

Classical, BERT-Based, and GPT-5 Models for Crime Type Classification in Turkish Criminal Law Texts

Eda Çetin , İlayda Kaya , Furkan Göz 

ARTICLE INFO

Dates:

Received: 21.01.2026

Accepted: 31.03.2026

Doi:

10.65206/pajes.1868918

Corresponding author:

Furkan Göz

(furkan.goz@kocaeli.edu.tr)

Author addresses:

Department of Computer Engineering, Faculty of Engineering, Kocaeli University, Kocaeli, 41310, Türkiye
(edacetin2001ebru@gmail.com; qkayailayda@gmail.com; furkan.goz@kocaeli.edu.tr)

ABSTRACT

Context—Artificial intelligence in the legal domain has been evolving both as supportive systems integrated into existing institutions and as next-generation legal technology initiatives centered entirely on AI. However, in Türkiye, studies in the field of legal technologies remain quite limited compared to the global literature. This limitation is largely associated with the agglutinative nature of the Turkish language, which poses complex linguistic challenges for natural language processing applications. Crime type classification plays a critical role in accelerating legal research processes and facilitating the identification of similar cases.

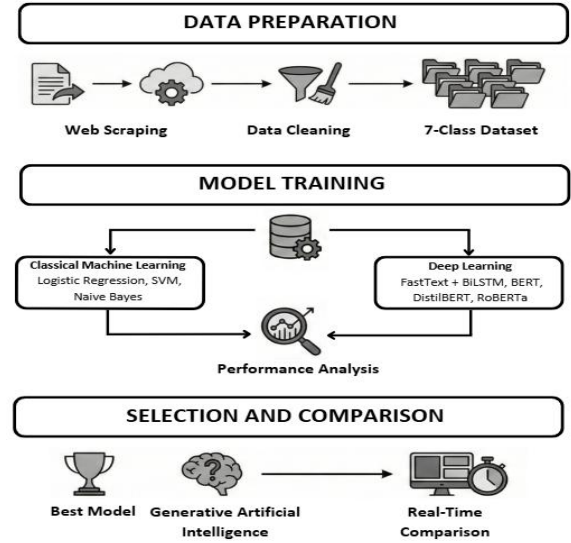
Objective—The objective of this study is to comparatively analyze the performance of classical machine learning, deep learning-based and GPT-5 models on Turkish legal terminology using open-source Turkish criminal court decision texts, and to determine the most suitable model for automatic text classification. Additionally, based on the analysis results, a real-time classification system is developed using the model that achieves the highest performance, and the applicability of the system is evaluated.

Method—A dataset consisting of Court of Criminal Chamber decision texts and covering seven different crime types was constructed using web scraping techniques. During dataset creation, the collected texts were first examined and filtered to retain only those containing crime type information. Subsequently, crime types were identified, and the texts were labeled accordingly. Some semantically similar classes were merged. In the final step, the labeled dataset was cleaned by removing embedded label information from the texts and excluding very short documents. The dataset was evaluated using different approaches: TFIDF + Naive Bayes, TFIDF + Logistic Regression, TFIDF + SVM, FastText + SVM, FastText + BiLSTM, BERT, RoBERTa, DistilBERT, GPT-5.2 and GPT-5.4. Model performances were compared using accuracy and F1-score metrics. According to the results, the TFIDF + SVM model and the generative AI-based GPT-5.4 model was tested in real time on data not included in the training dataset.



Results—Among traditional machine learning methods, TFIDF + SVM achieved the highest performance with an F1 score of 95.71%. Since legal court decisions contain strong keywords directly indicating the crime type, the TFIDF + SVM approach, which focuses on word frequency, produced more discriminative results compared to deep learning models that target complex semantic contexts. In the real-time testing phase, the trained TFIDF + SVM model was compared with the few-shot learning-based GPT-5.4 model. Both methods achieved 85.71% accuracy on test samples not included in the training set. However, while TFIDF + SVM generates responses in milliseconds, GPT-5.4 exhibit longer response times and incur API costs.

Conclusion—The findings indicate that the TFIDF + SVM method is effective for classifying crime types in Turkish court decision texts. This approach can be integrated into decision-support mechanisms within AI-assisted legal systems. Future studies may explore classification approaches covering a wider range of crime types using larger datasets derived from Turkish legal case texts.

Key Words—BERT, Turkish Legal NLP, Text Classification, Large Language Models, Real-Time Performance Analysis, GPT-5.2, GPT-5.4.



Türkçe Ceza Hukuku Metinlerinde Suç Türü Sınıflandırması için Klasik, BERT Tabanlı ve GPT-5 Modelleri

Eda Çetin , İlayda Kaya , Furkan Göz 

MAKALE BİLGİLERİ

Tarihler:

Geliş: 21.01.2026
Kabul: 31.03.2026

Doi:

10.65206/pajes.1868918

Sorumlu yazar:

Furkan Göz
(furkan.goz@kocaeli.edu.tr)

Yazar adresleri:

Bilgisayar Mühendisliği Bölümü,
Mühendislik Fakültesi, Kocaeli
Üniversitesi, Kocaeli, 41310,
Türkiye
(edacetin2001ebru@gmail.com;
qkayailayda@gmail.com;
furkan.goz@kocaeli.edu.tr)

ÖZ

Arka Plan—Hukuk alanında yapay zekanın hem mevcut kurumlara entegre edilen destekleyici sistemler hem de tamamen yeni nesil hukuk teknolojisi girişimleri şeklinde geliştiği, ancak Türkiye özelinde hukuk teknolojileri alanındaki çalışmaların dünya literatürüne kıyasla oldukça sınırlı olduğu görülmektedir. Bu sınırlılık, Türkçe'nin eklemeli yapısının doğal dil işleme uygulamaları için karmaşık dilsel zorluklar oluşturmasıyla ilişkilendirilmektedir. Suç türü sınıflandırması hem hukuki araştırma süreçlerini hızlandırmak hem de benzer davaların tespitini kolaylaştırmak açısından önemli bir rol oynamaktadır.

Amaç—Bu çalışmanın amacı, açık kaynaklı Türkçe ceza davası karar metinlerini kullanarak klasik makine öğrenmesi, derin öğrenme ve GPT-5 modellerinin Türkçe hukuk terminolojisi üzerindeki performanslarını karşılaştırmalı olarak analiz etmek, otomatik metin sınıflandırması için modellerin potansiyelini değerlendirmektir. Ayrıca analizler sonucunda en yüksek başarıyı gösteren modeller kullanılarak gerçek zamanlı bir sınıflandırma sistemi geliştirip sistemin uygulanabilirliği test etmektir.

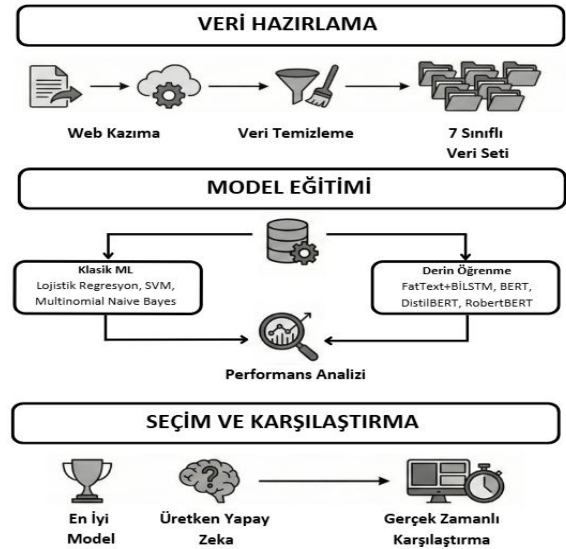
Yöntem—Web kazıma kullanılarak Ceza Dairesi karar metinlerinden oluşan ve yedi farklı suç türünü içeren bir veri seti oluşturulmuştur. Veri seti oluşturulurken ilk olarak ilgili web sitesinden elde edilen metinler incelenmiş ve yalnızca suç türü bilgisi içeren metinler filtrelenmiştir.

Ardından, suç türleri belirlenerek metinler etiketlenmiştir. Anlamsal olarak benzer olan sınıflar birleştirilmiştir. Son aşamada etiketli veri seti temizlenmiş, metinlerin içerisinde yer alan etiket bilgileri çıkarılmış ve çok kısa olan metinler veri setinden kaldırılmıştır. Veri kümesi, farklı yaklaşımlarla değerlendirilmiştir: TFIDF + Naive Bayes, TFIDF + Logistic Regression, TFIDF + SVM, FastText + SVM, FastText + BiLSTM, BERT, RoBERTa, DistilBERT, GPT-5.2 ve GPT-5.4. Model performansları, doğruluk ve F1-skor ölçütleri kullanılarak karşılaştırılmıştır. Elde edilen sonuçlara göre en yüksek doğruluğu sağlayan TFIDF + SVM modeli ile üretken yapay zekâ tabanlı GPT-5.4 modeli, eğitim veri kümesinde yer almayan veriler üzerinde gerçek zamanlı olarak test edilmiştir.

Bulgular—Geleneksel makine öğrenmesi yöntemlerinden TFIDF + SVM, %95.71 F1 skoru ile en yüksek performansı elde etmiştir. Hukuki karar metinlerinin, suç türünü doğrudan belirleyen güçlü anahtar kelimeler içermesi nedeniyle, kelime frekansına odaklanan TFIDF + SVM yapısı, karmaşık anlamsal bağlamlara odaklanan derin öğrenme modellerine kıyasla daha ayırt edici sonuçlar vermiştir. Gerçek zamanlı test aşamasında; eğitilmiş TFIDF + SVM modeli ile az örnekli öğrenme yaklaşımı kullanılan GPT-5.4 modeli karşılaştırılmıştır. Her iki yöntem de eğitim veri kümesinde yer almayan örnek vakalar üzerinde %85.71 doğruluk oranı sağlayarak birbirine yakın bir performans potansiyeli sergilemiştir. Ancak TFIDF + SVM milisaniyeler içinde yanıt üretirken, GPT-5.4'ün yanıt süresinin daha uzun olması ve API maliyetleri, gerçek zamanlı büyük ölçekli uygulamalar açısından önemli farklılık olarak öne çıkmıştır.

Sonuç—Yapay öğrenme modelleri kullanılarak yapılan sınıflandırmalarda TFIDF + SVM yönteminin Türk dava karar metinlerindeki suç türlerini sınıflandırmada etkili olduğu gözlemlenmiştir. Bu yöntem, yapay zekâ destekli hukuk sistemlerinde karar destek mekanizmalarına entegre edilebilir. Gelecekteki çalışmalar, Türk hukuk dava metinlerini kullanarak daha büyük veri setleri üzerinden daha geniş suç türlerini kapsayan sınıflandırma yaklaşımları araştırılabilir.

Anahtar Kelimeler—BERT, Türkçe Hukuk NLP, Metin Sınıflandırma, Büyük Dil Modelleri, Gerçek Zamanlı Performans Analizi, GPT-5.2, GPT-5.4.



I. GİRİŞ (INTRODUCTION)

Hukuk teknolojileri kavramı, ilk olarak 2000'li yılların başında dijital belge yönetimi ve çevrim içi hukuk veri tabanlarının geliştirilmesiyle ortaya çıkmıştır [1]. Zamanla doğal dil işleme (NLP), makine öğrenmesi ve derin öğrenme yöntemlerinin hukuk alanına uyarlanmasıyla önemli dönüşüm geçirmiştir [2]. Günümüzde belge arama ve belge yönetme ile sınırlı kalmayıp davaların analizi, karar tahmini ve metinlerin sınıflandırılması gibi daha karmaşık görevleri kapsayan yapay zekâ destekli sistemlere doğru evrilmiştir. Özellikle son yıllarda Büyük Dil Modelleri ve Transformer tabanlı yaklaşımlar hukuki metinlerin anlamsal olarak işlenmesinde büyük olanaklar sunmaktadır [3].

Hukuk teknolojilerindeki dönüşüm iki farklı yaklaşım üzerinden ilerlemektedir. Birinci yaklaşımda geleneksel hukuk büroları ve kurumsal hukuk hizmeti sunan yapılar, mevcut çalışma sistemlerine yapay zekâ tabanlı araçları entegre etmeye çalışmaktadır. Yapay zekâ sözleşme inceleme, metin özetleme ve belge yönetimi gibi süreçlerde avukatlara destek sağlayan bir yardımcı araç olarak konumlandırılmaktadır. Harvey AI'nin Allen & Overy uygulaması büyük hukuk bürolarında kullanılması bu entegrasyon yaklaşımına örnek olarak gösterilebilir [4]. Benzer şekilde Lexis+AI ve Westlaw Edge gibi köklü hukuk veri tabanları, klasik hukuki araştırma sistemlerini yapay zekâ destekli analiz ve öneri mekanizmalarıyla güçlendirmiştir [5], [6].

İkinci yaklaşım ise geleneksel hukuk bürosu yapısından bağımsız olarak, sıfırdan yapay zekâ sistemleri üzerine kurulan teknoloji odaklı hukuk girişimlerini kapsamaktadır. Bu tür sistemlerde yapay zekâ, sürecin merkezinde yer alan ana bileşen olarak tasarlanmıştır. DoNotPay, "Robot Lawyer" yaklaşımıyla hukuki süreçleri büyük ölçüde otomatikleştirmeyi hedefleyen örneklerden biri olarak öne çıkmıştır [7]. ROSS Intelligence ise içtihat analizi ve hukuki bilgiye erişimi yapay zekâ ile yeniden tanımlamayı amaçlamış ve hukuk yapay zekâ etkileşimi açısından önemli bir adım olmuştur [8]. Girişimler, hukuk hizmetlerinin sunum biçiminin teknolojik sistemler etrafında yeniden şekillenebileceğini göstermektedir. Bu iki yaklaşım birlikte değerlendirildiğinde, hukuk alanında yapay zekanın hem mevcut kurumlara entegre edilen destekleyici sistemler hem de tamamen yapay zekâ merkezli yeni nesil hukuk teknolojisi girişimleri şeklinde geliştiği görülmektedir.

Türkiye özelinde değerlendirildiğinde hukuk teknolojileri alanındaki çalışmaların dünya literatürüne kıyasla oldukça sınırlı olduğu görülmektedir. Bu sınırlılık, Türkçenin eklemeli yapısının doğal dil işleme uygulamaları için karmaşık dilbilimsel zorluklar oluşturmasıyla ilişkilendirilmektedir [9]. Ayrıca hukuki metinlerin uzun ve teknik bir yapıya sahip olması büyük ölçekli ve etiketli açık veri setlerinin kısıtlılığı Türkçe NLP çalışmalarının önünde önemli bir engel olmaktadır [10]. Türkiye'de dava dosyaları ve soruşturma evrakları kamuya açık şekilde paylaşılmadığından araştırmacıların erişebileceği başlıca kaynaklar Yargıtay ve Danıştay gibi yüksek yargı organları tarafından yayımlanan karar metinleriyle sınırlıdır. Bu durum Türkçe hukuk metinleri üzerinde yapılan çalışmaların genellikle karar metinleri üzerinden yürütülmesine neden olmaktadır [11].

Mevcut çalışmalarda ceza davalarına ait karar metinlerinde yer alan suç türlerinin otomatik olarak sınıflandırılması önemli bir problem olarak öne çıkmaktadır [12]. Suç türü sınıflandırması hem hukuki araştırma süreçlerini hızlandırmak hem de benzer davaların tespitini kolaylaştırmak açısından kritik bir rol oynamaktadır. Ayrıca bu tür sınıflandırmalar yapay zekâ destekli hukuk sistemlerinin temel unsurlarından birini oluşturmaktadır. Türkçe hukuk metinlerini klasik makine öğrenmesi algoritmalarından Büyük Dil Modellerine kadar uzanan yöntemlerle karşılaştırmalı olarak incelenmesi literatürdeki boşluğu doldurmak ve Türkiye'de hukuk teknolojilerinin gelişimine katkı sağlamak açısından önem taşımaktadır. Bu çalışma, açık kaynaklı karar metinlerini kullanarak klasik makine

öğrenmesi ve derin öğrenme modellerinin Türkçe hukuk terminolojisi üzerindeki performansını karşılaştırmalı olarak analiz ederek Türkiye'deki LegalTech literatüründeki boşluğu doldurmayı hedeflemektedir. Aynı zamanda analizler sonucunda en yüksek doğruluğa sahip modeli kullanarak gerçek zamanlı sistem oluşturup kullanılabilirliğini de test etmektedir.

Çalışmanın temel amacı, dili karmaşık ve uzun yapıya sahip Türkçe ceza davası karar metinlerini yapay zekâ yöntemleriyle otomatik olarak sınıflandırmada en uygun modeli belirlemektir. Projede farklı model yaklaşımlarının Türkçe hukuk metinlerindeki performansları karşılaştırmalı olarak incelenmiştir. Çalışma aşağıdaki hedeflere odaklanmaktadır: Geleneksel makine öğrenmesi yöntemleri olan TFIDF + Naive Bayes, TFIDF + Logistic Regression ve TFIDF + SVM, gömme tabanlı yöntemler olan FastText + SVM, derin öğrenme yaklaşımı olan FastText + BiLSTM ve Transformer tabanlı RobertaBERT, DistilBERT, BERT, GPT-5.2 ve GPT-5.4 modellerinin sınıflandırma performanslarını karşılaştırmalı olarak analiz etmek.

- Türkçe'nin sondan eklemeli morfolojik yapısını dikkate alarak FastText ve Transformer vektör temsil yöntemlerinin sınıflandırma başarısı üzerindeki etkisini değerlendirmek.
- Eğitim setinde en yüksek doğruluğu sağlayan öğrenme modeli ile az örneklili öğrenme (few shot learning) yaklaşımını kullanarak Üretken Yapay Zekâ sisteminin performanslarını kıyaslamak.
- Geliştirilen web arayüzü üzerinden farklı modellerin tahmin çıktılarını gerçek zamanlı olarak sunmak ve model kararlarının karşılaştırılabilirliğini sağlamak.
- Açık kaynaklı Türk yargı kararlarını kullanarak özgün bir veri seti oluşturup Türkiye'de kısıtlı olan hukuk teknolojileri literatürüne ve veri havuzuna katkı sağlamak.

II. İLGİLİ ÇALIŞMALAR (RELATED WORKS)

Doğal dil işleme teknolojilerinin hızlı gelişmesiyle birlikte çok sayıda metin verisi içeren hukuk kaynakları, hukuk alanındaki teknolojik gelişmeleri hızlandırmıştır. Literatür incelendiğinde sınıflandırma probleminin başlangıçta geleneksel makine öğrenmesi yöntemleriyle ele alındığı son yıllarda ise derin öğrenme ve Transformer tabanlı modellerin kullanımının arttığı görülmektedir. Çalışmalar, kullanılan veri setlerinin dili, boyutu ve uygulanan algoritmalar açısından farklılık göstermektedir.

Türkçe hukuk metinleri üzerine yapılan çalışmalarda Türkçe'nin sondan eklemeli yapısından kaynaklı çeşitli ön işleme ve modelleme stratejileri geliştirilmiştir. Tablo 1 bu alanda yapılan Türkçe ve diğer dillerde akademik çalışmalarda kullanılan veri setleri, uygulanan yöntemler ve elde edilen performans metrikleri karşılaştırmalı olarak özetlemektedir. Ayrıca her bir çalışmaya ait detaylar özet halinde sunulmuştur.

Görentaş ve Uçkan (2023) tarafından yapılan çalışmada, makine öğrenmesi yöntemlerini kullanarak Uyuşmazlık Mahkemesi kararlarının kümelenmesini incelemektedir [13]. Mahkeme kararları öncelikle TFIDF yöntemiyle NLP sürecinden geçirilmiş ve ardından CURE, K-MEANS, DBSCAN, AGNES, AFFINITY ve BIRCH algoritmalarıyla kümelenmiştir. Veri seti, Adli (4780 adet) ve İdari (1240 adet) olan olumsuz görev uyuşmazlıkları kararlarından oluşmaktadır. Kümeleme sonuçları Silhouette Skoru, Davies-Bouldin İndeksi (DBI) ve Calinski-Harabasz İndeksi (CHI) metrikleriyle değerlendirilmiştir. Yapılan değerlendirmeler sonucunda en iyi performansı gösteren algoritmanın BIRCH olduğu saptanmıştır.

Küçük ve Can (2024) tarafından yapılan derleme çalışması, yapay zekâ ve büyük dil modellerinin hukuki metinleri otomatik işleme potansiyelini incelemiştir [14]. Çalışma, derin öğrenme temelli araçların hukuk alanındaki kullanımına ve teknolojik gelişimine odaklanmaktadır. Yasal mevzuat, mahkeme kararları ve hukuki

Tablo 1. Literatür araştırması tablosu.

Table 1. Literature review table.

Kaynak	Çalışma Odak Noktası	Yöntem / Model	Veri Seti	Temel Sonuç
[13]	Mahkeme Kararı Kümelenme	TF-IDF + CURE, K-Means, BIRCH	6020 Uyuşmazlık Mahkemesi Kararı	BIRCH (0.90 Silhouette).
[14]	Hukuki Metin İşleme (Survey)	ML, DL, Üretken Yapay Zekâ	Kanun, Karar, Dilekçe	Otomatik belge üretimi ve tahmini analiz potansiyeli
[15]	Sınıflandırma & Açıklanabilirlik	ANN, RF, XGBoost + LIME/ SMOTE	6425 Yüksek Mahkeme Kararı	ANN + Alt Örnekleme %82.6 duyarlılık
[16]	Emsal Karar Sınıflandırması	Gradient Boosting	280000 UYAP Kararı	Doğruluk: %87 (İkili), %83 (Üçlü Sınıflandırma)
[17]	Suç Türü Sınıflandırma	FastText (n-gram)	531 Yargıtay Ceza Kararı	Kesinlik: %44.7 (78 farklı suç türü için)
[18]	Global Hukuki NLP (Survey)	Transformer, LLM, MTL/PEFT	24 Dilde Çoklu Veri Setleri	Hukuki iş yükünde %30-50 arası verimlilik potansiyeli
[19]	Duygu Analizi (Karar Tahmini)	CNN-LSTM (Hibrit)	2000 (Artırılmış 98k) Kanada Deniz Hukuku	Doğruluk: %98.05
[20]	Konu Modelleme	LDA, LSA, pLSA	4259 Brezilya Mahkeme Kararı	LDA 15 konu başlığında optimum ayırım
[21]	NLP Akış Hattı	LDA & Transformer Özetleme	Dilekçe, Sözleşme, NDA	Bilgi çıkarma ve özetlemede yüksek başarı
[22]	Bilgi Çıkarma	LDA, LSA, NMF	1000 Hindistan Mahkeme Kararı	LDA 0.82 F-Skoru ile en tutarlı model
[23]	Soru-Cevap Güvenilirliği	DeepEval, GPT-4o, NLI	573 Soru / 1558 Yanıt	Makro-F1 0.80 (GPT-4o ile F1 0.77)
[24]	Yasal metinlerde atf sınıflandırması	Lojistik Regresyon, Naive Bayes, Lineer SVM	Yasal metin sınıflandırma veri seti (25000 dosya)	SVM %95 Doğruluk 0.94 Makro F1
[25]	Türk Mahkeme Kararı Sınıflandırma	SVM, KNN, RF, Karar Ağaçları	Lexpera (Boşanma ve İmar Kararları)	SVM %90, KNN %90, RF %80, Karar Ağaçları %77.5 Doğruluk
[26]	Uzun Belge Sınıflandırma Karşılaştırması	Alan Konsepti + RF, BiLSTM+Att, TextCNN	30000 ABD Dava Belgesi (SigmaLaw)	Alan Konseptli RF %85.98 Doğruluk, Derin öğrenme (BiLSTM) %49.23 Doğruluk
[11]	Yüksek Mahkeme Karar Sonucu Tahmini	LSTM, BiLSTM+Att, SVM, RF	47796 Yargıtay ve 1290 AYM Kararı	BiLSTM+Att. %93.2 Doğruluk, AYM verilerinde LSTM+Att. 0,67 Makro F1
[27]	Hukuk metinlerinin analizi (ASKE)	BERT, Legal-BERT, RoBERTa, DeBERTa, Longformer ve CaseLaw- BERT	LexGlue (İngilizce hukuk dokümanları)	ASKE yöntemi UNFAIR-TOS %96.2, LEDGAR %88.7 ve EUR-LEX %73 F1 skoru

dilekçeler metinler yapay zekâ yöntemleriyle metinlerin otomatik sınıflandırılması, belgelerden bilgi çıkarımı ve bilgiye erişimin artırılması, mahkeme kararlarının önceden tahmini ve hukuki belge üretimi amacıyla kullanılmaktadır.

Okursoy ve İnkaya (2025) tarafından yapılan çalışma, Türkiye'deki iki Yüksek Mahkeme olan Yargıtay ve Danıştay'a ait 6425 adet kararı (%89'u bozma, %11'i onama) sınıflandırmak amacıyla, yapay zekâ tabanlı karar destek sistemi geliştirmeyi amaçlamıştır [15]. Çalışmada, metinlerden öznelikler çıkarıldıktan sonra veri dengesizliğini gidermek için SMOTE ve rastgele alt örnekleme teknikleri uygulanmıştır. Ardından KNN, DT, SVM, RF, LightGBM, XGBoost ve ANN sınıflandırma modelleri ile performansları değerlendirilmiştir. Deneysel sonuçlara göre, ANN modeli duyarlılık %82.6 ve özgülük %71.8 değerlerinde en iyi performansı sergilemiştir.

Demir, Aydemir ve Sönmez (2025), Türk yargı sistemindeki emsal kararları incelemiştir [16]. Çalışmada, UYAP veri tabanından elde edilen hukuki karar metinleri kullanılmıştır. Bu metinler üzerinde NLP ve makine öğrenmesi yöntemleriyle bir sınıflandırma modeli geliştirilmiştir. Çalışmanın odak noktası, en sık tekrar eden üç hüküm olan Reddi, Kabulü ve İadesi kararlarının sınıflandırılmasıdır. Metinler üzerinde ön işleme (kelime köklerine inme, stop word çıkarma) yapılmış ve TF-IDF, Word2Vec, Doc2Vec ile vektörleştirilmiştir. Naive Bayes, SVM, RF, ANN, Lojistik Regresyon ve Gradient Boosting modelleri kullanılmıştır. Gradient Boosting algoritması en başarılı performansı sergilemiştir. Bu algoritma, üçlü sınıflandırmada %83 ve ikili sınıflandırma senaryolarında %87 doğruluk oranına ulaşmıştır.

Kılıç ve Öner (2021), avukatların emsal karar arama süreçlerinde anahtar kelime tabanlı sistemlerin yetersizliğini ele almıştır [17]. Çalışmada, Yargıtay Ceza Genel Kurulu tarafından 2020 yılında karara bağlanan 531 karar kullanılmıştır. Bu kararlar, suç türlerine göre otomatik olarak sınıflandırılmıştır. Karar metinlerini işlemek ve vektörel gösterimlerini elde etmek için FastText kullanılmıştır. Modelin performansı farklı parametre

kombinasyonları ile test edilmiştir. Toplam 78 farklı suç türüne göre etiketlenen veri seti üzerinde yapılan denemelerde, FastText modelinin ulaştığı en yüksek kesinlik oranı %44,7 olmuştur.

Ariai, Mackenzie ve Demartini (2025) tarafından yapılan anket çalışması, Doğal Dil İşleme tekniklerinin hukuk alanındaki uygulamalarını görevler, veri setleri, modeller ve zorluklar çerçevesinde incelemiştir [18]. NLP'nin, hukuki süreçleri otomatikleştirmede iş yükünü azaltmada ve verimliliği artırmada büyük potansiyeli olduğu vurgulanmıştır. İncelenen temel görevler arasında Hukuki Soru Cevaplama (LQA), Hukuki Karar Tahmini (LJP), Metin Sınıflandırma (LTC), Belge Özetleme (LDS), Varlık Tanıma (NER) ve Argüman Madenciliği (LAM) bulunmaktadır. Gelişmiş sinir ağı mimarileri ve Transformer tabanlı Büyük Dil Modelleri yaygın olarak kullanılmaktadır. Çalışma, 24 dildeki legal korpuslarını ve veri setlerini incelemiştir.

Abimbola ve arkadaşları tarafından (2024) yapılan çalışmada, Kanada deniz hukuku davalarındaki mahkeme kararlarının duygusal analizini yaparak bu kararların sonucunu (olumlu/Affirmation veya olumsuz/Reversal) tahmin etmeyi amaçlamıştır [19]. Kanada Federal Yüksek Mahkemesi'nden toplanan 2000 adet deniz hukuku davası kararı veri seti olarak kullanılmıştır. Model olarak, Derin Öğrenme yaklaşımlarından LSTM + CNN hibrit modeli tercih edilmiş ve bu modelin performansı geleneksel makine öğrenmesi algoritmaları ile karşılaştırılmıştır. Yapılan analizlerde, LSTM + CNN modelinin %50 doğruluk oranına ulaştığı görülmüştür.

Braga ve Rodello (2025) tarafından yapılan çalışma, Brezilya São Paulo Adalet Mahkemesi'nden toplanan dava dosyalarını incelemiştir [20]. Çalışmada, özel eğitim politikalarıyla ilgili 4259 dava dosyası kullanılmıştır. Dosyaların bağlamını otomatik olarak yorumlamak için Olasılıksal Konu Modelleme teknikleri uygulanmıştır. Temel amaç, Brezilya'daki hukuki metinlerin içerdiği tematik yapıları ortaya çıkarmaktır. Araştırmada Latent Dirichlet Allocation(LDA), Latent Semantic Analysis (LSA) ve

Probabilistic Latent Semantic Analysis (pLSA) modelleri karşılaştırılmıştır. Analizler sonucunda LDA modelinin daha yüksek sayıda konuda (15 konu) en yüksek tutarlılığı verdiği ve hukuki konuları başarılı şekilde ayırt ettiği belirlenmiştir. Buna karşılık, pLSA daha az sayıda konuda (5 konu) daha geniş temaları yakalamada başarı göstermiştir. Çalışma konu modellemenin hukuk profesyonelleri için karar destek aracı olabileceğini göstermiştir.

Dipali Singh (2024) tarafından gerçekleştirilen çalışma, Hindistan'ın yargı sistemindeki önemli dava yığılmalarını azaltmak ve süreçleri daha verimli hale getirmek amacıyla NLP tekniklerinin potansiyelini araştırmaktadır [21]. Önerilen metodoloji, hukuki belgelerin analizini otomatikleştirmek için bir NLP akış hattı kullanmaktadır. Temel olarak bu akış hattı öncelikle belgeleri stopword kelimelerinden ve noktalama işaretlerinden temizler. Ardından, belgedeki gizli konuları belirlemek için LDA ve farklı kelime limitlerinde özlü özetler oluşturmak için Transformer tabanlı modelleri kullanır. Uygulama, örnek bir boşanma anlaşması belgesinde başarılı sonuçlar göstermiştir.

Zadgaonkar ve Agrawal (2024) tarafından yapılan çalışmada, Hint yargı sistemindeki yapılandırılmamış yasal kararlardan verimli bilgi çıkarımı için Konu Modelleme tekniklerini kullanan bir çerçeve sunulmaktadır [22]. Amaç, binlerce kararın hızlıca analiz ederek zaman alıcı manuel analiz ihtiyacını ortadan kaldırmaktır. Çalışma yasal kararları konulara ayırmak için LDA modelini kullanmaktadır. Süreç ön işleme, model uygulama ve tutarlılık skoru ile değerlendirme adımlarından oluşmaktadır. 1000 karardan oluşan bir veri kümesi üzerinde yapılan deneysel çalışmalarda, LDA'nın LSA ve Non-negative Matrix Factorization (NMF) yöntemlerine göre daha iyi bir tutarlılık skoru elde ettiği görülmüştür. Ayrıca, LDA tabanlı yaklaşım yasal kararlar arasındaki benzerliği ölçme görevinde mevcut yöntemlerden daha yüksek korelasyon katsayısı ve F1 skoru elde etmiştir.

Trautmann ve arkadaşları (2024), yasal soru cevap sistemleri alanında üretken yapay zekâ tarafından üretilen yanıtların güvenilirliğini ve doğruluğunu artırmak amacıyla, yanıtların sağlanan kaynak metinle uyumunu ifade eden gerekçelendirilme düzeyini ölçen yöntemleri kıyaslamaktadır [23]. Çalışma, semantik benzerlik, Doğal Dil Çıkarımı modelleri ve Prompt Zincirleme gibi yaklaşımları içeren bir karşılaştırma sunmuştur. Veri seti, Casetext'ten alınan yasal sorgular ve kasıtlı olarak oluşturulmuş halüsinasyonlar içeren yanıtlar kullanılarak oluşturulmuştur. Temel bulgular; prompt zincirleme yöntemi olan DeepEval Claims Verify'nin 0.80'lik Makro-F1 skoru ile en yüksek performans gösterdiğini, ancak yüksek performansın uzun gecikme süresi ile elde edildiğini ortaya koymuştur. Diğer yandan GPT-4o Direct Prompting ise 0.77 F1 Skor elde etmiştir.

Çelik ve arkadaşları (2026), makine öğrenme yöntemlerini kullanarak yasal metinlerde atıf sınıflandırması, anahtar kelime çıkarımı ve semantik analizi etik hususlarla birleştiren bir analiz yapmışlardır [24]. Çalışmada Kaggle'dan alınan Legal Text Classification Dataset adlı bir veri seti kullanılmıştır. Bu veri seti 25000 vaka dosyasından oluşmaktadır. Çalışmada Lojistik Regresyon, Naive Bayes ve Lineer SVM makine öğrenmesi modelleri kullanılmıştır. SMOTE ve GridSearchCV teknikleri kullanarak veri setindeki dengesizlik giderilmiştir. Model eğitim sonuçlarına göre Lineer SVM %95 doğruluk ve 0.94 makro F1 skoru ile en yüksek performansı göstermiştir.

Turan ve Küçüksille (2024), Türk mahkeme kararlarını kategorik olarak sınıflandırmayı amaçlamaktadır [25]. Bunun için Lexpera'dan alınan ilk derece mahkemesi kararlarını kullanarak boşanma ve imar davalarından oluşan bir veri kümesi hazırlanmış; imar kararları 1, boşanma kararları 0 olarak etiketlenmiştir. Davaların sınıflandırılması için KNN, SVM, Karar Ağaçları ve Rastgele Orman modelleri kurulmuş, her model için hiperparametre optimizasyonu yapılmış ve 10 katlı çapraz

doğrulama uygulanmıştır. Modeller doğruluk metriğine göre karşılaştırılmıştır. Sonuçlara göre SVM modeli ortalama %90, KNN modeli %90, Rastgele Ormanlar modeli %80, Karar Ağaçları modeli %77.5 doğruluk elde etmiştir.

Chen ve arkadaşları tarafından (2022) gerçekleştirilen çalışma, otomatik hukuki metin sınıflandırması için farklı yöntemleri karşılaştırmaktadır [26]. Deneyler, 78 hukuk alanını kapsayan ve yaklaşık 30000 dava belgesinden oluşan SigmaLaw veri kümesi üzerinde yapılmıştır. Çalışmada, alan konseptlerine dayalı Rastgele Orman modeli ile TFIDF tabanlı Rastgele Orman modeli ve Word2Vec, GloVe veya BERT vektörleri kullanan TextCNN, BiLSTM ve dikkat mekanizmalı BiLSTM modelleri karşılaştırılmıştır. Alan konsepti tabanlı Rastgele Orman modeli, %84.49 doğruluk ve 0.68 F1 skoru elde ederken, BiLSTM + attention yaklaşımı 0.49 doğruluk ve 0.44 F1 skoruna ulaşmıştır. PCA ile özellik seçimi yapıldığında, 400 alan konsepti kullanılarak doğruluk %85.98'e çıkarılmıştır.

Mumcuoğlu ve diğerleri (2021), Türkiye'de yüksek mahkeme kararlarının sonucunu dava açıklamalarına bakarak tahmin etmeyi amaçlamaktadır [11]. Problem, kararın kabul veya ret olarak sınıflandırıldığı ikili bir görev şeklinde ele alınmıştır. Çalışmada, Anayasa Mahkemesi'ne ait 1290 dava ile temyiz mahkemelerine ait 8870-47796 arasında değişen sayıda karar metni kullanılmıştır. Karar Ağaçları, Rastgele Orman ve SVM gibi klasik yöntemler ile LSTM, BiLSTM ve dikkat mekanizmalı derin öğrenme modelleri karşılaştırılmıştır. Deneyler sonucunda ortalama doğruluk %86.1, dengeli doğruluk %75.7 ve makro F1 skoru 0.75 olarak elde edilmiştir. Tüm deneyler içinde en yüksek doğruluk %93.2, en yüksek makro F1 skoru ise %87'dir. Vergi Temyiz Mahkemesi verilerinde LSTM modeli %87 makro F1 skoruna, BiLSTM + attention modeli %93.2 doğruluk değerine ulaşmıştır. Anayasa Mahkemesi verilerinde ise en iyi sonuçlar LSTM + attention modeliyle elde edilmiş, doğruluk %91.8 ve makro F1 skoru %67 olarak bulunmuştur.

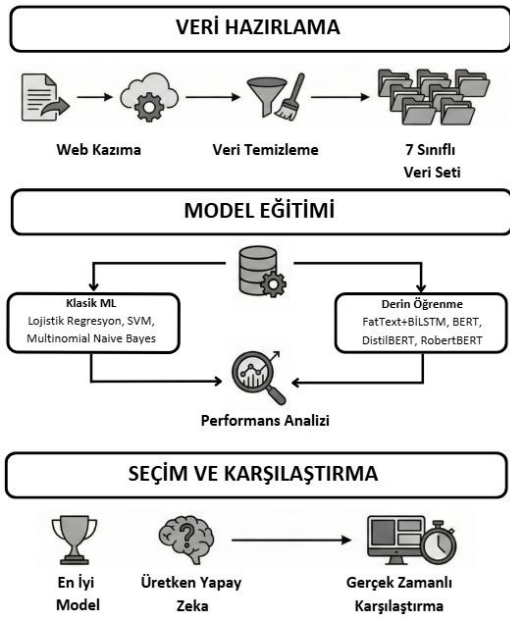
Khaliq ve arkadaşları (2026), tarafından geliştirilen çalışmada metinlerdeki gizli anlamsal yapıları yakalamak için tasarlanan ASKE adlı bilgi çıkarma yaklaşımı kullanılmıştır [27]. Temel mantığı etiketli veriye ihtiyaç duymayan sıfır örnekli öğrenme yöntemine dayanmaktadır. Metinlerin anlamlı parçalara bölünmesi, başlangıç kavramlarıyla eşleştirilmesi ve kavramların dokümanlardan çıkarılan terimlerle döngüsel olarak zenginleştirilmesi süreçlerini içermektedir. LexGLUE veri setleri kullanılarak yapılan testlerde ASKE mevcut gelişmiş modellerle benzer ya da daha iyi performans sergilemiştir.

III. YÖNTEM (METHODOLOGY)

Projede geliştirilen ve genel akışı Şekil 1'de sunulan yöntem kapsamında, Türkçe Ceza Dairesi kararları yedi farklı suç türüne göre sınıflandırılmıştır. Süreç; veri toplama, veri temizleme, öğrenme tabanlı modellerin eğitimi ve performans analizi aşamalarından oluşmaktadır. Çalışmanın son aşamasında en başarılı model ile az örnekli öğrenme yaklaşımı uygulanan Üretken Yapay Zekâ sisteminin doğruluk performansları geliştirilen bir arayüz üzerinden eş zamanlı olarak karşılaştırılmıştır.

A. Veri setinin oluşturulması (Dataset construction)

Projede kullanılan veri seti, Türkiye'deki yargı kararlarını dijital ortamda sunan sonkarar.com web platformu üzerinden web kazıma yöntemiyle elde edilmiştir. Çalışmanın temelini oluşturan Yargıtay karar metinleri, Yargıtay Başkanlığı tarafından anonim hale getirilerek halka açık bir şekilde yayınlanmaktadır. Bundan dolayı incelenen metinlerde tarafların isimleri, adresleri veya kişisel tanımlayıcı bilgileri bulunmamaktadır. Veri setinin anonim olması nedeniyle çalışmada ek bir anonimleştirme sürecine ihtiyaç duyulmamış ve veriler etik standartlara uygun şekilde doğrudan analiz edilmiştir.



Şekil 1. Çalışmanın genel akış şeması.

Figure 1. Overall workflow of the study.

1. Veri toplama ve filtreleme (Data collection and filtering)

Veri toplama süreci sonkarar.com web platformunda yer alan Yargıtay Kararları dizini altındaki ilk 700 sayfanın sistematik olarak taranmasıyla gerçekleştirilmiştir. İşlem yükünü azaltmak ve veri kaybını önlemek amacıyla tarama işlemi, her biri 100 sayfadan oluşan yedi blok halinde gerçekleştirilmiştir. Her bir sayfada ortalama 15 karar metni bulunmaktadır. Toplamda 10061 karar metni elde edilmiştir. Elde edilen veri kümesi, ön işleme ve etiketleme aşamalarında kullanılmıştır.

2. Etiketleme ve sınıf birleştirme (Labeling and class merging)

Ham veri seti üzerinde yapılan ilk incelemede, toplanan karar metninin tamamında suç türü bilgisinin yer almadığı tespit edilmiştir. Bu nedenle, yalnızca suç türü bilgisinin açık ve net biçimde yer aldığı Ceza Dairelerine ait kararlar veri setine dahil edilmiştir. Ceza dairesine ait olmayan veriler çıkarıldıktan sonra veri setindeki örnek sayısı 5097 seviyesine düşmüştür.

Ceza dairesi kararlarında yer alan suç türü ifadeleri temel alınarak metin içeriği ayrıntılı biçimde incelenmiştir. Belirlenen kurallar doğrultusunda suç türlerine göre etiketleme işlemi gerçekleştirilmiştir. Suç türü bilgisi içermeyen karar metinleri ise *no_crime* etiketi ile işaretlenmiştir.

Suç türlerine göre gerçekleştirilen otomatik etiketleme sonucunda yaklaşık 400 farklı suç kategorisi elde edilmiştir. Ancak modelin öğrenme performansını artırmak ve istatistiksel anlamlılığı korumak amacıyla örnek sayısı en yüksek olan yedi ana sınıf seçilmiştir. Seçilen sınıflar; *Uyuşturucu Madde Ticareti*, *Kullanmak İçin Uyuşturucu Madde Bulundurma*, *Yağma*, *Hakaret*, *Tehdit*, *Dolandırıcılık* ve *Hayata ve Vücut Dokunulmazlığına Karşı Suçlar* şeklindedir. Hayata ve Vücut Dokunulmazlığına Karşı Suçlar sınıfı; kasten öldürme, kasten yaralama ve insan hayatına yönelik benzer fiilleri kapsayan alt suç türlerinin birleştirilmesiyle oluşturulmuştur. İşlemler sonucunda yedi sınıftan toplam 3500 örnek veri elde edilmiştir. Etiketleme işlemi bittikten sonra karar metninin başında yer alan ve doğrudan suç türü bilgisini içeren bölümler metinlerden çıkarılmıştır. Böylece modelin sınıflandırma işlemini yalnızca karar metninin bağlamsal içeriğine dayanarak gerçekleştirmesi amaçlanmıştır.

3. Veri Temizleme ve Veri Kümesi (Data cleaning and dataset)

Model eğitimi öncesinde metin uzunlukları üzerine analiz yapılmıştır. Çok kısa metinlerin karakteristik suç unsurlarını barındırmadığı ve anlamsal bağlamdan yoksun olduğu

gözlemlenmiştir. Bu nedenle içerik kalitesini artırmak amacıyla 200 kelime ve altındaki metinler veri setinden çıkartılmıştır. Temizleme ve ön işleme adımının ardından 2208 adet nitelikli veri kalmıştır. Veri kümesi %80 eğitim, %20 test olacak şekilde ayrılmıştır. Düzenlenmiş veri setine ait sınıf bazlı örnek sayıları Tablo 2 'de gösterilmektedir.

B. Model seçimi (Model selection)

Çalışmada Türkçe hukuk metnlerinin sınıflandırılmasında başarıyı etkileyen faktörler göz önünde bulundurularak farklı yaklaşımlar denenmiştir. Geleneksel makine öğrenmesi yöntemleri, gömme tabanlı modeller ve Transformer tabanlı dil modelleri karşılaştırmalı olarak değerlendirilmiştir. Tablo 3'te modellerin temel özellikleri yer almaktadır.

Geleneksel makine öğrenmesi modellerinde metinler TFIDF yöntemiyle sayısal vektörlere dönüştürülmüştür. Ham veri setine ön işlem adımları uygulanmıştır. Metinler küçük harfe çevrilmiş, noktalama işaretleri ile Türkçe etkisiz kelimeler çıkarılmış ve ayırt ediciliği düşük olan TCK ve CMK gibi hukuk terimleri veri setinden temizlenmiştir. Sayısal temsil aşamasında TFIDF vektörleştirme kullanılmıştır. Vektörler üzerinde Naive Bayes, Logistic Regression ve SVM modelleri eğitilmiştir. Naive Bayes, kelime frekanslarına dayalı olasılıksal yapısı nedeniyle temel bir karşılaştırma modeli olarak kullanılmıştır. Logistic Regression, metinleri doğrusal bir karar fonksiyonu üzerinden sınıflandıran, basit ve yorumlanabilir bir yöntemdir. SVM ise yüksek boyutlu metin verilerinde sınıflar arasında maksimum ayrımı sağlayarak çalışmaktadır.

Gömme tabanlı sınıflandırmada, kelimelerin anlamsal ve morfolojik ilişkilerini temsil edebilmek amacıyla önceden eğitilmiş Türkçe FastText (cc.tr.300) vektörleri kullanılmıştır. Her bir karar metni, metni oluşturan kelimelere ait vektörlerin ortalaması alınarak 300 boyutlu bir vektör ile temsil edilmiştir. Elde edilen sayısal öznitelikler, SVM algoritmasına girdi olarak verilmiştir.

Hibrit derin öğrenme modellerinde kelimelerin anlamsal ilişkilerini vektör uzayında temsil eden kelime gömme ve derin öğrenme tabanlı modeller yer almaktadır. Metinler, maksimum 150 kelime (token) uzunluğunda olacak şekilde sınırlandırılmıştır. Ön işleme ve normalizasyon sonrası metinler,

Tablo 2. Veri setindeki sınıflar ve örnek sayıları.

Table 2. Classes in the dataset and number of samples.

Suç Türü	Veri Sayısı
Uyuşturucu Madde Ticareti	1107
Kullanmak için uyuşturucu madde bulundurma	310
Hayata ve Vücut Dokunulmazlığına Karşı Suçlar	289
Dolandırıcılık	81
Yağma	183
Hakaret	154
Tehdit	84
Toplam	2208

Tablo 3. Modeller ve temel özellikleri.

Table 3. Models and their key characteristics.

Model Grubu	Model	Temel Özellikler
Geleneksel ML	TFIDF+ Naive Bayes	Olasılık tabanlı, hızlı ve kelime frekansı odaklı
Geleneksel ML	TFIDF+Logistic Regression	Doğrusal ve yorumlanabilir model
Geleneksel ML	TFIDF+SVM	Maksimum marj ile güçlü ayrıştırma
Kelime Gömme	FastText + SVM	Alt kelime (n-gram) temsili
Derin Öğrenme	FastText+ BiLSTM	Çift yönlü bağlam öğrenme
Derin Öğrenme	BERT	Derin bağlamsal temsil
Derin Öğrenme	RobertaBERT	BERT'in gelişmiş ve optimize versiyonu
Derin Öğrenme	DistilBERT	Daha hızlı ve hafif BERT
Büyük Dil Modeli	GPT-5.2/GPT-5.4	Geniş dil bilgisi ve bağlamsal anlayış

çift yönlü bağlamı modelleyebilen BiLSTM mimarisine aktarılmıştır. FastText ve BiLSTM birleştirilerek hibrit bir yapı oluşturulmuştur. BiLSTM mimarisi, ardışık veriyi ileri ve geri yönde işleyen iki katmandan oluşmaktadır. Gizli durum boyutu 64 olarak belirlenmiştir. Her iki yönün son gizli durumları birleştirilmiş, aşırı öğrenmeyi azaltmak için %40 dropout uygulanmış ve ardından 128 boyutlu bir özellik vektörü tam bağlı katmana aktarılmıştır. Eğitim sürecinde batch size 16, epoch sayısı 10 ve öğrenme oranı 10^{-3} olarak kullanılmıştır. Optimizasyon sürecinde Adam algoritması tercih edilmiş ve modelin genelleme yeteneğini artırmak için 10^{-4} değerinde ağırlık azalımı (weight decay) uygulanmıştır. Eğitim sürecinde sınıf dengesizliğini gidermek için ağırlıklı örnekleme uygulanmıştır. FastText + BiLSTM modeli, hukuki metinlerde kelimeler arasındaki uzun mesafeli ilişkileri yakalayacağı düşünüldüğü için kullanılmıştır.

Modern sınıflandırma aşamasında, Türkçe NLP literatüründe yüksek başarımlar gösteren Transformer tabanlı mimariler kullanılmıştır. Türkçe için önceden eğitilmiş BERT, RoBERTa ve daha hafif bir mimari sunan DistilBERT modelleri proje veri seti üzerinde fine-tuning yöntemiyle yeniden eğitilmiştir. Metinler, her modelin kendi tokenizer yapısı kullanılarak en fazla 256 token olacak şekilde hazırlanmıştır. Sınıf dengesizliğini gidermek amacıyla, ters sınıf frekanslarına dayalı ağırlıklar kullanılarak özelleştirilmiş bir eğitim süreci uygulanmıştır. Modellerin ince ayar sürecinde kullanılan hiperparametreler literatürdeki standartlar temel alınarak farklı konfigürasyonların test edildiği deneme süreçleri sonucunda seçilmiştir. Eğitim aşamasında, modelin performansının gelişmediği ve hata payının artmaya başladığı noktayı belirlemek için erken durdurma mantığı gözetilmiş; aşırı öğrenmeyi önleyen 5 epoch tercih edilmiştir. Teknik donanım kapasitesi ve model mimarileri arasındaki farklar nedeniyle; yığın boyutu BERT için 16, RoBERTa için 8 ve DistilBERT için 32 olarak belirlenmiştir. Öğrenme oranı ise denemeler sonucunda BERT ve RoBERTa için 2×10^{-5} , DistilBERT modelinde 5×10^{-5} olarak uygulanmıştır [26], [12].

Ayrıca büyük dil modellerinin hukuki metin sınıflandırma üzerindeki performansını değerlendirmek için test veri seti üzerinde GPT-5.2 ve GPT-5.4 modelleri ile deneyler gerçekleştirilmiştir. Modellerde, az örneklili öğrenme tekniği kullanılarak sınıflandırma yapılmıştır. Modele sistem rolünde yalnızca kategori adını üretmesini isteyen bir talimat verilmiş ve farklı suç türlerini temsil eden örnek hukuk metinleri ile doğru sınıf etiketleri gösterilmiştir. Ardından test veri setindeki metinler modele verilerek yalnızca ilgili suç kategorisini üretmesi istenmiştir. Modelin çıktılarının tutarlılığını artırmak amacıyla modelin sıcaklık (temperature) değeri 0 olarak ayarlanmıştır. Test veri setinde yer alan toplam 442 karar metni değerlendirilmiş ve modelin ürettiği sınıf etiketleri gerçek etiketler ile karşılaştırılarak doğruluk değeri hesaplanmıştır.

Çalışmada, eğitilen denetimli makine öğrenmesi modellerinin performansı, herhangi bir ince ayar veya yeniden eğitim süreci uygulanmayan Üretken Yapay Zekâ modeli ile gerçek zamanlı olarak karşılaştırılmıştır. Karşılaştırma için OpenAI tarafından geliştirilen Transformer tabanlı GPT-5.4 büyük dil modeli tercih edilmiştir. Geleneksel yaklaşımı temsilen ise eğitim aşamasında en yüksek performansı gösteren TFIDF+SVM modeli sistemin yerel ortamına entegre edilmiştir.

Karşılaştırmanın tutarlılığını ve adilliğini sağlamak adına, eğitim setinde uygulanan ön işleme adımları (küçük harf dönüşümü, noktalama temizliği, stop-word eliminasyonu) ve TFIDF vektörleştirme süreci, gerçek zamanlı test sistemine birebir aktarılmıştır. Böylece TFIDF+SVM modeline giren verinin eğitim koşullarıyla tam uyumluluğu korunmuştur.

Üretken yapay zekâ modelinin hukuk metinlerine adaptasyonunu artırmak amacıyla az örneklili öğrenme tekniği uygulanmıştır. Sisteme her bir suç sınıfı için etiketlenmiş birer

örnek ve modelin bağlamı kavramasını sağlayan bir sistem rolü tanımlanmıştır. GenAI çıktılarının yapısal olarak karşılaştırılabilir olması için modele; "Sen bir hukuk metni sınıflandırma robotusun. Verilen karar metnini analiz et ve tanımlanan suç türleri arasından yalnızca bir tanesini seç. SADECE kategori ismini yaz, açıklama yapma" şeklinde katı bir kısıtlayıcı komut verilmiştir.

Test sürecinin yönetilmesi ve modellerin eğitim setinde bulunmayan verilerle manuel olarak sınanabilmesi için Gradio kütüphanesi ile web tabanlı etkileşimli bir arayüz geliştirilmiştir. Bu arayüz, araştırmacının girdiği metinlere her iki modelin verdiği yanıtları eş zamanlı görüntülemesine, yanıt sürelerini ölçmesine ve tahminlerin doğruluğunu anlık olarak kaydederek kümülatif başarı oranlarını raporlamasına olanak tanımıştır.

Elde edilen bulgular, üretken yapay zekâ modellerinin alan özgü eğitim olmaksızın bağlamsal tahmin yeteneğine sahip olduğunu ortaya koyarken; eğitilmiş makine öğrenmesi modellerinin yanıt süresi, alan özgüllüğü açısından daha kararlı ve hızlı çıktılar sunduğunu göstermiştir.

IV. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

Çalışmada, Türk hukuk sistemindeki dava karar metinlerinin suç türlerine göre otomatik sınıflandırılması amacıyla farklı yapay öğrenme modeli değerlendirilmiştir. Yapılan analizler sonucunda en yüksek doğruluk oranına sahip model seçilerek gerçek zamanlı sistem üzerinde test edilmiştir. Test aşamasında, seçilen modelin performansı, az örneklili öğrenme tekniği uygulanan GPT-5.4 ile karşılaştırmalı olarak incelenmiş ve elde edilen sonuçlar değerlendirilmiştir.

A. Model performans karşılaştırılması (Comparison of model performance)

Karar metinleri üzerinde gerçekleştirilen bu çalışmada, yedi farklı suç kategorisinin otomatik sınıflandırılması hedeflenmiştir. Geleneksel makine öğrenmesi algoritmaları kapsamında TFIDF + Naive Bayes, TFIDF + Lojistik Regresyon ve TFIDF + SVM incelenmiştir. Modellere ait doğruluk oranları ile F1 skorlarının sunulduğu Tablo 4 incelendiğinde ve elde edilen sonuçlar karşılaştırıldığında hem bu kategoride hem de tüm çalışma genelinde en yüksek performansın TFIDF + SVM modeli ile sağlandığı görülmüştür. TFIDF + SVM, %95.70 test doğruluğu ile diğer modellerden daha başarılı bir sonuç vermiştir. Ayrıca, F1 skoru metriğinde TFIDF + SVM %95.71 oranıyla bu çalışma kapsamında test edilen veri seti üzerinde en yüksek performansı sergileyen model olmuştur. Modeli sırasıyla %93.44 test başarıyla TFIDF + Lojistik Regresyon ve %88.69 test başarıyla TFIDF + Naive Bayes takip etmiştir.

Geleneksel yöntemlerde TFIDF+ SVM'in gösterdiği yüksek başarı üzerine kelime temsillerinin sınıflandırmaya katkısını ölçmek amacıyla FastText vektörleri ve SVM sınıflandırıcısının birlikte kullanıldığı hibrit bir yapı test edilmiştir. Ancak bu yaklaşım %83.48 test doğruluğu ve %83.50 F1 skoru ile beklenen performans artışını sağlayamamıştır. TFIDF tabanlı standart SVM modeline göre daha az doğruluk göstermiştir.

Derin öğrenme tabanlı modeller kategorisinde; FastText destekli BiLSTM mimarisi ile Transformer tabanlı BERT, RoBERTa ve DistilBERT modelleri analiz edilmiştir. Bu grupta en kararlı ve yüksek performansı BERT modeli sergilemiştir. Model %93.21 test doğruluğu ve %93.26 F1 skoru elde edilmiştir. FastText + BiLSTM modeli ise eğitim aşamasında %97.73 oranına ulaşmasına rağmen, test aşamasında %88.24 seviyesinde kalarak diğer modellere kıyasla aşırı öğrenme göstermiştir. Diğer Transformer modellerinden RoBERTa %89.36, DistilBERT ise %88.46 test doğruluğu ile BERT modelini takip etmiştir.

Büyük dil modellerinde kısmında ise test etmek için GPT-5.2 ve GPT-5.4 kullanılmıştır. Test verisinde bulunan 442 karar metni

Tablo 4. Modellerin doğruluk ve F1 skor değerleri.**Table 4.** Accuracy and F1-score values of the models.

Model	Eğitim Doğruluğu (%)	Test Doğruluğu (%)	F1 Skor (%)
TFIDF + Naive Bayes	90.77	88.69	89.19
TFIDF+Logistic Regression	95.53	93.44	93.52
TFIDF + SVM	99.55	95.70	95.71
FastText + SVM	86.69	83.48	83.5
FastText + BiLSTM	97.73	88.24	88
RobertBERT	91.53	89.36	89.4
DistilBERT	92.02	88.46	88.74
BERT	97.06	93.21	93.26
GPT-5.2	-	92.08	92.59
GPT-5.4	-	93.21	93.33

değerlendirilmiştir. GPT-5.2 bunların 407 tanesini doğru, 35 tanesini ise yanlış, GPT-5.4 ise 412 tanesini doğru 30 tanesini yanlış sınıflandırmıştır. Sonuçlara göre modellerin doğruluk değerleri sırasıyla %92.08, %93.21 olarak hesaplanmıştır. Bu sonuç, büyük dil modellerinin sınırlı örneklerle hukuki metinlerin sınıflandırılmasında iyi performans gösterebileceğini ortaya koymaktadır.

Genel bir değerlendirme yapıldığında çalışma kapsamındaki en başarılı sınıflandırma performansının %95.71 F1 skoru ile TFIDF + SVM tarafından sağlandığı, derin öğrenme yöntemleri arasında ise BERT modelinin öne çıktığı tespit edilmiştir.

B. Gerçek zamanlı sistem performansı (Real-time system performance)

Gerçek zamanlı sınıflandırma performansını değerlendirmek amacıyla çalışmada en yüksek başarıyı gösteren TFIDF + SVM algoritması ile az örnekli öğrenme tekniği uygulanan ChatGPT-5.4 modeli eş zamanlı olarak karşılaştırılmıştır. Modellerin tahmin süreçlerinin izlendiği ve veri girişinin yapıldığı arayüz ekranı Şekil 2'de sunulmuştur.

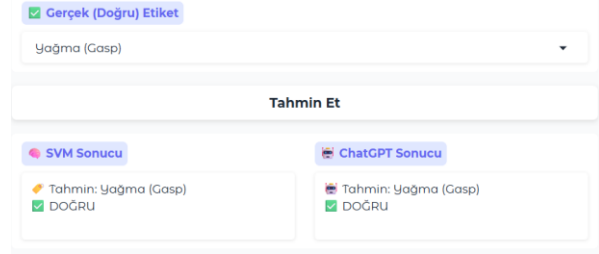
Performans analizi, geliştirilen web tabanlı arayüz üzerinden yapılmıştır. Modellerin veri setinde daha önce hiç görmediği özgün metinlerle gerçekleştirilmiştir. Çalışma kapsamındaki yedi farklı suç türünün her biri için sisteme birer adet örnek metin girilmiş ve modellerden tahmin üretmeleri istenmiştir.

Yapılan gerçek zamanlı testlere ait başarı durumlarını gösteren sonuçlar Şekil 3'te verilmiştir. Elde edilen bulgular incelendiğinde hem TFIDF + SVM modeli hem de GPT-5.4 modelinin yedi suç kategorisinin altısını doğru sınıflandırdığı, her ikisinin de yalnızca birer adet hatalı tahmin yaptığı görülmüştür. Sonuçlara göre her iki modelin de gerçek zamanlı senaryoda test edilen metin sayısına bağlı olarak %85.71 oranında benzer bir doğruluk başarısı sergilediği tespit edilmiştir.

V. TARTIŞMA (DISCUSSION)

Hukuk alanındaki Türkçe karar metinlerinin sınıflandırılması problemi, Geleneksel Makine Öğrenmesi, Derin Öğrenme ve Üretken Yapay Zekâ şeklinde üç farklı duruma göre ele alınmıştır.

Analizler sonucunda, Geleneksel Makine Öğrenmesi yöntemlerinden TFIDF + SVM %95.71 F1 skoru ile en yüksek performansı sergilemiştir. SVM'nin, TF-IDF gibi terim sıklığına dayalı özellik çıkarma yöntemiyle birleştiğinde derin öğrenme modellerini geride bırakması hukuk metinlerinin yapısıyla yakından ilişkilidir. Karar metinleri genellikle "beraat", "mahkûmiyet", "uyuşturucu" gibi ayırt edici gücü yüksek anahtar kelimeler içermektedir. Aynı zamanda karar metinleri genellikle dava metinlerinden daha kısa olduğu ve suç hakkında detaylı bilgi içermediği için anlamsal bağlama algoritmalarına göre veri seti kısıtları dahilinde daha başarılı sonuçlar verebilmektedir. Çalışmada kullanılan karar metinleri, gerçek dünya hukuk sisteminin doğası gereği sınıflar arası veri dengesizliği barındırmaktadır. Bu durumun modelleri çoğunluk sınıfını

**Şekil 2.** Veri girişinin yapıldığı arayüz.**Figure 2.** User interface for data Input.**Şekil 3.** Başarı durumlarını gösteren sonuçlar.**Figure 3.** Results showing performance outcomes.

ezberlemeye itmesini engellemek amacıyla başarı metriği olarak F1-Skoru da dikkate alınmıştır. TFIDF + SVM modelinin başarısı, modelin veri kümesi dengesizliğine karşı dirençli bir yapı sergilediğini ve azınlık sınıflarında kabul edilebilir bir ayırt edicilik sağladığını göstermektedir. Kelime gömmeleri kullanan FastText + SVM hibrit modelinin %83.50 F1 skoru ile daha düşük oran vermesi, genel amaçlı eğitilmiş vektörlerin hukuka özgü terminolojiyi TFIDF kadar keskin temsil edemediğini düşündürmektedir.

Transformer tabanlı BERT modeli %93.26 F1 skoru ile TFIDF + SVM'e en güçlü alternatif olmuştur. BERT'in çift yönlü bağlamı okuma yeteneği karmaşık ifadelerin çözülmesinde etkilidir. İnce ayar sürecinde tercih edilen hiper-parametreler doğrulama kaybının (validation loss) stabilize olduğu ve aşırı öğrenmenin (overfitting) engellendiği en kararlı noktayı temsil etmektedir. Ancak sınırlı ve dengesiz veri setinin bu doğruluğu aşağıya çektiği değerlendirilmektedir. Daha büyük veri seti ile başarıyı artırılabilir. Ayrıca, FastText + BiLSTM modelinde gözlemlenen eğitim %97.73 ve test %88.24 başarımlarındaki yaklaşık %10'luk farkın modelin veri setindeki gürültüyü ezberlemeye meyilli olduğunu ve genelleme yeteneğinin Transformer mimarilerine göre daha düşük kaldığını göstermiştir.

Çalışmanın gerçek zamanlı test aşamasında eğitilmiş TFIDF+SVM modeli ile az örnekli öğrenme uygulanan GPT-5.4 modeli karşılaştırılmıştır. Elde edilen sonuçlar, sınırlı sayıdaki test örneği kapsamında iki yaklaşımın benzer bir performans sergilediğini göstermektedir. Ancak TFIDF+SVM milisaniyeler içinde yanıt üretirken, LLM tabanlı sistemlerin yanıt süresinin daha uzun olması ve API maliyetleri, gerçek zamanlı büyük ölçekli uygulamalar için tartışma konusudur. Sonuç olarak başarı oranları, ceza hukuku kapsamındaki belirli suç tipleri için oldukça anlamlı olsa da bulguların veri seti kısıtları dahilinde yorumlanması ve farklı hukuk dallarındaki performans değişkenliğinin göz önünde bulundurulması gerekmektedir.

VI. KISITLAR ve GENELLENEBİLİRLİK (LIMITATIONS and GENERALIZABILITY)

Elde edilen yüksek sınıflandırma başarılarına rağmen, araştırmanın metodolojik ve veri odaklı bazı kısıtları bulunmaktadır. İlk olarak genellelenebilirlik açısından; eğitilen modeller özel olarak Türk Ceza Hukuku kapsamındaki yedi suç kategorisini içeren karar metinleri üzerinde optimize edilmiştir. Karar metinlerinin özet ve kalıplaşmış yapısı modellerin başarısını artırmış olsa da bu modellerin doğrudan Özel Hukuk

(örneğin borçlar hukuku, ticaret hukuku) dava metinlerine veya savcılık iddianameleri gibi daha uzun ve serbest metin formatlarına uyarlanması ek eğitim süreçleri gerektirecektir.

İkinci bir kısım olarak, hukuk terminolojisi ve kanun maddeleri yasal değişikliklerle sürekli güncellenmektedir. Statik veri setleri üzerinde eğitilen TFIDF+SVM ve BERT gibi modellerin bu yasal değişikliklere otomatik olarak uyum sağlayamayacağı göz önünde bulundurulmalıdır. Buna karşın, test edilen GPT-5.4 gibi üretken yapay zekâ tabanlı yaklaşımların, herhangi bir yeniden eğitim maliyeti gerektirmeksizin sadece komut bağlamının güncellenmesiyle yeni yasal durumlara daha esnek şekilde genelleneme potansiyeli taşıdığı değerlendirilmektedir. Bu durum, büyük dil modellerinin sürdürülebilir hukuk teknolojileri için bir avantaj sunduğunu göstermektedir.

VII. SONUÇ (CONCLUSION)

Türk yargı sistemine ait karar metinlerinin otomatik sınıflandırılması amacıyla yürütülen çalışmada, geleneksel makine öğrenmesi, derin öğrenme ve üretken yapay zekâ tabanlı yaklaşımların performansları karşılaştırmalı değerlendirilmiştir. Elde edilen deneysel bulgular TFIDF+SVM modelinin %95.71 F1 skoruna ulaşarak diğer modellere oranla daha yüksek bir sınıflandırma performansı sergilediğini göstermektedir. Hukuk terminolojisinin ayırt edici ve kalıplaşmış yapısı nedeniyle terim sıklığına dayalı geleneksel yöntemlerin bu veri seti özelinde karmaşık derin öğrenme mimarilerine kıyasla daha iyi sonuçlar ürettiği tespit edilmiştir. Derin öğrenme modelleri arasında Transformer tabanlı BERT modeli %93.26 F1 skoru ile en güçlü alternatif olarak öne çıkmıştır. Ancak FastText destekli hibrit modellerin veri setindeki gürültüyü ezberleme eğiliminde olduğu gözlemlenmiştir. Çalışmanın gerçek zamanlı uygulama aşamasında ise eğitilmiş TFIDF+SVM modeli ile az örneklili öğrenme tekniği uygulanan GPT-5.4 modeli veri setinde yer almayan toplam 7 yeni deneme verisi üzerinden karşılaştırılmıştır. Yapılan değerlendirmede her iki modelin de 7 örneğin 6'sını doğru tahmin ederek benzer bir doğruluk verdiği görülmüştür.

YAZAR BEYANI (AUTHOR STATEMENT)

İntihal Kontrolü (Plagiarism Check)—Makale iThenticate programı ile taranmış ve derginin intihal politikası ile uyumlu bulunmuştur.

Çıkar Çatışması (Conflict of Interest)—Makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

Etik Kurul Onayı (Ethics Committee Approval)—Makalede etik kurul izni alınmasına gerek yoktur.

Yapay Zekâ Araçlarının Kullanımı (Use of Artificial Intelligence Tools)—Makalede, dilbilgisi ve yazım denetimi için yapay zekâ araçları kullanılmış olup; tüm içerik yazarların özgün katkısını yansıtmaktadır.

Finansman ve Proje Desteği (Funding)—Bu çalışma için herhangi bir kurumsal/finansal destek alınmamıştır.

Veri Paylaşım (Data availability)—Bu çalışmanın bulgularını destekleyen veriler yazar(lar) tarafından üretilmiş olup; makul talepler doğrultusunda yazar(lar)dan temin edilebilir.

Yazar Katkısı (CRediT Author Contribution)—Yazma - ilk taslak, Yazılım, Yöntem, Doğrulama, Analiz Araçlarının Sağlanması, Veri Düzenleme, Görselleştirme (Eda Çetin, İlayda Kaya); Kavramsallaştırma, Yazma - inceleme ve düzeltme, Danışmanlık, Proje Yönetimi (Furkan Göz).

KAYNAKLAR (REFERENCES)

- [1] R. Susskind, *Tomorrow's lawyers: An introduction to your future*, Third Edition, Oxford, United Kingdom, Oxford University Press, 2023.
- [2] K. D. Ashley, *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge, United Kingdom, Cambridge University Press, 2017.
- [3] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androustopoulos, "LEGAL-BERT: The Muppets straight out of Law School", *Findings of the Association for Computational Linguistics: EMNLP*, Editors: T. Cohn, Y. He, Y. Liu, (Online), Association for Computational Linguistics, 2020, 2898-2904.

- [4] Harvey, "Harvey | Generative AI for Professional Services", <https://www.harvey.ai/> (20.12.2025).
- [5] LexisNexis, "Lexis+AI", <https://www.lexisnexis.com/en-int/products/lexis-plus-ai> (20.12.2025).
- [6] Thomson Reuters, "Westlaw Edge", <https://legal.thomsonreuters.com/en/products/westlaw-edge> (20.12.2025).
- [7] DoNotPay, "DoNotPay - Your AI Consumer Champion", <https://donotpay.com/> (20.12.2025).
- [8] ROSS Intelligence, "ROSS Intelligence - Legal Tech Corner", <https://blog.rossintelligence.com/> (20.12.2025).
- [9] C. Çetindağ, B. Yazıcıoğlu, A. Koç, "Named-entity recognition in Turkish legal texts", *Natural Language Engineering*, 29(3), (2023), 615-642.
- [10] O. Akça, "Natural language processings in legal domain: classification of Turkish legal texts", *Yüksek Lisans Tezi*, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2023.
- [11] E. Mumcuoğlu, C. E. Öztürk, H. M. Ozaktas, A. Koç, "Natural language processing in law: Prediction of outcomes in the higher courts of Turkey", *Information Processing and Management*, 58(5), (2021), 102684.
- [12] O. Akça, G. Bayrak, A. M. Issifu, M. C. Ganiz, "Traditional machine learning and deep learning-based text classification for Turkish law documents using transformers and domain adaptation", *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Biarritz, Fransa, 08-12 Ağustos 2022.
- [13] M. B. Görentaş, T. Uçkan, "Makine Öğrenmesi Yöntemleri Kullanılarak Mahkeme Kararlarının Kümelmesi", *Journal of Computer Science*, 8(2), (2023), 148-158.
- [14] D. Küçük, F. Can, "Exploiting artificial intelligence technologies for automatic processing of legal texts", *Bilişim Hukuku Dergisi*, 6(1), (2024), 1-23.
- [15] M. E. Okursoy, T. İnkaya, "Yüksek Mahkeme Kararlarının Sınıflandırılması için Veri Dengeleme ve Açıklanabilir Yapay Zekâ Tabanlı Karar Destek Sistemi", *Veri Bilimi*, 8(1), (2025), 11-23.
- [16] S. K. Demir, E. Aydemir, Y. Sönmez, "Türkiye'de Emsal Kararların Makine Öğrenmesi Algoritmaları ile Sınıflandırılması", *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 13(3), (2025), 1227-1239.
- [17] B. Kılıç, Y. Öner, "Yargıtay Kararlarının Suç Türlerine Göre Makine Öğrenmesi Yöntemleri İle Sınıflandırılması", *Veri Bilimi*, 4(3), (2021), 61-71.
- [18] F. Ariai, J. Mackenzie, G. Demartini, "Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges", *ACM Computing Surveys*, 58(6), (2025), 163.
- [19] B. Abimbola, Q. Tan, E. A. De La Cal Marín, "Sentiment analysis of Canadian maritime case law: a sentiment case law and deep learning approach", *International Journal of Information Technology*, 16(6), (2024), 3401-3409.
- [20] É. B. e Santos, I. A. Rodello, "Interpreting Lawsuits Contexts through Probabilistic Topic Modeling", *Revista Eletrônica de Iniciação Científica em Computação*, 23(1), (2025), 81-90.
- [21] D. Singh, "Legal documents text analysis using natural language processing (NLP)", *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India, 23-25 October 2024, 1302-1307.
- [22] A. Zadgaonkar, A. J. Agrawal, "An approach for analyzing unstructured text data using topic modeling techniques for efficient information extraction", *New Generation Computing*, 42(1), (2024), 109-134.
- [23] D. Trautmann, N. Ostapuk, Q. Grail, A. Pol, G. Bonifazi, S. Gao, M. Gajek, "Measuring the groundedness of legal question-answering systems", *Proceedings of the Natural Legal Language Processing Workshop 2024*, Miami, FL, USA, 15 November 2024, 176-186.
- [24] S. Çelik, B. Doğanlı, A. Cengiz, "Citation classification and key phrase extraction in legal texts using machine learning: an examination of the legal text classification dataset", *Quality & Quantity*, 60, (2026), 6645-6681.
- [25] T. Turan, E. U. Küçüksille, "Legal text classification in Turkey: A machine learning approach to divorce and zoning decisions", *Uluslararası Mühendislik Tasarım ve Teknoloji Dergisi*, 6(2), (2024), 53-63.
- [26] H. H. Chen, L. Wu, J. P. Chen, W. Lu, J. H. Ding, "A comparative study of automated legal text classification using random forests and deep learning", *Information Processing & Management*, 59(2), (2022), 102798.
- [27] A. A. Khaliq, D. Riva, S. Montanelli, "Evaluating knowledge-based approaches for legal text analysis: A benchmark study", *Computer Law & Security Review*, 61, (2026), 106279.