Destek Vektör Makineleri ve Multinomial Naive Bayes Sınıflandırma Algoritmaları ile İstenmeyen E-postaların(Spam) Tespiti ve Performans Karşılaştırması

Fatma OĞUZ
Fırat Üniversitesi, Yazılım Mühendisliği Bölümü
Elazığ, Türkiye
15290043@firat.edu.tr

Özet*—*E-posta; günümüzde sıkça kullanılan, çağın vazgeçilmez iletişim araçlarından biridir. Kişisel iletişimler, iş odaklı aktiviteler, pazarlama ve reklam gibi amaçlarla kullanılan e-postaların hızla artması, bazı sorunları da beraberinde getirmiştir. Bu sorunlardan en önemlisi alıcının isteği dışında gönderilen, reklam ya da güvenlik tehdidi amacıyla oluşturulmuş, istenmeyen e-postalardır. Bilgi güvenliği açısından önemli tehditler oluşturan bu e-postaların günden güne artması bu probleme çözüm olabilecek yöntemler geliştirilmesine neden olmuştur. Spam e-posta tespitinde kullanılan yöntemleri karşılaştıran literatürde birçok çalışma mevcuttur. Buna rağmen, çalışmaların çoğunda farklı veri kümeleri kullanıldığından hiçbir yöntem birbirleriyle karşılaştırılabilir nitelikte değildir. Bu çalışmanın amacı, spam e-postaların Destek Vektör Makineleri (Support Vector Machines-SVM) ve Multinomial Naive Bayes (MNB) yöntemleri kullanılarak karşılaştırmalı değerlendirmesini yapmaktır. Çalışma sonucunda SVM yönteminin başarımının %99.5436, MNB yönteminin başarımının ise %80.4173 olup SVM yönteminin daha iyi sonuç verdiği gözlenmiştir.

Anahtar Kelimeler*—*E-posta, Spam, Destek Vektör Makineleri, Multinomial Naive Bayes.

Abstract*— *Email; is one of the indispensable communication tools of the era. The rapid increase in the number of e-mails used for personal communications, business-oriented activities, marketing and advertising purposes has brought some problems. That has caused to the use of email for malicious purposes. Malicious emails, often called spam emails, are usually sent for advertising or security threatening purposes. Because of the increasing number of spam e-mails day by day which are important threats for information security new methods have been developed for the solution of this problem. There are many studies comparing methods that are used to detect spam in the literature. However, as different data sets were used in most of the studies, none of them are comparable with each other. The object of this study is to make a comparative evaluation of Support Vector Machines-(SVM) and Naive Bayes (NB) methods to recognize unsolicited e-mails from regular e-mails. As a result of the study, the performance of SVM method was % 99.5436, and the performance of NB method was% 80.4173, so it was shown that SVM method came up with better.

Key Words*—*E-mail, Spam, Support Vector Machines, Multinomial Naive Bayes.

1.GİRİŞ

E-posta, rahat ve kolay erişilebilir olmasından ötürü günümüzde sıkça kullanılan iletişim araçlarının başında gelmektedir (Ravikumar & Gandhimathi, 2017).

Başta sadece metin gönderme amacıyla kullanılan e-postalar, günümüzde çeşitli medya öğeleri ve çalışabilir program uygulamalarının iletimini de sağlamaktadır. İnternetin yaygınlaşması ve e-posta yoluyla daha rahat ve ucuz iletişim kurulması, e-postayı kullanan kullanıcı sayısını yüksek oranda arttırmıştır (Çalış, Gözdağı, & Yıldız, 2013). E-postanın gelişen ve sık kullanılan bir iletişim aracı haline gelmesi bazı sorunları da beraberinde getirmiştir. Spam e-posta olarak adlandırılan bu sorun, bilinmeyen veya güvenilmeyen adreslerden, kişinin isteği dışında, reklam ya da güvenlik tehdidi oluşturma amaçlı gönderilen e-postalardır (Şahin & Orhan, 2018). Spam e-postalar kötü amaçlı kullanıcılar tarafından, e-posta hizmetinin popülerliğinden ve düşük maliyetli olmasından yararlanılarak gönderilmektedir. Spam e-posta gönderimi için gerekli olan tek şey, hedef adreslerin listesidir. Bu e-postaları gönderenler, e-posta adreslerini Usenet postalarından, bilinen alanlarda kullanılan ortak adların tahmin edilmesinden, DNS listelerinden veya web sayfalarından toplama gibi çeşitli şekillerde elde etmektedir.

Yapılan araştırmalara göre 2002 yılında gönderilen günlük spam e-posta sayısının ortalama 2.4 milyar olduğu görülmektedir. 2010 yılında ise 300 milyara yükselen ve tüm e-postaların yaklaşık %90'ını oluşturan bu e-postalar, büyük bir sorun haline gelmeye başlamıştır (Almeida & Yamakami, 2012). Bu soruna çözüm olarak spam e-postaların tespitini sağlayan algoritmalar geliştirilmiştir (Goodman, Heckerman, & Rounthwaite, 2005).

Literatürde spam e-postaların sınıflandırılmasında çeşitli yöntemler kullanıldığı görülmektedir. Veri madenciliği yöntemlerinden Yapay Sinir Ağları, Karınca Koloni Optimizasyonu ve Kural Tabanlı Algoritmalar gibi yöntemlerle bu e-postaların tespiti yapılmıştır (Drucker, Wu, & Vapnik, 1999). Çalışmalarda en çok kullanılan makine öğrenme teknikleri Naive Bayes, Destek Vektör Makineleri ve Yapay Sinir Ağları olmuştur (Duda & Hart, 1973).

Bu çalışmanın amacı, farklı veri kümeleri üzerinde uygulanan yöntemleri ortak bir veri kümesi üzerinden eğitmek ve yöntemlerin başarı ve hatalarını hesaba katarak hangi yöntemin daha iyi sonuç ürettiğini saptamaktır. Kullanılacak yöntemler ise, Naive Bayes ile Destek Vektör Makineleridir (Eryiğit, Tantuğ, & Adalı, 2005).

2. VERİ TOPLAMA

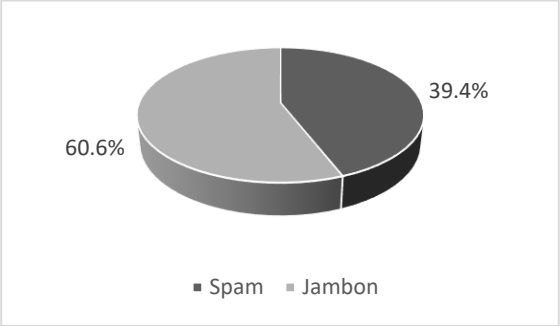
Spam e-postaların tespiti için Kaliforniya Üniversitesi'nin Spam7 adlı makine öğrenme veri seti kullanılmıştır. 1999 yılında Hewlett-Packard Laboratuvarlarından elde edilen bu veri seti 7 farklı özellikte 4601 e-postadan oluşmaktadır. Veri setinin son sütunu, e-postanın spam olup olmadığını(jambon) göstermektedir.

Tablo 2.1. E-postaların nitelik tanımı

Nitelik	**Aralık**
E-postadaki büyük harfle başlayan toplam kelime sayısı.	—
$ karakterinin kullanılma sıklığı.	[0,1]
'para' kelimesinin kullanılma sıklığı.	[0,1]
'000' dizisinin kullanılma sıklığı.	[0,1]
'make' kelimesinin kullanılma sıklığı.	[0,1]
E-postanın spam kabul edilip(y) edilmediği(n).	{y,n}

Veri setinde yer alan tüm özellikler Tablo 2.1'de gösterilmektedir. Burada belirtilen 7 farklı özellik, girdi değerleri ve e-postanın spam olup olmadığını belirten çıktı değeri olarak etiketlenmiştir. y, e-postanın spam olduğunu gösterirken, n değeri jambon olduğunu göstermektedir. Veri setindeki 4601 örnekten

1813'ü spam, 2788'i ise jambon e-postadır. Sınıfların veriler üzerindeki dağılımı Şekil 2.1'de gösterilmiştir.



Şekil 2.1. Sınıfların Dağılımı

3. ALGORİTMALAR VE YÖNTEM

Sınıflandırma yöntemi ile spam e-postaların tahmin edilmesi için iki farklı algoritma seçilmiştir. Bu algoritmalar Multinomial Naive Bayes ile Destek Vektör Makineleridir. Belirtilen algoritmaların açıklaması aşağıda detaylı bir şekilde yapılmıştır (Altay & Ulaş, 2018).

3.1. Destek Vektör Makineleri (SVM)

SVM; Cortes ve Vapnik tarafından önerilen istatiksel teori üzerine kurulu bir örüntü tanıma tekniğidir (Drucker, Wu, & Vapnik, 1999). Makine öğrenim topluluğu içinde iyi bir genelleme performansının olması ve yüksek boyutlu veriyi çekirdek kullanarak ele alması nedeniyle çok sık kullanılmaktadır (Lai & Tsai, 2004). İstatiksel Öğrenme Teorisi ile Yapısal Risk Azaltma prensiplerine dayanan SVM, çok çeşitli alanlarda başarıyla uygulanmıştır (Burges, 1998; Cristianini & Shawe-Taylor, 2000; Drucker, Wu, & Vapnik, 1999; Lee, 2008; Pontil & Verri, 1998; Ying, Lin, Lee, & Lin, 2010). SVM'nin örüntü sınıflandırılmasında kullanılmasındaki temel amaç, pozitif ve negatif ağırlıklar arasında maksimum marjı veren optimum ayırıcı hiper düzlemi bulmaktır (Lai, 2007). Verileri sınıflandırabilecek birçok hiper düzlem vardır ve en iyi hiper düzlem, iki sınıf arasındaki en büyük ayrımı veya marjı temsil edendir (Almeida & Yamakami, 2010).

SVM ile metin sınıflandırılmasında her kelime bir özelliği gösterir. Sistem tarafından tanınan kelimelerin listesine, diğer bir deyişle sözlüğe, özellik vektörü denir. Bir e-posta mesajı için, sözlükteki her bir kelimeye karşılık gelen x_i elemanlarını içeren özellik vektörü x ile temsil edilebilir (Woitaszek, Shaaban, & Czernikowski, 2003). Bir e-posta mesajı, mesajın özellik vektörü ile SVM model ağırlık vektörü arasında basit bir nokta ürünü gerçekleştirerek istenmeyen veya jambon e-posta olarak sınıflandırılabilir:

$$y=w.x-b \quad (1)$$

yukarıda belirtilen y sınıflandırmanın sonucu, w özellik vektöründekilere karşılık gelen ağırlık vektörü, b ise eğitim süreci tarafından belirlenen SVM modelinin sapma parametresidir (Lai & Tsai, 2004).

Deneysel sonuçlar, SVM yönteminin diğer sınıflandırıcılara eşit veya onlardan daha büyük bir performans gösterebildiğini ve böyle bir sonucu elde etmek için daha az eğitim verisi gerektirdiğini göstermektedir (Bassiouni, Ali, & El-Dahshan, 2018).

3.2. Multinomial Naive Bayes(NB)

Naive Bayes sınıflandırıcı, Bayes teoremini temel alan ve metin kategorizasyonunda sıkça kullanılan olasılık temelli bir yaklaşımdır (Karamollaoğlu, Doğru, & Dörterler, 2018; Lai & Tsai, 2004). Sade ve yüksek performansa sahip olmasından ötürü spam tespitinde sıkça kullanılan Naive bayes, metin içinde geçen terimleri Multinomial Model ve Multi-variate Bernoulli Model olmak üzere iki farklı şekilde ağırlıklandırır (Androutsopoulos, Paliouras, & Michelakis, 2004). Multi-variate Bernoulli Modeli bir kelimenin bir belgede bulunup bulunmamasıyla ilgilenir ve bu modelde kelimeler arasında herhangi bir bağ bulunmaz. Multinomial Model ise bir belgede geçen bir kelimenin birkaç kez kullanılabileceği durumunu göz önünde bulundurur ve kelimenin çoklu-oluşum olayının bağımsız olduğunu varsayar. Bu çalışmada Multinomial Model kullanılmıştır:

$$P(d_i|c_j;\theta) = P(|d_i|)|d_i|!\prod_{t=1}^{|V|}\frac{P(w_t|c_j;\theta)N_{it}}{N_{it}!} \quad (2)$$

Burada geçen d_i: eğitim belgelerini, |V|: toplam kelime sayısını, w_t: kelimeyi, N_{it} ise t kelimesinin j eğitim sınıfında görülme sıklığını ifade etmektedir (Karamollaoğlu, Doğru, & Dörterler, 2018). Kelime olasılığı tahmini(3) ile maksimum benzerlik tahmini(4) de ayrıca hesaplanmıştır.

$$\theta_{wt|cj} = P(w_t|c_j;\theta_j) = \frac{1+\sum_{i=1}^{|D|}N_{it}P(c_j|di)}{|V|+\sum_{s=1}^{|V|}\sum_{i=1}^{|D|}Nis\,P(cj|di)} \quad (3)$$

$$P(c_j|d_i;\theta) = \frac{P(c_j|\theta)P(d_i|c_i;\theta_j)}{P(d_i|\theta)} \quad (4)$$

3.3. Performans Değerlendirmesi

Bu makalede sınıflandırma algoritmalarının testi için değerlendirme ölçütleri kullanılmış ve iki algoritma içinde doğruluk (5), belirlilik (6), seçicilik(7) ve genel başarım (8) hesaplanmıştır.

Tablo 3.1. Karmaşıklık Matrisi

	Öngörülen Sınıf		
Gerçek Sınıf		*Spam*	*Jambon*
	Spam	(DP)	(YN)
	Jambon	(YP)	(DN)

$$\text{Doğruluk} = \frac{DP}{DP+YN} * 100\% \quad (5)$$

$$\text{Belirlilik} = \frac{DN}{DN+YP} * 100\% \quad (6)$$

$$\text{Seçicilik} = \frac{DP}{DP+YP} * 100\% \quad (7)$$

$$\text{Genel Başarım} = \frac{DP+DN}{DP+DN+YP+YN} \quad (8)$$

4. SONUÇLAR VE TARTIŞMA

Çalışmada spam e-posta tespiti için Multinomial Naive Bayes ile Destek Vektör Modeli kullanılmış, sınıflandırma performansını değerlendirmek ve sonuçları karşılaştırmak için çalışmalar yapılmıştır.

Sınırlandırma performansları: Doğruluk (5), Belirlilik (6), Seçicilik (7) ve Genel Başarım (8) şeklinde dört ölçüt kullanılarak değerlendirilmiştir. Burada belirtilen DP (Doğru Pozitif) sınıflandırıcı tarafından doğru sınıflandırılmış jambon e-posta sayısını, YP(Yanlış Pozitif) yanlış sınıflandırılmış jambon e-posta sayısını, DN(Doğru Negatif) doğru sınıflandırılmış spam e-posta sayısını, YN(Yanlış Negatif) yanlış sınıflandırılmış spam e-posta sayısını göstermekte olup Tablo 3.1'de gösterilmiştir.

Multinomial Naive Bayes yöntemi kullanılarak yapılan sınıflandırma sonucu elde edilen karmaşıklık matrisi Tablo 4.1'de gösterilmektedir.

Tablo 4.1. Multinomial Naive Bayes ile karmaşıklık matrisi

MNB	**Öngörülen Sınıf**		
Gerçek Sınıf		*Spam*	*Jambon*
	Spam	1104(DP)	709(YN)
	Jambon	192(YP)	2596(DN)

Tablo 4.1'deki karmaşıklık matrisi ile performans ölçütleri kullanılarak Multinomial Naive Bayes yönteminin sınıflandırma performansına ait elde edilen değerler ise Tablo 4.2'deki görülmektedir.

Tablo 4.2. MNB Yöntemine Ait Performans Değerleri

Ölçüt	**Değer(%)**
Duyarlılık	60.9
Belirlilik	93.1
Seçicilik	85.2
Genel Başarım	80.4

Buradan görüldüğü üzere Multinomial Naive Bayes yöntemi kullanılarak yapılan sınıflandırmada, spam e-posta tespitinde %60.9'luk, jambon e-posta tespitinde %93.1'lik doğrulukla sınıflandırma işlemi gerçekleştirildiği görülmüştür.

Destek Vektör Makineleri yöntemi kullanılarak yapılan sınıflandırma sonucu elde

edilen karmaşıklık matrisi Tablo 4.3'de gösterilmektedir.

Tablo 4.3. Destek Vektör Makineleri Karmaşıklık Matrisi

SVM	**Tahmini Sınıf**		
Gerçek Sınıf		*Spam*	*Jambon*
	Spam	1794(DP)	19(YN)
	Jambon	2(YP)	2786(DN)

Tablo 4.3'deki karmaşıklık matrisi ile performans ölçütleri kullanılarak Destek Vektör Makinelerinin sınıflandırma performansına ait elde edilen değerler Tablo 4.4'de görülmektedir.

Tablo 4.4. Destek Vektör Makinelerine Ait Performans Değerleri

Ölçüt	**Değer(%)**
Duyarlılık	98.9
Belirlilik	99.9
Seçicilik	99.9
Genel Başarım	99.5

Tablo 4.4'de görüldüğü üzere Destek Vektör Makineleri kullanılarak yapılan sınıflandırmada, spam e-posta tespitinde %98.9'luk, jambon e-posta tespitinde %99.9'luk doğrulukla sınıflandırma işlemi gerçekleştirilmiştir.

Multinomial Naive Bayes ve Destek Vektör Makineleri Modelinin sınıflandırma başarımlarının karşılaştırılmasıyla elde edilen değerler Tablo 4.5'de görülmektedir.

Tablo 4.5. Yöntemlerin Başarımının Karşılaştırılması

Ölçüt	**Değer(%)**		
	MNB Yöntemi	***Vektör Uzay Modeli***	***Fark***
Duyarlılık	60.9	98.9	38
Belirlilik	93.1	99.9	6.8
Seçicilik	85.2	99.9	14.7
Genel Başarım	80.4	99.5	19.1

Tablo 7'deki değerler dikkate alınarak kullanılan yöntemlerin karşılaştırılması sonucu Destek Vektör Makineleri yönteminin Multinomial Naive Bayes yönteminden daha başarılı olduğu görülmektedir.

5. SONUÇ

Spam e-postalar, internet topluluğu için ciddi bir sorun haline gelmekte olup ağ bütünlüğünü ve kullanıcı verimliliğini ciddi oranda tehdit etmektedir. Bu yazıda bu soruna çözüm olarak spam e-posta kategorizasyonunda sıkça kullanılan makine öğrenme algoritmalarından olan Multinomial Naive Bayes ile Destek Vektör Makineleri kullanılmış ve kullanılan yöntemlerin performansı belirli ölçütlere göre hesaplanarak karşılaştırılmıştır. Karşılaştırma sonucunda Destek Vektör Makineleri Modelinin %99.5436, Multinomial Naive Bayes Yönteminin ise %80.4173 doğrulukta çalıştığı ve Destek Vektör Makinelerinin daha başarılı olduğu saptanmıştır.

KAYNAKÇA

Akçetin, E., & Çelik, U. (2014). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması. *İnternet Uygulamaları ve Yönetimi, 5*(2), 45-56.

Almeida, T., & Yamakami, A. (2010). Content-based spam filtering. *In:Proceedings of the 23rd IEEE International Joint Conference on Neural Networks*, (s. 1-7). Barcelona.

Almeida, T., & Yamakami, A. (2012). Facing the Spammers: A Very Effective Approach to Avoid Junk E-mails. *Expert Systems with Applications, 39*(7), 6557-6561.

Altay, O., & Ulaş, M. (2018). Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbor in

Children. *2018 6th International Symposium on Digital Forensic and Security (ISDFS).* Antalya, Turkey: IEEE. doi:10.1109/ISDFS.2018.8355354

Androutsopoulos, L., Paliouras, G., & Michelakis, E. (2004). Learning to filter unsolicited commercial e-mail. *National Centre for Scientific Research "Demokritos".* Atina: Technical Report, 2004/2.

Bassiouni, M., Ali, M., & El-Dahshan, E. (2018). Ham and Spam E-Mails Classification Using Machine Learning Techniques. *Journal of Applied Security Research*, 315-331.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121-167.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines.* Cambridge: Cambridge University Press.

Çalış, K., Gözdağı, O., & Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. *Bilişim Teknolojileri Dergisi, 6*(1).

Drucker, H., Wu, D., & Vapnik, V. (1999). Support Vector Machines for Spam Categorization. *IEEE Transactions On Neural Networks, 10*(5), 1048-1054.

Duda, R., & Hart, P. (1973). *Pattern Classificafion and Scene Analysis.* New York: John Wiley & Sons.

Eryiğit, G., Tantuğ, C., & Adalı, E. (2005). Yaramaz E-Postaların Süzülmesinde, Karar Destek Makineleri, Naïve Bayes ve Bellek Tabanlı Öğrenme Yöntemlerinin Karşılaştırılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 27-36.

Goodman, B. J., Heckerman, D., & Rounthwaite, R. (2005). Stopping Spam. *Scientific American, 292*(4), 42-88.

Karamollaoğlu, H., Doğru, İ., & Dörterler, M. (2018). Makine Öğrenmesi Yöntemleri ile İstenmeyen E-postaların Tespiti. *2018 Akıllı Sistemlerde ve Uygulamalarda Yenilikler Konferansı (ASYU)* (s. 1-5). Adana, Türkiye: IEEE.

Lai, C.-C. (2007). An empirical study of three machine learning methods for spam filtering. *Knowledge-Based Systems, 20*(3), 249-254.

Lai, C.-C., & Tsai, M.-C. (2004). An Empirical Performance Comparison of Machine Learning Methods for Spam E-mail Categorization. *In Hybrid Intelligent Systems* (s. 44-48). USA: Fourth International Conference on IEEE.

Lee, Z. (2008). An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer. *Artificial Intelligence in Medicine, 42*(1), 81-93.

Naem, A., Ghali, N., & Saleh, A. (2018). Antlion optimization and boosting classifier for spam email detection. *Future Computing and Informatics Journal, 3*(2), 436-442.

Pontil, M., & Verri, A. (1998). Support vector machines for 3D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 637-646.

Ravikumar, K., & Gandhimathi, P. (2017). A Review on Different Spam Detection Approaches. *International Journal of Computer Techniques, 4*(4), 6-9.

Şahin, E., & Orhan, M. A. (2018). Makine Ögrenme Yöntemleri ve Kelime

Kümesi Teknigi ile Istenmeyen E-Posta/ E-Posta Sınıflaması.

Woitaszek, M., Shaaban, M., & Czernikowski, R. (2003). “Identifying junk electronic mail in Microsoft Outlook with a support vector machine. *Proceedings of the 2003 Symposium on Applications and the Internet*, (s. 166-169). Orlando,Fl.

Ying, K.-C., Lin, S.-W., Lee, Z.-J., & Lin, Y.-T. (2010). An ensemble approach applied to classify spam e-mails. *Expert Systems with Applications, 37*(3), 2197-2201.