International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 92-98
DOI: 10.36287/ijmsit.8.2.8
Received: October 16, 2024; Accepted: November 28, 2024

Metin Sınıflandırmaya Karşı Kriptografi Yöntemlerinin Kullanılması

Ahmet Emre ERGÜN [1*], Özgü CAN [2]

[1*] *İzmir Kâtip Çelebi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ahmetemreergun95@gmail.com) (ORCID: 0000-0002-3025-5640)*

[2] *Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ozgu.can@ege.edu.tr) (ORCID: 0000-0002-8064-2905)*

Özet **–** Bu makale, makine öğrenmesi sınıflandırma algoritmalarına karşı verilerin gizliliğini sağlamak için kriptografik tekniklerin nasıl kullanılabileceğini araştırmaktadır. Çalışma, ruh sağlığı sorunlarıyla ilgili metin ve etiket sütunlarını içeren Mental Health Corpus veri kümesine odaklanmaktadır. Metinleri sınıflandırmak için Rastgele Orman (*Random Forest*, RF), Karar Ağacı (*Decision Tree*, DT) ve Destek Vektör Makinesi (*Support Vector Machine*, SVM) sınıflandırma algoritmaları kullanılmıştır. Sınıflandırma doğruluğunu azaltmak için ise kriptografi yöntemi olan karakter kaydırma (*shift*) uygulanmaktadır. Sonuçlar, karakter kaydırmalarının sınıflandırıcı doğruluğunu büyük ölçüde azalttığını, 1 karakter kadar küçük kaydırmaların tüm modellerde doğruluğu %30'dan fazla azalttığını göstermektedir. Bulgular, kriptografik yöntemlerin, özellikle hassas bilgilerin söz konusu olduğu çeşitli alanlarda veri gizliliğini ve güvenliğini artırma potansiyelini göstermektedir.

Anahtar Kelimeler – gizlilik, makine öğrenmesi, kriptografi, kaydırma, ikame

Atıf: Ergün A.E., Can, Ö. (2024). Metin Sınıflandırmaya Karşı Kriptografi Yöntemlerinin Kullanılması. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 92-98.

Using Cryptography Methods Against Text Classification

Abstract – This article explores how cryptographic techniques can be used to ensure the confidentiality of data against machine learning classification algorithms. The study focuses on the Mental Health Corpus dataset, which contains text and tag columns related to mental health issues. Random Forest (RF), Decision Tree (DT) and Support Vector Machine (SVM) classification algorithms were used to classify the texts. To reduce classification accuracy, character shift, which is a cryptography method, is applied. Results show that character shifts greatly reduce classifier accuracy, with shifts as small as 1 character reducing accuracy by more than 30% across all models. The findings demonstrate the potential of cryptographic methods to increase data confidentiality and security in a variety of areas, especially where sensitive information is involved.

Keywords – *confidentiality, machine learning, cryptography, shift, substitution*

Citation: Ergün A.E., Can, Ö. (2024). Using Cryptography Methods Against Text Classification. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 92-98.

I. GİRİŞ

Büyük veri ve yapay zeka çağında, büyük veri kümelerini analiz etme ve bunlardan içgörü elde etme becerisi giderek daha değerli hale gelmiştir. Bu kabiliyet, veriye dayalı kararların önemli ilerlemelere ve verimliliklere yol açabileceği sağlık, finans ve pazarlama gibi alanlar için derin etkilere sahiptir [14], [17]. Bununla birlikte, özellikle ruh sağlığı kayıtları gibi hassas bilgilerle uğraşırken güvenlik riski artmaktadır. Bu tür verilere yetkisiz erişim, ayrımcılık, damgalama ve kimlik hırsızlığı gibi ciddi etik, yasal ve sosyal sonuçlara yol açabilmektedir [10], [12]. Kriptografik teknikler bu risklere karşı bir çözüm sunmaktadır. Kriptografi, verileri yetkisiz taraflar için anlamsız hale getirecek şekilde gizleyerek hassas bilgilerin istismar edilmesini önleyebilmektedir.

Ruh sağlığı verileri gibi hassas veriler mahremiyet açısından önemlidir. Sınıflandırma algoritlamaları kullanılarak hassas veriler sınıflandırılabilir ve bu verilere sahip kişilerin ruh sağlığı sorunu olup olmadığı analiz edilebilmektedir. Hassas metinler kriptografik yöntemler kullanılarak anlamsız hale getirilebilmektedir. Bu çalışma, metinsel verileri sınıflandırma saldırılarından korumak için basit ama etkili bir teknik olan kriptografik kaydırmaların (*substitution*) uygulanmasını araştırmaktadır. Karakter kaydırmaları ile, metinleri sistematik olarak değiştirerek, bu kriptografik yöntemlerin yaygın sınıflandırma algoritmalarının performansını nasıl

düşürebileceğini belirlemek ve böylece veri gizliliğini ve güvenliğini artırmak amaçlanmaktadır.

Bu çalışmadaki deneyler, ruh sağlığı sorunlarıyla ilgili metin ve etiket sütunlarını içeren Mental Health Corpus veri kümesine odaklanmaktadır. Veri seti İngilizce metinlerden oluştuğu için çalışmada kullanılan kaydırma yöntemi İngiliz alfabesine göre yapılmıştır. İngiliz alfabesinde toplam 26 harf bulunmaktadır. Bir harf 1 kere kaydırılırsa alfabedeki sıraya göre bir sonraki harfi almaktadır. Eğer bir harfe 0 veya 26 kere kaydırma yapılırsa kendisine geri döner ve harf değişmez. Bu sebeple 0 veya 26 harf kaydırma uygulanması güvenliği sağlamak açısından anlamsız olmaktadır. Bu çalışmadaki deneylerde kullanılan veri setinin %80'i eğitim, %20'si test verisi olarak kullanılmıştır. Deneylerdeki kaydırma işlemi 0'dan 25'e kadar olan sayılarda yapılmıştır. Bu sebeple, RF, DT ve SVM sınıflandırma algoritmalarının performansları 26 farklı deney üzerinde denenmiştir. Sonuçlar 4 farklı metrikte kıyaslanmıştır. Çalışmada, SVM algoritmasının DT ve RF algoritmalarına göre daha etkin olduğu görülmüştür.

Bu araştırmanın temel amacı, Mental Health Corpus veri kümesi üzerinde makine öğrenmesi sınıflandırıcılarının doğruluğunu azaltmada kriptografik yöntemlerin, özellikle de karakter kaydırmanın etkinliğini değerlendirmektir. Bu çalışma, karakter kaydırma yoluyla metinleri sistematik olarak değiştirerek, bu kriptografik tekniklerin yaygın sınıflandırma algoritmalarının performansını ne ölçüde düşürebileceğini belirlemeyi amaçlamaktadır. Nihai hedef, hassas verileri yetkisiz makine öğrenmesi analizinden koruyabilecek stratejiler geliştirmek ve böylece veri gizliliğini ve güvenliğini artırmaktır.

Bu çalışmadaki bölümler aşağıdaki şekilde yapılandırılmıştır: İkinci bölümde, mevcut literatürdeki çalışmalar ve ilgili araştırmalar gözden geçirilecektir. Üçüncü bölümde, araştırmada kullanılan metodoloji ve deneysel metrikler detaylandırılacaktır. Dördüncü bölümde, deneylerin sonuçları ve bulgular tartışılacaktır. Beşinci bölümde, elde edilen sonuçlar ışığında genel bir değerlendirme yapılacaktır. Altıncı bölümde, çalışmanın sonuçları ve gelecek araştırmalar için öneriler sunulacaktır. Son olarak çalışmada başvurulan kaynaklar listelenecektir.

II. LİTERATÜR

Literatürdeki bazı çalışmalar, ruh sağlığı bozuklukları için metin sınıflandırmasının kullanımını araştırmıştır. Sarno ve arkadaşları [16] ile Ameer ve arkadaşları [1], sosyal medya verilerine çok sınıflı sınıflandırma algoritmaları uygulamış, Sarno ve arkadaşları Mekanik Kontrol Tabanlı Makine Öğrenmesi (*Mechanical Control-Based Machine Learning*, MCML) algoritmasını kullanarak yüksek doğruluk elde ederken Ameer ve arkadaşları derin öğrenme ve transfer öğrenme modellerine odaklanmıştır. Abusaa ve arkadaşları [4], yazıya dökülmüş konuşma örneklerine dayanarak ruh sağlığı sorunlarını sınıflandırmak için makine öğrenmesi tekniklerini kullanmış ve şizofreni için yüksek doğruluk elde etmiştir. Ive ve arkadaşları [8], ruh sağlığıyla ilgili sosyal medya metinlerini sınıflandırmak için dikkat mekanizmalarına sahip hiyerarşik bir sinir modeli sunmuş ve geleneksel yöntemlere kıyasla daha iyi sonuçlar elde etmiştir. Bu çalışmalar toplu olarak ruh sağlığı analizinde metin sınıflandırmanın potansiyelini vurgulamaktadır.

Metin sınıflandırma, metin madenciliğinin önemli bir yönüdür ve bu amaçla çeşitli yöntemler ve sınıflandırıcılar kullanılmaktadır [7]. Zhang ve diğerleri [18], metin sınıflandırmada SVM ve Geri Yayılım Sinir Ağının (*Back Propagation Neural Network,* BPNN) performansını karşılaştırmış ve SVM'nin çok sınıflı sınıflandırmada daha iyi performans gösterdiğini bulmuştur. Kamruzzaman ve diğerleri [9], veri madenciliği kullanarak metin sınıflandırması için daha az eğitim gerektiren ve özellikleri türetmek için kelime ilişkilendirme kurallarını kullanan yeni bir algoritma tanıtmıştır. Arunachalam ve diğerleri [2] Bayesian sınıflandırması, Latent Dirichlet Allocation (*Gizli Dirichlet Tahsisi,* LDA) sınıflandırması, Dinamik Ontoloji Sınıflandırması ve Genetik Algoritma dahil olmak üzere duygu kutupluluğu tespiti için metin sınıflandırma tekniklerini tartışmıştır. Bu çalışmalar, metin sınıflandırmada kullanılan çeşitli yöntem ve tekniklere kapsamlı bir genel bakış sağlamaktadır.

Son araştırmalar, kaydırma ikamesi (*shifting substitution*) yoluyla metin şifrelemenin güvenliğini artırmak için çeşitli yöntemler önermiştir. Verma ve arkadaşları [13], Sezar şifresinin küçük bir dönüşümü olan modellenmiş kaydırma şifresini sunmaktadır. Shareef ve diğerleri [15], karakterleri bir anahtar değerine göre yeniden düzenleyen ve yaygın kriptografi saldırılarına karşı dirençli hale getiren bir şifreli metin kaydırma algoritması tasarlayarak bu tekniği daha da geliştirmiştir. Ambulkar [6], metin şifreleme için (Bilgi Değişimi İçin Amerikan Standart Kodlama Sistemi) (*American Standard Code for Information Interchange*, ASCII) değerleri oluşturmak üzere genetik algoritmalarla birlikte çoklu ikame (*shift*) yöntemlerinin kullanımını araştırmış ve şifreleme sürecine ekstra bir karmaşıklık katmanı eklemiştir. Bu çalışmalar toplu olarak, değişen ikame yöntemleri yoluyla metin şifrelemeyi geliştirme potansiyelini göstermektedir.

Metin manipülasyonunda ikame ve kaydırma tekniklerinin kullanımı çeşitli alanlarda yaygın bir uygulamadır. Borowiak ve diğerleri [3], karmaşık metin değiştirmeleri için PRXCHANGE işlevindeki düzenli ifadelerin gücünü vurgulamıştır. Li ve diğerleri [11], mobil cihazlarda metin revizyonu için değiştirme tabanlı bir teknik olan Swap'ı tanıtmış ve hassas işaret kontrolü ve tekrarlayan geri tuşuna basma ihtiyacını azaltmıştır. Son olarak, Pal ve diğerleri [5], metin mesajlarının güvenliğini sağlamak için kaydırma şifresi ve ikame şifresi gibi klasik kriptografi yöntemlerinin kullanımını araştırmıştır. Bu çalışmalar toplu olarak, metin manipülasyonu, çevirisi ve güvenliğinde kaydırma tekniklerinin öneminin altını çizmektedir.

III.METODOLOJİ

A. Veri Seti

Bu çalışmada kullanılan Mental Health Corpus veri kümesi, her biri bir metin ve buna karşılık gelen bir etiket içeren 27.977 girdiden oluşmaktadır. Metinler anksiyete, depresyon ve diğer ilgili durumlar gibi çeşitli ruh sağlığı sorunlarını tartışan bireylerin gönderilerini içermektedir. Veri kümesinde 0 ve 1 olmak üzere iki etiket bulunmaktadır. 0 ruh sağlığı sorunu olduğunu temsil eder, 1 ise ruh sağlığı sorunu olmadığını temsil etmektedir. Bu veri kümesi, yetkisiz erişim ve analizden korunması gereken hassas bilgiler içerdiğinden bu çalışma için özellikle uygundur.

B. Makine Öğrenmesi Algoritmaları

Saldırı senaryosunu simüle etmek için üç popüler makine öğrenmesi sınıflandırma algoritması kullanılmıştır:

- Rastgele Orman (RF): Eğitim sırasında birden fazla karar ağacı oluşturan ve sınıflandırma için sınıfların modunu çıkaran bir topluluk öğrenme yöntemidir. Bu algoritma sağlamlığı ve yüksek doğruluğu ile bilinir, bu da onu kriptografik tekniklerin etkinliğini test etmek için uygun bir aday haline getirmektedir.
- Karar Ağacı (DT): Kararların ağaç benzeri bir grafiğini ve tesadüfi olay sonuçları da dahil olmak üzere olası sonuçlarını kullanan bir modeldir. Karar ağaçlarının yorumlanması ve anlaşılması kolaydır, bu nedenle sınıflandırma görevlerinde sıklıkla kullanılırlar.
- Destek Vektör Makinesi (SVM): Bir veri kümesini sınıflara en iyi şekilde ayıran hiper düzlemi bularak sınıflandırma için verileri analiz eden denetimli bir öğrenme modelidir. SVM'ler yüksek boyutlu uzaylarda etkilidir ve metin sınıflandırma görevleri için yaygın olarak kullanılmaktadır.

Bu algoritmalar, metin sınıflandırma görevlerinde yaygın kullanımları ve etkinlikleri nedeniyle seçilmiştir. Çalışma, kriptografik tekniklerin bu sınıflandırıcılar üzerindeki etkisini değerlendirerek, yöntemlerin etkinliğinin kapsamlı bir değerlendirmesini sağlamayı amaçlamaktadır.

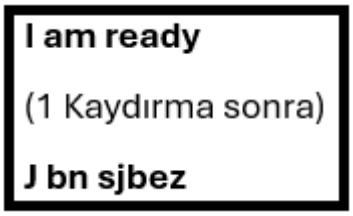
C. Kriptografi Yöntemleri

Bu çalışmada uygulanan kriptografik teknik karakter kaydırmadır. Bu, her bir karakteri alfabede sabit sayıda pozisyon kaydırarak metinleri değiştirmeyi içermektedir. Örneğin, 1'lik bir kaydırma 'a'yı 'b'ye, 'b'yi 'c'ye dönüştürür ve bu böyle devam eder. Tablo 1'de 1'er kaydırma uygulandığında harflerin alacağı yeni harfler gösterilmiştir.

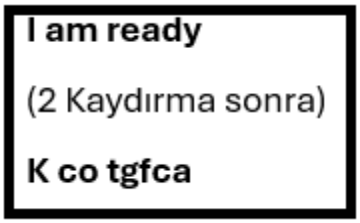
Tablo 1. *1'er Harf Kaydırmaya Göre Alfabetik Karşılık*

1 Kaydırmadan Önce	**1 Kaydırmadan Sonra**
a	b
b	c
c	d
d	e
e	f
f	g
g	i
i	j
j	k
k	l
l	m
m	n
n	o
o	p
p	q
q	r
r	s
s	t
t	u
u	v
v	w
w	x
x	y
y	z
z	a

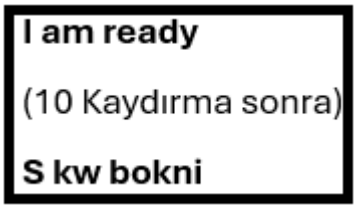
Şekil 1'de 1'er, 2'şer ve 10'ar kaydırma yapılmadan önce ve sonraki metinler gösterilmiştir. Metindeki her harf alfabetik olarak kaydırma sayısı kadar harf sonraki harfi almıştır.



a) 1'er Harf Kaydırma



b) 2'şer Harf Kaydırma



c) 10'ar Harf Kaydırma

Şekil 1. *Metindeki Harfleri Kaydırma*

Çalışmada sınıflandırma doğruluğu üzerindeki etkilerini değerlendirmek için 0 ila 25 arasında değişen kaydırmalar test edilmiştir. Bu basit ama etkili metin değiştirme yöntemi, farklı karakter kaydırma derecelerinin sınıflandırma algoritmalarını nasıl karıştırabileceğini değerlendirmek için kullanılmıştır.

D. Ölçü Metrikleri

Kriptografik yöntemlerin etkinliği dört temel ölçüt kullanılarak ölçülmüştür:

- Doğruluk: İncelenen toplam veri sayısı içindeki doğru sonuçların oranını ifade eder. Bu oran, hem Doğru Pozitifler (*True Positives*, TP), modelin doğru bir şekilde "pozitif" olarak tahmin ettiği durumlar hem de Doğru Negatifler (*True Negatives*, TN), modelin doğru bir şekilde "negatif" olarak tahmin ettiği durumlar göz önünde bulundurularak hesaplanır. Ayrıca, Yanlış Pozitifler (*False Positives*, FP), modelin "pozitif" olarak tahmin ettiği ancak gerçekte "negatif" olan durumlar ve Yanlış Negatifler (*False Negatives*, FN), modelin "negatif" olarak tahmin ettiği ancak gerçekte "pozitif" olan durumlar bu değerlendirmede dikkate alınır. Doğruluk, sınıflandırıcının etkinliğinin genel bir ölçüsünü sağlamaktadır. Doğruluk oranı hesaplama formülü (1)'de belirtilmiştir.

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Kesinlik (*Precision*): Pozitif olarak sınıflandırılan toplam vaka sayısı içinde doğru pozitiflerin oranıdır. Kesinlik, sınıflandırıcının ilgili vakaları tanımlamadaki performansı hakkında fikir vermektedir. Kesinlik oranı hesaplama formülü (2)'de belirtilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (2)$$

- Geri Çağırma (*Recall*): Gerçek pozitiflerin toplam sayısı içinde gerçek pozitiflerin oranıdır. Geri çağırma, sınıflandırıcının tüm ilgili verileri belirleme yeteneğini göstermektedir. Geri çağırma hesaplama formülü (3)'te belirtilmiştir.

$$Geri\ Çağırma = \frac{TP}{TP + FN} \quad (3)$$

- F1-Skoru: Kesinlik ve geri çağırmanın harmonik ortalamasıdır ve iki metrik arasında bir denge sağlamaktadır. F1-Skoru özellikle sınıf dağılımı dengesiz olduğunda kullanışlıdır. F1-Skoru hesaplama formülü (4)'te belirtilmiştir.

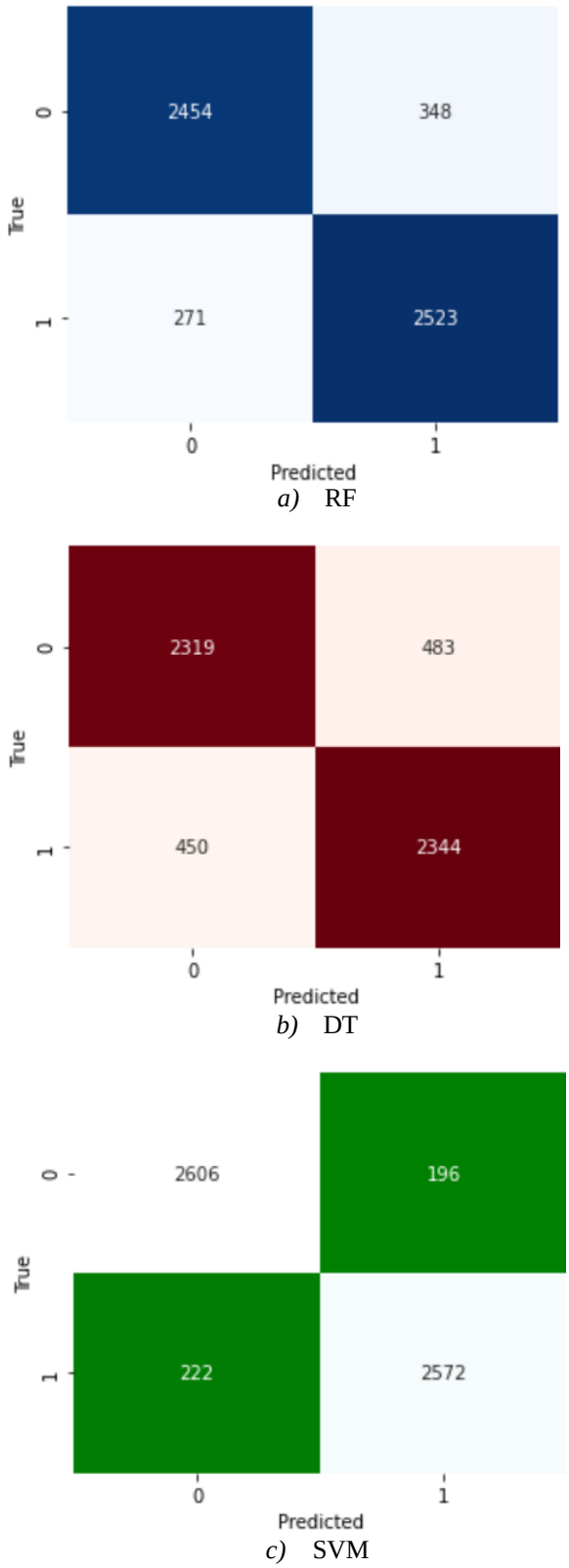
$$F1 - Skoru = 2.\frac{Kesinlik.Geri\ Çağırma}{Kesinlik + Geri\ Çağırma} \quad (4)$$

Bu metrikler, sınıflandırıcıların performansının kapsamlı bir değerlendirmesini sağlayarak kriptografik tekniklerin etkisinin ayrıntılı bir şekilde değerlendirilmesine olanak tanımaktadır.

IV. DENEY SONUÇLARI

Deneysel çalışmalar kapsamında 3 farklı makine öğrenmesi algoritması ve 26 farklı kaydırma durumu denenmiştir. Sonuçlar, RF, DT ve SVM sınıflandırıcılarının farklı kaydırmalardaki performansını göstermektedir. Değiştirilmemiş bir veri kümesi için (kaydırma 0), RF %88,94 doğruluk, %88,97 kesinlik, %88,94 geri çağırma ve %88,94 F1-Skoru elde etmiştir. DT %83,33 doğruluk, %83,33 hassasiyet, %83,33 geri çağırma ve %83,33 F1-Skoru elde etmiştir. SVM %92,53 doğruluk, %92,53 kesinlik, %92,53 geri çağırma ve %92,53 F1-Skoru elde etmiştir.

Harfler 1 karakter kaydırıldığında, tüm sınıflandırıcıların performansı önemli ölçüde düşmektedir. RF'nin doğruluğu %50,09'a, kesinliği %55,03'e, geri çağırma oranı %50,09'a ve F1-Skoru %33,52'ye düşmektedir. DT'nin doğruluğu %52,57'ye, kesinliği %62,92'ye, geri çağırma oranı %52,57'ye ve F1-Skoru %40,60'a düşmektedir. SVM'nin doğruluğu %52,36'ya, hassasiyeti %62,56'ya, geri çağırma oranı %52,36'ya ve F1-Skoru %40,11'e düşmektedir. Şekil 2'de makine öğrenmesi algoritmalarının 0 kaydırmada karışıklık matrisleri gösterilmektedir. SVM'in DT'ye RF'ye göre daha çok doğru pozitif ve doğru negatif sayısına sahip olduğu görülmektedir.



a) RF

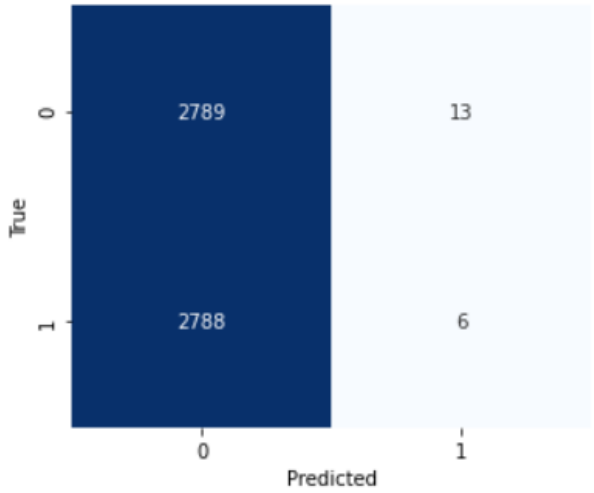
b) DT

c) SVM

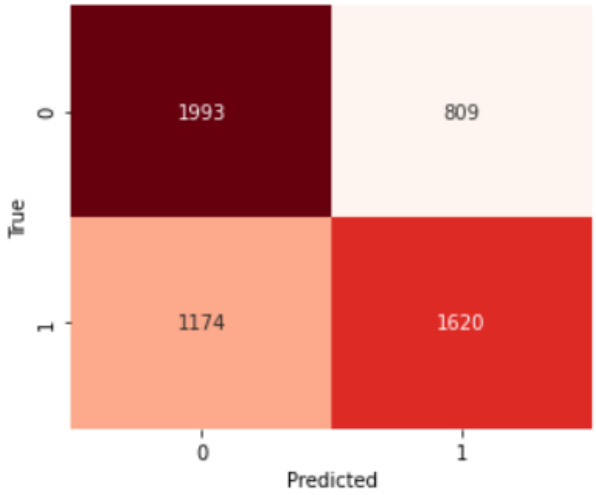
Şekil 2. *0 Kaydırma İçin Karışıklık Matrisleri*

Daha fazla kaydırma da sınıflandırıcıların performansını düşürmektedir. Örneğin, 5 karakterlik bir kaydırma ile RF %51,55 doğruluk, %68,13 kesinlik, %51,55 geri çağırma ve %37,11 F1-Skoru elde etmiştir. DT %51,89 doğruluk, %68,69 kesinlik, %51,89 geri çağırma ve %37,87 F1-Skoru elde etmektedir. SVM %52,11 doğruluk, %62,94 kesinlik, %52,11 geri çağırma ve %39,32 F1-Skoru elde etmektedir.

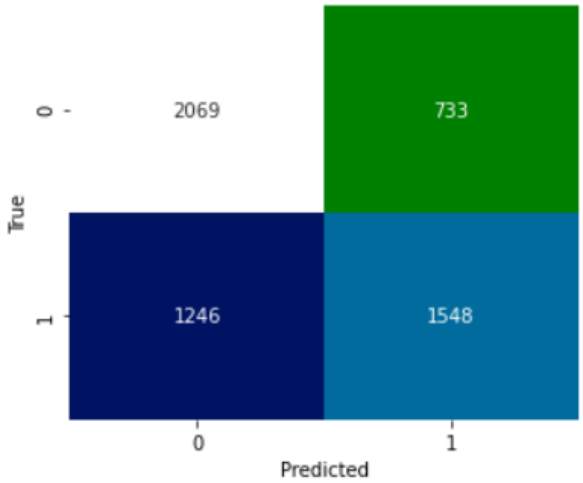
10 karakterlik bir kaydırmada DT'nin doğruluğu %64,56'ya, kesinliği %64,81'e, geri çağırma oranı %64,56'ya ve F1-Skoru %64,41'e düşmüştür. SVM %64,64 doğruluk, %65,14 kesinlik, %64,64 geri çağırma ve %64,33 F1-Skoru elde etmiştir. RF'nin performansı %49,95 doğruluk, %40,81 kesinlik, %49,95 geri çağırma ve %33,55 F1-Skoru ile nispeten sabit kalmaktadır. Şekil 3'te makine öğrenmesi algoritmalarının 10 kaydırmada karışıklık matrisleri gösterilmektedir. SVM ve DT'nin RF'ye göre daha çok doğru pozitif ve doğru negatif sayısına sahip olduğu görülmektedir.



a) RF



b) DT



c) SVM

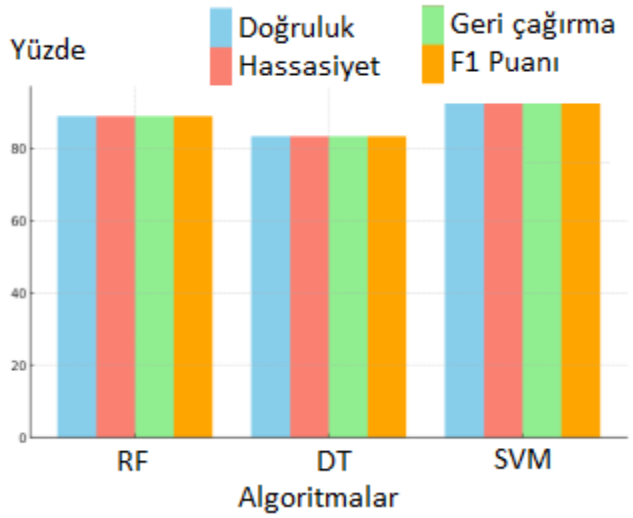
Şekil 3. 1*0 Kaydırma İçin Karışıklık Matrisleri*

Tablo 2'de 0'dan 25'e kadar kaydırmaya karşı makine öğrenmesi algoritmalarının performansları gösterilmektedir. SVM algoritması kaydırma olan ve olmayan durumlarda en yüksek doğruluk oranlarına sahiptir. Kaydırma kullanılan durumlarda her algoritmanın doğruluk oranı yaklaşık olarak %30 oranında düşürülmüştür. Genel olarak, kaydırma arttıkça sınıflandırıcıların performansı dalgalanmakta, ancak değiştirilmemiş veri kümesindeki performanslarından daha düşük kalmaktadır. Bu, karakter kaydırmanın makine öğrenmesi sınıflandırıcılarının doğruluğunu etkili bir şekilde azaltabileceğini ve hassas metinsel verileri yetkisiz sınıflandırma girişimlerinden koruyabileceğini göstermektedir.

Aşağıda Şekil 4'te, RF, DT ve SVM algoritmalarının kaydırmasız durumdaki doğruluk oranları karşılaştırılmaktadır. Kaydırmasız durumda, SVM %92,53 doğruluk ile en iyi performansı sergileyerek, doğrusal olmayan verileri sınıflandırma konusundaki güçlü yeteneğini ortaya koymaktadır. RF ise %88,94 doğrulukla oldukça etkili bir performans sunmaktadır; ağaç tabanlı yapısı, verinin önemli desenlerini ve ilişkilerini öğrenerek doğru sınıflandırmalar yapılmasını sağlamaktadır. DT'nin doğruluğu %83,33 ile RF ve SVM algoritmalarının gerisinde kalmaktadır; bu durum, karar ağaçlarının veri üzerinde aşırı uyum yaparak genelleme yeteneklerini zayıflatmasından kaynaklanmaktadır. Bununla birlikte, üç algoritma için de F1 skoru, hassasiyet, geri çağırma ve doğruluk oranları birbirine çok yakındır, bu da modellerin genel olarak iyi çalıştığını ve veriye uygun sınıflandırmalar gerçekleştirdiğini göstermektedir.

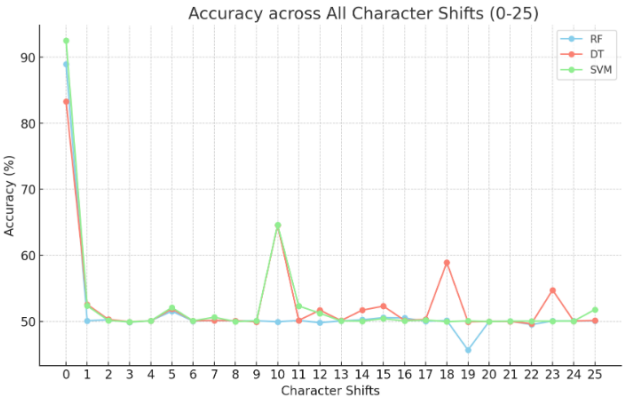
Tablo 2. *Algoritmaların Kaydırmalara Göre Doğruluk Oranları*

Kaydırma	RF Doğruluk Oranı	DT Doğruluk Oranı	SVM Doğruluk Oranı
0	%88,94	%83,33	**%92,53**
1	%50,09	%52,57	%52,36
2	%50,25	%50,30	%50,14
3	%49,89	%49,95	%49,93
4	%50,09	%50,07	%50,07
5	**%51,55**	%51,89	%52,11
6	%50,05	%50,09	%50,09
7	%50,13	%50,13	%50,63
8	%50,07	%50,11	%49,98
9	%50,09	%49,95	%50,07
10	%49,95	**%64,56**	**%64,64**
11	%50,13	%50,16	%52,34
12	%49,82	%51,70	%51,23
13	%50,07	%50,13	%50,07
14	%50,25	%51,70	%50,04
15	%50,57	%52,31	%50,41
16	%50,52	%50,16	%50,07
17	%50,02	%50,30	%50,21
18	%50,13	%58,92	**%49,96**
19	**%45,66**	%49,95	%50,07
20	%49,98	%50,00	%49,98
21	%50,05	%50,00	%50,04
22	%49,52	**%49,66**	%50,04
23	%50,07	%54,75	%50,04
24	%50,07	%50,05	%50,05
25	%50,07	%50,13	%51,79



Şekil 4. *Kaydırmasız Durumda Algoritmaların Performansı*

Şekil 5'te, RF, DT ve SVM algoritmalarının 0'dan 25'e kadar tüm kaydırma durumlarındaki doğruluk oranları karşılaştırılmaktadır. 1 kaydırmadan itibaren tüm algoritmaların doğruluğu hızla düşerek %50 seviyelerine gerilemiş, ancak bazı kaydırma noktalarında sınırlı da olsa farklılaşmalar gözlenmiştir. Özellikle 10. kaydırmada DT ve SVM doğruluk oranları %64 seviyelerine yükselmiş, diğer kaydırmalarda ise performans daha stabil bir şekilde düşük kalmıştır. Bu durum, kaydırmanın algoritmaların sınıflandırma performansı üzerinde dalgalı bir etki yarattığını ve genel olarak doğruluk oranlarının kaydırmasız duruma kıyasla ciddi ölçüde düştüğünü göstermektedir.



Şekil 5. *25 Karakterlik Kaydırmaya Kadar Doğruluk Oranları*

V. TARTIŞMA

Bu çalışmanın bulguları, hassas bilgilerin makine öğrenmesi tabanlı saldırılardan korunmasında kriptografik tekniklerin potansiyelini vurgulamaktadır. Ruh sağlığı bağlamında, mahremiyetin korunması özellikle önemlidir. Ruh sağlığı kayıtları, ifşa edilmesi halinde ayrımcılığa, damgalanmaya ve önemli kişisel sıkıntılara yol açabilecek son derece kişisel ve hassas bilgiler içerebilmektedir. Metinsel verilere karakter kaydırmaları uygulandığında, sınıflandırıcıların doğruluğu, kesinliği, geri çağırması ve F1 puanları önemli ölçüde azalır ve böylece verileri doğru bir şekilde sınıflandırma oranları düşürülmektedir. Bu yaklaşım, veri gizliliğini artırmanın basit ama etkili bir yolunu sunmaktadır.

Karakter kaydırma gibi kriptografik teknikler, metinsel verileri korumak için basit ancak güçlü bir yöntem sunmaktadır. Bu yöntemler, bir metindeki karakterleri değiştirerek, makine öğrenmesi sınıflandırıcılarının doğru tahminler yapmakta zorlanacağı noktaya kadar içeriği gizleyebilmektedir. Bu çalışma, basit kaydırmaların bile sınıflandırıcı performansını büyük ölçüde azaltabileceğini ve bu sayede veri gizliliğini artırmak için uygun bir seçenek olduğunu göstermektedir. Çalışmada veri setindeki her harf alfabede alabileceği tüm harflere kaydırılmış ve harf olarak alabileceği tüm değerlere göre kıyaslanmıştır. Bu da, yaygın kullanılan makine öğrenmesi algoritmaları olan RF, DT ve SVM sınıflandırıcılarının performans kıyaslamasına çok çeşitli bir perspektif katmıştır.

Karakter kaydırma, basitliği ve uygulama kolaylığı nedeniyle özellikle caziptir. Daha karmaşık kriptografik yöntemlerin aksine, sofistike algoritmalar veya kapsamlı hesaplama kaynakları gerektirmez. Bu da onu kişisel verilerin korunmasından kurumsal veri güvenliğine kadar geniş bir uygulama yelpazesi için erişilebilir kılmaktadır.

VI. SONUÇ

Bu çalışma, hassas metin verilerini yetkisiz makine öğrenmesi sınıflandırmasından korumak için kriptografik tekniklerin, özellikle de karakter kaydırmanın etkinliğini göstermektedir. Bu yöntemlerin uygulanması, Rastgele Orman, Karar Ağacı ve Destek Vektör Makinesi gibi yaygın

kullanılan sınıflandırıcıların doğruluğunu ve diğer performans ölçütlerini önemli ölçüde azaltmaktadır. Bu bulgular, kriptografik yöntemlerin çeşitli alanlarda veri gizliliği ve güvenliğini artırma potansiyelinin altını çizmektedir. Bu, özellikle sağlık ve finans gibi veri gizliliğinin çok önemli olduğu alanlarda önemlidir. Genel olarak, bu çalışmanın sonuçları, büyük veri ve yapay zeka çağında hassas verileri korumak için kriptografik yöntemlerin sürekli araştırılması ve geliştirilmesi ihtiyacını vurgulamaktadır. Çalışmadaki deneyler karakter kaydırmanın 26 farklı şekilde yapılması ve bunların sonuçlarının yaygın kullanılan sınıflandırma algoritmaları kullanılarak kıyaslanmasını içermektedir. Bu bağlamda incelenen literatürde benzer çalışma bulunamamıştır.

Gelecekteki çalışmalar kapsamında veri güvenliğini artırmak için birden fazla kriptografik tekniğin birleştirilmesi, yöntemlerin gerçek dünya senaryoları için genelleştirilebilmesi ve canlı ortama alınması hedeflenmektedir. Bu alanlara odaklanarak, gelecekteki araştırmalar hassas verileri yetkisiz makine öğrenmesi analizinden korumak için daha esnek ve etkili stratejilerin geliştirilmesine katkıda bulunabilir.

KAYNAKLAR

[1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.

[9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.