



Kurumsal Şikayet Yönetimi için Makine Öğrenmesi Tabanlı Otomatik E-posta Sınıflandırma Sistemi

İsmail Enes Parlak¹ 

¹ Endüstri Mühendisliği Bölümü, Mühendislik ve Doğa Bilimleri Fakültesi, Bursa Teknik Üniversitesi, Bursa, Türkiye
enes.parlak@btu.edu.tr

Öz

Kurumsal hediyeleşme süreçleri sonrasında, özellikle çalışanlara yan hak olarak sunulan Ramazan, yılbaşı gibi hediye paketlerinin dağıtımının ardından gelen şikayet e-postalarının manuel yönetimi, satın alma departmanları için ciddi bir operasyonel yük oluşturmaktadır. Bu çalışma, bu spesifik probleme odaklanarak, şikayet e-postalarının otomatik sınıflandırılması ve yönetimi için yapay zekâ (YZ) destekli, çok etiketli bir sistem geliştirmeyi amaçlamaktadır. Sistem, şikayetleri; ürün türü, paket türü, şikâyet türü ve sorumlu kişi olmak üzere dört ana kategoride sınıflandırmaktadır. Çalışmada, gerçek şirket verilerinin gizlilik kısıtları nedeniyle, büyük dil modelleri kullanılarak ve gerçek operasyonel senaryolar temel alınarak 1.870 adet sentetik şikâyet e-postası içeren özgün bir veri seti oluşturulmuş ve TF-IDF yöntemiyle metinler vektörleştirilmiştir. CatBoost, XGBoost ve Destek Vektör Makineleri (SVM) dâhil olmak üzere on farklı makine öğrenmesi algoritmasının performansı karşılaştırmalı olarak değerlendirilmiştir. Elde edilen bulgular, geliştirilen sistemin %87.75 ortalama F1-skoru başarıyla gösterdiğini ortaya koymuştur. Literatürdeki spam ve oltalama odaklı çalışmaların aksine, bu çalışmanın özgün katkısı; çok etiketli kurumsal şikâyet yönetimi problemine SHAP (SHapley Additive exPlanations) yöntemini uygulayarak model kararlarını şeffâf hale getirmesidir. Bu sayede, model tahminlerinin hangi kelimelere dayandığı açıklanarak sisteme duyulan güven artırılmıştır. Geliştirilen bu açıklanabilir YZ destekli sistem, kurumsal hediye paketleriyle ilgili şikâyet yönetimi süreçlerinde otomasyon ve verimlilik artışı için etkin ve yorumlanabilir bir çözüm sunmaktadır.

Anahtar kelimeler: Kurumsal Şikâyet Yönetimi, Çok Etiketli Sınıflandırma, Açıklanabilir Yapay Zekâ, Sentetik Veri Üretimi, E-Posta Sınıflandırma.

A Machine Learning-Based Automatic E-mail Classification System for Corporate Complaint Management

Abstract

Manual management of complaint emails following corporate gifting processes, specifically the distribution of employee benefits such as Ramadan or New Year packages, creates a significant operational burden for procurement departments. This study focuses on this specific problem and aims to develop an artificial intelligence (AI)-supported, multi-label system for the automatic classification and management of complaint emails. The system classifies complaints into four main categories: product type, package type, complaint type, and responsible person. In the study, to overcome privacy constraints associated with real-world corporate data, a unique dataset containing 1,870 synthetic complaint emails was generated using large language models based on actual operational scenarios. Texts were vectorized using the TF-IDF method, and the performance of ten machine learning algorithms, including CatBoost, XGBoost, and Support Vector Machines (SVM), was evaluated. The findings revealed that the developed system achieved an average F1-score of 87.75%. Distinguishing itself from the extensive literature primarily focused on spam and phishing detection, this study offers a unique contribution by applying the SHAP (SHapley Additive exPlanations) method to the domain-specific problem of multi-label corporate complaint management. In this way, trust in the system is increased by explaining which words the model predictions are based on. The developed explainable AI-supported system offers an effective and interpretable solution for increasing automation and efficiency in complaint management processes.

Keywords: Corporate Complaint Management, Multi-label Classification, Explainable Artificial Intelligence, Synthetic Data Generation, Email Classification.

* Sorumlu yazar.
E-posta adresi: enes.parlak@btu.edu.tr

Alındı : 19 Ekim 2025
Revizyon : 22 Aralık 2025
Kabul : 23 Ocak 2026

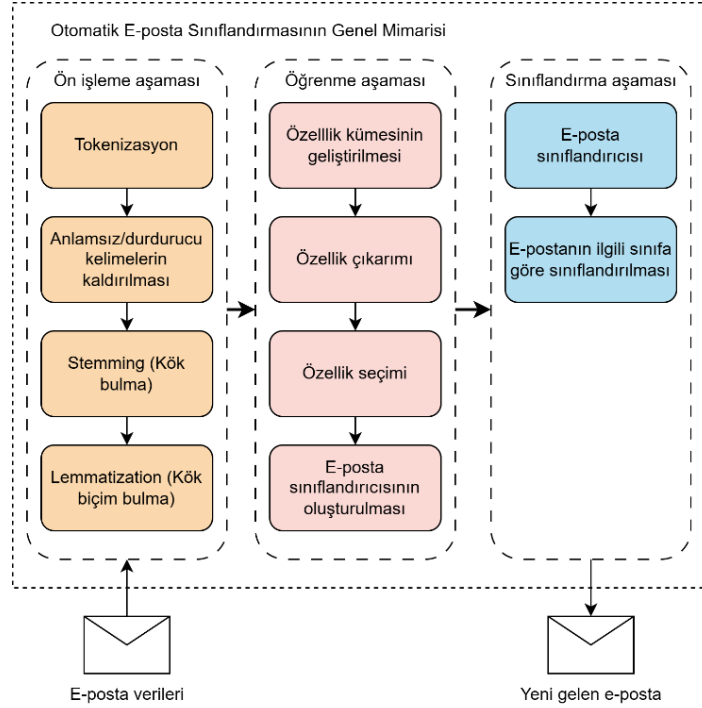
1. Giriş (Introduction)

Kurumsal hediyeleşme süreçlerinin ardından satın alma departmanlarına ulaşan yoğun şikayet e-postalarının manuel yönetimi, ciddi zaman kaybına ve insan kaynaklı hata riskine yol açarak operasyonel verimliliği düşürmektedir (Fung, 2014; Willcocks & Lacity, 2016). Artan dijital iletişim hacmiyle birlikte, bu süreci otomatikleştiren mekanizmalara (Brutlag & Meek, 2000) ve iletileri önceden tanımlanmış kategorilere ayıran araçlara (Cohen, 1996) duyulan ihtiyaç, makine öğrenmesi (ML) alanındaki gelişmelerle karşılanmaya başlanmıştır (Mitchell, 1997). Literatürde, özellikle SVM gibi algoritmaların metin sınıflandırmada üstün performans gösterdiği raporlanmış (Kiritchenko & Matwin, 2001) olsa da, günümüzde Derin Öğrenme (Goodfellow vd., 2016) ve Transformer tabanlı modern NLP mimarileri (Patwardhan vd., 2023) bu teknolojilerin en önemli uygulama alanları haline gelmiştir. E-posta sınıflandırma sistemlerinin genel mimarisi literatürde büyük ölçüde standartlaşmış olup (Mujtaba vd., 2017), otomatik sınıflandırma için önerilen mimari Şekil 1'de gösterilmiştir (Mujtaba vd., 2017).

Veri setlerinin zenginleştirilmesi ve özellikle dengesiz veri dağılımının giderilmesi amacıyla Büyük Dil Modellerinin (LLM) kullanımı literatürde giderek yaygınlaşmaktadır. Örneğin, Dai vd. (2025) tarafından geliştirilen AugGPT yaklaşımı, ChatGPT kullanılarak yapılan veri artırmanın metin sınıflandırma başarısını artırdığını ortaya koymuştur. Benzer şekilde, Gopali vd. (2024), dengesiz veri setlerinde LLM tabanlı sentetik veri üretiminin, özellikle azınlık sınıflar için F1 skorunu önemli ölçüde iyileştirdiğini raporlamıştır. ValizadehAslani vd. (2024) ise ChatGPT destekli veri artırma yönteminin, geleneksel yöntemlere kıyasla daha üstün bir sınıflandırma performansı sağladığını

göstermiştir. Bu güncel çalışmalar, veri gizliliği veya veri azlığı yaşanan durumlarda sentetik veri üretiminin bilimsel olarak geçerli ve etkili bir yöntem olduğunu doğrulamaktadır.

Ancak, yüksek performanslı algoritmaların karar süreçlerinin şeffaf olmaması (kara kutu problemi), kurumsal güveni zedeleyerek sistemlerin benimsenmesini zorlaştırmaktadır. Bu şeffaflık sorununu aşmak için son yıllarda Açıklanabilir Yapay Zeka (XAI) yaklaşımlarına olan ilgi artmıştır. Özellikle Mosca vd. (2022), SHAP tabanlı yöntemlerin Doğal Dil İşleme (NLP) alanındaki uygulanabilirliğini kapsamlı bir şekilde incelemiş ve karmaşık metin modellerinin yorumlanmasında kritik bir rol oynadığını belirtmiştir. Derin öğrenme mimarilerinde de bu arayış sürmektedir; örneğin Kocalj vd. (2021), Transformer tabanlı (BERT) modeller için SHAP açıklamalarını genişleten TransSHAP yöntemini önermiş, Zhao vd. (2020) ise CNN tabanlı metin sınıflandırma modelleri için yerel açıklanabilirlik sağlayan bir SHAP metodolojisi geliştirmiştir. Benzer şekilde Dewi vd. (2022), metin sınıflandırma ve duygu analizi görevlerinde model çıktılarını yorumlamak için SHAP tekniğini başarıyla uygulamıştır. Bu çalışmalar, metin sınıflandırmada sadece yüksek doğruluğun değil, aynı zamanda kararların açıklanabilirliğinin de modern literatürün ana odak noktalarından biri haline geldiğini göstermektedir. Bu çalışma, literatürdeki bu boşluğu doldurmak amacıyla, şikayet e-postalarını ürün, paket, şikayet ve sorumlu kişi türlerine göre çok etiketli sınıflandıran ve karar mekanizmalarını SHAP yöntemiyle şeffaf hale getiren (Lundberg & Lee, 2017) özgün bir sistem sunmaktadır. Geliştirilen bu yaklaşım, yalnızca sınıflandırma başarısına odaklanmakla kalmayıp, Açıklanabilir Yapay Zeka (XAI) alanına da katkı sunarak (Vilone & Longo, 2021) kurumsal şikayet yönetimi probleminin güvenilir ve yorumlanabilir bir çözüm getirmektedir.

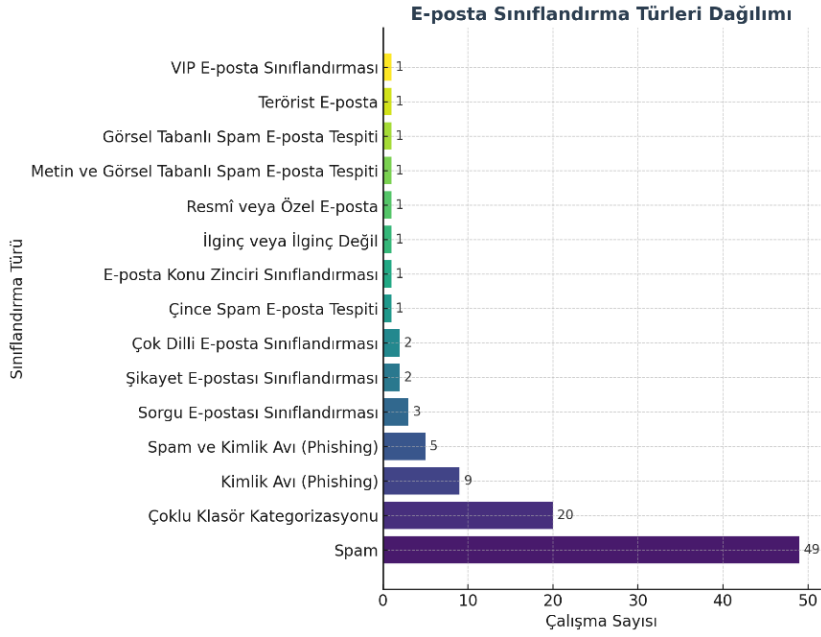


Şekil 1. Otomatik e-posta sınıflandırmasının genel mimarisi (General architecture of automatic email classification (Mujtaba vd., 2017))

E-posta sınıflandırması üzerine yapılan akademik çalışmalar incelendiğinde, araştırmaların büyük bir çoğunluğunun spam ve ortalama (phishing) tespiti üzerine yoğunlaştığı görülmektedir. Buna karşın, son dönemde müşteri geri bildirimlerinin ve şikayet metinlerinin analizi üzerine odaklanan çalışmalar artış göstermiştir. Örneğin Deniz vd. (2022), Türkçe e-ticaret yorumlarının çok etiketli sınıflandırılması üzerine makine öğrenmesi tabanlı bir yaklaşım sunarken, Cai vd. (2025) sivil havacılık alanındaki şikayet metinleri için derin öğrenme destekli bir model önermiştir. Benzer şekilde Kuyucuk ve Çallı (2022), kargo hizmetlerine yönelik şikayetleri analiz etmek için çok etiketli sınıflandırma yöntemlerini kullanmış, Türkmen vd. (2023) ise müşteri yorumları üzerinden çalışan davranışlarının analizine odaklanmıştır. Mujtaba vd. (2017) tarafından yapılan kapsamlı bir literatür taraması, incelenen 98 makalenin 49'unun doğrudan spam tespiti üzerine olduğunu, bunu çok klasörlü kategorizasyon ve ortalama tespitinin izlediğini göstermektedir (Şekil 2).

Buna karşın, "Şikayet E-postası Sınıflandırması (Complaint Email Classification)" gibi spesifik kurumsal ihtiyaçlara odaklanan çalışmaların sayısı oldukça sınırlıdır.

Kurumsal hediye paketleri sonrası gelen şikayet e-postalarının sınıflandırılması, spam tespitinden daha farklı ve karmaşık zorluklar barındırır. Bu e-postalar genellikle birden fazla sorunu aynı anda içerebilir (örneğin, hem ürünün eksik olması hem de ambalajının hasarlı olması), bu da çok etiketli (multi-label) bir sınıflandırma yaklaşımını zorunlu kılar. Ayrıca, kullanılan dilin çeşitliliği, kuruma özgü terminoloji ve metinlerdeki belirsizlikler, genel amaçlı sınıflandırıcılardan daha sofistike modellere ihtiyaç duyulmasına neden olmaktadır. Literatürdeki bu boşluk, mevcut çalışmanın temel motivasyonunu oluşturmaktadır: Kurumsal şikayet yönetimine odaklanan, çok etiketli ve yorumlanabilir bir sınıflandırma sistemi geliştirmek.



Şekil 2. E-posta sınıflandırma çalışmalarının uygulama alanlarına göre dağılımı (Distribution of articles by application area (Muftaba et al., 2017))

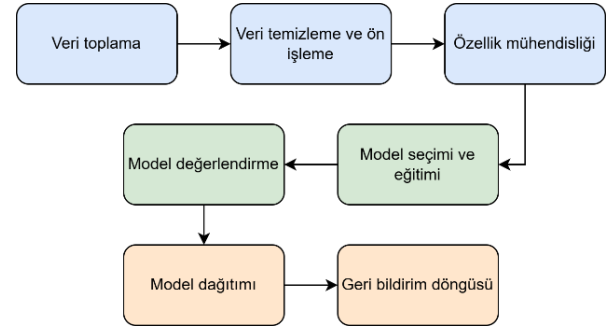
Gradyan artırma (gradient boosting) gibi modern ML modelleri yüksek tahmin doğruluğu sunsa da, karar mekanizmalarının anlaşılabilmesi nedeniyle "kara kutu" olarak nitelendirilmektedir. Kurumsal bir ortamda, bir şikayetin neden belirli bir departmana veya sorumlu kişiye atandığının gerçekleştirilememesi, bu sistemlere olan güveni zayıflatır ve pratik kullanımını engeller. Açıklanabilir Yapay Zeka (XAI), bu soruna çözüm olarak, modellerin tahminlerini insanlar için anlaşılır ve yorumlanabilir hale getirmeyi amaçlayan bir araştırma alanıdır (Adadi & Berrada, 2018). XAI, bir modelin belirli bir kararı 'neden' verdiğini gerçekleştirmek, modelin iç işleyişini kontrol etmek ve yeni bilgilerin keşfedilmesini sağlamak gibi amaçlara hizmet eder (Vilone & Longo, 2021).

Bu alandaki en popüler yöntemlerden biri olan SHAP (Lundberg & Lee, 2017), oyun teorisindeki Shapley Değeri kavramını kullanarak her bir özelliğin (örneğin metindeki bir kelimenin) modelin tahminine yaptığı katkıyı adil bir şekilde ölçer. Bu sayede, "ramazan" kelimesinin neden bir e-postayı belirli bir tedarikçiyle ilişkilendirdiğini veya "yırtık" kelimesinin neden "ambalaj hasarlı" şikayet türünü tetiklediğini sayısal olarak görmek mümkün hale gelir. Bu çalışmada SHAP yönteminin kullanılması, geliştirilen sınıflandırma sistemine literatürdeki birçok benzer çalışmada eksik olan şeffaflık ve güvenilirlik katmanını eklemiştir.

2. Materyal ve Metot (Materials and Methods)

Bu bölümde, kurumsal şikayet e-postalarının otomatik sınıflandırılması için geliştirilen sistemin metodolojik çerçevesi, kullanılan veri seti, uygulanan ön işleme teknikleri, model eğitim ve değerlendirme süreçleri ile model yorumlanabilirliği için kullanılan Açıklanabilir Yapay Zeka (XAI) yaklaşımı

detaylandırılmaktadır. Çalışmada izlenen iş akışının ana adımları Şekil 3'te özetlenmiştir.



Şekil 3. Metodolojide yer alan makine öğrenmesi uygulama çerçevesi (The machine learning application framework)

Şekil 3, çalışmanın metodolojik temelini oluşturan makine öğrenmesi iş akışını göstermektedir. Bu süreç, "Veri toplama", "Veri temizleme ve ön işleme" ve "Özellik mühendisliği" adımlarını içeren veri hazırlama aşamasıyla başlamaktadır. Ardından "Model seçimi ve eğitimi" ile "Model değerlendirme" adımlarını kapsayan modelleme aşamasına geçilmiştir. Son olarak, "Model dağıtımı" ile sistem operasyonel hale getirilirken, "Geri bildirim döngüsü" ile sistemin sürekli izlenmesi ve iyileştirilmesi hedeflenmektedir.

2.1. Veri kümesi ve ön işleme (Dataset and Preprocessing)

Gerçek dünya verilerinin gizlilik kısıtları ve etikleme maliyetleri göz önünde bulundurularak, bu çalışmada yapay bir veri seti oluşturulmuştur. Çalışanlara yan hak olarak dağıtılan Ramazan, yılbaşı ve hijyen paketlerine yönelik şikayet senaryolarını simüle eden toplam 1.870 adet şikâyet e-postası; ChatGPT, Claude ve Gemini gibi büyük dil modelleri (LLM)

kullanılarak yapay olarak üretilmiştir. Veri üretim sürecinin tekrarlanabilirliğini (reproducibility) sağlamak amacıyla, LLM'lere verilen komut istemi (prompt) yapısı, şirketin geçmiş şikayet kayıtlarından elde edilen gerçek senaryolar baz alınarak standardize edilmiştir. Oluşturulan komut şablonunda modele, 'kurumsal bir firmada çalışan ve hediye paketinden memnun olmayan bir personel' rolü atanmış ve 'satın alma departmanına hitaben bir şikayet e-postası yazması' görevi verilmiştir. Üretilen verinin çeşitliliğini sağlamak için istem içerisine; paket türü (Ramazan, Yılbaşı, Hijyen), şikayet sebebi (ürün ezik, tarihi geçmiş, eksik ürün, ambalaj yırtık), sorumlu tedarikçi (A/B Tedarikçisi, Kargo Firması) ve üslup (öfkeli, hayal kırıklığına uğramış, resmi, alaycı) gibi değişken parametreler rastgele kombinasyonlarla entegre edilmiştir. Ayrıca, çıktılarının tutarlılığını korumak adına metinlerin Türkçe olması ve 50-100 kelime aralığında uzunluğa sahip olması gibi yapısal kısıtlamalar da komut setine eklenmiştir. Bu yaklaşım, farklı dil modellerinin üslup çeşitliliğinden faydalanarak zengin ve gerçekçi bir veri seti elde edilmesini sağlamıştır.

Oluşturulan e-postalar, her birinin içeriği detaylı olarak incelenerek manuel olarak dört farklı kategoride etiketlenmiştir: Ürün türü (68 sınıf), paket türü (4 sınıf), şikâyet türü (12 sınıf) ve sorumlu kişi türü (5 sınıf). Bir e-postanın birden fazla ürün veya şikâyet türü içerebilmesi nedeniyle, problem çok etiketli (multi-label) bir sınıflandırma yapısına uygun olarak tasarlanmıştır.

Metin verileri, model eğitime hazırlanmadan önce kapsamlı bir ön işleme sürecinden geçirilmiştir. Bu süreçte; Türkçe karakterlerin standardizasyonu, tüm metinlerin küçük harfe dönüştürülmesi ve gereksiz boşlukların temizlenmesi gibi metin normalizasyon işlemleri uygulanmıştır.

Ön işleme adımlarının ardından, metin verilerini makine öğrenmesi algoritmalarının işleyebileceği sayısal bir formata dönüştürmek amacıyla Terim Frekansı-Ters Belge Frekansı (TF-IDF) vektörizasyon yöntemi kullanılmıştır. Bu yöntemde, hem yeterli ayırt edici bilgiyi korumak hem de hesaplama karmaşıklığını makul bir düzeyde tutmak amacıyla en önemli 3.000 özellik (kelime) seçilerek bir özellik uzayı oluşturulmuştur.

Çok etiketli sınıflandırma probleminin gerektirdiği yapıya uygun olarak, her bir kategoriye ait etiketler MultiLabelBinarizer kullanılarak ikili (binary) vektör formatına dönüştürülmüştür. Son olarak, veri seti %80 eğitim (1.496 örnek) ve %20 test (374 örnek) olmak üzere rastgele bölünerek model eğitimi ve değerlendirmesi için hazır hale getirilmiştir.

2.2. Model eğitimi ve değerlendirme (Model Training and Evaluation)

Bu çalışmada, metin sınıflandırma alanında yaygın olarak kullanılan on farklı makine öğrenmesi algoritmasının performansı karşılaştırmalı olarak

değerlendirilmiştir. Bu algoritmalar; XGBoost, LightGBM, CatBoost, Random Forest, AdaBoost, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression ve Naive Bayes'tir.

Model optimizasyon sürecinde, her bir algoritma için RandomizedSearchCV yöntemi kullanılarak 3 katlı çapraz doğrulama (3-fold cross-validation) ile en iyi hiperparametreler aranmıştır. Hesaplama maliyeti ve performans dengesi gözetilerek belirlenen arama uzayında; XGBoost için ağaç derinliği (max depth: 5, 7, 9), öğrenme oranı (learningrate: 0.01, 0.1) ve tahminci sayısı (n estimators: 100, 200); CatBoost için derinlik (depth: 6, 8) ve iterasyon (iterations: 100, 200); LightGBM için yaprak sayısı (num leaves: 50, 100); SVM için ceza parametresi (C: 0.1, 1, 10) ve lineer çekirdek fonksiyonu; Random Forest için ise ağaç sayısı (n estimators: 50, 100) parametreleri test edilmiştir. Ayrıca, TF-IDF vektörizasyonunda özellik sayısı (max features) 3.000 ile sınırlandırılmıştır. Bu eşik, kelime dağarcığının uzun kuyruk (long-tail) dağılımı incelenerek; çok nadir kelimelerin yarattığı gürültüyü engellemek ve özellik uzayının seyrekliğini optimize ederek aşırı öğrenmeyi önlemek amacıyla seçilmiştir.

Çok etiketli sınıflandırma yapısını desteklemek amacıyla, her bir temel model MultiOutputClassifier sarmalayıcısı (wrapper) içerisinde kullanılmıştır. Modellerin aşırı öğrenmesini önlemek ve en uygun parametreleri bulmak için RandomizedSearchCV yöntemiyle hiperparametre optimizasyonu yapılmıştır. Model performansının güvenilir bir şekilde ölçülmesi için veri setindeki etiket dağılımını koruyan MultilabelStratifiedKFold ile 3-katlı çapraz doğrulama (3-fold cross-validation) uygulanmıştır. Modellerin sınıflandırma başarısı, özellikle dengesiz veri setleri için güvenilir bir metrik olan F1-skoru (F1-score) ve bu metriğin micro, macro, weighted ve samples average gibi farklı ortalama stratejileri kullanılarak değerlendirilmiştir.

2.3. Model yorumlanabilirliği (Model Interpretability)

Geliştirilen modellerin "kara kutu" doğasını ortadan kaldırmak ve karar mekanizmalarını şeffaf hale getirmek amacıyla Açıklanabilir Yapay Zeka (XAI) tekniklerinden yararlanılmıştır. Bu kapsamda, Lundberg & Lee (2017) tarafından geliştirilen ve oyun teorisindeki Shapley Değerleri kavramına dayanan SHAP yöntemi kullanılmıştır. SHAP analizi, her bir özelliğin (bu çalışmada her bir kelimenin) modelin belirli bir sınıf için yaptığı tahmine olan pozitif veya negatif katkısını sayısal olarak hesaplamaktadır. Bu yöntem sayesinde, modellerin hangi kelime örüntülerini kullanarak karar verdiği anlaşılabilir sistemin yorumlanabilirliği ve güvenilirliği artırılmıştır.

3. Bulgular ve Tartışma (Results and Discussion)

Bu bölümde, geliştirilen otomatik e-posta sınıflandırma sisteminin performansı, dört ana kategori bazında sunulmakta ve seçilen en iyi modellerin karar mekanizmaları Açıklanabilir Yapay Zeka (XAI) yöntemleriyle analiz edilmektedir.

3.1. Modellerin karşılaştırmalı performansı (Comparative Performance of the Models)

Sistemin sınıflandırma başarımını değerlendirmek amacıyla on farklı makine öğrenmesi algoritması test edilmiştir. Her bir sınıflandırma kategorisi (ürün türü, paket türü, şikâyet türü, sorumlu kişi türü) için en yüksek F1-skoru değerini sağlayan model, nihai model olarak seçilmiştir. Dört kategori için elde edilen genel performans sonuçları ve seçilen en iyi modeller Tablo 1'de özetlenmiştir. Ayrıca, model performansını çok etiketli sınıflandırma bağlamında daha kapsamlı değerlendirmek amacıyla Hamming Loss ve Jaccard

Skoru metrikleri de analiz edilmiştir. Tablo 1'de görüldüğü üzere, 0.0037 ile 0.0706 arasında değişen düşük Hamming Loss değerleri, modellerin yanlış etiketleme oranının minimize edildiğini göstermektedir. Buna ek olarak, 0.74 ile 0.89 arasında gerçekleşen Jaccard skorları, tahmin edilen etiket kümeleri ile gerçek etiketler arasındaki yüksek örtüşmeyi ve modellerin tutarlılığını doğrulamaktadır.

Tablo 1'de görüldüğü üzere, geliştirilen sistem dört kategorinin genelinde %87.75'lik bir ortalama F1-skoru elde ederek yüksek bir başarıyı sergilemiştir. Her kategori için en iyi performansı gösteren algoritmanın farklılık göstermesi, problemin doğasına göre model seçiminin önemini ortaya koymaktadır. Ürün türü gibi kategorik özelliklerin yoğun olduğu sınıflandırmada CatBoost (%89) öne çıkarken, paket ve sorumlu kişi sınıflandırmasında XGBoost (sırasıyla %92 ve %82) en dengeli performansı sunmuştur. Yüksek boyutlu metin verisi içeren şikâyet türü sınıflandırmasında ise Destek Vektör Makineleri (SVM) (%88) en başarılı sonuçları vermiştir.

Tablo 1. Sınıflandırma kategorilerine göre seçilen en iyi modeller ve performans sonuçları (Selected best models and performance results by classification category)

Sınıflandırma kategorisi	Seçilen en iyi model	Micro ort. F1	Macro ort. F1	Ağırlıklı ort. F1	Örneklem ort. F1	Hamming kaybı (hata oranı)	Jaccard skoru
Ürün türü	CatBoost	0.89	0.91	0.89	0.84	0.0037	0.83
Paket türü	XGBoost	0.92	0.89	0.92	0.87	0.0374	0.89
Şikâyet türü	SVM	0.88	0.88	0.87	0.86	0.0332	0.74
Sorumlu kişi türü	XGBoost	0.82	0.80	0.81	0.76	0.0706	0.75

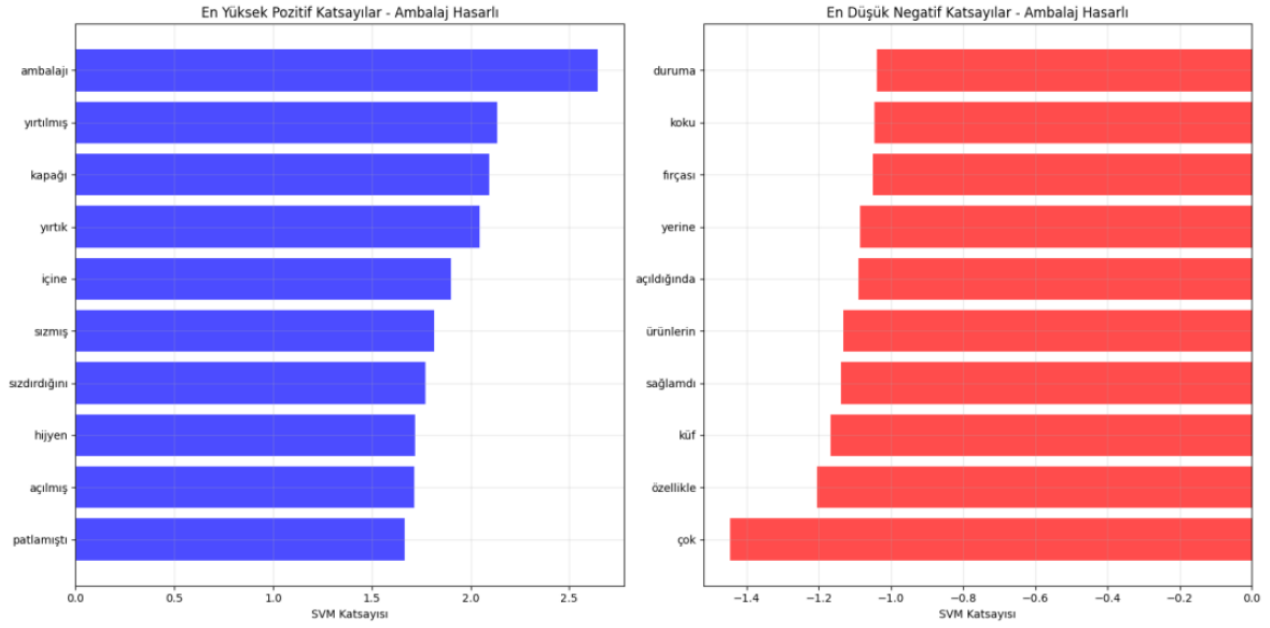
3.2. Model yorumlanabilirliği ve XAI analizi (Model Interpretability and XAI Analysis)

Sistemin sadece yüksek doğrulukta tahminler üretmesi değil, aynı zamanda bu tahminleri nasıl yaptığının anlaşılması da kritik öneme sahiptir. Bu amaçla, seçilen en iyi modellerin karar mekanizmaları XAI yöntemleri kullanılarak analiz edilmiştir.

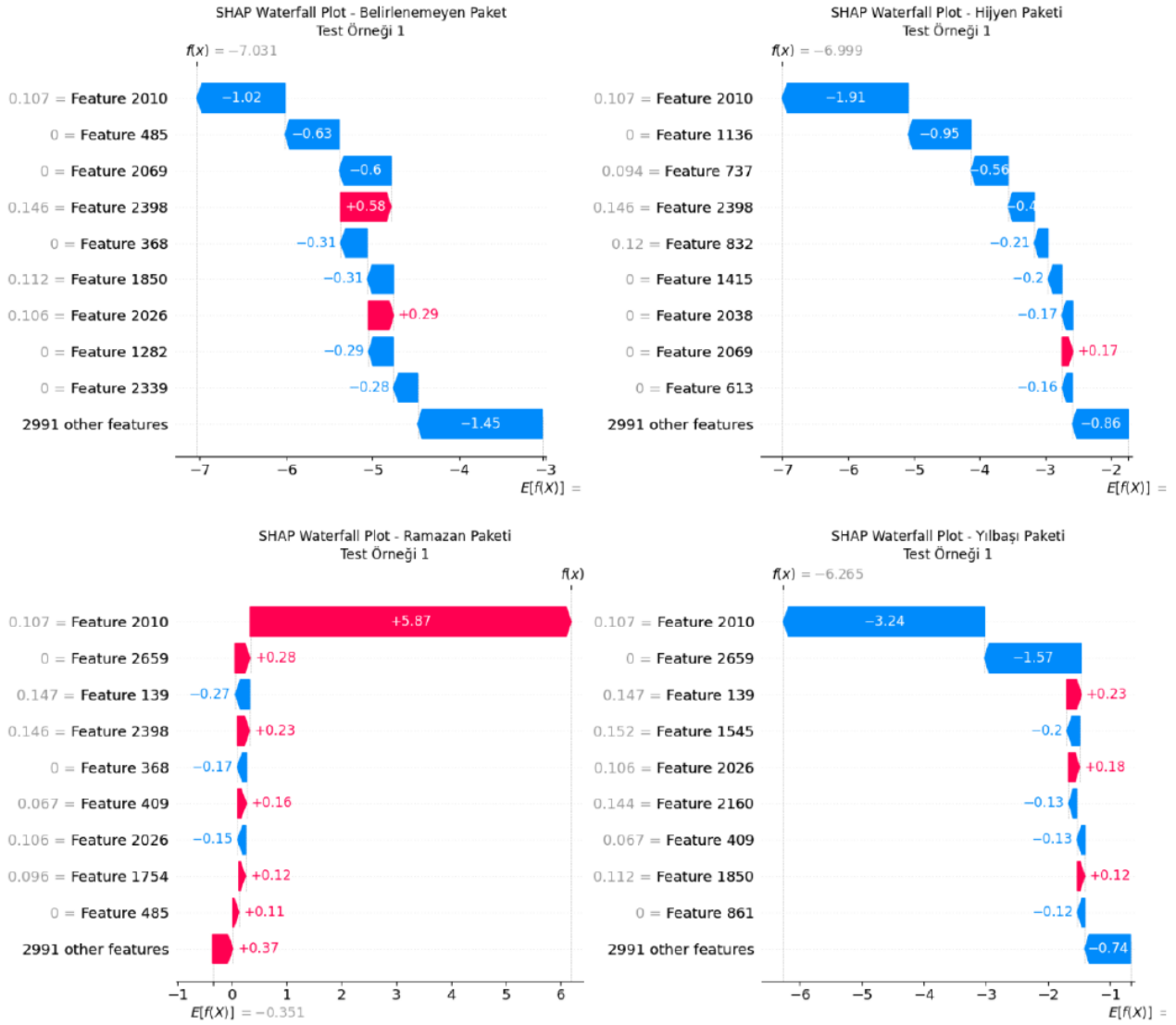
Şikâyet türü sınıflandırmasında en başarılı model olan SVM, doğrusal (linear) bir çekirdek ile eğitildiği için modelin özellik katsayıları doğrudan yorumlanabilmektedir. Şekil 4, "Ambalaj Hasarlı" şikâyet türü için modelin en yüksek pozitif ve en düşük (en negatif) katsayıları sahip kelimelerini göstermektedir. Modelin "ambalajı", "yırtılmış",

"kapağı" ve "yırtık" gibi kelimelere yüksek pozitif ağırlıklar ataması, karar mekanizmasının doğrudan problemle ilişkili ve mantıksal olduğunu kanıtlamaktadır.

Daha karmaşık olan paket türü sınıflandırması için en başarılı model olan XGBoost'un yorumlanması amacıyla SHAP yöntemi kullanılmıştır. Şekil 5, "Ramazan Paketi" tahmini için bir test örneği üzerinden oluşturulan SHAP waterfall plot grafiğini sunmaktadır. Grafikte görüldüğü gibi, "ramazan" kelimesi tek başına +5.87'lik bir SHAP değeri ile modelin tahminini pozitif yönde domine etmektedir. Diğer kelimelerin katkısı ise çok daha sınırlıdır. Bu analiz, modelin paket türünü belirlerken doğrudan paket adını içeren anahtar kelimelere ne kadar güçlü bir şekilde odaklandığını şeffaf bir biçimde ortaya koymaktadır.



Şekil 4. "Ambalaj Hasarlı" şikâyet türü için pozitif ve negatif SVM katsayılarının karşılaştırması (Comparison of positive and negative SVM coefficients for the "Packaging Damaged" complaint type)



Şekil 5. "Ramazan Paketi" tahmini için SHAP waterfall grafiği (SHAP waterfall plot for the "Ramadan Package" prediction)

3.3. API tabanlı alternatif yaklaşımın değerlendirilmesi (Evaluation of the API-Based Alternative Approach)

Geleneksel ML modellerine bir alternatif olarak, büyük dil modeli (LLM) tabanlı bir API yaklaşımının performansı da 5 adet test e-postası üzerinde değerlendirilmiştir. Elde edilen sonuçlar Tablo 2'de özetlenmiştir.

Tablo 2. LLM tabanlı API yaklaşımının değerlendirme sonuçları (Evaluation results of the LLM-based API approach)

Değerlendirme Kriteri	Başarı Oranı
Paket türü doğruluğu	%100
Sorumlu kişi doğruluğu	%100
Ürün türü doğruluğu	%100
Şikayet türü doğruluğu	%80

API tabanlı sistem, paket, sorumlu kişi ve ürün türü sınıflandırmalarında %100'lük bir başarı oranı elde etmiştir. Ancak, çok etiketli bir yapıya sahip olan şikayet türü sınıflandırmasında, bir örnekteki üç etiketten ikisini doğru tespit ederken birini gözden kaçırmış ve bu kategorideki başarısı %80'de kalmıştır. Bu bulgu, LLM'lerin yüksek anlama potansiyeline sahip olduğunu ancak çok etiketli senaryolarda dikkatli bir şekilde yapılandırılması gerektiğini göstermektedir.

3.4. Tartışma (Discussion)

Bu çalışmada, kurumsal hediye paketleri sonrası gelen şikayet e-postalarının otomatik sınıflandırılması için çok etiketli ve yorumlanabilir bir yapay zekâ sistemi geliştirilmiştir. Elde edilen bulgular, hem akademik literatür hem de pratik uygulama açısından önemli sonuçlar ve çıkarımlar sunmaktadır.

Çalışmanın en dikkat çekici bulgularından biri, her sınıflandırma kategorisi için en iyi performans gösteren makine öğrenmesi modelinin farklılık göstermesidir. Ürün türü gibi kategorik özelliklerin yoğun olduğu ve 68 farklı sınıf içeren karmaşık bir yapıda CatBoost'un en yüksek başarıyı (%89 F1-skoru) göstermesi, bu algoritmanın kategorik verileri işleme konusundaki üstünlüğünü doğrulamaktadır (Prokhorenkova vd., 2018). Buna karşılık, şikâyet türü sınıflandırmasında Destek Vektör Makineleri'nin (SVM) %88 F1-skoru ile öne çıkması, bu algoritmanın TF-IDF ile oluşturulan yüksek boyutlu ve seyrek metin verilerinde etkili karar sınırları bulabilme yeteneğini göstermektedir. Bu bulgu, SVM'nin metin sınıflandırmasındaki gücünü ortaya koyan klasik çalışmalarla tutarlıdır (Cortes & Vapnik, 1995). Paket türü ve sorumlu kişi gibi daha dengeli kategorilerde ise XGBoost'un (sırasıyla %92 ve %82 F1-skoru) hem yüksek doğruluğu hem de göreceli olarak daha hızlı eğitim süresi sunması, bu algoritmanın operasyonel verimlilik ve performans dengesi açısından neden endüstride sıkça tercih edildiğini açıklamaktadır (Chen & Guestrin, 2016).

Veri üretiminde LLM gibi ileri teknolojiler kullanılırken, sınıflandırma aşamasında TF-IDF gibi

geleneksel bir yöntemin tercih edilmesi bilinçli bir metodolojik seçimdir. Kurumsal şikayet yönetimi sistemlerinde hız, düşük hesaplama maliyeti ve açıklanabilirlik kritik öneme sahiptir. BERT veya GPT tabanlı sınıflandırıcılar (embedding-based) yüksek donanım kaynağı gerektirirken ve 'kara kutu' doğaları nedeniyle yorumlanmaları zorlaşırken; TF-IDF ve klasik makine öğrenmesi modelleri (özellikle SVM ve Boosting algoritmaları) şeffaf karar sınırları sunmaktadır. Bu çalışma, karmaşık LLM'lerin yaratıcılığında veri üretiminde faydalanırken, operasyonel süreçte SHAP analizi ile kelime bazlı doğrudan açıklanabilirlik sağlayan ve CPU üzerinde milisaniyeler içinde çalışabilen hafif (lightweight) modellerin yeterli ve etkili olduğunu göstermiştir.

Bu çalışmanın literatüre yaptığı en önemli ve özgün katkı, sınıflandırma modellerinin "kara kutu" yapısını Açıklanabilir Yapay Zekâ (XAI) yöntemleriyle şeffaf hale getirmesidir. Literatürdeki birçok e-posta sınıflandırma çalışması yalnızca tahmin doğruluğuna odaklanırken (Muftaba vd., 2017), bu araştırma model kararlarının "neden" verildiğini de ortaya koymaktadır. Örneğin, SHAP analizi, "Ramazan Paketi" tahmininin büyük ölçüde "ramazan" kelimesinin varlığına dayandığını sayısal olarak kanıtlamıştır (Lundberg & Lee, 2017). Benzer şekilde, SVM katsayıları, "Ambalaj Hasarlı" şikayetinin "yırtılmış" veya "kırılmış" gibi doğrudan hasarı belirten kelimelerle ilişkilendirildiğini göstermiştir. Bu yorumlanabilirlik seviyesi, sistemin sadece doğru tahminler yapmakla kalmadığını, aynı zamanda bu tahminleri insanlar için mantıklı ve doğrulanabilir gerekçelere dayandığını kanıtlar. Bu durum, kurumsal ortamlarda YZ sistemlerine duyulan güveni artırmak ve sistemin benimsenebilirliğini sağlamak için kritik öneme sahiptir (Adadi & Berrada, 2018).

Pratik uygulama açısından, geliştirilen sistem, şikayet yönetimi süreçlerinde önemli bir otomasyon potansiyeli sunmaktadır. Manuel e-posta okuma ve yönlendirme ihtiyacını ortadan kaldırarak yanıt sürelerini kısaltmakta ve satın alma departmanlarının iş yükünü azaltmaktadır. Bu yönüyle çalışma, Robotik Süreç Otomasyonu (RPA) hedefleriyle de uyumludur (Fung, 2014; Willcocks & Lacity, 2016). Geliştirilen bu akıllı sınıflandırma modülü, daha büyük bir RPA iş akışının karar verme mekanizması olarak entegre edilebilir ve şikayet alındığı andan çözüm anına kadar olan tüm sürecin uçtan uca otomasyonunu sağlayabilir.

Çalışmada kullanılan veri seti, şirket çalışanlarının ve müşterilerin gizliliğini korumak (KVKK uyumluluğu) amacıyla Büyük Dil Modelleri kullanılarak sentetik olarak üretilmiştir. Ancak bu üretim süreci rastgele bir simülasyon olmayıp, şirketin geçmiş dönemlerde aldığı gerçek şikayet e-postalarındaki yaygın örüntüler, üslup yapıları ve şikayet şablonları referans alınarak gerçekleştirilmiştir. Bu yaklaşım sayesinde, sentetik verinin gerçek operasyonel koşulları temsil etme yeteneği artırılmış ve literatürdeki 'sentetik veri yanlılığı' (synthetic data bias)

riski minimize edilmiştir. Buna rağmen, LLM tarafından üretilen metinlerin, gerçek insan yazımında sıkça rastlanan düşük seviyeli gürültüleri (örneğin; devrik cümleler, ciddi yazım hataları veya klavye sürçmeleri) tam olarak yansıtmama ihtimali, çalışmanın bir sınırlılığı olarak not edilmelidir. Gelecek çalışmalarda, modelin 'gürültülü' gerçek verilerle ince ayar (fine-tuning) yapılarak test edilmesi önerilmektedir. Ayrıca, hesaplama karmaşıklığı nedeniyle model optimizasyonunda 3 katlı çapraz doğrulama (3-fold cross-validation) ile yetinilmiştir; gelecek çalışmalarda istatistiksel anlamlılığı daha derinlemesine analiz etmek için kat sayısının artırılması ve McNemar testi gibi istatistiksel testlerin uygulanması hedeflenmektedir. Ek olarak, API tabanlı LLM yaklaşımının, özellikle çok etiketli sınıflandırma görevlerinde bazı etiketleri gözden kaçırabilmesi, bu modellerin tek başına kullanımından ziyade, geleneksel ML modellerinin düşük güvenle tahmin yaptığı durumları doğrulamak için hibrit bir yapıda kullanılmasının daha verimli olabileceğini düşündürmektedir.

4. Sonuçlar (Conclusions)

Bu çalışma, kurumsal şikayet yönetimi alanında karşılaşılan çok etiketli sınıflandırma problemi için açıklanabilir yapay zeka (XAI) destekli entegre bir çerçeve sunmuştur. Elde edilen bulgular, TF-IDF tabanlı vektörizasyon ile CatBoost ve XGBoost gibi topluluk öğrenme (ensemble learning) algoritmalarının kombinasyonunun, %87.75'lik ortalama F1-skoru ile yüksek performanslı ve hesaplama açısından verimli bir çözüm sağladığını göstermiştir. Bilimsel açıdan bu sonuçlar, metin sınıflandırma görevlerinde LLM tabanlı karmaşık mimariye kıyasla, geleneksel makine öğrenmesi modellerinin operasyonel hız ve yorumlanabilirlik avantajını koruduğunu doğrulamaktadır.

Çalışmanın metodolojik katkısı, veri gizliliği kısıtları nedeniyle gerçek veriye erişimin sınırlı olduğu durumlarda, gerçek senaryolara dayalı sentetik veri üretiminin geçerli bir alternatif olduğunu ortaya koymasındır. Ancak, sentetik verilerin dilbilgisel açıdan "idealize edilmiş" yapısı, gerçek dünya verilerindeki düzensiz gürültüyü tam olarak yansıtmayabilir. Bu nedenle, önerilen sistemin endüstriyel ölçekte genelleme yeteneğini artırmak için, gelecekteki çalışmalarda gürültülü gerçek verilerle ince ayar (fine-tuning) yapılması kritik bir gereklilik olarak belirlenmiştir. Ayrıca, SHAP analizi ile model kararlarının şeffaflaştırılması, bu tür sistemlerin kurumsal kaynak planlama (ERP) süreçlerine entegrasyonunda güven faktörünü artıran temel bir bileşen olarak öne çıkmaktadır.

Kaynaklar (References)

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.

- Brutlag, J.D., Meek, C., 2000. Challenges of the email domain for text classification. *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 103-110.
- Cai, H., Shao, X., Zhou, P., Li, H. (2025). Multi-Label Classification of Complaint Texts: Civil Aviation Service Quality Case Study. *Electronics*, 14(3), 434.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
- Cohen, W.W., 1996. Learning rules that classify e-mail. *Proceedings of the 1996 AAAI Spring Symposium on Information Access*, p. 25.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., ... & Li, X. (2025). Auggpt: Leveraging chatgpt for text data augmentation. *IEEE Transactions on Big Data*.
- Deniz, E., Erbay, H., & Coşar, M. (2022). Multi-label classification of e-commerce customer reviews via machine learning. *Axioms*, 11(9), 436.
- Dewi, C., Tsai, B. J., & Chen, R. C. (2022, November). Shapley additive explanations for text classification and sentiment analysis of internet movie database. In *Asian Conference on Intelligent Information and Database Systems* (pp. 69-80). Singapore: Springer Nature Singapore.
- Fung, H.P., 2014. Criteria, use cases, and effects of information technology process automation (ITPA). *Advances in Robotics & Automation*, 3(3), 1-11.
- Gopali, S., Abri, F., Namin, A. S., & Jones, K. S. (2024). The Applicability of LLMs in Generating Textual Samples for Analysis of Imbalanced Datasets. *IEEE Access*.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. The MIT Press.
- Kiritchenko, S., & Matwin, S. (2001). Email classification with co-training. *Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research*, p. 8.
- Kokalj, E., Škrlić, B., Lavrač, N., Pollak, S., & Robnik-Šikonja, M. (2021, April). BERT meets shapley: Extending SHAP explanations to transformer-based classifiers. In *Proceedings of the EACL hackashop on news media content analysis and automated report generation* (pp. 16-21).
- Kuyucuk, T., & Çallı, L. (2022). Using multi-label classification methods to analyze complaints against cargo services during the COVID-19 outbreak: Comparing survey-based and word-based labeling. *Sakarya University Journal of Computer and Information Sciences*, 5(3), 371-384.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30, pp. 4765-4774.
- Mitchell, T.M., 1997. *Machine learning*. McGraw-Hill.
- Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D., & Groh, G. (2022, October). SHAP-based explanation methods: a review for NLP interpretability. In *Proceedings of the 29th international conference on computational linguistics* (pp. 4593-4603).
- Mujtaba, G., Shuib, L., Raj, R.G., Majeed, N., Al-Garadi, M.A., 2017. Email Classification Research Trends: A Review and Open Issues. *IEEE Access*, 5, 9044-9064.

- Patwardhan, N., Marrone, S., Sansone, C., 2023. Transformers in the real world: A survey on NLP applications. *Information*, 14(4), 242.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems* 31, pp. 6638–6648.
- Türkmen, A., Bayram, B., & Aydın, G. (2023, July). Employee behavior analysis towards multi-label classification of customer reviews. In *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 511-517).
- ValizadehAslani, T., Shi, Y., Wang, J., Ren, P., Zhang, Y., Hu, M., ... & Liang, H. (2024). Two-stage fine-tuning with ChatGPT data augmentation for learning class-imbalanced data. *Neurocomputing*, 592, 127801.
- Vilone, G., Longo, L., 2021. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 89–106.
- Willcocks, L., Lacity, M., 2016. A new approach to automating services. *MIT Sloan Management Review*, 58(1), 40–49.
- Zhao, W., Joshi, T., Nair, V. N., & Sudjianto, A. (2020). Shap values for explaining cnn-based text classification models. *arXiv preprint arXiv:2008.11825*.