

Instagram Hesap Sınıflandırması: Fenomen ve Şirket Profillerinin Ayırıştırılması

Instagram Account Classification: Influencer vs. Company Profiles

Ozgun Yargi
Sabanci University
ozgun.yargi@sabanciuniv.edu

Onur Varol
Sabanci University
onur.varol@sabanciuniv.edu

Özetçe —Sosyal medya platformlarının kullanımının zaman içinde artmasıyla, bu platformlarda bulunan fenomen ve şirket gibi, kitleleri etkileyen hesapların sayısı da artmıştır. Şirketlerin, hedef kitlelere ulaşmalarındaki etkinliklerinden ötürü fenomenleri hızlıca tespit edebilecekleri sistemlere olan ihtiyacı artmıştır. Bu çalışmada, mevcut web kazıma yöntemleri ile çekilmiş limitli Instagram profil bilgilerini kullanarak, fenomen-şirket ayrımını yapabilen bir model geliştirdik. Buradan elde edilen veri üzerinde keşifsel analiz yapıldıktan sonra, öznitelik çıkarılması üzerine çalışıldı. Verisetindeki kategori bazlı öznitelikleri efektif kullanmak için CatBoost modeli kullanıldı ve %98.2'lık eğri altı alan skoruna ulaşan bir model elde edildi. Sınıflandırma problemiindeki başarımın elde edilmesine ek olarak modelin ürettiği sonuçların açıklanabilirliği için SHAP değerleri analiz edilmiş ve markaları fenomenlerden ayıran özellikler detaylı incelenmiştir.

Anahtar Kelimeler—Instagram, makina öğrenmesi, fenomen tahmini.

Abstract—As the use of social media platforms has increased over time, there has been a corresponding rise in the number of accounts influencing masses, such as influencers and companies, active on these platforms. Particularly in marketing efforts, the use of influencers and the need for brands to identify these influencers through automated systems have increased. In this study, we developed a model using existing web scraping methods and a limited data source to distinguish between influencers and companies based on Instagram profile data. After conducting exploratory analysis on the obtained data, the focus shifted to feature extraction. To effectively utilize category-based features in the dataset, we employed the CatBoost model, achieving a %98.2 auc rate with this model. In addition to achieving success in the classification problem, SHAP values were analyzed for interpretability of the model's results, exploring the features that distinguish brands from influencers in detail.

Keywords—Instagram, machine learning, influencer detection.

I. GİRİŞ

Dünya genelinde ucuz cihaz ve servislerin yaygınlaşmasıyla, internet trafiğindeki artış, her geçen yıl daha da belirgin hale gelmektedir [1]. İnternet kullanıcılarının artmasından ötürü, web geliştiricileri, farklı araçlar geliştirerek internetin

kullanımını da çeşitlendirmiştir. Haberleri takip etmekten, çevrimiçi alışverişe kadar pek çok faaliyet için bu sosyal medya platformlarına başvurulmaktadır. Bunun sonucu olarak da 2023 Mayıs ayında en çok ziyaret edilen ilk beş web sitesinden dördü sosyal medya platformlarından oluşmaktadır [2].

İnternet ve sosyal medya platformlarındaki bu hareketlilik, reklam ve pazarlama çalışmaları için önemli bir alan haline gelmiştir. Fenomen pazarı 2016 yılında toplamda 1.7 milyar dolarlık bir hacime sahipken, 2023 yılında bu hacim 21.1 milyar dolara yükselmiş ve şirketlerin %59'u, önümüzdeki yıl için ayrılan bütçeyi arttırmaya niyetli olduklarını belirtmişlerdir [3]. Bu artış, pazarın genişlemesiyle birlikte yeni fenomenlerin de platforma kaydolmasını sağlamıştır. Benzer bir artış şirket hesapları için de gözlemlenmiştir ve her iki türdeki artış, hesaplar arasındaki farkları tespit etmeyi zorlaştırmaya başlamıştır. Buna ek olarak sosyal medyada mevcut hesaplar arasında bot veya sahte takipçilerin tespit edilmesi [4], [5], yazılı metinden çıkarımlar yapacak dil modellerinin geliştirilmesi de gerekmektedir [6], [7].

Bu çalışmada, Instagram platformu üzerinde bulunan profil verilerini kullanarak bir fenomen-şirket sınıflandırma modeli oluşturulmuştur. Denetimli öğrenme (supervised learning) için, “Boomsocial” sitesinde bulunan ve manuel olarak oluşturulmuş etiketlerden faydalanılmıştır. Uygulanacak makine öğrenmesi modellerinin seçiminde, kategori bazlı özniteliklerin fazla olması nedeniyle CatBoost [8] modeli uygun görünmüştür. Seçtiğimiz modelin en iyi performansı verip vermediğini anlayabilmek için, farklı yaklaşımları kullanan modellerin F1 ve eğri altında kalan alan (AUC) skorları karşılaştırılmıştır. Eğittiğimiz makine öğrenmesi modellerinin doğruluk analizleri sonucunda, ortalama %98.2 eğri altı alan skoruna (AUC) sahip bir CatBoost modeli elde edilmiş olup, modelin açıklanabilirliği (model interpretation) için SHAP [9] değerleri incelenmiş ve modelin tahminlerinde en etkili olan özelliklerin “kategori ismi” ve “takip edilen hesap sayısı” olduğunu gözlemlenmiştir.

II. LİTERATÜR TARAMASI

Sosyal medya platformlarında otorite tespiti, pazarın büyümesi ile birlikte daha da önem kazanmıştır. Kullanıcılar arasında bir sıralama (ranking list) oluşturup, bu listenin üstünde yer alan N kişiyi otorite sahibi olarak etiketlemek mümkündür

[10]–[12] ancak sıralı kullanıcı listesinde kaç hesabın değerlendirilmeye alınacağı ek bir parametre olmaktadır. Öte yandan, özellik tabanlı yaklaşımlar (attribute based approach) [13]–[15], otoriter kullanıcıları tespit etmek için potansiyel olarak kullanılabilir bir dizi özelliği belirlemeyi amaçlar. Bu özellikler kullanıcıları yetkili veya yetkisiz olarak sınıflandırmak için denetimli öğrenme sürecinin nitelikleri olarak kullanılır. Bu yaklaşımın problemlerinden bir tanesi, kullanıldıkları etiketlenmiş veriye bağımlı kalmalarıdır. Buradaki sorun, etiketli örnek toplamanın etiketlenmemiş olanlardan daha zor, pahalı ve zaman alıcı olmasıdır.

Alternatif olarak yüksek kaliteli içerik üreten kullanıcılara otoriter kullanıcılar olarak tanımlanabilir [16]. Şirketlerin de kendi sosyal medya hesaplarını açması ile birlikte, otoriter hesaplar ayrımını tek başına yapmak yetersiz kalmaya başladı. Bu soruna bir çözüm olarak otoriter hesap tespit yaklaşımlarının yanı sıra, fenomen-şirket ayrımını yapabilen bir model geliştirmek üzerine bu çalışmadaki yaklaşımı geliştirdik.

III. VERİSETİ

Modelin denetimli öğrenme metodu kullanılarak tahminleme yapabilmesi için öncelikle etiketlenmiş bir verisetinin bulunması lazım. Bunun için, sosyal medya analizi ve ölçümleme araçları sağlayan *Boomsocial* adlı web sitesinden yararlandık. Bu site içerisinde 22,415 adet Türkiye ile bağlantılı olan Instagram hesabı bulunmakta. Bu hesaplar, Boomsocial sitesinin karar vermiş olduğu 24 kategori içerisinde ayrıştırılmış. Bu kategorilerden bir tanesi de “Fenomenler” kategorisi ve bu kategori içerisinde yer alan 1,858 hesabı kendi çalışmamız için kullandık. Şirket hesaplarının tespiti için de, şirketler için uygun olan ve genellikle hizmet sektöründeki firmalara yönelik etiketleri değerlendirdik. Bu kategoriler: Eğlence ve yaşam, alışveriş, enerji, finans, giyim, hızlı tüketim, internet, kurumsal, otomotiv, reklam ve pazarlama, sağlık, seyahat, spor, teknoloji, ürün hizmet ve yapı sektörü. Bu kategorileri ve alt sektörlerini kullanarak 5,447 hesabı şirket olarak etiketledik.

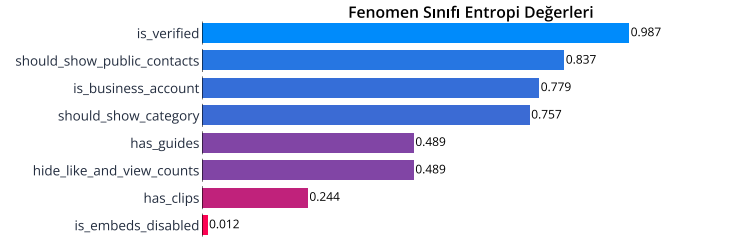
Eğitim aşamasında kullanılabilir olacak potansiyel hesapları belirledikten sonra, bu hesapların Instagram verilerini kazıma (scrape) sürecine geçildi. Bir hesabın şirket hesabı veya fenomen hesabı olup olmadığını anlayabilmek için, yalnızca profil bilgilerine yönelik veriyi topladık. Instagram’ın, her bir hesap için kullanmış olduğu `/api/v1/users/web_profile_info/?username=username` uzantısında dönen JSON dosyası sistematik olarak toplandı. Bu uzantının sağlamış olduğu dosya içerisinde, bir hesabın fenomen mi yoksa şirket mi olduğunu anlamaya yardımcı olabilecek 63 farklı anahtar-değer çifti bulunmakta ve öznelitliklerin bir kısmı bu veri kullanarak üretilmiştir.

IV. METODOLOJİ

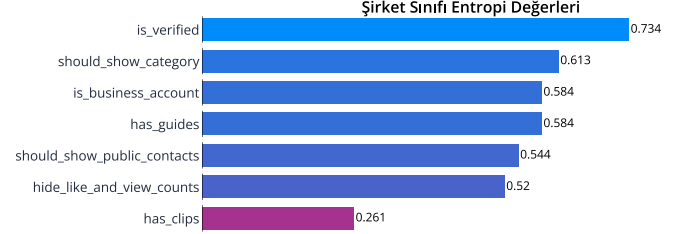
A. Keşif Amaçlı Veri Analizi ve Özellik Mühendisliği

Makine öğrenmesi modelleri üzerinde deney yapmaya başlamadan önce, keşifsel veri analizi ile model sonuçlarına olumlu katkı sağlayabilecek öznelitlikleri incelemek istedik ve bu doğrultuda analizler gerçekleştirdik.

Kullandığımız veriseti daha çok ikili (binary) tipte veriler barındırdığı için analizlerimize bilgi teorisi (information theory) metodlarını kullanarak başladık. İlk olarak, veri setini

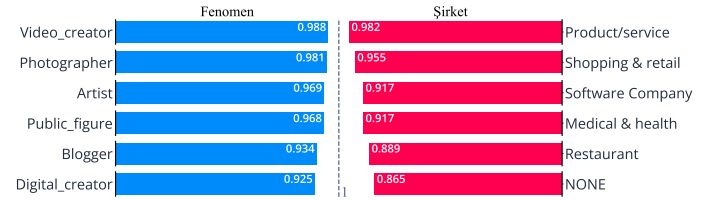


((a)) Mantıksal tipli özelliklerin fenomen sınıfı üstündeki kararsızlık durumları.



((b)) Mantıksal tipli özelliklerin şirket sınıfı üstündeki kararsızlık durumları.

Şekil 1: Öznelitliklerin hesaplar üzerinde dağılım entropileri. Tekil öznelitliklerin sınıflandırma için kullanılabilirliğini incelemek üzere entropi değerleri karşılaştırılmıştır.



Şekil 2: Fenomen ve Şirket sınıfları en iyi temsil eden 6 kategori. Hesaplara Instagram tarafından tanımlanan kategorilerin gözlemlenme istatistikleri.

etiketlerine ayırarak her bir etiket üstünden entropi analizi oluşturduk. Şekil 1, öznelitliklerin her bir sınıf içerisinde ne kadar kararsız durumda olduklarının ölçütlerini gösteriyor. Buradan yola çıkarak, “is_verified” özelliğinin her iki sınıfta da diğer özelliklere göre daha kararlı bir konumda olduğunu söyleyebiliriz. Bu bilgi her ne kadar sınıflar tek başlarına düşünüldüğünde önemli bir ölçütmüş gibi gözükse de, her iki sınıf için de “is_verified” özelliğinin kararlı konumda olması, o özelliği ayırt edici bir pozisyonda bulundurmaktan çıkartıyor. Bunun yanı sıra, hali hazırda her bir sınıf için kararsız bir konumda bulunan “has_clips”, “is_embeds_disabled” özellikler, bu sınıflar için çok ayırt edici bir bilgi taşımamakta ve tahminleme modeli içerisinde kullanılmaya aday gösterilmeyebilirler.

Bir özelliğin, iki sınıfı ayırt etmedeki kullanılabilirliğini anlamak için bilgi kazanımı (information gain) gibi daha güçlü bir metrik kullanılabilir. Bu verisetindeki özelliklerin sınıfları ayırt etmekteki bilgi kazanımlarına baktığımızda, “is_business_account” özelliğinin 0.312 değeri ile diğer özelliklerden daha iyi bir konumda olduğunu gördük. Entropi değerlerine baktığımızda her bir sınıf için en kararlı olarak gözükten “is_verified” özelliği ise, her iki sınıf için de yüksek oranda kararlı olduğu için, bilgi kazanımı “is_business_account” özelliğinden daha düşük çıktı ve değeri 0.101 olarak hesaplandı.

TABLO I: Verisetinden çıkartılan öznitelikler

Öznitelik ismi	Açıklama
bio boş mu	"bio_links" alanında text var mı bilgisi
İş adresi yok mu	"business_address_json" özelliği boş mu bilgisi
takipçi takip eden oranı	"edge_followed_by_count"/"edge_follow_count"
isim sayısı	"full_name" içerisindeki Türkçe isim sayısı
isimi oluşturan eleman sayısı	"full_name" içindeki toplam kelime sayısı
isimdeki büyük harf oranı	"full_name" içerisindeki büyük harf sayısının ismin uzunluğuna oranı
isimdeki noktalama işareti sayısı	"full_name" içerisindeki toplam noktalama sayısının uzunluğuna oranı
isimdeki rakam sayısı	"full_name" içindeki toplam rakam sayısının uzunluğuna oranı
isim uzunluğu	"full_name" karakter sayısı

TABLO II: Model Performans Karşılaştırmaları

Modeller	Metrikler	
	AUC(%)	F1(%)
CatBoost	98.2 ± 0.415	94.5 ± 0.77
XGBoost	92.8 ± 0.933	85.2 ± 1.465
Logistic Regression	81.7 ± 1.563	72.4 ± 1.890
SVM	79.3 ± 1.651	70.4 ± 1.892

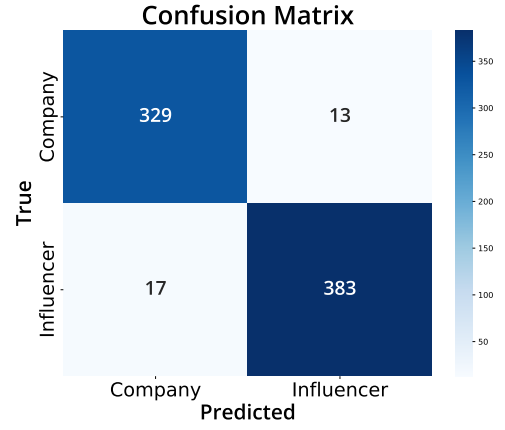
Her ne kadar verisetini oluşturan özelliklerin çoğu ikili (binary) tipli olsa da, hesabın kategorisi ile alakalı özellikler de mevcut ve bu özellikler tahminlemede etkin bir rol oynayabilirler. Şekil 2 kategorilerin tahminlemede etkin olarak kullanılabileceğine dair kanıtlar vermektedir. Verisetinde kategorisi "Video_creator" olarak gözüken 171 hesaptan 169 tanesi Fenomen olarak etiketlenmiş. Bunun yanı sıra, kategorisi "Product/service" olarak gözüken 172 hesaptan 168 tanesi Şirket olarak etiketlenmiş. Bu kesin ayırım, kategorilerin etiketlemede etkin olarak kullanılabileceğini destekliyor.

Var olan özellikleri kullanmanın yanı sıra, elde bulunan veriyi kullanarak yeni öznitelikleri çıkartmaya çalışmak da tahminlemede etkin bir rol oynayabilir. Örneğin verisetinde "full_name" bilgisi bulunuyor. Her ne kadar "full_name" tahminlemede direkt olarak kullanılamasada, bu veri üstünden tahminlemeye etki edebilecek yeni öznitelikler çıkartılabilir. Tablo I eldeki verileri kullanarak çıkartılmış öznitelikleri gösteriyor. Bu öznitelikleri ve keşif amaçlı veri analizi çıktılarını kullanarak toplam 16 öznitelikten oluşan bir set hazırladık.

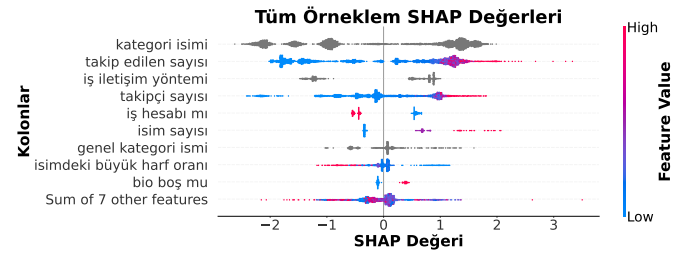
B. Model Eğitim ve Değerlendirme

Öznitelik seti belirlendikten sonra, model seçimini yaparken, göz önünde bulundurduğumuz bir konu da, özniteliklerin 8 tanesinin ikili (binary) tipte olmasıydı. Metin tipli de toplamda 6 tane özellik bulunmaktadır ve bunlardan "kategori ismi" özelliği, içerisinde 307 farklı kategori barındırıyor. Eğitim setindeki kategori miktarının, örneklem büyüklüğüne kıyasla çok büyük olması, bu özelliğin one-hot encoding yöntemiyle temsil edilmesini zorlaştırıyor. Fazla miktarda kategori barındıran verisetlerinde başarılı olan CatBoost yaklaşımı, metin tipli özniteliklerle çalışabiliyor ve eğitim sürecinde bunları Ordered Boosting metoduyla numerik olarak ele alabiliyor.

Modelin iyi bir performans gösterebilmesi için her iki sınıf için de dengeli bir örneklem topluluğu oluşturmak önemlidir. Elimizde fenomenler için 1,858 ve şirketler için ise 5,447 adet hesap bulunmakta. Her iki etiket arasında denge sağlamak amacıyla 5,447 adet hesap arasından rastgele 1,858 tanesi eğitilmek ve valide etmek amacıyla kullandık. Verisetinin %80'i eğitim, geri kalanı da validasyon için kullanılmıştır.



Şekil 3: Karışıklık Matrisi. Sınıflandırma sonucunda diagonal ekseninde doğru tahminlerin fazlalığı model kalitesi için olumlu bir işaret. Her iki kategori için de dengeli sayıda hatalı tahminleme yapıldığını görmek de mümkün.

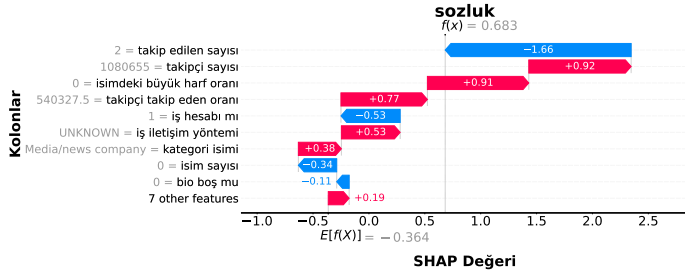


Şekil 4: Beeswarm Plot. Özniteliklerin aldığı değerlere göre SHAP dağılımları.

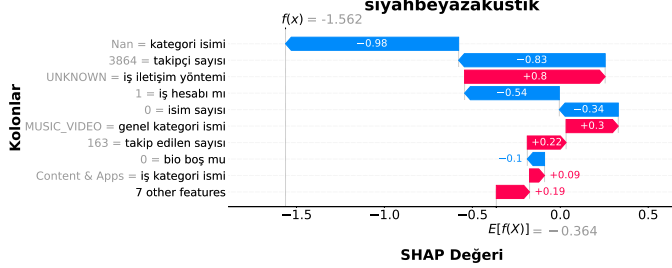
Modelin eğitilmesinde 1,000 iterasyon ve 0.1 öğrenme oranı kullanılmaktadır ve derinlik hiper parametresi 2 seçilirken kayıp fonksiyonu için logloss tercih edilmiştir. Modeli bu ayarlarda eğittiğimizde, validasyon setindeki eğri altında kalan alanı (AUC) %98.2 olarak buluyoruz. Seçilen modelin en iyi model olduğunu anlayabilmek adına, aynı eğitim seti kullanılarak XGBoost, Support Vector Machines (SVM) ve Logistic Regression modelleri de karşılaştırıldı. Kategorik verileri bu modeller özel olarak analiz edemediğinden one-hot encoding ile öznitelikler oluşturuldu. Tablo II farklı modellerin validasyon seti üstündeki F1 ve AUC skorlarının karşılaştırılmasını gösteriyor. Eldeki veriler kendi aralarında 100 defa karıştırılarak ortalama AUC ve F1 skorları çıkartıldı. Çalışma prensibi olarak Catboost'a en yakın model olan XGBoost, %92.8 AUC skoru ile Catboost'tan sonra en iyi performans gösteren model oldu. Bunun yanı sıra, Lojistik Regresyon ve SVM modellerinin performanslarına bakılarak da, lineer modellerin bu iş için ne kadar uygun olduğu anlaşılmaya çalışıldı. Validasyon seti üstünden Catboost modelinin tahminlerinin karışıklık matrisini (confusion matrix) oluşturduğumuzda, Şekil 3'de iki sınıf için yapılan yanlış tahmin sayısı birbirlerine yakın çıkıyor.

C. Model Açıklaması

Modelin, tahminleme performansının yüksek olmasına ek olarak modelin tahminlemeyi nasıl yaptığını anlayabilmek de özellikle uygulamalar için büyük önem taşır. CatBoost modeli, lineer karar sınırları üreten bir model olmadığı için model tahmini ile öznitelikler arasında birebir bir bağlantı yakalamak zor



((a)) "sozluk" hesabının özniteliklerinin SHAP değerleri



((b)) "siyahbeyazakustik" hesabının özniteliklerinin SHAP değerleri

Şekil 5: Özniteliklerin sınıflandırmaya etkisi. Her iki etiket için yanlış tahminlenmiş hesapların özelliklerinin SHAP değerleri temsil edilmiştir. Sınıflandırma kararını destekleyen (mavi) ve bununla çelişen (kırmızı) öznitelikler temsil edilmiştir.

olabilir. Bu bağlamda, model tahminlerinin, veri seti üzerindeki SHAP değerlerine bakacağız.

Şekil 4. verisetindeki özniteliklerin tahminlemelere ne kadar etki ettiğine dair bilgi veriyor. Burada dikkat çeken durumlardan biranesi, "iş hesabı mı," "bio boş mu," "isim sayısı," "isimdeki büyük harf oranı" özniteliklerinde kesin ayrımlar görebiliyor olmamız. "iş hesabı mı" özneliliğinin model üstünde etkili olabileceğini bilgi kazanımının yüksek olmasından ötürü bekliyorduk. SHAP değerlerinde de modelin, bu özelliğe önem verdiğini görebiliyoruz. "isim sayısı" özelliğinin arttığı durumda SHAP değerinin de artıyor olması modelin bu özelliğin yüksek olduğu durumları, fenomen özelliği olarak kabul ettiğini gösteriyor. Bunun yanı sıra, "isimdeki büyük harf oranı" özelliğinin yüksek olması da, o hesabın daha çok şirket olabileceği yönünde bir fikir veriyor. Model üstünde en çok etkisi olduan iki özellik "kategori ismi" ve "takip edilen sayısı" olmuştur. Takip edilen sayısının çok yüksek olduğu durumları model, fenomen özelliği olarak kabul etmiştir.

Model tahminlerinin neden yanlış olduğunu anlamak, sonraki süreçlerde geliştirmeler yapabilmek için önemlidir. Şekil 5. Şirket ve Fenomen etiketleri için yanlış tahminlenen örneklerin SHAP değerlerini gösterir. Şekil 5(a). Şirket olmasına rağmen model tarafından Fenomen olarak tahminlenmiş. SHAP değerlerine baktığımızda, "takipçi sayısı," "isimdeki büyük harf oranı," "takipçi takip eden oranı" gibi öznitelikler, aldığı değerlerden ötürü modelin, Şirket olarak tahminlemesini zorlaştırmışlar. Şekil 5(b). incelendiğinde ise, tahmine etki eden özelliklerin çoğunun, Şirket yanlısı olduğunu görebiliyoruz.

V. SONUÇ

Sosyal medya platformlarının kullanıcı sayısının artmasıyla doğan ihtiyaçlara sunulacak çözümlerde hem akademiye hem endüstride artmaktadır. Ticari amaçla kullanımın artmasıyla

birlikte, sosyal ağlarda marka işbirlikleri yapan fenomenlere ek olarak ürün ve hizmetlerini daha yaygın kitlelere ulaştırmaya çalışan markalar için bu platformlar önemli hale gelmiştir.

Bu çalışmada, Instagram profillerini analiz etmek amacıyla influencer ve markalar gibi fazla sayıda takipçili ve aktif hesapları ayıracak bir model geliştirdik. Veri çeşitliliği ve az sayıda veriye sahip olma sorunlarını özellikle uygulama alanlarını arttırmak amacıyla ele aldık ve sınırlı veri ile çalışacak bir model geliştirmeyi hedefledik. Modelimizin yaklaşık %98.2 doğruluk oranında başarıyla influencer ve markaları ayırdığını gözlemledik. Bu kararda etkili olan faktörleri ve yanlış sınıflandırmaları ise SHAP değerlerine bakarak inceledik.

KAYNAKLAR

- [1] J. Degenhard, "Number of internet users worldwide from 2014 to 2029," <https://www.statista.com/forecasts/1146844/internet-users-in-the-world>, January 2024, accessed: February 23, 2024.
- [2] D. H. Anwar, "July 2023 rankings: The world's most visited websites around the globe," <https://www.digitalinformationworld.com/2023/07/july-2023-rankings-worlds-most-visited.html>, 2023, accessed: February 22, 2024.
- [3] W. Geyser, "The state of influencer marketing 2024: Benchmark report," <https://influencermarketinghub.com/influencer-marketing-benchmark-report/>, Februray 2024, accessed: February 22, 2024.
- [4] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, 2016.
- [5] Y. Zouzou and O. Varol, "Unsupervised detection of coordinated fake-follower campaigns on social media," *arXiv preprint arXiv:2310.20407*, 2023.
- [6] A. Najafi and O. Varol, "Turkishbertweet: Fast and reliable large language model for social media analysis," *arXiv preprint arXiv:2311.18063*, 2023.
- [7] —, "VRLlab at HSD-2Lang 2024: Turkish Hate Speech Detection Online with TurkishBERTweet," in *Proc. of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, 2024, pp. 185–189.
- [8] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.
- [9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
- [10] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proc. of the 19th Intl. Conf. on World wide web*, 2010, pp. 591–600.
- [11] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proc. of ACM Intl. Conf. on Web search and data mining*, 2010, pp. 261–270.
- [12] H. Zhu, H. Cao, H. Xiong, E. Chen, and J. Tian, "Towards expert finding by leveraging relevant categories in authority ranking," in *Proc. of ACM Intl. Conf. on Information and knowledge management*, 2011, pp. 2221–2224.
- [13] A. Pal, R. Farzan, J. A. Konstan, and R. E. Kraut, "Early detection of potential experts in question answering communities," in *Intl. Conf. on User Modeling, Adaption and Personalization*. Springer, 2011, pp. 231–242.
- [14] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proc. of the ACM Intl. Conf. on Web search and data mining*, 2011, pp. 45–54.
- [15] Y. M. Cetinkaya, M. Gürlek, I. H. Toroslu, and P. Karagöz, "Twitter account classification using account metadata: organizationvs. individual," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 4, pp. 1404–1418, 2022.
- [16] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, "Social network analysis and mining for business applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–37, 2011.