Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 47, S. 1-6, Ocak 2023

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 47, pp. 1-6, January 2023

Research Article

Türkçe E-postalarda Spam Tespiti için Makine Öğrenme Yöntemlerinin ve Dil Modellerinin Analizi

Zekeriya Anil Guven[1*]

[1*] Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye, (ORCID: 0000-0002-7025-2815), anilguven1055@gmail.com

(3rd International Conference on Engineering and Applied Natural Sciences ICEANS 2023, January 14 - 17, 2023)

(**DOI:** 10.31590/ejosat.1234079)

ATIF/REFERENCE: Guven, Z. A. (2023). Türkçe E-postalarda Spam Tespiti için Makine Öğrenme Yöntemlerinin ve Dil Modellerinin Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (47), 1-6.

Öz

Son zamanlarda teknolojinin ve sosyal ağların gelişmesiyle çevrimiçi karşılıklı etkileşim, herhangi konuda fikirlerini paylaşma oldukça önem kazanmıştır. Bu etkileşimlerin olumlu yanı olsa da olumsuz yanı da oldukça fazladır. Sosyal ağlarda kullanıcıların bilgilerini elde edip kullanıcıları taklit etmek güvenlik açısından büyük bir problemdir. Böylelikle kullanıcılar üzerinden dolandırıcılık vs. yapılabilmektedir. Kullanıcıları taklit edebilmek için en yaygın yol spam mesajların, e-postaların, vs. atılmasıdır. Güvenlik probleminin üstesinden gelmek için spam filtreleme, spam tespiti yöntemi geliştirme gibi işlemler uygulanmaktadır. Bu çalışmada Türkçe e-postalarda spam içeren e-postaların tespiti için Rastgele Orman, Lojistik Regresyon, Naive Bayes, Yapay Sinir Ağları makine öğrenme yöntemleri ve BERT, ELECTRA, ALBERT, DistilBERT dil modelleri analiz edilmiştir. Böylece dil modellerinin Türkçe için spam e-postaları sınıflandırmadaki etkisi gösterilmek istenmiştir. Deneysel çalışmaların sonucunda, spam e-postaları sınıflandırmada tüm dil modelleri makine öğrenme yöntemlerine göre daha başarılı olmuştur. Makine öğrenme yöntemlerinden yapay sinir ağları %90.15 doğrulu değeri elde ederken, en başarılı dil modelleri %94.08 doğruluk değeri ile BERT ve ELECTRA olmuştur.

Anahtar Kelimeler: Siber Güvenlik, Spam Tespiti, Dil Modeli, Makine Öğrenmesi, Doğal Dil İşleme, Metin Sınıflandırma.

Analysis of Machine Learning Methods and Language Models for Spam Detection in Turkish Emails

Abstract

Recently, with the development of technology and social networks, online interaction, sharing ideas on any subject has gained importance. While there are positive aspects to these interactions, there are also many negative aspects. Obtaining users' information and impersonating users in social networks is a big problem in terms of security. Thus, fraud etc. can be done by under cover of users. The most common way to impersonate users is by sending spam messages, emails, etc. In order to overcome the security problem, processes such as spam filtering and spam detection method development are applied. In this study, Random Forest, Logistic Regression, Naive Bayes, Artificial Neural Networks machine learning methods and BERT, ELECTRA, ALBERT, DistilBERT language models were analyzed to detect e-mails containing spam in Turkish e-mails. Thus, it is aimed to show the effect of language models in classifying spam e-mails for Turkish. As a result of experimental studies, all language models were more successful than machine learning methods in classifying spam emails. While artificial neural networks from machine learning methods achieved 90.15% accuracy, the most successful language models were BERT and ELECTRA with 94.08% accuracy.

Keywords: Cyber Security, Spam Detection, Language Model, Machine Learning, Natural Language Processing, Text Classification.

[*] Sorumlu Yazar: anilguven1055@gmail.com

1. Giriş

Teknolojinin yaşamamıza girmesiyle beraber akıllı cep telefonu, bilgisayar gibi birçok teknolojik cihaz günlük hayatta kullanılmaktadır. İnsanların birbiri ile etkileşime geçmesini sağlayan bu gelişmeden dolayı birçok sosyal ağ ortaya çıkmaktadır. Kişilerin birbirine e-posta ve mesaj göndermesi, sosyal medyada paylaşımlara yorumlar yapılması, kullanıcının etkileşim alması gibi süreçler sosyal ağlarda mevcuttur.

Sosyal veya profesyonel bağlantılar, etkileşimler, büyüyen iş itibarı, kamuoyu görüşlerini bilmek, haberler, çevrimiçi öğrenme, afet yönetimi, sağlık hizmetleri, öneri sistemleri sosyal ağların temel uygulamalarıdır. Ancak kullanıcılar, kişisel bilgilerini ara sıra bu sosyal ağlarda bilmeden de olsa yabancılara paylaşabilmektedir. Bu da güvenlik açısından sorun oluşturmaktadır. Spam gönderenler, botlar, bilgisayar korsanları, siber suçlular ve üçüncü taraf şirketler gibi kötü niyetli kullanıcılar, dolandırıcılık gibi sorun çıkaran faaliyetleri gerçekleştirmek için bu kullanıcıların bilgilerden yararlanmaktadır. Spam gönderenler ayrıca insanların güvenini kullanarak reklam, kimlik avı, casusluk, kadına yönelik şiddet, siber zorbalık gibi yasa dışı faaliyetlerde bulunabilmektedir. Genel olarak spam gönderenler, spam yaymak için meşru hesaplar yerine sahte/kopyalanan hesaplar, otomatik botlar vs. kullanmaktadır (Rao vd., 2021). Böylece spam gönderenler kendilerini güvene almaktadır. Spam gönderme mesaj veya e-posta yoluyla, yorum yapılma vs. gibi olaylarla yapılabilmektedir. Yorum ile iletilen spamlar, tüketicinin güveninde kayba neden olduğundan işletmeleri de olumsuz etkilemektedir. Örneğin, BBC ve New York Times, "spam yorumların Web'de yaygın bir sorun haline geldiğini ve yakın zamanda bir fotoğraf şirketinin yüzlerce karalayıcı tüketici spam yorumuna maruz kaldığını" belirtmektedir (Crawford vd., 2015).

Spam mesajlar veya e-postalar, kaçınılması veya ortadan kaldırılması gereken bir konudur. Çünkü bu tür spam e-postalar, oluşturulan zararlı olaylarla genel e-postaları etkilemektedir. Bu tür spam e-postaların tespit edilmesi, yasa dışı faaliyetler sorunundan kaçınmak için oldukça önemlidir. Spam e-postaları filtreleme için kara liste, beyaz liste, içerik tabanlı veri filtresi gibi metodolojiler yer almaktadır. Ancak, bunlar tamamen başarılı bir metodoloji olmamaktadır. E-posta kara liste mekanizması, çok sayıda spam e-posta kimliğini saklayarak kullanıcılara mesajların bu saklanan e-posta kimliklerinden gelip gelmediğini kontrol etmektedir. İçerik tabanlı veri filtresinde ise şüpheli IP adreslerinden gelen mesajları engelleyen ve mevcut sisteme makul bir şekilde güvenlik sağlayan IP tabanlı filtreleme bulunmaktadır. Beyaz listeye alınan e-postalar ise bilinen kimlik e-postalarıdır. Bu liste ile kullanıcıya bilinen posta kimliklerini engelleme, gelecek e-postalardan korumayı sağlamaktadır (Ismail vd., 2022).

Spam e-postalarının tespit edilebilmesi için birçok İngilizce dilinde çalışma bulunmaktadır. Bu çalışmada Türkçe açısından literatüre katkı sağlamak için Türkçe spam e-postalarını sınıflandırma için analiz uygulanmaktadır. Analiz aşamasında makine öğrenme yöntemlerinin yanı sıra günümüzde oldukça sık kullanılan dil modelleri (DM) de tercih edilmektedir. Böylece makine öğrenme yöntemleri ile DM arasında Türkçe dili için karşılaştırma yapılması da amaçlanmaktadır.

Çalışmanın başlıca katkıları şunlardır:

- Türkçe dili için spam tespiti analiziyle literatüre katkı sağlamaktadır.
- Dil modellerinin Türkçe için kullanıldığı çalışma sayısı oldukça azdır. Bundan dolayı dil modelleri ile analiz önem taşımaktadır.
- Makine öğrenme yöntemleri ile dil modellerinin başarısı kıyaslanarak dil modellerinin Türkçe dili için etkisi gösterilmektedir.

Çalışmanın akışı incelendiğinde ikinci başlıkta Türkçe spam tespiti için ilgili çalışmalar anlatılmaktadır. Materyal ve Yöntem başlığı altında ise kullanılan veri seti, uygulanan ön işlemler, makine öğrenme yöntemleri ve dil modelleri kısaca açıklanmaktadır. Dördüncü başlık içeriğinde ise ön işlem analizi, makine öğrenme yöntemleri ve dil modellerinin analizi gerçekleştirilmektedir. Tartışma bölümünde ise makine öğrenme ve dil modellerinin spam tespitine etkisi açıklanmaktadır. Son başlıkta ise sonuçlara ve gelecek çalışmalara değinilmektedir.

2. İlgili Çalışmalar

Spam tespitinde Türkçe içinde belli çalışmalar yapılmıştır. Bu çalışmalar genel olarak makine öğrenme ve derin öğrenme yöntemlerinin kullanımına dayalıdır. Deniz vd. (2019) Türkçe spam tespitinde e-postaların sayısallaştırılması için öznitelik çıkarımında Doc2Vec kütüphanesinin algoritmalarını kullanmışlardır. Ardından sınıflandırma algoritmaları ile Doc2Vec'in başarısını ölçmüşlerdir. Eryılmaz vd. (2020) Keras derin öğrenme kütüphanesi ile Türkçe spam verileri analiz etmişlerdir. Uzun kısa süreli bellek (LSTM) yöntemi ile analizin başarılı olduğunu göstermişlerdir. Şimsek ve Aydemir (2022), yeni bir Türkçe spam e-posta veri seti oluşturmuşlar ve Weka programındaki algoritmalar kullanılarak Türkçe spam veya normal e-postaların sınıflandırılmasını analiz etmişlerdir. Karasoy ve Balli (2022) makine öğrenmesi ve derin öğrenme yöntemleri kullanarak içerik tabanlı SMS sınıflandırması uygulamışlardır. Elde edilen özelliklere ek Word2Vec ile çıkarılan ayrı özellikler ile yöntemlerin başarılarını analiz etmişlerdir. Dedetürk ve Atay (2020) yapay arı kolonisi algoritmasını bir lojistik regresyon sınıflandırma modeliyle birleştiren yeni bir spam tespit yöntemi önermişlerdir. Türkçe e-posta veri seti dahil üç veri seti üzerinde bu yöntemi analiz etmişlerdir. Önerilen yöntemi destek vektör makinesi, lojistik regresyon ve Naive Bayes algoritmaları ile karşılaştırmışlar ve yöntemleri ile daha iyi sonuç elde etmişlerdir. Isik vd. (2020) Karşılıklı Bilgi ve Ağırlıklı Karşılıklı Bilgi özellik seçme yöntemlerini kullanarak farklı derin öğrenme yöntemlerini Türkçe e-posta spam tespiti için analiz etmişlerdir. Özellik seçim işleminden sonra, kelime çantası yöntemi ile özellik vektörleri elde etmişlerdir. Ardından, Yapay Sinir Ağı (YSA), LSTM ve Çift Yönlü LSTM yöntemleri kullanılarak sistemin performansını ölçmüşlerdir. Ekici ve Takcı (2021), diğer dillerde sıkça kullanılan Word2Vec ve Terim Frekansı-Ters Terim Frekansı (TF-IDF) yöntemlerini Türkçe spam tespiti için karşılaştırmışlar ve başarıyı artırmışlardır.

Türkçe dili için dil modelleri genellikle duygu analizi (Acikalin vd. (2020), Guven (2021a), Sigirci vd. (2020)) alanında, metin sınıflandırmada (Çelikten ve Bulut (2021), Şahin ve Diri (2021)) kullanılmıştır.

3. Materyal ve Metotlar

Çalışmada kullanılan veri seti, uygulanan ön işlemler, uygulanan yöntemler bu kısımda açıklanmaktadır. Uygulanan makine öğrenme yöntemleri ve dil modelleri analiz için ayrı olarak kullanılmaktadır.

2.1. Veri Seti

Kaggle platformunda yüklü olan Türkçe e-posta spam veri seti[†] analiz için kullanılmaktadır. Veri seti 502 normal (ham) mailden, 517 ise spam mailden oluşmaktadır. 20 kişiye gelen normal ve spam maillerden veri seti elde edilmiştir (Şimsek ve Aydemir, 2022). Veri setine ait kelime bulutu Şekil 1'de gösterilmiştir. Veri setinin, eğitim ve test olarak kullanımı açıkça belirtilmediği için karşılaştırma amaçlı kullanılamamıştır.

Şekil 1. Veri setine ait kelime bulutu

2.2. Ön İşlemler

Analiz öncesinde veri seti üzerinde belli ön işlemler uygulanmaktadır. Ön işlemlerin uygulanması sırayla gerçekleştirilmektedir:

- E-posta küçük harfe dönüştürülmektedir.
- Metin içindeki email ve web adresleri düzenli ifadeler (RegEx) ile tespit edilerek sırasıyla "email, website" olarak güncellenmiştir. Bu işlem ile farklı adreslerin aynı yapıda olmasını sağlamak amaçlanmıştır.
- E-posta içindeki sayısal ifadeler ve noktalama işaretleri silinmektedir.
- E-postadaki tüm metin için Zeyrek[‡] kütüphanesi aracılığıyla kök alma (lemmatization) işlemi uygulanmıştır. Böylece aynı anlama gelen farklı yapıdaki kelimeler aynı yapıya dönüştürülmüştür.
- Kök alma işlemi sonrası e-posta içinde etkisiz kelimeler silinmiştir. Bu işlemin sonucunda analize etki etmeyen kelimeler silinerek veri hacmi azalmıştır.

Ön işlemler sonrasında veri seti güncel haliyle kaydedilerek analiz aşamasında kullanılmıştır. Ön işlemlere ait bir örnek Tablo 1'de verilmiştir.

2.3. Makine Öğrenme Yöntemleri

Çalışmada makine öğrenme yöntemleri olarak Rastgele Orman (RO), Naive Bayes (NB), Lojistik Regresyon (LR) ve Yapay Sinir Ağları (YSA) kullanılmaktadır. Etkisiz kelime temiz-

Tablo 1. Makine öğrenme yöntemlerinin başarısı

Ön İşlemler	Metnin Değişimi
Mailin ilk hali	Ödemen İade Edildi **kitapsec.com** web sitesinde 16/02/2022 22:00 tarihinde gerçekleştirmiş olduğun 39.39 TRY tutarındaki işleminin iadesi iyzico tarafından başarıyla tamamlanmıştır. İade işlemleri, kredi kartları ile yapılan …
Küçük harfe dönüştürme + Email ve web adresi düzenleme	ödemen iade edildi **website** web sitesinde **16/02/2022 22:00** tarihinde gerçekleştirmiş olduğun **39.39** try tutarındaki işleminin iadesi iyzico tarafından başarıyla tamamlanmıştır. iade işlemleri, kredi kartları ile yapılan …
Sayısal ifade ve noktalama işaretleri silme	ödemen iade edildi website web sitesinde tarihinde gerçekleştirmiş olduğun try tutarındaki işleminin iadesi iyzico tarafından başarıyla tamamlanmıştır iade işlemleri kredi kartları ile yapılan …
Kök alma işlemi	ödemek iade edildi website web site tarihî gerçek olmak try tutar işlem iade iyzico tarafından başar tamam iade işlem **ile** kredi kart yapılanmak …
Etkisiz kelimelerin silinmesi	ödemek iade edildi website web site tarihî gerçek olmak try tutar işlem iade iyzico tarafından başar tamam iade işlem kredi kart yapılanmak …
Mailin son hali	ödemek iade edildi website web site tarihî gerçek olmak try tutar işlem iade iyzico tarafından başar tamam iade işlem kredi kart yapılanmak …

leme, noktalama işaretleri kaldırma gibi ön işlemler sonrasında Terim Frekansı-Ters Terim Frekansı (TF-IDF) yöntemi ile veri setinin özellikleri çıkarılmaktadır. TF-IDF ile çı-karılan özellikler, makine öğrenme yöntemlerinin eğitimi aşamasında k-çapraz doğrulama tekniği ile kullanılmaktadır. Teknikteki k değeri 5 olarak belirlenmiştir. Böylece veri seti 5 kümeye ayrılarak her aşamada farklı 4'lü küme eğitim, kalan tek küme ise test olarak eğitime verilmektedir. Sonuçlar bu 5'li varyasyona göre aşağıdaki makine öğrenme yöntemler için elde edilmektedir:

- **RO:** Karar ağacı yapısını kullanan bir topluluk öğrenmesi yöntemidir. Çok sayıda karar ağacı üretip bir modele dahil edilen her bir değerlendiricinin önemini belirlemektedir (Probst ve Boulesteix, 2017).
- **NB:** Olasılıksal bir makine öğrenme yöntemidir. Sınıfa verilen özelliklerin birbirinden bağımsız olduğu varsayarak bir sınıf için yeni olan örneğin olasılığını hesaplamaktadır (Chen vd., 2020).
- **LR:** İki veya daha fazla sınıf için ayrım yapabilmeyi sağlayan model oluşturmada kullanılmaktadır. Girdinin doğrusal fonksiyonu kullanılarak belirlenen sayıda sınıfların olasılıklarını modellemektedir (Chen vd., 2018).

[†] https://www.kaggle.com/datasets/emrahaydemr/turkish-mail-dataset-normalspam

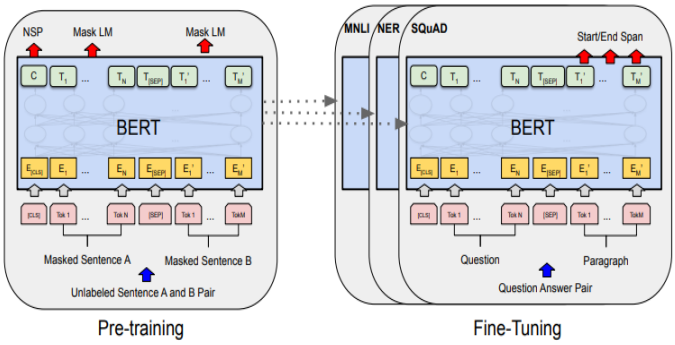
[‡] https://zeyrek.readthedocs.io/en/latest/#

- **YSA:** Sinir ağları insanların beyin yapısından etkilenerek oluşturulmuştur. Girdi, gizli ve çıktı katmanlarından oluşmaktadır. Modelin hesaplanmasında değişken ağırlık çarpanları, toplam fonksiyonu, tanımlama (aktivasyon) fonksiyonu kullanılmaktadır (Taşar vd., 2018).

2.4. Dil Modelleri

DM, kelime tahmini, sınıflandırma gibi görevler için metinlerin gövdelerini analiz etmektedir. Metinlerin gövdelerine ait kelime dizilerini girdi olarak kullanarak görev için olasılık dağılımı hesaplamaktadır (Guven, 2021b). DM, tek yönlü ve çift yönlü olarak iki modele ayrılmaktadır. Tek yönlü DM, girdi dizisi için diziyi çarpanlarına ayırıp bu diziye bir olasılık atamaktadır. Çift yönlü DM için ise kelimenin konumu ile kelimenin sol ve sağ bağlamını da analiz etmektedir (Petroni vd., 2019). Çalışmada BERT, ELECTRA, ALBERT ve DistilBERT DM'leri kullanılmaktadır.

BERT: Transformer temelli çift yönlü kodlayıcı gösterimi olarak tanımlanmaktadır. Kelimenin sadece tek bir yönüne odaklanmayarak sol ve sağ bağlamını incelemektedir. BERT modeli maskelenmiş dil modeli adı verilen yapıyı eğitim aşamasında kullanmaktadır. Bu yapıda girişteki bazı belirteçler rastgele seçilerek maskelenmektedir, model ise bu maskelenen belirteçleri bağlama göre doğru tahmin etmeyi hedeflemektedir (Devlin vd., 2018).

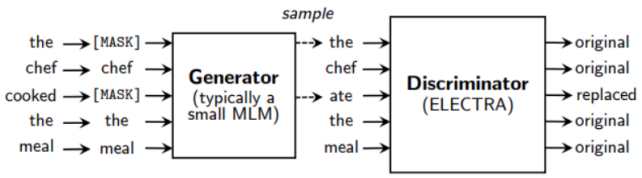


Şekil 2. BERT modelinin işleyişi

Şekil 2'de BERT modelinin işleyişi gösterilmiştir. BERT modeli ön eğitim ve ince ayar olmak üzere iki aşamadan oluşmaktadır. Ön eğitim (pre-training) aşamasında, model farklı ön eğitim görevleri için etiketlenmemiş veriler kullanılarak eğitilmektedir. İnce ayar (fine-tuning) aşamasında ise önceden eğitilmiş parametreler ile başlatılan modelde tüm parametreler seçilen göreve göre etiketli veriler ile uyarlanmaktadır (Devlin vd., 2018). BERT modeli, sınıflandırma, görüntü işleme, boşluk doldurma gibi birçok görevde kullanılabilmektedir. Bunun için eğitilmiş ve kullanılabilecek birçok ön-eğitilmiş DM mevcuttur. Türkçe içinde ön-eğitilmiş BERT-TR[§] modeli kullanılarak bu çalışmada analiz gerçekleştirilmektedir.

DistilBERT: BERT modelinin farklı bir versiyonu olarak önerilmiş bir modeldir. BERT modelinden farklı olarak eğitim öncesi bilgi ayrıştırma işleminden yararlanarak BERT'e göre boyutu daha küçük ve daha hızlı DM elde edilmiştir (Sanh vd., 2019). Çalışmada analiz için DistilBERT-TR[**] ön eğitilmiş modeli kullanılmaktadır.

ELECTRA: BERT modeline göre daha az hesaplama uygulayarak önceden eğitmek için kullanılmaktadır. Modelin amacı "gerçek" olan belirtecin "sahte" belirteç ile değiştirilip değiştirilmediğini tespit etmektir (Clark vd., 2020). Şekil 3'te ELECTRA modelinin yapısı gösterilmiştir. Şekil incelendiğinde "cooked" kelimesi maskelenmiştir ve küçük MLM aracılığıyla tahmin edilen kelime "ate" olmuştur. Bu kelimenin değiştirildiğini ise ELECTRA modeli "replaced" olarak tespit etmiştir.



Şekil 3. ELECTRA modelinin yapısı

ELECTRA modelinin, BERT modelinin maskelenmiş dil modellemesi yapısına göre, hesaplama açısından daha verimli olduğu gösterilmiştir (Clark vd., 2020). Çalışmada spam e-postaları sınıflandırma görevi için ELECTRA-TR[††] ön eğitilmiş modeli kullanılmaktadır.

ALBERT: BERT modelinde bellek limiti ve iletişim ek yükünü sorunu eğitim aşamasında sorun oluşturmaktadır. Bundan dolayı BERT'e göre daha az parametreye sahip ALBERT modeli geliştirilmiştir. Geliştirme aşamasından iki tane parametre azaltma (çarpanlara ayırarak gömme, katmanlar arası parametre paylaşımı) tekniği uygulanmıştır. Bu iki teknikle beraber ALBERT için parametre sayısı oldukça azaltmaktadır (Lan vd., 2019). Böylece model daha az parametre ile hızlı eğitilebilmektedir. Çalışmada ALBERT-TR[‡‡] ön eğitilmiş modeli kullanılarak analiz işlemi gerçekleştirilmektedir.

3. Deneysel Çalışmalar

Çalışmanın ilk aşamasında veri seti üzerine ön işlemler uygulanmıştır. Uygulanan ön işlemler sonucunda veri setinin yapısı, hacmi değişmiştir. Bu ön işlemler arasında email ve website adreslerinin belli formatta yazılması, kelimelerin kökünün alınması işlemleri veri setinin yapısını değiştirirken, etkisiz kelimelerin silinmesi işlemi veri hacmini azaltmaktadır. Tablo 2'de veri seti üzerindeki kelime sayısının ön işlem sonrası değişimi gösterilmiştir. Tablodaki veriler incelendiğinde, ön işlem sonrası kelime sayısında düşüş gözlemlenmiştir. Bu düşüş veri hacmi için toplam kelime sayısında yaklaşık %25 iken, eşsiz (unique) kelime sayısında ise yaklaşık %64 olmuştur. Eşsiz kelime sayısında aşırı düşüşün nedeni kelimelerin kökünün alınması ile kelimelerin aynı biçime dönüştürülmesidir.

Tablo 2. Makine öğrenme yöntemlerinin başarısı

İşlem		Kelime Sayısı
	Toplam	66870

[§] https://huggingface.co/dbmdz/bert-base-turkish-uncased

[**] https://huggingface.co/dbmdz/distilbert-base-turkish-cased

[††] https://huggingface.co/dbmdz/electra-base-turkish-cased-discriminator

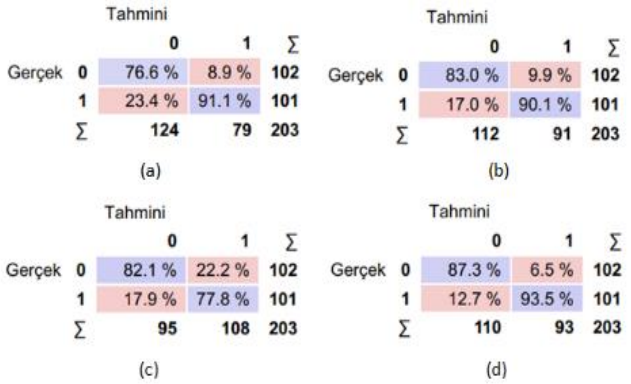
[‡‡] https://huggingface.co/loodos/albert-base-turkish-uncased

Ön İşlem Olmadan	Eşsiz	19162
Ön İşlem Sonrası	Toplam	50112
	Eşsiz	6892

Çalışmada eğitim ve geçerleme olarak ayrılan veri setlerinden, eğitim seti makine öğrenme yöntemleriyle eğitilerek geçerleme seti ile test edilmiştir. Türkçe spam e-postaların tespitinde kullanılan makine öğrenme yöntemlerine ait sonuçlar Tablo 3'te gösterilmiştir. Sonuçlar incelendiğinde en başarılı makine öğrenme yöntemi doğruluk ve F1-ölçüm değerleri için YSA olmuştur. Ayrıca yöntemlerin ayrı ayrı etiketler için analizlerini içeren karışıklık matrisi Şekil 4'te verilmiştir. Makine öğrenme yöntemleri arasında RO hariç diğer yöntemler 1 numaralı (spam) etikete sahip verileri daha doğru sınıflandırmıştır. RO ise 0 numaralı (ham) etikete sahip verileri daha başarılı sınıflandırmıştır.

Tablo 3. Makine öğrenme yöntemlerinin başarısı

Yöntem	Eğitim Süresi(sn)	Doğruluk (%)	F1-Ölçüm (%)
RO	6.90	79.80	79.78
YSA	9.76	90.15	90.13
NB	13.78	86.21	86.17
LR	37.96	82.27	82.05

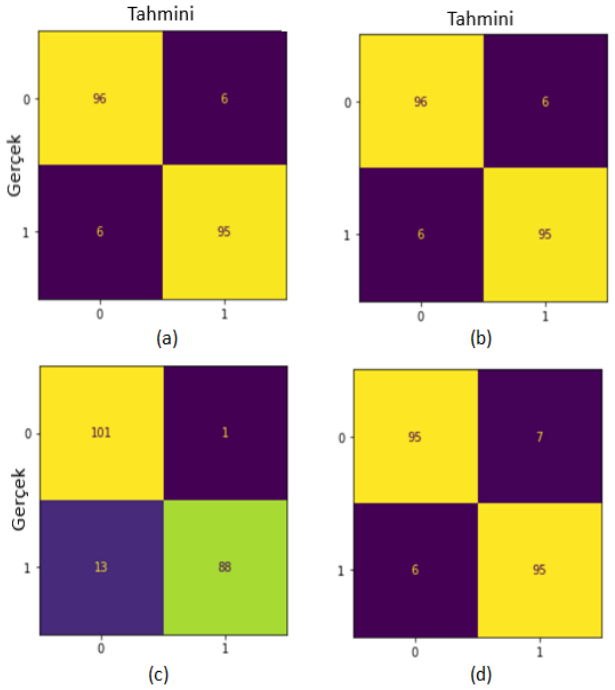


Şekil 4. Makine öğrenme yöntemlerinin karışıklık matrisleri (a) LR (b) NB (c) RO (d) YSA

İkinci aşamada aynı eğitim ve geçerleme seti ile DM'nin analizi gerçekleştirilmiştir. Türkçe için ön-eğitilmiş BERT-TR, ALBERT-TR, ELECTRA-TR, DistilBERT-TR modelleri için analiz sonuçları Tablo 4'te verilmiştir. Tablodaki veriler incelendiğinde BERT-TR ve ELECTRA-TR modelleri en başarılı DM olmuştur. Ayrıca DM için hangi etiketleri daha doğru yanıtladığına dair karışıklık matrisi Şekil 5'te gösterilmiştir. Şekil 5'teki sonuçlar gösteriyor ki ALBERT-TR modeli haricinde tüm modeller her iki etiketi de dengeli şekilde doğru sınıflandırmıştır. Ancak ALBERT-TR modeli 0 etiketini (ham), 1 etiketine (spam) göre daha doğru tahmin etmiştir.

Tablo 4. Türkçe spam e-postaların tespiti için dil modellerinin başarısı

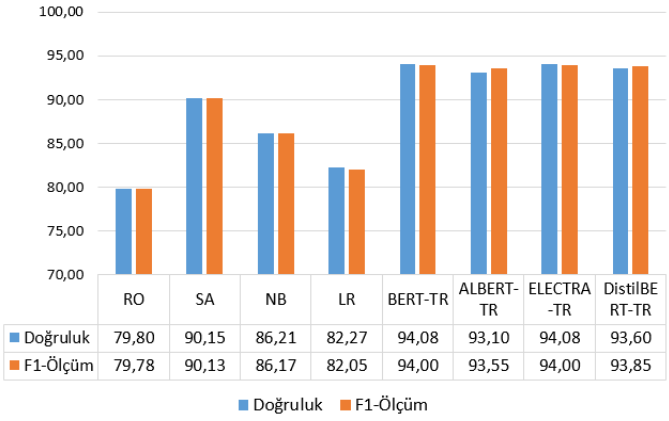
Yöntem	Doğruluk (%)	F1-Ölçüm (%)
BERT-TR	94.08	94.0
ALBERT-TR	93.10	93.55
ELECTRA-TR	94.08	94.0
DistilBERT-TR	93.60	94.0



Şekil 5. Dil modellerinin karışıklık matrisleri (a) BERT-TR (b) ELECTRA-TR (c) ALBERT-TR (d) DistilBERT-TR

4. Tartışma

Türkçe dili ile alakalı çalışmalar az olduğu ve makine öğrenmesi yöntemlerine ek olarak DM'nin Türkçe spam e-posta sınıflandırmada etkisinin gösterilmesi için bu çalışma uygulanmıştır. Veri seti ilk olarak çok fazla sayıda çalışmada kullanılan makine öğrenme yöntemleri ile analiz edilmiştir. Analizler sonucunda makine öğrenme yöntemlerinde YSA %90.15'lik doğruluk değeri elde etmiştir. Bu analize ek olarak, DM'nin Türkçe spam e-posta sınıflandırmadaki başarısı analiz edilmiştir. Aynı yapıdaki veri seti (eğitim, geçerleme) ile önceden belirtilen ön-eğitilmiş DM analizde kullanılmıştır. DM, derin öğrenme ve anlamsal yapı içermesinin katkısıyla spam e-postaları sınıflandırmada makine öğrenme yöntemlerine göre %3 ile %4 arasında doğruluk değeri artışı sağlamıştır. Makine öğrenme yöntemlerinin ve DM'nin doğruluk değerleri ve F1-ölçüm değerleri Şekil 6'da verilmiştir. Grafik incelendiğinde tüm DM, makine öğrenme yöntemlerinden daha iyi sonuç vermiştir. Makine öğrenme yöntemleri arasında YSA en iyi doğruluk değeri elde ederken, dil modelleri arasında BERT-TR ve ELECTRA-TR



Şekil 6. Tüm makine öğrenme yöntemleri ve dil modellerinin analiz sonuçları

modelleri %94.08 doğruluk elde ederek en başarılı sonuçları vermiştir. Bu açıdan bakıldığında, çalışma Türkçe spam tespiti için DM'nin olumlu etkisini göstermede öncülük etmektedir.

5. Sonuç ve Gelecek Çalışmalar

Spam tespiti, email, twitter gibi sosyal ağlarda kullanıcıların kandırılmasını ve suçlara karıştırılmasını önlemek için büyük önem taşımaktadır. İngilizce dilinde bu konuyla ilgili birçok çalışma bulunmaktadır. Bu çalışmada Türkçe spam e-postaların tespiti için makine öğrenme yöntemleri ve DM kullanılarak başarıları analiz edilmiştir. Türkçe spam e-postaların tespiti için DM içeren çalışma olmadığı için DM, Türkçe literatüre katkı sağlama açısından tercih edilmiştir. Böylece DM makine öğrenme yöntemleri ile karşılaştırılmış ve DM'nin sınıflandırmadaki etkisi gösterilmiştir. Ayrıca makine öğrenme yöntemleri ve DM'nin başarıya etkisi karşılaştırılmıştır. Deneysel çalışmaların sonucunda tüm DM, makine öğrenme yöntemlerinden daha iyi doğruluk değeri vermiştir.

Gelecek çalışmalarda duygu analizi, haber sınıflandırma, iğneleme tespiti gibi farklı sınıflandırma görevleri için Türkçe dilinde dil modelleri ile çalışmalar yapılması ve Türkçe literatüre katkı sağlanması hedeflenmektedir.

Kaynakça

Acikalin, U. U., Bardak, B., & Kutlu, M. (2020, October). Turkish sentiment analysis using bert. In 2020 28th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. Knowledge-Based Systems, 192, 105361.

Chen, H., Gilad-Bachrach, R., Han, K., Huang, Z., Jalali, A., Laine, K., & Lauter, K. (2018). Logistic regression over encrypted data from fully homomorphic encryption. BMC medical genomics, 11(4), 3-12.

Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.

Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. Journal of Big Data, 2(1), 1-24.

Çelıkten, A., & Bulut, H. (2021, June). Turkish Medical Text Classification Using BERT. In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Dedeturk, B. K., & Akay, B. (2020). Spam filtering using a logistic regression model trained by an artificial bee colony algorithm. Applied Soft Computing, 91, 106229.

Deniz, E., Erbay, H., & Coşar, M. (2019, November). Classification of Turkish E-Mails with Doc2Vec. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). IEEE.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ekici, B. & Takcı, H. (2021). Spam Tespitinde Word2Vec ve TF-IDF Yöntemlerinin Karşılaştırılması ve Başarı Oranının Artırılması Üzerine Bir Çalışma. Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi, 8 (2), 646-655.

Eryılmaz, E. E., Şahin, D. Ö., & Kılıç, E. (2020, June). Filtering turkish spam using LSTM from deep learning techniques. In 2020 8th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.

Guven, Z. A. (2021a). Comparison of BERT models and machine learning methods for sentiment analysis on Turkish tweets. In 2021 6th International Conference on Computer Science and Engineering (UBMK) (pp. 98-101). IEEE.

Guven, Z. A. (2021b). The Effect of BERT, ELECTRA and ALBERT Language Models on Sentiment Analysis for Turkish Product Reviews. In 2021 6th International Conference on Computer Science and Engineering (UBMK) (pp. 629-632). IEEE.

Isik, S., Kurt, Z., Anagun, Y., & Ozkan, K. (2020). Spam E-mail Classification Recurrent Neural Networks for Spam E-mail Classification on an Agglutinative Language. International Journal of Intelligent Systems and Applications in Engineering, 8(4), 221-227.

Ismail, S. S., Mansour, R. F., El-Aziz, A., Rasha, M., & Taloba, A. I. (2022). Efficient E-Mail Spam Detection Strategy Using Genetic Decision Tree Processing with NLP Features. Computational Intelligence and Neuroscience, 2022.

Karasoy, O., & Ballı, S. (2022). Spam SMS detection for Turkish language with deep text analysis and deep learning methods. Arabian Journal for Science and Engineering, 47(8), 9361-9377.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., & Riedel, S. (2019). Language models as knowledge bases?. arXiv preprint arXiv:1909.01066.

Probst, P., & Boulesteix, A. L. (2017). To tune or not to tune the number of trees in random forest. The Journal of Machine Learning Research, 18(1), 6673-6690.

Rao, S., Verma, A. K., & Bhatia, T. (2021). A review on social spam detection: challenges, open issues, and future directions. Expert Systems with Applications, 186, 115742.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Siğirci, İ. O., Özgür, H., Oluk, A., Uz, H., Çetiner, E., Oktay, H. U., & Erdemir, K. (2020, September). Sentiment Analysis of Turkish Reviews on Google Play Store. In 2020 5th International Conference on Computer Science and Engineering (UBMK) (pp. 314-315). IEEE.

Şahin, G., & Diri, B. (2021, June). The Effect of Transfer Learning on Turkish Text Classification. In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Şimşek, H. & Aydemir, E. (2022). Classification of Unwanted E-Mails (Spam) with Turkish Text by Different Algorithms in Weka Program. Journal of Soft Computing and Artificial Intelligence, 3 (1), 1-10.

Taşar, B., Fatih, Ü. N. E. Ş., Demirci, M., & Kaya, Y. Z. (2018). Yapay sinir ağları yöntemi kullanılarak buharlaşma miktarı tahmini. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 9(1), 543-551.