
GAZİ ÜNİVERSİTESİ
Gazi
1926

AVRUPA İNSAN HAKLARI MAHKEMESİ

YARGI KARARLARININ TAHMİN EDİLMESİ

Nagihan ÜNAL

YÜKSEK LİSANS TEZİ
ADLİ BİLİŞİM ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

ŞUBAT 2025

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Nagihan ÜNAL

...../……./……

AVRUPA İNSAN HAKLARI MAHKEMESİ YARGI KARARLARININ TAHMİN EDİLMESİ

(Yüksek Lisans Tezi)

Nagihan ÜNAL

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Şubat 2025

ÖZET

Son yıllarda makine öğrenmesi, doğal dil işleme ve büyük veri analitiği gibi teknolojilerdeki ilerlemeler, hukuki süreçlerin otomasyonuna yönelik çeşitli uygulamaların geliştirilmesini mümkün kılmıştır. Avrupa İnsan Hakları Mahkemesi kararlarının tahmin edilmesi, hukuki analizin derinleştirilmesi ve karar alma süreçlerinin hızlandırılması açısından önem arz etmektedir. Hukuki metinlerin kendine özgü terminolojisi ve karmaşık yapısı, bu metinlerin sistematik olarak işlenmesini zorunlu kılmaktadır. Bu bağlamda, doğal dil işleme teknikleri, hukuki belgelerin analiz edilmesini, sınıflandırılmasını ve anlamlandırılmasını sağlayarak yargı süreçlerine katkıda bulunmaktadır. Bu çalışmada, Avrupa İnsan Hakları Mahkemesi'nin Avrupa İnsan Hakları Sözleşmesi'nin çeşitli maddeleri çerçevesinde verdiği ihlal kararlarının tahmin edilmesine yönelik modeller geliştirilmiştir. Tahmin sürecinde, Avrupa İnsan Hakları Mahkemesi'nin resmi veri tabanıdan elde edilen dava metinleri kullanılarak bir veri seti oluşturulmuştur. Çalışmada, metin ön işleme teknikleri uygulanmış ve çeşitli makine öğrenmesi algoritmaları ile derin öğrenme modelleri kıyaslanmıştır. Veri setindeki sınıf dengesizliği problemi, dengelenmiş veri kümeleri ve veri artırma yöntemleriyle giderilmiş, bu doğrultuda azınlık sınıfına ait örneklerin temsil gücü artırılmıştır. Veri artırma sürecinde, geri çeviri yöntemi kullanılarak "ihlal" ve "ihlal değil" sınıflarına ait örnekler çoğaltılmış ve modellerin genelleme kapasitesi iyileştirilmiştir. Deneysel çalışmalar kapsamında, geleneksel makine öğrenmesi algoritmalarından Lojistik Regresyon, Multinomial Naive Bayes ve Destek Vektör Makineleri gibi yöntemlerin yanı sıra, hukuki alanda özelleşmiş derin öğrenme modeli de kullanılmıştır. Yapılan performans analizleri çeşitli metrikler üzerinden değerlendirilmiş olup, LegalBERT modelinin Avrupa İnsan Hakları Mahkemesi kararlarının sınıflandırılmasında %94 doğruluk ve %94 F1 skoru ile en yüksek başarı oranını sunduğu gözlemlenmiştir. Bu bulgular, doğal dil işleme ve makine öğrenmesi tekniklerinin hukuki metin analizi ve sınıflandırması için güçlü araçlar sunduğunu ortaya koymaktadır. Bu tez çalışması, hukuk alanında veri odaklı karar destek sistemlerinin geliştirilmesine yönelik önemli katkılar sunmaktadır. Özellikle, hukuki alan için özel olarak eğitilmiş derin öğrenme modellerinin, yargı süreçlerindeki belirsizlikleri azaltmada ve öngörülebilirliği artırmada kritik bir rol oynadığı görülmektedir. Bu doğrultuda, hukuki karar tahminine yönelik yapılan bu çalışma hem akademik hem de uygulamalı çalışmalar için yol gösterici nitelikte olup, gelecekte hukuki süreçlerin daha şeffaf ve erişilebilir hale gelmesine katkı sunabilecek bir temel oluşturmaktadır.

Bilim Kodu : 92444
Anahtar Kelimeler : Avrupa İnsan Hakları Mahkemesi, doğal dil işleme, hukuki karar tahmini, makine öğrenmesi
Sayfa Adedi : 76
Danışman : Doç. Dr. Oktay YILDIZ

PREDICTING THE JUDGMENTS OF THE EUROPEAN COURT OF HUMAN RIGHTS

(M. Sc. Thesis)

Nagihan ÜNAL

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

February 2025

ABSTRACT

In recent years, advancements in technologies such as machine learning, natural language processing, and big data analytics have enabled the development of various applications aimed at automating legal processes. Predicting decisions of the European Court of Human Rights is crucial for enhancing legal analysis and accelerating decision-making processes. The unique terminology and complex structure of legal texts necessitate their systematic processing. In this context, natural language processing techniques facilitate the analysis, classification, and interpretation of legal documents, thereby contributing to judicial procedures. In this study, models were developed to predict violation decisions issued by the European Court of Human Rights under various articles of the European Convention on Human Rights. During the prediction process, a dataset was created using case texts obtained from the official European Court of Human Rights database. Text preprocessing techniques were applied, and various machine learning algorithms were compared with deep learning models. The issue of class imbalance in the dataset was addressed through balanced datasets and data augmentation techniques, thereby improving the representation of minority class examples. The data augmentation process involved back-translation to generate additional examples for both the “violation” and “non-violation” classes, enhancing the generalization capacity of the models. In experimental phase, traditional machine learning algorithms such as Logistic Regression, Multinomial Naive Bayes, and Support Vector Machines were used alongside a domain-specific deep learning model. Performance analyses were conducted using various evaluation metrics, revealing that the LegalBERT model achieved the highest accuracy and effectiveness in classifying European Court of Human decisions, with 94% accuracy and a 94% F1 score. These findings demonstrate that natural language processing and machine learning techniques provide powerful tools for legal text analysis and classification. This thesis makes significant contributions to the development of data-driven decision support systems in the legal domain. In particular, domain-specific deep learning models play a critical role in reducing uncertainty and increasing predictability in judicial processes. Accordingly, this study on legal decision prediction serves as a valuable reference for both academic research and practical applications. Moreover, it establishes a foundation for future developments that could contribute to making legal processes more transparent and accessible.

Science Code : 92444
Key Words : European Court of Human Rights, natural language processing, legal decision prediction, machine learning
Page Number : 76
Supervisor : Assoc. Prof. Dr. Oktay YILDIZ

TEŞEKKÜR

Yüksek lisans eğitimim boyunca değerli rehberliği, sabrı ve sürekli desteğinden dolayı danışmanım Sayın Doç. Dr. Oktay YILDIZ hocama şükranlarımı sunarım.

Bu araştırma sürecinde benden desteğini esirgemeyen ve beni sürekli cesaretlendiren aileme de minnettarım. Bana gösterdikleri sabır için teşekkür ederim. Onların varlığı ve desteği, bu sürecin her aşamasında bana çok büyük motivasyon kaynağı olmuştur.

İÇİNDEKİLER

Sayfa

ÖZET……... İV

ABSTRACT…… Vİ

TEŞEKKÜR…… Vİİİ

İÇİNDEKİLER……. İX

ÇİZELGELERİN LİSTESİ……. İV

ŞEKİLLERİN LİSTESİ…… XV

SİMGELER VE KISALTMALAR……. XVİ

1.GİRİŞ……..... 1

2. LİTERATÜR ARAŞTIRMASI……..... 7

2.1. Hukuki Karar Tahmini Alanında Yapılan Çalışmalar……. 7

2.2. Hukuki Metin Veri Kümeleri…….. 11

3. AVRUPA İNSAN HAKLARI MAHKEMESI İŞLEYIŞI VE AİHS İLE KORUNAN HAKLAR..... 13

4. MATERYAL VE YÖNTEM..... 17

4.1. HUDOC Veri Tabanı..... 17

4.2. Verinin Hazırlanması (Metin Önişleme)..... 20

4.3. Metin Temsili..... 22

4.3.1. Terim Frekansı-Ters Doküman Frekansı (TF-IDF)..... 23

4.3.2. N-Grams..... 23

4.3.3. Word2vec..... 24

4.3.4. GloVe..... 25

4.4. Sınıflandırma Algoritmaları..... 26

4.4.1. Destek Vektör Makineleri..... 26

4.4.2. Multinomial Naive Bayes..... 26

4.4.3. Lojistik Regresyon..... 27
4.4.4. XGBoost..... 28
4.4.5. Yinelemeli Sinir Ağı (Recurrent Neural Networks-RNN)..... 29
4.4.6. Dönüştürücü Modeller (Transformers)..... 32
4.5. Hiperparametre Optimizasyonu..... 34
4.6. Veri Seti Dengeleme Teknikleri..... 35
4.6.1. Veri Artırma Teknikleri..... 35
4.6.2. Aşağı Örnekleme (Downsampling)..... 37
4.7. Model Değerlendirme..... 37
4.7.1. Performans Metrikleri..... 37
4.7.2. K-Kat Çapraz Geçerleme..... 39
5. DENEYSEL ÇALIŞMA VE SONUÇLARI..... 411
5.1. Veri Kümesinin Hazırlanması..... 41
5.2. Dava Sonuçlarının Tahmin Edilmesi..... 45
5.2.1. Deneysel Çalışma 1..... 46
5.2.2. Deneysel Çalışma 2..... 51
5.2.3. Deneysel Çalışma 3..... 55
6. SONUÇ..... 577
KAYNAKLAR..... 611
EKLER..... 66
EK-1. Grafikler ve Karmaşıklık Matrisleri (LSTM-Madde 6 (Artırılmış veri))..... 67
EK-2. Grafikler ve Karmaşıklık Matrisleri (LegalBERT-Madde 6 (Orijinal Dengeli Veri))..... 70
EK-3. Grafikler ve Karmaşıklık Matrisleri (XLNet-Madde 6(Orijinal Dengeli Veri))..... 72
EK-4. Grafikler ve Karmaşıklık Matrisleri (BERT-Madde 6(Orijinal Dengeli Veri))..... 74

ÖZGEÇMİŞ..... 76

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 4.1. HUDOC veri tabanından çekilen dava sayıları..... 19

Çizelge 4.2. Orijinal ve önişleme adımlarından geçirilmiş veri örneği..... 22

Çizelge 4.3. LegalBERT eğitim korpusu..... 34

Çizelge 5.1. Dengelenmiş veri seti doküman sayıları..... 46

Çizelge 5.2. Dengelenmiş veri SVM sonuçları..... 47

Çizelge 5.3. Dengelenmiş veri Lojistik Regresyon sonuçları..... 47

Çizelge 5.4. Dengelenmiş veri XGBoost sonuçları..... 48

Çizelge 5.5. Dengelenmiş veri Multinomial Naive Bayes sonuçları..... 48

Çizelge 5.6. Dengelenmiş veri LSTM sonuçları..... 50

Çizelge 5.7. Artırılmış veri doküman sayıları..... 52

Çizelge 5.8. Artırma yapılmış veride SVM sonuçları..... 53

Çizelge 5.9. Artırma yapılmış veride Lojistik Regresyon sonuçları..... 53

Çizelge 5.10. Artırma yapılmış veride XGBoost sonuçları..... 54

Çizelge 5.11. Artırma yapılmış veride Multinomial Naive Bayes sonuçları..... 54

Çizelge 5.12. Artırma yapılmış veride LSTM sonuçları..... 55

Çizelge 5.13. Dengelenmiş veride Dönüştürücü modellerin sonuçları..... 56

Çizelge 6.1. Literatür ve bu çalışmanın karşılaştırılması..... 59

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 4.1. HUDOC veri tabanından veri çekme süreci..... 17

Şekil 4.2. Metin önişleme adımları..... 20

Şekil 4.3. N-grams..... 24

Şekil 4.4. Word2Vec Mimarisi..... 25

Şekil 4.5. Destek Vektör Makineleri..... 26

Şekil 4.6. Sigmoid fonksiyonunun grafiği..... 28

Şekil 4.7. XGBoost algoritmasının çalışma prensibi..... 29

Şekil 4.8. LSTM katmanları..... 31

Şekil 4.9. Transformer mimarisi..... 32

Şekil 4.10. Metin verisi artırma teknikleri..... 36

Şekil 4.11. İkili sınıflandırma için karışıklık matrisi..... 38

Şekil 4.12. K-kat çapraz geçerleme süreci..... 39

Şekil 5.1. Kapsama alınan AİHS maddelerine göre sınıf dağılımı..... 41

Şekil 5.2. Dokümanlardaki İhlal ve İhlal Değil sınıflarına ait ortalama kelime sayısı... 42

Şekil 5.3. Metinlerde kapsama alınan Olaylar bölümü (İngilizce ve Türkçe)..... 43

Şekil 5.4. Orijinal veride yapılan önişleme faaliyetleri..... 44

Şekil 5.5. Madde 6 tüm veri üzerinde SVM ve XGBoost Karmaşıklık Matrisleri..... 45

Şekil 5.6. Önerilen BiLSTM mimarisi..... 50

Şekil 5.7. Geri çeviri (Backtranslation) süreci..... 51

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
AGNES	Kümeleri Birleştirerek Yerleştirme (Agglomerative Nesting)
AİHS	Avrupa İnsan Hakları Sözleşmesi
AİHM	Avrupa İnsan Hakları Mahkemesi
AUC	Eğrinin Altındaki Alan (Area Under The Curve)
BERT	Transformer'dan Çift Yönlü Kodlayıcı Temsilcileri (Bidirectional Encoder Representations from Transformers)
CART	Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
CBOW	Sürekli Kelime Çantası (Continuous Bag of Words)
CNN	Evrişimsel Sinir Ağı (Convolutional Neural Network)
CSV	Virgülle Ayrılmış Değerler (Comma-Separated Values)
DT	Karar Ağaçları (Decision Trees)
GloVe	Küresel Vektörlerle Kelime Temsili (Global Vectors for Word Representation)
HTML	HyperText Markup Language (Hiper Metin İşaretleme Dili)
HUDOC	Avrupa İnsan Hakları Mahkemesi Veri tabanı (Human Rights Documentation)
KNN	K En Yakın Komşu (K Nearest Neighboor)
LR	Lojistik Regresyon
LSTM	Uzun Kısa Süreli Bellek (Long Short-Term Memory)
NB	Naive Bayes

NER	Adlandırılmış Varlık Tanıma (Named Entitiy Recognition)
NLP	Doğal Dil İşleme (Natural Language Processing)
MCC	Matthew Korelasyon Katsayısı (Matthew Correlation Coefficient)
PDF	Taşınabilir Belge Formatı (Portable Document Format)
RF	Rastgele Orman Algortiması (Random Forest)
RNN	Yinelemeli Sinir Ağı (Recurrent Neural Network)
SVM	Destek Vektör Makineleri (Support Vector Machines)
XGBoost	Ekstrem Gradyan Artırma (eXtreme Gradient Boosting)

1. GİRİŞ

Bilgisayar teknolojilerinin gelişimi, hukuk alanında köklü değişimlerin yaşanmasını sağlamıştır. 1960'lı yıllardan itibaren, araştırmacılar bilgisayarların hukuk dünyasında nasıl bir değişim sağladığını tartışmaktadır. Bu dönemde, bilgisayar teknolojilerinin hukuk alanında kullanılmaya başlanması, hukuki analizlerin hızlanmasını ve daha verimli hale gelmesini mümkün kılmak için önemli bir ilk adım olmuştur [1]. Hukuki metinlerin ve yargı kararlarının bilgisayar destekli analizine ilişkin ilk girişimler, yargıçların karar alma süreçlerinde etkili olan faktörlerin belirlenmesi ve karar desenlerinin istatistiksel olarak modellenmesi üzerine olmuştur. 1960'lı yıllarda, bilgisayarların hukuki analizlerde kullanılma potansiyeli, daha çok sayısal verilerin işlenmesiyle sınırlıydı [2]. 1962 yılında bu alanda yapılan ilk çalışmalardan birinde yargıçların sosyal, dini ve mesleki geçmişlerinin, kararlar üzerindeki etkileri istatistiksel yöntemlerle araştırılmıştır [3].

Geçtiğimiz yüzyılın sonlarına doğru internetin hızla yaygınlaşması, pek çok alanda büyük değişimlere yol açmıştır. Bu değişim, özellikle hukuk alanında belirgin bir etki yaratmış, artan veri hacmi ile birlikte “büyük veri” kavramı, hukuk dünyasında da kendine önemli bir yer edinmiştir. Hukuk alanında, büyük veri kavramı, davaların daha hızlı ve etkili bir şekilde çözülmesi adına veri analizinin önemini arttırmış ve bu alanda yeni teknolojilerin gelişmesine olanak sağlamıştır. Veri madenciliği, bu süreçte, büyük veri setlerinde gizli kalmış desenleri, anlamlı ilişkileri ve bilgileri ortaya çıkarmak için kullanılan güçlü bir araç olarak öne çıkmaktadır [4]. Metin madenciliği, özellikle hukuki metinler gibi yapılandırılmamış verilerin işlenmesinde kritik bir teknik haline gelmiştir. Bu teknik, verilerin daha anlamlı hale getirilmesini sağlar, bu sayede hukuki veriler üzerinde yapılacak analizlerin daha verimli olmasına olanak tanır.

Metin madenciliği, yapılandırılmamış metin verilerinden anlamlı yapılar çıkararak bu verileri analiz edilebilir hale getirmeyi amaçlamaktadır. Hukuki verilerin analiz edilmesinde, özellikle hukukçuların sürekli olarak karşılaştığı PDF’ler, Word belgeleri, HTML sayfaları gibi veri biçimleri, metin madenciliğinin kapsamına girmektedir. Metin madenciliğinin temel amacı, metinlerden sayısal özellikler çıkararak bunları veri madenciliği algoritmaları için uygun hale getirmektir [5]. Hukuki verilerin hızlı bir şekilde işlenmesi ve anlamlı sonuçlar elde edilmesi, bu alandaki en büyük gereksinimlerden biridir. Artan hukuki veri

hacmiyle birlikte, metin madenciliği tekniklerinin kullanılması, hukuki süreçlerin daha hızlı ve etkili bir şekilde yönetilmesine olanak tanımaktadır. Özellikle hukuki belgelerin özetlenmesi, tasnif edilmesi, örüntülerin belirlenmesi gibi konular, metin madenciliği tekniklerinin yaygın kullanım alanları arasında yer almaktadır [6].

Hukuki belgeler, teknik ve karmaşık bir terminolojiye sahip olduğundan, bu belgelerin sistematik olarak analiz edilmesi ve sınıflandırılması kritik bir ihtiyaç haline gelmiştir. Hukuki dilin karmaşıklığı, metinlerin doğru bir şekilde analiz edilmesini ve anlamlandırılmasını zorlaştırmaktadır. Özellikle yargı kararlarının tahmin edilmesi, hukukçular ve yargı mensupları için önemli bir çalışma alanı haline gelmiştir. Bu bağlamda, doğal dil işleme (NLP) teknolojileri, metinlerin anlamlandırılması ve işlenmesinde kilit bir rol oynar. NLP, bilgisayarların insan dilini anlamasını ve işlemesini sağlar [7]. Hukuki metinlerin NLP teknikleriyle analiz edilmesi, karar alma süreçlerini hızlandırarak hukuk profesyonellerinin bilgiye erişimlerini kolaylaştırır. Yapay zekâ ile entegre edildiğinde, NLP'nin sağladığı avantajlar daha da artar. Yapay zekâ, bilgisayar sistemlerinin karmaşık görevleri öğrenme ve uygulama becerisini ifade ederken, NLP bu sistemlerin dil ile ilgili görevleri yerine getirmesini sağlar.

Yapay zekâ ve NLP entegrasyonu, hukuk alanında daha derin analizlerin yapılmasını ve daha karmaşık problemlerin çözülmesini mümkün kılmaktadır. Bu entegrasyonun en dikkat çekici uygulama alanlarından biri ise hukuki karar tahminidir. Hukuki karar tahmini, geçmiş yargı kararları ve içtihatlardan öğrenilen bilgilerle yeni davaların sonuçlarını öngörmeyi amaçlar. Bu süreç, makine öğrenimi algoritmaları ile NLP tekniklerinin bir araya getirilmesini gerektirir [8]. Bu teknolojiler, hukuki süreçlerdeki belirsizliği azaltarak karar alma süreçlerini hızlandırmakta ve hukuk profesyonellerine önemli bir destek sağlamaktadır.

Sonuç olarak, yapay zekâ ve NLP tabanlı sistemler, hukuk alanındaki veri analizinde yeni bir dönemin kapısını aralamaktadır. Özellikle hukuki karar tahmini gibi uygulamalar hem araştırma dünyası hem de pratik hukuk açısından büyük bir potansiyel sunmaktadır. Hukuki karar tahmin sistemleri, hukuk profesyonellerine dava sonuçları hakkında otomatik öngörüler sunarak zaman tasarrufu sağlar ve kaynakların daha etkin bir şekilde kullanılmasını mümkün kılar. Bu sistemler, profesyonellerin karmaşık ve stratejik görevlerine odaklanmalarını kolaylaştırırken, rutin işlemleri daha hızlı bir şekilde tamamlamalarını sağlar. Özellikle Avrupa İnsan Hakları Mahkemesi (AİHM veya

Mahkeme) gibi geniş çaplı başvuruların incelendiği yüksek mahkemelerde, iş yükünü hafifletmek ve daha verimli bir süreç oluşturmak açısından önem taşır. Karar tahmin sistemleri, benzer davaların analiz edilmesi ve öngörülerden faydalanılması yoluyla, hukuki süreçlerin daha hızlı ve etkili bir şekilde yönetilmesine katkıda bulunur.

AİHM, Avrupa Konseyi'ne üye ülkelerde temel hakların ihlaline yönelik bireysel başvuruları değerlendirir. Bu süreçte, davalara ilişkin geçmiş kararların sistematik bir şekilde incelenmesi, içtihatların analiz edilmesi ve benzer durumlar için öngörüler sunulması büyük önem taşır. Bu noktada, veri odaklı tahmin modelleri devreye girer. Bu modeller, AİHM'nin önceki kararlarından öğrenilen bilgiler ışığında, yeni hukuki sorunlar için muhtemel sonuçlar üretir. Böylelikle, hukuki profesyoneller ve mahkemeler, daha bilinçli ve veriye dayalı karar alma süreçlerine sahip olur [9].

Son yıllarda, hukuki kararların analizine yönelik geliştirilen model ve yöntemlerde kayda değer bir ilerleme kaydedilmiştir. Bu modeller, NLP teknikleriyle güçlendirilmiş olarak, metin verilerinin temsili, özellik çıkarımı ve makine öğrenmesi algoritmalarının uygulanması gibi adımları kapsamaktadır [10]. Hukuki metinlerin analizine ve tahminine yönelik bu yaklaşımlar, büyük çaplı hukuki veri setlerinin erişilebilirliğinin artması ve makine öğrenmesi alanındaki gelişmelerle desteklenmektedir.

Bu tez çalışmasında, Bölüm 2'de, konuya dair mevcut literatür taranmış ve farklı yaklaşımlar ele alınmıştır. Bölüm 3'te, çalışmada kullanılan veri setine kaynaklık eden AİHM işleyişi ve bu mahkemeye dayanak olan AİHS ile korunan haklar irdelenmiştir. Bölüm 4'te çalışmada kullanılan veri ve çalışmanın temelini oluşturan teknolojiler sunulmuştur. Bölüm 5'te gerçekleştirilen deneysel çalışmalar ve sonuçları verilmiştir. Bölüm 6'da deney sonuçları yorumlanmış, çıkarılan anlamlı bulgular tartışılmış ve gelecekteki araştırmalar için öneriler getirilmiştir.

Problem durumu / Konunun tanımı

AİHM kararlarının makine öğrenmesi teknikleriyle tahmin edilmesi, hukuki süreçlerin veriye dayalı bir şekilde analiz edilerek yeni başvuruların olası sonuçlarının öngörülmesini amaçlayan bir problemdir. Temel hedef, AİHM'nin içtihatlarına ve geçmiş kararlarına dayanarak belirli bir davanın Avrupa İnsan Hakları Sözleşmesi'nin (AİHS veya Sözleşme)

ilgili maddelerinin ihlal edilip edilmediğini tahmin edebilecek modeller geliştirmektir. Bu, mahkeme süreçlerini hızlandırabilir, iş yükünü azaltabilir ve hukuk profesyonellerine karar alma süreçlerinde yardımcı olabilir.

Tezin amacı, kapsamı ve önemi

Bu tezin temel amacı, AİHM kararlarının metin madenciliği ve makine öğrenmesi yöntemleri kullanılarak analiz edilmesi ve kararların tahmin edilmesine yönelik modeller geliştirilmesidir. Bu süreçte, hukuki metinlerin anlamlandırılması, yapılandırılması ve analiz edilmesi için NLP teknikleri ile makine öğrenmesi algoritmalarının bir araya getirilmesi hedeflenmektedir. Çalışma, AİHM kararlarının içtihatlarından yola çıkarak benzer hukuki durumlar için olası sonuçları öngörmeyi amaçlamaktadır.

Bu tez çalışmanın önemi maddeler halinde açıklanmıştır:

- AİHM gibi yoğun iş yüküne sahip kurumlarda, davaların daha hızlı analiz edilmesi ve karar süreçlerinin desteklenmesi için etkili bir çözüm sunar
- Mahkeme kararlarının öngörülebilirliğinin artırılması, hukukçuların benzer davalarda daha bilinçli ve veriye dayalı kararlar almasını sağlar
- Yapılandırılmamış hukuki metinlerin otomatik işlenmesi, bu alandaki bilgiye erişimi kolaylaştırır ve yeni iç görüler ortaya çıkarır.
- Hukuki belgelerin, makine öğrenmesi ve NLP teknikleri kullanılarak analiz edilmesi, büyük ölçekli hukuki veri setlerinin anlamlandırılmasını sağlar.
- Tezde geliştirilen modeller, mahkeme kararlarının tahmin edilmesine olanak tanıyarak, hukukçulara ve araştırmacılara zaman kazandırır, bu durum onların daha karmaşık hukuki problemlere odaklanmalarını mümkün kılar.

Varsayımlar / Sayıltılar

Çalışmada AİHM tarafından yayınlanan halka açık dava verileri kullanılacaktır. AİHM, 1959 yılında kurulan uluslararası bir mahkemedir. Mahkeme, AİHM'de belirtilen çeşitli hakların ihlali iddiasında bulunan bireysel ve devlet başvurularını inceler ve karara bağlar. Davaların görülebilmesi için kabul edilebilirlik aşamasını geçmesi gerekmektedir. Mahkemenin kararları çevrimiçi olarak erişilebilir durumdadır [11]. Yapılacak çalışmada

yayımlanmış olan kararlardan alınan metinler kullanılmıştır. Tahmin görevi için esas olması gereken Mahkemeye taraflarca sunulan argümanları içeren başvuru metinlerinin kullanılarak sonuç tahmini yapılmasıdır. Ancak bu başvuru metinleri yayımlanmamaktadır. Dolayısıyla burada ana varsayım yayımlanmış karar metinlerindeki belirli bölümler (Olaylar) ile Mahkeme'ye sunulan başvurular arasında yeterli benzerlik olduğudur. Eğer bu iki metin arasında yeterli benzerlik varsa bu çalışma başvuru metinleri analiz edilerek sonuçların tahmin edilmesinde kullanılabilir.

Sınırlılıklar

AİHM'ye yapılan başvuruların büyük çoğunluğu bireyler tarafından yapılmaktadır. Başvurular, öncelikle kabul edilebilirlik kriterlerine dayalı olarak ön aşamada değerlendirilmektedir. Bu kriterler arasında en önemlisi, etkili iç hukuk yollarının tüketilmiş olmasıdır. Eğer bir başvuru bu ilk aşamayı geçerse, ya başvuruyu kabul edilemez olarak ilan edip Mahkeme'nin dava listesinden çıkarabilecek bir tek hâkime atanır ya da başvuru bir Komite veya Daire'ye sevk edilir. Mahkeme istatistiklerine göre, başvuruların büyük bir kısmı bu ilk kabul edilebilirlik aşamasını geçememektedir. Örneğin AİHM 2024 [12] yılına ait yıllık rapora göre 31,329 adet başvuru kabul edilemez olarak ilan edilmiş veya dava listesinden çıkarılmıştır. Aynı yıl içinde ise yalnızca 6,931 başvuru esas hakkında bir hükme bağlanmıştır. Kabul edilemez bulunan veya dava listesinden çıkarılan davalar raporlanmamaktadır. Bu noktayı akılda tutmak önemlidir, çünkü analiz yalnızca Mahkeme'nin elektronik veri tabanı olan HUDOC (Avrupa İnsan Hakları Mahkemesi Veritabanı) üzerinden erişilebilen davalar üzerinde gerçekleştirilmiştir. Analiz edilen davalar, yalnızca ilk kabul edilebilirlik aşamasını geçmiş ve Mahkeme'nin bir oluşumu tarafından esasları hakkında karar verilen davalardır.

2. LİTERATÜR ARAŞTIRMASI

2.1. Hukuki Karar Tahmini Alanında Yapılan Çalışmalar

NLP ile yapay zekânın birleştirilmesi, hukuk profesyonellerinin günlük iş akışını dönüştürme potansiyeline sahip, hızla gelişen bir araştırma alanıdır. Bu entegrasyonun en dikkat çekici uygulamalarından biri, hukuki kararların öngörülmesidir. Hukuki karar tahmini, geçmiş yargı kararları ve mahkeme içtihatlarını analiz ederek, belirli bir dava bağlamında mahkemenin verebileceği olası kararları tahmin etmeye odaklanır. Bu teknoloji, hukuk profesyonellerine vakaların sonuçlarına ilişkin otomatik öngörüler sunarak zaman ve kaynak yönetimini optimize etme imkânı tanır ve onların daha kritik ve karmaşık görevlere yönelmelerini kolaylaştırır. Özellikle yüksek mahkemelerde, çok sayıda başvurunun olduğu karmaşık hukuki durumları çözmek zor olabilir. Hukuki karar tahmini, bu tür mahkemelerde iş yükünü hafifletebilir ve hakimlere, avukatlara ve diğer hukuki profesyonellere, benzer önceki davaların incelenmesi ve tahminlerin dikkate alınması yoluyla daha bilinçli kararlar almalarına yardımcı olabilir.

Son dönemlerde, araştırmacılar hukuki kararların temsil edilmesine yönelik modeller ve yöntemler geliştirme konusunda önemli mesafeler kat etmiştir. Bu modeller, metin temsilini, özellik çıkarımını ve makine öğrenimi algoritmalarını içeren çeşitli NLP tekniklerinden faydalanarak hukuki davaların sonuçlarını analiz etmek ve tahmin etmek amacıyla tasarlanmıştır. Geniş çaplı hukuki veri setlerine erişim imkanlarının olması ve makine öğrenimindeki yenilikler, bu modellerin başarısında önemli rol oynamıştır. Literatüre bakıldığında, doğal dil işleme yöntemleriyle hukuki karar tahmini gerçekleştiren çok sayıda çalışmaya rastlanmaktadır.

AİHM'nin yargı kararlarının doğal dil işleme yöntemleriyle tahmin edildiği ilk sistematik çalışmada [8] AİHS'nin bir maddesinin ihlal olup olmadığı şeklinde iki adet sınıfa sahip olan bir problem tanımlanmıştır. SVM modeli kullanılmıştır. Sınıflandırıcıya verilen girdiyi olarak davadan seçilen metin içerikleri oluşturmaktadır. Davada olaylar (The Facts) bölümünün mahkeme kararını tahmin ederken en önemli bölüm olduğu tespit edilmiştir. Mahkeme kararlarının kendine özgü bir yapısı vardır ve bu onları metne dayalı analiz için uygun hale getirmektedir. Sözleşme Madde 3, 6 ve 8'e göre 250, 80 ve 254 dava veri setini

oluşturmaktadır. Veriler HUDOC [11] veri tabanından çekilmiştir. Metin verileri n-gramlar kullanılarak temsil edilmiştir. Mahkeme kararları %79 doğruluk oranıyla tahmin edilmiştir.

Farklı yargı alanlarındaki hukuki kategorilerin tahminin yapıldığı çalışmada [13] metin temsili için temsili için iki tür n-gram kullanılmış: char-gram ve n-gram. Veri seti, İspanyol yasal sisteminden alınan 96.163 karardan oluşmaktadır. Veri seti 42 farklı hukuk kategorisine sahiptir. Sınıflandırma için SVM, DT, RF ve Gradyan Artırma sınıflandırıcıları kullanılmıştır. Doğruluk, duyarlılık ve F1 metriklerine göre sonuçlar değerlendirilmiş ve tüm bu metriklerde %90'ın üzerinde oranlar ölçülmüştür.

Bir başka çalışmada [9] Fransız Yüksek Mahkemesi davalarının sonuçlarını tahmin etmek için destek vektör makineleri kullanılmıştır. Çalışma kapsamında, Fransız Yüksek Mahkemesi'nin kararlarını ve bir davanın ait olduğu hukuk alanını yüksek doğruluk oranıyla tahmin etmek için makine öğrenimi tekniklerini uygulayan çeşitli deneyler gerçekleştirilmiştir. 126.000'den fazla belgeden oluşan veri setinde, her vakadan unigramlar ve bigramlar çıkarılmış ve karşılık gelen etiketler atanmıştır. Birden fazla SVM sınıflandırıcısının çıktısını birleştiren bir "ortalama olasılık toplu(ensemble) sistem" geliştirilmiştir. Çalışmada, dava kararlarının tahmini için %98 ortalama F1 skoru, bir davanın hukuk alanını tahmin etmek için %96 F1 skoru ve karar tarihini tahmin etmek için %87.07 F1 skoru elde edilmiştir.

Bir başka çalışmada [14] metin madenciliği yöntemleri kullanılarak mobbing içerikli yargı kararları hakkında yüksek mahkemece verilen kararların tahmini gerçekleştirilmiştir. Oluşturulan model ile %80 ve üzeri başarı sağlanmıştır.

Başka bir çalışmada [15] AİHM kararlarının tahmini için makine öğrenmesi algoritması olan SVM kullanılmıştır. AİHS'nin 9 maddesi için "ihlal" ya da "ihlal değil" şeklinde sınıflandırma yapmak için HUDOC veri tabanından davalara ilişkin metinler alınmıştır. N-gram kullanılarak metin temsili yapılmıştır. Davaların yaklaşık %75 inde kararlar doğru tahmin edilmiştir.

Bir başka çalışmada [16] makine öğrenmesi yöntemleri kullanılarak mahkeme kararları TF-IDF metin sayısallaştırma yöntemi ve CURE, K-MEANS, AGNES, AFFINITY ve BIRCH

gibi yapay zekâ yöntemleri ile kümelenmiştir. Performans metriklerine göre en iyi sonucu BIRCH vermiştir.

Bir diğer çalışmada [17] Delhi Bölge Mahkemesi ceza davalarından oluşan veri seti kullanılarak davaların "Beraat" veya "Mahkûmiyet" olarak sınıflandırıldığı ikili sınıflandırma çalışması yapılmıştır. Bu çalışmada 86 davadan (2017 ve 2018 yıllarına ait) oluşan bir veri seti kullanılmıştır, öncelikle davaların sonuçlarını etkileyen önemli faktörler belirlenmiş ve bu faktörlerin sonuç tahmini için tanımlayıcı olarak kullanılmıştır. Sınıflandırıcıların performansı doğruluk, kesinlik, duyarlılık ve F1 skoru metrikleri kullanılarak ölçülmüştür. LR, KNN, CART, NB, SVM, Bagging, RF ve Boosting algoritmaları kullanılmış ve CART hem doğruluk (%91,86) hem de F1 skoru (%91,76) açısından en iyi sonuçları vermiştir. Bununla birlikte diğer algoritmalar da %85 ve %92 arasında değişen doğruluk ve F1 skoru vererek başarılı performans göstermiştir.

Başka bir çalışmada [18] Türk Anayasa Mahkemesi'nin 2012 ve 2020 yılları arasında verdiği kamu ahlakı ve ifade özgürlüğü ile ilgili bireysel başvuru kararları incelenmiştir. Araştırma, bu başvuruların ihlal, ihlal yok ya da kabul edilemez kararlarıyla sonuçlanıp sonuçlanmadığını tahmin etmeyi amaçlamıştır. Mahkeme kararlarından türetilen metinlerle MLP (Multilayer Perceptron) modeli kullanılarak başarılı tahminler elde edilmiştir. Model, davaların "konu" ve "gerekçe" bölümlerinin karar sonuçlarını oldukça iyi bir şekilde tahmin edebildiğini (ortalama %90 doğruluk) göstermiştir.

Bir diğer çalışmada [19] yazarlar Brezilya'nın 5. Bölgesel Federal Mahkemesi'ne ait 612.961 Federal Küçük Davalar temyiz dosyasını kullanarak üç farklı derin öğrenme modelini (ULMFiT, BERT ve Big Bird) eğitmişlerdir. Ardından, bu modellerin tahmin performanslarını 22 deneyimli hukuk uzmanının tahminleriyle karşılaştırmışlardır. Elde edilen sonuçlar, tüm derin öğrenme modellerinin, hukuk uzmanlarını geride bıraktığını ve en iyi modelin Matthew Korelasyon Katsayısı (MCC) değeri olarak 0.3688 elde ettiğini, oysa insan uzmanlarının MCC değerinin sadece 0.1253 olduğunu göstermektedir. Bu durum, doğal dil işleme ve makine öğrenimi tekniklerinin hukuki sonuçları tahmin etme konusunda oldukça umut verici bir yaklaşım sunduğunu ortaya koymaktadır.

Bir başka çalışmada [20] yazarlar 35.000 den fazla Hindistan Yüksek Mahkemesi davalarını içeren ve orijinal mahkeme kararlarıyla etiketlenmiş bir veritabanı olan ILDC (Indian Legal

Documents Corpus)' yi kullanarak mahkeme kararı tahmini yapmışlardır. Bu görev, bir otomatik sistemin davanın sonucunu tahmin etmesini ve bu sonucu açıklanabilir şekilde sunmasını kapsamaktadır. En iyi tahmin modeli, insan hukuk uzmanlarının %94 doğruluk oranına karşılık %78 doğruluk elde etmiştir. Bu durum, tahmin görevinin zorluğunu ve karmaşıklığını göstermektedir. Ayrıca, önerilen algoritmanın açıklamaları üzerinde yapılan analiz, algoritmanın ve hukuk uzmanlarının mahkeme kararlarını açıklamak için farklı bakış açılarına sahip olduklarını ortaya koymuştur, bu da gelecekteki araştırmalar için bir potansiyel alanın varlığını göstermektedir.

Mahkeme kararlarının tahminin yapıldığı başka bir çalışmada [21] LSTM+CNN hibrit modeli kullanılmıştır. Mahkeme kararlarını tahmin etmek için öncelikle kullanılan modelde etkili özellikleri belirlemek amacıyla RFE (Recursive Feature Elimination) testi kullanılmıştır. Bu test, modelin kullanacağı en önemli özellikleri belirlemek ve tahmin performansını artırmak için kullanılmıştır. Sonuç olarak, önerilen teknik, geçmiş yargı kayıtlarından mahkeme kararlarını tahmin etmek için hem optimum özellik seçimi hem de LSTM ve CNN katmanlarının avantajlarını sunmaktadır. Hedefte 3 sınıf vardır (onay, red ve diğer). ABD Yüksek Mahkemesi Kararlarından 120,506 adet dava, veri setini oluşturmaktadır. Veri seti, eğitim seti için yaklaşık olarak %80 ve test seti için %20 olacak şekilde ayrılmıştır. Her dava için 10 öznitelik bulunmuştur. %92 doğruluk, %93 kesinlik, %94 duyarlılık, %93 F1skoru elde edilmiştir.

2022 yılında yapılan bir başka çalışmada [22] AİHM davalarının sınıflandırılması için transformer tabanlı modellerin kullanımı araştırılmıştır. Çalışmada, BERT ve türevleri gibi transformer tabanlı modellerin metin sınıflandırmadaki etkinliği incelenmiş, ancak bu modellerin uzun belgeleri sınıflandırmada bellek ve hesaplama gücü açısından maliyetli olduğu ve genellikle genel dil modelleri üzerine önceden eğitildikleri için hukuk gibi özel alanlarda daha az etkili olabileceği belirtilmiştir. Bu nedenle, çalışmada, modelin maksimum dizi uzunluğunu artırmak için kaydırma penceresi (sliding window) yaklaşımı kullanılarak yasal belgelerin sınıflandırılması önerilmiştir. Ayrıca, dengesiz olan AİHM veri setini dengelemek için web'den dava makaleleri toplanmış ve veri çıkarımı yapılmıştır. Daha sonra, SVM, Karar Ağaçları, Naive Bayes, AdaBoost gibi geleneksel makine öğrenimi teknikleri ve BERT, Legal-BERT, RoBERTa, BigBird, ELECTRA ve XLNet gibi transformer tabanlı modeller sınıflandırma görevi için kullanılmıştır. Deneysel bulgular, RoBERTa modelinin diğer BERT versiyonlarından daha iyi performans gösterdiğini ve

sırasıyla %89,1, %86,2 ve %86,7 doğruluk, duyarlılık ve F1 skoru elde ettiğini göstermiştir. Geleneksel makine öğrenimi teknikleri arasında ise AdaBoost, SVM, Karar Ağaçları ve Naive Bayes'ten daha iyi performans göstererek sırasıyla %81,9, %81,5 ve %81,7 doğruluk, duyarlılık ve F1 skoru elde etmiştir.

2.2. Hukuki Metin Veri Kümeleri

Mahkeme kararlarını tahmin etmeye yönelik etkili bir sistem tasarlamak, güvenilir ve doğru sonuçlar elde etmek için özenle yapılandırılması gereken bir süreçtir. Bu süreçte, en temel adımlardan biri veri toplama aşamasıdır. Tahmin modellerinin eğitimi ve test edilmesi için kullanılacak güvenilir ve kapsamlı bir veri setinin oluşturulması hayati bir öneme sahiptir. Veri seti, tahmin modelini eğitmek ve test etmek için kullanılır. Veri setleri çeşitli hukuki konularda olabilir. Bunlar ülkelerin anayasa mahkemesi kararları, ceza mahkemesi kararları, sivil mahkeme kararları, ticaret ve iş hukuku kararları, yüksek mahkeme kararları, uluslararası mahkeme kararları olabilir. Veri toplama aşamasında, hukuk kurumlarından, kamuya açık kaynaklardan veya özel veri tabanlarından yararlanılarak mahkeme kararlarına erişim sağlanabilir. Bu aşamada, modelin performansının genel geçerliliğini artırmak amacıyla veri çeşitliliğine ve kalitesine önem gösterilmelidir. Aşağıda literatürde sıklıkla kullanılan bazı hukuki metin veri kümeleri verilmiştir.

ECtHR (European Court of Human Rights)

AİHM, Avrupa Konseyi'ne üye 47 ülkenin vatandaşlarının, AİHS'ne aykırı hareket eden devletler aleyhine başvurabileceği bir yargı organıdır. Özellikle AİHS'ne taraf devletlerin uymadığı kararlar ve ihlallerle ilgili davaları incelemektedir. Veri tabanı [11] ise, mahkeme kararları, başvurular ve bunlarla ilgili yargı süreçlerini içeren çok kapsamlı bir kaynaktır.

EURLEX

Avrupa Birliği'nin (AB) resmi hukuki veri tabanıdır. EURLEX [23], AB'nin hukuki belgelerini, düzenlemelerini, direktiflerini, mahkeme kararlarını ve diğer hukuki metinlerini içeren kapsamlı bir veri setidir. Bu veri seti, AB'nin hukuk sistemine dair geniş bir bilgi sunar ve hukuki profesyoneller, araştırmacılar ve ilgilenen diğer taraflar için önemli bir kaynaktır.

US Supreme Court Database

US Supreme Court Database [24], ABD Yüksek Mahkemesi'nin kararlarına dair kapsamlı, detaylı ve erişilebilir bir kaynak sunar. Karar metinleri, davaların geçmişi, taraflar, hukuk alanları ve dava sonuçları hakkında bilgi sağlar.

Caselaw Access Project

Harvard Hukuk Fakültesi Kütüphanesi tarafından geliştirilen ve sağlanan büyük bir dijital veritabanıdır [25]. Bu proje, Amerika Birleşik Devletleri'nin federal ve eyalet mahkemelerinin kararlarını dijital ortamda toplayan ve bunlara erişim sağlayan bir kaynaktır.

Legifrance

Fransa'da yürürlükte olan hukuk metinlerini ve yargı kararlarını içeren resmi bir internet sitesidir [26]. Legifrance, Fransız hukuk sistemiyle ilgili bir dizi kaynağı çevrimiçi olarak sağlar. Fransa'da yürürlükte olan yasaların metinleri, değişiklikler ve güncellemeler, Fransız Anayasa metni ve anayasa değişiklikleri, Fransa'nın taraf olduğu uluslararası anlaşmalar bu veri tabanından erişilebilir durumdadır.

European Parliment Proceedings

Avrupa Parlamentosu'nun toplantılarının ve işleyişinin kayıtlarını içeren bir veri setini ifade eder [27]. Bu kayıtlar, Avrupa Parlamentosu'nun çeşitli oturumlarında yapılan konuşmaları, oylamaları, kararları ve diğer önemli olayları içerir.

Open Australian Legal Corpus

Avustralya'da verilen mahkeme kararlarını, yasal metinleri ve diğer hukuki belgeleri dijital ortamda bir araya getirir [28]. Avustralya'daki yasama ve yargı süreçlerini kapsayan çok sayıda hukuki belgeyi içerir.

3. AVRUPA İNSAN HAKLARI MAHKEMESI İŞLEYIŞI VE AİHS İLE KORUNAN HAKLAR

"Avrupa İnsan Hakları Sözleşmesi" olarak bilinen ve İnsan Hakları ile Temel Özgürlüklerin Korunmasını amaçlayan uluslararası anlaşma, 4 Kasım 1950'de Roma'da imzalanmış ve 3 Eylül 1953'te yürürlüğe girmiştir. Bu sözleşme, İnsan Hakları Evrensel Bildirgesi'nde yer alan temel hakları daha somut bir çerçeveye oturtmuş ve bu hakların ihlali durumunda devletlerin sorumluluğunu ortaya koyabilecek bir uluslararası yargı mekanizması oluşturmuştur. Bu mekanizma, Avrupa Konseyi'ne üye devletler arasında Sözleşme'nin uygulanmasını izleyen ve bireysel başvuruları değerlendiren AİHM olarak faaliyet göstermektedir. Bu mahkeme 1959 yılında kurulmuştur. AİHM 'nin üye sayısı, 47 Avrupa Konseyi üye ülkesinin her biri tarafından bir yargıç atanmasıyla belirlenir. Bu nedenle, AİHM'nin toplam üye sayısı 47'dir [29].

AİHM, Sözleşme 'ye taraf devletlerin yükümlülüklerini yerine getirip getirmediğini denetleyen bağımsız bir organ olarak, bireylerin ve devletlerin başvurularını ele alır. Sözleşme, iki temel başvuru türüne izin verir. İlk olarak, bireyler, tüzel kişiler, gruplar veya sivil toplum kuruluşları tarafından yapılan başvurular, bireysel hak ihlallerinin Mahkeme önüne getirilmesini sağlar. İkinci tür başvurular ise devletler arası nitelikte olup, bir devletin başka bir devleti Sözleşme hükümlerini ihlal etmekle suçladığı durumları kapsamaktadır.

Mahkeme'ye yapılan başvurular yalnızca Sözleşme'yi onaylayan devletlere karşı yöneltilebilir. Sözleşme'ye taraf olmayan devletlere ya da bireylere karşı yapılan başvurular, usulen geçersiz sayılır ve kabul edilmez. Kuruluşundan bu yana Mahkeme'nin gündemindeki başvuruların büyük çoğunluğunu bireysel başvurular oluşturmaktadır. Bu başvurular genellikle, Sözleşme'nin ihlal edildiğini iddia eden kişiler tarafından yapılmakta ve Mahkeme'nin iş yükünün büyük bir bölümünü temsil etmektedir. AİHM, bu bireysel başvurular aracılığıyla, temel hakların korunmasında önemli bir rol oynamaktadır [30].

Mahkemeye yapılan başvurular ilk olarak bazı kabul edilebilirlik kriterlerine göre değerlendirilir (hususiyetle iç hukuk yollarının tüketilmesi) .
Mahkeme'ye taşınan davalar iki temel aşamada değerlendirilir: kabul edilebilirlik aşaması ve esas inceleme aşamasıdır (diğer bir ifadeyle şikâyetlerin detaylı incelenmesi aşamasıdır).

Tek yargıç sistemi, açıkça kabul edilemez olduğu belirlenen başvuruları reddetme yetkisine sahiptir. Bu tür kararlar nihai olup, itiraz hakkı bulunmamaktadır.

Kabul edilebilirlik şartları

Başvuruların, AİHM tarafından kabul edilebilmesi için belirli şartları taşıması gerekir. Bunlar arasında, başvurulan ülkede iç hukuk yollarının tükenmiş olması zorunludur; yani, başvurucu, önce kendi ülkesindeki en yüksek mahkemeye başvurmalı ve devlete, hak ihlalini ulusal düzeyde düzeltme fırsatı tanımalıdır. Başvurucu, Sözleşme 'deki haklardan biri veya birkaçı ihlal edildiğini ileri sürmelidir. Mahkemeye başvurular, en son verilen kararın ardından dört ay içinde yapılmalıdır. Başvurucu, ihlalin doğrudan mağduru olmalı ve önemli bir zarara uğramış olmalıdır. Ayrıca, yalnızca Sözleşme 'ye taraf olan devletler aleyhine başvurulabilir; diğer ülkelere veya kişilere yönelik başvurular kabul edilmez.

AİHS ile sağlanan haklar

Sözleşmeye taraf devletler kendi yetki alanları içindeki herkesin Sözleşme' de açıklanan hak ve özgürlüklerden yararlanmalarını sağlar [31]. Bu haklar şunlardır:

- Madde 2: Yaşam hakkı
- Madde 3: İşkence yasağı
- Madde 4: Kölelik ve zorla çalıştırma yasağı
- Madde 5: Özgürlük ve güvenlik hakkı
- Madde 6: Adil yargılanma hakkı
- Madde 7: Kanunsuz ceza olmaz
- Madde 8: Özel ve aile hayatına saygı hakkı
- Madde 9: Düşünce, vicdan ve din özgürlüğü
- Madde 10: İfade özgürlüğü
- Madde 11: Toplantı ve dernek kurma özgürlüğü
- Madde 12: Evlenme hakkı
- Madde 13: Etkili başvuru hakkı
- Madde 14: Ayrımcılık yasağı
- Madde 17: Hakları kötüye kullanma yasağı
- Madde 18: Haklara getirilecek kısıtlanmaların sınırlanması

- 1. Ek Protokol: Mülkiyet hakkı (m. 1), eğitim hakkı (m. 2), serbest seçim hakkı (m.3)
- 4. Ek Protokol: Borçtan dolayı özgürlükten yoksun bırakılma yasağı (m. 1), serbest dolaşım özgürlüğü (m. 2), vatandaşların sınır dışı edilmeleri yasağı (m. 3), yabancıların topluca sınır dışı edilmeleri yasağı (m. 4)
- 6. Ek Protokol: Savaş ve yakın savaş tehlikesi zamanında işlenen fiiller hariç ölüm cezasının kaldırılması (m.1 ve 2)
- 7. Ek Protokol: Yabancıların sınır dışı edilmelerine ilişkin usuli güvenceler (m. 1), cezai konularda iki dereceli yargılanma hakkı (m. 2), adli hata halinde tazminat hakkı (m. 3), aynı suçtan iki kez yargılanmama ve cezalandırılmama hakkı (m. 4), eşler arasında eşitlik (m. 5)
- 12. Ek Protokol: Ayrımcılığın genel olarak yasaklanması (m. 1)
- 13. Ek Protokol: Ölüm cezasının kaldırılması (m. 1)

Mahkeme karar metni yapısı

Mahkemenin kararları çevrimiçi olarak erişilebilir durumdadır. HUDOC, AİHM kararları, içtihatları ve diğer ilgili belgelerinin toplandığı bir veri tabanıdır. Çalışmada bu veri tabanından elde edilen yayımlanmış kararlar kullanılmıştır. Tahmin görevi için esas olması gereken, Mahkemeye taraflarca sunulan argümanları içeren başvuru metinleridir. Ancak bu başvuru metinleri yayımlanmamaktadır. Bu nedenle ana varsayım, yayımlanmış karar metinlerindeki belirli bölümler (Olaylar) ile Mahkeme'ye sunulan başvurular arasında yeterli benzerlik olduğudur. Eğer bu iki metin arasında yeterli benzerlik varsa, bu çalışma başvuru metinleri analiz edilerek sonuçların tahmin edilmesinde kullanılabilir.

AİHM İç Tüzüğü'nün [32] 74. Maddesine göre AİHM karar metinleri standart bir formatı izler ve aşağıdaki ana bölümleri içerir:

- Giriş (Introduction): Dava adı, başvuru numarası, kararın tarihi, yapılan atamalar (yargıçlar, başkan, kayıt memuru vb.),
- Prosedür (Procedure): Başvurunun yapıldığı andan Mahkeme tarafından verilen karara kadar gerçekleşen işlemler,
- Olaylar (The Facts): Dava ile ilgili olayların ayrıntıları ve maddi gerçekler,
- Hukuk (The Law): Mahkeme'nin hukuki gerekçeleri, tarafların argümanlarının değerlendirilmesi ve Sözleşme maddelerinin uygulanması,

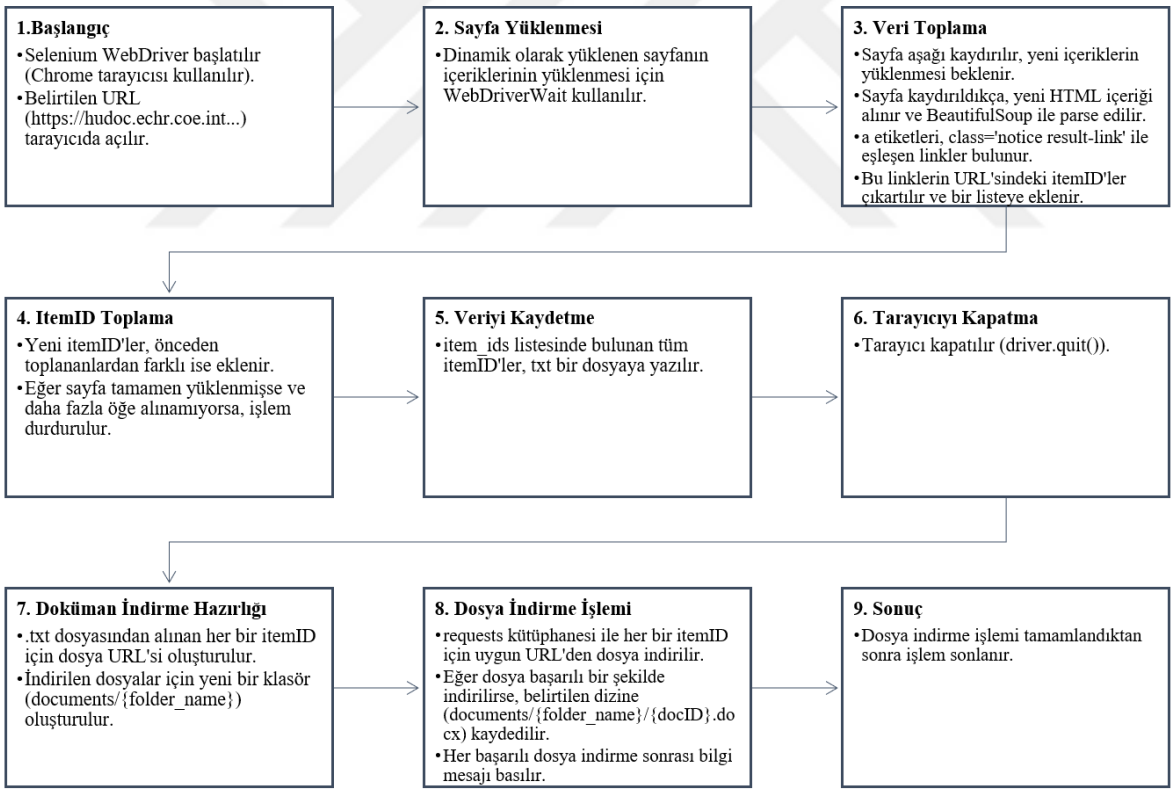
- Karar (Judgement): Mahkeme’nin kararı ve hükümler.

Bu yapı, kararların analizinde standartlaştırmayı ve bölümlerin birbirinden ayrılmasını kolaylaştırır.

4. MATERYAL VE YÖNTEM

4.1. HUDOC Veri Tabanı

AİHM kararları, İngilizce ve Fransızca dillerinde yayımlanmaktadır. Çalışmada dili İngilizce dilinde yayımlanmış olan davalar kullanılmıştır. HUDOC veri tabanında yayımlanan dava metinleri hazırlanan bir otomasyon ile çekilmiştir. Bunun için bu veri tabanında Şekil 4.1.'deki maddeler için İngilizce dilinde yayımlanan davalar için "ihlal" ve "ihlal değil" durumları ayrı ayrı filtrelenmiş ve dava metinleri .docx formatında indirilerek klasörlere kaydedilmiştir. İlgili veri tabanında her bir dava metninin "İhlal" ve "İhlal değil" durumları için filtreleme yapılarak elde edilen sayfalarda yapılan veri kazıma süreci aşağıda verilmiştir.



Şekil 4.1. HUDOC veri tabanından veri çekme süreci

Bir davada AİHS'nin bir maddesi veya madde fıkrası için hem "ihlal" hem de "ihlal değil" kararlarının verilmesi durumu söz konusu olabilmektedir. Bu durum genellikle birden fazla şikâyet veya hukuki argümanın aynı madde veya aynı fıkra kapsamında değerlendirildiği

durumlarda ortaya çıkar. Örneğin 6. Maddenin 1. Fıkrası hakkaniyete uygun yargılama hakkını düzenler. Aşağıda bir dava metninden alınan karar bölümünde bu fıkra ile ilgili hem "ihlal" hem "ihlal değil" kararı verilmiştir.

(İngilizce)

…
FOR THESE REASONS, THE COURT, UNANIMOUSLY,
…
Holds that there has been a violation of Article 6 § 1 of the Convention on account of the lack of objective impartiality;
Holds that there has been no violation of Article 6 §§ 1 and 3 of the Convention on account of the Supreme Court's reversal of the applicant's acquittal by means of written proceedings and the reasoning of the relevant judgment;
…

(Türkçe)

…
BU NEDENLERLE, MAHKEME, OYBİRLİĞİYLE,
…
Mahkemenin, objektif tarafsızlık eksikliği nedeniyle Avrupa İnsan Hakları Sözleşmesi'nin 6. Maddesi'nin 1. Fıkrasının ihlaline karar vermiştir;
Yüksek Mahkeme'nin başvurucunun beraatını yazılı yargılama ile bozması ve ilgili kararın gerekçesi nedeniyle Avrupa İnsan Hakları Sözleşmesi'nin 6. Maddesi'nin 1. ve 3. Fıkralarının ihlal edilmediğine karar vermiştir;
…

Mahkeme, yukarıdaki örnekte bir konuda (objektif tarafsızlık) Sözleşme'ye aykırılık tespit etmiş. Bu, adil yargılanma hakkının ihlali anlamına gelir. Ancak diğer bir konuyu (Yargıtay'ın yazılı yargılama süreci ve gerekçe açıklaması), Sözleşme' ye uygun bulmuştur.

Çalışma, bir davada aynı madde için sadece tek bir kararın bulunduğu davalara odaklanmış ve bir madde için her iki kararın da bulunduğu davalar veri çekme esnasında elenmiştir. Bunun sebebi bu iki kararı (ihlal ve ihlal değil) etkileyen kalıpları daha net belirlemektir.

Çalışmada karar metinlerinin tüm bölümleri değil "Olaylar (The Facts)" bölümü kullanılmıştır. Çünkü diğer alanlar, AİHM kararlarında genellikle hukuki değerlendirme, başvuruculların ve hükümetin argümanları, mahkemenin hukuki gerekçeleri gibi teknik ve yorumsal bilgiler içerir. "Olaylar" bölümü daha somut olaylara, davada yaşanan duruma odaklanır. Burada, olayların ne şekilde geliştiği, tarafların ne yaptıkları gibi net bilgiler verilir, bu da modelin hangi olayların ihlale yol açtığını öğrenmesi için daha uygun bir veri

seti sağlar. Dolayısıyla çalışma kapsamındaki her Sözleşme maddesi için ihlal ve ihlal olmayan karar metinlerini içeren dokümanlar içerisinden "Olaylar" bölümü düzenli ifadelerle (regular expressions) çıkarılmıştır. Bunun için "THE FACTS(Olaylar)" ve "THE LAW (Hukuki Dayanak)" başlıkları arasındaki metin alınmış, diğer bölümler dahil edilmemiştir. Dahil edilen metin bölüm CSV dosyalarına kaydedilmiştir. "İhlal" kararı verilenler 1, "ihlal yoktur" kararı verilenler 0 olarak etiketlenmiştir.

Bütün bu işlemleri gerçekleştirdikten sonra sözleşme maddelerine ilişkin dava sayısı aşağıdaki çizelgede verilmiştir.

Çizelge 4.1. HUDOC veri tabanından çekilen dava sayıları

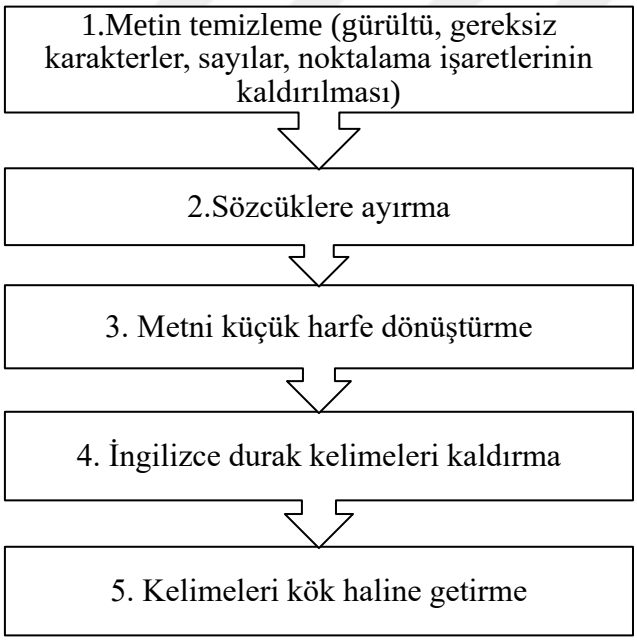
Sözleşme Maddesi	İhlal kararı verilen dava sayısı (1 etiketli)	İhlal yok kararı verilen dava sayısı (0 etiketli)	Toplam Sayı
Madde 2	578	98	676
Madde 3	334	739	1760
Madde 5	1489	223	1702
Madde 6	5010	731	5765
Madde 7	42	64	106
Madde 8	526	440	1329
Madde 10	558	186	730
Madde 11	205	50	255
Madde 13	1289	149	1445
Madde 14	265	256	521
Protokol 1 (1)	734	128	872
Protokol 1 (2)	18	10	28
Protokol 1 (3)	41	35	76
Protokol 4 (2)	71	24	95
Protokol 4 (4)	51	30	81
Protokol 7 (1)	11	24	35

Protokol 7 (2)	22	4	26
Protokol 7 (4)	47	16	63
Protokol 12 (1)	14	6	20

Her madde için oluşturulan ihlal ve ihlal yok etiketli koleksiyonlardan anlaşılacağı üzere (Bkz. Çizelge 4.1.) Mahkeme tarafından incelenen çoğu kabul edilebilir davanın 'ihlal' kararıyla sonuçlandığı görülmektedir. Dolayısıyla sınıflar arasında genel olarak büyük bir dengesizlik mevcuttur.

4.2. Verinin Hazırlanması (Metin Önişleme)

Metin verisi önişleme, doğal dil işleme uygulamalarında metin verisinin temizlenmesi, düzenlenmesi ve hazırlanması sürecidir. Bu süreç, verinin daha etkili bir şekilde analiz edilmesini ve makine öğrenimi modellerine beslenmesini sağlar. Geliştirilen sistemde gerçekleştirilen önişleme adımları aşağıda gösterilmiştir.



Şekil 4.2. Metin önişleme adımları

Metin temizleme

Metin verisini analiz için daha düzenli ve uygun bir hale getirmek amacıyla yapılan işlemlerdir. Noktalama işaretlerinin, rakamların, özel karakterlerin, gereksiz boşlukların kaldırılması, eksik alanlı değerlerin kaldırılması süreçleri gerçekleştirilir.

Sözcüklere ayırma (Tokenization)

Metni sözcüklere ayırma (tokenization), doğal dil işleme süreçlerinde metni daha küçük anlamlı birimlere (genellikle kelimelere veya alt kelimelere) bölme işlemidir. Analiz ve modelleme için temel bir adımdır.

Metni küçük harfe dönüştürme

Bu işlem, metindeki büyük/küçük harf farklılıklarını kaldırarak tutarlılık sağlar.

İngilizce durak kelimeleri kaldırma (Stopword filtering)

Durak kelimeler, genellikle bağlaçlar, zamirler ve sıkça kullanılan, ancak analize anlamlı bir katkı sağlamayan kelimelerdir. Durak kelimeleri, her sınıftaki metinlerde yoğun ve benzer sıklıkta kullanıldığı için sınıflandırma sürecine anlamlı bir katkı sağlamaz. Ayrıca, öznitelik vektörünün boyutunu gereksiz yere büyüterek modeli karmaşıklaştırabilir. Bu nedenle, durak kelimelerin metinlerden çıkarılması, yaygın olarak tercih edilen bir ön işleme adımıdır [33].

Kelimeleri kök haline getirme (Lemmatization)

Kelimeleri dil bilgisel köklerine, yani temel hallerine indirgeme işlemidir. Kelimenin bağlamı ve dil bilgisel türü dikkate alınarak belirlenir. Böylece, daha az, ama daha anlamlı özniteliklerle daha verimli modeller oluşturulması sağlanır.

Bu tez çalışmasında metinler üzerinde gerçekleştirilen önişleme adımları ve bu adımlar sonucunda elde edilen metinler aşağıdaki şekilde verilmiştir.

Çizelge 4.2. Orijinal ve önişleme adımlarından geçirilmiş veri örneği

Orijinal veri	**Metin temizleme işlemi yapılmış**	**Durak kelimeler kaldırılmış**	**Durak kelimeler kaldırılmış ve kelimeler kök haline getirilmiş**
1. The applicant was born in 1935 and lives in Greece. 2. On 13 November 2002 the ship Prestige, flying the flag of the Bahamas, was sailing in the Spanish exclusive economic zone off the coast of Galicia, carrying 70,000 tonnes of fuel oil. At a distance of 28 miles from Cape Finisterre it sent out an SOS after sustaining sudden and severe damage which produced a leak and caused the contents of its tanks to spill into the Atlantic Ocean. 3. As the Prestige was in danger of sinking, the maritime authorities launched a large-scale operation to rescue its crew. The ship was adrift and was approaching the coast, spilling its cargo into the sea.	the applicant was born in and lives in greece on date the ship prestige flying the flag of the bahamas was sailing in the spanish exclusive economic zone off the coast of galicia carrying tonnes of fuel oil at a distance of miles from cape finisterre it sent out an sos after sustaining sudden and severe damage which produced a leak and caused the contents of its tanks to spill into the atlantic ocean as the prestige was in danger of sinking the maritime authorities launched a large scale operation to rescue its crew the ship was adrift and was approaching the coast spilling its cargo into the sea	applicant born lives greece date ship prestige flying flag bahamas sailing spanish exclusive economic zone coast galicia carrying tonnes fuel oil distance miles cape finisterre sent sos sustaining sudden severe damage produced leak caused contents tanks spill atlantic ocean prestige danger sinking maritime authorities launched large scale operation rescue crew ship adrift approaching coast spilling cargo sea	applicant bear live greece date ship prestige fly flag bahama sail spanish exclusive economic zone coast galicia carry tonne fuel oil distance mile cape finisterre send sos sustain sudden severe damage produce leak cause content tank spill atlantic ocean prestige danger sink maritime authority launch large scale operation rescue crew ship adrift approach coast spill cargo sea

4.3. Metin Temsili

Metin temsili, metinsel ifadelerin makine öğrenmesi veya derin öğrenme algoritmaları için anlamlı bir hale getirilmesi gerekir. Bu sebeple kelimelerin, sayısal olarak ifade edilmesi gerekir. TF-IDF, Word2Vec, FastText gibi algoritmalar, bu tarz problemlerin çözümü için kullanılır ve bu yöntemler kelimelerin matematiksel olarak ifade edilebilmesini sağlar.

4.3.1. Terim Frekansı-Ters Doküman Frekansı (TF-IDF)

Bu yöntem, metni vektörlere dönüştürmek için ağırlıklandırma kullanır. Ağırlık, bir kelimenin bir korpustaki önemini değerlendirmek için kullanılan istatistiksel bir ölçüttür. TF-IDF ağırlığı iki terimden oluşur. TF (Term Frequency), bir kelimenin bir belgede görülme sayısının o belgedeki toplam kelime sayısına bölünmesiyle elde edilir. IDF (Inverse Document Frequency) ise bir terimin korpusta ne kadar seyrek olduğunu ölçer.

$$TF(t) = \frac{Belgedeki\ terim\ t'ninsayısı}{Belgedeki\ toplam\ terim\ sayısı} \tag{1}$$

$$IDF(t) = log\ \left(\frac{Toplam\ Belge\ Sayısı}{Belgede\ terim\ t'nin\ bulunduğu\ belge\ sayısı}\right) \tag{2}$$

TF-IDF, belgeler arasındaki kelime önemini belirlemek için kullanılır. Bir kelime, belirli bir belgede sıkça görünüyorsa ve genel korpusta nadirse, TF-IDF ağırlığı yüksek olacaktır. Bu, bir kelimenin belirli bir belge için ne kadar anlamlı olduğunu gösterir. Aşağıdaki şekilde hesaplanır:

$$TF - IDF(t) = TF(t) \times IDF(t) \tag{3}$$

4.3.2. N-Grams

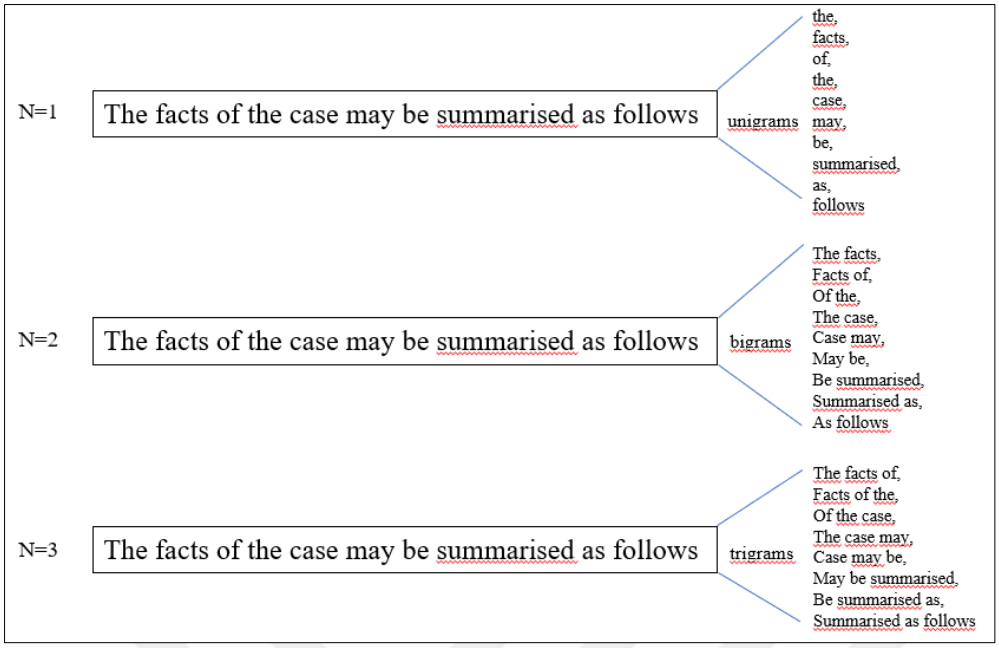
N-Grams yöntemi, doküman sınıflandırması için kullanılan bir başka ön işleme tekniğidir. N-gram, bir dizi kelimeden veya karakterden oluşan bir grup veya alt dizedir. Metin verilerini anlamak ve özellik çıkarmak için yaygın olarak kullanılan bir yöntemdir.

1-gram (unigram): Tek bir kelime veya karakter.

2-gram (bigram): İki ardışık kelime veya karakter.

3-gram (trigram): Üç ardışık kelime veya karakter.

4-gram (four-gram): Dört ardışık kelime veya karakter.



Şekil 4.3. N-grams

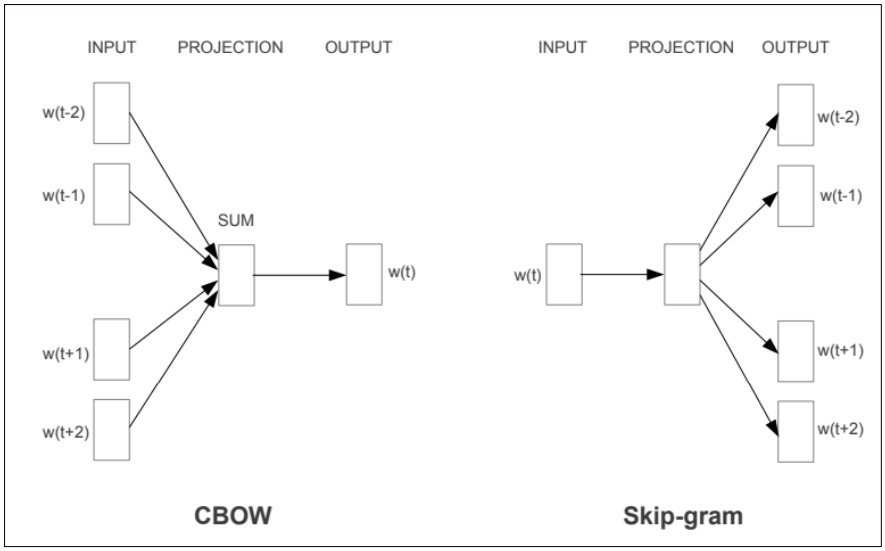
4.3.3. Word2vec

Word2vec (Kelime gömme), kelimeleri vektör uzayında temsil etmeye çalışan, etiketsiz (unsupervised) öğrenme yöntemiyle çalışan ve tahmine dayalı (prediction-based) bir modeldir. İstatistiksel temsillerden farklı olarak, kelime gömme yöntemleri, kelimeler arasındaki anlamsal ilişkileri ve bağlama dayalı bilgiyi öğrenmeye odaklanır. Word2vec'in iki temel yaklaşımı vardır: Sürekli kelime torbası (CBOW) ve Skip-gram.

Bu yöntemler birbirine benzese de işleyiş açısından farklılık gösterir:

- CBOW yönteminde, hedef kelimeyi tahmin etmek için giriş olarak o kelimenin çevresindeki (komşu) kelimeler kullanılır.
- Skip-gram yönteminde ise süreç tersine işler; giriş olarak hedef kelimenin kendisi kullanılır ve bu kelimenin çevresindeki komşu kelimeler tahmin edilir.

Her iki yöntem de kelimelerin anlamlarını bağlama göre öğrenmeye çalışarak dil modelleri oluşturmakta etkili bir yol sunar.



Şekil 4.4. Word2Vec Mimarisi [34]

4.3.4. GloVe

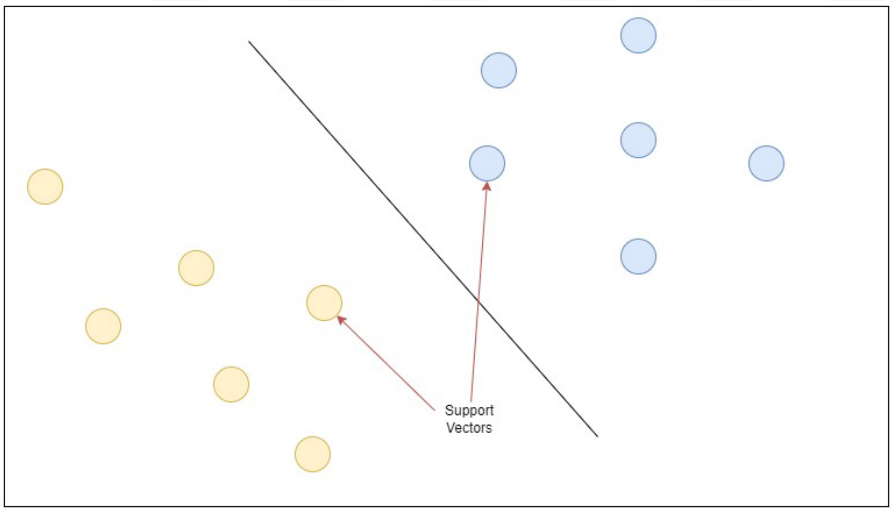
GloVe (Global Vectors for Word Representation) [35], Stanford Üniversitesi'nden bir araştırma ekibi tarafından 2014 yılında geliştirilen, kelimelerin anlamını vektörlerle temsil etmek için kullanılan bir kelime gömme (word embedding) yöntemidir. GloVe, kelimelerin çok boyutlu vektörler şeklinde temsil edilmesi amacıyla metinlerdeki sözcük-sözcük ilişkilerini kullanır.

GloVe, büyük bir metin korpusunda kelimelerin birbirleriyle kaç kez birlikte geçtiğini birlikte görülme matrisi (co-occurrence matrix) oluşturur. Matrisin her hücresi, bir kelimenin diğer bir kelime ile bağlamda kaç kez geçtiğini temsil eder. Ortak görüm matrisinden, kelimeler arasındaki anlam ilişkileri çıkarılır. Bu matristen, kelimeleri temsil eden yoğun (dense) vektörler elde edilir. Bu vektörler, kelimelerin anlamsal ve dil bilgisel ilişkilerini yakalar. GloVe, kelimeler arasındaki ilişkileri sadece yerel bağlamda (örneğin, bir cümlede) değil, tüm korpus boyunca öğrenir. Bu nedenle "global vektörler" adını almıştır.

4.4. Sınıflandırma Algoritmaları

4.4.1. Destek Vektör Makineleri

Destek vektör makineleri (Support Vector Machines) önceden tanımlanmış sınıflara dayalı olarak veri noktaları arasındaki sınırları belirleyen optimum veri dönüşümleri gerçekleştirerek karmaşık sınıflandırma, regresyon ve aykırı değer algılama sorunlarını çözmek için denetimli öğrenme modellerini kullanan bir makine öğrenimi algoritmasıdır [36]. Teknik olarak SVM algoritmasının temel amacı, farklı sınıfların veri noktalarını ayırt edilebilir şekilde ayıran bir hiperdüzlem tanımlamaktadır. Hiperdüzlem en büyük marjın söz konusu sınıfları ayıracağı şekilde belirlenir. Karmaşık ama küçük ve orta ölçekli veri setler için uygundur. Şekil 4.6'da resmedilmiştir.



Şekil 4.5. Destek Vektör Makineleri [37]

4.4.2. Multinomial Naive Bayes

Naive Bayes (NB) sınıflandırma algoritması, olasılık ilkelerine dayalı bir öğrenme yöntemi olup, sisteme sunulan verilerin ait olduğu sınıfı belirlemeyi amaçlayan bir yaklaşımdır [38]. Naive Bayes modelleri, özellikle büyük veri setlerinde verimli ve hızlı sonuçlar üretebilmeleri nedeniyle yaygın olarak kullanılmaktadır. Bu algoritma, Bayes Teoremi'ni [39] temel alır ve özellikler arasındaki bağımsızlık varsayımı ile çalışır.

Bayes Teoremi, belirli bir olayın gerçekleşme olasılığını, mevcut bilgiler doğrultusunda güncellemeye olanak tanıyan bir matematiksel prensiptir. Olasılık teorisinin temel

taşlarından biri olarak, özellikle koşullu olasılık hesaplamalarında kullanılır. Bu teorem, yeni veriler ışığında bir hipotezin doğruluğunu yeniden değerlendirmek için uygulanır.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4}$$

Verilen formülde:

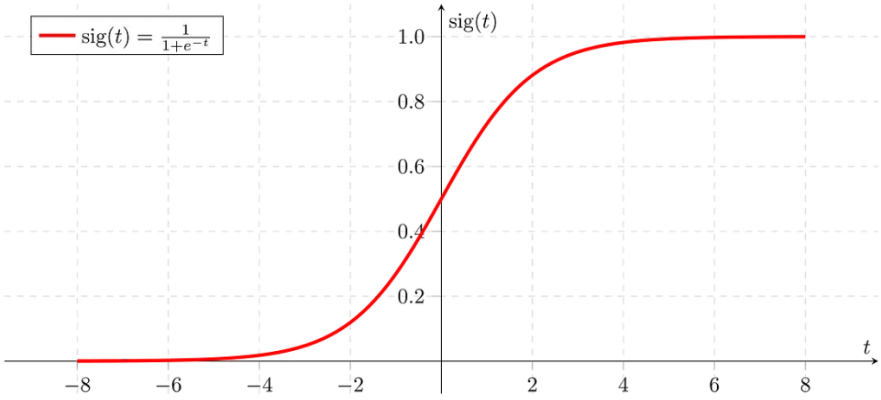
- Veri (B) gözlemlendiğinde A sınıfının olasılığı (posterior olasılık)
- A sınıfına ait olunduğunda B verisinin ortaya çıkma olasılığı (olabilirlik)
- A sınıfının öncül (prior) olasılığı
- B verisinin ortaya çıkma olasılığı (tüm sınıflar üzerinden normalize edilen olasılık)

Naive Bayes algoritması, özellikle metin sınıflandırmada yaygın olarak kullanılır çünkü yüksek boyutlu veri kümeleriyle çalışmaya uygundur ve kelimelerin birbirinden bağımsız olduğu varsayımıyla çalışarak hesaplamaları kolaylaştırır.

Multinomial Naive Bayes, özellikle kelime frekanslarının önemli olduğu metin sınıflandırma problemlerinde yaygın olarak kullanılan bir Naive Bayes çeşididir. Model, her bir özelliğin (kelimenin) sınıf koşullu olasılıklarını hesaplar ve en olası sınıfı tahmin eder. Her kelimenin belirli bir sınıfta gözlemlenme olasılığı, eğitim verisi üzerinden hesaplanan olasılıklara dayalıdır.

4.4.3. Lojistik Regresyon

Lojistik Regresyon, genellikle ikili (binary) veya çoklu sınıflandırma problemlerinde kullanılan bir istatistiksel modeldir [40]. Temel amacı, bağımsız değişkenler ile sınıf olasılıklarını ilişkilendirmektir. Özellikle doğrusal olmayan ilişkilere dayanarak sınıf tahmini yapar. Lojistik regresyon, doğrusal regresyonun bir genellemesidir, ancak çıktıyı sınıflandırma yapacak şekilde dönüştürür. Lojistik regresyon, sınıflandırma yapmak için Sigmoid (Lojistik) Fonksiyonu kullanır.



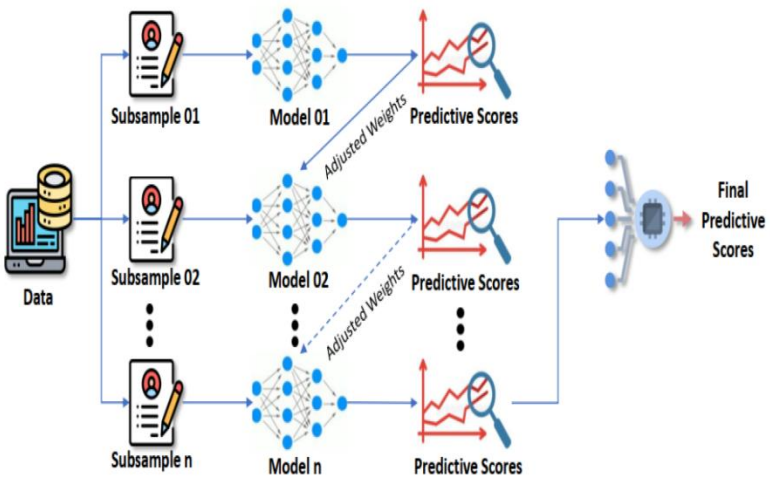
Şekil 4.6. Sigmoid fonksiyonunun grafiği [41]

Sigmoid fonksiyonu, modelin tahminini 0 ile 1 arasındaki bir aralığa ölçeklendirmek için kullanılan bir matematiksel araçtır. Bu işlem, bir olayın gerçekleşme olasılığının hesaplanmasını sağlar. Lojistik regresyonda, model tarafından üretilen çıktı doğrudan sınıf olasılığına karşılık gelir ve bu olasılık değerine göre sınıflandırma yapılır. Derin öğrenme modellerinde ise sigmoid fonksiyonu, aktivasyon fonksiyonu olarak kullanılır ve verilerin 0 ile 1 arasına sıkıştırılmasını sağlar. Özellikle Modelin çıktısının bir olasılık değeri olarak yorumlanması gerektiğinde bu fonksiyon kullanışlıdır.

4.4.4. XGBoost

XGBoost (eXtreme Gradient Boosting), Gradient Boosting yönteminin geliştirilmiş ve optimize edilmiş bir versiyonu olup, yüksek verimlilik ve performans sunan bir algoritmadır [42]. XGBoost, her adımda mevcut hataları düzeltmeye odaklanarak, modelin tahminlerini giderek daha doğru hale getiren bir makine öğrenimi tekniğidir. Bu yaklaşım, modelin eksik kaldığı noktaları belirleyip, bu eksiklikleri hedefleyerek optimize eder. Özellikle karmaşık ve yüksek doğruluk gerektiren problemlerde, güçlü bir performans sergileyerek, doğru sonuçlar elde etmeyi sağlar.

XGBoost algoritmasının çalışma mantığı aşağıdaki şekilde verilmiştir.



Şekil 4.7. XGBoost algoritmasının çalışma prensibi [43]

XGBoost algoritmasında veriler farklı alt kümelere bölünür, her alt küme modelin daha iyi genelleştirme yapmasını sağlamak için kullanılır. Bu süreç, XGBoost'un bir özelliği olan Stokastik Gradient Boosting yöntemine dayanır. Burada her iterasyonda verinin sadece bir alt kümesi kullanılır. Her bir alt örnekleme, bir model eğitmek için kullanılır. Bu modeller karar ağaçlarına dayalıdır. Her model, bir önceki modelin hatalarını minimize etmeye çalışır. Bu şekilde, her yeni model önceki modellerin hatalarını öğrenerek genel performansı artırır. Modellerin ürettiği tahmin sonuçlarına dayalı olarak, verilerdeki her örnek için ağırlıklar güncellenir. Hatalı tahmin edilen örneklerin ağırlıkları artırılır, böylece bu örnekler sonraki modelde daha fazla dikkate alınır. Her bir model, kendi alt örnekleminden tahmin skorları üretir. Tüm modellerin tahmin skorları birleştirilir. Sonuçta, tüm modellerin bilgisi birleştirilerek nihai tahmin oluşturulur. Bu skorlar, modelin bir gözlemin hangi sınıfa ait olduğunu belirlemede kullanılır.

4.4.5. Yinelemeli Sinir Ağı (Recurrent Neural Networks-RNN)

Yinelemeli Sinir Ağı (RNN), doğal dil işleme ve zaman serisi tahmini gibi ardışık verilerle çalışan makine öğrenimi görevlerinde sıklıkla kullanılan bir yapay sinir ağıdır [44]. RNN'lerin temel özelliği, önceki adımlardan gelen bilgileri hatırlayabilmesi ve bu sayede sıralı verileri etkili bir şekilde işleyebilmesidir. Bununla birlikte, RNN'ler bazı sınırlamalara sahiptir ve bu sorunları gidermek için Long Short Term Memory (LSTM) ve Gated Recurrent Unit (GRU) gibi gelişmiş varyantları daha yaygın olarak tercih edilmektedir. RNN'ler, eğitimi sırasında önceki adımlarda öğrenilen bilgileri belleğinde saklar ve bu

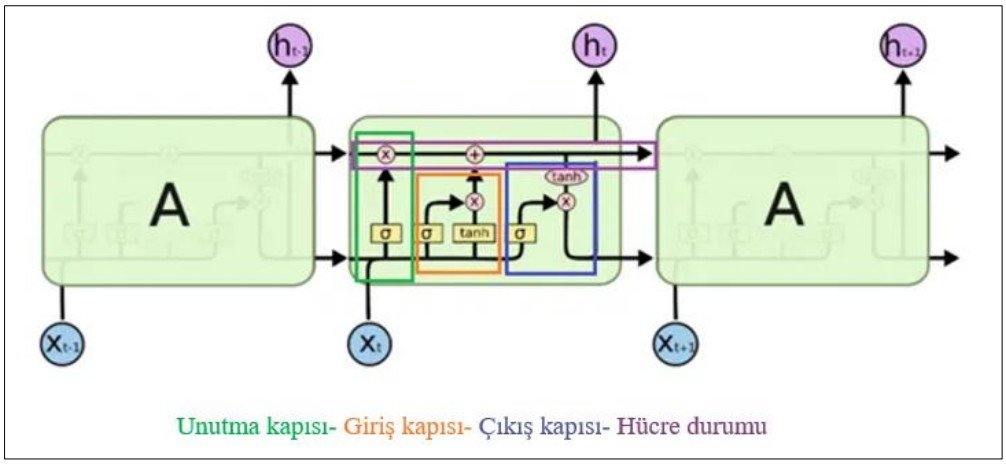
bilgileri yeni girişleri işlerken kullanır. Bu sayede, sıralı ilişkiler ve zaman içindeki değişimlerin etkili bir şekilde modellenmesini sağlar.

Günümüzde RNN'lerin bazı sınırlamaları vardır. Özellikle bağlamın çok uzun olduğu durumlarda, yani modelin karar verebilmesi için çok eski bilgilere ihtiyaç duyulduğunda, RNN'ler verimli olmayabilir. Ayrıca, birden fazla GPU'nun aynı anda paralel eğitimde kullanılmaması nedeniyle büyük modellerin eğitilmesi uzun süre alabilir. Bu sebeple, transformer tabanlı modeller popüler bir alternatif olarak öne çıkmaktadır.

- Gradyan kaybolması: Bu sorun, bir derin öğrenme modelinin eğitiminde kullanılan geri yayılım algoritması sırasında, gradyanın zamanla küçülmesi veya sıfıra yaklaşması durumudur. Gradyan kaybolması, ağın uzun vadeli bağımlılıkları doğru şekilde öğrenememesine yol açar ve bu da modelin eski bilgileri hatırlamakta zorluk çekmesine neden olur.
- Gradyan patlaması: Bu problemde ise gradyanlar aşırı büyüyerek model parametrelerinin gereksiz bir şekilde aşırı güncellenmesine yol açar. Genellikle çok sayıda katmanın üst üste eklenmesi veya tekrarlı yapılar kullanılması sonucunda ortaya çıkar. Gradyan patlaması, modelin aşırı büyük güncellemeler yaparak stabil olmayan sonuçlar üretmesine neden olabilir.
- Bu tür problemleri azaltmak veya ortadan kaldırmak için gradyan kesme (gradient clipping), aktivasyon fonksiyonlarının doğru seçimi (örneğin, ReLU), iyi başlangıç ağırlıkları seçimi ve özel RNN hücreleri (LSTM, GRU gibi) gibi yöntemler kullanılabilir.

Uzun Kısa Süreli Bellek (Long Short Term Memory)

LSTM (Long Short-Term Memory), RNN (Recurrent Neural Network) ailesine ait, sıralı verilerle çalışmak için kullanılan bir yapay sinir ağı mimarisidir. LSTM, özellikle uzun bağımlılıklara sahip dizileri işlemek için tasarlanmıştır. RNN'lerin "uzun vadeli bağımlılıkları öğrenmekte" yaşadığı problemleri (örneğin gradyan kaybı) çözmek üzere geliştirilmiştir [45]. Çalışma şekli aşağıda verilmiştir. LSTM, hafıza hücreleri (memory cells), giriş kapıları (input gates), çıkış kapıları (output gates) ve unutma kapıları (forget gates) gibi yapılar kullanarak sıralı veriler üzerinde daha güçlü öğrenme sağlar.



Şekil 4.8. LSTM katmanları

- Zaman t'deki girdi alınır.
- Unutma kapısı: Hücre durumundaki hangi bilgilerin unutulacağı belirlenir.
- Giriş kapısı: Yeni bilgiler alınır ve güncellenir.
- Hücre durumu: Hücre durumu güncellenir (eski bilgiler unutulur, yenileri eklenir).
- Çıkış kapısı: Çıkış oluşturulur ve bir sonraki zaman adımına (t+1) aktarılır.

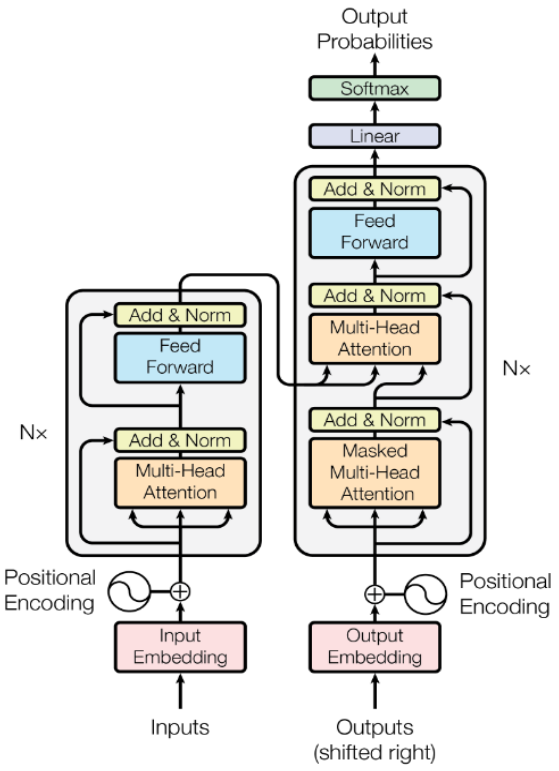
LSTM, kapılar aracılığıyla bilgi akışını kontrol eder ve hücre durumunu güncelleyerek uzun vadeli bağımlılıkları daha verimli bir şekilde öğrenir. Zaman içindeki bu uzun vadeli ilişkileri başarıyla modelleyebilmesi sayesinde, derin öğrenme uygulamalarında sıklıkla tercih edilen bir yapıdır.

BiLSTM (Bidirectional Long Short-Term Memory), klasik LSTM'nin çift yönlü bir versiyonudur. LSTM'ler, zaman serisi verilerinde veya metin gibi sıralı verilerde uzun vadeli bağımlılıkları öğrenmek için kullanılan geri beslemeli sinir ağlarıdır. BiLSTM, verinin sırasını hem ileri hem de geri yönde işleyerek bağlam bilgisini daha iyi öğrenir. Her iki yönden elde edilen çıktı vektörleri birleştirilerek modelin son katmanına iletilir. Bu yaklaşım, bir kelime veya zaman adımı için hem önceki hem de sonraki bağlam bilgisinin hesaba katılmasını sağlar.

4.4.6. Dönüştürücü Modeller (Transformers)

Dönüştürücüler, NLP ve diğer yapay zekâ alanlarında kullanılan derin öğrenme modelleridir. 2017'de Google tarafından geliştirilen ve özellikle dil anlama ve üretme görevlerinde büyük başarı elde eden bir mimari olan "Transformer" modeli, büyük bir dönüm noktası olmuştur [46].

Dönüştürücüler bir kodlayıcı (encoder), çözücü (decoder) yapısından oluşmaktadır. Burada amaç giriş matrisini kodlayıcıdan geçirerek barındırdığı tüm bilgileri sıkıştırılmış bir vektörle ifade etmek, başka bir deyişle giriş matrisinin sahip olduğu tüm bilgileri kodlamaktır. Çözücünün görevi ise bu kodlanmış vektörü kullanarak hedef matrisini oluşturmaktır. Transformer mimarisi Şekil 4.9.'te gösterilmektedir.



Şekil 4.9. Transformer mimarisi [47]

Geleneksel doğal dil işleme modelleri, uzun mesafeli bağımlılıkları ele almak için sınırlıdır ve büyük metinleri işlerken zorlanabilirler. Ancak, Transformer mimarisi ve dikkat mekanizması, bu uzun mesafeli bağımlılıkları daha etkili bir şekilde ele alır. Dikkat mekanizması sayesinde, bir kelimenin diğer kelimelerle olan ilişkisi daha iyi anlaşılır ve model, bağlamı daha doğru bir şekilde değerlendirebilir. Bu, uzun metinlerde daha iyi

performans sağlarken, karmaşık dil yapılarıyla daha iyi başa çıkılmasını sağlar. Transformer mimarisi, dikkat mekanizması sayesinde, büyük veri setleri ve karmaşık dil konularında daha üstün sonuçlar elde edilmesine olanak tanır.

BERT

BERT algoritması Google tarafından sorguları daha iyi anlamak ve kullanıcılarına daha doğru sonuçlar sunmak adına geliştirilmiştir. "Bidirectional Encoder Representations from Transformers" ifadelerinin baş harflerinden oluşan BERT algoritması, yapay zekâ ve makine öğrenimi teknolojilerini bir arada kullanan bir doğal dil işleme tekniğidir. BERT dil modeli Wikipedia'daki metinler kullanılarak önceden eğitilmiştir [47].

Google bu algoritma ile yapılan sorgulardaki kelimeleri tek tek değerlendirmek yerine önündeki ve arkasındaki ya da benzer ve eş anlamlı kelimeleri dikkate alarak değerlendirir. Google BERT algoritması, sorgularda kullanılan bağlaç ve edatların da ne anlama geldiğini daha iyi anlamakta ve bu bilgiyi sonuçların kalitesini artırmak için kullanmaktadır.

LegalBERT

LegalBERT, doğal dil işleme alanında kullanılan ve özellikle hukuk metinleri üzerinde çalışan bir derin öğrenme modelidir. BERT (Bidirectional Encoder Representations from Transformers), Google tarafından geliştirilen ve genel doğal dil işleme görevlerinde yüksek başarı elde eden bir modeldir. LegalBERT ise, bu genel modelin hukuk metinleri için özelleştirilmiş bir versiyonudur.

LegalBERT, hukuki metinlerin özelliklerine odaklanarak, bu tür metinlerin anlamını daha iyi yakalamak üzere eğitilmiştir. Hukuk metinleri, genellikle teknik terimler, uzun cümleler ve belirli bir bağlam içinde kullanılan dil özellikleri bakımından standart metinlerden farklılık gösterir. LegalBERT, bu özellikleri dikkate alarak, hukuki belgelerin daha etkili bir şekilde anlaşılması ve analiz edilmesi için tasarlanmıştır [48].

Kamuya açık kaynaklardan elde edilen mevzuat, sözleşme ve davalardan oluşan İngilizce yasal belgeler (toplam 12 GB veri) kullanılarak eğitilmiş olan LegalBERT eğitim korpusu Çizelge 4.3.'te verilmiştir.

Çizelge 4.3. LegalBERT eğitim korpusu [48]

Korpus	**Doküman Sayısı**
AB Mevzuatı (EURLEX)	116,062
Birleşik Krallık Mevzuatı	61,826
Avrupa Adalet Divanı Kararları	19,867
AİHM Kararları	12,554
ABD Mahkeme Kararları	164,141
ABD Sözleşmeleri	76,366

XLNet

XLNet, 2019 yılında Google tarafından geliştirilmiş bir dil modelidir ve BERT'in (Bidirectional Encoder Representations from Transformers) gelişmiş bir versiyonudur. XLNet, dil modelleme görevlerinde ve doğal dil işleme uygulamalarında başarılı sonuçlar elde etmek için kullanılır. XLNet, Transformer mimarisine dayalıdır. Transformer, özellikle dikkat (attention) mekanizmaları sayesinde, dil modellemesi ve diğer NLP görevlerinde güçlü performans gösterir. XLNet, bu yapıyı kullanarak dilin bağlamını öğrenir [49].

BERT, maskelenmiş dil modellemesi (MLM) kullanır, yani bazı kelimeler metinden çıkarılır ve model bu kelimeleri tahmin etmeye çalışır. XLNet, bu yaklaşımdan farklı olarak, tüm kelimeleri sırasız olarak permütasyonlarla öğrenir, bu da onun daha güçlü bir dil modellemesi yapmasına yardımcı olur [50].

4.5. Hiperparametre Optimizasyonu

Modelin hiperparametreleri öğrenme süreci başlamadan önce kullanıcı tarafından ayarlanabilen parametrelerdir. Hiperparametreler, bir modelin performansını artırmak için ayarlanması gereken parametrelerdir. Bunlar arasında öğrenme oranı, iterasyon sayısı, aktivasyon fonksiyonları, ağırlıklar, düzenlileştirme (regularization) parametreleri ve çıkış fonksiyonları gibi unsurlar bulunur [51]. Sınıflandırıcıyı daha da geliştirmek için her algoritma için en uygun parametrelerin seçildiği bir hiperparametre optimizasyonu işlemi gerçekleştirilmiştir. Bu süreç, modelin performansını artıran en iyi kombinasyonu bulmak

için önceden belirlenmiş bir hiperparametre uzayında arama yapılmasını içerir. Hiperparametre ayarı, modelin performansı üzerinde önemli bir etkiye sahiptir.

Rastgele Arama (Randomized SearchCV) makine öğrenimi modelleri için en uygun hiperparametreleri bulmak amacıyla kullanılan bir hiperparametre optimizasyon yöntemidir ve modelin performansını artırmak için en uygun hiperparametrelerin bulunmasına yardımcı olur. Izgara Arama (GridSearchCV) ile benzer bir mantıkta çalışır, ancak GridSearchCV'den farklı olarak tüm olası hiperparametre kombinasyonlarını denemek yerine, belirli sayıda rastgele seçilmiş kombinasyonları test eder [52]. Bu yaklaşım, özellikle büyük hiperparametre uzaylarında daha hızlı ve verimli sonuçlar elde etmeye olanak tanır.

Hyperopt, Python tabanlı bir açık kaynaklı optimizasyon kütüphanesidir. Temel amacı, makine öğrenimi modelleri için en uygun hiperparametreleri bulma sürecini kolaylaştırmak ve hızlandırmaktır. Hyperopt, farklı optimizasyon algoritmalarını destekler, ancak en popüler ve etkili olanı Bayesyen optimizasyonudur. Bayesyen optimizasyonu, karmaşık ve bilinmeyen fonksiyonların (bu durumda, modelin performansını ölçen fonksiyon) en iyi değerini bulmak için kullanılan güçlü bir yöntemdir. Bu yöntem, olasılık ve istatistik temelli bir yaklaşım olan Bayes teoremini kullanır. Amaç, deneme sayısını en aza indirerek, en iyi sonucu (genellikle en yüksek performansı veya en düşük hatayı) elde etmektir [53].

4.6. Veri Seti Dengeleme Teknikleri

4.6.1. Veri Artırma Teknikleri

Veri artırma, verileri gerçekten toplamadan, gerçek verilerin farklı versiyonlarını oluşturarak veri boyutunu yapay olarak artırmamızı sağlayan bir süreçtir. Bu işlem yetersiz veri problemini çözmeye çalışarak sınıflandırma performansının artırılması için yapılır. Oluşturulan artırılmış verilerin orijinaline ne çok benzer ne de çok farklı olmalıdır. Aksi takdirde bu durum aşırı uyum ya da düşük performansa yol açabilir. Metin artırmada kullanılan teknikler aşağıdaki şekilde verilmiştir.



Şekil 4.10. Metin verisi artırma teknikleri [54]

Basit veri artırımı

Veri artırma, metin verilerini zenginleştirip çeşitlendirmek amacıyla kullanılan etkili bir tekniktir ve genellikle metin sınıflandırma işlemlerinde veri setini genişletmek için tercih edilir. Bu yöntem, modelin performansını iyileştirmeyi hedefler.

- Eş anlamlı kelime ile değiştirme: Bu yöntem, her bir cümledeki rastgele seçilen kelimelerin anlamca benzer olan eş anlamlılarla değiştirilmesi işlemidir.
    - Kelime Gömmeye Dayalı Değiştirme: Kelime gömme (word embedding) modelleri (örneğin, Word2Vec, GloVe, FastText, BERT, ELMo vb.) kullanılarak her kelime, anlamını ve ilişkilerini belirten yoğun vektörler biçiminde temsil edilir. Bu vektörler arasındaki benzerlikler (örneğin, kosinüs benzerliği) kullanılarak bir kelimenin anlamına en yakın olan eşdeğerleri belirlenir ve bu kelimeler orijinal metinde uygun şekilde değiştirilir. Bu sayede metnin anlamı bozulmadan çeşitlilik sağlanmış olur.
    - Sözcük Tabanlı Değiştirme: Bu teknik, kelimeleri, alt ve üst kavramlarıyla birlikte içeren, o dile özgü veri tabanlarını kullanarak yapılır. Örneğin, İngilizce metinlerde, WordNet gibi bir kaynak kullanılarak bir kelimenin eş anlamlıları tespit edilir ve orijinal metinde istenilen kelimenin yerine bu eş anlamlılar eklenir.
- Rastgele Ekleme, Silme, Değiştirme: Bu yöntem, metindeki kelimeler üzerinde rastgele değişiklikler yaparak veri çeşitliliği sağlar. Örneğin, belirli bir kelimenin eş

anlamlısı metne rastgele bir yere eklenebilir, metindeki kelimelerden bazıları rastgele silinebilir veya iki kelimenin yerleri değiştirilebilir.

Geri Çeviri (Backtranslation)

Geri çeviri, bir metnin önce hedef dile çevrilip sonra tekrar kaynak dile çevrilmesiyle veri setinin çeşitlendirilmesini sağlayan bir tekniktir. Bu yöntem, modelin daha etkili bir şekilde genelleme yapabilmesini sağlamak amacıyla veri setindeki çeşitliliği artırır ve modelin daha güçlü bir şekilde eğitilmesine yardımcı olur.

Üretken Modeller

Üretken dil modelleri, metin üretme görevinde kullanılan güçlü araçlardır. BERT, RoBERTa, BART ve T5 gibi modeller, belirli bir metin sınıfını koruyarak yeni metin örnekleri üretmek için eğitim alır. Bu modeller, belirli bir sınıf etiketini ve o sınıfa ait örnek metinleri kodlayarak, yeni örnekler oluşturur. Böylece, metinler daha zengin ve çeşitli hale gelir, bu da modelin daha esnek ve genelleyici olmasına katkı sağlar.

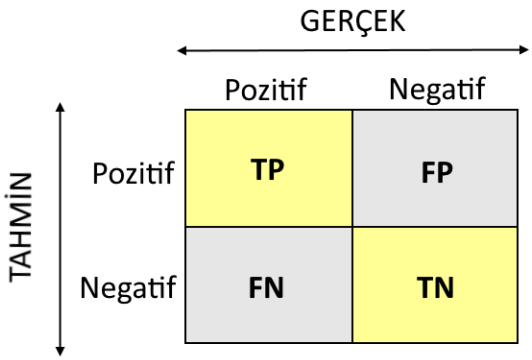
4.6.2. Aşağı Örnekleme (Downsampling)

Aşağı örneklemede çoğunluk sınıfındaki örneklerin sayısı azaltılarak veri seti dengelenmeye çalışılır. Çoğunluk sınıfından rastgele örnekler silinir veya alt küme alınır. Azınlık sınıfıyla eşit veya daha dengeli bir dağılım oluşturulur. Modelin çoğunluk sınıfını aşırı öğrenmesi önlenir, ancak silinen veriler değerli olabileceği için önemli bilgiler kaybedilebilir.

4.7. Model Değerlendirme

4.7.1. Performans Metrikleri

Sınıflandırma modellerinin tahmin başarısının ölçülmesi için çeşitli sınıflandırma metrikleri kullanılır. Bu metrikler hesaplanırken karışıklık matrisi dikkate alınır. Karışıklık matrisi aşağıdaki şekildedir.



Şekil 4.11. İkili sınıflandırma için karışıklık matrisi

- Doğru Pozitif (DP): Gerçekte ihlal olan ve model tarafından doğru bir şekilde ihlal olarak tahmin edilen dava sayısı.
- Yanlış Pozitif (YP): Gerçekte ihlal olmayan ve model tarafından ihlal olarak tahmin edilen dava sayısı.
- Doğru Negatif (DN): Gerçekte ihlal olmayan ve model tarafından doğru bir şekilde ihlal değil olarak tahmin edilen dava sayısı.
- Yanlış Negatif (YN): Gerçekte ihlal olan ve model tarafından ihlal değil olarak tahmin edilen dava sayısını ifade etmektedir.

Yukarıda verilen tanımlamalara göre performans değerlendirme metriklerine ait hesaplama formülleri aşağıda verilmiştir:

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

$$Duyarlılık = \frac{DP}{DP + YN} \quad (2)$$

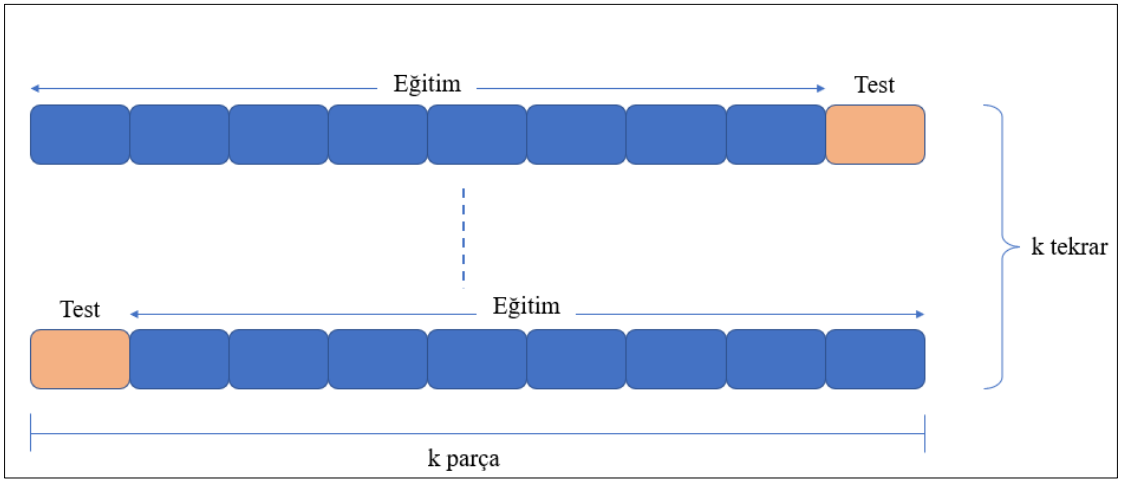
$$Kesinlik = \frac{DP}{DP + YP} \quad (3)$$

$$F1 = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik \ + Duyarlılık} \quad (4)$$

4.7.2. K-Kat Çapraz Geçerleme

Çapraz geçerleme, makine öğrenimi modellerinin doğruluğunu ve genelleme yeteneğini değerlendirmek için kullanılan bir istatistiksel tekniktir. Veri kümesi eğitim ve test seti olarak iki ayrı parçaya bölünüp model performansı değerlendirilebilir ancak bu yöntemde modelin performansı yalnızca bir kez değerlendirilmiş olur. Elde edilen sonuçların şans eseri olup olmadığı doğrulanamaz. Çapraz geçerleme yönteminde model birden fazla kez değerlendirerek model performansı hakkında daha fazla güvence elde edilir.

K-katlı çapraz geçerleme, en yaygın çapraz geçerleme tekniklerinden biridir. Bu yöntem, genellikle veri kümesinin kaç gruba bölüneceğini belirten tek bir parametreye sahiptir: k. K için belirli bir değer seçildiğinde, bu değer yönteme atıfta bulunurken kullanılabilir; örneğin, k=5 seçildiğinde, yöntem 5-katlı çapraz geçerleme olarak ifade edilir. Veri seti k farklı alt kümeye bölünür. Her yinelemede bir küme test için kullanılırken, kalan kümeler eğitim için kullanılır. İşlem k defa tekrarlanır ve her küme bir kez test seti görevi görür. K ayrı deney sonucunda nihai model performansı için elde edilen performans değerlerinin ortalaması alınır.

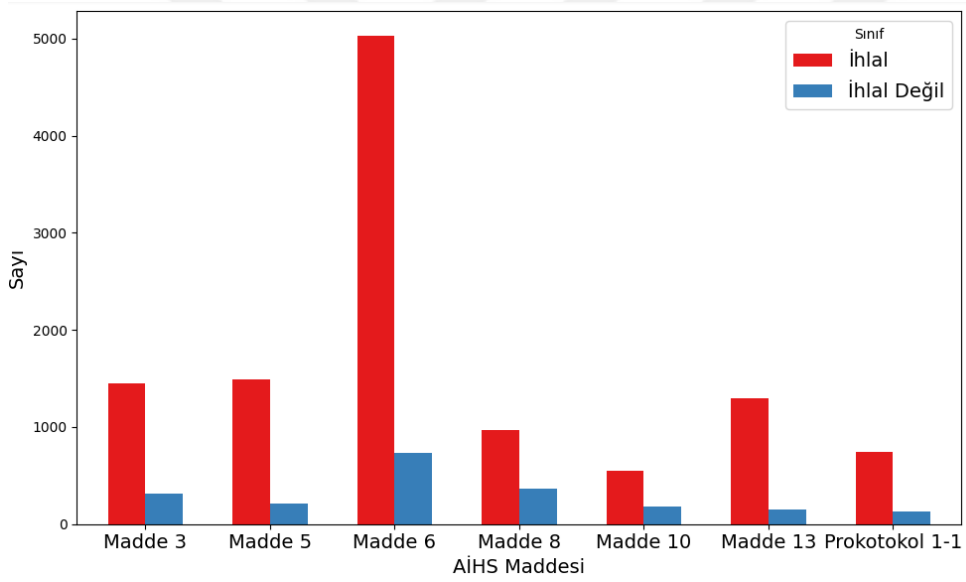


Şekil 4.12. K-kat çapraz geçerleme süreci

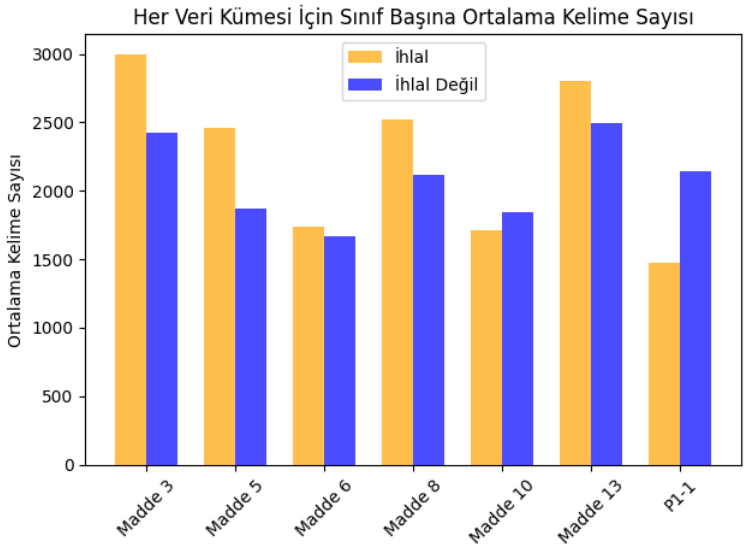
5. DENEYSEL ÇALIŞMA VE SONUÇLARI

5.1. Veri Kümesinin Hazırlanması

Çizelge 4.1. de her sözleşme maddesi için HUDOC veri tabanından elde edilen dava sayıları görülmektedir. İhlal ve ihlal değil olarak etiketlenen dava sayılarının oldukça dengesiz olduğu görülmektedir. Çalışma kapsamına dava sayısının en fazla olduğu Madde 3, 5, 6, 8, 10, 13 ve Protokol 1'in 1.maddesine ilişkin davalar alınmıştır. Kapsama yalnızca İngilizce dilinde yayımlanmış olan metinler dahil edilmiştir. Kapsama alınan AİHS maddelerine göre elde edilen dava sayısı Şekil 5.1.'de ve bu maddeler kapsamında yer alan dokümanlardaki İhlal ve ihlal değil sınıflarındaki ortalama kelime sayıları ve Şekil 5.2.'de verilmiştir.



Şekil 5.1. Kapsama alınan AİHS maddelerine göre sınıf dağılımı



Şekil 5.2. Dokümanlardaki İhlal ve İhlal Değil sınıflarına ait ortalama kelime sayısı

Dava dokümanları önceki bölümlerde açıklandığı üzere HUDOC veri tabanı üzerinde yayımlanmış olan dava metinlerinden alınarak oluşturulmuştur. Bu metinler ilgili dava metinlerdeki THE FACTS (Olaylar) bölümlerinden oluşmaktadır. Dolayısıyla metinlerden çalışma için gerekli olmayan diğer bölümler (Giriş, Prosedür, Karar vs. bölümleri) çıkartılmıştır. Aşağıda HUDOC veri tabanındaki dava metinlerinden bir tanesi için kapsama alınan bölüm ve Türkçe çevirisi aşağıda yer almaktadır.

THE FACTS

2. The applicant was born in 1966 and lives in Istanbul. He was represented by Mr E.U. Yıldırım, a lawyer practising in Istanbul.
3. The Government were represented by their Agent, Mr Hacı Ali Açıkgül, Head of the Department of Human Rights of the Ministry of Justice of the Republic of Türkiye.
4. The facts of the case may be summarised as follows.
5. On 16 August 2006 the Istanbul Assize Court convicted the applicant and his wife of aiding and abetting a terrorist organisation. The Assize court sentenced each of them to three years and nine months' imprisonment.
6. On 5 January 2009 the applicant's wife started to serve her prison sentence in Bakırköy Prison. On 2 March 2009 she was transferred to Gebze Prison together with their four-year-old son. At her request, they were placed in cell no. 5 of Gebze Prison, which could accommodate twelve persons, instead of a cell where those convicted of terrorism-related offences were held.
7. On 5 January 2009 the applicant started to serve his sentence in Metris Prison. On 29 January 2009 he was transferred to Amasya Prison.
8. On an unspecified date, in order to have contact with his wife and son, the applicant applied to the General Directorate of Prisons and Detention Centres of the Ministry of Justice to be transferred to Gebze Prison, where they were staying. The directorate allowed his request and on 2 November 2009 the applicant was transferred to Gebze Prison. He was placed in cell no. B-1, which was designated for those convicted of terrorism-related offences.
9. On 13 November 2009 the applicant filed a petition with the Gebze enforcement judge. In his petition, he noted that he had been transferred, upon his request, to Gebze penitentiary institution where his wife was being accommodated with their son in order to have more frequent meetings with them and that in this institution, he was able to receive non-contact visits from his wife and his son and he would also be able to receive contact visits from them once a month. Nevertheless, he complained that while his child, who was in prison with his mother because he needed care at the time, was allowed to stay with his mother in the women's section of the same prison under the relevant national regulation, there was no provision for the child to spend time with his father. The applicant contended that this legislative approach was sexist and ran counter to the logic and, in his view, failed to take into account the psychological well-being of the child. The applicant further argued that the parents' inability to spend time with their son all together in a common area would have negative consequences for the child. In the light of the foregoing, the applicant requested that the enforcement judge order the prison administration to create an appropriate environment in which he could meet with his child during working hours (during the day) and to organise regular sessions when the child could have contact with both of his parents at the same time.
10. On the same day the administration and observation board of Gebze Prison ("the observation board"), which consisted of the prison governor, deputy governor, psychologist, doctor, teacher and chief officer of the execution guard, examined the applicant's request. In its decision, the observation board took note of Article 70 § 1 of the Regulation on prison management and the execution of sentences and preventive measures, which stated that children under the age of six and in need of care could stay with their mother in prison (see paragraph 17 below). Having regard to that provision, the observation board decided not to allow the child to stay in the cell where his father (the applicant) was staying. However, on humane grounds, it ordered that the father be allowed to have physical contact with his child in the visiting area for one hour during his non-contact visits, which were to take place every Thursday. Furthermore, the observation board decided that the child and the parents could meet for one hour during monthly visits.
11. Following the observation board's above-mentioned decision, in a decision of 4 December 2009, the Gebze enforcement judge examined the applicant's petition of 13 November 2009 (see paragraph 9 above). The enforcement judge dismissed his request, holding that the observation board's decision had been in compliance with the law.
12. On 16 December 2009 the applicant lodged an objection against the enforcement judge's decision with the Gebze Assize Court. In his objection, the applicant stated that he had not requested that his son stay in his cell or in his care, but merely that he be allowed to meet with his son at certain times of the day during the week in appropriate areas, such as in visiting facilities or recreation rooms of the prison. He submitted that the regulation providing that children should stay with their mother in prison did not explicitly prevent their staying with their fathers. Thus, it was possible to interpret the legislation in his favour and allow him to spend more time with his son. He stated that as a convicted person, he was entitled to spend ten hours a week socialising with other convicted prisoners in an open area of the prison, and as he was not allowed to enjoy that right because of the prison regime which he was subject to, he could spend that time with his child instead. He lastly asked that his request be assessed not only in respect of his situation, but also taking into account the rights of his child.

OLAYLAR

2. Başvuran, 1966 doğumlu olup; İstanbul'da ikamet etmektedir. Başvuran, İstanbul Barosuna kayıtlı Avukat E.U. Yıldırım tarafından temsil edilmiştir.
3. Hükümet, kendi görevlisi Türkiye Cumhuriyeti Adalet Bakanlığı İnsan Hakları Dairesi Başkanı Hacı Ali Açıkgül tarafından temsil edilmiştir.
4. Dava konusu olaylar aşağıdaki gibi özetlenebilir.
5. İstanbul Ağır Ceza Mahkemesi, 16 Ağustos 2006 tarihinde, başvuran ve eşini, terör örgütüne yardım ve yataklık etme suçundan mahkûm etmiştir. Ağır Ceza Mahkemesi, başvuran ve eşinin her birine üç yıl dokuz ay hapis cezası verilmesine hükmetmiştir.
6. Başvuranın eşinin hapis cezası, 5 Ocak 2009 tarihinde, Bakırköy Ceza İnfaz Kurumunda infaz edilmeye başlanmıştır. Başvuranın eşi, 2 Mart 2009 tarihinde, dört yaşındaki oğluyla birlikte Gebze Ceza İnfaz Kurumuna nakledilmiştir. Başvuranın eşinin talebi üzerine, oğlu ile birlikte, Gebze Ceza İnfaz Kurumunda, terörle ilgili suçlardan hüküm giyenlerin tutulduğu hücre yerine, on iki kişinin barındırılabileceği 5 no.lu hücreye yerleştirilmiştir.
7. Başvuranın cezası, 5 Ocak 2009 tarihinde, Metris Ceza İnfaz Kurumunda infaz edilmeye başlanmıştır. Başvuran, 29 Ocak 2009 tarihinde, Amasya Ceza İnfaz Kurumuna nakledilmiştir.
8. Sonrasında, başvuran, eşi ve oğluyla görüşebilmek için, kaldıkları Gebze Ceza İnfaz Kurumuna nakledilmek üzere Adalet Bakanlığı Ceza ve Tevkifevleri Genel Müdürlüğüne başvuruda bulunmuştur. Genel Müdürlük başvuranın talebini kabul etmiş ve başvuran, 2 Kasım 2009 tarihinde, Gebze Ceza İnfaz Kurumuna nakledilmiştir. Başvuran, terörle ilgili suçlardan hüküm giyenler için ayrılan B-1 no.lu hücreye yerleştirilmiştir.
9. Başvuran, 13 Kasım 2009 tarihinde, Gebze İnfaz Hâkimliğine bir dilekçe sunmuştur. Başvuran, dilekçesinde, talebi üzerine, eşiyle ve oğluyla daha sık görüşebilmek için kaldıkları Gebze Ceza İnfaz Kurumuna nakledildiğini ve bu kurumda eşi ve oğluyla kapalı görüş yapabildiğini, ayrıca ayda bir kez de onlarla açık görüş yapabileceğini belirtmiştir. Bununla birlikte, başvuran, söz konusu zamanda bakıma ihtiyacı olduğu için annesiyle birlikte ceza infaz kurumunda bulunan çocuğunun, ilgili Tüzük uyarınca aynı ceza infaz kurumunun kadınlar bölümünde annesiyle birlikte kalmasına izin verilmesine karşın, çocuğun babasıyla vakit geçirmesine yönelik herhangi bir hüküm bulunmadığından şikâyetçi olmuştur. Başvuran, bu yasal yaklaşımın cinsiyetçi ve mantığa aykırı olduğunu ve kendi kanaatine göre çocuğun psikolojik sağlığını dikkate almadığını ileri sürmüştür. Başvuran, ayrıca, ebeveynlerin oğullarıyla birlikte ortak bir alanda vakit geçirememesinin çocuk açısından olumsuz sonuçlar doğuracağını iddia etmiştir. Yukarıdaki hususlar ışığında, başvuran, infaz hâkimliğinden, ceza infaz kurumu idaresine kendisinin mesai saatleri içerisinde (gün içinde) çocuğuyla görüşebileceği uygun bir ortam sağlanması ve çocuğun her iki ebeveyniyle aynı anda iletişim kurabileceği düzenli görüşmeler yapılması yönünde talimat vermesini talep etmiştir.
10. Aynı gün, ceza infaz kurumu müdürü, müdür yardımcısı, psikolog, doktor, öğretmen ve infaz koruma baş memurundan oluşan Gebze Ceza İnfaz Kurumu İdare ve Gözlem Kurulu ("Gözlem Kurulu"), başvuranın talebini incelemiştir. Gözlem Kurulu, kararında, altı yaşından küçük ve bakıma muhtaç çocukların ceza infaz kurumunda anneleriyle birlikte kalabileceğini düzenleyen Ceza İnfaz Kurumlarının Yönetimi ile Ceza ve Güvenlik Tedbirlerinin İnfazı Hakkında Tüzük'ün 70 § 1 maddesini dikkate almıştır (bk. aşağıda 17. paragraf). Bu hükmü göz önünde bulunduran Gözlem Kurulu, çocuğun babasının (başvuranın) kaldığı hücrede kalmasına izin vermemeye karar vermiştir. Bununla birlikte, Gözlem Kurulu, insani gerekçelerle, babanın her Perşembe gerçekleşecek olan kapalı görüş sırasında bir saat boyunca ziyaret bölümünde çocuğuyla fiziksel temas kurmasına izin verilmesine karar vermiştir. Ayrıca, Gözlem Kurulu, çocuk ve ebeveynlerin aylık görüşler sırasında bir saat görüşebileceklerine karar vermiştir.
11. Gözlem Kurulunun yukarıda belirtilen kararının ardından, 4 Aralık 2009 tarihli bir kararla, Gebze İnfaz Hâkimliği, başvuranın 13 Kasım 2009 tarihli dilekçesini incelemiştir (bk. yukarıda 9. paragraf). İnfaz Hâkimliği, Gözlem Kurulunun kararının hukuka uygun olduğuna hükmederek başvuranın talebini reddetmiştir.
12. Başvuran, 16 Aralık 2009 tarihinde, İnfaz Hâkimliğinin kararına karşı Gebze Ağır Ceza Mahkemesine itirazda bulunmuştur. Başvuran, itirazında, oğlunun kendi hücresinde ya da bakımı altında kalmasını talep etmediğini, sadece hafta boyunca günün belirli saatlerinde, ceza infaz kurumunun ziyaret bölümleri ya da hobi odaları gibi uygun alanlarda oğluyla görüşmesine izin verilmesini talep ettiğini belirtmiştir. Başvuran, çocukların ceza infaz kurumunda anneleriyle birlikte kalmalarını öngören tüzüğün, çocukların babalarıyla birlikte kalmalarını açıkça engellemediğini ileri sürmüştür. Dolayısıyla, mevzuatın kendi lehine yorumlanması ve oğluyla daha fazla zaman geçirmesine izin verilmesi mümkündü. Başvuran, hükümlü bir kişi olarak, haftada on saatini ceza infaz kurumunu açık bir alanında diğer hükümlülerle sosyalleşerek geçirme hakkına sahip olduğunu ve tabi olduğu ceza infaz kurumu rejimi nedeniyle bu haktan yararlanamaması nedeniyle, bu zamanı çocuğuyla geçirebileceğini belirtmiştir. Başvuran, son olarak, talebinin yalnızca kendi durumu açısından değil, aynı zamanda çocuğunun hakları da dikkate alınarak değerlendirilmesini istemiştir.

Şekil 5.3. Metinlerde kapsama alınan Olaylar bölümü (İngilizce ve Türkçe)

Her bir dava metni Bölüm 4.2. de bahsedilen önişleme adımlarından geçirilmiştir. Çalışmada hem geleneksel makine öğrenmesi yöntemleri hem derin öğrenme yöntemleri kullanılmıştır. Özellikle derin öğrenmede fazla önişleme yapmak sınıflandırma performansında olumsuz sonuçlara yol açabilmektedir. Örneğin metinde kelimeleri kök haline getirmek modelin anlamlı ayırt edici özellikler bulmasını zorlaştırabilmektedir. Derin öğrenme modelleri genellikle ham veriden önemli özellikleri öğrenebilir. Ön işlemeyi fazla yapmak, verinin anlamını bozabilmektedir. Dolayısıyla kullanılacak derin öğrenme modelleri için kullanılacak veriler de düşünülerek Şekil 5.4.'te görüldüğü üzere veride yapılan önişleme faaliyetleri ayrı sütunlar halinde kaydedilmiştir. Metinlerin ilk çekildiği (Extracted Text), kelimelerin kök haline getirildiği (extracted_text_lem), durak kelimelerin kaldırıldığı (extrated_text_removeSW), durak kelimelerin kaldırılıp kelimelerin kök haline getirildiği (extracted_text_removeSW_lem) halleri verilmiştir. Derin öğrenme yöntemlerinde extracted_text_removeSW sütunundaki metinler kullanıldığında en iyi sonuçlar alınmıştır, dolayısıyla bu yöntemlerde bu sütundaki veriler kullanılmıştır. Diğer makine öğrenmesi yöntemlerinde extracted_text_removeSW_lem sütunundaki veriler kullanıldığında en iyi sonuçlar elde edilmişitir. Derin öğrenme modelleri, veriyi kendi başına işleyerek önemli özellikleri öğrenebilir ve bu nedenle elle yapılan bu ön işleme adımlarına duyulan ihtiyaç azalmaktadır. Önişleme fazla olduğunda özellik kaybı olmakta ve derin öğrenme modellerinin başarısı düşmektedir.

File Name	Extracted Text	extracted_text_lem	extracted_text_removeSW	extracted_text_removeSW_lem	Sınıf
001-100036.docx	the applicant was born in and lives in iauliai...	applicant bear live iauliai fact case submit p...	applicant born lives iauliai facts case submit...	applicant bear live iauliai fact case submit p...	0
001-100393.docx	the applicant was born in and lives in he stud...	applicant bear live study educational science ...	applicant born lives studied educational scien...	applicant bear life study educational science ...	0
001-100407.docx	the applicant was born in and lives in ple na ...	applicant bear live ple na applicant brother s...	applicant born lives ple na applicant brother ...	applicant bear live ple applicant brother leas...	0
001-100461.docx	the applicant was born in and lives in rostov ...	applicant bear live rostov don applicant russi...	applicant born lives rostov applicant russian ...	applicant bear live rostov applicant russian a...	0
001-100472.docx	the applicant was born in and lives in uspensk...	applicant bear live uspenskoye krasnodar regio...	applicant born lives uspenskoye krasnodar regi...	applicant bear live uspenskoye krasnodar regio...	0

Şekil 5.4. Orijinal veride yapılan önişleme faaliyetleri

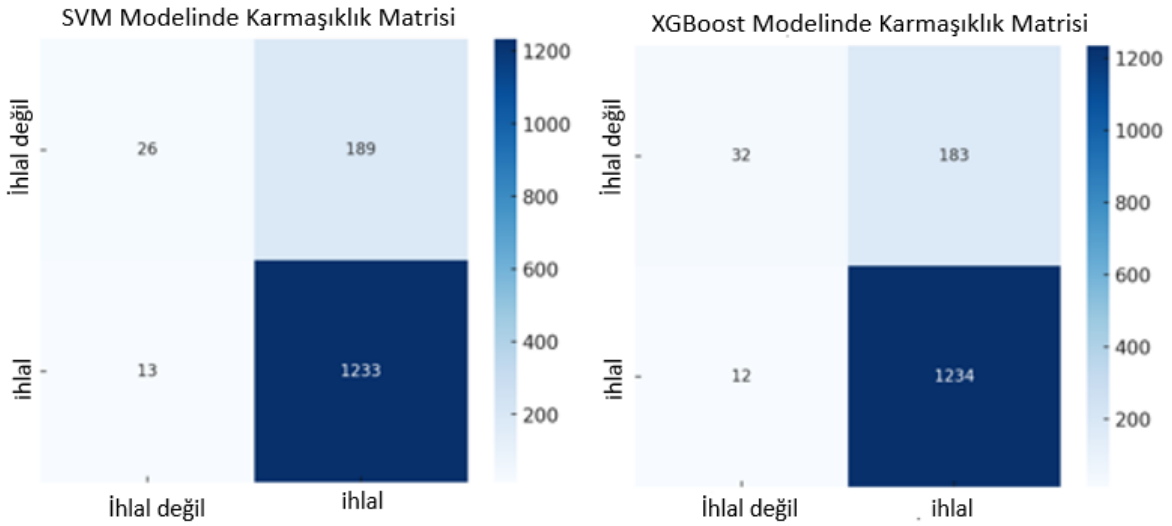
Veri setinde (Madde 3, 5, 6, 8, 10, 13 ve Protokol 1'in 1.maddesine ilişkin davalar) Sınıf değişkeni, tahmin için kullanılacaktır ve "ihlal" ile sonuçlanan davalar için 1, "ihlal değil" olarak sonuçlanan davalar için 0 olarak tanımlanmıştır. Nihai olarak her madde için İhlal ve İhlal Değil şeklinde 2 etiketi olan Sınıf bağımlı değişkeni ve önişleme yapılmış dava metni bağımsız değişkeni olmak üzere toplamda 2 değişkenli veri kümeleri (her madde için ayrı

olarak) sınıflandırma algoritmalarında girdi olarak kullanılmıştır.

5.2. Dava Sonuçlarının Tahmin Edilmesi

Yapılan deneysel çalışmalarda metin önişleme yöntemi olarak SVM, Lojistik Regresyon, XGBoost, Multinomial Naive Bayes TF-IDF; LSTM için GloVe metin temsil yöntemi kullanılmış ve modeller eğitilmiştir. XGBoost algoritması için HyperOpt hiperparametre optimizasyonu kullanılmış, diğer sınıflandırma algoritmaları için RandomizedSearchCV yöntemi kullanılmıştır. 5 katlı çapraz geçerleme kullanılarak model performansları çeşitli metriklerle değerlendirilmiştir. LSTM algoritmasında uzun optimizasyon süresi dikkate alınarak hiperparametre optimizasyon yöntemleri uygulanmamış, modelin eğitim sürecini optimize etmek amacıyla öğrenme oranını (learning rate) dinamik olarak değiştiren bir teknik (LR Scheduler) kullanılmıştır. Ayrıca modelin geçerleme kaybı iyileşmediği takdirde eğitim durdurulmuştur.

Elde edilen veri kümesinde sınıflar arası dengesizlik mevcuttur. Madde 6 ait tüm veriler kullanılarak bir ön çalışma yapılmış ve dengesiz veri setinin modellerin başarısı üzerindeki etkisi gözlemlenmiştir. (Şekil 5.5.)



Şekil 5.5. Madde 6 tüm veri üzerinde SVM ve XGBoost Karmaşıklık Matrisleri

Şekil 5.5. de görüldüğü üzere her iki model de Sınıf 1 (ihlal) için iyi performans göstermiştir, ancak Sınıf 0 (ihlal değil) üzerinde her iki model de zayıf kalmıştır. Bunun sebebi Sınıf 0'ın az sayıda örneğe sahip olması ve modelin Sınıf 1'i öğrenmeye odaklanmasıdır. Modelin Sınıf 1 örneklerini neredeyse her zaman doğru sınıflandırdığı Sınıf 0 için ise modelin bu sınıfı çoğunlukla göz ardı ettiği görülmektedir.

Dengesiz veri setinden kaynaklanan bu sonuçlardan hareketle modelin sınıf 0 üzerindeki başarısını artırmak için farklı stratejiler denenmiştir. İlk yöntemde downsampling yapılarak, veri setindeki çoğunluk sınıfının (Sınıf 1) azaltılıp modelin daha dengeli bir veri setiyle eğitilmesi sağlanmıştır. Bu, özellikle dengesiz veri setlerinde modelin daha adil tahminler yapmasına yardımcı olmaktadır. Yöntemlerden ikincisinde ise nadir örneğe sahip olan sınıfta (Sınıf 0) veri artırma yaparak modelin daha fazla çeşitlendirilmiş verilerle eğitilmesi sağlanmıştır. Sonraki bölümlerde bu yöntemlere ilişkin gerçekleştirilen çalışmalar verilmiştir.

5.2.1. Deneysel Çalışma 1

Bu deneyde 0 etiketli veriler 1 etiketli verilerden daha az olduğundan veri setini dengelemek için pozitif sınıfa ait örneklerden negatif sınıfa ait örnekler kadar örnek alınarak (downsampling) veri setleri dengelenmiştir. Nihai durumda her Sözleşme maddesi için kullanılan doküman sayıları aşağıda verilmiştir.

Çizelge 5.1. Dengelenmiş veri seti doküman sayıları

Sözleşme Maddesi	**Doküman Sayısı**	
	İhlal	**İhlal Değil**
Madde 3	334	334
Madde 5	223	223
Madde 6	733	733
Madde 8	440	440
Madde 10	186	186
Madde 13	149	149
Protokol 1(1. madde)	128	128

Kullanılan yöntemlerle her sözleşme maddesi için elde edilen optimum parametrelerle yapılan 5-kat çapraz geçerlemede elde edilen ortalama performans metrikleri ve eğitim hiperparametre optimizasyon süresi toplamı aşağıdadır.

Destek Vektör Makineleri Sonuçları

Çizelge 5.2. Dengelenmiş veri SVM sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	359	0.71	0.70	0.74	0.72	0.78
Madde 5	174	0.82	0.79	0.88	0.83	0.86
Madde 6	1040	0.69	0.70	0.67	0.68	0.74
Madde 8	515	0.70	0.70	0.70	0.71	0.76
Madde 10	134	0.61	0.61	0.59	0.60	0.63
Madde 13	127	0.73	0.74	0.73	0.78	0.78
Protokol 1 (1)	90	0.72	0.72	0.74	0.73	0.81

Lojistik Regresyon Sonuçları

Çizelge 5.3. Dengelenmiş veri Lojistik Regresyon sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	147	0.75	0.74	0.77	0.76	0.83
Madde 5	78	0.72	0.72	0.74	0.73	0.80
Madde 6	216	0.74	0.76	0.70	0.73	0.81
Madde 8	158	0.71	0.72	0.69	0.71	0.77
Madde 10	70	0.60	0.60	0.59	0.60	0.64
Madde 13	64	0.80	0.81	0.79	0.80	0.86
Protokol 1 (1)	39	0.74	0.75	0.72	0.74	0.80

XGBoost Sonuçları

Çizelge 5.4. Dengelenmiş veri XGBoost sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	217	0.70	0.70	0.73	0.71	0.78
Madde 5	126	0.71	0.72	0.70	0.71	0.77
Madde 6	374	0.71	0.71	0.69	0.70	0.76
Madde 8	185	0.63	0.63	0.62	0.62	0.68
Madde 10	87	0.55	0.55	0.62	0.58	0.57
Madde 13	175	0.71	0.72	0.70	0.71	0.76
Protokol 1 (1)	55	0.71	0.70	0.75	0.72	0.77

Multinomial Naive Bayes Sonuçları

Çizelge 5.5. Dengelenmiş veri Multinomial Naive Bayes sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	203	0.74	0.70	0.87	0.77	0.84
Madde 5	106	0.74	0.72	0.79	0.75	0.79
Madde 6	293	0.72	0.73	0.70	0.72	0.80
Madde 8	217	0.70	0.71	0.68	0.69	0.74
Madde 10	100	0.61	0.61	0.61	0.61	0.62
Madde 13	84	0.79	0.79	0.79	0.79	0.87
Protokol 1 (1)	52	0.72	0.71	0.74	0.73	0.81

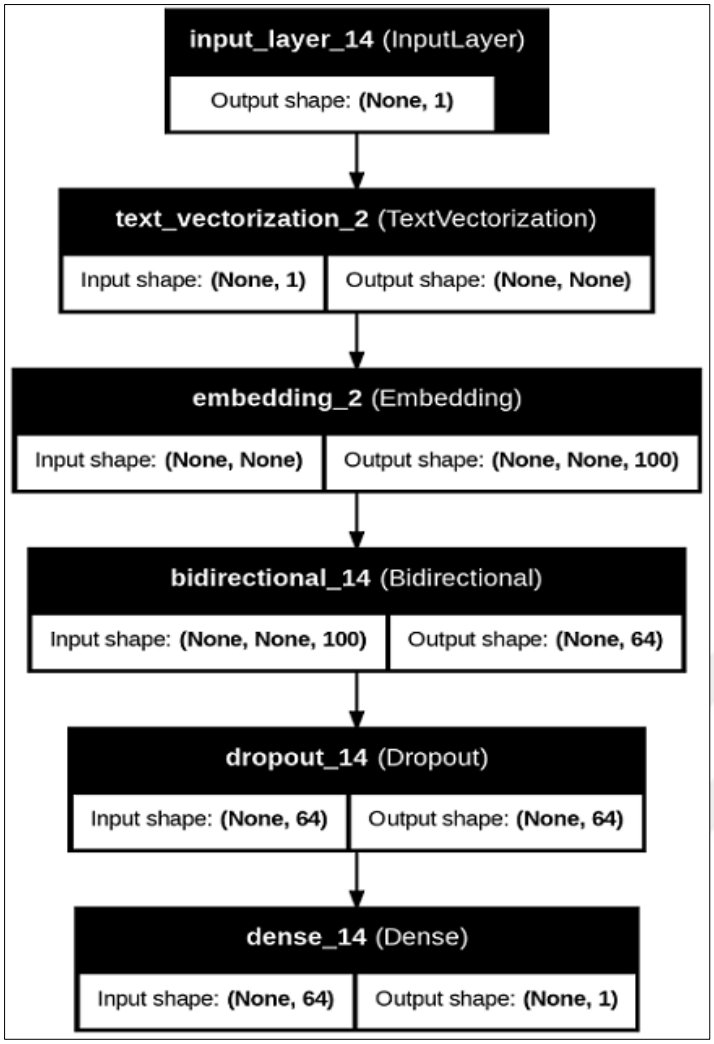
LSTM Sonuçları

Bu aşamada GloVe önceden eğitilmiş kelime gömme (embedding) vektörleri kullanılarak bir derin öğrenme modeli oluşturulmuş ve 5-kat çapraz geçerleme yöntemiyle performansı

test edilmiştir. GloVe, önceden eğitilmiş kelime vektörleri sağladığından bu, modelin her bir kelimeye ait semantik anlamları öğrenmesine yardımcı olmaktadır. Her bir kelimeyi sayısal bir vektör olarak ifade etmek için 100 boyutlu vektör karşılığı kullanılmıştır. Model yapısı aşağıda verilmiştir. Eğitim sırasında, model doğrulama başarımı düşmeye başladığında durdurulmuştur, bu da aşırı öğrenmeyi engellemektedir.

- İlk katman: TextVectorization (Metinleri vektör hâline getiriyor).
- Gömme Katmanı: Embedding (Kelime vektörlerini GloVe'den alıyor).
- İki yönlü LSTM: Bidirectional LSTM (Metnin hem baştan hem sondan anlamını öğreniyor).
- Dropout: Aşırı öğrenmeyi (overfitting) önlemek için belirli nöronları rastgele devre dışı bırakır.
- Çıkış Katmanı: Sigmoid aktivasyon fonksiyonu ile bir sınıflandırma sonucu üretiyor (0 veya 1).

Metin girişlerinin alındığı, vektörlere dönüştürüldüğü ve LSTM katmanında bağlamsal özellikleri öğrenerek bir sınıflandırma sonucu (ihlal var/yok) ürettiği modele ilişkin mimari aşağıda verilmiştir.



Şekil 5.6. Önerilen BiLSTM mimarisi

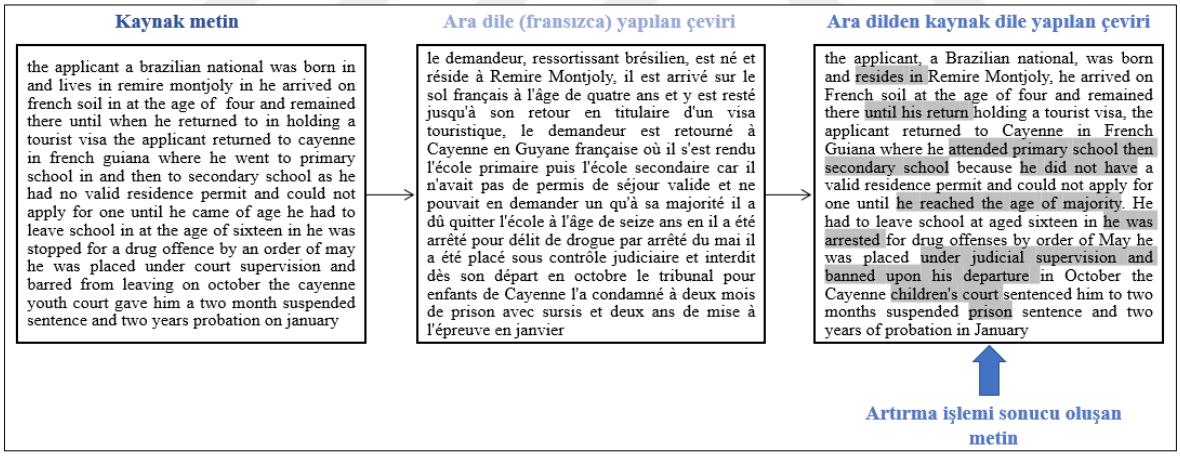
Çizelge 5.6. Dengelenmiş veri LSTM sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	776	0.75	0.73	0.81	0.77	0.75
Madde 5	222	0.93	0.95	0.91	0.93	0.96
Madde 6	238	0.68	0.67	0.76	0.71	0.68
Madde 8	942	0.74	0.78	0.67	0.72	0.74
Madde 10	179	0.55	0.56	0.71	0.60	0.60
Madde 13	74	0.63	0.57	0.67	0.60	0.60
Protokol 1 (1)	83	0.55	0.54	0.50	0.50	0.55

5.2.2. Deneysel Çalışma 2

Veri dengesizliği problemine çözüm olarak sunulan veri artırma yaklaşımı kullanılmıştır. Çalışmada 5-kat çapraz geçerleme sırasında her kat için eğitim verileri artırılmıştır. Bunun nedeni, modelin performansını ve genelleme yeteneğini artırmaktır. Bu özellikle dengesiz veri setlerinde önemlidir. Çalışmada nadir olan sınıfa (ihlal değil) ait veriler artırılmış, bunun için metinlerde geri çeviri (back translation) yöntemi kullanılmıştır. Bu yöntem, mevcut metni bir ara dile çevirip tekrar orijinal dile çevirerek metin varyasyonu oluşturmayı hedeflemektedir. Google Translate API'si [55] kullanılmıştır. Bunun için metin önce ara bir dile (Fransızca), ardından tekrar İngilizceye çevrilmektedir. Geri çeviri, metni yapısal olarak farklılaştırır ve veri çeşitliliği sağlar. Bu sayede, metnin anlamını bozmadan yeni ve çeşitlendirilmiş veri örnekleri oluşturulmuştur.

Geri çeviri işleminde orijinal metin, ara dile yapılan çeviri ve tekrar İngilizceye yapılan çeviri için bu işlemin yapıldığı örnek bir metinden alınan kesit Şekil 5.6.'da verilmiştir. Artırma sonucu oluşan metinde değişen bölümler vurgulanmıştır.



Şekil 5.7. Geri çeviri (Backtranslation) süreci

Sözleşme maddelerine ilişkin metinler 5-kat çapraz geçerleme yapılırken her katta eğitim setindeki 0 (ihlal değil) etiketli veriler, eğitim setindeki 1 etiketli verilerle aynı sayıda olacak şekilde artırılarak etiketlere ilişkin denge kurulmaya çalışılmıştır. Nihai durumda etiketlere ilişkin veri sayıları aşağıdaki gibi olmuştur. Veri artırma sadece eğitim setinde yapılmıştır, test seti orijinal haliyle bırakılmıştır. Çünkü eğitim verisi, modelin öğrendiği ve genelleme yaptığı kaynaktır. Veri artırma, eğitim verisinin çeşitliliğini artırarak modelin daha iyi genelleme yapmasını sağlar. Ancak test verisi, modelin doğruluğunu ölçmek için kullanılır

ve bu veri üzerinde yapılan herhangi bir değişiklik, modelin test setine ne kadar iyi genelleme yaptığına dair yanıltıcı sonuçlar verebilir. Test seti, eğitim sırasında görülmemesi gereken ve modelin daha önce karşılaşmadığı verileri temsil eder. Eğer test verisi üzerinde veri artırma yapılırsa, test seti modelin eğitim sürecine dahil edilmiş olur, bu da aşırı öğrenme (overfitting) riskini artırabilir. Bu durumda model, gerçek dünyadaki yeni ve farklı verilerle karşılaştığında daha düşük performans gösterebilir.

Çizelge 5.7. Artırılmış veri doküman sayıları

Sözleşme Maddesi	**Etiket**	**Doküman Sayısı**	
		Eğitim	**Test**
Madde 3	İhlal	534	134
	İhlal değil	267+**267**	67
Madde 5	İhlal	1170	292
	İhlal değil	179+**178**	44
Madde 6	İhlal	1169	293
	İhlal değil	585+**584**	146
Madde 8	İhlal	421	105
	İhlal değil	352+**69**	88
Madde 10	İhlal	297	75
	İhlal değil	149+**148**	37
Madde 13	İhlal	238	60
	İhlal değil	119+**119**	30
Protokol 1 (1)	İhlal	205	51
	İhlal değil	103+**102**	25

Destek Vektör Makineleri Sonuçları

Çizelge 5.8. Artırma yapılmış veride SVM sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	574	0.76	0.78	0.90	0.84	0.78
Madde 5	159	0.86	0.89	0.91	0.90	0.93
Madde 6	916	0.77	0.80	0.87	0.83	0.82
Madde 8	473	0.70	0.71	0.75	0.73	0.76
Madde 10	174	0.58	0.73	0.58	0.65	0.62
Madde 13	0.78	0.78	0.83	0.84	0.84	0.84
Protokol 1 (1)	47	0.70	0.71	0.93	0.80	0.79

Lojistik Regresyon Sonuçları

Çizelge 5.9. Artırma yapılmış veride Lojistik Regresyon sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	224	0.75	0.77	0.90	0.83	0.76
Madde 5	76	0.86	0.85	0.94	0.90	0.92
Madde 6	130	0.74	0.79	0.83	0.81	0.80
Madde 8	178	0.68	0.70	0.72	0.71	0.73
Madde 10	63	0.62	0.68	0.82	0.74	0.61
Madde 13	44	0.78	0.81	0.87	0.84	0.83
Protokol 1 (1)	27	0.71	0.74	0.85	0.79	0.77

XGBoost Sonuçları

Çizelge 5.10. Artırma yapılmış veride XGBoost sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	290	0.75	0.76	0.91	0.85	0.80
Madde 5	140	0.87	0.88	0.93	0.92	0.84
Madde 6	420	0.76	0.76	0.74	0.75	0.83
Madde 8	205	0.69	0.70	0.69	0.69	0.75
Madde 10	105	0.65	0.64	0.70	0.67	0.66
Madde 13	190	0.76	0.77	0.75	0.76	0.83
Protokol 1 (1)	65	0.74	0.77	0.80	0.78	0.77

Multinomial Naive Bayes Sonuçları

Çizelge 5.11. Artırma yapılmış veride Multinomial Naive Bayes sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	255	0.78	0.75	0.89	0.81	0.87
Madde 5	135	0.86	0.85	0.92	0.91	0.93
Madde 6	355	0.75	0.76	0.74	0.75	0.83
Madde 8	275	0.74	0.75	0.72	0.73	0.80
Madde 10	130	0.67	0.66	0.67	0.66	0.68
Madde 13	105	0.82	0.82	0.82	0.82	0.89
Protokol 1 (1)	55	0.72	0.70	0.92	0.81	0.80

LSTM Sonuçları

Çizelge 5.12. Artırma yapılmış veride LSTM sonuçları

Sözleşme Maddesi	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**	**AUC**
Madde 3	364	0.78	0.70	0.85	0.79	0.81
Madde 5	355	0.97	0.96	0.98	0.96	0.96
Madde 6	375	0.82	0.71	0.92	0.85	0.85
Madde 8	556	0.74	0.60	0.90	0.70	0.70
Madde 10	269	0.64	0.70	0.90	0.79	0.65
Madde 13	117	0.77	0.70	0.93	0.80	0.71
Protokol 1 (1)	125	0.69	0.69	0.92	0.79	0.60

Madde 6 için artırılmış veride LSTM sonuçlarına ilişkin grafikler ve karmaşıklık matrisleri Ekler bölümünde sunulmuştur.

5.2.3. Deneysel Çalışma 3

Bu deneysel çalışmada performans artışı hedeflenerek dönüştürücü yöntemler kullanılmış ve sonuçları gözlemlenmiştir. Modellerde kullanılan parametreler aşağıdadır:

- max_len: Metinlerin pad edileceği veya kesileceği maksimum uzunluk. Bu parametre, metinlerin modelin anlayabileceği belirli bir uzunluğa getirilmesini sağlar. 500 olarak belirlenmiştir.
- lr (learning rate): AdamW optimizasyonu için öğrenme hızı (learning rate). Burada 2e-5 olarak belirtilmiş.
- batch_size: Eğitim ve doğrulama sırasında her bir küme için kullanılan örnek sayısı. 8 olarak belirlenmiştir.
- num_labels: Çıktı sınıf sayısı.2 olarak belirlenmiştir.
- n_splits: Çapraz geçerleme için katman sayısı. 5 kat çapraz geçerleme kullanılmıştır.
- epochs: Her bir katmanda modelin eğitim iterasyon sayısı (epoch). 10 olarak belirlenmiştir.
- patience: Erken durdurma için limit parametresi. Eğer doğrulama kaybı belirtilen

sayıda iterasyon boyunca iyileşmezse, eğitim durdurulmaktadır. 3 olarak ayarlanmıştır.

Dönüştürücü Modellerin Sonuçları

Algoritmalar en fazla veriye sahip olan Madde 6 için uygulanmıştır. Bu algoritmalar yüksek maliyetler ve eğitim sürelerinden dolayı yalnızca Madde 6 verileri kullanılmış ve performans metriklerine ilişkin sonuçlar aşağıda sunulmuştur.

Dönüştürücülere ilişkin karmaşıklık matrisleri ve grafikler Ekler bölümünde verilmiştir.

Çizelge 5.13. Dengelenmiş veride Dönüştürücü modellerin sonuçları

Sözleşme Maddesi	**Model**	**Toplam Eğitim Süresi (s)**	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1 Skor**
Madde 6	LegalBERT	2.266,32	**0.94**	**0,94**	**0,94**	**0,94**
Madde 6	XLNet	4.683,01	0.93	0,93	0,93	0,93
Madde 6	BERT	2.369,33	0.93	0,93	0,93	0,93

6. SONUÇ

Bu tez çalışması, AİHM kararlarının makine öğrenmesi ve doğal dil işleme NLP teknikleri kullanılarak tahmin edilmesine odaklanmıştır. AİHM kararlarının tahmini, hukuki süreçlerin otomatikleştirilmesi ve karar alma mekanizmalarının desteklenmesi açısından büyük bir potansiyel taşımaktadır. Kullanılan makine öğrenimi modelleri, AİHM kararlarının tahmin edilmesinde potansiyel bir uygulama alanı sunmuştur. Çalışma, özellikle doğal dil işleme ve makine öğrenimi algoritmalarının, hukuki metinlerin analizinde etkin bir şekilde kullanılabileceğini göstermiştir. Ayrıca elde edilen sonuçlar, AİHM kararlarının, ikili sınıflandırma yöntemleriyle doğru bir şekilde tahmin edilebileceğini ve bu tür tekniklerin hukuki alanda kullanımını mümkün kılabileceğini ortaya koymaktadır. AİHM kararlarının tahmin edilmesinin, mahkemelerin daha hızlı ve doğru kararlar almasına yardımcı olacağı, ayrıca hukuki sistemin şeffaflığını ve erişilebilirliğini artırılmasını sağlayacağı değerlendirilmektedir.

AİHM kararlarının veri setinde, ihlal kararları ile ihlal olmayan kararlar arasında belirgin bir dengesizlik bulunmaktadır. Tez çalışması kapsamında ilk olarak dengesiz veri seti kullanılarak elde edilen modellerde nadir olan sınıfa ait performansların düşük olduğu gözlemlenmiş (Şekil 5.5.) ve sınıf dengesizliği problemine çözüm aranmıştır.

Downsampling yapılarak uygulanan deney çalışmasında AİHM kararlarının ikili sınıflandırılmasında dengelenmiş veri kullanılarak makine öğrenimi modelleriyle gerçekleştirilen bir tahmin analizi gerçekleştirilmiştir. Deneyde veri dengesizliği sorununu çözmek için, pozitif sınıfa ait örneklerin sayısının negatif sınıfa eşitlenmesi amacıyla veri seti dengelenmiştir. Bu yaklaşım, her iki sınıfın (ihlal ve ihlal değil) temsil edilme oranlarının eşitlenmesini sağlamış ve modelin öğrenme sürecinde daha iyi genelleme kapasitesi oluşturmuştur. Lojistik Regresyon ve Multinomial Naive Bayes, hızlı eğitim süreleriyle dikkat çekmektedir. SVM bazı maddelerde (örneğin Madde 13 ve Protokol 1(1)) oldukça iyi sonuçlar vermiştir. Dezavantajı, daha uzun eğitim süresi gerektirmesidir.

Veri artırılarak uygulanan deney çalışmasında veri dengesizliği problemine çözüm olarak sunulan geri çeviri (back-translation) yaklaşımı kullanılmıştır. 5 kat çapraz geçerleme yapılmış ve her katta eğitim setindeki daha az örneğe sahip olan 0 etiketli veriler ilgili yöntem ile artırılmıştır. Böylece, modelin daha fazla ve çeşitlendirilmiş eğitim örneği

üzerinde öğrenmesini sağlanarak, genelleme kapasitesini artırması amaçlanmıştır. Verilerin artırılarak dengelenmesi, veri yetersizliği kaynaklı performans sorunlarını azaltmış ve çoğu modelin doğruluk ve diğer metriklerinde iyileşme sağlamıştır. Geri çeviri yöntemi, metinlerin anlamını bozmadan yeni örnekler üretmiş ve bu da modellerin genelleme yeteneğini güçlendirmiştir. Veri artırma sonrası en başarılı model, genel olarak LSTM algoritması olmuştur. Bu durum, LSTM algoritmasının ardışık verilerdeki uzun vadeli bağımlılıkları öğrenme yeteneği sayesinde hukuki metinlerin yapısal ve anlamsal özelliklerini daha iyi yakalamasından kaynaklanmaktadır.

Son deneysel çalışmada dengeli veride dönüştürücü modeller uygulanmış ve bu modellerin oldukça güçlü performans sergilediği görülmüştür. Kullanılan LegalBERT algoritması hukuki bağlamda özelleşmiş bir modeldir ve Madde 6 verilerinde %94 doğruluk ve %94 F1 skoru ile oldukça yüksek bir performans sergilemiştir. Ayrıca duyarlılık ve kesinlik arasındaki denge dikkat çekicidir. Ancak bu modellerin eğitim süresi geleneksel makine öğrenmesi yöntemlerine göre çok daha uzundur. Çünkü bunlar milyonlarca parametreye sahip büyük modellerdir. Bu modeller, her bir kelimenin bağlamını anlamak için dikkat mekanizması kullanır. Bu mekanizma, her kelimenin diğer tüm kelimelerle etkileşimlerini hesaplar, bu da işlem yükünü artırır.

Yapılan deneyler sonucunda dönüştürücü modellerin başarısı göze çarpmaktadır. Özellikle LegalBERT gibi hukuki bağlamda özelleşmiş modellerin yüksek doğruluk ve F1 skorları ile başarılı sonuçlar verdiği görülmüştür. Bununla birlikte, bu modellerin eğitim sürelerinin geleneksel makine öğrenmesi algoritmalarına göre çok daha uzun olması, dikkat edilmesi gereken bir faktördür.

Çizelge 6.1, bu çalışmanın ve farklı çalışmaların yöntemlerini, kullanılan veri setlerini ve başarı oranlarını karşılaştırmalı olarak sunmaktadır.

Çizelge 6.1. Literatür ve bu çalışmanın karşılaştırılması

Çalışma	**Yıl**	**Kullanılan Veri Seti**	**Kullanılan Yöntemler**	**Başarı Oranı**
Aletras et al. (2016) [8]	2016	AİHM kararları	SVM, N-gram temsili	%79 doğruluk
Sulea et al. (2017) [9]	2017	Fransız Yüksek Mahkemesi kararları	SVM, ensembler modeller	%98 F1 skoru
Medvedeva et al. (2020) [15]	2020	AİHM kararları	Derin öğrenme modelleri	%75 doğruluk
Delhi Bölge Mahkemesi Çalışması (2020) [17]	2020	Delhi Bölge Mahkemesi davaları	CART, SVM, RF, Boosting	%91 doğruluk, %91 F1 skoru (CART modeli)
Brezilya Federal Mahkemesi Çalışması (2021) [19]	2021	Brezilya Federal Küçük Davalar Temyiz Dosyaları (612.961 dava)	ULMFiT, BERT, Big Bird	MCC değeri: 0.3688 (modeller), 0.1253 (hukuk uzmanları)
Henrik Hodnefjeld (2022) [22]	2022	AİHM kararları	Transformer tabanlı modeller (BERT, RoBERTa, XLNet)	%88 doğruluk
Bu tez çalışması (2025)	2025	AİHM kararları	SVM, Naive Bayes, LSTM, XGBoost, Lojistik Regresyon, Transformer tabanlı modeller (BERT, LegalBERT, XLNet)	LegalBERT: %94 doğruluk, %94 F1 skoru

Tez çalışmasında ulaşılan en önemli bulgulardan biri, veri artırma tekniklerinin model başarısını olumlu yönde etkilediğidir. Geri çeviri veri artırma tekniğinin kullanılması, modelin daha fazla örnekle eğitilmesini sağlamış ve başarıyı yükseltmiştir.

Elde edilen bulgular, makine öğrenimi ve doğal dil işleme tekniklerinin hukuki metinlerin analizinde etkin bir şekilde kullanılabileceğini ve AİHM kararlarının tahmin edilmesinin hukuki sistemin şeffaflık ve erişilebilirliğini artırabileceğini göstermektedir. Bu çalışma, gelecekteki araştırmalara, hukuki alanda daha verimli ve hızlı karar destek sistemlerinin geliştirilmesine yönelik önemli bir temel sunmaktadır.

Bu çalışmada, AİHM kararlarının "Olaylar" bölümü kullanılmıştır. Ancak gelecekteki çalışmalar, AİHM'ne yapılan başvuru metinleri üzerinde tahmin yapılmasına odaklanabilir, bu nedenle, AİHM tarafından başvuru metinlerinin erişilebilir hale getirilmesi, gelecekteki

çalışmalar için önemli bir adım olacaktır. AİHM kararlarının yalnızca ikili sınıflandırma değil, çok etiketli sınıflandırma (multi-label classification) teknikleri ile incelenmesinin, farklı karar türleri ve sonuçları hakkında daha ayrıntılı tahminler yapılmasına olanak tanıyabileceği değerlendirilmektedir. Ayrıca AİHM'nin İngilizce dışındaki resmi dili olan Fransızca dilinde yayımlanmış dava metinleri de kapsama alınarak veri seti zenginleştirilebilir. Şeffaflık ve hesap verebilirlik açısından Açıklanabilir Yapay Zekâ (Explainable AI-XAI) yöntemleri, modellerinin nasıl karar verdiğini anlamak için kullanılabilir. Modellerin ırk, cinsiyet, ülke gibi değişkenlere dayalı ayrımcılık yapıp yapmadığı test edilebilir. Örneğin, model bazı ülkelerden gelen başvurular için daha yüksek "ihlal var" tahmini yapıyorsa, modelin önyargılı olup olmadığı araştırılmalıdır.

Bu tez çalışması, AİHM kararlarının tahmin edilmesinde makine öğrenmesi ve doğal dil işleme tekniklerinin etkin bir şekilde kullanılabileceğini göstermiştir. Elde edilen bulgular, hukuki süreçlerde veri madenciliği yöntemlerinin daha yaygın kullanımını teşvik edebilir ve hukuk teknolojileri alanında yeni araştırmalara yol açabilir. Özellikle LegalBERT gibi hukuki alanda özelleşmiş modeller, hukuki metinlerin analizi ve sınıflandırılmasında büyük bir potansiyel sunmaktadır. Gelecekteki çalışmalar, bu tür modellerin daha da geliştirilmesi ve hukuki süreçlerde yaygın olarak kullanılması için önemli bir temel oluşturmaktadır.

KAYNAKLAR

1. Arthur, R. S. (1977). The computer and the practice of law: Litigation support. American Bar Association Journal, 63(12), 1737–1742.

2. Lawlor, R. C. (1963). What computers can do: Analysis and prediction of judicial decisions. American Bar Association Journal, 49(4), 337–344.

3. Nagel, S. S. (1963). Applying correlation analysis to case prediction. Tex. L. Rev., 42, 1006.

4. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

5. Hearst, M. A. (1999, June). Untangling text data mining. In Proceedings of the 37th Annual meeting of the Association for Computational Linguistics (pp. 3-10).

6. Ashley, K. D. (2017). Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press.

7. Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT press.

8. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ Computer Science, 2, e93.

9. Sulea, O. M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., & Van Genabith, J. (2017). Exploring the use of text classification in the legal domain. arXiv preprint arXiv:1710.09306.

10. Lippi, M., & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. ACM Transactions on Internet Technology (TOIT), 16(2), 1-25.

11. Avrupa İnsan Hakları Mahkemesi. (2023). HUDOC-Avrupa İnsan Hakları Mahkemesi Karar Veritabanı. https://hudoc.echr.coe.int. Son Erişim Tarihi: 23 Haziran 2023.

12. Avrupa İnsan Hakları Mahkemesi. (2023). AİHM 2023 yıllık raporu. Erişim adresi: https://www.echr.coe.int/documents/d/echr/annual-report-2024-eng. Son Erişim Tarihi: 3 Şubat 2025

13. González-González, J., d'Arriba-Pérez, F., García-Méndez, S., Busto-Castiñeira, A., & González-Castaño, F. J. (2023). Automatic explanation of the classification of Spanish legal judgments in jurisdiction-dependent law categories with tree estimators. Journal of King Saud University, 35(7), 1541-1551.

14. Aydın, Ö. (2020). Mobbing içerikli yargı kararlarının makine öğrenmesi algoritmaları ile sınıflandırılması (Yayınlanmamış yüksek lisans tezi). Balıkesir Üniversitesi, Balıkesir, Türkiye.

15. Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. Artificial Intelligence and Law, 28(2), 237-266.

16. Görentaş, M. B., & Uçkan, T. (2023). Makine öğrenmesi yöntemleri kullanılarak mahkeme kararlarının kümelenmesi. Journal of Computer Science, 8(2), 148-158.

17. Shaikh, R. A., Sahu, T. P., & Anand, V. (2020). Predicting outcomes of legal cases based on legal factors using classifiers. Procedia Computer Science, 176, 2393-2402.

18. Sert, M. F., Yıldırım, E., & Haşlak, I. (2021). Using artificial intelligence to predict decisions of the Turkish Constitutional Court. Social Science Computer Review, 40(6), 1416-1435.

19. D. Menezes-Neto, E. J., & M. B. M. Clementino. (2022). Using deep learning to predict outcomes of legal appeals better than human experts: A study with data from Brazilian federal courts. PLoS ONE, 17(7), e0269675.

20. Malik, V., Sanjay, R., Nigam, S. K., Ghosh, K., Guha, S. K., Bhattacharya, A., & Modi, A. (2021). ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (pp. 141-151).

21. Alghazzawi, D., Bamasag, O., Albeshri, A., Sana, I., Ullah, H., & Asghar, M. Z. (2022). Efficient prediction of court judgments using an LSTM+CNN neural network model with an optimal feature set. Mathematics, 10(5), 1345.

22. Henrik, H. (2022). Classifying European Court of Human Rights cases using transformer based models (Master's thesis, NTNU).

23. European Union. EUR-Lex: Access to European Union law. Retrieved February 12, 2025, from https://eur-lex.europa.eu/homepage.html. Son Erişim Tarihi: 23 Haziran 2024.

24. Supreme Court of the United States. Supreme Court of the United States. Retrieved February 12, 2025, from https://www.supremecourt.gov/search_center.aspx. Son Erişim Tarihi: 20 Temmuz 2024.

25. Harvard Law School. Caselaw Access Project. Retrieved February 12, 2025, from https://case.law/. Son Erişim Tarihi: 10 Temmuz 2024.

26. Le service public de la diffusion du droit. Legifrance. Retrieved February 12, 2025, from https://www.legifrance.gouv.fr/. Son Erişim Tarihi: 1 Eylül 2024.

27. European Parliament. European Parliament Plenary. Retrieved February 12, 2025, from https://www.europarl.europa.eu/plenary/en/home.html. Son Erişim Tarihi: 4 Temmuz 2024.

28. Open Australian Legal Corpus. OpenAusLC: Open Australian Legal Corpus. Retrieved February 12, 2025, from https://huggingface.co/datasets/isaacus/open-australian-legal-corpus. Son Erişim Tarihi: 5 Ağustos 2024.

29. Anayasa Mahkemesi. Avrupa İnsan Hakları Mahkemesi yapısı. Erişim adresi: https://www.anayasa.gov.tr/tr/yayinlar/insan-haklari-bilgi-bankasi/avrupa-insan-haklari-mahkemesi/aihm-yapisi/.

30. European Court of Human Rights. (2021). 50 questions about the European Court of Human Rights: A guide for the general public. Erişim adresi: https://www.echr.coe.int/documents/d/echr/50Questions_TUR.

31. European Court of Human Rights. (1950, November 4). İnsan Hakları ve Temel (Avrupa İnsan Hakları Sözleşmesi). Erişim adresi: https://www.echr.coe.int/documents/d/echr/convention_tur. Son Erişim Tarihi: 4 Eylül 2024.

32. European Court of Human Rights. Rules of Court. Erişim adresi: https://www.echr.coe.int/documents/d/echr/Rules_Court_ENG. Son Erişim Tarihi: 14 Ekim 2024.

33. Alparslan, G. (2023). Makine öğrenmesi kullanarak doküman sınıflandırma. Ankara, Türkiye.

34. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Proceedings of the International Conference on Learning Representations.

35. Stanford University. (n.d.). GloVe: Global vectors for word representation. Retrieved December 20, 2024, from https://nlp.stanford.edu/projects/glove/. Son Erişim Tarihi: 15 Ağustos 2024.

36. Cortes, C. (1995). Support-Vector Networks. Machine Learning.

37. Imran, A. S., & Kastrati, Z. (2022). Classifying European Court of Human Rights cases using transformer-based models. arXiv preprint arXiv:2202.02375.

38. McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In AAAI-98 workshop on learning for text categorization (Vol. 752, No. 1, pp. 41-48).

39. Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.

40. Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B: Statistical Methodology, 20(2), 215-232.

41. Medium. (2020, March 15). Derivative of the sigmoid function. Erişim adresi: https://towardsdatascience.com/derivative-of-the-sigmoid-function-536880cf918e.

42. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

43. Rahaman, M. N., Chaki, S., Biswas, M. S., Biswas, M., Ahmed, S., Mahi, M. J. N., & Faruqui, N. (2022, April). Identifying the signature of suicidality: A machine learning approach. In THEETAS 2022: Proceedings of The International Conference on Emerging Trends in Artificial Intelligence and Smart Systems, THEETAS 2022, 16-17 April 2022 (p. 279). European Alliance for Innovation.

44. Goodfellow, I. (2016). Deep learning.

45. Hochreiter, S. (1997). Long Short-term Memory. Neural Computation MIT-Press.

46. Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

47. [Hugging Face. Hugging Face Model Hub. Erişim adresi: https://huggingface.co/. Son Erişim Tarihi: 23 Kasım 2024.

48. Hugging Face. Legal BERT-based uncased model. Erişim adresi: https://huggingface.co/amanbawa96/legal-bert-based-uncase. Son Erişim Tarihi: 23 Kasım 2024.

49. Yang, Z. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. arXiv preprint arXiv:1906.08237.

50. Hugging Face. XLNet-base-cased model. Erişim adresi: https://huggingface.co/xlnet/xlnet-base-cased. Son Erişim Tarihi: 23 Eylül 2024.

51. Emeç, M., & Özcanhan, M. H. Makine öğrenmesi algoritmalarında hiperparametre belirleme.

52. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of machine learning research, 13(2).

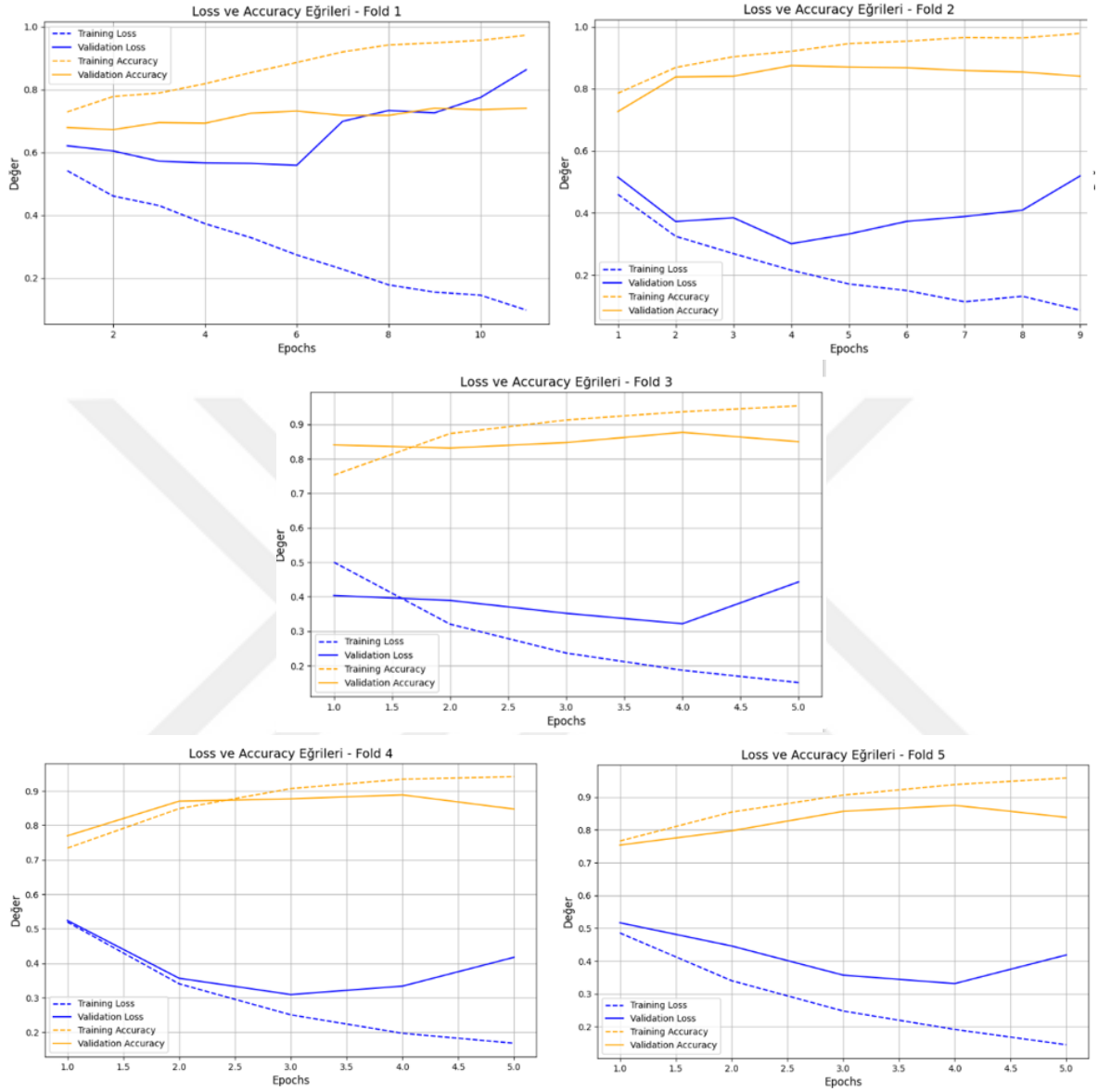
53. Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. Automated machine learning: Methods, systems, challenges, 3-33.

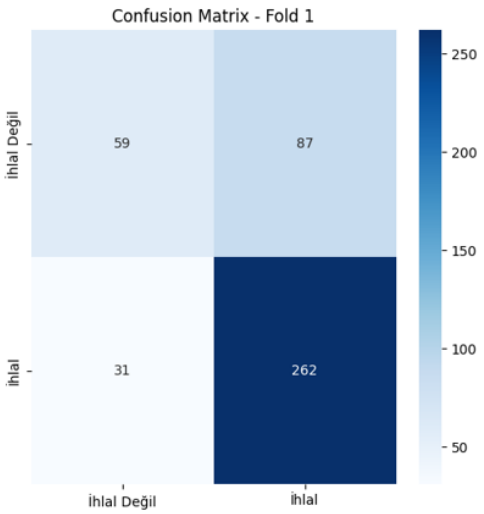
54. GeeksforGeeks. Text augmentation techniques in NLP. Erişim adresi: https://www.geeksforgeeks.org/text-augmentation-techniques-in-nlp/. Son Erişim Tarihi: 24 Aralık 2024.

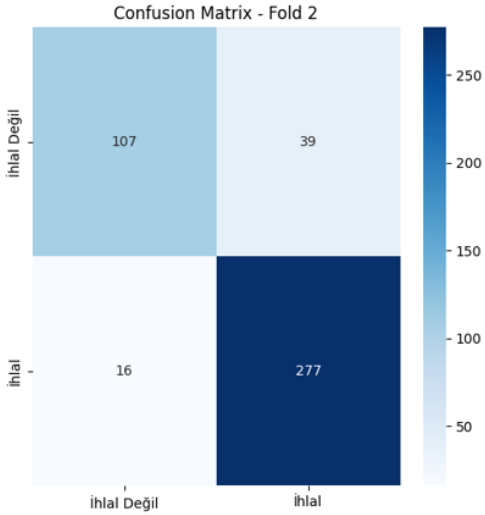
55. Google. Google Translate API. Erişim adresi: https://cloud.google.com/translate. Son Erişim Tarihi: 15 Ocak 2024.

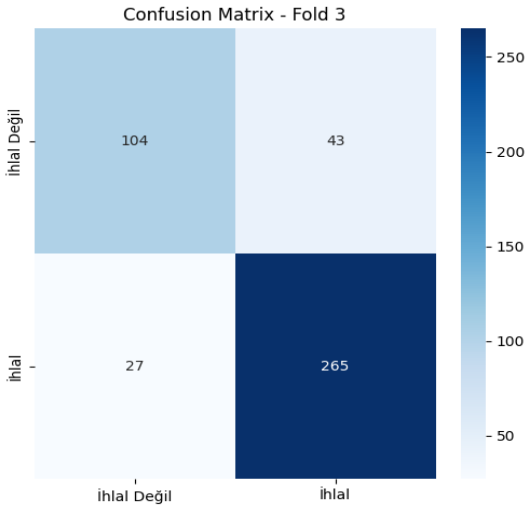
EKLER

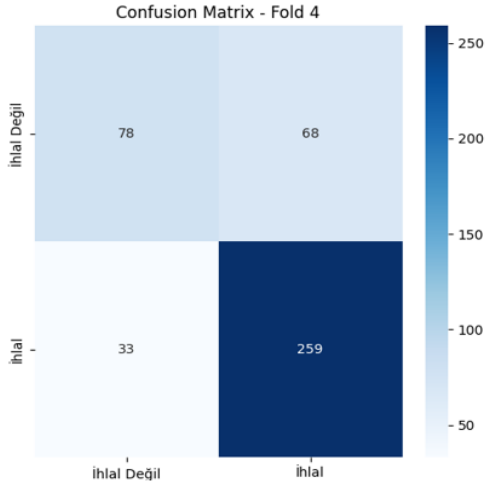
EK-1. Grafikler ve Karmaşıklık Matrisleri (LSTM-Madde 6 (Artırılmış veri))

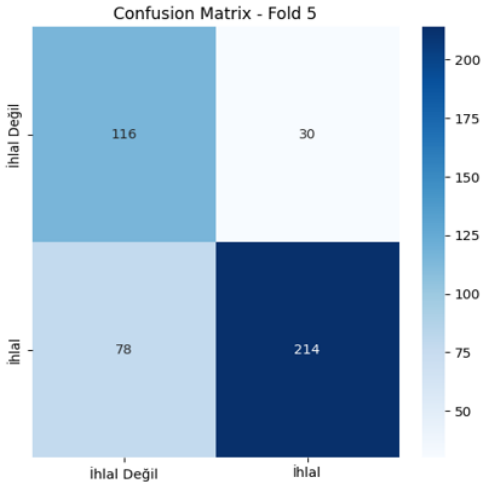


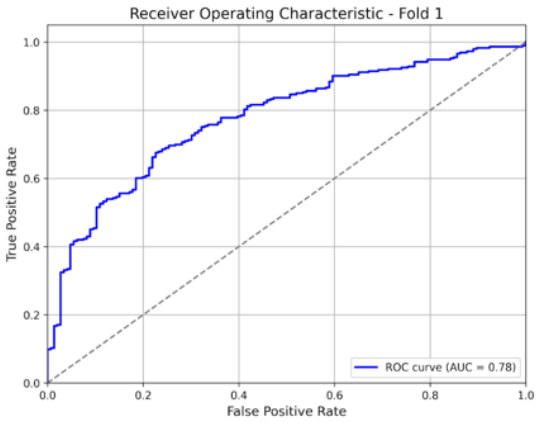

Confusion Matrix - Fold 1
İhlal Değil
İhlal
59
87
31
262
İhlal Değil
İhlal
250
200
150
100
50

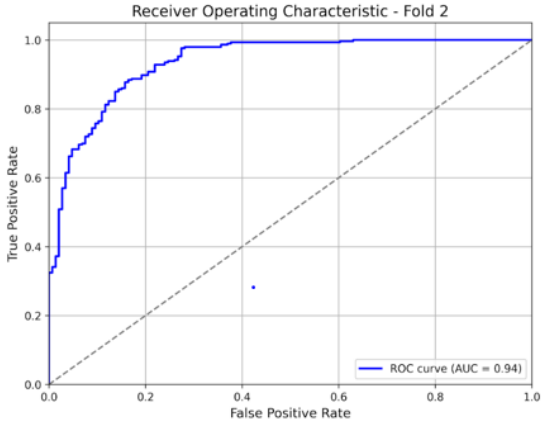

Confusion Matrix - Fold 2
İhlal Değil
İhlal
107
39
16
277
İhlal Değil
İhlal
250
200
150
100
50

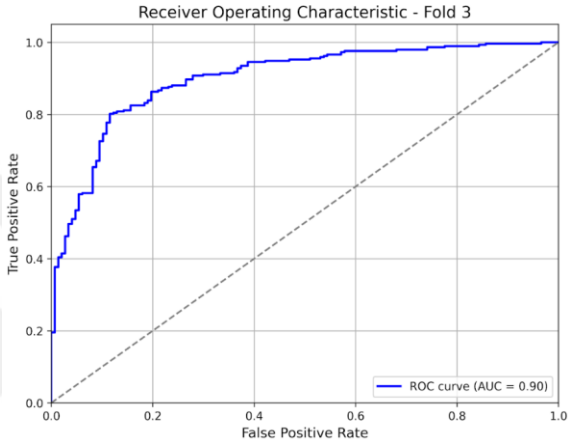

Confusion Matrix - Fold 3
İhlal Değil
İhlal
104
43
27
265
İhlal Değil
İhlal
250
200
150
100
50

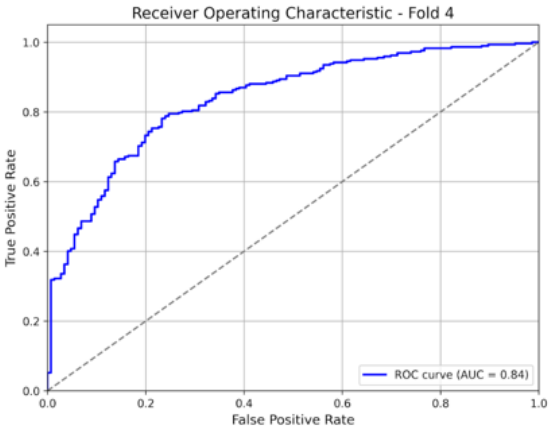

Confusion Matrix - Fold 4
İhlal Değil
İhlal
78
68
33
259
İhlal Değil
İhlal
250
200
150
100
50

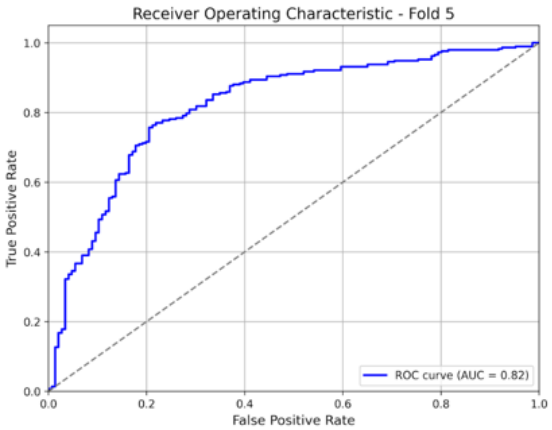

Confusion Matrix - Fold 5
İhlal Değil
İhlal
116
30
78
214
İhlal Değil
İhlal
200
175
150
125
100
75
50


Receiver Operating Characteristic - Fold 1
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.78)

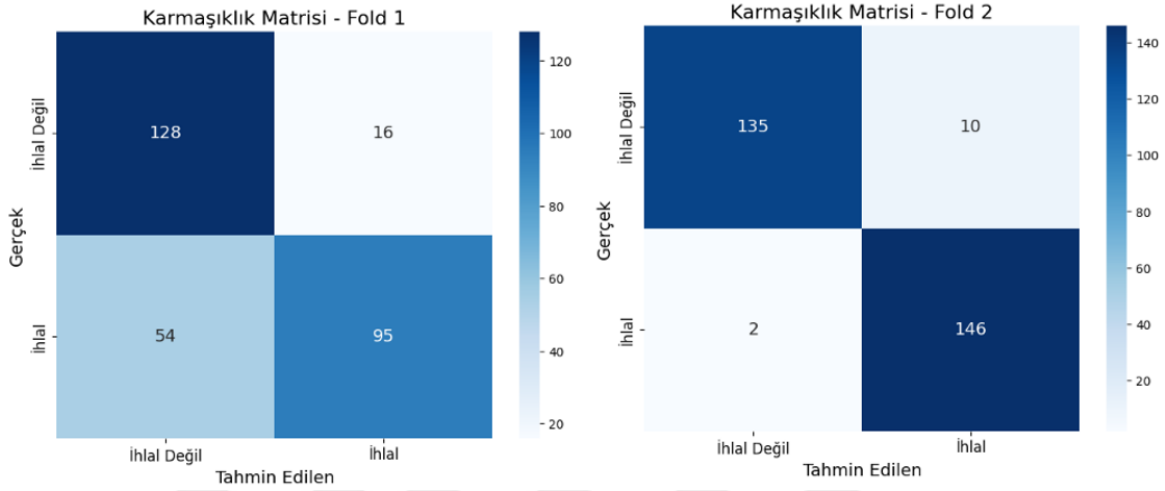

Receiver Operating Characteristic - Fold 2
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.94)

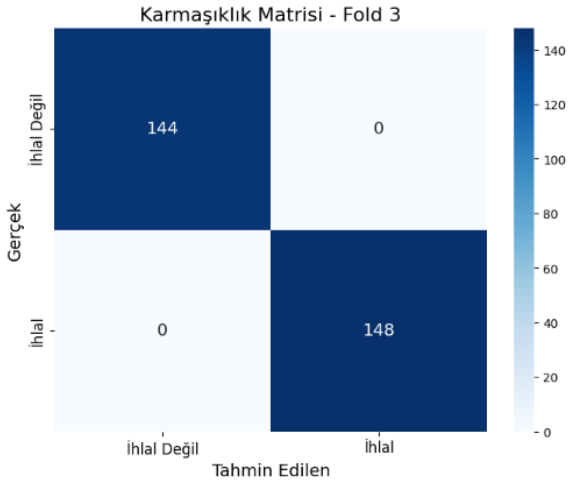

Receiver Operating Characteristic - Fold 3
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.90)

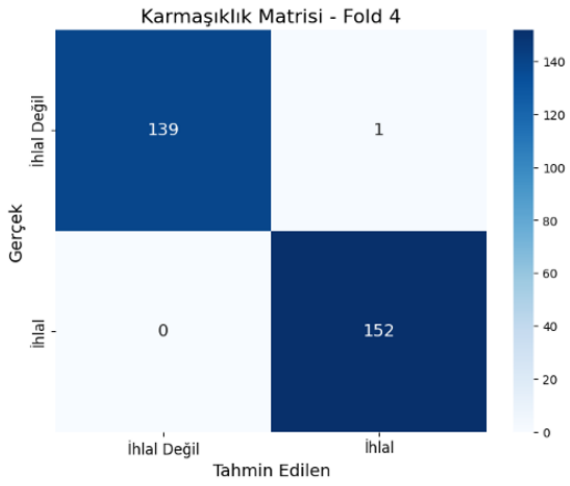

Receiver Operating Characteristic - Fold 4
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.84)

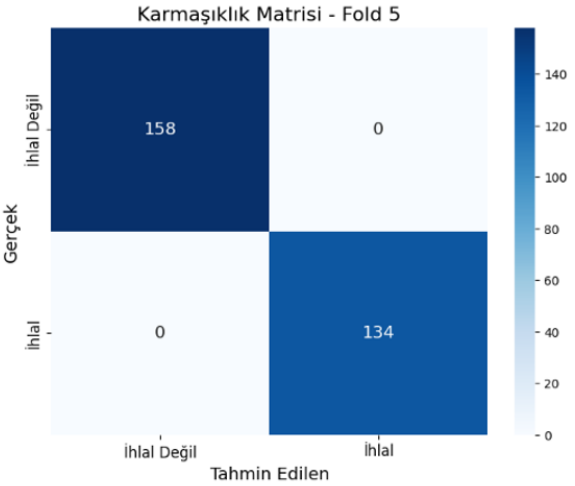

Receiver Operating Characteristic - Fold 5
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.82)

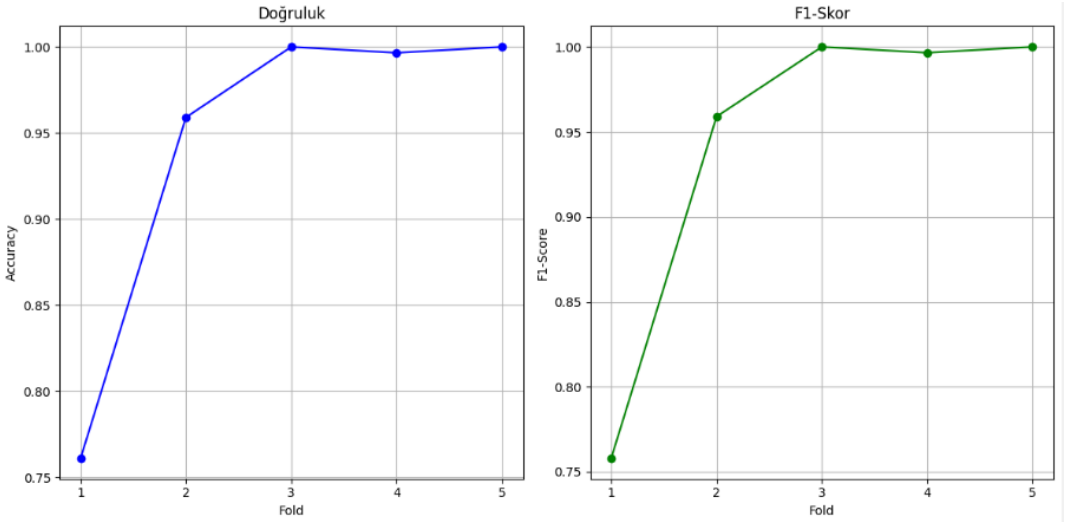
EK-2. Grafikler ve Karmaşıklık Matrisleri (LegalBERT-Madde 6(Orijinal Dengeli Veri))

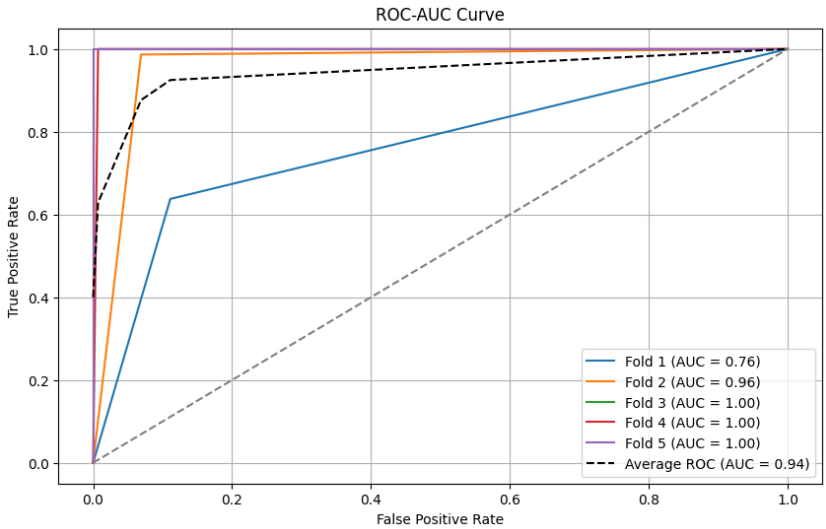




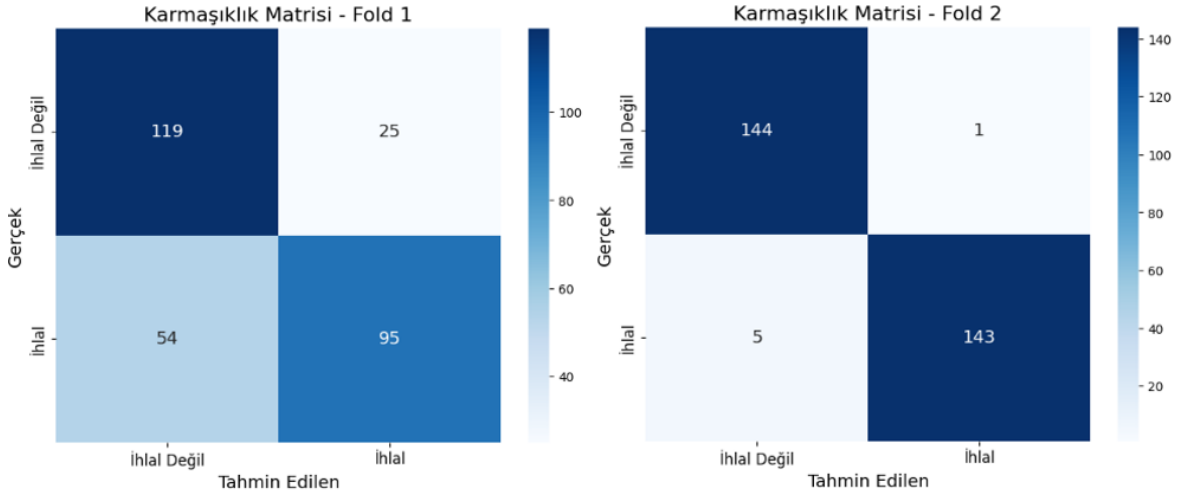


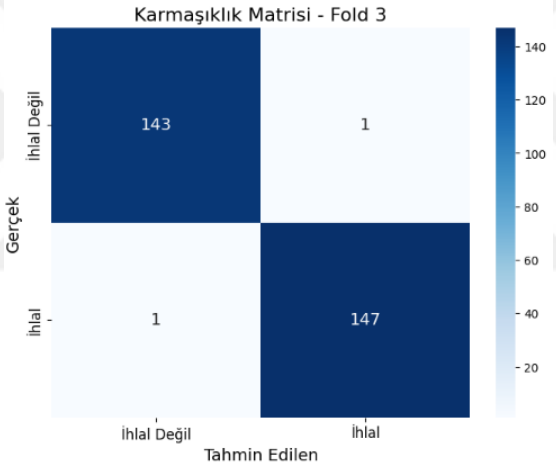


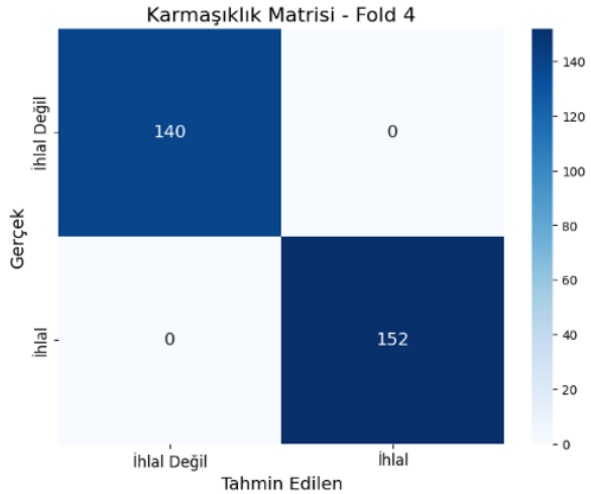

Doğruluk
Accuracy
Fold
F1-Skor
F1-Score
Fold

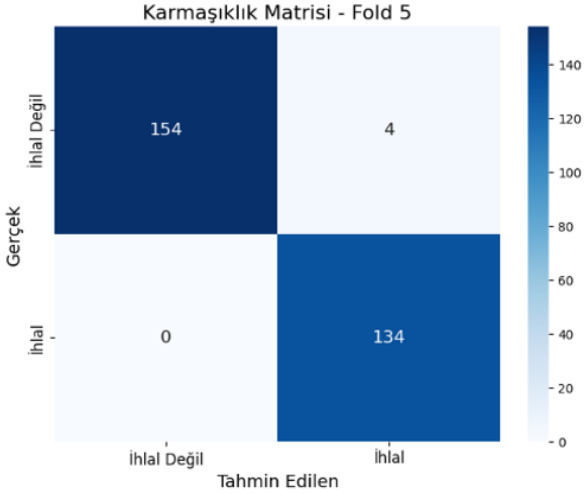

ROC-AUC Curve
True Positive Rate
False Positive Rate
Fold 1 (AUC = 0.76)
Fold 2 (AUC = 0.96)
Fold 3 (AUC = 1.00)
Fold 4 (AUC = 1.00)
Fold 5 (AUC = 1.00)
Average ROC (AUC = 0.94)

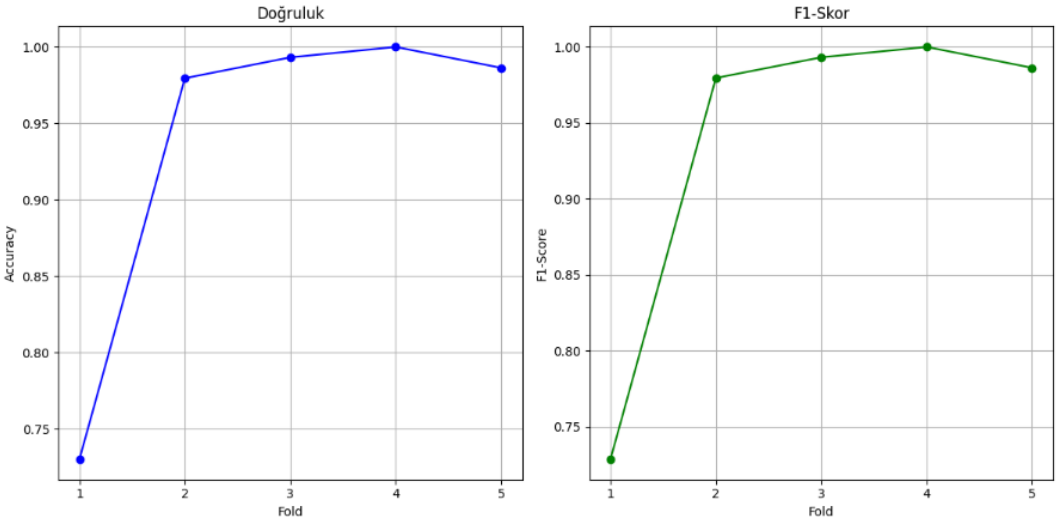
EK-3. Grafikler ve Karmaşıklık Matrisleri (XLNet-Madde 6(Orijinal Dengeli Veri))

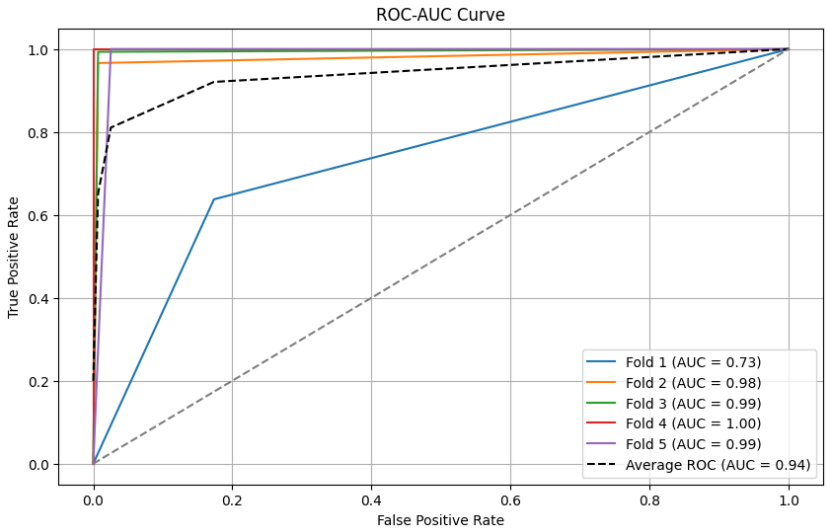




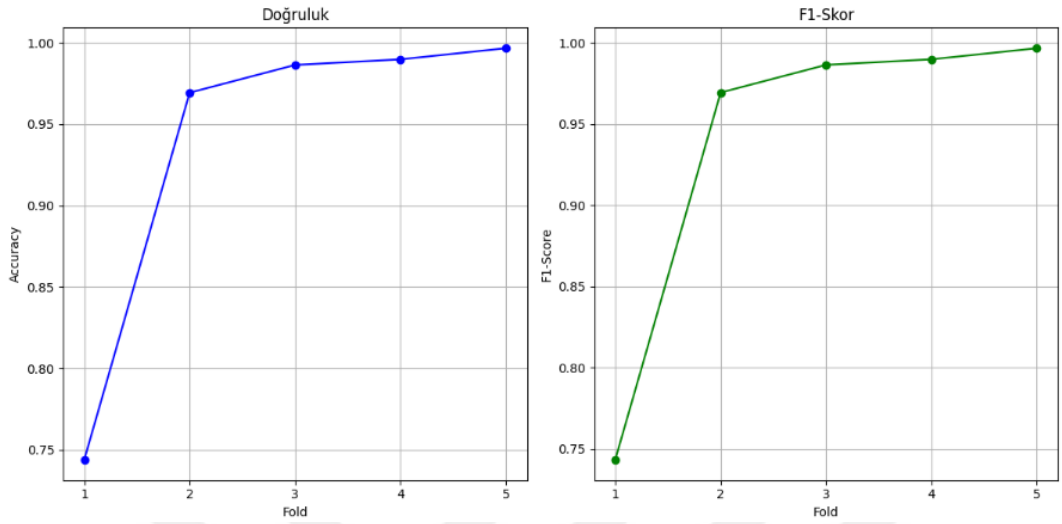


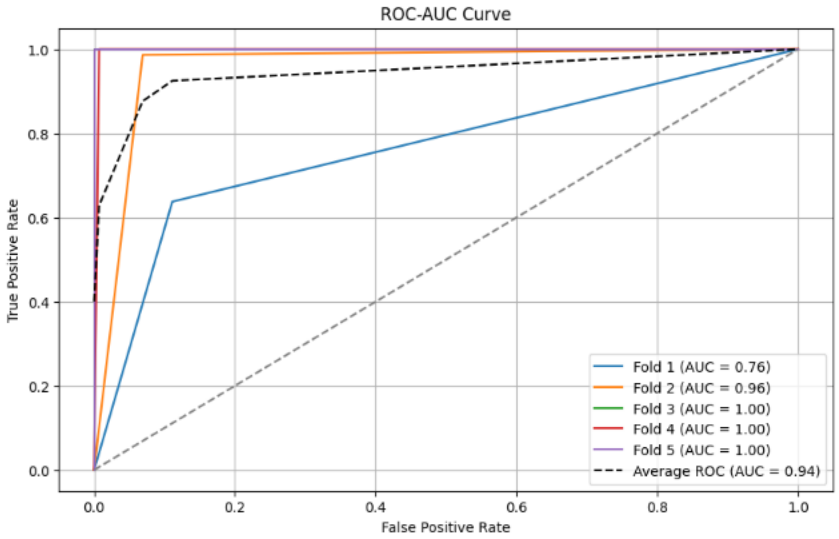


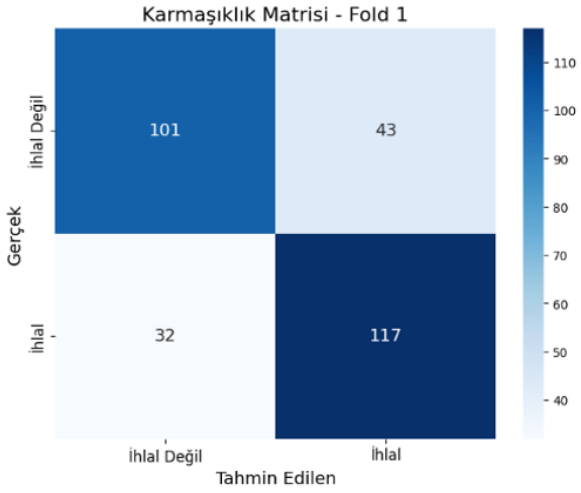

Doğruluk
Accuracy
Fold
F1-Skor
F1-Score
Fold

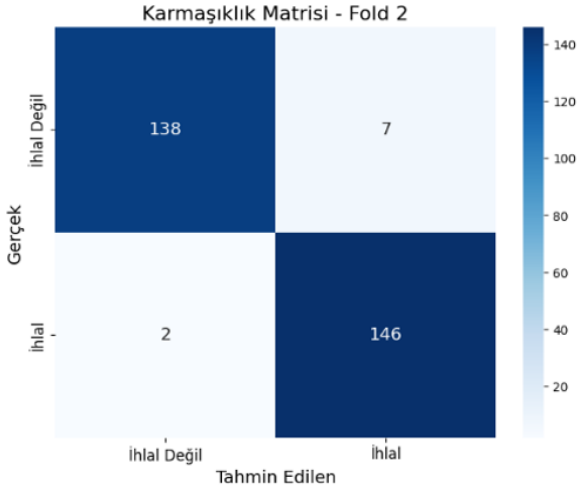

ROC-AUC Curve
True Positive Rate
False Positive Rate
Fold 1 (AUC = 0.73)
Fold 2 (AUC = 0.98)
Fold 3 (AUC = 0.99)
Fold 4 (AUC = 1.00)
Fold 5 (AUC = 0.99)
Average ROC (AUC = 0.94)

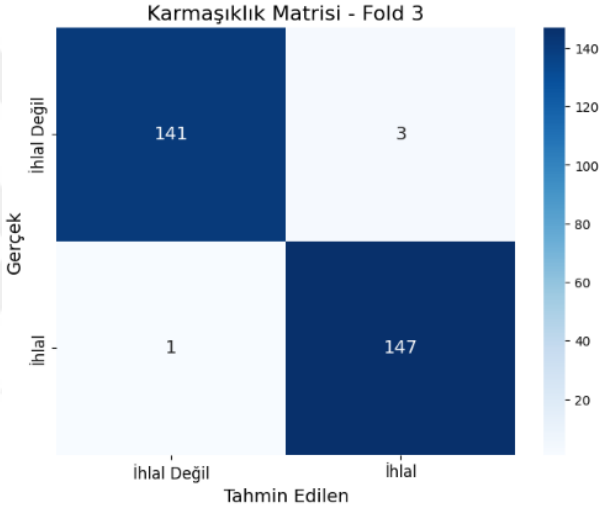
EK-4. Grafikler ve Karmaşıklık Matrisleri (BERT-Madde 6(Orijinal Dengeli Veri))

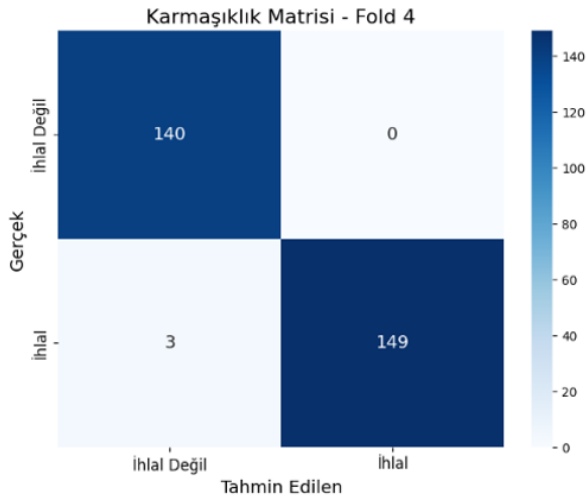


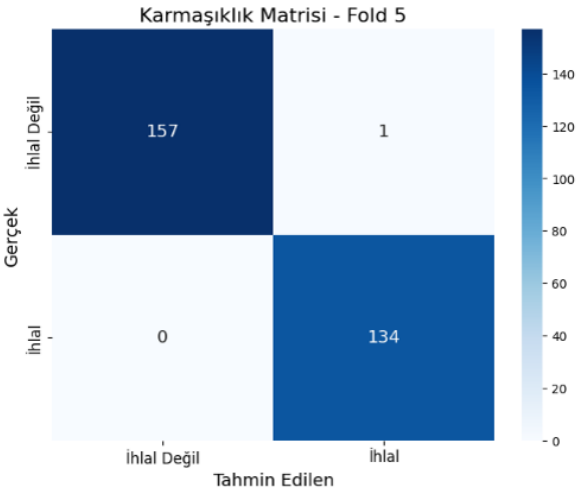



Karmaşıklık Matrisi - Fold 1
101
43
32
117
İhlal Değil
İhlal
Gerçek
İhlal Değil
İhlal
Tahmin Edilen
110
100
90
80
70
60
50
40


Karmaşıklık Matrisi - Fold 2
138
7
2
146
İhlal Değil
İhlal
Gerçek
İhlal Değil
İhlal
Tahmin Edilen
140
120
100
80
60
40
20


Karmaşıklık Matrisi - Fold 3
141
3
1
147
İhlal Değil
İhlal
Gerçek
İhlal Değil
İhlal
Tahmin Edilen
140
120
100
80
60
40
20


Karmaşıklık Matrisi - Fold 4
140
0
3
149
İhlal Değil
İhlal
Gerçek
İhlal Değil
İhlal
Tahmin Edilen
140
120
100
80
60
40
20
0


Karmaşıklık Matrisi - Fold 5
157
1
0
134
İhlal Değil
İhlal
Gerçek
İhlal Değil
İhlal
Tahmin Edilen
140
120
100
80
60
40
20
0

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı	: ÜNAL, Nagihan
Uyruğu	: T.C.
e-mail	:

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / Adli Bilişim	-
Lisans	Karadeniz Teknik Üniversitesi / Bilgisayar Mühendisliği	2013
Lise	Gölbaşı Anadolu Lisesi	2008

İş Deneyimi

Yıl	Yer	Görev
2015-Halen	Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumu	Uzman

Yabancı Dil

İngilizce

Yayınlar

Avrupa İnsan Hakları Mahkemesi Yargı Kararlarının Metin Madenciliği Yöntemleriyle Tahmin Edilmesi (Tezden üretilmiştir), International Data Science and Statistics Congress:Artificial Intelligence (2024).



GAZİLİ OLMAK AYRICALIKTIR.