MIRToolbox ile Çıkarılabilecek Müzik Öznitelikleri* 35
Şekil 3.5. *Destek Vektör Makineleri Temel Yapısı* 43
Şekil 4.1. *Karışıklık Matrisi (DVM-MFCC Öznitelik)* 54
Şekil 4.2. *Karışıklık Matrisi (DVM-MFCC Öznitelik)(Yüzde Değeri)* 54
Şekil 4.3. *Analiz Sonuçları (10-kat Çapraz Doğrulama)* 55
Şekil 4.4. *Karışıklık Matrisi (LR-391Veriseti)* 59
Şekil 4.5. *Karışıklık Matrisi (LR-391Veriseti)(Normalize Edilmiş)* 59
Şekil 4.6. *Karışıklık Matrisi (LR-862Veriseti)* 62
Şekil 4.7. *Karışıklık Matrisi (LR-862Veriseti)(Normalize Edilmiş)* 62
Şekil 4.8. *Karışıklık Matrisi (LR-1099 Veriseti)* 64
Şekil 4.9. *Karışıklık Matrisi (LR-1099 Veriseti)(Normalize Edilmiş)* 64

GİRİŞ

İnsan iletişiminin en hızlı ve organik yollarından biri konuşmadır. Bu sebeple araştırmacılar, konuşma sinyalini kullanmaya yönelmiştir. Konuşma sinyali, konuşurken ortaya çıkan sesin içerisinde kişiye dair birçok önemli bilgi barındırır; cinsiyet, fizyolojik durum, lehçe, duygusal durum, yaş gibi unsurlar bunlardan sadece birkaçıdır. Konuşmadan duygu tanıma çalışmaları, konuşma sinyalinden anlam çıkarmaya odaklanır ve bu da çeşitli uygulama alanlarında fayda sağlayabilir.

Duygu, bireyin sevinçli, üzgün, öfkeli, heyecanlı ya da korkmuş olduğunu ifade eden bir terimdir. Tüm duygular, bir düşünceden kaynaklanır.

İnsan-bilgisayar etkileşiminin ve iş birliğinin en üst düzeyde gerçekleşmesi için önemli bir alan, konuşma seslerinden duyguların tanınmasıdır. Konuşma sinyalleri farklı frekansta ve karakteristikte özellikler içerdiğinden, bu verilerin analizi oldukça zorlayıcıdır. Bu nedenle, sınıflandırma algoritmaları ile konuşma verilerinden duyguları doğru şekilde çıkarmak oldukça karmaşık bir sorundur.

Literatürde, konuşmadan duygu tanıma sistemleri için çok sayıda çalışma yapılmıştır. Bu çalışmalar makine ve derin öğrenme mimarileri üzerine kurgulanmış ve başarılı tanıma oranlarına ulaşmışlardır.

Tez kapsamında önceki çalışmalardan farklı olarak kendi veri setimiz oluşturuldu. Bunun için Türk yapımı dizi ve filmler incelendi. Toplamda 3 farklı versiyon ile 391,862 ve 1099 kayıttan oluşan veriseti hazırlandı. Bu kayıtlar 5 saniyeden oluşmakta ve 4 duyguyu içermektedir. Ses analizinde kullanılan duygular mutluluk, üzüntü, kızgınlık ve nötr duygudur.

Ayrıca tezi diğer çalışmalardan ayıran önemli bir özelliği de konuşmacı bağımsız olmasıdır. Literatürde incelemiş olduğumuz çalışmalar genelde benzer veri setleri

üzerinde ve konuşmacı bağımlı yapıya sahipti. Özniteliklerin belirlenmesi, özniteliklerin seçimi ve sınıflandırma süreçlerinde çalışmalar farklılık göstermekteydi.

Temelde tez çalışmamız 4 ana kısımdan oluşmaktadır. Bu kısımlar özniteliklerin çıkarılması, ön işleme süreci, özniteliklerin seçimi ve sınıflandırma olarak tanımlanır.

Özniteliklerin etkili bir şekilde seçilmesi, sese özgü özniteliklerin başarılı bir şekilde belirlenmesi, genel performansı büyük ölçüde etkileyebilir. Ancak sesin özelliklerinin çıkarılması ve sınıflandırılmasında farklı yöntemler kullanılsa da elde edilen performans sonuçları veri setine, duygusal durumlara, dillere ve eğitim setinin özelliklerine bağlı olarak farklılık gösterebilir.

Sınıflandırma sürecinde ise farklı sınıflandırıcılar tekil (Hibrit bir model olmayacak) olarak kullanılıp performansları karşılaştırılmıştır.

Bu çalışma da insan sesinin analiz edilerek duyguların (4 duygu mutlu, üzgün, kızgın ve nötr) otomatik olarak tanımlanması amaçlandı. Bu amaç doğrultusunda dizi ve filmlerden 5 saniyelik kesitler alınıp veri seti oluşturulmuştur. Projede bu şekilde hazırladığımız kendi veri setimiz kullanıldı ve makine öğrenmesi sınıflandırıcıları ve derin öğrenme algoritmaları kullanılarak modeller oluşturuldu ve algoritma ile yöntemler birbirleri ile karşılaştırıldı.

Duygu tanıma konusu birçok şekilde yapılabilir. Bunlardan biri görüntü üzerinde mimikler üzerinden duygu tanıma işlemidir. Bunun için hem bir görüntü gerekir hem de sesin olmadığı bir yerde mimikler yetersiz kalacaktır.

Bir diğeri metinden duygu tanımadır. Yani konuşulan sesin metne dönüştürülmesi ve metnin analiz edilmesidir. Ancak imla kuralları uygulanmadığı zaman bir metin başka türlü de yorumlanabilir.

Bu noktada ses ve konuşma analizi büyük önem taşımaktadır. Özellikle çağrı merkezi gibi sadece sesin olduğu yerlerde hem görüntüden hem de metinden çok daha etkili bir sonuç elde edilebilir.

Ayrıca bu noktada Türkçe dili üzerinde duygu analizi yapan çok az sayıda çalışma bulunmaktadır.

Çalışma da konuşmacı bağımsız bir model oluşturulmuştur. Bu da aslında çalışma performansını zorlayacak bir durumdur. İncelenen çalışmalarda düşük tanıma oranları nedeniyle genelde tercih edilmemiştir.

Bu çalışma genel olarak bu boşluğu doldurabilecek mahiyettedir.

BİRİNCİ BÖLÜM

TEMEL KAVRAMLAR

1.1. Konuşma

İnsanlar arasındaki iletişimin en doğal ve hızlı yöntemi konuşmaktır. Konuşma sinyalleri konuşmanın içeriği haricinde konuşan kişinin ruh hali, yaşı, cinsiyeti, fizyolojisi ve lehçesi gibi birçok veriyi de barındırmaktadır [1]. Bu bilgileri ses sinyallerinden alabilmek adına birçok çalışma yapılmış ve yenileri de devam etmektedir. Konuşarak insanlar sadece birbirlerine söylemek istediklerini değil aynı zamanda hangi ruh hali içinde olduklarını da isteyerek ya da istemeden paylaşmış olurlar. Fiziksel olarak karşı karşıya olmasak bile, iletişim partnerimizin duygusal durumunu konuşma sinyalinden kolaylıkla algılayabiliriz [2].

Bir iletişimin temelini ses oluşturur. Konuşmanın özellikleri arasında tını, ton, vurgu, kesinti, duraklama, ses seviyesi, süre, tonlama ve melodi yer alır [2]. Konuşma sırasında oluşan sesler ise ses tellerinin titreşimi, artikülerlerin hareketi ve akciğerlerden havanın solunması ile ortaya çıkar. Sesler ağız, dudak ve dilin farklı pozisyonları ile oluşur. Oluşan bu ses sinyalleri analogdur. Bilgisayar sistemlerinin bu verileri anlayabilmesi için sinyalin dijitale dönüştürülmesi gerekir. Bu işlem ise bir ses sinyalinin kodlanması, bir saniye sürecinde birden çok kez örneklenmesi ve ses yüksekliğinin kaydedilmesi ile sağlanır [3].

1.2. Duygu

Duygular davranışlarımıza yön verir ve o doğrultuda hareket etmemizi sağlar. Duygular mutluluk, öfke, heyecan, üzüntü ve sakinlik gibi isimler altında gruplandırılır [4].

Duygu insanları diğer canlılardan ve makinelerden ayıran temel özelliklerden biridir. İnsanlar arasındaki iletişimde çok önemli bir yere sahiptir [5]. Hissedilen duygu aslında beynimizdeki bir düşünceden ortaya çıkan durumdur.

Bireyler duygularını hem beden dili yoluyla hem de sesli iletişim ile birbirlerine aktarırlar. Ancak bazı durumlarda beden dilini yorumlamak ya da görebilmek mümkün olmamaktadır. Böyle durumlarda sesten duygunun çıkarımı daha fazla önem kazanmaktadır [6].

Duygu tanıma sistemlerinde hem işitsel hem de görsel verilerden yararlanılır. Ayrıca bunlara ek olarak fizyolojik parametrelerinde (kalp atışı, kan basıncı vb.) sürece dâhil edilebilmesi mümkündür [4]. Ancak duyguların en net ifadesi yüzde görülen tepkilerdir. Bireyin yüzüne baktığımızda genellikle hangi duygu durumunda olduğunu çıkarabiliriz. Benzer biçimde, insan sesleri de duygusal durumları anlamak için kullanılabilir [2].

Duygu tanıma alanında ilk çalışmalar 80'li yıllarda başlamış ve 90'lı yıllara gelindiğinde oldukça hız kazanmıştır. İlk çalışmalarda sadece konuşmanın olumlu ya da olumsuz olmasına odaklanılmış ve analizler bu şekilde ilerlemişken sonraları teknolojik gelişmelere de paralel olarak çalışma alanı pek çok duyguyu ve stres derecesini de kapsar hale gelmiştir [5].

Duygu tanıma da kullanılan iki temel yöntem konuşma ya da yüz ifadelerinin analiz edilmesidir [7]. Son yıllarda her iki alanda da çok fazla çalışma yapılmıştır.

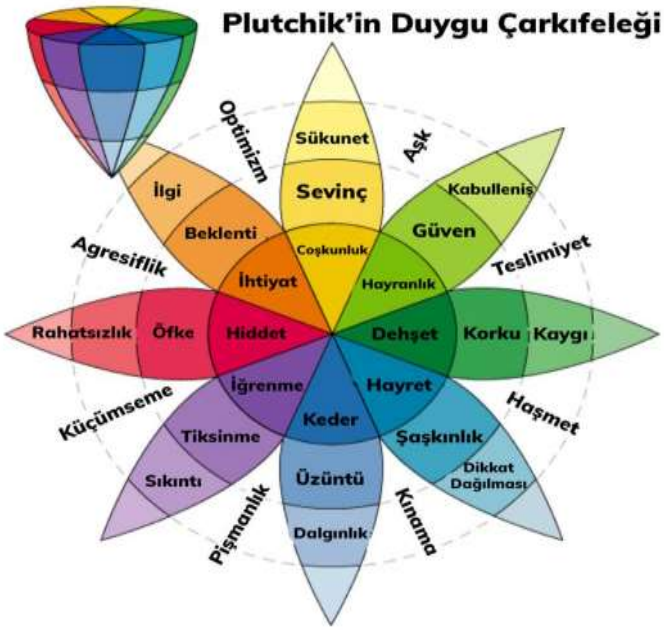
1.3. Duygu Sınıflandırma

İnsanlar seneler boyu duyguları anlamlandırmaya çalışmışlardır. Bunun sonucunda ise duygular çeşitli şekillerde kategorize edilmiştir.

Bugüne kadar birçok psikolog bu alanda çalışmış ve duygular ile ilgili çeşitli sistemler geliştirmiştir. Bunlar arasında Paul Ekman, William James, Richard ve Bernice Lazarus gibi isimler vardır. Albert Einstein Tıp Koleji ve Güney Florida Üniversitesi'nden Robert Plutchik ise 1980 yılında "Duygu çarkıfeleği" ismi verilen bir çalışma yapmıştır [8].

Plutchik e göre duygular temelde 8 kategoriye ayrılır: korku, şaşkınlık, tiksinme, öfke, sevinç, güven, üzüntü ve beklenti. Daha sonra ise bu 8 temel duyguyu 3'lü seviyelere ayırmıştır. Şekil 1.1'deki görselin her yaprağında tüm temel duyguların 3 ayrı seviyesi görülmektedir [8].

Şekil 1.1. *Plutchik'in Duygu Çarkıfeleği* [9]



Her bir yaprağın en dışındaki kısım bir duygunun en ılıman duygu seviyesini gösterir. Örneğin kabulleniş güvenin en ılıman halidir. Her yaprağın en ortasındaki duygular daha önce bahsedilen 8 temel duyguyu oluşturur. En iç kısım ise yaprağın bir üstündeki duygunun yoğun halidir. Örneğin hiddet öfkenin en yoğun halidir.

Yapraklar karşılıklı olarak da zıt duyguları ifade ederler. Örneğin ılıman bir duygu olan dikkat dağılmasının zıttı diğer ılıman duygu olarak ilgidir.

Ayrıca bu yaprakların arasında kalan duygularda o iki yaprağın birleşimi ile ifade edilir. Buna ikililer adı verilmiştir. Örneğin güven ve sevinç bir araya geldiğinde aşk oluşur [8].

Tabi ki temelde duyguları bu şekilde %100 olarak genellemek hatalı olacaktır. Ancak bu tarz çalışmaları duyguları kategorize etmek adına bir girişim olarak görmek, duyguların nasıl oluştuğu, birleşerek nasıl yeni duygular oluşturabildiği gibi konuları düşünmek ve yorumlayabilmek adına başlangıç noktaları olarak görebiliriz [8].

1.4. Ses Sinyali

Ses katı, sıvı ve gaz ortamlarda oluşan mekanik titreşimlerdir. Bu titreşimlerin ses olarak algılanabilmesi için kulak hassasiyeti, enerji ve yayılma ortamına sahip olması gerekir.

Ses bir dalgadır ve zaman ile boşluk içinde ilerler. Temelde 3 çeşit akustik sinyal vardır. Bunlar

- Periyodik sinyaller
- Rastgele sinyaller
- Darbe sinyaller

Periyodik sinyaller zaman içinde tekrar eden sinyallerdir ve süreklidir. Rastgele sinyaller beyaz veya pembe gürültü olarak isimlendirilen sinyallerdir. Periyodik değillerdir. Darbeler ise zamana bağlı olarak tekrarlı değillerdir ama şekilleri de bellidir [10].

Ses sinyali ise sesin elektriksel formudur. İnsan kulağı bunu doğrudan anlamlandıramaz. Ses sinyalleri analog ve dijital olmak üzere iki gruba ayrılır. Analog ses sinyali sesin bire bir kopyalanmış halidir. Dijital ses sinyali ise analog ses sinyalinden birim zamanda belirli örneklerin alınması ve bu örneklerin 0 ve 1'lerden oluşan dijital sisteme dönüştürülmesi ile elde edilen bir içerik olup sesin sayısal halidir. Sesin boyu, periyodu, hızı ve frekansı bulunmaktadır.

1.5. Konuşmadan Duygu Tanıma

İnsanlarda duygu analizinde kullanılan yaklaşımları genel olarak tekil ve çoklu kip olmak üzere iki grupta toplayabiliriz. Tekil yaklaşımda işlenen veri ses ya da görüntü sinyalidir. Çoklu yaklaşımda ise iki farklı kaynaktan gelen veri birleştirilir ve duygu tanıma başarısı artırılmış olur. Geçmiş yıllardaki ilk çalışmalar genellikle görüntü sinyallerinden duygu analizi üzerine olmuştur. Son yıllarda ise ses sinyali üzerinden duygu analizi çalışmaları hız kazanmaktadır [11].

Ses sinyalinde duygu analizi özellikle görsel bilginin alınamadığı ya da kısıtlı olduğu durumlarda daha büyük önem taşımaktadır [11].

Konuşmadan duygu tanıma sistemleri kişilerin konuşmalarına yansıyan duygu durumlarının ses sinyalinden tespit edilmesini amaçlamaktadır [7].

Son yıllarda konuşmadan duygu tanıma çalışmaları önem kazanmaya başlamıştır. Kullanım alanlarına bakıldığında ise sürüş yardımı, sağlık, eğlence, çağrı merkezleri, eğitim ve internet tabanlı öğrenme karşımıza en çok çıkan başlıklardır.

Konuşma sinyallerinden duygu tespiti aslında karmaşık ve zor bir süreçtir. Çünkü konuşma sinyalleri farklı frekans ve özellikleri barındırır. Bu da analizi zorlu kılan nedenlerden biridir [2]. Yine bu çalışmalarda başarıyı etkileyen diğer faktörlerde dil, lehçe, kullanılan verinin kalitesi, örnekleme sayısı, grup çeşitliği ve eşit ağırlıklı dağılım

olma durumu, sınıflandırılacak duygu sayısı, konuşmacı bağımlı ya da bağımsız analizlerdir.

Değişkenlik açısından daha düşük seviyede olan konuşmacı bağımlı tanıma sistemlerinde bile bir sözcüğün söyleniş şekli başarı oranını etkileyebilmektedir. Bu tez kapsamında ise konuşmacı bağımsız bir model geliştirilmeye çalışılmıştır. Bu durumda yaş, cinsiyet, konuşma tarzı, farklı anatomik özellikler ve ağız özellikleri gibi faktörler konuşmadan duygu tanıma da başarı sürecini etkileyecektir [12].

1.6. Konuşmadan Duygu Tanımada İzlenen Aşamalar

1.6.1. Veriseti Seçimi

Veri makine öğrenim sürecinin ana bileşenidir. Bu nedenle konuşmadan duygu tanıma çalışmalarında en önemli ve başlangıç kabul edilecek nokta veri setinin belirlenmesidir. Veri seti etiketlenmiş, güncel, elde edilmesi kolay, literatürde aynı veri seti kullanılmış çalışmaların bulunması (başarılı karşılaştırma yapabilmek adına), örnekleme sayısının yeterli olması, grup dağılımının eşit olması, örneklerin etiketlenmiş duyguyu tereddüt oluşturmadan tam olarak ve sinyal boyunca aynı duygu içeriğinde kalacak şekilde yansıtıyor olması çalışmanın verimi açısından önemli noktadır.

Makine öğrenim ve yapay zekâ modellerinde yüksek kaliteli veriler ile çalışmak modelin performansı açısından önemlidir. Çünkü modeller eğitildikleri veri kadar iyidir. Bu nedenle eğitim sürecinin ilk basamağı kaliteli, doğru, eksiksiz verileri kullanmaktır [13].

Makine öğrenim (ML) algoritmaları geçmiş veriler ile beslenir. Yani ML modelleri daha iyi sonuçlara ulaşabilmek adına geçmiş verilere ve eğitim veri setlerine ihtiyaç duyar. Mümkün olduğunca olası tüm ihtimalleri bünyesinde bulunduran eğitim verileri ile tahmin yürütülebilecek ML algoritmaları üretilir [13].

Modelin ürettiği sonuçların doğruluğunu ve performansını değerlendirmek için test verileri kullanılır. Bu aslında üretilen modelin tarafsız bir değerlendirmesidir. Ancak burada dikkat edilecek nokta eğitim ve test verilerini süreç başında ayrıştırmaktır. Çünkü testi eğitim verisi içinden yaparsak modelde yüksek doğruluk elde ederiz ancak modelin güvenirliği düşük olacaktır. O nedenle model tarafından görülmemiş veri seti üzerinden test sürecinin yürütülmesi modelin tarafsız değerlendirilmesini sağlar.

Bir sınıflandırma sürecinde gerekli veri miktarı kullanılacak algoritmaya ve probleme göre değişebilir. Genel geçer bir miktar tanımı o nedenle yoktur. Ancak genel olarak

bilinen bir nokta derin öğrenme (DL) algoritmalarının eğitim sürecinde büyük veri setlerine ihtiyaç duyduğu görülürken ML algoritmalarında daha az veriler ile eğitim süreci başarılı bir şekilde tamamlanabilir [13].

1.6.2. Öznitelik Çıkarma

Duyguları sınıflandırmak adına ses sinyallerinden çıkarılan özelliklere öznitelik denilmektedir. Konuşmadan duygu tanıma sistemleri oluşturulurken karşılaşılan önemli problemlerden biri de farklı duygu sınıflarını karakterize edebilen uygun özniteliklerin çıkarılmasıdır. Bu seçimin uygun yapılması sınıflandırma performansını ciddi şekilde etkilemektedir [14].

1.6.3. Ön İşleme Süreci

Büyük veri kümelerinde çoğunlukla kirli, hatalı ya da eksik veri noktaları bulunabilir. Veri kalitesi veri madenciliğinde en önemli konudur. Yapılan çalışmalarda güvenilirliği ve başarı oranlarını artırmak adına verinin ön işleme tabi tutulması gerekir. Temelde yapılacak işlemler yarı otomatiktir ve zaman alır [15].

Çok sayıda veri ön işleme teknikleri vardır. Bunlar ana başlık olarak 4 grupta toplanır [15].

- **Veri Temizleme:** Verideki tutarsızlıkların ve gürültünün temizlenmesi düzeltilmesidir.
- **Veri Birleştirme:** Farklı kaynaklardan gelen verilerin ortak bir veritabanında birleştirilmesidir.
- **Veri Dönüştürme:** Normalleştirme, Standardizasyon, Resample, Smote gibi dönüştürme yöntemleri kullanılarak yapılan işlemlerdir.
- **Veri İndirgeme: Fazla** ve gereksiz olan bazı değişkenlerin atılması, birleştirilmesi veya kümeleme yoluyla veri büyüklüğünün azaltılmasıdır. Discretize buna bir örnek olabilir.

Bu işlemler sınıflandırma sürecinden önce yapılarak verinin kalitesi ve güvenirliği artırılmış olur [15].

Kullanılan veri madenciliği programları genellikle ön işleme tekniklerinin birçoğunu gerçekleştirebilir. Ancak bu işlemi yapmak adına başka programlarda kullanılabilir.

Temelde veri ön işleme sürecinde kullanılan uygulamalar Python, R, Excel, SQL, KNIME, WEKA, OpenRefine, RapidMiner'dır.

1.6.4. Öznitelik Seçme

Öznitelik seçme konuşma sinyallerinden elde edilmiş öznitelikler içinden doğru duyguyu en iyi ifade edebilen özniteliklerin tespit edilmesi ve veri boyutunun buna göre azaltılması esasına dayanır [16].

Birçok çalışmada gözlemlenen sınıflandırılması istenen veri setlerinin her zaman gerekli ve verimli içeriklerden oluşmamasıdır. Bu durum da başarı oranını olumsuz etkilemektedir. Bu sorunu çözebilmek adına ise öznitelik seçme algoritmaları kullanılabilmektedir. Bu algoritmalar sınıflandırma sürecini hem iş yükü hem de başarı artışı açısından desteklemektedir [17].

Duygu tanıma ve duygu boyutlarının tahmini konusunda kullanılacak olan özniteliklerin belirlenmesi ve seçilecek sınıflandırma yönteminin sistemin başarısı üzerindeki etkisi büyük bir öneme sahiptir.

Kullanılacak olan sınıflandırıcının, sınıflandırılmak istenen veriyi yüksek başarımla ayırabilmesi için seçilecek öznitelikler sınıflar arasındaki farkları ortaya koyacak nitelikte olmalıdır.

Sınıflandırılmak istenen duygu kategorilerini en iyi şekilde ayırabilen özniteliklerin seçilmesi bunlara uygun bir sınıflandırıcı kullanılarak model karmaşıklığı azaltılabileceği gibi aynı zamanda sınıflandırma performansını da arttırılabilmektedir.

1.6.5. Sınıflandırma

Veri madenciliği kapsamında, sınıflandırma kavramı temel olarak bir veri kümesinde tanımlı farklı sınıflar arasında veriyi ayırmayı ifade eder. Sınıflandırma algoritmaları, eğitim verilerinden bu sınıf ayrımını öğrenirler ve daha sonra test verilerinde hangi sınıfa ait olduğunu doğru bir şekilde tahminlemeye çalışırlar.

Bu sınıfları tanımlayan değerlere etiket adı verilir ve hem eğitim sürecinde hem de test aşamasında verinin hangi sınıfa ait olduğunu belirlemek için önemli bir yere sahiptirler.

Sınıflandırma algoritmalarının çalışması iki aşamadan oluşur:

- Eğitim verisi üzerinden öğrenme
- Öğrenilen değerlerle test verisi üzerinden sınıflandırma

Konuşmadan duygu tanıma sistemlerinde çeşitli sınıflandırıcılar kullanılmaktadır. Bu sınıflandırıcılar, tek başına kullanılabileceği gibi, aynı zamanda birden fazla sınıflandırıcının bir araya gelerek oluşturduğu hibrit sınıflandırıcılar şeklinde de kullanılabilmektedir.

1.6.5.1. Makine Öğrenmesi Nedir?

İnsanlar için öğrenme yaşantıları boyunca kazandıkları tecrübelerden çıkarım yapabilme ve sonrasında bu çıkarımlara göre davranışlarını değiştirebilme sürecidir. Benzer şekilde makine öğrenmesi de makinelerin kendilerine verilen örnekleri ve varsa sonuçları incelemesi ve sonrasında bu işlemi nasıl yapacağını öğrenmesi ile yeni gelen durumlar hakkında da genelleme yapabilmesidir. ML örüntü tanıma, istatistik ve veri madenciliği gibi birçok alanla ilgili ve iç içe geçmiş durumdadır [11].

ML yöntemlerini aşağıdaki gibi sınıflandırabiliriz:

- Sınıflandırma
- Kümeleme
- Regresyon
- Öznitelik Seçimi/Çıkarımı
- Veride ilişkinin belirlenmesi

ML, çeşitli alanlarda aşağıdaki amaçlarla kullanılmaktadır: sanal kişisel asistanlar (örneğin Google Assistant gibi), bilgisayar oyunları, çevrimiçi müşteri destekleri için chatbotlar (BoTs), hisse senedi piyasası tahmini, video gözetim sistemi aracılığıyla suç tahmini, sosyal medya hizmetleri, tıbbi teşhis, örüntü tanıma, doğal dil işleme, veri madenciliği, trafik tahmini, çevrimiçi ulaşım ağları, ürün önerileri, çevrimiçi dolandırıcılık tahmini, robotik, tarım danışmanlığı, e-posta spam filtrelemesi [18].

1.6.5.2. ML Sınıflandırıcı Türleri

Öğrenme süreci, sistemin bilenen veya bilinmeyen veriye göre eğitilmesine ya da kendi kendine veri içinden sonuç çıkarmasına göre denetimli ve denetimsiz olmak üzere ikiye ayrılır [11].

Bu iki grup arasındaki temel fark ise denetimli sınıflandırma algoritmalarında veri setinin sonuçları bilinerek model oluşturulmakta iken; denetimsiz sınıflandırma algoritmalarında

veri setinin sonucu belli olmadan sınıflandırma yapılmasıdır. Genel olarak denetimli sınıflandırma algoritmalarında başarı oranı daha yüksektir [17].

Denetimli öğrenme de eğitim verileri etiketli olarak sisteme verilir ve öğrenim süreci başlatılır. Öğrenmeyi gerçekleştirmek adına bilgisayar en iyi sonucu elde edecek optimum fonksiyonu bulmayı amaçlar. Eğitim süreci tamamlandığında yine etiketli olan test verisi üzerinden model teste tabi tutulur. Test sonucunda doğru grupta belirlenmiş kayıtların sayısı toplam örneklem sayısına bölünerek doğruluk oranı elde edilmiş olur. Burada amaç ezberleme değil öğrenmedir. Denetimsiz öğrenmeye göre öğrenmesi daha hızlıdır. Hızlı olmasının nedeni ise hatanın sisteme öğrenme olarak girilmesi, sistemin yaptığı hatadan ders çıkarması ve bu dersleri öğrenme sürecine artı katkı sağlayacak şekilde kullanmasıdır [11].

Denetimsiz öğrenme de eğitim sürecinde etiketsiz veriler sisteme verilir. Etiketlenmemiş verideki bilinmeyen ilişkiyi bulma işlemidir. Verilerin arasında bilinen ama belirtilmemiş bağın çıkarılması işlemidir. Tabii veriler arasında ilişki olmayabilir bu da çıkarılır.

Sınıflandırma denetimli öğrenme yöntemlerindendir. Kümeleme ise denetimsiz öğrenme yöntemidir [11].

Konuşmadan duygu tanıma sürecinde yaygın kullanılan makine öğrenmesi algoritmaları Destek Vektör Makineleri (DVM), Lojistik Regresyon (LR), K-En Yakın Komşu (KNN), Lineer Diskriminant Analizi (LDA), Karar Ağacı (KA) ve Rastgele Orman (RO)'dır.

1.6.5.3. Derin Öğrenme Nedir?

Derin öğrenme (DL) algoritmalarının çoğu yapay sinir ağlarına dayanır. Klasik bir sinir ağında sınırlı sayıda (genelde iki ya da üç) gizli katman bulunurken derin öğrenme algoritmalarında yüzlerce gizli katman bulunmaktadır. Derin öğrenme algoritmalarının en büyük avantajı öznitelik çıkarma, seçme ve sınıflandırma aşamalarının tek seferde gerçekleşmesidir [19].

Sinir ağlarındaki katman sayısının artışı veri miktarının artması ve işlemi yürütecek bilgisayarların performanslarının artmasının bir sonucudur. Ayrıca yapay sinir ağlarından ayrışan kısımlarından biri de yeni tekniklerin geliştirilmiş olmasıdır. Bu yeni tekniklere; genelleştirme hatalarını düşürücü yöntemler, ağ ağırlıklarını ilklendirme yöntemleri, aktivasyon fonksiyonları, eniyileme algoritmaları gibi öğrenme sürecini iyileştiren teknikler örnek olabilir. Ayrıca üç işlem sürecinin (öznitelik çıkarma, seçme ve

sınıflandırma) birleştiğinin kabulü ise her katmanda öğrenilen özniteliklerin bir sonraki katmana girdi olarak alınmasından ileri gelir. Bu şekilde giriş katmanından çıkış katmanına kadar geçen süreçte özniteliklerin öğrenildiği bir yapı kurulmuş olur [12].

Konuşmadan duygu tanıma sürecinde yaygın olarak kullanılan derin öğrenme algoritmaları Derin Sinir Ağı (DNN), Evrişimli Sinir Ağı (CNN), Uzun-Kısa Süreli Bellek (LSTM) ve Özyinelemeli Sinir Ağı (RNN)'dir.

1.6.5.4. Model Performans Değerlendirme Ölçütleri

Sınıflandırma algoritmaları kullanılarak oluşturulan modellerin performansının değerlendirilmesinde bazı metrikler kullanılır. Buna genel olarak Karışıklık Matrisi denir. Karışıklık matrisi DL, ML ve istatistiksel sınıflandırmalarda sınıflandırıcının başarısını gösterir. Satırları gerçek değeri ve sütunları tahmin edilen değeri ifade eder [20].

Tablo 1.1. *Karışıklık Matrisi*

		Tahmin Edilen Sınıf	
	Sınıf	Pozitif	Negatif
Gerçek Sınıf	Pozitif	DP	YN
	Negatif	YP	DN

Tablo 1.1'e göre isimlendirmeleri yorumlarsak;

- DP (Doğru Pozitif): Gerçekte pozitif olup sınıflandırma sonucunda da pozitif olarak işaretlenmiş kayıtların sayısıdır.
- DN (Doğru Negatif): Gerçekte negatif olan ancak sınıflandırma sonucunda da negatif olarak belirlenen örneklerin sayısını ifade eder.
- YN (Yanlış Negatif): Gerçekte pozitif olan ancak sınıflandırma sonucunda negatif olarak işaretlenen örneklerin sayısını temsil eder.
- YP (Yanlış Pozitif): Gerçekte negatif olan ancak sınıflandırma sonucunda pozitif olarak işaretlenen örneklerin sayısını ifade eder.

Bu matriste üretilen değerler yardımıyla sınıflandırma performansını belirlemek adına kullanılan metrikler ise aşağıdaki şekildedir:

Doğruluk (Accuracy): Sınıflandırıcı tarafından doğru tahmin edilen veri sayısının tüm verisetindeki kayıt sayısına oranıdır.

$$Doğruluk = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

Kesinlik (Precision): Sınıflandırma işlemi sonucunda pozitif olarak tahmin edilenlerin ne oranda doğru olarak tahmin edildiğini gösterir.

$$Kesinlik = Pozitif\ Tahmin\ Değeri = \frac{DP}{DP+YP} \quad (2)$$

Hassasiyet (Recall): Sınıflandırıcının gerçekte pozitif sınıfa ait verileri doğru olarak tahmin etme oranını gösterir.

$$\text{Duyarlılık} = \text{Anma} = \text{DPO} = \frac{\text{DP}}{\text{DP+YN}} \quad (3)$$

F1-ölçüsü: Kesinlik ve Hassasiyet metriklerinin harmonik ortalamasıdır.

$$F1 - \text{ölçüsü} = 2 * \frac{Kesinlik*Hassasiyet}{Kesinlik+Hassasiyet} \quad (4)$$

İKİNCİ BÖLÜM

LİTERATÜR TARAMASI

2.1. Sık Kullanılan Veri Setleri

EmoDB veritabanı ücretsiz olarak kullanılması mümkün Almanca dilinde duygu içeren veri setidir. 5 Kadın ve 5 erkek profesyonel konuşmacı tarafından kayıtlar oluşturulmuştur. Toplam 535 örnek veri bulunmaktadır. Bu veriseti 7 duyguyu içermektedir. Bu duygular mutluluk, nötr, öfke, üzüntü, can sıkıntısı, kaygı ve iğrenmedir. Verisetindeki tüm veriler 48 kHz örnekleme hızına sahiptir. Sonrasında ise bu değer 16 kHz'e indirilmiştir. [21] . Son durumda bu ses dosyaları 16 khHz hızında, 16 bit boyutunda, yaklaşık 2-3 sn uzunluğunda ve 256 kbps bit hızında örneklenen konuşma sinyallerinden oluşmaktadır [19]. Veri setinde duygu bazında homojen bir dağılım söz konusu değildir. Duygu dağılımlarının detayı Tablo 2.1'de gösterilmektedir.

Tablo 2.1. *EmoDB'de Duygu Sayıları*
(K: Kızgın, N: Nötr, M: mutlu, Ü: Üzgün, S: Sıkkın, T: Tiksinti, Ko: Korku)

	K	**N**	**M**	**Ü**	**S**	**T**	**Ko**
Erkek	60	39	27	25	35	11	36
Kadın	67	40	44	37	46	35	33
Toplam=535	127	79	71	62	81	46	69

Cümleler anlamlı ve anlamsız 5 kısa ve 5 uzun olmak üzere 10 cümledir [19].

IEMOCAP veriseti oturum başına 2 konuşmacı tarafından oluşturulmuş 302 videodan oluşmaktadır. Toplamda 5 çift konuşmacı ile 5 oturumda kayıt işlemi gerçekleşmiştir. Veri seti 9 duyguyu içermektedir. Bunlar kızgın, heyecan, korku, üzgün, şaşırma, hüsrana uğrama, mutlu, hayal kırıklığı ve nötr dür.

RAVDESS veri setinde 1.440 konuşma ve 1.012 şarkı dosyası yer almaktadır. Bu veriseti Kuzey Amerika aksanıyla 2 farklı cümleyi seslendirme yapmış 24 profesyonel oyuncunun kayıtlarından oluşmaktadır. (12 erkek ve 12 kadın). Konuşma veriseti sakin,

mutlu, üzgün, kızgın, korkulu, şaşkınlık, tarafsız ve tiksinti duygularını ve şarkı veriseti de mutlu, üzgün, korkulu, kızgın ve sakin duygularını içermektedir [1].

SAVEE veri seti İngiltere'deki Surrey Üniversitesinde öğrenci olan dört erkek tarafından seslendirilen verilerden oluşmaktadır. Temelde 15 İngilizce cümleden oluşmaktadır. 7 farklı duyguyu (üzüntü, nötr, iğrenme, şaşkınlık, korku/kaygı, neşe ve öfke) içeren 480 kayıt bulunmaktadır. Bu verisetindeki tüm kayıtlar wav formatında, 44.1 kHz hızında, 16 bit boyutunda ve 705 kbps bit hızındadır. Nötr için 120 ve diğer duygular için ise 60 örnek bulunmaktadır [19]. Bu veri tabanı da duygu dağılımına göre dengesizdir.

TESS verisetinde 24 ve 64 yaşlarında ana dilleri İngilizce olan üniversite ve müzik eğitimi almış 2 kadın aktristen 200 farklı kelimeyi 7 farklı duygu ile ifade etmesi istenmiştir. Bu 7 duygu ise kızgın, iğrenme, şaşırma, üzgün, tarafsız, korku ve mutluluktur. Bu şekilde toplam 2.800 kayıt oluşturulmuştur [22].

2.2. Yöntem ve Modelleme Süreci

Akustik özellikler, büyüklük ve faz bilgilerini içerir. Bununla birlikte, geleneksel konuşmadan duygusu tanıma yöntemleri sadece büyüklük bilgisine odaklanır ve faz verilerini yok sayar ve böylece bazı bilgileri kaçırmış oluruz. Lili Guo ve diğerleri [23], çalışmalarında faz özelliklerinin doğru şekilde çıkarılmasını ve etkin kullanımını amaçlıyor. Kullanılan veri setleri EmoDB ve IEMOCAP'dır. Kullanılan öznitelikler Değiştirilmiş Grup Gecikmesi Kepstral Katsayıları (MGDCC), DRP, RP ve büyüklük spektrogramıdır. 2 model geliştirilmiştir. Tek kanallı model SCM ve çok kanallı model MCMA dir. ELM aşırı öğrenme makineleridir. (KELM), orijinal ELM'nin bir modifikasyonudur ve önceki çalışmalar, KELM'nin ELM'den daha iyi performansa sahip olduğunu göstermektedir. Bu çalışmada KELM, duyguları ayırt etmek için bir sınıflandırıcı olarak kullanılmaktadır. Büyüklük spektrogramlarından V1, MGDCC V2 ve DRP V3'ten büyüklük ve faz ile ilgili özellikleri çıkarmak için CNN kullanılmıştır. Hem EmoDB hem de IEMOCAP da en yüksek başarı oranları çok kanallı model MCMA model ile elde edilmiştir. Bu başarı oranında dâhil edilen özniteliklerde büyüklük, MGDCC, DRP dir. EmoDB için (WA %94,02 ve UA %93,66) ve IEMOCAP için (WA %57,58 ve UA %58,07)'dir.

Ziping Zhao ve diğerlerinin [24], çalışmasında, FAU-AEC ve IEMOCAP veri setleri kullanılmıştır. Kullanılan öznitelikler log-Mels, delta ve delta-delta olarak belirlenmiştir. Özellik çıkarma bölümünde, paralel olarak iki boyutlu SENet ile birleştirilmiş Conv.

kullanılarak bir model oluşturulmuştur. Ayrıca, uzun menzilli bağımlılıkları modellemek amacıyla Öz-Dikkat Genişletilmiş Artık Ağ (SADRN) kullanılmıştır. Duygusal etiketleri otomatik olarak hizalamak için Bağlantıcı Zamansal Sınıflandırma (CTC) katmanı içeren bir model de yer almaktadır. Çalışmada iki farklı model geliştirilmiştir. Bu iki model arasındaki temel fark, ilk modelde Genişletilmiş Artık Ağ katmanının (DRN) yerine ikincisinde Çift Yönlü Uzun Kısa Süreli Bellek (BLSTM) katmanının kullanılmasıdır. Sonuçlar incelendiğinde, IEMOCAP veri seti için en iyi sonuç PCNSE-SADRN-CTC modeliyle elde edilmiştir (%73,1). FAU-AEC veri seti için ise en iyi sonuç sırasıyla PCN-SADRN ve PCNSE-SADRN-CTC modelleriyle elde edilmiştir (%41,1).

N. Senthilkumar ve diğerlerinin [25], çalışmasında, IEMOCAP, Emo-DB ve RAVDESS veri setleri kullanılmıştır. Öznitelik olarak Spektrogram tercih edilmiştir. Öznitelik çıkarma sürecinde farklı tipte CNN'ler kullanılmıştır. Sınıflandırıcı olarak çok katmanlı BLSTM modeli kullanılmıştır. Elde edilen sonuçlar incelendiğinde, bu çalışmada geliştirilen modelin hem konuşmacı bağımsız hem de konuşmacı bağımlı durumlar için IEMOCAP, Emo-DB ve RAVDESS veri setlerinde en yüksek başarı oranlarını sağladığı görülmüştür. Ayrıca, veri setleri içinde Emo-DB (%86 konuşmacı bağımsız) ve Emo-DB (%85,5 konuşmacı bağımlı) en yüksek başarı oranlarına sahip veri setleri olarak belirlenmiştir.

Husam Ali Abdulmohsin ve diğerlerinin [26], çalışmalarında kullanılan veri setleri RAVDESS, Emo-DB ve SAVEE olarak belirtilmiştir. Bu çalışmada kullanılan öznitelikler Sıfır geçiş (ZC), entropi, enerji, Sıfır geçişin sapması, enerjinin sapması, Haar, Fourier fonksiyonu, MATLAB uygunluk fonksiyonu, ses yüksekliği fonksiyonu, saha fonksiyonu, zamana ve frekansa göre MFCC işlevi, zamana ve frekansa göre Gamaton Kepstrum Katsayısı (GTCC) ve harmonik orandır. Öznitelik seçimi f-puanı özellik seçimi ile yapılmıştır. Sınıflandırıcı olarak sinir ağı tercih edilmiştir. Veriseti bazında elde edilen en iyi sonuçlar: %86,1 (RAVDESS), %96,3 (Emo-DB) ve %91,7 (SAVEE).

Stavros Ntalampiras [27], çalışmasında, Log-mel spektrogramı ve zamansal modülasyon özellikleri gibi özniteliklerin kullanıldığı belirtilmektedir. Çalışmada Emo-DB veri seti kullanılmış ve sınıflandırma işlemi için Siyam Sinir Ağı (SNN) kullanılmıştır. Bu çalışmada, 9 konuşmacı ile gerçekleştirilen deneyler sonucunda en yüksek tanıma oranının %82,1 olduğu tespit edilmiştir.

Serhat Hızlısoy ve Zekeriya Tüfekci [28] çalışmalarında, Türkçe duygusal veri setini kullanmışlardır. Öznitelikler OpenSMILE, MIRtoolbox ve JAudio araçları kullanılarak elde edilmiştir. Öznitelik seçimi aşamasında korelasyon tabanlı öznitelik seçme yöntemi kullanılmıştır. Sınıflandırma işlemi için ise WEKA uygulaması tercih edilmiştir. Sınıflandırma da sıralı minimal optimizasyon (SMO), LR, KA ve Bayes ağlarından sonuçlar elde edilmiş ve karşılaştırılmıştır. 10 kat çapraz doğrulama tercih edilmiştir. En iyi sonuç MIRtoolbox+Jaudio+OpenSMILE+Bayes kombinasyonu ile %95,96 olarak elde edilmiştir. Öznitelik seçimi olmadan SMO ile en iyi sonuç elde edilmişken, öznitelik seçimi sonrası Bayes ağları en yüksek başarıyı elde etmiştir.

Zengwei Yao ve diğerleri [29], çalışmalarında, IEMOCAP veri setini kullanmışlardır. Bu çalışmada dört farklı duygu (kızgın, mutlu, normal ve üzgün) için tanıma süreci planlanmıştır. Çalışmada hibrit bir yapı tercih edilmiştir. Bu hibrit yapıda farklı girdi türlerine sahip DNN, CNN ve RNN kullanılmıştır. Bu farklı girdi türleri ile modeller ayrı ayrı eğitilmiş ve sonuca ulaşmak için karar düzeyinde füzyon stratejisi kullanılmıştır. Ses özniteliklerini elde etmek için librosa kullanılmıştır ve Düşük Düzeyli Açıklayıcı (LLD) ve Yüksek Düzeyli Özellik (HSF) özniteliklerini çıkarmak için ise openSMILE kullanılmıştır. Çalışmanın sonucunda, önerilen füzyon yöntemi ile en iyi sonuç elde edilmiş ve %57,1 WA (Ağırlıklı Doğruluk) ve %58,3 UA (Ağırlıksız Doğruluk) başarı oranlarına ulaşılmıştır.

Shadi Langari ve diğerleri [30], Kesirli Fourier dönüşümü kullanılarak uyarlanabilir bir zaman-frekans öznitelik çıkarımına dayalı bir model tasarlamıştır. Öznitelik seçiminde Cuckoo Arama ve Genetik Algoritma beraber kullanılmıştır. Kullanılan sınıflandırıcı ise 2D CNN dir. Veri seti olarak ise PDREC, SAVEE, EMO-DB kullanılmıştır. Veriseti bazında en iyi tanıma oranları şu şekilde elde edilmiştir: SAVEE (%80, %80), PDREC (%91,46, %87,31), EMO-DB (%97,57, %97,21) (doğruluk oranları WA, UA).

Ning Jia ve Chunjun Zheng [31], çalışmalarında IEMOCAP ve özel bir verisetini kullanmışlardır. Mel-Frekans Kepstral Katsayıları (MFCC) çalışma da tercih edilen özniteliktir. Sınıflandırıcı ise BLSTM modeldir. Her iki veri setinde de beşli çapraz doğrulama yapılmıştır. Veri seti %80-20 şeklinde eğitim ve test olarak ayrılmıştır. Elde edilen en yüksek tanıma oranları: (%78 (SER), %70 (IEMOCAP)).

J. Ancilin ve A. Milton [32], çalışmalarında veri seti olarak Berlin, Ravdess, SAVEE, EMOVO, eNTERFACE ve Urdu kullanılmıştır. Kullanılan öznitelikler Mel frekansı

kepstral katsayısı, Mel frekans büyüklük katsayısı, log frekansı güç katsayısı ve lineer tahmin kepstal katsayısıdır. Bu yazıda MFCC'nin çıkarım süreci değiştirilmiş ve duyguları MFCC ve log frekans güç katsayısı dâhil olmak üzere geleneksel spektral özelliklerden daha iyi bir doğrulukla tanıyan Mel frekans büyüklük katsayısını (MFMC) çıkarmışlardır. Sınıflandırıcı olarak da DVM kullanılmıştır. Sınıflandırma sürecinde 10 kat çapraz doğrulama yapılmıştır. MFMC özelliği altı verisetinde de duyguları tanıma da diğer özniteliklerden daha iyi performans göstermiştir. Aynı zamanda MFCC, LFPC ve LPCC'den daha yüksek tanıma oranlarıyla veri setlerinin bireysel duygularını da tanımıştır. En iyi sonuçlar ise: SAVEE (%75,63), Berlin (%81,50), Ravdess (%64,31), eNTERFACE (%56,41), EMOVO (%73,30), Urdu (%95,25).

Süha Gökalp ve İlhan Aydın [1], RAVDESS ve Toronto Duygusal Konuşma Seti (TESS) çalışmada veri seti olarak yer almaktadır. Kullanılan öznitelikler MFCC ve Spektogramdır. Sınıflandırıcılar ise DVM, KA ve otomatik kodlayıcıdır (AE). Yapay Sinir Ağları (YSA) olarak de AlexNet (türü CNN), MobileNet (türü CNN), Squeezenet (türü DNN) ve Resnet50 (türü CNN) kullanılmıştır. En iyi sonuç ise %98 (AlexNet+Oto Kodlayıcı+TESS) olarak elde edilmiştir.

Turgut Özseven [7], yılında yaptığı çalışmada, EmoDB veri setini kullanmıştır. Bu çalışmada 149 farklı öznitelik kullanılmıştır. Bu özniteliklerin 45'i prozodik özelliklerden oluşurken, 104'ü spektral özelliklerden oluşmaktadır. MFCC öznitelikleri openSMILE aracılığıyla elde edilirken, diğer öznitelikler Praat kullanılarak elde edilmiştir. Öznitelik seçim aşamasında aykırı değer algılama (ADA), ileri doğru seçim (IDS) ve temel bileşen analizi (TBA) gibi yöntemler tercih edilmiştir. Sınıflandırıcı olarak DVM, ÇKA (Çoklu Katman Algılayıcı) ve KNN yöntemleri kullanılmıştır. En iyi sonuç %90,3 (z-puan normalizasyon+ÇKA+Yüksek geçiren filtre+ADA) olarak elde edilmiştir.

Anjali Bhavan ve diğerlerinin [33], çalışmasında, MFCC, Delta ve Delta–Delta MFCCs, Spektral Merkezler gibi öznitelikler kullanılmıştır. Bu çalışmada RAVDESS, IITKGP-SEHSC ve EmoDB veri setleri kullanılmıştır. Ayrıca Boruta yöntemi ile öznitelik seçimi gerçekleştirilmiştir. Sınıflandırıcı olarak ise DVM kullanılmıştır. Modeldeki torbalama topluluğu, 20 adet DVM modelini içermektedir. Eğitim ve test verilerinin dağılımı 90:10 şeklinde belirlenmiştir. Çalışma kapsamında veri üzerinde 10 kat çapraz doğrulama yapılmıştır. En iyi sonuç, EmoDB veri setinde elde edilmiştir (%92,45).

Kun-Yi Huang ve diğerlerinin [34],çalışmasında kullanılan öznitelikler sıfır geçiş oranı, kök kare ortalama, temel frekans, Harmonik-Gürültü-Oranı ve MFCC dir. Veri seti olarak Multimedya İnsan Makine İletişimi duygu veriseti (MHMC) ve CHI-MEI duygu durum bozukluğu veriseti kullanılmıştır. CNN ve LSTM ise sınıflandırma sürecinde yer almıştır. En iyi sonuç ise HSC + EP + LSTM kombinasyonu ile elde edilmiştir (%75,56).

Dias Issa ve diğerleri [35] ,çalışmalarında, Mel-frekans Kepstrum Katsayıları, Mel ölçekli spektrogram, Kromogram, Spektral kontrast özelliği ve Tonnetz sunumu gibi öznitelikler kullanılmıştır. Veri setleri olarak RAVDESS, EMO-DB ve IEMOCAP tercih edilmiştir. Öznitelik çıkarımında librosa ses kitaplığı kullanılmıştır. Sınıflandırmada ise 1D CNN (1 Boyutlu Evrişimli Sinir Ağı) yöntemi kullanılmıştır. Çalışma boyunca farklı analizler gerçekleştirilmiş ve bu analizlere farklı isimler verilmiştir. "Model B" olarak adlandırılan analiz, en yüksek başarıya sahip olmuş ve EmoDB veri setinde %96,34 başarı oranına ulaşılmıştır.

Hadhami Aouani ve Yassine Ben Ayed [36] çalışmasında, 39 MFCC, Sıfır geçiş oranı (ZCR), Teager Enerji operatörü (TEO) ve Harmonik gürültü oranı (HNR) gibi öznitelikler kullanılmıştır. Öznitelik seçimi için AE yöntemi kullanılmıştır. Sınıflandırma aşamasında ise DVM tercih edilmiştir. Kullanılan veri seti RML duygu veritabanıdır. En iyi sonuç, AE ile elde edilmiş ve %74,07 başarı oranına ulaşılmıştır. Ayrıca öznitelik seçimi işlemi sonrasında performansta artış gözlemlenmiştir.

Kunxia Wang ve diğerleri [37], çalışmalarında, öznitelik olarak MFCC ve WPC kullanılmıştır. EMODB ve EESDB veri setleri üzerinde yapılan analizlerde, konuşmacıdan bağımsız olarak duygu tanımanın amaçlandığı bir yaklaşım benimsenmiştir. Bu amaç doğrultusunda, dalgacık paket analizi kullanılarak duyguların özellikleri çıkarılmıştır. Öznitelik seçimi için Sıralı Kayan İleri Arama (SFFS) yöntemi tercih edilmiştir. Sınıflandırma aşamasında ise DVM kullanılmıştır. En yüksek başarı oranı ise WPC+SFFS+RDVM (%79,5) kombinasyonu ile EMODB veri setinde elde edilmiştir.

Yousef Pourebrahim ve diğerleri [38], çalışmalarında etiketli ve etiketsiz veriler birlikte kullanılmıştır. Bu veri üzerinde konuşmadan duygu tanıyabilmek için AE dayalı yarı denetimli bir öğrenme modeli oluşturulmaya çalışılmıştır. Kullanılan veri setleri IEMOCAP, FAUAEC, PESD, EmoDB ve RAVDESS'dir. Önerilen yöntem ile %61,6 başarı elde etmiştir.

Ziping Zhao ve diğerlerinin [39], çalışmasında öznitelik olarak log-Mel spektrogram, delta ve delta-delta kullanılmıştır. IEMOCAP veri seti olarak çalışma da yer almaktadır. Modeller 100 devir için eğitilmiştir. Sınıflandırıcı olarak 5 katmanlı 1D evrişimli modüllere sahip TCN ve kod çözücü için LSTM kullanılmıştır. Karşılaştırma için de BLSTM ağı tercih edilmiştir. En iyi sonuç test setinde WA (%65,0) ve UA (%66,1) ve eğitim setinde WA (%68,6) ve UA'nın (%69,5)'tir. Dikkat temelli TCN model ile bu sonuçlar elde edilmiştir.

Orhan Atila ve Abdulkadir Şengür [40], tarafından gerçekleştirilen çalışmada, öznitelik olarak MFCC, koleagram ve fraktal boyut kullanılmıştır. Sınıflandırma işlemi ise 3D CNN-LSTM modeli ile gerçekleştirilmiştir. Kullanılan veri setleri ise RAVDESS, SAVEE ve RML'dir. Sinyalden görüntüye dönüştürme işleminde Python kullanılmış, DL model ise MATLAB ile geliştirilmiştir. Çalışmada 10 kat çapraz doğrulama yöntemi uygulanmıştır. Her bir veri seti için elde edilen başarı oranları sırasıyla RAVDESS (%96,18), SAVEE (%87,5) ve RML (%93,2) olarak belirlenmiştir.

Cevahir Parlak ve Banu Diri [5] , tarafından yürütülen çalışmada, EmoSTAR ve EmoDB veri setleri kullanılmıştır. Öznitelik çıkarımı için OpenSMILE tarafından sunulan Emobase ve Emo_large öznitelik setleri kullanılmıştır. Öznitelik seçimi ve sınıflandırma aşamaları ise Weka aracıyla gerçekleştirilmiştir. Kullanılan sınıflandırıcılar ise Naive Bayes (NB) ve Sıralı Minimal Optimizasyon (SMO) olarak belirlenmiştir. Elde edilen en iyi sonuçlar ise EmoSTAR veri seti için InfoG+Rank (6.755) (%97,2) ve Chi+Rank (6.755) (%97,20) olarak elde edilmiştir.

Cevahir Parlak ve Banu Diri [41], tarafından yürütülen çalışmada, televizyon kanallarından elde edilen konuşma örnekleri ile Berlin EmoDB veri setleri kullanılmıştır. Bu çalışma, konuşmadan bağımsız bir modeli ele almaktadır. Çalışmada, f0'un (ses frekansı) konuşma sinyali boyunca gösterdiği değişim ve sessiz duraklamalar kullanılmıştır. Bu faktörler duygu tanıma sürecinde kullanılmıştır. En iyi sonuçlar, EmoDB veri seti için %82 (kızgın) ve %76 (nötr) olarak elde edilmiştir.

Yixiong Pan ve diğerleri [42] tarafından yürütülen çalışmada, Berlin duygu veritabanı (5 farklı duygu sınıfı içeren ancak 3'ü kullanılmış) ve SJTU Çince duygu veritabanı (kendileri tarafından oluşturulmuş) veri setleri kullanılmıştır. Özellik çıkarma aşamasında enerji, perde, formant, LPCC, MFCC ve MEDC gibi özellikler elde edilmiş ve bu özelliklerin farklı kombinasyonları kullanılarak duygu tanıma başarısı karşılaştırılmıştır.

Sınıflandırma işleminde DVM yöntemi kullanılmıştır. En iyi sonuç, MFCC+MEDC+Enerji kombinasyonu ile SJTU veritabanında %95,08 olarak elde edilmiştir.

Gökhan Polat ve Halis Altun [43], tarafından yapılan çalışmada, Berlin Veri Seti kullanılmıştır. Öznitelik çıkarımı aşamasından sonra 38 öznitelikten oluşan vektörler elde edilmiştir. Sınıflandırma işlemi için çok katmanlı bir Yapay Sinir Ağı (YSA) kullanılmıştır. En iyi sonuç, 9 gizli katmanlı model ile %81,65 olarak elde edilmiştir.

Haytham M. Fayek ve diğerleri [44], tarafından yapılan çalışmada, IEMOCAP veri seti kullanılmıştır. Bu çalışmada, ileri beslemeli sinir ağları için Rektifiye Doğrusal Birim (ReLU) aktivasyon fonksiyonu kullanılmıştır. Önerilen model, derin ve çok katmanlı bir sinir ağı yapısına sahiptir. Kullanılan sınıflandırıcılar arasında DNN, LSTM-RNN ve CNN bulunmaktadır. En iyi sonuç, CNN+FC+Softmax kombinasyonu ile elde edilmiş ve %64,78 olarak belirlenmiştir.

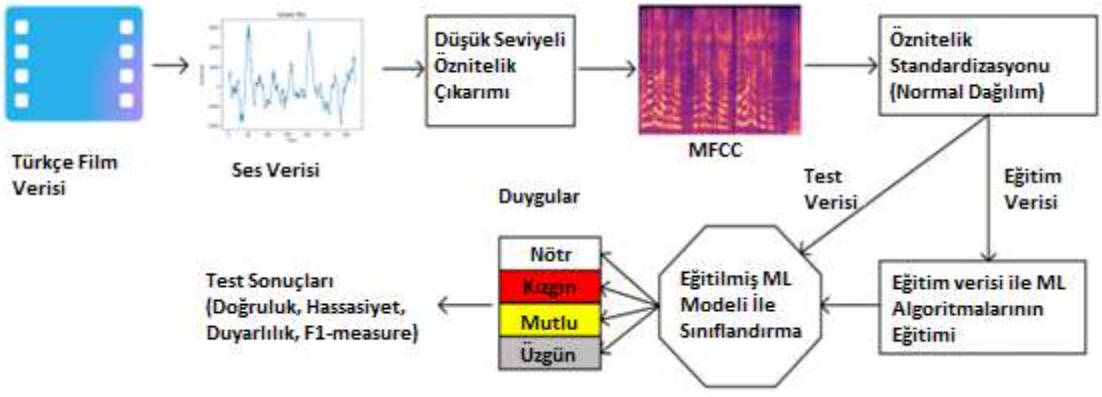
ÜÇÜNCÜ BÖLÜM

METODOLOJI

3.1. Sistem Mimarisi

Çalışma kapsamında 3 farklı modelden bahsedebiliriz. Bunlardan ilki makale kapsamındaki çalışma olan 391 kayıttan oluşan verisetinden MFCC özniteliklerin çıkarılması, sonrasında bu veriseti üzerinde standardizasyon uygulanması, verinin test ve eğitim olarak %90-10 şeklinde bölümlenmesi, eğitim ve test sürecinin tamamlanması, sonrasında 4 duygu için başarı oranlarının belirlenmesidir. Genel hatlarıyla bahsedilen bu mimari Şekil 3.1'de görsel olarak gösterilmiştir.

Şekil 3.1. *İlk Modelin Mimari Yapısı*



Bu model kapsamında veri seti hazırlanırken 58 tekil Türkçe film ve dizi incelenmiştir. Buradan 5 saniyelik 391 ses kaydı oluşturulmuştur. Ayrıca performansını incelemek adına farklı 10 saniyelik 403 kayıt da oluşturulmuştur. Bu kayıtlar alınırken Camtasia Recorder uygulaması kullanılmıştır. Ayrıca kayıt sonrası ses dosyalarında kesme ve düzeltme işlemleri de Camtasia Studio uygulaması ile yapılmıştır. Hazırlanan verisetinde tüm kayıtlar wav formatında, tek kanallı ve örnekleme frekansı 22.050'dir. Bu şekilde veri seti oluşturulmuştur. Çalışmanın diğer adımında ise öznitelik çıkarma işlemi

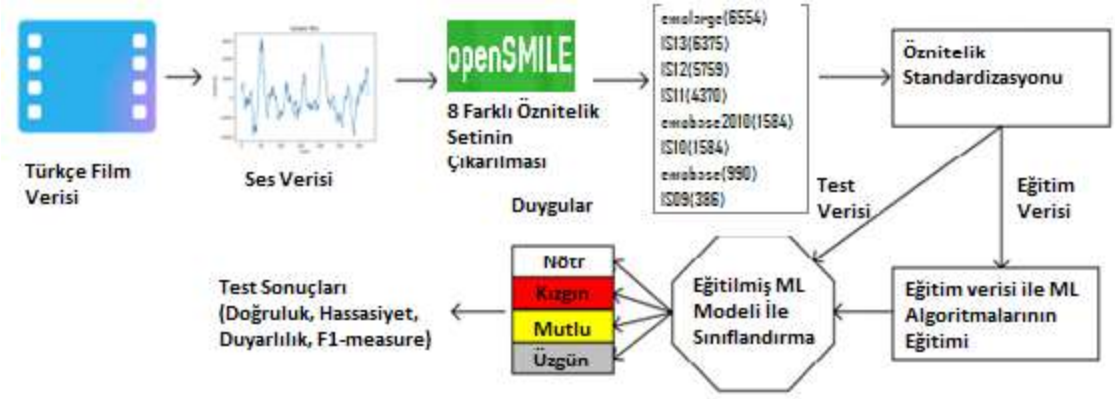
uygulanmıştır. Bu işlem sonrasında MFCC öznitelikleri verilerden ayıklanmıştır. Daha sonra ise öznitelikler üzerine normal dağılımlı standardizasyon uygulanmıştır.

Bu modelde ML algoritmalarının hazırladığımız veri seti üzerinde sınıflandırma sürecindeki başarısı incelenmiştir. Veri seti %90-10 eğitim ve test verisi olarak ayrılmıştır. Eğitim verisi farklı sınıflandırma algoritmaları ile eğitilmiş ve sonrasında yine aynı algoritmaları ile test süreci tamamlanmıştır.

Sınıflandırma işlemi sırasında Gradyan Arttırma (GA), DVM, LR, XGBoost (XGB), KNN, KA, Gauss NB (GNB), Ekstra Ağaç (EA), LDA, Ada Boost (ADA), ve RO algoritmaları test edildi ve bu algoritmalar arasından en iyi sonuçlar DVM ile sağlanmıştır. Sınıf bazlı incelendiğinde ise duygu tanıma işleminde en iyi tanıma oranı öfkeli ve nötr duygularda görüldü, ancak mutluluk duygusunda en düşük tanıma başarısı elde edildi.

İkinci modelde ise farklı veriseti ve farklı öznitelik kümeleri analiz sürecine dâhil edilmiştir. Mimari yapı Şekil 3.2'de detaylı olarak gösterilmektedir.

Şekil 3.2. *İkinci Modelin Mimari Yapısı*



Şekil 3.2'ye göre mimari yapı incelendiğinde öncelikli olarak ilk hazırlanan 391 kayıtlık veriseti ve sonrasında hazırlanan 862 ve 1.099 kayıtlık veri setleri üzerinde makine öğrenim algoritmaları ile sınıflandırma işleminin sonuçları analiz edilmiştir. Bu üç veri setinin de ortak özelliği 5 saniyelik kayıtlardan oluşması, tek kanallı ve 22.050 örnekleme frekansına sahip olmasıdır. İlk analizden farklı olarak burada öznitelik çıkarma sürecinde OpenSMILE uygulaması kullanılmıştır. OpenSMILE ile gelen hazır öznitelik kümelerinden emolarge, IS13_ComParE, IS12_speaker, IS11_speaker, emobase2010, IS10_paraling, emobase ve IS09_emotion setleri analiz edilmiştir. Bu şekilde 3 veriseti

için toplam 8 farklı öznitelik kümesi oluşturulmuştur. Yani analiz sürecinde hepsini sonucunu alıp karşılaştırdığımız 24 farklı öznitelik seti hazırlanmış oldu. Elde edilen öznitelik setleri üzerinde ise 3 farklı filtreleme işlemi uygulanmış ve ayrı ayrı sonuçları elde edilmiştir. Yapılan işlemler sıralama-standardizasyon; sıralama-standardizasyon-resample; sıralama-standardizasyon-discretize şeklindedir. Sıralama ve resample filtreme işlemleri weka uygulaması ile yapılmıştır. Diğer filtreler ise Python tarafında yapılan geliştirme sürecine dâhildir.

Eğitim ve test sürecinde ise öznitelik setleri %90-10, %80-20 ve 10 kat çapraz doğrulama şeklinde bölümlere ayrıldı. Her bir grubun eğitim sürecinde 8 farklı sınıflandırıcı ile eğitim süreci tamamlandı. Yine 8 farklı sınıflandırıcı ile de test edildi. Tez kapsamında ise bu 8 öznitelik setinden en iyi sonuç alınan 4 öznitelik setinin sonuçları paylaşılmıştır.

Veri seti olarak ise 391 kayıtlık veriseti V1, 862 kayıtlık veriseti V2 ve 1.099 kayıtlık veriseti V3 olarak tanımlandı ve tezin kalan kısmında bu şekilde kullanılmıştır.

Özetleyecek olursak; V1 verisetinde en yüksek başarı oranı %91,14'tür. V2 verisetinde ise en yüksek başarı oranı %94,25 olarak elde edildi. V3 verisetinde en yüksek başarı oranı %92,63'tür.

Her iki modelin de geliştirme sürecinde Python dili kullanılmıştır. Son yıllarda, Python makine öğrenimi analizlerinde sıkça tercih edilen bir dil haline gelmiştir. Veri analizi, sınıflandırma, veri görselleştirme, doğal dil işleme ve DL gibi analizlerde Python oldukça zengin bir kütüphaneye sahiptir. Çalışma kapsamında makine öğrenim sürecinde Python da kullandığımız kütüphaneler Pandas, OS, Matplotlib, Sklearn, Numpy, Imblearn ve Random dır. Bunların içinde ML için temel olan kütüphane sklearn'dır. Bu kütüphane birçok ML algoritması içermektedir. Dokümantasyon tarafı da oldukça detaylıdır. Ayrıca elde edilen sonuçları grafiksel olarak görüntüleyebilmek adına da Matplotlib kullanılmıştır.

Üçüncü modelde ise DL algoritmaları sınıflandırma sürecinde analiz edilmiş ve elde edilen sonuçlar incelenmiştir. ML sürecinde elde edilen sonuçlar incelendiğinde genel olarak en iyi performansın emo_large ile elde edildiğini gözlemlendi. DL algoritmaları analiz sürecinde daha uzun eğitim ve test süreçlerine gereksinim duyması nedeniyle veriseti burada tekilleştirildi ve emo_large öznitelik seti kullanıldı. Ayrıca hem veriseti üzerinden hem ses dosyaları üzerinden hem de ses dosyalarından elde edilmiş spektogram resimleri üzerinden sonuçlar alındı ve karşılaştırıldı. Hazırlanan tüm verisetlerine

standardizasyon uygulandı. Sınıflandırıcı olarak da DNN, CNN (1D), RNN, LSTM, CNN-LSTM kullanıldı.

3.2. Veriseti

Bu tez kapsamında kullanılan veri seti bizim tarafımızdan hazırlanmıştır. Bu veri setini hazırlarken Türkçe dizi ve filmler incelenmiştir. Bu inceleme sonucunda 4 farklı duygu sınıfına göre (mutlu, kızgın, üzgün ve nötr) kayıtlar alınmıştır. Kayıt işlemi Camtasia Recorder uygulaması ile yapılmıştır. Ayrıca kayıt sonrası tüm düzenleme, kesme, konuşma sinyallerinden görüntü dosyalarının ayrılması, gürültü temizliği, sesin yükseltilmesi, boşlukların temizlenmesi gibi işlemler de Camtasia Studio ile yapılmıştır. Veri setindeki tüm kayıtlar tek kanallı, wav formatında ve 22.050 örnekleme frekansına sahiptir.

Kayıtlar hazırlanırken görüntü ve ses dosyaları beraber incelenerek duygu etiketlendirmesi yapılmıştır. Ancak görüntü dosyalarından etkilenme durumunu engellemek adına sonrasında ses dosyaları tekrar farklı kontrol kişileri tarafından dinlenmiş ve etiketleri kontrol edilmiştir. Bu şekilde verisetinin tarafsızlığı ve güvenilirliği yükseltilmiştir. Bu süreç içinde ise uygun görülmeyen tüm kayıtlarda da verisetinden elenmiş ve varsa etiket hataları da çoklu kontrol ile düzeltilmiştir.

İlk analizde kullanılan V1 veri seti için 58 tekil film, dizi serisi incelenmiştir. Bu sayı içinde ilgili dizinin bölümleri tek olarak sayılmıştır. 391 kayıtlık veri setinde (V1) duygu dağılımı Tablo 3.1'de gösterilmektedir.

Tablo 3.1. *391 Kayıtlık Veri Setinde Duygu Dağılımı*

Kızgın	Mutlu	Üzgün	Nötr
100	93	103	95

V2 ve V3 isimli veri setlerini hazırlarken 68 tekil film, dizi serisi incelenmiştir. Bu sayı kapsamında aynı dizinin farklı bölümleri ise yine tek sayılmıştır. Her bir örneğin uzunluğu sabit ve 5 saniyedir.

862 (V2) ve 1.099 (V3) kayıtlık veri setlerinde duygu dağılımları Tablo 3.2 ve Tablo 3.3'te gösterilmektedir.

Tablo 3.2. *862 Kayıtlık Veri Setinde Duygu Dağılımı*

Kızgın	Mutlu	Üzgün	Nötr
220	209	220	213

Tablo 3.3. *1.099 Kayıtlık Veri Setinde Duygu Dağılımı*

Kızgın	Mutlu	Üzgün	Nötr
276	274	278	272

Ayrıca 10 saniyelik 403 kayıtlık bir veriseti daha oluşturuldu. 5 saniyelik ve 10 saniyelik verisetlerinde elde edilen sonuçlar karşılaştırıldığında tanıma oranında anlamlı bir fark elde edilmedi. O nedenle ilk analizden sonra 10 saniyelik veri üzerinde çalışmaya devam edilmemiştir.

Veri setini hazırlarken izlenen dizi ve filmlerde müzik olmamasına, sesin kalitesine, aynı anda tüm kayıt boyunca tek konuşmacının olmasına, kayıt boyunca sadece tek bir duygu yansıtacak konuşmaların seçilmesine, konuşmacı bağımsız yapıya destek olması adına farklı aktör ve aktrislerin kayıtlarının kullanılmasına, konuşma metninin farklılığına, boşlukların olmaması ya da duyguyu algılamayı engellemeyecek kadar az olmasına dikkat edildi. Bunu sağlayabilmek adına birçok dizi ve film izlenip elendi.

Çalışmayı literatürdeki çalışmalardan farklılaştıran en önemli özelliklerinden biri de konuşmacı bağımsız bir model geliştirilmiş olmasıdır. Çünkü literatür incelemesinde genelde konuşmadan duygu tanıma için hazırlanmış özel veri setleri tercih edilmekte ve başarı oranları da oldukça yüksektir. Bu tür çalışmalarda, konuşmacıdan bağımsız modellerin başarı için bir dezavantaj olarak kabul edilmesi yaygın bir durumdur. Bu çalışma ise konuşmacı bağımsız olmasına rağmen yüksek tanıma oranlarına ulaşmıştır.

3.3. Özniteliklerin Çıkarılması

Konuşma sinyallerinden özniteliklerin elde edilmesinde hazır araçlar kullanılabildiği gibi araştırmacıların geliştirdiği uygulamalar da kullanılabiliyor. Normalde bir ses sinyalinden çıkarılabilecek öznitelikler 4 grupta toplanabilir. Bunlar akustik öznitelikler, bağlamsal bilgiler, dilsel öznitelikler ve bu 3 grubun karması olarak oluşan hibrit özniteliklerdir.

Ama genel olarak konuşmadan duygu tanıma çalışmaları incelendiğinde çalışmaların çoğunda akustik özniteliklerin tercih edildiği görülmektedir. Akustik öznitelikler de kendi içinde 2 gruba ayrılır. Bunlar prozodik ve spektral özniteliklerdir [7].

3.3.1. Prozodik Öznitelikler

Konuşma temelde anlamlı seslerin sıralı olarak bir araya getirilmesi ile oluşur. Normalde arka arkaya gelen bu seslerin konuşma olabilmesi için doğal da olması gerekir. Bu noktada bir konuşmayı doğal yapan özelliklerin tümüne prozodi denir. Perde değişimi konuşmaya melodik özellik ekler. Perdeme değişimi ile tonlama oluşur. Hece ve fonem boyutundaki ses birimleri ise konuşmaya ritmik özellikler katabilmek adına uzatılıp kısaltılabilir. Ayrıca bir konuşma içerisinde belirli kelime ya da heceler diğerlerine göre daha vurgulu söylenip belirginleştirilebilir. Tonlama, vurgu, ritim ve sıra gibi prozodik öznitelikler verilmek istenen mesajın anlaşılırlığını artırır. Bu özniteliklerin hepsi algısal düzeyde ipuçlarıdır. Fiziksel olarak ise prozodik özniteliklerin ana bileşenleri perde, süre ve enerjidir [45].

3.3.1.1. Perde

Bir sesin işitme sisteminde uyandırdığı tizlik/peslik duygusuna perde (pitch) adı verilir. Bir duygunun ölçüsü ses kaynağının titreşim frekansıdır. Bir sesin frekansı arttıkça perdesi yükselir (tizleşir). Sesin frekansı azaldıkça ise perdesi düşer (pesleşir). Ancak frekans ile perde arasındaki ilişki doğrusal değildir. Bundan dolayı da MEL adı verilen yeni bir birim ortaya çıkmıştır. MEL birimi perdedeki eşit artışlara karşılık gelir ve bu artışları frekans ile ilişkilendirir [10].

3.3.1.2. Enerji

Enerji, konuşmanın ses yoğunluğu veya gücünü ifade eder. Konuşmacının sesinin ne kadar yüksek veya düşük bir seviyede olduğunu kontrol etmesini sağlar. Enerji, iletişimin duygusal tonunu ve vurgusunu etkileyebilir. Örneğin yüksek enerjili bir konuşma, coşku, öfke veya dikkat çekme amacıyla kullanılabilirken; düşük enerjili bir konuşma daha sakinlik veya ciddiyeti ifade etmek için kullanılabilir.

3.3.1.3. Süre

Süre, sesin ne kadar uzun veya kısa bir süre boyunca sürdüğünü ifade eder. Konuşma sırasında her bir ses veya hece belirli bir süre boyunca devam eder. Süre, bir kelimenin veya bir cümlenin içindeki seslerin süresini ve vurgusunu belirleyebilir. Örneğin, bir kelime veya hece daha uzun süreyle söylendikçe, vurgu daha belirgin hale gelebilir. Aynı şekilde, bazı durumlarda hızlı veya kısa süreli bir konuşma, heyecanı veya aceleyi ifade etmek için kullanılabilir.

3.3.2. Spektral Öznitelikler

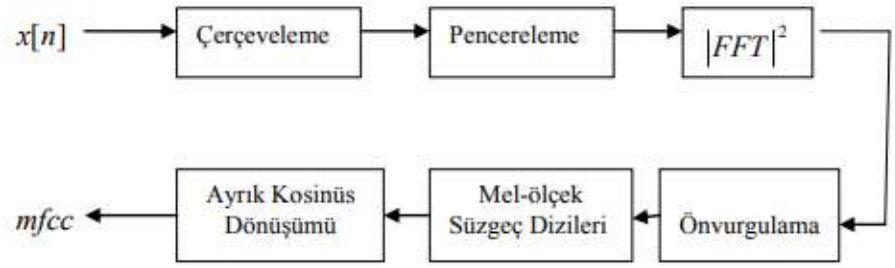
Spektral öznitelikler, ses sinyallerinin frekans içeriğini temsil etmek ve analiz etmek için kullanılan özelliklerdir. Ses sinyallerinin spektral içeriği, sesin frekans bileşenlerinin dağılımını ifade eder. Sesin frekans bileşenleri, sesin yüksek veya düşük tonlarda nasıl duyulduğunu belirler.

Spektral öznitelikler, ses sinyallerinin spektral içeriğini bir dizi matematiksel işleme tabi tutarak elde edilir. Bu öznitelikler, genellikle ses işleme, konuşma tanıma, konuşma sentezi ve diğer ses analizi uygulamalarında kullanılır.

3.3.2.1. MFCC

MFCC ses sinyallerinde öznitelik elde edilirken ve özellikle de sesten duygu tanıma çalışmalarında en sık kullanılan spektral özniteliklerin başında gelir. MFCC'de insan kulağı modellenir ve bu nedenle duygu tanıma çalışmalarında başarı oranı yüksektir. MFCC'deki işlem adımları Şekil 3.3'te gösterilmektedir.

Şekil 3.3. *MFCC İşlem Adımları*



Çerçeveleme: Konuşma işaretleri küçük parçalara ayrılmalıdır. Çünkü tüm işaret boyunca FFT hesaplanırsa, farklı fenomenlere ait spektral bilgilerin tutulmasında kayıplar oluştur. Burada ilgili çerçevenin FFT si hesaplanır [46]. Çerçeve uzunluğu 20-30 ms arasında olmalıdır.

Pencereleme: Her bir çerçevenin başında ve sonundaki süreksizliği önlemek için yapılan işleme denir [47]. Yaygın kullanılan pencere yapısı Hamming'dir. Fonksiyonu ise Formül 5'teki şekildedir.

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{(N-1)}\right] \qquad N-1 \geq n \geq 0 \tag{5}$$

Hızlı Fourier Dönüşümü (FFT): N örneğe sahip her çerçevenin zaman bölgesinden frekans bölgesine çevrilmesi için FFT işlemi uygulanır. Fonksiyonu Formül 6'da açıklanmıştır [47].

$$X_n = \sum_{k=0}^{N-1} x_k e^{\frac{-2\pi jkn}{N}} \qquad ,n = 0,1,2,\dots N-1 \tag{6}$$

Ön vurgulama: Ünlü seslerde ses yolunda toplam -6 dB/oktav'lık zayıflama oluşur. Ünlü seslerde bu zayıflamayı gidermek için genellikle birinci dereceden yüksek geçiren filtre kullanılır. Ünsüz seslerde ise ön vurgulamaya ihtiyaç yoktur. Çünkü spektrumları düzgündür [46].

Mel-ölçek Süzgeç Dizileri: Bu basamakta daha önce spektrumu alınan işaret frekans ölçeklerinden (Bark, Mel, Doğrusal, ERB) birine göre düzenlenmiş üçgen süzgeç dizilerinden geçirilip elde edilen işaretin logaritması alınır [48].

Ayrık Kosinüs Dönüşümü: Burada ise en son işarete ayrık kosinüs dönüşümü uygulanarak kepstrum katsayıları olarak bilinen öznitelik vektörleri elde edilir [48].

3.3.2.2. LPCC (Doğrusal Öngörü Kepstral Katsayıları)

LPCC (Doğrusal Öngörü Kepstral Katsayıları) yöntemi de spektral özniteliklerden biridir. LPCC, ses sinyallerinin frekans içeriğini temsil etmek ve analiz etmek için kullanılan bir özellik çıkarma yöntemidir. LPCC, öncelikle doğrusal öngörü analizini kullanarak ses sinyalinin spektral içeriğini modeller.

LPCC'nin temel amacı, bir ses sinyalini kapsayan aralığı hesaplamak için lineer öngörü yöntemini kullanmaktır. Bu yöntemde, ses sinyali parçalara bölünür ve her bir parça üzerinde doğrusal öngörü analizi uygulanır. Sonuç olarak, LPCC, ses sinyalinin frekans bileşenlerini ve spektral özelliklerini temsil etmek için elde edilen katsayıları kullanır.

LPCC, özellikle ses tanıma, konuşma sentezi ve diğer ses işleme uygulamalarında kullanılan bir özellik çıkarma yöntemidir. Bu öznitelikler, sesin frekans bileşenlerinin dağılımını temsil ederek, ses sinyallerinin spektral içeriğini analiz etmek ve tanımlamak için kullanılır.

3.3.2.3. Formantlar

Ses yolu rezonansının bir ses kaynağıyla etkileşiminin bir sonucudur. Formant frekansları sınırsız sayıda tanımlanmıştır. Ancak bunların içinde ilk üçü yani F1, F2 ve F3 değerleri konuşmacı tanıma çalışmalarında yaygın olarak kullanılmaktadır. Daha yüksek

formantlar ise aynı konuşma içeriğini aynı frekansa çıkaran konuşmacıları ayırt eden tını algısında kullanılır. Formant frekansı titreşimli ve titreşimsiz seslere göre değişiklik gösterir [19].

Formantlar, özellikle insan sesi ve konuşma analiziyle ilişkilendirilir. Konuşma sırasında, konuşmacının vokal kanalı (boğaz, ağız ve burun) rezonans yapar ve belirli frekanslarda yoğunlaşma gösteren formantlar oluşur. Bu formantlar, konuşmacının ürettiği sesin karakteristiklerini, sesin yükseklik, vokal ayırımı ve diğer akustik özelliklerini belirler.

3.3.3. Öznitelik Çıkarılmasında Kullanılan Uygulamalar

3.3.3.1. OpenSMILE

OpenSMILE sinyal işleme ve ML için modüler ve esnek bir öznitelik çıkarım uygulamasıdır. Ana odak noktası, ses sinyallerinden özniteliklerin elde edilmesi üzerine yoğunlaşmaktadır. Ancak fiziksel sinyaller, görsel sinyaller ve diğer fiziksel sensörlerden elde edilen sinyallerin analiz edilmesinde de kullanılabilir. Tamamen C++ da yazılmış hızlı, verimli ve esnek bir uygulamadır. Windows, Linux ve MacOS gibi çeşitli sistemlerde çalışabilir [16].

OpenSMILE, veri madenciliği ve ML alanında yaygın olarak kullanılan çeşitli veri biçimlerini okuma ve yazmayı destekleyerek birlikte çalışabilirliği kolaylaştırır. Bu biçimler arasında ses dosyaları için PCM WAVE, metin tabanlı veri dosyaları için CSV (Virgülle Ayrılmış Değer, elektronik tablo biçimi) ve ARFF (Weka Veri Madenciliği), HTK (Hidden-Markov Toolkit) parametre dosyaları ve basit bir ikili kayan matris biçimi bulunur [16].

OpenSMILE, güçlü bir C API'yi içerir ve diğer uygulamalarla sorunsuz entegrasyon sağlar. Android, Python, iOS ve .NET platformları için de kullanılabilir. Bu API, ses veya özellik verilerini gerçek zamanlı olarak openSMILE'a iletmeyi içeren tüm komut satırı işlevlerine programlı erişim imkânı sunar [16].

Tez kapsamında özniteliklerin çıkarılmasında hem Python hem de OpenSMILE kullanılmıştır. İlk modeldeki analizde kullanılan MFCC öznitelik Python ile gelen kütüphaneler yardımıyla elde edilmiştir. İkinci modeldeki tüm öznitelik kümeleri için ise OpenSMILE kullanılmıştır.

İkinci modelde 3 farklı veri seti için 8 farklı öznitelik grubu oluşturulmuştur. Böylece analiz sürecinde kullanılan 24 farklı veri seti bulunmaktadır. OpenSMILE uygulaması

Python arayüzü üzerinden entegre edildi ve bu şekilde kullanıldı. OpenSMILE ile gelen ve çalışma sürecinde kullanılan öznitelik setleri Interspeech 2009 (384), Interspeech 2010 (1.582), Interspeech 2011 (4.368), Interspeech 2012 (5.757), Interspeech 2013 (6.373), emobase (990), emobase2010 (1.582) ve emo_large (6.552)'dir.

Burada bahsedilen 8 farklı öznitelik grubundan en iyi sonuç alarak belirlediğimiz 4 öznitelik grubunun öznitelik detay dağılımları aşağıda listelenmiştir:

a) Emobase_large: Bu veri öznitelik setinde 6.552 öznitelik yer almaktadır. Konfigürasyon dosyası emo_large.conf'dır [49]. Tablo 3.4'te özniteliklerin dağılımı yer almaktadır.

Tablo 3.4. *Emo_large Öznitelikler*

Prozodik
Log Enerji
Seslendirme Olasılığı
Sıfır Geçiş Oranı
F0
Spektral
MFCC (0-12)
Spektral Maks ve Min
Mel Spektrum (0-25)
Spektral Akı
Spektral yuvarlanma noktası (50%)
Spektral yuvarlanma noktası (75%)
Spektral yuvarlanma noktası (90%)
Spektral Merkez

b) The InterSpeech 2013 ComParE: Konfigürasyon dosyası IS13_ComParE.conf'dır. Toplam 6.373 öznitelik içerir. 64 düşük seviye tanımlayıcının (LLD) listesi Tablo 3.5'te gösterilmiştir [50].

Tablo 3.5. *64 Düşük Seviyeli Tanımlayıcı (IS13_Compare.Conf)*

4 Enerjiyle İlgili LLD
İşitsel spektrumun toplamı (ses yüksekliği)
RASTA tarzı filtrelenmiş işitsel spektrumun toplamı
RMS enerjisi, sıfır geçiş oranı
55 Spektral LLD
RASTA tarzı işitsel spektrum, 1-26(0-8 kHz) bantları
MFCC 1-14
Spektral enerji 250-650 Hz, 1 k-4 kHz
Spektral kayma noktası 0,25, 0,50, 0,75, 0,90
Spektral akı, ağırlık merkezi, entropi, eğim
Psikoakustik keskinlik, uyum
Spektral varyans, çarpıklık, basıklık
6 seslendirmeyle ilgili LLD
F0 (SHS ve vitrerbi yumuşatma)
Olasılık (Sesin)
Log. HNR, Jitter (yerel, delta), Parıltı (yerel)
Logaritmik HNR, Titreşim (yerel, delta), Parıltı (yerel)

Bu veri setinde kullanılan fonksiyonlar ise Tablo 3.6'da yer almaktadır.

Tablo 3.6. *Interspeech 2013 Öznitelik Seti Fonksiyonları*

LLD/Δ LLDD'ye Uygulanan İşlevler
Çeyrekler 1-3, 3 çeyrekler arası aralık
%1 Yüzdelik dilim (≈min), %99 yüzdelik dilim (≈maks)
Yüzdelik aralığı %1-99
Min/maks konumu, aralık (maks – min)
Aritmetik ortalama, ikinci dereceden kök ortalama
Kontur merkezi, düzlük
Standart sapma, çarpıklık, basıklık
İlgili süre LLD, %25/50/75/90 aralığının üzerinde
LLD nin arttığı ilgili süre
LLD nin pozitif eğriliği olan ilgili süre
Doğrusal tahmin kazancı (LP), LP katsayıları 1-5
Ortalama, Maksimum, minimum, Segment uzunluğunun SD si
Yalnızca LLD'ye Uygulanan İşlevler
Zirvelerin ortalama değeri
Zirvelerin ortalama değeri- aritmetik ortalama
Zirveler arası mesafelerin ortalama/sd'si
Zirvelerin, minimumların genlik ortalaması
Pik genlik aralığı
Yükselen/düşen eğimlerin ortalaması/SD'si
Doğrusal regresyon eğimi, ofset, ikinci dereceden hata
İkinci dereceden regresyon a, b, ofset, ikinci dereceden hata
Sıfır olmayan karelerin yüzdesi3

c) The InterSpeech 2011: Bu veri setinin konfigürasyon dosyası IS11_speaker_state.conf'dur. Toplam 4.368 öznitelik içerir. 60 adet düşük seviyeli tanımlayıcının listesi Tablo 3.7'de gösterilmiştir [51].

Tablo 3.7. *60 Düşük Seviyeli Tanımlayıcı (IS11_speaker_state.conf)*

4 enerji ile ilgili LLD
RMS Enerji
RASTA tarzı filtrelenmiş işitsel spektrumun toplamı
İşitsel spektrumun toplamı (ses yüksekliği)
Sıfır Geçiş Oranı
50 spektral LLD
MFCC 1-12
Spektral Kayma Noktası 0,25, 0,50, 0,75, 0,90
Spektral enerji 25-650 Hz, 1 k-4 kHz
RASTA tarzı filtre. İşitsel spektrum, bantlar 1-26(0-8 kHz)
Spektral Akı, Entropi, Varyans, Çarpıklık, Basıklık, Eğim
5 sesle ilgili LLD
Titreşim (yerel, delta)
F0, Seslendirme olasılığı
Parıltı (yerel)

Bu veri setinde kullanılan fonksiyonlar ise Tablo 3.8'de yer almaktadır.

Tablo 3.8. *InterSpeech 2011 Veri Setinde Kullanılan Fonksiyonlar*

33 temel fonksiyon
Çeyrekler 1-3
3 çeyrekler arası aralık
%1 yüzdelik dilim (≈ min), %99 yüzdelik dilim (≈maks)
İkinci dereceden regresyon a ve b ve ikinci dereceden hata
Doğrusal regresyon eğimi ve ikinci dereceden hata
Çarpıklık, basıklık
Tepe mesafelerinin ortalaması
Tepe mesafelerinin standart sapması
Zirvelerin ortalama değeri
Doğrusal tahmin kazancı (LP)
Zirvelerin ortalama değeri- aritmetik ortalama
Yüzdelik aralığı %1- %99
Kontur merkezi
Süre sinyali %25 aralığının altında
Süre sinyali %90 aralığının üzerinde
Yükselen/düşen süre sinyali
LP Katsayıları 1-5
6 F0 işlevi
Ortalama, maksimum, minimum, std. dev. Segment uzunluğu
Sıfır olmayan karelerin yüzdesi
Saniye cinsinden giriş süresi

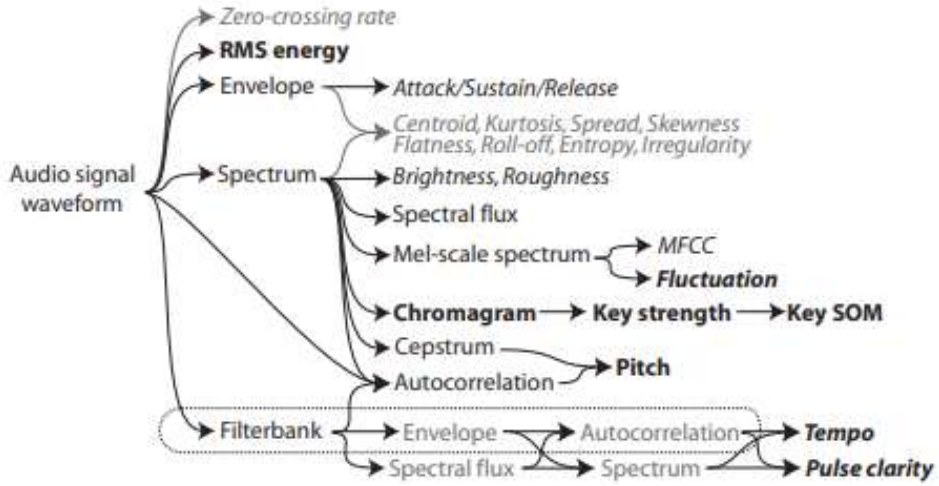
d) Emobase2010: Konfigürasyon dosyası emobase2010.conf'dır. Toplamda 1.582 öznitelik içerir. 34 düşük seviyeli tanımlayıcı (LLD) ve bu 34 tanımlayıcının delta katsayısını içerir. Bu 68 LLD değeri için ise 21 adet fonksiyon kullanılmış ve hesaplanan her değer yeni öznitelik oluşturmuştur (1.428 öznitelik). Son olarak da 4 perde tabanlı LLD ye ve bunların deltasına 19 fonksiyon hesaplaması yapılmıştır (152). En son olarak da 2 yeni öznitelik eklenmiştir (Perde başlangıçlarının sayısı ve toplam giriş süresi). Bu öznitelik seti, INTERSPEECH 2010 Paralinguistic Challenge setinden ayıran nokta, minPos ve maxPos değerlerinin normalleştirilmiş olmasıdır.

3.3.3.2. MIRtoolbox

MIRtoolbox sesten müzikal özelliklerin çıkarılması için kullanılan bir yazılımdır. MATLAB dilinde yazılmıştır ve açık kaynaklıdır. MATLAB kullanıcıları tarafından kullanılabilir.

Parametrik yapısı ve modüler kullanımı sayesinde esnek bir yapıya sahiptir. Bu sayede ses dosyalarından elde edilen özelliklerin başka bir özellik çıkarımı için kullanılmasına imkân sağlar. MIRToolbox ile çıkarılabilecek müzik öznitelikleri Şekil 3.4'te gösterilmiştir [52].

Şekil 3.4. *MIRToolbox ile Çıkarılabilecek Müzik Öznitelikleri* [52]



Bu araç kutusunun en büyük avantajı MATLAB ile entegre olması ve çok sayıda yüksek seviyeli müzikal öznitelik çıkarım araçları bulundurmasıdır [53].

3.3.3.3. JAudio

JAudio açık kaynak kodlu bir kütüphanedir. Hem kullanıcı arayüzü ile hem de komut satırı modunda öznitelik çıkartabilir. Güncellemelerin kütüphaneye dâhil edilmesi kolaydır [53].

JAudio, ses dosyalarını girdi olarak kabul eder ve çıktı olarak ACE XML ve ARFF dosyalarını oluşturur. ARFF dosyasına dışa aktarma işlemi, tüm çok boyutlu özelliklerin tek tek özelliklerin koleksiyonları olarak işlenerek gerçekleştirilir. Java'nın platformlar arası taşınabilirliğinden ve tasarım avantajlarından yararlanmak için Java ile yazılmıştır. Java'nın sınırlı çekirdek ses özelliklerini desteklemek amacıyla, jAudio'nun özelliklerini yazan kişilerin düşük seviyeli sorunlarla -örneğin bellekle ilgili ve format dönüşümleri gibi- uğraşmak yerine, özel bir düşük seviyeli ses katmanı uygulanmıştır. Bu sayede, doğrudan örnek değer dizileriyle çalışmalarına olanak tanınmıştır [54].

Hesaplamaların karmaşıklığını azaltmak amacıyla, genellikle daha önceki bir hesaplamanın sonuçlarını diğer modüllerde yeniden kullanmak avantajlı olabilir. Ayrıca, diğer sistemlerden farklı olarak, kullanıcının seçilen özelliklerin bağımlılıklarını bilmesine gerek yoktur. Bağımlılıkları olan çıktı için seçilen herhangi bir özellik, bağımlı özellikleri çoğaltma olmadan gerektiğinde otomatik ve sessiz bir şekilde hesaplar [54].

JAudio, rastgele sayıda boyut sağlayan özellikleri kabul etme kapasitesine sahiptir. Bu, örneğin MFCC gibi bir kez hesaplanan ilgili özellikleri gruplandırmanın son derece kullanışlı bir yaklaşımıdır. Bu, tüm özelliklerin tek boyutlu olduğu Weka'nın ARFF formatının aksine gelir. Ayrıca, her özelliğin boyutu dışa aktarılır. Bu, tüm özellikler için aynı kodun kullanılmasına rağmen, türevlerin ve diğer meta özelliklerin hesaplandığı özelliklerle aynı boyutta olmasına izin verir [54].

3.4. Ön İşleme Süreci

3.4.1. Sıralama

Ön işleme sürecinde “sort” işlemi, veri kümesindeki gözlemleri belirli bir özelliğe veya sıraya göre sıralama işlemidir. Veri kümesindeki gözlemlerin belirli bir sıralamaya göre düzenlenmesi, bozuk veya eksik verileri tespit edilmesi, verinin analiz edilmesi veya ML algoritmalarına uygulamak için bazen gereklidir.

Sort işlemi, veri setindeki gözlemleri artan veya azalan bir şekilde sıralamak için kullanılır. Sort işlemi, veri kümesindeki gözlemleri sadece belirli bir özelliğe göre

sıralamakla kalmaz, aynı zamanda birden fazla özelliğe göre sıralama yapabilme yeteneğine de sahiptir

3.4.2. Resample

"Resample" terimi, istatistiksel veri analizi ve makine öğrenimi alanlarında kullanılan bir veri dönüştürme yöntemini ifade eder. Resampling, mevcut bir veri setinin boyutunu veya örneklemini değiştirmek için kullanılır. Bu yöntem, özellikle dengesiz veri setleriyle çalışırken veya örneklem boyutunu değiştirerek model performansını iyileştirmek için kullanılır.

ML algoritmalarında Resampling işlemi modeli aşağıdaki şekillerde geliştirebilir [55]:

- Çapraz doğrulama ve önyükleme gibi Resample teknikleri modelin görünmeyen veriler üzerindeki performansının daha doğru bir tahminini sağlayarak aşırı öğrenmeyi azaltmaya yardımcı olabilir [55].
- Resample teknikleri, bir modelin daha büyük ve daha çeşitli bir veri kümesi üzerinde eğitilmesine olanak tanıyarak modelin performansını da artırabilir. Bu, modelin daha genelleştirilebilir desenler öğrenmesine ve doğruluğunu artırmasına yardımcı olabilir [55].
- Resample teknikleri, bir sınıfın diğerinden önemli ölçüde daha az örneğe sahip olabileceği dengesiz veri kümelerini işlemek için kullanılabilir. Azınlık sınıfını aşırı örnekleyerek veya çoğunluk sınıfını az örnekleyerek, yeniden örnekleme veri kümesinin dengelenmesine yardımcı olabilir ve modelin çoğunluk sınıfına karşı önyargılı olmasını önleyebilir [55].
- Resample, tek tek özelliklerin veya değişkenlerin önemi hakkında daha fazla bilgi sağlayarak modelin yorumlanabilirliğini artırmaya da yardımcı olabilir. Verilerin alt kümelerini rastgele seçerek ve modelin her bir alt kümedeki performansını değerlendirerek, yeniden örnekleme, modelin tahminleri için hangi özelliklerin en önemli olduğunu belirlemeye yardımcı olabilir [55].

3.4.3. Standardizasyon

Standardizasyon, farklı değişkenleri aynı ölçeğe getirme işlemi ve adımıdır. Bu yaklaşım, farklı değişken türleri arasındaki puanları karşılaştırmak için kullanılır. Standardizasyon

(Z-Skor Normalizasyon) kavramı, tamamen Standart Sapma ve Varyans matematiksel terimlerine dayanır. Formül 7'deki şekilde hesaplanır [56].

$$Z = \frac{(x-\mu)}{\sigma} \quad (7)$$

Standardizasyon, özellik değerlerini belirli bir ölçekte ifade ederek, farklı özellikler arasında doğru karşılaştırmalar yapılmasını sağlar. Özellikler genellikle farklı birimlerde veya farklı ölçeklerde olabilir. Bu durum, çeşitli ML algoritmalarının başarılarını etkileyebilme durumunu içerebilir. Standardizasyon işlemi, veri setinin her bir özelliği için ayrı ayrı uygulanır. Bu işlemde, veri kümesinin eğitim ve test kümeleri gibi farklı parçalara bölünmesi durumunda, aynı standardizasyon parametrelerinin her bir parçaya uygulanması gerektiğini unutmamak önemlidir.

3.4.4. Discretize

Discretize, bir veri kümesindeki sürekli (continuous) özellikleri, belirli aralıklara bölerek kategorik (categorical) veya düzeyli (ordinal) değerlerle temsil etmek anlamına gelir. Bu işlem, sürekli değerlerin daha az sayıda ve daha anlamlı kategorilere dönüştürülmesini sağlar.

Discretize, sürekli değerlere sahip bir özelliği kategorik veya düzeyli değerlere dönüştürerek veri kümesini dönüştürdüğü için veri ön işleme adımı olarak kabul edilir.

Discretize işlemi, özellikle ML algoritmalarının sürekli değerlerle çalışması gerektiğinde veya bazı veri analiz tekniklerinde kullanılır. Bazı algoritmalar, sürekli değerlerle doğrudan çalışamaz veya sürekli verileri daha iyi anlamak ve analiz etmek için kategorik veya düzeyli verilere ihtiyaç duyar. Discretize işlemi, veri analizinde önemli bir adımdır. Ancak, discretize işlemi bazen bilgi kaybına yol açabilir, bu nedenle dikkatlice seçilmeli ve uygulanmalıdır.

Tez sürecinde çalışma kapsamında sıralama, resample, standardizasyon ve discretize ön işleme işlemleri kendi içinde farklı kombinasyonlar ile kullanılmış ve sonuçlar karşılaştırılmıştır.

3.5. Özniteliklerin Seçilmesi

Öznitelik seçme metotlarını temelde 3 grupta toplayabiliriz. Bunlar Filtreler, Sarmalayıcılar ve Gömülü yöntemlerdir [57].

Filtreler öznitelik seçiminde kullanılan en yaygın yöntemlerdir; farklı veri tiplerine ve öznitelik özelliklerine göre kullanılmaları uygun olabilir. Filtrelerde veri setindeki nitelikler arasında istatistiksel yöntemler kullanılarak öznitelik ile sonuç arasında bir ilişki oluşturulur. Yüksek ilişki değerine sahip olan nitelik ise süreçte değerli bilgi olarak kabul edilir [17]. Filteler aşağıdaki şekilde gruplandırılır:

- **Information Gain filtresi:** Özniteliklerin hedef değişken üzerindeki etkilerini ölçmek için entropi ölçüsünü kullanır.
- **Anova filtresi:** Sayısal öznitelikler arasındaki farklılıkları ölçmek için varyans analizini kullanır.
- **Chi-kare filtresi:** Kategorik özniteliklerin hedef değişken üzerindeki etkilerini ölçmek için chi-kare testini kullanır.
- **Korelasyon filtresi:** Özniteliklerin birbirleriyle olan ilişkisini ölçerek, yüksek korelasyona sahip öznitelikleri filtreler.
- **Varyans filtresi:** Varyansı düşük olan öznitelikleri filtreler.

Makine öğrenmesinde öznitelik seçiminde kullanılan sarmalayıcı yöntemlerinden bazıları ise şunlardır:

- **Geriye Dönük Seçim (Backward Elimination):** Başlangıçta tüm özniteliklerin bulunduğu bir model oluşturulur ve daha sonra önemsiz öznitelikler sırayla çıkarılarak en iyi öznitelik kümesi elde edilir.
- **İleri Seçim (Forward Selection):** Başlangıçta hiçbir öznitelik olmayan bir model oluşturulur ve en iyi öznitelikler sırayla eklenerek en iyi öznitelik kümesi elde edilir.
- **Melez Seçim (Hybrid Selection):** İleri seçim ve geriye dönük seçim yöntemleri birleştirilerek, daha hızlı ve verimli bir öznitelik seçimi sağlanır.
- **Yerleşik Öznitelik Seçimi (Embedded Feature Selection):** Özellikle karmaşık öğrenme algoritmaları için kullanılan bu yöntemde, öznitelik seçimi öğrenme işlemi sırasında doğrudan gerçekleştirilir.

Sarmalayıcı yöntemler, filtre yöntemlerine göre daha zaman alıcı ve hesaplama açısından daha maliyetlidir, ancak daha iyi sonuçlar verebilirler. Seçilecek yöntem, veri setinin özelliklerine ve amaçlarına göre belirlenmelidir.

Gömülü yöntemler, öznitelik seçimi işlemini öğrenme algoritmasının kendisi içinde gerçekleştiren yöntemlerdir. Bu yöntemlerden bazıları şunlardır:

- **Karar Ağaçları (Decision Trees):** KA, verileri belirli bir hedef değişkene göre sınıflandıran bir ağaç yapısıdır. KA, özniteliklerin önemini belirlemek için kullanılabilir.
- **Lasso (Least Absolute Shrinkage and Selection Operator):** Öznitelik seçimi konusunda sıkça tercih edilen bir yöntem olarak Lasso regresyonu yaygın bir şekilde kullanılır. Lasso, regresyon katsayılarını sıfıra yaklaştırarak önemsiz öznitelikleri filtreler.
- **Ridge Regresyonu:** Ridge regresyonu da Lasso regresyonu gibi, öznitelik seçimi için kullanılan bir yöntemdir. Ridge regresyonu, regresyon katsayılarının büyüklüğünü sınırlayarak önemsiz öznitelikleri filtreler.
- **LDA:** LDA, öznitelik seçiminde kullanılan bir yöntemdir. LDA, sınıflandırma doğruluğunu artırmak için en iyi öznitelik kümesini seçer.
- **Özellik Çıkarma Yöntemleri (Feature Extraction Methods):** Özellik çıkarma yöntemleri, öznitelik seçimi işlemini öğrenme algoritmasının bir parçası olarak gerçekleştiren yöntemlerdir. Bu yöntemler arasında, Prensip Bileşen Analizi (Principal Component Analysis- PCA) ve Yerleştirme Yöntemleri (Embedding Methods) yer almaktadır. PCA da yine analiz sürecinde değerlendirilmiş ve sonuçları incelenmiştir.

Gömülü yöntemler, diğer yöntemlere göre daha az hesaplama maliyetine sahip olabilir, ancak aynı zamanda öğrenme algoritmasının performansına da bağlıdır. Seçilecek yöntem, veri setinin özelliklerine ve amaçlarına göre belirlenmelidir.

Öznitelik seçme kısmında farklı uygulamalar ve metrikler ile çeşitli analizler yapıldı. İlk olarak Weka Programı ile gelen öznitelik seçme algoritmalarından BestFirst (CfsSubsetEval) ve Ranker (InfoGainAttributeEval, ClassifierAttributeEval, CorrelationAttributeEval) arama metodu ve öznitelik değerlendiricileri ile veriseti üzerinde öznitelik seçme işlemi uygulandı. 8 farklı sınıflandırıcı ve 8 farklı öznitelik kümesinde öznitelik seçme öncesi ve sonrası tüm değerler karşılaştırıldı. Genel olarak yüksek tanıma oranı elde edilmiş sınıflandırıcılarda öznitelik seçme işlemi sonrası başarı oranlarında ciddi bir artış gözlemlenmedi.

İkinci olarak ise Orange uygulaması üzerinde “Information Gain Ratio” metrik değerine göre farklı öznitelik kümeleri oluşturulup performansa etkileri analiz edildi. Bu analiz sonucunda da daha yüksek tanıma oranlarına ulaşılmadığı için çalışma kapsamına dâhil edilmemiştir.

Son olarak ise Python sklearn kütüphanesi ile gelen öznitelik seçme yöntemlerinden SelectKBest, SelectFromModel ve RFE öznitelik seçme işlemleri analiz edildi. SelectKBest yöntemi, en yüksek puanlı k özniteliği seçer. ‘Score_func’ parametresi değiştirilerek, bu yöntem sınıflandırma ve regresyon problemleri için uyarlanabilir [58]. Yapılan analizde ‘score_func’ varsayılan değeri kullanılarak emo_large (6.552) öznitelik seti ve V2 veriseti üzerinde k değerini 3.000, 3.500, 3.700, 4.000, 4.500, 5.000 ve 5.500 atayarak sonuçları analiz edildi. Burada emo_large 6.552 öznitelik setinden elde edilen sonuçlardan daha yüksek bir başarı oranı gözlemlenmedi. Scikit-learn API, veri kümesinin en iyi özelliklerini ağırlıklarına dayalı olarak seçmek için SelectFromModel sınıfını sunar. SelectFromModel, verilen eşik değeriyle ağırlık önemini karşılaştırarak belirleyen bir meta-tahminci işlevi görür [59]. Burada bu yapıyı kullanabilmek için bir tahmin edici seçilmesi gerekmektedir. Çalışma kapsamında lineer model olarak Lasso, Ridge, ElasticNet; ağaç yapısı olarak Karar Ağacı Regresörü (DecisionTreeRegressor), Karar Ağacı Sınıflandırıcısı (DecisionTreeClassifier); topluluk modelleri olarak da Rastgele Orman Regresörü (RandomForestRegressor), Rastgele Orman Sınıflandırıcısı (RandomForestClassifier), Gradyan Arttırıcı Regresörü (GradientBoostingRegressor), Gradyan Arttırıcı Sınıflandırıcısı (GradientBoostingClassifier) kullanılmıştır. Lineer ve ağaç yapısı modellerinin tahmin edici seçildiği analizlerde tanıma performansının ciddi oranda düştüğü gözlemlenmiştir. Ancak topluluk modellerine bakıldığında ise öznitelik seçme işlemi olmadan elde edilen sonuçlar ile yakın değerlerin alındığı görülmektedir. RFE ise özyinelemeli öznitelik eleme yöntemidir. RFE, yapılandırılması ve kullanılması kolay olduğu ve bir eğitim veri kümesinde hedef değişkeni tahmin etmede daha çok veya en alakalı olan özelliklerin (sütunların) seçilmesinde etkili olduğu için popülerdir. RFE kullanırken iki önemli yapılandırma seçeneği vardır: seçilecek özellik sayısı seçimi ve özelliklerin seçilmesine yardımcı olmak için kullanılan algoritma seçimi [60]. Bu çalışmada RFECV kullanıldı. RFECV (Recursive Feature Elimination with Cross-Validation), RFE (Recursive Feature Elimination) yönteminin geliştirilmiş bir versiyonudur. Her ikisi de öznitelik seçimi yöntemleridir; ancak RFECV, RFE'nin aşamalarını daha da iyileştirerek daha güvenilir sonuçlar elde etmeyi amaçlar. RFECV

çapraz doğrulama (cross-validation) yöntemini içerir. Çalışmada tahmin edici olarak Karar Ağacı Sınıflandırıcısı kullanılmıştır. Öznitelik seçme yapılmadan elde edilen sonuçlar ile benzer başarı oranlarına ulaşılmıştır.

Yine test sürecinde LDA ve PCA yöntemleri de öznitelik seçme sürecinde analiz edilmiş modeller içerindedir.

Sonuç olarak tüm bu analizler neticesinde öznitelik seçme işlemi sonrası ek bir başarı oranı elde edilmediği için öznitelik seçme süreci çalışma kapsamına dâhil edilmemiştir.

3.6. Sınıflandırma

Çalışmanın sınıflandırma süreci 2 farkı kırılımda ilerlemiştir. Bu çalışmada konuşmadan duygu analizi problemi ML ve DL algoritmaları ile analiz edilmiştir. ML algoritmaları ile yapılan analiz kısmında SVM, KNN, LDA, LR, KA, Naive Bayes (NB), EA ve RO sınıflandırıcıları kullanılmıştır.

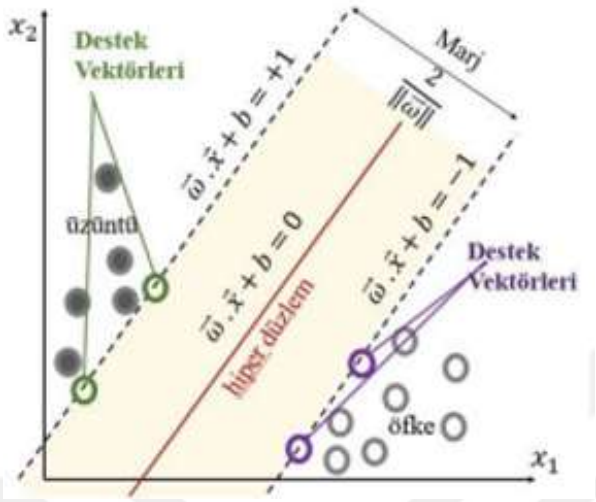
3.6.1. Makine Öğrenmesi

3.6.1.1. Destek Vektör Makineleri (DVM)

DVM istatistiksel öğrenme teorisine dayanan bir algoritmadır. DVM küçük ölçekli verilerin sınıflandırılmasında daha uygun olup daha başarılı sonuçlar elde edilmesini sağlamaktadır. Ayrıca DVM konuşmacı bağımsız veri setlerinde ve doğrusal olmayan öznitelik setlerinde daha yüksek performans sağlamaktadır [19]. DVM, denetimli optimizasyon temelli bir makine öğrenimi algoritmasıdır ve hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. Esneklik ve genelleme performansı oldukça yüksektir. Birçok konuşmadan duygu tanıma çalışmalarında da başarılı sonuçlar verdiği gözlemlenmiştir.

DVM girdileri yüksek boyutlu bir özellik uzayına dönüştürmek için çekirdek fonksiyonu (kernel) kullanır. Bu fonksiyon girdi değerlerini doğrusal bir şekilde ayrılabilir hale getirir. DVM nin temel yapısını Şekil 3.5'te görebilirsiniz.

Şekil 3.5. *Destek Vektör Makineleri Temel Yapısı*



Burada beyaz ve siyah yuvarlaklar farklı iki sınıfı göstermektedir. Sınıfları birbirinden ayıran bir doğru ve hiper düzlem oluşturulmuştur. Marj değeri ne kadar fazla olursa sınıflar o kadar başarılı bir şekilde ayırt edilmiş olur [19].

3.6.1.2. Lojistik Regresyon (LR)

LR, ML alanında sınıflandırma problemlerini çözmek için kullanılan bir istatistiksel modeldir. LR, bağımlı bir değişkenin (hedef değişkeni) iki veya daha fazla kategorik sınıfa ait olduğu durumlarda kullanılır.

LR analizi, bağımsız değişkenlerin sonuç değişkenine etkilerini olasılık olarak hesaplayan bir yöntemdir. Aynı şekilde risk faktörlerini de olasılık olarak belirler. Bu analizin temel amacı tıpkı doğrusal regresyon analizinde olduğu gibi bağımsız değişkenler ile sonuç değişkeni arasında bir model oluşturmaktır. Diğer regresyon yöntemlerinde sonuç değişkeni sürekli değer alır ancak bu yöntemde sonuç değişkeni kesikli değerler alır. Ayrıca diğer analiz yöntemlerine göre daha esnek varsayımlar içerir. Bu da tercih edilmesinde bir başka faktördür [61].

LR bağımsız değişkenin sürekli ya da süreksiz olabildiği, sürekli ya da süreksiz bağımsız değişkenlerin bir arada bulunabildiği ve bağımlı değişkenin süreksiz olduğu bir analiz yöntemidir. LR'na benzeyen analiz türleri Çoklu Doğrusal Regresyon Analizi, Diskriminant analizi ve kümeleme analizidir. Ancak Çoklu Regresyon ve Diskriminant Analizi sahip olduğu varsayımlar nedeniyle ve kümele analizi grup sayıları ve üyelikleri bilinmediği için LR analizinden ayrılmaktadır [62].

LR matematiksel olarak ise olasılık, odds ve odds'un logaritmasına dayanır [62].

Bağımlı değişkenin iki gruptan oluştuğu regresyon modeline İkili Lojistik Regresyon modeli, bağımlı değişkenin ikiden fazla kategoriye sahip olduğu regresyon modeline Çoklu Lojistik Regresyon modeli adı verilir [61].

Bir değişkenin bağımsız değişkenlere bağlı olarak alacağı değeri tahminlemek doğrusal regresyon analizinin alanı iken, iki sonuçlu bir değişkenin bağımsız değişkenlere bağlı olarak alacağı değerin olasılığının tahminlenmesi LR analizinin çalışma alanıdır [61].

3.6.1.3. K-en Yakın Komşu (KNN)

KNN sınıflandırıcısı tembel öğrenme ya da mesafe tabanlı öğrenme olarak adlandırılan denetimli bir öğrenme algoritmasıdır. KNN, test edilen vektörlere en yakın olan k eğitim vektörlerine dayalı kararlar veren çok basit bir yöntemdir. KNN verilerden üretilen bir modelden ziyade doğrudan eğitim verilerini kullanan bir sınıflandırıcıdır [19].

KNN yönteminin avantajları; yüksek başarı oranları, gürültülü verilerde etkili basitlik ve verimliliktir. KNN yönteminin dezavantajları ise büyük boyutlu verilerde düşük doğruluk değerleri, mesafe ölçüsünün belirsizliği, yüksek donanım gereksinimi ve yüksek hesaplama maliyeti olarak sıralanabilir [63].

Temelde çalışma şekline baktığımızda eğitim veri seti üzerinden süreç başlar. KNN Algoritması veri noktaları arasındaki benzerliği ölçmek için genellikle öklidyen mesafesini kullanır. Öklidyen iki nokta arasındaki doğrusal mesafeyi hesaplar. KNN algoritması, komşuluk sayısını belirleyen bir parametre olan K değerini gerektirir. K değeri, bir veri noktasının sınıflandırılmasında kullanılacak en yakın komşu sayısını belirtir. K genellikle tek sayıdan oluşur. Yapılan literatür incelemesinde K değerinin genellikle 3, 5 ya da 10 olarak kullanıldığı görülmüştür. KNN algoritması, tahmin yapmak veya sınıflandırma yapmak için bir veri noktasını işler. İşlem sırasında, veri noktasının K en yakın komşusu belirlenir. Belirlenen K en yakın komşunun sınıfları dikkate alınarak, sınıflandırma veya tahmin yapılır. Sınıflar arasında çoğunluk oyu uygulanır. Yani, K en yakın komşunun sınıfları arasında hangi sınıfın daha fazla sayıda olduğu belirlenir ve veri noktası bu sınıfa atanır.

3.6.1.4. Karar Ağacı (KA)

KA, denetimli makine öğrenimi algoritmalarından biridir. Sınıflandırma ve regresyon problemlerini çözmede kullanılır. Kararlar yaprak düğümlerde bulunur ve veri düğümlere bölünür. Sınıflandırma ağaçlarında karar değişkeni genellikle kategorik iken, regresyon ağaçlarında bu değişken sürekli bir değere sahiptir. KA'nın avantajları hem sınıflandırma hem de regresyon problemlerine uygun olması, yorumlama kolaylığı, kategorik ve kantitatif değerlerin kullanım kolaylığı, en olası değer ile eksik değerleri doldurabilmek ve yüksek performans değerleridir [18].

Dezavantajları ise kararsız olabilmeleri, ağacın boyutunu kontrol etmenin zorluğu, örnekleme hatası ve genel bir çözüm yerine yerel bir çözüm sunmasıdır [18].

3.6.1.5. Naive Bayes (NB)

Bu sınıflandırıcı algoritması Naive Bayes teoremine dayanır. Genel olarak veri madenciliği, makine öğrenmesi ve duygu analizi gibi alanlarda yaygın kullanıma sahiptir. Denetimli öğrenme grubunda olan bir algoritmadır. Temelde olasılık hesaplamalarını kullanır ve en yüksek olasılığa sahip olan kararı seçer. Bu hesaplama da her bir niteliğin aynı sınıftaki diğer niteliklerden farklı olduğu kabul edilir [64].

Bayes Teoreminin matematiksel ifadesi Formül 8'de gösterilmiştir [6].

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{8}$$

P(X|Y); Y olayı gerçekleştiği durumda X olayının meydana gelme olasılığı,

P(Y|X); X olayı gerçekleştiği durumda Y olayının meydana gelme olasılığı

P(X) ve P(Y); X ve Y olaylarının olasılıklarıdır.

NB sınıflandırma algoritmasının matematiksel ifadesi Formül 9'da gösterilmiştir.

$$P\left(Class_j|x\right) = \frac{P\left(x|Class_j\right)P(Class_j)}{P(x)} \tag{9}$$

Burada verilmiş bir x tahminci olasılığı, j sınıfına ait olasılık ve eğitim kümesinden bulunan bu sınıf tahminci olasılığı ile x tahmincisinin bu sınıfa ait olma olasılığı hesaplanabilir [6].

3.6.1.6. Lineer Diskriminant Analizi (LDA)

Konuşma ve sınıflandırma örneklerinin boyutsallığının azaltılması amacıyla kullanılan bir istatistiksel yöntemdir. Bu algoritma, optimal bir dönüşüm matrisini hesaplamak için sınıflar arasındaki kovaryans matrisini maksimize ederken, konuşma örneklerinin sınıf içi kovaryans matrisini ise en aza indirir [19].

LDA, verilere sınıf koşullu yoğunluklarının uyum sağlandığı ve Bayes kuralının uygulandığı bir yöntemdir. Bu işlem sonucunda doğrusal bir karar sınır sınıflandırıcısı oluşturulur. LDA aynı zamanda doğrusal diskriminantın genelleştirilmiş bir formudur [65].

Bu yaklaşım, veri kümesini daha düşük boyutlu bir formata dönüştürmeyi amaçlar. Diğer bir ifadeyle, modelin karmaşıklığını ve hesaplama maliyetlerini düşürmek için boyutların azaltılması hedeflenir. Normal LDA modeli, Gauss yoğunluğuna sahip herhangi bir sınıfa uyarlanabilir ve her sınıfın aynı kovaryans matrisini kullandığını varsayar [65].

3.6.1.7. AdaBoost (ADA)

“Adaptive Boosting” ifadesinin kısaltmasıdır. Amacı ise sınıflandırma problemlerine odaklanıp zayıf olan sınıflandırıcıları güçlü sınıflandırıcılara dönüştürmektir. Başlarken her bir örnek için eşit dağılım yapılır ve sonuçta ise sınıflandırma performansına göre en zayıf sınıflandırıcı belirlenir. Ağırlıklar güncellenir. Belli bir tekrar sonucunda ise en güçlü sınıflandırıcılar bir araya getirilir ve güçlü bir sınıflandırıcı oluşturulmuş olur. Bu şekilde başarı oranı da artırılır [66].

3.6.1.7. Gradyan Artırma (GA)

Güçlendirme temel olarak zayıf öğrencileri güçlü öğrenenlere dönüştürmenin bir yöntemidir. Bunu özyinelemeli, aşamalı olarak yapar. Birden fazla yükseltme algoritması vardır. Temel fark, zayıf öğrencinin eksikliğini nasıl tanımladıklarında ortaya çıkar. GA’da ilk yaprak önce oluşturulur. Sonrasında, tahmin hataları dikkate alınarak yeni ağaçlar inşa edilir. Bu süreç belirtilen ağaç sayısına veya istenen performans seviyesine kadar tekrar eder. Başlangıçta, hedef değerin ortalaması alınır. Bu, ilk tahmin olarak kabul edilen ilk yapraktır. Daha sonra elde edilen değer hedef değer ile karşılaştırılıp hata miktarı tespit edilir [67].

3.6.1.8. Ekstra Ağaç (EA)

EA sınıflandırıcısı, topluluk ağacı tabanlı bir makine öğrenimi algoritmasıdır. Randomizasyonu kullanarak hesaplama maliyetini ve varyansı azaltmayı amaçlar. Hem sınıflandırma hem de regresyon problemleri için kullanılabilir. EA algoritması, birçok karar ağacını eğitir ve bu ağaçlardan elde edilen sonuçları birleştirerek tahminler oluşturur.

RO ile benzerlik taşıyan EA algoritması, eğitim verilerinin örneklemesinde bazı farklılıklar içerir. RO, her ağaç için örneklem seçiminde torbalama yöntemini kullanırken, EA tüm veri kümesini kullanır. Ayrıca, EA karar ağaçları oluştururken bir özelliğin bölünme değerlerini ve alt düğümleri rastgele seçer, fakat RO'da bu işlem için belirli bir algoritma kullanılır.

EA algoritması, RO'na göre daha hızlı çalışabilme özelliğine sahiptir [68].

3.6.1.9. XGBoost (XGB)

XGB, karar ağacı algoritmasına dayanan ve optimize edilmiş performans sunan bir ML algoritmasıdır. Bu algoritmanın en gelişmiş versiyonu olarak kabul edilir ve özellikle karar ağacı tabanlı uygulamalarda mükemmel bir performans gösterir. Yüksek tahmin gücü ve hızı ile dikkat çekerken, aynı zamanda aşırı öğrenme riskini azaltabilir.

XGB nin temel çalışma prensibi, önce bir tahmin yaparak başlar. Bu tahmin, ardışık adımlarda yapılan düzeltmelerle daha doğru bir sonuca yaklaşmak için kullanılır. İlk tahmin için varsayılan bir değer genellikle 0,5 olarak alınır. Daha sonra, modelin hatalı tahminleri incelenir ve bu hatalar, gözlemlenen değer ile tahmini değeri arasındaki fark olarak hesaplanır. Model, bu hataları düzelterek daha iyi sonuçlar elde etmeye çalışır.

XGB ve GA'nın temel çalışma mantığı aynıdır, ancak detaylarda bazı farklılıklar bulunmaktadır. XGB, özellikle hız ve performans konusunda optimize edilmiş bir GA sürümüdür. Bu nedenle, büyük veri setlerinde ve karmaşık modellerde daha iyi sonuçlar elde etmek için tercih edilir [69].

3.6.1.10. Rastgele Orman (RO)

RO algoritması, denetimli sınıflandırma ve regresyon problemlerinde kullanılan bir ML yöntemidir. Hem sınıflandırma hem de regresyon görevlerinde etkilidir. Temel olarak, bu algoritmada, birden fazla bağımsız çalışan karar ağacı bir araya gelir ve en yüksek puanı

alan seçilir. Bu senaryoda, ağaç sayısının artmasıyla başarı oranının yükseldiği görülür. Temel farkı, karar ağaçlarına kıyasla, rasgele ormanda kök düğümünün seçimi ve düğümlerin bölünmesi işlemlerinin rastgele gerçekleşmesidir [70].

Kullanılan 8 sınıflandırıcı da varsayılan değerden farklı olarak kullanılan parametre değerleri Tablo 3.9'da listelenmektedir.

Tablo 3.9. *Sınıflandırıcılarda Varsayılan Değerden Farkı Atanan Parametreler*

Sınıflandırıcı	Varsayılan Değerden Farkı Atanan Parametreler
DVM	Random_state=0, tol=0.001
KNN	N_neighbors=5
LDA	-
LR	Solver='lbfgs', max_iter=1.000, random_state=0
KA	Random_state=0
NB	-
EA	N_estimators=100, random_state=0
RO	Max_depth=2, random_state=0

Derin öğrenme tarafında ise DNN, CNN (1D), CNN (Alexnet), CNN (Resnet), LSTM, RNN, CNN+LSTM sınıflandırıcıları ile konuşmadan duygu tanıma süreci analiz edilmiştir.

3.6.2. Derin Öğrenme

3.6.2.1. Derin Sinir Ağları (DNN)

DNN temelde "Çok katmanlı algılayıcılar" ya da "Derin İleri Besleme Ağları" olarak da tanımlanabilir.

İleri beslemeli sinir ağları denetimli öğrenme modelinde çalışır. Bu sinir ağında (x, y) şeklinde gelen eğitim verisi kullanılır ve x ile y'yi ilişkilendiren bir model oluşturulur. Bu ilişki doğrusal ya da doğrusal olmayabilir.

İleri beslemeli sinir ağlarında, giriş ve çıkış katmanları arasında en az bir tane veya daha fazla gizli katman bulunmaktadır. Bu katmanlardaki işlem birimlerine nöron denir. Bu tür sinir ağları tek yönlü olduğu için içeride döngü oluşmaz ve akış girişten çıkışa doğrudur. Her katmanın çıktısı, bir sonraki katmanın girdisi olarak işlev görür. Genelde ise ikiden fazla gizli katmanı bulunan sinir ağları "derin" olarak tanımlanır. Yalnız gizli katman sayısında kabul edilmiş bir adette yoktur [12].

DL sürecinde öğrenme durumu ağın çıktı katmanında görülen hatayı azaltacak şekilde ağırlıkların değiştirilmesi sürecidir. Eğitim sürecinin başında bu değerler rastgele belirlenir. Sonrasında ise girdi ile ağırlıklar çarpılıp toplanarak bir fonksiyondan geçirilir ve sürekli değere sahip bir çıktı elde edilir. Bu işlem çıkış katmanında da tekrar edilerek girdi için sonsal olasılık değeri hesaplanır. Ağın ürettiği bu olasılık değeri gerçek çıktı ile karşılaştırılarak aralarındaki hata miktarı bulunur. Bu hata değerinin yüksek olması ağın başarısız olarak değerlendirilmesi demektir. Bu durumda bu hata değerini düşürebilmek adına ağırlık değerlerinde değişiklik yapılır. Bu işlem için hatanın her bir ağırlık değerine göre türevi alınır. Buna göre de ağırlık değeri biraz artırılır ya da azaltılır. Bu yeni değerler ile ağ tekrar eğitilir. Hata değeri en düşük değere ulaşıncaya kadar bu işlemlere devam edilir. En düşük değere ulaşmak demek x ile y arasındaki modelin kurgulanmış olması demektir. Eğitim süreci de bu şekilde tamamlanır [12].

Tez kapsamında oluşturulmuş DNN modeli 128 nörona sahip bir girdi katmanı, 64 nörondan oluşan 5 gizli katman ve 4 nörona sahip çıkış katmanından oluşmaktadır. Giriş ve gizli katmanlarda ReLU aktivasyon fonksiyonu kullanılmıştır. Çıkış katmanında ise Softmax aktivasyon fonksiyonu kullanılmıştır. Eğitim sırasında kullanılan optimize edici algoritma Adam'dır. Adam gradient tabanlı bir optimize edici algoritmadır. Adam, önceki gradyanların momentumunu kullanarak gradient tabanlı güncellemeler yapar. Modelin eğitimi sırasında kullanılan kayıp fonksiyonu ise "sparse_categorical_crossentropy"'dır. Kayıp fonksiyonu, modelin tahminlerinin gerçek değerlerden ne kadar farklı olduğunu ölçer ve modelin bu farkı azaltması için geriye doğru yayılım ile kullanılır. "Sparse_categorical_crossentropy" çok sınıflı sınıflandırma problemlerinde kullanılan bir kayıp fonksiyonudur. V1 verisetinde batch_size=64 ve V2, V3 verisetinde batch_size=128 alınmıştır. Epoch değeri de 100 olarak belirlenmiştir. Bu kapsamda emo_large öznitelik seti üzerinde V1, V2 ve V3 verisetleri için %90-10, %80-20 ve 10 kat çapraz doğrulama ile model analiz edilmiş ve sonrasında en yüksek tanıma oranı %92,57 ile V2 kayıtlık veri setinde, 10 kat çapraz doğrulama ile elde edilmiştir.

3.6.2.2. Evrişimli Sinir Ağları (CNN)

CNN, genellikle görüntü tanıma ve işleme görevleri için özellikle uygun bir DL algoritmasıdır. Evrişimli katmanlar, havuzlama katmanları ve tam bağlantılı katmanlar gibi farklı katmanlardan oluşur. Evrişimli katmanlar, giriş görüntüsüne filtrelerin uygulandığı ve kenarlar, dokular ve şekiller gibi özellikleri çıkarmak için kullanılan bir

CNN'nin temel bileşenidir. Evrişimli katmanların çıktısı, daha sonra uzamsal boyutları azaltarak en önemli bilgileri koruyan özellik haritalarını elde etmek için kullanılan havuzlama katmanlarına geçirilir. Havuzlama katmanlarının çıktısı ise bir veya daha fazla tam bağlantılı katmana geçirilerek bir tahminde bulunma veya görüntüyü sınıflandırma amacıyla kullanılır [71].

AlexNet, LeNet, VGGNet, Resnet, MobileNet, GoogleNet, DenseNet ve EfficientNet DL ve CNN alanında önemli modellerdir. Bu modeller genellikle büyük miktarda veri ile eğitilmiş ve çeşitli görsel tanıma, nesne tespiti, yüz tanıma, doğal dil işleme ve benzer görevlerde yüksek performans gösteren ağ yapılarıdır.

Çalışma kapsamında 1 boyutlu CNN ile model oluşturulmuştur. Model 8 adet 1D Evrişimli katman, 2 adet BatchNormalization, 2 adet 1D Havuzlama, 2 adet Dropout, 1 Flatten ve çıktı katmanından oluşmaktadır. Tasarladığımız modelin girdi katmanı 256 nörondur. Sonraki evrişimli katmanlarda nöron sayısı 2 üstel olacak şekilde azaltılmıştır. Girdi ve diğer katmanlarda ReLU aktivasyon fonksiyonu olarak kullanılmıştır. Modelin çıktı katmanında ise aktivasyon fonksiyonu olarak Softmax kullanılmıştır. 'categorical_crossentropy' ise kayıp fonksiyonudur. Optimize edici algoritma ise RMSprop'tır. Epoch sayısı 200 ve batch_size 64'tür. Modele erken durdurma eklenmiştir. Erken durdurma, modelin aşırı öğrenme (overfitting) riskini azaltmak ve daha iyi genelleştirme yapmasını sağlamak için kullanılan bir yöntemdir. Erken durdurmada kullandığımız monitor='loss' ifadesi, erken durdurma işlevinin izleme ölçütünün kayıp (loss) değeri olduğunu belirtir. "Patience" ifadesi ise, erken durdurmanın kaç epoch boyunca izleme ölçütünde iyileşme görmemesi durumunda eğitimi durduracağını belirtir.

Modelde en yüksek tanıma oranı %90,25 ile V2 verisetinde elde edilmiştir (10 kat çapraz doğrulama ile).

3.6.2.3. Tekrarlayan Sinir Ağları (RNN)

RNN önceki adımın çıktısının mevcut adıma girdi olarak beslendiği bir sinir ağı türüdür. Geleneksel sinir ağlarında tüm girdi ve çıktılar birbirinden bağımsız olarak ele alınır. Ancak bir cümlenin sonraki kelimesini tahmin etmek gibi durumlarda, önceki kelimelere ihtiyaç duyulur ve bu da önceki kelimelerin hatırlanmasını gerektirir. Bu ihtiyacı karşılamak amacıyla RNN ortaya çıkmıştır. RNN'nin temel ve önemli özelliği, bir dizi hakkında bilgi saklayabilen gizli durumudur, aynı zamanda "Hafıza Durumu" olarak adlandırılır. RNN, her girdide veya gizli katmanda aynı görevi yerine getirerek çıktıyı

üretir. Bu, diğer sinir ağlarının aksine parametrelerin karmaşıklığını azaltarak işlemi daha etkili hale getirir [72].

Çalışma kapsamında RNN ile de model oluşturulup analiz edilmiştir. Hem model katmanları hem de parametrelerde değişiklik yapılarak farklı kontroller yapılmış ancak yüksek tanıma oranına ulaşılamamıştır.

3.6.2.4. Uzun Kısa Süreli Bellek (LSTM)

LSTM ağları, bir tür tekrarlayan sinir ağı (RNN) olarak kabul edilir. LSTM, zaman serileri, konuşma ve metin gibi sıralı verileri özellikle işlemek üzere tasarlanmış bir RNN türüdür. LSTM ağları, sıralı verilerdeki uzun vadeli bağımlılıkları öğrenme yeteneğine sahip olup, bu özellikleri sayesinde dil çevirisi, konuşma tanıma ve zaman serisi tahmini gibi görevler için özellikle uygundur [73]. Geleneksel bir RNN, tek bir gizli durumunun zaman içinde geçmiş bilgileri tutmasını zorlaştıran bir yapısı vardır. Bunun aksine, LSTM'ler, uzun vadeli bağımlılıkları öğrenmeyi kolaylaştırmak amacıyla bir bellek hücresi kullanır. Bu bellek hücresi, uzun süre boyunca bilgi saklayabilen bir kap anlamına gelir ve üç farklı kapı ile kontrol edilir: giriş kapısı, unutma kapısı ve çıkış kapısı. Bu kapılar, bellek hücresine hangi bilgilerin eklenmesi gerektiğini, hangi bilgilerin unutulması gerektiğini ve hangi bilgilerin çıkış olarak kullanılacağını belirler. Bu sayede LSTM'ler, sıralı verilerdeki uzun vadeli bağımlılıkları daha etkili bir şekilde öğrenme yeteneğine sahiptir [73].

Çalışma kapsamında sadece LSTM, CNN-LSTM şeklinde 2 farklı model oluşturulmuştur. LSTM modelde düşük tanıma oranları elde edildi. CNN+LSTM modelinde ise emo_large öznitelik setinde 1.099 kayıt için %83,64 başarı oranına ulaşılmıştır (%80-20 eğitim ve test verisi için). Bu CNN-LSTM model ise 3 adet 1D evrişimli katman, 3 adet 1D havuzlama katmanı, 3 LSTM katmanı, 1 Dropout, 1 Flatten ve çıktı katmanından oluşmaktadır. Modelin girdi katmanı 32 nörondur. Sonraki evrişimli ve LSTM katmanlarındaki nöron sayısı 64'tür. Girdi ve diğer katmanlarda ReLU aktivasyon fonksiyonu olarak kullanılmıştır. Modelin çıktı katmanında ise aktivasyon fonksiyonu olarak Softmax tercih edilmiştir. 'categorical_crossentropy' ise kayıp fonksiyonudur. Optimize edici algoritma ise Adam'dır. Epoch sayısı 100 ve batch_size 32'dir.

DÖRDÜNCÜ BÖLÜM

DENEYSEL SONUÇLAR

4.1. Analiz Sonuçları ve Karşılaştırmalar

Çalışma kapsamında farklı veri seti sayıları, farklı öznitelik kümeleri ve farklı sınıflandırıcılar ile analizler yapılmıştır.

İlk yapılan analizde 5 saniyelik kesitlerden oluşan 391 ses kaydı oluşturuldu. Öznitelik olarak literatür taramasında da karşılaşılan ve en çok tercih edilen MFCC özniteliği kullanıldı. Sonrasında ise veriler üzerinde normal dağılımlı standardizasyon uygulandı. Elde edilen veriler farklı kırılımlarda (%90-10 ve 10-fold) eğitim ve test verisi olarak ayrıldı. Sınıflandırma aşamasının sonunda LR, GB, KNN, GNB, DVM, LDA, ADA, EA, KA, XGB ve RO gibi algoritmalar test edilmiştir. Bu testler sonucunda en iyi sonuçlar DVM ile elde edilmiştir. Ayrıca duygu bazında değerlendirme yapılırsa en yüksek tanıma oranına sahip olan duygu kızgın ve nötr iken en düşük tanıma oranı ise mutlu duygusunda elde edildi. Tüm sınıflandırıcılar ile elde edilen sonuçlar Tablo 4.1'de gösterilmektedir.

Tablo 4.1. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V1 Veriseti ve MFCC)*

Sınıflandırıcı	Doğruluk	Hassasiyet	Duyarlılık	F1-measure
DVM	**82,5%**	**84,2%**	**82,5%**	**82,5%**
LR	57,5%	57,3%	57,5%	57,3%
KNN	45,0%	54,0%	45,0%	44,6%
KA	35,0%	34,7%	35,0%	34,3%
GNB	50,0%	49,6%	50,0%	43,3%
LDA	62,5%	63,8%	62,5%	62,8%
ADA	47,5%	53,7%	47,5%	47,6%
GB	70,0%	70,8%	70,0%	70,2%
EA	75,0%	80,5%	75,0%	73,9%
XGB	60,0%	61,0%	60,0%	59,4%
RO	77,5%	77,9%	77,5%	77,2%

Tablo 4.1'e göre en iyi sonuç DVM ile elde edilmiş olup doğru tanıma başarı oranı %82,5'tir. Kullanılan veriseti %90-10 eğitim ve test verisi olarak ayrılmıştır. Uygun parametreleri belirlemek adına GridSearchcv kullanılmıştır. V1 için en yüksek başarı

oranı elde edilen DVM sınıflandırıcısı için kullanılan hiper parametre listesi Tablo 4.2'de gösterilmiştir.

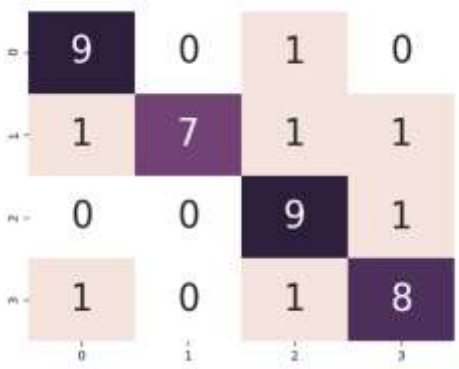
Tablo 4.2. *Hiper Parametreler (DVM Sınıflandırıcısı İçin)*

Parametre	**Kullanılan Değer**	**Parametre Aralığı**	**Varsayılan Değer**
Probability	True	True, False	False
C	1.0	0.1, 1, 10, 100, 1.000	1.0
Kernel	rbf	Linear, rbf, poly, sigmoid, precomputed	rbf
Gamma	scale	Scale, auto or float	scale
Break_ties	False	True, False	False
Decision_function_shape	ovr	Ovr, ovo	ovr
Class_weight	None	Dict, Balanced, None	None
Tol	$1e^{-3}$	Float values	$1e^{-3}$
Shrinking	False	True, False	True
Degree	6	0, 1, 2, 3, 4, 5, 6	3

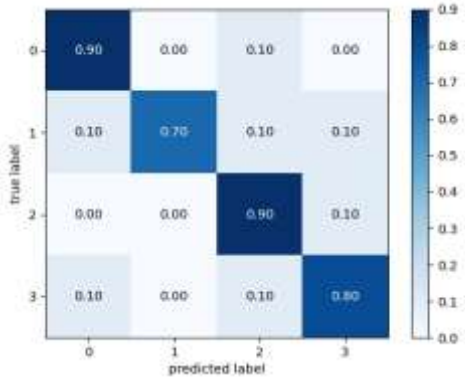
Tablo 4.2 'yi incelediğimizde varsayılan değerden farklı olarak kullandığımız 3 parametre bulunmaktadır. Bu parametreler arasında "probability", "shrinking" ve "degree" bulunmaktadır. "Probability", olasılık tahminlerinin etkinleştirilip etkinleştirilmeyeceğini belirler. "Shrinking", sezgisel bir yöntemin kullanılıp kullanılmayacağını belirtir. "Degree" ise polinom çekirdek fonksiyonunun derecesini ifade eder. Ayrıca, önemli olan birkaç parametreden biri de "gamma"'dır. Düşük bir "gamma" değeri seçilirse, eğitim veri kümesine sıkışık bir şekilde uyarlanırken, daha yüksek bir değer tam bir uyarlamaya yol açabilir. Ancak bu durum aşırı öğrenmeye neden olabilir.

Sınıflandırıcılar içinden en iyi sonucun elde edildiği DVM için Karışıklık Matrisi Şekil 4.1 ve Şekil 4.2'de gösterilmektedir.

Şekil 4.1. *Karışıklık Matrisi (DVM-MFCC Öznitelik)*



Şekil 4.2. *Karışıklık Matrisi (DVM-MFCC Öznitelik) (Yüzde Değeri)*



Şekil 4.1 ve Şekil 4.2'ye göre bakıldığında kızgın ve nötr duyguda %90 oranında tanıma sağlanırken, en düşük tanıma oranı %70 ile mutlu duygusunda gözlemlenmiştir. Genel sonuç ise %82,5'tir.

Ayrıca yine aynı modelleme yapısı içinde eğitim ve test verisinin ayrılması sürecinde 10 kat çapraz doğruma işlemi uygulandı. Bu yaklaşımda, veriler belirli bir "k" değerine göre eşit parçalara ayrılır. Bu şekilde, her bir parça hem eğitim hem de test için kullanılır. Bu yaklaşım, dağılma ve parçalanmanın sebep olduğu hataları ve sapmaları minimize etmek amacıyla kullanılır [74]. Ancak bu yöntemin dezavantajı ise eğitim ve test süresinin uzamasıdır. Bu küçük veriler için bir sorun olmasa da büyük verilerde hesaplama ve zaman açısından maliyetli olabilir.

Tablo 4.3'te k-kat çapraz doğrulama için çalışmada kullanılan parametreler, ilgili parametrenin varsayılan değerleri listelenmiştir.

Tablo 4.3. *K-kat Çapraz Doğrulama Parametreleri*

Parametreler	**Kullanılan Değer**	**Parametre Aralığı**	**Varsayılan Değer**
n_splits	10	Int values	5
n_repeats	3	Int values	10
random_state	123	RandomState instance or None	None

K-kat çapraz doğrulama işleminde en iyi sonuç kullanılan değer sütunundaki veriler ile sağlanmıştır. N_splits kart numarası değeridir. 2'den az olamaz. N_repeats, çapraz doğrulayıcının kaç kez tekrarlaması gerektiğidir. Random_state, tekrarlanan tüm çapraz doğrulama örneklerinin rastgeleliğini denetler.

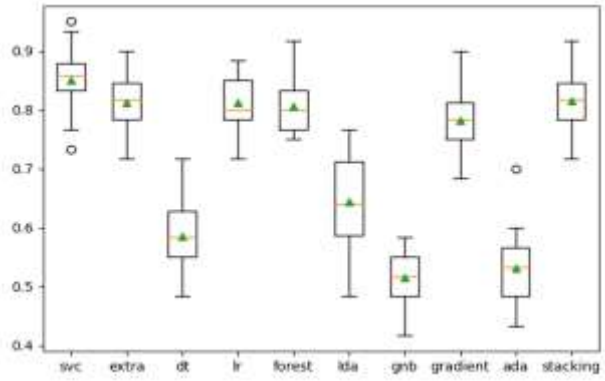
Tablo 4.4. *Sonuçlar ve Tüm Algoritmaların Karşılaştırılması (10 kat çapraz doğrulama ile Karşılaştırma)*

Sınıflandırıcı	Doğruluk	Doğruluk (10 kat Çapraz Doğrulama)
DVM	**82,5%**	**85,1%**
LR	57,5%	81,3%
KNN	45%	-
KA	35%	58,6%
GNB	50%	51,6%
LDA	62,5%	64,5%
ADA	47,5%	53,1%
GB	70%	78,3%
EA	75%	81,3%
XGB	60%	-
RO	77,5%	80,7%

Tablo 4.4'e göre incelediğimizde 10 kat çapraz doğrulama işlemi sonrası tüm sınıflandırıcılarda tanıma performansının arttığı gözlemlendi. Ancak yine de en yüksek doğruluk değeri DVM de elde edilmiştir (%85,1). En yüksek yüzdesel artış ise (iki model arasında) LR görüldü (%57,5'ten %81,3'e).

Şekil 4.3'te ise 10 kat çapraz doğrulama sonucu grafiksel olarak gösterilmiştir.

Şekil 4.3. *Analiz Sonuçları (10 kat Çapraz Doğrulama)*



Ayrıca farklı bir analiz olarak da birinci katmanda EA ve ikinci katmanda DVM olacak şekilde hibrit bir model de test edildi. Bu hibrit modelde doğruluk oranı %81,5 olmuştur. Bu durumda tek katmanlı DVM'de performansın daha iyi olduğu göstermiştir.

İkinci analizde ise hem farklı veri seti kümeleri hem de farklı öznitelik kümeleri çalışma sürecine dâhil edilmiştir. Üç farklı veri seti bu çalışma da analiz edildi. Bunlar V1 391, V2 862 ve V3 1.099 kayıttan oluşmaktadır. Tüm veri seti kümelerinin ortak özelliği 5 saniyelik ses kayıtlarından oluşması, tek kanallı olacak şekilde düzenlenmesi ve örnekleme frekansının 22.050 Hz olmasıdır. Özniteliklerin çıkarılmasında Python dili ile

OpenSMILE kütüphanesi sürece dâhil edildi. Öznitelik kümesi olarak ise emolarge, IS13_ComParE, IS12_speaker, IS11_speaker, emobase2010, IS10_paraling, emobase ve IS09_emotion setleri kullanıldı. Toplamda 3 veri seti için 8 farklı öznitelik kümesi oluşturuldu. Böylece 24 farklı öznitelik seti elde edilmiş oldu. Bu öznitelik setleri 8 farklı sınıflandırıcı ile analiz edildi.

İlk olarak V1 veriseti üzerinde en iyi sonuç alınan 4 öznitelik seti için 8 farklı sınıflandırıcı ile elde edilen sonuçlar analiz edildi. Çalışmanın yukarıda bahsedilen ilk analizinde sadece MFCC öznitelikleri üzerinden sonuçlar elde edilmişti. Burada ise yine aynı veri seti üzerinden farklı öznitelik kümeleri ile sonuçlar alındı ve karşılaştırıldı. Bu analizde eğitim ve test verileri %80-20, %90-10 ve 10 kat çapraz doğrulama ile elde edilmiştir. Tüm analizlerin sonuçları Tablo 4.5'te listelenmektedir.

Tablo 4.5. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V1 Veriseti)*

		Sort+Std			Sort+Std+Resample			Sort+Std+Disc		
Öznitelik Seti	**Sınıflandırıcı**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**
Emolarge	DVM	**80,00**	79,75	76,73	80,00	86,08	88,25	80,00	78,48	75,21
(6.552)	KNN	70,00	72,15	69,85	80,00	67,09	71,37	67,50	70,89	72,38
	LDA	77,50	82,28	75,95	82,50	81,01	82,86	**82,50**	**84,81**	76,22
	LR	**80,00**	**87,34**	**77,73**	82,50	**88,61**	**90,29**	80,00	**84,81**	**77,74**
	KA	55,00	49,37	51,67	80,00	77,22	78,54	32,50	51,90	47,58
	NB	67,50	64,56	59,60	65,00	58,23	63,97	60,00	55,70	51,93
	EA	77,50	78,48	71,62	**85,00**	82,28	89,51	72,50	78,48	72,14
	RO	75,00	69,62	60,62	82,50	65,82	73,68	72,50	65,82	62,14
IS13	DVM	65,00	72,15	73,16	85,00	86,08	88,75	67,50	74,68	74,19
(6.373)	KNN	62,50	64,56	64,45	67,50	69,62	68,03	62,50	67,09	65,97
	LDA	75,00	81,01	75,96	85,00	75,95	76,98	75,00	**83,54**	78,77
	LR	**77,50**	**83,54**	**78,53**	**87,50**	**89,87**	**89,78**	**80,00**	**83,54**	**79,04**
	KA	55,00	58,23	54,24	82,50	72,15	78,53	57,50	60,76	50,65
	NB	52,50	45,57	48,09	57,50	54,43	64,96	45,00	46,84	46,79
	EA	70,00	75,95	74,68	82,50	86,08	88,24	67,50	75,95	74,69
	RO	70,00	64,56	69,83	80,00	69,62	75,47	70,00	63,29	64,97
IS11	DVM	67,50	69,62	70,86	**87,50**	87,34	88,49	67,50	73,42	73,41
(4.368)	K-En	60,00	65,82	62,93	72,50	69,62	70,62	62,50	67,09	61,64
	LDA	72,50	75,95	75,19	**87,50**	73,42	81,09	**77,50**	81,01	76,72
	LR	67,50	**82,28**	**77,47**	**87,50**	**88,61**	**90,56**	72,50	**82,28**	**79,02**
	KA	42,50	54,43	51,69	75,00	87,34	83,39	37,50	56,96	54,19
	NB	50,00	51,90	51,17	60,00	62,03	67,01	50,00	45,57	49,62
	EA	**80,00**	77,22	72,91	80,00	86,08	87,47	72,50	75,95	72,15
	RO	62,50	59,49	66,50	85,00	68,35	79,06	62,50	68,35	68,55
Emobase	DVM	75,00	70,89	70,33	87,50	86,08	89,01	**80,00**	72,15	69,83
2010	K-En	65,00	70,89	63,92	67,50	65,82	70,87	62,50	68,35	66,22
(1,582)	LDA	67,50	**74,68**	66,21	87,50	75,95	82,62	72,50	**75,95**	68,81
	LR	**80,00**	**74,68**	**72,63**	**90,00**	**91,14**	**90,29**	**80,00**	74,68	72,11
	KA	55,00	56,96	52,69	80,00	81,01	79,81	45,00	54,43	50,64
	NB	65,00	59,49	61,38	62,50	65,82	66,01	60,00	56,96	60,37
	EA	80,00	**74,68**	71,62	82,50	87,34	88,24	77,50	72,15	**72,89**
	RO	65,00	58,23	61,62	77,50	69,62	71,39	72,50	58,23	63,92

Tablo 4.5'e göre sonuçları incelediğimizde veri seti üzerinde sıralama ve standardizasyon ön işlemleri sonrası yapılan sınıflandırma da en iyi sonuç emo_large öznitelik kümesi ve %80-20 eğitim ve test verisi bölümlemesi ile elde edilmiştir. Başarı oranı %87,34 olup ilk analizdeki sadece MFCC kullanarak elde edilmiş en yüksek başarı oranı olan %85,1 değerinin üzerinde bir sonuç elde edilmiştir. Ayrıca emo_large toplamda 6.552 öznitelik içermektedir. Sıralama ve standardizasyon uygulanmış verisetine resample filtesi

uygulanması sonrasında yapılan sınıflandırma çalışmasında ise en iyi sonuç emobase2010 öznitelik setinde elde edilmiştir. Bu grupta ise en iyi sınıflandırıcı LR olup başarı oranı %91,14'tür. Ayrıca tüm öznitelik setlerinde resample işlemi sonrası LR sınıflandırıcısında doğruluk başarı oranında artış gözlenmiştir. Diğer bir filtre olarak da sıralama ve standardizasyon işlemi uygulanmış veriseti discretize işlemine tabii tutulmuştur. 391 veri seti kapsamında bu işlem önceki iki ön işleme göre yüksek başarı oranına sahip sınıflandırıcılarda ek bir performans artışı sağlamadığı gözlemlenmiştir. Bu kırılımda ise en yüksek başarı oranı emo_large verisetinde LR ve LDA'da aynı değer olacak şekilde gözlemlenmiştir (%84,81) (%80-20).

V1 analizinde en yüksek başarı oranına sahip LR sınıflandırıcısında kullanılan parametreler Tablo 4.6'da listelenmiştir.

Tablo 4.6. *Lojistik Regresyon Parametreleri*

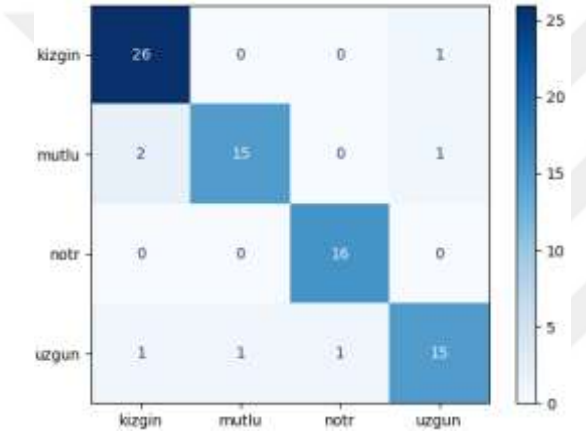
Parametre Listesi	**Kullanılan Değer**	**Parametre Aralığı**	**Varsayılan Değer**
penalty	L2	'L1', 'L2', 'elasticnet',none	L2
dual	False	True, False	False
tol	1e-4	float	1e-4
C	1.0	float	1.0
fit_intercept	True	True, False	True
intercept_scaling	1	float	1
class_weight	none	'Dict','balanced', none	none
random_state	**0**	Int RandomState instance, none	none
solver	**lbfgs**	'newton-cg', 'lbfgs', 'liblinear', 'saag', 'saga'	liblinear
max_iter	**1000**	int	100
multi_class	ovr	'ovr', 'multinomial', 'auto'	ovr
verbose	0	int	0
warm_start	false	True, false	false
n_jobs	None	Int, None	None
l1_ratio	None	Float, None	None

Tablo 4.6'da gösterilen parametre listesinde varsayılan değerden farklı kullanılan üç parametre bulunmaktadır. Bunlar random_state, solver ve max_iter dir. Random_state parametresi verileri karıştırırken kullanılan sözde rasgele sayının çekirdeğini temsil eder. Solver parametresi, bir optimizasyon problemi için hangi algoritmanın kullanılacağını belirtir. Diğer bir deyişle, bu parametre hangi optimizasyon yönteminin kullanılacağını gösterir. Örneğin, "lbfgs", "sag" veya "adam" gibi farklı algoritmaları

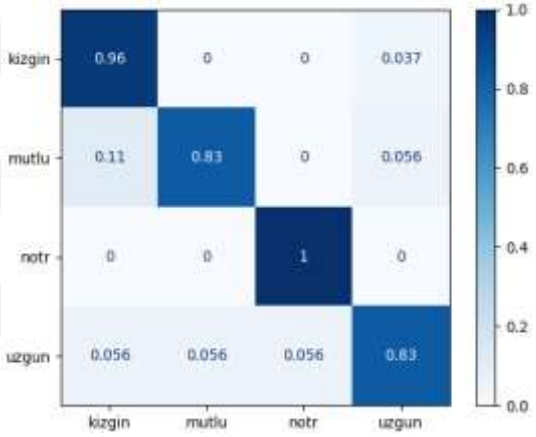
ifade edebilir. Max_iter parametresi ise çözücülerin yakınsamasını kontrol etmek için kullanılır. Bu parametre, çözücünün maksimum kaç kez yineleme yapacağını belirler. Yani çözümün istenen hassasiyete ne kadar sürede ulaşacağını belirler. Bu değeri artırarak veya azaltarak algoritmanın ne kadar süre çalışacağını ayarlayabilirsiniz.

Tablo 4.5'e göre en iyi sonucun (%91,14) emobase2010 öznitelik setinde %80-20 eğitim ve test verisi ayrımında, LR sınıflandırıcısı ile elde edildiği görülmektedir. Bu sonucun Karışıklık Matrisi Şekil 4.4 ve Şekil 4.5'te gösterilmiştir.

Şekil 4.4. Karışıklık Matrisi (LR-391Veriseti)



Şekil 4.5. Karışıklık Matrisi (LR-391Veriseti) (Normalize Edilmiş)



Şekil 4.4 ve Şekil 4.5'teki sonuçlara bakıldığında nötr duygusu en yüksek tanıma oranına sahip olup ve başarı yüzdesi %100'dür. Duygu tanıma da elde edilmiş en düşük başarı oranı ise mutlu ve üzgün duygusundadır ve değeri %83'tür. Genel sonuç ise %91,14'tür.

İkinci veri seti olan V2 veriseti üzerinde de en iyi sonuç alınan 4 öznitelik seti için yine 8 farklı sınıflandırıcı ile elde edilen sonuçların karşılaştırması yapıldı. V2 veriseti için sonuçlar Tablo 4.7'de listelenmiştir.

Tablo 4.7. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V2 Veriseti)*

		Sort+Std			Sort+Std+Resample			Sort+Std+Disc		
Öznitelik Seti	**Sınıflandırıcı**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**
Emolarge	DVM	**86,21**	86,13	81,32	90,80	87,86	92,11	**86,21**	**85,55**	79,58
	KNN	67,82	70,52	71,57	77,01	67,63	76,58	64,37	71,10	73,78
	LDA	75,86	78,61	80,27	89,66	86,13	88,17	79,31	78,61	83,29
	LR	82,76	**87,86**	**85,50**	**94,25**	88,44	**93,73**	79,31	84,39	**84,21**
	KA	47,13	56,07	53,24	85,06	78,03	83,18	52,87	46,24	52,91
	NB	57,47	60,69	59,51	62,07	64,16	64,03	62,07	61,85	59,05
	EA	73,56	73,41	74,71	87,36	84,97	90,02	74,71	74,57	75,41
	RO	55,17	60,12	59,04	62,07	67,05	65,89	57,47	58,96	60,78
IS13	DVM	77,01	73,99	76,68	91,95	86,71	90,71	75,86	71,68	75,40
	KNN	66,67	64,74	64,97	70,11	68,21	70,65	63,22	64,74	62,99
	LDA	75,86	79,19	76,32	85,06	83,82	83,52	75,86	79,77	78,30
	LR	**83,91**	**82,08**	**83,17**	**93,10**	**87,28**	**92,80**	**83,91**	**80,92**	**82,82**
	KA	48,28	57,23	51,63	83,91	79,77	78,88	51,72	54,91	52,32
	NB	57,47	52,60	52,67	57,47	59,54	58,12	45,98	42,77	48,84
	EA	71,26	68,79	71,22	89,66	84,39	89,21	72,41	69,94	72,04
	RO	57,47	57,23	61,59	77,01	72,83	66,82	60,92	56,07	61,71
IS11	DVM	75,86	72,83	75,98	89,66	86,13	89,55	77,01	72,25	75,52
	KNN	54,02	57,80	63,23	73,56	69,36	69,84	50,57	57,80	63,46
	LDA	71,26	72,25	72,26	86,21	79,77	83,29	78,16	73,99	75,16
	LR	**80,46**	**78,03**	**81,19**	**91,95**	**87,28**	**91,76**	**80,46**	**78,03**	**80,96**
	KA	52,87	49,71	53,83	85,06	74,57	80,61	56,32	50,29	50,35
	NB	50,57	49,13	52,67	56,32	60,69	61,36	49,43	47,40	48,14
	EA	70,11	69,36	72,27	90,80	81,50	89,33	71,26	70,52	72,85
	RO	58,62	59,54	64,04	71,26	70,52	67,75	60,92	61,27	61,83
Emobase	DVM	79,31	**76,30**	**80,05**	**93,10**	88,44	**90,84**	**79,31**	**76,30**	**77,74**
2010	KNN	65,52	62,43	67,87	77,01	69,36	70,65	63,22	63,58	65,66
	LDA	59,77	58,38	56,96	90,80	85,55	84,80	62,07	61,85	65,66
	LR	**80,46**	75,14	79,24	**93,10**	**90,75**	90,72	77,01	74,57	77,38
	KA	51,72	50,29	54,86	83,91	78,03	81,33	52,87	57,23	54,87
	NB	56,32	57,23	62,88	67,82	67,05	65,88	57,47	60,12	62,18
	EA	71,26	67,63	73,56	89,66	82,08	88,28	67,82	67.05	72.04
	RO	60.92	59.54	58.70	64.37	65.90	64.39	55.17	60.69	58.46

Tablo 4.7'de listelenen sonuçlar incelendiğinde sort ve standardizasyon ön işlemi sonrası yine en yüksek başarı oranı emo_large öznitelik setinde ve LR sınıflandırıcısı ile elde edilmiş olup değeri %87,86'dır. Bu değer %80-20 eğitim ve test verisi ayrımında elde edilmiştir. %90-10 kırılımında ise en yüksek başarı oranı DVM sınıflandırıcısı ile elde edilmiştir (%86,21). Ön işlem olarak sıralama, standardizasyon ve resample işlemi uygulanması sonrasında yapılan analizde ise en yüksek tanıma oranı emo_large öznitelik

setinde, %90-10 veri kırılımında ve LR sınıflandırıcısında elde edilmiştir (%94,25). %80-20 veri kırılımında ise en yüksek performans emobase2010 öznitelik setinde elde edilmiş ve LR yine en başarılı sınıflandırıcı olmuştur (%90,75). Ön işlem olarak sıralama, standardizasyon ve discretize işlemi uygulanmış veride yapılan analizde ise en yüksek tanıma oranı DVM sınıflandırıcısı ve emo_large öznitelik setinde elde edilmiş olup değeri %86,21'dir (Veri kırılımı %90-10). Son ön işlemde 10 kat çapraz doğrulama işlemi sonrası ise en yüksek başarı oranı LR sınıflandırıcısında elde edilmiştir (%84,21).

Genel olarak en yüksek başarı oranına sahip LR sınıflandırıcısında kullanılan parametreler 391 veriseti incelemesinde (Tablo 4.6) listelenmişti. Tablo 4.8'de ise diğer yüksek tanıma oranına sahip olan DVM sınıflandırıcısı için kullanılan parametreler listelenmiştir.

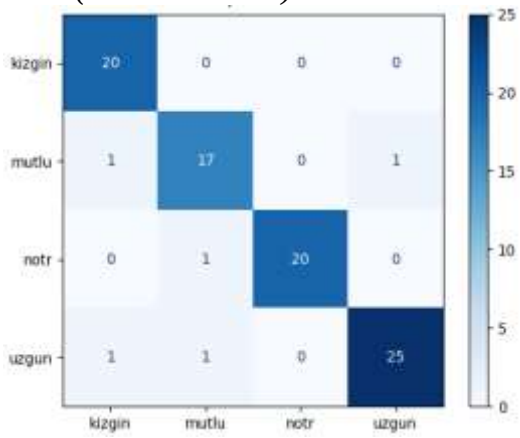
Tablo 4.8. *DVM Parametreleri*

Parametre Listesi	**Kullanılan Değer**	**Parametre Aralığı**	**Varsayılan Değer**
C	1.0	float	1.0
kernel	rbf	'linear', 'poly', 'rbf', 'sigmoid', 'precomputed' ya da callable	rbf
degree	3	int	3
Gamma	scale	'scale', 'auto' ya da float	scale
coef0	0.0	Float	0.0
Shrinking	True	True, False	True
Probability	False	True, False	False
Tol	1e-3	Float	1e-3
cache_size	200	Float	200
class_weight	None	dict or 'balanced'	None
Verbose	False	True, False	False
max_iter	-1	İnt	-1
decision_function_shape	Ovr	'ovo', 'ovr'	Ovr
break_ties	False	True, False	False
random_state	**0**	**Int or None**	**None**

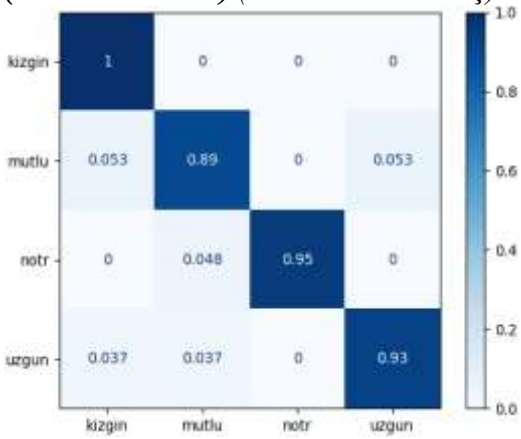
Tablo 4.8'de gösterilen parametre listesinde varsayılan değerden farklı kullanılan sadece bir parametre bulunmaktadır. Bu da random_state'dir. Random_state parametresi olasılık tahminleri için verileri karıştırmak için sözde rasgele sayı üretimini kontrol eder.

Tablo 4.7'ye göre en iyi sonucun (%94,25) emo_large öznitelik setinde %90-10 eğitim ve test verisi ayrımında, sıralama+Standardizasyon+Resample ön işlemi süreci ve LR sınıflandırıcısı ile elde edildiği görülmektedir. Bu sonucun Karışıklık Matrisi Şekil 4.6 ve Şekil 4.7'de gösterilmiştir.

Şekil 4.6. *Karışıklık Matrisi (LR-862Veriseti)*



Şekil 4.7. *Karışıklık Matrisi (LR-862Veriseti) (Normalize Edilmiş)*



Şekil 4.6 ve Şekil 4.7'deki sonuçlara bakıldığında kızgın duygusu en yüksek tanıma oranına sahip olup ve başarı yüzdesi %100'dür. Duygu tanıma da elde edilmiş en düşük başarı oranı ise mutlu duygusundadır ve değeri %89'dur. Genel sonuç ise %94,25'tir.

Üçüncü veriseti olan V3 veri setinde yine en iyi sonuç alınan 4 öznitelik seti için sonuçların karşılaştırılması Tablo 4.9'da gösterilmektedir.

Tablo 4.9. *Tüm Algoritmaların Sonuçları ve Karşılaştırılması (V3 Veriseti)*

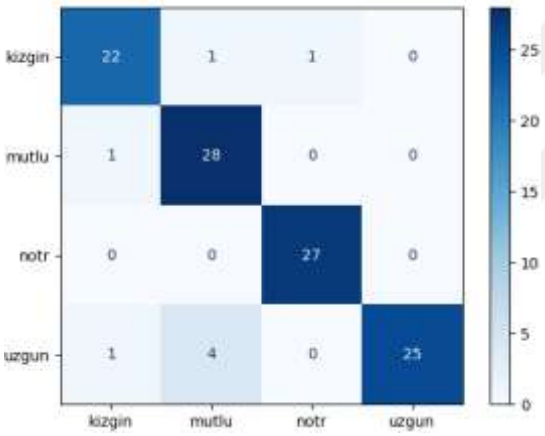
		Sort+Std			Sort+Std+Resample			Sort+Std+Disc		
Öznitelik Seti	**Sınıflandırıcı**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**	**90-10 %**	**80-20 %**	**10 Kat**
Emolarge	DVM	80,00	81,36	79,53	**88,18**	**86,36**	90,53	80,00	82,27	77,44
	KNN	70,91	69,55	71,70	70,00	72,27	74,70	70,00	71,36	70,97
	LDA	**86,36**	79,09	73,53	81,82	81,36	86,72	84,55	85,00	77,98
	LR	85,45	**86,82**	**80,44**	**88,18**	**86,36**	**92,63**	**85,45**	**86,36**	**80,25**
	KA	53,64	55,45	51,05	75,45	75,00	79,43	50,91	50,91	50,77
	NB	56,36	57,27	52,96	50,91	52,27	56,96	57,27	55,45	53,77
	EA	77,27	78,64	74,43	87,27	87,27	87,99	73,64	75,00	72,88
	RO	60,91	64,55	53,96	59,09	63,64	59,24	54,55	61,82	53,59
IS13	DVM	82,73	77,73	75,61	**89,09**	**90,45**	88,90	82,73	77,73	73,70
	KNN	70,91	61,36	60,23	54,55	62,27	65,24	70,00	62,73	59,96
	LDA	78,18	75,00	74,62	80,91	77,73	80,80	80,91	78,64	76,98
	LR	**83,64**	**81,82**	**76,89**	**89,09**	**90,45**	**90,45**	**86,36**	**81,36**	**78,26**
	KA	51,82	53,18	50,96	72,73	77,27	79,17	55,45	45,45	47,77
	NB	54,55	51,82	49,68	54,55	50,00	53,50	48,18	44,09	46,22
	EA	79,09	72,73	68,97	86,36	88,64	87,08	76,36	73,18	69,61
	RO	67,27	60,00	57,14	62,73	59,09	61,42	68,18	58,64	54,87
IS11	DVM	**84,55**	76,82	74,98	**89,09**	**90,45**	88,35	**81,82**	76,36	73,70
	KNN	68,18	60,00	61,14	58,18	65,45	68,97	66,36	58,18	59,87
	LDA	71,82	70,91	68,43	77,27	79,09	81,90	78,18	73,64	72,89
	LR	81,82	**81,82**	**77,62**	88,18	87,73	**89,81**	**81,82**	**81,36**	**78,16**
	KA	55,45	47,27	48,77	73,64	80,00	80,71	54,55	51,36	47,05
	NB	50,91	47,27	47,95	51,82	50,45	54,69	47,27	40,45	44,22
	EA	77,27	74,55	70,06	85,45	88,64	86,99	78,18	69,55	70,52
	RO	69,09	62,27	56,41	62,73	65,00	62,43	66,36	61,36	57,06
Emobase 2010	DVM	**81,82**	**79,09**	**78,07**	90,00	**90,45**	89,35	**81,82**	**79,55**	**75,16**
	KNN	72,73	65,45	66,06	68,18	70,45	71,61	71,82	66,36	65,06
	LDA	49,09	56,82	46,67	84,55	84,55	84,26	66,36	61,36	57,06
	LR	72,73	74,09	74,98	**92,73**	**90,45**	**89,72**	73,64	73,64	74,43
	KA	54,55	51,36	48,77	77,27	79,55	81,89	52,73	56,36	49,87
	NB	69,09	60,45	55,69	60,00	60,91	61,69	66,36	59,55	55,05
	EA	75,45	73,64	70,88	87,27	88,64	86,99	80,91	73,18	71,07
	RO	65,45	62,73	53,05	60,91	61,82	59,70	67,27	60,91	53,51

Tablo 4.9'a göre sonuçlar incelendiğinde sort ve standardizasyon filtresi sonrası en iyi sonuç yine emo_large öznitelik setinde elde edilmiştir. %80-20 eğitim ve veri seti bölümlemesinde en yüksek değer %86,82 olup LR ile elde edilmiştir. İlk kısmında 4 öznitelik seti bazında genel olarak incelendiğinde ise LR ve DVM sınıflandırıcılarında tanıma oranı yüksektir. Ön işlem olarak sıralama, standardizasyon ve resample işlemi uygulanması sonrasında yapılan analizde ise en yüksek tanıma oranı emobase2010

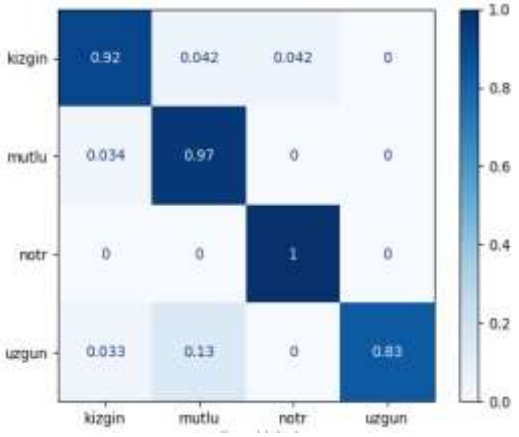
öznitelik setinde, %90-10 veri kırılımında ve LR sınıflandırıcısında elde edilmiştir (%92,73). Ön işlem olarak sıralama, standardizasyon ve discretize işlemi uygulanmış veride yapılan analizde ise en yüksek tanıma oranı LR Sınıflandırıcısı ile elde edilmiştir. Emo_large ve IS13_ComParE öznitelik setinin ikisinde de aynı sonuç elde edilmiştir (%86,36).

V3 veri seti kapsamında en yüksek tanıma oranına %92,73 ile emobase2010 öznitelik seti, LR sınıflandırıcısı, %90-10 eğitim test veri kümelemesi ile ulaşılmıştır. Tablo 4.6'da daha önce LR için kullanılan parametreler paylaşılmıştı. Elde edilen iyi sonuç için Karışıklık Matrisi Şekil 4.8 ve Şekil 4.9'da gösterilmiştir.

Şekil 4.8. *Karışıklık Matrisi (LR-1099 Veriseti)*



***Şekil* 4.9.** Karışıklık Matrisi (LR-1099 Veriseti) (Normalize Edilmiş)



Şekil 4.8 ve Şekil 4.9'daki sonuçlara bakıldığında nötr duygusu en yüksek tanıma oranına sahip olup ve başarı yüzdesi %100'dür. Duygu tanıma da elde edilmiş en düşük başarı oranı ise üzgün duygusundadır ve değeri %83'tür. Genel sonuç ise %92,73'tür.

Son analiz ise DL Algoritmaları ile yapılmıştır. Burada tek öznitelik seti kullanıldı (emo_large). Sonuçlar hem kendi arasında hem de ML algoritmaları ile karşılaştırılmıştır. Tablo 4.10'da detayı listelenmektedir.

Tablo 4.10. *ML ve DL Algoritmaları Sonuçların Karşılaştırılması*

			Sort			Resample		
Sınıflandırıcı	**Veri Seti**	**Öznitelik Adı**	**%90-10**	**%80-20**	**10-Kat**	**%90-10**	**%80-20**	**10-Kat**
DNN	862	emo_large	82,76	85,55	82,25±5,16	91,95	**90,17**	**92,57±3,30**
CNN (1D)	862	emo_large	**86,21**	82,08	77,83±3,11	88,51	83,24	**90,25±1,82**
DVM	862	emo_large	**86,21**	86,13	81,32±4,92	90,80	87,86	92,11±3,93
KNN	862	emo_large	67,82	70,52	71,57±4,03	77,01	67,63	76,58±4,63
LDA	862	emo_large	75,86	78,61	80,27±4,74	89,66	86,13	88,17±2,63
LR	862	emo_large	82,76	**87,86**	**85,50±2,92**	**94,25**	88,44	**93,73±3,05**
KA	862	emo_large	47,13	56,07	53,24±4,88	85,06	78,03	83,18±2,70
NB	862	emo_large	57,47	60,69	59,51±4,32	62,07	64,16	64,03±4,04
EA	862	emo_large	73,56	73,41	74,71±3,75	87,36	84,97	90,02±3,73
RO	862	emo_large	55,17	60,12	59,04±5,82	62,07	67,05	65,89±3,84
DNN	1.099	emo_large	82,73	85,91	**81,97±3,97**	**90,91**	**89,55**	91,81±2,99
CNN (1D)	1.099	emo_large	83,64	82,27	76,88±3,10	87,27	85,91	88,81±3,02
DVM	1.099	emo_large	80,00	81,36	79,53±3,12	88,18	86,36	90,53±2,10
KNN	1.099	emo_large	70,91	69,55	71,70±2,92	70,00	72,27	74,70±3,56
LDA	1.099	emo_large	**86,36**	79,09	73,53±3,51	81,82	81,36	86,72±3,40
LR	1.099	emo_large	85,45	**86,82**	80,44±3,28	88,18	86,36	**92,63±3,08**
KA	1.099	emo_large	53,64	55,45	51,05±2,71	75,45	75,00	79,43±2,77
NB	1.099	emo_large	56,36	57,27	52,96±3,98	50,91	52,27	56,96±5,37
EA	1.099	emo_large	77,27	78,64	74,43±3,43	87,27	87,27	87,99±1,53
RO	1.099	emo_large	60,91	64,55	53,96±2,42	59,09	63,64	59,24±7,30

Tablo 4.10'a göre sonuçları incelediğimizde DNN modelinde en yüksek tanıma oranı %92,57 dir (±3,30 standart sapma). CNN modelinde ise %90,25 (±1,82 standart sapma) ile en yüksek değere ulaşılmıştır. Her iki modelinde en iyi sonuç aldığı veriseti 862 kayıttan oluşmakta ve 10 kat çapraz doğrulama sonrası bu değerler elde edilmiştir.

Genel olarak ML ve DL modellerini karşılaştırdığımızda ise bazı ML algoritmalarının DL Algoritmalarına göre daha iyi tanıma oranına sahip olduğu görülmektedir. Ancak DNN ve CNN sonuçları da genel olarak sıralama ön işlemi için %80 ve üzeri; resample ön işlemi için ise %90 sınırlarında gözlemlenmektedir. Tablo 4.10'a göre sıralama ve standardizasyon için en yüksek tanıma oranı LR ile elde edilmiştir (%87,86). Resample ve standardizasyon için ise yine en yüksek tanıma oranı LR sınıflandırıcısındadır (%94,25).

4.2. Tartışma

Çalışma kapsamında elde edilen tüm sonuçlar incelendiğinde ML algoritmaları kısmında en yüksek tanıma oranı %94,25 ile LR ile elde edilmiştir. Kızgın duygusu en yüksek

tanıma oranına sahip olup başarı yüzdesi %100'tür. En düşük tanıma oranı ise %89 ile mutlu duygusundadır. Derin öğrenme tarafında ise en yüksek tanıma oranı %92,57 ile DNN algoritmasında elde edilmiştir. Her iki değerde emo_large öznitelik setinde ve 862 kayıtlık veri setinde alınan sonuçlardır.

Tablo 4.11'de konuşmadan duygu analizinde yapılan benzer çalışmaların karşılaştırma tablosu yer almaktadır.

Tablo 4.11. *Önceki Çalışmalar ile Karşılaştırmalar*

Çalışma	VeriSeti	Öznitelikler	Sınıflandırıcı	Tanıma Oranı
S. Gokalp and I. Aydin (2021)	TESS RAVDESS	MFCC, spektrogram	KA, DVM, AE	98,00%
S. Langari et al (2020)	EmoDB SAVEE PDREC	DFrFT, Kepstral öznitelikler	2D CNN	97,57%
D.Issa et al (2020)	RAVDESS, EMO-DB IEMOCAP	Kromagram, MFCCs, Mel ölçekli spektrogram, Tonnetz gösterimi, Spektral kontrast özelliği	1D CNN	96,34%
H.A. Abdulmohsin et al (2021)	EmoDB RAVDESS SAVEE	Enerji, Harmonik Oran, ZC, Entropi, MFCC, GTCC, Perde, Ses Yüksekliği, Uygunluk, Fourier, Haar, Enerji sapması, Sıfır geçiş sapması	Sinir Ağı	96,30%
J. Ancilin and A. Milton. (2021)	Urdu eNTERFACE Berlin EMOVO Ravdess SAVEE	MFCC,MFMC,LFPC, LPCC	DVM	95,25%
L.Guo et al (2022)	EmoDB IEMOCAP	MGDCC, DRP, RP, büyüklük spektrogramı	KELM	94,02%
N. Senthilkumar et al (2021)	EmoDB RAVDESS	Spektrogram	BLSTM	86,00%
S. Ntalampiras (2021)	EmoDB	Log-mel spektrogramı, Zamansal modülasyon	Siamese Neural Network	82,10%
K. Wang et al (2020)	EmoDB	MFCC,WPC,SFFS	DVM	79,50%

Tablo 4.11. *Önceki Çalışmalar ile Karşılaştırmalar (Devamı)*

Çalışma	VeriSeti	Öznitelikler	Sınıflandırıcı	Tanıma Oranı
N. Jia and C. Zheng (2021)	Özel SER	MFCC	BLSTM	78,00%
H. Aouani and Y.B. Ayed (2020)	RML	TEO, 39 MFCC, HNR, ZCR	DVM	74,07%
Z. Zhao et al (2021)	IEMOCAP	3 boyutlu log-mels spektrogramı	1D evrişimli modüllere sahip bir TCN, LSTM,BLSTM	66,1%
Y. Pourebrahim et al (2021)	IEMOCAP FAUAEC PESD EmoDB RAVDESS	-	AE	61,60%
Z. Yao et al (2020)	IEMOCAP	-	DNN,CNN,RNN	58,30%
Önerilen Yöntemimiz (ML)	**Veri Setimiz**	**Emo_large Öznitelik Seti**	**DVM, LR, KNN, KA, GNB, LDA, ADA, GB, EA, XGB, RO**	94,25 %
Önerilen Yöntemimiz (DL)	**Veri Setimiz**	**Emo_large Öznitelik Seti**	**DNN, CNN(1D)**	92,57%

Tablo 4.11'de görülebileceği gibi, son yıllarda konuşmadan duygu tanıma konusunda çok sayıda çalışma yapılmıştır. Çeşitli sınıflandırıcıların eğitiminde farklı veri setlerinden elde edilen farklı özellikler kullanılmış ve sonuçlar tabloda verildiği gibi olmuştur. Elde edilen sonuçlar incelendiğinde, çalışmanın amacına göre hazırlanan hazır veri setlerinde (konuşma duygularının tanınması) tanıma oranları oldukça yüksektir. Tez kapsamında özellikle emo_large öznitelik kümesinde yapılan çalışmamızda, özgün Türk filmleri veri seti üzerinde farklı sınıflandırıcılarla testler gerçekleştirildi. Bu testler sonucunda yüksek bir sınıflandırma başarısı elde edildi. Türk dili için böyle bir çalışmanın yapılması ve sonuç olarak yüksek bir performans elde edilmesi oldukça değerlidir. Bu alanda daha fazla veri kullanıldığı takdirde performansın artacağı da düşünülmektedir.

4.3. Sonuç

Çalışma çerçevesinde, çeşitli makine ve derin öğrenme algoritmaları kullanılarak konuşmadan duygu tanıma analizi yapılmıştır. Yapılan çalışmayı literatürdeki incelenen çalışmalardan ayıran en önemli özelliklerden biri yüksek tanıma oranına sahip ve konuşmadan duygu analizinde sıklıkla kullanılan hazır ve bu amaca yönelik oluşturulmuş verisetlerinin tercih edilmemesidir. Tez kapsamında hazırlamış olduğumuz kendi verisetimiz kullanıldı. Hazırlanan veriseti konuşmacıdan ve metinden bağımsız birçok örnekten oluşmaktadır. Tanıma oranı konuşmacı bağımsız modeller ile yapılan önceki çalışmalara göre oldukça iyidir.

MFCC öznitelikleri ile en yüksek tanıma oranı DVM sınıflandırıcısı ile elde edildi. İlk analizde kızgın ve nötr duygular en başarılı, mutlu duygusu ise en düşük tanıma oranına sahiptir. Genel olarak da yapılan çalışmalarda mutlu duygusunun tanıma oranları düşük olmaktadır.

OpenSMILE ile elde edilen öznitelikler kullanıldığında ise V1, V2 ve V3 verisetlerinde LR sınıflandırıcısı ile yüksek tanıma oranına ulaşılmıştır. OpenSMILE ile yapılan analizlerde ise 3 verisetinde farklılık olmakla beraber nötr ve kızgın duygusu daha yüksek tanıma oranına sahipken mutlu ve üzgün duygusu daha düşük değerlere sahiptir.

Genel sonuç olarak ise öznitelik sayısının artması başarı oranını artırmıştır. Veriseti örnek sayısının artırılması başarıyı doğrusal artıran bir durum olmamıştır. Buna etki başka faktörlerde bulunmaktadır. Ön İşleme sürecinde kullanılan filtreler performansı olumlu etkilemektedir. Öznitelik seçme sürecinin her problem ve veriseti için kullanılmasının her zaman olumlu etki oluşturmadığı da gözlemlenmiştir. Parametre optimizasyonu da yine performansı olumlu etkileyen bir durum olmuştur.

Derin öğrenme tarafında ise DNN ve CNN (1D) modelleri analiz edilmiş. DNN modeli 128 nörona sahip bir girdi katmanı, 64 nörondan oluşan 5 gizli katman ve 4 nörona sahip çıkış katmanından oluşmaktadır. DNN model bu yapı ile en yüksek tanıma oranına ulaşmıştır.

Sonraki çalışmalarda veri setindeki örnek sayısını artırmak ve derin öğrenme tarafındaki analizi daha da derinleştirmek adına çalışmalar yapmayı planlıyoruz.

KAYNAKÇA

[1] S. Gökalp ve A. İlhan, "Comparison of Different Deep Neural Network Models in Emotion Recognition" Journal of Muş Alpaslan University Faculty of Engineering and Architecture, no. 2(1), ss. 35-43, 2021.

[2] E. Çolakoğlu, S. Hızlısoy ve R. S. Aslan, "T-ser: An Efficient Speech Emotional Recognition Model For Turkish Language Based On Machine Learning Algorithms" Innovations And Technologies In Engineering, Eğitim Yayınevi, 2022, ss. 106-127.

[3] S. Hızlısoy, E. Çolakoğlu ve R. S. Arslan, "Speech-to-Gender Recognition Based on Machine Learning Algorithms" International Journal Of Applied Mathematics Electronics And Computers, no. 10(4), ss. 084-092, 2022.

[4] S. Hızlısoy ve Z. Tüfekçi, "Emotion Recognition From Turkish Music" European Journal of Science and Technology, Cilt %1 / %26-12, no. Ejosat Special Issue 2020 (ICCEES), 2020.

[5] C. Parlak ve B. Diri, "Farklı Veri Setleri Arasında Duygu Tanıma Çalışması" Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik dergisi, cilt 16, no. 48, ss. 21-29, 2014.

[6] A. Tepecik ve E. Demir, "Türkçe Ses Kayıt Verilerinin CountVectorizer ve TF-IDFVectorizer Yöntemleri ile BERT Modelleri Olarak Google Colab Platformunda ve RapidMiner'da Makine Öğrenmesi Algoritmalarıyla Analizi" Fırat Üniversitesi Fen Bilimleri Dergisi, cilt 34(1), ss. 19-29, 2022.

[7] T. Özseven, "Konuşma Tabanlı Duygu Tanımada Ön İşleme ve Öznitelik Seçim Yöntemlerinin Etkisi" DÜMF Mühendislik Dergisi, cilt 10:1, ss. 99-112, 2019.

[8] Ç. M. Bakırcı, "Evrim Ağacı" 22 05 2018. [Çevrimiçi]. Available: https://evrimagaci.org/plutchikin-duygu-carkifelegi-duygularinizin-kokenlerini-daha-yakindan-taniyin-5407. [Erişildi: 31 05 2023].

[9] Ö. Koçak, "Psikoloji Yazıları" 29 04 2021. [Çevrimiçi]. Available: https://psikolojiyazilari.com/robert-plutchikin-duygu-cemberi-kavrami/. [Erişildi: 19 01 2023].

[10] L. Güner ve İ. Ergenç, "Sesin Doğasi Ve Oluşumu".

[11] E. Erdem, “Ses Sinyallerinde Duygu Tanima Ve Geri Erişimi” Başkent Üniversitesi Fen Bilimleri Enstitüsü, 2014.

[12] G. Işık, “Türkçe Ağizlarin Taninmasinda Derin Öğrenme Tekniğinin Kullanilmasi” Hacettepe Üniversitesi, Ankara, 2019.

[13] “İnnova” 29 09 2021. [Çevrimiçi]. Available: https://www.innova.com.tr/tr/blog/buyuk-veri-blog/egitim-verileri-makine-ogreniminde-nasil-kullanilir. [Erişildi: 12 06 2023].

[14] O. E. Korkmaz, “Ses Sinyalinden Duygu Tanima” Trabzon, 2016.

[15] A. Oğuzlar, “Veri Ön İşleme” Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, no. 21, ss. 67-76, 2003.

[16] “OpenSMILE Dökümantasyonu” [Çevrimiçi]. Available: https://audeering.github.io/opensmile/index.html. [Erişildi: 13 01 2023].

[17] A. Gümüşçü, “Yüksek Boyutlu Veri Kümelerindeki Özniteliklerin Hibrit Yöntemle Seçilmesi” Şanlıurfa, 2016.

[18] S. Ray, “A Quick Review of Machine Learning Algorithms” 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, Faridabad, India, 2019.

[19] Y. Ülgen Sönmez, “Konuşma Seslerinden Duygulari Tanimak İçin Yerel İkili Ve Üçlü Örüntülere Dayali Yeni Bir Modelin Geliştirilmesi” Firat Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Elazığ, 2021.

[20] A. Alan ve M. Karabatak, “Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi” Fırat Üniversitesi Mühendislik Bilimleri Dergisi, cilt 32(2), ss. 531-540, 2020.

[21] P. Agnihotri, “EmoDB Veri Kümesi” Kaggle, [Çevrimiçi]. Available: https://www.kaggle.com/datasets/piyushagni5/berlin-database-of-emotional-speech-emodb. [Erişildi: 09 06 2023].

[22] K. Donuk ve D. Hanbay, “Konuşma Duygu Tanıma için Akustik Özelliklere Dayalı LSTM Tabanlı Bir Yaklaşım” Journal of Computer Science, cilt 2, no. 7, ss. 54-67, 2022.

[23] L. Guo, L. Wang, J. Dang, E. S. Chng ve S. Nakagawa, "Learning affective representations based on magnitude and dynamic relative phase information for speech emotion recognition" Speech Communication, no. 136, ss. 118-127, 2022.

[24] Z. Zhao, Q. Li, Z. Zhang, N. Cummins, H. Wang, J. Tao ve B. W. Schuller, "Combining a parallel 2D CNN with a self-attention Dilated Residual Network for CTC-based discrete speech emotion recognition" Neural Networks, no. 141, ss. 52-60, 2021.

[25] N. Senthilkumar, S. Karpakam, M. Gayathri Devi, R. Balakumaresan ve P. Dhilipkumar, "Speech emotion recognition based on Bi-directional LSTM architecture and deep belief networks" Materials Today: Proceedings., 2021.

[26] H. A. Abdulmohsin, H. B. Abdul Wahab ve A. M. J. Abdul hossen, "A new proposed statistical feature extraction method in speech emotion recognition" Computers and Electrical Engineering, no. 93, 2021.

[27] S. Ntalampiras, "Speech emotion recognition via learning analogies" Pattern Recognition Letters, no. 144, ss. 21-26, 2021.

[28] S. Hızlısoy ve Z. Tüfekçi, "Türkçe Müzikten Duygu Tanıma" Avrupa Bilim ve Teknoloji Dergisi, Cilt %1 / %26-12, no. Ejosat Özel Sayı 2020 (ICCEES), 2020.

[29] Z. Yao, Z. Wang, W. Liu, Y. Liu ve J. Pan "Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN, MS-CNN and LLDRNN" Speech Communication, no. 120, ss. 11-19, 2020.

[30] S. Langari, H. Marvi ve M. Zahedi, "Efficient speech emotion recognition using modified feature extraction" Informatics in Medicine Unlocked, no. 20, 2020.

[31] N. Jia ve C. Zheng, "Two-level discriminative speech emotion recognition model with wave field dynamics: A personalized speech emotion recognition method" Computer Communications, no. 180, ss. 161-170, 2021.

[32] J. Ancilin ve A. Milton, "Improved speech emotion recognition with Mel frequency magnitude coefficient" Applied Acoustics, no. 179, 2021.

[33] A. Bhavan, P. Chauhan, Hitkul ve R. R. Shah, "Bagged support vector machines for emotion recognition from speech" Knowledge-Based Systems, no. 184, 2019.

[34] K.-Y. Huang, C.-H. Wu ve M.-H. Su, "Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses" Pattern Recognition, no. 88, ss. 668-678, 2019.

[35] D. Issa, M. F. Demirci ve A. Yazici, "Speech emotion recognition with deep convolutional neural networks" Biomedical Signal Processing and Control, no. 59, 2020.

[36] H. Aouani ve Y. B. Ayed, "Speech emotion recognition with deep learning" Procedia Computer Science, no. 176, ss. 251- 260, 2020.

[37] K. Wang, G. Su, L. Liu ve S. Wang, "Wavelet packet analysis for speaker-independent emotion recognition" Neurocomputing, no. 398, ss. 257-264, 2020.

[38] Y. Pourebrahim, F. Razzazi ve H. Sameti, "Semi-supervised parallel shared encoders for speech emotion" Digital Signal Processing, no. 118, 2021.

[39] Z. Zhao, Z. Bao, Z. Zhang, N. Cummins, S. Sun, H. Wang, J. Tao ve B. W. Schuller, "Self-attention transfer networks for speech emotion recognition" Virtual Reality & Intelligent Hardware, no. 3, ss. 43-54, 2021.

[40] O. Atila ve A. Şengür, "Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition" Applied Acoustics, no. 182, 2021.

[41] C. Parlak ve B. Diri, "İnsan Sesinden Duygu Çıkarma" Sinyal İşleme ve Uygulamaları Kurultayı, 2013.

[42] Y. Pan, P. Shen ve L. Shen, "Speech emotion recognition using support vector machine" International Journal of Smart Home, cilt 2, no. 6, ss. 101-108, 2012.

[43] G. Polat ve H. Altun, "Ses Öznitelik Vektörlerinin Duygusal Durum Sınıflandırılmasında Kullanımı" 2008.

[44] H. M. Fayek, M. Lech ve L. Cavedon, "Evaluating Deep learning architectures for Speech Emotion Recognition" Neural Networks, no. 92, ss. 60-68, 2017.

[45] G. Işık ve H. Artuner, "Turkish dialect recognition in terms of prosodic by long short-term memory neural networks" Journal of the Faculty of Engineering and Architecture of Gazi University, no. 35:1, ss. 213-224, 2020.

[46] Ö. Eskidere ve F. Ertaş, "Mel Frekansı Kepstrum Katsayılarındaki Değişimlerin Konuşmacı Tanımaya Etkisi" Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi Cilt 14, Sayı 2, 2009.

[47] S. Karasartova, "Metinden Bağımsız Konuşmacı Tanıma Sistemlerinin İncelenmesi ve Gerçekleştirilmesi" Ankara, 2011.

[48] Ö. Eskidere ve F. Ertaş, "The Effects of Filter Frequency Scale Variability On Speaker Identification Performance" Journal of Engineering and Natural Sciences, ss. 197-207, 09 2009.

[49] H. Perez ve A. A. Torres, "Evaluation of quantitative and qualitative features for the acoustic analysis of domestic dogs' vocalizations" Journal of Intelligent & Fuzzy Systems, p. 5051–5061, 2019.

[50] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro ve K. R. Scherer, "On the acoustics of emotion in audio: what speech, music, and sound have in common" Frontiers in Psychology, no. 4, 2013.

[51] B. Schuller, S. Steidl, A. Batliner, F. Schiel ve J. Krajewski, "The INTERSPEECH 2011 Speaker State Challenge" Florence, Italy, 2011.

[52] T. Ş. Acuner, "Sinyal İşleme Yöntemleri Tabanli Keman Ses Analizi İçin Bir Kullanici Arayüz Gerçekleştirilmesi" Abant İzzet Baysal Üniversitesi, Bolu, 2017.

[53] M. B. Er, "Makine Öğrenmesi İle Türk Müziğinde Duygu Analizi" Maltepe Üniversitesi, İstanbul, 2019.

[54] D. McEnnis, C. McKay, I. Fujinaga ve P. Depalle, "Jaudio: A Feature Extraction Library" 6th International Conference on Music Information Retrieval, London, UK, 11-15 September 2005, Proceedings, 2005.

[55] S. Mozaffari, "The Power of Resampling: Techniques for Boosting Your ML Model Performance" 19 03 2023. [Çevrimiçi]. Available: https://www.linkedin.com/pulse/power-resampling-techniques-boosting-your-ml-model-sadaf-mozaffari/.[Erişildi: 09 03 2023].

[56] Editör Ekibi, "Matematik ile Makine Öğrenimi Standardizasyonu (Z-Score Normalization)" Yapay Zekaya Yönelik — Multidisipliner Bilim Dergisi, 2020.

[57] V. Bolón, N. Sánchez ve A. Alonso, "A review of feature selection methods on synthetic data" 2012, p. 483–519.

[58] "Python'da SelectKBest Özellik Seçimi Örneği" 2021. [Çevrimiçi]. Available: https://www.datatechnotes.com/2021/02/seleckbest-feature-selection-example-in-python.html. [Erişildi: 14 04 2023].

[59] “Python'da SelectFromModel Özellik Seçimi Örneği” 2021. [Çevrimiçi]. Available: https://www.datatechnotes.com/2021/04/selectfrommodel-feature-selection.html. [Erişildi: 03 04 2023].

[60] J. Brownlee, “Python'da Özellik Seçimi için Özyinelemeli Özellik Eleme (RFE)” 2020. [Çevrimiçi]. Available: https://machinelearningmastery.com/rfe-feature-selection-in-python/. [Erişildi: 23 05 2023].

[61] N. Yüksel, “Üniversite Öğrencilerinin Yurtlari Tercih Etme Nedenlerinin Lojistik Regresyon Analizi İle Belirlenmesi: Kirşehir Ahi Evran Üniversitesi Örneği” Sivas Cumhuriyet Üniversitesi, Sivas, 2019.

[62] S. Şenel ve B. Alatlı, “Lojistik Regresyon Analizinin Kullanıldığı Makaleler Üzerine Bir İnceleme” Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, cilt 5, no. 1, ss. 35-52, 2014.

[63] D. Anggraeni, W. M. Sanjaya, M. Munawwaroh, M. Y. S. Nurasyidiek ve I. P. Santika, “Control of Robot Arm based on Speech Recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and K-Nearest Neighbors (KNN) Method” 2017 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA), 2017.

[64] Ö. Şahinaslan, H. Dalyan ve E. Şahinaslan, “Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi” Bilişim Teknolojileri Dergisi, cilt 15, no. 2, 2022.

[65] I. H. Sarker, “Machine Learning Algorithms, Real-World Applications and Research Directions” SN Computer Science, no. 160, 2021.

[66] S. Emanet, G. K. Baydogmus ve Ö. Demir, “Öznitelik Seçme Yöntemlerinin Makine Öğrenmesi Tabanlı Saldırı Tespit Sistemi Performansına Etkileri” Dicle University Journal of Engineering, no. 12:5, ss. 743-755, 2021.

[67] E. R. Muratlar, “Gradient Boosted Regression Trees” 2020. [Çevrimiçi]. Available: https://www.veribilimiokulu.com/gradient-boosted-regresyon-agaclari/.[Erişildi: 19 05 2023].

[68] K. Thankachan, “What? When? How?: ExtraTrees Classifier” Aug 2022. [Çevrimiçi]. Available: https://towardsdatascience.com/what-when-how-extratrees-classifier-c939f905851c. [Erişildi: 13 03 2023].

[69] E. R. Muratlar, "XGBoost Nasıl Çalışır? Neden İyi Performans Gösterir?" 2020. [Çevrimiçi]. Available: https://www.veribilimiokulu.com/xgboost-nasil-calisir/.[Erişildi: 16 03 2023].

[70] E. Akdağlı, "Makine Öğrenmesinde Random Forest Algoritması" 4 Mar 2021. [Çevrimiçi]. Available: https://ece-akdagli.medium.com/makine-%C3%B6%C4%9Frenmesinde-random-forest-algoritmas%C4%B1-a79b044bbb31. [Erişildi: 21 02 2023].

[71] Goelaparna1520, "Convolutional Neural Network (CNN) in Machine Learning" 2023. [Çevrimiçi]. Available: https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/.[Erişildi: 27 03 2023].

[72] Aishwarya, "Introduction to Recurrent Neural Network" 2023. [Çevrimiçi]. Available: https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/. [Erişildi: 17 05 2023].

[73] aakarsha_chugh, "Deep Learning | Introduction to Long Short Term Memory" 2023. [Çevrimiçi]. Available: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/. [Erişildi: 07 05 2023].

[74] E. Şirin,"Veri Bilimi Okulu" [Çevrimiçi]. Available: https://www.veribilimiokulu.com/hata-matrisini-confusion-matrix-yorumlama/.[Erişildi: 17 01 2023].

[75] S. Gökalp ve İ. Aydın, "Farklı Derin Sinir Ağı Modellerinin Duygu Tanımadaki Performanslarının Karşılaştırılması" Muş Alparslan Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, no. 2(1), ss. 35-43, 2021.

[76] B. Schuller, S. Steidl, A. Batliner, E. Noth, A. Vinciarelli, F. Burkhardt, R. v. Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi ve B. Weiss, "The INTERSPEECH 2012 Speaker Trait Challenge" Proceedings INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association, 2012.

[77] W. Styler, Using Praat for Linguistic Research, 2022.

[78] Q. Kong, T. Siauw ve A. Bayen, "24.1 The Basics of Waves | Contents | 24.3 Fast Fourier Transform (FFT)" Python Programming and Numerical Methods - A Guide for Engineers and Scientists, Elsevier, 2020.

[79] “Ne Demek”[Çevrimiçi]. Available: https://www.nedemek.page/kavramlar/ayr%C4%B1k%20fourier%20d%C3%B6n%C3%BC%C5%9F%C3%BCm%C3%BC. [Erişildi: 19 03 2023].

[80] O. Lartillot, T. Eerola ve P. Toiviainen, “A Matlab Toolbox for Music Information Retrieval” Data Analysis, Machine Learning and Applications - Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg,, 2007.

[81] S. Jain, S. Shukla ve R. Wadhvani, “Dynamic selection of normalization techniques using data complexity measures” Expert Systems with Applications, no. 106, ss. 252-262, 2018.

[82] S. Türüdü, “Praat ve Adobe Audition ile Ses Analizleri” Yildirim Beyazit Üniversitesi Odyoloji, Ses Ve Konuşma Bozukluklari, 2018.

ÖZ GEÇMİŞ

1. İLETİŞİM BİLGİLERİ

Adı Soyadı : Emel ÇOLAKOĞLU

Uyruğu : Türkiye Cumhuriyeti

2. ÖĞRENİM DURUMU

Derece	**Alan**	**Üniversite**	**Yıl**
Lisans	Mühendislik Fakültesi Bilgisayar Mühendisliği	Erciyes Üniversitesi	2001-2005
YüksekLisans	Lisansüstü Eğitim Enstitüsü Hesaplamalı Bilimler ve Mühendislik Ana Bilim Dalı	Kayseri Üniversitesi	2022-Devam

3. YABANCI DİL : İngilizce

4. İŞ DENEYİMLERİ (varsa)

Yıl	**Kurum**	**Görev**
2005-2011	Merkez Çelik A.Ş (Kayseri)	Yazılım Mühendisi
2011-2020	Kayserigaz	BW&BI Yöneticisi

5. YAYINLAR

- E. Çolakoğlu, S. Hızlısoy ve R. S. Arslan, "Konuşmadan Duygu Tanıma Üzerine Detaylı Bir İnceleme: Özellikler ve Sınıflandırma Metotları", *Avrupa Bilim ve Teknoloji Dergisi*, sayı. 32, ss. 471-483, Ara. 2021, doi:10.31590/ejosat.1039403

- S. Hızlısoy, E. Çolakoğlu ve R. S. Arslan, "Speech-to-Gender Recognition Based On Machine Learning Algorithms", International Journal of Applied Mathematics Electronics and Computers, c. 10, sayı. 4, ss. 84-92, Ara. 2022, doi:10.18100/ijamec.1221455

- E. Çolakoğlu, S. Hızlısoy ve R. S. Arslan, "T-Ser: An Efficient Speech Emotional Recognition Model For Turkish Language Based On Machine Learning Algorithms" Innovations And Technologies In Engineering (ss.106-127) Chapter: 9 Publisher: Eğitim Yayınevi

TEZ İNTİHAL RAPORU

Yeni Türkçe Duygusal Veri Seti Üzerinde Konuşmadan Duygu Tanıma

Yazar Emel Çolakoğlu

Gönderim Tarihi: 06-Eyl-2023 04:44PM (UTC+0400)
Gönderim Numarası: 2159004352
Dosya adı: Y_ksekLisansTez_4034620006-EmelColakoglu.pdf (1.73M)
Kelime sayısı: 18363
Karakter sayısı: 111532

Yeni Türkçe Duygusal Veri Seti Üzerinde Konuşmadan Duygu Tanıma

ORİJİNALLİK RAPORU

%9	%8	%1	%2
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	dergipark.org.tr İnternet Kaynağı	%1
2	www.asafvarol.com İnternet Kaynağı	%1
3	acikerisim.baskent.edu.tr İnternet Kaynağı	%1
4	www.egitimyayinevi.com İnternet Kaynağı	%1
5	docplayer.biz.tr İnternet Kaynağı	<%1
6	acikbilim.yok.gov.tr İnternet Kaynağı	<%1
7	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<%1
8	evrimagaci.org İnternet Kaynağı	<%1
9	adudspace.adu.edu.tr:8080 İnternet Kaynağı	<%1

34	acikerisim.mersin.edu.tr İnternet Kaynağı	<%1
35	izmirodm.meb.gov.tr İnternet Kaynağı	<%1
36	opendata.dspace.ceu.es İnternet Kaynağı	<%1
37	"TDAH: análises, compreensões e intervenções clínicas e pedagógicas", Editora Cientifica Digital, 2023 Yayın	<%1
38	iranarze.ir İnternet Kaynağı	<%1
39	katalogtarama.ikcu.edu.tr İnternet Kaynağı	<%1
40	openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<%1
41	storage.googleapis.com İnternet Kaynağı	<%1
42	www.tgm.gov.tr İnternet Kaynağı	<%1

Alıntıları çıkart Kapat Eşleşmeleri çıkar < 7 words
Bibliyografyayı Çıkart üzerinde