

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SOHBET ROBOTU VERİ SETİ İLE NİYET
SINIFLANDIRMASINDA GELENEKSEL
VE MODERN YAKLAŞIMLARIN
KARŞILAŞTIRILMASI

Kardel Rûveyda ÇETİN

YÜKSEK LİSANS TEZİ

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Danışman

Prof. Dr. Fatma AYDIN AKGÜN

Nisan, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SOHBET ROBOTU VERİ SETİ İLE NİYET
SINIFLANDIRMASINDA GELENEKSEL VE MODERN
YAKLAŞIMLARIN KARŞILAŞTIRILMASI

Kardel Rûveyda ÇETİN tarafından hazırlanan tez çalışması (10.04.2025) tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Mühendisliği Anabilim Dalı, Matematik Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Fatma AYDIN AKGÜN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Fatma AYDIN AKGÜN, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. İbrahim EMİROĞLU, Üye
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Oğuzhan KIVRAK, Üye
Bandırma Onyedli Eylül Üniversitesi

Danışmanım Prof. Dr. Fatma AYDIN AKGUN sorumluluğunda tarafımca hazırlanan “Sohbet Robotu Veri Seti İle Niyet Sınıflandırılmasında Geleneksel ve Modern Yöntemlerin Karşılaştırılması” başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Kardel R veyda ETİN

İmza



Bu çalışma, “...Yıldız Teknik Üniversitesi Bilimsel Araştırma Proje Koordinatörlüğü’nün 1234A...” numaralı projesi ile desteklenmiştir.



Anneme

TEŐEKKÜR

Bu tez alıőmasında bilgi ve tecrübesiyle bana her zaman yardımcı olan saygıdeęer danıőman hocam Prof. Dr. Fatma AYDIN AKGÜN'e, veri setinin oluőturulmasında ve yöntemlerin incelenmesinde desteklerini esirgemeyen alıőma arkadaşlarım Alper ULUSOY ve Hamza AĖAR'a, tez süreci boyunca yönlendirmeleri ve daimi desteęi için Tarlan MAHOUTI'ye, yüksek lisans eęitimim süresince ders aldığım tüm hocalarıma ve öğretim hayatım boyunca destekleri ile her zaman yanımda olan sevgili anneme sonsuz teşekkürlerimi sunarım.

Kardel Ruvéyda ETİN

İÇİNDEKİLER

SİMGE LİSTESİ	ix
KISALTMA LİSTESİ	x
ŞEKİL LİSTESİ	xiii
TABLO LİSTESİ	xiv
ÖZET	xv
ABSTRACT	xvii
1 GİRİŞ	1
1.1 Literatür Özeti.....	4
1.2 Tezin Amacı.....	10
1.3 Hipotez.....	11
2 VERİ VE VERİ SETİ	12
2.1 Veri	12
2.2 Veri Seti	13
2.3 Veri Kalitesi.....	13
2.4 Veri Kategorilendirme	15
2.5 Veri Ön İşleme Süreçleri	15
2.5.1. Veri Temizleme	16
2.5.2. Veri Birleştirme	16
2.5.3. Veri İndirgeme	16
2.5.4. Veri Dönüştürme	16

2.6 Verilerin Maskelenmesi	16
2.7 Veri Ürettirme İşlemleri.....	17
3 TEMEL NLP KAVRAMLARI	18
3.1. FiT.....	18
3.2 Dönem Sayısı (Epoch)	18
3.3 Oft	19
3.4 Eğitim ve Test (Train-Test)	19
3.5. IC	20
4 PERFORMANS METRİKLERİ	22
4.1. Doğruluk (Accuracy)	22
4.2. Kesinlik (Precision)	23
4.3. Duyarlılık (Recall)	23
4.6.4 F1-Skoru	24
4.5. ROC eğrisi ve AUC	24
4.6. RMSE ve MAE	25
5 GELENEKSEL YÖNTEMLER	26
5.1. Word2vec	26
5.2. GloVe.....	27
5.3. FastText	28
6 MODERN DİL İŞLEME YÖNTEMLERİ	29
6.1 BERT ve BERTurk.....	29
6.2 ConvBERT ve ConvBERTurk.....	30
6.3 Electra	30
6.4 RAG.....	31
7 VERİ TOPLAMA VE TARAMA YÖNTEMLERİ	33
7.1 Web Kazıma Yöntemleri	34

7.1.1 BeautifulSoup	34
7.1.2 Scrapy	34
7.1.3 Selenium	34
8 SOHBET ROBOTU VERİ SETİ	35
8.1 Üretilen Verinin RAG ile İşlenmesi	35
8.2 Veri Setinde Dikkat Edilen Hususlar	36
8.3 Veri Seti Versiyonları: V1 ve V2 Arasındaki Farklar ve Ön İşleme Süreçleri	37
9 BULGULAR	40
9.1 V1 Veri Seti ile Geleneksel Yöntemlere Ait Bulgular.....	40
9.2 V1 Veri Seti ile Modern Yöntemlere Ait Bulgular.....	44
9.3 V2 Veri Seti ile Geleneksel Yöntemlere Ait Bulgular.....	49
9.4 V2 Veri Seti ile Modern Yöntemlere Ait Bulgular.....	53
10 SONUÇ	60
KAYNAKÇA	64

SİMGE LİSTESİ

\hat{y}^i	Tahmin Edilen Değer
y_i	Gerçek Değer
n	Veri Sayısı
X	Matris
x_{ij}	Kelime i 'nin kelime j ile birlikte görünme sıklığı
W_i	Kelime i için öğrenilen vektör
W_j	Kelime j için öğrenilen vektör
b_i	i Bias Terimi
b_j	j Bias Terimi
J	Kayıp Fonksiyonu

KISALTMA LİSTESİ

AI	Yapay Zeka
AUC	Eđri Altındaki Alan
ALBERT	Hafifletilmiş BERT
BERT	Çift Yönlü Kodlayıcı Temsilleri
CBOW	Sürekli Kelimeler Torbası
CNN	Konvolüsyonel Sinir Ağ
CNN-LSTM	Konvolüsyonel Sinir Ağı ve Uzun Kısa Süreli Bellek Ağı
CNN-SVM	Konvolüsyonel Sinir Ağı ve Destek Vektör Makinesi
ConvBERT Transformatör Modeli	Konvolüsyonel Çift Yönlü Kodlayıcı Temsilleri
CPU	Merkezî İşlem Birimi
DeBERTa	Çözümleme Geliştirilmiş BERT ve Ayrık Dikkat
DL	Derin Öğrenme
Doc2Vec	Belgeyi Vektöre Dönüştürme
Electra	Verimli Kodlayıcı Temsilleri
Faiss	Facebook Yapay Zeka Benzerlik Araması
FastText	Hızlı Metin Temsili
FiT	İnce Ayar
FP	Yanlış Pozitif
FN	Yanlış Negatif
FPR	Yanlış Pozitif Oranı
GloVe	Küresel Vektörler ile Kelime Temsili
GPT	Üretken Önceden Eğitilmiş Transformatör
GPT-3.5	Üretici Önceden Eğitilmiş Dönüştürücü-3.5
GPT-4	Üretken Önceden Eğitilmiş Transformatör-4
IC	Intent Classification

JSON	JavaScript Nesne Gösterimi
LangChain	Dil Zinciri
LLM	Büyük Dil Modelleri
LSTM	Uzun Kısa Süreli Bellek Ağı
LSA	Gizli Anlamsal Analiz
MAE	Ortalama Mutlak Hata
MBART Dönüştürücüler	Çok Dilli İki Yönlü ve Otomatik Geri Beslemeli
ML	Makine Öğrenmesi
MLM	Maskeli Dil Modelleme
MLIC	Çok Etiketli Niyet Sınıflandırma
MT5	Çok Dilli T5
NLP	Doğal Dil İşleme
OCR	Optik Karakter Tanıma
OFT	Aşırı Uyum
RAG	Geri Alma Artırılmış Nesil
RL	Pekiştirmeli Öğrenme
RMSE	Ortalama Hata Karekökü
ROC	Alıcı İşletim Karakteristik Eğrisi
RTD	Değiştirilen Belirteç Algılama
SGM	Atlama Dilbilgisi Modeli
SF	Alan Doldurma
SLIC	Tek Etiketli Niyet Sınıflandırması
SVM	Destek Vektör Makineleri
SVD	Tekil Değer Ayrıştırması
TF-IDF	Terim Sıklığı - Ters Belge Sıklığı
TFRC	TensorFlow Araştırma Bulutu
TRBART Dönüştürücüler	Türkçe İki Yönlü ve Otomatik Geri Beslemeli
TRT5	Türkçe Transformatör 5
TP	Gerçek Pozitif
TN	Gerçek Negatif
ULMFiT	Evrensel Dil Modeli İnce Ayarı

UFT	Yetersiz Öğrenme
WebBaseLoader	Web Tabanlı Veri Yükleyici
WS	Web Kazıma
WC	Web Tarama
Word2Vec	Kelimeden Vektöre
XML	Genişletilebilir İşaretleme Dili



ŞEKİL LİSTESİ

Şekil 2.1 Veri Ürettirme İşlemleri	17
Şekil 8.1 Üretilen Verinin RAG ile İşlenmesi	35
Şekil 9.2 V1 (50 dönem sayısı) ve V1 (100 dönem sayısı) FastText Yöntemlerin Sonuçlarının Karşılaştırılması	44
Şekil 9.3 V1 (50 dönem sayısı) Modern Yöntemlerin Süreleri	46
Şekil 9.4 V1 (100 dönem sayısı) Modern Yöntemlerin Sonuçlarının Değerlendirilmesi	48
Şekil 9.5 V1 ve V2 (dönem sayısı 50) Geleneksel Yöntemlerin F1 Sonuçlarının Değerlendirilmesi	50
Şekil 9.7 V2 (100 dönem sayısı) Modern Yöntemlerin Sonuçları	58
Şekil 9.8 En İyi Modellerin Test Performanslarının Karşılaştırılması	59

TABLO LİSTESİ

Tablo 8.1	Versiyon 1 (V1) ve Versiyon 2 (V2) Detayları	38
Tablo 9.1	FastText -Glove -Word2Vec V1(50 dönem sayısı) Sonuçları.....	41
Tablo 9.2	FastText- Glove- Word2vec V1 (100 dönem sayısı) Sonuçları	43
Tablo 9.3	BERT-ConvBert-Electra V1 (50 dönem sayısı) Sonuçları.....	45
Tablo 9.4	RAG V1 Sonuçları.....	46
Tablo 9.5	BERT-ConvBert-Electra V1 (100 dönem sayısı) Sonuçları.....	47
Tablo 9.6	FastText-Glove-Word2Vec V2 (50 dönem sayısı) Sonuçları.....	49
Tablo 9.8	BERT-ConvBert-Electra V2 (50 dönem sayısı) Sonuçları.....	54
Tablo 9.9	RAG V2 Sonuçları.....	55
Tablo 9.10	BERT-ConvBert-Electra V2-100 Epoch Sonuçları	56

Sohbet Robotu Veri Seti İle Niyet Sınıflandırılmasında Geleneksel ve Modern Yöntemlerin Karşılaştırılması

Kardel R veyda ETİN

Matematik M hendisliđi Anabilim Dalı

Matematik M hendisliđi. Programı

Y ksek Lisans Tezi

Danıřman: Prof. Dr. Fatma AYDIN AKG N

Bu tez alıřmasında, sohbet robotu (chatbot) verileri kullanılarak geleneksel ve modern dil modellerinin kapsamlı bir karřılařtırması yapılmaktadır. Geleneksel y ntemler arasında; b y k metin verileri  zerinde eđitilerek kelime vekt rleri oluřturan ve kelimeler arası iliřkileri temsil eden Kelimedenden Vekt re (Word to Vector, Word2Vec), K resel Vekt rler ile Kelime Temsili (Global Vectors for Word Representation, GloVe) ve Hızlı Metin Temsili (Fast Text Representation, FastText) gibi modeller incelenmiřtir. Literat rde bu y ntemlerin sınırlılıkları nedeniyle alıřmada ayrıca modern y ntemlere de yer verilmiřtir. Bu kapsamda, derin  đrenme temelli dil anlama ve  retme kabiliyetine sahip Transformat rlerden ift Y nl  Kodlayıcı G sterimleri (Bidirectional Encoder Representations from Transformers, BERT), maskeleyme tekniđiyle verimli eđitim sađlayan Token Deđiřimlerini Dođru Őekilde Sınıflandıran Bir Kodlayıcıyı Verimli Őekilde  đrenme (Efficiently Learning an Encoder that Classifies Token Replacements Accurately, ELECTRA) ve konvol syonel yapısıyla  ne ıkan Transformat rlerden Evriřimli ift Y nl  Kodlayıcı G sterimleri (Convolutional

Bidirectional Encoder Representations from Transformers, ConvBERT) gibi modern modeller değerlendirilmiştir.

Çalışmada, Türkçe diline özel olarak Türkçe BERT, ConvBERT ve ELECTRA modelleri üzerinde detaylı ince ayar (fine-tuning, FiT) süreçleri uygulanmış ve performansları karşılaştırılmıştır. Ayrıca, bilgiyi geri getirme ile nesil modelini birleştiren yeni bir yaklaşım olan Geri Alma Artırılmış Nesil (Retrieval-Augmented Generation, RAG) yöntemi de analiz edilmiştir. Bu yöntem kapsamında, Dil Zinciri (Language Chain, LangChain) aracılığıyla Üretken Önceden Eğitilmiş Transformatör-4 (Generative Pre-trained Transformer-4, GPT-4) Büyük Dil Modeli (Large Language Model, LLM) kullanılarak sınıf tahmini yapılmıştır.

Veri seti, Web Tabanlı Veri Yükleyici (Web Base Loader, WebBaseLoader) kütüphanesi aracılığıyla yapılan web kazıma (web scraping, WS) işlemleri ve RAG destekli üretimle oluşturulmuştur. Performans değerlendirmesi; doğruluk (accuracy), hassasiyet (precision), geri çağırma (recall) ve F1 skoru (F1 Score) gibi temel metriklerle yapılmıştır. Literatürde genellikle hazır veya sentetik İngilizce ve Türkçe veri setleri kullanılırken, bu çalışmada gerçek bir sohbet robotu uygulaması için özel geliştirilen Türkçe veri seti kullanılmıştır. Bu yönüyle çalışma, hem veri niteliği hem de uygulama odaklı yapısıyla literatüre katkı sunmaktadır.

Veriler hem ham hem de ön işlemden geçirilmiş biçimleriyle modellere uygulanmış; Türkçe'nin morfolojik yapısı ve kelime türetme kuralları göz önünde bulundurularak modellerin başarımı değerlendirilmiştir. Bu çalışma, Türkçe dilinde daha etkili ve başarılı sohbet robotları geliştirilmesine katkı sağlamayı ve gelecekte yapılacak araştırmalara yol göstermeyi amaçlamaktadır.

Anahtar Kelimeler: Büyük Dil Modeli, İnce Ayar, Bilgi Destekli Metin Üretimi, Web'den Veri Toplama, Sohbet Botu

Comparison of Traditional and Modern Methods in Intent Classification Using a Chatbot Dataset

Kardel Rveyda ETİN

Department of Mathematical Engineering
Master of Science

Supervisor: Prof. Dr. Fatma AYDIN AKGN

In this thesis, a comprehensive comparison of traditional and modern language models is conducted using chatbot data. Traditional methods examined include Word to Vector (Word2Vec), Global Vectors for Word Representation (GloVe), and Fast Text Representation (FastText), which generate word embeddings by training on large-scale text corpora and aim to represent semantic relationships between words. Due to the limitations frequently noted in the literature, the study also incorporates modern approaches. These include Bidirectional Encoder Representations from Transformers (BERT), which enables deep learning-based language understanding and generation; Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA), which improves training efficiency through token replacement strategies; and Convolutional Bidirectional Encoder Representations from Transformers (ConvBERT), which enhances performance through convolutional architectures.

In this study, extensive fine-tuning (FiT) was applied to Turkish-specific versions of BERT, ConvBERT, and ELECTRA, and their performances were comparatively evaluated. Additionally, a recent approach called Retrieval-Augmented Generation (RAG), which combines retrieval mechanisms with generative models, was also analyzed. As part of this approach, class prediction was performed using the

Generative Pre-trained Transformer 4 (GPT-4) Large Language Model (LLM) via the Language Chain (LangChain) framework.

The dataset was created through web scraping (WS) using the Web Base Loader (WebBaseLoader) library and was further enriched with data generated through the RAG-supported pipeline. Performance evaluation was carried out using fundamental metrics such as accuracy, precision, recall, and F1 score. While pre-built or synthetic English and Turkish datasets are commonly used in the literature, this study utilizes Turkish dataset specifically developed for a real-world chatbot application. In this regard, the study contributes to the literature both in terms of data quality and practical applicability.

The models were tested with both raw and preprocessed versions of the dataset. Considering the morphological complexity and derivational rules of the Turkish language, the performance of the models was thoroughly evaluated. This research aims to support the development of more effective and high-performing Turkish-language chatbots and to guide future academic and industrial studies in this field.

Keywords: LLM, Fine-Tuning, RAG, Web Scraping, Chatbot.

1 GİRİŞ

Yapay zeka (Artificial Intelligence, YZ), insan zekasını taklit edebilen ve öğrenme ile uyum sağlama yeteneği sayesinde minimum insan müdahalesiyle karmaşık görevleri bağımsız bir şekilde gerçekleştiren sistemlerin geliştirilmesini amaçlayan bir teknoloji alanı olarak tanımlanır [1], [2]. Bu sistemler, büyük veri kümelerine maruz kalarak öğrenir ve zamanla gelişir; bu süreçte, insanların fark edemeyeceği örüntüleri ve ilişkileri belirler [3]. YZ'nin temel amacı, makinelerin insan benzeri düşünme, öğrenme ve problem çözme becerileri kazanmasını sağlamaktır [4]. Bu öğrenme süreci genellikle algoritmalar aracılığıyla gerçekleştirilir. Algoritmalar, yapay zekanın analiz ve karar verme süreçlerini yönlendiren kurallar ve talimatlardan oluşur [5].

Derin öğrenme (Deep Learning, DL), yapay sinir ağlarını kullanarak bilgiyi işleyen ve insan beyninin yapısını, işleyişini taklit eden bir uzmanlık alanıdır. Bu yöntem; ağ tasarımı, eğitimi ve değerlendirilmesi gibi pratik unsurları kapsayan geniş bir çerçeve sunarak, derin öğrenme araştırmalarının mevcut ve gelecekteki yönlerini anlamak için sağlam bir temel oluşturur [6].

Makine öğrenmesi (Machine Learning, ML) ise, YZ'nin popüler bir alt dalı olarak bilinir ve algoritmaların etiketli veya etiketlenmemiş veriler üzerinde eğitilmesiyle tahminlerde bulunmayı veya bilgileri kategorize etmeyi amaçlar. Makine öğrenimi genellikle üç ana bileşen etrafında şekillenmektedir. Bu süreçler; karar süreci, hata fonksiyonu ve model optimizasyon sürecidir.

Karar süreci, makine öğrenimi algoritmalarının tahmin veya sınıflandırma yaptığı aşamadır. Bu aşamada algoritma, etiketli ya da etiketlenmemiş verilerden yararlanarak verideki örüntüler hakkında öngörülerde bulunur.

Hata fonksiyonu, modelin tahminlerini değerlendirerek modelin doğruluğunu belirlemede kullanılır. Eğer bilinen örnekler varsa, hata fonksiyonu bu örneklerle karşılaştırmalar yaparak modelin doğruluğunu ölçer.

Model optimizasyon süreci ise eğitim setindeki veri noktalarına daha iyi uyum sağlamak için modelin ağırlıklarını ayarlayarak modelin tahminleri ile bilinen örnekler arasındaki farkı azaltmaya çalışır. Algoritma, bu iteratif "değerlendir ve optimize et" sürecini, belirli bir doğruluk eşiğine ulaşıncaya kadar ağırlıkları otomatik olarak güncelleyerek tekrarlar. Bu sayede bilgisayarlar, karmaşık veri kümelerinden faydalı bilgiler çıkarabilir ve gelecekteki veri örnekleri üzerinde tahminlerde bulunabilir [8]. ML, verilerden örüntüleri ve ilişkileri öğrenerek kredi kartı dolandırıcılığını tespit etmek, sürücüsüz araçları yönlendirmek ve kullanıcıların ilgi alanlarına uygun elektronik içerik önerileri sunmak gibi birçok önemli alanda aktif olarak kullanılmaktadır [7]. Makine öğrenimi yöntemleri de kendi içerisinde denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme olmak üzere dört ana kategoriye ayrılır.

Denetimli öğrenme, bir modelin belirli bir hedefi tahmin etmesine veya sınıflandırmasına olanak tanır. Örneğin denetimli öğrenme çalışmasında, bir istenmeyen e-posta filtreleme uygulamasında kullanıcıdan gelen örnek e-postalar ile birlikte bu e-postaların sınıfını (istenmeyen e-posta veya ham) öğrenerek yeni gelen e-postaları sınıflandırmayı öğrenebilir. Ayrıca bir regresyon probleminde, arabanın fiyatını tahmin etmek için araç özelliklerini (kilometre, yaş, marka vb.) kullanarak hedef bir sayısal değer tahmin eder. Bu öğrenme sürecinde etiketli veri setleri kullanılarak işlemler gerçekleştirilmektedir [9].

Denetimsiz öğrenme, algoritmanın gruplar arasındaki ilişkileri veya desenleri keşfetmesine olanak tanıyabilen bir öğrenme çeşididir. Örneğin, bir blog ziyaretçi verisini analiz ederek benzer ziyaretçi gruplarını tanımlayan bir kümeleme algoritması kullanılabilir ve oluşturulabilir. Bu tür algoritmalar, verilerin etiketli olarak belirlenmediği durumlarda bile, örüntüleri keşfederek kullanıcıların davranışları hakkında bilgi sağlayabilir ve bu kullanıcı davranışlarıyla ilgili bir analiz çıkarılmasında yardımcı rol oynayabilir [9].

Yarı denetimli öğrenme, genellikle etiketlenmiş ve etiketlenmemiş veri karışımını kullanır. Örneğin, bir fotoğraf paylaşım servisinde yüklenen fotoğraflardaki kişilerin otomatik olarak tanıma işlemi gerçekleştirilebilir. Buna benzer bir işlemde, kişinin kim olduğunu öğrenmek için yalnızca bir etiket eklenilmesi yeterlidir [9].

Pekiştirmeli öğrenme (Reinforcement Learning, RL), kullanan bir ajan, çevreyi gözlemleyip yaptığı hareketlere göre ödüller veya cezalar alarak en iyi stratejiyi öğrenmeye çalışır. Bu süreçte, robotlar gibi sistemler çevreleriyle etkileşime girerek, daha karmaşık görevleri yerine getirmek amacıyla insan geri bildirimlerine dayalı hedefler belirleyebilirler [9]. İnsan geri bildiriyle robotlara yürüme gibi karmaşık davranışları öğretmek için pekiştirmeli öğrenme algoritmalarının etkili bir şekilde kullanıldığı gösterilmiştir. Bu yöntem, ödül fonksiyonuna doğrudan erişim olmadan, insan gözetimi gereksinimini azaltarak robotların gelişimini hızlandırabilmektedir [10].

Makine öğrenmesindeki çeşitlilik, makine öğreniminin çok yönlü doğasını ve farklı uygulama alanlarındaki etkisini ortaya koymaktadır. Özellikle gerçek dünya uygulamalarında, farklı öğrenme türlerinin etkili bir şekilde kullanılması, sistemlerin daha verimli ve etkili hale gelmesini sağlamaktadır. Makine öğrenme projeleri, sistematik bir yaklaşım gerektirir ve her aşama, başarılı bir sonuç elde etmek için kritik öneme sahiptir. Proje genel bir bakış açısıyla başlar; burada problemin tanımlanması, kullanılacak verilerin belirlenmesi ve beklenen sonuçların netleştirilmesi gibi adımlar gerçekleştirilir. Ardından, verilerin toplanması aşamasına geçilir; bu aşamada toplanan verinin kalitesi, projenin başarısını doğrudan etkiler.

Doğal Dil İşleme (Natural Language Processing, NLP), bilgisayarların insan dilini anlamasını ve işlemeyi sağlayan bir teknoloji alanıdır [11]. Bu alan, metin ve konuşma verilerini analiz ederek anlam çıkarmayı, sınıflandırmayı ve tepki vermeyi amaçlar. Özellikle Niyet Sınıflandırma (Intent Classification, IC), kullanıcıların girdilerini analiz ederek ne istediklerini belirlemeye çalışır. Bu teknoloji; müşteri hizmetleri, sohbet robotları ve dil çeviri sistemleri gibi birçok alanda kullanılır ve insanlarla daha doğal bir iletişim kuran sistemlerin geliştirilmesine yardımcı olur.

Sohbet robotları, insanlarla doğal dilde etkileşim kurabilen yazılımlar olarak bilinmektedir. Bu teknolojiler, kullanıcıların sorularına anında cevap verebilecek kapasitede yazılımlar olduğu için özellikle müşteri hizmetleri, bilgi edinme ve günlük görevlerde önemli bir yer tutmaktadır. İnsan benzeri bir diyalog yeteneği sundukları için işletmelerin verimliliğini artırır ve kullanıcı deneyimini geliştirir. Ayrıca sürekli öğrenme ve gelişim süreçleri sayesinde daha karmaşık görevleri

yerine getirebilir hale gelirler. Bu yönleriyle sohbet robotları, dijital dönüşümün vazgeçilmez bir parçası haline gelmiştir [12,13].

1.1 Literatür Özeti

Bu tez çalışması, YZ'nin NLP alanındaki uygulamalarını WS yöntemi ile elde edilmiş bir sohbet robotu verisi özelinde incelenmekte ve özellikle IC problemleri üzerine odaklanmaktadır. Bu bağlamda, geleneksel ve modern yöntemlerin karşılaştırılması yapılarak en uygun yöntemlerin belirlenmesi hedeflenmektedir. IC problemlerinde başarıyı değerlendirmek için çeşitli yöntemler ve modeller karşılaştırılmaktadır. Tez kapsamında Türkçe dilinde çeşitli modellerin başarılarının değerlendirildiği literatür çalışmaları detaylı bir şekilde incelenmiş olup bu çalışmaların ayrıntıları kronolojik sıralama ile aşağıda sunulmuştur:

- 2020 yılında Dođuhan Yeke [14] belge sıralaması performansını iyileştirmek için bağlamsal kelime düğümüne dayanan bir sorgu genişletme yaklaşımı sunmuştur. Çalışmada, orijinal sorguyu anlamsal olarak benzer terimlerle genişletmek için BERT kelime düğümünden Çift Yönlü Enkoder Temsilleri kullanılmıştır. BERT'den kelime düğümünü çıkarmanın en iyi yöntemine karar verildikten sonra, sorgu en iyi aday terimleri ile genişletilmiştir. Birincil hedef olarak, BERT'nin vektör uzayındaki terimleri temsil etmek için en yaygın işlem olarak bilinen Word2Vec modeli üzerinde nasıl performans gösterdiği gösterilmiştir. Deneyler ile BERT'in ingilizce veri setinde bilgi erişim sistemlerinde en popüler sorunları ele aldığı ve değerlendirme metriklerinde Word2Vec'ten daha iyi performans gösterdiği ortaya konulmuştur
- 2020 yılında Dođançan Kımık [15] tarafından haber metinleri üzerinde metin sınıflandırma başarısını artırmaya yönelik olarak Terim Sıklığı - Ters Belge Sıklığı (Term Frequency-Inverse Document Frequency, TF-IDF) terim ağırlıklandırma ve Belgeyi Vektöre Dönüştürme, (Document to Vector, Doc2Vec) kelime gömme yöntemleri kullanılmıştır. Çalışma, ardışık kelime gruplarının tespiti ile sınıflandırma doğruluğunun nasıl etkilendiğini analiz etmeyi amaçlamaktadır. 1150Haber, Hürriyet6c-1k, Milliyet9c-1k ve AA Haber veri kümeleri üzerinde 28 farklı haber kategorisi için yaklaşık 35.000 haber metni kullanılmıştır. Öncelikle, TF-IDF ve Doc2Vec

teknikleri yalnızca bireysel sınıflandırıcılar kullanılarak uygulanmış, ardından bu yöntemler ardışık kelime grubu tespiti ile desteklenmiş ve karşılaştırmalı analizler yapılmıştır. Doc2Vec yönteminin alt teknikleri test edilmiş ve hibrit sonuçlar elde edilmiştir. Sonuçlarda, TF-IDF için %94, Doc2Vec için %95 doğruluğa ulaşarak ardışık kelime gruplarının sınıflandırma doğruluğunu genel olarak artırdığını gösterilmiştir.

- 2020 yılında Şeyhmus Yılmaz [16], Türkçe'de soru sınıflandırma problemini çözmek amacıyla derin öğrenme tekniklerinden faydalanmıştır. Türkçenin aglutinatif yapısına uygun olarak Word2Vec temsilleri kullanılarak kelime vektörleri üretilmiştir. Derin öğrenme modelleri olarak Konvolüsyonel Sinir Ağı (Convolutional Neural Network, CNN), Uzun Kısa Süreli Bellek Ağı (Long Short-Term Memory, LSTM), Konvolüsyonel Sinir Ağı ve Uzun Kısa Süreli Bellek Ağı (Convolutional Neural Network-Long Short-Term Memory, CNN-LSTM) ve Konvolüsyonel Sinir Ağı ve Destek Vektör Makinesi (Convolutional Neural Network and Support Vector Machine, CNN-SVM) yapıları kullanılmış, %94 doğruluk oranı elde edilmiştir. Bu çalışmada, Sürekli Kelimeler Torbası (Continuous Bag of Words, CBOW) ve Atlamalı Dilbilgisi Modeli (Skip-Gram, SGM) yöntemleri ile çeşitli derin öğrenme mimarileri kıyaslanmıştır. Sonuçlar, SGM yönteminin CNN, CNN-LSTM ve CNN-SVM modellerinde CBOW'a göre daha yüksek doğruluk verdiğini, ancak CBOW yönteminin LSTM yapısında daha başarılı olduğunu göstermiştir. En iyi sonuç %94 doğrulukla 300 boyutlu özellik vektörlü SGM kullanılarak CNN modelinde elde edilmiştir. Soru sınıflandırma veritabanı için uygun korpus seçiminin önemine dikkat çekilmiştir. İngilizce dilinde yapılan benzer çalışmalarla kıyaslandığında, Türkçedeki sınıflandırma başarısı dil yapısının etkisiyle daha düşük çıkmıştır.
- 2021 yılında Bahadır Özkan [17], sohbet robotları için niyet sınıflandırma alanında önemli bir çalışma sunmuştur. Bu tezde, sohbet robotlarının konuşmalarında niyet tahminini geliştirmek amacıyla çeşitli NLP teknikleri kullanılmıştır. Çalışmada, TF-IDF, Doc2Vec ve BERT gibi hem geleneksel hem de modern kelime gömme yöntemleri incelenmiş; çoklu sınıf ve çoklu etiket tahmini için lojistik regresyon, rastgele orman ve yapay sinir ağları gibi algoritmalar uygulanmıştır. Veriler, sinema bileti rezervasyonu,

restoran rezervasyonu ve taksi çağırma gibi üç farklı alandan toplanmıştır. Çalışmanın bulguları, BERT modelinin özellikle sinema bileti veri kümesinde diğer modellere göre üstün performans sergilediğini göstermiştir. Sinir ağları ile TF-IDF kombinasyonu, performans açısından BERT'in sinir ağları üzerinde eğitilmiş haliyle benzer sonuçlar vermiştir ancak genel olarak BERT, diğer modelleri geride bırakmıştır. BERT ve TF-IDF'in birleşimi tüm veri kümelerinde, yalnızca BERT kullanılan modele göre bir miktar daha yüksek doğruluk sağlamış ve eğitim süresini artırmadan performans metriklerini iyileştirmiştir. Doc2Vec ise genel olarak en düşük performansı göstermiştir; hatta bazı durumlarda naif tahminlerin bile gerisinde kalmıştır. Çalışmanın sonuçları, sohbet robotlarının etkinliğini artırmak için ileri düzey NLP yöntemlerinin önemini vurgulamaktadır.

- 2022'de Anıl Özberk [18] tarafından yapılan yüksek lisans tez çalışması, Türkçe Twitter verilerinde saldırgan dilin tespiti amacıyla önceden eğitilmiş ve ince ayarlı BERT modellerinin yanı sıra lojistik regresyon, karar ağacı, rastgele orman ve Destek Vektör Makineleri (Support Vector Machine-SVM) gibi klasik makine öğrenimi yöntemlerinin etkilerini araştırmaktadır. Özberk, BERT modellerinin klasik yöntemlere göre belirgin bir üstünlük sağladığını, bu yöntemler arasında SVM'nin en yüksek F1 skorunu elde ettiğini belirtmektedir. Çalışma sonucunda, Türkçe dilinde önceden eğitilmiş büyük veri setleriyle daha fazla iterasyona sahip BERT modellerinin performans açısından üstün olduğu, ayrıca Türkçe tweet verilerine özel ince ayar tekniklerinin sınıflandırma performansını artırdığı gözlemlenmiştir.
- 2022 yılında Onur Emre Atıcı [19] tarafından kaleme alınan " Röportajcı: Dönüşüm modellerini kullanarak yetkinlik bazlı mülakatları değerlendiren bir chatbot (Interviewster: A chatbot evaluating competency based interviews using transformer models)" başlıklı tez, mülakat uygulamalarının genel bir tanımını yaparak yetkinlik bazlı mülakatların yöntemlerini ve uygulamalarını detaylandırmaktadır. Tezde, Interviewster isimli sohbet botunun mimarisi, kullanılan teknolojiler, kütüphaneler ve makine öğrenmesi teknikleri ayrıntılı bir şekilde açıklanmıştır. Transformatör tabanlı modellerden BERT, Çözümleme Geliştirilmiş BERT

ve Ayrık Dikkat (Decoding-enhanced BERT with disentangled attention, DeBERTa) ve Electra'nın gerçek adayların yetkinlik bazlı mülakat sonuçlarına uygulanarak elde edilen değerlendirme sonuçları incelenmiştir. Elde edilen sonuçlara göre, modellerin doğruluk oranları sırasıyla %86, %78 ve %85 olup, en yüksek doğruluk oranı %87 ile en iyi sonucu vermiştir. Bu çalışma, Türkçe dilinde niyet sınıflandırma problemini çözerek yetkinlik bazlı mülakat gerçekleştirebilen bir sohbet robotu geliştirilmesiyle alanda bir ilk olmuştur.

- 2023 yılında Huda Alfigi [20] tarafından yapılan yüksek lisans tezinde, NLP alanında çok etiketli ve tek etiketli metin sınıflandırma problemleri ele alınmıştır. Çalışma, standart makine öğrenimi algoritmaları ve önceden eğitilmiş BERT transformatör modelleri kullanılarak gerçekleştirilmiştir. İngilizce ve Arapça metin veri kümeleri üzerinde yapılan bu çalışma, metin sınıflandırma tekniklerinin etkinliğini artırmayı amaçlamıştır. Sonuçlar, TF-IDF yönteminin tüm diğer teknikleri geride bıraktığını ve metin verisi üzerinde çok etkili olmadığını göstermektedir. BERT modeli, hem küçük hem de büyük veri setlerinde mükemmel sonuçlar elde ederken, Arapça BERT %95 başarı oranına ulaşmıştır. Bu sonuçlar, Arapça BERT'in, dengesiz ve çok etiketli veri setleri yerine dengeli ve tek etiketli veri setleriyle çalışmasının etkisiyle daha yüksek başarı gösterdiğini ortaya koymaktadır.
- 2023 yılında Suat Erkan [21], e-ticaret platformlarında yer alan müşteri yorumlarının BERT ve Hafifletilmiş BERT (A Lite BERT, ALBERT) dil modelleri kullanılarak analizini ele almıştır. Çalışma, çevrimiçi alışverişte müşteri memnuniyetini artırmak ve ürün geliştirme süreçlerine katkı sağlamak amacıyla duygu analizi gerçekleştirmiştir. Elde edilen bulgular, BERT modelinin %94 doğruluk oranıyla, ALBERT modelinin ise %92 doğruluk oranıyla performans gösterdiğini ortaya koymaktadır. Naive Bayes ve Destek Vektör Makineleri algoritmaları %90 doğruluk oranına ulaşırken, Rastgele Orman algoritması %89 oranında bir doğruluk elde etmiştir. Sonuçlar, BERT ve ALBERT modellerinin geleneksel makine öğrenimi algoritmalarına göre daha etkili olduğunu ve dil modellerinin duygu analizi tekniklerindeki önemini vurgulamaktadır.

- 2023 yılında Ayşegül Albayrak Doğan [22], ortaokul ve lise seviyesindeki ders sorularının sınıflandırılması amacıyla BERTurk modelini kullanmıştır. Veriler, EasyOCR kütüphanesi ile resimlerden metin formatına dönüştürülmüş ve ardından çeşitli veri ön işleme adımları uygulanarak sınıflandırma süreci gerçekleştirilmiştir. Bu süreçte, modelin Optik Karakter Tanıma (Optical Character Recognition, OCR) kaynaklı veri gürültüsüne ve veri ön işleme tekniklerine karşı performansı analiz edilmiştir. Çalışma sonucunda, Türkçe BERT modelinin özellikle eğitim adımlarının artırılması ve soru setinden gereksiz cevap şıklarının kaldırılması gibi düzenlemelerle %98 doğruluk oranına ulaştığı belirtilmiştir. Ayrıca, farklı dersler arasında doğru sınıflandırma oranları karşılaştırılmış; örneğin, Fizik alanında bazı hatalı tahminler yapılırken, Matematik sorularında hiç yanlış sınıflandırma yapılmadığı gözlemlenmiştir. Gelecekte, veri ön işleme süreçlerinin çeşitlendirilmesi ve OCR sonrası Türkçe karakter hatalarının giderilmesi gibi geliştirmelerle model performansının iyileştirilebileceği önerilmiştir.
- 2023 yılında Halil Hakan Sarıçiçek [23], yüksek lisans tezinde bilimsel materyaller için soru-cevap sohbet robotu tasarımını ve uygulamasını geliştirmiştir. Çalışmada, NLP ve LLM kullanılarak, kullanıcılara bilimsel sorularına yanıt verebilen bir sohbet robotu oluşturulması amaçlanmıştır. Anket yöntemine dayalı analizlerde, sohbet robotunun psikoloji alanında %78 başarı oranı ile etkili bir performans gösterdiği, ayrıca bilgisayar mühendisliği, tıp, uluslararası ticaret ve fen bilgisi öğretmenliği alanlarında da başarılı sonuçlar elde ettiği görülmüştür. Elde edilen bulgular, modelin akademik soru-cevap uygulamalarında yüksek performans sağladığını ve büyük dil modellerinin akademik sohbet robotları için potansiyelini ortaya koymaktadır.
- 2024 yılında Eyüp Halit Yılmaz [24], yazılım hata raporu öncelik tespiti alanında dönüştürücü tabanlı sınıflandırıcıların BERT ve LLMlerin performansını incelemiştir. Araştırmada, RAG ile bütünleşmiş bir vektör veritabanı kullanılarak LLM'lerin performansının artırıldığı ve BERT tabanlı sınıflandırıcıların, LLM'lere göre daha iyi sonuçlar elde ettiği sonucuna ulaşılmıştır. Bu bulgular, BERT'in yazılım hata raporu öncelik tespiti görevinde daha etkili olduğunu göstermektedir.

- 2024 yılında Aleyna Er [25], turizm sektöründe hizmet veren Setur'un müşteri geri bildirimlerini analiz ederek, memnun olunan ve gelişime açık alanları belirlemek amacıyla anahtar kelime üretimi ve konu modelleme çalışması yapılmıştır. Çalışmada, müşteri yorumlarını temsil eden anahtar kelimeler Üretici Önceden Eğitilmiş Dönüştürücü-3,5 (Generative Pre-trained Transformer 3,5, GPT-3.5), GPT-4 ve Gemini Pro gibi büyük dil modelleriyle üretilmiş, Türkçe Transformatör 5 (Turkish Transformer 5, TRT5), Çok Dilli T5 (Multilingual T5, MT5), Türkçe İki Yönlü ve Otomatik Geri Beslemeli Dönüştürücüler (Turkish Bidirectional and Auto-Regressive Transformers, TRBART) ve Çok Dilli İki Yönlü ve Otomatik Geri Beslemeli Dönüştürücüler (Multilingual Bidirectional and Auto-Regressive Transformers, MBART) modelleriyle öğrenme transferi gerçekleştirilmiştir. Anahtar kelimelerin, yorumlarla olan benzerlikleri ve anlamsal çeşitlilikleri incelenmiş, uzunlukları yorumlarla kıyaslanarak sıkıştırma oranları hesaplanmıştır. Konu modelleme aşamasında ise, üretilen anahtar kelimeler ile yorumların vektörleri kümelenecek konu başlıkları oluşturulmuş, her bir küme bir konu olarak ele alınarak, kullanıcıların hizmet sürecine dair yoğunlaştıkları alanlar tespit edilmiştir.

Literatürde yer alan yabancı kaynaklar incelendiğinde, modern tekniklerin sağladığı hız, doğruluk ve verimlilik gibi avantajlarla geleneksel yöntemlerin dayanıklılık, basitlik ve düşük hesaplama maliyetleri gibi güçlü yönleri detaylı bir şekilde ele alındığı görülmektedir. Bu tür kapsamlı analizler, yöntemlerin belirli senaryolardaki etkilerini değerlendirerek araştırmacılara daha derinlemesine bir bakış açısı kazandırmakta ve alandaki yenilikçi uygulamaların gelişimine katkı sunmaktadır. Son yıllarda bu alanda yer alan çalışmalar geleneksel yöntemlerden modern yaklaşımlara kadar geniş bir yelpazeyi kapsamaktadır, [26], [27], [28], [29], [30], [31], [32] bunlardan bazılarıdır. Geçmişten günümüze kadar olan gelişmeleri göz önünde bulundurarak yapılan kronolojik incelemeler, bu alandaki ilerlemeleri ve model performanslarındaki değişimleri anlamada kritik bir rol oynamaktadır.

1.2 Tezin Amacı

Sohbet robotu geliştirme alanında, geleneksel ve modern yöntemlerin karşılaştırılması önemli bir araştırma konusudur. Özellikle Türkçe'nin dil yapısının karmaşıklığı ve eklemeli bir yapıya sahip olması bu alandaki çalışmaları zorlaştırmaktadır.

Literatürde daha önce yapılan çalışmalarda, sohbet robotu geliştirme sürecinde ya yalnızca eski yöntemlerin (Word2Vec, TF-IDF, vb.) ya da yeni yöntemlerin (BERT, ELECTRA, vb.) performansları incelenmiş ve genellikle genelleştirilmiş veri setleri üzerinde çalışılmıştır. Bazı çalışmalar İngilizce veri setleri üzerinde gerçekleştirilmiş ve bu veri setlerinin, dilin analitik yapısı ve geniş çapta bulunan dengeli veri kaynakları nedeniyle daha yüksek başarı oranları sunduğu gözlemlenmiştir. Ancak, eski ve yeni yöntemlerin birlikte kapsamlı bir şekilde karşılaştırıldığı, Türkçe'nin karmaşık dil yapısına uygun ve gerçek dünyada kullanılabilir bir sohbet botu için özel olarak hazırlanmış bir veri seti üzerinde çalışılmadığı fark edilmiştir. Bu çalışmada, hem geleneksel hem de modern yöntemlerin yanı sıra ince ayar dışında RAG yöntemi de kullanılarak, Türkçe dil yapısına uyum düzeyleri ve performansları karşılaştırılmıştır.

Bu amaçla, çalışmada Word2vec, GloVe ve FastText gibi geleneksel yöntemlerle BERT, ConvBERT ve ELECTRA gibi modern yöntemlerin yanı sıra RAG yöntemi karşılaştırılarak, yöntemlerin Türkçe dil yapısı üzerindeki etkilerinin kapsamlı bir şekilde değerlendirilmiştir. Çalışmanın hedefleri şöyledir;

- Türkçe'nin karmaşık dil bilgisi kurallarının ve ek yapısının sohbet robotu modelleri üzerindeki etkisinin incelenmesi ve yöntemlerin bu zorluklara ne derece uyum sağlayabildiğinin belirlenmesi,
- LangChain çatısı altında kullanılan WS işlemi ile ilgili konuyla ilgili verilerin toplanması,
- RAG yöntemi ile ilgili sonuçlar kullanılarak LangChain aracılığıyla GPT-4 modeline veri gönderilerek, sınıf tahminlemesinin gerçekleştirilmesi,
- Hem ön işleme aşamasından geçmiş hem de ön işleme aşamasından geçmemiş veri setleri üzerinde ele alınan, tüm geleneksel ve modern

yöntemlerle testler yapılarak, elde edilen sonuçların karşılaştırılması ve performanslarının analiz edilmesi,

- Sonuçları ortaya koyan kapsamlı başarı tablolarının oluşturulmasıdır.

Çalışma, Türkçe sohbet robotu geliştirme süreçlerinde en uygun modellerin belirlenmesi ve bu modellerin performanslarının objektif bir şekilde değerlendirilmesi açısından önemlidir.

1.3 Hipotez

Bu tez çalışması kapsamında bir teknoloji firması tarafından geliştirilmiş ve hali hazırda kullanılan bir canlı sohbet robotu için, hazırlanmış veri seti üzerinde geleneksel ve modern yöntemlerin karşılaştırılması amaçlanmıştır. Çalışmada, öncelikle sohbet robotu uygulamalarında yaygın olarak kullanılan geleneksel ve modern yöntemler ele alınmıştır. Veri seti ön işlemeli ve ön işlemeli olmayan şeklinde iki versiyon olarak ele alınmıştır. Her iki versiyon da derin öğrenme ve makine öğrenmesi modellerine uygulanmıştır. Veri seti, canlı bir sohbet robotuna ait olduğundan, kullanıcı geri bildirimleri ve etkileşimleri içermektedir. İlk aşamada, veri setinin ön işleme tabi tutulmamış hali, ardından ön işleme tutulmuş hali modellere sokulmuştur. Bu iki veri seti arasındaki performans farklılıkları incelenmiş ve karşılaştırmalar yapılmıştır. Veri setleri üzerinde ön işlemlerin performans üzerindeki etkileri analiz edilmiştir.

Model performans değerlendirmeleri için doğruluk, hassasiyet, geri çağırma ve F1 skoru gibi temel ölçütler kullanılmıştır. Elde edilen sonuçlar ile her iki veri seti için model performanslarını karşılaştırmalı olarak sunulmuş ve hangi yöntemlerin daha etkili olduğunu belirlemeye yönelik önemli bulgular elde edilmiştir. Böylece, bu çalışmanın sohbet robotu geliştirme süreçlerinde hangi yöntemlerin ve veri işleme tekniklerinin daha başarılı olduğunu ortaya konularak, gelecekteki uygulamalar için yön gösterici bir kaynak oluşturması hedeflenmiştir.

2

VERİ VE VERİ SETİ

2.1 Veri

Veri, ham ögeler veya işlenmemiş gerçekler olarak tanımlanır ve sayılar, semboller, metinler ve görseller gibi çeşitli biçimlerde olabilir. Başka bir ifadeyle, veri toplandığı veya gözlemlendiği anda henüz yorumlanmamış ve anlamlandırılmamış bir bilgi parçası olarak kalır; bu nedenle yalnızca organize edilmemiş ve basit unsurlar bütünü olarak kabul edilir [33]. Veriler; ölçümler, gözlemler veya görüşler biçiminde toplanır ve bu ham bilgilerin analiz edilmesi, verilerden anlam çıkarılmasına, bunun sonucunda da kararlar alınmasına veya sonuçlar çıkarılmasına olanak sağlar [34]. İşlenmeden önce “ham veri” olarak bilinen bu bilgi, araştırmacılar tarafından ayıklanıp doğruluğu sağlanmadıkça sadece bir sayı veya karakter dizisi olarak kalır. Ham verinin işlenmesi sırasında hatalar, cihaz sapmaları veya veri giriş hataları giderilir; bu şekilde bir aşamanın “işlenmiş verisi” bir sonraki aşamanın “ham verisi” olarak kabul edilebilir [35].

Verinin anlamlandırılması süreci, veri bilimi ve analitik yöntemlerle sağlanır. Veri bilimi; matematiksel bilgiler, programlama becerileri, bilimsel yöntemler, algoritmalar ve sistemleri bir araya getirerek hem yapılandırılmış hem de yapılandırılmamış verilerden işlevsel ve anlamlı bilgi elde edilmesini sağlar [36].

2.2 Veri Seti

Veri seti, yapay zeka ve makine öğrenimi alanlarında önemli bir temel yapı taşıdır ve bilgisayar tarafından tek bir birim olarak ele alınan veri topluluğu olarak tanımlanmaktadır [37]. Veri setleri, birbirinden bağımsız çok sayıda veriyi bir araya getirir. Günümüzde veri setleri, farklı endüstrilerde analiz ve karar alma süreçlerini desteklemek amacıyla yaygın olarak kullanılmaktadır. Özellikle, müşterilerin alışkanlıkları, ürün kullanım oranı ve popülerliği gibi iş verilerini kapsayan veri setleri, iş dünyasında büyük bir değer yaratır hale gelmiştir. Veri setleri, gelişmiş modelleme çalışmaları ve analiz süreçleri için yapılandırılmış ham veri içerir ve büyük veri miktarlarıyla örüntülerin tespit edilmesine olanak sağlar. Veri setlerinin doğru bir şekilde yapılandırılması için veriler kullanım amacına uygun hale getirilmelidir.

2.3 Veri Kalitesi

Veri kalitesi, veri madenciliğinde anahtar bir konudur ve bir veri kümesinin amacına uygun şekilde doğru ve güvenilir bilgi sağlama yeteneğini ifade eder [38]. Bir verinin kalitesi; doğruluk, tutarlılık, geçerlilik, tamamlık, güncellik ve benzersizlik gibi belirli kriterlerle değerlendirilir. Düşük kaliteli veriler, işletmelerde operasyonel aksaklıklar, hatalı analizler ve yanlış iş stratejileri gibi kritik sorunlara yol açabilir.

- **Doğruluk:** Verinin temsil etmesi gereken varlıkları veya olayları doğru şekilde yansıttığını ve güvenilir, doğrulanabilir kaynaklardan alındığını gösteren bir unsurdur [39].
- **Tutarlılık:** Sistemler ve veri setleri arasında uyumlu olduğunu, aynı verilere ait değerlerin farklı sistemlerde veya veri setlerinde çelişmediğini gösteren bir kavramdır. Örneğin, bir yaş kolonunda "250" gibi gerçekçi olmayan değerlerin yer almaması gerekir [39, 40].
- **Geçerlilik:** Verinin belirli iş kurallarına ve parametrelere uygun olduğu, yapılandırmalarının doğru bir şekilde yapıldığı değerleri barındırmalıdır. Örneğin, "doğum tarihi" alanı için sadece geçerli bir tarih formatında

(örneğin, 01/07/1995) bilgi girilmesi gereklidir. "krc1234" gibi bir giriş, geçersizdir ve sistemde hata oluşturur [39, 40].

- **Tamamlık:** Veri olması gereken tüm bilgileri ve veri türlerini içermelidir. Eğer gerekiyorsa meta veriler de veri setine dahil olmalıdır. Örneğin; müşteri bilgilerinin kayıt işleminde ad, soyad, telefon gibi alanlar isteniyor ise bu alanların eksiksiz bir şekilde doldurulması gerekir. [39, 40].
- **Güncellik:** Veri ihtiyaçlara uygun şekilde güncel olmalıdır [39]. Örneğin; Bir işletmeye şirket dokümantasyonlarından oluşan bir sohbet robotu tasarlanıyorsa bu robotun sorulara doğru ve güncel bir şekilde cevap vermesi gerekir. Veri güncel tutulmadığı takdirde sohbet robotunun verdiği sorular da güncel olmayacaktır. Bu da güven ve itibar kaybına sebebiyet verebilir.
- **Benzersizlik:** Veri seti içinde tekrar eden kayıtlar olmamalıdır. Her bir kayıt benzersiz şekilde tanımlanması gerekir [39]. Bir e-ticaret platformunda aynı müşteri, iki kez kayıt edilmişse bu da veri kalitesini olumsuz yönde etkiler ve karışıklığa yol açar. Her müşterinin benzersiz bir müşteri numarasıyla tanımlanması gerekir.

Veri setinin kalitesini düşüren bazı olumsuz durumlarda söz konusudur. Bunlar tutarsız veriler ve kasıtlı problemlerdir.

Tutarsız veriler genellikle veri tabanlarında veya veri ambarlarında kaydedilen bilgilerdeki farklılıklardan ve uyumsuzluklardan kaynaklanır. Bu durumun nedenleri arasında bazı verilerin silinmesi, zaman içerisinde kayıt ve düzenleme süreçlerindeki farklılıklar yer alabilir [41]. Örneğin, bir satış veri tabanında bir müşterinin adres bilgisi bulunurken, başka bir tabloda aynı müşterinin adres bilgilerinin eksik veya farklı formatlarda oluşu tutarsızlığa yol açabilir. Bu tür tutarsızlıklar beraberinde doğru analiz ve raporlama süreçlerini olumsuz etkileyebilir. Bu nedenle, tutarsız veriler doğru bir şekilde tespit edilmeli ve uyumluluk sağlanarak düzeltilmelidir [42].

Kasıtlı problemler, verilerin bilinçli bir şekilde eksik, yanlış veya yanıltıcı şekilde kaydedilmesinden kaynaklanır. Verilerin eksik girilmesi, bir kısmının önemsiz bulunarak kayıt edilmemesi veya yanlış anlaşılma sonucunda verilerin yanlış girilmesi gibi durumlar, eksik ve tutarsız verilere neden olabilir [41]. Bu

problemler, özellikle veri toplama süreçlerinde insan veya sistem hataları ile birleşerek verilerin kalitesini daha da düşürebilir. Kasıtlı ya da bilinçli şekilde oluşturulan bu sorunlar, doğru teşhis edilip çözülmediği takdirde veri analizi ve karar alma süreçlerini ciddi şekilde etkileyebilir [42].

2.4 Veri Kategorilendirme

Veri kategorilendirme, verilerin duyarlılığına ve ihlal edilmesi durumunda yaratacağı potansiyel etkiye dayanarak kategorilere ayrılması sürecidir. Siber güvenlikte, özellikle işletmeler için veri kategorilendirme, korunması gereken hassas bilgilerin belirlenmesine ve önceliklendirilmesine yardımcı olur. Verilerin kategorize edilmesi, şirketlerin kritik bilgileri korumak için uygun güvenlik önlemleri uygulamalarını sağlar. Bu süreç, sadece güvenliği sağlamaktan daha fazlasını yapar; aynı zamanda işletmelerin verileri daha verimli bir şekilde yönetmelerine yardımcı olur [43].

Veri kategorilendirme, verileri duyarlılığına göre düzenlemenin yanı sıra, yetkisiz şekilde erişilmesi, değiştirilmesi veya yok edilmesi durumunda doğacak potansiyel etkileri de göz önünde bulundurur. Bu süreç, hassas bilgilerin korunmasını sağlamanın yanı sıra, organizasyonların uygun güvenlik önlemleri uygulamasını da temin eder.

Örneğin, IBM'in raporlarına göre, etkili veri kategorilendirme politikaları uygulayan işletmelerin veri ihlali maliyetlerini %30'a kadar azalttığı belirtilmiştir. Bu da, işletmelerin güvenlik açıklarını minimize etmelerini ve yasal riskleri azaltmalarını sağlar. Aynı zamanda verilerin yönetilmesi de daha verimli hale gelmektedir. Yapılan bir araştırmaya göre, verilerin doğru bir şekilde yönetilmemesi nedeniyle, 2023'te bir veri ihlalinin ortalama maliyeti 4,45 milyon dolar olarak kaydedilmiştir ve bu maliyetin gelecek senelerde daha da artması beklenmektedir [44].

2.5 Veri Ön İşleme Süreçleri

Veri ön işleme, veri madenciliği süreçlerinin güvenilirliği ve verimliliği için kritik bir adımdır. Hatalı veya eksik verilerin yanlış çıktılara yol açmasını engellemek amacıyla uygulanır. Bu süreçteki işlemler genellikle yarı otomatik olup büyük veri

kümelerinde etkin tekniklerin önemini artırır. Veri ön işlemenin temel amaçları arasında veri problemlerini çözmek, anlamlı analizler yapmak için daha değerli bilgiler elde etmek yer alır [38].

2.5.1. Veri Temizleme

Eksik, gürültülü ve tutarsız verilerin düzenlenmesini içerir. Eksik veriler için kayıtların atılması, ortalama değerlerin kullanılması veya regresyon gibi yöntemler uygulanmaktadır. Gürültülü veriler; binning, kümeleme veya regresyon yöntemleri ile düzeltilmektedir. Tutarsızlıklar ise dışsal referanslarla giderilir [38].

2.5.2. Veri Birleştirme

Farklı veri tabanlarının veri ambarlarında birleştirilmesini ifade eder. Şema birleştirme hatalarını azaltmak için meta veri kullanılır, fazla değişkenler ise korelasyon analiziyle tespit edilip veri kümesinden çıkarılabilir [38].

2.5.3. Veri İndirgeme

Büyük veri kümelerinin hacmini azaltmayı amaçlar. Boyut indirgeme, veri sıkıştırma veya kesikli hale getirme gibi yöntemlerle uygulanabilir.

2.5.4. Veri Dönüştürme

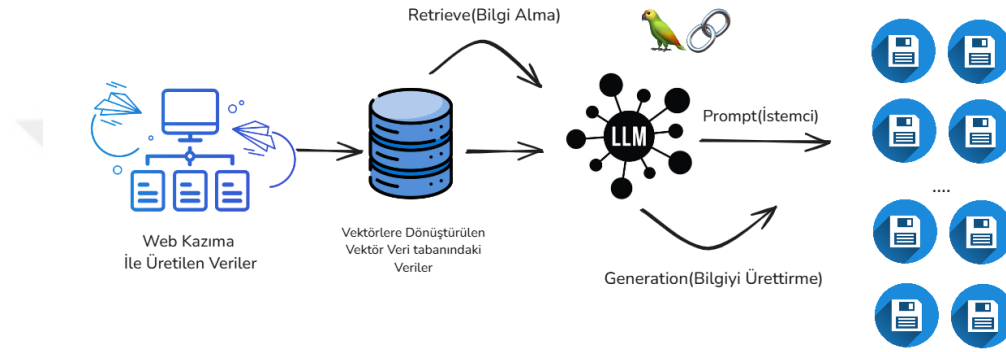
Verilerin, veri madenciliği için uygun formlara dönüştürülmesi sürecidir. Normalleştirme gibi işlemleri içerir. Örneğin, min-maks, z-skor ve ondalık ölçekleme teknikleri uygulanabilir [38].

2.6 Verilerin Maskelenmesi

Veri madenciliği, büyük veri tabanlarından yeni ve değerli bilgiler elde etmeyi amaçlayan bir alandır. Ancak bu süreç, gizlilikle ilgili sorunları da beraberinde getirebilir. Elde edilen bilgiler, niyet edilen kullanımdan bağımsız olarak hassas verileri ortaya çıkarabilir ve gizliliği tehlikeye atabilir. Bu nedenle paylaşılmadan önce veri setlerinde yer alan hassas bilgilerin değiştirilmesi, yani maskelenmesi önemlidir. Hassas verilerin korunması için farklı maskeleyme teknikleri kullanılmaktadır. Bu tür yöntemler, veri gizliliğini sağlamak ve bilgilerin kötüye kullanımını önlemek için kritik bir öneme sahiptir [45].

2.7 Veri Ürettirme İşlemleri

Modern ve geleneksel NLP yöntemlerinin performansını değerlendirmek ve karşılaştırmak amacıyla uygun bir veri setinin hazırlanması, bu yöntemlerin etkinliğini ölçebilmek için kritik bir adımdır. Bu bağlamda, veri toplama süreçlerinde WS yöntemleri, büyük miktarda veri elde etmek için etkili bir yöntem olarak öne çıkmaktadır. WS ile elde edilen veri setini daha da genişletmek amacıyla RAG ve LLM gibi yöntemler kullanılabilir.



Şekil 2.1 Veri Ürettirme İşlemleri

İlk aşamada, belirli bir miktarda veri WS yöntemiyle toplanır. Bu işlem gerçekleştirilirken web sitelerinden, forumlardan veya sosyal medya platformlarından faydalanılabilir. Toplanan bu veriler metin formatındadır. WS yöntemiyle toplanan bu veriler, metin formatında olduğu için LLM ile kullanılması için öncelikle sayısal verilere yani vektörlere dönüştürülmelidir. Vektörlere dönüşüm gerçekleştikten sonra elde edilen sayısal veriler bir vektör veri tabanında toplanmalıdır. Facebook Yapay Zeka Benzerlik Araması (Facebook AI Similarity Search, Faiss), Chroma gibi vektör veri tabanları verilerinin etkin bir şekilde saklanmasını sağlar. Genellikle bir sorunun cevabı arandığında yanıt ürettirebilmek için LLMlerden faydalanılabilir. Ancak, LLM'in daha önce eğitildiği veri kümesinde bulunan bilgilerin dışında bazı sohbet robotları spesifik bilgilere cevap veremeyebilir. Bu durumda RAG devreye girerek LLM'in eksik bilgisini tamamlar ve sadece eğitim verisine dayalı bilgileri değil, dışarıdaki veri kaynaklarından aldığı ek bilgilerin de kullanılmasını sağlar. RAG ile LLM'e öğretilen bu bilgilere istinaden ilgili bir istemci (prompt) ile ilgili konuda LLM yardımı ile daha fazla veri üretilmesi sağlanabilir.

3.1. FiT

FiT, soru yanıtlama [46], uzak denetimli duygu analizi [47] ve makine öğrenmesi [48] gibi alanlarda, benzer görevler arasında bilgi transferi sağlamak için yaygın olarak kullanılan bir tekniktir. Ancak, bu teknik, ilişkili olmayan görevler arasında kullanıldığında etkinliği düşebilmektedir [49]. Dil modelleriyle ilgili olarak farklı projelerde FiT çalışmaları denenmiştir ve iyi bir performans için milyonlarca etki alanı verisi gerekliliği vurgulanmıştır [50].

Evrensel Dil Modeli İnce Ayarı (Universal Language Model Fine-tuning, ULMFiT), genel alanın önceden eğitimi ve yeni ince ayar tekniklerini kullanarak aşırı öğrenmeyi engelleyebilmektedir. Bu yaklaşım, yalnızca 100 etiketli örnekle bile aşırı öğrenmeyi önleyebilmekte ve küçük veri setlerinde bile son teknoloji düzeyinde performans elde edebilmektedir. Bu nedenle, sınırlı etiketli verilerle yapılan görevlerde oldukça faydalıdır [51].

3.2 Dönem Sayısı (Epoch)

FiT işlemi sırasında, modelin öğrenme sürecini optimize etmek için dönem sayısı adı verilen kavram kritik bir rol oynar. Dönem sayısı, modelin eğitim veri seti üzerinde kaç kez tamamen eğitim aldığına işaret eder. Yetersiz dönem sayıları modelin öğrenmesini sınırlayabilir iken, çok fazla dönem sayısı aşırı öğrenme/uyum (Overfitting, Oft) neden olabilir. Bu bağlamda, uygun dönem sayısının belirlenmesi, özellikle sınırlı veriyle çalışıldığında performansın artırılmasında önemlidir.

3.3 Oft

Oft, makine öğreniminde modelin eğitim verisindeki ayrıntılara fazlasıyla odaklanıp yeni ve görülmemiş verilere genelleme yapamaması durumudur [52]. Bu, modelin eğitim setinde harika sonuçlar verirken, test setinde başarısız olmasıyla kendini gösterir. Bunun sebebi, modelin verilere dair genel desenleri öğrenmek yerine, eğitim setindeki her detayı yanlış veya rastgele ezberlemesidir.

Oft nin birkaç temel sebebi vardır. Bunlar;

- Eğitim verisinin az ya da temsil etme seviyesinin düşük olduğu durumlarda modelin gürültüyü öğrenmesidir [53],
- Modelin fazla karmaşık olmasıdır ki bu, modele birçok değişken eklenmesiyle doğruluğun artmasına rağmen tutarlılığın azalmasına yol açabilmektedir [53],
- Özellikle yapay zeka algoritmalarında görülen, farklı seçeneklerin karşılaştırılması sırasında yanlış seçimlerin yapılmasıdır. Bu durum bazen modelin doğruluğunu düşürebilmektedir [54].

Oft'yi önlemek için çeşitli yöntemler geliştirilmiştir ancak bunlar her zaman gerçek hayattaki sorunlar karşısında etkili olamayabilir.

3.4 Eğitim ve Test (Train-Test)

Makine öğrenimi projelerinde eğitim ve test veri setlerinin doğru bir şekilde ayrılma işlemi model performansının değerlendirilmesinde temel bir adımdır. Eğitim verisi, modelin veri içindeki örüntüleri öğrenmesi için kullanılırken, test verisi modelin bu öğrendiklerini yeni ve görülmemiş verilere nasıl genelleyebildiğini ölçmektedir [55]. Test veri seti ile amaçlanan elde edilen modelin tahmin edebilme yeteneğini ölçmektir. Bu süreç de modelin Oft veya Oft'nin tam tersi yetersiz öğrenme (underfitting, Uft) gibi istenmeyen durumlarını önlemek için kritik bir rol oynamaktadır.

Veri setinin eğitim ve test olarak bölünmesi genellikle %70-80 eğitim veri seti ve %20-30 test veri seti oranında yapılır [56]. Eğitim sürecinde model, veri setindeki örüntüleri keşfeder ve bu bilgilere dayanarak tahmin fonksiyonları oluşturur. Test sürecinde ise modelin bu tahmin fonksiyonları test verisi üzerindeki doğruluğu ölçülür. Bu yaklaşım ise modelin gerçek dünya problemlerine genelleyebilme

kapasitesini anlamak için önem arz eder. Örneğin, bir mağazadaki müşteri alışkanlıkları üzerine veri toplayarak müşterilerin gelecekteki alışveriş davranışlarını tahmin eden bir model oluşturulabilir [57]. Ya da bir e-ticaret platformunda kullanıcıların ürün inceleme ve puanlama alışkanlıkları analiz edilerek hangi ürünlerin daha fazla ilgi göreceğini tahmin eden bir model tasarlanabilir. Bu örnekler hem gerçek dünya uygulamalarının başarımlarını artırabilir hem de kullanıcıların isteklerini karşılama noktasında daha faydalı ürünler çıkmasını sağlayabilir.

Doğru bir eğitim ve test ayrımı yapılmadığında model performans ölçümleri yanıltıcı olabilir. Eğitim verileri ile yüksek doğruluk sağlayan bir model, test verileri üzerinde düşük performans sergileyebilir. Bu tür durumlar, genellikle modelin eğitim verilerine fazla bağımlı hale gelmesinden kaynaklanır. Bu nedenle modelin performansını değerlendirmek için k-katlı çapraz doğrulama (k-fold cross-validation) ve katmanlı k-katlı çapraz doğrulama gibi yöntemler de sıklıkla kullanılan yöntemler arasındadır [55].

3.5. IC

IC, kullanıcı sorgularının önceden tanımlanmış niyetlere eşlenerek uygun yanıtlar veya eylemler sağlanmasını hedefleyen bir süreçtir. Özellikle sohbet robotlarında kullanıcı niyetlerinin doğru anlaşılması, sistemin etkinliğini artırmada kritik bir rol oynamaktadır [58].

IC, Tek Etiketli Niyet Sınıflandırması (Single Label Intent Classification, SLIC) ve Çok Etiketli Niyet Sınıflandırma (Multi Label Intent Classification, MLIC) olarak ikiye ayrılabilir. SLIC, her örneğin yalnızca bir sınıfa ait olduğu durumlarda kullanılırken, MLIC, bir örneğin birden fazla sınıfa ait olabileceği durumları ele alır. MLIC birçok etiketin örtüşmesi nedeniyle daha karmaşık sınırlara sahip olduğu için bu durum sınıflandırma performansını zorlaştırmaktadır. Ayrıca, çoklu etiket sınıflandırma görevlerinin nasıl gerçekleştirilmesi gerektiğine dair bilgi eksikliği bu zorlukları artırmaktadır [59].

Geleneksel IC yaklaşımlarında genellikle SLIC problemleri kullanılmaktadır. Bu durum özellikle birden fazla niyetin bir arada bulunabileceği karmaşık eğitim senaryolarında yetersiz kalabilmektedir. Yanıtların kullanıcı niyetlerini

karşılammaması, sohbet botlarının yanlış niyet etiketlerine dayalı olarak yanıt verememesine yol açabilmektedir. MLIC verilerindeki sorun ise örneklerin birden fazla sınıfa ait olabilmesi ve bu durumun etiket sınırlarını bulanıklaştırmasıdır. Bu sınıflandırma türüne yönelik bilgi eksikliği, bu zorlukları daha da artırmaktadır [58].

Günümüzde derin öğrenme teknolojilerindeki ilerlemeler ve konuşma yapay zekâsına olan yüksek talep, görev odaklı diyalog sistemlerinde kullanıcı ihtiyaçlarını anlamayı amaçlayan yaklaşımları hızla geliştirmiştir. IC ve Alan Doldurma (Slot Filling, SF) bu bağlamda temel görevler olarak öne çıkmaktadır. Örneğin, bir kullanıcının "The Beatles'tan Hey Jude şarkısını dinlemek istiyorum" ifadesinde, sistemin kullanıcının bir şarkıyı çalma komutunu verdiğini anlaması ve "Hey Jude" ile "The Beatles" bilgilerini doğru şekilde eşlemesi gereklidir. Bu bilgiler genellikle bir semantik çerçeve yapısı içinde temsil edilir ve bu yapının çıkarımı IC ve SF görevlerini içerir [60].

IC problemlerinde bağımsız modeller, ortak modeller ve transfer öğrenme modelleri gibi farklı sinir ağı mimarileri kullanılmaktadır. Özellikle transformatör tabanlı modeller (BERT, Electra vb.), SF ve IC görevlerinde dikkat temelli mekanizmaları ve bağlamsal kelime gömme yöntemlerini kullanarak performansı artırmaktadır. Bu gelişmeler, görev odaklı diyalog sistemlerinde bağlam modelleme, uzun mesafeli kelime bağımlılıklarının ele alınması ve SF ile IC'nin sinerjisini daha etkili kullanmayı mümkün kılmaktadır [60].

Makine öğrenmesi modellerinin başarısını değerlendirmenin amacı, geliştirilen modelin ne kadar etkili çalıştığını anlamak ve gerçek dünya verileri üzerinde nasıl performans gösterdiğini belirlemektir. Performans metrikleri, modelin doğruluğunu, genelleme yeteneğini ve veri uyumunu değerlendirmemize olanak tanır. Bu metrikler, algoritmaların ve yapılandırmaların ne kadar etkili olduğunu anlamamıza yardımcı olur ve doğru seçimler yapabilmemiz için hayati öneme sahiptir. Performans metriklerinin doğru seçilmesi, makine öğrenmesi projelerinde modelin ne kadar başarılı olduğunu doğru bir şekilde değerlendirmemize olanak tanır. Doğruluk (accuracy), hassasiyet (precision), geri çağırma (recall), F1-Skoru (F1 skoru), Alıcı İşletim Karakteristik Eğrisi (Receiver Operating Characteristic Curve, ROC eğrisi), Eğri Altındaki Alan (Area Under the Curve, AUC), Ortalama Hata Karekökü (Root Mean Square Error, RMSE) ve Ortalama Mutlak Hata (Mean Absolute Error, MAE) gibi metrikler, modelin çeşitli yönlerinin değerlendirilmesine yardımcı olur. Bu metriklerin uygun şekilde seçilmesi, projelerin başarısını ve modelin gerçek dünya verileri üzerindeki etkisini doğru bir şekilde ölçmeyi sağlar [61], [62], [63].

4.1. Doğruluk (Accuracy)

Doğruluk, en yaygın kullanılan performans ölçütlerinden biridir ve modelin doğru tahmin ettiği örneklerin oranını gösterir. Bu metrik sınıflandırma problemlerinde modelin genel performansını hızlıca değerlendirmek için kullanılır, ancak dengesiz veri setlerinde yanıltıcı olabilir. Özellikle bazı sınıfların çok fazla olduğu veri setlerinde, yüksek doğruluk oranı bile modelin başarısını yansıtmayabilir. [61].

Hata matrisi üzerinden hesaplanan doğruluk, aşağıdaki formülle ifade edilir:

$$\text{Doğruluk(Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1)$$

Burada;

- **Gerçek Pozitif (True Positive, TP):** Gerçek değeri pozitif olup model tarafından pozitif tahmin edilen örnekler,
- **Gerçek Negatif (True Negative, TN):** Gerçek değeri negatif olup model tarafından negatif tahmin edilen örnekler,
- **Yanlış Pozitif (False Positive, FP):** Gerçek değeri negatif olup model tarafından pozitif tahmin edilen örnekler,
- **Yanlış Negatif (False Negative, FN):** Gerçek değeri pozitif olup model tarafından negatif tahmin edilen örneklerdir [61].

4.2. Kesinlik (Precision)

Kesinlik, modelin pozitif olarak tahmin ettiği örneklerin gerçekte ne kadarının doğru olduğunu ölçmektedir. Kesinlik, özellikle yanlış pozitiflerin yüksek maliyetlere yol açtığı durumlarda önemli bir rol oynar. Örneğin, tıbbi teşhislerde yanlış pozitif sonuçlar ciddi sonuçlara yol açabilir. Bu yüzden, kesinlik metriği modelin güvenilirliğini değerlendirmede kritik bir rol oynamaktadır [62]. Bu hesaplama yöntemi, yanlış pozitif tahminlerin sayısını azaltmaya odaklanır.

Precision aşağıdaki formüldeki gibi hesaplanır:

$$\text{Kesinlik(Precision)} = \frac{TP}{TP + FP} \quad (4.2)$$

4.3. Duyarlılık (Recall)

Duyarlılık, modelin gerçek pozitifleri ne kadar doğru bir şekilde tespit ettiğini ölçer. Bu metrik, yanlış negatif tahminlerin sayısını azaltmaya odaklanır. Duyarlılık, özellikle yanlış negatif tahminlerin önemli sonuçlar doğurabileceği durumlarda; örneğin güvenlik uygulamaları veya hastalık teşhislerinde, büyük önem taşır. Modelin duyarlılığını artırmak, gerçek pozitifleri kaçırma riskini en aza indirir [62],[63],[64].

Duyarlılık aşağıdaki formüldeki gibi hesaplanır:

$$\text{Duyarlılık(Recall)} = \frac{TP}{TP+FN} \quad (4.3)$$

4.6.4 F1-Skoru

F1-Skoru daha önceki paragraflarda anlatılan kesinlik ve duyarlılık metriklerinin birleştirilmiş bir ölçüsüdür. Bu hem yanlış pozitiflerin hem de yanlış negatiflerin etkisini dengeler ve modelin genel performansını daha iyi bir şekilde değerlendirmeye olanak tanır. Özellikle dengesiz sınıflama problemlerinde ve farklı maliyetlerin olduğu durumlarda kullanışlıdır. Bu metrik, modelin doğruluğunu ve gerçek pozitifleri kaçırma riskini dikkate alarak daha sağlıklı bir performans değerlendirmesi sağlar [61], [63].

F1-Skoru aşağıdaki formüldeki gibi hesaplanır:

$$\text{F1 Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (4.4)$$

4.5. ROC eğrisi ve AUC

ROC eğrisi, sınıflandırma modelinin performansını görsel olarak değerlendirilmesine olanak tanır. Bu eğri, farklı kesme noktalarında duyarlılık ile yanlış pozitif oranı (False Positive Rate, FPR) arasındaki ilişkiyi gösterir. ROC eğrisinin altındaki alan ise AUC ile ölçülür. AUC değeri, modelin genel performansını tek bir sayı ile özetler ve şu şekilde yorumlanır:

- AUC = 1: Mükemmel model,
- AUC = 0.5: Rastgele tahmin.

AUC, modelin çeşitli duyarlılık ve özgüllük düzeylerinde nasıl performans gösterdiğini özetleyen önemli bir metriktir [62], [63].

4.6. RMSE ve MAE

RMSE, regresyon modellerinin tahmin performansını değerlendirmek için kullanılır. Bu metrik, modelin tahminleri ile gerçek değerler arasındaki farkların karelerinin ortalamasının karekökünü alır. MAE, RMSE'ye benzer şekilde, tahmin hatalarının mutlak değerlerinin ortalamasını hesaplar. Her iki metrik de tahmin hatalarını ölçmekte kullanılır ancak MAE, özellikle aykırı değerlerin etkisini daha az hissettirir [58].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}^i - y_i)^2} \quad (4.5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}^i - y_i| \quad (4.6)$$

Burada; \hat{y}^i = Tahmin Edilen Değer, y_i = Gerçek Değer ve n = Veri Sayısı'nı temsil etmektedir.

5

GELENEKSEL YÖNTEMLER

NLP alanında, kelime temsillerini öğrenme ve anlamını belirleme konusunda kullanılan çeşitli geleneksel yöntemler bulunmaktadır. Bu yöntemler, genellikle temel istatistiksel ve matematiksel yaklaşımlara dayanır ve kelimelerin anlamsal ilişkilerini vektörel uzayda temsil etmeye odaklanır.

Çalışmanın bu bölümünde, Word2vec, GloVe ve FastText yöntemlerinin temel özellikleri ve avantajları ele alınmıştır.

5.1. Word2vec

Word2Vec, metin verilerini matematiksel temsillere dönüştürerek NLP uygulamalarında kullanılan geleneksel bir yöntemdir.

Word2Vec'in başarısı, kullanılabilirlik ve performans avantajlarından kaynaklanır. Modelin basitliği, araştırmacıların ve uygulayıcıların kolayca kodu indirip kullanabilmelerine olanak tanır.

Word2Vec, kelimeleri, sürekli vektör uzaylarında temsil ederek semantik ilişkileri geometrik olarak ifade eder. Örneğin, "Kral" - "Adam" + "Kadın" işlemi, "Kraliçe" kelimesine en yakın vektörü döndürebilir. Bu, dağıtım hipotezine dayanan bir yaklaşımla kelimelerin bağlamsal bilgilerinden semantik bağlantıların çıkarılabileceğini göstermektedir. Bu yöntem, çok boyutlu vektör uzayları oluşturarak kelimeler arasındaki benzerlikleri farklı düzeylerde yakalamayı mümkün kılar [65].

Bu model, sıklıkla kullanılan iki temel öğrenme yaklaşımını benimser: Bu yöntemler CBOW VE SGM yöntemleridir. CBOW modeli, çevresel kelimelere dayanarak hedef kelimeyi tahmin etmeye çalışırken, SGM hedef kelimedenden yola çıkarak çevresel kelimeleri tahmin eder. SGM, literatürde en çok kullanılan yöntemlerden biridir. Tekil Değer Ayrıştırması (Singular Value Decomposition, SVD) ve Gizli Anlamsal Analiz (Latent Semantic Analysis, LSA) gibi klasik

tekniklerle literatürde modern yöntemlere göre analogiler üzerinde daha iyi performans gösterdiği örnekler bulunmaktadır. [64].

Word2Vec'in öğrenme süreçlerinde bazı sınırlamalar bulunmaktadır. Örneğin, modelin kelimeler arasındaki dil bilgisel bağlamı öğrenemediği, Oft problemi yaşamadığı, ancak bazı hiper parametre seçimlerinin sonuçlar üzerindeki etkisinin yeterince analiz edilmediği bilinmektedir. Özellikle, modelin öğrenme oranı ve dönem sayısına bağlı davranışı gibi faktörler, detaylı analiz gerektirmektedir. Bu eksiklikler, modelin sınırlamalarını anlamak ve performansını artırmak için yeni yöntemlerin geliştirilmesini teşvik etmektedir [65].

Bu model, büyük veri kümeleri üzerinden semantik vektör alanları üreterek farklı NLP senaryolarında etkin bir başlangıç noktası olarak kullanılabilir. Bununla birlikte, modelin performansını artırmak için özel eğitim yöntemleri geliştirilmiştir. Örneğin, analogiler üzerine ön eğitim(pre-train) edilmiş bir Word2Vec modeline küçük iyileştirmeler yapılması, çok yüksek doğruluk oranlarına ulaşılmasını mümkün kılmıştır [64].

5.2. GloVe

GloVe, kelimeler için vektör temsilleri elde etmek amacıyla kullanılan denetimsiz bir öğrenme algoritmasıdır. Bu algoritma bir metin kümesindeki kelimeler arasındaki küresel kelime-kelime eşzamanlılık istatistikleri üzerinde eğitilir ve elde edilen temsiller kelime vektör alanındaki ilginç doğrusal alt yapıları sergiler. Pennington ve arkadaşlarının [66], 2014 yılında geliştirdiği GloVe modeli, kelime temsili öğrenmede kullanılan bilinen geleneksel yöntemlerden biridir. GloVe'nin başarısı, küresel kelime ilişkilerini öğrenme biçimine dayalıdır. Bu modelde, kelimelerin eş zamanlılıkları, kelimeler arasındaki anlamlı ilişkilerin belirlenmesinde önemli bir yer tutmaktadır.

GloVe modelinin amacı, X matrisinin logaritmasını ve kelime vektörlerinin iç içe geçmiş ilişkilerini optimize ederek anlamlı bir vektör temsili elde etmektir. [67]

X matrisi, metin bütünündeki kelimeler arasındaki eşzamanlılıkları temsil eden bir matris, her x_{ij} , kelime i 'nin kelime j ile birlikte görünme sıklığını ifade etsin. W_i ve W_j , kelime i ve kelime j için öğrenilen vektörler, b_i ve b_j , her bir kelime için öğrenilen bias terimleri olmak üzere; kayıp fonksiyonu şu şekilde hesaplanır:

$$J = \sum (W_i^T W_j + b_i + b_j - \log x_{ij}) \quad (5.1)$$

Bu temsiller kelimeler arasındaki semantik ilişkileri yansıtır ve kelime vektörleri üzerinde çeşitli doğrusal ilişkiler oluşturur [66].

GloVe'nin paralel olarak eğitilebilmesi, büyük veri kümesi üzerinde daha hızlı eğitim yapılmasına olanak tanırken, kelimeler arasındaki semantik ilişkilerin doğru bir şekilde öğrenilmesini sağlar [67].

5.3. FastText

FastText, metin sınıflandırma ve kelime temsilde kullanılan, hızlı ve verimli bir derin öğrenme tabanlı modeli olarak geçmektedir. Facebook AI tarafından geliştirilen FastText, metin sınıflandırma görevlerinde derin öğrenme modellerine kıyasla benzer doğruluk oranlarına ulaşırken, hem eğitim hem de değerlendirme süreçlerinde çok daha yüksek hız sunar. Model, çok büyük veri setlerini standart bir çok çekirdekli Merkezî İşlem Birimi (Central Process Unit, CPU) kullanarak kısa sürede işleyebilir. Örneğin, bir milyardan fazla kelimeyi on dakikadan kısa sürede eğitebilir ve yüz binlerce sınıfa ayrılmış veriyi dakikalar içinde sınıflandırabilir [68]. Ayrıca, FastText, farklı dillerdeki metinlerin semantik anlamlarını temsil etme yeteneğiyle öne çıkar. Özellikle az kaynaklı dillerde, kelime seviyesindeki temsil yöntemleri yerine alt kelime bilgilerini kullanan yapısı sayesinde daha yüksek performans gösterebilir. TF-IDF veya BERT gibi yöntemlerin daha yüksek doğruluk sağladığı durumlarda bile, FastText'in eğitimin ve tahminin hızını artırma potansiyeli, onu hala tercih edilen bir seçenek haline getirebilmektedir [69].

FastText algoritması, metin sınıflandırma problemlerinde geleneksel makine öğrenme algoritmalarının yaşadığı boyutsallık sorunu ve düşük performans gibi problemlere çözüm sunabilmektedir. Bazı çalışmalarda, FastText tabanlı bir modelin, geleneksel algoritmalara kıyasla daha düşük boyutlu ve yüksek kaliteli metin temsilleri elde ettiği, dahası duygusal kutupluluk değerlendirme görevinde elde ettiği doğruluk, geri çağırma ve F1 değerlerinin, geleneksel makine öğrenimi yöntemlerini aşarak üstün bir sınıflandırma performansına sahip olduğu gösterilmiştir [70].

6.1 BERT ve BERTurk

BERT derin öğrenme tabanlı, iki yönlü dil temsili oluşturan bir dil modelidir. Bu model, son dönemde popüler hale gelen dil temsili modellerinden farklı olarak, her iki yönü de (sol ve sağ bağlam) dikkate alarak etiketlenmemiş metin üzerinde derin, iki yönlü temsiller eğitmek üzere tasarlanmıştır. Bu özellik, BERT'in herhangi bir özel görev için mimariyi büyük ölçüde değiştirmeye gerek duymadan, yalnızca bir ek çıktı katmanı ekleyerek fine-tuning yapılabilmesini sağlar. Bu sayede BERT, soru-cevap ve dil çıkarımı gibi çok çeşitli doğal dil işleme (NLP) görevlerinde son derece yüksek performans sergileyebilir.

Yapılan çalışmalar, BERT'in çok sayıda NLP görevinde yeni birinci sınıf sonuçlar elde ettiğini göstermektedir [71].

Türkçe NLP uygulamaları için de BERT'in çeşitli uyarlanmış modelleri mevcuttur. Bu modellerden biri BERTurk olup, Türkçe dilindeki metinleri işlerken yüksek başarı gösteren bir dil modelidir. BERTurk, Türkçe OSCAR korpusu, Wikipedia verisi, OPUS korpusları ve Kemal Oflazer tarafından sağlanan özel bir korpus kullanılarak eğitilmiştir. Bu model, farklı kelime dağarcığı boyutlarına sahip versiyonlara sahiptir. BERTurk'ün çeşitli versiyonları arasında büyük/küçük harf duyarlı (cased) ve büyük/küçük harf duyarsız (uncased) olmak üzere iki ana versiyon bulunur. Her iki versiyon için de 32k ve 128k boyutlarında kelime dağarcığına sahip modeller bulunmaktadır. Bu modellerin eğitim verisi 35GB boyutunda olup, toplamda 44 milyardan fazla token içermektedir [72].

BERTurk modelleri şu şekilde sunulmaktadır [73]:

- BERTurk (cased, 32k)
- BERTurk (uncased, 32k)
- BERTurk (cased, 128k)
- BERTurk (uncased, 128k)

BERTurk, Türkçe dilindeki metinler üzerinde yapılan birçok doğal dil işleme görevinde kullanıma uygun hale getirilmiştir ve araştırmacılar tarafından sürekli olarak geliştirilmekte olan bir modeldir. TensorFlow Araştırma Bulutu (TensorFlow Research Cloud, TFRC) ile yapılan eğitim süreci, modelin başarıya ulaşmasında önemli bir etken olmuştur. Şu anda yalnızca PyTorch uyumlu model ağırlıkları mevcuttur; ancak TensorFlow uyumlu ağırlıklara da erişim sağlanması için çalışmalar yapılabilir [73].

6.2 ConvBERT ve ConvBERTurk

ConvBERT, BERT modelinin verimliliğini artırmayı hedefleyen bir dil işleme modelidir. Geleneksel BERTin, küresel kendiliğinden dikkat (self-attention) bloklarına dayanması, modelin yüksek bellek tüketimi ve hesaplama maliyeti gibi zorluklar doğurmasına neden olmaktadır. ConvBERT, bazı dikkat başlıklarını span-tabanlı dinamik evrişimle (dynamic convolution) değiştirerek, yerel bağımlılıkları doğrudan modellemeye olanak tanır. Bu yeni evrişim başlıkları, diğer kendiliğinden dikkat başlıklarıyla birleşerek, hem küresel hem de yerel bağlam öğreniminde daha verimli bir yapı oluşturur. ConvBERT, BERT ve türevlerinden çok daha verimli olup, daha düşük eğitim maliyetleri ve daha az parametre ile çeşitli doğal dil işleme görevlerinde daha iyi sonuçlar elde etmektedir. [74].

Mevcut ConvBERTurk Modelleri aşağıdaki gibidir [75].

- ConvBERTurk (cased): 35GB
- ConvBERTurk mC4 (cased): 242GB
- ConvBERTurk mC4 (uncased): 242GB

Bu modeller, Hugging Face model hub'ı üzerinden Türkçe NLP görevleri için kullanılabilir ve ince ayar yapılabilir.

6.3 Electra

Electra, doğal dil işleme alanında, geleneksel Maskeli Dil Modelleme (Masked Language Modeling, MLM), yöntemlerine alternatif olarak geliştirilmiş bir ön eğitim tekniğidir. BERT gibi MLM yöntemleri, girdi metninde bazı tokenları “[MASK]” ile değiştirerek orijinal tokenları yeniden tahmin etmeye çalışır. Ancak bu yöntemler büyük hesaplama kaynakları gerektirir. Electra, bunun yerine Replaced Token Detection (Değiştirilen Belirteç Algılama, RTD) adı verilen daha

verimli bir ön eğitim görevi sunar. Bu yöntem, girdi metnindeki bazı tokenları küçük bir jeneratör ağı tarafından üretilen alternatiflerle değiştirir. Ardından, model, her token'ın gerçek mi yoksa değiştirilmiş mi olduğunu tahmin eder. Bu yaklaşım, MLM'e göre daha verimlidir çünkü sadece maskelenmiş girdi tokenları üzerinde değil, tüm girdi tokenları üzerinde işlem yapılır. Yapılan çalışmalar, Electra'nın aynı model boyutu, veri ve hesaplama ile BERT'ten daha iyi performans gösterdiğini ortaya koymaktadır. Özellikle küçük modellerde bu fark belirginleşmektedir. Örneğin, Electra, 30 kat daha fazla hesaplama kaynağı kullanılan Üretken Önceden Eğitilmiş Transformatör (Generative Pre-trained Transformer, GPT)'yi geride bırakmıştır. Ayrıca, ELECTRA büyük ölçekli modellerle de iyi performans gösterir; RoBERTa ve XLNet ile karşılaştırıldığında daha az hesaplama kaynağı kullanarak benzer sonuçlar elde eder ve aynı hesaplama kaynağı kullanıldığında onları geride bırakır [76].

ELECTRA'nın Türkçe versiyonu da Hugging Face üzerinde mevcuttur ve şu modelleri içerir [77]:

- ELECTRA Small (cased): 35GB
- ELECTRA Base (cased): 35GB
- ELECTRA Base mC4 (cased): 242GB
- ELECTRA Base mC4 (uncased): 242GB

Eğitim korpusu 35GB büyüklüğünde olup, 44,04 milyon token içermektedir. Model, TPU v3-8 üzerinde 1 milyon adımda eğitilmiştir [77].

6.4 RAG

RAG, son yıllarda LLMlerin bilgiye dayalı ve alan odaklı görevlerdeki sınırlamaları aşmak için geliştirilmiş bir teknolojidir. Özellikle, modelin eğitim verilerinin dışındaki bilgilere erişim sağlamak ve bu verileri kullanarak doğru sonuçlar üretmek için önemlidir. LLMler, eğitim verileri dışında yeni, güncel veya özel bilgi gerektiren durumlarla karşılaştıklarında, genellikle "halüsinasyonlar" yani hatalı ve gerçeğe dayanmayan sonuçlar üretebilirler. Bu durum, özellikle yoğun bilgi ve uzmanlık gerektiren alanlarda büyük bir sorun oluşturur. RAG, bu sorunu aşmak için LLM'lere dışarıdan bilgi çekme yeteneği ekler. Model, ilgili döküman parçalarını, genellikle büyük bir bilgi tabanından, semantik benzerlik hesaplamaları yaparak alır ve ardından bu bilgiyi kullanarak doğru ve gerçekçi içerik üretir. Bu

süreç, RAG'ın doğru ve güvenilir içerik üretme yeteneğini artırır, çünkü model sadece kendi eğitim verilerine değil, dışarıdan alınan bilgilere de başvurur. Başlangıçta, RAG, Transformer mimarisi ile birlikte gelişti ve dil modellerinin daha fazla bilgiyle güçlendirilmesi üzerine yoğunlaştı. Erken dönem çalışmaları, genellikle önceden eğitim (pre-training) tekniklerinin iyileştirilmesi üzerinedir. Daha sonra, ChatGPT'nin çıkışıyla birlikte RAG araştırmaları; LLM'lerin daha karmaşık ve bilgi yoğun görevlerde başarılı olabilmesi için daha iyi bilgi sağlamaya odaklandı. Zamanla, RAG sadece çıkarım (inference) aşamasında değil, aynı zamanda LLMlerin fine-tuning süreçlerine de entegre edilmeye başlandı. Bu sayede, RAG'ın gelişimi ve modüler yapısı daha da güçlendi. Bugün RAG, büyük dil modellerinin gerçek dünya uygulamalarına daha uygun hale gelmesini sağlayan önemli bir teknoloji olarak kabul edilmektedir. RAG teknolojisinin başarılı bir şekilde işleyebilmesi için üç temel bileşen bulunur: "Bilgi alma" (Retrieval), "Üretim" (Generation) ve "Artırma" (Augmentation). Bu bileşenler, birlikte çalışarak RAG'ın etkili ve verimli bir şekilde çalışmasını sağlar. "Retrieval", doğru bilgi parçalarını almak, "Generation", bu bilgileri anlamlı ve doğru bir şekilde sentezlemek ve "Augmentation" ise mevcut dil modelini dışarıdan alınan bilgi ile zenginleştirmektir [78].

7

VERİ TOPLAMA VE TARAMA YÖNTEMLERİ

WS ve Web Tarama (Web Crawling, WC), çevrimiçi verilerin otomatik olarak toplanmasını sağlayan iki farklı ancak birbiriyle ilişkili süreçlerdir. WC, HTML gibi yapılandırılmamış metinlerden veri çıkarmaya odaklanırken, genellikle bu veri makine tarafından okunabilir formatlara JavaScript Nesne Gösterimi (JavaScript Object Notation, JSON) veya Genişletilebilir İşaretleme Dili (Extensible Markup Language, XML) olarak dönüştürülür. Özellikle veri sağlanmadığında veya erişim kısıtlı olduğunda, WS sayesinde kullanıcılar yapılandırılmış veriler elde edebilir. Örneğin, e-ticaret sitelerinden gerçek zamanlı fiyat bilgileri toplamak veya karanlık ağdaki yasa dışı ticaret verilerini analiz etmek gibi uygulamalarda büyük bir öneme sahiptir [79].

WC ise daha geniş bir süreçtir ve belirli bir web sitesi ya da web alanında bulunan tüm sayfaların sistematik olarak keşfedilmesini içerir ve bu sürecin bir alt kümesi olarak düşünülebilir. Bu kapsamda yapılan çalışmalarda toplanan verilere odaklanır. WS ve WC'nin başarılı bir şekilde uygulanması için örümcekler (spiders), desen eşleme (pattern matching) gibi teknolojiler kullanılır. Bu teknikler, manuel veri girişi ile karşılaştırıldığında daha tutarlı, kapsamlı ve doğru sonuçlar sunar. Ancak, bu süreçlerin etik ve yasal boyutlarının göz önünde bulundurulması önemlidir.

WS, veri toplama süreçlerinde etkili bir yöntem olmakla birlikte, bu yöntemin yasal ve etik boyutları sıklıkla göz ardı edilmektedir. Bu durum, ciddi etik tartışmalara ve hukuki ihtilaflara neden olabilmektedir [80]. Özellikle web sitelerinin kullanım koşulları, otomatik erişime izin verilip verilmediği ve toplanan verilerin nasıl kullanılacağı gibi hususlar kritik öneme sahiptir. Kullanım koşullarının uygulanabilirliği, teknolojik koruma araçlarının varlığı, içeriklerin telif hakkı durumu ve erişim yasalarının ihlal edilip edilmediği dikkatle değerlendirilmelidir. WS faaliyetlerinin yasal ve etik çerçevede yürütülmesi, hem veri toplayıcılar hem de web sitesi sahipleri açısından olası uyuşmazlıkları en aza indirmek için önemli bir gereklilik olarak öne çıkmaktadır. Yasal düzenlemelerin ve mahkeme

kararlarının sürekli evirildiği bu alanda hem veri toplayıcıların hem de web sitesi sahiplerinin güncel gelişmeleri takip etmesi büyük önem taşımaktadır. Konuyla ilgili daha detaylı araştırmalar farklı çalışmalarda gerçekleştirilmiştir [79-81].

7.1 Web Kazıma Yöntemleri

Veri toplama işlemi için Python'daki popüler WS kütüphaneleri kullanılabilir. Bu kütüphaneler, verileri internet üzerindeki sayfalardan çekmek ve LangChain'den faydalanarak RAG ile işlemek için temel bir yöntem sağlayabilir.

7.1.1 BeautifulSoup

BeautifulSoup, HTML ve XML dosyalarındaki verileri ayrıştırmak için yaygın olarak kullanılan bir Python kütüphanesidir. Genellikle basit sayfalardan veri çekmek için kullanılır. Web sayfalarındaki etiketleri, metinleri ve bağlantıları kolayca ayıklamak için idealdir [82]. Ayrıca BeautifulSoup LangChain ile entegrasyonu sayesinde, WS süreci daha etkili hale gelir ve harici veriler RAG tabanlı yöntemlerle işlenebilir [83].

7.1.2 Scrapy

Scrapy, Python ile WS işlemleri gerçekleştirmek için güçlü bir çerçevedir. Web sayfalarını gezip veri çekmek için verimli ve hızlı bir yapıya sahiptir. Scrapy, verilerin toplanmasının yanı sıra, verilerin düzenlenmesi ve farklı formatlarda (JSON, CSV, XML vb.) kaydedilmesi işlemlerini de yönetebilir [84].

7.1.3 Selenium

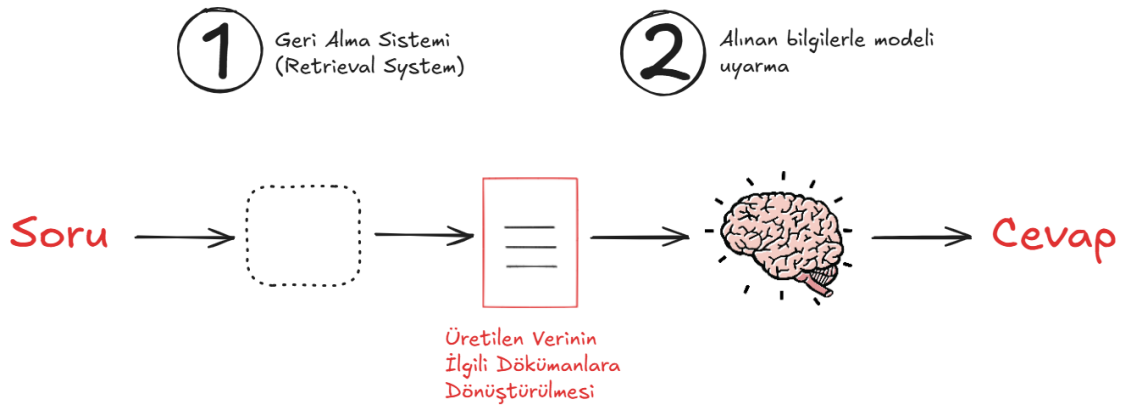
Selenium, özellikle dinamik web sayfalarından veri çekmek için kullanılır. JavaScript ile yüklenen içerikleri almak için idealdir. Çoğu modern web sitesi, içeriklerini dinamik olarak JavaScript ile yüklediğinden, bu tür sayfalardan veri çekebilmek için Selenium kullanılabilir [85].

8

SOHBET ROBOTU VERİ SETİ

Çalışmada, gizliliği koruma amacıyla adının açıklanmasına izin vermeyen bir teknoloji firması tarafından geliştirilen sohbet robotu için tez kapsamında hazırlanmış olan Türkçe veri seti kullanılmıştır. Veri seti, çeşitli sosyal ve kültürel faktörlere odaklanarak geniş kapsamlı bilgiler toplanarak oluşturulmuştur. Veriler, kullanıcıların restoran ve mağaza tercihlerinden ulaşım ve etkinlik kullanımına kadar geniş bir yelpazeyi kapsayan konularda elde edilmiştir. Veri toplama süreci, etkileşimli sohbetler aracılığıyla bilgi edinilmesi ile gerçekleştirilmiştir. Bu sayede, katılımcıların günlük yaşamlarıyla ilgili çeşitli tercihler ve alışkanlıklar üzerine derinlemesine veriler toplanmıştır. Ayrıca, Bölüm 7’de bahsedilen WC yöntemi kullanılarak, belirli sosyal medya platformlarından ve çevrimiçi forumlardan, kullanıcıların günlük yaşamlarına dair tercihleri, yorumları ve alışkanlıkları analiz edilmiştir. Elde edilen veriler, sohbet robotunun daha etkili bir şekilde geliştirilmesine olanak tanıyacak şekilde düzenlenmiştir. Bu botun amacı, kullanıcıların günlük yaşamlarıyla ilgili daha iyi önerilerde bulunmak, onların alışkanlıklarını analiz ederek kişiselleştirilmiş deneyimler sunmak ve kullanıcıların çeşitli konularda daha hızlı ve verimli çözümler bulmalarına yardımcı olmaktır.

8.1 Üretilen Verinin RAG ile İşlenmesi



Şekil 8.1 Üretilen Verinin RAG ile İşlenmesi

Doğal dil işleme modelleri, eğitildikleri veri setine dayalı olarak çalışır ve bu veriler genellikle sabittir. Bu da, modelin güncel olaylara ve özgül alan bilgilerine dair sorulara yanıt verme yetisini kısıtlar. RAG, böyle bir durumda, ilgili bilgileri dış bir bilgi tabanından alarak modelin yanıtlama kapasitesini genişletir. Bu sayede modelin cevapları daha güncel ve doğru olur. Ayrıca, RAG ile sağlanan dış kaynaklar, özel bilgi sunarak modelin uzmanlık alanlarını genişletir ve yanıtların tutarlılığını artırır.

İlk aşamada, RAG, modelin sınırlı ve genellikle güncellenmeyen dahili bilgilerini aşmak için harici bilgi kaynaklarından veri alır. İkinci aşama ise, bu harici bilgiler modele entegre edilir. Bu adımda, öncelikle modelin aldığı giriş sorgusu üzerinden bilgi alma sistemi, uygun verileri arar ve elde eder. Elde edilen bu veriler, modelin sorguya verdiği yanıtı eklenir. Bu sayede model, yalnızca dahili bilgisiyle değil, aynı zamanda dış kaynaklardan aldığı verilerle güçlendirilmiş bir yanıt üretir. Bu yaklaşım, yanlış bilgi üretme (hallucination) oranını düşürür ve daha güvenilir cevaplar elde edilmesini sağlar. RAG yöntemi, özellikle modelin eğitim maliyetlerini azaltan ve bilgi güncellemeleri sağlamak için pahalı fine-tuning yerine daha verimli bir alternatif sunar. Ayrıca, dış veri kaynaklarının entegrasyonu sayesinde, model daha az hata yapar ve daha doğru sonuçlar verir. Bu süreç, dil modellerinin sürekli değişen dünyaya uyum sağlamasına ve daha geniş bir bilgi yelpazesinde işlem yapmasına olanak tanır [86].

8.2 Veri Setinde Dikkat Edilen Hususlar

Veriler arasında sicil numarası, telefon numarası, e-posta adresi gibi kişisel veri içerebilecek herhangi bir bilgi bulunmamaktadır. Kurum ve kuruluşların kimliğini ya da kullanıcıların kimliğini herhangi bir şekilde belli edebilecek hiçbir veri üretilmemiş ve kullanılmamıştır. Ayrıca, tüm veriler anonimleştirilmiş ve maskeleyme teknikleri kullanılarak, kişisel verilerin korunmasına ilişkin yasal düzenlemelere ve etik kurallara uygun olarak işlenmiştir. Veri seti yalnızca yasal izinler çerçevesinde ve telif hakkı ihlali oluşturmayacak şekilde oluşturulmuş ve kullanılmıştır. Bu doğrultuda, KVKK (Kişisel Verilerin Korunması Kanunu) ve Telif Hakları Kanunu'na uygun hareket edilmiştir [87, 88].

8.3 Veri Seti Versiyonları: V1 ve V2 Arasındaki Farklar ve Ön İşleme Süreçleri

Veri seti, başlangıçta Versiyon 1 (V1) olarak 118 alt kategori altında sınıflandırılmıştır. Ancak, daha yalın bir yapı oluşturmak ve analiz süreçlerini kolaylaştırmak amacıyla, Versiyon 2 (V2)'de bu alt kategoriler birleştirilerek 63 temel kategoriye indirgenmiştir. Çalışmada, karmaşıklığı azaltmak ve daha net sonuçlar elde edebilmek adına MLIC yerine SLIC yöntemi tercih edilmiştir.

Çalışmada, veri seti ön işleme süreçlerinin kapsamına göre iki ayrı versiyonda değerlendirilmiştir. İlk versiyon olan V1 veri seti, herhangi bir ön işleme işlemi gerçekleştirilmeden ham haliyle bırakılmıştır. Bu versiyon, hatalı ve eksik kategorilendirmelerin varlığı nedeniyle, analiz süreçlerinde anlam karmaşasına yol açma potansiyeli göz önünde bulundurularak, sonuçları karşılaştırmak amacıyla kullanılmıştır. Benzer kategorilerin birleştirilmemesi, bazı verilerin yanlış sınıflarda yer almasının veri kalitesini olumsuz yönde etkileyebileceği ve modelin genel performansında risk yaratma ihtimali göz önünde bulundurulmuştur. V1 veri seti, iki versiyon arasındaki farklılıkları incelemek ve analizlerdeki belirgin sorunları gözlemek amacıyla bilinçli olarak seçilmiştir.

İkinci versiyon olan V2 veri seti, daha tutarlı ve doğru analizler yapabilmek için kapsamlı bir ön işleme sürecine tabi tutulmuştur. Bu süreçte ham verilerin detaylı bir şekilde incelenmiş, eksik veya hatalı etiketler uygun kategorilerle değiştirilmiş ve veri setindeki kategori ile sınıflandırma tutarlılığı sağlanmıştır. Ayrıca, yorumlardaki nümerik değerler, gereksiz terimler, özel karakterler, noktalama işaretleri, bağlantı adresleri ve gereksiz boşluklar temizlenmiştir ve tüm harfler küçük harfe dönüştürülmüş ve yazım hataları düzeltilmiştir.

Son zamanlarda duygu analizi çalışmalarında durak kelimelerinin önemine dair tartışmalar bulunduğundan [89], bu süreçte veri ön işleme aşamasında durak kelimeleri bilerek kaldırılmamıştır. Öte yandan, durak kelimelerinin etkisini gözlemek amacıyla veri seti, hem durak kelimeleri ile birlikte hem de durak kelimeleri olmadan iki farklı versiyon olarak değerlendirilmiştir.

Tablo 8.1 Versiyon 1 (V1) ve Versiyon 2 (V2) Detayları

Veri Seti Versiyonu	Eğitim Veri Sayısı	Test Veri Sayısı	Eğitim Veri Oranı	Test Veri Oranı	Toplam Veri Sayısı	Toplam Kategori Sayısı
Versiyon 1 (V1)	3359	903	%78.8	%21.2	4262	118
Versiyon 2 (V2)	3403	880	%79.4	%20.6	4283	63

Eğitim veri sayısı açısından her iki versiyon da benzer büyüklükte verilere sahiptir. V1'de 3359 eğitim verisi bulunurken, V2'de bu sayı 3403'e çıkmıştır. Bu durum, V2'nin veri ön işleme işlemi sonrasında biraz daha fazla eğitim verisi içerdiğini göstermektedir.

Test veri sayısı açısından ise, V1'in test veri sayısı 903 iken, V2'de bu sayı 880'e düşmüştür. Bu küçük fark, V2'nin biraz daha az sayıda test verisi kullandığını, ancak genel veri setinin hala geniş olduğunu ve bu farkın modelin doğruluğunu etkileyebilecek bir seviyede olmadığını göstermektedir.

Eğitim veri oranı V1'de %78.8, V2'de ise %79.4 olarak kaydedilmiştir. Bu oranlar arasındaki küçük fark, her iki versiyonun eğitim seti için oldukça benzer oranlara sahip olduğunu ve veri setindeki eğitim verilerinin büyük kısmının her iki versiyonda da kullanılacağını göstermektedir.

Test veri oranı ise sırasıyla V1'de %21.2 ve V2'de %20.6 olarak belirlenmiştir. Bu da her iki versiyon için de veri setlerinin büyük kısmını eğitim için ayırırken, test verisi oranlarının birbirine yakın seviyelerde olduğunu göstermektedir.

Toplam kategori sayısı açısından, V1 veri setinde 118 kategori bulunurken V2'de bu sayı 63'e düşürülmüştür. Bu azalma, veri setindeki kategorilerin daha yalın bir yapıya indirgenmesi ve analiz süreçlerinin daha kolay yönetilmesi amacıyla yapılan ön işleme sürecinin bir sonucu olarak değerlendirilebilir. V2'nin daha az kategoriyle

işlem yapılması, verilerin daha tutarlı ve anlaşılır bir şekilde kategorize edilmesine olanak sağlamıştır.

Her iki veri setinde de makine öğrenmesi uygulamalarında yaygın olarak önerilen %80 eğitim ve %20 test oranına göre veri setlerinin ayrılması bu çalışmada da benimsenmiştir. Önerilen bu oranlar, modelin genelleme yeteneğini değerlendirmede önemli bir rol oynamakta ve benzer çalışmalarda sıklıkla kullanılmaktadır [90-92].

Bu çalışmada, veri setlerinin yönetimi ve ön işleme süreçlerinde çeşitli Python kütüphanelerinden yararlanılmıştır. Özellikle, pandas kütüphanesi, veri analizi ve işleme işlemlerinde etkin bir şekilde kullanılarak veri çerçevelerinin (DataFrame) oluşturulması ve manipülasyonu için tercih edilmiştir. Pandas, kullanıcıların veri kümesi üzerinde kolayca işlem yapmasına olanak tanırken, eksik verilerin tespiti, veri tiplerinin değiştirilmesi ve temel istatistiklerin hesaplanması gibi işlemleri hızlandırmaktadır [93].

Ayrıca, datasets kütüphanesi, farklı veri setlerinin yüklenmesi ve yönetilmesi için kullanılmıştır. Bu kütüphane, kullanıcıların veri setlerini kolayca yükleme, dönüştürme ve analiz etmelerine olanak tanırken, modelleme sürecinde veri yönetimini basitleştirmektedir [94]. Veri setlerinin eğitim ve test kümesine ayrılması işlemi için ise sklearn.model_selection modülünden train_test_split fonksiyonu tercih edilmiştir. Bu fonksiyon, kullanıcıların veri setlerini belirli oranlarda (%80 eğitim, %20 test) bölerek modelin genelleme yeteneğini değerlendirmelerini sağlamaktadır [95].

Veri ön işleme aşamasında ise özel karakterlerin ve noktalama işaretlerinin temizlenmesi gibi işlemler için Python'un yerleşik string kütüphanesi kullanılmıştır. Bu kütüphane, metin verilerinin işlenmesi sırasında gereksiz karakterlerin kaldırılmasını ve metinlerin standartlaştırılmasını sağlayarak veri kalitesini artırmaktadır [96]. Bu kütüphanelerin kombinasyonu, veri setlerinin daha iyi bir biçimde işlenmesini sağlamanın yanı sıra sonuçların güvenilirliğini de artırmaktadır. Veri analizi ve makine öğrenimi alanında bu kütüphanelerin etkin kullanımı, doğru ve geçerli sonuçlar elde edilmesine önemli katkılar sunmaktadır.

Çalışma kapsamında, modellerin eğitimi ve değerlendirilmesi sırasında, yukarıda belirtilen yüksek performanslı donanım özelliklerine sahip bir sistem kullanılmıştır.

- 3. İşlemci:** Intel® Alder Lake Core™ i7-12700H, 14 çekirdekli ve 20 izlekli yapısıyla yüksek işlem gücü sunmaktadır. Bu, veri işlemede ve model eğitiminde düşük süreler elde edilmesine katkı sağlamıştır.
- 4. RAM:** 32GB DDR4 3200MHz bellek, büyük veri setlerinin işlenmesi ve modellerin eğitimi sırasında bellek sıkıntılarını önlemiştir
- 5. Ekran Kartı:** nVIDIA® GeForce® RTX3050 TI, grafik işlemci tabanlı hızlandırma gerektiren modellerde ek bir performans avantajı sağlamıştır.

Kullanılan donanım özellikleri makine öğrenimi modellerinin hızlı bir şekilde eğitilmesine ve değerlendirilmesine olanak tanımış; eğitim süreleri, model türüne bağlı olarak dönem sayısı başına değişkenlik göstermiştir. Bu tür bir donanım seçimi, büyük veri setleriyle çalışan modellerin eğitim sürelerini kısaltırken aynı zamanda işlem doğruluğunda da istikrar sağlayabilmiştir.

9.1 V1 Veri Seti ile Geleneksel Yöntemlere Ait Bulgular

V1 veri seti FastText, Word2Vec ve GloVe gibi geleneksel tekniklerle işlenmiş ve bu tekniklerin performansları değerlendirilmiştir. Bu sürecin temel amacı modellerin dilsel örüntüleri yakalayıp metinleri daha etkili bir şekilde temsil etmesini sağlamaktır. Her model öncelikle 50 dönem sayısı boyunca eğitilmiş ve bu süreçte dilsel özellikleri öğrenilerek anlamlı vektör temsilleri oluşturmuştur. Eğitim sürecinin ardından, modellerin performansı test verisi üzerindeki tahminler üzerinden değerlendirilmiş ve test sonuçları, modellerin gerçek dünya koşullarındaki başarısını ölçmek için ana kriter olarak kullanılmıştır. Eğitim verisine ait sonuçlar aşağıda tablolar halinde sunulmuştur ve modellerin öğrenme süreçleri ile performansları detaylı bir şekilde incelenmiştir. Modellerin başarıları; doğruluk, F1 skoru, kesinlik ve hatırlama gibi metrikler kullanılarak ölçülmüştür. Bu metrikler, hem eğitim hem de test aşamalarındaki performansları kapsamlı bir

şekilde analiz etmek amacıyla kullanılmıştır. Her bir model için bu metrikler, sonuçların güvenilirliğini değerlendirmek ve modellerin farklı veri setlerinde nasıl bir performans sergilediğini anlamak açısından kritik bir rol oynamıştır.

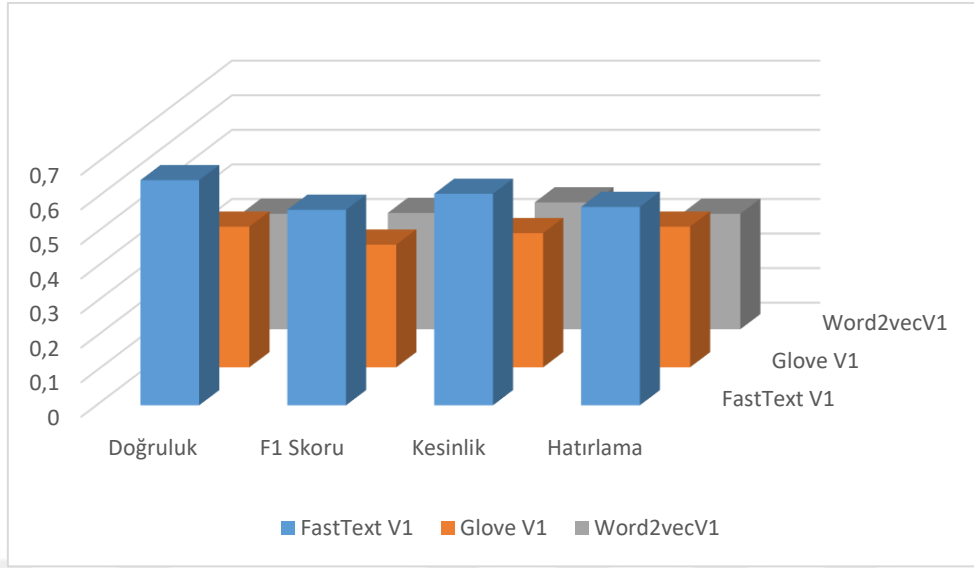
Tablo 9.1 FastText -Glove -Word2Vec V1(50 dönem sayısı) Sonuçları

Model (V1)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre (sn)
FastText	Train	0.8893	0.8858	0.8920	0.9077	50	0.7
	Test	0.6512	0.5654	0.6117	0.5739	50	
Glove	Train	0.4843	0.4369	0.5197	0.4843	50	0,2
	Test	0.4067	0.3547	0.3881	0.4067	50	
Word2vec	Train	0.9034	0.8989	0.9165	0.9034	50	1.2
	Test	0.3341	0.3365	0.3670	0.3341	50	

Tablo 9.1’de görüldüğü üzere FastText modeli V1 için, hem eğitim hem de test aşamalarında en iyi sonuçları veren model olarak öne çıkmıştır. Eğitim setinde %88.93 doğruluk ve %88.58 F1 skoru elde etmiş, test setinde ise %65.12 doğruluk ve %56.54 F1 skoru ile en iyi performansı sergilemiştir. Eğitim süresi 0.7 saniye olan FastText, hızlı eğitimin yanı sıra test verisi üzerinde de başarılı bir genelleme yeteneği göstermektedir. Bu sonuçlar, modelin hızlı ve etkili bir şekilde çalıştığını göstermiştir.

GloVe modeli V1 için, 0.2 saniyelik eğitim süresiyle en hızlı model olmasına rağmen performans açısından geride kalmıştır. Eğitim setinde %48.43 doğruluk ve %43.69 F1 skoru elde ederken, test setinde %40.67 doğruluk ve %35.47 F1 skoru ile düşük performans sergilemiştir. GloVe'nin veri seti üzerindeki başarısı sınırlı kalmış ve özellikle test verisinde yeterli bir genelleme yapamamıştır.

Word2Vec modeli ise V1 için, 1.2 saniye ile en uzun eğitim süresine sahip modeldir. Eğitim setinde %90.34 doğruluk ve %89.89 F1 skoru ile oldukça iyi sonuçlar elde etmiş olmasına rağmen, test setinde %33.41 doğruluk ve %33.65 F1 skoru ile en düşük performansı göstermiştir. Bu durum, Word2Vec'in eğitim verisinde Oft nedeniyle test verisi üzerinde iyi bir performans sergileyemediğini göstermektedir.



Şekil 9.1 V1 (50 dönem sayısı) Geleneksel Yöntemlerin Sonuçlarının Değerlendirilmesi

Elde edilen sonuçlar, özellikle Word2Vec ve FastText modellerinde Oft problemini açıkça göstermektedir. GloVe modeli ise düşük performansı ile daha çok Uft sorunu yaşamaktadır. Bu durum, V1 veri seti üzerinde çalışmanın sınırlamalarını ve daha iyi bir veri yapısının gerekliliğini vurgulamaktadır.

50 dönem sayısı boyunca elde edilen sonuçların ardından, modellerin genelleme yeteneklerini daha kapsamlı bir şekilde analiz etmek amacıyla bir sonraki aşamaya geçilmiştir. Bu aşamada, veri setinde herhangi bir değişiklik yapılmadan dönem sayısı artırılmış ve modellerin performansları değerlendirilmiştir. Her bir model, 100 dönem sayısı boyunca eğitilmiş ve bu süreçte öğrenme kapasitelerinde bir iyileşme olup olmadığı ile test verisi üzerindeki genelleme yeteneklerinde herhangi bir değişiklik meydana gelip gelmediği gözlemlenmiştir. Özellikle FastText, GloVe ve Word2Vec modellerinin daha uzun eğitim sürelerinde gösterdiği performans, modellerin hem eğitim hem de test veri setleri üzerindeki davranışlarının daha iyi anlaşılmasını sağlamıştır.

Tablo 9.2 FastText- Glove- Word2vec V1 (100 dönem sayısı) Sonuçları

Model (V1)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Süre (sn)
FastText	Train	0.9018	0.8913	0.8970	0.8964	1.3
	Test	0.6445	0.6590	0.6372	0.6445	
Glove	Train	0.5174	0.4873	0.6027	0.5174	0,4
	Test	0.4232	0.4160	0.3779	0.4232	
Word2vec	Train	0.9034	0.8989	0.9165	0.9034	1.7
	Test	0.3481	0.3479	0.3825	0.3481	

Tablo 9.2'deki sonuçlara göre, FastText modeli V1 için, hem eğitim hem de test aşamalarında diğer modellere kıyasla üstün bir performans sergilemiştir. Eğitim setinde %90.18 doğruluk ve %89.13 F1 skoru elde eden model, test setinde de %64.45 doğruluk ve %65.90 F1 skoru ile başarılı bir genelleme yapabildiğini göstermiştir. Ayrıca, 1.3 saniyelik kısa eğitim süresi, modelin hızlı ve etkili bir sonuç sunduğunu ortaya koymaktadır. Bu veriler, FastText'in kelime vektörlerini etkili bir şekilde temsil edebilme yeteneğini desteklemektedir.

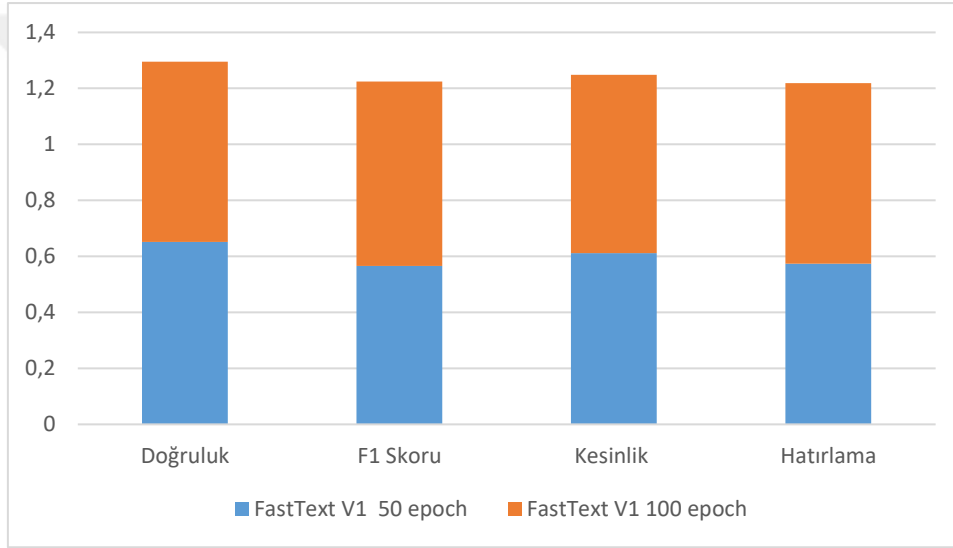
GloVe modeli V1 için, 0.4 saniye ile en kısa eğitim süresine sahip olmasına karşın, performans açısından en zayıf sonuçları sunmuştur. Eğitim setinde elde edilen %51.74 doğruluk ve %48.73 F1 skoru, modelin veri setindeki özellikleri yeterince iyi öğrenemediğini göstermektedir. Test setinde elde edilen %42.32 doğruluk ve %41.60 F1 skoru, GloVe'nin test verisi üzerinde düşük genelleme yeteneğine sahip olduğunu ve gerçek dünya verileri üzerinde yetersiz kaldığını işaret etmektedir.

Word2Vec modeli V1 için, 1.7 saniye ile en uzun eğitim süresine sahip modeldir. Eğitim setinde elde edilen %90.34 doğruluk ve %89.89 F1 skoru, modelin eğitim verisine oldukça iyi uyum sağladığını göstermektedir. Ancak, test setindeki %34.81 doğruluk ve %34.79 F1 skoru, Word2Vec'in Oft problemine sahip olduğunu ve test verisi üzerinde yeterli genelleme yapamadığını ortaya koymaktadır. Bu durum, modelin eğitim verisiyle güçlü bir şekilde uyum sağlarken, farklı verilerle çalışmada başarısız olduğunu göstermektedir.

Tablo 9.1 ve Tablo 9.2 den FastText modelinin V1 veri setinde, hem eğitim süresi hem de test seti üzerindeki performansıyla diğer modellere göre her iki dönem

sayısı içinde daha başarılı olduğu görülmektedir. GloVe, kısa eğitim süresine rağmen yeterli başarıyı sağlayamazken, Word2Vec uzun sürede eğitildiği halde test seti üzerinde dönem sayısının yükseltilmesine rağmen en zayıf performansı sergilemiştir. Bu sonuçlar ışığında eğitim süresi ile performans arasında her zaman doğrudan bir ilişki olmadığını ve genelleme yeteneğinin modelin başarısı açısından kritik olduğunu ortaya koymaktadır.

Şekil 9.2 den görüldüğü gibi, bu analizler, dönem sayısı sayısındaki artışın yalnızca belirli durumlarda etkili olduğunu ve genelleme yeteneği zayıf olan modellerin bu tür ayarlamalardan sınırlı fayda sağladığını göstermektedir.



Şekil 9.2 V1 (50 dönem sayısı) ve V1 (100 dönem sayısı) FastText Yöntemlerin Sonuçlarının Karşılaştırılması

9.2 V1 Veri Seti ile Modern Yöntemlere Ait Bulgular

Bu bölümde, V1 veri seti kullanılarak modern yöntemlerle gerçekleştirilen analiz sonuçlarına yer verilmiştir. Geleneksel yöntemlerin yeterince yüksek başarı oranlarına ulaşamaması üzerine, aynı veri seti üzerinde daha gelişmiş modellerin performansı incelenmiştir. Bu kapsamda, BERT, ConvBERT ve Electra gibi derin öğrenme modellerine FiT uygulanarak niyet sınıflandırma başarıları değerlendirilmiştir. Ayrıca, çalışmayı daha kapsamlı hale getirmek ve karşılaştırma çeşitliliğini artırmak amacıyla, güncel tekniklerden biri olan RAG yöntemi de

deneysel analizlere dahil edilmiştir. Bu yöntemlerin kullanımıyla elde edilen sonuçlar, modern yaklaşımların niyet sınıflandırma alanında sağladığı katkıları somut verilerle ortaya koymayı hedeflemiştir.

Tablo 9.3 BERT-ConvBert-Electra V1 (50 dönem sayısı) Sonuçları

Model (V1)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre
BERT	Train	0.9035	0.8910	0.8934	0.8962	50	01:18:59
	Test	0.7297	0.6797	0.6897	0.6878	50	
ConvBert	Train	0.9035	0.8902	0.8927	0.8955	50	01:13:54
	Test	0.7242	0.6515	0.6511	0.6714	50	
Electra	Train	0.9035	0.8923	0.8964	0.8943	50	01:02:28
	Test	0.7165	0.6494	0.6564	0.6665	50	

Tablo 9.3'te görüldüğü üzere, BERT modeli V1 için, %72.97 doğruluk ve %67.97 F1 skoru ile test verisi üzerinde en başarılı sonuçları elde eden modeldir. Eğitim verisi üzerinde %90.35 doğruluk ve %89.10 F1 skoru kaydeden model, veriyi etkili bir şekilde öğrenmiş ve test verisi üzerinde de güçlü bir genelleme yeteneği sergilemiştir. Eğitim süresinin 1 saat 18 dakika gibi uzun bir süre olmasına rağmen, elde edilen sonuçlar, BERT'in derin yapısının ve FiT süreçlerinin etkili olduğunu ortaya koymaktadır. Bu sonuçlar, modelin Oft belirtileri göstermediğini ve hem eğitim hem de test verileri üzerinde dengeli bir performans sergilediğini işaret etmektedir.

ConvBERT modeli V1 için, test verisi üzerinde %72.42 doğruluk ve %65.15 F1 skoru ile BERT'e oldukça yakın bir performans göstermiştir. Eğitim verisindeki %90.35 doğruluk ve %89.02 F1 skoru, modelin veriyi iyi öğrendiğini ortaya koymaktadır. Eğitim süresinin 1 saat 13 dakika olması, BERT'e kıyasla daha kısa bir sürede benzer sonuçlar elde edilebileceğini göstermektedir. Ancak test performansındaki küçük farklar, ConvBERT'in genelleme yeteneğinin BERT kadar güçlü olmadığını düşündürmektedir. Yine de, Oft belirtilerine rastlanmamış olması modelin dengeli bir öğrenme gerçekleştirdiğini ortaya koymaktadır.

Electra modeli V1 için, 1 saat 2 dakika ile en hızlı eğitilen modeldir. Eğitim verisi üzerinde %90.35 doğruluk ve %89.23 F1 skoru elde etmesine rağmen, test verisi üzerindeki %71.65 doğruluk ve %64.94 F1 skoru, diğer modellere göre biraz daha

düşük bir performansı işaret etmektedir. Eğitim ve test sonuçları arasındaki bu fark, modelin eğitim verisine biraz daha fazla odaklandığını ve genelleme yeteneğinin BERT ve ConvBERT kadar güçlü olmadığını göstermektedir. Electra, hız avantajına sahip olsa da, bu avantajın test verisi üzerindeki performansa yeterince yansımadağı görülmektedir.

Tablo 9.4 RAG V1 Sonuçları

Model	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Süre
RAG V1	Train	0.7351	0.6952	0.7514	0.7351	10:05
	Test	0.6091	0.5626	0.5651	0.6090	

Tablo 9.4'te görüldüğü üzere, RAG modeli V1 için, diğer modellerden farklı olarak FiT yapılmadan çalıştırılmıştır ve dönem sayısı kullanılmamıştır. Eğitim süresinin yalnızca 10 dakika 5 saniye gibi oldukça kısa olması, modelin hızlı bir şekilde çalıştırılabilir olması açısından önemli bir avantajdır. Ancak, bu hıza karşılık, modelin performansının diğer modellere kıyasla düşük kaldığı görülmektedir. Eğitim setinde %73.51 doğruluk ve %69.52 F1 skoru elde eden RAG, test setinde ise %60.91 doğruluk ve %56.26 F1 skoru ile en düşük sonuçları sergilemiştir. RAG modeli eğitim setinde nispeten yüksek bir doğruluk oranı elde etmiş olmasına rağmen, test setindeki düşük performans, modelin test verisine yeterince uyum sağlayamadığını ve bu durumun da yetersiz uyum eğilimi olarak değerlendirilebileceğini göstermektedir.



Şekil 9.3 V1 (50 dönem sayısı) Modern Yöntemlerin Süreleri

BERT modeli, test verisi üzerindeki en iyi performansı sunan modeldir ve Şekil 9.3'te eğitim süresi uzun olmasına rağmen, genelleme yeteneği oldukça güçlüdür. ConvBERT ve Electra, BERT'e yakın sonuçlar verse de, performans açısından biraz daha geride kalmışlardır. Özellikle Electra, hız açısından avantajlı olsa da test verisinde aynı başarıyı sağlayamamıştır. RAG modeli ise Şekil 9.3'te görüldüğü üzere modern ve en hızlı yöntem olmasına rağmen, test verisi üzerinde en düşük sonuçları vermiştir.

Tablo 9.5 BERT-ConvBert-Electra V1 (100 dönem sayısı) Sonuçları

Model (V1)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre
BERT	Train	0.9035	0.8916	0.8984	0.8997	100	02:14:30
	Test	0.7320	0.6818	0.6876	0.6918	100	
ConvBert	Train	0.9035	0.8879	0.8965	0.8953	100	02:51:29
	Test	0.7297	0.6663	0.6782	0.6778	100	
Electra	Train	0.9035	0.8904	0.8989	0.8954	100	02:13:01
	Test	0.7032	0.6515	0.6681	0.6602	100	

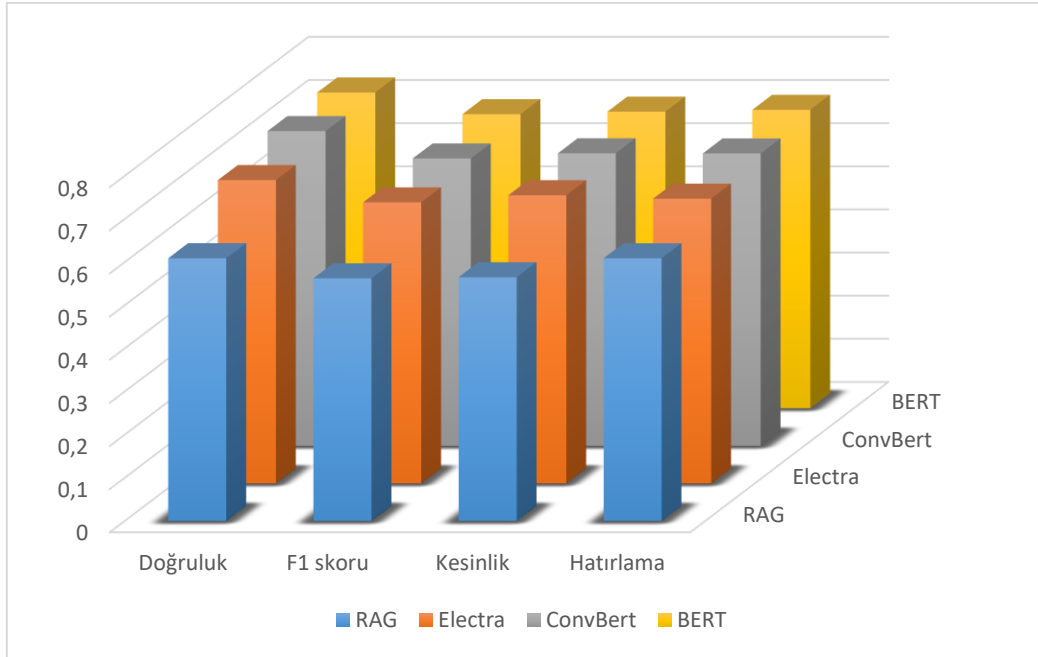
50 dönem sayısı için verilen sonuçların ardından, modellerin genelleme yeteneklerini daha kapsamlı bir şekilde analiz edebilmek için dönem sayısını artırarak modern yöntemlerin performanslarını değerlendirmeye yönelik bir sonraki adıma geçilmiştir. Tablo 9.5 de görüldüğü üzere; BERT model V1 için, 100 epoch sonucunda eğitim setinde %90.35 doğruluk ve %89.16 F1 skoru ile güçlü bir performans sergilemiştir. Test verisi üzerinde ise %73.20 doğruluk ve %68.18 F1 skoru ile önceki 50 dönem sayısı sonuçlarına kıyasla sadece küçük bir iyileşme görülmüştür. Eğitim süresi 2 saat 14 dakikada tamamlanmıştır, yani önceki sürelerle göre oldukça uzun bir eğitim süreci geçirmiştir. Ancak, test verisi üzerindeki bu küçük artış, eğitim süresinin uzatılmasının BERT üzerinde ciddi bir performans artışı sağlamadığını göstermiştir.

ConvBERT modeli V1 için, eğitim setinde %90.35 doğruluk ve %88.79 F1 skoru elde etmiş. Test seti üzerinde %72.97 doğruluk ve %66.63 F1 skoru ile 50 dönem sayısı sonucuna kıyasla yine hafif bir artış gözlemlenmiş, ancak eğitim süresi ciddi şekilde uzayarak 2 saat 51 dakikada tamamlanmıştır. Bu sonuçlar, ConvBERT'in uzun bir eğitim süresine rağmen genelleme yeteneğinde belirgin bir fark

yaratmadığını ortaya koymuştur. ConvBERT modeli BERT ile neredeyse benzer sonuçlar almasına rağmen, eğitim süresinin uzaması dikkate değer bir avantaj sunmamıştır.

Electra modeli ise V1 için, yine 100 dönem boyunca eğitim sürecinde %90.35 doğruluk ve %89.04 F1 skoru ile benzer bir performans göstermiş, ancak test verisi üzerinde %70.32 doğruluk ve %65.15 F1 skoru elde ederek diğer modellere göre daha düşük sonuçlar vermiştir. Eğitim süresi 2 saat 13 dakika olup BERT'e yakın bir hızda gerçekleşmiş, ancak test verisindeki sonuçları BERT ve ConvBERT'in gerisinde kalmaya devam etmiştir.

100 dönem sayısıyla elde edilen model sonuçlarını Tablo 9.5'te sonuçları verilen RAG yöntemiyle karşılaştıracak olursak RAG'nin herhangi bir epoch sayısına bağlı kalmayan yapısı ile test verisi üzerinde %60.91 doğruluk ve %56.26 F1 skoru ile en düşük performansı sergilediği görülmektedir. RAG, çok kısa bir eğitim süresi (10 dakika 5 saniye) ile öne çıkmış olsa da, test seti üzerindeki sonuçlar açısından diğer modellere kıyasla geride kaldığı aşikârdır. RAG'nin hız avantajı olsa da, genelleme yeteneği 100 dönem sayısıyla ile FiT edilen BERT, ConvBERT ve Electra ile kıyaslandığında zayıf kalmaktadır.



Şekil 9.4 V1 (100 dönem sayısıyla) Modern Yöntemlerin Sonuçlarının Değerlendirilmesi

Şekil 9.4'te görüleceği üzere BERT modeli 100 epoch boyunca en iyi sonuçları vermeye devam ederken ConvBERT ve Electra de BERT'e yakın performanslar sergilemiştir ancak eğitim süreleri uzadıkça bu modellerdeki performans artışı sınırlı kalmıştır. RAG yöntemi ise çok daha hızlı bir eğitim süresi sunmasına rağmen, doğruluk ve genelleme açısından diğer modellere göre geride kalmıştır. Bu sonuçlar, dönem sayısının artırılmasının modellerin performansı üzerinde her zaman anlamlı bir fark yaratmadığını göstermiş ve özellikle eğitim süresinin uzamasına rağmen test verisi üzerindeki iyileşmeler sınırlı kalmıştır.

9.3 V2 Veri Seti ile Geleneksel Yöntemlere Ait Bulgular

V1 veri setinde, gerekli ön işlemler yapılmamış bir veri setiyle karşılaştırmalar yapılmıştı. Gerçek dünyada kullanılan bir veri seti olmasına rağmen, bu ham veri seti, genelleme süreçlerinde beklenen performansı gösterememiştir. Özellikle veri ön işleme ve doğru kategorilendirme eksiklikleri, modellerin test verisinde istenilen başarıya ulaşmasını engellemiştir. Bu durumu iyileştirmek için, veri setine dördüncü bölümde bahsi geçen ön işleme aşamaları uygulandı ve veriler doğru kategorilere ayrıldı. V2 veri seti üzerinde yapılan bu iyileştirmelerin, geleneksel yöntemler üzerinde değerlendirilmesi Tablo 7'de olduğu gibi gösterilmiştir.

Tablo 9.6 FastText-Glove-Word2Vec V2 (50 dönem sayısı) Sonuçları

Model (V2)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre Sn
FastText	Train	0.9815	0.9815	0.9820	0.9815	50	0.5
	Test	0.8261	0.7749	0.8179	0.7599	50	
GloVe	Train	0.5811	0.5833	0.7210	0.5811	50	54
	Test	0.5204	0.5087	0.5886	0.5204	50	
Word2vec	Train	0.9845	0.9844	0.9846	0.9845	50	1.1
	Test	0.5005	0.5031	0.5257	0.5005	50	

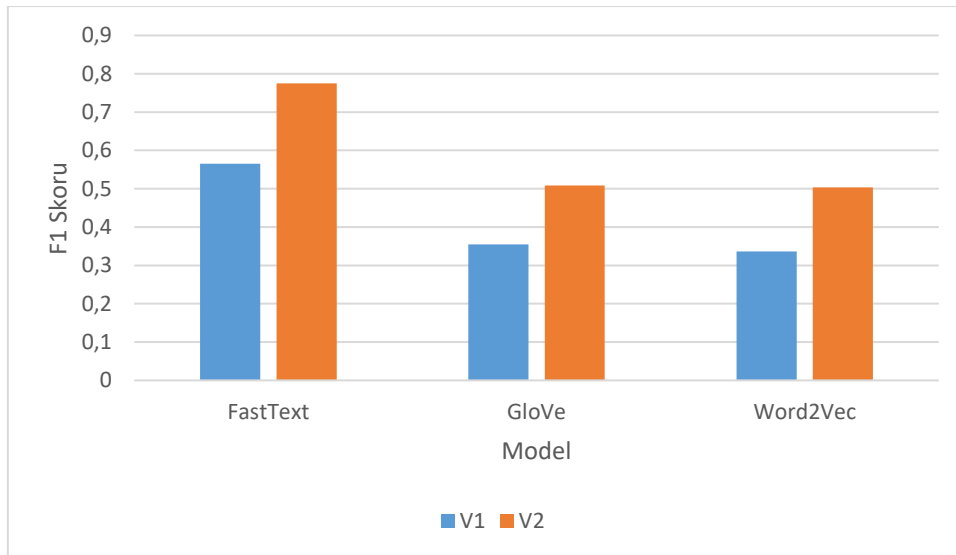
Tablo 9.6'da, V2 veri seti üzerinde gerçekleştirilen çalışmaların sonuçları, veri ön işleme ve doğru kategorilendirme işlemlerinin, modellerin genel performansını artırmada önemli bir rol oynadığını göstermektedir.

FastText modeli V2 için, eğitim verisinde %98.15 doğruluk ve %98.15 F1 skoru ile dikkat çekerken, test verisinde %82.61 doğruluk ve %77.49 F1 skoru elde etmiştir.

Eđitim süresi sadece 0.5 saniye olup, modelin hızlı bir şekilde eğitilebilmesi, yüksek doğruluk oranlarıyla birleřtiđinde oldukça etkili bir performans sergilediđini göstermektedir. FastText'in test setindeki yüksek doğruluđu, modelin veri üzerinde iyi genelleme yapabildiđini ve test verisi ile uyum sađladığını ortaya koymaktadır.

GloVe modeli V2 için, eğitim süresi 54 saniye olmasına rağmen, eğitim setindeki performansı daha düşük seviyelerde kalmıřtır. Eğitim verisinde %58.11 doğruluk ve %58.33 F1 skoru elde edilmesine rağmen, test verisinde bu deđerler %52.04 doğruluk ve %50.87 F1 skoruna düşmüřtür. GloVe, veri üzerinde başarılı sonuçlar elde etmekte zorlanmış ve genelleme yeteneđi sınırlı kalmıřtır. Bu durum, modelin test verisinde daha düşük başarı göstermesinin yanı sıra, veri ön işleme aşamalarının etkisinin sınırlı olabileceđini de göstermektedir.

Word2Vec modeli V2 için, eğitim verisinde %98.45 doğruluk ve %98.44 F1 skoru ile yüksek performans göstermiřtir. Ancak, test verisinde %50.05 doğruluk ve %50.31 F1 skoru ile diđer modellerin gerisinde kalmıřtır. Bu düşük performans, modelin eğitim verisine aşırı uyum sađladığını ve test verisi üzerinde genelleme yapmada zorlandıđını göstermektedir. Eğitim süresi ise 1.1 saniye olarak belirlenmiřtir.



řekil 9.5 V1 ve V2 (dönem sayısı 50) Geleneksel Yöntemlerin F1 Sonuçlarının Deđerlendirilmesi

V2 veri setinin ön işlemden geçirilmiş ve daha iyi kategorilendirilmiş yapısı, modellerin performansına önemli bir katkı sağlamıştır. Özellikle FastText modeli bu iyileştirmelerden en fazla fayda sağlayarak hem hız hem de doğruluk açısından en yüksek performansı sergilemiştir. Eğitim süresinin kısalığına rağmen, test verisinde elde ettiği yüksek F1 skoru Şekil 9.5'te görüleceği üzere V2'de Fasttext özelinde gerçekleşmiştir, bu da modelin genelleme kabiliyetinin güçlendiğini göstermektedir. GloVe ve Word2Vec modelleri ise veri setindeki iyileştirmelerden sınırlı ölçüde faydalanmış ve genel olarak test verisinde hala beklenen performansı gösterememiştir. Bu durum, veri setinin iyileştirilmiş yapısının, GloVe ve Word2Vec gibi kelime tabanlı gömme yöntemlerinin bağlamı tam anlamakta ve karmaşık ilişkileri modellemekteki sınırlamalarını aşmaya yetmediğini açıkça ortaya koymaktadır.

Tablo 9.7 FastText -Glove -Word2Vec V2 (100 dönem sayısı) Sonuçları

Model (V2)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre sn
FastText	Train	0.9818	0.9902	0.9883	0.9928	100	0.8
	Test	0.8295	0.8236	0.8423	0.8295	100	
GloVe	Train	0.6085	0.6186	0.7599	0.6085	100	90
	Test	0.5250	0.5149	0.6088	0.5250	100	
Word2vec	Train	0.9848	0.9847	0.9849	0.9848	100	1.4
	Test	0.5262	0.5228	0.5349	0.5262	100	

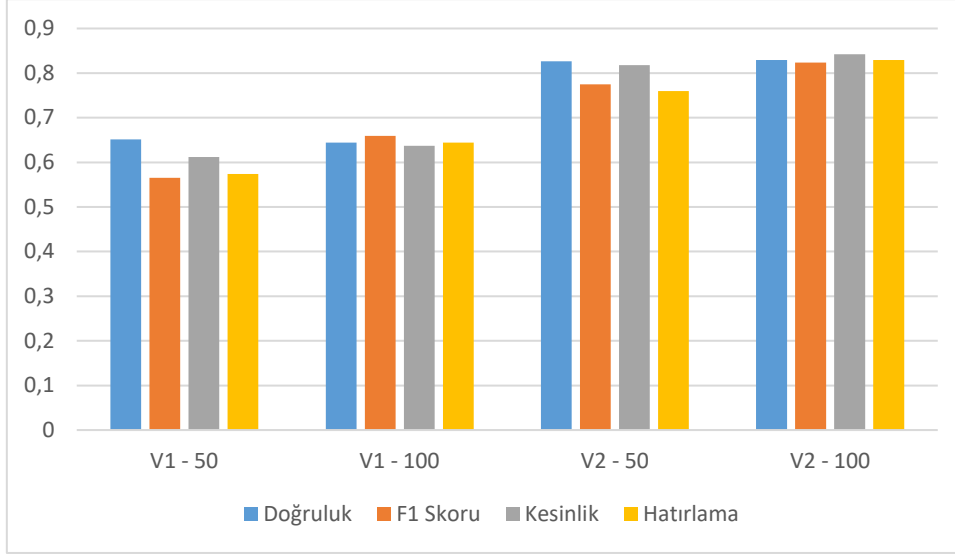
Tablo 9.7'deki 100 dönem sayısı sonuçları incelendiğinde, önceki denemelere kıyasla modellerin performansında bazı iyileşmeler gözlemlenmiştir. Özellikle eğitim seti üzerindeki doğruluk ve F1 skorlarında belirgin artışlar görülmektedir, bu da modellerin veriyle daha derin bir şekilde çalışarak öğrenme sürecini geliştirdiklerini göstermektedir. Ancak, her modelin performansı birbirinden farklıdır; bazı modeller eğitim seti üzerinde güçlü sonuçlar verirken test seti üzerinde bu başarıyı genellemekte zorlanmıştır.

FastText modeli V2 için, hem eğitim hem de test setinde en güçlü performansı sergileyen model olarak öne çıkmıştır. Eğitim verisinde %98.18 doğruluk ve %99.02 F1 skoru ile oldukça başarılı sonuçlar elde etmiştir. Test verisinde ise %82.95 doğruluk ve %82.36 F1 skoru ile test seti üzerinde de iyi bir genelleme

yapabilmiştir. FastText modeli V2 veri setinde en iyi sonuçları veren model olmaya devam etmiştir. Hem eğitim süresi kısa hem de test seti üzerinde en yüksek performansı sergileyen modeldir. Ek olarak bu çalışmada, FastText modelinin Oft veya Uft yapmadığını belirlemek için çapraz doğrulama tekniği kullanılmıştır. Eğitim ve test setleri üzerinde yapılan analizler, modelin eğitim verisine aşırı uyum sağlamadığını ve genelleme yeteneğini başarıyla gösterdiğini ortaya koymuştur. Eğitim setindeki yüksek doğruluk ve F1 skoru ile test setindeki benzer sonuçlar, modelin dengeli bir performans sergilediğini ve hem Oft hem de Uft belirtilerini göstermediğini kanıtlamaktadır

GloVe modeli ise V2 için, eğitim setinde %60.85 doğruluk ve %61.86 F1 skoru ile önceki denemelere göre bir gelişme kaydetmiş, ancak hala istenen seviyede bir sonuç vermemiştir. Test setinde ise %52.50 doğruluk ve %51.49 F1 skoru ile performans açısından beklenen sıçramayı yapamamıştır. 100 dönem sayısı ile eğitim süresi 1 dakika 30 saniye ile oldukça uzun, ancak bu uzun süreye rağmen test sonuçlarında ciddi bir gelişme sağlanamamıştır. GloVe'nin veri seti üzerindeki genelleme yeteneği zayıf kalmış ve modelin eğitim süresiyle test başarısı arasında yeterli bir ilişki kurulamamıştır.

Word2Vec modeli ise V2 için, eğitim setinde %98.48 doğruluk ve %98.47 F1 skoru ile iyi sonuçlar elde etmiş, ancak test setine bakıldığında %52.62 doğruluk ve %52.28 F1 skoru ile yine beklentilerin gerisinde kalmıştır. Eğitim süresi 1.4 saniye ile hızlı olsa da, test sonuçları açısından modelin genelleme yeteneği yeterli seviyeye ulaşmamıştır. Word2Vec modeli, eğitimde başarılı sonuçlar verse de, test seti üzerinde hala düşük performans sergileyerek genelleme yeteneği açısından beklenen düzeye ulaşamamıştır.



Şekil 9.6 V1-V2 Veri Setlerinde FastText Sonuçlarının Karşılaştırılması

FastText modeli, V2 veri setinde dönem sayısının artırılmasıyla test doğruluğu ve F1 skorunda iyileşme göstermiştir. 50 dönem sonuçlarına kıyasla 100 dönem sonucunda performans artarken, eğitim süresi yalnızca 0.8 saniye ile oldukça kısa kalmıştır. Bu durum, modelin hız-performans dengesini koruyarak verimli bir şekilde çalıştığını göstermektedir. FastText'in kelime temsili ve genelleme yeteneği, V2 veri setinde de en iyi sonuçları sağlayarak üstünlüğünü sürdürmüştür.

9.4 V2 Veri Seti ile Modern Yöntemlere Ait Bulgular

V1 veri seti üzerinde yapılan önceki denemelerin ardından, bu çalışmada V2 veri seti kullanılarak modellerin performansları tekrar değerlendirilmiştir. Tablo 9'da yer alan sonuçlar, her bir modelin eğitim süreçleri, doğruluk oranları ve test verisi üzerindeki performansını detaylı bir şekilde göstermektedir.

Tablo 9.8 BERT-ConvBert-Electra V2 (50 dönem sayısı) Sonuçları

Model (V2)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Dönem Sayısı	Süre
BERT	Train	0.9835	0.9916	0.9946	0.9892	50	01:05:55
	Test	0.9238	0.9059	0.9257	0.9012	50	
ConvBert	Train	0.9835	0.9916	0.9939	0.9899	50	02:02:18
	Test	0.9250	0.9141	0.9261	0.9142	50	
Electra	Train	0.9835	0.9913	0.9931	0.9902	50	01:06:26
	Test	0.9181	0.9141	0.9197	0.9226	50	

Tablo 9.8’de görüldüğü üzere BERT modeli V2 için, eğitim verisinde %98.35 doğruluk, %99.16 F1 skoru, %99.46 kesinlik ve %98.92 hatırlama elde edilmiştir. Bu sonuçlar, modelin eğitim verisine oldukça iyi uyum sağladığını ve doğru öğrenme süreci geçirdiğini göstermektedir. Test verisi üzerinde ise %92.38 doğruluk, %90.59 F1 skoru, %92.57 kesinlik ve %90.12 hatırlama elde edilmiştir. BERT’in test performansı oldukça yüksek olup, modelin eğitim setinde öğrendiği bilgiyi genelleme yeteneği güçlüdür. Test sonuçlarındaki iyileşme, BERT’in Oft yapmadığını ve veriye iyi uyum sağladığını gösterir. Çünkü eğitim ve test sonuçları arasında büyük bir fark bulunmamaktadır. Ayrıca, Oft olmaması, modelin genelleme yeteneğini kaybetmeden eğitim verisi üzerinde yüksek başarı gösterdiğini doğrulamaktadır. Eğitim süresi 1 saat 5 dakika olmasına rağmen, BERT’in test seti üzerindeki başarısı modelin sadece eğitim verisinde değil, gerçek dünya senaryolarında da güçlü bir genelleme yeteneğine sahip olduğunu göstermektedir.

ConvBERT modeli V2 için, eğitim setinde %98.35 doğruluk, %99.16 F1 skoru, %99.39 kesinlik ve %98.99 hatırlama ile iyi sonuçlar elde etmiştir. Bu da, modelin eğitim verisi üzerinde oldukça başarılı olduğunu göstermektedir. Test verisinde ise %92.50 doğruluk, %91.41 F1 skoru, %92.61 kesinlik ve %91.42 hatırlama elde edilmiştir. ConvBERT’in eğitim ve test performansı birbirine oldukça yakın olup, bu da modelin Oft yapmadığını gösterir. Eğitim setindeki yüksek başarıyı test setine başarılı bir şekilde genellebilemiştir. Eğitim süresi açısından bakıldığında,

ConvBERT'in eğitimi yaklaşık 2 saat sürmüştür, bu da BERT'e kıyasla daha uzun bir eğitim süresi anlamına gelmektedir. ConvBERT'in yapısal olarak daha karmaşık bir model olması, bu uzun eğitim süresinin nedenlerinden biridir. Derin öğrenme modellerinin karmaşıklığı arttıkça, modelin veri üzerindeki öğrenme süreci de daha fazla zaman alabilmektedir. Bu çalışma özelinde ConvBERT'in uzun eğitim süresine rağmen test verisi üzerinde güçlü sonuçlar vermesi, bu ek zamanın modelin performansına pozitif bir şekilde yansıdığını göstermektedir.

Electra modeli V2 için, performans açısından BERT ve ConvBERT ile oldukça yakın sonuçlar elde etmiştir. Eğitim sürecinde %98.35 doğruluk ve %99.13 F1 skoru yakalayarak, veri setini başarılı bir şekilde öğrenip yüksek doğruluk oranlarına ulaşmaktadır. Bu sonuç, modelin eğitim aşamasında veri setini derinlemesine kavrayarak sınıflandırma görevini çok iyi yerine getirdiğini göstermektedir. Eğitim verisinde elde edilen bu güçlü performans, modelin öğrenme yeteneğinin üst düzeyde olduğunu vurgulamaktadır. Test setine geçildiğinde ise Electra V2 %91.81 doğruluk ve %91.41 F1 skoru ile etkileyici bir genelleme yeteneği ortaya koymaktadır. Test verisi üzerinde bu denli yüksek sonuçlar elde etmesi, modelin yeni ve görülmemiş veriler üzerinde başarılı olduğunu kanıtlar niteliktedir. Bu da, modelin yalnızca eğitim verisinde değil, gerçek dünya senaryolarına daha yakın olan test verisinde de öğrenme kapasitesini koruyabildiğini göstermektedir. %91.41 F1 skoru, modelin sınıflandırma görevlerinde yüksek hassasiyet ve doğrulukla çalıştığını, yanlış pozitif ve yanlış negatif oranlarını düşük tutarak dengeli bir performans sergilediğini işaret etmektedir. Eğitim süresi açısından Electra, 1 saat 6 dakika ile BERT'e oldukça yakın bir süre içerisinde eğitilmiştir.

Tablo 9.9 RAG V2 Sonuçları

Model	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Süre
RAG V2	Train	0.8716	0.8799	0.8954	0.8716	13:05
	Test	0.7602	0.7677	0.8023	0.7602	

Tablo 9.9'da görüldüğü üzere RAG modeli V2 için, diğer modern yöntemlerle kıyaslandığında daha düşük sonuçlar vermektedir. Eğitim aşamasında %87.16 doğruluk ve %87.99 F1 skoru elde ederek belirli bir başarı sağlasa da, test setinde

%76.02 doğruluk ve %76.77 F1 skoru ile beklenen performansı gösterememektedir. Ancak V1'deki sonuçlara göre başarısını arttırmıştır. RAG'nin eğitim süresi sadece 13 dakika 5 saniye sürmüştür, yani diğer modellere kıyasla çok daha hızlı bir eğitim sunmaktadır. Bu hız, modelin yapılandırılmasından ve FiT aşamasının daha az zaman almasıyla açıklanabilmektedir. RAG ile V2'nin eğitim süresi büyük bir avantaj sağlasada, test verisi üzerinde beklenen başarıyı yakalayamamıştır. Diğer modeller, daha uzun eğitim süreçleri gerektirirse, genelleme yeteneği ve doğruluk açısından çok daha iyi sonuçlar vermiştir. RAG'nin hızlı olmasının avantajı, belirli uygulamalarda etkili olabilir, ancak bu çalışmada diğer modern yöntemlerle kıyaslandığında daha düşük performans sergilediğini söylemek mümkündür. Farklı durumlarda hızlı bir sonuç almak isteniyorsa iyi ön işleme aşamalarından geçmiş bir veri seti ile RAG modeli de tercih edilebilir gözükmektedir.

Eğitim ve test verisi üzerindeki başarıların birbirine yakın olması, modellerin düzgün şekilde öğrenme süreci geçirdiğini ve her birinin genelleme yeteneğinin yüksek olduğunu kanıtlamaktadır. Ayrıca, her üç model için de çapraz doğrulama işlemleri yapılmış ve sonuçların tutarlı olduğu gözlemlenmiştir.

Tablo 9.10 BERT-ConvBert-Electra V2-100 Epoch Sonuçları

Model (V2)	Veri Seti	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Epoch	Süre
BERT	Train	0.9835	0.9912	0.9944	0.9888	100	02:11:43
	Test	0.9193	0.9023	0.9122	0.9055	100	
ConvBert	Train	0.9835	0.9915	0.9952	0.9886	100	02:21:18
	Test	0.9295	0.9218	0.9386	0.9222	100	
Electra	Train	0.9835	0.9915	0.9951	0.9886	100	02:08:08
	Test	0.9227	0.9114	0.9160	0.9199	100	

Tablo 9.10'da görüldüğü üzere V2 veri seti üzerinde model performansını arttırmak amacıyla 100 dönem sayısı ile yapılan denemelerde BERT modelinin V2 için 50 ve 100 dönem sayısı sonuçları kıyaslandığında, dönem sayısının 50'den 100'e çıkarılması eğitim süresini yaklaşık iki katına çıkarırken (%98.35 doğruluk, %99.16

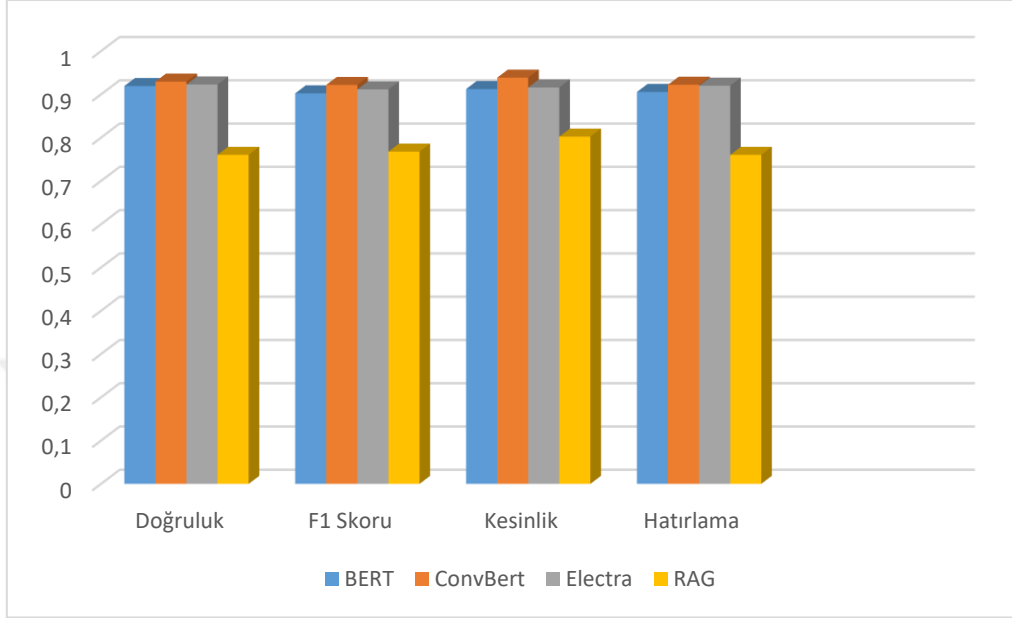
F1 skoru), test setinde hafif bir performans düşüşüne yol açmıştır (%92.38 doğruluk ve %90.59 F1 skorundan %91.93 doğruluk ve %90.23 F1 skoruna). Eğitim seti üzerinde sonuçlar benzer kalırken, test verisinde genelleme yeteneği zayıflamış ve Oft işaretleri görülmüştür. Bu nedenle, 50 dönem sayısı eğitimin BERT V2 için daha verimli olduğu ve 100 dönem sayısına çıkarılmasının zaman maliyetine değmediği sonucuna varılabilmektedir.

ConvBERT modelinin V2 için 50 ve 100 dönem sayısı sonuçlarını kıyasladığımızda, her iki durumda da eğitim setinde %98.35 doğruluk ve %99.16 F1 skoru ile başarılı bir performans sergilediği görülmektedir. Ancak, test setinde 50 dönem sayısında %92.50 doğruluk ve %91.41 F1 skoru elde eden model, 100 dönem sayısında %92.95 doğruluk ve %92.18 F1 skoru ile test verisi üzerinde daha da iyi sonuçlar vermiştir. Eğitim süresi 50 dönem sayısında 2 saat 2 dakika iken, 100 dönem sayısında 2 saat 21 dakikaya çıkarak en uzun süreyi almıştır. 100 dönem sayısı ile birlikte, ConvBERT V2, test setinde en iyi performansı sergileyen model haline gelmiştir.

Electra modelinin V2 için 50 ve 100 dönem sayısı sonuçlarını kıyasladığımızda, modelin her iki durumda da başarılı sonuçlar verdiği, ancak 100 dönem sayısında test performansında küçük bir iyileşme olduğu gözlemlenmektedir. 50 dönem sayısında %91.81 doğruluk ve %91.41 F1 skoru ile etkileyici bir genelleme yeteneği sergileyen Electra V2, 100 dönem sayısında bu değerleri %92.27 doğruluk ve %91.14 F1 skoruna çıkarmaktadır. Eğitim setindeki doğruluk ve F1 skorları 50 dönem sayısı ile neredeyse aynı (%98.35 doğruluk ve %99.15 F1), ancak 100 dönem sayısında eğitim süresi 1 saat 6 dakikadan 2 saat 8 dakikaya çıkmıştır. Bu ekstra süre, test setinde elde edilen küçük iyileşme ile karşılaştırıldığında zaman-performans dengesi açısından dikkatle değerlendirilmesi gereken bir durumdur. 100 dönem sayısı ile Electra V2, genelleme yeteneğini bir miktar daha iyileştirmiş olsa da, daha uzun eğitim süresiyle elde edilen kazançların sınırlı olduğu ve 50 dönem sayısının zaman verimliliği açısından daha avantajlı olabileceği söylenebilir.

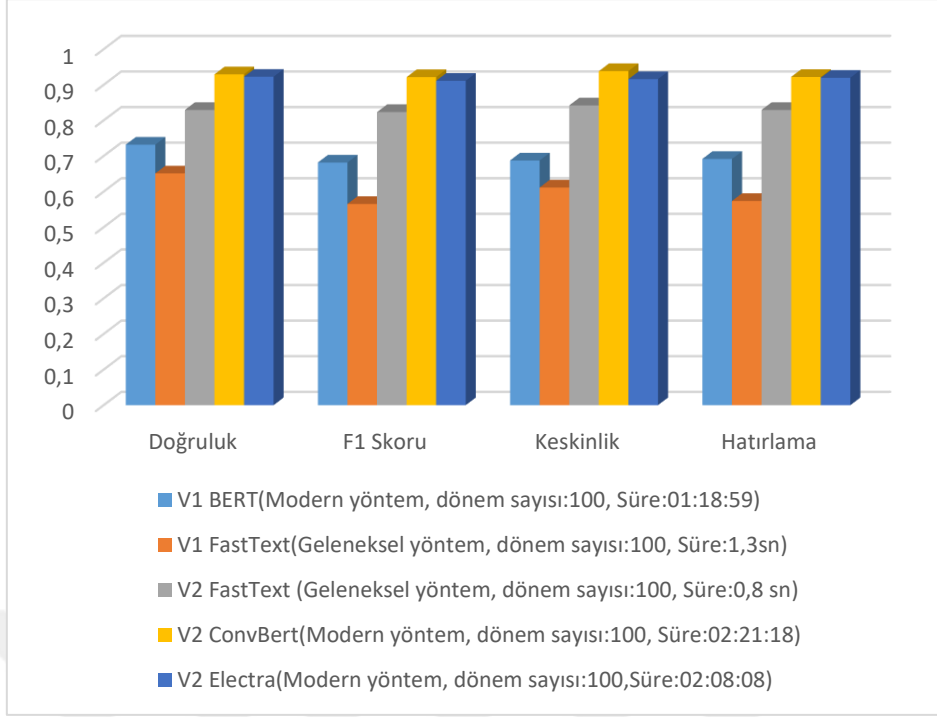
Tablo 9.10'da incelenen RAG modeli V2 için, diğer modellerden farklı olarak dönem sayısı kullanmadan çalışarak eğitim süresi çok kısa sadece 13 dakika 5 saniye sürmektedir. Eğitim setinde %87.16 doğruluk ve %87.99 F1 skoru elde

etmesine rağmen, test setinde %76.02 doğruluk ve %76.77 F1 skoru ile en düşük sonuçları vermektedir. Bu hızlı eğitim süresi, diğer modellerin uzun eğitim süreleriyle kıyaslandığında büyük bir avantaj sunsa da, test verisinde genelleme yeteneği açısından önemli bir performans eksikliği yaratmaktadır.



Şekil 9.7 V2 (100 dönem sayısı) Modern Yöntemlerin Sonuçları

100 dönem sayısı V2 sonuçlarına genel olarak baktığımızda, ConvBERT V2 test setinde %92.95 doğruluk ve %92.18 F1 skoru ile en iyi performansı sergileyen model olarak öne çıkmaktadır. Electra V2 de test setinde %92.27 doğruluk ve %91.14 F1 skoru ile oldukça güçlü bir performans sunmaktadır. Ancak eğitim süresi bakımından daha kısa ve dengeli bir model olarak dikkat çekmektedir. BERT V2, test setinde %91.93 doğruluk ve %90.23 F1 skoru ile hala yüksek bir performans sergilemesine rağmen, eğitim süresinin uzunluğu ve küçük iyileşmeler göz önüne alındığında daha verimli seçeneklerin mevcut olduğu görülmektedir. RAG ile V2, yalnızca 13 dakikalık eğitim süresiyle en hızlı model olsa da, test setinde %76.02 doğruluk ile diğer modellere göre belirgin şekilde geride kalmaktadır.



Şekil 9.8 En İyi Modellerin Test Performanslarının Karşılaştırılması

Sonuç olarak şekil 9.8’de görüldüğü üzere V2 veri seti, hem geleneksel hem de modern yöntemler için daha iyi sonuçlar sunmuş, özellikle modern yöntemlerin performansını önemli ölçüde artırmıştır. En yüksek test performansı ise ConvBERT V2 tarafından sağlanmaktadır ancak Eğitim süresi ve test performansı dengesi açısından Electra V2 en verimli seçenek olarak öne çıkmaktadır.

10

SONUÇ

Bu çalışmanın sonuçları, hem V1 hem de V2 veri setleri (bkz. Bölüm 8) üzerinde yapılan geleneksel ve modern yöntemlerle elde edilen model performanslarını derinlemesine analiz etmektedir. V1 veri seti, yeterli ön işleme ve kategorilendirme yapılmamış ham veriyle çalışıldığı için, hem geleneksel hem de modern yöntemler açısından beklenen başarıya ulaşamamıştır.

Özellikle geleneksel yöntemler açısından değerlendirildiğinde, V1 veri setinde FastText, %65.12 doğruluk ve %56.54 F1 skoru ile diğer geleneksel yöntemlere kıyasla daha başarılı bir performans sergilemiştir. Ancak, Glove ve Word2Vec yöntemleri sırasıyla %40.67 ve %33.41 doğruluk oranları elde ederek genelleme yeteneklerinin zayıf olduğunu göstermiştir. Bu düşük performans, veri setindeki eksikliklerden kaynaklanmış ve modellerin test verisi üzerinde yeterince iyi sonuç verememesine yol açmıştır.

Modern yöntemler açısından V1 veri setinde, BERT, ConvBERT ve Electra gibi modellerin, geleneksel yöntemlere kıyasla daha başarılı sonuçlar verdiği gözlemlenmiştir. BERT, %72.97 doğruluk ve %67.97 F1 skoru ile modern yöntemler arasında en başarılı model olarak öne çıkmıştır. ConvBERT ve Electra ise %72.42 ve %71.65 doğruluk oranlarıyla BERT'e yakın performans sergilemiştir. Bu sonuçlar, modern yöntemlerin geleneksel yöntemlere kıyasla daha gelişmiş bir yapıya sahip olduğunu ve dönem sayısının artırılmasıyla daha iyi sonuçlar verebildiğini göstermektedir. Özellikle BERT ve ConvBERT modelleri, 100 dönem sayısı ile eğitildiğinde, 50 sayısına göre daha iyi genelleme yetenekleri göstermiştir. BERT, 100 dönem sayısı sonrasında %73.20 doğruluk ve %68.18 F1 skoruna ulaşırken, ConvBERT %72.97 doğruluk ve %66.63 F1 skoru elde etmiştir. Ancak, bu modellerin eğitim sürelerinin 2 saat 14 dakikaya kadar çıkması, daha uzun eğitim süreçlerinin performansı artırdığını fakat zaman açısından maliyetli olduğunu ortaya koymuştur. RAG modeli ise V1 veri seti üzerinde yapılan denemelerde eğitim süreci toplamda 10 dakika 5 saniye sürmüş olmasına rağmen, test verisi üzerinde elde edilen sonuçlar beklenenin altında kalmıştır. Model, eğitim

verisinde %73.51 doğruluk, %69.52 F1 skoru, %75.14 kesinlik ve %73.51 hatırlama elde etmiştir. Ancak, test verisi üzerindeki sonuçlar daha düşük olup, %60.91 doğruluk, %56.26 F1 skoru, %56.51 kesinlik ve %60.90 hatırlama ile performans açısından oldukça geride kalmıştır.

V2 veri setiyle yapılan çalışmalar, verilerin ön işleme aşamasından geçip daha doğru bir şekilde kategorilendirilmesi ile model performanslarında büyük iyileşmeler sağlandığını göstermektedir. V2 veri seti için; FastText, %82.61 doğruluk ve %77.49 F1 skoru ile geleneksel yöntemler arasında en iyi sonucu vermiştir. Bu, V1'e kıyasla çok büyük bir iyileşme olup, veri setinin işlenmiş olmasının model başarısını nasıl doğrudan etkilediğini net bir şekilde göstermektedir. Glove ve Word2Vec ise sırasıyla %52.04 ve %50.05 doğruluk oranlarıyla, geleneksel yöntemlerde FastText'in gerisinde kalmış, ancak V1 sonuçlarına göre daha iyi performans sergilemiştir. Geleneksel yöntemlerde dönem sayısının artırılması, özellikle FastText V2'de %82.95 doğruluk ve %82.36 F1 skoru ile daha başarılı sonuçlar elde edilmesine olanak sağlamıştır. Ancak, Glove ve Word2Vec modelleri için 100 epoch artışı anlamlı bir iyileşme sağlamamış, bu modellerin test verisi üzerindeki performansı %52 doğruluk civarında kalmıştır. Modern yöntemler için V2 veri setiyle yapılan deneyler bu modellerin V1'e göre çok daha iyi sonuçlar verdiğini göstermektedir. V2 veri setinde; BERT, %92.38 doğruluk ve %90.59 F1 skoru ile V1'deki sonuçlarından çok daha iyi bir performans sergilemiştir. Bu, modelin veri seti işlenmiş olduğunda ne kadar güçlü bir genelleme yeteneği gösterebildiğini açıkça ortaya koymaktadır. ConvBERT V2, %92.50 doğruluk ve %91.41 F1 skoru ile BERT'e çok yakın performans göstermiştir, hatta daha fazla dönem sayısı ile eğitildiğinde test setinde %92.95 doğruluk ve %92.18 F1 skoru ile en yüksek başarıyı elde etmiştir. Electra da bu iki modele oldukça yakın sonuçlar elde etmiş, %91.81 doğruluk ve %91.14 F1 skoru ile güçlü bir genelleme yeteneği sergilemiştir.

Dönem sayısının artırılması, modern yöntemlerde anlamlı bir fark yaratmıştır. V2 veri setinde; BERT, ConvBERT ve Electra modelleri, 100 dönem sayısı sonrasında daha yüksek doğruluk ve F1 skorlarıyla performanslarını geliştirmiştir. Ancak, bu iyileşme eğitim sürelerinde ciddi bir artışa yol açmıştır. Örneğin, ConvBERT, 100 dönem sayısı sonrasında 2 saat 21 dakika eğitim süresiyle en uzun sürede eğitilen

model olmuştur, ancak test setindeki %92.95 doğruluk oranı ile bu sürenin karşılığını vermiştir. Electra V2, 100 dönem sayısı sonrası %92.27 doğruluk ile BERT ve ConvBERT modellerine oldukça yakın sonuçlar elde etmiş, eğitim süresi açısından dengeli bir model olarak öne çıkmıştır.

RAG modeli ise V2 veri setinde modern yöntemler arasında en kısa eğitim süresine sahip olmasına rağmen, test verisi üzerinde %76.02 doğruluk ve %76.77 F1 skoru ile diğer modern modellerin gerisinde kalmıştır. RAG'ın eğitimi sadece 13 dakika sürmüştür, ancak test verisi üzerinde istenen başarıyı sağlayamamıştır. Bu da, modern yöntemlerde daha uzun eğitim sürelerinin, özellikle derin öğrenme modelleri söz konusu olduğunda, performansı önemli ölçüde artırdığını ortaya koymaktadır. Eğitim süresi ve performans dengesine bakıldığında, Electra genel verimlilik açısından en avantajlı model olarak dikkat çekmiştir ancak en yüksek doğruluk ve F1 skorlarını ise ConvBERT sağlamıştır.

Tez çalışmasının bulguları ışığında, gelecekteki araştırmalara yönelik birkaç öneri öne çıkmaktadır. Öncelikle, veri setinin doğru şekilde işlenmesi ve kategorilere ayrılması, model performansının artırılması için temel bir adımdır. Bu süreç, veri temizleme, normalizasyon, tokenizasyon gibi detaylı ön işleme adımlarını içerir. Ayrıca, dönem sayısının dikkatli bir şekilde optimize edilmesi de büyük önem taşımaktadır. Her modelin öğrenme kapasitesi ve veri setine uyum yeteneği farklıdır; bu nedenle, dönem sayısının modelin gereksinimlerine göre belirlenmesi, Oft ya da Uft durumlarının önüne geçmek açısından kritik olacaktır. Ayrıca bu tez kapsamında literatüre kazandırılmış farklı bir karşılaştırma yöntemi olan RAG sonuçlarının da eğitim süresi açısından FiT yapılan modellere oranla başarısız kalmadığı ve sahip olduğu genelleme yeteneğiyle kullanılabilir bir yöntem olabileceği gözükmemektedir.

Sonuç olarak, bazı modeller daha kısa eğitim süreleriyle yüksek doğruluk oranlarına ulaşırken, bazıları ise daha uzun sürede eğitilmesine rağmen daha güçlü bir genelleme yeteneği sunabilmektedir. Bu denge dikkate alındığında, farklı veri setleri ve uygulama senaryoları için en verimli modellerin belirlenmesi, gerçek dünya uygulamalarında önemli bir avantaj sağlayacaktır. Gelecekte, model yapılandırmalarının daha da geliştirilmesi ve daha büyük veri setleri üzerinde test edilmesi, yapay zeka tabanlı sistemlerin başarısını artırmaya devam edecektir.



Veri seti paylaşılmayacaktır.

KAYNAKÇA

- [1] S. J. Russell, P. Norvig, Artificial intelligence: a modern approach. Pearson, 2016.
- [2] Google Cloud, <https://cloud.google.com/learn/what-is-artificial-intelligence>, 03.03.2025.
- [3] A. Ethem, Introduction to machine learning. MIT Press, 2020.
- [4] N. Barbara, M. Bianca, C. Krzysztof, “The impact of artificial intelligence on event experiences: a scenario technique approach,” *Electronic Markets*, vol. 31, no. 3, pp. 601–617, 2021.
- [5] B. Jason, Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End. Machine Learning Mastery, 2016.
- [6] T. J. Sejnowski, The Deep Learning Revolution. MIT Press, 2018..
- [7] M. Tom, “Does machine learning really work?,” *AI Magazine*, vol. 18, no. 3, pp. 11–11, 1997.
- [8] IBM Machine Learning, <https://www.ibm.com/topics/machine-learning>, 03.03.2025.
- [9] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Reilly Media, 2019.
- [10] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, “Deep reinforcement learning from human preferences,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [12] Amazon Web Services, <https://aws.amazon.com/tr/what-is/nlp>, 03.03.2025.
- [13] T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, and S. Bisen, “Implementation of a chatbot system using AI and NLP,” *International Journal of Innovative Research in Computer Science & Technology*, vol. 6, no. 3, 2018.

- [14] D. Yeke, "BERT Word Embeddings'i Temel Alan Sorgu Geniřletme ile Belge Sıralamasını Geliřtirme," Yüksek Lisans Tezi, Orta Doęu Teknik Üniversitesi, 2020.
- [15] D. Kınık, "TF-IDF ve Doc2Vec Tabanlı Metin Sınıflandırma Sisteminin Başarım Deęerinin Ardışık Kelime Grubu Tespiti ile Arttırılması," Yüksek Lisans Tezi, Doęuş Üniversitesi, 2020.
- [16] ř. Yılmaz, "Word2Vec Temsillerini Kullanarak Türkçe'de Soru Sınıflandırmasında Derin Öğrenme Analizi," Doktora Tezi, Düzce Üniversitesi, 2020.
- [17] B. Özkan, "Dialog Intent Classification Using NLP Methods," Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, 2021.
- [18] A. Özberk, "BERT Modelleriyle Türkçe Twitter Verilerinde Saldırgan Dil Tespiti," Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2022.
- [19] A. Özberk, "BERT Modelleriyle Türkçe Twitter Verilerinde Saldırgan Dil Tespiti," Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2022.
- [20] H. Alfigi, "Multi-Label and Single-Label Text Classification Using Standard Machine Learning Algorithms and Pre-Trained BERT Transformer," Yüksek Lisans Tezi, Çankaya Üniversitesi, 2023.
- [21] S. Erkan, "E-ticaret sitelerinde yer alan yorumların BERT ve ALBERT dil modelleri ile analizi," Yüksek Lisans Tezi, İstanbul Beykent Üniversitesi, 2023.
- [22] A. Doęan, "Ortaokul ve Lise Ders Sorularının BERTurk Dönüřtürücü Modeli ile Sınıflandırılması," Yüksek Lisans Tezi, Kocaeli Üniversitesi, 2023.

- [23] H. H. Sarıççek, “Bilimsel Materyaller İçin Soru-Cevap Sohbet Robotu Tasarımı ve Uygulaması Geliştirilmesi,” Yüksek Lisans Tezi, Ankara Üniversitesi, 2023.
- [24] E. H. Yılmaz, “Yazılım Hatalarında Otomatik Öncelik Tespiti: Arttırılmış Verimlilik İçin Karşılaştırmalı Öğrenme, Büyük Dil Modelleri ve Vektör Veritabanları ile Transformatör Tabanlı Kodlayıcılar Üzerine Kapsamlı Bir Çalışma,” Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, 2024.
- [25] A. Er, “Büyük Dil Modelleri Kullanılarak Anahtar Kelime Üretimi ve Konu Modelleme: Müşteri Geri Bildirimlerinden İçgörü Edinimi,” Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, 2024.
- [26] A. Souha, C. Ouaddi, L. Benaddi, and A. Jakimi, “Pre-trained models for intent classification in chatbot: Comparative study and critical analysis,” in Proceedings of the 2023 6th International Conference on Advanced Communication Technologies and Networking (CommNet), IEEE, 2023, pp. 1–6.
- [27] H. K. Chaubey, G. Tripathi, and R. Ranjan, “Comparative analysis of RAG, fine-tuning, and prompt engineering in chatbot development,” in Proceedings of the 2024 International Conference on Future Technologies for Smart Society (ICFTSS), IEEE, 2024, pp. 169–172.
- [28] A. F. Adoma, N. M. Henry, and W. Chen, “Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition,” in Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), IEEE, 2020, pp. 117–121.
- [29] Q. Zhong, L. Ding, J. Liu, B. Du, and D. Tao, “Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT,” arXiv preprint, arXiv:2302.10198, 2023.
- [30] L. Gomes, R. da Silva Torres, and M. L. Côrtes, “BERT-and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: A comparative study,” Information and Software Technology, vol. 160, p. 107217, 2023.

- [31] C. A. A. Silitonga, M. D. Dermawan, F. Adeta, and N. Nadia, “Comparative study of BERT-CNN, TRANS-BLSTM, and RoBERTa models for sentiment analysis,” in Proceedings of the 2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, 2024, pp. 358–363.
- [32] A. D. Susanto, S. A. Pradita, C. Stryadhi, K. E. Setiawan, M. F. Hasani, “Text vectorization techniques for trending topic clustering on Twitter: A comparative evaluation of TF-IDF, Doc2Vec, and Sentence-BERT,” in Proceedings of the 2023 5th International Conference on Cybernetics and Intelligent System (ICORIS), IEEE, 2023, pp. 1–7.
- [33] Bloomfire, <https://bloomfire.com/blog/data-vs-information/> , 03.03.2025.
- [34] DataCamp, <https://www.datacamp.com/blog/data-demystified-what-exactly-is-data>, 03.03.2025.
- [35] GeeksforGeeks, <https://www.geeksforgeeks.org/what-is-data/> , 03.03.2025.
- [36] Simplilearn, <https://www.simplilearn.com/what-is-data-article>, 03.03.2025
- [37] GeeksforGeeks, <https://www.geeksforgeeks.org/what-is-dataset/>, 03.03.2025
- [38] A. Oğuzlar, “Veri Ön İşleme,” Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, no. 21, pp. 67–76, 2003.
- [39] TechTarget, <https://www.techtarget.com/searchdatamanagement/definition/data-quality>, 03.03.2025.
- [40] S. E. Şeker, <https://youtu.be/iButBaRPPBM?si=mY-BI2x6yAOpic2a>, 03.03.2025.
- [41] W. Kim, B. Choi, E. Hong, S. Kim, and D. Lee, “A taxonomy of dirty data,” Data Mining and Knowledge Discovery, vol. 7, pp. 81–99, 2003.
- [42] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.
- [43] Keepnet Labs, <https://keepnetlabs.com/blog/what-is-data-classification> , 03.03.2025.

- [44] IBM Security, <https://nationalcioreview.com/articles-insights/information-security/special-report-the-cost-of-a-data-breach-in-2023/> , 03.03.2025.
- [45] S. Vijayarani and A. Tamilarasi, “An efficient masking technique for sensitive data protection,” in Proceedings of the 2011 International Conference on Recent Trends in Information Technology (ICRTIT), IEEE, 2011, pp. 1245–1249.
- [46] H. Min, M. G. M. D. A. Ferreira, and C. Cardie, “Question answering with knowledge base,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017.
- [47] A. Severyn and A. Moschitti, “Twitter sentiment analysis with deep convolutional neural networks,” in Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, 2015.
- [48] R. Sennrich, B. Haddow, and A. Birch, “Neural machine translation of rare words with subword units,” arXiv preprint, arXiv:1508.07909, 2015.
- [49] A. Subramaniam, M. Chatterjee, and A. Mittal, “Deep neural networks with inexact matching for person re-identification,” Advances in Neural Information Processing Systems, vol. 29, 2016.
- [50] A. M. Dai and Q. V. Le, “Semi-supervised sequence learning,” Advances in Neural Information Processing Systems, vol. 28, 2015.
- [51] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” arXiv preprint, arXiv:1801.06146, 2018.
- [52] Elite Data Science, <https://elitedatascience.com/overfitting-in-machine-learning> , 03.03.2025.
- [53] P. G. Paris, D. Robilliard, and C. Fonlupt, “Exploring overfitting in genetic programming,” in Proceedings of the Artificial Evolution, International Conference, Evolution Artificielle (EA 2003), DBLP, 2004, pp. 267–277.
- [54] D. D. Jensen and P. R. Cohen, “Multiple comparisons in induction algorithms,” Machine Learning, vol. 38, no. 3, pp. 309–338, 2000.

- [55] M. Emeç and M. H. Özcanhan, “Veri ön işleme ve öznitelik mühendisliğinin yapay zekâ yöntemlerine uygulanması,” in *Mühendislikte Öncü ve Çağdaş Çalışmalar*, pp. 33–54, 2023.
- [56] C. Karakuş, *Makine Öğrenmesi Algoritmaları*. Balbiti, 2021.
- [57] D. Zilyas and A. Yılmaz, “Makine öğrenmesi yöntemleri ile eğitim başarısının tahmini modeli,” *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, vol. 14, no. 3, pp. 437–447, 2023.
- [58] A. Isa, W. Hamzah, M. K. Yusof, I. Ismail, M. Makhtar, “Multi-Label Intent Classification for Educational Chatbot: A Comparative Study Using Problem Transformation, Adapted Algorithm and Ensemble Method,” *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 4, 2024.
- [59] N. A. Sajid, A. Rahman, M. Ahmad, D. Musleh, M. I. Basheer Ahmed, R. Alassaf, S. Chabani, “Single vs. Multi-label: The Issues, Challenges and Insights of Contemporary Classification Schemes,” *Applied Sciences*, vol. 13, no. 11, p. 6804, 2023.
- [60] F. Hengst, R. Wolter, P. Altmeyer, A. Kaygan, “Conformal Intent Classification and Clarification for Fast and Accurate Intent Recognition,” *arXiv preprint*, arXiv:2403.18973, 2024.
- [61] M. Sokolova and J. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, no. 5, pp. 427–444, 2009.
- [62] E. İlge, <https://ilge.com.tr/makine-ogrenmesinde-basariyi-tartmak--temel-metrikler-ve-anlamlari> , 03.03.2025.
- [63] Veri Manisa Celal Bayar Üniversitesi, <https://veri.mcbu.edu.tr/article-detail/12/> , 03.03.2025.
- [64] K. W. Church, “Word2Vec,” *Natural Language Engineering*, vol. 23, no. 1, pp. 155–162, 2017.
- [65] G. Di Gennaro, A. Buonanno, and F. A. N. Palmieri, “Considerations about learning Word2Vec,” *The Journal of Supercomputing*, pp. 1–16, 2021.

- [66] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.
- [67] A. K. Nandanwar and J. Choudhary, “Semantic features with contextual knowledge-based web page categorization using the GloVe model and stacked BiLSTM,” *Symmetry*, vol. 13, no. 10, p. 177, 2021.
- [68] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” arXiv preprint, arXiv:1607.01759, 2016.
- [69] S. A. Sazan, M. H. Miraz, and A. B. M. Rahman, “Enhancing depressive post detection in Bangla: A comparative study of TF-IDF, BERT and FastText embeddings,” arXiv preprint, arXiv:2407.09187, 2024.
- [70] T. Yao, Z. Zhai, and B. Gao, “Text classification model based on FastText,” in Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS), IEEE, 2020, pp. 154–157.
- [71] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long and Short Papers), 2019, pp. 4171–4186.
- [72] Github, <https://github.com/stefan-it/turkish-bert> , 03.03.2025.
- [73] Hugging Face, <https://huggingface.co/dbmdz/bert-base-turkish-cased> , 03.03.2025.
- [74] Z. H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, “ConvBERT: Improving BERT with span-based dynamic convolution,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 12837–12848, 2020.
- [75] ConvBERTurk, <https://huggingface.co/dbmdz/convbert-base-turkish-cased> , 03.03.2025.
- [76] K. Clark, M. T. Luong, Q. V. Le, and C. D. Manning, “ELECTRA: Pre-training text encoders as discriminators rather than generators,” arXiv preprint, arXiv:2003.10555, 2020.

- [77] Hugging Face, <https://huggingface.co/dbmdz/electra-base-turkish-cased-discriminator> , 03.03.2025.
- [78] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang, “Retrieval-augmented generation for large language models: A survey,” arXiv preprint, arXiv:2312.10997, 2023.
- [79] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application,” *International Journal of Advances in Soft Computing & Its Applications*, vol. 13, no. 3, 2021.
- [80] C. Fiesler, N. Beard, and B. C. Keegan, “No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, no. 1, pp. 187–196, 2020.
- [81] V. Krotov and L. Silva, “Legality and ethics of web scraping,” in *Proceedings of the Twenty-Fourth Americas Conference on Information Systems*, New Orleans, 2018.
- [82] Crummy, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> , 03.03.2025
- [83] LangChain, https://python.langchain.com/docs/integrations/document_transformers/beautiful_soup/ , 03.03.2025.
- [84] Scrapy, <https://docs.scrapy.org/en/latest/> , 03.03.2025.
- [85] Selenium Documentation, <https://www.selenium.dev/documentation/en/>, 03.03.2025
- [86] LangChain, <https://python.langchain.com/docs/tutorials/rag/> , 03.03.2025
- [87] Türkiye Cumhuriyeti Resmi Gazete, <https://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf> , 03.03.2025.
- [88] Türkiye Cumhuriyeti Resmi Gazete, <https://www.mevzuat.gov.tr/mevzuatmetin/1.3.5846.pdf> , 03.03.2025.
- [89] H. Saif, M. Fernandez, Y. He, and H. Alani, “On stopwords, filtering and data sparsity for sentiment analysis of Twitter,” in *Proceedings of the Ninth*

International Conference on Language Resources and Evaluation (LREC'14),
May 2014, pp. 810–817.

- [90] Towards Data Science, <https://towardsdatascience.com/evaluating-train-test-split-strategies-in-machine-learning-beyond-the-basics-c3e84b58ddce>, 03.03.2025.
- [91] Scikit-learn, https://scikit-learn.org/stable/modules/cross_validation.html#train-test-split ,03.03.2025.
- [92] Wikipedia, https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets , 03.03.2025.
- [93] Pandas, <https://pandas.pydata.org/docs/> , 03.03.2025.
- [94] Hugging Face, <https://huggingface.co/docs/datasets/> , 03.03.2025.
- [95] Scikit-learn, <https://scikit-learn.org/stable/documentation.html> , 03.03.2025.
- [96] Python, <https://docs.python.org/3/library/string.html> , 03.03.2025.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. Çetin, K. R., & Akgün, F. A. (2025, Mayıs). Comparison of traditional and modern approaches on chatbot dataset. In Proceedings of the 4th International Congress on Engineering and Natural Sciences Studies (ICENSS) (p. 104). ISBN 978-625-372-169-5

