



**T.C.
YALOVA ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**TÜRKÇE SES KAYITLARINDAKİ KONUŞMALARDAN ELDE EDİLEN
VERİLERDEN DUYGU TESPİTİ VE ANALİZİ**

**YÜKSEK LİSANS TEZİ
Engin DEMİR
195105005**

Bilgisayar Mühendisliği Anabilim Dalı Adı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Abdülkadir TEPECİK

ARALIK 2021



**T.C.
YALOVA ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**TÜRKÇE SES KAYITLARINDAKİ KONUŞMALARDAN ELDE EDİLEN
VERİLERDEN DUYGU TESPİTİ VE ANALİZİ**

**YÜKSEK LİSANS TEZİ
Engin DEMİR
195105005**

Bilgisayar Mühendisliği Anabilim Dalı Adı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Abdülkadir TEPECİK

ARALIK 2021

ÖNSÖZ

Tez çalışmamda ve öncesinde akademik çalışmalar için beni teşvik eden, desteklerini ve zamanını esirgemeyen danışmanım Prof. Dr. Abdülkadir TEPECİK’e ve bu süreçte beni destekleyen aileme teşekkürlerimi sunarım.

Aralık 2021

Engin DEMİR
(Bilgisayar Mühendisi)



ÖNSÖZ	i
İÇİNDEKİLER	iii
KISALTMALAR	viii
ÇİZELGE LİSTESİ.....	xii
ŞEKİL LİSTESİ.....	xvi
ÖZET.....	xviii
ABSTRACT.....	xx
1. GİRİŞ	1
1.1 Tezin Amacı.....	2
1.2 Literatür Çalışmaları	3
2. MATERYAL VE YÖNTEM	17
2.1 Veri Madenciliği	17
2.2 Duygu Analizi	19
2.2.1 BERT	19
2.2.2 TextBlob.....	21
2.3 Makine Öğrenmesi	21
2.3.1 Naive Bayes	23
2.3.2 Rastgele Orman.....	24
2.3.3 Destek Vektör Makinesi.....	25
2.3.4 Karar Ağacı	26
2.3.5 K-En Yakın Komşu.....	26
3. DENEYSEL BULGULAR	29
3.1. Veri Setinin Elde Edilmesi.....	29
3.2 Türkçe Ses Kayıtlarından Elde Edilen Veriler Üzerinde Python Programlama Dili ile Duygu Tespiti	30
3.2.1 İki Duygu Etiketine Sahip BERT Modeli ile Duygu Tespiti	31
3.2.2 Üç Duygu Etiketine Sahip BERT Modeli ile Duygu Tespiti.....	32
3.2.3 TextBlob Yöntemi ile Duygu Tespiti.....	34



3.3 CoLab’da Python Kullanarak Makine Öğrenim Algoritmalarıyla Veri Analizi..	36
3.3.1 İki Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi	37
3.3.2 Üç Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi	38
3.3.3 TextBlob Kütüphanesi ile Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi	39
3.4 Makine Öğrenim Algoritmalarının Analiz Sonuçlarında Elde Edilen Karmaşıklık Matrislerine Ait Sonuçlar.....	40
3.4.1 İki Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri	41
3.4.2 Üç Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri.....	42
3.4.3 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisleri	43
3.5 Python Programlama Dili ile Eğitim ve Test Veri Setleri Oranlarının Analizi ...	44
3.5.1 BERT Modeli (Pozitif - Negatif)	44
3.5.2 BERT Modeli (Pozitif – Negatif – Nötr)	44
3.5.3 TextBlob.....	44
3.6 RapidMiner’da Makine Öğrenim Algoritmalarıyla Veri Analizi	45
3.6.1 İki Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi.....	46
3.6.2 Üç Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi.....	50
3.6.3 TextBlob ile Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi.....	53
3.7 Rapid Miner ile Analiz Sonucunda Elde Edilen Karmaşıklık Matrisleri.....	56
3.7.1 İki Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri	56
3.7.2 Üç Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri.....	56
3.7.3 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisleri	57
3.8 Rapid Miner ile Eğitim ve Test Veri Setleri Oranlarının Analizi	57



3.8.1 BERT Modeli (Pozitif - Negatif)	57
3.8.2 BERT Modeli (Pozitif – Negatif – Nötr)	57
3.8.3 TextBlob.....	58
4. SONUÇLAR VE ÖNERİLER	59
4.1 Sonuçlar	59
4.2 Öneriler	73
KAYNAKLAR	75
ÖZGEÇMİŞ	83



KISALTMALAR

BERT	: Bidirectional Encoder Representations from Transformers (Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri - TÇYKT)
NLP	: Natural Language Processing (Doğal Dil İşleme - DDİ)
EmoDB	: Berlin Database of Emotional Speech (Berlin Duygusal Konuşma Veritabanı - BDKV)
RAVDESS	: Ryerson Audio-Visual Database of Emotional Speech and Song (Duygusal Konuşma ve Şarkının Ryerson Görsel-İşitsel Veritabanı - DKŞRGİV)
K-NN	: K-Nearest Neighbour (K-En Yakın Komşu -KEYK)
SVM	: Support Vector Machine (Destek Vektör Makinesi -DVM)
LR	: Logistic Regression (Lojistik Regresyon - LR)
GNB	: Gaussian Naïve Bayes
IEMOCAP	: Interactive Emotional Dyadic Motion Capture (Etkileşimli Duygusal İkili Hareket Yakalama- EDİHY)
SAVEE	: Surrey Audio-Visual Expressed Emotion (Surrey Görsel-İşitsel İfade Edilmiş Duygu - SGİİED)
LSTM	: Long Short-Term Memory (Uzun Kısa Süreli Bellek - UKSB)
MFCC	: Mel-Frequency Cepstral Coefficient (Mel-Frekans Cepstrum Katsayısı - MFCK)
LMT	: Logistic Model Tree (Lojistik Model Ağacı - LMT)
USC	: University of Southern California (Güney Kaliforniya Üniversitesi - GKÜ)
TESS	: Toronto Emotional Talk Set (Toronto Duygusal Konuşma Seti - TDKS)
LMT	: Logistic Model Tree (Lojistik Model Ağacı - LMA)
VAM	: Vera am Mittag- German Audio-Visual Emotional Speech Database (Alman Görsel-İşitsel Duygusal Konuşma Veritabanı - AGİDKV)
AVEC	: Audio/Visual Emotion Challenge (Sesli/Görsel Duygu Yarışması - SGDY)
SPD	: Secreted Protein Database (Salgılanmış Protein Veritabanı - SGV)
CRF	: Conditional Random Field (Koşullu Rastgele Alan - KRA)
IITKGP- SEHSC	: Indian Institute of Technology Kharagpur Simulated Emotion Hindi Speech Corpus (Hint Teknoloji Enstitüsü Kharagpur Simüle Edilmiş Duygu Hintçe Konuşma Derlemi - HTEKSEDHKD)
KL	: Kullback-Leibler
PEAQ	: Perceptual Assessment Of Sound Quality (Ses Kalitesi Algısal Değerlendirmesi - SKAD)
CNN	: Convolutional Neural Network (Evrişimli Sinir Ağı - ESA)
LDC	: Linguistic Data Consortium (Dilbilimsel Veri Konsorsiyumu - DVK)
LPCC	: Linear Predictive Cepstral Coefficients (Doğrusal Öngörülü Cepstral Katsayıları - DÖCK)
OAA	: One Against All (Herkes Karşı Bir - HKB)



GAD	: Gizli Anlambilimsel Dizinleme (GAD)
VADER	: Valence Aware Dictionary and Sentiment Reasoner (Değer Duyarlı Sözlük ve Duygu Akıl Yürütücüsü - DDSDAY)
ASR	: Automatic Speech Recognition (Otomatik Konuşma Tanıma - OKT)
RBF	: Radial Basis Function (Radyal Temel Fonksiyon - RTF)
RNN	: Recurrent Neural Network (Tekrarlayan Sinir Ağı - TSA)
MLP	: Multilayer Perceptron (Çok Katmanlı Algılayıcı - ÇKA)
NB	: Naive Bayes
RF	: Random Forest (Rastgele Orman - RO)
SMO	: Sequential Minimal Optimization (Sıralı Minimum Optimizasyon - SMO)
IB1	: 1-Nearest Neighbors (1-En Yakın Komşular – 1EYK)
J48	: Decision Tree (Karar Ağacı - KA)
MLM	: Masked Language Modeling (Maskeli Dil Modelleme - MDM)
NSP	: Next Sentence Prediction (Sonraki Cümle Tahmini - SCT)
VC	: Vapnik-Chervonenkis
Onxx	: Open Neural Network Exchange (Açık Sinir Ağı Değişimi - ASAD)
TF-IDF	: Term Frequency–Inverse Document Frequency (Terim Frekansı x Ters Belge Frekansı – TF-TBF)
RML	: Ryerson Multimedia Research Lab (RML Duygu Veritabanı - RMLDV)
UGA	: University of Georgia (Georgia Üniversitesi - GÜ)



ÇİZELGE LİSTESİ

Sayfa No

Çizelge 3.1 BERT (Pozitif -Negatif) Modeliyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi.....	38
Çizelge 3.2 BERT (Pozitif -Negatif-Nötr) Modeliyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi	39
Çizelge 3.3 TextBlob Kütüphanesiyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi	40
Çizelge 3.4 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Karmaşıklık Matrisi	41
Çizelge 3.5 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Karmaşıklık Matrisi	43
Çizelge 3.6 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisi	43
Çizelge 3.7 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Doğruluk Oranları.....	44
Çizelge 3.8 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Doğruluk Oranları	44
Çizelge 3.9 TextBlob Kütüphanesi ile Elde Edilen Doğruluk Oranları	45
Çizelge 3.10 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Analiz Sonuçları	49
Çizelge 3.11 BERT (Pozitif -Negatif-Nötr) Modeliyle Elde Edilen Analiz Sonuçları	52
Çizelge 3.12 TextBlob Kütüphanesi ile Elde Edilen Analiz Sonuçları.....	55
Çizelge 3.13 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Karmaşıklık Matrisi ..	56
Çizelge 3.14 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Karmaşıklık Matrisi	56
Çizelge 3.15 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisi	57
Çizelge 3.16 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Doğruluk Oranları.....	57
Çizelge 3.17 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Doğruluk Oranları	57
Çizelge 3.18 TextBlob Kütüphanesi ile Elde Edilen Doğruluk Oranları	58



Çizelge 4.1 BERT (Pozitif-Negatif) Modelinde En İyi Sonuçlara Ulaşan Algoritmalar	61
Çizelge 4.2 BERT (Pozitif-Negatif-Nötr) Modelinde En İyi Sonuçlara Ulaşan Algoritmalar	63
Çizelge 4.3 TextBlob Kütüphanesinde En İyi Sonuçlara Ulaşan Algoritmalar	65
Çizelge 4.4 İki Duygu Etiketine Sahip BERT Modelinde En İyi Oranlara Ulaşan Algoritmalar	68
Çizelge 4.5 Üç Duygu Etiketine Sahip BERT Modelinde En İyi Oranlara Ulaşan Algoritmalar	69
Çizelge 4.6 TextBlob ile En İyi Oranlara Ulaşan Algoritmalar	70





ŞEKİL LİSTESİ

Sayfa No

Şekil 1.1 Çalışmanın Şematik Gösterimi	3
Şekil 2.1 BERT Model Yapısı.....	20
Şekil 2.2 Makine Öğrenmesi, Öğrenme Yöntemleri.....	22
Şekil 2.3 Öklid Mesafe Fonksiyonu	27
Şekil 3.1 Veri Setinin İlk 19 Verisi	30
Şekil 3.2 İki Duygu Etiketine Sahip BERT Modeli Kod Yapısı.....	31
Şekil 3.3 İki Duygu Etiketine Sahip BERT Modeli ile Yapılan Duygu Tespiti Sonuçları	32
Şekil 3.4 Üç Duygu Etiketine Sahip BERT Modeli Kod Yapısı	33
Şekil 3.5 Üç Duygu Etiketine Sahip BERT Modeli ile Yapılan Duygu Tespiti Sonuçları	34
Şekil 3.6 TextBlob Kütüphanesinin Veri Seti Üzerinde Uygulanması	35
Şekil 3.7 TextBlob Kütüphanesi ile Yapılan Duygu Tespiti Sonuçları	35
Şekil 3.8 Naive Bayes (Kernel) Algoritması.....	47
Şekil 3.9 Rastgele Orman Algoritması.....	47
Şekil 3.10 Destek Vektör Makinesi Algoritması	48
Şekil 3.11 Karar Ağacı Algoritması.....	48
Şekil 3.12 K-En Yakın Komşu Algoritması.....	49
Şekil 3.13 Naive Bayes (Kernel) Algoritması.....	50
Şekil 3.14 Rastgele Orman Algoritması.....	50
Şekil 3.15 Destek Vektör Makinesi Algoritması	51
Şekil 3.16 Karar Ağacı Algoritması.....	51
Şekil 3.17 K-En Yakın Komşu Algoritması.....	52
Şekil 3.18 Naive Bayes (Kernel) Algoritması.....	53
Şekil 3.19 Rastgele Orman Algoritması.....	53
Şekil 3.20 Destek Vektör Makinesi Algoritması	54
Şekil 3.21 Karar Ağacı Algoritması.....	54
Şekil 3.22 K-En Yakın Komşu Algoritması.....	55



TÜRKÇE SES KAYITLARINDAKİ KONUŞMALARDAN ELDE EDİLEN VERİLERDEN DUYGU TESPİTİ VE ANALİZİ

ÖZET

Duygu, insanın ruh halinde içsel ve çevresindeki etkilerle etkileşiminden doğan fiziksel değişimlerdir. Duygu veya duygular jest ve mimikler dışında ses ve konuşmalarla da aktarabilmektedir. Sesteki duygu tespiti, veri bilimindeki gelişmelerle beraber önem kazanmaya başlamıştır. Özellikle beden dilinin etkin olmadığı zaman ve yerlerde, örneğin ses kayıtları gibi verilerde konuşmadan yapılan duygu tespiti ve analizleri gelişen teknolojiyle birlikte bu alanla ilgili çalışma konularında üst sıralara yükselmiştir.

Duygu tespiti, iki şekilde ifade edilmektedir. Bunlardan ilki daha çok insani duyguları ifade eden öfke, mutluluk, üzüntü vb. duygulardır. Diğer duygu ifade biçimi ise bu duygulardan sevinç, mutluluk gibi olumlu durumları ifade eden pozitif terimi, öfke, mutsuzluk gibi olumsuz durumları ifade negatif terimi ve kişide hiçbir tepki yaratmayan durumlar için kullanılan nötr terimidir. Pozitif, negatif ve nötr terimleri ile ifade edilen duygu aktarımları makine dili için daha uygun görülmektedir. Makine diline insani duyguları öğretmek için makine öğrenim modellerinin yanı sıra, yapay zeka algoritmaları da kullanılmaktadır. Böylelikle yapılan duygu tespitlerinin yanı sıra makine öğrenim modelleri ile duygu tespitinin analizi yapılmaktadır.

Bu çalışmada da, Türkçe ses kayıtlarından oluşan bir veri setinin, duygu tespiti ve analizi gerçekleştirilerek, Türkçe duygu veri tabanı oluşturma ve çalışmanın içeriğiyle, yapılış yöntemiyle diğer Türkçe çalışmalara öncü olması amaçlanmıştır. Çalışmada kullanılan veri seti, Common Voice açık kaynak ses veri kümelerinden elde edilmiştir. Veri setindeki ses kayıtları üzerinden direkt duygu tespiti gerçekleştirilmemiş, ses kayıtları metin haline getirilerek, bu metinler üzerinden duygu tespiti gerçekleştirilmiştir. Bu durum çalışmayı farklı kılan hususlardan biri olmuştur. Duygu tespiti için BERT modelinin iki farklı varyasyonu ve TextBlob kütüphanesi kullanılmıştır. Çalışmanın bir diğer önemli noktası ise veri seti üzerinde yapılan duygu tespiti sonrasında makine öğrenim modelleri ile yapılan analiz olmuştur. Literatür araştırması sonrası çalışmalarda en çok kullanılan beş makine öğrenim modeli ile bu analizler gerçekleştirilmiştir. Analizler, hem Python programlama dilinde hem de RapidMiner veri analiz platformunda gerçekleştirilmiştir. Çalışmadaki duygu tespitinin üç farklı yöntemle gerçekleştirilmesi, analizlerin iki farklı şekilde ve kendi içinde de çeşitlendirilmesi çalışmayı özgün kılan hususlar olmuştur.

Anahtar Kelimeler: Türkçe Ses Kayıtları, Duygu Analizi, Makine Öğrenmesi, Ses Kayıt Analizi, Veri Madenciliği, Veri Elde Edilmesi



EMOTION DETECTION AND ANALYSIS FROM DATA OBTAINED FROM SPEECHES IN TURKISH AUDIO RECORDINGS

ABSTRACT

Emotions are physical changes in a person's mood resulting from his interaction with internal and environmental influences. Emotions or emotions can be conveyed by voices and speeches besides gestures and gestures. Emotion detection in sound has become important with the developments in data science. Especially at times and places where body language is not effective, for example, in data such as voice recordings, the emotion detection and analysis made from speaking has risen to the top in the subjects of study about this field with the developing technology.

Emotion detection is expressed in two ways. The first of these is anger, happiness, sadness, etc., which express more human emotions. are emotions. The other emotion expression form is the positive term that expresses positive situations such as joy and happiness, the negative term that expresses negative situations such as anger and unhappiness, and the neutral term used for situations that do not cause any reaction in the person. Emotional transfers expressed with the terms positive, negative and neutral are considered more suitable for machine language. Besides machine learning models to teach human emotions to machine language, artificial intelligence algorithms are also used. In this way, in addition to emotion detection, emotion detection is analyzed with machine learning models.

In this study, it is aimed to create a Turkish emotion database by performing emotion detection and analysis of a data set consisting of Turkish voice recordings, and to be a pioneer in other Turkish studies with the content and method of the study. The dataset used in the study was obtained from Common Voice open-source audio datasets. Emotion detection was not carried out directly through the voice recordings in the data set, the voice recordings were converted into text, and emotion detection was carried out through these texts. This was one of the things that made the study different. Two different variations of the BERT model, and the TextBlob library were used for emotion detection. Another important point of the study was the analysis made with machine learning models after the emotion detection on the data set. After the literature research, these analyzes were carried out with the five most used machine learning models in studies. Analyzes were performed in both Python programming language and RapidMiner data analysis platform. The fact that the feeling of emotion in the study has been issues that make the fact that the analysis is organized in two different ways and in their own.

Keywords: Turkish Voice Recordings, Sentiment Analysis, Machine Learning, Voice Recording Analysis, Data Mining, Data Acquisition



1. GİRİŞ

İnternet ağının yaygınlaşmasıyla beraber kişiler görüşlerini, değerlendirmelerini sosyal mecralarda, videolarda ve ses kayıtları aracılığıyla aktarmaya başlamıştır. Kişilerin bu alanlarda fikir sahibi olmalarına, bir konu, durum ve eşya hakkında bilgi sahibi olmasına veya bilgilerini aktarmada katkı sağlamıştır. Bu durum beraberinde yeni araştırma alanlarını da getirmiştir.

İnsanların ruhsal etkenlerle ve çevrelerindeki olaylara göre verdikleri tepkilerin tanımına duygu denilmektedir [1]. Beden dilinin yanı sıra konuşmadaki duygu da önem arz etmektedir. Özellikle insanların duygularına yönelik analizler, her geçen gün artan bu verilere ait araştırma alanlarını ve tekniklerinin de gelişimlerini sağlamıştır.

Duygu analizine baktığımızda ise insanlar, olumlu ve olumsuz duygularını rahatlıkla ifade edilebilmeyi imkân buldukları andan itibaren bu alana ilgi gün geçtikçe artmış ve bu konudaki çalışma sayısı artmıştır.

Duygu analizi terimin tarihine baktığımızda ilk olarak 2003 yılında bu ifade ortaya çıkmıştır. Aslında dilbilim ve doğal dil işleme tarih boyunca olmasına rağmen insanların duygularına yönelik analizler 2000'li yıllarından başlarından itibaren bulunmaktadır [2]. Bu yıllardan itibaren duygu analizi, dilbilim ve doğal dil işlemenin ilgi çeken bir kolu haline gelmiştir. Bu durumun nedenlerine bakıldığında üç temel sebep gösterilebilir. Bunlar; doğal dil işlemenin geniş bir yelpazeye sahip olması, hiç incelenmemiş araştırma alanları ve sosyal medyada çok büyük miktarda verilerin bulunmasıdır. Bu nedenlere bakıldığında verilerin araştırmalar ve analizler için önemli bir kaynak olduğu göze çarpmaktadır.

Duygu tespiti, duygu analizi gibi terimler, makine öğrenmesiyle beraber daha çok ön plana çıkmış ve veri bilimi ile ilgili araştırma alanlarının odak noktası haline gelmiştir. Makine öğrenmesinin tanımına bakıldığında ise bir taraftan insan zekâsını ve algısını taklit ederken, diğer taraftan da kişinin yorumlayıp elle gireceği kurallara ihtiyaç duymayan algoritmalar bütünü olarak ifade edilmektedir [3]. Duygu tespiti sonrası makine öğrenim algoritmaları ile yapılan analizler, yapılacak olan birçok çalışma için de yeni fikirlerin ortaya çıkmasında katkı sağlamıştır.

1.1 Tezin Amacı

Tez çalışması, duygu tespiti, duygu analizi ve makine öğrenmesi terimlerini içeren bir çalışma olmuştur. Yapılan çalışmayla Türkçe bir duygu veri tabanı oluşturmak, metot ve materyallerle yapılacak olan çalışmalara öncü olmak, duygu tespitinde kullanılan BERT modelinin Türkçe yapısını geliştirmek ve bu model ile geliştirilecek çalışmalar için fikir oluşturmak, makine öğrenim algoritmalarının Türkçe dil yapısı üzerindeki başarı oranlarının ölçümü amaçlanmıştır.

Çalışmada öncelikle ses kayıtlarından bir veri seti oluşturmak hedeflenmiştir. Ses kayıtlarından direkt duygu tespiti yerine ses kayıtlarının metne dönüştürülmüş haliyle bir veri seti oluşturulup hazır hale getirilmiştir. Böylelikle metinler üzerinden bir duygu tespiti gerçekleştirilmiştir.

Çalışmanın duygu tespiti aşamasında ise Jacob Devlin ve arkadaşları tarafından 2018 yılında Google'da doğal dil işleme alanındaki çalışmalar için geliştirilen BERT modeli ile Python programlama dilinde bulunan TextBlob kütüphanesi kullanılmıştır. BERT modelinin Türkçe dil yapısına uygun hazırlanan ve Türkçe dil yapısını barındıran model yapısı ile iki farklı duygu tespiti gerçekleştirilmiştir. Bunlardan ilkinde pozitif ve negatif duygu etiketi yer alırken, ikincisinde bunlara ek olarak nötr duygu etiketi de eklenmiştir. TextBlob kütüphanesinde Türkçe dil desteği bulunmadığından veriler öncelikle İngilizce'ye çevrilmiş, sonrasında duygu tespiti işlemi gerçekleştirilmiştir. Böylelikle üç farklı yöntemle yapılan, üç farklı duygu tespiti içeren üç veri seti ortaya çıkmıştır.

Duygu analizi, çalışmanın son aşaması olmuştur. Veri setinde bulunan cümleler üzerinde işlem yapabilmek için metni kelime sayısı vektörlerine dönüştüren CountVectorizer ve kelime frekansı vektörlerine dönüştüren TF-IDFVectorizer vektörizasyon yöntemleri kullanılmıştır. Analiz için iki platform belirlenmiştir. Bunlardan ilki olan Python programlama dili, zengin kütüphane içeriğiyle, kolay yazım diliyle ve makine öğrenim modelleri eklentileriyle veri analizi için önemli bir platform haline gelmiştir. Bir diğeri de RapidMiner'dır. RapidMiner platformu, veri analizi için kolay ara yüzü, eklenti zenginliği ve duygu tespiti ve analizi için barındırdıklarıyla veri bilimi için önemli bir konumdadır.

Literatür araştırmasından yararlanılarak, çalışmalarda en çok kullanılan makine öğrenim yöntemleri tespit edilip ve bunlardan beş tanesinin kullanımına karar verilmiştir. Python programlama dili ile yapılan analizde hem CountVectorizer hem de TF-IDFVectorizer vektörizasyon yöntemleri kullanılmıştır ve her makine öğrenim algoritmasına uygulanmıştır. RapidMiner’da ise sadece TF-IDFVectorizer vektörizasyon yöntemi kullanılmıştır ve her makine öğrenim algoritmasına uygulanmıştır. Çalışmanın aşamaları Şekil 1’de şema halinde gösterilmiştir.



Şekil 1.1 Çalışmanın Şematik Gösterimi

1.2 Literatür Çalışmaları

İnsan duygularına yönelik çalışmalar ve analizler son yıllarda özellikle gelişen teknolojiyle beraber bir ivme kazanmıştır. Makine diline öğretimi sağlayan makine öğrenmesi bilim alanının çeşitli platformlara entegrasyonu bu analizler ve çalışmalar giderek artmıştır. Bireylerin internet teknolojisi aracılığıyla sağladıkları veri aktarımı, araştırma alanlarındaki çeşitliliğe katkı sağlamıştır.

Bu araştırma alanlarından biri olan sesteki duygu analizi de son yıllarda üzerinde yapılan çalışmalarla önem kazanmıştır. Ses özelliklerindeki çıkarımlardan yapılan duygu analizinin yanı sıra ses kayıtlarının metin hali üzerinden duygu analizi de yapılmaktadır.

Tezin hedefine uygun bir şekilde ilerlemesi için konuşma veya ses kayıtlarından duygu tespitini ve analizini içeren çalışmalara dair literatür özeti aşağıda verilmiştir. Literatür özeti dört kısımda incelenmiştir. Bunlar; videolardan elde edilen ses kayıtlarını inceleyen çalışmalar, EmoDB, RAVDESS vb. hazır veri seti kullanılan çalışmalar, Türkçe duygu analizi ile ilgili çalışmalar ve ses kayıtlarıyla ilgili diğer çalışmalardır. İlk olarak videolardan elde edilen ses kayıtlarını içeren literatür özeti verilmiştir.

A.Murarka ve arkadaşları [4], kamuya açık veri tabanlarında ve videolardaki seslerden elde ettikleri metinler üzerinde duygu analizi yapmışlardır. Yaptıkları çalışmada arama motorlarının, kullanıcıların ilgili bilgi arama yeteneklerini geliştirmeyi amaçlamışlardır. Çalışmada Hidden Markov ve Destek Vektör Makinesi algoritmasını kullanmışlardır. Sonuç kısmında ise çalışmanın açık uçlu analizleri ortaya çıkardığını ve bununla ilgili birçok tartışmayı beraberinde getirdiğini belirtmişlerdir. E. Cambria ve arkadaşları [5], çevrim içi videolarda bulunan jest ve sesleri uçtan uca öğrenen bir model sunmuşlardır. Tensor Fusion Network adını verdikleri bu model ile seslerde duygu analizi gerçekleştirmişlerdir. Duygu analizinde ise Evrişimli Sinir Ağı ve Destek Vektör Makinesi sınıflandırma algoritmalarını kullanmışlardır. Çalışmanın sonuç kısmında ise kullandıkları füzyon modelinin üzerinde çalıştıkları veri setinde başarılı olduğunu ifade etmişlerdir. M. H. R. Pereira ve arkadaşları [6], haber videolarını içeren bir veri setinde duygu analizi gerçekleştirmek için bir yaklaşım sunmuşlardır. Çalışmada haberlerin gerginliği ve duyguların anlaşılabilirliği üzerine durulmuştur. Sonuç kısmında ise yüksek gerginlik içeren haber videolarında yaklaşımın yüksek doğruluk sağladığı fakat düşük gerginliğe sahip videolarda aynı başarı oranını sağlanamadığı ifade edilmiştir. J. Y. R. Cornejo ve arkadaşları [7], videolardaki ses ve görüntü özelliklerini çıkarmak için hibrit bir derin konvolüsyonel sinir ağı modeli ortaya çıkarmışlardır. Çalışmada, RML, Enterface05 ve Baum-1S veri kümelerini kullanmışlardır ve K-En Yakın Komşu, Destek Vektör Makinesi, Lojistik Regresyon ve Gaussian Naïve Bayes sınıflandırma algoritmalarını duygu tanıma için bu veri setleri üzerinde uygulamışlardır. Sonuç kısmında ise önerilen modelin Baum-1S veri seti üzerinde etkin sonuçlar ortaya koyduğunu ve başka veri setleri üzerinde de deneyeceklerini ifade etmişlerdir.

EmoDB, RAVDESS, IEMOCAP gibi hazır duygu veri tabanları ile yapılan çalışmalarla ilgili literatür özeti aşağıda verilmiştir. A. Demir ve arkadaşları [8], insanların duygularını işitsel bilgilerden elde etmek için bir otomatik duygu tanıma sistemi önermişlerdir. Çalışmada SAVEE, RAVDESS ve RML veri setleri kullanılarak bu veri setlerinde bulunan ses sinyallerinden özellikler çıkarılmış ve sınıflandırma için derin öğrenme tabanlı LSTM algoritması kullanılmıştır. Sonuç kısmında ise en başarılı sonuçların SAVEE veri seti için elde edildiğini ve bu durumun SAVEE veri setinin, ana dili İngilizce olan erkek konuşmacılardan elde edilmiş olmasından kaynaklandığını ifade etmişlerdir. M. G. Pinto ve arkadaşları [9], sesteki duygu analizinin tespiti için Evrişimli Sinir Ağlarına dayanan bir sınıflandırma modeli sunmuşlardır. Çalışmada 8 duygunun tespiti için RAVDESS veri seti kullanılmıştır. Sonuç kısmında ise F1-skoruna göre değerlendirme yapıldığını ve en yüksek skorun kızgın duygusuna, en düşük skorun üzgün duygusuna ait olduğunu belirtmişlerdir. K. M. Kudiri ve arkadaşları [10], insan konuşmalarındaki sesleri kullanarak duygu tanımlama sistemi için yeni bir yaklaşım sunmuşlardır. Çalışmada dört duygu (kızgın, üzgün, mutlu ve nötr) türünün tahmini sağlanmaya çalışılmış ve bunun içinde Berlin duygu veri tabanını kullanmışlardır. Sınıflandırma için Sinir Ağlarını ve Destek Vektör Makinesi makine öğrenim algoritmaları kullanılan çalışmanın sonuç kısmında göreceli bin frekans katsayılı yaklaşımın umut verici sonuçlar verdiğini belirtmişlerdir. S. Hasan ve arkadaşları [11], konuşmadaki duyguların tespitini ve sınıflandırılmasını amaçlamışlardır. EmoDB ve RAVDESS veri setleri üzerinde yaptıkları çalışmada, konuşmalardaki Mel Frekans Cepstrum Katsayısı özellikleri çıkarılmış ve bu özellikler kullanılarak Lojistik Model Ağacı modelleri eğitilmiştir. Çalışmanın sonuç kısmında ise 7 farklı duygunun tespitinde doğruluk oranının %70 olduğunu ve gelecek çalışmalar için konuşma sinyallerindeki bağlamsal bilgilerin çıkarılması gerektiğini ifade etmişlerdir. S. Suganya ve E. Y. A. Charles [12], ses kaydına derin sinir ağı uygulayan uçtan uca bir derin öğrenme yaklaşımı sunmuşlardır. Çalışmada, USC-IEMOCAP ve EmoDB veri setleri kullanılmıştır. Evrişimli Sinir Ağı modeli uygulanan bu veri setlerinde dört duygu üzerinde IEMOCAP için %68.6 ve yedi duygu üzerinde EmoDB için %85.62'lik doğruluk oranı sonucuna ulaşmışlardır.

Sonuç kısmında ise, önerdikleri modelin, derin öğrenme yaklaşımına göre daha iyi sonuç verdiğini belirtmişlerdir. K. Venkataramanan ve H. R. Rajamohan [13], konuşma tabanlı duygu tanıma sistemlerine çeşitli yaklaşımların kapsamlı bir karşılaştırmasını yapmayı amaçlamışlardır. Çalışma RAVDESS veri seti üzerinde gerçekleştirilmiş ve log-mel spektrogramı, Mel Frekans Cepstral Katsayıları, pitch ve enerji gibi özellikler baz alınmıştır. Yapılan deneylerde Evrişimli Sinir Ağları, Gizli Markov Modelleri ve Derin Sinir Ağları yöntemleri karşılaştırılmıştır. Sonuç kısmında cinsiyete bağlı sınıflandırmanın daha yüksek performans sağladığı, duygu tanımda ses özellik seçimlerinin model karmaşıklığından daha etkin olduğunu ve Evrişimli Sinir Ağları modeliyle %68 oranında doğruluk sağladığını belirtmişlerdir. S. Dani ve arkadaşları [14], seslerdeki duyguyu tespit etmeyi amaçlamışlardır. Çalışmada, K-En Yakın Komşu ve Karar Ağaçları makine öğrenim tekniklerini, Toronto Duygusal Konuşma Seti adlı veri seti üzerinde 7 duygu için uygulamışlardır. Sonuç kısmında ise K-En Yakın Komşu tekniğinin %98, Karar Ağaçlarının %92 ve Ekstra-Ağaç sınıflandırıcının %99 doğruluk oranı sağladığını belirtmişlerdir. U. Garg ve arkadaşları [15], sesli duygu tanıma modeli önermişlerdir ve bu önerilen model ile duygu tahminlerinde önemli bir etken olan sınıflandırma algoritmaları arasında en iyi olanı bulmayı amaçlamışlardır. Çalışmada RAVDESS veri seti kullanılmıştır ve bu veri seti üzerinde K-En Yakın Komşu, Destek Vektör Makinesi, Rastgele Orman ve Karar Ağacı sınıflandırma algoritmaları uygulanmıştır. Sonuç kısmında ise karmaşıklık matrisine göre yapılan değerlendirmeye göre en iyi sonucu Rastgele Orman sınıflandırma algoritmasının verdiğini belirtmişlerdir. C. Sezgin ve arkadaşları [16], önerdikleri model içerisinde yer alan 9 boyutlu algısal ses öznitelik vektörlerini kullanarak duygu sınıflandırma gerçekleştirmişlerdir. Çalışma için EmoDB ve VAM duygu veri tabanlarını kullanmışlardır. Sınıflandırma algoritması olarak Destek Vektör Makinesi, Gauss Karışım Modelleri ve Vektör Niceleme ile öğrenme yöntemleri kullanılmış ve çıkan sonuçlar OpenEAR ve GerDa platformlarında karşılaştırılmıştır. Çalışmanın sonuç kısmında ise, doğal ses kayıtlarında öğrenmeye dayalı sınıflandırma yöntemiyle daha yüksek başarı sağlandığı ve en yüksek başarı oranının da Destek Vektör Makinesi yöntemiyle elde edildiği belirtilmiştir. R. Matin ve D. Valles [17], otizm spektrum bozukluğu olan çocuklara, sosyal etkileşimdeki duyguları tanımlamak için bir konuşma duygu tanıma modeli geliştirmeyi amaçlamışlardır. Destek Vektör Makinesi algoritmasına dayalı olan önerilen model Python programlama dili ile geliştirilmiştir.

Çalışmada RAVDESS veri setiyle beraber Destek Vektör Makinesi, Karar Ağacı ve Lojistik Regresyon sınıflandırma algoritmaları kullanılmıştır. Sonuç kısmında ise yapılan deneylerin sonucunda Destek Vektör Makinesi algoritması %77 oranda doğruluk sağladığı ve önerilen modelin arka plan gürültüsü ile ses için iyi sınıflandırma doğruluğu sağladığı belirtilmiştir. M. R. Amer ve arkadaşları [18], konuşmada duyguları tespit etmek için, hibrit bir model sunmuşlardır. Önerilen hibrit model, zamansal üretken ve sınıflandırma için kullanılan ayırt edici bir modelden oluşmaktadır. AVEC, VAM ve SPD veri setleri üzerinde uygulanan bu modelin, zamansal bağımlılıkların modellenmesini sağlayan CRF'nin katkısıyla üstün bir sınıflandırma performansı sağladığı sonuç kısmında belirtilmiştir. S. Yoon ve arkadaşları [19], konuşma verilerinin daha iyi anlaşılmasını sağlamak için aynı anda metin verilerini ve ses sinyallerini kullanan yeni bir Derin Çift Tekrarlayan Enkoder Modeli önermişlerdir. Çalışmada IEMOCAP veri seti üzerinde deneyler ve duygu tahmini yapılmıştır. Sonuç kısmında ise önerilen modelin %68 ile %71.8 oranları arasında doğruluk sağladığı ve nötr duygusunun oluşturduğu sıkıntıları giderdiği belirtilmiştir. M. C. Sezgin ve arkadaşları [20], Ses Kalitesi Algısal Değerlendirmesine dayanan yeni bir ses duygu tanıma sistemi sunmuşlardır. Önerilen model, duygusal farklılıkların istatistiksel özelliklerini öğrenerek veri normalleştirme ihtiyacını ortadan kaldırmaktadır. VAM ve EmoDB veri setleri üzerinde uygulanan model, sınıflandırıcı olarak Destek Vektör Makinesi algoritması ve Gauss Karışım modelini kullanmıştır. Çalışmanın sonuç kısmında ise hem doğal hem de hareketli duygusal veriler için yapılan kıyaslamalarda yüksek doğruluk oranları elde ettiklerini belirtmişlerdir. V. C. Govoreanu ve M. Neghină [21], endüstriyel eğitim istasyonu için bir konuşma duygu tanıma sistemi tasarlamaya yönelik bir yaklaşım sunmayı amaçlamışlardır. Önerilen yöntem RAVDESS veri seti üzerinde değerlendirilmiş ve sınıflandırıcı olarak Yapay Sinir Ağı modeli kullanılmıştır. Sonuç kısmında ise önerilen yöntemin diğer yöntemlerle rekabet oluşturduğu ve olumlu sonuçlar elde ettiği belirtilmiştir. A. Bhavan ve arkadaşları [22], üç konuşma duygu veri tabanı üzerinde duygu sınıflandırması yapmak için torba bir model sunmayı amaçlamışlardır. Çalışmada, Berlin EmoDB, RAVDESS ve IITKGP-SEHSC veri setleri kullanılmıştır. Kullanılan veri setleri üzerinde Destek Vektör Makinesi algoritmasının performansı değerlendirilmiştir.

Çalışmanın sonuç kısmında konuşma duygu veri tabanları üzerinde yapılan deneyler sonucunda önerilen yöntemin doğruluk açısından etkin sonuçlar ortaya koyduğunu belirtmişlerdir. S. R. Kadir ve P. Alku [23], otomatik bir duygu algılama sistemi gerçekleştirmek için duygusal konuşma ve nötr bir konuşma arasındaki sapmaları araştırmayı amaçlamışlardır. Çalışmada, Kullback-Leibler mesafesi ile duygusal ve nötr konuşmanın özellik dağılımları arasındaki benzerliği ölçmek için hesaplanmasını sağlamışlardır. Önerilen yöntem, Berlin EmoDB veri seti üzerinde uygulanmıştır. Çalışmanın sonuç kısmında ise önerilen yöntemin en büyük avantajının, yöntemi duygusal konuşma ile eğitmeye gerek olmadığı ve yöntemin performansının referans yöntemlerden daha iyi olduğu belirtilmiştir. S. Banga ve arkadaşları [24], gerçek hayat senaryolarında duygu analizi için bir ifadenin içinde hem sözel (konuşma komutları) hem de sözel olmayan sesleri (arka plan sesleri) ele almayı amaçlamışlardır. Bu çalışma için Hint EmoSpeech komut veri kümesini oluşturmuşlardır. Çalışmada benzer duygu tabanlı veri setleriyle (RAVDESS, LibriSpeech vs.) karşılaştırmalar yapılmıştır ve kalabalık ve gürültülü ortamlarda ses analizini sürdürmek için veri setini arka plan seslerine karşı sağlam hale getirmişlerdir. Çalışmanın sonuç kısmında ise %3.3'lük bir performans artışı yakaladıklarını belirtmişlerdir. S. Tripathi ve arkadaşları [25], konuşma özelliklerine ve konuşma metinlerine dayalı bir konuşma duygu tanıma yöntemi önermişlerdir. Çalışmanın amacını, konuşma özelliklerini, metinden anlamsal bilgilerle birlikte, konuşma özelliklerini kullanarak duygusal özellikleri yakalamak ve duygu algılama hassasiyetlerini geliştirmek için derin bir öğrenme tabanlı bir duygu sınıflandırıcı oluşturmak olarak açıklamışlardır. Çalışmada IEMOCAP veri seti ve konuşma özelliklerini kullanarak duygu sınıflandırması için bir Evrişimli Sinir Ağı modeli kullanmışlardır. Sonuç kısmında ise önerilen yöntemin mevcut yöntemlere göre %5 oranında sınıf doğruluğunda, %7 oranında genel doğrulukta artış sağladığı ayrıca önerilen yöntemin konuşma sohbetleri, sosyal robotlar vb. gibi duygularla ilgili uygulamalar için kullanılabileceğini belirtmişlerdir. J. Bhaskar ve arkadaşları [26], konuşmacının duygularını tespit etmek için hem konuşmayı hem de ilgili metin bileşenini analiz etmeyi amaçlamışlardır. Önerilen yöntem, hem ses hem de metnin özelliklerini, daha sonra sınıflandırıcıya verilen tek bir özellik vektöründe birleştirerek duygu sınıflandırmasının verimliliğini arttırmayı amaçlamaktadır.

Çalışmada Semval-2007 ve eNTERFACE'05 EMOTION veri setleriyle birlikte doğal dil işleme, Destek Vektör Makinesi, WordNet etkisi ve SentiWordNet gibi yöntemler kullanılmıştır. Çalışmanın sonuç kısmında ise önerilen yöntemin, metin madenciliği veya konuşma madenciliği ile karşılaştırıldığında daha iyi bir doğruluk sağladığı ve doğruluğunun, yalnızca ses veya metin sınıflandırmasına kıyasla önemli ölçüde daha yüksek olduğunu belirtmişlerdir. M. Jain ve arkadaşları [27], konuşmayı üzüntü, öfke, korku ve mutluluk olmak üzere dört duygudan biri olarak sınıflandırmayı ve Destek Vektör Makinesi yöntemini kullanarak daha yüksek sınıflandırma doğruluğu elde etmeyi amaçlamışlardır. Çalışmada, Linguistic Data Consortium ve UGA veri setleri kullanılmıştır ve enerji, pitch, MFCC katsayıları, LPCC katsayıları ve hoparlör hızı gibi özellikler baz alınmıştır. One against All ve cinsiyete bağlı sınıflandırma yapılmıştır ve LPCC ve MFCC algoritmaları arasında karşılaştırmalı bir analiz yapılmıştır. Çalışmanın sonuç kısmında ise LDC veri kümeleri gürültüsüz bir ortamda ve profesyonel aktörler tarafından kaydedildiğinden, bu veri kümesinin doğruluğunun daha fazla olduğu, cinsiyete bağlı sınıflandırıcının genel doğruluğunun OAA sınıflandırıcısından daha yüksek olduğu ve LPCC algoritmasının doğruluğunun, MFCC algoritmasının doğruluğundan daha az olduğu ifade edilmiştir. P. Harár ve arkadaşları [28], bir kişinin duygusal durumunu 20 milisaniye segmentine bölünmüş kısa bir ses kaydından tahmin etmeyi amaçlamışlardır. Çalışmada, bağlı katmanlarla Derin Sinir Ağı mimarisini ve Berlin duygusal konuşma veri tabanı ve 3 sınıf alt kümesini (kızgın, nötr, üzgün) kullanmışlardır. Sonuç kısmında ise önerilen yöntemin %96.97'lik bir doğruluk elde ettiği ve içeriğe bağlı yaklaşımın bu sonuçları önemli ölçüde artırabileceğini belirtmişlerdir. O. E. Korkmaz [29], ses sinyalinin duyguyu tespiti yapmayı amaçlamıştır. Çalışmada, EmoDB ve çeşitli veri setleri kullanılmıştır ve Destek Vektör Makinesi ve K-En Yakın Komşu sınıflandırıcılarıyla çapraz doğrulama yapılarak sonuçlar elde edilmiştir. Sonuç kısmında ise Destek Vektör Makinesi algoritmasıyla %98,7 oranında başarı elde edildiğini ve Temel Bileşenler Analizi kullanılarak işlem zamanında ve performansta iyi sonuçlar alındığını belirtmiştir.

Türkçe duygu analizi çalışmalarına dair literatür özeti aşağıda verilmiştir.

F. S. Çetin ve G. Eryiğit [30], Türkçe hedef tabanlı duygu analizi konusunda yapılan incelemeleri sunmuşlardır. Çalışmada, Türkçe restoran yorumları veri kümesi kullanılmıştır ve kelime vektörleri ve doğal dil işleme çıktıları kullanan koşullu rastgele alanlara dayalı bir dizilim etiketleme algoritması tasarlanmıştır. Çalışmanın sonuç kısmında, duygu sınıfı belirleme için cümle analizi sonucunda hedef terime yakın kelimelerden özellik seçimine dayalı bir lineer sınıflandırma yöntemi sunulduğunu ve literatürde sınırlı sistemler tarafından raporlanan en başarılı sonuç - %76,1 F1-skoru - elde edildiğini belirtmişlerdir. H. Göker ve H. Tekedere [31], FATİH projesine yönelik internet ortamında yer alan görüşlerin metin madenciliği yöntemleri ile otomatik tespitin yapılması amaçlanmıştır. Çalışmada, geliştirilen metin madenciliği veri ön işleme yazılımının en önemli ayırt edici özelliğinin, yazılımın sadece FATİH projesine yönelik görüşlerinin veri ön işleme basamağında değil, istenilen amaca yönelik metin sınıflandırma işleminin veri ön işleme basamağında konudan bağımsız bir şekilde kullanılabilir olması ifade edilmiştir. Çalışmanın sonuç kısmında, FATİH projesine yönelik 444 görüş içeren metin dosyasındaki metinler TF-IDF ağırlıklandırma yöntemi ile vektörel olarak temsil edilerek sınıflandırma algoritmalarının model başarımleri ölçütleri karşılaştırıldığı ve en yüksek başarımın Ardışık Minimal Optimizasyon Algoritmasına ait olduğu belirtilmiştir. Y. Safalı [32], popüler sosyal ağlardan olan Facebook, Twitter, Instagram ve LinkedIn kullanıcılarının “Cumhur İttifakı” hakkındaki görüşlerini sınıflandırmayı amaçlamıştır. Çalışmada bu sınıflandırma için bu platformlardan veriler elde edilmiştir ve veri seti oluşturulmuştur. K-En Yakın Komşu, Karar Ağacı, Sıralı Minimum Optimizasyon ve Bayes Sınıflandırma algoritmaları ile oluşturulan modeller kullanıcıları Cumhur İttifakı’nı destekleyen ve desteklemeyen kullanıcılar olarak sınıflandırılmıştır. Çalışmanın sonuç kısmında Kappa İstatistik Testi ile yapılan karşılaştırma sonucunda en başarılı sonucu veren K-En Yakın Komşu sınıflandırma algoritmasının sosyal ağ kullanıcılarının Cumhur İttifakı hakkındaki görüşlerini %97 başarı oranı ile sınıflandırdığını belirtmişlerdir. İ. N. Yoldaş [33], yoruma açık cümleler üzerinde insanların yapmış oldukları sınıflandırma ve sözlük tabanlı yaklaşım ile yapılmış olan sınıflandırmanın karşılaştırılmasını amaçlamıştır. Çalışmada 63 farklı insan cümlelerini içeren veri seti kullanılmıştır.

Çalışmanın sonuç kısmında insanların kararsız kaldığı durumlarda hedeflenmiş olan sistemin nasıl bir sonuç üreteceği görüldüğü ve Türkçe doğal dil işleme çalışması yapmak isteyenleri cesaretlendirmek hedeflendiği belirtilmiştir. F. Doyranlı Tayşi [34], ses verisi kullanılarak duygusal durum tespitini hedefleyen, Türkçe'ye özgü bir sistem geliştirmeyi amaçlamıştır. Çalışmada kızgınlık, üzüntü, heyecan, mutluluk, çaresizlik ve nötr duygularına ait 2194 adet kayıttan oluşan bir veri seti oluşturulmuştur ve çeşitli sınıflandırma algoritmaları kullanılmıştır. Çalışmanın sonuç kısmında veri setleri üzerinde yapılan deneyler sonucunda Rastgele Orman algoritmasının daha iyi sonuçlar elde ettiği belirtilmiştir. A. Güven [35], Türkçe belgeler üzerinde belge madenciliği yapmak amacıyla, Gizli Anlambilimsel Dizinleme yöntemini kullanarak ve kelimelere uygulanan n-gram yaklaşımını bu yöntemle birleştirmeyi amaçlamıştır. Çalışmada, Türkçe ve İngilizce belgeleri içeren veri kümeleri kullanılmıştır ve n-gram destekli GAD ve GAD yöntemlerinin karşılaştırmaları yapılmıştır. Çalışmanın sonucunda n-gram destekli GAD yönteminin daha seçici bir belge madenciliği yöntemi olduğunu belirtmişlerdir.

Ses kayıtlarıyla ilgili duygu analizi yapılan diğer çalışmalara ait literatür özeti aşağıda verilmiştir.

Maghilnan S. ve Rajesh Kumar M. [36], belirli bir hipotezi baz alan konuşma kayıtlarını ve konuşma tanıma araçlarını kullanmışlardır. Çalışmada insanların neden bahsettiklerini ve neler hissettiklerini anlamak yani duygu analizi için yeni bir model sunmayı amaçlamışlardır. Yapılan duygu analizinde Naive Bayes, Lineer Destek Vektör Makinesi ve VADER yöntemlerini kullanmışlardır. Sonuç kısmında ise modelin, yapay olarak oluşturulan veri kümesiyle iyi sonuçlar elde ettiğini fakat daha büyük bir veri kümesi için çalışmalarını sürdürdüklerini ifade etmişlerdir.

R. R. Sehgal ve arkadaşları [37], çağrı merkezini arayan müşterilerin duygu analizini yapmayı hedeflemişlerdir. Otomatik Konuşma Tanıma sistemlerinin de kullanıldığı bu çalışmada duygu analizi için Naive Bayes ve Max Entropy sınıflandırma algoritmaları kullanılmıştır. Çalışmanın sonuç kısmında ise yapılan analizde başarılı sonuçlar elde edildiği fakat gelişmekte olan ülkelerde yeterli kaynağın olmaması nedeniyle her ülkede aynı sonucu vermeyeceği belirtilmiştir.

K. Napier ve L. Shamir [38], 1950'lerden sonraki şarkılarda olan duyguları analiz etmeyi amaçlamışlardır. 1951'den 2016'ya kadar olan süreçte Billboard Hot 100'ün 6150 şarkısını bu analiz için kullanmışlardır. Çalışmada yöntem olarak Destek Vektör Makinesi algoritmasını ve Pearson Korelasyonunu kullanmışlardır. Sonuç kısmında ise öfke, iğrenme, korku, üzüntü ve vicdan gibi duyguların önemli ölçüde azaldığını ayrıca bununla beraber pop şarkı sözlerinde ifade edilen sevinç, güven duygularının azaldığını ifade etmişlerdir.

S. Cunningham ve arkadaşları [39], uluslararası duygusal dijital ses veri setini kullanarak, ses üzerindeki duyguların etkilerini ve makine öğrenim tekniklerini kullanarak sonuçları üzerinde analizler yapılmasını hedeflemişlerdir. Çalışmada makine öğrenim yöntemlerinden Regresyon modelleri ve Yapay Sinir Ağı modelleri kullanılmıştır. Sonuç kısmında ise Yapay Sinir Ağı modelinin, Regresyon modellerine göre daha iyi sonuçlar ortaya koyduğunu belirtmişlerdir.

O. M. Nezami ve arkadaşları [40], Farsça için ShEMO adı verilen duygusal konuşma veri tabanı sunmuşlardır ve bu veri tabanında sınıflandırıcılar karşılaştırılmış, çıkan sonuçlar diğer dillerle karşılaştırılmıştır. Veri tabanı, 3 saat 25 dakikalık konuşma verilerine eşdeğer 3000 yarı doğal ifadeyi ve beş duyguyu içermektedir. Çalışmada sınıflandırma algoritmalarından Destek Vektör Makinesi, K-En Yakın Komşu ve Karar Ağacı kullanılmıştır. Sonuç kısmında ise Destek Vektör Makinesinin en iyi sonuçları verdiği, gelecekteki çalışmalar için Gizli Markov Modeli ve Derin Sinir Ağları gibi sınıflandırma yöntemlerini kullanarak daha etkili bir tespit elde etmeyi amaçladıklarını ifade etmişlerdir.

H. Kaya ve arkadaşları [41], 3 ile 7 yaş arasındaki çocuklardan toplanan Rusça'daki ilk çocuk duygusal konuşma veri tabanını sunmuşlardır. Bu veri tabanı ile duygusal durumun ses ve konuşma özelliklerine yansımaları ve ontogeneze duygusal durumların oluşumunu incelemeyi amaçlamışlardır. Çalışmada bu veri tabanı üzerinde insan ve makine tahminlerinin karşılaştırmaları yapılmıştır. Yapılan deneylerin sonucunda yaş tahmininde, akustik tabanlı otomatik sistemlerin daha yüksek performans gösterdiği, konfor durumu ve cinsiyet sınıflandırmasında insan algı seviyelerinin daha iyi olduğu ifade edilmiştir.

C. Supriyanto ve arkadaşları [42], Mel-Frekans Cepstral Katsayılarının, Endonezya konuşma duygu tanıma üzerindeki performansını araştırmayı amaçlamışlardır. Çalışmada kullanılan veri seti Endonezya filmlerinden elde edilmiştir. Elde edilen ses kayıtları kızgın, üzgün, mutlu ve nötr diye dört sınıfa ayrılmıştır ve Destek Vektör Makinesi sınıflandırma algoritmasıyla sınıflandırılmıştır. Çalışmanın sonuç kısmında ise doğrusal çekirdek, polinom çekirdeği ve Radyal Temel Fonksiyon gibi çeşitli DVM çekirdek işlevlerinde yapılan karşılaştırmada doğrusal çekirdeğin en yüksek doğruluk oranını sağladığı belirtilmiştir.

S. Yaslan ve B. Günsel [43], sayısal ses sinyalleri kullanılarak duygu algılama gerçekleyen bir sistem sunmuşlardır. Bu çalışma için, 11 farklı kişiden 4 farklı duygu modunda söylenen 11 farklı cümleden oluşan bir ses veri tabanı ve Gauss RBF kernel kullanan Destek Vektör Makinesi sınıflandırıcısı kullanılmıştır. Sonuç kısmında ise %70'in üzerinde bir doğruluk oranı sağlandığı belirtilmiştir.

S. K. Bhakre ve A. Bang [44], ses sinyallerini dört duygu üzerinde sınıflandırarak bir duygu tanıma modeli oluşturmayı amaçlamışlardır. 2000 veriye sahip bir veri seti üzerinde Naive Bayes sınıflandırıcısını kullanmışlardır. Naive Bayes sınıflandırıcısını sesteki özellikler için pitch, enerji ve MFCC ile eğitmişlerdir. Yapılan deneyler sonucunda kızgın, mutlu, üzgün ve nötr duygularında sırasıyla %81, %78, %76, %77 doğruluk oranı elde etmişler ve önerilen sistemin gerçek zamanlı duygu tanınması yeteneğine sahip olduğunu belirtmişlerdir.

A. Davletcharova ve arkadaşları [45], insan konuşmasından gelen duyguları tanımak için deneysel bir çalışma yapmayı amaçlamışlardır. Yapılan deneyler için nötr, öfke, neşe ve üzüntü duyguları dikkate alınmıştır ve 30 denek üzerinden bir veri seti oluşturulmuştur. Yapılan duygu sınıflandırması sonrasında çalışmanın sonuç kısmında bir grup insandan elde edilen verileri dikkate almak yerine, bir kişiden toplanan verileri dikkate alınması gerektiğini belirtmişlerdir.

J. Gondohanindijo ve arkadaşları [46], duygusal konuşmayı tanıma problemini çözmeyi, ses işleme yoluyla sınıflandırmayı, duyguların sınıflandırılmasına dayanan ses kalıplarını tanımayı ve kullanılan bazı sınıflandırıcıların doğruluk düzeyini karşılaştırmayı amaçlamışlardır. Çalışmada kullanılan 639 veriden oluşan veri seti üzerinde Destek Vektör Makinesi, Rastgele Orman, Naive Bayes ve Sinir Ağı sınıflandırıcıları kullanılmıştır. Sonuç kısmında ise Destek Vektör Makinesi sınıflandırıcısının en iyi doğruluk ve hassasiyet seviyesine sahip olduğu ve F1-skorunda diğer üç sınıflandırıcıya göre daha iyi olduğu belirtilmiştir.

C. M. Lee ve S. S. Narayanan [47], konuşma sinyallerindeki duyguların akustik korelasyonları ile birlikte dil ve söylem bilgilerini kullanarak alana özgü duyguların tespitini araştırmayı amaçlamışlardır. Çalışmada bir çağrı merkezinden alınan veriler ile konuşmadaki olumlu ve olumsuz duyguları tespit etmeyi hedeflemişlerdir. Yapılan deneyler sonucunda akustik ve dil bilgileri birleştirildiğinde en iyi sonuçların elde edildiğini ifade etmişlerdir. Sonuç kısmında ise duygu sınıflandırmasında erkekler için %40.7 ve kadınlar için %36.4 oranında iyileştirme sağlandığını belirtmişlerdir.

L. Jie ve arkadaşları [48], MFCC ve RNN-LSTM'ye dayanan yeni bir konuşma duygu tanıma yöntemi sunmuşlardır. Çalışmada, konuşma duyguları örnekleri yeniden sınıflandırılmış ve öğretmenlerin konuşma duyguları veri tabanı oluşturulmuştur. Yapılan deneyler sonucunda önerilen yöntemle gerçekleştirilen iyileştirmeler sonucunda doğruluk oranı %85.32'ye ulaşmıştır. Çalışmanın sonucunda önerilen yöntemle, geleneksel yöntemlerden daha etkin bir sonuç elde ettiklerini ifade etmişlerdir.

A. Tripathi ve arkadaşları [49], ses kayıtlarını kullanarak duyguların algılanması ile ilgili yapılan çalışmalar hakkında bir derleme çalışma sunmuşlardır. Çalışmada, konuşma yoluyla duygu analizi ile ilgili son gelişmeler ve bununla ilgili çeşitli sorunlar ele alınmıştır. Çalışmanın sonuç kısmında duygu tanımlama alanında, veri tabanı seçiminin sonuçtaki iyi doğrulukta önemli bir rol oynadığı, MFCC algoritmasının, özellik çıkarma için çalışmaların çoğunda yaygın olarak kullanıldığını, çünkü gürültü durumunda diğer tekniklerden daha iyi performans gösterdiği ve Destek Vektör Makinesinin, duygu analizi sorununu çözmek için en sık kullanılan makine öğrenim algoritması olduğunu belirtmişlerdir.

S. Tuzcu [50], çevrimiçi bir kitap satış sitesinin kullanıcı yorumları üzerinde duygu analizi için öncelikle Python programlama dili kullanılarak Çok Katmanlı Algılayıcı algoritması kullanmayı amaçlamıştır. Çalışmada, Naive Bayes, Destek Vektör Makinesi ve Lojistik Regresyon algoritmaları kullanılmıştır ve sonuçları karşılaştırılmıştır. Çalışmanın sonuç kısmında ise Destek Vektör Makinesi, pozitif yorumları sınıflandırmada tüm algoritmalar içinde en iyi sonucu vermiş fakat negatif yorumları sınıflandırmada diğer algoritmalarından oldukça geride kaldığı, Naive Bayes ise bu üç algoritmadan farklı olarak negatif yorumları pozitif yorumlara göre daha iyi sınıflandırmış fakat genel sınıflandırma başarısı bu algoritmaların altında kalmış olduğu belirtilmiştir.

H. Nizam ve S. S. Akın [51], makine öğrenmesi yöntemlerinden denetimli öğrenme yaklaşımı kullanılarak sosyal medyada duygu analizi çalışması yapmayı amaçlamışlardır. Çalışmada, bazı gıda firmalarının çeşitli ürünlerine ait yapılan yorumlardan oluşturulan veri seti Twitter üzerinden elde edilmiştir. Veri setindeki veriler pozitif, negatif ve nötr olarak etiketlenmiştir ve Weka kütüphanesinde yer alan Naive Bayes, Rastgele Orman, Sıralı Minimum Optimizasyon, Karar Ağacı ve 1-En Yakın Komşular sınıflandırma algoritmalarının gösterdikleri başarımlarına etkisi incelenmiştir. Çalışmanın sonuç kısmında, dengeli veri setinin, dengesiz veri setine göre daha iyi performans sonuçları elde ettiği ve en iyi performansı gösteren sınıflandırma algoritmasının %72.3 ortalama doğruluk başarı oranıyla Sıralı Minimum Optimizasyon algoritmasının olduğu belirtilmiştir.

R. Aşlıyan ve K. Günel [52], En Yakın Komşu ve K-En Yakın Komşu metotlarını kullanarak metin içerikli dokümanların sınıflandırılmasını gerçekleştiren bir sistem geliştirmeyi amaçlamışlardır. Çalışmada, eğitim, otomobil, sağlık, spor ve teknoloji sınıflarından oluşan toplam 125 tane doküman ve sistemin test edilmesi için kullanılan her bir sınıf için 50, toplam 250 doküman içeren veri seti oluşturulmuştur. Çalışmanın sonuç kısmında En Yakın Komşu metodunun, K-En Yakın Komşu metoduna göre daha başarılı olduğu ve bütün sınıflar için %88.4 oranında başarı elde ettiği belirtilmiştir.

İ. O. Yiğit ve arkadaşları [53], görüşme kaydı metinlerini kullanarak görüşmenin içeriğinin duygu yönünden değerlendirilmesi, müşteri memnuniyetinin ve müşteri temsilcisi performansının ölçülmesini amaçlamışlardır. Çalışmada 400 ses kaydı olumlu/olumsuz olarak etiketlenmiştir ve Karar Ağacı, Destek Vektör Makinesi, K-En Yakın Komşu, Lojistik Regresyon ve Rastgele Orman sınıflandırma algoritmaları kullanılmıştır. Çalışmanın sonucunda çağrı merkezlerindeki görüşme kaydı metinlerin işlenmesi için piyasada kullanılan bir ürünün varlığının tespit edilemediğini belirtmişlerdir.

E. S. Erdem [54], tam ve genişletilebilir bir ses tabanlı duygu tanıma ve geri erişim çatısı yapısı sunmuştur. Çalışmada, 20 ses özniteliği üzerinde, Destek Vektör Makinesi kullanılarak kapsamlı analizler yapılmıştır ve çıkan sonuçlar değerlendirilmiştir. Sonuç kısmında ise yerli TV dizilerinden yeni bir veri kümesi oluşturulduğu ve ses sinyalinde duygu tanımda, öznitelik çıkarımı ve sınıflandırıcı optimizasyonu sonucunda, %11'in üzerinde ciddi bir başarı artışı sağlandığı belirtilmiştir.

C. Parlak [55], yeni bir duygu veri tabanı oluşturmayı ve insan sesinin akustik özniteliklerine dayanarak konuşmadaki duyguyu konuşmacı ve dilden bağımsız olarak tanımlamayı amaçlamıştır. Çalışmada kullanılan veri tabanları üzerinde Naive Bayes, SVM_SMO ve Bagging sınıflandırıcılar 988 ve 6669 özellik içeren iki farklı özellik konfigürasyonu ile uygulanmıştır. Sonuç kısmında ise tek setli çalışmalarda en başarılı sınıflandırıcı SMO, en yavaş sınıflandırıcı Bagging, başarı oranı en düşük sınıflandırıcı Naive Bayes sınıflandırma algoritmasının olduğu belirtilmiştir.

2. MATERYAL VE YÖNTEM

2.1 Veri Madenciliği

Teknolojinin gelişimiyle beraber gündelik hayatımızdaki birçok işlem internet ortamında gerçekleşmektedir. Her geçen gün artan internet kullanımı da bu ortamdaki veri sayısını artırmıştır. İşlenmeyen ve bir bilgi oluşturmeyen verilerin, anlamlı bir hal alması ve bir bilgi aktarımında bulunması için ortaya ‘ Veri Madenciliği ‘ kavramı çıkmıştır. Veri madenciliği, birden fazla verinin bulunduğu alandan anlamlı ve yararlı bilgilerin ortaya çıkarılması için işlenmesi ve kullanılması demektir. Bu kavram ilk kez 1983 yılında Michael C. Lovell tarafından kullanılmıştır [56].

Veri madenciliği, multi-disipliner bir alandır. İstatistik, makine öğrenmesi, yapay zekâ, veri tabanı ve görselleştirme gibi disiplinleri bir arada kullanma yetisine sahip bir alandır. Günümüzde veri madenciliği için birçok platform bulunmaktadır. Bu platformlar aracılığıyla veriler işlenip anlamlı sonuçlar elde edilebilmektedir. Programlama dili ile veri madenciliği yapan platformlardan Python ve R programlama dilleri günümüzde kodlama ile yapılan veri madenciliğinde popüler olan platformlardandır. Programlama dillerinin dışında sadece veri bilimi, veri madenciliği için kullanılan platformlarda bulunmaktadır. Bu platformlardan en bilindikleri Weka ve RapidMiner’dir. Veri madenciliği için kolay kullanımları, eklentileri ve performansları ile önemli bir yer edinmişlerdir.

Verinin işlenip anlamlı hale getirilmesi için veri madenciliği süreçlerine tabii tutulması gerekmektedir. Bu süreçler ve açıklamaları aşağıda verilmiştir.

- **Veri Filtreleme:** Veri madenciliğinde kullanılacak olan verinin veya veri setinin belirlenmesidir.
- **Veri Temizliği:** Verilerin içinden anlamsız, herhangi bir şey ifade etmeyen verilerin ayıklanmasıdır.
- **Veri Birleştirme:** Veriler arasında birbiriyle ilişkili olan, bütünlük barındıran verilerin birleştirilmesidir.
- **Veri Seçme:** Verilerin temizlenme aşamasından sonra bu veriler arasından yapılacak olan veri madenciliği için verilerin seçilmesi aşamasıdır.

- **Veri Dönüştürme:** Elde edilen verilerin veri madenciliğine uygun formata dönüştürülmesidir.
- **Veri Madenciliği:** Hazır olan veriler üzerinde veri madenciliği algoritmalarının uygulanmasıdır.
- **Örüntü Değerlendirme:** Bilgiyi temsil eden farklı örüntüleri tanımlamaktır.
- **Bilgi Sunumu:** Tüm bu işlemler sonucunda elde edilen sonuçların ve değerlendirmelerin kullanıcıya sunulmasıdır [56].

Yukarıda ifade edilen veri madenciliği süreçlerinden veri filtrelemeden veri madenciliğine kadar olan kısım ön işleme olarak da adlandırılmaktadır. Yani verinin, veri madenciliği sürecine hazır hale getirilme durumu olarak ifade edilebilir. Veri madenciliği sürecinden itibaren artık verinin algoritmalar ve yöntemlerle analiz işlemleri sonrasında kullanıcıya bilgi aktarımı aşamaları olarak belirtilebilir.

Veri madenciliğinde süreçlerin yanı sıra yöntemleri de önem arz etmektedir. Bu yöntemler ve açıklamaları aşağıda verilmiştir.

- **Sınıflandırma:** Bu yöntemle ile veriye uygun niteliklere göre önceden belirlenen sınıflardan veriye uygun olanı bulunup verinin oraya aktarılmasıdır.
- **Birliktelik Kuralları:** Veri tabanlarında veya veri kümelerinde verilerin birbirleriyle aralarında bağlantı olanları ve aralarındaki bağlantıyı saptamaya yarayan yöntemdir.
- **Kümeleme:** Verilerin daha iyi analizi için kendi aralarında ilişkilere göre alt sınıflara ayrılma durumudur.
- **Tahminleme:** Veri seti içerisinde eksik olan bilgilerin tahmin edilmesini sağlayan yöntemdir.
- **Aykırılık Analizi:** Veri kümesi ya da veri seti içerisinde aykırılık durumu gösteren verilerin tespit edilmesini sağlayan yöntemdir [57].

Veri madenciliği içerdiği yöntem ve süreçlerin sağladığı olumlu etkilerle günümüzde birçok alanda kullanılmaktadır. Bu alanlara örnek olarak bankacılık, e-ticaret, sigortacılık gibi alanlar gösterilebilir. Veri madenciliği ile çalışmalar hız kesmeden sürdürülmektedir.

2.2 Duygu Analizi

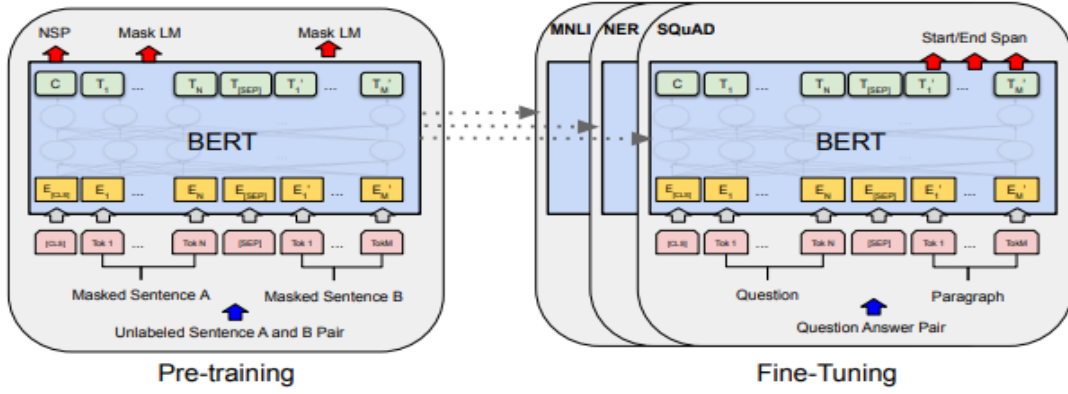
Duygu analizi, verinin olumlu, olumsuz veya tarafsız olup olmadığının analizini yapan yöntemdir. Veri bilimi için son zamanlarda önemli bir araştırma alanı haline gelmiştir. Araştırmacıların çalışma alanıyla ilgili varsayımları arttıkça kendi içinde de alt sınıflara ayrılmıştır. Bu sınıflar; Doküman Seviyesinde, Cümle Seviyesinde, Özellik Temelli ve Karşılaştırmalı'dır [58].

Duygu analizi için doküman sınırlaması bulunmamaktadır. Veriler, ses verisi, metin verisi vb. türde olabilir. Verilerin ön işlem adımlarından sonra yapılan analiz sonrasında elde edilen sonuçlarla farklı varsayımlar ortaya atılabilmektedir. Örneğin; kişinin bir alışveriş sitesinde aldığı bir ürün için yaptığı yorum ürün hakkında bilgi verebilir ve ürünü alacak kişiler için bir fikir oluşturabilir. Duygu analizinin bu noktada öneminden bahsedilebilir.

Tez çalışmasında, duygu analizi için BERT modeli ve TextBlob kütüphanesi kullanılmıştır. Bu bölümde de bu yöntemler ile ilgili tanımlamalara ve açıklamalara yer verilmiştir.

2.2.1 BERT

BERT çok çeşitli doğal dil işleme görevleri hakkında en gelişmiş sonuçları elde eden yeni bir eğitim öncesi dil temsili yöntemidir. 2018 yılında Google'da Jacob Devlin ve arkadaşları tarafından geliştirilen bu model yapısı cümleyi hem sağdan sola hem de soldan sağa taraftan değerlendirmektedir. Böylelikle kelimelerin birbirleriyle olan ilişkilerini daha iyi ortaya koymaktadır. Modelin yapısında, BookCorpus ve Wikipedia veri setleri bulunmaktadır [59]. BERT model yapısı Şekil 2.1'de verilmiştir.



Şekil 2.1 BERT Model Yapısı [60]

Model, Masked Language Modeling (MLM) ve Next Sentence Prediction (NSP) yöntemleri ile eğitilmektedir. MLM tekniğinde, maskelenen kelime, açık şekilde beslenen kelimelerden yararlanılarak tahmin edilmeye çalışılır. MLM yönteminde daha çok kelimeler arasında ilişkiler üzerinde durulmaktadır. Bir diğer yöntem olan NSP’de ise cümleler arasındaki ilişki üzerinde durulmaktadır. NSP ile model eğitim aşamasındayken iki cümle yapısı arasındaki ilişkiye bakılarak ikinci cümlenin ilkinin devamı olup olmadığına bakılır ve tahmin yapılmaya çalışılır. Eğitim aşamasında gerçekleştirilen optimizasyon ile bu iki yöntem kullanılırken ortaya çıkan kaybın minimuma indirilmesi amaçlanmıştır [59]. BERT model yapısı verilen metin dokümanı hem sağdan hem soldan incelemesi, yapısında bulunan yöntemlerle eğitim aşamasında daha iyi bir öğrenme sağlamaktadır.

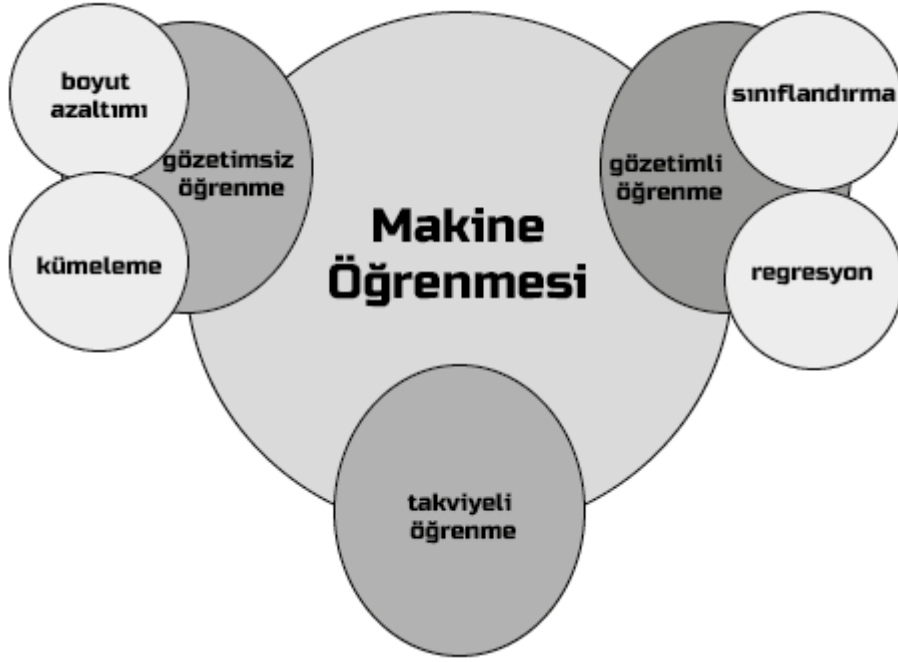
BERT modeli ilk çıktığı zamanlarda dünya üzerinde en çok kullanılan dilleri kapsayan bir yapı içermekteydi. Zaman içerisinde çeşitli dil yapıları da eklenerek BERT modelinin çok dilli yapısı üzerinde geliştirmeler devam etmiştir. BERT modelinin Türkçe dili ilgili geliştirdiği modelin adı BERTurk’tür. Bu model yapısının gelişimini Prof. Dr. Kemal OFLAZER, Reyhan YENİTERZİ ve Stefan SCHWETER’in oluşturduğu ekip yapmıştır. Böylelikle Türkçe dilini içeren doğal dil işleme çalışmaları BERT modeli aracılığıyla da yapılmaya başlanmıştır. Türkçe BERT modeli aracılığıyla ad-varlık tanımlayıcı, soru cevaplama, doküman sınıflandırma ve duygu analizi gibi doğal dil işleme çalışmaları yapılmaktadır.

2.2.2 TextBlob

TextBlob, metin verilerini işlemek için kullanılan Python programlama dili kütüphanesidir. Etiketleme, isim öbeği çıkarma, duygu analizi gibi birçok doğal dil işleme görevlerini yerine getirmektedir [61]. TextBlob kütüphanesi Python programlama diline kolay entegrasyonu ile duygu analizi işlemlerinde avantaj sağlamaktadır. Duygu analizi için duygu (sentiment) nesnesi kullanılmaktadır. TextBlob'un duyarlılık işlevi iki özellik, kutuplaşma (polarity) ve öznellik döndürür. Bu nesneye ait bileşenlerden kutuplaşma, cümlelerin negatif, pozitif ve nötr durumlarını belirlemek amacıyla kullanılmıştır. Duygu analizinde verinin polarite değeri aralığı -1 ile 1 arasındadır. -1 değeri negatif, 0 değeri nötr ve 1 değeri de pozitif olarak belirlenmiştir. Böylelikle verilerin skor değerleri yapılan duygu analizi sonucunda görülebilmektedir.

2.3 Makine Öğrenmesi

Makine öğrenmesi, istatistiksel ve matematiksel işlemler aracılığıyla veriler üzerinden çıkarımlarda bulunarak tahminler yapılmasını sağlayan sistemlerin bilgisayarlarla modellenmesidir [62]. Bir diğer tanımı ise bilgisayarların tıpkı insanlar gibi öğrenmesini sağlamak amacıyla çeşitli algoritma ve yöntemlerin geliştirilmesi için çalışılan bilim alanıdır. Makine öğrenmesi, öğrenme tekniklerine göre üç gruba ayrılmaktadır. Bunlar; Supervised (Gözetimli), Unsupervised (Gözetimsiz) ve Reinforcement (Takviyeli)'dir. Bu öğrenme tekniklerini gösteren görsel Şekil 2.2'de verilmiştir.



Şekil 2.2 Makine Öğrenmesi, Öğrenme Yöntemleri [63]

Gözetimli Öğrenme: Etiketlenmiş veriler üzerinden yapılan gözetimli öğrenme sürecidir [64]. Bu sayede algoritmaya diğer verilerin nasıl etiketlenmesi gerektiğini belirtir. Gözetimli öğrenme de kendi içinde, verilerin türüne göre ikiye ayrılmaktadır. Veri çıktıları kategorik ise sınıflandırma (classification), nümerik ise regresyon (regression) algoritmaları kullanılır [62].

Gözetimsiz Öğrenme: Bu öğrenme yönteminde de etiketlenmemiş veri üzerinden bilinmeyen yapıları veya verileri tahmin etme yöntemidir. Kullanılan algoritmaların kendi kendilerine çıkarımlar yapması beklenir. Gözetimsiz öğrenme yöntemi içyapısında iki unsuru barındırır. Bunlar; kümeleme (clustering) ve boyut azaltımı (dimensionally reduction)'dır. Kümeleme; Gözlemleri, verileri homojen bölgelere ayırır. Boyut azaltımı ise gözlemlerdeki, verilerdeki mevcut bulunan özellik sayısını azaltarak daha iyi bir öğrenme sunar [64].

Takviyeli Öğrenme: Davranış psikolojisinden hareketle yola çıkan öğrenme yöntemidir [62]. Çevreden alınan geri bildirimlere göre davranışların belirlendiği yöntem olarak da tanımlanabilir. Bu özelliği ile diğer iki öğrenme yönteminden ayrılmaktadır.

Tez çalışmasında da makine öğrenim algoritmalarından beş tanesi kullanılmıştır. Gözetimli öğrenmenin sınıflandırma yönteminde kullanılan bu algoritmalar, çalışmada da bu amaçla ve sonuç verilerinde elde edilen ölçütler üzerinde değerlendirmelerde bulunmak için kullanılmıştır. Bu algoritmalar; Naive Bayes, Rastgele Orman, Destek Vektör Makinesi, Karar Ağacı ve K-En Yakın Komşu algoritmalarıdır. Bu algoritmalara ait açıklamalar aşağıda verilmiştir.

2.3.1 Naive Bayes

Naive Bayes algoritması, Bayes teoremine dayalı bir öğrenme algoritmasıdır. Tembel (lazy) yapıya sahip olmasına rağmen düzensiz yapıdaki veri setlerinde de çalışabilmektedir [65].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Bayes teoreminin matematiksel ifadesi yukarıdaki gibidir. $P(A|B)$; B olayı gerçekleştiği durumda A olayının meydana gelme olasılığı, $P(B|A)$; A olayı gerçekleştiği durumda B olayının meydana gelme olasılığı ve $P(A)$ ve $P(B)$; A ve B olaylarının olasılıklarıdır.

Naive Bayes sınıflandırma algoritması, tüm koşullu olasılıkların çarpımıdır [66]. Bu sınıflandırma algoritmasıyla verilen eğitim ve test verilerinin doğru bir şekilde sınıflandırılması yapılabilmektedir. NB sınıflandırma algoritmasına ait formül Denklem 2'deki gibi verilebilir.

$$P(Class_j | x) = \frac{P(x | Class_j)P(Class_j)}{P(x)} \quad (2)$$

Burada verilmiş bir x tahminci olasılığı, j sınıfına ait olasılık ve eğitim kümesinden bulunan bu sınıf tahminci olasılığı ile x tahmincisinin bu sınıfa ait olasılığı hesaplanabilir. Naive Bayes algoritmasının kullanım alanları ise şunlardır; spam filtreleme, duyarlılık analizi, çok sınıflı tahmin, öneri sistemleri ve benzeri alanlardır [65].

2.3.2 Rastgele Orman

Rastgele Orman sınıflandırma algoritması, birden fazla karar ağacının avantajlarını kullanarak daha uyumlu ve daha iyi sonuçlar üreten modeller ortaya koyarak sınıflandırma işlemini en iyi şekilde yapmaya çalışan bir sınıflandırma algoritmasıdır. Rastgele Orman algoritması çoğu zaman büyük bir sonuç üreten, esnek, kullanımı kolay bir makine öğrenim algoritmasıdır. Rastgele Orman algoritmasının en önemli avantajı ise hem sınıflandırma hem de regresyon işlemlerinde kullanılabilen bir algoritma olmasıdır [67]. RO algoritmasına ait matematiksel ifade aşağıda verilmiştir. Denklem 3'te $fi_{sub}(i)$: i özelliğinin önemi ve $ni_{sub}(j)$: j düğümünün önemi olmak üzere;

$$fi_i = \frac{\sum_{j:i \text{ özelliğine ayrılan } j \text{ düğümü}} ni_j}{\sum_{k \in \text{bütün düğümler}} ni_k} \quad (3)$$

Bunlar bütün özellik önem değerlerinin toplamına bölünerek 0 ve 1 arasında bir değere normalize edilirse yani;

$$norm \ fi_i = \frac{fi_i}{\sum_{j \in \text{bütün özellikler}} fi_j} \quad (4)$$

RO düzeyindeki nihai özelliğin önemi, tüm ağaçların ortalamasıdır. Her ağaçtaki özelliğin önem değerinin toplamı hesaplanır ve toplam ağaç sayısına bölünür:

$$RFfi_i = \frac{\sum_{j \in \text{bütün ağaclar}} normfi_{ij}}{T} \quad (5)$$

Burada, RFf_i RO modelindeki tüm ağaçlardan hesaplanan özelliğinin önemi, $^{norm} f_{ij}$ j ağacındaki i için normalleştirilmiş özellik önemi, T ise toplam ağaç sayısıdır. Bu algoritma aynı zamanda sınıflandırıcı kategorik değerler için değerlendirilebilmektedir. Rastgele Orman algoritmasıyla ayrıca over-fitting (aşırı uyum) problemlerini de gidermektedir.

2.3.3 Destek Vektör Makinesi

Destek Vektör Makinesi, 1995 yılında Vapnik tarafından istatistiksel öğrenme teorisi ve VC-boyut teorisinden türetilmiştir. DVM, ikili sınıflandırma problemini çözmek için geliştirilmiştir. Destek Vektör Makinesi, karmaşık verilerin yüksek doğrulukla işleme özelliği olarak da tanımlanmaktadır. DVM'deki ana amaç ise eğitilmiş veri örneklerini önceden tanımlanmış sayıda sınıfa ayıran hiper düzlem bulmayı amaçlamaktadır [68] [69]. DVM'nin matematiksel ifadeleri Denklem 6'da verilmiştir.

$$\begin{aligned} \vec{x}_i * \vec{w} + b &= 1; \text{pozitif sınıflar için} \\ \vec{x}_i * \vec{w} + b &= -1; \text{negatif sınıflar için} \end{aligned} \quad (6)$$

Sabit marjlı DVM, hiper düzlemi geçen destek vektörleri konusunda çok kesindir. Hiper düzlemin marjını maksimize etmek için, sabit marjlı DVM optimizasyon problemine dönüşür yani;

$$\begin{aligned} \min_{\vec{w}, b} \quad & \frac{1}{2} \|\vec{w}\|^2 \\ y_i (\vec{x}_i \cdot \vec{w} + b) & \geq 1; \\ i &= 1, \dots, n \end{aligned} \quad (7)$$

Bu sınıflandırma algoritmasının avantajları olarak şunlar söylenebilir; yüksek boyutlu uzayda diğer algoritmalara göre daha etkinlerdir, karar fonksiyonu için farklı çekirdek fonksiyon yapıları kullanmaktadırlar ve karar noktalarında kullandıkları eğitim fonksiyonları sayesinde belleğin daha verimli kullanılmasına olanak sağlamaktadırlar [70].

Algoritma kendi içinde de doğrusal ve doğrusal olmayan olarak ikiye ayrılmaktadır. Doğrusal olmayan yöntemde çekirdek fonksiyonları yöntemi kullanılmaktadır. Destek Vektör Makinesi algoritması, metin ve görüntü sınıflandırmada, biyolojik bilim alanlarında ve elle yazılmış karakterlerin tanınması gibi alanlarda kullanılmaktadır.

2.3.4 Karar Ağacı

Karar Ağacı, özellik, hedef ve sınıflamaya göre karar düğümlerinden ve yaprak düğümlerinden oluşan ağaç yapısı şeklinde bir model yapısı oluşturan gözetimli sınıflandırma algoritmasıdır [71]. Çeşitli bilim alanlarına uyarlanabilmektedir.

Yapısında bulunan en üst düğüme root (kök) ve diğer düğüm yapılarına da leaf (yaprak) adı verilmektedir. Karar Ağacı algoritmaları veri setini küçük parçalara bölerek geliştirilmektedir. Böylelikle algoritma yapısında büyük kayıpların önüne geçilerek daha küçük kayıpların olması sağlanmıştır.

Algoritma yapısı, ağaç yapısına benzetildiği için kullanılan matematiksel ifadelerde de bu durum göz önünde bulundurulmuştur. Belirli bir $T=t$ ağacı ve $D=d$ eğitim veri seti için, t 'nin d üzerinde ne kadar iyi çalıştığına dair olasılıksal tahmin yani $P(T=t|D=d)$ bulunabilir. İdeal ağaç ise maksimum $P(T=t|D=d)$ değerine sahip olmalıdır.

$$P(Y = y_1 | X = x, D = d) = \sum_{\text{Bütün Ağaçlar}} P(Y = y_1 | T = t, X = x, D = d) P(T = t | D = d) \quad (8)$$

Buna göre, bir ağaç yapısı eğitim verilerini ne kadar iyi ayırırsa $P(T=t|D=d)$ ifadesi de maksimum değeri alır.

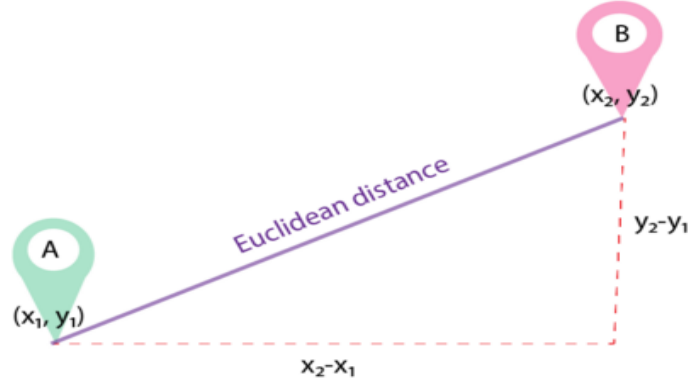
Karar ağaçlarının avantajları ise şöyledir; hem kategorik hem de nümerik verileri işleyebilirler, birden fazla çıktısı olan problemler için de çözüm üretebilirler, yorumlanması ve anlaşılması kolaydır ve kullanılan ağaç yapıları görsel halinde sunulabilir [72].

2.3.5 K-En Yakın Komşu

K-En Yakın Komşu algoritması, eklenecek olan yeni verinin mevcut veri kümelerine uzaklığını hesaplayarak, k sayıda yakın komşuluğa bakarak sınıflandırma işlemini gerçekleştirmektedir. Uzaklık hesaplamaları için, Öklid (Euclidean), Minkowski ve Manhattan uzaklık hesaplamalarını kullanmaktadır. Algoritma yapısı eski, gürültülü eğitim verilerine karşı dirençli olduğundan günümüzde halen kullanılmakta olan popüler bir sınıflandırma algoritmasıdır. Algoritma işleyişi ise şu şekildedir; ilk olarak k parametresi belirlenir, uzaklık hesaplamaları gerçekleştirilir, komşular bulunur ve veri etiketlenmesiyle işleyiş tamamlanır. Algoritmanın hesaplamalarında kullanılan Öklid mesafe fonksiyonuna göre (x_1, y_1) ve (x_2, y_2) noktaları arasındaki mesafe Denklem 9'a göre hesaplanır.

$$dist((x_2, y_2), (x_1, y_1)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (9)$$

Bu formülün iki boyutlu grafiği Şekil 2.3’de verilmiştir.



Şekil 2.3 Öklid Mesafe Fonksiyonu

KEYK algoritmasında, bir K değeri için, algoritma veri noktasının K’ya en yakın komşularını bulacak ve K’nın komşularını en fazla veriye sahip noktalara atayacaktır. Matematiksel ifadesi Denklem 10’da verilmiştir.

$$P(y = j | X = x) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j) \quad (10)$$

K-En Yakın Komşu algoritması, hem regresyon hem sınıflandırma işlemlerinin yanı sıra birçok alanda sınıflandırma işlemleri içinde kullanılmaktadır [73].



3. DENEYSEL BULGULAR

3.1. Veri Setinin Elde Edilmesi

Veri, bilgisayar ortamındaki bilgilerin, programlar aracılığıyla işlenebilmesi için derlenmiş ve formüle edilmiş halidir [74]. Günümüzde teknolojik gelişmelerin artmasıyla beraber büyük bir veri dünyası da oluşmaya başlamıştır. İnsanların çeşitli konularda internet üzerinden yaptıkları e-işlemler ve kullandıkları sosyal mecralardan yaptıkları paylaşımlar, elde edilen verilerin büyük kümeler haline gelmesine olanak sağlamıştır.

Tez çalışmasında Türkçe ses kayıtlarını içeren bir veri seti ele alınmıştır. Bu veri seti, Mozilla tarafından başlatılan ses ve konuşma tanıma yazılımları için ücretsiz bir veri tabanı oluşturulmak üzere geliştirilen Common Voice platformundan elde edilmiştir. İnsanların nasıl konuştuklarını makinalara öğretmek amacıyla oluşturulan bu platformda çeşitli diller için ses kayıtları ve bu ses kayıtlarına ait metin dokümanları bulunmaktadır.

Kişiler bu platform aracılığıyla seslerini bu platformun veri tabanına kaydedebilmekte ve kaydın doğruluğunu test edebilmektedir. Common Voice’da bir ses kaydının süreci şu şekildedir;

- Cümle veri tabanında olan cümlelerden belirlenen bir cümlenin ses kaydı yapılır,
- Dinleyiciler tarafından dinlenmek ve doğrulanmak üzere kayıt kuyruğa eklenir,
- Dinleyici kaydın doğruluğunu kontrol eder ve ‘Evet’ oyu aldığında geçerli kabul edilir,
- Ses kaydının doğruluğu iki farklı kullanıcı tarafından onaylandığında kayıt, veri tabanına eklenir eğer kayıt onay almazsa geçersiz kayıtlar veri tabanına eklenir [75].

Çalışmamızda da bu ses kayıtlarında Türkçe olanları bu platform aracılığıyla elde edilmiştir. Türkçe ses kayıtlarını içeren 592 MB boyutundaki dosya yaklaşık 20.760 ses verisini içermektedir. Bu veri setine ait bilgiler ise şu şekildedir; Yaş aralığı ve oranları; 19-29 %47, 30-39 %23, <19 %4, 40-49 %3, 50-59 %1 şeklindedir. Bu veri seti için %71 erkek , %6 kadın şeklinde bir cinsiyet oranı bilgisi de paylaşılmıştır. Birçoğu tekrar eden kayıtlarından oluşan bu ses kayıtlarının tüm metin dokümanları birleştirilmiştir. Elde edilen metin dokümanları üzerinde eleme işlemi gerçekleştirilmiştir. Bir anlam ifade etmeyen, tekrar eden ve sadece sayı belirten

veriler ayrıştırıldıktan sonra 5001 adet metin verisi elde edilmiştir. Böylelikle çalışma için kullanılacak olan veri seti hazır hale getirilmiştir. Veri setinin ilk 19 verisine ait görsel Şekil 3.1’de verilmiştir.

	A	B	C	F	G	I
1	client_id	path	sentence	age	gender	locale
2	f05abfcb4f3020	common_voice	Üye sayımız otuz iki, otuz üç, hatta belki otuz dörde çıkacak.			tr
3	f05abfcb4f3020	common_voice	Hepsi de hasta adamlar, çocuklar ve yaşlılardı.			tr
4	f05abfcb4f3020	common_voice	İlk tur seçimler yirmi yedi Aralık Pazar günü yapılacak.			tr
5	f05abfcb4f3020	common_voice	Yüz binlerce insan tasarruflarını kaybetti.			tr
6	f05abfcb4f3020	common_voice	Sonuçların beş Aralık'ta belli olması bekleniyor.			tr
7	f05abfcb4f3020	common_voice	Öte yandan Priştine ise sevincini ifade etti.			tr
8	f05abfcb4f3020	common_voice	Bazı durumlarda maaşlar yüzde doksan arttı.			tr
9	f05abfcb4f3020	common_voice	Yargı reformu özellikle endişe yaratan bir konu.			tr
10	1b80f76a6bca3e	common_voice	Bu programlar da onlara bu şansı verecek.	twenties	male	tr
11	1b80f76a6bca3e	common_voice	Otopark şirketleri karara ateş püskürüyorlar.	twenties	male	tr
12	1b80f76a6bca3e	common_voice	İyi günler dilemenin çok güzel bir yolu.	twenties	male	tr
13	1b80f76a6bca3e	common_voice	Ve tabii ki İnternet.	twenties	male	tr
14	1b80f76a6bca3e	common_voice	Yasada korsanlık suç sınıfına sokuluyor.	twenties	male	tr
15	1b80f76a6bca3e	common_voice	Saraybosna'ya bir niyet mektubu gönderildi bile.	twenties	male	tr
16	1b80f76a6bca3e	common_voice	Bazı planlar üzerinde çalışılıyor.	twenties	male	tr
17	1b80f76a6bca3e	common_voice	Hayvanlar kimsenin iyi niyetine bağlı değil.	twenties	male	tr
18	1b80f76a6bca3e	common_voice	Seçimlerde toplam on iki aday yarıştı.	twenties	male	tr
19	1b80f76a6bca3e	common_voice	Proje yaklaşık sekiz yüz milyon avroya mal olacak.	twenties	male	tr
20	4131b422c65a5c	common_voice	Bugüne dek bunu kimse yapamadı.	twenties	male	tr

Şekil 3.1 Veri Setinin İlk 19 Verisi

Aşağıdaki bölümlerde öncelikle hem BERT modelinde hem de TextBlob kütüphanesinde duygu tespitleri yapılmıştır. Bu duygu tespiti sonrasında üç duygu veri seti elde edilmiştir. Bu üç duygu veri seti üzerinde önce Python programlama dilinde ile sonra RapidMiner platformunda, beş makine öğrenim algoritmasını kullanarak analizler yapılmıştır ve sonuçları verilmiştir.

3.2 Türkçe Ses Kayıtlarından Elde Edilen Veriler Üzerinde Python Programlama Dili ile Duygu Tespiti

Duygu analizi / tespiti, verinin olumlu, olumsuz veya nötr olup olmadığını belirlemektedir. Tez çalışmasında elde edilen veri seti üzerinde üç farklı yöntemle duygu tespiti / analizi yapılmıştır.

Bu yöntemler; BERT modelinin iki farklı varyasyonu ve TextBlob kütüphanesidir. Bu yöntemler Python programlama dili aracılığıyla geliştirilmiştir. Python, Guido Van Rossum tarafından geliştirilmiştir. Nesne yönelimli ve yüksek seviyeli bir programlama dili olan Python, kolay yazım dili ve çoğu işletim sistemine uyarlabilir yapısıyla günümüzde popüler bir programlama dili haline gelmiştir [76]. Özellikle son yıllarda veri bilimi ve makine öğrenmesi için önemli bir yer

edinen Python programlama dili, bu alanlarla ilgili geliřtirmelerin odak noktası haline gelmiřtir.

Python programlama diline ait kodların tarayıcı üzerinde çalışma imkânı sunan Google Colaboratory platformu üzerinde bu uygulamalar gerekleřtirilmiřtir. Bu platform, kod bloklarının hızlı bir řekilde alıřtırılmasına, eklenti ve kütüphanelerin kolayca yüklenmesine ve paylařılmasına olanak saęlamaktadır. alıřmamızda da duygu tespiti ařaması, bu platform aracılıęıyla ve Python programlama dili ile gerekleřtirilmiřtir.

3.2.1 İki Duygu Etiketine Sahip BERT Modeli ile Duygu Tespiti

alıřmamızda, Türke duygu tespiti gerekleřtirmek için oluřturulan BERT modelinden yararlanılmıřtır. Model yapısı içinde cümleler pozitif ve negatif řeklinde deęerlendirilmektedir. alıřmada ilk olarak gerekli kütüphanelerin ve veri setinin yüklenmesi iřlemleri gerekleřtirilmiřtir. Bu iřlem adımının sonrasında cümle yapıları tokenlerine-belirtelerine ayrıldıktan sonra modele aktarılır. Burada cümlelerin pozitiflik veya negatiflik skorları ile etiketleri belirlenir.

alıřmanın kod kısmında pipeline nesnesi aracılıęıyla veriler, peř peře seri biimde tanımlanmıř birtakım iřlemlerden geer. Bu iřlemler sırasıyla verilerin tokenlerine ayrılması ve sonrasında yüklenen BERT modeline aktarılmasını ifade eder. Bu iřlemlere ait kod görseli řekil 3.2’de verilmiřtir.

```
#Burada BERT'in Türke Duygu Tespiti yapabilmesi için oluřturulan model yüklenir.  
#Model yapısı içinde iki duygu etiketi bulunmaktadır.Bunlar; pozitif ve negatiftir.  
  
from transformers import AutoModelForSequenceClassification, AutoTokenizer, pipeline  
model = AutoModelForSequenceClassification.from_pretrained("savasy/bert-base-turkish-sentiment-cased")  
  
#BERT yapısı içinde bulunan tokenizer nesnesi kullanılır ve verileri tokenlerine ayırma iřlemi gerekleřtirilir.  
tokenizer = AutoTokenizer.from_pretrained("savasy/bert-base-turkish-sentiment-cased")  
  
#pipeline nesnesi verinin peřpeře seri biimde tanımlanmıř birtakım iřlemlerden gemesine olanak saęlar.  
#Önce tokenleřtirme iřlemi gerekleřtiriliyor ve veriler modele verilerek duygu tespitinin gerekleřtirilmesi saęlanıyor.  
sa= pipeline("sentiment-analysis", tokenizer=tokenizer, model=model)
```

řekil 3.2 İki Duygu Etiketine Sahip BERT Modeli Kod Yapısı

Kod kısmında çıkan sonuçlar etiket, skor ve cümle sütunlarına ayrılarak Excel dosyasına aktarılmıştır. Veri setinde bulunan cümlelerin karşısında duygu etiketleri (pozitif - negatif) ve skorları yazdırılmıştır. Gerçekleştirilen duygu tespiti sonrasında 2956 negatif ve 2045 pozitif duygu etiketine sahip cümle elde edilmiştir. Veri seti içinde oranlamaya bakıldığında %59 oranında negatif ve %41 oranında pozitif cümle bulunmaktadır. Çıktı sonuçlarının bir kısmını gösteren görsel Şekil 3.3'te verilmiştir.

```
[{'label': 'positive', 'score': 0.9095399975776672, 'sentence': 'Etkinlik yirmi iki Mayıs'a kadar açık kalacak.'}]
[{'label': 'negative', 'score': 0.9859170317649841, 'sentence': 'İnsan doğasının kusurundan doğan bir durum.'}]
[{'label': 'negative', 'score': 0.9882176518440247, 'sentence': 'Kasedin etkisi büyük oldu.'}]
[{'label': 'negative', 'score': 0.9988734126091003, 'sentence': 'Bunun değişmesini beklemiyorum.'}]
[{'label': 'negative', 'score': 0.9816184043884277, 'sentence': 'Aralarında pek çok mülteci var.'}]
[{'label': 'positive', 'score': 0.5698550343513489, 'sentence': 'Bir sonraki ay bir anayasa kabul edildi.'}]
[{'label': 'positive', 'score': 0.7273528575897217, 'sentence': 'Biz realiteyi dikkate alıyoruz.'}]
[{'label': 'positive', 'score': 0.6747391223907471, 'sentence': 'Barış müzakereleri hiçbir zaman kolay olmaz.'}]
[{'label': 'positive', 'score': 0.9280513525009155, 'sentence': 'Yaşanan gelişmeler pek çok kimseyi şaşırtıyor.'}]
[{'label': 'positive', 'score': 0.5782375931739807, 'sentence': 'Dört yıl sonra dava gün ışığına çıktı.'}]
[{'label': 'positive', 'score': 0.5698100328445435, 'sentence': 'Bağış önerileri açık rekabetle seçiliyor.'}]
[{'label': 'negative', 'score': 0.8981885313987732, 'sentence': 'Saldırının sorumluluğunu henüz üstlenen olmadı.'}]
[{'label': 'negative', 'score': 0.8290243744850159, 'sentence': 'Proje Mart ayında başlatıldı.'}]
[{'label': 'negative', 'score': 0.5087615251541138, 'sentence': 'Erdoğan eleştirileri reddetti.'}]
[{'label': 'positive', 'score': 0.8988512754440308, 'sentence': 'Anlaşma bir Ocak'ta yürürlüğe girecek.'}]
[{'label': 'positive', 'score': 0.7391859889030457, 'sentence': 'Projeler altı bilimsel alanı kapsıyorlar.'}]
```

Şekil 3.3 İki Duygu Etiketine Sahip BERT Modeli ile Yapılan Duygu Tespiti Sonuçları

3.2.2 Üç Duygu Etiketine Sahip BERT Modeli ile Duygu Tespiti

Türkçe duygu analizi için geliştirilen bu model yapısında duygu etiketlerine nötr duygu etiketinin eklenmesi verilerin duygu tespitinde farklı sonuçlar ortaya koymuştur. Model yapısında negatif duygu etiketi -1, nötr duygu etiketi 0 ve pozitif duygu etiketi 1 ile ifade edilmiştir. Çalışmanın kod kısmında Onnx çalışma zamanı kütüphanesi kullanılmıştır. Onnx çalışma zamanı, bir transformatör modeli yüklenirken çoğu optimizasyonu otomatik olarak uygular. Bu özelliği aracılığıyla model yapısında gerekli optimizasyon işlemlerinin gerçekleştirilmesine olanak sağlar. Türkçe duygu analizi için oluşturulan model yapısı ve 5001 veriden oluşan veri setinin yüklenmesine ait işlem adımları gerçekleştirilir. Bu işlem adımlarından sonra, öncelikle veriler üzerinde tokenleştirme işlemleri gerçekleştirilir. Modele belirlenen boyutlarda veri gönderimi sağlanır. Maskeleye işlemleri gerçekleştirilir ve belirlenen veri gönderimi boyutu ile veri seti üzerinde işlemler yapılır.

Veri seti üzerinde yapılan işlemler sonrasında *final_results* değişkenine verilerin duygu etiketleri aktarılır. Veri setinde bulunan cümleler üzerinde yapılan tokenleştirme, maskeleme ve modelde yapılan işlemler sonucunda duygu tespiti gerçekleştirilir. Bu işlemlere ait kod yapıları Şekil 3.4'te verilmiştir.

```
ort_session = ort.InferenceSession("distilbert.onnx") # Türkçe Duygu Analizi için oluşturulan Model ve Tokenizer yüklenir.
tokenizer = BertTokenizerFast.from_pretrained('dbmdz/distilbert-base-turkish-cased')

#Öncelikle veriler üzerinde tokenleştirme işlemleri gerçekleştirilir.
#Modele belirlenen boyutlarda veri gönderimi sağlanır.
#Maskeleme işlemleri gerçekleştirilir ve belirlenen veri gönderim boyutu ile veri seti üzerinde işlemler yapılır.
#Veri seti üzerinde yapılan işlemler sonrasında final_results değişkenine verilerin duygu etiketleri aktarılır.
tokenized_sentences = tokenizer.batch_encode_plus(data, padding="max_length", truncation=True, max_length=128)
batch_size = 40
N = len(data)
final_results = []
for idx in range(math.ceil(N/batch_size)):
    print("Inference:", len(final_results), "/", N)
    inp_ids = tokenized_sentences["input_ids"][idx*batch_size:min((idx+1)*batch_size, N)]
    att_mask = tokenized_sentences["attention_mask"][idx*batch_size:min((idx+1)*batch_size, N)]
    real_size = len(inp_ids)
    if len(inp_ids)<batch_size:
        inp_ids.extend([inp_ids[0]]*(batch_size-real_size))
        att_mask.extend([att_mask[0]]*(batch_size-real_size))
    x = {'input_ids': np.array(inp_ids), 'input_atns': np.array(att_mask)}
    outputs = ort_session.run(None, x)
    values = np.argmax((outputs[0]), 1)[:real_size]
    final_results.extend([x-1 for x in values.tolist()])
```

Şekil 3.4 Üç Duygu Etiketine Sahip BERT Modeli Kod Yapısı

Sentence ve sentiment şeklinde iki sütuna ayrılan çıktılar, Excel dosyasına aktarılır. Yapılan duygu tespiti sonucunda 2430 nötr, 1989 negatif ve 582 pozitif duygu etiketine sahip cümle elde edilmiştir. Veri seti içerisinde oranlamaya bakıldığında %48,7 oranında nötr, %39,7 oranında negatif ve %11,6 oranında pozitif cümle bulunmaktadır. Çıktı sonuçlarının bir kısmını gösteren görsel Şekil 3.5'te verilmiştir.

	Sentence	Sentiment
0	Hayatta küçük şeyleri kovalıyor ve yine küçük ...	-1
1	Seydiu şimdi iki mevkiyi de kaybetti.	-1
2	Fakat öte yandan, hayatta gelişim sağlayabilme...	0
3	Seferler Haziran ayında başlıyor.	0
4	Siyasette temiz insanlara ihtiyacımız var.	1
...	...	...
4996	Kamuoyu ikiye bölünmüş durumda.	-1
4997	Ancak Hiseni suçlamaktan da geri kalmadı.	-1
4998	Kosova başkentindeki yolcu sayısı arttı.	0
4999	Bu tamamen Kosova yönetimine bağlı.	-1
5000	Bu olay daha büyük yankılara yol açacak mı?	-1

Şekil 3.5 Üç Duygu Etiketine Sahip BERT Modeli ile Yapılan Duygu Tespiti Sonuçları

3.2.3 TextBlob Yöntemi ile Duygu Tespiti

TextBlob, duygu tespiti için çalışmamızda kullanılan bir diğer yöntemdir. TextBlob' un Türkçe dil desteği olmadığı için öncelikle veri setimiz İngilizce' ye çevrilmiştir. Python dilinde yazılan kod kısmında ise öncelikle gerekli kütüphanelerin yükleme işlemleri gerçekleştirilmiştir. Veri setinin İngilizce 'ye çevrilmiş hali yüklendikten sonra TextBlob'a ait nesneler kullanılmıştır ve cümlelerin skorlarıyla beraber cümleler bir Excel dosyası içine aktarılmıştır. TextBlob'ta skoru 0 çıkan cümleler nötr olarak değerlendirilmektedir. Çalışmamızda skoru 0'a yakın çıkan cümlelerinde nötr olarak değerlendirilmesi için bu aralığı -0,1 ve 0,1 arasına çekilmiştir. 0,1'den büyük skora sahip cümleler pozitif, -0,1'den küçük cümleler ise negatif olarak değerlendirilmiştir. Bu işlemlere ait kod görseli Şekil 3.6'da verilmiştir.

```

for str in df.values:
    sentence = str[0]
    blob = TextBlob(sentence) #TextBlob kütüphanesi kullanılarak veri setinde bulunan cümleler blob nesnesine aktarılmıştır.
    row = []
    polarity = blob.sentiment.polarity #Cümlelerin duygu analizini gerçekleştirmek için sentiment ve polarity bileşenleri kullanılmıştır.

    #TextBlob'ta cümleler negatif ise -1, nötr ise 0 ve pozitif ise 1 ile ifade edilmektedir.
    #Çıkan duygu analizi skorlarındaki değerlendirmede -0,1 ve 0,1 arasındaki skorlar nötr olarak değerlendirilmiştir.
    #0,1'den büyük skorlara sahip cümleler pozitif, -0,1'den küçük skorlara sahip cümleler negatif olarak değerlendirilmiştir.

    print("Sentence:",sentence)
    if (polarity >= (-0.1)) and (polarity <= 0.1):
        print("neutral:",polarity)
        polarity = 0
    elif polarity > 0.1:
        print("positive",polarity)
        polarity = 1
    else:
        print("negative",polarity)
        polarity = -1

```

Şekil 3.6 TextBlob Kütüphanesinin Veri Seti Üzerinde Uygulanması

Yapılan duygu tespiti sonucunda 3614 nötr, 974 pozitif ve 413 negatif duygu etiketine sahip cümle elde edilmiştir. Veri seti içinde oranlamasına bakıldığında %72 oranında nötr, %19 oranında pozitif ve %9 oranında negatif cümle bulunmaktadır. TextBlob ile yapılan duygu tespiti sonucunda veri setinde bulunan verilerin büyük oranda nötr olduğu gözlemlenmiştir. TextBlob ile elde edilen sonuçların bir kısmına ait görsel Şekil 3.7’de verilmiştir.

```

Sentence: Turkish invention was produced in Germany.
neutral: 0.0
Sentence: However, the international community is waiting for concrete steps.
neutral: 0.07500000000000001
Sentence: The relevant negotiation chapter has not yet been opened.
positive 0.4
Sentence: Altun believes more is needed.
positive 0.5
Sentence: The festival has come a long way.
neutral: -0.05
Sentence: Is the name dispute coming to an end?
neutral: 0.0
Sentence: The ancient building was found during archaeological excavations.
neutral: 0.0
Sentence: The event will continue until June sixth.
neutral: 0.0
Sentence: The portal consists of seven sections.
neutral: 0.0
Sentence: According to Skopje, Kosovo is dealing with this issue.
neutral: 0.0
Sentence: Other parties could not reach the required threshold.
negative -0.125
Sentence: How do you see the relations between Albania and Serbia?
neutral: 0.0

```

Şekil 3.7 TextBlob Kütüphanesi ile Yapılan Duygu Tespiti Sonuçları

3.3 CoLab’da Python Kullanarak Makine Öğrenim Algoritmalarıyla Veri Analizi

Çalışmada duygu tespiti yapılan veri setinin analizi için günümüzde popüler olan makine öğrenmesi bilimi ele alınmıştır. Çalışmanın bir diğer aşaması olan makine öğrenim modelleri ile analiz için beş adet makine öğrenim modeli belirlenmiştir. Bu modeller, yapılan literatür özetindeki kullanım yoğunluğuna göre belirlenmiştir. Literatür çalışmalarından yapılan çıkarımda en çok kullanılan makine öğrenim modeli Destek Vektör Makinesi modeli (19 kez) olmuştur. Diğer makine öğrenim modelleri ise şunlardır: K-En Yakın Komşu (9 kez), Naive Bayes (8 kez), Karar Ağacı (8 kez) ve Rastgele Orman algoritmasıdır(5 kez). Çalışmada elde edilen veri setinde, duygu tespiti yapıldıktan sonra ortaya çıkan 3 veri seti için ayrı ayrı kod blokları oluşturulmuştur. Bunlar BERT yöntemiyle elde edilen veri setleri ve TextBlob ile elde edilen veri setidir. Kullanılacak olan veri setleri üzerinde veriyi temizleme ön işlemleri gerçekleştirilmiştir ve %85 eğitim verisi ve %15 test verisi olarak ayrılmıştır.

Bilgisayarlarda makine öğrenimi için verilerin vektörizasyon işlemlerine tabi tutulması gerekmektedir. Çalışmamızda vektörizasyon işlemlerini gerçekleştiren CountVectorizer ve TF-IDFVectorizer faktörleri kullanılmıştır. Bu iki yöntemin daha iyi anlaşılması için yöntemlere ait açıklamalar aşağıda verilmiştir.

CountVectorizer: Bir metin belgesi koleksiyonunu bir belirteç sayısı matrisine dönüştürür. Vektör gösteriminden önce metinlerin işlenmesini de yapan bu vektörizasyon yöntemi metnin işlevselliğini de artırmaktadır [77]. CountVectorizer yöntemi, özellikle metin sınıflandırma ve analizlerinde etkin bir şekilde kullanılmaktadır.

TF-IDFVectorizer: Bir kelimenin bulunduğu doküman içindeki değerini gösteren istatistiki yöntemle hesaplanan ağırlık faktörüdür [78]. Belge içinde sıklıkla geçen terimler, belge içinde baskın hale gelirler ve diğer terimler önüne geçerler. Bu durumun önüne geçmek için ve çözüm üretmek için TF-IDFVectorizer yöntemi kullanılır [79]. TF-IDFVectorizer yöntemi, durak kelimeleri filtrelemede, metin sınıflandırması gibi işlemlerde kullanılmaktadır.

Bu iki yöntem arasındaki farkı şöyle ifade edebiliriz: CountVectorizer’de terim farklı dökümanlarda geçse dahi o kelimenin geçtiği miktarı matriste belirtir. TF-IDFVectorizer’ın amacı ise, bir belgede sık görülen ancak belgeler arasında olmayan sözcükleri vurgulamaktır. Bu iki yöntem her modelde ayrı ayrı uygulanmıştır. Elde edilen sonuçlar arasında farklı değerlendirmeler için katkı sağlamışlardır.

Analiz sonucunda ortaya çıkan doğruluk (accuracy), duyarlılık değeri (precision), anma değeri (recall) ve F1-skoru (F1-score) terimlerine ait açıklamalar aşağıda verilmiştir.

Doğruluk: Doğru sınıflandırmanın toplama bölümüdür. Yani; doğrular (negatif ve pozitif) / toplam veri olarak da nitelendirebilir.

Duyarlılık Değeri: Yapılan testler sonucunda ölçümlerin birbirine ne derecede yakın olduğunu gösteren başarımlar ölçütü terimidir. Bu değer hesaplanırken sınıflandırılmış pozitif örnek sayısının, toplam pozitif örnek sayısına bölünmesiyle bu değere ulaşılır. Bu değer aralığı 0 ile 1 arasındadır.

Anma Değeri: Hedefi tutturma oranı olarak bilinmektedir. Yani gerçekte ulaşılması gereken bilgilere ne oranda ulaşılmış olduğunu gösteren bir değerdir. Bu değer hesaplanırken doğru sınıflandırılmış ilgili pozitif örnek sayısının, toplam ilgili belgelerin sayısına bölünmesiyle bu değere ulaşılır.

F1-Skoru: Duyarlılık ve anma değerlerinin harmonik ortalamaları hesaplanarak elde edilir. Bu değer 0 ile 1 arasında olur. Yapılmış olan testin doğruluğunu ifade eden değerdir.

3.3.1 İki Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi

Çizelge 3.1’de pozitif ve negatif duygu etiketine sahip BERT modeliyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla gerçekleştirilen analiz sonuçları verilmiştir. Çizelgede duyarlılık değeri, anma değeri, F1-skor ve doğruluk metrikleri sütunları ikiye ayrılmıştır. Sütunun sol tarafındaki sonuçlar CountVectorizer ile, sağ tarafındaki sonuçlar TF-IDFVectorizer ile elde edilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.1 BERT (Pozitif -Negatif) Modeliyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
NB	Pozitif	0.68	0.71	0.55	0.44	0.61	0.54	%70	%68
	Negatif	0.71	0.68	0.81	0.87	0.75	0.76		
RO	Pozitif	0.69	0.64	0.42	0.49	0.52	0.56	%67	%67
	Negatif	0.66	0.68	0.86	0.80	0.75	0.73		
DVM	Pozitif	0.63	0.67	0.56	0.56	0.59	0.61	%68	%69
	Negatif	0.70	0.71	0.76	0.79	0.73	0.75		
KA	Pozitif	0.59	0.54	0.42	0.51	0.49	0.52	%63	%61
	Negatif	0.65	0.65	0.78	0.68	0.71	0.67		
KEYK	Pozitif	0.68	1.00	0.17	0.07	0.27	0.13	%61	%60
	Negatif	0.60	0.59	0.94	1.00	0.74	0.74		

3.3.2 Üç Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi

Çizelge 3.2’de pozitif, negatif ve nötr duygu etiketine sahip BERT modeliyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla gerçekleştirilen analiz sonuçları verilmiştir. Çizelgede duyarlılık değeri, anma değeri, F1-skor ve doğruluk metrikleri sütunları ikiye ayrılmıştır. Sütunun sol tarafındaki sonuçlar CountVectorizer ile, sağ tarafındaki sonuçlar TF-IDFVectorizer ile elde edilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar, mavi renk ile belirtilen hücreler, nötr veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.2 BERT (Pozitif -Negatif-Nötr) Modeliyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
NB	Pozitif	0.44	0.00	0.11	0.00	0.17	0.00	%67	%65
	Negatif	0.64	0.64	0.69	0.64	0.67	0.64		
	Nötr	0.71	0.66	0.77	0.80	0.74	0.73		
RO	Pozitif	0.40	0.45	0.19	0.20	0.25	0.28	%64	%64
	Negatif	0.67	0.65	0.50	0.54	0.57	0.59		
	Nötr	0.64	0.65	0.85	0.82	0.73	0.73		
DVM	Pozitif	0.39	0.40	0.31	0.29	0.34	0.34	%66	%67
	Negatif	0.66	0.67	0.66	0.66	0.66	0.67		
	Nötr	0.70	0.71	0.73	0.75	0.71	0.73		
KA	Pozitif	0.35	0.35	0.21	0.29	0.26	0.32	%61	%58
	Negatif	0.59	0.56	0.54	0.50	0.56	0.53		
	Nötr	0.64	0.63	0.74	0.71	0.69	0.67		
KEYK	Pozitif	0.56	0.33	0.07	0.01	0.12	0.03	%54	%58
	Negatif	0.50	0.63	0.54	0.39	0.52	0.48		
	Nötr	0.58	0.57	0.64	0.86	0.61	0.68		

3.3.3 TextBlob Kütüphanesi ile Yapılan Duygu Tespitinin Makine Öğrenmesi Algoritmalarıyla Analizi

Çizelge 3.3'te TextBlob kütüphanesiyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla yapılan analiz sonuçları verilmiştir. Çizelgede duyarlılık değeri, anma değeri, F1-skor ve doğruluk metrikleri sütunları ikiye ayrılmıştır. Sütunun sol tarafındaki sonuçlar CountVectorizer ile, sağ tarafındaki sonuçlar TF-IDFVectorizer ile elde edilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar, mavi renk ile belirtilen hücreler, nötr veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.3 TextBlob Kütüphanesiyle Duygu Tespiti Yapılan Verilerin Makine Öğrenim Algoritmalarıyla Analizi

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
NB	Pozitif	0.72	0.89	0.55	0.24	0.63	0.38	%80	%78
	Negatif	0.36	0.00	0.08	0.00	0.14	0.00		
	Nötr	0.83	0.77	0.94	0.99	0.88	0.87		
RO	Pozitif	0.81	0.83	0.69	0.69	0.74	0.75	%86	%86
	Negatif	0.67	0.67	0.37	0.30	0.47	0.41		
	Nötr	0.88	0.87	0.95	0.96	0.91	0.91		
DVM	Pozitif	0.74	0.77	0.65	0.66	0.69	0.71	%83	%84
	Negatif	0.59	0.64	0.33	0.23	0.43	0.34		
	Nötr	0.87	0.86	0.93	0.95	0.90	0.90		
KA	Pozitif	0.79	0.54	0.72	0.75	0.75	0.74	%85	%84
	Negatif	0.55	0.73	0.47	0.50	0.50	0.52		
	Nötr	0.89	0.90	0.92	0.90	0.91	0.90		
KEYK	Pozitif	0.76	0.89	0.09	0.06	0.17	0.11	%75	%75
	Negatif	0.50	1.00	0.02	0.02	0.03	0.02		
	Nötr	0.75	0.75	0.99	1.00	0.85	0.85		

3.4 Makine Öğrenim Algoritmalarının Analiz Sonuçlarında Elde Edilen Karmaşıklık Matrislerine Ait Sonuçlar

Makine öğrenim algoritmaları ile yapılan analiz sonuçlarında karmaşıklık matrisleri de elde edilmiştir. Karmaşıklık matrislerinin daha iyi anlaşılması için bu matrislerin ne olduğu ve nasıl yorumlanması gerektiğine dair açıklamalar aşağıda verilmiştir.

Karmaşıklık Matrisi: Elde edilen sonuçlar üzerinde bilgi veren ölçüm tablosudur [80]. Özellikle sınıflandırma işlemlerinde sıklıkla kullanılan karmaşıklık matrisleri sonuçlar hakkında da detay bilgilere ulaşılmasına olanak sağlamaktadır. Elinizde bulunan durumlar için gerçek değerler ile tahmin edilen değerler için belli sonuçlar üretir. Bu sonuçlar için değerlendirmelerde kullanılan algoritmanın ne kadar doğru ne kadar yanlış veya ne kadar tarafsız sonuç ürettiği hakkında bilgi sahibi olmayı sağlamaktadır.

3.4.1 İki Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri

Karmaşıklık matrisinde negatif ve pozitif duygu etiketlerine ait gerçek değerler ve tahmin değerleri gösterilmiştir. Bu sonuçları yorumlamak için karmaşıklık matrislerinde kullanılan terimler aşağıda verilmiştir.

Gerçek Negatif: Gerçek değeri negatif ve tahmin edilen değeri de negatif olan verilerdir.

Gerçek Pozitif: Gerçek değeri pozitif ve tahmin edilen değeri de pozitif olan verilerdir.

Yanlış Pozitif: Gerçek değeri negatif ve tahmin edilen değeri pozitif olan verilerdir.

Yanlış Negatif: Gerçek değeri pozitif ve tahmin edilen değeri negatif olan verilerdir.

Çizelge 3.4'te pozitif ve negatif duygu etiketine sahip BERT modeliyle yapılan analiz sonucunda ortaya çıkan karmaşıklık matrisidir. Tahmin Edilen Değerler altında bulunan Negatif ve Pozitif duygu durumları iki sütuna ayrılmıştır. Sütunun sol tarafı CountVectorizer ile, sağ taraf ise TF-IDFVectorizer ile elde edilen sonuçları göstermektedir.

Çizelge 3.4 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler			
			Negatif		Pozitif	
NB		Negatif	348	374	83	57
		Pozitif	143	180	177	140
RO		Negatif	371	343	60	88
		Pozitif	187	162	133	158
DVM		Negatif	328	341	103	90
		Pozitif	141	141	179	179
KA		Negatif	337	295	94	196
		Pozitif	184	158	136	162
KEYK	Negatif	405	431	26	0	
	Pozitif	266	297	54	23	

3.4.2 Üç Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri

Karmaşıklık matrisinde negatif, pozitif ve nötr duygu etiketlerine ait gerçek değerler ve tahmin değerleri gösterilmiştir. Üç duygu etiketine sahip bu karmaşıklık matrisindeki terim sayısı da artmıştır. Ortaya çıkan yeni terimlere ait tanımlar aşağıda verilmiştir.

Gerçek Nötr: Hem tahmin edilen değerlerde hem de gerçek değerlerde nötr olan verilere verilen terimdir.

Yanlış Nötr-Negatif: Gerçek değeri nötr olup tahmin edilen değeri negatif olan verilerdir.

Yanlış Nötr-Pozitif: Gerçek değeri nötr olup tahmin edilen değeri pozitif olan verilerdir.

Yanlış Pozitif-Nötr: Gerçek değer pozitif ve tahmin edilen değer nötr olan verilerdir.

Yanlış Negatif-Nötr: Gerçek değer negatif ve tahmin edilen değer nötr olan verilerdir.

Çizelge 3.5'te pozitif, negatif ve nötr duygu etiketine sahip BERT modeliyle yapılan analiz sonucunda ortaya çıkan karmaşıklık matrisidir. Tahmin Edilen Değerler altında bulunan Negatif, Pozitif ve Nötr duygu durumları iki sütuna ayrılmıştır. Sütunun sol tarafı CountVectorizer ile, sağ taraf ise TF-IDFVectorizer ile elde edilen sonuçları göstermektedir.

Çizelge 3.5 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler					
			Negatif		Pozitif		Nötr	
NB		Negatif	213	213	7	7	87	87
		Pozitif	37	37	8	8	30	30
		Nötr	83	83	3	3	283	283
RO		Negatif	152	165	18	16	137	126
		Pozitif	24	24	14	15	37	36
		Nötr	52	65	3	2	314	302
DVM		Negatif	202	203	20	20	85	84
		Pozitif	20	22	23	22	32	31
		Nötr	85	78	16	13	268	278
KA		Negatif	165	154	21	28	121	125
		Pozitif	28	27	16	22	31	26
		Nötr	86	95	9	13	274	261
KEYK		Negatif	165	119	3	1	139	187
	Pozitif	35	18	5	1	35	56	
	Nötr	131	52	1	1	237	316	

3.4.3 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisleri

TextBlob kütüphanesi ile yapılan analizler sonucunda üç duygu etiketine sahip karmaşıklık matrisi elde edilmiştir.

Çizelge 3.6’da TextBlob kütüphanesiyle yapılan analiz sonucunda ortaya çıkan karmaşıklık matrisidir. Tahmin Edilen Değerler altında bulunan Negatif, Pozitif ve Nötr duygu durumları iki sütuna ayrılmıştır. Sütunun sol tarafı CountVectorizer ile, sağ taraf ise TF-IDFVectorizer ile elde edilen sonuçları göstermektedir.

Çizelge 3. 6 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler					
			Negatif		Pozitif		Nötr	
NB		Negatif	5	0	4	0	51	60
		Pozitif	2	0	76	33	59	104
		Nötr	7	0	25	4	522	550
RO		Negatif	22	18	3	2	35	40
		Pozitif	4	3	94	94	39	40
		Nötr	7	6	19	17	528	531
DVM		Negatif	20	14	5	4	35	42
		Pozitif	3	3	89	90	45	44
		Nötr	11	5	26	23	517	526
KA		Negatif	28	30	3	3	29	27
		Pozitif	5	4	99	103	33	30
		Nötr	18	22	24	35	512	497
KEYK		Negatif	1	1	0	0	59	59
	Pozitif	0	0	13	8	124	129	
	Nötr	1	0	4	1	549	553	

3.5 Python Programlama Dili ile Eğitim ve Test Veri Setleri Oranlarının Analizi

3.5.1 BERT Modeli (Pozitif - Negatif)

BERT'in iki duygu etiketine sahip modelinden elde edilen veri seti üzerinde, makine öğrenim algoritmalarıyla analiz yapılmıştır. Bu analizlerde doğruluk oranı üzerinde durulmuştur. Eğitim ve test veri setlerindeki oranların değişimine göre makine öğrenim algoritmalarının doğruluk oranları gözlemlenmiştir.

Çizelge 3.7 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15		Eğitim:%90 Test:%10		Eğitim:%91 Test:%9		Eğitim:%92 Test:%8		Eğitim:%93 Test:%7	
	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF
NB	0.70	0.68	0.71	0.70	0.70	0.70	0.70	0.69	0.71	0.70
RO	0.67	0.66	0.69	0.66	0.66	0.65	0.67	0.65	0.66	0.63
DVM	0.70	0.70	0.69	0.70	0.69	0.69	0.70	0.70	0.72	0.72
KA	0.64	0.62	0.64	0.61	0.65	0.61	0.64	0.62	0.65	0.60
KEYK	0.62	0.67	0.63	0.68	0.63	0.67	0.61	0.68	0.61	0.69

3.5.2 BERT Modeli (Pozitif – Negatif – Nötr)

BERT'in üç duygu etiketine sahip modelinden elde edilen veri seti üzerinde, makine öğrenim algoritmalarıyla analiz yapılmıştır. Bu analizlerde doğruluk oranı üzerinde durulmuştur. Eğitim ve test veri setlerindeki oranların değişimine göre makine öğrenim algoritmalarının doğruluk oranları gözlemlenmiştir.

Çizelge 3.8 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15		Eğitim:%90 Test:%10		Eğitim:%91 Test:%9		Eğitim:%92 Test:%8		Eğitim:%93 Test:%7	
	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF
NB	0.67	0.66	0.69	0.66	0.68	0.67	0.70	0.68	0.72	0.70
RO	0.63	0.62	0.62	0.63	0.62	0.65	0.62	0.65	0.63	0.64
DVM	0.66	0.67	0.65	0.67	0.64	0.66	0.63	0.66	0.65	0.68
KA	0.62	0.59	0.62	0.61	0.60	0.60	0.62	0.59	0.62	0.58
KEYK	0.49	0.59	0.50	0.49	0.49	0.48	0.46	0.47	0.50	0.48

3.5.3 TextBlob

TextBlob ile elde edilen veri seti üzerinde, makine öğrenim algoritmalarıyla analiz yapılmıştır. Bu analizlerde doğruluk oranı üzerinde durulmuştur. Eğitim ve test veri setlerindeki oranların değişimine göre makine öğrenim algoritmalarının doğruluk oranları gözlemlenmiştir.

Çizelge 3.9 TextBlob Kütüphanesi ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15		Eğitim:%90 Test:%10		Eğitim:%91 Test:%9		Eğitim:%92 Test:%8		Eğitim:%93 Test:%7	
	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF	CountVec	TF-IDF
NB	0.80	0.78	0.78	0.75	0.76	0.74	0.77	0.73	0.76	0.72
RO	0.86	0.86	0.85	0.85	0.85	0.85	0.84	0.85	0.85	0.86
DVM	0.84	0.84	0.82	0.81	0.83	0.80	0.83	0.81	0.82	0.81
KA	0.85	0.85	0.84	0.83	0.84	0.84	0.84	0.85	0.83	0.83
KEYK	0.76	0.75	0.74	0.73	0.72	0.72	0.73	0.72	0.71	0.71

3.6 RapidMiner’da Makine Öğrenim Algoritmalarıyla Veri Analizi

RapidMiner, veri madenciliği ve makine öğrenmesi gibi günümüzde popüler olan alanlar için geliştirmeler ve analizlerin gerçekleştirilebileceği bir ortam sağlar. Kolay eklenti seçenekleri, basit ara yüzü ve kullanımıyla birçok kullanıcı tarafından tercih edilmiştir. Tez çalışmasının da bir diğer analiz programı olan RapidMiner’ın seçiminde bu avantajlar etkili olmuştur. Python programlama dilinde olduğu gibi RapidMiner’da da beş makine öğrenim modeli ile analiz yapılmıştır. Bu analizlerde her bir algoritma için model ve çıktı yapısı, sonuç ve veri sayılarını gösteren görseller elde edilmiştir. Oluşturulan model yapısının anlaşılabilir olması için gerekli açıklamalar aşağıda verilmiştir.

Retrieve: Bu operatör depoda depolanan bilgilere erişebilmekte ve bunları işleme yükleyebilmektedir.

Set Role: Bir veya daha fazla niteliğin rolünü değiştirmek için kullanılmaktadır.

Nominal to Text: Seçilen nominal özniteliklerin türünü metinle değiştirmektedir. Ayrıca, bu özniteliklerin tüm değerlerini karşılık gelen dize değerleriyle eşleştirir.

Process Documents from Data: Tokenize, durak kelimelerin çıkarımı ve verilerin küçük harfe çevirimi gibi işlemleri yapan kısaca model içindeki ön işlem operatörüdür.

Split Data: Verilerin eğitim ve test veri oranlarının belirlenmesini sağlayan operatördür.

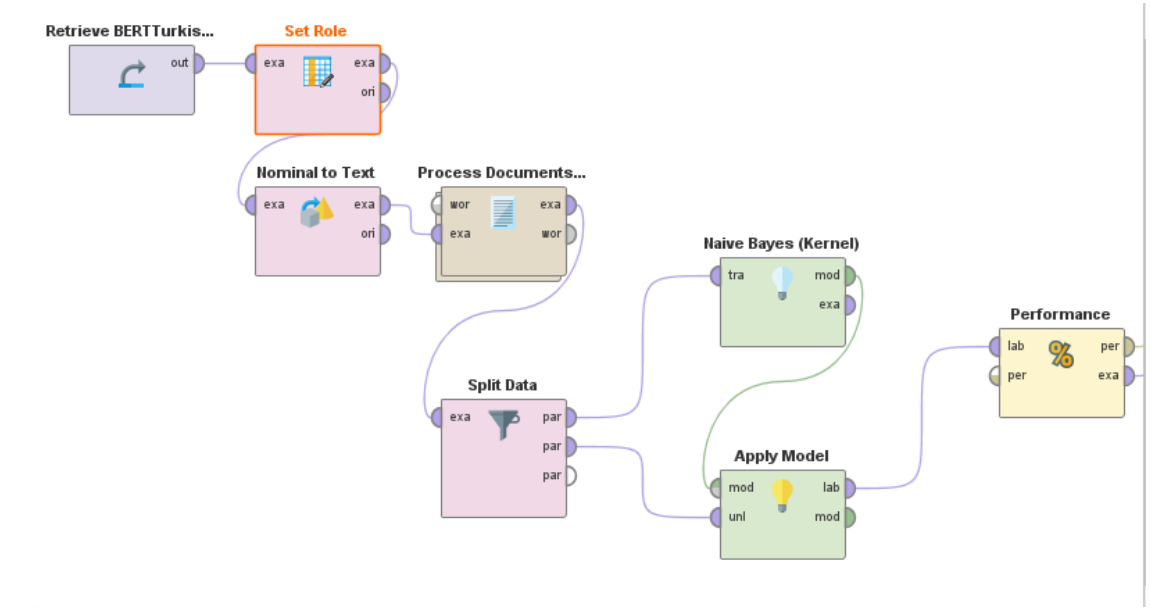
Apply Model: Bu operatör aracılığıyla seçilen algoritmanın modelin veri seti üzerinde uygulanması sağlanır.

Performance: Sınıflandırma görevlerinin istatistiksel performans değerlendirmesi için kullanılır. Sınıflandırma görevinin performans ölçütü değerlerinin bir listesini sunar.

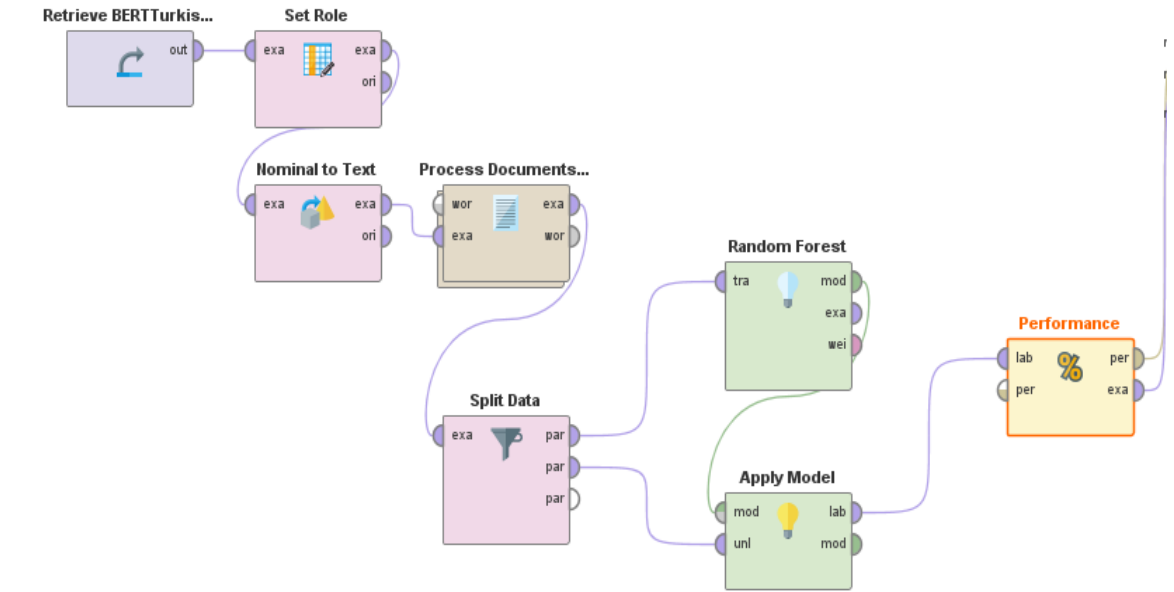
RapidMiner ile gerçekleştirilen analizde vektörizasyon yöntemi olarak TF-IDFVectorizer yöntemi kullanılmıştır. Verinin temizlenmesi analize hazır hale getirilme işlemlerinde ise durak kelimelerde Türkçe durak kelimeler eklentisi desteği RapidMiner’da bulunmamaktadır. BERT modelleriyle yapılan analizlerde, bu işlemin gerçekleştirilebilmesi için Türkçe durak kelimelerin olduğu bir .txt dosyası yüklenmiştir. Makine öğrenim algoritmalarından Naive Bayes yerine daha iyi performans gösteren Naive Bayes algoritmasının bir türü olan Naive Bayes (Kernel) kullanılmıştır. Naive Bayes operatörünün aksine, Naive Bayes (Kernel) operatörü sayısal niteliklere uygulanabilmektedir. Elde edilen sonuçların sayısal biçimde ifade edilmesi bu algoritmanın kullanımında da etkili olmuştur. RapidMiner’da yapılan analiz sonucunda duyarlılık değeri ve anma değeri metriklerine ait sonuçlar elde edilmiştir ve tablolarda gösterilmiştir. Duygu tespiti gerçekleştirilen her model ve algoritma için sonuçlar görsel ve tablo halinde verilmiştir.

3.6.1 İki Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi

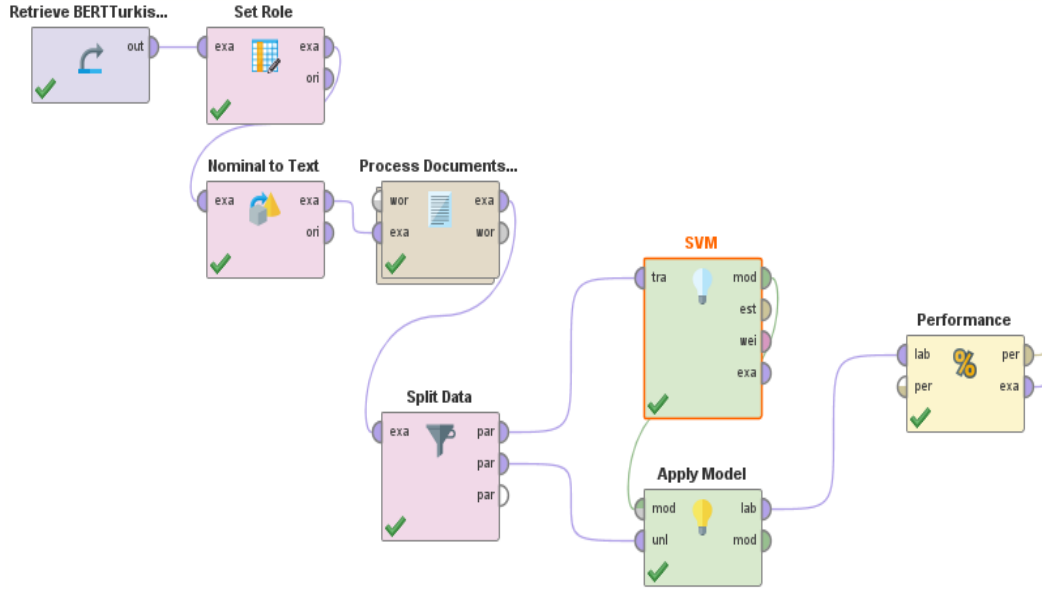
RapidMiner’da, beş makine öğrenim algoritmasıyla yapılan analiz model yapıları aşağıdaki görsellerde her algoritma için ayrı ayrı verilmiştir.



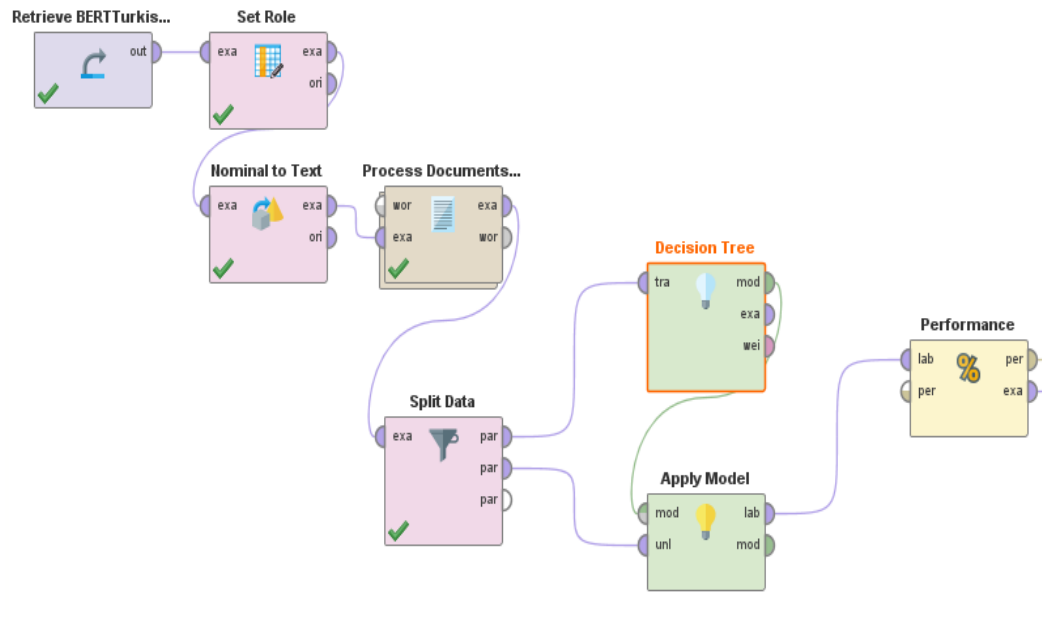
Şekil 3.8 Naive Bayes (Kernel) Algoritması



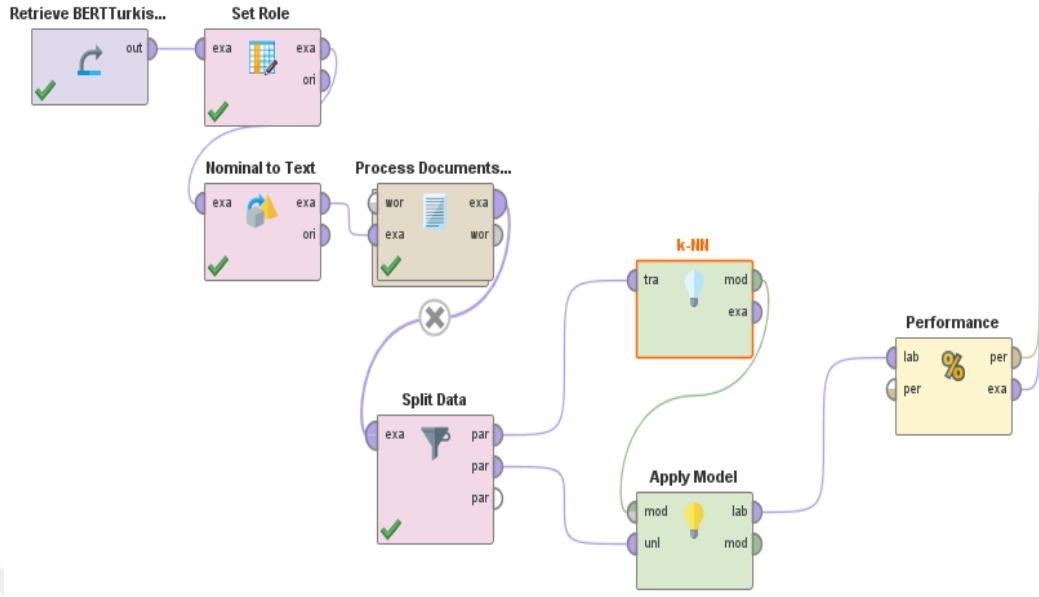
Şekil 3.9 Rastgele Orman Algoritması



Şekil 3.10 Destek Vektör Makinesi Algoritması



Şekil 3.11 Karar Ağacı Algoritması



Şekil 3.12 K-En Yakın Komşu Algoritması

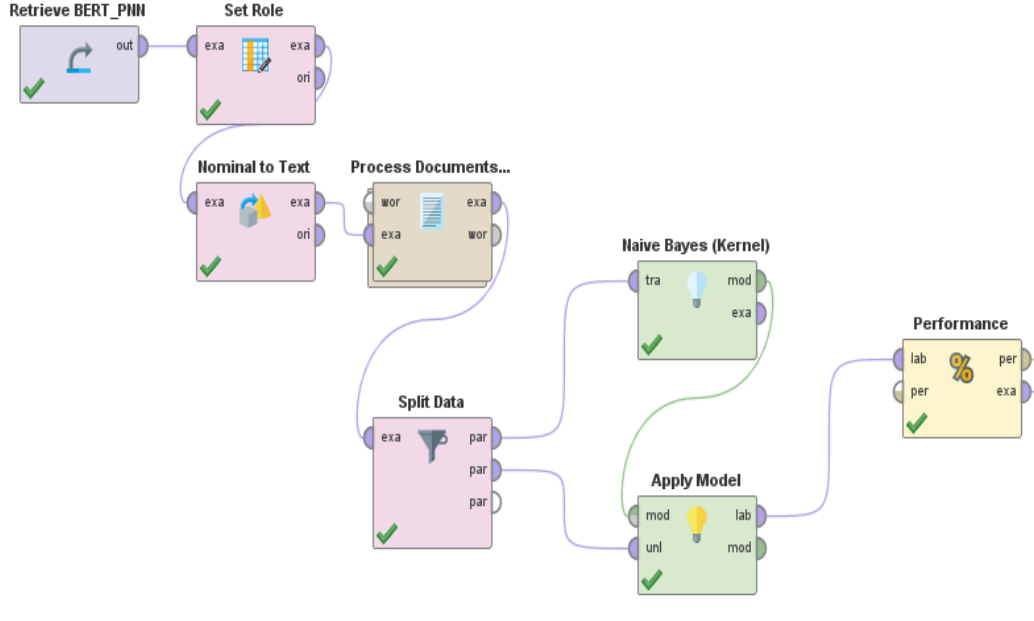
Çizelge 3.10’da pozitif ve negatif duygu etiketine sahip BERT modeliyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla gerçekleştirilen analiz sonuçları verilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.10 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Analiz Sonuçları

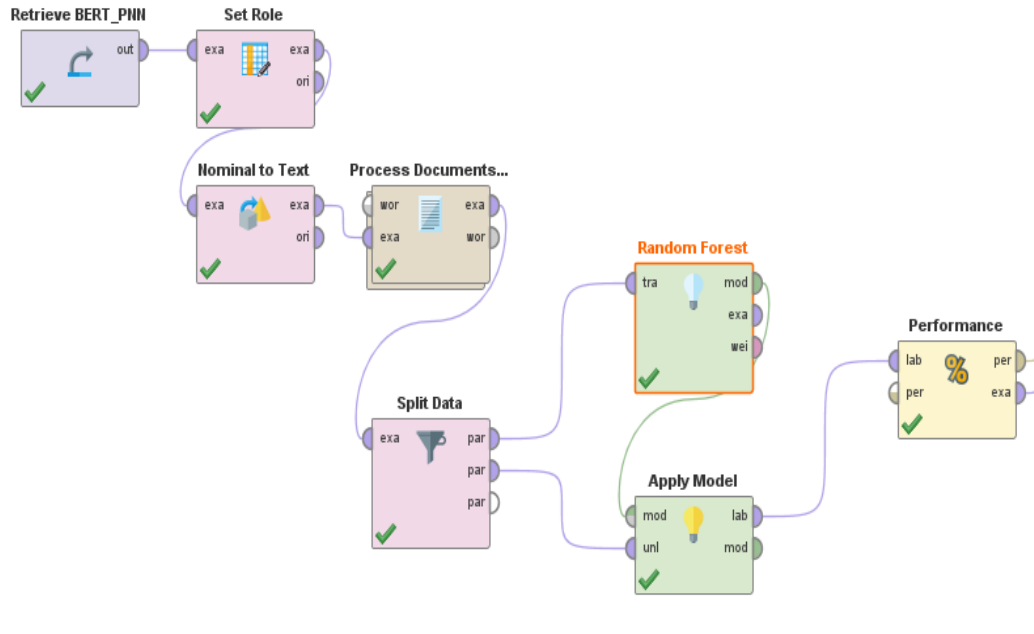
Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
NB (Kernel)	Pozitif	%62.35	%51.79	%67,47
	Negatif	%70.10	%78.33	
RO	Pozitif	%0	%0	%59,07
	Negatif	%59.07	%100	
DVM	Pozitif	%90.29	%39.74	%69,60
	Negatif	%68.38	%73.94	
KA	Pozitif	%85.71	%3.91	%60,40
	Negatif	%59.92	%99.55	
KEYK	Pozitif	%64.94	%53.09	%69,07
	Negatif	%71.14	%80.14	

3.6.2 Üç Duygu Etiketine Sahip BERT Modeliyle Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi

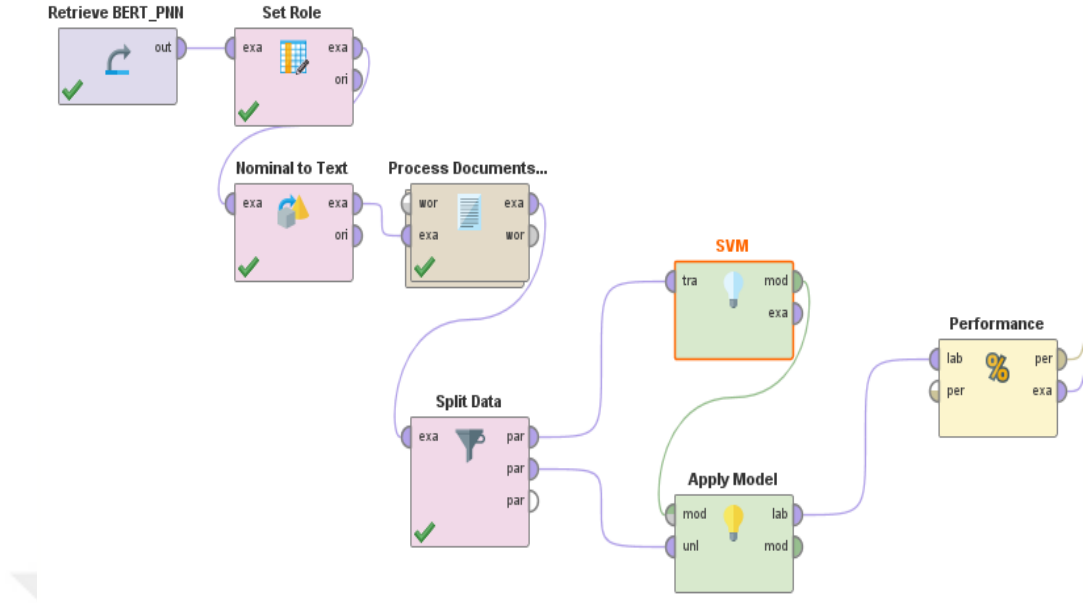
RapidMiner’da üç duygu etiketine sahip BERT modeliyle yapılan duygu tespiti sonrasında beş makine öğrenim algoritmasıyla analizler yapılmıştır. Yapılan analizlere ait model yapıları aşağıdaki görsellerde verilmiştir.



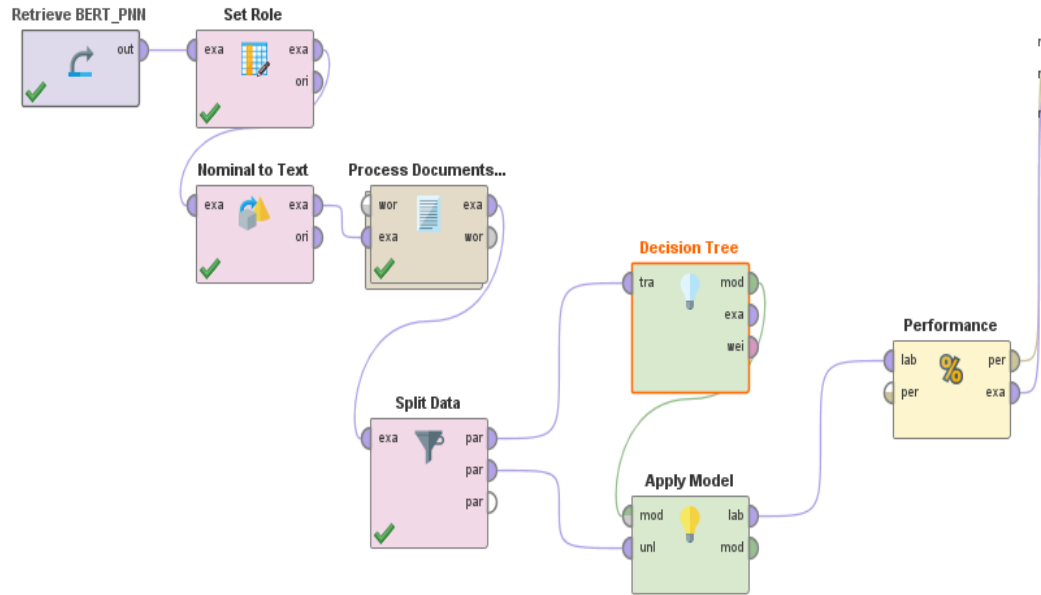
Şekil 3.13 Naive Bayes (Kernel) Algoritması



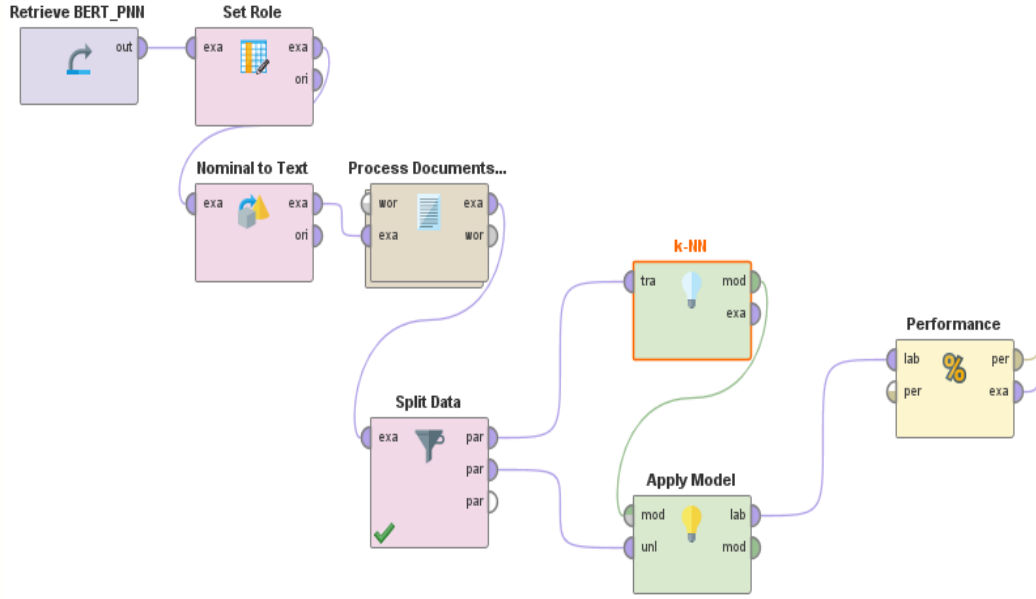
Şekil 3.14 Rastgele Orman Algoritması



Şekil 3.15 Destek Vektör Makinesi Algoritması



Şekil 3.16 Karar Ağacı Algoritması



Şekil 3.17 K-En Yakın Komşu Algoritması

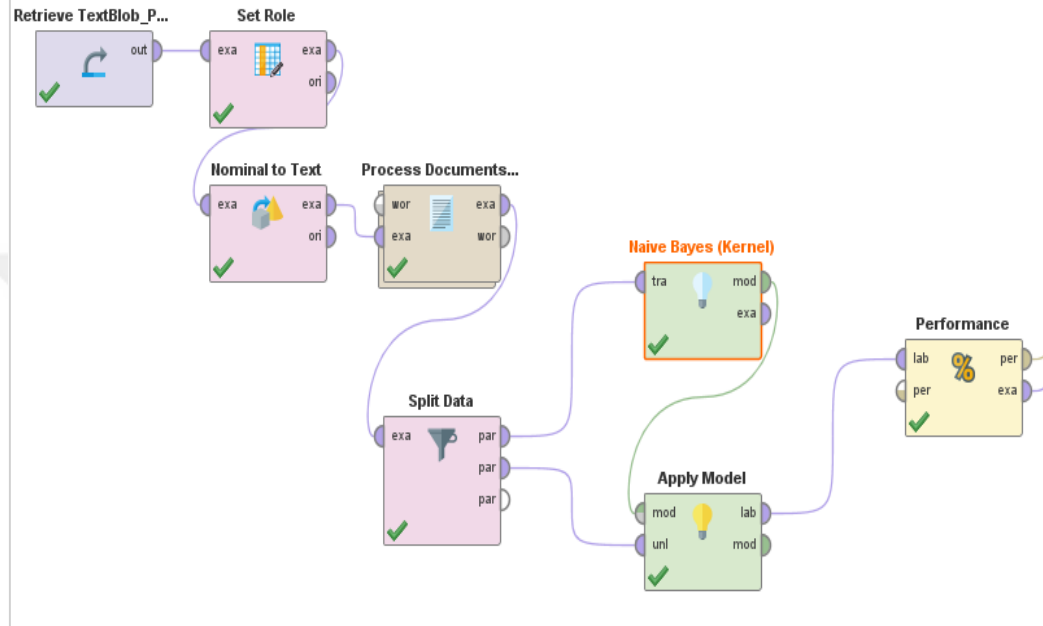
Çizelge 3.11’de pozitif, negatif ve nötr duygu etiketine sahip BERT modeliyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla gerçekleştirilen analiz sonuçları verilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar, mavi renk ile belirtilen hücreler, nötr veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.11 BERT (Pozitif -Negatif-Nötr) Modeliyle Elde Edilen Analiz Sonuçları

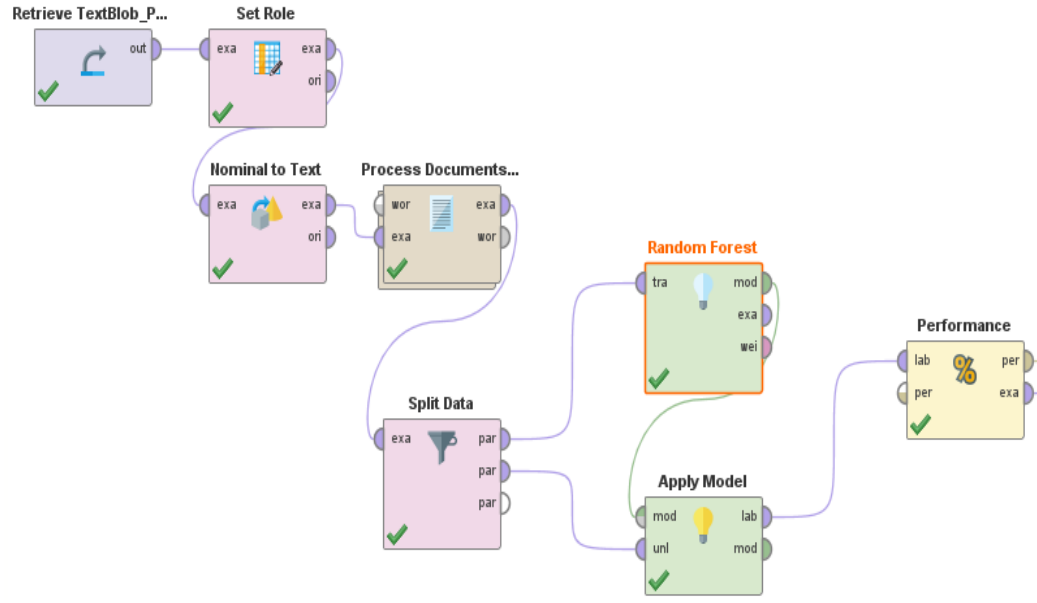
Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
NB (Kernel)	Pozitif	%31.34	%24.14	%60.61
	Negatif	%59.67	%61.07	
	Nötr	%66.58	%68.96	
RO	Pozitif	%0	%0	%48.60
	Negatif	%0	%0	
	Nötr	%48.60	%100	
DVM	Pozitif	%0	%0	%48.60
	Negatif	%0	%0	
	Nötr	%48.60	%100	
KA	Pozitif	%83.33	%5.75	%49.27
	Negatif	%0	%0	
	Nötr	%48.99	%100	
KEYK	Pozitif	%43.59	%19.54	%58.34
	Negatif	%64.56	%34.23	
	Nötr	%57.61	%87.36	

3.6.3 TextBlob ile Yapılan Duygu Tespitinin Makine Öğrenim Algoritmalarıyla Analizi

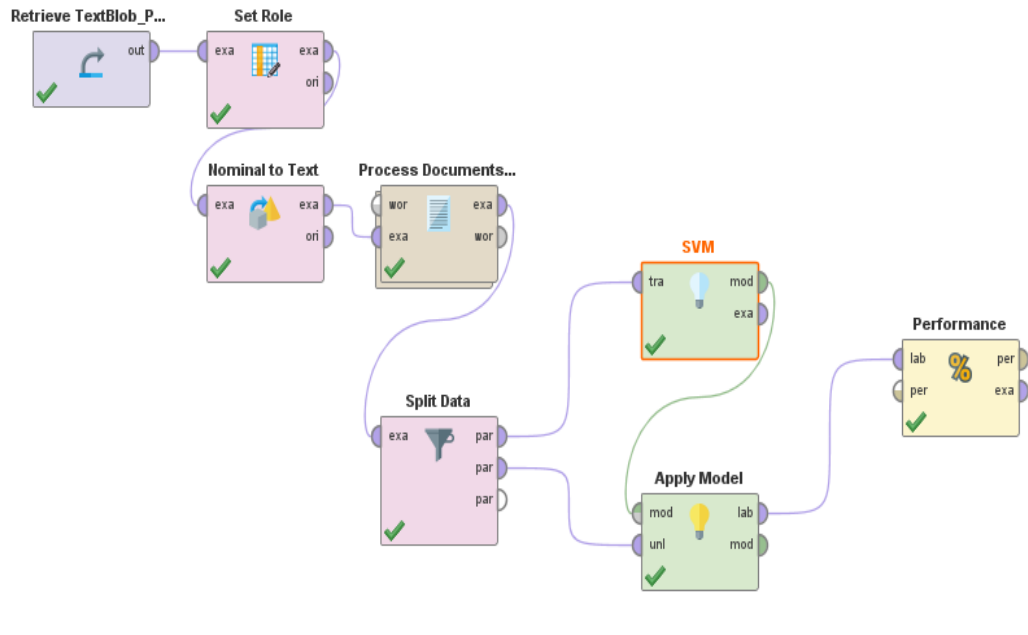
TextBlob kütüphanesi ile duygu tespiti yapılan verilerin, RapidMiner’da beş makine öğrenim algoritmasıyla yapılan analiz sonuçları verilmiştir. Analiz model yapıları her algoritma için ayrı ayrı olarak aşağıdaki görsellerde verilmiştir.



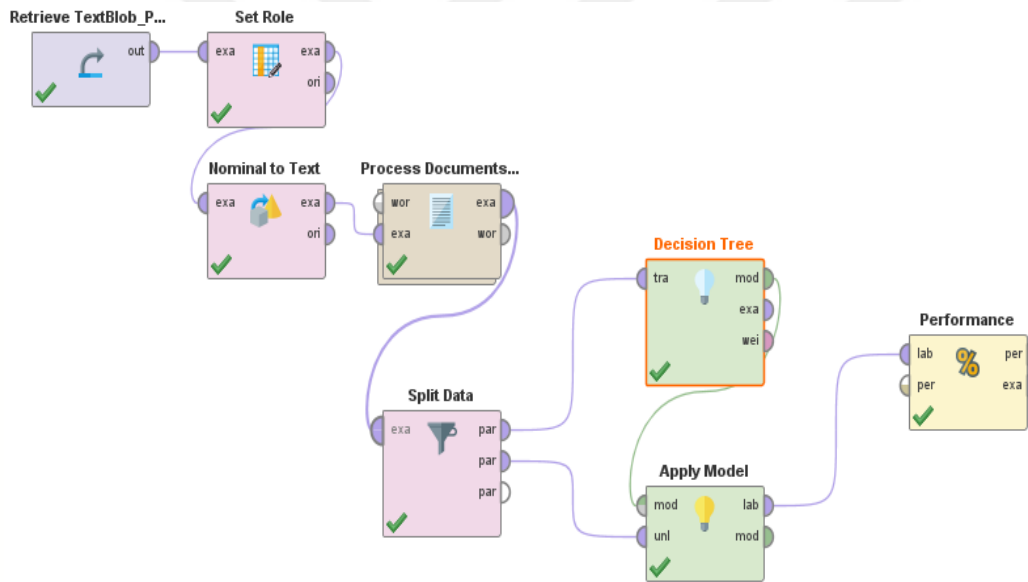
Şekil 3.18 Naive Bayes (Kernel) Algoritması



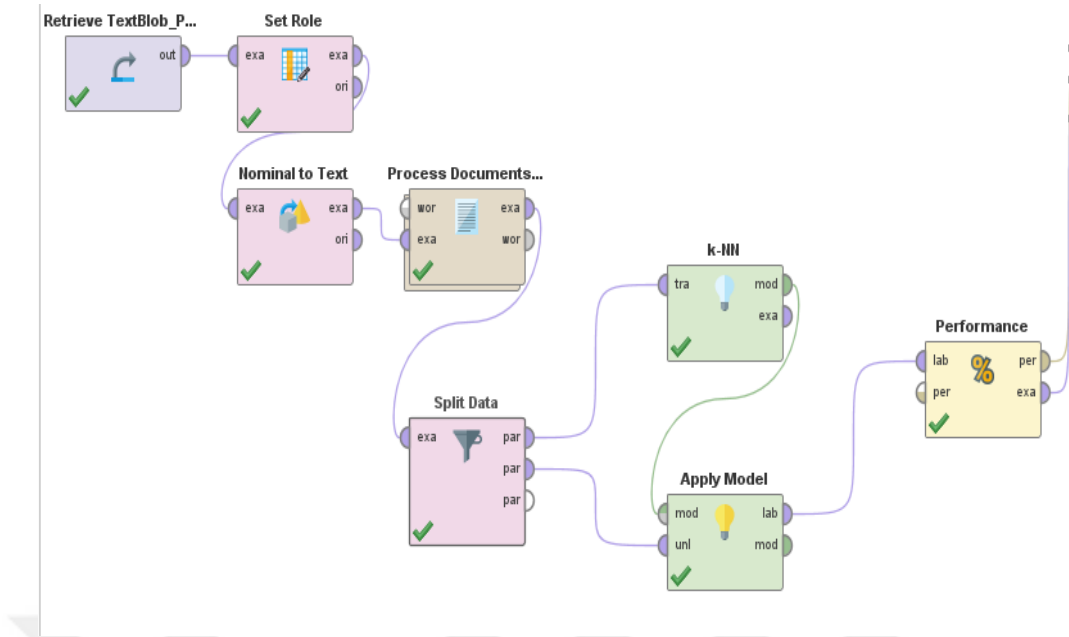
Şekil 3.19 Rastgele Orman Algoritması



Şekil 3.20 Destek Vektör Makinesi Algoritması



Şekil 3.21 Karar Ağacı Algoritması



Şekil 3.22 K-En Yakın Komşu Algoritması

Çizelge 3.12’de TextBlob kütüphanesiyle duygu tespiti yapılan verilerin, beş makine öğrenim algoritmasıyla yapılan analiz sonuçları verilmiştir. Çizelgede turuncu renk ile belirtilen hücreler, pozitif veriler arasındaki en iyi sonuçlar, sarı renk ile belirtilen hücreler, negatif veriler arasındaki en iyi sonuçlar, mavi renk ile belirtilen hücreler, nötr veriler arasındaki en iyi sonuçlar ve yeşil renk ile belirtilen hücreler ise doğruluk metriğindeki en iyi sonuçlardır.

Çizelge 3.12 TextBlob Kütüphanesi ile Elde Edilen Analiz Sonuçları

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
NB (Kernel)	Pozitif	%49.5	%34.25	%73.20
	Negatif	%41.3	%30.65	
	Nötr	%79.6	%88.56	
RO	Pozitif	%0	%0	%72.27
	Negatif	%0	%0	
	Nötr	%72.27	%100	
DVM	Pozitif	%0	%0	%72.27
	Negatif	%0	%0	
	Nötr	%72.27	%100	
KA	Pozitif	%0	%0	%73.20
	Negatif	%75	%14.52	
	Nötr	%73.17	%99.63	
KEYK	Pozitif	%80	%2.74	%73.33
	Negatif	%100	%6.45	
	Nötr	%73.14	%100	

3.7 Rapid Miner ile Analiz Sonucunda Elde Edilen Karmaşıklık Matrisleri

3.7.1 İki Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri

Çizelge 3.13 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler	
			Negatif	Pozitif
NB (Kernel)		Negatif	347	148
		Pozitif	96	159
RO		Negatif	443	307
		Pozitif	0	0
DVM		Negatif	400	185
		Pozitif	43	122
KA		Negatif	441	295
		Pozitif	2	12
KEYK		Negatif	355	144
		Pozitif	88	163

3.7.2 Üç Duygu Etiketine Sahip BERT Modeli ile Elde Edilen Karmaşıklık Matrisleri

Çizelge 3.14 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler		
			Negatif	Pozitif	Nötr
NB (Kernel)		Negatif	182	20	96
		Pozitif	36	21	30
		Nötr	87	26	251
RO		Negatif	0	0	298
		Pozitif	0	0	87
		Nötr	0	0	364
DVM		Negatif	0	0	298
		Pozitif	0	0	87
		Nötr	0	0	364
KA		Negatif	0	1	297
		Pozitif	0	5	82
		Nötr	0	0	364
KEYK		Negatif	102	14	182
		Pozitif	18	17	52
		Nötr	38	8	318

3.7.3 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisleri

Çizelge 3.15 TextBlob Kütüphanesi ile Elde Edilen Karmaşıklık Matrisi

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler		
			Negatif	Pozitif	Nötr
NB (Kernel)		Negatif	19	7	36
		Pozitif	9	50	87
		Nötr	87	44	480
RO		Negatif	0	0	62
		Pozitif	0	0	146
		Nötr	0	0	542
DVM		Negatif	0	0	62
		Pozitif	0	0	146
		Nötr	0	0	542
KA		Negatif	9	0	53
		Pozitif	1	0	145
		Nötr	2	0	540
KEYK		Negatif	4	1	57
	Pozitif	0	4	142	
	Nötr	0	0	542	

3.8 Rapid Miner ile Eğitim ve Test Veri Setleri Oranlarının Analizi

3.8.1 BERT Modeli (Pozitif - Negatif)

Çizelge 3.16 BERT (Pozitif-Negatif) Modeli ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15	Eğitim:%90 Test:%10	Eğitim:%91 Test:%9	Eğitim:%92 Test:%8	Eğitim:%93 Test:%7
	TF-IDF	TF-IDF	TF-IDF	TF-IDF	TF-IDF
NB	%67.47	%67.60	%68.67	%68.25	%69.14
RO	%59.07	%59.20	%59.11	%59	%59.14
DVM	%69.60	%69.60	%68.67	%68.25	%68
KA	%60.40	%60.20	%60.44	%60.50	%60.86
KEYK	%69.07	%68.20	%67.33	%66.75	%67.43

3.8.2 BERT Modeli (Pozitif – Negatif – Nötr)

Çizelge 3.17 BERT (Pozitif-Negatif-Nötr) Modeli ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15	Eğitim:%90 Test:%10	Eğitim:%91 Test:%9	Eğitim:%92 Test:%8	Eğitim:%93 Test:%7
	TF-IDF	TF-IDF	TF-IDF	TF-IDF	TF-IDF
NB	%60.61	%58.80	%58.44	%59.25	%59.71
RO	%48.60	%48.60	%48.67	%48.50	%48.57
DVM	%66.22	%65.80	%64.44	%63.25	%62.29
KA	%49.27	%49.40	%49.33	%49	%49.14
KEYK	%58.34	%60.20	%59.78	%59.25	%58.86

3.8.3 TextBlob

Çizelge 3.18 TextBlob Kütüphanesi ile Elde Edilen Doğruluk Oranları

Algoritmalar	Eğitim:%85 Test:%15	Eğitim:%90 Test:%10	Eğitim:%91 Test:%9	Eğitim:%92 Test:%8	Eğitim:%93 Test:%7
	TF-IDF	TF-IDF	TF-IDF	TF-IDF	TF-IDF
NB	%73.20	%73.95	%73.78	%73.25	%73.14
RO	%72.27	%72.34	%72.22	%72.25	%72.29
DVM	%78.80	%78.96	%78.89	%79.25	%79.43
KA	%73.20	%73.75	%73.56	%73.50	%73.71
KEYK	%73.33	%72.95	%73.11	%73.25	%73.43

4. SONUÇLAR VE ÖNERİLER

4.1 Sonuçlar

Türkçe ses kayıtlarından elde edilen veriler üzerinde yapılan çalışmalar sonrasında, bu veri seti üzerinde üç model yapısı ile duygu tespiti yapılmış ve beş farklı makine öğrenim algoritmasıyla analizler yapılmıştır. Bu bölümde ise duygu tespitinden başlayarak sonrasında yapılan analizleri kapsayan sonuçlar ve değerlendirmelerden bahsedilecektir. Çalışmanın birden fazla yöntemle gerçekleştirilmesi sonuçlar açısından geniş kapsamlı değerlendirmeler ve analizler yapmaya olanak sağlamıştır. İlk olarak Python programlama dili ile yapılan işlemler sonrasında ortaya çıkan sonuçlar ve değerlendirmeler verilmiştir.

Türkçe ses kayıtlarından elde edilen metinler üzerinde duygu tespiti için üç yöntem uygulanmıştır. Uygulanan BERT modellerinden, sadece pozitif ve negatif duygu etiketine sahip olanı TextBlob ile yapılacak olan karşılaştırmaya dâhil edilmemiş olup, elde edilen sonuçlar üzerindeki değerlendirmeye dahil edilmiştir. Pozitif, negatif ve nötr duygu etiketini içeren BERT modelinde, 2430 nötr, 1989 negatif ve 582 pozitif cümle elde edilmiştir. Veri setinin İngilizce diline çevrilip TextBlob’da yapılan duygu tespitinin sonucunda ise 3614 nötr, 974 pozitif ve 413 negatif cümle elde edilmiştir. İki yöntemde de çıkan sonuçlar karşılaştırıldığında veri setinde nötr duygu etiketine sahip cümlelerin yüksek oranda çıktığı görülmektedir. Veri seti içinde oranlamalara bakıldığında BERT modeli, %48,7 nötr, %39,7 negatif ve %11,6 pozitif cümle içermektedir. TextBlob yönteminde ise veri seti, %72 nötr, %19 pozitif ve %9 negatif cümle içermektedir. Yarı yarıya nötr duygu etiketine sahip olan bir Türkçe duygu veri seti edilmiştir. Her iki yöntemde de elde edilen sonuçlar, belirtilen farklılıklar üzerine negatif ve pozitif duygu etiketine sahip verilerde farklı sonuçlar ortaya çıkarsa da genel olarak veri setinin %50 oranına yakınının nötr duygu etiketine sahip olduğunu ortaya koymuştur.

Pozitif ve negatif duygu etiketine sahip BERT modeli nötr duygu etiketine sahip olmadığı için çıkan sonuçlar üzerinden yapılan karşılaştırmaya dahil edilmemiştir. Bu BERT modelinde skor yapısı 0 ile 1 arasında değerlendirilmiştir. Veri seti üzerinde uygulanan bu BERT modeli sonucunda 2956 negatif ve 2045 pozitif cümle elde edilmiştir. Veri setinde bulunan verilerde negatif duygu etiketine sahip verilerin etkin olduğu gözlemlenmiştir. Pozitif, negatif ve nötr duygu etiketine sahip olan

BERT modelinde çıkan sonuçlara bakıldığında nötr duygu etiketine sahip verilerden sonra en çok negatif duygu etiketine sahip veriler olduğu görülmektedir.

BERT modeliyle yapılan duygu tespiti sonucunda pozitif ve negatif veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmaları ile Python programlama dili ile analizi gerçekleştirilmiştir. Analiz hem pozitif hem de negatif için ayrı ayrı gerçekleştirilmiştir. Ayrıca vektörizasyon işlemlerindeki farklılıkta göz önünde bulundurulmuştur. Veri seti %85'e - %15 olarak ayrılmıştır. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 431 negatif ve 320 pozitif olmak üzere 751 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonucu NB algoritması CountVectorizer ile yapılan analiz sonucunda %70 olarak elde etmiştir. En düşük doğruluk oranını ise %60 ile KEYK algoritması, TF-IDFVectorizer işlemiyle elde etmiştir. Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %71 oranla en iyi duyarlılık değeri oranını NB algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi duyarlılık oranını %71 ile DVM algoritması elde etmiştir. Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %69 oranla en iyi duyarlılık oranı RO algoritması elde edilmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %100 oranla (1.00) en iyi duyarlılık oranını KEYK algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %94 oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi anma değeri oranı %100 oranla (1.00) KEYK algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %56 oranla en iyi anma değeri oranı DVM algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %56 oranla en iyi anma değeri oranını DVM algoritması elde etmiştir.

F1-Skoru için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %75 oranla hem NB hem de RO algoritmalarında elde edilmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi F1-Skoru oranını %71 ile DVM algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %61 oranla en iyi F1-skoru oranı NB algoritması ile elde edilmiştir. Pozitif veriler üzerinde TF-IDF vektörizasyon yöntemiyle elde edilen sonuçlara bakıldığında %61 oranla en iyi F1-skoru oranını DVM algoritması elde etmiştir.

Çizelge 4.1 BERT (Pozitif-Negatif) Modelinde En İyi Sonuçlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
Negatif	NB	DVM	KEYK	KEYK	NB, RO	DVM	NB	DVM
Pozitif	RO	KEYK	DVM	DVM	NB	DVM		

CountVectorizer yöntemiyle elde edilen sonuçlarda Naive Bayes algoritmasının daha çok ön plana çıktığı görülmektedir. TF-IDFVectorizer yönteminde ise Destek Vektör Makinesi ikili sınıflandırmadaki etkinliğini ortaya koymuştur. Türkçe verilerle yapılan analizlerde bu iki model yapısının kullanımıyla daha etkin sonuçlar alınabileceği söylenebilir.

Üç duygu etiketine sahip BERT modeliyle yapılan duygu tespiti sonucunda pozitif, negatif ve nötr veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmaları ile Python programlama dili ile analizi gerçekleştirilmiştir. Analiz hem pozitif hem negatif hem de nötr veriler için ayrı ayrı gerçekleştirilmiştir. Ayrıca vektörizasyon işlemlerindeki farklılıkta göz önünde bulundurulmuştur. Veri seti %85'e - %15 olarak ayrılmıştır. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 307 negatif, 75 pozitif ve 369 nötr olmak üzere 751 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonuca NB ve DVM makine öğrenim algoritmalarında ulaşılmıştır. Doğruluk oranı 0.67 (yüzde 67)'dir. Naive Bayes'de, CountVectorizer yöntemiyle, DVM'de ise TF-IDFVectorizer yöntemiyle vektörizasyon işlemine tabii tutulan verilerle elde edilmiştir.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %67 oranla en iyi duyarlılık değeri oranını RO algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi duyarlılık oranı %67 ile DVM algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %56 oranla en iyi duyarlılık oranını KEYK algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %45 oranla en iyi duyarlılık oranını RO algoritması elde etmiştir.

Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %71 oranla en iyi duyarlılık oranını NB algoritması elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %71 oranla en iyi duyarlılık oranını DVM algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %69 oranla en iyi anma değeri oranını NB algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi anma değeri oranı %66 oranla DVM algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %56 oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %29 oranla en iyi anma değeri oranını DVM ve KA algoritmaları elde etmiştir.

Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %85 oranla en iyi anma değeri oranını RO algoritması elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %86 oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir.

F1-skoru için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %67 oranla NB algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi F1-skoru oranı %67 ile DVM algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %34 oranla en iyi F1-skorumu DVM algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %34 oranla en iyi F1-skoru oranını DVM algoritması elde etmiştir.

Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %74 oranla en iyi F1-skoru oranını NB algoritması elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %73 oranla en iyi F1-skoru oranını DVM, RO ve NB algoritmaları elde etmiştir.

Çizelge 4.2 BERT (Pozitif-Negatif-Nötr) Modelinde En İyi Sonuçlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
Negatif	RO	DVM	NB	DVM	NB	DVM	NB	DVM
Pozitif	KEYK	RO	KEYK	DVM, KA	DVM	DVM		
Nötr	NB	DVM	RO	KEYK	NB	DVM, RO, NB		

CountVectorizer yöntemiyle elde edilen sonuçlarda Naive Bayes algoritmasının daha çok ön plana çıktığı görülmektedir. TF-IDFVectorizer yönteminde ise Destek Vektör Makinesi algoritması ön plana çıkmıştır. Vektörizasyon yöntemleri içinde CountVectorizer’da Naive Bayes’in, TF-IDFVectorizer’da Destek Vektör Makinesi algoritmalarının verimli sonuçlar ürettiği ifade edilebilir. Türkçe verilerle yapılan analizlerde bu iki model yapısının kullanımıyla daha etkin sonuçlar alınabileceği söylenebilir.

TextBlob ile yapılan duygu tespiti sonucunda pozitif, negatif ve nötr veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmaları ile Python programlama dili ile analizi gerçekleştirilmiştir. Analiz hem pozitif hem negatif hem de nötr veriler için ayrı ayrı gerçekleştirilmiştir. Ayrıca vektörizasyon işlemlerindeki farklılıkta göz önünde bulundurulmuştur. Veri seti %85'e - %15 olarak ayrılmıştır. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 60 negatif, 137 pozitif ve 554 nötr olmak üzere 751 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonuca RO algoritmasında ulaşılmıştır. Doğruluk oranı 0.86 (yüzde 86)'dır. RO'da bu doğruluk oranı hem CountVectorizer ile hem de TF-IDFVectorizer ile vektörizasyon işlemine tabii tutulan verilerle elde edilmiştir.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %67 oranla en iyi duyarlılık değeri oranını RO algoritması elde etmiştir.

Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi duyarlılık oranını %100 (1.00) ile KEYK algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %81 oranla en iyi duyarlılık oranını RO algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %89 oranla en iyi duyarlılık oranını NB ve KEYK algoritmalarında elde etmiştir.

Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %89 oranla en iyi duyarlılık oranını KA algoritması elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %90 oranla en iyi duyarlılık oranını KA algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %47 oranla en iyi anma değeri oranını KA algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi anma değeri oranı %50 oranla KA algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %72 oranla en iyi anma değeri oranını KA algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %75 oranla en iyi anma değeri oranını KA algoritması elde etmiştir.

Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %99 oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %100 (1.00) oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir.

F1-skoru için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %50 oranla KA algoritması elde etmiştir. Negatif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında en iyi F1-skoru oranı %52 ile KA algoritması elde etmiştir.

Pozitif veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %75 oranla en iyi F1-skoru oranını KA algoritması elde etmiştir. Pozitif veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %75 oranla en iyi F1-skoru oranını RO algoritması elde etmiştir. Nötr veriler için ilk değerlendirme CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %91 oranla en iyi F1-skoru oranını RO ve KA algoritmaları elde etmiştir. Nötr veriler üzerinde TF-IDFVectorizer ile elde edilen sonuçlara bakıldığında %91 oranla en iyi F1-skoru oranı oranını RO algoritması elde etmiştir.

Çizelge 4.3 TextBlob Kütüphanesinde En İyi Sonuçlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skor		Doğruluk	
Negatif	RO	KEYK	KA	KA	KA	KA	RO	RO
Pozitif	RO	KEYK, NB	KA	KA	KA	RO		
Nötr	KA	KA	KEYK	KEYK	KA, RO	RO		

CountVectorizer yöntemiyle elde edilen sonuçlarda Karar Ağacı algoritmasının daha çok ön plana çıktığı görülmektedir. TF-IDFVectorizer yönteminde ise K-En Yakın Komşu ile Karar Ağacı algoritmaları ön plana çıkmıştır. Vektörizasyon yöntemleri içinde CountVectorizer'da Karar Ağacı'nın, TF-IDFVectorizer'da Karar Ağacı ve K-En Yakın Komşu algoritmalarının verimli sonuçlar ürettiği söylenebilir.

Eğitim ve test verilerinin oranları üzerinde yapılan değişikliklerin sonuçları her yöntem için ayrı olarak aşağıda verilmiştir.

BERT modelinin iki duygu etiketine sahip veri setiyle, eğitim ve test verilerinin oranlarında değişiklikler yapılarak doğruluk oranında farklılıklar gözlemlenmiştir. Eğitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiştir.

En yüksek doğruluk oranı, eğitim verisinin %93, test verisinin %7 olduğu analizde elde edilmiştir. Destek Vektör Makinesi algoritması hem CountVectorizer hem de TF-IDFVectorizer işlemine tabii tutulan veriler üzerinde %72 doğruluk oranı sağlamıştır.

Eğitim verisi arttıkça yani eğitilen veri setinin artması, modellerdeki doğru tahmin olasılığını arttırmaktır. %15 oranındaki test verisi sayısının %7 oranına düşürülmesi, %2'lik bir doğruluk oranı artışında etkili olmuştur. Vektörizasyon işlemlerine bakıldığında ise CountVectorizer vektörizasyonu ile yapılan analizlerde daha iyi sonuçlar elde edildiği söylenebilir.

BERT modelinin üç duygu etiketine sahip veri setiyle, eğitim ve test verilerinin oranlarında değişiklikler yapılarak doğruluk oranında farklılıklar gözlemlenmiştir. Eğitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiştir.

En yüksek doğruluk oranı, eğitim verisinin %93, test verisinin %7 olduğu analizde elde edilmiştir.

Naive Bayes algoritması, CountVectorizer vektörizasyon işlemine tabii tutulan veriler üzerinde %72 doğruluk oranı sağlamıştır.

Eğitim verisi arttıkça yani eğitilen veri setinin artması, modellerdeki doğru tahmin olasılığını arttırmaktır. %15 oranındaki test verisi sayısının %7 oranına düşürülmesi, %5'lik bir doğruluk oranı artışında etkili olmuştur. Vektörizasyon işlemleri açısından yapılacak olan değerlendirme de ise bir denge olduğu söylenebilir.

TextBlob ile elde edilen üç duygu etiketine sahip veri setiyle, eğitim ve test verilerinin oranlarında değişiklikler yapılarak doğruluk oranında farklılıklar gözlemlenmiştir. Eğitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiştir.

En yüksek doğruluk oranı, eğitim verisinin %85 ve %93, test verisinin %15 ve %7 olduğu analizlerde elde edilmiştir.

Rastgele Orman algoritması, hem CountVectorizer hem de TF-IDFVectorizer işlemine tabii tutulan veriler üzerinde %86 doğruluk oranı sağlamıştır.

İngilizce diline çevrilen veri setinde %15 test verisinin yeterli olduğu, daha az test veri seti oranlarında ciddi bir artış veya azalma görülmediği fakat daha az test veri seti oranlarında %1 ila %5 arasında doğruluk oranında azalma görüldüğü elde edilen sonuçlara göre söylenebilir.

Vektörizasyon işlemleri açısından yapılacak olan değerlendirme de ise CountVectorizer vektörizasyon işleminin olduğu analizlerde daha iyi doğruluk oranları elde edildiği söylenebilir.

RapidMiner ile gerçekleştirilen analiz sonuçları, duygu tespiti yapılan her model için ayrı olarak aşağıda verilmiştir.

BERT modeliyle yapılan duygu tespiti sonucunda pozitif ve negatif veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmaları ile RapidMiner'da analizi gerçekleştirilmiştir. Analiz hem pozitif hem de negatif için ayrı ayrı gerçekleştirilmiştir. Veri seti %85'e - %15 olarak ayrılmıştır.

Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 443 negatif ve 307 pozitif olmak üzere 750 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonucu DVM algoritması ile yapılan analiz sonucunda %69.60 olarak elde etmiştir. En düşük doğruluk oranını ise %59.07 ile RO algoritması elde etmiştir.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %71.14 oranla en iyi duyarlılık değeri oranını KEYK algoritması elde etmiştir. Pozitif veriler için %85.71 oranla en iyi duyarlılık oranını KA algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %100 oranla en iyi anma değeri oranını RO algoritması elde etmiştir. Pozitif veriler için %53.09 oranla en iyi anma değeri oranını KEYK algoritması elde etmiştir.

Çizelge 4.4 İki Duygu Etiketine Sahip BERT Modelinde En İyi Oranlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
Negatif	KEYK	RO	DVM
Pozitif	KA	KEYK	

Elde edilen sonuçlara bakıldığında doğruluk oranında Destek Vektör Makinesi algoritması iyi bir sonuç elde etmesine rağmen aynı durum değerlendirilen metrikler üzerinde gerçekleşmemiştir. Hem duyarlılık değeri hem de anma değeri metriklerinde K-En Yakın Komşu algoritması öne çıkmıştır.

Üç duygu etiketine sahip BERT modeliyle yapılan duygu tespiti sonucunda pozitif, negatif ve nötr veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmaları ile RapidMiner’da analizi gerçekleştirilmiştir. Analiz hem pozitif hem negatif hem de nötr veriler için ayrı ayrı gerçekleştirilmiştir. Veri seti %85’e - %15 olarak ayrılmıştır. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 298 negatif, 87 pozitif ve 364 nötr olmak üzere 749 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonuca NB algoritmasıyla ulaşılmıştır. Doğruluk oranı %60.61’dir.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %64.56 oranla en iyi duyarlılık değeri oranını KEYK algoritması elde etmiştir.

Pozitif veriler için %83.33 oranla en iyi duyarlılık oranını KA algoritması elde etmiştir.

Nötr veriler için %66.58 oranla en iyi duyarlılık oranını NB algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %61.07 oranla en iyi anma değeri oranını NB algoritması elde etmiştir.

Pozitif veriler için %24.14 oranla en iyi anma değeri oranını NB algoritması elde etmiştir.

Nötr veriler için %100 oranla en iyi anma değeri oranını RO, DVM ve KA algoritmaları elde etmiştir.

Çizelge 4.5 Üç Duygu Etiketine Sahip BERT Modelinde En İyi Oranlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
Negatif	KEYK	NB	NB
Pozitif	KA	NB	
Nötr	NB	RO, DVM, KA	

Elde edilen sonuçlara bakıldığında doğruluk oranında en iyi sonucu elde eden Naive Bayes (Kernel) algoritması özellikle anma değeri metriğinde de etkili olmuştur ve öne çıkmıştır. Bunun yanı sıra Karar Ağacı algoritması da her iki metrikte de iyi sonuçlar elde etmiştir.

TextBlob ile yapılan duygu tespiti sonucunda pozitif, negatif ve nötr veriler / cümleler elde edilmiştir. Bu verilerin, belirlenen makine öğrenim algoritmalarıyla RapidMiner’da analizi gerçekleştirilmiştir. Analiz hem pozitif hem negatif hem de nötr veriler için ayrı ayrı gerçekleştirilmiştir. Veri seti %85’e - %15 olarak ayrılmıştır. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 62 negatif, 146 pozitif ve 542 nötr olmak üzere 750 veri vardır.

Doğruluk oranında elde edilen sonuçlar göz önüne alındığında en iyi sonuca KEYK algoritmasında ulaşılmıştır. Doğruluk oranı %73.33’tür.

Duyarlılık değeri için ilk değerlendirme negatif veriler yapılmıştır. %100 oranla en iyi duyarlılık değeri oranını KEYK algoritması elde etmiştir.

Pozitif veriler için %80 oranla en iyi duyarlılık oranını KEYK algoritması elde etmiştir.

Nötr veriler için %79.60 oranla en iyi duyarlılık oranını NB algoritması elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %30.65 oranla en iyi anma değeri oranını NB algoritması elde etmiştir.

Pozitif veriler için. %34.25 oranla en iyi anma değeri oranını NB algoritması elde etmiştir.

Nötr veriler için %100 oranla en iyi anma değeri oranını KEYK, RO ve DVM algoritmaları elde etmiştir.

Çizelge 4.6 TextBlob ile En İyi Oranlara Ulaşan Algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
Negatif	KEYK	NB	KEYK
Pozitif	KEYK	NB	
Nötr	NB	KEYK, DVM, RO	

Türkçe veri setinin İngilizce diline çevrilip makine öğrenim algoritmalarının analizi sonucunda hem doğruluk oranında hem de değerlendirilen ölçütlerdeki oranlarda artışa sebep olmuştur. Doğruluk oranında en iyi sonuca ulaşan K-En Yakın Komşu algoritması değerlendirilen metrikler üzerinde de etkili olmuştur. Benzer şekilde Naive Bayes (Kernel) algoritması da özellikle anma değeri metriği olmak üzere elde ettiği sonuçlarla öne çıkmıştır.

Eğitim ve test verilerinin oranları üzerinde yapılan değişikliklerin sonuçları her yöntem için ayrı olarak aşağıda verilmiştir.

BERT modelinin iki duygu etiketine sahip veri setiyle, eğitim ve test verilerinin oranlarında değişiklikler yapılarak doğruluk oranında farklılıklar gözlemlenmiştir.

Eđitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiřtir.

En yksek dođruluk oranı, eđitim verisinin hem %85 hem de %90, test verisinin hem %15 hem de %10 oranında olduđu analizlerde elde edilmiřtir. %69.60 oranıyla Destek Vektr Makinesi algoritması elde etmiřtir.

Eđitim ve test verilerinin oranlarındaki deđiřim ok yksek oranlarda deđiřimlere sebep olmamıřtır. En yksek dođruluk oranının bulunduđu eđitim ve test oranına bakıldıđında, test oranında azalmanın yeterli bir artıřa sebep olmadıđı ifade edilebilir.

BERT modelinin  duygu etiketine sahip veri setiyle, eđitim ve test verilerinin oranlarında deđiřiklikler yapılarak dođruluk oranında farklılıklar gzlemlenmiřtir. Eđitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiřtir.

En yksek dođruluk oranı, eđitim verisinin %85, test verisinin %15 olduđu analizde elde edilmiřtir. Destek Vektr Makinesi algoritması ile %66.22 oranında bir dođruluk sađlamıřtır.

Eđitim ve test verilerinde azalma dođruluk oranında artıřa neden olmamıřtır.

TextBlob ile elde edilen  duygu etiketine sahip veri setiyle, eđitim ve test verilerinin oranlarında deđiřiklikler yapılarak dođruluk oranında farklılıklar gzlemlenmiřtir. Eđitim veri setindeki oranlar %85, %90, %91, %92, %93 olarak belirlenirken test verilerindeki oranlar %15, %10, %9, %8 ve %7 olarak belirlenmiřtir.

En yksek dođruluk oranı, eđitim verisinin %93, test verisinin %7 olduđu analizde elde edilmiřtir.

Destek Vektr Makinesi algoritması, %79.43 dođruluk oranı sađlamıřtır.

İngilizce diline evrilen veri setinde, test veri seti oranında azalma dođruluk oranında ciddi bir artıř veya azalıř sađlamamıřtır. Belirlenen makine đrenim algoritmaları arasında en iyi dođruluk oranı sađlayan Destek Vektr Makinesi algoritması diđer algoritmalar arasında, belirlenen her eđitim ve test veri setinde en iyi sonucu veren makine đrenim algoritması olmuřtur.

İki duygu etiketine sahip BERT modelinde, Python programlama dili ile yapılan analizde CountVectorizer yöntemiyle en iyi doğruluk oranı elde edilmiştir. Doğruluk metriğinde en iyi sonucu Naive Bayes algoritması elde etmiştir. Destek Vektör Makinesi algoritması da anma değeri, F1-skor gibi metriklerde en iyi sonuçları elde ederek ön plana çıkmıştır. RapidMiner ile yapılan analizde, doğruluk metriğinde en iyi sonucu Destek Vektör Makinesi elde etmiştir. Duyarlılık değeri ve anma değeri metriklerinde aldığı iyi sonuçlarla K-En Yakın Komşu algoritması öne çıkmıştır. RapidMiner ile analizde hem vektörizasyon yönteminin tek olması hem de değerlendirilen metriklerin eşit olmaması RapidMiner ile yapılan analizin dezavantajı olmuştur. İki duygu etiketine sahip bu veri setinde Destek Vektör Makinesi algoritmasının her iki analiz sonucunda da öne çıktığı söylenebilir.

Veri seti üzerinde, üç duygu etiketine sahip BERT modeli ve TextBlob kütüphanesiyle duygu tespiti yapılmıştır. Duygu tespiti sonucunda elde edilen veri setlerinde nötr verilerin ağırlıkta olduğu görülmüştür. Duygu tespitinde her iki yöntemde benzer sonuçlar elde etmiştir. Python programlama dilinde yapılan analizde, BERT modeli her iki vektörizasyon yönteminde aynı doğruluk oranını farklı algoritmalarla elde etmiştir. TextBlob ise en yüksek doğruluk oranını aynı algoritmayla her iki vektörizasyon yönteminde elde etmiştir.

BERT modelinde, üç değerlendirme metriğinde en iyi sonuçların büyük kısmını elde eden Destek Vektör Makinesi algoritması olmuştur. TextBlob'da ise Karar Ağacı algoritması alınan en iyi sonuçlarda ön plana çıkmıştır. RapidMiner platformunda yapılan analizde, BERT modelinde hem en iyi doğruluk oranını hem de diğer metriklerde de iyi sonuçlar elde eden Naive Bayes algoritması öne çıkmıştır. TextBlob'da ise K-En Yakın Komşu algoritması doğruluk oranında en iyi sonucu elde ederken metriklerde Naive Bayes algoritmasıyla iyi sonuçlar elde etmiştir. Üç duygu etiketine sahip veri seti üzerinde yapılan işlemlerde hem Destek Vektör Makinesi hem de Naive Bayes algoritmasının başarılı olabileceği ifade edilebilir.

4.2 Öneriler

Türkçe ses kayıtlarından elde edilen veriler üzerinde duygu tespiti ve analiz çalışmaları gerçekleştirilmiştir. Çalışmanın duygu tespiti aşamasında BERT modeli ve TextBlob kütüphanesi kullanılmıştır. BERT modelinde Türkçe çalışmalar için daha iyi sonuçlar elde edilebilmesi adına BERT modelinin yapısına Türkçe durak kelimeler, Türkçe kelimelerin ve skor yapılarının eklenmesi ve geliştirilmesi için çalışmalar yapılabilir. TextBlob içinde benzer geliştirmeler yapılabilir. Özellikle Türkçe duygu analizi için dil seçenekleri arasına Türkçe dili eklenebilir ve bu yönde geliştirmeler yapılabilir. Analiz kısmında kullanılan iki platform sonuçlar açısından çeşitlilik sağlamıştır. Python programlama dili ile yapılan analizlerde daha etkin sonuçlar alınmıştır. Kullanılan makine öğrenim algoritmalarının Türkçe dili ile ilgili çalışmalarda daha iyi sonuçlar ortaya koyabilmesi için dil yapısına uygun hale getirecek çalışmalar yapılabilir. RapidMiner platformu ile yapılan analiz sonuçları, Python programlama dili ile yapılan analiz sonuçları kadar başarılı olmamıştır. RapidMiner’da vektörizasyon yöntemindeki eksiklikler ve ön işlem adı verilen işlemlerin yeterli olmaması sonuçları etkileyen unsurlar olmuştur. RapidMiner’da bu yapıların geliştirilmesi analiz sonuçlarının daha başarılı olmasında etkili olacaktır.



KAYNAKLAR

- [1] Duygu. <https://tr.wikipedia.org/wiki/Duygu>, 2021.
- [2] D. Tam Hoang, Charles University in Prague Sentiment Analysis: Polarity Dataset, 2014.
- [3] Yapay Zeka, Makine Öğrenmesi ve Derin Öğrenme Kavramları Arasındaki Fark Nedir? <https://evrimagaci.org/yapay-zeka-makine-ogrenmesi-ve-derin-ogrenme-kavramlari-arasindaki-fark-nedir>, 2021.
- [4] A. Murarka, K. Shivarkar, V. Gupta, P. Sankpal, IJARCCE Sentiment Analysis of Speech,” *Int. J. Adv. Res. Comput. Commun. Eng. ISO*, vol. 3297, no. 11, 2007
- [5] E. Cambria, L. Morency, S. Poria, A. Zadeh , M. Chen, Tensor Fusion Network for Multimodal Sentiment Analysis, 2017.
- [6] M. H. R. Pereira, F. L. C. Pádua, A. C. M. Pereira, F. Benevenuto, D. H. Dalip, Fusing Audio, Textual and Visual Features for Sentiment Analysis of News Videos, *Proc. 10th Int. Conf. Web Soc. Media, ICWSM 2016*, pp. 659–662, Apr. 2016.
- [7] J. Yarif, R. Cornejo, H. Pedrini, Audio-Visual Emotion Recognition Using a Hybrid Deep Convolutional Neural Network based on Census Transform, 2015.
- [8] A. Demir, O. Atila, A. Sengur, Deep Learning and Audio Based Emotion Recognition, Sep. 2019.
- [9] M. G. De Pinto, M. Polignano, P. Lops, G. Semeraro, Emotions Understanding Model from Spoken Language using Deep Neural Networks and Mel-Frequency Cepstral Coefficients, *IEEE Conference on Evolving and Adaptive Intelligent Systems*, May 2020.
- [10] K. M. Kudiri, A. M. Said, M. Y. Nayan, Emotion detection using average relative amplitude features through speech, *Proceedings - 2012 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2012*, 2012, pp. 115–118.

- [11] A. A. A. Zamil, S. Hasan, S. M. Jannatul Baki, J. M. Adam, I. Zaman, Emotion detection from speech signals using voting mechanism on classified frames, *1st International Conference on Robotics, Electrical and Signal Processing Techniques, ICREST 2019*, Feb. 2019, pp. 281–285.
- [12] S. Suganya, E. Y. A. Charles, Speech Emotion Recognition Using Deep Learning on audio recordings, Sep. 2019.
- [13] K. Venkataramanan, H. R. Rajamohan, “Emotion Recognition from Speech, *SpringerBriefs Speech Technol.*, pp. 31–32, Dec. 2019.
- [14] S. Dani, A. A. Mande, S. Telang, Z. Shao, Emotion Detection Using Audio Data Samples, *Int. J. Adv. Res. Comput. Sci.*, vol. 10, no. 6, pp. 13–20, Dec. 2019.
- [15] U. Garg, S. Agarwal, S. Gupta, R. Dutt, D. Singh, Prediction of Emotions from the Audio Speech Signals using MFCC, MEL and Chroma, 2020.
- [16] C. Sezgin, B. Günsel, C. Hacıoğlu, Algısal Öznelikler Kullanılarak Sesten Duygu Durum Tanıma, 2012.
- [17] R. Matin, D. Valles, A Speech Emotion Recognition Solution-based on Support Vector Machine for Children with Autism Spectrum Disorder to Help Identify Human Emotions, 2020.
- [18] M. R. Amer, B. Siddiquie, C. Richey, A. Divakaran, Emotion detection in speech using deep networks, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2014, pp. 3724–3728.
- [19] S. Yoon, S. Byun, K. Jung, Multimodal Speech Emotion Recognition Using Audio And Text, 2019.
- [20] M. C. Sezgin, B. Günsel, G. K. Kurt, A novel perceptual feature set for audio emotion recognition, 2011.
- [21] V. C. Govoreanu, M. Neghină, Speech Emotion Recognition method using time-stretching in the Preprocessing Phase and Artificial Neural Network Classifiers, 2020.

- [22] A. Bhavan, P. Chauhan, Hitkul, R. R. Shah, Bagged support vector machines for emotion recognition from speech, 2019.
- [23] S. Kadiri, P. Alku, Excitation Features of Speech for Speaker-Specific Emotion Detection, 2020.
- [24] S. Banga, U. Upadhyay, P. Agarwal, A. Sharma, P. Mukherjee, “Indian EmoSpeech Command Dataset: A dataset for emotion based speech recognition in the wild, 2019.
- [25] S. Tripathi, A. Kumar, A. Ramesh, C. Singh, P. Yenigalla, Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions, 2019.
- [26] J. Bhaskar, Sruthi K, P. Nedungadi, Hybrid Approach For Emotion Classification Of Audio Conversation Based On Text And Speech Mining, 2015.
- [27] M. Jain, S. Narayan, P. Balaji, Bharath K P, A. Bhowmick, Karthik R, R. K. Muthu, Speech Emotion Recognition using Support Vector Machine, *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 4, pp. 1212–1216, Feb. 2020.
- [28] P. Harár, R. Burget, M. K. Dutta, Speech Emotion Recognition with Deep Learning, 2017.
- [29] O. E. Korkmaz, Ses Sinyalinden Duygu Tanıma, 2016.
- [30] F. S. Çetin, G. Eryiğit, Türkçe Hedef Tabanlı Duygu Analizi için Alt Görevlerin İncelenmesi – Hedef Terim, Hedef Kategori ve Duygu Sınıfı Belirleme, 2018.
- [31] H. Göker, H. Tekedere, FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi, 2017.
- [32] Y. Safalı, Sosyal Medya Kullanıcılarının Cumhuriyet İttifakı Hakkındaki Görüşlerinin Makine Öğrenmesi Teknikleri ile Sınıflandırılması, 2020
- [33] İ. N. Yoldaş, Türkçe Metinlerde Duygu Analizi: Sözlük Tabanlı Yaklaşım ve İnsanların Tepkilerinin Karşılaştırılması , 2021.

- [34] F. D. Tayşı, *Konuşma Verisinden Duygu Durum Tespiti*, 2019.
- [35] A. Güven ,*Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciliği*, 2007.
- [36] S. Maghilnan, M. R. Kumar, Sentiment analysis on speaker specific speech data, *Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017*, 2018, vol. 2018-Janua, pp. 1–5.
- [37] R. R. Sehgal, S. Agarwal, G. Rai, Interactive Voice Response using Sentiment Analysis in Automatic Speech Recognition Systems, *Proceedings on 2018 International Conference on Advances in Computing and Communication Engineering, ICACCE 2018*, Aug. 2018, pp. 213–218.
- [38] K. Napier, L. Shamir, Quantitative sentiment analysis of lyrics in popular music, *J. Pop. Music Stud.*, vol. 30, no. 4, pp. 161–176, Dec. 2018.
- [39] S. Cunningham, H. Ridley, J. Weinel, and R. Picking, Supervised machine learning for audio emotion recognition: Enhancing film sound design using audio features, regression models and artificial neural networks, *Pers. Ubiquitous Comput.*, pp. 1–14, Apr. 2020.
- [40] O. Mohamad Nezami, P. Jamshid Lou, M. Karami, ShEMO: a large-scale validated database for Persian speech emotion detection, *Lang. Resour. Eval.*, vol. 53, no. 1, Mar. 2019.
- [41] H. Kaya, A. A. Salah, A. Karpov, O. Frolova, A. Grigorev, E. Lyakso, *Emotion, Age, and Gender Classification in Children’s Speech by Humans and Machines*, 2017.
- [42] C. Supriyanto, Muljono, M. R. Prasetya, A. Harjoko, *Speech Emotion Recognition of Indonesian Movie Audio Tracks based on MFCC and SVM*, 2019.
- [43] S. Yaslan, B. Günsel, Sayısal Ses Sinyallerinden Duygu Algılama, *Proceedings of the IEEE 13th Signal Processing and Communications Applications Conference, SIU* vol. 2005, pp. 656–659.

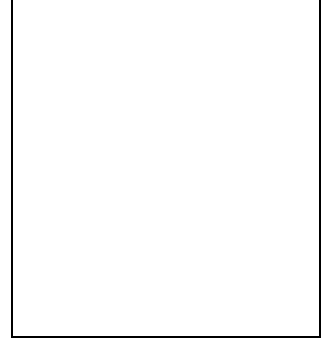
- [44] S. K. Bhakre, A. Bang, Emotion recognition on the basis of audio signal using Naive Bayes classifier, 2016.
- [45] A. Davletcharova, S. Sugathan, B. Abraham, A. P. Jame, Detection and analysis of emotion from speech signals, 2015.
- [46] J. Gondohanindijo, E. Noersasongko, Pujiono, Muljono, A. Z. Fanani, Affandy, R. S. Basuki, Comparison Method in Indonesian Emotion Speech Classification, 2019.
- [47] C. Min Lee, S. Member, S. S. Narayanan, S. Member, Toward Detecting Emotions in Spoken Dialogs, *IEEE Trans. SPEECH AUDIO Process.*, vol. 13, no. 2, p. 293, 2005.
- [48] L. Jie, Z. Xiaoyan, Z. Zhaohui, Speech Emotion Recognition of Teachers in Classroom Teaching, 2020.
- [49] A. Tripathi, U. Singh, G. Bansal, R. Gupta, A. K. Singh, A Review on Emotion Detection and Classification using Speech, 2020.
- [50] S. Tuzcu, Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması, 2020.
- [51] H. Nizam, S. S. Akın, Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, 2014.
- [52] R. Aşlıyan, K. Günel, Metin İçerikli Türkçe Dokümanların Sınıflandırılması, 2010.
- [53] İ. O. Yiğit, A. F. Ateş, M. Güvercin, H. Ferhatosmanoğlu, B. Gedik, Çağrı Merkezi Metin Madenciliği Yaklaşımı, *2017 25th Signal Processing and Communications Applications Conference, SIU 2017*, Jun. 2017, pp. 1–4.
- [54] E. Sonat Erdem, Ses Sinyallerinde Duygu Tanıma Ve Geri Erişimi, 2014.
- [55] C. Parlak, İnsan Sesinden Duygu Tanıma, 2015.

- [56] Veri Madenciliği Nedir?
https://www.beyaz.net/tr/guvenlik/makaleler/veri_madenciligi_nedir.html, 2021.
- [57] Veri Madenciliği Nedir? Nasıl Yapılır? Süreç ve Yöntemleri
<https://www.gtech.com.tr/veri-madenciligi-nedir-ve-nasil-yapilir>, 2021.
- [58] E.A.Yurt, Türkçe Metinlerde Duygu Analizi, 2015.
- [59] BERT Modeli ile Türkçe Metinlerde Sınıflandırma Yapmak
<https://medium.com/@toprakucar/bert-modeli-ile-türkçe-metinlerde-sınıflandırma-yapmak>, 2021.
- [60] J.Devlin, M. Chang, K.Lee, K.Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019.
- [61] Tutorial: Quickstart — TextBlob
<https://textblob.readthedocs.io/en/dev/quickstart.html#create-a-textblob>, 2021
- [62] Makine Öğrenmesi Nedir ? <https://medium.com/türkiye/makine-öğrenmesi-nedir>, 2021.
- [63] COGNITIVE COMPUTING AND MACHINE LEARNING
<http://www.cognub.com/index.php/cognitive-platform>, 2021.
- [64] Makine Öğrenmesi Nedir ·<https://www.emrealadag.com/makine-ogrenmesi-nedir.html>. 2021.
- [65] Machine Learning — Classification — Naive Bayes
<https://medium.com/@ekrem.hatipoglu/machine-learning-classification-naive-bayes-part-11>, 2021.
- [66] Naive Bayes Classifier. https://erdincuzun.com/makine_ogrenmesi/naive-bayes-classifier, 2021.
- [67] K. Kirasich, T. Smith, B. Sadler, Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets, 2018.
- [68] L. He, F. Kong, Z. Shen, Multiclass SVM based land cover classification with multisource data, 2005.

- [69] A. D. Kulkarni, B. Lowe, Random Forest Algorithm for Land Cover Classification International Journal on Recent and Innovation Random Forest Algorithm for Land Cover Classification, 2016.
- [70] Makine Öğrenimi (Destek Vektör Makineleri) <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-4-destek-vektor-makineleri>, 2021.
- [71] Decision Tree (Karar Ağacı) https://erdincuzun.com/makine_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama, 2021.
- [72] Makine Öğrenimi (Karar Ağaçları) <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agac-lari>, 2021.
- [73] Makine Öğrenimi (k-En Yakın Komşuluk) <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-2>, 2021.
- [74] Veri Nedir? <https://www.iienstitu.com/blog/veri-nedir>, 2021.
- [75] Common Voice. <https://commonvoice.mozilla.org/tr/about>, 2021.
- [76] Python Programlama Dili Nedir ? Nerelerde Kullanılır ? <https://muhendistan.com/python-nedir/>, 2021.
- [77] CountVectorizer in Python. <https://www.educative.io/edpresso/countvectorizer-in-python>, 2021.
- [78] TF-IDF (Term Frequency — Inverse Document Frequency) <https://medium.com/algorithms-data-structures/tf-idf-term-frequency-inverse-document-frequency>, 2021.
- [79] Metin İşleme 1 — Eski Tarz Yöntemler (Bag of Words ve TFXIDF) <https://medium.com/deep-learning-turkiye/metin-isleme-1-eski-tarz-yontemler-bag-of-words-ve-tfxidf>, 2021.
- [80] Confussion Matrix (Karmaşıklık Matrisi) <http://tolgaoguz.net/2018/06/26/confusion-matrix>, 2021.



ÖZGEÇMİŞ



Ad Soyad: Engin DEMİR

Doğum Yeri ve Tarihi:

E-Posta:

Lisans: Karabük Üniversitesi, Bilgisayar Mühendisliği (İngilizce)

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

[1] E. DEMİR, A. TEPECİK, “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.

[2] E. DEMİR, A. TEPECİK, “Evaluation of Time Performance of Artificial Bee Colony and Particle Swarm Optimization Algorithms”, 6th Global Conference on Computer Engineering, (Kabul Edildi).