BŞEÜ Fen Bilimleri Dergisi
8(2), 646-655, 2021

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.935247

e-ISSN:2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Araştırma Makalesi – Research Article

Spam Tespitinde Word2Vec ve TF-IDF Yöntemlerinin Karşılaştırılması ve Başarı Oranının Artırılması Üzerine Bir Çalışma

A Study on Comparing Word2Vec and TF-IDF Methods and Increasing Success Rate for Spam Detection

Burak Ekici[1*], Hidayet Takcı[2]

Geliş / Received: 09/05/2021 ***Revize / Revised: 22/09/2021*** ***Kabul / Accepted: 22/10/2021***

ÖZ

Elektronik posta, internet üzerinden gönderilen bir tür dijital mektuptur. Elektronik postalar aracılığı ile belge, resim, video, müzik gibi her türlü dosya gönderilip alınabilmektedir. Düşük maliyeti nedeniyle sıklıkla tercih edilmektedir. Elektronik postalar zaman ve para tasarrufu sağladığı için etkili bir iletişim yoludur. Düşük maliyetinden ve kullanımının kolaylığından dolayı reklam yapmak isteyenler tarafından etkin bir şekilde kullanılmaktadır. Bunun yanında siber saldırganlar da kurbanlarına bu tür elektronik postalar göndererek onlara zarar verebilmektedirler. Bu durumların önüne geçebilmek için, günümüzde makine öğrenmesi algoritmalarıyla spam elektronik postaları sınıflayan modeller tasarlanmaktadır. Bu çalışmanın amacı da spam tespiti konusunda literatürde sıklıkla yer alan Word2Vec ve Term Frequency – Inverse Document Frequency(TF-IDF) yöntemlerinin karşılaştırılmasını Türkçe bir veri seti üzerinde yapmak ve daha önce bahsedilen veri seti üzerinde yapılan çalışmalara göre başarı oranını artırmaktır. Bu amaç doğrultusunda, daha önce yapılan çalışmalar incelendiğinde, çalışmaların genellikle İngilizce veri setleri üzerinde yoğunlaştığı görülmektedir. Bu konudaki eksiği gidermek adına, Türkçe veri seti üzerinde yapılan bu çalışmada bahsedilen özellik çıkarma yöntemlerinin karşılaştırılması yapılarak iki farklı model oluşturulmuştur. Bu modellerde farklı sınıflayıcılar da kullanılarak en etkili yöntemin öne çıkarılması hedeflenmiştir.

Anahtar Kelimeler- Spam Tespiti, E-posta, Word2vec, Tf-idf

ABSTRACT

Electronic mail is a kind of digital letter sent over the Internet. A lot of documents such as, images, videos, and music can be transferred via electronic mail. E-mails are often preferred due to their cheapness and easy usage. E-mail is an effective way of communication as it saves time and money. E-mails are used due to its easy usage and low cost by the people who want to advertise their products. Also, hackers can hurt their victims by sending e-mails to them. Nowadays, to prevent these situations, classifiers of the spam electronic mails with some machine algorithms are designed. The aim of this study is to compare Word2Vec and Term Frequency – Inverse Document Frequency (TF-IDF) methods which are frequently included in the literature on Spam Detection, on a Turkish data set and to increase the success rate over previous studies on the related data set. For this purpose, when the previous studies are examined, it is seen that studies generally focus on English data sets. In order to eliminate the lack in this matter, by comparing the mentioned feature extraction methods, two different models are created on a Turkish

[1*]Sorumlu yazar iletişim: bekici391@gmail.com (https://orcid.org/ 0000-0002-2455-2454)
Savunma Sanayi Teknolojileri Yüksek Lisans Programı, Sivas Bilim ve Teknoloji Üniversitesi, 58070, Sivas, Türkiye
[2]İletişim: htakci@cumhuriyet.edu.tr (https://orcid.org/0000-0002-4448-4284)
Bilgisayar Mühendisliği Bölümü, Sivas Cumhuriyet Üniversitesi, 58140, Sivas, Türkiye

data set in this study. It is aimed to highlight the most effective method by using different classifiers in these models.

Keywords- Spam Detect, E-mail, Word2vec, Tf-idf

I. GİRİŞ

Günümüzde haberleşme amacıyla sık sık kullandığımız elektronik posta (e-posta) ilk kez 1971 yılında Raymond Samuel Tomlinson tarafından denenmiştir.

E-Posta kullanımı günümüzde çok yaygın olarak kullanılmaktadır. Neredeyse her internet kullanıcısının bir e-posta hesabı vardır. Kişisel olarak da artık her türlü sosyal medya, video-içerik paylaşımı, oyun uygulamaları gibi dünya genelinde popüler uygulamalara üye olabilmek için bir e-posta adresinin gerekli olması, artık her internet kullanıcısı için neredeyse en az bir e-posta hesabı açmasını zorunlu kılmaktadır. E-postalar bunun yanında kurumlar, kuruluşlar ve şirketlerin gündelik iş hayatındaki vazgeçilmez iletişim aracıdır. Öğrenciler için ise ödevlerine ilgili yerlere iletmelerindeki en büyük gereçtir. Günlük hayatta artık yaygınlaşan e-ticaret sitelerinden yapılan alışveriş sonrasında oluşan e-faturalar müşterilere e-posta aracılığıyla gönderilmektedir.

E-posta kullanımı yaygınlaşması bazı problemleri de beraberinde getirmiştir. Bu sorunların en önemlilerinden bir tanesi istenmeyen (spam) e-postalardır. Spam e-posta; istenmeyen önemsiz, yaramaz veya gereksiz e-posta olarak tarif edilmektedir. Genellikle talep edilmeden çok sayıdaki kullanıcıya bir reklam veya alakasız bir içerik şeklinde gönderilmektedir. Şirketler için bu şekilde reklam yapmak daha az maliyetli olduğundan dolayı sık sık bu şekilde e-postalarla kullanıcıları rahatsız etmektedir. Ayrıca reklamdan farklı olarak, kötü niyetli kişilerce internet kullanıcılarını aldatıp onların hassas bilgilerini ele geçirmek amaçlı spam e-postalar gönderilebilmektedir. Bu şekilde kullanıcıların kişisel bilgileri, kredi kartı bilgileri ve hatta sosyal medya hesaplarının ele geçirilmesi için bile bu yöntem kullanılmaktadır.

Spam e-postalar ve mesajlar internet kullanıcıların zamanını çalmakta ve cihazlarında veya e-posta kutularında gereksiz şekilde yer işgal etmektedir. Ayrıca ağ trafiğini de gereksiz şekilde meşgul etmektedir. Güvenlik açısından da kötü niyetli internet korsanlarına hizmet edebildiği için kullanıcıların özel bilgilerinin, özel hesaplarının ve kredi kartı bilgileri gibi önemli bilgileri için de bir tehdit oluşturmaktadır. Bu nedenlerden dolayı, spam e-postaların ve kısa mesajların tespit edilerek kullanıcılara sunulmadan engellenmesi önemlidir. Günümüzde bu tür mesajların tespit edilmesi çeşitli makine öğrenme algoritmalarıyla başarılı bir şekilde sağlanabilmektedir. Uygulanan bu tür yöntemlerin her geçen gün daha doğru bir şekilde spam mesajları tespit edilmesi için bu tür çalışmalar devam etmektedir.

Literatür incelendiğinde ise Türkçe bir veri seti üzerinde yapılan spam tespiti çalışmalarının sayısının çok az olduğu görülmüştür. Bu eksikliğin giderilmesi için yapılan bu çalışmada, Türkçe bir veri seti üzerinde farklı özellik çıkarım yöntemleri ve farklı sınıflayıcılar kullanmak suretiyle ortaya çıkarılan modellerin karşılaştırılması ve en başarılı modelin belirlenmesi hedeflenmiştir. Bu hedef doğrultusunda Özdemir ve arkadaşları (2018) tarafından oluşturulan Türkçe veri seti kullanılarak, ilgili veri seti üzerinde daha önce yapılan çalışmalara göre daha başarılı modeller ortaya koyarak bu modellerin karşılaştırılması amaçlanmıştır.

II. LİTERATÜR TARAMASI

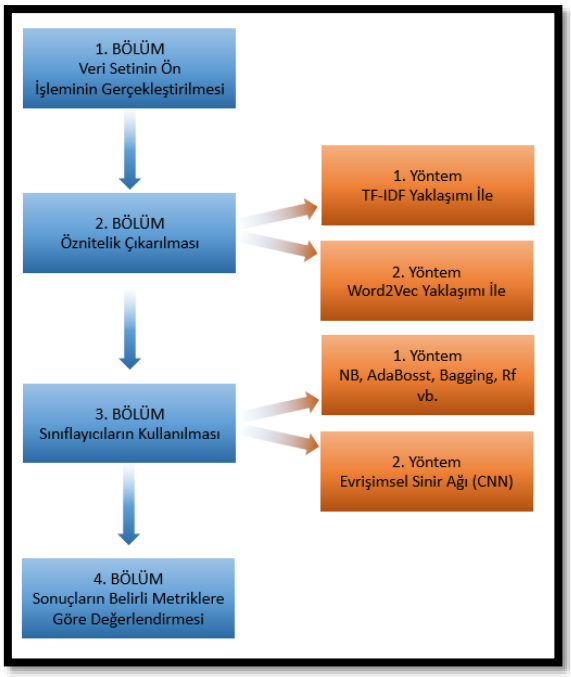
Spam tespiti konusunda bugüne kadar yapılmış çok sayıda çalışma yapılmıştır. Bu çalışmalardan bir kısmı aşağıda özetlenmiştir.

Akçetin ve Çelik [1], spam e-posta tespiti için karar ağaçlarının performansını incelemişlerdir. Çalışmada kullanılan 12 farklı karar ağacı WEKA makine öğrenemesi yazılımı kullanılarak 10 katlı çapraz doğrulama ile sınıflandırılmıştır. Sonuçlar Roc analizi yöntemi ile değerlendirilmiştir. 12 farklı sınıflayıcının başarı oranı %94.68 ile %91 aralığında değişirken, Rastgele Orman Algoritması (RF) %94.68'lik doğruluk oranı ile en iyi sınıflandırıcı olarak önerilmiştir. Sharaff ve ark. [2], karşılaştırmalı bir çalışma yaparak spam filtrelemede kullanılan J48, Destek Vektör Makineleri (SVM), Bayes Ağı (BN) ve K-En Yakın Komşuluk (KNN) algoritmalarının performanslarını incelemişlerdir. Spamların doğru tespitinde J48 algoritması %93.31'lik bir başarı oranı yakalarken, BN %93.08, SVM %88.39 ve KNN ise %89.24'lük başarı oranları elde edebilmişlerdir. Bozkır ve ark. [3], N-gram yöntemi kullanarak bir elektronik posta kümesinin özniteliklerini çıkardıktan sonra Naive Bayes (NB) algoritmasını kullanarak spam sınıflandırma çalışması yapmışlardır. Nazlı [4], Makine öğrenmesi tabanlı spam filtreleme yöntemlerinin F1 metriğine göre karşılaştırılmaları üzerinde çalışmıştır. Bu çalışmada NB ve SVM

algoritmalarının F1 metriğine göre spam filtrelemede başarılı oldukları görülmüştür. Shajideen ve Bindu [5], yaptıkları çalışmada spam sınıflandırılmasında kullanılan SVM, NB ve J48 (C4.5) sınıflayıcılarını karşılaştırmışlardır. Yapılan çalışmada SVM algoritması en başarılı sınıflandırmayı gerçekleştirmiştir. Özdemir ve ark. [6], yaptıkları çalışmada elektronik postaların sınıflandırılması için Motif örüntüler yöntemi kullanılarak öznitelik çıkarma işlemi üzerinde durulmuştur. Yazarlar kendi oluşturdukları Türkçe veri setini kullanmışlardır. Bu çalışmada oluşturulan motif örüntüleri J48 algoritmasına göre spam sınıflandırılması yapılabileceği gösterilmiştir. Dada ve Joseph [7], yaptıkları çalışmada elektronik postaların mesaj gövdesi, konu, mesajın boyutu, kelimelerin tekrar sayısı, alıcının yaşı-cinsiyeti-ülkesi, mesaj içeriğinden kelime çantasının (bag of words) oluşturarak, RF algoritmasına göre spam sınıflandırması yapmışlardır. Sonuçları F1 metriğine göre değerlendirmişlerdir. Aydoğan ve Karcı [8] yaptıkları çalışmada Apache Spark üzerinde makine öğrenmesi kütüphanelerinden biri olan NB yöntemi kullanarak bir spam elektronik posta sınıflandırması uygulaması geliştirmiştir. Yapılan bu çalışma büyük verilerin işlenmesinde Apache Spark'ın etkili ve yeterince hızlı olduğunu, NB yöntemi kullanılarak yapılan sınıflandırma çalışmasının da başarılı olduğunu göstermiştir. Dewangan ve Gupta [9], spam elektronik postaları, içeriklerinden etkili bir şekilde tanımlayabilen bir spam tespit sistemi geliştirmek için SVM algoritmasından faydalanmıştır. Yapılan bu çalışmada %98'lik bir oranla spam e-postalar doğru şekilde tespit edilmiştir. Gupta ve ark. [10], Kagglebenchmark veri seti üzerindeki çalışmalarında, spam filtreleme konusunda en etkili algoritmanın NB olduğunu vurgulamışlardır. Yaptıkları çalışmada %95.56 'lık doğruluk oranı yakalamışlardır. Çalışma sonucunda spam filtrelemede geleneksel yaklaşımlara göre NB yaklaşımının daha başarılı olduğu ortaya konulmuştur. Popovac ve ark. [11], yaptıkları çalışmada Tiago'nun veri seti olarak bilinen spam ve spam olmayan kısa mesajların bulunduğu veri seti üzerinde sınıflandırma yapabilmek için Evrimli Sinir Ağı modelini önermişlerdir. Önerilen model sonucunda %98.4'lük bir spam tespit başarı oranı sağlanmıştır. Deniz ve ark. [12], yaptıkları çalışmada Türkçe elektronik postalar için makine öğrenmesi teknikleri ile sınıflandırma uygulaması geliştirerek spam olan Türkçe elektronik postaları tespit etmeyi amaçlamışlardır. Bu amaç doğrultusunda, çalışmalarında Turkish Email veri setini eğitim ve test için kullanmışlardır. Bu veri setindeki elektronik postaları Doc2Vec kütüphanesine ait algoritmalar kullanarak sayısallaştırdıktan sonra özellik çıkarımı yapılarak çeşitli sınıflandırma algoritmaları ile sınıflandırmışlardır. Oluşturulan modele göre sonuçta en başarılı sınıflayıcının SVM olduğu ifade edilmiştir. Krause ve ark. [13], yaptıkları çalışmada elektronik postaların sadece başlıklarından elde edilen meta veri özelliklerine bakılarak spam algılama yaklaşımı önermişlerdir. Elektronik postaların meta verilerinin içerdiği tüm ek bilgileri kullanarak spam elektronik postaları sınıflandıran yaygın bir kullanım yöntemi olmadığı için Krause ve arkadaşları böyle bir çalışma yapma ihtiyacı duymuşlardır. Krause ve arkadaşları, oluşturdukları modeli, CDMC2010 veri seti kullanarak SVM, Karar Ağaçları (DT) ve Adaboost gibi makine öğrenmesi algoritmalarıyla çalıştırmışlardır. Sonuçta da hem SVM hem de Adaboost algoritmalarıyla başarılı spam sınıflandırma sonuçları elde etmişlerdir (%99). Kumar ve ark. [14], Spam elektronik postaların sınıflandırılmasında grup halindeki öğrenme metotlarından yararlanarak AdaBoost algoritmasıyla başarılı bir spam sınıflandırması çalışması gerçekleştirmişlerdir. Eryılmaz ve ark. [15], literatürdeki çalışmalarda Türkçe veri setiyle yapılan yeterince çalışma bulunmadığını kaydederek, Türkçe e-postaların oluşturduğu veri seti üzerinde TD-IDF yaklaşımıyla öznitelik çıkardıktan sonra Sıralı Minimum Optimizasyon (SMO) algoritmasıyla %90 başarı oranıyla spam postaları tespit edebilmişlerdir. Eryılmaz ve Kılıç [16], yaptıkları çalışmada spam tespitinde kullanılan yöntemleri inceleyerek, klasik makine öğrenmesi algoritmalarının bu alandaki başarısını vurgulamışlardır. Derin öğrenme temelli yaklaşımlar kullanılarak temel makine öğrenmesi algoritmalarında elde edilen başarı oranının ve oluşturulan modellerdeki performansın artacağı ortaya konulmuştur. Ahi ve Soğukpınar [17], kullanıcıların hassas bilgilerini ele geçirmeye yönelik kimlik avı saldırılarına karşı derin öğrenme modellerini kullanan bir yöntem önermişlerdir. Yazarlar, yaptıkları çalışmada spam e-postaları başlık ve gövde olarak ayırarak, bu kısımlar için özellikler matrisleri oluşturmuşlardır. Word2Vec yaklaşımının da kullanıldığı bu çalışmada, yazarlar %96 oranında başarılı bir şekilde bu tip saldırıları tespit etmişlerdir. Yağanoğlu ve Irmak [18], yaptıkları çalışmada İngilizce e-postaları içeren bir veri seti üzerinde makine öğrenmesi yöntemleri kullanarak spam e-postaların ayrıştırılmasını sağlamışlardır. Çalışmada K-En Yakın Komşu, Destek Vektör Makineleri ve Karar Ağaçları gibi bilinen yöntemler kullanılarak %98'lik bir başarı oranı sağlanmıştır.

III.YÖNTEM

Yapılan bu çalışmada aynı veri seti üzerinde iki farklı yöntem uygulanmak suretiyle oluşturulan modeller üzerinden çeşitli karşılaştırmalar yapılmıştır. İki yöntem arasında; öznitelik çıkarma işlemlerinde kullanılan algoritmalar ve kullanılan sınıflayıcılar bakımında farklılıklar mevcuttur. Çalışmanın bölümleri ve yöntemlerin farklılaşmasına ilişkin diyagram Şekil 1'de gösterilmektedir.



Şekil 1. Yöntem Detayları

A. Veri Seti

Çalışmada, Kaggle platformunda erişime açık olan "Turkish Spam Dataset" isimli Türkçe veri seti kullanılmıştır [19]. Kullanılan veri setinde 330 spam ve 496 spam olmayan elektronik posta örneği bulunmaktadır. Veri seti içeriği ".csv" formatında olup iki sütundan oluşmaktadır. İlk sütunda elektronik posta içeriği, diğer sütunda ise "ham" veya "spam" şeklinde, o örneğin spam mı yoksa normal mi olduğunu belirten etiket bulunmaktadır.

Çalışmada kullanılan veri seti, daha önce Özdemir ve arkadaşlarının (2018) çalışmasında da kullanılmıştır [6]. Yazarlar kendi oluşturdukları bu veri setini, Motif Örüntüler (MÖ) yöntemi ve farklı algoritmalar ile sınıflandırarak %90'lık bir başarı oranı elde etmişlerdir. Literatürde, ilgili çalışma haricinde bahsedilen veri setini kullanan farklı çalışmalara rastlanamamıştır.

Çalışmada ilgili veri setinin %80'i eğitim, %20'si ise test verileri olarak kullanılmıştır.

B. Ön İşleme

Spam tespitinde başarılı bir model oluşturmanın ilk ve temel adımı elde bulunan verilen ön işlemden geçirilmesidir. Ön işlem bünyesinde, veri setini oluşturan tüm elektronik posta kayıtlarında aşağıdaki işlemler gerçekleştirilmiştir:

- Özel karakterlerin silinmesi,
- Noktalama işaretlerinin temizlenmesi,
- Harflerin tamamının küçük harflere dönüştürülmesi,
- Durak kelimelerin çıkarılması,
- Kelime köklerinin bulunması.

Veri setindeki özel karakterlerin ve noktalama işaretlerinin silinmesi için uygulamanın geliştirildiği ortam olan python'da "string" kütüphanesi kullanılmıştır. Türkçe' deki "ve, ile, bazı, belki, ayrıca …vb." durak kelimelerin veri setinden çıkarılması için "Natural Language Toolkit (NLTK) - Corpus" modülünün "Stop Words" kütüphanesi kullanılmıştır. Bu kütüphane Türkçe durak kelimeleri de bünyesinde barındırmaktadır. Ön işlem

aşamasının son adımında da elde kalan kelimelerin eklerini atmak suretiyle kökleri bulunur. Bunun için de "snowballstemmer" modülünün "Turkish Stemmer" kütüphanesi kullanılmıştır.

Snowball, özellikle kök bulma amacıyla tasarlanan bir kelime işleme dilidir. Birçok dil için kök bulma algoritmalarının geliştirilmesinde Snowball kullanılmaktadır [20]. Türkçe için Snowball kullanılarak geliştirilen kök bulma algoritmaları "Evren (Kapusuz) Çilden" tarafından yürütülmektedir [21].

Bu aşamada yapılan işlemler elektronik postaların "spam" ya da "normal" olarak sınıflandırılmasındaki başarı oranının artırılmasına yöneliktir. Harflerin standartlaştırılarak hepsinin küçük harfe dönüştürülmesinin nedeni, sınıflandırmada büyük/küçük harf ayrımının sonuca olan etkisini ortadan kaldırmaktır.

C. Öznitelik Çıkartma

Öznitelik çıkartma, bazı kriterlere dayanarak mevcut bilgilere bir dönüştürme işlemi uygulayarak yeni bir öznitelik uzayı oluşturmak iken öznitelik seçme, mevcut öznitelikler arasından, bazı kriterlere dayanarak öznitelik seçme yani o örneği temsil edebilecek en iyi özniteliği seçmektir [22].

Yapılan bu çalışmada, veri setinin boyutunu indirgeyip sınıflandırma için kullanılabilecek en anlamı özellikleri ortaya çıkarıp oluşturulan modelin sınıflandırmadaki başarısını artırmak amaçlı öznitelik çıkarım işlemi uygulanmıştır. Literatürde spam tespiti çalışmalarında kullanılan öznitelik çıkartma işlemleri incelendiğinde "Word2Vec" ve "TF-IDF" yaklaşımlarının ön plana çıkarıldığı görülmüştür. Yapılan bu çalışma kapsamında, her iki öznitelik çıkarım yaklaşımları da oluşturulan farklı modellerde kullanılmıştır.

TF-IDF: Bir terimin doküman içerisindeki önemini gösteren istatistiki yöntem ile hesaplanmış ağırlık faktörüdür. Terim Sıklığı ve Ters Doküman Sıklığı kavramlarını bünyesinde barındırır. Terim sıklığı; seçili terimin, metin içinde bulunan toplam terimler sayısına bölümünü ifade eder. Ters Doküman Sıklığı ise metinlerin kaç tanesinde aranılan terimin bulunduğunu gösterir. Toplam metin sayısının, terimi içeren metin sayısına bölümünün logaritması ile hesaplanır. Bu iki değerin çarpımı ile de TF-IDF değeri elde edilmiş olur.

Word2Vec: kelimeleri vektör uzayında ifade etmeye çalışan denetimsiz öğrenmeye dayalı ve tahmin temelli bir modeldir. Bu yaklaşımın özünde, kelimeler arasındaki uzaklığın vektörel olarak hesaplanması yatar. Daha açık bir ifadeyle birbirlerine yakın kelimeleri ortaya çıkarır. CBOW (ContinuousBag of Words) ve Skip Gram alt algoritmalarını kullanır.

D. Sınıflandırma

Çalışma, öznitelik çıkarma işleminden itibaren iki farklı uygulama şeklinde ilerlemiştir. İlk uygulamada TF-IDF yaklaşımıyla çıkarılan özniteliklere göre Gaussian NaiveBayes (GNB), RF, AdaBoost, Gradient Boosting ve Bagging algoritmalarıyla sınıflandırma işlemleri gerçekleştirilmiştir. Ayrıca her sınıflandırma, 5 katlı çapraz geçerlilik testine göre doğrulanmıştır.

İkinci uygulamada ise Word2Vec yaklaşımıyla çıkarılan özniteliklere göre Keras Kütüphanesi yardımıyla oluşturulan Evrişimsel Sinir Ağı (CNN) modeli ile sınıflandırılması gerçekleştirilmiştir. 5 devirli bir eğitim iterasyonu sonrasında yine ilk uygulamadaki gibi bu uygulamada da 5 katlı çapraz geçerlilik testi kullanılmıştır. Oluşturulan model Şekil 2'de görülmektedir.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, None, 500)	19842500
dropout (Dropout)	(None, None, 500)	0
conv1d (Conv1D)	(None, None, 50)	75050
conv1d_1 (Conv1D)	(None, None, 50)	7550
max_pooling1d (MaxPooling1D)	(None, None, 50)	0
dropout_1 (Dropout)	(None, None, 50)	0
conv1d_2 (Conv1D)	(None, None, 100)	15100
conv1d_3 (Conv1D)	(None, None, 100)	30100
max_pooling1d_1 (MaxPooling1	(None, None, 100)	0
dropout_2 (Dropout)	(None, None, 100)	0
conv1d_4 (Conv1D)	(None, None, 200)	60200
conv1d_5 (Conv1D)	(None, None, 200)	120200
global_max_pooling1d (Global	(None, 200)	0
dropout_3 (Dropout)	(None, 200)	0
dense (Dense)	(None, 200)	40200
activation (Activation)	(None, 200)	0
dropout_4 (Dropout)	(None, 200)	0
dense_1 (Dense)	(None, 2)	402
activation_1 (Activation)	(None, 2)	0

Total params: 20,191,302
Trainable params: 20,191,302
Non-trainable params: 0

Şekil 2. CNN Modeli

IV. BULGULAR

İlk uygulamada TF-IDF yaklaşımıyla öznitelik çıkarım işlemi gerçekleştirilmiş ve GNB, Gradient Boosting, RF, Adaboost ve Bagging algoritmaları ile sınıflandırma işlemi tamamlanmıştır. Yapılan sınıflandırmaların başarılarını ölçümünde Karmaşıklık Matrisi, doğruluk, hassasiyet, geri çekilme ve F1 metrikleri kullanılmıştır. Bu metriklerden doğruluk puanı, doğru sınıflandırılan verilerin toplam verilere oranını ifade eder. Hassasiyet metriği doğru tahmin edilen pozitif gözlem sayısının, pozitif olarak nitelendirilen tüm gözlem sayısına bölünmesiyle bulunur. Geri çekilme metriği ise doğru tahmin edilen pozitif gözlem sayısının toplam doğru tahmin edilmesi gereken gözlem sayısına bölünmesiyle elde edilir. F1 metriği ise hassasiyet ve geri çekilme metriklerinin harmonik ortalamasını ifade eder. Bu metriklere göre ortaya çıkan sonuçlar Tablo 1'de gösterilmektedir.

Tablo 1. Birinci sınıflandırma uygulamasına ait sonuçlar

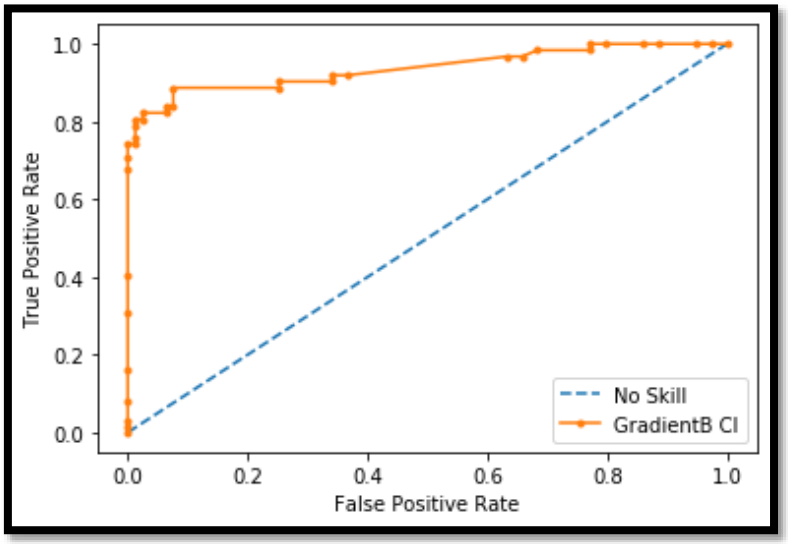
Sınıflayıcı	Karmaşıklık Matrisi	Doğruluk (AccuracyScore)	Kesinlik (Precision Score)	Duyarlılık (RecallScore)	F1 Score
GNB	[79 0 16 46]	0,89	0,83	1.0	0,90
GradientBoosting	[78 1 13 49]	0,90	0,86	0,99	0,92
RF	[79 0 16 46]	0,89	0,83	1.0	0,91
AdaBoost	[72 7 11 51]	0,87	0,87	0,91	0,89
Bagging	[79 0 17 45]	0,88	0,82	1.0	0,90

İlk uygulama için yapılan sınıflandırmada genel olarak en iyi sonuçların elde edildiği GradientBoosting algoritmasıyla sınıflandırılması yapılan model için çapraş doğrulama işlemi uygulanmış ve ROC analizi yapıldıktan sonra ortaya çıkan sonuçlar Tablo 2'de gösterilmiştir.

Tablo 2. Birinci sınıflandırma uygulamasında en iyi sonuç alınan sınıflandırıcıya ait çapraz doğrulama ve ROC değerleri

Sınıflayıcı	Maks. Doğruluk Değeri	Min. Doğruluk Değeri	Ortalama Doğruluk Değeri	Standart Sapma Değeri	ROC AUC Değeri
GradientBoosting	0,95	0,86	0,91	0,03	0,94

Birinci uygulamada Gradient Boosting algoritmasıyla yapılan sınıflandırmaya ait ROC Eğrisi Şekil 3'de gösterilmektedir. ROC (Receiver Operating Characteristic) analizi, yapılan sınıflandırmanın doğruluğu hakkında bilgi veren önemli bir metriktir. Temel olarak gerçek doğru oranı (TPR)'nın yanlış doğru oranına (FPR) bölünmesiyle elde edilir.



Şekil 3. Gradient Boosting için ROC eğrisi

İkinci uygulamada ise Word2Vec yaklaşımıyla öznitelik çıkarım işlemi gerçekleştirilmiş ve Şekil 2'de gösterilen CNN modeli ile sınıflandırma işlemi tamamlanmıştır. Ortaya çıkan sonuçlar Tablo 3'de gösterilmektedir.

Tablo 3.İkinci sınıflandırma uygulamasına ait sonuçlar

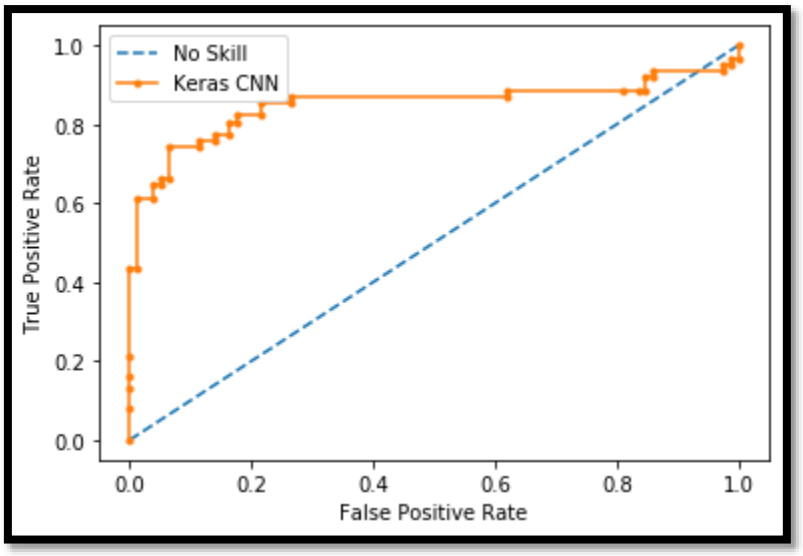
Sınıflayıcı	Doğruluk (AccuracyScore)	Kesinlik (Precision Score)	Duyarlılık (RecallScore)	F1 Score
CNN	0,94	0,91	0,89	0,93

İkinci uygulamanın sonunda yapılan çapraz doğrulama işlemi ve ROC analizine ilişkin sonuçlar Tablo 4'te gösterilmektedir.

Tablo 4.ikinci sınıflandırma uygulamasına ait çapraz doğrulama ve ROC değerleri

Sınıflayıcı	Maks. Doğruluk Değeri	Min. Doğruluk Değeri	Ortalama Doğruluk Değeri	Standart Sapma Değeri	ROC AUC Değeri
CNN	0,91	0,82	0,87	4	0,85

Evrişimsel Sinir Ağı (CNN) ile oluşturulan ikinci uygulamaya ait ROC Eğrisi Şekil 4'te gösterilmektedir.



Şekil 4. CNN için ROC Eğrisi

V. TARTIŞMA

Elde edilen bulgular incelendiğinde, TF-IDF yaklaşımıyla oluşturulan modelde kullanılan GNB, Gradient Boosting, RF, Adaboost, Bagging sınıflayıcıları içerisinde en başarılı sonuçların, F1 metriğinde %92 oranında ve doğruluk metriğinde %90 oranında, Gradient Boosting algoritmasıyla elde edildiği gözlemlenmiştir. Bu doğrultuda Gradient Boositng algoritmasına yapılan çapraz doğrulama sonucu ortalama doğruluk %91 ve ROC analizi sonucu %94'lük bir başarı oranı yakalanmıştır.

İkinci uygulamada ise CNN tabanlı bir model ile oluşturulan spam tespit uygulamasının doğruluk, hassasiyet ve F1metriklerine göre sonuçları incelendiğinde %90 üzerinde başarı sağlandığı görülmüştür. Çapraz doğrulama işlemi sonucunda ortalama doğruluk değeri %87 ve ROC analiz sonucu ise %85 olarak ölçülmüştür.

Sonuçlara göre, TF-IDF yaklaşımıyla özniteliklerin çıkarılarak Gradient Boosting algoritmasıyla yapılan spam tespiti uygulamasının çapraz doğrulama ve ROC analiz sonuçlarının, word2vec yaklaşımıyla özniteliklerin çıkarılarak CNN modeliyle spam tespitinin yapıldığı uygulamaya göre daha iyi olduğu görülmüştür. Her ne kadar CNN modelinde doğruluk ve F1 değerlerinin, Gradient Boosting'e göre birkaç puanlık fazlalığı olsa da Gradient Boosting'in daha kararlı olarak spam tespitinde etkili olduğu görülmüştür.

VI.SONUÇ VE ÖNERİLER

Yapılan bu çalışmada Türkçe elektronik posta veri seti üzerinde iki farklı öznitelik çıkarma yaklaşımları ile farklı sınıflayıcıların ve Evrişimsel Sinir Ağının performansları karşılaştırılarak en iyi şekilde spam tespiti yapabilecek bir model ortaya çıkarılmaya çalışılmıştır. Literatürdeki Türkçe elektronik posta veri seti ile yapılan çalışmalar konusundaki yetersizlik göz önüne alındığında, yapılan bu çalışma ile söz konusu eksiklik giderilmeye çalışılmıştır.

Çalışma sonucunda ortaya çıkan sonuçlara göre TF-IDF yaklaşımı ile Gradient Boosting algoritmasıyla oluşturulan uygulama, Word2Vec yaklaşımı ve CNN modelinde oluşturulan uygulamaya göre bir nebze daha başarılı olarak değerlendirilmiştir.

Bu konuda çok daha başarılı sonuçlar elde edebilmek için Türkçe e-posta veri setleri genişletilmelidir. Literatürde çok sayıda İngilizce e-postalar ile hazırlanan spam veri seti bulunmasına rağmen Türkçe e-postaların oluşturduğu veri seti oldukça az sayıdadır. Az sayıda bulunan bu veri setinin içinde de veri miktarı çok azdır. Daha başarılı modellerin oluşturulabilmesi, oluşturulan modellerin daha iyi eğitilmesi öncelikle daha nitelikli ve daha çok sayıda Türkçe veri setinin oluşturulmasına bağlıdır.

Sonuç olarak gelecekte bu konuda çalışacak olan araştırmacıların oluşturacağı dengeli ve kaliteli Türkçe veri setleri ile spam elektronik postaların tespiti benzer yaklaşımlarla çok daha başarılı olacaktır. Veri setindeki

yetersizliğe rağmen bu çalışmada oluşturulan modellerin yaklaşık %90'lık başarı ortalamaları bu konudaki potansiyeli ortaya koymaktadır.

KAYNAKLAR

[1] Akçetin, E. & Çelik, U. (2015). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması. *İnternet Uygulamaları ve Yönetimi Dergisi, 5(2), 43-56.*

[2] Sharaff A., Nagwani N. K. & Dhadse A. (2016). *Comparative Study of Classification Algorithms for Spam Email Detection. Emerging Research in Computing, Information, Communication and Applications.* Springer, New Delhi, India.

[3] Bozkır, A. S., Şahin, E., Aydos, M., Akçapınar Sezer, E. & Orhan, F. (2017). Spam E-Mail ClassificationbyUtilizing N-Gram Fatures of HyperlinkTexts. *The 11th IEEE International Conference AICT2017.* 20-22 September, Moscow, Russia, 1-5.

[4] Nazlı, N. (2018). *Analysis of Machine Learning-Based Spam Fİlter Techniques.* Yüksek Lisans Tezi, Çankaya Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

[5] Shajideen, N. M. & Bindu, V. (2018). Spam Filtering: A Comparison Between Different Machine Learning Classifiers. *Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).* 29-31 March, Coimbatore, India ,1919-1922.

[6] Özdemir, C., Kaya, Y. & Minaz, M. R. (2018). Motif Örüntüler Yöntemi ile Spam E-Postaların Filtrelenmesi. *Uluslararası Mühendislik ve Teknoloji Sempozyumu (IETS'18).* 3-5 Mayıs, Batman, 755.

[7] Dada, E.G. & Joseph, S.B. (2018). Random Forests Machine Learning Technique for Email Spam Filtering. *University of Maiduguri Seminar Series,9(1).*

[8] Aydoğan, M. & Karcı, A. (2018). Apache Spark ile Naïve Bayes Yöntemi Kullanarak Spam Mail Tespiti. *International Conference on Artificial Intelligence and Data Processing (IDAP 2018).* 28-30 Eylül, Malatya, 1-6.

[9] Dewangan, D. K. & Gupta, P. (2018). Email Spam Classification Using Support Vector Machine Algorithm. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 6(6), 6-10.*

[10] Gupta, A., Mohan, K. M. & Shidnal, S. (2018). Spam Filter using Naïve Bayesian Technique. *International Journal of Computational Engineering Research (IJCER),* 8(6), 26-32.

[11] Popovac, M., Karanovic, M., Sladojevic, S., Arsenovic, M. & Anderla, A. (2018). Convolutional Neural Network Based SMS Spam Detection. *26th Telecommunications forum TELFOR 2018. Belgrade, Serbia.*

[12] Deniz, E., Erbay, H. & Coşar M. (2019). Türkçe E-Postaların Doc2Vec ile Sınıflandırılması. *1st International Informatics and Software Engineering Conference (UBMYK).* 6-7 Kasım, Ankara,1-4.

[13] Krause, T., Uetz, R. & Kretschmann, T. (2019). Recognizing Email Spam from Meta Data Only. *IEEE Conference on Communications and Network Security 2019.* 10-12 June, Washington DC, USA, 178-186.

[14] Kumar, N., Sonowal, S. & Nishant. (2020). Email Spam Detection Using Machine Learn Algorithms. *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (ICIRCA 2020).* 15-17 July, Coimbatore, India, 108-113.

[15] Eryılmaz, E. E., Şahin, D. Ö. & Kılıç, E. (2020). Türkçe Yaramaz E-postaların Farklı Öznitelik Seçim Yöntemleri Kullanılarak Makine Öğrenmesi Algoritmaları İle Tespit Edilmesi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 13(2),* 77.

[16] Eryılmaz, E. E. & Kılıç, E. (2020). İstenmeyen Epostaların Tespiti için Kullanılan Yöntemlerin İncelenmesi. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 11 (3) , 977-987.*

[17] Ahi, Ş. & Soğukpınar, İ. (2020). Derin Öğrenme Modelleri ile Kimlik Avı E-posta Tespiti. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 13 (2), 17-31.*

[18] Yağanoğlu, M. & Irmak, E. (2021). Separation of Incoming E-Mails Through Artificial Intelligence Techniques. *Avrupa Bilim ve Teknoloji Dergisi, (21), 690-696.*

[19] Özdemir, C. (2019).Turkish Spam Dataset. Kaggle. https://www.kaggle.com/cuneytdemir/turkish-spam-dataset. (18.12.2020)

[20] Çilden, E. (2006) Stemming Turkish Words Using Snowball. http://snowball.tartarus.org/algorithms/turkish/stemmer.html, (11.04.2021).

[21] Yüksel, M. E., Turna, Ö. C. & Ertürk, M. A. (2009). Bilgiye Erişim Sistemlerinde Veri Arama ve Eşleştirme. *XII. Akademik Bilişim Konferansı Bildirileri Kitapçığı.* 10-12 Şubat, Muğla.

[22] Küçüksille, E. U. & Ateş, N. (2013). Destek Vektör Makineleri ile Yaramaz Elektronik Postaların Filtrelenmesi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 6(1), 81-87.*

[23] Agarwal, K. & Kumar, T. (2018). Email Spam Detection using integrated approach of Naïve Bayes and Particle Swarm Optimization. *Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018).* 14-15 June, Madurai, India, 685-690.

[24] Anihta, P. U., Guru Rao, C.R. & Babu, S. (2017). Email Spam Classification using Neighbor Probablity based Naive Bayes Algorithm. *2017 7th International Conference on Communication Systems and Network Technologies (CSNT).* 11-13 November, Nagpur, India, 350-355.

[25] Annareddy, S. & Tammina, S. (2019). A Comparative Study of Deep Learning Methods for Spam Detection. *Proceedings of the Third International Conference on I-SMAC (IoT in Social, Mobile, AnalyticsandCloud) (I-SMAC 2019).* 12-14 December, Palladam, India, 66-72.

[26] Dewangan, D. K. & Gupta, P. (2018). Email Spam Classification Using Support Vector Machine Algorithm. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 6(6), 6-10.*

[27] Harisinghaney, A., Dixit, A., Gupta, S. & Arora A. (2014). Text and Image Based Spam Email Classification using KNN, Naive Bayes and Reverse DBSCAN Algorithm. *2014 International Conference on Reliability, Optimization and Information Technology - ICROIT 2014.* 6-8 February, India, 153-155.

[28] Huang, T. (2019). A CNN Model for SMS Spam Detection. *2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE).* 25-27 October, Hohhot, China, 851.

[29] Liu, G. & Yang, F. (2012). The Application of Data Mining in the Classification of Spam Messages. *2012 International Conference on Computer Scienceand Information Processing (CSIP).* 24-26 August, Shaanxi, China, 1315-1317.

[30] Octaviani, N. L., Rachmawanto, E. K., Setiadi, I. M. & Sari, C. A. (2020). Comparison of Multinomial Naive Bayes Classifier, Support Vector Machine, and Recurrent Neural Network to Classify Email Spams. *2020 International Seminar on Application for Technology of Information and Communication (iSemantic).* 19-20 September, Semerang, Indonesia, 17-21.

[31] Oskuie, M. D. & Razavi, S. N. (2014). A Survey of Web Spam Detection Techniques. *International Journal of Computer Applications Technology and Research, 3(3), 180-185.*

[32] Örnek, Ö. (2019). Orange 3 İle Türkçe ve İngilizce SMS Mesajlarında Spam Tespiti. *ESTUDAM Bilişim Dergisi, 1(1), 1-4.*

[33] Shrivastava, A. & Dubey, R. (2018). Classification of Spam Mail Using Different Machine Learning Algorithms. *2018 International Conference on Advanced Computation and Telecommunication (ICACAT).* 28-29 December, Bhopal, India, 1-10.