HARRAN ÜNİVERSİTESİ MÜHENDİSLİK DERGİSİ

HARRAN UNIVERSITY JOURNAL OF ENGINEERING

e-ISSN: 2528-8733 (ONLINE)

Çevrimiçi Sosyal Ağlarda Spam Hesap Tespiti

Spam Account Detection in Online Social Networks

Yazar(lar) (Author(s)): *Esma Elma[1], Nagehan İLHAN[2]*

[1] ORIC ID: 0009-0009-1715-8342
[2] ORIC ID: 0000-0002-1367-9230

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): Elma E., İlhan N., "Çevrimiçi Sosyal Ağlarda Spam Hesap Tespiti", **Harran Üniversitesi Mühendislik Dergisi,** 9(2): 71-89, (2024).

DOI: 10.46578/humder.1442237

HRU Muh Der, 9(2): 71-89 (2024)

Harran Üniversitesi
Mühendislik Dergisi

Araştırma Makalesi

Çevrimiçi Sosyal Ağlarda Spam Hesap Tespiti

Esma ELMA[1,*], Nagehan İLHAN[2]

[1,2]*Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 63050, Haliliye/Şanlıurfa*

Makale Bilgisi

Başvuru:24/02/2024
Kabul:30/08/2024

Anahtar Kelimeler

Sosyal Medya, Çevrimiçi Sosyal Ağlar, Instagram

Keywords

Social Media, Online Social Networks, Instagram

Öz

Yeni iletişim teknolojilerinin ortaya çıkışı, bireylerin haber ve bilgiye erişim biçimlerinde derin bir dönüşüme yol açmıştır. Sosyal ağlar bir konuda haber ve bilgi kaynağı olarak en çok öne çıkan ortamlardan biri haline gelmiştir. Bu nedenle, sosyal medya olgusunu güvenilirlik kavramı bağlamında incelemek zorunludur.Sosyal ağlarda takipçi sayısı bir başarı göstergesi olarak algılanmaktadır. Ancak, sahte hesaplar kendi gibi sahte takipçileri satın alarak kendi kimliğini gizleyebilmekte ve yanlış bilgileri rahatlıkla paylaşabilmektedir. Çalışmamızın amacı sahte hesapları tespit ederek sosyal ağların daha güvenilir hale gelmesini sağlamaktır. Bu çalışmada Instagram verileri üzerinde Rastgele Orman (Random Forest), Ekstra Ağaçlar (Extra Trees), Gradyan Arttırma (Gradient Boosting), Karar Ağacı (Decision Tree), AdaBoost (AdaptiveBoosting) gibi ağaç yapısına sahip algoritmalar kullanılmıştır. Algoritmalarımızın doğruluk oranları sırasıyla %90, %91, %91, %82, %89 olarak elde edilmiştir. Ayrıca, tekil olarak başarımları ölçülen algoritmaların performansını artırmak için tüm kullanılan algoritmaları "birleştiren" yeni bir entegre yaklaşımımız olan VotingClassifier ve StackingClassifier algoritmaları uygulanmış, eğitim ve test verilerimiz için %91 doğruluk elde edilmiştir. Ayrıca, bir arayüz ile sunulan modelin son kullanıcı tarafından test edilebilmesi sağlanmıştır.

Spam Account Detection in Online Social Networks

Abstract

The emergence of new communication technologies has led to a profound transformation in the way individuals access news and information. Social networks have become one of the most prominent media as a source of news and information. Therefore, it is imperative to analyse the social media phenomenon in the context of the concept of reliability. The number of followers in social networks is perceived as an indicator of success. However, fake accounts can hide their identity by buying fake followers like themselves and can easily share false information. The aim of our study is to make social networks more reliable by detecting fake accounts. In this study, algorithms with tree structure such as Random Forest, Extra Trees, Gradient Boosting, Decision Tree, AdaBoosting and AdaptiveBoosting were used on Instagram data. The accuracy rates of our algorithms are 90%, 91%, 91%, 91%, 82%, 89% respectively. In addition, in order to improve the performance of the algorithms whose performance was measured individually, a new integrated approach, VotingClassifier and StackingClassifier algorithms, which 'combine' all the algorithms used, were applied and 91% accuracy was obtained for our training and test data. In addition, an interface was provided to allow end-user testing of the presented model.

1. GİRİŞ (INTRODUCTION)

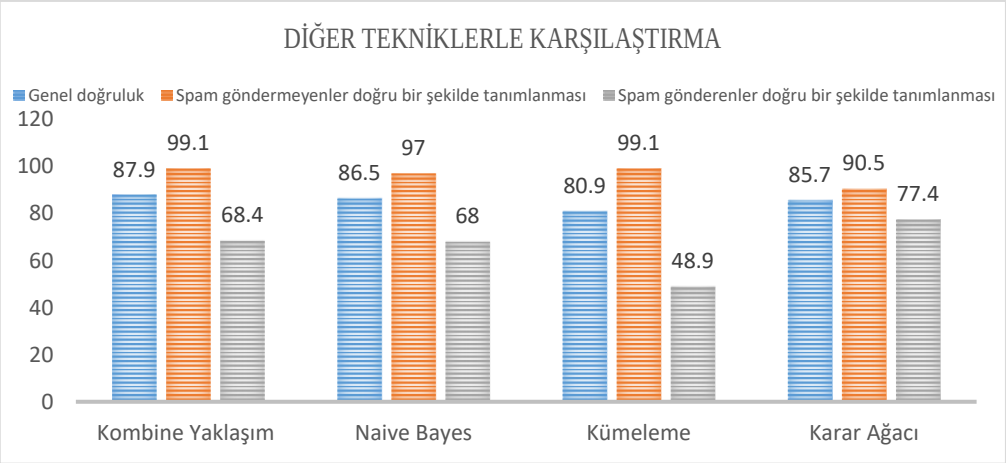
Teknolojinin gelişimi ile zamandan ve mekândan bağımsız internet yaygın halde kullanılmaya başlanmıştır. İnternet kullanımının artması ile insanların kendi içeriklerini başkalarıyla paylaşıp etkileşim içerisine girmesi için çevrimiçi sosyal ağlar oluşturulmuştur. Sosyal Ağlar insanların vakitlerinin büyük bir kısmını geçirdikleri mecralar haline gelmiştir. Şimdiki nesilde, herkesin sosyal hayatı çevrimiçi sosyal ağlarla ilişkilendirilir hale geldi. Sosyal ağlar (Orbit Showtime Network veya kısaca OSN'ler), insanların etkileşime girdikleri yararlı bilgiler paylaştıkları ve iletişim kurduğu bir platformlar olarak büyük önem kazanmaktadır. Eski dönemlerde gazete, dergi, televizyon gibi

*İletişimyazarı, e-mail:esma.elma15@gmail.com

geleneksel yöntemlerle reklam uygulamaları yayınlanmaktayken, günümüzde çevrimiçi sosyal ağların yayılması ile reklam işlemleri yeni bir biçim kazanmıştır. Bu nedenle haber ajansları için sosyal medya üzerinden bilgi edinmek kolay ve avantajlı gelmektedir. Dünyadaki tüm haberlere anında ulaşabilmeyi sağlamaktadır. İş açısından bakıldığında, bu platformlar hem devlet hem de ticari kurumlara pazarlama, müşteri davranış analizi ve fikir madenciliği için etkili yollar sunar.

Geçtiğimiz yıllarda, ABD'deki bazı bankalar ve finans kurumları, krediyi vermeden önce kredi başvuru sahiplerinin sosyal medya hesaplarını analiz etmeye karar verdi. Bu nedenle, popüler bir hesaba sahip olmak, başvuru sahibinin kredibilitesini ve güvenilirliğini etkili bir şekilde artırmaya yardımcı olabilir. Bu nedenle, bir spam gönderici sahte takipçileri kabul ederse, meşru bir kullanıcı olarak etkili bir şekilde hareket edebilir ve daha yetkili mesajlar gönderebilir böylece çeşitli etkili reklam kampanyaları başlatabilir [1]. Tüm bu özellikler, sosyal ağları kullanmayı herkesin ilk tercihi haline getirdi. Sosyal ağlarda şu an kayıtlı 150 milyondan fazla kullanıcı hesabı bulunmaktadır. Ancak, muhteşem özellikleri olan sosyal ağlar, sahte ve gerçek hesapların bulunduğu bir pazar haline geldi [2]. Sosyal ağların bir parçası olan Instagram'da kitleler arasında bu kadar popüler hale gelmesine rağmen, burayı daha güvenli ve özgün hale getirmek için çok az araştırma yapıldı. Kötü amaçlı kişiler tarafından yanlış bilgilendirme, söylentiler, sahte haberler vb. gibi spam içerikler yayılmaya başlandı. Spam içerik ve mesajlardan edinilen bilgilerin yanlışlığı bu ağları kullananların güvenliği açısından büyük bir tehdit oluşturmaktadır. Spam hesap ve içeriklerin artmasıyla sosyal medya platformları, tüm kullanıcıların tek tek incelenmesi mümkün olmadığı için, spam gönderileri azaltmak amacıyla veri madenciliği yöntemlerine başvurarak sahte takipçi temizleme işlemi yapmaktadır. Bu süreç boyunca yüksek sayılarda sahte hesap temizliği sağlanmış olsa da sonuç olarak ne kadar başarıya ulaştıkları hakkında genel bir bilgi bulunmamaktadır.

Literatür incelendiğinde, yaygın olarak Twitter verilerinin ele alındığı ve hazır veri setleri üzerinde işlemler yapıldığı görülmektedir. Twitter üzerinde yapılan bir çalışmada Fabricio Benevenuto ve ark. tarafından elde edilen veri kümesi kullanılmıştır. Veri seti 1064 Twitter kullanıcısının etiketli kaydından oluşmaktadır. Veri kümesi, kullanıcıya ve tweet'e özgü bilgileri içeren 62 özellikten oluşur. Veri kümesinin %36'sı spam hesaplardan oluşmaktadır. Takip ve takipçi sayısı, URL, spam kelimeler, cevaplar, hashtag'ler ele alınarak işlem yapılmıştır. Çalışmada öğrenme algoritmaları olarak Naive Bayes, kümeleme ve karar ağaçları kullanılmıştır. Her ne kadar bu yaklaşımların her biri kullanıcı hesaplarını sınıflandırmak için kullanılabilir olsa da, doğruluğu artırmak için bu yaklaşımlar çalışmada entegre olarak birleştirilmiştir. Önerilen algoritmanın, bir hesabı %87,9 doğrulukla spam veya spam olmayan olarak başarılı bir şekilde tanımlayabildiği açıktır [3]. Çalışma sonucunda dört öğrenme yaklaşımı kullanılarak Spam Tespitindeki iyileşmelerin karşılaştırılması Şekil 1'de gösterilmiştir.



Şekil 1. *Dört öğrenme yaklaşımı kullanılarak spam tespitindeki iyileşmelerin karşılaştırılması [3]*

Bu algoritmanın spam göndermeyenlerin tespit doğruluğunun (%99,1), spam gönderenlerin tespit doğruluğuna (%68,4) kıyasla daha yüksek olduğu gözlemlenmiştir. Sonuçlarda ise, Kümeleme algoritmasının spam olmayan hesapları tespit etmede daha iyi performans gösterdiği, ancak spam hesaplarını tespit etmede çok zayıf olduğu gösterilmiştir [3].

Başka bir çalışmada ise Twitter'daki spam botlarını tanımlamak için Karar Ağacı, Sinir Ağı, Destek Vektör Makineleri ve K-en Yakın Komşular gibi farklı sınıflandırma yöntemleri uygulanmıştır. Bu algoritmalar arasında, Bayes sınıflandırıcı birkaç nedenden dolayı en iyi performansa sahip olduğu tespit edilmiştir [4]. İlk olarak, Bayes sınıflandırıcının gürültüye karşı dayanıklı olduğu gözlemlenmiştir. Bayes sınıflandırının daha iyi bir performansa sahip olmasının bir başka nedeni ise, sınıf etiketinin kullanıcının özel modeline göre tahmin edilmesinden kaynaklanmaktadır. Veri kümesi olarak, 3 Ocak-24 Ocak 2010 tarihleri arasında 3 hafta boyunca veri toplanmış, halka açık verilerden toplam 25.847 kullanıcı, yaklaşık 500 bin Tweet ve yaklaşık 49 milyon takipçi/arkadaş ilişkisi elde edilmiştir. Sonuç olarak Naive Bayes sınıflandırıcısının, diğer algoritmalara kıyasla en iyi genel performansa sahip olduğu gözlemlenmiştir [4].

Farklı bir çalışmada beş adet denetimli makine öğrenme tekniği olan Rastgele Orman, Sinir Ağı, Lojistik Regresyon, Naive Bayes ve J48 Karar Ağacı kullanılmıştır. Bu yöntemlerin çoğu araştırmada en iyi doğruluğu sağladığı gözlemlenmiştir. Önerilen model, %91,76'ya kadar doğruluk oranıyla sahte kullanıcıları ortadan kaldırmaya ve daha sağlıklı bir sosyal medya ortamı oluşturmaya yardımcı olmuştur. Instagram'ı ele alan çalışmalarda kullanılan mevcut meta veriler kullanıcı adı, tam ad, biyografi, bağlantı, profil resmi, mesaj sayısı, takipler, takipçilerdir [5].

Bir başka çalışmada Instagram'daki sahte, bot ve gerçek hesapları makine öğrenmesi algoritmasıyla tespit etmek amaçlanmıştır. Instagram'dan 970 bot hesap, 959 gerçek hesap verileri kullanılmıştır. KNN, MLP, NB, SVM, DT ve RF metotları kullanılmıştır. Rastgele Orman yöntemi %90,2 ile en yüksek doğruluk değerini vermiştir ve diğer sınıflandırıcılardan daha iyi performansa sahip olduğu gözlemlenmiştir [6].

Diğer bir çalışmada ise Facebook verileri RapidMiner'da (makine öğrenmesi, veri madenciliği amaçlarına yönelik olarak geliştirilmiş bir yazılım) incelenmiştir. Veri kümesi 889 satır ve 23 sütun içermektedir. Naive Bayes, Karar Ağacı ve Lojistik Regresyon kullanarak, genellikle lise ve üniversite düzeyindeki öğrencilerin sahte veya gerçek hesaplar hakkında tahminleri ele alınmıştır. Naive Bayes %51 doğruluk sağlarken, Lojistik Regresyon %91 doğruluk ve Karar Ağacı'nın %36 doğruluk sağladığı gözlemlenmiştir [7].

Instagram üzerinde yapılan bir çalışmada Kaggle platformunda oluşturulan veri seti ele alınmıştır. Veri seti 576 eğitim 120 test verilerinden oluşmaktadır. Lojistik Regresyon ve Rastgele Orman öğrenme algoritmaları sahte hesap tespiti için kullanılmıştır. Veri setinde aynı olan değerlerin ortalaması hesaplanarak ortancası aldıktan sonra, tüm aynı değerlerin hesaplanan ortanca değeri ile değiştirme işlemi yapılarak aynı değerle de temizlenme işlemi gerçekleştirilmiştir. Sonuç olarak, Lojistik Regresyon ve Rastgele Orman modellerinde sırasıyla %90,8 ve %92,5 doğruluk oranları elde edilmiştir [8].

Instagram verilerini ele alan başka çalışmada yedi farklı algoritma kullanılmış; Lojistik Regresyon (LR), Destek Vektör Makinesi (SVM), Naive Bayes (NB), K-en Yakın Komşular (KNN), Karar Ağacı, Rastgele Orman (RF), Çok Katmanlı Algılayıcı (MLP). Veri setinde 2799 veri ele alınmıştır. Son aya ait verilerdeki ortalama medya beğeni sayısı, son medya sayısı ve medya miktarına ulaşılamadığı için bu veriler eksik veri olarak değerlendiriliyordu. Bu çalışmada eksik veriler yerine uygun değerleri kullanılmış. Sınıflandırıcıların başarımını değerlendirmek için doğruluk, F-puanı, kesinlik ve geri çağırma metriklerini kullanılmış. Çalışmada RF sınıflandırıcının bot hesapların %92,8'ini doğru şekilde tanımladığını, bot hesapların %4,4'ünün sahte olarak, %2,8'inin ise gerçek hesap olarak yanlış sınıflandırıldığını gözlemlenmiştir. Tüm hesap türleri içinde RF sınıflandırma başarısı oldukça yüksek olduğu görülmüştür. Ancak RF sınıflandırıcının gerçek ve sahte hesapları ayırt etmede zorluk yaşadığı söylenebilir [6].

Çalışmaların genelinde, Rastgele Orman'ın sürekli olarak diğer algoritmalardan daha iyi performans gösterdiği ve en yaygın kullanılanı olduğu görülmüştür. Doğru girdiler mevcut olduğunda ve eksik girdi olmadığında, Rastgele Orman harika çalışmaktadır. Yapılan araştırmalara bakıldığında, mevcut çalışmaların çoğunun özellik tabanlı algılama yöntemlerini kullandığı ve Twitter'ı platform olarak kullanan bir sınıflandırıcı modeli önerdiği görülürken, sınırlı sayıda Instagram sahte hesap tespiti çalışması mevcuttur. Bunun nedenlerinden biri, Instagram'dan veri çekmenin Twitter'dan daha karmaşık olmasıdır. Bir diğer neden ise Instagram'ın Twitter'dan 4 yıl sonra kullanıma sunulmasıdır [6].

Çalışmamızda, Instagram platformunda sınırlı sayıda çalışma olması nedeniyle Instagram tercih edilmiştir. Genellikle yüksek başarı oranlarından bahsedilen Rastgele Orman algoritması, entegre yaklaşımımızda başarıyı artırmak için kullanılmıştır. Rastgele Orman algoritmasının yanı sıra, sık tercih edilen K-En Yakın Komşular (KNN), Naive Bayes ve Kümeleme gibi mevcut sistemli algoritmalar yerine, nadiren kullanılan ve yeni yöntemler denendi.

Bu kullanılan algoritmalar, Rastgele Orman'a benzer bir şekilde ağaç yapısı ile çalışan algoritmalardır, bu nedenle tüm sonuçlarımızda yüksek başarı elde edilmiştir. Gözlemlerimize göre, bu algoritmaların düşük hatalı hipotezler üretmeye çalıştığını gördük. Tüm algoritmalarımızın sonuçlarında, sahte ve gerçek hesaplar için yüksek doğruluk elde ettik. Ayrıca, diğer çalışmalardan farklı olarak, yüksek başarı oranları elde edilen algoritmaları entegre ederek yeni algoritmalar oluşturduk ve bu algoritmaların arayüzde girilen tüm sahte ve gerçek hesaplar için yüksek doğrulukla sonuç verdiğini gözlemledik.

2. MATERYAL VE METOT (MATERIAL AND METHOD)

2.1. MATERYAL

Veri seti üzerinde yinelenen satırların silinmesi, sürekli ve ikili özelliklerin gösterilmesi, verilerdeki özelliklerin kullanım oranının kontrol edilerek gereksiz özelliklerin bulunması gibi ön işlemler yapıldıktan sonra sınıflandırma algoritmaları ile analiz gerçekleştirilerek çıkan sonuçlar karşılaştırılmıştır. Ön işleme aşamasında kullanılan sürekli özellikler, herhangi bir değer alabilecek gerçek sayılardan oluşan verilerdir (ör. sıcaklık, ağırlık). İkili özellikler ise belirli kategorilerden oluşan verileridir (ör. cinsiyet, şehir).

2.1.1. Veri Seti

Bu çalışmada Kaggle platformunda bulunan “Instagram fakespammergenuineaccounts” veri seti kullanılmıştır [9]. Veri setindeki yinelenen satırlar silinmiş olup, %30 test ve %70 eğitim olarak iki kısma ayrılmıştır. Train setimiz 574, test setimizde 118 veri noktasından oluşmaktadır. Veri setinde bulunan nitelikler aşağıdaki gibidir ;

- profilepic(profil fotografı) : Profil resminin olup olmadığını gösteren özellik.
- nums/lengthusername (sayılar/kullanıcı adı uzunluğu) : Bir hesabın isminin toplam karakter uzunluğunu gösteren özellik.
- fullnamewords (tam ad sözcükleri) : Hesap sahibinin adına ait kelimelerin toplamını içeren özellik.
- nums/length fullname (sayılar/tam ad uzunluğu) : Kişinin tam adının toplam uzunluğuna oranını belirten özellik.
- name == username (ad ==kullanıcı adı): Kişinin adının kullanıcı adı ile eşleşip eşleşmediğini gösteren özellik.
- descriptionlength (açıklama uzunluğu) : Profil açıklaması özelliği.
- external URL (harici url) : Biyografide harici bir web url adresi bilgisini gösteren özellik.
- private (özel) : Profilin gizli olup olmadığını gösteren özellik.
- posts (gönderiler) : Hesabın toplam gönderi sayısı.
- followers (takipçiler): Hesabın toplam takipçi sayısı.
- follows (takipler) :Hesabın toplam takip ettiği kişi sayısı.

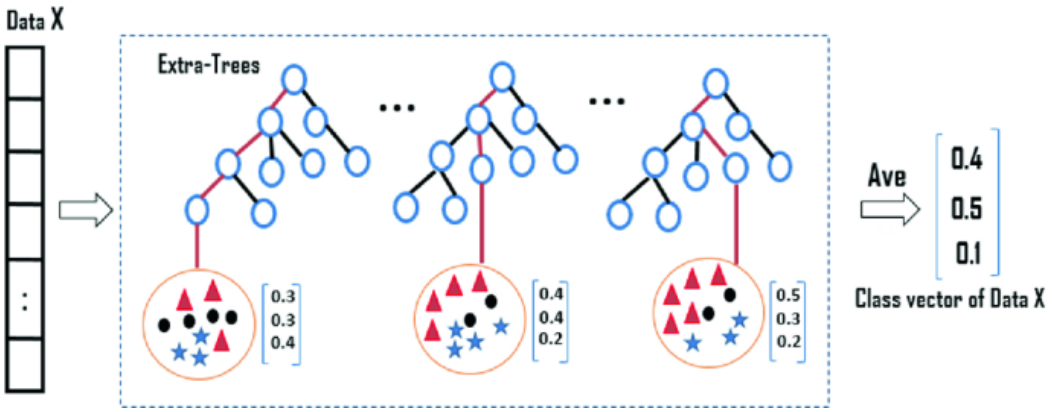
- fake (sahte): Bir hesabın gerçek veya sahte olduğunu gösteren özellik burada 0 gerçek bir hesabı, 1 ise sahte bir hesabı göstermektedir.

2.2. YÖNTEM

Uygulamaların gerçekleştirilmesinde, Python programlama dili tercih edilmiştir. Kodlama ortamı olarak veri eğitimi Kaggle platformunda tamamlanıp Visual Studio programında da kullanıcı ara yüzü oluşturulup proje bir bütün haline getirilmiştir. Python, Guidovan Rossum tarafından ilk sürümü 1991'de ortaya konan genel amaçlı bir programlama dilidir. Diğer dillere göre öğrenim kolaylığı ve geniş kütüphane desteğiyle oldukça yaygın kullanıcı kitlesine ulaşmıştır [10].
Kodlama için kullanılan Kaggle**;** veri bilimcilerin ve makine öğrenmesi ile ilgilenenlerin kullandığı çevrimiçi bir platformdur. Kaggle, 2010 yılında Anthony Goldbloom ve Jeremy Howard tarafından kurulmuştur. Ardından, 2017'de Google tarafından satın alınmıştır. 8 milyondan fazla kullanıcısı olan Kaggle'ın asıl amacı**,** veri bilimine önem veren kişileri bir araya getirmek ve onları amaçlarına ulaşmalarını sağlamaktır. Kaggle'ın veri bilimciler tarafından bu kadar popüler olmasının bir nedeni de içerisinde yarışmalara yer vermesidir [11]. Daha sonra ara yüz tasarımı için kullanılan Visual Studio, Microsoft'un geliştirdiği entegre bir geliştirme ortamıdır (IDE). Kod yazmak, düzenlemek, hata ayıklamak ve derlemek ve ardından uygulamanızı dağıtmak için kullanabileceğiniz kapsamlı bir tümleşik geliştirme ortamdır. Visual Studio, farklı programlama dilleriyle uyumludur ve geniş bir eklenti ve araç yelpazesi sunar [12]. Bu sayede yazılım geliştiricilerin verimliliğini artırmak için birçok özellik sunar.
Çalışmamızda ele alınan Rastgele Orman, Ekstra Ağaçlar, Gradyan Arttırma, Karar Ağacı, AdaBoost gibi ağaç yapısına sahip algoritmaların sınıflandırma yöntemlerine kısaca bakacak olursak;

2.2.1. Ekstra Ağaçlar

Ekstra Ağaçlar (Extra Trees), makine öğrenimi alanında kullanılan bir topluluk öğrenme algoritmasıdır. Karar Ağaçları yöntemine dayalı olan bu algoritma Rastgele Orman yöntemine benzer bir şekilde çalışır. Fakat Rastgele Orman'dan farklı olarak ağaçların oluşturulma şeklinde daha fazla rastgelelik mevcuttur [13]. Böylece modelin daha yüksek bir hesaplama verimliliği ve daha iyi bir performans elde edilmesi sağlanmıştır. Bu algoritmada Şekil 2'de görüldüğü gibi veri setinde rastgele alt kümeler oluşturularak her alt küme için ağaçlar oluşturulur. Daha sonra bu ağaçların tahminleri birleştirilerek sonuç elde edilir. Veri setinin karmaşıklığını ve aşırı uyum sorununu önlemek amaçla bu algoritma kullanılabilir. Bu algoritma, genellikle sınıflandırma ve regresyon problemlerinde kullanılmaktadır.



Şekil 2.*Ekstra ağaçlar [14]*

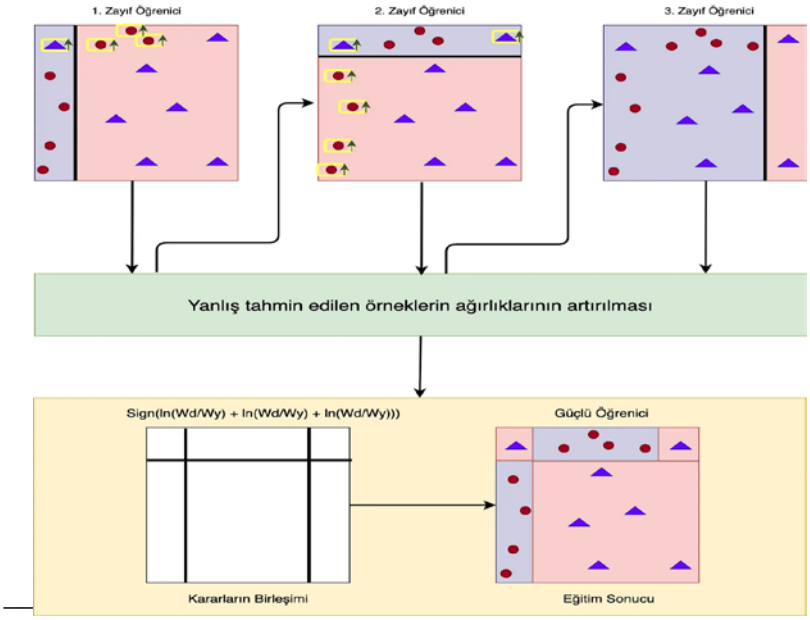
2.2.2. Gradyan Arttırma

Gradyan Arttırma yöntemi zayıf öğrenicileri (genellikle karar ağaçları bir araya getirerek) güçlü öğrenicilere dönüştürme yöntemidir. Her zayıf öğrenici, önceki öğrenicinin hatalarını düzeltmeye

odaklanarak eğitilir. Bu nedenle, Gradyan Arttırma modelleri genellikle yüksek doğruluk sağlayan güçlü bir tahmin modeli oluşturur [15]. Özellikle büyük ve karmaşık veri kümelerinde tahmin hızı ve doğruluğuyla öne çıkan bir yöntemdir. Kaggle yarışmalarından üretilen makine öğrenimi çözümlerine kadar bu algoritmanın en iyi sonuçları ürettiği gözlemlenmişti.

2.2.3. Adaboost

Adaboost olarak kısaltılmış Adaptive Boosting makine öğrenme alanında toplu öğrenme yapan bir algoritmadır. Burada da asıl amaç zayıf öğrenicileri bir araya getirip güçlü bir öğrenici elde etmektir. Bu algoritmada her bir zayıf öğrenici tek tek eğitilerek her bir modelin performansına göre ağırlıklar vererek çalışır. Sonraki eğitim sırasında ilk tahminde yanlış öğrenilen veriler daha fazla öncelik verilerek eğitilir. Sonuç olarak zayıf öğrenicileri bir araya getirip daha başarılı güçlü sonuçlar elde edilir. Adaboost ilk boosting algoritmalarından sayılmaktadır [16]. Adaboost'un çalışma prensibi Şekil 3'de gösterilmiştir.



Şekil 3. *AdaBoost çalışma prensibi [16]*

2.2.4. Karar Ağacı

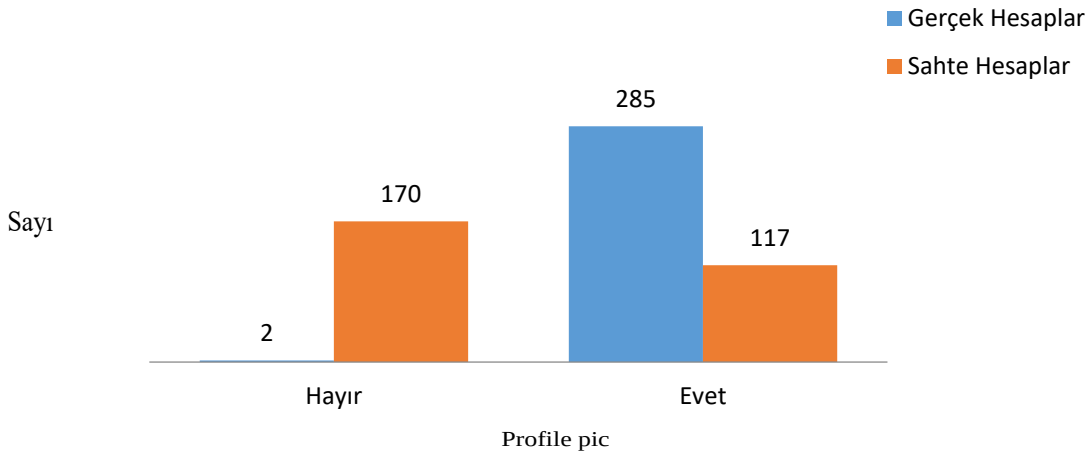
Karar Ağacı gözetimli öğrenmede yaygın olarak kullanılan bir algoritma türüdür [16]. Makine öğrenme metotlarında, "Karar Ağacı" kavramı hem sınıflandırma hem de regresyon modelleri ile birlikte kullanılabilir. Öte yandan, yöneylem araştırmasında "Karar Ağacı" kavramı kararların silsile olarak sıralanmasını ve bu kararların sonuçlarını göstermek için kullanılmaktadır. Karar Ağaçları sınıflandırma için kullanılıyorsa sınıflandırma ağacı, regresyon için kullanılıyorsa regresyon ağacı olarak adlandırılmalıdır [18]. Karar Ağacı, finans, pazarlama, mühendislik ve tıp alanında kullanılan makine algoritmalarındandır [19].

2.2.5. Rastgele Orman

Rastgele Orman (RO) algoritması başarılı sınıflandırma yöntemlerinden biri olarak bilinir. Doğası gereği çok farklı disiplinlere hitap etmesinden dolayı, RO farklı alanlarda çalışan araştırmacıların dikkatini çekmektedir [20]. Rastgele Orman, birden fazla karar ağacının oluşturulup birlikte kullanılmasıyla yapılan bir makine öğrenimi yöntemidir. Her ağaç, rastgele seçilen veri alt kümeleriyle eğitilir ve sonuç olarak ağaçların oylarıyla nihai tahmin yapılır. Bu yöntem, genelde modelin doğruluğunu artırır ve aşırı öğrenmeyi azaltır.

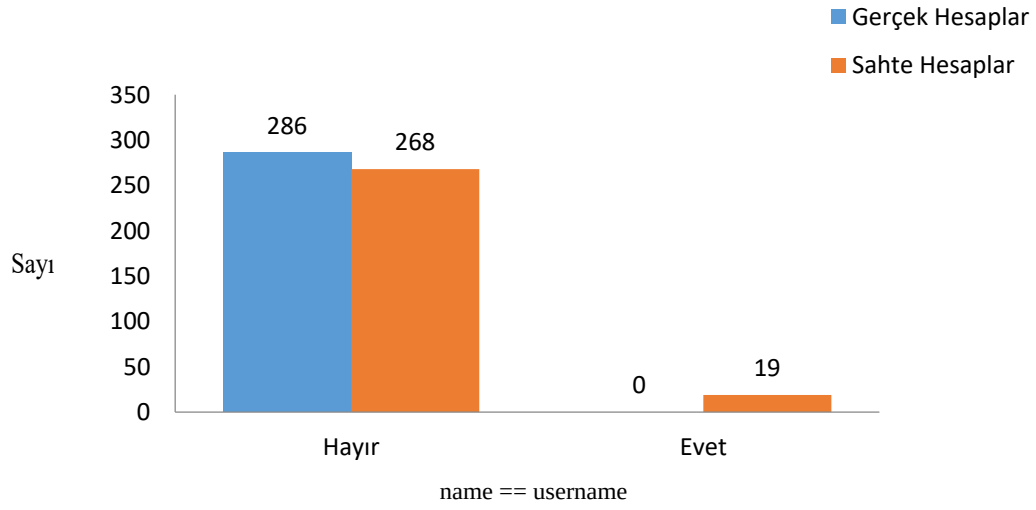
3. ARAŞTIRMA SONUÇLARI VE TARTIŞMA (RESEARCH RESULTS AND DISCUSSION)

Deneysel çalışmalarda Rastgele Orman, Ekstra Ağaçlar, Gradyan Arttırma, Karar Ağacı, AdaBoost ve bu algoritmalar kullanılarak topluluk öğrenimi (Ensemble Learning) yaklaşımında yararlanılmıştır. Bu algoritmaların ele alınmasının nedeni sık kullanılan algoritmalar kullanmak yerine, farklı bir çalışma sunmak ve ağaç yapıları nedeniyle zayıf öğrenicilerin tekrar tekrar eğitilerek güçlü birer öğrenici haline gelip yüksek başarı sonuçları elde edilmesini sağlamaktır. Bu algoritmalar ile büyük ve karmaşık veri kümelerinde bile yüksek sonuçlar elde etmek mümkündür. Veri setinin eğitimi ve sınıflandırma sürecinde ilk olarak verilerimize ön işleme aşamaları uygulanmıştır. Ön işleme aşamasında veri setinde bulunan veriler ikili ve sürekli özellikler olarak kategorize edilmiştir. İkili özellik sonuçlarının grafikleri Şekil 4 ve Şekil 7 arasında gösterilmiştir.



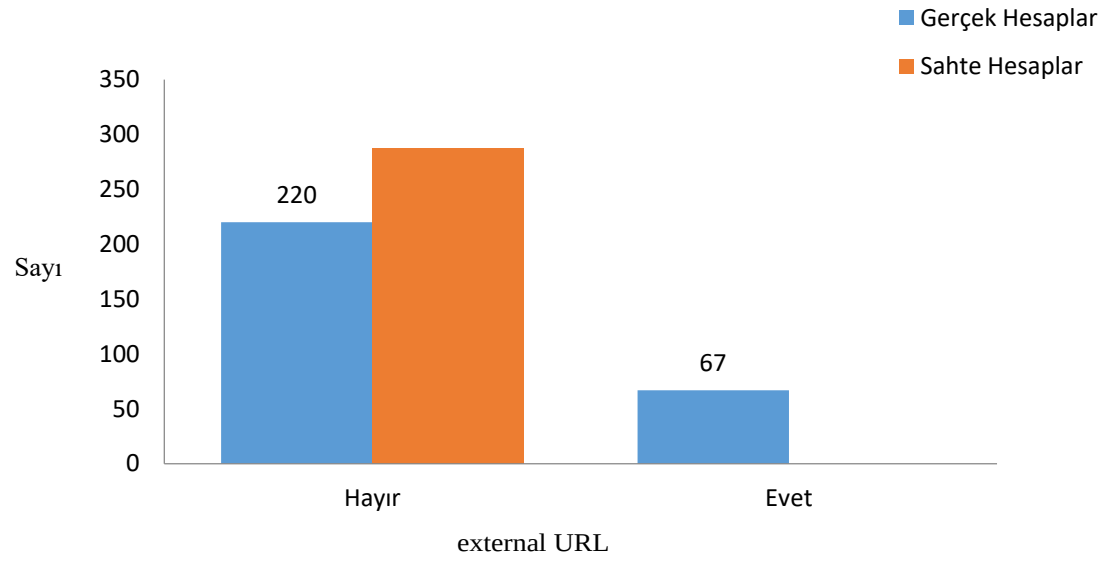
Şekil 4. *Hesabın profil resmi var mı?*

Şekil 4'te profil resmi olmayan hesapların çoğunluğunun (%98) sahte hesaplar olduğu gözlemlenmiştir.



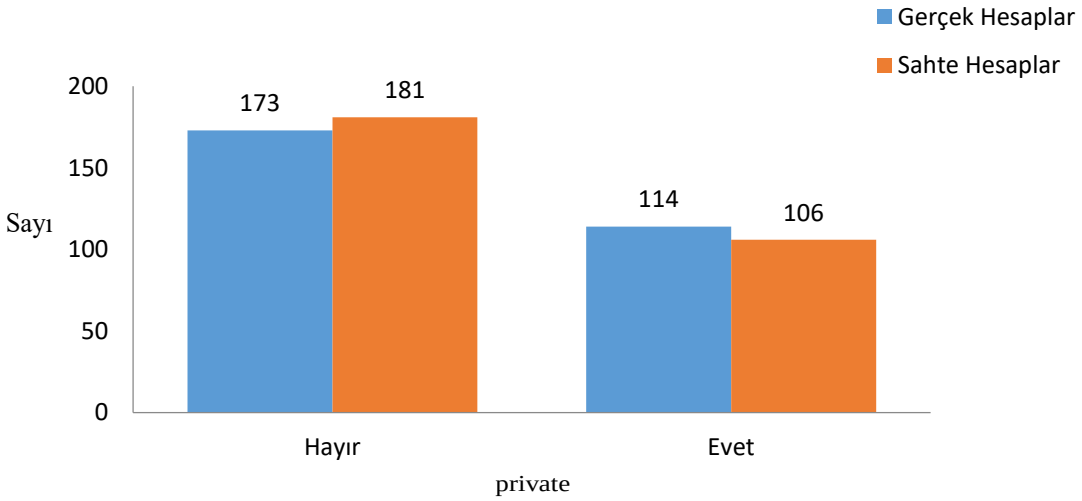
Şekil 5. *Hesabın adı kullanıcı adı ile eşleşiyor mu?*

Şekil 5'te genellikle, adı kullanıcı adıyla aynı olan hesapların da sahte hesap olduğu gözlemlenmiştir.



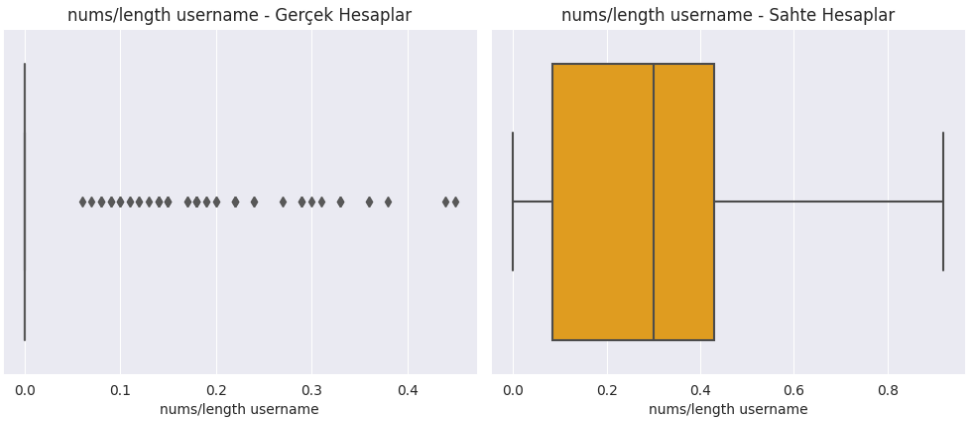
Şekil 6. *Hesap harici bir URL'ye erişim sağlıyor mu?*

Şekil 6'da şaşırtıcı bir şekilde, biyografilerinde dış URL'ye sahip olan hesapların tümü gerçek hesaplarmış gibi görünmektedir. Genel olarak, çoğu hesabın dış bir URL'si bulunmadığı gözlemlenmiştir.



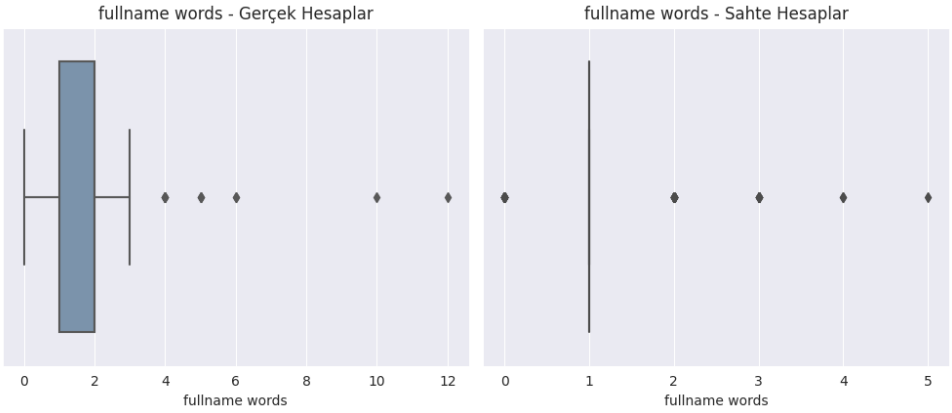
Şekil 7. *Özel hesaplar gerçek mi?*

Şekil 7'de özel olmayan hesapların çoğunun sahte hesaplar olduğu görülmektedir, grafikteki gibi gerçekten de özel hesapların çoğunun gerçek hesaplar olduğu bilinmektedir. Ve aynı şekilde takipçi ile takip edilen sayısı eşit olan hesapların da gerçek hesaplar olduğu gözlemlenmiştir.
Tüm ikili özellikleri grafikledikten sonra sürekli özellikler ele alınmıştır. Sürekli özelliklere ait grafikler Şekil.8 ve Şekil.13 aralığında gösterilmiştir. Bu şekilde sahte hesapların Instagram'daki gerçek hesaplarla karşılaştırıldığında ne kadar farklı davrandığını görebiliriz.



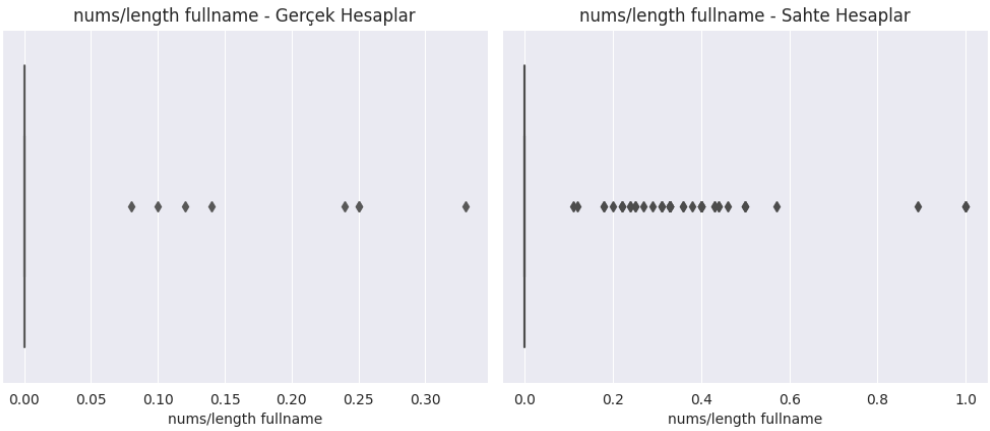
Şekil 8. *Sayısal karakterler/kullanıcı adının uzunluğu*

Şekil 8'de görüldüğü üzere sahte hesapların kullanıcı adlarında genellikle daha fazla sayısal karakter oranı bulunduğu gözlemlenmiştir.



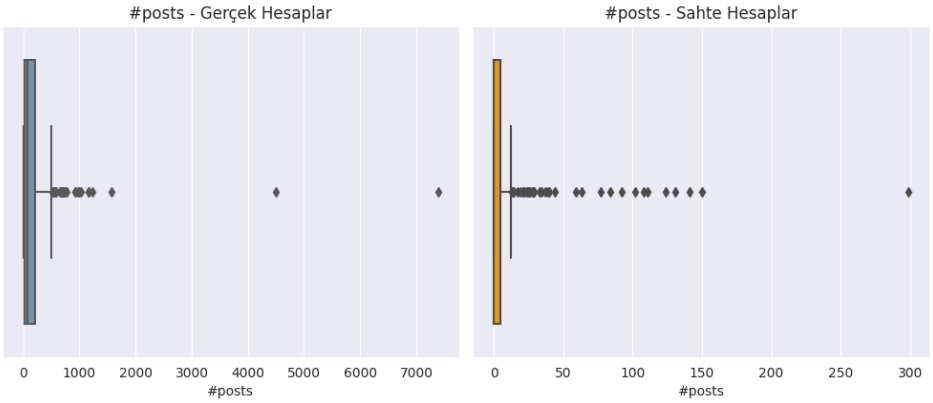
Şekil 9. *Tam isimdeki kelime sayısı*

Şekil 9'da ise sahte hesapların, tam adlarında daha az kelimeye sahip olma eğiliminde olduğu gözlemlenmiştir.



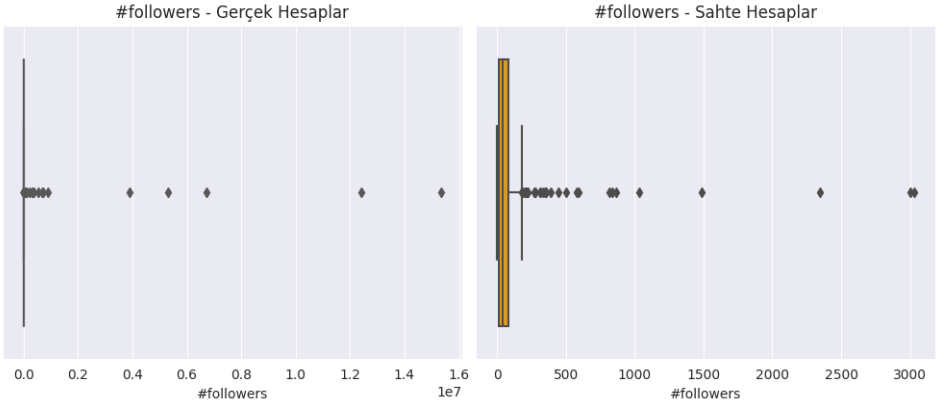
Şekil 10. *Sayısal karakterler/tam adın uzunluğu*

Şekil 10'da aykırı değer (outliers) hariç tutulduğunda, sahte hesapların açıklamaları/biyografilerinin genellikle çok daha kısa olduğu görülmektedir.



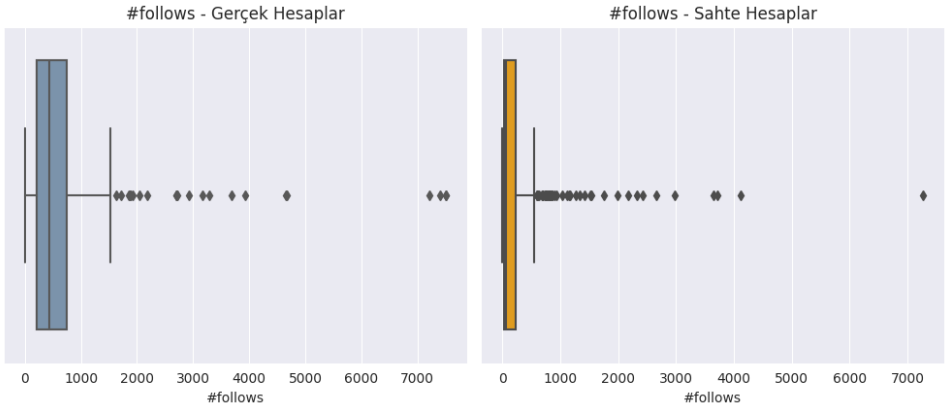
Şekil 11. *Toplam gönderi sayısı*

Şekil 11'e baktığımız zaman gerçek hesapların, sahte hesaplara göre çok daha fazla gönderiye sahip olduğu görülmektedir.



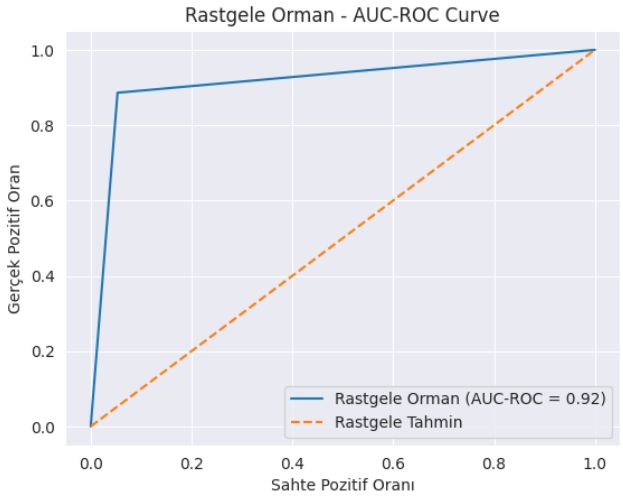
Şekil 12. *Toplam takipçi sayısı*

Şekil 12'de ise gerçek hesapların, sahte hesaplardan çok daha fazla takipçiye sahip olduğu gözlemlenmiştir.

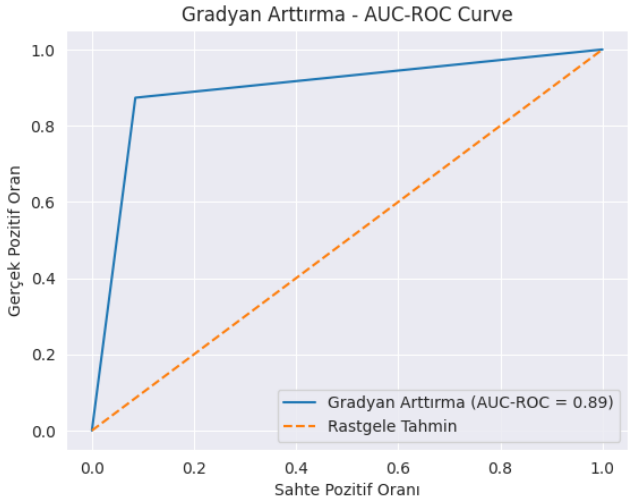


Şekil 13.*Hesabın takip ettiği toplam kişi sayısı*

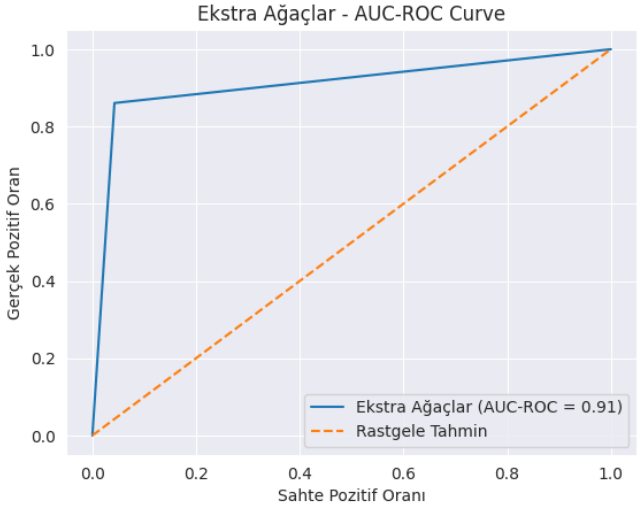
Şekil 13'de görüldüğü üzere genel olarak, aykırı değerler (outliers)çıkarıldığında, ortalama olarak çoğu sahte hesabın gerçek hesaplara kıyasla daha az kişiyi takip ettiği görülmektedir.



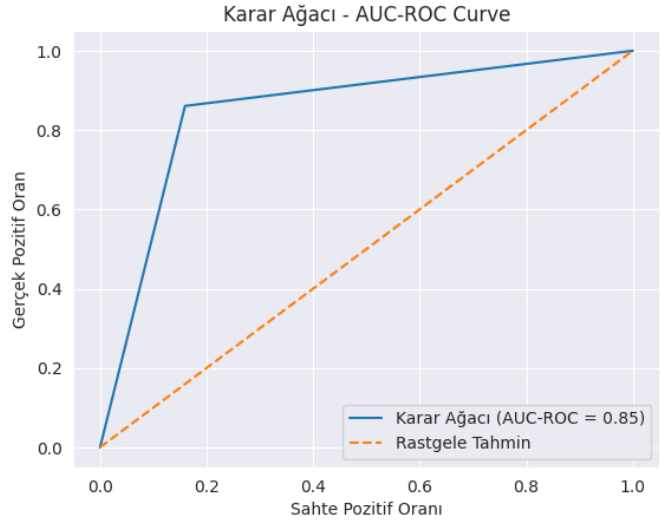
Şekil 14.*Rastgele orman Auc-Roc eğrisi*



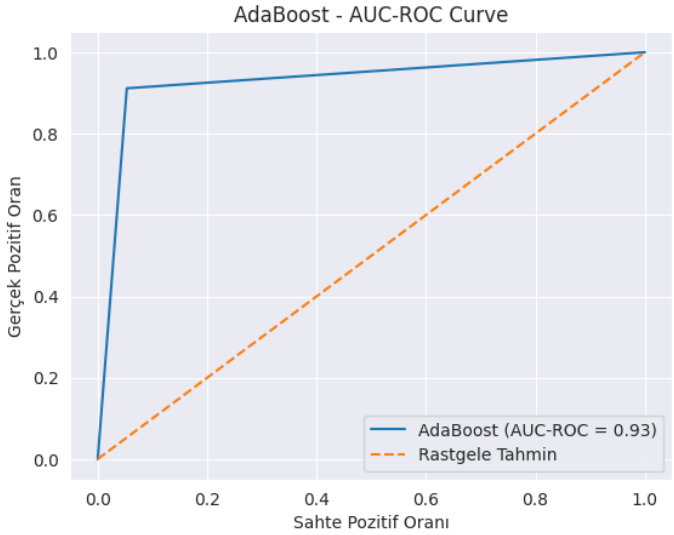
Şekil 15. *Gradyan arttırma Auc-Roc eğrisi*



Şekil 16. *Ekstra ağaçlar Auc-Roc eğrisi*



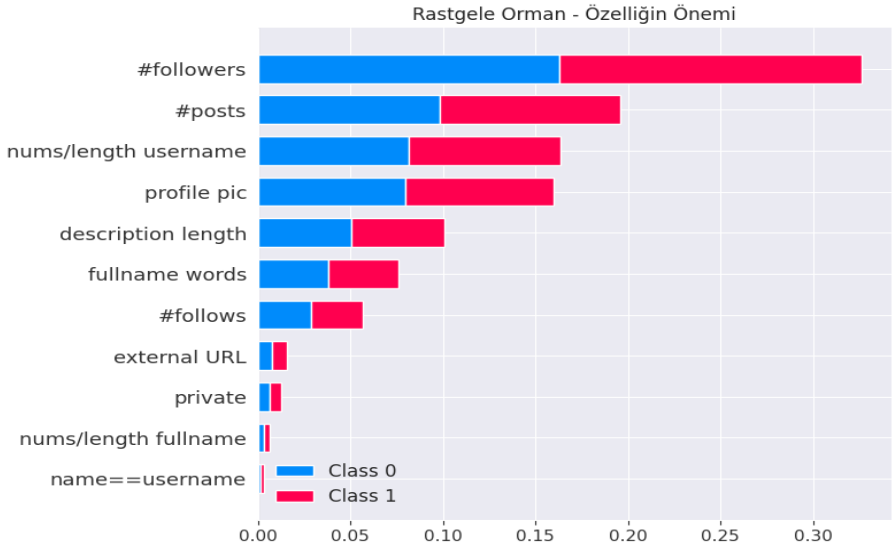
Şekil 17. *Karar ağacı Auc-Roc eğrisi*



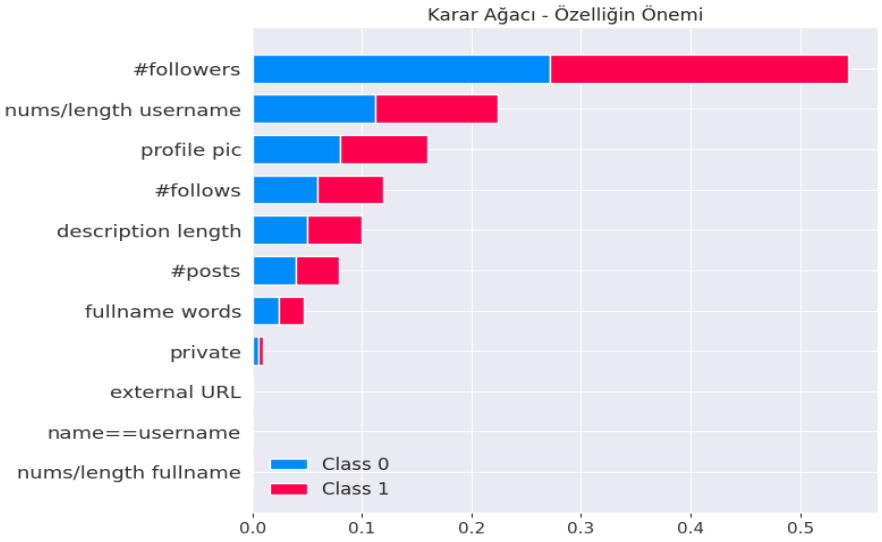
Şekil 18. *AdaBoostAuc-Roc eğrisi*

AUC-ROC eğrisi sınıflandırma problemlerinde modelin performansını ölçmek için kullanılan bir tekniktir. Mükemmel sınıflandırma modeli için 1'e, rastgele tahmin eden bir model için ise 0,5'e yakın değerler alır. Bizim modelimizde de temel olarak amaçladığımız 1,0 yakın bir AUC-ROC eğrisidir.

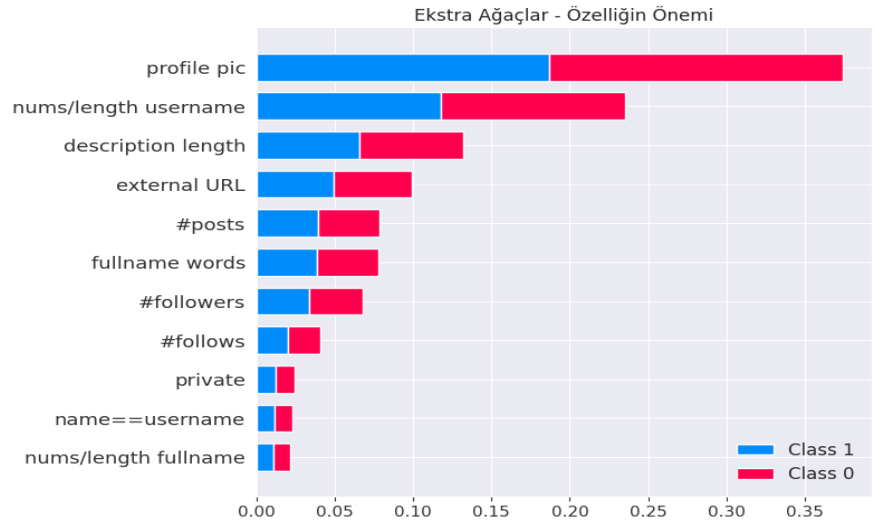
Şekil 14-18 aralığında beş modelin Auc-Roc sonuçlarına baktığımızda, her birinin tahmin başarısı ve özelliklerin önem sıralaması açısından farklılık gösterdiği görülmektedir. "Rastgele Orman" modelinin, özellikle sahte hesapları tespit etmede yüksek doğrulukla öne çıktığı gözlemlenmiştir. "Karar Ağacı" ise Karar Ağacının doğası gereği, belli özelliklerin üzerinde yoğunlaşarak tahminlerde bulunduğu gözlemlenmiştir. Genel olarak hepsinin başarılı sonuç verdiği görülmektedir. Şimdi, özellik seçimi için, hedef değişkeni tahmin etmede her bir özelliğin gösterdiği önemi içeren grafiği çizmek üzere SHAP kütüphanesini kullanabiliriz. SHAP kütüphanesi ile model sonuçlarını etkileyen en önemli özellikleri ve bunların katkı seviyelerini gösterebiliriz.



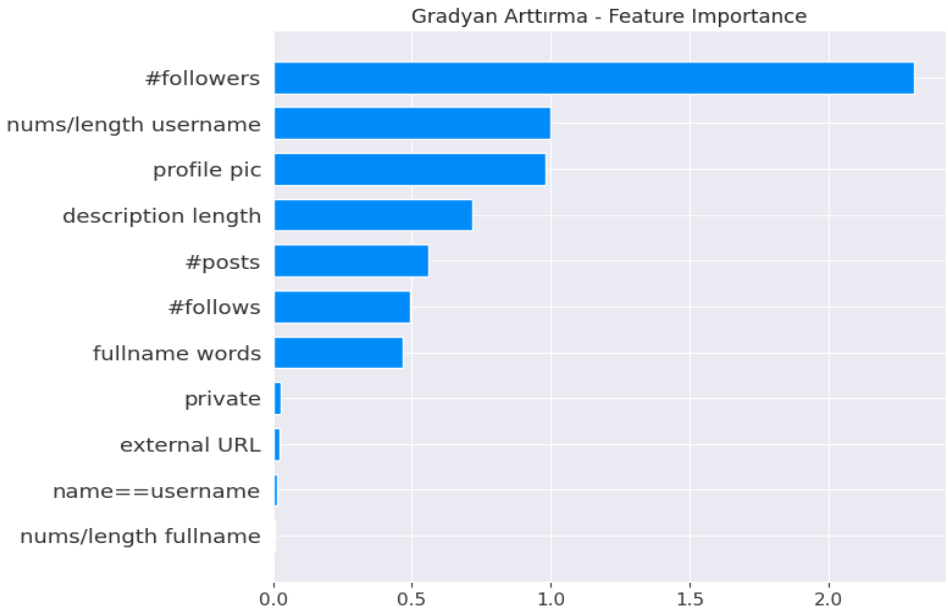
Şekil 19. *Rastgele orman özellik analizi*



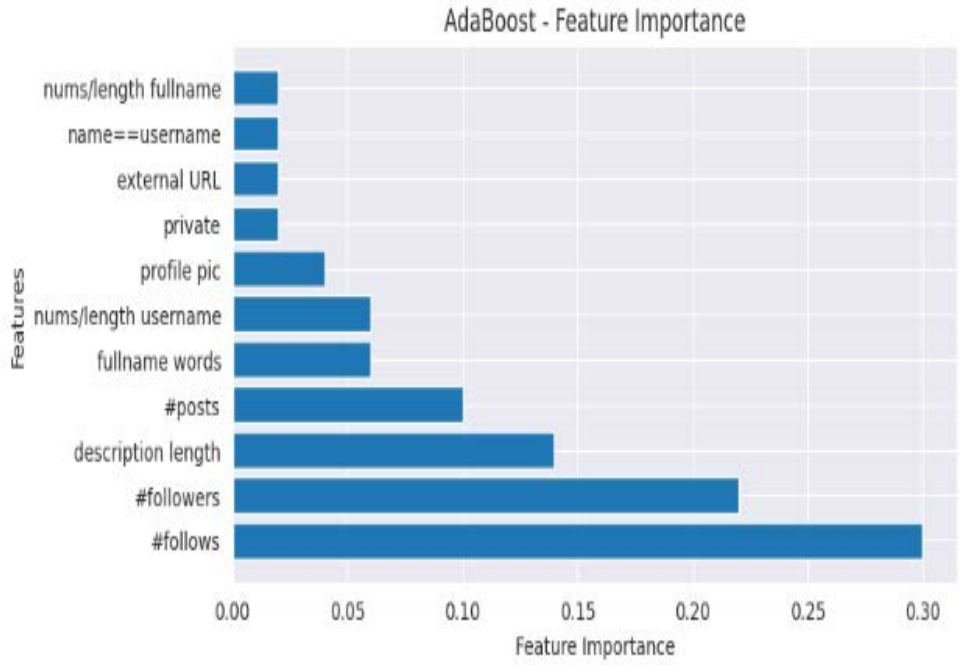
Şekil 20. *Karar ağacı özellik analizi*



Şekil 21. *Ekstra ağaçlar özellik analizi*



Şekil 22. *Gradyan arttırma özellik analizi*



Şekil 23. *AdaBoost özellik analizi*

Şekil 19-23 arasında yer alan özellik analizi sonuçlarında görüldüğü gibi genel olarak, tüm özelliklerin çıktıyı bir şekilde etkilediği görülüyor, bu nedenle hiçbir özelliğin gereksiz olmadığını söyleyebiliriz. En önemli özellikleri öğrendiğimize göre, şimdi bu özelliklerden bazılarını kullanarak yeni özellikler oluşturabiliriz. Amacımız, son modelimizin tahmin yeteneklerini daha da güçlendirecek özellik mühendisliği yapmaktır.

Özellik Mühendisliğinde yeni özellikler oluşturmak için öncelikle takipçi ve takip edilen oranına bakıldığında, sahte hesapların çoğunun kendilerinin takip ettiği kişi sayısından çok daha az takipçiye sahip olduğu görülmektedir. Bu nedenle bu ikili özellik, her hesap için 0 veya 1 olarak etiketlenecektir. Burada 1, belirli bir hesabın takip ettiğinden daha fazla takipçisi olduğunu

gösterecektir. Bu da gösteriyor ki takipçi_takip edilen hedef tahmini için önemli bir özellik oluşturabilir.

Tablo 1. *Veri Üzerinde özellik analiz sonucu*

	profile pic	nums/length username	Full name words	nums/lengthfullname	name=username	Description length	external URL	Private	Posts	Followers	follows	fake	aktivite_orani	takipci_takip Edilen
571	1	0.55	1	44	0	0	0	0	33	166	596	1	0.20	0
572	1	0.38	1	0.33	0	21	0	0	44	66	75	1	0.67	0
573	1	0.57	2	0.00	0	0	0	0	4	96	339	1	0.04	0
574	1	0.57	1	0.00	0	11	0	0	0	57	73	1	0.00	0
575	1	0.27	1	0.00	0	0	0	0	2	150	487	1	0.01	0

Çalışmamızda Tablo 1'de sonucu gösterilen özellik mühendisliğini gerçekleştirdikten sonra modelleme öncesi son aşama olarak her bir özelliğin ortalaması ve standart sapması Tablo 2'deki gibi hesaplanmıştır.

Tablo 2. *Veri özellik analiz standart sapması*

	Ortalama	**Standart Sapma**
profilepic (profil fotoğrafı)	0.72	0.45
nums/lengthusername (sayı/kullanıcı adı uzunluğu)	0.16	0.21
fullnamewords (tam ad kelimesi)	1.47	1.06
nums/lengthfullname (sayı/tam ad uzunluğu)	0.04	0.13
name==username (ad=kullanıcıad)	0.03	0.18
descriptionlength (açıklama uzunluğu)	23.27	38.13
external URL (harici URL)	0.12	0.33
private(özel)	0.37	0.48
#posts(paylaşım)	110.94	408.69
followers(takipçi)	88366.21	926256.64
follows(Takip)	519.88	930.82
fake(sahte)	0.48	0.50
aktivite_oranı	0.28	1.64
takipci_takipEdilen	0.43	0.5

Son olarak hiperparametre ayarlamaları ile modelin uygun değerleri test ve eğitim seti için ayrı olarak elde edilmiştir. Yeni yaklaşımımız için eğittiğimiz beş modelin iki farklı makine öğrenim tekniği olan VotingClassifier ve StackingClassifier yöntem ile entegrasyonu sağlandı. Bu yöntemler, birden fazla modelin tahminlerini bir araya getirerek çoğunluğun (veya ağırlıklı oyların) kararını nihai tahmin olarak belirler.

Bu şekilde, tek tek doğruluğunu elde ettiğimiz farklı modellerin güçlü yanları bir araya getirilerek daha iyi bir genelleme performansı elde edilmiştir. Sonuç olarak entegre modellerimizde elde edilen tahmin değerlerinin eğitim verisi kullanılarak VotingClassifier %90 ve StackingClassifier %91'lik, test

verisi kullanılarak VotingClassifier %91 ve StackingClassifier %91'lik olacak şekilde yüksek başarı sonuçları elde ettiği gözlemlenmiştir.

Tablo 3. *Eğitim ve test verisi üzerinde AUC-ROC değeri hesaplaması sonucu*

Algoritmalar	Eğitim Verisi İle AUC-ROC değeri	Test Verisi İle AUC-ROC değeri
Rastgele Orman	0.90	0.94
Gradyan Arttırma	0.91	0.92
Ekstra Ağaçlar	0.91	0.92
Karar Ağacı	0.82	0.88
AdaBoost	0.89	0.90
VotingClassifier	0.90	0.91
StackingClassifier	0.91	0.91

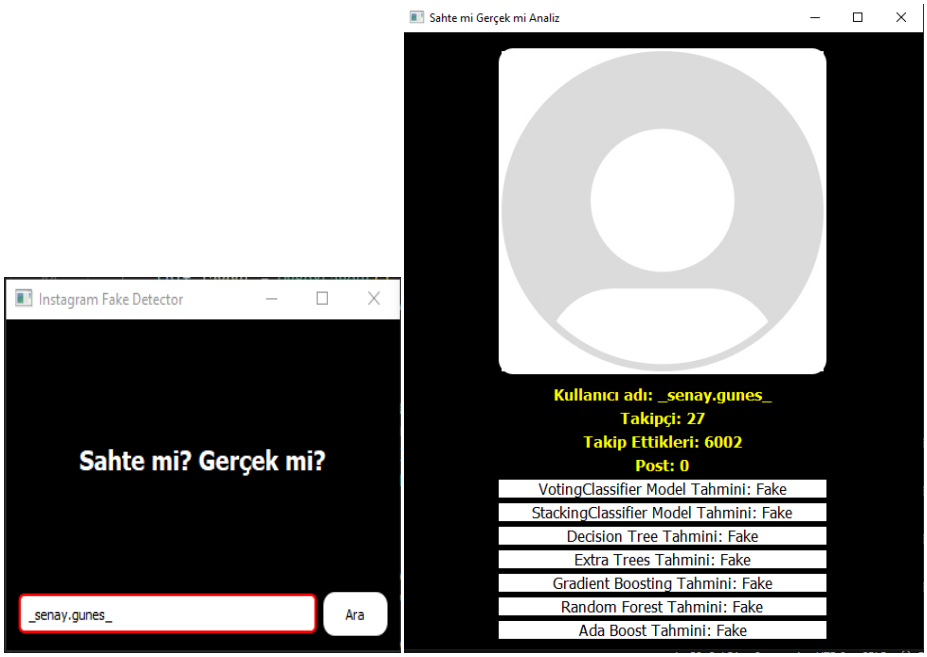
Tablo 3'te AUC-ROC grafiğinden elde edilen değerlerde tüm algoritmalarımızda başarılı sonuçlar elde edildiği gözlemlendi. Bu nedenle tüm algoritmalarımızı entegre ederek oluşturduğumuz VotingClassifier ve StackingClassifier AUC-ROC değerlerini uygulamamızda kullanabileceğimizin kararını verebiliriz.

4. SONUÇ (CONCLUSION)

Sonuç olarak, çalışmamızda çok sayıda veri işleme ve veri analizi gerçekleştirilmiştir. Gerçek hesapların, sahte hesaplardan çok daha fazla gönderi paylaştığı, sahte hesapların takip etme oranının gerçek hesaplardan daha yüksek olduğu, sahte hesapların açıklamalarının (biyografilerinin) genellikle daha kısa olduğu ve sahte hesapların kullanıcı adlarında genellikle daha fazla sayısal karakter bulunduğu gibi özelliklerle gerçek ve sahte hesaplar arasındaki farklar başarılı bir şekilde tahmin edilmiştir. Ayrıca, daha önce yapılmış çalışmalardan farklı olarak, sistem dışında yüksek başarı sonuçları veren sıklıkla kullanılmayan güncel algoritmalar kullanılarak literatüre katkı sağlanması hedeflenmiştir. Yüksek başarım gösteren Rastgele Orman ile aynı mantıkta çalışan algoritmalar ele alındı. Sonuç olarak, tüm algoritmalar%80doğruluk değerinin üzerinde sonuç vermiştir. Eğitim verileri üzerinde Ekstra Ağaçlar, Gradyan Arttırma ve StackingClassifier kullanılarak %91 ile en yüksek başarı sonucu elde edildiği gözlemlenmiştir. Test verileri üzerinde ise Rastgele Orman %94 oranıyla en yüksek başarımı göstermiştir. Son olarak, ara yüz ile sunduğumuz model gerçek zamanlı olarak son kullanıcı tarafından test edilmiş, sahte/gerçek olup olmadığı belirlenmek istenen hesabın kullanıcı adı girilerek belirlenmesi sağlanmıştır. Bu sonuçlarda da, Şekil 23 ve Şekil 24'te gösterildiği gibi sahte ve gerçek hesapların yüksek doğrulukla tespit edildiği gözlemlenmiştir.

Sahte hesap tespitinde kullanılan algoritma teknikleri oldukça önemli bir yer almaktadır. Ancak veri seti içeriğinin de (takip-takipçi sayısı, biyografi bilgileri, kullanıcı adı, url linki, profil fotoğrafları vs.) algoritmalar kadar önemli olduğu gözlemlenmiştir. Bu nedenle yapılacak çalışmalarda sahte Instagram hesabı tespit modeli oluşturmak için gerçek ve sahte hesapların etiketli olarak yer aldığı büyük ve çeşitli bir veri setine, dikkatli bir özellik mühendisliğine ve model seçimine ihtiyaç duyulmaktadır. Ayrıca, sahte hesapların zaman içinde evrimleşip değiştiği göz önüne alındığında, modelin sürekli olarak güncellenmesi başarı sonucunu arttırmak için önemlidir. Daha büyük ve güncel bir veri seti kullanımı sahte/gerçek hesap tespiti başarımını arttıracaktır. Sıklıkla kullanılan Kümeleme, Rastgele Orman, Naive Bayes, Karar Ağaçları, Lojistik Regresyon, Destek Vektör Makineleri gibi algoritmaların topluluk öğrenimi yöntemleriyle daha yüksek performans göstermesi sağlanabilir.

Şekil 23. *Gerçek hesap sonucu*



Şekil 24. *Sahte hesap sonucu*

ÇIKAR ÇATIŞMASI (CONFLICTOFINTEREST)

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

KAYNAKÇA (REFERENCES)

[1] Y. S. Wu, S. Bagchi, N. Singh, R. Wita, Spam detection in voice-over-ip calls through semi-supervised clustering. Proceedings of the 2009 Dependable Systems Networks, 307–316, 2009.

[2] H. Uzun, Sosyal Medyanın Bilgi Kalitesine Etkisi: Sahte Hesaplar, Akademia Doğa ve İnsan Bilimleri Dergisi, 2:1(2016) 1-31.

[3] A. Gupta, R. Kaushal, Improving spam detection in Online Social Networks,2015 International Conference on Cognitive Computing and Information Processing (CCIP)**,** Noida, India, 1-6. doi: 10.1109/CCIP.2015.7100738, 2015.

[4] A. H. Wang, Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach. In S. Foresti ve S. Jajodia (Eds.), Data and Applications Security and Privacy XXIV. DBSec 2010. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. 6166 (2010) 25-39. doi:10.1007/978-3-642-13739-6_25.

[5] K. R. Purba, K. D. Asirvatham, R. K. Murugesan, Classification of Instagram Fake Users Using Supervised Machine Learning Algorithms, International Journal of Electrical and Computer Engineering (IJECE)**,** 10:3, (2020) 2763-2772. doi: 10.11591/ijece.v10i3.

[6] Ü. Tunç, E. Atalar, M. S. Gargı, Z. Ergül Aydın, Classification of fake, bot, and real accounts on ınstagram using machine learning, Politeknik Dergisi**,** 2024

[7] A. Anwaar, A. Wajid, A Simplified Study of Fake and Real Accounts: Using Prediction, Cluster and Outlier Detection Methods, 2020.

[8] A. Dey, H. Reddy, M. Dey, N. Sinha, Detection of Fake Accounts in Instagram Using Machine Learning, International Journal of Computer Science and Information Technologies**,** 10:10 (2019) 82-86. doi: 10.5121/ijcsit.2019.1150783.

[9] Free4ever1, Instagram Fake, Spammer, and Genuine Accounts Dataset. Kaggle. Erişim adresi:https://www.kaggle.com/datasets/free4ever1/instagram-fake-spammer-genuine-accounts, 2024.

[10] B. Malkoç, Temel bilimler ve mühendislik eğitiminde programlama dili olarak Python, XIV. Akademik Bilişim Konferansı Bildirileri, 201, 2012.

[11] CoderSpace,Kaggle. Erişim tarihi: https://coderspace.io/sozluk/kaggle, 2023.

[12] Visual Studio,Visual Studio ile çalışmaya başlama,Erişim tarihi: https://learn.microsoft.com/tr-tr/visualstudio/get-started/visual-studio-ide?view=vs-2022, 2022.

[13] N. Bhandari, Extra Trees Classifier, Medium: https://medium.com/@namanbhandari/extratreesclassifier8e7fc052c7, 2018

[14] A. Berrouachedi, R. Jaziri, G. Bernard, Ekstra ağaçların derin çağlayanı. In U. Lauw, H. (Eds.), Bilgi Keşfi ve Veri Madenciliğinde Eğilimler ve Uygulamalar. PAKDD 2019. Bilgisayar Bilimlerinde Ders Notları. 11607 (2019) 25-39. Springer, Cham. doi:10.1007/978-3-030-26142-9_11.

[15] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, 2013.

[16] K. Güzel, Boosting Nedir? Adım Adım Adaboost Algoritması, Medium: https://kadirguzel.medium.com/boosting-nedir-ad%C4%B1m-ad%C4%B1m-adaboost-algoritmas%C4%B1-439cce20ab9a, 2022.

[17] T. Kalaycı, Kimlik hırsızı web sitelerinin sınıflandırılması için makine öğrenmesi yöntemlerinin karşılaştırılması, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi**,** 24:5 (2018) 870-878. doi:10.9733/jgg.241212.1t.

[18] S. Atan, KNN, Naive Bayes ve karar ağacı makine öğrenme algoritmaları, Bu algoritmaların sosyal bilimlerde kullanım imkânları. 2020. doi:10.31235/osf.io/8r5pu.

[19] L. Rokach, O.Z. Maimon, Data Mining with Decision Trees: Theory and Applications. World Scientific Publishing Co., Inc., Singapore, 2008.

[20] Ö. Akar, O. Güngör, Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. Jeodezi ve Jeoinformasyon Dergisi, 106 (2012) 139-146.