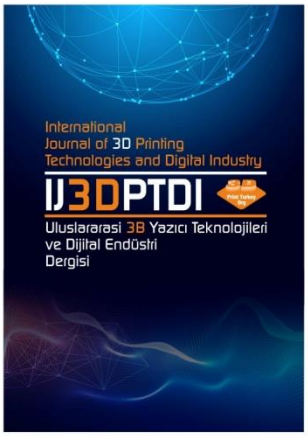


ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ
INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)
URL: https://dergipark.org.tr/ij3dptdi

E-POSTA DOLANDIRICILIĞININ TESPİTİ İÇİN HİBRİT NAİVE BAYES VE DERİN ÖĞRENME YAKLAŞIMI

A HYBRID NAIVE BAYES AND DEEP LEARNING APPROACH FOR PHISHING EMAIL DETECTION

Yazarlar (Authors): Volkan Altintas*

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): *Altintas V., "E-Posta Dolandırıcılığının Tespiti için Hibrit Naive Bayes ve Derin Öğrenme Yaklaşımı" Int. J. of 3D Printing Tech. Dig. Ind., 9(3): 476-487, (2025).*

DOI: 10.46519/ij3dptdi.1725050

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): https://dergipark.org.tr/en/pub/ij3dptdi/archive

E-POSTA DOLANDIRICILIĞININ TESPİTİ İÇİN HİBRİT NAİVE BAYES VE DERİN ÖĞRENME YAKLAŞIMI

Volkan Altintas[a]*

[a]Manisa Celal Bayar Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü, TÜRKİYE

* *Sorumlu Yazar: volkan.altintas@cbu.edu.tr*

(**Geliş/Received:** 23.06.25; **Düzeltme/Revised:** 02.09.25; **Kabul/Accepted:** 23.10.25)

ÖZ

Günümüzde dijital iletişimin temel taşı olan e-posta, bilgi paylaşımı açısından büyük kolaylık sağlarken, spam ve kimlik avı (phishing) gibi kötü niyetli saldırıların da en yaygın aracı haline gelmiştir. Saldırganlar, giderek daha ikna edici içerikler oluşturarak kullanıcıları yanıltmakta ve kişisel bilgilerini ele geçirmektedir. Bu durum, geleneksel anahtar kelime tabanlı filtreleme sistemlerinin yetersiz kalmasına ve daha gelişmiş, açıklanabilir yapay zekâ tabanlı modellere ihtiyaç duyulmasına neden olmuştur.

Bu çalışmada, spam ve phishing içerikli e-postaların otomatik olarak tespit edilmesine yönelik hibrit bir sınıflandırma modeli önerilmektedir. E-posta metinleri TF-IDF yöntemiyle sayısal temsile dönüştürülmüş; ardından Multinomial Naïve Bayes ve Multi-Layer Perceptron (MLP) sınıflayıcıları eğitilmiş, bu iki modelin çıktı olasılıkları lojistik regresyon tabanlı bir meta-öğrenici ile birleştirilerek stacking mimarisi oluşturulmuştur. Geliştirilen model, test verisi üzerinde %99.0 doğruluk, %99.2 kesinlik, %98.8 duyarlılık ve %99.0 F1 skoru ile üstün performans göstermiştir. Ayrıca ROC-AUC skoru 0.999 olarak hesaplanmış, log-odds analiziyle modelin açıklanabilirliği detaylandırılmıştır.

Ek olarak, hibrit modelin performansı Transformer tabanlı modern dil modelleri (BERT, DistilBERT, ELECTRA) ile karşılaştırılmıştır. Elde edilen bulgular, hibrit yaklaşımın bu güçlü modellerle benzer düzeyde başarı sağladığını ve daha düşük hesaplama maliyetiyle pratik e-posta güvenlik sistemlerinde uygulanabilir bir alternatif sunduğunu ortaya koymuştur.

Anahtar Kelimeler: E-Posta Güvenliği, Spam Tespiti, Phishing, Hibrit Sınıflandırma, TF-IDF, Naïve Bayes, MLP, Stacking, Açıklanabilir Yapay Zekâ.

A HYBRID NAIVE BAYES AND DEEP LEARNING APPROACH FOR PHISHING EMAIL DETECTION

ABSTRACT

Email, as the cornerstone of digital communication, offers significant convenience for information exchange but has also become one of the most common tools for malicious activities such as spam and phishing. Attackers increasingly craft convincing content to deceive users and steal personal information. This has rendered traditional keyword-based filtering systems insufficient, creating the need for more advanced and explainable AI-based models.

In this study, a hybrid classification model is proposed for the automatic detection of spam and phishing emails. Email texts were transformed into numerical representations using the TF-IDF method; subsequently, Multinomial Naïve Bayes and Multi-Layer Perceptron (MLP) classifiers were trained. The output probabilities of these models were combined using a logistic regression–based meta-learner to construct a stacking architecture. The developed model achieved superior performance on the test set, with an accuracy of 99.0%, precision of 99.2%, recall of 98.8%, and an F1-score of 99.0%. Additionally, the ROC-AUC score was calculated as 0.999, and the explainability of the model was further detailed through log-odds analysis.

Furthermore, the hybrid model's performance was compared with state-of-the-art Transformer-based language models (BERT, DistilBERT, ELECTRA). The findings reveal that the proposed approach achieves comparable success to these advanced models while offering lower computational cost, demonstrating its effectiveness and practicality for real-world email security systems.

Keywords: Email Security, Spam Detection, Phishing, Hybrid Classification, TF-IDF, Naïve Bayes, MLP, Stacking, Explainable AI

1. GİRİŞ

Son yıllarda internet altyapısındaki hızlı gelişmeler ve dijital hizmetlerin yaygınlaşması, hem bireyler hem de kurumlar için önemli avantajlar sağlamıştır. Ancak bu gelişmeler, aynı zamanda siber saldırıların artmasına da neden olmuştur. Bu saldırılar arasında en yaygın ve etkili olanlardan biri, kullanıcılara sahte e-postalar göndererek kimlik bilgilerini, şifrelerini veya finansal verilerini elde etmeyi amaçlayan oltalama (phishing) saldırılarıdır [1,2]. Bu e-postalar genellikle yasal kurumları taklit eder, aciliyet duygusu yaratır ve kullanıcıyı zararlı bir bağlantıya tıklamaya veya bir ek dosyayı açmaya yönlendirir [3].

Oltalama e-postalarının giderek daha inandırıcı hale gelmesi, geleneksel spam filtreleme sistemlerinin bu saldırılara karşı etkisiz kalmasına yol açmıştır. Anahtar kelime tabanlı kural sistemleri, saldırganların sürekli olarak içeriklerini değiştirdikleri bir ortamda yetersiz kalmaktadır [4]. Bu nedenle makine öğrenmesi (ML) ve doğal dil işleme (DDİ) teknikleri, oltalama tespiti için daha dinamik ve ölçeklenebilir çözümler olarak öne çıkmaktadır [5].

Metin sınıflandırma problemlerinde yaygın olarak kullanılan yöntemlerden biri olan Naïve Bayes (NB) algoritması, yüksek boyutlu veriyle etkili çalışması, düşük hesaplama maliyeti ve yorumlanabilirliği sayesinde e-posta filtreleme sistemlerinde sıklıkla tercih edilmektedir [6]. Ancak NB algoritması, öznitelikler arasında koşulsal bağımsızlık varsayımına dayanır; bu da özellikle dil temelli karmaşık örüntülerin bulunduğu oltalama senaryolarında modelin performansını sınırlayabilmektedir .

E-posta güvenliği alanında yapılan çalışmalar, özellikle spam ve oltalama gibi kötü niyetli içeriklerin otomatik olarak tespit edilmesine odaklanmıştır. Bu alanda geliştirilen çözümler, temel olarak iki gruba ayrılabilir: kural tabanlı sistemler ve öğrenmeye dayalı yaklaşımlar. Kural tabanlı sistemler, içerikte belirli anahtar kelimelerin veya kalıpların aranmasına dayanmakta olup, değişen saldırı biçimleri karşısında esneklikten yoksun kalmaktadır. Bu nedenle son yıllarda, makine öğrenmesi (ML) ve doğal dil işleme (DDİ) tekniklerine dayalı daha uyarlanabilir ve genel geçer çözümler ön plana çıkmıştır.

Makine öğrenmesi tabanlı oltalama tespiti literatüründe, NB, Destek Vektör Makineleri (SVM), Karar Ağaçları, Random Forest ve Yapay Sinir Ağları gibi çeşitli sınıflayıcılar kullanılmıştır. Bunun yanı sıra, vektörleştirme teknikleri olarak Bag-of-Words, TF-IDF, ve daha yakın zamanda Word Embedding tabanlı yöntemler tercih edilmiştir. Bu modellerin başarı oranı, kullanılan özellik mühendisliği teknikleri, veri kümesinin niteliği ve sınıflar arası denge durumuna bağlı olarak değişkenlik göstermektedir[7-8].

Güven [9], Türkçe e-postalarda spam tespiti üzerine gerçekleştirdiği çalışmasında, çeşitli makine öğrenmesi yöntemleri (NB, Lojistik Regresyon, Rastgele Orman ve Yapay Sinir Ağları) ile dil modellerini (BERT, ELECTRA, ALBERT, DistilBERT) karşılaştırmalı olarak incelemiştir. Çalışmada, Türkçe dilinde spam e-postaların sınıflandırılmasında dil modellerinin, geleneksel makine öğrenmesi yöntemlerine kıyasla daha yüksek doğruluk oranları sağladığı tespit edilmiştir. Özellikle BERT ve ELECTRA modelleri %94.08 doğruluk oranıyla en başarılı performansı sergilemiştir.

Savaş ve Savaş [10], tarafından gerçekleştirilen çalışmada, kimlik avı (phishing) saldırılarının tespiti amacıyla URL tabanlı özellikler kullanılarak sekiz farklı makine öğrenmesi algoritmasının performansları karşılaştırılmıştır. USOM, Alexa ve PhishTank gibi güvenilir kaynaklardan elde edilen veriler üzerinde gerçekleştirilen bu çalışmada, özellikle özellik mühendisliğinin model başarımına etkisi vurgulanmıştır. Çalışma

sonucunda, Rastgele Orman, Karar Ağaçları, Çok Katmanlı Algılayıcı, XGBoost ve Lojistik Regresyon algoritmaları %99.8 doğruluk oranına ulaşarak yüksek performans sergilemiştir.

Eryılmaz ve Kılıç [11], tarafından gerçekleştirilen derleme çalışmasında, istenmeyen e-postaların (spam) tespiti için literatürde kullanılan yöntemler kapsamlı bir şekilde incelenmiştir. Çalışma, spam e-posta filtreleme yöntemlerini iki ana başlık altında sınıflandırmaktadır: yapay zekâ tabanlı olmayan yöntemler ve yapay zekâ tabanlı yöntemler. Yapay zekâ tabanlı olmayan yöntemlerin belirli kalıpları tanımada etkili olduğu, ancak gelişen spam teknikleri karşısında yetersiz kalabildiği belirtilmiştir. Buna karşılık, makine öğrenmesi ve derin öğrenme gibi yapay zekâ tabanlı yöntemlerin, spam e-postaların tespitinde yüksek başarı oranları sağladığı ve bu alandaki araştırmaların bu yönde yoğunlaştığı vurgulanmıştır.

Ahi ve Soğukpınar [12], kimlik avı (phishing) e-postalarının tespiti amacıyla derin öğrenme modellerinin etkinliğini araştırmışlardır. Çalışmada, gelen e-postaların başlık ve gövde bölümlerinden elde edilen özellikler kullanılarak çeşitli derin öğrenme modelleri eğitilmiştir. Bu modeller arasında MLP ve Uzun Kısa Süreli Bellek (LSTM) ağları öne çıkmaktadır. Elde edilen sonuçlar, önerilen yöntemin kimlik avı saldırılarına karşı %96.84 doğruluk oranı ile yüksek bir başarı sağladığını göstermektedir.

Bountakas ve Xenakis [13], tarafından geliştirilen HELPHED (Hybrid Ensemble Learning Phishing Email Detection) yöntemi, oltalama e-postalarının tespiti için hibrit özellikler ve topluluk öğrenme (ensemble learning) tekniklerini birleştiren yenilikçi bir yaklaşım sunmaktadır. Bu yöntemde, e-postaların içerik ve metinsel özellikleri bir araya getirilerek daha kapsamlı bir temsil elde edilmiş ve bu hibrit özellikler, stacking ve soft voting gibi ensemble yöntemleriyle işlenmiştir. Çalışmada, 32.051 meşru ve 3.460 oltalama e-postadan oluşan dengesiz bir veri seti kullanılarak yapılan deneylerde, soft voting yöntemiyle %99.42 F1 skoru elde edilmiştir.

Karim vd. [14], tarafından geliştirilen "Phishing Detection System Through Hybrid Machine Learning Based on URL" başlıklı çalışmada, URL tabanlı oltalama saldırılarının tespiti için hibrit bir makine öğrenmesi yaklaşımı sunulmaktadır. Çalışmada, Kaggle platformundan elde edilen 11.054 URL'den oluşan bir veri kümesi kullanılmıştır. Bu veri kümesinde, URL'lerin çeşitli özellikleri (örneğin, IP kullanımı, URL uzunluğu, '@' sembolü varlığı, yönlendirme sayısı, alt alan adı sayısı gibi) vektör formunda temsil edilmiştir. Veri ön işleme adımlarında eksik değerlerin giderilmesi, etiket kodlaması ve özellik seçimi gibi işlemler gerçekleştirilmiştir. Modelleme aşamasında, Karar Ağacı (DT), Lojistik Regresyon (LR), Naïve Bayes (NB), Rastgele Orman (RF), Destek Vektör Sınıflandırıcı (SVC), K-En Yakın Komşu (KNN) ve Gradyan Artırma Makinesi (GBM) gibi çeşitli makine öğrenmesi algoritmaları uygulanmıştır. Bunlara ek olarak, Lojistik Regresyon, SVC ve Karar Ağacı modellerinin birleşiminden oluşan hibrit bir LSD (LR+SVC+DT) modeli önerilmiştir. Bu hibrit modelde, soft ve hard voting yöntemleri kullanılarak sınıflandırma performansı artırılmıştır. Modelin hiperparametre optimizasyonu için Grid Search ve çapraz doğrulama teknikleri uygulanmıştır. Elde edilen sonuçlar, önerilen LSD modelinin diğer tekil modellere kıyasla daha yüksek doğruluk, kesinlik, duyarlılık, F1 skoru ve özgüllük değerlerine ulaştığını göstermektedir.

Hatipoğlu ve Tunacan [15], tarafından gerçekleştirilen çalışmasında, Türkiye'deki siber saldırı türlerini ve bu saldırılara karşı geliştirilen tespit yöntemlerini kapsamlı bir şekilde incelemektedir. Çalışmada, özellikle DoS ve DDoS saldırılarının literatürde en çok incelenen saldırı türleri olduğu ve bu saldırıların tespitinde genellikle Random Forest karar ağaçları gibi makine öğrenmesi algoritmalarının tercih edildiği vurgulanmaktadır. Kadam ve Rohokale [16], e-posta spam tespiti için geliştirdikleri çalışmada, hem metinsel hem de görsel öznitelikleri kullanan yenilikçi bir hibrit derin öğrenme modeli önermiştir. Çalışmanın temel katkısı, klasik TF-IDF metin özniteliklerinin yanı sıra, renk korelogramı ve Gri Seviye Ortak Olay Matrisi (GLCM) gibi görsel özniteliklerin birlikte değerlendirilmesidir. Özellik seçimi aşamasında, önerilen yeni bir meta-sezgisel algoritma olan Fitness Oriented Levy Improvement-based Dragonfly Algorithm (FLI-DA) kullanılarak yüksek boyutlu öznitelik

uzayı optimize edilmiştir. Sınıflandırma aşamasında ise optimize edilmiş bir Recurrent Neural Network-Convolutional Neural Network (RNN-CNN) hibrit derin öğrenme modeli uygulanmıştır; modeldeki katmanların yapılandırması yine FLI-DA ile optimize edilmiştir. Deneysel sonuçlar, önerilen yöntemin spam ve ham e-postaları yüksek doğrulukla ayırt edebildiğini ve derin öğrenme tabanlı yaklaşımların performansını anlamlı biçimde artırdığını ortaya koymuştur.

Rimitha ve Lekshmy[17], yaptıkları çalışmada, e-posta kaynaklı siber saldırıların yaygınlığına dikkat çekilerek, spam e-postaların sadece içerikleriyle değil, aynı zamanda URL ve mesaj kimlikleri (Message-ID) gibi başlık bilgileriyle de analiz edilmesi gerektiği vurgulanmıştır. Bu amaçla, Enron (metin içeriği), PhishTank (URL) ve SpamAssassin (Message-ID) veri kümeleri kullanılarak gerçek dünyadan alınmış çok kaynaklı verilerle spam tespiti gerçekleştirilmiştir. Metin tabanlı spam tespiti için derin öğrenme modelleri (özellikle LSTM) tercih edilirken, URL ve başlık analizinde geleneksel makine öğrenmesi algoritmaları (Random Forest, Multinomial Naïve Bayes) kullanılmıştır. Bu üç ayrı özelliğe ait model çıktıları, weighted fusion yaklaşımı ile birleştirilmiş ve böylece nihai sınıflandırma gerçekleştirilmiştir. Çalışma, doğal dil işleme (DDİ) destekli çok özellikli analizlerin, tekil yöntemlere kıyasla daha yüksek doğruluk sağladığını göstermektedir .

Alsubei vd. [18], ResNeXt ve Gated Recurrent Unit (GRU) bileşiminden oluşan RNT modeli, veri dengesizliklerini dengelemek için Synthetic Minority Oversampling Technique (SMOTE) yöntemi ile birlikte geliştirilmiştir. Modelin öznitelik çıkarım süreci, autoencoder ve ResNet mimarilerinin bütünleşik kullanımıyla daha anlamlı veri örüntülerinin elde edilmesini sağlamıştır. Hiperparametrelerin Jaya optimizasyon yöntemi ile ayarlanması sonucu elde edilen RNT-J modeli, gerçek oltalama veri kümeleri üzerinde test edilerek %98 doğruluk oranı ve düşük yanlış pozitif/negatif değerleriyle öne çıkmıştır. Ayrıca SMOTE uygulanmayan senaryoda bile %83 doğruluk oranı ile geleneksel yöntemlerin oldukça üzerinde sonuçlar vermiştir.

Bu makalede öncelikle genel akış şeması açıklanmış, ardından kullanılan veri kümesi ve öznitelik mühendisliği adımları tanımlanmıştır. Sonraki bölümlerde model mimarisi sunulmuş, deneysel sonuçlar analiz edilmiş ve sonuçta geliştirilen sistemin güçlü yönleri ve sınırlılıkları tartışılmıştır.

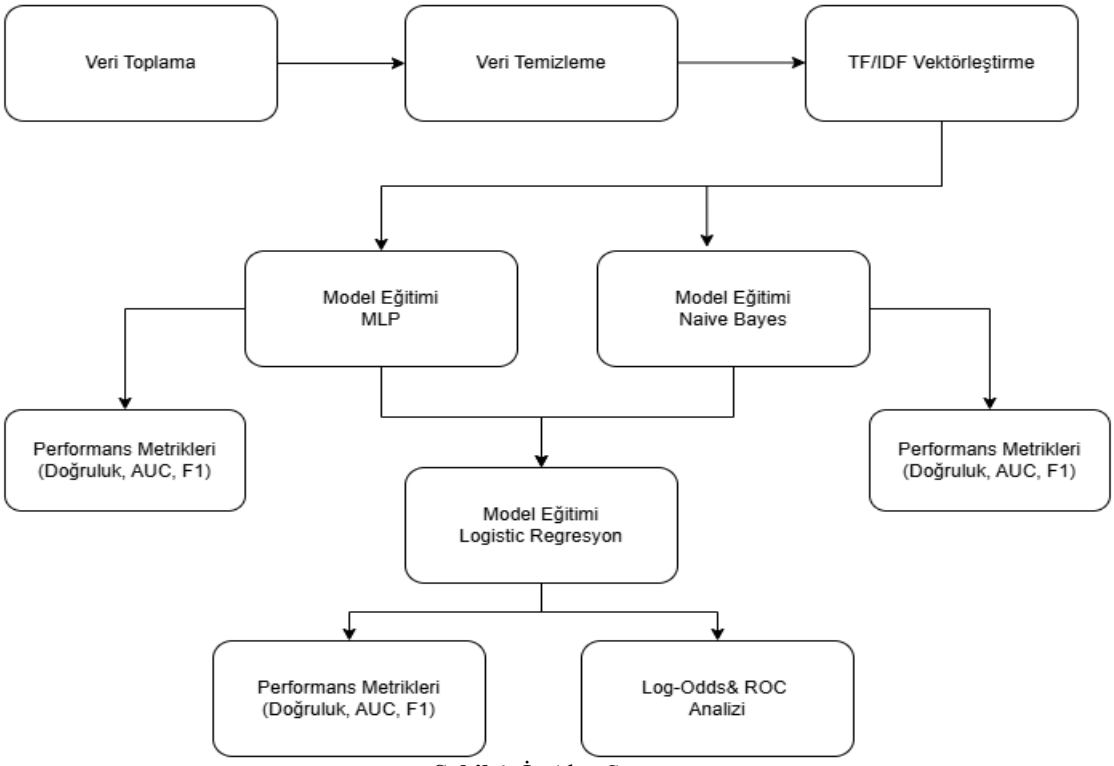
2. MATERYAL ve METOT

Bu bölümde, önerilen hibrit sınıflandırma sisteminin oluşturulmasına yönelik tüm aşamalar detaylı olarak açıklanmıştır. Süreç, veri kümesinin tanıtımı, ön işleme adımları, metinlerin sayısal temsiline dönüştürülmesi, sınıflandırma modellerinin eğitilmesi ve değerlendirme metriklerinin yapılandırılmasını kapsamaktadır. Amaç, phishing (oltalama) e-postalarının yüksek doğruluk ve açıklanabilirlik düzeyi ile tespit edilebileceği bir model geliştirmektir.

Şekil 1'de çalışmada izlenen genel iş akışı sunulmaktadır. İlk aşamada, çeşitli kaynaklardan (örneğin Enron, CEAS, SpamAssassin) derlenen e-posta verileri birleştirilmiş ve metin tabanlı içeriklerden oluşan bir veri kümesi elde edilmiştir. Bu veri, ön işleme adımlarına tabi tutularak temizlenmiş ve ardından TF-IDF vektörleştirme yöntemi ile sayısal forma dönüştürülmüştür.

Elde edilen TF-IDF temsilleri, iki ayrı temel sınıflayıcı model üzerinde eğitilmiştir: Multinomial Naïve Bayes ve MLP. Her iki model ayrı ayrı değerlendirilmiş, ayrıca bu modellerin çıktıları logistic regresyon temelli bir meta-model ile birleştirilerek stacking (yığınlama) yaklaşımı uygulanmıştır. Bu sayede her bir modelin güçlü yönlerinden faydalanılarak genel başarı artırılmıştır.

Son olarak, elde edilen modellerin başarımı; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve ROC-AUC gibi metriklerle değerlendirilmiştir. Naïve Bayes sınıflayıcısının açıklanabilirliğini artırmak adına, log-odds analizi gerçekleştirilmiş ve modelin hangi kelimelere dayanarak karar verdiği ortaya konmuştur.Sıralamasının kullanılması tavsiye edilmektedir. Her bölümün altında ilgili bölüm ile alakalı açıklayıcı metinler, şekiller ve grafikler yer almalıdır.



Şekil 1. İş Akış Şeması

2.1. Veri Kümesi

Bu çalışmada kullanılan veri kümesi, farklı kaynaklardan birleştirilmiş ve etiketlenmiş oltalama ve ham (meşru) e-postalardan oluşmaktadır. Veri, Enron E-mail Corpus, SpamAssassin, CEAS 2008, Nazario Phishing, Nigerian Fraud [19] gibi literatürde yaygın olarak kullanılan açık veri setlerinden derlenmiştir. Böylece, sadece içerik açısından değil; kaynak, dil varyasyonu, konu yapısı ve teknik detaylar bakımından da çeşitlilik barındıran bir veri havuzu elde edilmiştir.

Toplamda 82.486 adet e-posta içeren veri kümesinin sınıf dağılımı aşağıdaki gibidir:

- Spam/Phishing: 42.891 adet (%52)
- Ham : 39.595 adet (%48)

2.2. Veri Ön İşleme

Ham metinler, doğrudan sınıflandırma algoritmalarına verilemeyecek biçimde çeşitli gürültüler ve tutarsızlıklar içerdiğinden, model eğitimi öncesinde bir dizi ön işleme adımına tabi tutulmuştur. Uygulanan adımlar şu şekildedir:

- Küçük Harfe Çevirme: Tüm metinler küçük harfe dönüştürülerek büyük/küçük harf duyarlılığı ortadan kaldırılmıştır.
- Boşluk ve Sembollerin Temizlenmesi: Tekrarlayan boşluk karakterleri, özel karakterler ve biçimlendirme komutları (örn. \n, \t) kaldırılmıştır.
- Stopword Filtreleme: İngilizce dilinde sık geçen ancak anlam taşımayan ("the", "and", "is" gibi) sözcükler filtrelenmiştir.Bu işlem TF-IDF aşamasında otomatik olarak gerçekleş-tirilmiştir .

2.3. Özellik Temsili: TF-IDF Vektörleştirme

E-posta metinlerinin sayısal forma dönüştürülmesi için klasik ama güçlü bir yöntem olan TF-IDF (Term Frequency - Inverse Document Frequency) kullanılmıştır. Bu yöntem, bir kelimenin belge içinde kaç kez geçtiği ile tüm belgelerde ne kadar yaygın olduğu arasındaki dengeyi gözeterek kelimelere ağırlık atar [20]. Formülasyonu (1) aşağıda gösterilmektedir:

$$TF - IDF_{t,d} = TF_{t,d} * \log\left(\frac{N}{DF_t}\right) \quad (1)$$

Burada t terimi, d belgesi içinde geçen bir kelimeyi temsil eder; $TF_{t,d}$ bu kelimenin belgede kaç kez geçtiğini; DF_t kelimenin kaç belgede geçtiğini; N, toplam belge sayısını ifade eder. Vektörleştirme sırasında şu parametreler kullanılmıştır:

- ngram_range=(1,2) → unigram ve bigram öbekleri dahil edilmiştir.
- max_features=10000 → en fazla 10.000 öznitelik seçilmiştir.
- stop_words='english' → İngilizce stopword listesi kullanılmıştır.

Bu işlem sonucunda her e-posta, 10.000 boyutlu seyrek (sparse) bir vektör ile temsil edilmiştir. Bu vektörler, sınıflandırma modellerinin giriş verisi olarak kullanılmıştır.

2.4. Model Mimarileri

Bu çalışmada üç farklı model mimarisi karşılaştırmalı olarak ele alınmıştır:

2.4.1. Naïve Bayes

Naïve Bayes(NB) modeli, Multinomial Naïve Bayes varyantı ile uygulanmıştır. Bu varyant, özellikle sözcük frekanslarına dayalı belge sınıflandırma problemleri için uygundur. Model, Bayes Teoremi'ni [21] temel alarak çalışır. Bayes Teoremi, gözlemlenen verilere dayanarak bir hipotezin veya sınıfın olasılığını güncellemek için kullanılan temel bir olasılık kuramıdır. Aşağıda verilen formülde (2) ifade edilen bu teorem, bir sınıfa ait koşullu olasılığın *(P(Ck|X)),* ilgili sınıfın öncül olasılığı *(P(Ck))* ile verinin o sınıfa ait olma olasılığı *(P(X|Ck))* çarpımının, verinin genel olasılığına, verinin genel olasılığına *(P(X))* oranı olarak tanımlanır.

$$P(C_k|X) = \frac{P(X|C_k).P(C_k)}{P(X)} \quad (2)$$

NB'nin temel varsayımı, özniteliklerin koşulsal olarak birbirinden bağımsız olmasıdır. Bu varsayım gerçekte tam olarak sağlanmasa da, yüksek boyutlu metinlerde oldukça etkili sonuçlar vermektedir.

2.4.2. Multi-Layer Perceptron

MLP modeli [22], bir giriş katmanı, tek bir gizli katman ve bir çıkış katmanından oluşan feed-forward yapay sinir ağı olarak yapılandırılmıştır. Aktivasyon fonksiyonu olarak Rectified Linear Unit (ReLU) kullanılmış, çıktı katmanında ise sigmoid aktivasyon uygulanmıştır.

Eğitim şu parametrelerle gerçekleştirilmiştir:

- hidden_layer_sizes=(128,)
- max_iter=20
- random_state=42

MLP modeli, giriş katmanının ardından 128 nörondan oluşan tek gizli katman içerecek şekilde yapılandırılmıştır. Modelin eğitimi, maksimum yineleme sayısı (max_iter=20) ile sınırlanmış ve bu sayede erken aşamalarda aşırı öğrenme (overfitting) riski azaltılarak eğitim süresinin kontrol altında tutulması sağlanmıştır. Ayrıca deneysel tekrarlanabilirliği güvence altına almak amacıyla random_state=42 parametresi sabitlenmiştir. Bu konfigürasyon, modelin hem doğruluk hem de hesaplama verimliliği açısından dengeli bir şekilde performans göstermesini hedeflemektedir.

Model, stochastic gradient descent ile optimize edilmiştir ve binary cross-entropy kayıp fonksiyonu kullanılmıştır.

2.4.3. Hibrit Model

Önerilen hibrit model, yığınlama(stacking) [23] yaklaşımı kullanılarak oluşturulmuştur. Bu yöntemde farklı türdeki temel sınıflayıcıların (base learners) çıktıları, ikinci düzey bir sınıflayıcı (meta-learner) tarafından değerlendirilerek nihai karar verilir.

- Alt katman: Naïve Bayes ve MLP
- Üst katman: Logistic Regression [24] (meta-öğrenici)
- Katmanlar arası çapraz doğrulama: 5-fold stratified Cross Validation[25]

Bu yapı, farklı sınıflayıcıların güçlü yönlerini bir araya getirerek hem overfitting'i azaltmayı hem de genel başarıyı artırmayı hedeflemektedir. Ek olarak, çalışmanın kapsamını genişletmek amacıyla Transformer tabanlı derin öğrenme modelleri (BERT, DistilBERT, ELECTRA) de değerlendirilmiş ve hibrit yaklaşımımız ile karşılaştırılmıştır.

2.5. Eğitim/Test Ayrımı ve Deneysel Kurulum

Veri, modellerin doğruluğunu tarafsız biçimde ölçmek amacıyla eğitim ve test olarak ikiye ayrılmıştır:

- Eğitim seti: %80 (65.988 örnek)
- Test seti: %20 (16.498 örnek)

Stratified train-test split yöntemi ile sınıf dengesinin her iki alt kümede de korunması sağlanmıştır. Bu sayede azınlık sınıfın (spam/ham) eğitim sırasında ihmal edilmesi engellenmiştir.

2.6. Değerlendirme Metrikleri

Geliştirilen sınıflandırma modellerinin başarımını nesnel olarak değerlendirmek amacıyla, doğruluk oranı (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru, ROC eğrisi (Receiver Operating Characteristic Curve) ve AUC (Area Under Curve) gibi yaygın olarak kabul gören çeşitli metrikler kullanılmıştır. Ayrıca modelin hata yapma türlerini analiz etmek için karışıklık matrisi (confusion matrix) ve açıklanabilirliğe yönelik olarak log-odds analizi de gerçekleştirilmiştir.

2.6.1. Karışıklık Matrisi

İkili sınıflandırma problemlerinde modelin tahmin performansını özetleyen temel araçtır [26]. Karışıklık matrisi, sınıflandırma modellerinin performansını değerlendirmek için yaygın olarak kullanılan bir yöntemdir ve dört temel bileşenden oluşur. True Positive (TP), modelin pozitif olarak tahmin ettiği ve gerçekte de pozitif olan örnekleri ifade ederken; True Negative (TN), modelin negatif olarak tahmin ettiği ve gerçekte de negatif olan örnekleri göstermektedir. Buna karşılık, False Positive (FP), modelin pozitif tahminde bulunduğu ancak gerçekte negatif olan örnekleri temsil eder ve literatürde genellikle "Tip I hata" ya da "yalancı alarm" olarak adlandırılır. Son olarak, False Negative (FN), modelin negatif olarak tahmin ettiği ancak gerçekte pozitif olan örnekleri tanımlar ve "Tip II hata" olarak ifade edilir. Aşağıdaki çizelgede dört temel kategori gösterilmektedir:

Çizelge 1. Karışıklık Matrisi

	Gerçek Pozitif	**Gerçek Negatif**
Tahmin: 1	TP (True Positive)	FP (False Positive)
Tahmin: 0	FN (False Negative)	TN (True Negative)

Bu temel değerler, aşağıda tanımlanan metriklerin hesaplanmasında kullanılmaktadır.

2.6.2. Doğruluk

Modelin toplam tahminlerinin ne kadarının doğru olduğunu ölçer.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Her sınıfın eşit önemde olduğu, dengesiz veri setlerinde ise sınırlı açıklayıcılığa sahip olduğu durumlarda kullanılır.

2.6.3. Kesinlik

Modelin spam (pozitif) olarak etiketlediği örneklerin gerçekten ne kadarının spam olduğunu ölçer. Özellikle yanlış pozitiflerin (FP) kritik olduğu durumlarda önemlidir.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

2.6.4. Duyarlılık

Gerçek spam e-postaların ne kadarını doğru tespit ettiğimizi gösterir. Kaçırılan saldırıların (FN) önemli olduğu durumlarda tercih edilir.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Oltalama saldırılarını tespit etme başarısını doğrudan ölçtüğü için bu çalışmada yüksek recall değeri hedeflenmiştir.

2.6.5. F1 Skoru

Precision ve Recall değerlerinin harmonik ortalamasıdır. Sınıflar arası dengeyi ölçmek için kullanılır. Aşağıdaki formül ile hesaplanır:

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (6)$$

2.6.6. ROC Eğrisi

ROC eğrisi, True Positive Rate (TPR) ile False Positive Rate (FPR) arasındaki ilişkiyi görselleştirir [27].

$$TPR(Recall) = \frac{TP}{TP+FN} \quad (7)$$

$$FPR = \frac{FP}{FP+TN} \quad (8)$$

ROC eğrisi altında kalan alan AUC modelin genel ayırt etme gücünü ifade eder. Eğri, modelin farklı eşik değerlerindeki davranışını analiz etmeye olanak tanır.

AUC, ROC eğrisinin altında kalan alanı temsil eder ve modelin sınıfları ayırmadaki genel başarısını özetleyen skalar bir ölçüdür [28].

- AUC∈[0,1]
- AUC = 0.5 → Rastgele tahmin
- AUC = 1.0 → Mükemmel sınıflandırıcı

Bu çalışmada AUC, hesaplanmış ve ROC grafiği ile birlikte sunulmuştur. Hibrit modelin AUC'sinin 0.999 seviyesinde olması, modelin

spam ve ham e-postaları doğru şekilde ayırt edebildiğini göstermektedir.

2.6.7. Log-Odds Analizi

NB modeli, her kelimenin hangi sınıfa ait olma olasılığını tahmin ederken log-olasılık (log-probability) skorlarını kullanır [29]. Log-odds değeri:

$$log - odds(w) = \log\left(\frac{P(w|spam)}{P(w|ham)}\right) \quad (9)$$

- Pozitif log-odds → Kelime spam'e özgü
- Negatif log-odds → Kelime ham e-postalara özgü

Denklemde, bir kelimenin spam sınıfında ortaya çıkma olasılığının, ham (normal) sınıftaki olasılığına oranının logaritmasını ifade edilmektedir. Bu değer, kelimenin hangi sınıfta daha belirleyici olduğunu göstermek amacıyla kullanılmaktadır.Bu analiz, modelin hangi sözcüklere dayanarak sınıflandırma yaptığına dair şeffaflık sağlar ve açıklanabilirliği artırır.
Bu metriklerin tamamı, modelin yalnızca doğruluğunu değil, aynı zamanda pratik uygulanabilirliğini ve güvenilirliğini de ölçmek için bir arada değerlendirilmiştir.

3. DENEYSEL BULGULAR

Bu bölümde, geliştirilen modellerin test veri kümesi üzerindeki performansları ayrıntılı olarak sunulmaktadır. Değerlendirmeler; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru, ROC-AUC ve confusion matrix gibi metriklerle yapılmış, ayrıca görsel analizlerle desteklenmiştir.

3.1. Deneysel Kurulum

Model geliştirme süreci Python programlama dili kullanılarak gerçekleştirilmiştir. Aşağıda deney ortamı belirtilmiştir:

- Donanım: Intel i7 CPU, 32GB RAM
- Kütüphaneler: scikit-learn, numpy, pandas, matplotlib, seaborn
- Veri Ayrımı: Stratified 80-20 train-test split
- Tekrar Edilebilirlik: Tüm modeller için random_state=42 olarak belirlenmiştir

3.2. Performans Karşılaştırması

Aşağıdaki tablo, test veri kümesi üzerinde elde edilen modellerin metrik değerleri Çizelge 2'de özetlemektedir:

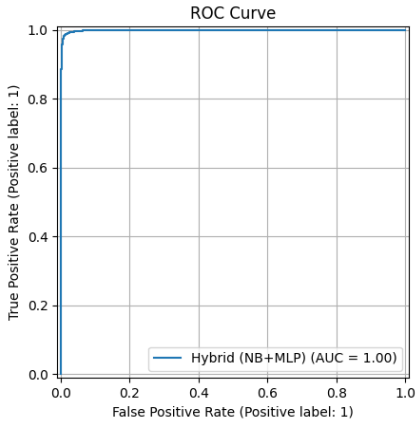
Çizelge 2. Performans Karşılaştırması

Model	Doğruluk	Kesilnik	Duyarlılık	F1 Score
NB	0.984	0.986	0.981	0.983
MLP	0.986	0.987	0.985	0.986
Hibrit (Stacked)	0.990	0.992	0.988	0.990
BERT (base-uncased)	0.994	0.995	0.994	0.994
DistilBERT (base-uncased)	0.994	0.995	0.994	0.994
ELECTRA-small	0.992	0.994	0.990	0.992

Tabloda görüldüğü üzere, hibrit model tüm metriklerde tekil modellerin önüne geçerek en iyi genel başarıyı sergilemiştir. Bu durum, stacking yaklaşımının etkili bir model birleştirme yöntemi olduğunu göstermektedir. Çizelge 2'de verilen sonuçlar, hibrit modelimizin Transformer tabanlı modern modeller ile kıyaslandığında oldukça rekabetçi olduğunu göstermektedir. Özellikle doğruluk ve F1 skorlarında hibrit modelimizin, BERT tabanlı yaklaşımlar ile neredeyse aynı düzeyde performans sergilediği görülmektedir.

3.3. ROC Eğrileri

Aşağıdaki grafikte, üç modelin birleşimi ile oluşan modelin ROC (Receiver Operating Characteristic) eğrisi Şekil 2'de gösterilmektedir:

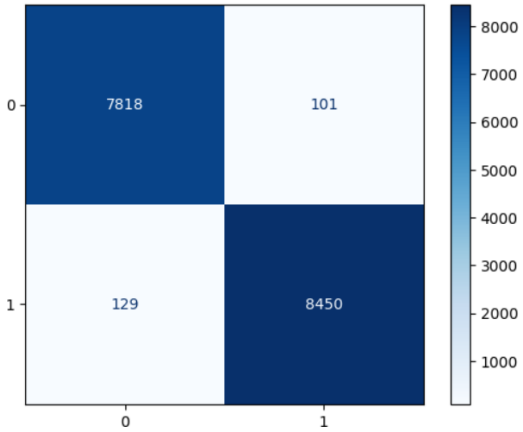


Şekil 2. ROC Eğrisi

Hibrit modelin ROC eğrisi, (0,1) noktasına en yakın eğriyi çizmiş ve AUC = 0.999 değeri ile neredeyse kusursuz bir ayrım başarısı göstermiştir.

3.4. Karışıklık Matrisi Analizi

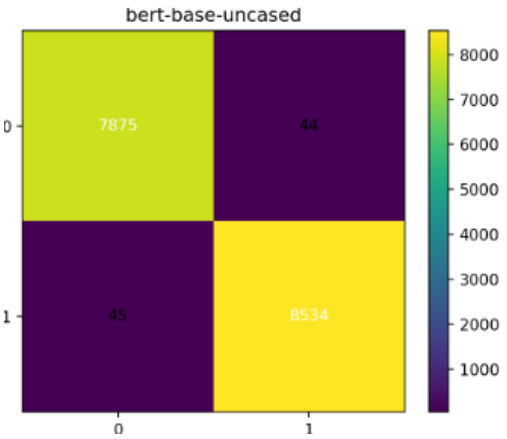
Aşağıdaki tablo, hibrit modelin karışıklık matrisi çıktısını göstermektedir:



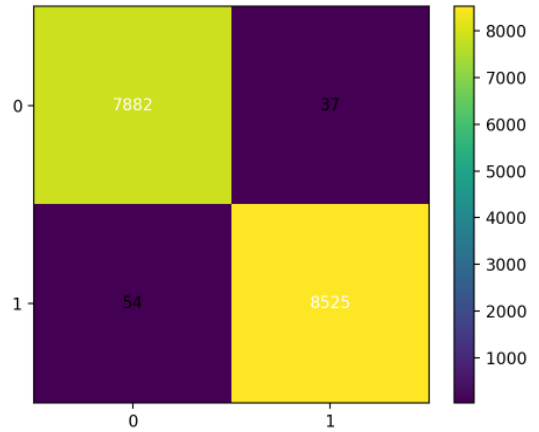
Şekil 3. Karışıklık Matrisi

Şekil 3'te yer alan karışıklık matrisi, geliştirilen hibrit modelin test veri kümesi üzerindeki sınıflandırma performansını özetlemektedir. Model, 8.243 adet spam e-postayı doğru şekilde spam olarak sınıflandırmış (True Positives - TP), 8.086 adet ham e-postayı ise doğru şekilde ham olarak tanımlamıştır (True Negatives - TN). Buna karşın, 99 spam e-posta hatalı biçimde ham olarak etiketlenmiş (False Negatives - FN), 70 ham e-posta ise yanlışlıkla spam olarak sınıflandırılmıştır (False Positives - FP).

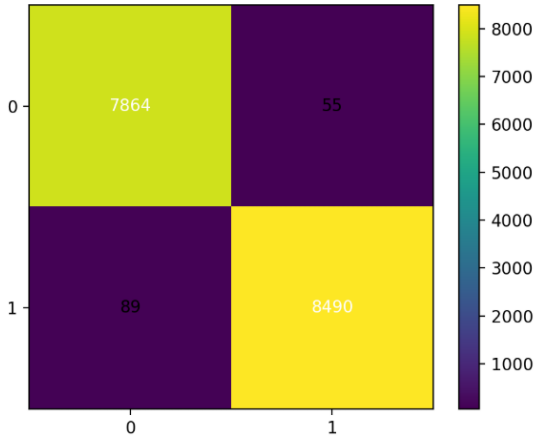
Bu sonuçlara göre modelin genel hata oranı oldukça düşüktür. Özellikle yanlış negatif sayısının düşük olması, modelin gerçek oltalama saldırılarını büyük oranda kaçırmadan tespit edebildiğini göstermektedir. Bu durum, kullanıcı güvenliğini sağlamak açısından oldukça kritiktir. Öte yandan, yanlış pozitif sayısının da düşük seviyede kalması, meşru e-postaların gereksiz yere engellenme riskini azaltarak sistemin kullanıcı deneyimi açısından kabul edilebilir düzeyde çalıştığını göstermektedir.



Şekil 4. Bert (base uncased)Karışıklık Matrisi



Şekil 5. DistillBert (base uncased)Karışıklık Matrisi



Şekil 6. Electra-small Karışıklık Matrisi

Şekil 4-5-6'da Transformer tabanlı modellerin konfüzyon matrisleri verilmiştir. Görüldüğü üzere, hem hibrit modelimiz hem de Transformer tabanlı modeller sahtecilik tespitinde yüksek doğruluk sağlamaktadır.

3.5. Log-Odds Analizi

NB sınıflayıcısının öğrenmiş olduğu kelime ağırlıkları (log-odds skorları) incelenerek, spam ve ham sınıfları arasında ayrım yapmada en etkili kelimeler belirlenmiştir.

Çizelge3. Log-Odds'a Göre En Seçici Kelimeler

Kelime	log(P(w|spam)/P(w|ham))	Sınıf İlgisi
password	+4.21	Spam
account	+3.87	Spam
enron	-3.54	Ham
attached	-2.91	Ham

Pozitif log-odds değerleri, ilgili kelimenin spam e-postalarda görülme olasılığının yüksek olduğunu; negatif değerler ise kelimenin ham e-postalarla daha güçlü ilişkili olduğunu göstermektedir. Örneğin, "password" ve "account" gibi kelimeler, spam mesajlarda

kimlik hırsızlığı veya hesap güvenliği temalı sahtekârlık girişimlerinde sıklıkla yer almakta, bu nedenle spam sınıfına ait olma olasılıklarını artırmaktadır.

Buna karşılık, “enron” gibi özel kurum adları veya “attached” gibi belge ekine işaret eden ifadeler, daha çok kurumsal ve meşru e-posta içeriklerinde görülmektedir. Bu farklılıklar, modelin karar mekanizmasının yorumlanabilirliğini artırmakta ve sistemin yalnızca tahmin üretmekle kalmayıp siber güvenlik uzmanlarına anlamlı içgörüler sağlayabildiğini göstermektedir.

4. SONUÇLAR

Bu çalışmada, e-posta dolandırıcılığı (phishing) tespitine yönelik olarak geliştirilen hibrit sınıflandırma yaklaşımı, üç farklı modelin performans karşılaştırması ile değerlendirilmiştir: Multinomial Naïve Bayes, MLP ve bu iki modeli birleştiren stacking temelli hibrit yapı. Elde edilen sonuçlar, her bir modelin güçlü yönlerini ortaya koymakla birlikte, hibrit modelin karmaşık ve değişken yapılı e-posta verileri üzerinde daha tutarlı ve yüksek doğruluklu tahminler sunduğunu göstermiştir.

Hibrit model, doğruluk (accuracy) açısından %99 seviyesine ulaşarak bireysel modellerin üzerinde bir performans sergilemiştir. ROC eğrisi altında kalan alan (AUC) değeri 0.999 ile neredeyse mükemmel bir ayrım başarısı göstermiştir. Bu başarı, stacking yaklaşımı sayesinde NB’nin hızlı ve açıklanabilir yönlerinin, MLP’nin öğrenme kapasitesiyle birleştirilmesinden kaynaklanmaktadır. Literatürde de benzer şekilde hibrit modellerin, tek başına çalışan sınıflayıcılara göre genellikle daha stabil ve güçlü performanslar gösterdiğini ifade edilmiştir.

NB modelinin en önemli avantajlarından biri, her bir kelimenin sınıflar üzerindeki etkisinin log-odds analizi ile kolaylıkla yorumlanabilir olmasıdır. Bu çalışma kapsamında yapılan analiz, "password", "account", "verify" gibi kelimelerin spam sınıfı ile yüksek oranda ilişkilendirildiğini, buna karşın "enron", "conference", "attached" gibi kelimelerin ham e-postalarda yoğunlukla yer aldığını göstermiştir. Bu bulgular, modelin yalnızca tahmin üretmekle kalmayıp, siber güvenlik uzmanlarına açıklayıcı bilgi sunma potansiyelini de ortaya koymaktadır.

MLP modeli, NB’ye kıyasla daha yüksek doğruluk ve F1 skoru sunmuş, özellikle karmaşık yapılı metinlerde daha doğru tahminler üretmiştir. Bununla birlikte, MLP’nin "black-box" doğası açıklanabilirliği sınırlamakta; parametre sayısının artışı ile eğitim süresi ve kaynak tüketimi de yükselmektedir. Bu durum, uygulamada yüksek performans ile açıklanabilirlik arasında bir denge kurulması gerektiğini göstermektedir.

Confusion matrix analizi, hibrit modelin FN oranının oldukça düşük olduğunu, yani gerçek oltalama e-postalarının büyük ölçüde doğru sınıflandırıldığını göstermektedir. Ancak az sayıda spam e-posta, ham sınıfa yanlış atanmıştır. Bu örneklerin ortak özellikleri aşağıda özetlenmiştir:

- Gövdesi kısa, bağlamdan yoksun içerikler (“see attached”, “check invoice” gibi)
- Yalnızca görsel veya bağlantı içeren, metin tabanlı olmayan e-postalar
- Kötü amaçlı ancak meşru gibi yazılmış e-posta gövdeleri (örneğin müşteri hizmeti dilinde yazılmış dolandırıcılıklar)

Bu durum, derin öğrenme tabanlı modellerin yanı sıra görsel veya bağlantı analizlerini de kapsayan multimodal sistemlerin geliştirilmesine ihtiyaç olduğunu göstermektedir.

Kullanılan veri seti, farklı kaynaklardan gelen 82.000’den fazla e-posta içerdiğinden dolayı hem içerik hem de biçim bakımından zengin bir örneklem sunmuştur. CEAS, Enron ve SpamAssassin gibi çeşitli kaynaklardan gelen veriler, modelin genellenebilirliğini artırmış, overfitting riskini azaltmıştır. Ancak, veri kümesinin zamanla güncelliğini kaybedebileceği ve yeni saldırı türlerini içermeyebileceği göz önünde bulundurulmalıdır. Bu nedenle, ileriye dönük çalışmalarda güncel verilerle yeniden eğitim yapılması önerilmektedir.

Modelin doğruluk düzeyine bakıldığında, ticari e-posta servislerinde, kurum içi güvenlik duvarlarında ya da e-posta filtreleme sistemlerinde uygulanabilir olduğu açıktır. Ayrıca, NB bileşeninin açıklanabilir doğası

sayesinde, bu sistemler yalnızca otomatik engelleme yapmaktan ziyade, güvenlik uzmanlarına neden-sonuç ilişkileri sunarak karar verme sürecine destek olabilir.

Transformer tabanlı modellerin son yıllarda doğal dil işleme alanında güçlü bir performans sergilediği bilinmektedir. Ancak çalışmamızın sonuçları, önerilen hibrit yaklaşımımızın bu modellerle yarışabilecek düzeyde olduğunu ortaya koymaktadır. Hibrit model, hem görece daha düşük hesaplama maliyeti hem de yüksek başarı oranları ile pratik uygulamalarda güçlü bir alternatif sunmaktadır. Bu durum, özellikle kurumsal e-posta filtreleme ve gerçek zamanlı sahtekârlık tespit sistemlerinde hibrit modelimizin daha verimli bir çözüm olabileceğini göstermektedir. Çalışmada kullanılan hibrit model, Transformer tabanlı modern yaklaşımlar kadar başarılı sonuçlar elde etmiş ve düşük maliyetli bir çözüm alternatifi olarak öne çıkmıştır.

Bu çalışmanın temel sınırlılıkları aşağıdaki gibi özetlenebilir:

- Dil Sınırlılığı: Veri yalnızca İngilizce e-postaları kapsamaktadır. Farklı dillerde (ör. Türkçe, İspanyolca) oltalama analizleri yapılmamıştır.
- Metin Temelli Yaklaşım: Görseller, bağlantıların URL yapıları veya başlık meta verileri analiz dışı bırakılmıştır.
- Dinamik Güncelleme Eksikliği: Model eğitildikten sonra statik kalmakta, yeni tehdit türlerine karşı güncellen-memektedir.

Gelecek çalışmalarda, LSTM, Transformer tabanlı modeller, multimodal veri (resim + metin) kullanımı ve sıfır örnekli öğrenme (zero-shot phishing detection) gibi yaklaşımlar değerlendirilebilir.

KAYNAKLAR

1. Jakobsson, M., Myers, S., “Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft”, Sayfa 545-556,Springer, 2006.

2. Abu-Nimeh, S., Nappa, D., Wang, X., Nair, S., “A Comparison of Machine Learning Techniques for Phishing Detection”, In Proceedings of the Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, Pages 60–69, 2007.

3. Khonji, M., Iraqi, Y., Jones, A., “Phishing Detection: A Literature Survey”, IEEE Communications Surveys & Tutorials, Vol. 15, Pages 2091–2121, 2013.

4. Bergholz, A., De Beer, J., Glahn, S., Moens, M.F., Paaß, G., Strobel, S., “New Filtering Approaches for Phishing Email”, J Comput Secur, Vol. 18, Pages 7–35, 2010.

5. Geerthik, S., Anish, T. P. , “Filtering spam: Current trends and techniques”, International Journal of Mechatronics, Electrical and Computer Technology Austrian E-Journals of Universal Scientific Organization, Vol. 3, Pages 208-223, 2013.

6. Sahami, M., Dumais, S., Heckerman, D., Horvitz, E.,“A Bayesian Approach to Filtering Junk E-Mail”, In Proceedings of the Learning for Text Categorization: Papers from the 1998 workshop. Vol. 62, Pages 98–105, 1998.

7. Ngartera, L., Issaka, M.A., Nadarajah, S., “Hybrid Naïve Bayes Models for Scam Detection: Comparative Insights From Email and Financial Fraud”, IEEE Access, Vol. 13, Pages 85207–85216, 2025.

8. Zhou, Z.H., “Ensemble Methods: Foundations and Algorithms”, Sayfa 101-102 CRC press, 2025;

9. Güven, Z.A., “Türkçe E-Postalarda Spam Tespiti Için Makine Öğrenme Yöntemlerinin ve Dil Modellerinin Analizi”, European Journal of Science and Technology, Sayı 47, Sayfa 1-6 2023.

10. Savaş, T., Savaş, S., “Tekdüzen Kaynak Bulucu Yoluyla Kimlik Avı Tespiti Için Makine Öğrenmesi Algoritmalarının Özellik Tabanlı Performans Karşılaştırması”, Politeknik Dergisi, Sayı 25, Sayfa 1261–1270,2022.

11. Eryılmaz, E.E., Kılıç, E., “İstenmeyen Epostaların Tespiti Için Kullanılan Yöntemlerin İncelenmesi”, DÜMF Mühendislik Dergisi, Cilt 11, Sayı 3, Sayfa 977-987, 2020.

12. Ahi, Ş., Soğukpınar, İ., “Derin Öğrenme Modelleri Ile Kimlik Avı E-Posta Tespiti”, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi Sayı 13, Sayfa 17–31, 2020.

13. Bountakas, P., Xenakis, C., “HELPHED: Hybrid Ensemble Learning PHishing Email Detection”, Journal of Network and Computer Applications, Sayı 210, Article 103545, 2023.

14. Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S.B., Joga, S.R.K., "Phishing Detection System Through Hybrid Machine Learning Based on URL", IEEE Access, Vol. 11, Pages 36805–36822, 2023.

15. Hatipoğlu, C., Tunacan, T., "Türkiye'de Siber Saldırı ve Tespit Yöntemleri: Bir Literatür Taraması", Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi, Sayı 8, Sayfa 430–445, 2021.

16. Samarthrao, K.V., Rohokale, V.M., "Enhancement of Email Spam Detection Using Improved Deep Learning Algorithms for Cyber Security", J Comput Secur , Vol. 30, Pages 231–264, 2022.

17. Shajahan R., Lekshmy, P.L., "Hybrid Learning Approach for E-Mail Spam Detection and Classification", In Proceedings of the Intelligent Cyber Physical Systems and Internet of Things; Hemanth Jude and Pelusi, D. and C.J.I.-Z., Pages 781-794 ,Springer International Publishing: Cham, 2023.

18. Alsubaei, F.S., Almazroi, A.A., Ayub, N., "Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics", IEEE Access, Vol. 12, Pages 8373–8389, 2024.

19. Al-Subaiey, A., Al-Thani, M., Alam, N. A., Antora, K. F., Khandakar, A.,Zaman, S. A. U., " Novel interpretable and robust web-based AI platform for phishing email detection. " Computers and Electrical Engineering, 120, 109625, 2024.

20. Zhang, C., Zuo, W., Peng, T., He, F., "Sentiment Classification for Chinese Reviews Using Machine Learning Methods Based on String Kernel", In Proceedings of the 2008 Third international conference on convergence and hybrid information technology, Vol. 2, Pages 909–914, 2008.

21. Webb, G.I., Keogh, E.,Miikkulainen, R. "Naive Bayes. Encyclopedia of machine learning", Vol. 15, Pages 713–714, 2010.

22. Bisong, E., "The Multilayer Perceptron (MLP). In Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners", Pages 401-405 Springer, 2019.

23. Divina, F., Gilson, A., Goméz-Vela, F., Garcia Torres, M., Torres, J.F., "Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting", Energies, Vol. 11, Pages 949-965, 2018.

24. LaValley, M.P., "Logistic Regression", Circulation, Vol. 117, Pages 2395–2399, 2008.

25. Allgaier, J., Pryss, R., "Cross-Validation Visualized: A Narrative Guide to Advanced Methods", Mach Learn Knowl Extr, Vol 6, Pages 1378–1388, 2024.

26. Susmaga, R.,"Confusion Matrix Visualization",In Proceedings of the Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM '04 Conference, Zakopane, Poland, Sayfa 107–116, 2004.

27. Nahm, F.S., "Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians", Korean J Anesthesiol, Vol. 75, Pages 25–36, 2022.

28. Turner, J.R., "Area under the Curve (AUC)", Encyclopedia of Behavioral Medicine, Sayfa 146-150, 2020.

29. Breslow, N.,"Regression Analysis of the Log Odds Ratio: A Method for Retrospective Studies", Biometrics, Pages 409–416, 1976.