شناسایی حساب‌های کاربری جعلی با استفاده از یادگیری ماشین: مطالعه‌ای روی شبکهٔ اجتماعی X (توییتر سابق)

محمد رجبیان

کارشناس ارشد، گروه فناوری اطلاعات، دانشکده صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

منیره حسینی*[1]

دانشیار، گروه فناوری اطلاعات، دانشکده صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

مدیریت اطلاعات

دوره ۹، شماره ۲

پاییز و زمستان ۱۴۰۲

چکیده: محبوبیت شبکه‌های اجتماعی آنلاین (OSNs) هر روز افزایش می‌یابد و تعداد کاربران و میزان استفاده از این شبکه‌های اجتماعی، روزبه‌روز بیشتر می‌شود. این موضوع باعث شده است که سودجویانی از این فضا سوءاستفاده کنند و به اقدامات مخرب و حتی غیرقانونی نظیر، کلاه‌برداری، انتشار اخبار جعلی، جعل هویت و فریب کاربران بپردازند. اغلب این افراد از حساب‌های کاربری جعلی، برای اقدامات مخربانه خود استفاده می‌کنند. در بیشتر پژوهش‌های گذشته، پژوهشگران ابتدا با استفاده از روش‌های کاهش ویژگی، انتخاب ویژگی یا استخراج ویژگی، تعداد ویژگی‌های مجموعه داده را کاهش می‌دادند؛ سپس با استفاده از روش‌های ترکیبی یادگیری ماشین، به ارائه روشی جهت افزایش صحت در شناسایی حساب‌های کاربری جعلی اقدام می‌کردند. در پژوهش جاری، ۶۷ ویژگی برای شناسایی حساب‌های کاربری جعلی انتخاب شده‌اند که مقادیر عددی دارند و این امر باعث می‌شود که نیاز به نرمال‌سازی و تحلیل متن وجود نداشته باشد و مدل آموزش‌دیده توسط یادگیری ماشین، به زبان خاصی وابسته نباشد و بتواند واقعی و جعلی بودن حساب‌های کاربری‌ای را که با زبان‌های مختلف ایجاد شده‌اند، شناسایی کند. همچنین روشی ترکیبی جدیدی ارائه شده است که در آن، مقدار پارامتر C الگوریتم ماشین بردار پشتیبان (SVM) با الگوریتم‌های بهینه‌سازی ازدحام ذرات (PSO) و بهینه‌ساز گرگ خاکستری (GWO) و شبکهٔ عصبی (NN) بهینه‌سازی می‌شود. در پژوهش جاری بهترین میانگین صحت به‌دست‌آمده، در صد مرتبه اجرای الگوریتم‌ها ۹۸/۲۸درصد بوده که هنگام ترکیب الگوریتم SVM با NN اتفاق افتاده است.

کلیدواژه‌ها: حساب کاربری جعلی، شبکهٔ اجتماعی X (توییتر)، یادگیری ماشین، الگوریتم ماشین بردار پشتیبان، بهینه‌سازی

---

۱. نویسنده مسئول: hosseini@kntu.ac.ir

مقدمه

شبکه‌های اجتماعی آنلاین[1] توجه زیادی را در رابطه با ارتباط بین کاربران و به اشتراک‌گذاری اطلاعات بین یکدیگر به خود جلب کرده‌اند. افراد از سایت‌های شبکه‌های اجتماعی آنلاین برای برقراری ارتباط با دوستان، دوستان و دوستان، ملاقات با افراد جدید، ایجاد روابط حرفه‌ای، اشتراک‌گذاری اخبار، سازمان‌دهی رویدادها و حتی ادارهٔ تجارت الکترونیکی خود استفاده می‌کنند (ساهو و گوپتا[2]، ۲۰۱۹؛ خالد، التازی و مختار[3]، ۲۰۱۸). این موضوع زمینه را برای انجام فعالیت‌هایی مانند کلاه‌برداری، انتشار اخبار جعلی، جعل هویت و فریب کاربران توسط سودجویان، فراهم کرده است. این افراد برای انجام فعالیت‌های مخربانهٔ خود، معمولاً از حساب‌های کاربری جعلی استفاده می‌کنند. امروزه حساب‌های جعلی به یک تهدید بسیار جدی در شبکه‌های اجتماعی آنلاین تبدیل شده‌اند. بنابراین، استفاده از روش‌هایی به‌منظور شناسایی این حساب‌های کاربری جعلی بسیار حائز اهمیت است (ماجیب و گوپتا[4]، ۲۰۲۲).

اغلب حساب‌های کاربری جعلی شبکه‌های اجتماعی، به‌دلیل ماهیت خودکار بودن، رفتارهایی را از خود بروز می‌دهند که در مقایسه با کاربران واقعی متفاوت‌اند. همچنین با توجه به پیچیدگی‌های تحلیل ویژگی‌های حساب‌های کاربری و ماهیت خودکار بودن حساب‌های کاربری جعلی، استفاده از الگوریتم‌های طبقه‌بندی[5]، می‌تواند کمک شایانی در شناسایی حساب‌های کاربری جعلی داشته باشد (کرسکی، دی پیترو، پتروچی، اسپوگناردی و تسکونی[6]، ۲۰۱۵). پژوهشگران سعی داشته‌اند تا با استفاده از الگوریتم‌های مختلف طبقه‌بندی و روش‌های ترکیبی متفاوت، روشی را ارائه کنند که بهینه و کارآمد باشد و ضعف‌های موجود در این زمینه را مرتفع سازند.

این پژوهش با هدف ارائه روشی کاربردی و مبتنی بر یادگیری ماشین به‌منظور بهبود بخشیدن به شناسایی حساب‌های کاربری جعلی شبکه‌های اجتماعی انجام شده است و در آن به مطالعه و بررسی روی شبکهٔ اجتماعی X (توییتر سابق) که یکی از پرطرفدارترین شبکه‌های اجتماعی بوده و توجه تعداد زیادی از کاربران (ساهو و گوپتا، ۲۰۱۹) و پژوهشگران را به خود اختصاص داده است، پرداخته می‌شود. این پژوهش در پنج بخش تدوین شده است، در بخش‌های بعدی، ابتدا به بیان پیشینه پژوهش پرداخته می‌شود که در آن مروری بر کلیدی‌ترین پژوهش‌های پیشین در این زمینه انجام می‌شود، پس از بررسی پژوهش‌های پیشین، روش پژوهش جاری ابتدا به‌صورت مبانی بیان و بررسی می‌شود؛ سپس در محیط آزمایشگاهی پیاده‌سازی می‌شود. نتایج حاصل از مدل ارائه شده در این پژوهش، در بخش تجزیه‌وتحلیل یافته‌ها بیان و مورد تجزیه‌وتحلیل قرار می‌گیرد و در نهایت، در بخش نتیجه‌گیری، نتیجه‌گیری نهایی صورت می‌گیرد و پیشنهادهایی برای انجام کارهای آتی ارائه می‌شود.

---

1. Online Social Networks (OSNs)
2. Sahoo & Gupta
3. Khaled, El-Tazi & Mokhtar
4. Mujeeb, Gupta
5. Classification
6. Cresci, Di Pietro, Petrocchi, Spognardi & Tesconi

پیشینهٔ پژوهش

پژوهشگران فراوانی به شناسایی حساب‌های کاربری جعلی شبکه‌های اجتماعی مختلف با استفاده از روش‌ها و الگوریتم‌های مختلف یادگیری ماشین پرداخته‌اند و برخی از آن‌ها از روش‌های ترکیبی نیز جهت بهبود صحت[1] در شناسایی حساب‌های کاربری جعلی استفاده کرده‌اند. در زیر به شرح پژوهش‌های پیشین در این زمینه می‌پردازیم.

جیا، ونگ و گانگ[2] (۲۰۱۷) از روش یادگیری ماشین و الگوریتم‌های SybilWalk، SybilWalk-Var، CIA[3] و SybilRank برای شناسایی حساب‌های کاربری جعلی شبکه‌های اجتماعی توییتر، فیسبوک، انرون[4] و اپینیونز[5] استفاده کردند. مجموعه داده‌های استفاده شده در این پژوهش برای شبکه‌های اجتماعی فیسبوک، انرون و اپینیونز از مجموعه داده «SNAP» استفاده شده و برای شبکهٔ اجتماعی توییتر نیز از گراف هدایت شده پژوهش کواک، لی، پارک و مون[6] (۲۰۱۰) استفاده شده است. پژوهشگران با استفاده از الگوریتم SybilWalk توانستند حساب‌های کاربری جعلی را با صحت ۹۶ درصد شناسایی کنند.

ملیگی، ابراهیم و ترکی[7] (۲۰۱۷) پژوهشی به‌منظور شناسایی حساب‌های کاربری جعلی سه شبکهٔ اجتماعی توییتر، فیسبوک و گوگل پلاس انجام داده‌اند. در این پژوهش از سه مجموعه داده با نام‌های «Ego-Facebook»، «Ego-Google+» و «Ego-Twitter» استفاده شده است که فایل‌های آن‌ها در کتابخانه «SNAB» است. این پژوهش با استفاده از روش یادگیری ماشین و الگوریتم‌های RE[8] و DFA[9] انجام شد و توانست با دقت ۸۲/۸۸ درصد حساب‌های کاربری جعلی را شناسایی کند.

خالد و همکاران (۲۰۱۸) با استفاده از ترکیب الگوریتم SVM[10] و NN[11] و مجموعه داده MIB شبکهٔ اجتماعی توییتر که در پژوهش کرسکی و همکاران (۲۰۱۵) استفاده شده است، توانستند حساب‌های کاربری جعلی را با صحت حدود ۹۸ درصد شناسایی کنند.

رواتی و سوریاکالا[12] (۲۰۱۸) با استفاده از یادگیری ماشین، الگوریتم‌های SVM، KNN[13] و NSCM[14] و مجموعه داده‌ای از شبکهٔ اجتماعی فیسبوک که توسط پژوهشگران جمع‌آوری شده بود، به شناسایی

1. Accuracy
2. Jia, Wang & Gong
3. Criminal account Inference Algorithm
4. Enron
5. Epinions
6. Kwak, Lee, Park & Moon
7. Meligy, Ibrahim & Torky
8. Regression
9. Deterministic Finite Automata
10. Support Vector Machine
11. Neural Network
12. Revathi & Suriakala
13. K-Nearest Neighbors
14. Node Similarity Communication Matching

حساب‌های کاربری جعلی فیسبوک اقدام کردند و توانستند حساب‌های کاربری جعلی را با صحت ۹۳/۸۷درصد شناسایی کنند.

سوریاکالا و رواتی (۲۰۱۸) با استفاده از روش یادگیری ماشین و الگوریتم‌های RF[1]، SVM، SMO[2] و PPS[3] به شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی فیسبوک پرداختند. در این پژوهش از یک مجموعه دادهٔ عمومی منتشر شده در Github استفاده شده است. پژوهشگران توانستند حساب‌های کاربری تکراری ایجاد شده توسط یک شخص را با دقت ۹۷/۳۰درصد شناسایی کنند.

وانی، آگراوال، جابین و حسین[4] (۲۰۱۹) با استفاده از الگوریتم‌های SVM، RF، JRip و NB[5] و مجموعه داده‌ای از شبکهٔ اجتماعی فیسبوک که توسط پژوهشگران این پژوهش با ایجاد دو ظرف عسل[6] به‌منظور جذب حساب‌های کاربری واقعی و جعلی تهیه شده بود، به شناسایی حساب‌های کاربری جعلی اقدام کردند و با استفاده از الگوریتم RF توانستند با صحت ۹۰ درصد حساب‌های کاربری جعلی را شناسایی کنند.

پاکایا، ابراهیم و بودی[7] (۲۰۱۹) با استفاده از روش یادگیری ماشین، الگوریتم‌های LR[8]، AdaBoost[9]، XGBoost و RF و مجموعه داده‌ای که در پژوهش کرسکی و همکاران (۲۰۱۷) جمع‌آوری شده و شامل ۷۵۴۳ حساب کاربری Spambot، ۳۳۵۱ دنبال کننده جعلی و ۳۴۷۴ حساب کاربری واقعی است، به شناسایی حساب‌های کاربری جعلی پرداختند. در نتیجه این پژوهش، پژوهشگران توانستند با حداکثر دقت ۹۵/۵۵درصد، حساب‌های کاربری جعلی را شناسایی کنند.

بالا اناند و همکاران[10] (۲۰۱۹) از روش یادگیری ماشین و الگوریتم‌های EGSLA[11]، Game theory، KNN، SVM و DT[12]، برای شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی توییتر استفاده کردند. مجموعه داده مورد استفاده در این پژوهش توسط پژوهشگران با خزش وب به وجود آمده است که شامل اطلاعات ۲۱۴۷۳ کاربر و ۲۹۱۵۱۴۷ توییت است. نتایج این پژوهش نشان داد که الگوریتم EGSLA امکان شناسایی کاربران جعلی با صحت ۹۰/۳ درصد را دارد.

بهارتی و پاندی[13] (۲۰۲۱) از روش یادگیری ماشین و الگوریتم LR بهینه‌سازی با الگوریتم PSO[14] و الگوریتم‌های NB، DT و LR بدون بهینه‌سازی، برای شناسایی حساب‌های کاربری جعلی استفاده کردند.

1. Random Forests
2. Sequential Minimal Optimization
3. Privacy Protected System
4. Wani, Agarwal, Jabin & Hussain
5. Naive Bayes
6. Honey pot
7. Pakaya, Ibrohim & Budi
8. Logistic Regression
9. Adaptive Boosting
10. Bala Anand et al.
11. Enhanced Gravitational Search-Like Algorithm
12. Decision Tree
13. Bharti & Pandey
14. Particle Swarm Optimization

مجموعه داده استفاده شده در این پژوهش، ترکیبی از مجموعه داده‌ای به نام «TheFakeProject» که در پژوهش کرسکی و همکاران (۲۰۱۵) ایجاد شده و داده‌های جمع‌آوری شده با استفاده از REST API است که در این مجموعه داده ۳۴۷۴ حساب کاربری واقعی و ۴۹۱۲ حساب کاربری جعلی وجود دارد. در نتیجه این پژوهش، الگوریتم LR بهینه‌سازی شده با PSO بهترین نتیجه را داشت و در آن میزان صحت ۹۶/۲درصد، F-Score ۸۹ درصد، PPR[1] حدود ۷۹/۹درصد و NPR[2] حدود ۹۷/۳ درصد به‌دست آمد.

جباردی و هادی[3] (۲۰۲۱) با استفاده از روش یادگیری ماشین و ontology construction، ایجاد قوانین SWRL[4] و طبقه‌بندی‌کننده دلیل معنایی[5]، به شناسایی حساب‌های کاربری جعلی اقدام کردند. مجموعه داده این پژوهش، مجموعه داده استفاده شده در «Fake Project» است که توسط مؤسسهٔ انفورماتیک و تله ماتیک آزمایشگاه شورای ملی پژوهش‌های ایتالیا (IIT-CNR) برای شبکهٔ اجتماعی توییتر ایجاد شده است (کرسکی و همکاران، ۲۰۱۵). این مجموعه داده شامل ۸۲۶۳ حساب کاربری جعلی و ۳۴۷۴ حساب کاربری واقعی است. در نهایت، روش ارائه شده در این پژوهش موفق شد که حساب‌های کاربری جعلی را با دقت ۹۷/۵ درصد شناسایی کند.

کوندتی، یرامردی، پرادان و سواین[6] (۲۰۲۱) در پژوهش خود به شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی توییتر پرداختند. در این پژوهش از روش‌های Z-Score و Min–Max برای نرمال‌سازی و الگوریتم‌های SVM، LR، RF و K-NN برای طبقه‌بندی استفاده شده است. مجموعه داده استفاده شده در این پژوهش توسط پژوهشگران جمع‌آوری شده است و در نتیجه این پژوهش، حساب‌های کاربری جعلی شبکهٔ اجتماعی توییتر با صحت ۹۳/۴ درصد شناسایی شدند.

ماجیب و گوپتا (۲۰۲۲) با استفاده از دو روش یادگیری ماشین و تحلیل کلان‌داده روی الگوریتم‌های GB[7]، RF و DT و مجموعه داده‌ای عمومی از شبکهٔ اجتماعی توییتر که در Github منتشر شده است، به شناسایی حساب‌های کاربری جعلی پرداختند. در نتیجه این پژوهش، با روش یادگیری ماشین و الگوریتم GB، حساب‌های کاربری جعلی با صحت ۹۷/۰۴ درصد و با روش تحلیل کلان داده و الگوریتم GB، حساب‌های کاربری جعلی با صحت ۹۹/۵ درصد شناسایی شدند.

تانگ، ژنگ، لیانگ، لی و سوخیجا[8] (۲۰۲۲) به شناسایی حساب‌های کاربری روی مجموعه داده‌ای متشکل از تمامی شبکه‌های اجتماعی پرداختند. این مجموعه داده، مجموعه داده‌ای خصوصی است که توسط یک شرکت فناوری هوش مصنوعی ارائه شده و شامل ۵میلیون ۳۸۰ هزار دادهٔ رفتاری برای

1. Positive Predictive Rate
2. Negative Predictive Rate
3. Jabardi, Hadi
4. Semantic Web Rule Language
5. Semantic Reasoner Classifier
6. Kondeti, Yerramreddy, Pradhan & Swain
7. Gradient Boosting
8. Tang, Zhang, Liang, Li & Sukhija

۱میلیون و ۳۸۰ هزار کاربر است. در این پژوهش از روش‌های ترکیبی M-Zoom، CPD[1]، MAF[2]، HoloScope و MFML[3] توانست روشی ارائه دهد که دقت و سرعت را بهبود بخشد.

پرابهو کاوین و همکاران[4] (۲۰۲۲) با استفاده از سه الگوریتم SVM، ANN[5] و RF و مجموعه داده‌ای متشکل از ترکیب مجموعه داده پروژهٔ The Fake Project و داده‌هایی که به‌صورت دستی و با استفاده از Rest API جمع‌آوری شده‌اند، به شناسایی حساب‌های کاربری جعلی اقدام کردند و دریافتند الگوریتم SVM بیشترین میزان دقت را نسبت به دو الگوریتم دیگر به‌دست آورد.

الحسان و رسام[6] (۲۰۲۲) با استفاده از دو روش یادگیری ماشین و یادگیری عمیق، الگوریتم‌های SVM، DT، NB، LR، LSTM[7] و روش ترکیبی CNN[8] و NN به شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی توییتر اقدام کردند. مجموعه دادهٔ استفاده شده در این پژوهش، توسط پژوهشگران ایجاد شد و آن‌ها توانستند با استفاده از روش ترکیبی CNN و NN حساب‌های کاربری جعلی را با صحت ۹۴/۲۷ درصد شناسایی کنند.

شن و همکاران[9] (۲۰۲۳) با استفاده از یادگیری ماشین و الگوریتم‌های CART[10]، SVM، DNN[11]، RF، AdaBoost و XGBoost به شناسایی حساب‌های کاربری جعلی یک پورتال دوست‌یابی به‌نام «Xie Hou Si Nian» پرداختند. در این پژوهش مجموعه داده‌ای با ۶۱۵۲ حساب کاربری حقیقی و ۷۸ حساب کاربری جعلی توسط پژوهشگران استخراج شد. در نهایت مدل پیشنهادی این پژوهش توانست با دقت ۵۹/۱۶ درصد و بازخوانی ۷۳ درصد، حساب‌های کاربری جعلی را شناسایی کند.

سای راجا، آدیتیا و موهانتی[12] (۲۰۲۳) با استفاده از یادگیری ماشین و الگوریتم‌های SVM، NB، LR، DT و ترکیب الگوریتم‌های LR و GD[13]، به شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی توییتر پرداختند. در این پژوهش از ترکیب مجموعه داده عمومی که در کاگل منتشر شده بود و داده‌های جمع‌آوری شده به‌وسیلهٔ APIهای اینستاگرام استفاده شده است. در نتیجهٔ این تحقیق، پژوهشگران توانستند حساب‌های کاربری جعلی را با صحت ۹۲/۷۰ درصد، دقت ۸۹ درصد و F1-Score ۹۲ درصد شناسایی کنند.

ناگا پراوینا، آدیتیا و موهانتی[14] (۲۰۲۴) از روش پردازش متن در یادگیری ماشین و الگوریتم‌های SVM، NB و KNN برای شناسایی حساب‌های کاربری جعلی شبکهٔ اجتماعی فیسبوک استفاده کردند.

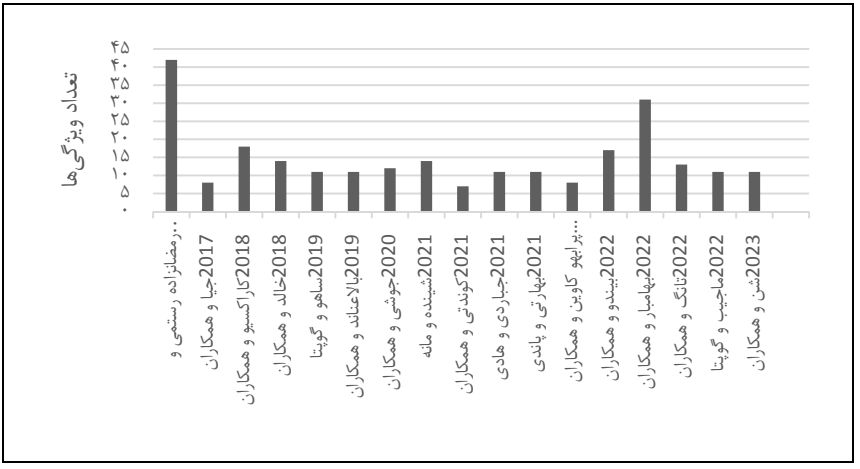
1. Customer Data Platforms
2. Masked Autoregressive Flow
3. Multimodal Fusion Machine Learning
4. Prabhu Kavin
5. Artificial Neural Network
6. Alhassun, Rassam
7. Long Short-Term Memory
8. Convolutional Neural Network
9. Shen et al.
10. Classification and Regression Trees
11. Deep neural networks
12. Sai Raja, Aditya & Mohanty
13. Gradient Descent
14. Naga Praveena

مجموعه داده استفاده شده در این پژوهش، توسط پژوهشگران جمع‌آوری شدند و با روش ارائه شده در این پژوهش، توانستند حساب‌های کاربری جعلی را با دقت ۵۳/۱۷ درصد شناسایی کنند.

همان طور که در بررسی پژوهش‌های کلیدی بیان شد، بسیاری از پژوهشگران از الگوریتم طبقه‌بندی SVM جهت شناسایی حساب‌های کاربری واقعی و جعلی استفاده کرده‌اند و در بسیاری از موارد، این الگوریتم نتایج رضایت بخشی داشته است. در میان پژوهش‌های بررسی شده، کمتر دیده شده است که پژوهشگران به بهینه‌سازی الگوریتم SVM بپردازند تا نتایج حاصل از آن را بهبود ببخشند.

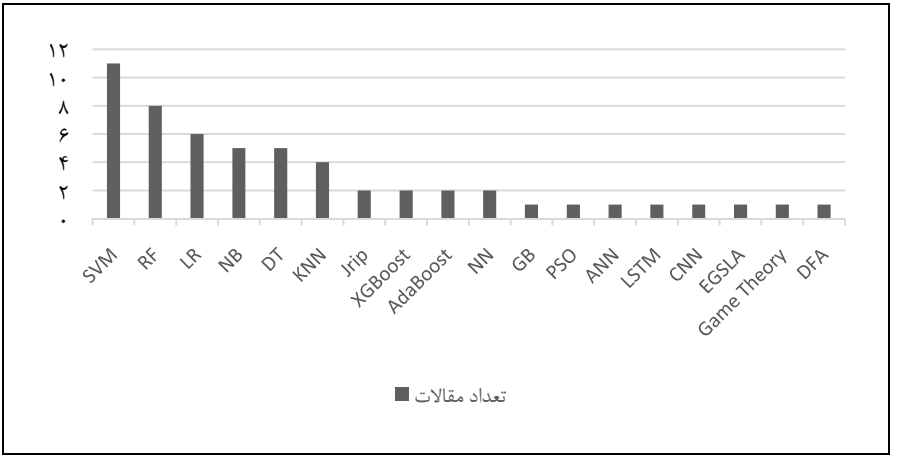
شکل ۱ تعداد ویژگی‌های حساب‌های کاربری را که برای شناسایی واقعی و جعلی بودن حساب کاربری استفاده می‌شود، در ۱۷ مورد از مقاله‌های کلیدی گذشته مقایسه کرده است. همان طور که در این شکل مشاهده می‌شود، پژوهش رمضان‌زاده رستمی و کرباسی (۱۳۹۷)، با توجه به ۴۲ ویژگی، بیشترین تعداد ویژگی و پژوهش کوندتی و همکاران (۲۰۲۱) با استفاده کردن از ۷ ویژگی، کمترین تعداد ویژگی را بررسی کرده‌اند.

همان طور که در این شکل مشاهده می‌شود، تعداد ویژگی‌های بررسی شده توسط پژوهشگران پیشین، به‌طور عمده کمتر از ۲۰ ویژگی است که اغلب ویژگی‌های کلی حساب‌های کاربری شبکه‌های اجتماعی هستند و پژوهشگران اندکی به تحلیل ویژگی‌های جزئی برروی حساب‌های کاربری شبکه‌های اجتماعی پرداخته‌اند.



شکل ۱. نمودار تعداد ویژگی‌های مورد بررسی قرار گرفته برای شناسایی حساب‌های کاربری جعلی

در شکل ۲ مشخص شده است که در ۲۳ مورد از کلیدی‌ترین پژوهش‌های انجام شده در زمینهٔ شناسایی حساب‌های کاربری جعلی شبکه‌های اجتماعی، پژوهشگرها بیشتر از کدام الگوریتم‌ها استفاده کرده‌اند. همان طور که در این شکل مشخص است، سه الگوریتم SVM، RF و LR پُراستفاده‌ترین الگوریتم‌ها در این زمینه بوده‌اند.



شکل ۲. نمودار الگوریتم‌های استفاده شده در تحقیقات پیشین

بر اساس بررسی‌های انجام شده در این بخش، دریافتیم که الگوریتم SVM از پرکاربردترین الگوریتم‌های استفاده شده در پژوهش‌های پیشین بوده است؛ اما کمتر به بهینه‌سازی این الگوریتم توجه شده است. همچنین در پژوهش‌های پیشین، بیشتر ویژگی‌های عمومی حساب‌های کاربری مورد بررسی قرار گرفته‌اند و فقط برخی از پژوهشگران به ویژگی‌های جزئی حساب‌های کاربری توجه کرده‌اند. در این پژوهش سعی بر آن است تا بهترین مقدار پارامتر C را که یکی از مهم‌ترین پارامترهای الگوریتم SVM است با استفاده از الگوریتم‌های بهینه فرا ابتکاری پُرطرفدار PSO و GWO[1] و همچنین الگوریتم شبکه‌های عصبی، به‌صورت خودکار پیدا کنیم و از الگوریتم SVM با پارامترهای دقیق‌تر و به‌صورت بهینه‌تر استفاده کنیم.

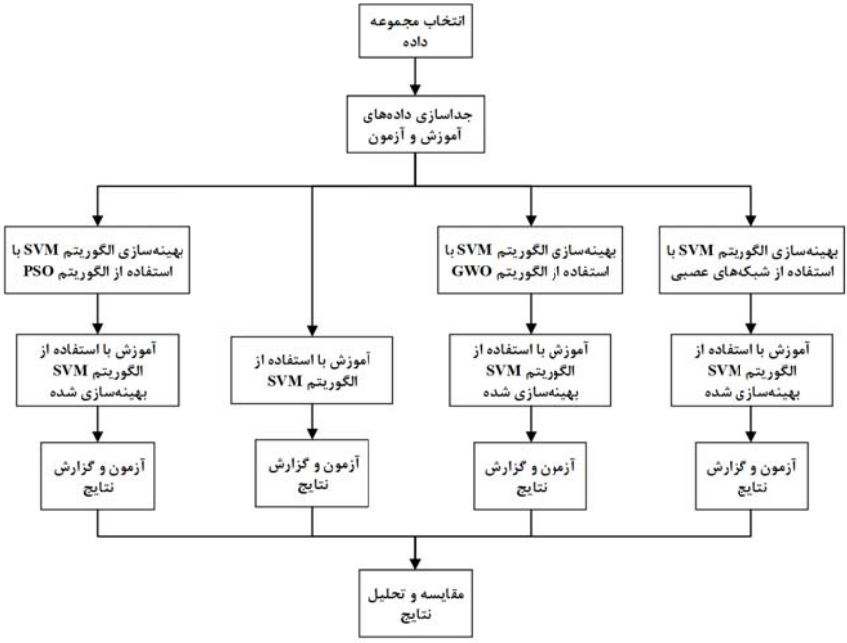
روش‌شناسی پژوهش

همان طور که در بخش قبل بیان شد، پژوهش‌های بسیاری در زمینه شناسایی حساب‌های کاربری جعلی با استفاده از روش‌های یادگیری ماشین انجام شده است، در بیشتر این پژوهش‌ها، پژوهشگران از مجموعه داده‌هایی با تعداد ویژگی‌های کم و کلی استفاده کرده‌اند. با توجه به اینکه الگوریتم طبقه‌بندی SVM در فضاهای با ابعاد بالا مؤثر است (بن ساسی و بن یحیا[2]، ۲۰۲۱) و همچنین بژوهشگران بسیاری از این الگوریتم برای شناسایی حساب‌های کاربری جعلی استفاده کرده‌اند، به نظر می‌رسد در صورت استفاده از این الگوریتم و مجموعه داده‌ای که تعداد زیادی از ویژگی‌های کلی و جزئی را دارند، بتوان شناسایی حساب‌های کاربری جعلی را با دقت بالایی انجام داد. همچنین برای بهبود عملکرد الگوریتم SVM

1. Gray Wolf Optimization
2. Ben Sassi, Ben Yahia

می‌توان به بهینه‌سازی پارامترهای این الگوریتم با استفاده از الگوریتم‌های بهینه‌سازی فراابتکاری PSO و GWO که جزء محبوب‌ترین الگوریتم‌های بهینه‌سازی هستند و الگوریتم NN که در پژوهش خالد و همکاران (۲۰۱۸) عملکرد خوبی در شناسایی حساب‌های کاربری جعلی داشته است، پرداخت. در صورتی که داده‌های مورد استفاده مقادیر عددی داشته باشند، مشکلاتی از قبیل نیازمندی به نرمال‌سازی، تا حد زیادی مرتفع خواهد شد و مزایایی از قبیل عدم وابستگی به زبانی را که کاربران در ساخت حساب کاربری استفاده کرده‌اند، به ارمغان خواهد آورد.

همچنین در این پژوهش قصد بر آن داریم که الگوریتم Linear-SVM را با استفاده از الگوریتم‌های PSO، GWO و NN بهینه‌سازی کنیم؛ در این بهینه‌سازی توسط هر یک از این الگوریتم‌ها مقادیری در بازهٔ مجاز به پارامتر C که از پارامترهای کلیدی الگوریتم SVM است داده می‌شود و میزان صحت مدل آموزش‌دیده با مقادیر مختلف پارامتر C ارزیابی می‌شود تا بهترین مقدار پارامتر C که بیشترین میزان صحت را دارد، مشخص شود. شایان ذکر است که در روش مقادیر پارامتر C توسط الگوریتم‌های PSO، GWO و NN به‌صورت خودکار برای پارامتر C تعیین می‌شود، در این روش مقادیر پارامتر C بر اساس مقادیر قبلی آن و نتایج به‌دست‌آمده از آن تعیین می‌شود.



شکل ۳. مدل ارائه شده جهت شناسایی حساب‌های کاربری جعلی

شکل ۳ نمایانگر مدل ارائه شده برای پژوهش جاری است، همان طور که در این شکل مشاهده می‌کنید، گام‌های پژوهش بدین صورت است که ابتدا مجموعه داده مورد نیاز تهیه و تدوین می‌شود، سپس داده‌ها به‌صورت تصادفی به دو بخش آموزش و آزمون تقسیم می‌شود، داده‌های آموزش در چهار حالت الگوریتم SVM بدون بهینه‌سازی و الگوریتم SVM بهینه‌سازی شده با الگوریتم‌های PSO، GWO و NN مورد آموزش قرار می‌گیرند و توسط داده‌های آزمون مورد ارزیابی قرار می‌گیرند و پارامترهای ارزیابی مورد نظر برای آن‌ها محاسبه می‌شود. در نهایت نتایج به‌دست‌آمده مورد تجزیه‌وتحلیل قرار می‌گیرند تا مناسب‌ترین حالت آموزش جهت شناسایی حساب‌های کاربری جعلی مشخص شود.

اجرا در محیط آزمایشگاهی

بر اساس مدل ارائه شده در این پژوهش، اولین نیاز ما داشتن یک مجموعه داده که شامل حساب‌های کاربری واقعی و جعلی شبکهٔ اجتماعی X است که تعداد زیادی از ویژگی‌های حساب کاربری را در خود داشته باشد و این ویژگی‌ها دارای مقادیر عددی باشند. مجموعه داده مورد استفاده قرار گرفته در این پژوهش یک مجموعه داده عمومی است که در پژوهش سیا[1] (۲۰۲۱) جمع‌آوری شده است و این مجموعه داده دارای ۳۴۷۴ حساب کاربری واقعی، ۴۹۱۲ حساب کاربری جعلی و ۶۷ ویژگی با مقادیر عددی برای هر حساب کاربری است. فهرست تمامی ویژگی‌هایی که از این مجموعه داده، در این پژوهش استفاده شده‌اند، به همراه توضیحات در جدول ۱ به تفصیل بیان شده است.

جدول ۱. فهرست ویژگی‌های حساب‌های کاربری در مجموعه داده مورد استفاده به همراه توضیحات هر کدام

ویژگی	توضیحات
class_bot	واقعی یا جعلی بودن حساب کاربری را نشان می‌دهد.
statuses_count	تعداد توییت‌های انجام شده را بیان می‌کند.
followers_count	تعداد دنبال‌شوندگان را مشخص می‌کند.
friends_count	تعداد دنبال‌کنندگان را نمایش می‌دهد.
favourites_count	تعداد لایک‌های انجام‌شده را مشخص می‌کند.
listed_count	نشان‌دهنده تعداد فهرست‌های عمومی حساب کاربری است.
default_profile	بیانگر این است که حساب کاربری از پروفایل پیش‌فرض استفاده می‌کند یا خیر.
default_profile_image	مشخص می‌کند که حساب کاربری از عکس پروفایل پیش‌فرض استفاده می‌کند یا خیر.
geo_enabled	نشان می‌دهد که در حساب کاربری مجوز انتشار موقیت مکانی به همراه توییت‌ها داده شده است یا خیر.
profile_use_background_image	نشان می‌دهد که آیا پروفایل از عکس پس‌زمینه استفاده کرده است یا خیر.

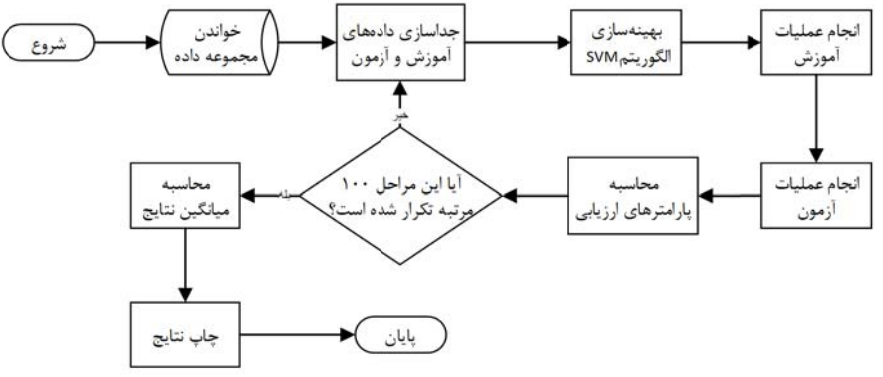
1. Cea

ویژگی	توضیحات
profile_background_tile	مشخص می‌کند که عکس پس‌زمینه پروفایل کاربر دارای کاشی است یا خیر.
utc_offset	موقعیت زمانی منطقه‌ای را مشخص می‌کند.
Protected	بیانگر وضعیت محافظت و یا عدم محافظت از حساب کاربری است.
Verified	وضعیت تأیید یا عدم تأیید حساب کاربری را نشان می‌دهد.
min_tweet_length	حداقل تعداد کاراکترهای توییت‌های انجام شده را بیان می‌کند.
max_tweet_length	حداکثر تعداد کاراکترهای توییت‌های انجام شده را مشخص می‌سازد.
avg_tweet_length	میانگین تعداد کاراکترها در توییت‌های انجام شده را نشان می‌دهد.
min_urls	حداقل تعدادURL های استفاده شده در توییت‌ها را نشان می‌دهد.
max_urls	حداکثر تعدادURL های استفاده شده در توییت‌ها را نشان می‌دهد.
avg_urls	میانگین تعدادURL های استفاده شده در توییت‌ها را نشان می‌دهد.
min_hashtags	بیانگر حداقل تعداد هشتگ‌های استفاده شده در هر توییت است.
max_hashtags	بیانگر حداکثر تعداد هشتگ‌های استفاده شده در هر توییت است.
avg_hashtags	بیانگر میانگین تعداد هشتگ‌های استفاده شده در هر توییت است.
max_mentions	حداکثر تعداد منشن در هر توییت را مشخص می‌کند.
avg_mentions	میانگین تعداد منشن در هر توییت را مشخص می‌کند.
max_retweets	حداکثر تعداد ریتوییت‌ها را نشان می‌دهد.
avg_retweets	میانگین تعداد ریتوییت‌ها را نشان می‌دهد.
min_favorite	حداقل تعداد لایک‌های توییت‌های انجام شده، است.
max_favorite	حداکثر تعداد لایک‌های توییت‌های انجام شده، است.
avg_favorite	میانگین تعداد لایک‌های توییت‌های انجام شده، است.
num_reply	تعداد پاسخ‌ها را مشخص می‌کند.
num_retweet	تعداد ریتوییت‌ها را مشخص می‌کند.
digits_name	تعداد عددهای موجود در فیلد نام را بیان می‌کند.
digits_screen_name	تعداد عددهای موجود در نام صفحه را بیان می‌کند.
name_length	تعداد کاراکترهای فیلد نام را نشان می‌دهد.
screen_name_length	تعداد کاراکترهای نام صفحه را نشان می‌دهد.
screen_name_length_ name_length_ratio	نسبت تعداد کاراکترهای نام صفحه به فیلد نام را نشان می‌دهد.
name_contains_bot	بیانگر این است که در فیلد نام کلمه «Bot» وجود دارد یا خیر.
screen_name_contains_bot	بیانگر این است که در نام صفحه کلمه «Bot» وجود دارد یا خیر.
name_entropy	نمایانگر آنتروپی نام است.

ویژگی	توضیحات
screen_name_entropy	نمایانگر آنتروپی نام صفحه است.
name_screen_name_lev_sim	میزان شباهت نام صفحه به فیلد نام را مشخص می‌کند.
description_length	نمایشگر تعداد کاراکترهای متن توضیحات است.
description_contains_bot	بیان می‌کند که در متن توضیحات از کلمه «Bot» استفاده شده است یا خیر.
profile_has_url	نشان‌دهندهٔ این است که پروفایل دارای آدرس URL است یا خیر.
profile_pic_frec	تعداد دفعاتی که عکس پروفایل تغییر کرده است را نشان می‌دهد.
friends_followers_ratio	نسبت تعداد دنبال‌کنندگان به دنبال‌شوندگان را بیان می‌کند.
friends_followers _ratio_beq_50	نشان می‌دهد که آیا نسبت دنبال‌کنندگان به دنبال‌شوندگان برابر و یا بزرگ تر از ۵۰٪ است؟
friends_followers_square_ratio	نسبت تعداد دنبال‌کنندگان به مجذور تعداد دنبال‌شوندگان.
friends_followers+friends_ratio	نسبت تعداد دنبال‌کنندگان به مجموع تعداد دنبال‌کنندگان و دنبال‌شوندگان.
2_followers_minus_friends	نشان‌دهنده اختلاف تعداد دنبال‌شوندگان و دنبال‌کنندگان است.
2_followers_beq100	مشخص می‌کند که تعداد دنبال‌شوندگان بزرگ‌تر یا برابر ۱۰۰ است یا خیر.
lists_followers_ratio	نمایانگر نسبت تعداد فهرست‌ها به تعداد دنبال‌شوندگان است.
retweet_followers_ratio	تعداد ریتوییت‌ها به تعداد دنبال‌شوندگان را نشان می‌دهد.
favorites_followers_ratio	تعداد لایک‌ها به تعداد دنبال‌شوندگان را نمایش می‌دهد.
lists_status_ratio	تعداد فهرست‌ها به تعداد توییت‌ها را نشان می‌دهد.
retweet_status_ratio	تعداد ریتوییت‌ها به تعداد توییت‌ها را نمایان می‌سازد.
favorites_status_ratio	تعداد لایک‌ها به تعداد توییت‌ها را بیان می‌کند.
reply_status_ratio	تعداد پاسخ‌ها به تعداد توییت‌ها را نشان می‌دهد.
avg_urls_status_ratio	نمایانگر نسبت میانگین تعداد URLهای موجود در هر توییت به تعداد تمامی توییت‌هاست.
avg_mentions_status_ratio	میانگین تعداد منشن‌های هر توییت به تعداد کل توییت‌های کابر است.
avg_favorite_status_ratio	میانگین تعداد لایک‌های هر توییت به تعداد تمامی توییت‌ها است.
account_age	سن حساب کاربری را نشان می‌دهد.
followers_account_age_ratio	نسبت تعداد دنبال‌شوندگان به سن حساب کاربری را مشخص می‌کند.
friends_account_age_ratio	نبست تعداد دنبال‌کنندگان به سن حساب کاربری را بیان می‌کند.
statuses_account_age_ratio	نسبت تعداد توییت‌ها به سن حساب کاربری را نشان می‌دهد.
favourites_account_age _ratio	نسبت تعداد لایک‌ها به سن حساب کاربری مشخص می‌کند.
lists_account_age_ratio	نسبت تعداد فهرست‌ها به سن حساب کاربری را نشان می‌دهد.

برای پیاده‌سازی مدل ارائه شده در این پژوهش، از یک سیستم با پردازنده Intel Core i7 4500 و مقدار رم ۸ گیگابایت استفاده شده است. همچنین این مدل با استفاده از زبان برنامه‌نویسی پایتون پیاده‌سازی شده است که در آن ابتدا مجموعه داده را به‌صورت تصادفی و با نسبت‌های ۷۰درصد به ۳۰درصد به دو مجموعه داده آموزش و آزمون تقسیم می‌کنیم، سپس به بهینه‌سازی پارامتر C الگوریتم SVM اقدام می‌کنیم. پس از بهینه‌سازی، مدل یادگیری روی الگوریتم SVM بدون بهینه‌سازی و بهینه‌سازی شده با الگوریتم‌های PSO، GWO و NN اجرا می‌شود و درنهایت مدل‌های یادگیری با استفاده از داده‌های آزمون، مورد آزمایش قرار می‌گیرند و پارامترهای ارزیابی روی آن‌ها محاسبه می‌شود.

با توجه به اینکه جداسازی داده‌ها به‌صورت تصادفی موجب می‌شود که نتایج در هر بار اجرای کد تغییر کند و همچنین برای جلوگیری ارزیابی یک حالت خاص، یک حلقه تکرار مطابق با شکل ۴، طراحی شده است که باعث می‌شود مراحل جداسازی داده‌ها به‌صورت تصادفی، بهینه‌سازی SVM، اجرای مدل‌های یادگیری، آزمون و ارزیابی مدل‌های یادگیری، ۱۰۰ مرتبه تکرار شود و در نهایت میانگین ارزیابی‌های آن محاسبه شود.



شکل ۴. حلقهٔ تکرار طراحی شده در مدل پژوهش جاری

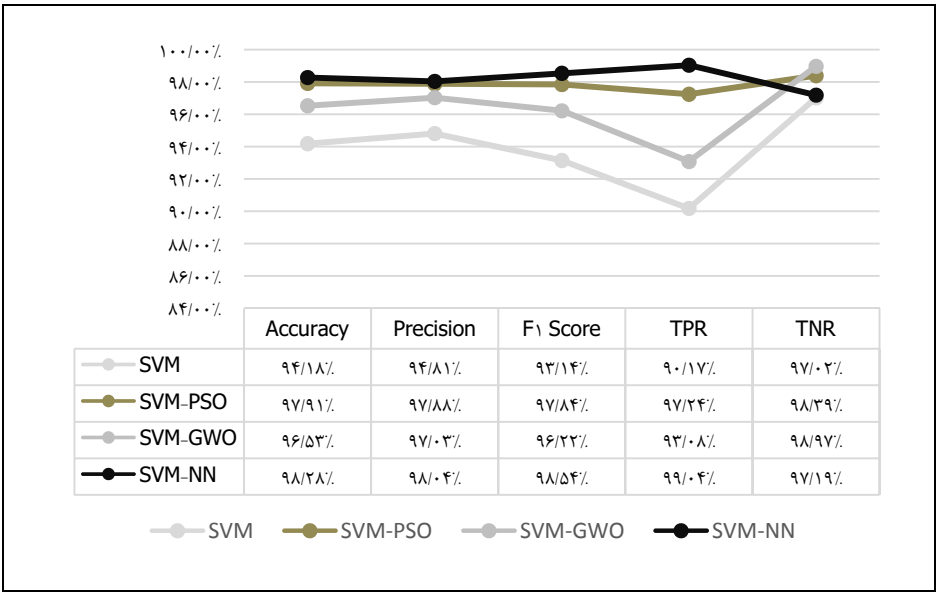
استفاده از میانگین پارامترهای ارزیابی منجر می‌شود که نتایج به‌دست‌آمده به نتایجی که با پیاده‌سازی مدل‌ها در محیط‌های کسب‌وکار اتفاق می‌افتند، نزدیک‌تر شود.

تجزیه‌وتحلیل یافته‌ها

مطابق با توضیحاتی که در بخش قبل ارائه شد، مدل پیشنهادی در یک حلقه تکرار ۱۰۰ مرتبه تکرار شد و میانگین پارامترهای ارزیابی برای آن محاسبه شد. شکل ۵ به مقایسه میانگین نتایج به‌دست‌آمده در هر یک از حالت SVM بدون بهینه‌سازی و SVM بهینه‌سازی شده توسط PSO، GWO و NN می‌پردازد. همان طور که مشاهده می‌شود، الگوریتم SVM بهینه‌سازی شده توسط NN، در تمامی پارامترهای

ارزیابی به‌جز نرخ منفی واقعی[1] بهترین نتیجه را داشته است و الگوریتم SVM بهینه‌سازی شده توسط الگوریتم GWO، بالاترین میزان نرخ منفی واقعی را داشته است.

با توجه به نتایج ارزیابی به‌دست‌آمده، مشخص شدکه در حالت کلی و زمانی که تشخیص صحیح حساب‌های کاربری جعلی برای ما اهمیت داشته باشد یا میزان اهمیت تشخیص حساب‌های کاربری جعلی و واقعی به یک اندازه باشد، الگوریتم SVM بهینه‌سازی شده با NN بهتر از سایر حالات الگوریتم SVM می‌تواند حساب‌های کاربری جعلی را شناسایی کند؛ اما در صورتی که اهمیت تشخیص درست حساب‌های کاربری واقعی برای ما اهمیت بیشتری داشته باشد، الگویتم SVM بهینه‌سازی شده با GWO عملکرد بهتری خواهد داشت.



	Accuracy	Precision	F۱ Score	TPR	TNR
SVM	۹۴/۱۸٪	۹۴/۸۱٪	۹۳/۱۴٪	۹۰/۱۷٪	۹۷/۰۲٪
SVM-PSO	۹۷/۹۱٪	۹۷/۸۸٪	۹۷/۸۴٪	۹۷/۲۴٪	۹۸/۳۹٪
SVM-GWO	۹۶/۵۳٪	۹۷/۰۳٪	۹۶/۲۲٪	۹۳/۰۸٪	۹۸/۹۷٪
SVM-NN	۹۸/۲۸٪	۹۸/۰۴٪	۹۸/۵۴٪	۹۹/۰۴٪	۹۷/۱۹٪

شکل ۵. مقایسه میانگین نتایج به‌دست‌آمده در ۱۰۰ مرتبه اجرای مدل پیشنهادی

باتوجه به آمار منتشر شده توسط دیکسون[2] (۲۰۲۳)، در سال ۲۰۲۲، شبکهٔ اجتماعی X حدود ۳۶۸میلیون ۴۰۰ هزار حساب کاربری فعال داشته است. با توجه به تعداد زیاد کاربران این شبکهٔ اجتماعی، بهبود بخشیدن به شناسایی حساب‌های کاربری جعلی بسیار حائز اهمیت می‌شود؛ به گونه‌ای که حتی در صورتی که بتوان مدل را ۱درصد بهبود بخشید، نتیجهٔ آن روی ۳میلیون و ۶۸۰ هزار حساب کاربری تأثیر می‌گذارد.

1. True Negative Rate (TNR)
2. Dixon

نتیجه‌گیری

در پژوهش جاری، ۶۷ ویژگی حساب‌های کاربری جعلی و واقعی که مقادیر تمامی ویژگی‌ها به‌صورت عدد بودند، با استفاده از ترکیب الگوریتم طبقه‌بندی SVM و الگوریتم‌های PSO، GWO و NN، تحلیل شدند تا بتوان الگوی مناسبی جهت شناسایی حساب‌های کاربری جعلی به‌دست آورد. این تعداد ویژگی و الگوریتم‌های ترکیبی مورد استفاده در پژوهش جاری یک کار جدید بوده است که در پژوهش‌های کلیدی گذشته کمتر مورد توجه قرار گرفته است.

مدل ارائه شده در این پژوهش می‌تواند توسط مدیران شبکه‌های اجتماعی و روی مجموعه داده‌های بزرگ‌تر و کامل‌تر مورد استفاده قرار گیرد و با استفاده از این مدل، حساب‌های کاربری جعلی را به‌صورت خودکار و با دقت بالایی شناسایی کنند و با جلوگیری از فعالیت حساب‌های کاربری جعلی، امنیت و محبوبیت شبکه‌های اجتماعی خود را افزایش دهند. همچنین پلیس‌های امنیت سایبری با استفاده از مدل ارائه شده در پژوهش جاری، می‌توانند به شناسایی خودکار حساب‌های کاربری جعلی شبکه‌های اجتماعی بپردازند و از این طریق متخلفان و کلاه‌برداران را شناسایی کنند.

شایان ذکر است که مجموعه داده مورد استفاده در پژوهش جاری، یک مجموعه داده عمومی از شبکهٔ اجتماعی X است. پژوهش‌های آینده می‌توانند ویژگی‌های حساب‌های کاربری استفاده شده در پژوهش جاری را روی حساب‌های کاربری جعلی و واقعی سایر شبکه‌های اجتماعی استخراج کنند و این روش را روی داده‌های شبکه‌های اجتماعی مختلف مورد ارزیابی قرار دهد. علاوه‌براین پژوهش‌های آینده می‌توانند به بهینه‌سازی سایر پارامترهای الگوریتم SVM نیز بپردازند. همچنین پیشنهاد می‌شود که در پژوهش‌های آینده با استفاده از الگوریتم‌های استخراج ویژگی، ویژگی‌هایی را که تأثیر کمی در طبقه‌بندی حساب‌های کاربری واقعی و جعلی دارند، پیدا و حذف شود.

فهرست منابع

رمضان‌زاده رستمی، رضا و کرباسی، سهیلا (۱۳۹۷). شناسایی حساب‌های جعلی در شبکهٔ اجتماعی توییتر با استفاده از رویکرد انتخاب ویژگی ترکیبی چندهدفه. *ششمین همایش ملی مدیران فناوری اطلاعات*، تهران.

Alhassun, A. S. & Rassam, M. A. (2022). A Combined Text-Based and Metadata-Based Deep-Learning Framework for the Detection of Spam Accounts on the Social Media Platform Twitter. *Processes*, 10(3). https://doi.org/10.3390/pr10030439

Bala Anand, M., Karthikeyan, N., Karthik, S., Varatharajan, R., Manogaran, G. & Sivaparthipan, C. B. (2019). An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter. *The Journal of Supercomputing*, 75(9), 6085-6105. https://doi.org/10.1007/s11227-019-02948-w

Ben Sassi, I. & Ben Yahia, S. (2021). Malicious accounts detection from online social networks: a systematic review of literature. *International Journal of General Systems*, 50(7), 741-814. https://doi.org/10.1080/03081079.2021.1976773

Bharti, K. K. & Pandey, S. (2021). Fake account detection in twitter using logistic regression with particle swarm optimization. *Soft Computing*, 25(16), 11333-11345. https://doi.org/10.1007/s00500-021-05930-y

Cea, C. (2021). *Dataset for supervised bot detection on Twitter.* https://doi.org/10.5281/zenodo.5574403

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems*, 80, 56-71. https://doi.org/https://doi.org/10.1016/j.dss.2015.09.003

Dixon, S.J. (2023). *Global integrated drought monitoring and prediction system (GIDMaPS) data sets.* Availavble: https://www.statista.com/statistics/303681/twitter-users-worldwide

Jabardi, M. & Hadi, A. (2021). Ontology Meter for Twitter Fake Accounts Detection. *International Journal of Intelligent Engineering and Systems*, 14, 410-419. https://doi.org/10.22266/ijies2021.0228.38

Jia, J., Wang, B. & Gong, N. Z. (2017, 26-29 June 2017). Random Walk Based Fake Account Detection in Online Social Networks. *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN).*

Khaled, S., El-Tazi, N. & Mokhtar, H. M. O. (2018, 10-13 Dec. 2018). Detecting Fake Accounts on Social Media. *2018 IEEE International Conference on Big Data (Big Data)*.

Kondeti, P., Yerramreddy, L. P., Pradhan, A. & Swain, G. (2021, 2021//). Fake Account Detection Using Machine Learning. *Evolutionary Computing and Mobile Sustainable Networks*, Singapore.

Kwak, H., Lee, C., Park, H. & Moon, S. (2010, April). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (pp. 591-600).

Meligy, A., Ibrahim, H. & Torky, M. (2017). Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks. *International Journal of Computer Network and Information Security*, 9, 31-39. https://doi.org/10.5815/ijcnis.2017.01.04

Mujeeb, S. & Gupta, S. (2022, 2022//). Fake Account Detection in Social Media Using Big Data Analytics. *Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems*, *Singapore*.

Naga Praveena, T.A. M., Singh, R., Manogna, S. & Pragna, B. (2024). Fake profile identification in social network using machine learning and NLP. *Journal of Engineering Sciences*, 15(01), 1-11.

Pakaya, F. N., Ibrohim, M. O. & Budi, I. (2019, 16-17 Oct. 2019). Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning. *2019 Fourth International Conference on Informatics and Computing (ICIC).*

Prabhu Kavin, B., Karki, S., Hemalatha, S., Singh, D., Vijayalakshmi, R., Thangamani, M., Haleem, S. L. A., Jose, D., Tirth, V., Kshirsagar, P. R. & Adigo, A. G. (2022). Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks. *Wireless Communications and Mobile Computing*, 2022, 6356152. https://doi.org/10.1155/2022/6356152

Revathi, S. & Suriakala, D. M. (2018, 20-22 Dec. 2018). Profile Similarity Communication Matching Approaches for Detection of Duplicate Profiles in Online Social Network. *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS).*

Sahoo, S. R. & Gupta, B. B. (2019). Hybrid approach for detection of malicious profiles in twitter. *Computers & Electrical Engineering*, 76, 65-81. https://doi.org/10.1016/j.compeleceng.2019.03.003

Sai Raja, E. V., Aditya, B. L. & Mohanty, S. N. (2023). Fake Profile Detection Using Logistic Regression and Gradient Descent Algorithm on Online Social Networks. EAI Endorsed Transactions on Scalable Information Systems, 11(1). https://doi.org/10.4108/eetsis.4342

Shen, X., Lv, W., Qiu, J., Kaur, A., Xiao, F. & Xia, F. (2023). Trust-Aware Detection of Malicious Users in Dating Social Networks. *IEEE Transactions on Computational Social Systems*, 10(5), 2587-2598. https://doi.org/10.1109/TCSS.2022.3174011

Tang, Y., Zhang, D., Liang, W., Li, K. C., & Sukhija, N. (2021, December). Active malicious accounts detection with multimodal fusion machine learning algorithm. In *Inernational conference on ubiquitous security* (pp. 38-52). Singapore: Springer Singapore.

Wani, M. A., Agarwal, N., Jabin, S. & Hussain, S. Z. (2019, 7-11 Jan). Analyzing Real and Fake users in Facebook Network based on Emotions. *2019 11th International Conference on Communication Systems & Networks (COMSNETS).*

Detecting Fake Accounts with Machine Learning Techniques: A Case Study on the X Social Network (Formerly Twitter)

Mohammad Rajabian
MSc., Department of Information Technology, Faculty of Industry, K.N. Toosi Universityof Technology, Tehran, Iran

Monireh Hosseini [*1]
Associate Prof., Department of Information Technology, Faculty of Industry, K.N. Toosi Universityof Technology, Tehran, Iran

Abstract
The popularity of online social networks (OSNs) is increasing daily. This has led to malicious individuals exploiting these platforms. Often, these individuals use fake accounts to carry out their malicious actions. Many researchers have focused on identifying fake user accounts on various social networks using machine learning methods. Most previous studies have used feature reduction, selection, or extraction methods to reduce the number of features and increase the speed when using ensemble machine learning methods to accurately identify fake user accounts. In the current research, 67 numerical features were selected to identify fake user accounts, which ensures high classification speed and eliminates the need for feature reduction. Additionally, a new hybrid method has been proposed in which the Support Vector Machine (SVM) algorithm is optimized with Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Neural Networks (NN). In the current research, the best average accuracy achieved over one hundred runs of the algorithms was 98.28%, which occurred when combining the SVM algorithm with NN.

Keywords: Fake account, X social network, Machine learning, Support Vector Machines (SVM), Optimization.

1. Corresponding Author: hosseini@kntu.ac.ir