Sosial şəbəkələrdə saxta profillərin aşkarlanması metodları haqqında

Yadigar İmamverdiyev[1], Xəyalə Əhmədova[2]
[1,2]AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
[1]*yadigar@iit.science.az*, [2]*ehmedovaxeyale.97@mail.ru*

***Xülasə*— Sosial şəbəkələr genişləndikcə onların insanlar üzərində təsiri də artmaqdadır. Sosial şəbəkələrdə fərdi məlumatlar da paylaşıldığından istifadəçilərə yönəlmiş təhdidlərdən olan saxta profillər bu məlumatlara təhlükə yaradır. Real istifadəçidən birbaşa və ya dolayı yolla alınan fərdi məlumatlar daha sonra müxtəlif məqsədlər üçün (məsələn, şantaj yolu ilə pul tələb etmək, həmin şəxs adından başqa bir profil yaradıb onun adından yalan məlumatlar yaymaq və s.) istifadə edilir. Saxta profillərin məqsədləri gün keçdikcə daha təhlükəli hal aldığından onların təyin edilməsi günümüzün aktual məsələlərindən biridir. Bu məqalədə saxta profillərin növləri, saxta profillərin təyinində istifadə edilə biləcək əlamətlər, saxta profillərin təyinində maşın təlimi yanaşmaları və avtomatik aşkarlanma üçün təklif edilən yanaşmaya baxılmışdır.**

Açar sözlər— sosial şəbəkə; saxta profil; trol profil; aşkar edilmə; maşın təlimi

I. GİRİŞ

Sosial şəbəkələr həyatımızın ayrılmaz parçasına çevrilmişdir. Belə ki dünya əhalisinin 7.7 milyard olduğunu nəzərə alsaq, onlardan 2 milyarda yaxını sosial media istifadəçisidir. Təbiidir ki, bunun səbəblərindən biri saxta hesab istifadəsidir. 2018-ci ildə Facebook sosial şəbəkəsində edilmiş araşdırmaya əsasən 500 milyon qadın adı altında fəaliyyət göstərən saxta profil vardı. Bu tip saxta profil açmaqda bədniyyətlininin məqsədi çirkli pulların yuyulması, qaçaqmalçılıq və bəzi digər anti-sosial davranışlar ola bilir [1]. Bu zaman baxılacaq əsas məsələlərdən biri ilk növbədə paylaşım edənin cinsini müəyyən etməkdir. Bu halda, adətən, Təbii Dilin Emalı (ing. Natural Language Processing, NLP) texnologiyaları istifadə edilir.

Sosial şəbəkələrdə əks cinsin adı altında fəaliyyət göstərən saxta profillərlə yanaşı daha mürəkkəb işlər icra edən saxta profil növləri də vardır. Saxta profillərdən insanları manipulyasiya etmək üçün də istifadə edilir. İnsanları idarə etməkdə, yönləndirməkdə birinci adddım onların fikirlərinə təsir etməkdir. Bu məqsədlə sosial şəbəkələrdən istifadə edilir. Sosial şəbəkələrdən zərərli məqsədlərçün istifadə edilməsinin başlıca səbəbi bu platformaların geniş izləyici kütləsinə malik olmasıdır.

II. SOSİAL ŞƏBƏKƏLƏRDƏ BƏZİ TƏHDİDLƏR

Sosial şəbəkələrin funksionallığı artdıqca ondan istifadə edən insanların sayı da artır, bu isə sosial şəbəkələrin daha da böyüməsinə təkan verir. 2019-cu il üçün ən böyük 7 sosial şəbəkə arasında aktiv istifadəçi sayı Facebook üçün 2.38 milyard, Twitter üçün 321 milyon, LinkedIn üçün 303 milyon, Instagram üçün 1 milyard, Snapchat üçün 330 milyondan artıq, Pinterest üçün 291 milyon, Reddit üçün isə 330 milyon olmuşdur. Bu saylar sosial medianın yetərincə böyük izləyici kütləsinə sahib olduğunu göstərir [2].

Sosial şəbəkələrdən istifadə zamanı üzləşilən təhdidlər aşağıdakılardır:

- *Sosial mühəndislik* − kiber cinayətkarların zamanla hədəf təşkilatın işçilərilə qurduğu əlaqələr nəticəsində hədəf təşkilat haqqında konfidensial məlumatların ələ keçirilməsi məqsədilə ən geniş istifadə edilən sosial media təhdididir.
- *Fişinq hücumları* − pul və ya şəxsi məlumatların oğurluğu üçün istifadə edilən, əsasən, tanışlıq yönümlü soial şəbəkələrdə geniş yayılmış təhdid növüdür.
- *Saxta hesablar* − kimi və ya nəyisə gözdən salmaq, nüfuz qazandırmaq, insanları manipulyasiya etmək, saxta xəbər yaymaq kimi məqsədlərin icrası üçün istifadə edilən təhdid növüdür.
- *Məşhurların adından sui-istifadə* − Məşhur şəxs adından saxta profil yaradılaraq onun adından saxta paylaşımlar edilməsi və ya izləyici sayının artırılması üçün istifadə edilən təhdid növüdür.
- *Saytın ələ keçirilməsi* − bədniyyətlinin zərərli kodlar yerləşdirdiyi reklamları sosial şəbəkədə yayması, saxta məlumat paylaşması, öz gücünü sübut etməsi kimi bir sıra zərərli məqsədləri icra etmək üçün yerinə yetirdiyi təhdid növüdür.
- *Spam və zərərli proqram təminatının yayılması*− Twitter, Facebook və Instagram kimi sosial şəbəkələr spam və zərərli proqram təminatlarının yayılması üçün ideal platformalardır. Kiber cinayətkarlar, əsasən, qısa uzunluqlu URL istifadə etməklə öz zərərli linklərini gizlədirlər ki bu da saytın qanuni və ya zərərli olmasının ayırd edilməsini

DOI: 10.25045/NCInfoSec.2019.10

çətinləşdirir [3].

III. SAXTA PROFİLLƏRİN NÖVLƏRİ

Sosial şəbəkələrdə fəaliyyət göstərən saxta profilləri ümumi şəkildə üç böyük qrupa bölmək olar:

- insan tərəfindən yaradılan saxta profillər;
- proqram tərəfindən yaradılan saxta profillər;
- yarı insan, yarı proqram tərəfindən yaradılan saxta profillər (Cyborgs).

İnsan tərəfindən yaradılan saxta hesablara trol profillər, proqram tərəfindən yaradılan saxta profillərə botlar misal olaraq göstərilir. Cyborg-lar isə insan tərəfindən yaradılıb, sonrakı fəaliyyəti avtomatlaşdırılmış saxta profillərdir [4].

Sosial şəbəkələrdə olan saxta profillər yaradılma və istifadə tərzinə görə aşağıdakı kimi təsnifləndirilir:

- *Ələ keçirilmiş profillər* − real istifadəçinin profilinin oğurlanması nəticəsində fəaliyyət göstərən saxta profil növüdür.
- *Klonlanmış profillər* − real profilin nüsxəsinin yaradılması nəticəsində fəaliyyət göstərən saxta profil növüdür. Real profilin yaradıldığı sosial şəbəkədə onun verilənlərindən istifadə edilərək yaradılan profil saytdaxili, fərqli sosial şəbəkədə yaradıldıqda isə saytlararası klon profil adlanır.
- *Kuklalar* − saxta ad altında özünü əslində mövcud olmayan bir şəxs kimi göstərərək kimi və ya nəyisə gözdən salmaq, nüfuzunu artırmaq üçün istifadə edilən saxta profil növüdür.
- *Sibil hesablar (ing. Sybil Accounts)*− bir şəxs tərəfindən yaradılan birdən çox profilin əl ilə idarə olunması nəticəsində fəaliyyət göstərən saxta profil növüdür.
- *Botlar saxta profil kimi* − fəaliyyəti Sibil hesablara oxşar, spam, bəyənmə, təsir, botnet kimi növləri olan saxta profil növüdür. Sibil hesablardan fərqli olaraq insan tərəfindən deyil, avtomatik idarə olunur [5].
- *Trol profillər*− qarşıya qoyulmuş hədəfə nail olmaq üçün sistemli şəkildə “troll fabriki” şəklində fəaliyyət göstərən saxta profil növüdür. Trol fermalara misal olaraq 2016-cı ildə ABŞ-da keçirilən prezident seçkilərinə təsiri olan İnternet Araşdırma Agentliyini (ing. Internet Research Agency, IRA) göstərmək olar [6].

IV. SOSİAL ŞƏBƏKƏLƏRDƏ SAXTA PROFİLLƏRİN AŞKARLANMASINDA MAŞIN TƏLİMİ METODLARI

Sosial şəbəkələrdə saxta profillərin aşkarlanması problemlərindən biri tədqiqatçıların xüsusi tip saxta profilin aşkarlanması üçün metod işləməsidir.

Twitter sosial şəbəkəsində Spam botlarının aşkarlanması üçün Naïve Bayes (NB) və k ortalar (ing. K-means) klassifikasiya alqoritmləri tətbiq edilmişdir [7]. Facebook sosial şəbəkəsində ələ keçirilmiş və sibil hesabların aşkarlanması üçün k ən yaxın qonşu (ing. K Nearest Neighbor, k-NN) və Əsas Komponent Analizi (ing. Principal Component Analysis, PCA) metodları istifadə edilmişdir [8]. Vikipediya sosial şəbəkəsində kukla hesabların aşkarlanması üçün Support Vector Machine (SVM) metodu tətbiq edilmişdir [9]. Facebook, Twitter, MySpace sosial platformalarında spam botlarının aşkarlanması üçün Random Forest algorithm tətbiq edilmişdir [10].

[11]-də tədqiqatçılar Facebook sosial şəbəkəsində saxta profillərin aşkarlanması üçün “FBCecker” ağıllı sistemini təklif edilmişdir. Təklif edilən yanaşma 3 pillədən ibarətdir:

1. Facebook istifadəçi profillərindən “CRAWLER” adlı moduldan istifadə edilərək seçilmiş əlamətlərə uyğun məlumatların toplanması;
2. k-NN sxemi və filtrləmə operatorunun köməyilə ötürülmüş verilənlər probleminin həll edilməsi;
3. 4 supervizorlu öyrənmə alqoritminin: Qərar ağacı (ing. Decision Tree, DT), k-NN, SVM və NB təlim və tətbiqi.

Alqoritmin tətbiqi zamanı profil şəkli, iş yeri, təhsil, yaşayış yeri, münasibət statusu və s. kimi əlamətlərdən istifadə edilmişdir.

K-NN sxeminin istifadəsilə ötürülmüş verilənlərin qiymətləndirilməsinin tətbiqi nəticəsində Accuracy Qərar ağacı DT üçün 0.965, k-NN üçün 0.84, SVM üçün 0.985, NB üçün 0.975 olmuşdur.

Filtrləmə operatorlarının köməyilə itkin verilənlərin istisna edilməsinin tətbiqi nəticəsində Accuracy DT üçün 0.946, k-NN üçün 0.844, SVM üçün 0.988, NB üçün 0.964 olmuşdur.

[12]-də müəlliflər sosial mediada kuklaların avtomatik aşkarlanması və qruplaşdırılması üçün SocksCatch adlanan model təklif edirlər. Model 3 mərhələdən ibarətdir:

1. Mərhələ verilənlərin toplanması və seçilməsi addımlarından ibarətdir.
2. Mərhələ əlamətlərin seçilməsi və maşın təlimi metodlarının tətbiqi addımlarından ibarətdir.
3. Mərhələ qraflar nəzəriyyəsindən istifadə etməklə eyni istifadəçi tərəfindən yaradılan kukla hesabların fəaliyyət və hesab əlamətləri əsasında qruplaşdırılması addımlarından ibarətdir.

Bu məqalədə müəlliflər Ingilis vikipediyasında kukla və aktiv hesabların arasındakı fərqləri müəyyənləşdirmək üçün əlamətləri 3 qrupa: fəaliyyət davranışı, hesab fəaliyyətinə nəzərən digər hesabların davranışı, hesab xüsusiyyətləri bölürlər.

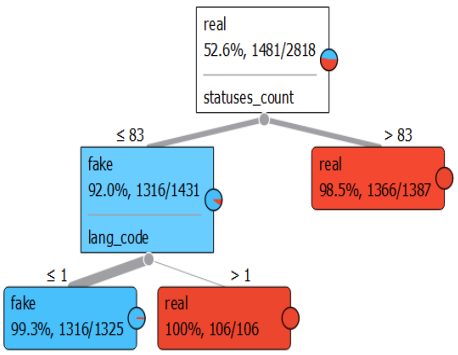
Seçilmiş əlamətlər əsasında təklif edilən proses üçün saxta profillərin aşkarlanmasında 6 maşın təlimi yanaşmasından SVM 92.6%, Random Forest 94.8%, NB 89.1%, k-NN 90.9%,

Bayesian network 89.1% və Adaptive boosting 92.6% accuracy göstərmişdir.

V. APARILMIŞ EKSPERİMENTLƏR

Aparılmış tədqiqat nümunələrindən göründüyü kimi saxta profillərin aşkarlanmasında maşın təlimi yanaşmaları, xüsusilə, Supervizorlu maşın təlimi yanaşmaları geniş istifadə edilir. Buna səbəb profillərin saxta və ya real olaraq sinifləndiriləcəyinin məlum olmasıdır.

Təcrübələr “Orange” proqram təminatında aparılmışdır. 2818 sayda profil (1337 saxta, 1481 real) status sayı, izləyici sayı, dostlar sayı, favorit sayı, izlənənlər sayı (ing. listed_count), cins kodu və dil kodu əlamətlərindən istifadə edilərək DT, NB və SVM alqoritmlərinin tətbiqi ilə klassifikasiya edilmişdir. DT tətbiqinin vizual nəticəsi şəkil 1-də göstərilmişdir.



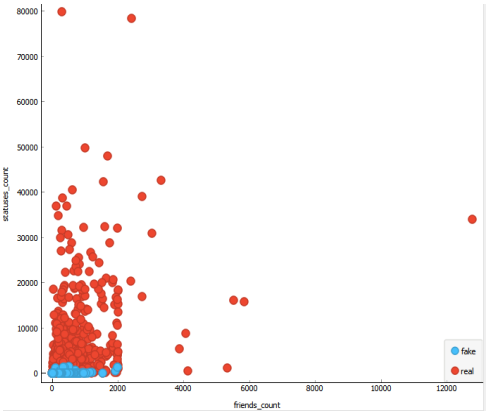
Şəkil1. DT tətbiqinin vizual görüntüsü

DT tətbiqi nəticəsində ‘statusların sayı’ əlamətinin profilin saxta olub olmamasının müəyyən edilməsində ən vacib əlamət olması müşahidə edilmişdir. Belə ki, status sayı ağacın kökü olaraq qəbul edilmişdir. Status sayı 83-dən böyük olan profillər 98,5% göstəricisi ilə real profil olaraq qəbul edilmişdir. Daha sonra “dil kodu” əlamətindən istifadə edilərək müşahidə edilmişdir ki, dil kodu 1-dən böyük olan profillər real profillər, digərləri saxtadır.

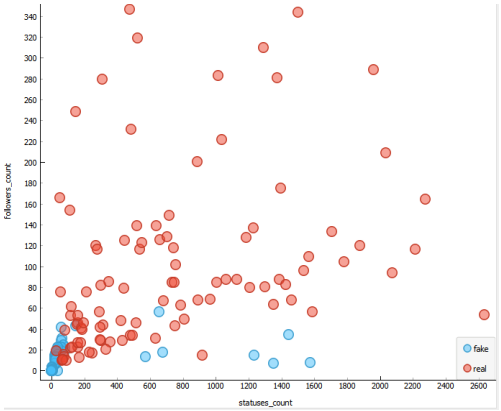
NB tətbiqinin Scatter Plot vasitəsilə vizual nəticəsi şəkil 2-də verilmişdir. Qrafikdə status sayı ilə dostlar sayı arasında asılılıq təsvir edilmişdir. Qırmızı nöqtələr real, mavi nöqtələr saxta profilləri əks etdirir. Qrafikə əsasən qeyd etmək olar ki, real profillər saxta profillərə nisbətən daha çox status paylaşır.

SVM tətbiqinin Scatter Plot vasitəsilə vizual nəticəsi şəkil 3-də təsvir edilmişdir. Qrafikdə izləyici sayının status sayından asılılığı əks etdirilmişdir. Qrafikə əsasən qeyd etmək olar ki, izləyici və status sayı çox olan profillər real profillər olaraq qəbul edilə bilər.

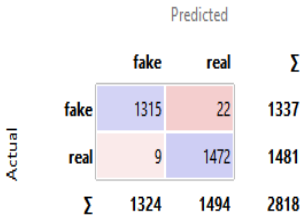
DT, NB və SVM üçün olan xəta matrisləri (ing. confusion matrix) uyğun olaraq şəkil 4-6-da verilmişdir.



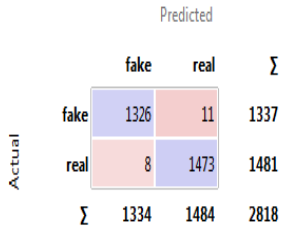
Şəkil2. NB tətbiqinin vizual görüntüsü



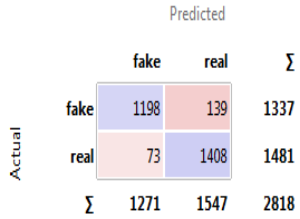
Şəkil 3. SVM tətbiqinin vizual görüntüsü



Şəkil 4. DT tətbiqi nəticəsində əldə edilən xətalar matrisi



Şəkil 5. NB tətbiqi nəticəsində əldə edilən xətalar matrisi



Şəkil 6. SVM tətbiqi nəticəsində əldə edilən xətalar matrisi

Tətbiq edilmiş modellərin müqayisəsi cədvəl 1-də əks etdirilmişdir.

CƏDVƏL I. Tətbiq edilmiş modellərin müqayisəsi

Model	***AUC***	***CA***	***F1***	***Precision***	***Recall***
DT	0.987	0.989	0.989	0.989	0.989
SVM	0.972	0.925	0.925	0.926	0.925
NB	**0.999**	**0.993**	**0.993**	**0.993**	**0.993**

Nəticələrdən də göründüyü kimi saxta profillərin qeyd edilmiş bazada seçilmiş əlamətlər əsasında aşkarlanması üçün tətbiq edilən maşın təlimi yanaşmalarından ən yaxşı nəticə göstərən Naive Bayes, ən pis nəticə göstərən isə SVM olmuşdur.

NƏTİCƏ

Tədqiqatçıların, əsasən, saxta profilin hər hansısa bir növünü aşkarlamaq üzrə tədqiqatlar aparmalarına baxmayaraq, sosial şəbəkələrdə saxta profillərin ümumi şəkildə təyin edilməsi üçün də bir sıra tədqiqatar vardır. Burada əsas məsələ sosial şəbəkədə olan profillərin saxta və ya real profil qrupuna daxil edilməsidir. Bu məqsədlə, əsasən, klassifikasiya metodlarından istifadə edilir. Buna səbəb profilləri daxil edəcəyimiz qrupların əvvəlcədən məlum olmasıdır. Yəni baxılan məsələ profili klassifikasiya edib hansı qrupa daxil olacağını müəyyən etməkdir.

Bununla yanaşı klasterizasiya alqoritmləri də saxta profillərin aşkarlanmasında istifadə olunmaqdadır. Saxta profillərin təyin edilməsində klasterizasiya alqoritmlərinin klassifikasiya alqoritmlərinə nisbətən zəif nəticə göstərməsinə baxmayaraq, onların da üstün cəhətləri vardır. Məsələn, saxta profillərin aşkarlanmasında klassifikasiya alqoitmlərinin tətbiqi zamanı profillər tək-tək analiz edilib qruplara daxil edilirdisə klasterizasiya məsələsində oxşar şəkildə fəaliyyət göstərən və ya oxşar əlamətlərə sahib profillər toplu şəkildə qruplaşdırılır.

İSTİNADLAR

[1] R. Raturi, “Machine Learning Implementation for Identifying Fake Accounts in Social Network”, International Journal of Pure and Applied Mathematics, vol. 118(20), pp. 4785-4797, 2018.

[2] “The 7 Biggest Social Media Sites in 2019” https://www.searchenginejournal.com/biggest-social-media-sites/308897/

[3] “Security threats we face while using social media” https://www.novalisit.com/2019/04/security-threats-we-face-while-using-social-media/

[4] E. Van Der Walt, J. Eloff, “Using machine learning to detect fake identities: bots vs humans”, IEEE Access, vol. 6, pp. 6540-6549, 2018.

[5] M. A. Wani , S. Jabin , G. Yazdani , et al “Sneak into devil’s colony – A study of fake profiles in Online Social Networks and the cyber law”, 2018.

[6] C. Llewellyn, L. Cram, A. Favero, et al “For whom the bell trolls: Troll behaviour in the Twitter Brexit debate”, 2018.

[7] J. S. Alowibdi, U. A. Buy, S. Y. Philip, S. Ghani and M. Mokbel, “Deception detection in Twitter”, Social Network Analysis and Mining, vol. 5(1), pp. 1-16, 2015.

[8] B. Viswanath, M. A. Bashir, M. Crovella, et al, “Towards detecting anomalous user behavior in online social networks”, In 23rd USENIX Security Symposium, pp. 223-238, 2014.

[9] A. H. Wang, “Detecting spam bots in online social networking sites: a machine learning approach”, In Data and Applications Security and Privacy XXIV, pp. 335-342, Heidelberg, 2010.

[10] G. Stringhini, C. Kruegel, G. Vigna, “Detecting spammers on social networks”, Proceedings of the 26th Annual Computer Security Applications Conference, pp. 1-9, December, 2010.

[11] M. B. Albayati, A. M. Altamimi, “An Empirical Study for Detecting Fake Facebook Profiles Using Supervised Mining Techniques”, Informatica, vol. 43(1), 2019.

[12] Z. Yamak, J. Saunier, L. Vercouter, “SocksCatch: Automatic detection and grouping of sockpuppets in social media”, Knowledge-Based Systems, 149, pp. 124-142, 2018.

[13] “Fake-Profile-Detection-using-ML” https://github.com/harshitkgupta/Fake-Profile-Detection-using-ML

ABOUT METHODS OF FAKE PROFİLES DETECTION IN SOCIAL NETWORKS

Yadigar İmamverdiyev[1], Khayala Ahmadova[2]

[1,2]Institute of Information Technology of ANAS, Baku, Azerbaijan

[1]yadigar@iit.science.az, [2]ehmedovaxeyale.97@mail.ru

Abstract– As social networks expand, so does their influence over people. Because personal information is also shared on social networks, fake profiles, which are threats to users, endanger this information. Personal information obtained directly or indirectly from a real user is then used for various purposes (for example, extorting money, creating another profile on behalf of that person, spreading false information on his behalf, etc.). As the purpose of fake profiles becomes more and more dangerous, their identification is one of the most pressing issues of our day. This article looks at the types of fake profiles, the features that can be used to identify fake profiles, the machine learning approach to fake profiles, and the proposed approach for automatic detection.

Keywords– *social network; fake profile; troll profile; detection; machine learning*