Journal of Electronical & Cyber Defence
Vol. 10, No. 1, Spring 2022, Serial No. 37

The enhancement of online social network security by detecting and preventing fake accounts through machine learning

S. Siadat*, V. Rahmaty, S. F. Noorani

*Assistant Professor, Computer Department, Tehran Payam Noor University, Tehran, Iran

(Received: 09/05/2021, Accepted: 03/10/2021)

ABSTRACT

Today, with the pervasiveness of social networks, the security of this environment is considered as one of the most important network issues. One of the security challenges is creating fake accounts that harass social media users. The owners of these fake accounts pursue goals such as creating likes and followers or distributing misinformation for political, cultural and economic purposes. In this study, with the aim of improving security in social networks and improving the security of cyberspace, a method for investigating and detecting fake accounts is presented. This method proposes an algorithm that combines the decision tree, the nearest neighbor and Bayes methods. The results of this combined algorithm demonstrate an accuracy of 95.34%. This method has stability and does not suffer from overfit, as proved in the conclusion. The results of this research can be used to provide solutions to prevent the creation of fake accounts and increase account security and lead to the recognition and use of new data mining techniques and also data analysis in social networks. One of the achievements of this research is the method of detecting the falsity of the account and identifying the factors affecting its detection, which has been done using a hybrid algorithm that obtains correct results.

Keywords: fake accounts, data mining, detection and prevention of fake accounts, cyber security, social networks.

* Corresponding Author Email: safieh.siadat@gmail.com

نشریه علمی "پدافند الکترونیکی و سایبری"
سال دهم، شماره ۱، بهار ۱۴۰۱، ص ۹۷-۸۵

علمی-پژوهشی

دنا: استفاده از اهداف شبکه اجتماعی و یادگیری ماشین به‌منظور تشخیص حساب‌های جعلی و بهبود امنیت شبکه‌های اجتماعی

وحید رحمتی[1*]، سیده صفیه سیادت[2]، سیده فاطمه نورانی[3]

۱- دانشجوی دکتری، ۲ و ۳- استادیار، گروه کامپیوتر و فناوری اطلاعات، دانشگاه پیام‌نور، تهران، ایران

(دریافت: ۱۴۰۰/۰۲/۲۹، پذیرش: ۱۴۰۰/۰۹/۲۲)

چکیده

امروزه همگام با فراگیر شدن شبکه‌های اجتماعی، امنیت این محیط یکی از مسائل مهم و پراهمیت تلقی می‌شود. یکی از چالش‌های امنیتی، ایجاد حساب‌های کاربری جعلی است که موجب آزار و اذیت کاربران شبکه‌های اجتماعی می‌شود. صاحبان این حساب‌های جعلی، اهدافی مانند ایجاد لایک و دنبال‌کننده و یا توزیع اطلاعات غلط با اهداف سیاسی، فرهنگی و اقتصادی را دنبال می‌کنند. در این پژوهش، با هدف بهبود امنیت در شبکه‌های اجتماعی و ارتقای امنیت فضای سایبری، روشی برای بررسی و تشخیص حساب‌های جعلی ارائه خواهد شد. روش پیشنهادی به نام «دنا»، از یک جهت از اهداف شبکه اجتماعی و از طرف دیگر از روش الگوریتم ترکیبی با درخت تصمیم، نزدیک‌ترین همسایه و بیز بهره خواهد گرفت. نتایج از اجرای روش پیشنهادی با الگوریتم ترکیبی، میزان صحت ۹۵/۳۴ درصد را نشان می‌دهد. پایداری و نداشتن overfit از دیگر ویژگی‌های روش پیشنهادی است که در قسمت نتایج اثبات شده است. نتایج این تحقیق می‌تواند در ارائه‌ی راهکارهای پیشگیری از ایجاد حساب‌های جعلی و افزایش امنیت آن به کار رود و منجر به شناخت و بهره‌گیری از تکنیک‌های جدید داده‌کاوی در شبکه‌های اجتماعی گردد و در حوزه‌ی تحلیل داده و داده‌کاوی در شبکه‌های اجتماعی مورد استفاده قرار گیرد.

کلیدواژه‌ها: حساب‌های جعلی، داده‌کاوی، تشخیص و پیشگیری از ایجاد حساب جعلی، امنیت سایبری، شبکه‌های اجتماعی

۱- مقدمه

نخستین بار مفهومی با عنوان شبکه‌های اجتماعی اینترنتی با قالب امروزی در سال ۱۹۶۰ در دانشگاه ایلی نویز در ایالت متحده‌ی آمریکا مطرح شد. سپس با رشد تکنولوژی و توسعه زیرساخت‌ها به‌صورت شبکه‌های امروزی تغییر چهره داد [۱]. در عصر حاضر و با پیشرفت تکنولوژی و دسترسی بیشتر افراد به شبکه‌های اجتماعی و تأثیری که این شبکه‌ها در زندگی روزمره‌ی افراد دارند، سبب گرایش افراد زیادی به این شبکه‌ها شده است تا بسیاری از کارهای روزمره‌ی خود را از این طریق انجام دهند. کارهایی از قبیل به اشتراک‌گذاری اطلاعات، برقراری ارتباط با یکدیگر، ارسال عکس و ویدئو و موارد مشابه بسیاری که هر روز نیز بر توسعه و تنوع آن افزوده می‌شود[۲]. شبکه‌های اجتماعی آنلاین[1] بستر بسیار خوبی را برای کاربران خود جهت تعامل، همکاری و به اشتراک گذاشتن محتوا در وب فراهم می‌نماید[۳]. جوامع مجازی به افراد در ایجاد ارتباطات آنلاین کمک می‌کند و امکان برقراری ارتباط آن‌ها را با دوستان و اعضای خانواده‌ی خود تسهیل می‌نماید[۴]. فیس‌بوک به‌عنوان یکی از شبکه‌های اجتماعی مقبول، بستری را در اختیار کاربران قرار می‌دهد تا بتوانند اطلاعات را با دوستان خود به اشتراک بگذارند[۵]. توییتر به اعضای خود پیشنهاد می‌دهد تا افکار، عقاید، و پیشنهادهای خود را انتشار دهند. برخی از سایت‌های شبکه‌های اجتماعی به‌طور خاص طراحی شده‌اند تا روشی آسان و راحت برای به اشتراک‌گذاری فیلم‌ها و عکس‌ها فزاهم کنند؛ مانند **یوتیوب** و **فلیکر** [۶]. علاوه‌براین، برنامه‌های شبکه‌ی آنلاین وجود دارد که در درجه‌ی اول برای رشد حرفه‌ای کاربران آن طراحی شده‌اند. **لینکدین** بزرگ‌ترین شبکه‌ی خبرگان آنلاین است که گزینه قدرتمندی برای ارتباط با افراد مرتبط در حوزه‌ی شغلی آن‌ها فراهم می‌کند [۷]. یک انجمن گفت‌وگو نیز یکی از انواع رسانه‌های اجتماعی است. کورا (Quora) که به کاربر اجازه می‌دهد تا سؤالات مربوط به موضوعی را از مردم بپرسد و به آن‌ها پاسخ دهد و دانش خود را به اشتراک بگذارد. برنامه‌های همسریابی آنلاین نوع دیگری از OSN است که ارتباط با غریبه‌ها را برای کاربران آسان می‌کند [۸]. شبکه‌های اجتماعی با هم متفاوت است؛ اما برخی کارکردهای شبکه‌ای در تمامی جوامع با هم مشترک است. مهم‌ترین کارکرد شبکه‌های اجتماعی ایجاد گروه‌ها و دسته‌های ارتباطی و انجمن[2] پیرامون ویژگی یا ویژگی‌های خاص است. همچنین کارکردهای اقتصادی، مبتنی بر بازاریابی اجتماعی نیز از دیگر کارکردهای این شبکه‌ها است. کارکرد دیگری که برای این شبکه‌ها متصور است کارکرد سیاسی است. ایجاد کمپین‌های سیاسی در یک فضای اجتماعی اینترنتی از کارکردهای شبکه‌های اجتماعی است[۳].

* رایانامه نویسنده مسئول: rhmtywhyd39@gmail.com

[1] Online Social Network (OSNs)

[2] Community

OSNs را می‌توان بر اساس زمینه‌ی قابلیت‌هایی که در اختیار اعضای آن قرار می‌دهد طبقه‌بندی نمود؛ مانند برنامه‌های کاربردی طراحی شده برای ساخت و حفظ ارتباطات اجتماعی، برنامه‌های تسهیل‌کننده‌ی اشتراک رسانه‌ها، و انجمن‌هایی که به کاربر اجازه می‌دهد دانش، اخبار و ایده‌های خود را به اشتراک بگذارد[1][10]. گسترش و توسعه‌ی روزافزون شبکه‌های اجتماعی با اهداف مختلف صورت می‌گیرد. ایجاد فضای آزاد اطلاع‌رسانی توسط این شبکه‌ها و ایجاد بسترهای مناسب برای کسب‌وکار الکترونیکی و کنترل بیماری‌های همه‌گیر مانند کرونا از موارد توسعه‌ی این شبکه‌ها است.

با وجود اینکه شبکه‌های اجتماعی زندگی مردم را بهتر کرده‌اند، اما چندین مسئله‌ی اجتناب‌ناپذیر و جدی در ارتباط با آن‌ها وجود دارد مانند سوءاستفاده از اطلاعات شخصی و حرفه‌ای، مهندسی اجتماعی[1]، قلدری آنلاین[2]، جعل هویت آنلاین و مقاصد سیاسی، فرهنگی و اقتصادی. وجود پروفایل‌های جعلی در شبکه یکی از مهم‌ترین نگرانی‌های ارائه‌دهندگان خدمات OSNs و کاربران آن است که با توسعه‌ی حوزه‌های کاربردی OSNsها به نگرانی امنیتی و اجتماعی نیز تبدیل شده است[۱۱]. در شبکه‌های اجتماعی، نامعلوم بودن هویت افراد باعث ایجاد انگیزه برای افراد سودجو شده است تا به دنبال راهکارهایی جهت اختلال و دستیابی به اطلاعات افراد در این شبکه‌ها باشند[۱۲]. ایجاد پروفایل‌های جعلی به نام افراد یا حملات فیشینگ و اسپم‌ها از موارد این سودجویی‌ها است. نمایه‌ی جعلی[3] نشان‌دهنده‌ی هویت افرادی است که ادعا می‌کنند شخصی هستند که نیستند. حساب‌های جعلی معمولاً برای انجام فعالیت‌های مختلف غیرقانونی، گمراه‌کننده، مخرب یا تبعیض‌آمیز در شبکه ایجاد می‌شود و این امر تهدیدی برای شبکه و همچنین کاربران آن است[۱۳]. تعداد فزاینده‌ای از هکرها در حال ایجاد هویت جعلی در شبکه‌هایی مانند فیس‌بوک و توییتر برای دسترسی به اطلاعات شخصی کاربران، تأیید یک مارک خاص یا شخص خاص، بدنام کردن کاربر و موارد مشابه هستند. دشمنان ممکن است سایت‌های حرفه‌ای مانند **لینکدین** را با هدف ردیابی فعالیت اعضا یا جلب اعتماد متخصصان تجارت، هدف قرار دهند. مهاجمان اغلب وب‌سایت‌های همسریابی را هدف قرار می‌دهند تا از افرادی که برای اهداف مالی، هدایا یا اطلاعات شخصی به دنبال همراه خود می‌گردند، بهره ببرند[۱] [۱۴]. بنابراین تشخیص انگیزه‌ی مهاجم، شناسایی حساب جعلی و یافتن راهکاری مؤثر برای مقابله با آن از چالش‌های فضای شبکه‌های اجتماعی است [۱۱].

[1] Social Engineering
[2] Bullying online
[3] Fake account

با اینکه تحقیقات گوناگونی در مورد شناسایی حساب‌های جعلی انجام پذیرفته است اما هریک از زاویه‌ای خاص به این موضوع پرداخته‌اند. برخی محققان مانند مدیسیتی، رهیت و اسلام از منظر یادگیری عمیق و شبکه‌های عصبی به تشخیص حساب جعلی پرداخته‌اند [۱۵] [۱۶] [۱۷]، بعضی محققان ایرانی مانند محمد رضایی، قادری و اقوم، حوزه‌ی یادگیری ماشین را در روش‌های شباهت کسینوسی و چگالی هسته، گراف شبکه و سایر طبقه‌بندی‌ها، برای تشخیص حساب‌های جعلی به کار برده‌اند، بعضی محققان هم اقدام به رده‌بندی جوامع مجازی نموده‌اند تا با بررسی اهداف تشکیل و کاربردهای شبکه، اقدام به شناسایی حساب‌های جعلی نمایند.

ضروری به نظر می‌رسد که با استفاده از روشی جامع، ضمن بررسی اهداف شبکه‌ی اجتماعی و یادگیری ماشین، به شناسایی حساب‌های مخرب و جعلی بپردازیم تا ضمن اینکه به دقت بالاتری در تشخیص حساب‌های جعلی دست می‌یابیم، به روش‌های مؤثر با قابلیت فراگیرتر خواهیم رسید که حساسیت کمتری به دیتاست داشته باشند؛ به این نحو که برای سایر دیتاست‌های یک شبکه‌ی اجتماعی دقت تشخیص مؤثری داشته باشند و ما در این تحقیق به دنبال ارائه‌ی روشی ترکیبی با این خصوصیات هستیم.

در این پژوهش، با هدف افزایش دقت در شناسایی حساب‌های جعلی، روشی جدید به نام «دنا» جهت شناسایی حساب‌های جعلی ارائه شده است. در این روش با استفاده از اهداف یک شبکه‌ی اجتماعی ، و بهره‌گیری از روش ترکیبی در یادگیری ماشین، روشی جهت افزایش دقت تشخیص حساب‌های جعلی ارائه شده است.

در ادامه‌ی مقاله، در بخش دوم تحقیقات پیشین مورد بررسی قرار گرفته است، روش پیشنهادی در بخش سوم معرفی می‌گردد و نتایج در بخش چهارم و در نهایت در بخش پنجم نتیجه‌گیری مقاله ارائه شده است.

۲- مرور تحقیقات پیشین

با بررسی تحقیقات گذشته، مشاهده شد که سه شیوه در تشخیص و مقابله با حساب‌های جعلی پیشنهاد و استفاده شده‌اند:

- ارائه دستورالعمل‌های امنیتی به کاربران
- تشخیص هدف شبکه‌ی اجتماعی
- بهره‌بردن از الگوریتم‌های یادگیری ماشین

۲-۱- ارائه‌ی دستورالعمل‌های امنیتی به کاربران

شبکه‌های اجتماعی خطرات خاص خود را دارند که ممکن است با رعایت نکردن نکات امنیتی برای کاربر دردسرساز شوند. هکرها می‌توانند از طریق رسانه‌های اجتماعی به اطلاعات شخصی افراد دست پیدا کنند و یا ویروس‌هایی را وارد سیستم کاربر نمایند و با دستیابی به اطلاعات شخصی آن‌ها، کاربران را مورد آزار و اذیت قرار دهند. این اتفاقات ممکن است حتی تا جایی پیش برود که شهرت و اعتبار و موقعیت شغلی و درنهایت زندگی افراد را نشانه برود و آن‌ها را نابود کند. برخی از دستورات امنیتی جهت مصون ماندن از این مخاطرات عبارت است از [۶]:

- احراز هویت دو عامل
- رمز عبور با مدیریت رمز عبور
- استفاده از ایمیل جداگانه برای جوامع مجازی مختلف
- استفاده از شماره‌تلفن در هنگام ایجاد حساب کاربری
- خصوصی‌سازی حساب
- تغییر گذرواژه به‌طور منظم
- بررسی ایمیل‌ها برای ورودهای جعلی
- کلیک نکردن بر روی URL های کوتاه
- بررسی و حذف برنامه‌های مشکوک بر روی سخت‌افزار
- توجه و استفاده از آدرس‌های با پیشوند HTTPS
- حذف حساب‌هایی که مدتی استفاده نشده است.
- به‌روزرسانی برنامه‌های سیستمی

۲-۲- تشخیص هدف شبکه اجتماعی

شبکه‌های اجتماعی آنلاین (**OSNs**) به برنامه‌های تحت وب گفته می‌شود که در درجه اول برای تسهیل، تعامل، همکاری و اشتراک مطالب میان کاربران طراحی شده‌اند. این برنامه‌ها بستری برای کاربران فراهم می‌کنند تا کاربران با علایق متفاوت را بتوانند جذب نمایند. با بررسی شبکه‌های اجتماعی مشخص می‌شود که تشکیل این شبکه‌ها با اهداف گوناگونی انجام می‌شود. وانی و همکاران [۱] هدف از تشکیل یک شبکه‌ی اجتماعی را در پنج گروه تقسیم می‌کند. این تقسیم‌بندی در جدول (۱) به‌صورت خلاصه ارائه گردیده است.

شایان ذکر است که این رده‌بندی جوامع مجازی با هدف دستیابی به تحلیل در ویژگی‌هایی است که حساب‌های مخرب به دلیل آن شکل‌گرفته‌اند. انتخاب این ویژگی‌ها در داده‌کاوی و تحلیل داده در افزایش دقت الگوریتم در یادگیری ماشین، نقش بسزایی دارد. باید به این نکته توجه نمود که هر روز بر قابلیت‌های شبکه‌های اجتماعی افزوده می‌شود و امکانات بیشتری را برای کاربران آن فراهم می‌نماید. همچنین کاربری‌های ذکر شده در جدول (۱) به‌عنوان اهداف تشکیل آن شبکه است و نافی کاربری‌های دیگر آن نیست.

جدول (۱): اهداف تشکیل شبکه‌های اجتماعی [1]

هدف	نمونه	ویژگی
شبکه‌های صرفاً اجتماعی آنلاین (POSNs)[1]	Facebook, Twitter, Friend Orkut, ester	ایجاد ارتباطات اجتماعی، به اشتراک گذاشتن اطلاعات
شبکه‌ی اجتماعی اجتماع خبرگان	LinkedIn, gate, esearch Academia, Opportunity	ایجاد ارتباطات حرفه‌ای، پیداکردن فرصت‌های شغلی، همکاران تحقیقاتی و غیره
بحث و گفت‌وگو انجمن‌ها و وبلاگ‌ها	Baidu Quora, Sky Tieba, rock	بحث و گفت‌وگو، پرسش‌وپاسخ
اشتراک رسانه	Instagram, YouTube, etc. Snapchat,	اشتراک عکس، ویدئو
دوست‌یابی	Badoo, match.com, Beautiful etc. People,	ایجاد روابط شخصی، مطابقت ایده‌آل و غیره.

برای مثال در فیس‌بوک هم فیلم و عکس به اشتراک گذاشته می‌شود. رده‌بندی جوامع مجازی هر روز توسعه می‌یابد و تفکیک این شبکه‌ها هر روز دشوارتر خواهد شد. در تحقیقات وانی و همکاران [۱] ضمن بیان این پنج گروه، انواع حساب‌های مخرب به این صورت لیست شده است:

- با هدف زورگویی[2]
- با هدف استخراج شماره ایمیل و تلفن برای انجام فعالیت‌های خرابکارانه.
- با هدف تهمت و بدنامی[3]
- با هدف کسب شهرت با افزودن تعداد مشترکان با آدرس ایمیل جعلی. با هدف انجام فعالیت‌های غیرقانونی در حوزه‌های فرهنگی و سیاسی.
- با هدف ارسال محتوای نامناسب.
- با هدف اقتصادی که دیگر کاربران باور کنند، سرمایه‌گذاری روی یک محصول خوب است؛ درصورتی‌که واقعیت ندارد.
- با هدف دسترسی به اطلاعات طبقه‌بندی شده.
- با هدف بازی با احساسات و باج‌خواهی، مانند کلاه‌برداری‌های عاشقانه.
- با هدف کسب درآمد از آگهی دزدیده‌شده از کاربران دیگر

[1] Pure OSNs
[2] Bullying
[3] Defamtion

با بررسی کارهای انجام شده در این گروه، به تحقیقات انجام شده در منابع [۱, ۲,۲۲] می‌توان اشاره کرد. در سه ردیف آخر جدول (۲)، برخی از فعالیت‌های تحقیقاتی انجام شده با استفاده از این روش نشان داده شده است. وانی و همکاران **[۱]** ضمن پرداختن به رده‌بندی جوامع مجازی علل شکل‌گیری حساب مخرب را به روش کیفی بررسی نموده‌اند. استب و همکاران **[۲]** ضمن رده‌بندی جوامع مجازی و بررسی این بستر برای تبلیغات و تجارت، با بررسی چگونگی پیوند پردازش اطلاعات با زمینه‌های اجتماعی، توپولوژی و هم‌بندی شبکه را تعریف می‌نماید. وگار و همکاران**[۲۲]** به بررسی جوامع مجازی و نقش آن در تبلیغات انتخاباتی پرداخته‌اند و از این منظر، اهداف جوامع مجازی به‌خصوص توییتر را مورد تحلیل قرار داده‌اند. مشخصات کارهای انجام شده در سه ردیف آخر جدول (۲) نیز مشخص شده است. در حوزه‌ی تحقیقات داخلی، آقای محمد رضایی و همکاران در روش پیشنهادی برای آموزش ماشین، از ویژگی‌های شباهت مختلفی مانند شباهت کسینوس، شباهت جاکارد، شباهت شبکه‌ی دوستی و معیارهای مرکزیت استفاده می‌شود که همه‌ی این ویژگی‌ها از گراف متناظر با شبکه‌ی اجتماعی استخراج می‌شود. سپس با استفاده از دسته‌بندهای تخمین چگالی هسته و الگوریتم شبکه‌ی عصبی خودسازمان‌ده داده‌ها دسته‌بندی می‌شود و روش پیشنهادی با دقت ۹۹/۶۴٪ اکانت‌های جعلی را تشخیص می‌دهد **[۱۱] [۲۱]**.

جدول (۲): تحقیقات پیشین

محققین/ سال	نحوه تشخیص حساب‌های جعلی	نتیجه تحقیق
کالیار و همکاران [۱۸] (۲۰۲۰)	یادگیری با شبکه‌ی عصبی	برای طبقه‌بندی اخبار جعلی با دقت ۹۸/۳۶٪ از طریق لایه‌های پنهان چندگانه ساخته‌شده در شبکه‌ی عصبی عمیق طراحی شده است
اسلام و همکاران[۱۷] (۲۰۲۰)	یادگیری عمیق	عمدتاً بر روی سه دسته‌ی گسترده از اطلاعات غلط متمرکز شده است: اطلاعات غلط، اخبار جعلی و تشخیص شایعات
التنویر [۲۰] (۲۰۱۹)	یادگیری ماشین و یادگیری عمیق	با توجه به الگوریتم یادگیری تقویتی اخبار جعلی در مجموعه‌داده‌های توییتر، مدلی را برای شناخت پیام‌های جعلی خبری، از پست‌های توییتر ارائه داده است
محمد رضایی و همکاران[۲۱] (۲۰۱۸)	یادگیری ماشین	مدل مبتنی بر شباهت بین شبکه‌های دوستان کاربران ارائه شده است. معیارهای شباهت مانند دوستان مشترک، کسینوس، Jaccard ، L1-size و شباهت وزن از ماتریس مجاورت نمودار مربوط در شبکه‌ی اجتماعی محاسبه شده است
ahmad Mudsir و wani همکاران[۱] (۲۰۱۷)	تشخیص هدف شبکه‌ی اجتماعی	بحث در مورد دسته‌بندی‌های مختلف شبکه‌های اجتماعی آنلاین با مزایای مربوط به آن‌ها برای کاربرانشان است. سپس، مقاله دلایل مختلفی را ایجاد می‌کند که مهاجم را برای ایجاد پروفایل‌های جعلی در یک نوع خاص از شبکه تحریک می‌کند
Vegeer و همکاران [۲۲] (۲۰۱۳)	تشخیص هدف شبکه‌ی اجتماعی	به تقسیم‌بندی شبکه‌های اجتماعی بر اساس ارائه‌ی خدمات به کاربران می‌پردازد
Staab و همکاران[۲] (۲۰۰۵)	تشخیص هدف شبکه‌ی اجتماعی	نوع و گسترش شبکه‌های اجتماعی بر اساس شکل کاربری از موارد پرداخته شده است

۲-۳- بهره بردن از الگوریتم‌های یادگیری ماشین

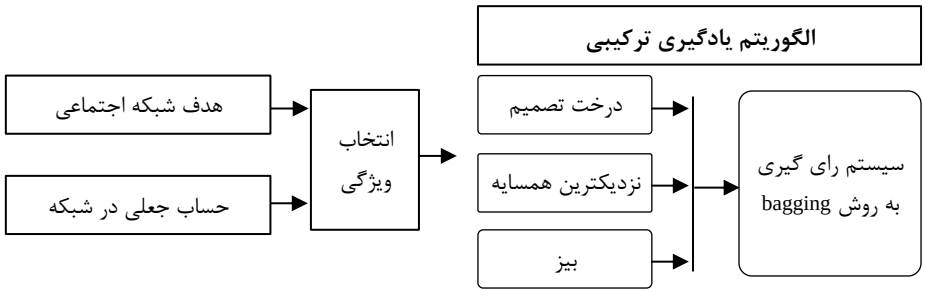
در سومین روش شناسایی و مقابله با حساب‌های جعلی، الگوریتم‌های یادگیری ماشین به خدمت گرفته می‌شوند. ردیف‌های اول تا چهارم جدول ۲، گزارشی از تحقیقات انجام شده در حوزه تشخیص حساب‌های جعلی و اهداف حساب‌های مخرب با استفاده از الگوریتم‌های یادگیری ماشین را نشان می‌دهد. ستون اول محققان و سال انجام تحقیق را توضیح می‌دهد. در ستون دوم ابزار تحلیل تحقیقات مشخص شده است و در ستون سوم نتایج تحقیقات تشریح می‌گردد.

۳- دنا: روشی برای شناسایی حساب جعلی با توجه به هدف شبکه‌ی اجتماعی و استفاده از روش ترکیبی یادگیری ماشین

پژوهشگران این مقاله، با بررسی تحقیقات پیشین، روشی که از مزایای توأم هدف شبکه‌ی اجتماعی و ابزار یادگیری ماشین استفاده نماید را مشاهده ننمودند؛ بنابراین هدف اصلی مقاله‌ی حاضر، افزایش دقت شناسایی حساب‌های جعلی از دو جهت است:

* جهت اول: بررسی شبکه‌ی اجتماعی و استخراج ویژگی‌های جدید با توجه به هدف شبکه‌ی اجتماعی در کنار سایر ویژگی‌ها

* جهت دوم: استفاده از یک روش ترکیبی از روش‌های درخت تصمیم، نزدیک‌ترین همسایگی و نیز بیز به‌منظور آموزش مدل و تشخیص مؤثرتر حساب جعلی.



شکل (۱): دنا: روشی برای تشخیص حساب جعلی (پیشنهادی در این مقاله)

شکل (۱)، مراحل مختلف روش پیشنهادی را نشان می‌دهد. روش پیشنهادی در این مقاله «دنا» نام‌گذاری شده است که از اختصار اول سه نام الگوریتم به‌کاررفته در الگوریتم ترکیبی حاصل یعنی درخت تصمیم، نزدیک‌ترین همسایه، ناوی بیز گرفته شده است. در ادامه جزئیات بیشتر دنا توضیح داده خواهد شد.

۱-۳- استفاده از هدف شبکه‌ی اجتماعی در مرحله‌ی انتخاب ویژگی

یکی از مراحل زمان‌بر در یادگیری ماشین، تجزیه و تحلیل داده‌ها، پیش‌پردازش و انتخاب ویژگی‌ها است. نظر به اهمیت هدف شکل‌گیری یک شبکه‌ی اجتماعی، در این مقاله به استخراج برخی ویژگی‌ها از درون هدف شکل‌گیری یک شبکه اجتماعی و استفاده از آن در آموزش مدل خواهیم پرداخت. ضمن اینکه در تعیین و تشخیص ویژگی‌های مؤثر انتخابی، دردست‌داشتن این اطلاعات، سبب انتخاب صحیح صفات و تشکیل صفات ترکیبی مؤثر در کوتاه‌ترین زمان می‌شود که مؤید اصل بهینه‌سازی در پیش‌پردازش داده‌ها است. به‌عنوان مثال در فیس‌بوک و اینستاگرام که جعل هویت افراد نیز از شاخص‌های موردتوجه در حساب جعلی است، «نسبت دنبال‌کننده به دنبال شونده» صفت ترکیبی مؤثری خواهد بود. یا به‌عنوان مثال دیگر، هدف ایجاد شبکه‌ی توییتر در حالت کلی ایجاد توییت است و در این فضا هدف از تشکیل یک حساب جعلی، پخش توییت‌های نادرست است. به‌عنوان مثال، صاحب یک حساب جعلی برای نشر توییت‌های نادرست، حجم زیادی از توییت‌ها را با هشتگ‌گذاری بیش از حد به دنبال‌کننده‌ها ارسال می‌کند. پس دو ویژگی «نسبت تعداد توییت در روز» و ویژگی «نسبت تعداد توییت به دنبال‌کننده» می‌تواند با توجه به هدف شبکه‌ی مذکور، در زمره‌ی ویژگی‌های مورداستفاده در آموزش مدل مورد استفاده قرار گیرد. همچنین نکته‌ای که وجود دارد این است که حساب‌های جعلی علاقه‌ی زیادی به جذب مخاطب بیشتر دارند و اقدام به دنبال‌کردن حساب‌های کاربری بیشتری می‌نمایند تا پیام‌های مخرب به آن‌ها ارسال کنند. از طرف دیگر تعداد کمتری حساب آن‌ها را دنبال می‌کنند. این حساب‌ها، همچنین توییت‌های زیادی ارسال می‌کنند و در متن از هشتگ و خطاب‌های بیشتری استفاده می‌کنند[۱]. از آنجا که اکثر این پیام‌ها حاوی لینک‌های مخرب هستند و معمولاً بعد از مدتی از طریق نرم‌افزارهای تشخیص حساب مخرب شناخته و حذف می‌شوند، عمر کوتاهی دارند. این صفات با استفاده از هدف شبکه‌ی اجتماعی و حساب جعلی آن به دست می‌آید که باعث افزایش دقت در مرحله‌ی پیش‌پردازش داده‌ها و انتخاب صفات شده است و منطقی است که پیرو آن دقت یادگیری ماشین افزایش خواهد یافت.

۳-۲- یک روش ترکیبی از روش‌های درخت تصمیم، نزدیک‌ترین همسایگی و بیز به‌منظور آموزش مدل و تشخیص حساب جعلی

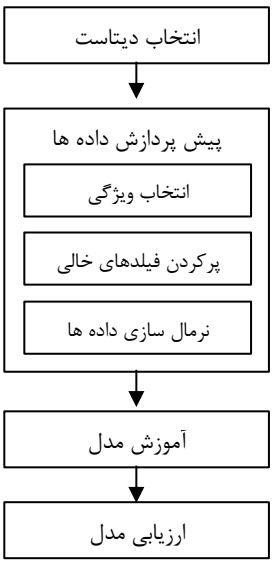
در این مقاله برای آموزش مدل و همچنین تشخیص حساب جعلی در مجموعه‌ی تست، از روش ترکیبی شامل سه روش درخت تصمیم، نزدیک‌ترین همسایگی و نیز بیز استفاده شده است. بدیهی است، انتخاب نوع الگوریتم با توجه به نیاز مسئله، حصول دقت بیشتری در آموزش مدل را به همراه خواهد داشت. از آنجا که در روش پیشنهادی در این مقاله، از هدف شبکه‌ی اجتماعی نیز استفاده می‌شود، به‌عنوان اولین الگوریتم، از درخت تصمیم استفاده می‌شود.

به‌طور مثال در پاسخ به سؤالاتی که مربوط به تعلق داشتن یا نداشتن یک حساب به «کلاس حساب جعلی» است، درخت تصمیم می‌تواند پاسخگو باشد. درخت تصمیم یک ساختار درختی است که در آن گره داخلی ویژگی را نشان می‌دهد. شاخه نشان‌دهنده‌ی یک قانون تصمیم‌گیری است و هر گره برگ نتیجه را نشان می‌دهد. زمان آموزش آن در مقایسه با الگوریتم شبکه‌ی عصبی سریع‌تر است. زمان پیچیدگی درختان تصمیم تابعی از تعداد سوابق و تعداد صفات است. نزدیک‌ترین همسایه (KNN)[1] یک الگوریتم یادگیری غیرپارامتری و تنبل است. یعنی برای مدل نیاز به هیچ الگوریتمی ندارد و با استفاده از داده‌ها و فاصله همسایگی اقدام به طبقه‌بندی می‌نماید. یکی از مشخصات حساب‌های جعلی در جوامع مجازی شباهت آن‌ها به یکدیگر است. به‌منظور کشف حساب‌های جعلی مشابه، الگوریتم نزدیک‌ترین همسایه به همین دلیل انتخاب می‌شود.

Bayes Naive یک تکنیک طبقه‌بندی آماری است که بر اساس قضیه‌ی Bayes ساخته شده است. این یکی از ساده‌ترین الگوریتم‌های یادگیری تحت نظارت است. طبقه‌بندی بیز الگوریتم سریع، دقیق و قابل‌اعتماد است. طبقه‌بندی‌کننده‌های بیز در مجموعه داده‌های بزرگ دقت و سرعت بالایی دارند. طبقه‌بندی بیز فرض می‌کند که تأثیر یک ویژگی خاص در یک کلاس مستقل از سایر ویژگی‌ها است. این فرض محاسبات را ساده می‌کند. اما اهداف جوامع مجازی هر روز گسترش می‌یابد و پیرو آن تنوع و گسترش حساب‌های جعلی را خواهیم داشت. این بدین معنی است که ما باید از Overfit یا بیش‌برازش مدل و الگوریتم بکاهیم و به همین دلیل از ترکیب الگوریتم‌های مذکور و در حالت رأی‌گیری استفاده نماییم. باید تعادل معقولی بین bias و variance یعنی بین سادگی زیاد طبقه‌بند و پیچیدگی زیاد آن برقرار شود تا یک الگوریتم طبقه‌بندی خوب داشته باشیم. با توجه به اهداف شبکه و موارد جعل حساب، استفاده از این الگوریتم ترکیبی همین مفهوم تعادل است و به همین نام نیز شهرت دارد (off Trade Variance and Bias). طبقه‌بندهای ترکیبی از ترکیبِ چندین طبقه‌بند[2] استفاده می‌کنند. در واقع این طبقه‌بندها، هرکدام مدل خود را بر روی داده‌ها ساخته و این مدل را ذخیره می‌کنند. مقادیر پیش‌بینی هر یک از الگوریتم‌ها که به‌صورت مجزا است به بخش رأی‌گیری وارد و با استفاده از تکنیک اکثریت آرا جواب معین می‌شود. به دلیل کاهش بایاس و واریانس از روش bagging استفاده نمودیم. به این صورت که یک زیرمجموعه از مجموعه‌ی داده‌ی اصلی به هر کدام از طبقه‌بندها داده می‌شود. یعنی هر طبقه‌بند یک قسمت از مجموعه‌ی داده (داده‌ی آموزشی) را مشاهده می‌کند و باید مدل خود را بر اساس همان قسمت از داده‌ها که در اختیارش قرار گرفته است، بسازد. در نهایت برای طبقه‌بندی نهایی یک رأی‌گیری در بین این طبقه‌بندها انجام می‌شود و آن طبقه‌ای که بیشترین میزان رأی را بیاورد، طبقه‌ی نهایی محسوب می‌شود. برای الگوریتم‌هایی مانندِ شبکه‌های عصبی یا درخت‌های تصمیم که با تغییرِ کمِ نمونه‌ها ممکن است طبقه‌های مختلفی ایجاد کنند، روش bagging (این الگوریتم‌ها به الگوریتم‌های غیرثابت[3] نیز شناخته می‌شوند) می‌تواند مفید باشد که در این مقاله استفاده شد.

۴- روش‌شناسی

در این مقاله، روشی جهت تشخیص حساب جعلی در شبکه‌ی اجتماعی ارائه شده است. به‌منظور امکان‌سنجی و ارزیابی روش «دنا» از داده‌های توییتر استفاده شده است. شکل ۲، روش ارزیابی مدل را نشان می‌دهد. برای این منظور پس از انتخاب دیتاست و پیش‌پردازش داده‌ها، با استفاده از روش یادگیری ترکیبی، مدل آموزش داده می‌شود و سپس نتایج به‌دست‌آمده ارزیابی خواهند شد.



شکل (۲): روش‌شناسی جهت ارزیابی روش پیشنهادی در این مقاله

[1] K-Nearest Neighbor

[2] Classifier
[3] Unstable

۴-۱- انتخاب دیتاست و پیش‌پردازش داده‌ها

دیتاست مورد استفاده از توییتر و با استفاده از ابزار API از توییتر در سال ۲۰۱۹ استخراج شده است [۲۳].

این مجموعه حاوی دودسته داده، واقعی و جعلی است. در مرحله‌ی اول: با استفاده از پایتون و توسعه‌دهنده[1] آن Jupiter که دارای کتابخانه‌های ارزشمندی است، داده‌ها را با استفاده از کتابخانه Pandas خواندیم که در این مرحله ۳۴ ویژگی مشاهده شد. در مرحله‌ی بعد، با استفاده از تابع describe و Coefficient Correlation Standard (corr_matrix) در پایتون تعداد صفات مؤثر را به ۹ ویژگی کاهش دادیم. این ویژگی‌ها در جدول ۳ با F1 تا F9 نشان داده شده است. جهت بررسی بیشتر صفات جدول (۲) از تابع scatter_matrix نیز استفاده می‌شود. صفاتی که حالت توزیع نرمال دارند، مناسب جهت یادگیری ماشین هستند. همچنین در تابع corr_matrix نیز، صفاتی که به عدد یک نزدیک هستند، صفات مؤثرتر و با رابطه‌ی مستقیم با صفت برچسب هستند و صفاتی که به عدد منفی یک نزدیک هستند، صفات با تأثیر بیشتر هستند اما با رابطه عکس.

جدول ۳: ویژگی‌های حساب کاربر توییتر

ویژگی	توضیح ویژگی
F1	تعداد وضعیت
F2	تعداد دنبال‌کنندگان
F3	تعداد دوستان
F4	تعداد علاقه‌ها
F5	تعداد لیست‌شده
F6	آدرس
F7	منطقه‌ی زمانی
F8	گزارش ها(retweet)
F9	GEO
F10	نسبت تعداد توییت به دنبال‌کننده
F11	نسبت هشتگ در توییت
F12	نسبت خطاب در توییت

در مرحله‌ی بعد، با استفاده از اهداف شبکه‌ی مجازی که در بخش قبل توضیح داده شد، صفات F10 تا F12 که به‌صورت نوآوری هستند به دست می‌آیند. F10 از تقسیم ستون حاوی صفت تعداد توییت بر ستون حاوی صفت تعداد دنبال‌کننده به دست می‌آید و این ویژگی معناداری است؛ زیرا حساب جعلی با تعداد دنبال‌کننده‌ی کمتر، تعداد توییت بسیار بیشتری ارسال می‌کند و این به دلیل انتشار خبر جعلی بیشتر است که از اهداف حساب مخرب است و به‌همین ترتیب صفات F11 و F12 که از تقسیم ستون صفات هشتگ و خطاب بر ستون حاوی تعداد توییت به دست می‌آید؛ زیرا حساب جعلی برای انتشار هر چه بیشتر خبر، از تعداد خطاب و هشتگ بالاتری در متن استفاده می‌نماید. این صفات به مجموعه‌ی ویژگی‌ها اضافه شدند که صفات بسیار بهتری نسبت به هریک از این ویژگی‌ها به‌تنهایی هستند. جهت حصول اطمینان از درستی ایجاد این صفات ترکیبی، آن‌ها را با توابع تحلیلی ذکر شده در پایتون، بررسی می‌نماییم، مشخصات به دست آمده، مناسب بودن این صفات جهت یادگیری ماشین را تأیید می‌کنند و نسبت به ۹ صفت ابتدای جدول (۳) از کارآمدی بالاتری برخوردار هستند.

همان‌طور که در شکل (۲) مشاهده می‌شود مراحل پیش‌پردازش داده شامل حذف مقادیر مفقود و استانداردسازی بر روی صفات متنی و عددی است. از کلاس sklearn تابع SimpleImputer را برای حذف values Miss به کار می‌بریم و جهت نرمال‌سازی از StandardScaler استفاده می‌نماییم. پس از پایان انتخاب ویژگی و آماده‌سازی داده‌ها به توضیح مرحله‌ی آموزش می‌پردازیم.

به‌دلیل اینکه مرحله‌ی پیش‌پردازش و آماده‌سازی داده‌ها وقت‌گیر است و جهت استفاده‌ی آسان‌تر از مجموعه‌های دیتاست‌های مختلف برای توسعه و آزمون الگوریتم، ما از داده‌ها با استفاده از تابع pipeline یک ذخیره از کلیه‌ی این مراحل تهیه می‌نماییم و در استفاده‌های بعدی با فراخوانی این تابع با چند خط برنامه‌نویسی کوتاه، قادر خواهیم بود کلیه‌ی این مراحل را با سرعت انجام دهیم و به این ترتیب امکان تست‌های دیگر با مجموعه‌داده‌های متفاوت و همچنین استفاده از الگوریتم‌های متفاوت فراهم می‌آید و امر توسعه‌ی تحقیقات نیز میسر خواهد شد.

۴-۲- تعیین الگوریتم

مراحل اجرایی جهت انتخاب الگوریتم ترکیبی دنا در چهار گام انجام می‌پذیرد

گام اول: الگوریتم با نظارت و یا بدون نظارت

برای تشخیص پروفایل جعلی در شبکه‌های اجتماعی آنلاین، ابتدا باید شبکه‌ی مورد نظر را از لحاظ کاربری و اهداف تشکیل و خدماتی که به کاربران ارائه می‌دهد مورد تجزیه و تحلیل کافی قرار داد تا مشخص گردد چه اهدافی در این شبکه وجود خواهد داشت که برای هکرها و صاحبان حساب‌های مخرب و جعلی می‌تواند مفید باشد، یعنی به‌نوعی با تشکیل جدول اهداف شبکه به جدول اهداف حساب جعلی در آن شبکه خواهیم رسید. مسئله‌ی دوم نگاه ک

[1] Integrated Development Enviroment

لی به داده‌های شبکه است و دریافت این مورد که مجموعه‌ی داده از نظر ویژگی‌ها و صفات، با کدامین روش و مدل داده‌کاوی هم‌خوانی بهتری دارد و با کدام الگوریتم به خواسته‌ی خود بیشتر دست می‌یابیم. به همین دلیل است که گام اول باید درست طی شود و اهداف شبکه و حساب مخرب در آن مشخص گردد تا با نگرش درست روش داده‌کاوی مؤثر انتخاب شود.

شبکه‌ی اجتماعی انتخابی در این تحقیق توییتر است و همان‌طور که در بخش قبل توضیح داده شد، هدف اصلی حساب جعلی در آن پخش و توزیع اخبار جعلی است که برای انجام این منظور، حساب جعلی باید تعداد زیادی توییت به افراد زیاد ارسال نماید. پس ما در اینجا صفت برچسب داریم و آن جعلی و واقعی بودن کاربر است و گرایش ما به سمت الگوریتم‌های با نظارت و طبقه‌بندی‌کننده است.

گام دوم : تعیین الگوریتم بر اساس زمان و نوع یادگیری:

همچنین در این تحقیق، ما از مجموعه‌داده‌ی مشخص که از توییتر با روش API استخراج شده است استفاده می نماییم و به همین دلیل روش‌های غیرافزایشی و Batch انتخاب بهتری هستند؛ زیرا روش‌های Online زمانی مؤثرتر خواهند بود که الگوریتم به‌صورت افزایشی ورودی داده داشته باشد و به این ترتیب هرچه زمان می‌گذرد، الگوریتم بهتر و بر دقت آن نیز افزوده خواهد شد. پس استفاده از یادگیری عمیق با شبکه‌های عصبی در این حالت، منطقی به نظر نمی‌رسد؛ زیرا توسعه‌ای در ورودی مجموعه‌داده وجود ندارد و مجموعه‌داده ثابت است. این مورد نیز باید در نظر گرفته شود که مجموعه دارای برچسب کلاس است و اگر هدف تشخیص خبر جعلی بود باید از معیارهای مشابهت و تعیین توابع مناسب و روش ماشین بردار پشتیبان استفاده می‌گردید تا الگوریتم الگوهای خبر جعلی را تشخیص دهد. استفاده نکردن از روش ماشین بردار همچنین شامل دلایل زیر نیز است:

• این نوع الگوریتم‌ها، محدودیت‌های ذاتی دارند مثلاً هنوز مشخص نشده است که به ازای یک تابع نگاشت ، پارامترها را چگونه باید تعیین کرد.

•ماشین‌های مبتنی بر بردار پشتیبان به محاسبات پیچیده و زمان‌بر نیاز دارند و به دلیل پیچیدگی محاسباتی، حافظه‌ی زیادی نیز مصرف می‌کنند.

•داده‌های گسسته و غیرعددی هم با این روش سازگار نیستند و باید تبدیل شوند.

همچنین نباید جهت افزایش دقت، ایجاد حالت overfit در مدل بشود. تا اینجا الگوریتم ما از طبقه‌بندها و با نظارت و Batch انتخاب می‌شود.

گام سوم: تعیین الگوریتم از نظر -Base Instance بودن یا -Base Model بودن:

با نگاهی به مجموعه‌ویژگی‌ها در مجموعه داده و اهداف حساب جعلی هم‌زمان که روش داده‌کاوی معین می‌شود معلوم می‌گردد که شیوه‌ی انتخابی Base Model است و جعلی و واقعی بودن حساب با مدل تعیین می‌گردد. همچنین جهت جلوگیری از overfit شدن مدل، که به نوعی تأیید وابستگی نداشتن بیش از حد مدل به مجموعه‌داده است به الگوریتم ترکیبی می‌رسیم.

گام چهارم: کدام ترکیب جهت الگوریتم مؤثرتر است؟

الف- به دلیل اینکه تفکیک بین دو خصوصیت جعلی و واقعی انجام می‌شود؛ بنابراین با استفاده از امتیاز Gini وزن‌دار (score Gini weighted) مربوط به هر گره از آن تفکیک، Gini index را محاسبه می‌کنیم. الگوریتم درخت طبقه‌بندی و رگرسیون (CART) با استفاده از متد Gini ، تفکیک دودویی را تولید می‌کند. ما از درخت تصمیم طبقه‌بندی به‌عنوان الگوریتم طبقه‌بند با نظارت استفاده می‌کنیم. جنگل تصادفی از تعدادی درخت تصمیم تشکیل شده است که با توجه به تعدد صفات و بزرگی مجموعه‌های داده در OSNs می‌توانند الگوریتم را برای پیش‌بینی‌های جهان واقعی کند و غیرمؤثر کنند.

ب- با توجه به نیاز ما به الگوریتم با خاصیت تشخیص مشابهت بین حساب‌های جعلی، از KNN استفاده می‌کنیم.

در ابتدا با کمک روش اقلیدسی [1]، فاصله‌ی بین داده‌ی تست و هر سطر از داده‌ی آموزشی را اندازه‌گیری می‌کنیم، سپس بر اساس مقدار فاصله، آن‌ها را به‌صورت صعودی مرتب می‌نماییم، آنگاه الگوریتم k سطر بالاتر از آرایه‌ی مرتب شده را انتخاب می‌کنیم و در آخر بر اساس متداول‌ترین کلاس از این سطرها، الگوریتم، یک کلاس به نقطه‌ی تست تخصیص می‌دهد. در این مرحله به پایان عملکرد الگوریتم می‌رسیم. مقدار K در این پژوهش عدد هشت مشخص گردید.

ج – نکته‌ی دیگر در ویژگی‌ها میزان تأثیر آن‌ها در تشخیص حساب جعلی است؛ یعنی ارتباط هر ویژگی با صفت برچسب به چه نحوی است و با توجه به اینکه ویژگی‌های مؤثر (به‌صورت منفرد و یا ترکیب چند ویژگی) در اهداف جوامع مجازی و پیش‌پردازش داده‌ها انجام شده است، بیز ساده قابلیت این شناسایی را دارا است و نیاز به شبکه‌ی باور بیزین [2] است و در نظر گرفتن این نکته در طراحی روش ضروری است که افزایش دقت باید با کمترین افت سرعت همراه باشد.

[1] Euclidean
[2] Bayesian Belief Network

به این ترتیب الگوریتم‌های روش دنا مشخص می‌شود توضیحات ویژگی‌های ذکر شده‌ی الگوریتم‌های مطرح شده در متن، مؤید مشخصات مورد نیاز ما در تعیین الگوریتم داده‌کاوی است. به نحوی که الگوریتم دنا منطبق با نیاز داده‌کاوی در تعیین حساب جعلی است و به ایجاد الگوریتم با دقت بالا و با قابلیت تعمیم به مجموعه‌داده‌ی مختلف و ایجاد نشدن overfit در مدل دست یافته است. همان‌طور که ذکر شد، رأی‌گیری [1] ابتدا سه مدل مستقل از مجموعه‌داده‌ی آموزشی می‌سازد، و سپس یک طبقه‌بندی‌کننده‌ی رأی‌گیر [2]، مدل را به همراه میانگین پیش‌بینی‌های زیر مدل [3]، در هر زمان که به داده‌ی جدید نیاز باشد، بسته‌بندی می‌کند. Kfold عدد ده در نظر گرفته شده است. در این راهکار در پایتون، با استفاده از کلاس VotingClassifier از sklearn، مدل گروهی رأی‌گیری را برای طبقه‌بندی، روی مجموعه داده توییتر، خواهیم ساخت. که در مقایسه با روش‌های تحقیقات اخیر بهتر توانسته است اهداف خود را محقق نماید.

۴-۳- آموزش مدل (یادگیری ترکیبی)

در حوزه‌ی یادگیری ماشین برای یادگیری نظارتی روش‌های متنوعی ارائه شده است که دارای مزایا و معایب خاص خود هستند؛ بنابراین به جهت رفع مشکل بایاس و واریانس در خروجی‌های الگوریتم‌های طبقه‌بندی و بهبود عملکرد آن‌ها از مدل های یادگیری ترکیبی استفاده می‌شود. برای تشخیص حساب جعلی در چارچوب داده‌کاوی از ترکیب طبقه‌بندی‌ها در حالت مدل رأی‌گیری استفاده شده است. در این روش مجموعه‌داده به ۲۰٪ تست و ۸۰٪ داده جهت آموزش با استفاده از train_test_split از کلاس sklearn تقسیم می‌شود که در آن داده‌ی آموزشی به‌صورت جداگانه به الگوریتم درخت تصمیم و نزدیک‌ترین همسایه و بیز داده می‌شود تا عمل یادگیری انجام شود. سپس داده‌های آزمایشی جهت پیش‌بینی نمونه‌ها به الگوریتم‌ها داده می‌شود. جهت ارزیابی مدل از ماتریس درهم‌ریختگی [4] برای دست‌یافتن به تصویری جامع‌تر در ارزیابی عملکرد مدل استفاده می‌شود. این ماتریس، در شکل ۳ نشان داده شده است.

		دسته‌ی پیش‌بینی‌شده	
		+	-
دسته‌ی واقعی	+	**TP** درست مثبت	**FN** منفی نادرست
	-	**FP** مثبت نادرست	**TN** درست منفی

شکل (۳): ماتریس درهم‌ریختگی

جدول (۴): معیارهای ارزیابی[5] عملکرد یک مدل در داده‌کاوی

معیار	فرمول
صحت	$\frac{TP+TN}{TP+TN+FN+FP}$
دقت	$\frac{TP}{TP+FP}$
فراخوانی	$\frac{TP+TN}{TP+FN}$
ویژگی	$\frac{TN}{TN+FP}$
score F1	$\frac{2TP}{2TP+FN+FP}$

در داده‌کاوی به‌منظور ارزیابی عملکرد یک مدل داده‌کاوی، از معیارهای صحت[6]، دقت[7]، فراخوانی[8]، ویژگی[9] و F1 Score استفاده می‌شود. این معیارها و نحوه‌ی اندازه‌گیری آن‌ها در جدول ۴ نشان داده شده است.

معیار صحت در واقع میزان تعداد پیش‌بینی‌های درست در هر دو طبقه منفی و مثبت را نسبت به کل مجموعه‌ها نشان می‌دهد و برای مقایسه‌ی کلی الگوریتم‌ها استفاده شده است. معیار فراخوانی، بیان‌کننده نسبت تعداد داده‌های درست دسته‌بندی شده در یک کلاس خاص، به تعداد کل داده‌هایی است که باید در همان کلاس خاص دسته‌بندی شوند. مقدار بالا برای معیار فراخوانی، بیانگر تعداد کم داده‌هایی است که به‌اشتباه، در آن کلاس خاص دسته‌بندی نشده‌اند.

معیار دقت، نسبت تعداد پیش‌بینی‌های صحیح انجام شده برای نمونه‌های یک کلاس خاص، به تعداد کل پیش‌بینی‌ها برای نمونه‌های همان کلاس خاص را ارزیابی می‌کند. مقدار بالا برای معیار دقت، بیانگر تعداد کم داده‌هایی است که به‌اشتباه، در کلاس خاص دسته‌بندی شده‌اند.

معیار امتیاز F1، پارامترهای دقت و فراخوانی را با هم ترکیب می‌کند تا مشخص شود یک مدل دسته‌بند تا چه حد عملکرد خوبی از خود نشان می‌دهد.

۵- نتایج

به‌منظور ارزیابی نتایج، الگوریتم‌ها به‌تنهایی و در حالت ترکیبی با یکدیگر مقایسه می‌شود و در نهایت به بررسی صحت در انواع مدل‌ها می‌پردازیم.

جدول (۵)، میزان دقت، فراخوانی و ویژگی را در مورد اجرای سه الگوریتم نزدیک‌ترین همسایه، بیز و درخت تصمیم بر روی داده‌ها نشان می‌دهد.

[1] Voting
[2] Voting Classifier
[3] Sub-Model
[4] Confusion matrix
[5] Evaluation Metrics
[6] Accuracy
[7] Precision
[8] Recall
[9] Specificity

جدول ۵: ارزیابی عملکرد سه مدل به‌صورت جداگانه

نوع مدل	دقت	فراخوانی	score F1
نزدیک‌ترین همسایه	۰/۸۶	۰/۸۳	۰/۸۴
بیز	۰/۶۸	۰/۵۷	۰/۶۲
درخت تصمیم	۰/۹۲	۰/۸۷	۰/۸۹

در مرحله‌ی بعد، صحت هر کدام از الگوریتم‌ها به‌تنهایی و به‌صورت ترکیبی محاسبه گردید. نتیجه‌ی محاسبه در جدول (۶) نشان داده شده است.

جدول ۶ : مقایسه میزان صحت الگوریتم‌های درخت تصمیم، نزدیک‌ترین همسایه، بیز و دنا (روش پیشنهادی این مقاله)

الگوریتم	صحت
درخت تصمیم	٪۹۱
نزدیک‌ترین همسایه	٪۸۵
بیز	٪۶۵
الگوریتم ترکیبی	٪۹۵,۳۴

همان‌طور که جدول ۶ نشان می‌دهد، روش الگوریتم ترکیبی دارای عملکرد بالاتر و بهتری برابر با ٪۵/۹۳۴ نسبت به سایر الگوریتم‌ها است و پس از آن درخت تصمیم با ٪۹۱ قرار دارد. نتایج فوق تأییدکننده دقت روش پیشنهادی در این مقاله است

در ادامه، ابتدا مقایسه‌ای از نتایج مقاله حاضر و تحقیقات اخیر ارائه خواهیم نمود و سپس میزان پایداری و صحت روش پیشنهادی را به‌وسیله‌ی اجرای دنا بر روی دیتاست‌های دیگر مورد ارزیابی قرار خواهیم داد. آقای محمد رضایی و همکاران به ارائه‌ی روش تشخیص حساب جعلی با استفاده از ویژگی‌های شباهت مختلفی مانند شباهت کسینوس، شباهت جاکارد، شباهت شبکه‌ی دوستی و معیارهای مرکزیت می‌پردازند که همه‌ی این ویژگی‌ها از گراف متناظر با شبکه‌ی اجتماعی استخراج می‌شوند. سپس با استفاده از دسته‌بندهای تخمین چگالی هسته و الگوریتم شبکه‌ی عصبی خودسازمان‌ده داده‌ها دسته‌بندی می‌شوند. دقت این روش ٪۹۹/۶۴ بیان شده است[۱۱]. همچنین آقای محمد رضایی و همکاران بیان داشتند: برای کشف حساب‌های جعلی در شبکه‌های اجتماعی ، مدل جدیدی مبتنی بر شباهت بین شبکه‌های دوستان کاربران ارائه شده است. معیارهای شباهت مانند دوستان مشترک، کسینوس، Jaccard ، اندازه L1 و شباهت وزن از ماتریس مجاورت نمودار مربوط شبکه‌ی اجتماعی محاسبه شد. برای ارزیابی مدل پیشنهادی، تمام مراحل در مجموعه‌داده‌های تویتر اجرا شد. مشخص شد که الگوریتم SVM Gaussian Medium حساب‌های جعلی را با مساحت بالا در زیر منحنی ۱ و نرخ مثبت کاذب کم ۰/۰۲ پیش بینی می‌کند[۲۱]. اختلاف دقت این روش با دنا به دلایل زیر است:

این دقت مرتبط با مجموعه‌داده‌ی خاص استفاده‌شده در تحقیق مذکور است. هر شبکه‌ی اجتماعی مانند فیس‌بوک را می‌توان با یک گراف مدل‌سازی کرد. کاربران رأس‌های این گراف هستند و هر رابطه‌ی دوستی نیز یک یال است که دو رأس را به یکدیگر متصل می‌کند. البته در برخی شبکه‌ها ارتباطات به‌صورت دوطرفه نیست. برای مثال اینستاگرام مبتنی بر دنبال کردن است که در این موارد باید از یال‌های جهت‌دار استفاده کرد. گراف یک شبکه‌ی اجتماعی واقعی ممکن است بسیار بزرگ باشد. برای نمونه گراف شبکه‌ی فیس‌بوک شامل بیش از یک میلیارد رأس (متناظر با هر کاربر) است که مسلماً تعداد بسیار زیادی یال (متناظر با هر رابطه‌ی دوستی) در آن موجود است. می‌توان گراف‌های موجود را تحلیل کرد و به نتایج ارزشمندی در این زمینه رسید. در تحقیق آقای محمد رضایی این مدل بسیار مناسب، جهت مجموعه‌داده‌ی ایشان است. به‌طور کلی جهت تشخیص ناهنجاری‌های ساختاری و رفتاری استفاده از این مدل‌ها با معیارهای شباهت مذکور مناسب و دارای دقت بالا است.

اسلام و همکاران در مورد اطلاعات غلط، اخبار جعلی و تشخیص شایعات جهت پردازش خودکار داده‌ها و تطبیق اطلاعات با اطلاعات درست و تشخیص محتوای جعلی از یادگیری عمیق و شبکه‌های عصبی استفاده نموده‌اند[۱۷]. به دلیل نداشتن صفت کلاس، بهترین روش، الگوریتم‌های بدون نظارت است و الگوریتم باید به مرور با آموزش آنلاین به‌دقت بالای خویش برسد و به نظر مؤثرترین روش در متن‌کاوی است. اما در این مقاله با بررسی صفات و داشتن صفت برچسب حساب جعلی تشخیص داده می‌شود که مدل محور و با الگوریتم‌های با نظارت است؛ بنابراین می‌توان از ترکیب این دو روش به روشی جامع‌تر جهت تشخیص حساب و اخبار جعلی رسید که می‌تواند محور پژوهش‌های آینده قرار گیرد. حبیب ایزدخواه و همکاران با ایده‌ی کشف و پیشگیری از اخبار جعلی اقدام به خوشه‌بندی کاربران جعلی نموده‌اند و در مجموعه‌داده‌ی خویش به دقت ٪۹۷ دست یافته‌اند که دارای مجموعه‌داده‌ای متفاوت با این تحقیق است. اما با توجه به تعیین میزان مشابهت در الگوریتم خوشه‌بندی به نظر می‌رسد این تأییدی مضاعف بر وجود الگوریتم KNN در الگوریتم دنا باشد.

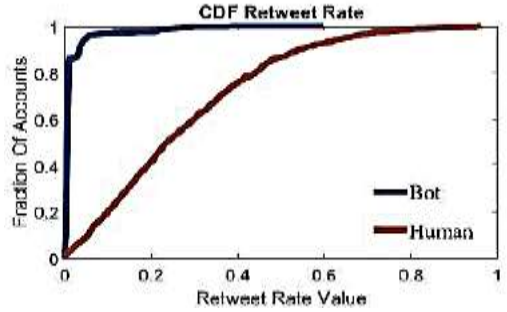
در مقاله‌ی [۲۲]، الگوریتمی جهت تشخیص حساب جعلی با استفاده از شبکه‌ی عصبی ارائه شده است و با همان دیتاست مورد استفاده در دنا میزان صحت ٪۹۸ گزارش شده است. با

بررسی دقیق‌تر الگوریتم مقاله [۲۲] دریافتیم که ویژگی‌های استفاده شده در آموزش مدل تعداد «وضعیت»، «تعداد دنبال‌کنندگان»، «تعداد دوستان»، «تعداد علائق»، «تعداد لیست‌شده‌ها»، «آدرس»، و «منطقه‌ی زمانی» هستند که همان ویژگی‌های F1 تا F7 در جدول ویژگی شماره(۳) در روش پیشنهادی دنا هستند.

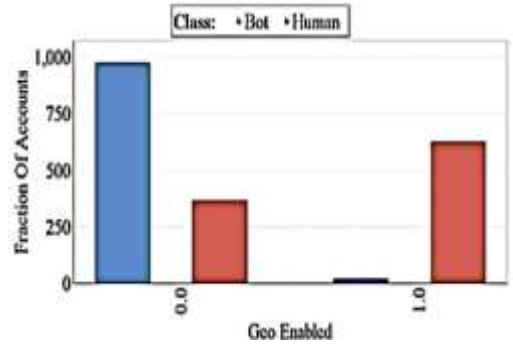
در حقیقت دو ویژگی retweet و GEO و همچنین سه ویژگی مربوط به هدف شبکه‌ی اجتماعی یعنی نسبت «تعداد توییت به دنبال‌کننده»، «نسبت هشتگ در توییت»، و «نسبت خطاب در توییت» در مقاله‌ی [۲۳] در نظر گرفته نشده است. به منظور مقایسه صحت، بار دیگر الگوریتم دنا را با هفت ویژگی استفاده‌شده در مقاله‌ی [۲۳] اجرا نمودیم و مشاهده شد که دنا (الگوریتم پیشنهادی در این مقاله) نیز با این هفت ویژگی دارای صحت ٪۹۸ در تشخیص حساب‌های جعلی است.

در کنار سه ویژگی انتخابی با توجه به هدف شبکه‌ی اجتماعی، مقادیر دو ویژگی retweet و GEO دارای اهمیت هستند و می‌توانند در تمایز بین رفتار حساب جعلی و حساب واقعی استفاده شوند. نکته مهم این است که حساب‌های جعلی به‌اندازه‌ی کافی هوشمند نیستند و نمی‌توانند رفتار تولید محتوا را عیناً و تماماً همانند کاربران انسانی شبیه‌سازی کنند. این نوع حساب‌ها، غالباً توییت‌های ارسالی کاربران دیگر را بازنویسی و ارسال می‌کنند و یا با استفاده از روش‌های احتمالی مانند الگوریتم زنجیره‌ای مارکوف، توئیت‌هایی ایجاد می‌کنند و یا از پایگاه داده‌های خاص برای ارسال محتوا استفاده می‌کنند. نویسندگان مقاله [۲۵] با استفاده از شکل ۳، نشان می‌دهند که تابع توزیع تجمعی retweet در حساب‌های جعلی، مقادیری بیشتری نسبت به حساب‌های واقعی و انسانی دارد. بنابراین همان‌طور که در روش دنا نیز انجام شده، وجود این ویژگی در مرحله‌ی آموزش مدل ضروری و دارای اهمیت است.

ویژگی F9 یا Tag GEO ویژگی دیگر است که در روش پیشنهادی این مقاله استفاده شده است و نقطه‌ی تمایز از مقاله‌ی [۲۳] است. این ویژگی مشخص می‌کند که این توییت از کدام منطقه ارسال شده است. همان طور که نمودار هیستوگرامِ شکل ۴ نشان می‌دهد حدود ٪۹۸ حساب‌های جعلی این ویژگی غیرفعال است، درحالی‌که در ٪۶۳/۷ حساب‌های واقعی و کاربران انسانی این ویژگی را فعال می‌کنند. بنابراین، همان‌طور که در روش پیشنهادی در این مقاله انجام شده، این صفت نیز باید در جمع ویژگی‌های تشخیص حساب جعلی منظور شود.



شکل ۳: عملکرد توزیع تجمعی ویژگی Rate Retweet در انسان و حساب‌های جعلی [۲۵]



شکل ۴: هیستوگرام فعال و غیر فعال بودن ویژگی Tag GEO در حساب‌های جعلی و انسانی [۲۵]

همان‌طور که در شکل ۳ و ۴ ملاحظه می‌شود دو داده‌ی مؤثر در این تحقیق نسبت به مقاله‌ی [۲۳] اضافه شده است. در قدم بعدی، صحت الگوریتم مقاله‌ی [۲۳] با استفاده از ویژگی‌های مورد استفاده در جدول ۳ که در دنا استفاده شده است، مورد بررسی قرار گرفت. با استفاده از ۱۲ ویژگی ذکر شده در جدول ۳، الگوریتم نویسندگان [۲۳] صحت ٪۹۳/۴۲ را نشان داد که کمتر از صحت روش دنا با مقدار ٪۹۵/۳۴ است.

به منظور بررسی پایداری الگوریتم و عدم overfit روش دنا، یک بار دیگر الگوریتم با دیتاست [۲۴] اجرا گردید. این دیتاست را چائوچان و همکاران با ابزار API و در سال ۲۰۱۲ از توییتر استخراج کرده‌اند.

همان‌طور که در بخش روش‌شناسی بیان شد، از آنجا که از تابع pipeline برای ذخیره‌ی کلیه‌ی مراحل استفاده شده است، با یک تغییر کوتاه در کدهای برنامه‌نویسی، امکان تست روش پیشنهادی در این مقاله با دیتاست‌های دیگر امکان‌پذیر است. با استفاده از دیتاست[۲۴] صحت ٪۹۴/۶۳ (نزدیک به همان نتیجه ٪۹۵/۳۴) حاصل گردید. با استفاده از همین دیتاست و الگوریتم مقاله‌ی [۲۳]، صحت حدود ٪۹۲ درصد حاصل گردید. این نتیجه مؤید این است که الگوریتم دنا با تأکید بر انتخاب ویژگی‌های مؤثر و کامل(یعنی تعداد بیشتری از ویژگی‌های مؤثر)، پایداری بیشتری دارد و حساسیت کمتری نسبت به بایاس و واریانس دارد.

بنـابراین روش دنـا (روش پیشـنهادی در ایـن مقالـه) بـا بهره‌گیری از اهداف شبکه‌ی اجتماعی و نیز ویژگـی‌هـای مـؤثر و کامل یک روش تشخیص حساب جعلی با دقـت و پایـدار را ارائـه می‌دهد.

۶- نتیجه‌گیری و پیشنهادها

در ایـن مقالـه بـه رده‌بنـدی و تقسـیم انـواع جوامـع مجـازی بـا انگیزه‌های شکل‌گیری حساب‌های مخرب پرداخته شـد و پـس از آن به روش ترکیبی دنا پرداخته شد که نتایج صحت آن بـالاتر از این الگوریتم‌ها به‌تنهایی است. حاصل ترکیـب دو روش فـوق بـه دست آوردن صفات و ویژگی صحیح جهت داده‌کاوی در شـبکه‌ی اجتماعی است و می‌تواند مبنایی جهـت سـایر تحقیقـات در ایـن حوزه قرار گیرد. این روش ترکیبی، در ایجاد الگوریتمی جـامع در بررسی مجموعه‌دیتاست‌های مختلف، با دقتـی قابـل‌قبـول مـورد ارزیابی قرار گرفت.

مدل دنا دارای شرایط زیر است:

- دارای مقاومت بهتر در برابر واریانس و بایاس.
- انجـام داده‌کـاوی مفهـومی‌تـر و اسـتفاده از اهـداف جوامـع مجازی در تعیین ویژگی.
- کاهش زمان مورد نیاز در تحلیل و پیش‌پردازش داده
- افزایش میزان صحت
- استفاده از تابع pipeline و امکـان اسـتفاده از دیتاسـت‌هـای دیگر در تست و توسعه‌ی الگوریتم

در نهایـت بـا مـدل ارائـه‌شـده مـی‌تـوان در تشـخیص حساب‌های جعلی موفق‌تر عمل نمود و با بالابردن امنیـت در فضای شبکه‌های مجازی برای کاربر و همچنین برای جامعـه، در ایجاد امنیت در فضـای سـایبری موفـق‌تـر عمـل نمـاییم، همچنین ایـن روش مـی‌توانـد بـه‌عنـوان راهکـاری مـؤثر در استفاده و توسعه در الگوریتم‌های ترکیبی در تحقیقات آینـده مورد استفاده قرار گیرد.

۷- مراجع

[1] Mudasir Ahmad Wani, Muzafar Ahmad Sofi, Suheel Yousuf Wani., "Why Fake Profiles: A Study of Anomalous Users in Different", categories of Online Social NetworksInternational Journal of Engineering Technology Science and Research, ISSN 2394 – 3386Volume 4, Issue 9 September 2017.

[2] Staab, S., Domingos, P., Mike, P., Golbeck, J., Ding, L., Finin, T., ... & Vallacher, R. R. "Social Networks Applied", IEEE Intelligent systems, 20(1), 80-93, 2005.

[3] Ed Grabianowski, "How Online Dating Works", http://people.howstuffworks.com/online-dating.htm, last accessed20-07-2016.

[4] Howard, B., "Analyzing Online Social Networks", Communications of the ACM,51(11), 2008.

[5] Kirman, B., Lawson, S., & Linehan, C., "Gaming on and off the Social Graph The Social Structure of Facebook Games", In Computational Science and Engineering, 2009. CSE'09. International Conference on (Vol. 4, pp.627-632). IEEE, August 2009.

[6] Aichner, T., & Jacob, F., "Measuring the Degree of Corporate Social Media Use", International Journal of Market Research,57(2), 257-275, 2015.

[7] Skeels, M. M., & Grudin, J., "When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn", In Proceedings of the ACM 2009 international conference on Supporting group work (pp. 95-104). ACM, May 2009.

[8] Dan Kaplan (January 23, 2012)Twilio Blog , "Match.com Lets Online Daters Call or Text Message without Revealing Their Phone Numbers", last accessed 15-07-2016.

[9] "Sky Rock", http://www.skyrock.com/, last accessed 10-07-2016.

[10] David Matthews, "The world University ranking, Do academic social networks share academics'interests?", https://www.timeshighereducation.com/features/do-academic-social networks-share-academics-interests, last accessed 21-07-2016.

[11] Mohammad Rezaei, Mohammad Reza. Detection of fake accounts on social networks using principal component analysis and kernel density estimation algorithm (Case study on Twitter social network). Electronic and Cyber Defense, 1399; (): - , (in Persian).

[12] Ghaderi Piraqom, Saeed, Sakhainia, Mehdi, Mansoorizadeh, Muharram. Diagnosis of anomalies in dynamic social networks based on DOR behavioral assessment: 20.1001.1.23224347.1400.9.1.9.6. Electronic and Cyber Defense, 1400; 9 (1): 115-123 , (in Persian).

[13] Ghesmati, Simin. Manage spam on social media using content tagging. Electronic and Cyber Defense, 2014; 2 (2): - , (in Persian).

[14] Snapfish, https://www.snapfish.com/photo-gift/home.

[15] S. Madisetty and M. S. Desarkar, "A Neural Network-Based, Ensemble Approach for Spam Detection in Twitter," IEEE,Trans. Comput. Soc. Syst., vol. 5, no. 4, pp. 973–984, 2018.

[16] Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, Soumendu Sinha, FNDNet – A deep convolutional neural network for fake news detection, Cognitive Systems Research,Volume 61,,2020 Pages 32-44,

[17] Islam, M.R., Liu, S., Wang, X. et al. Deep learning for misinformation detection on online social networks: a survey and new perspectives. Soc. Netw. Anal. Min. 10, 82 (2020). https://doi.org/10.1007/s13278-020-00696-x

[18] Mohammadi, Bahman and Izadkhah, Habib, 1398, Discovery of fake news on social networks using clustering of fake users, Fifth National Conference on Distribution Computing and Big Data Processing , ،https://civilica.com/doc/961918, (in Persian)

[19] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp. 1-5, doi: 10.1109/ICSCC.2019.8843612.

[20] Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, Soumendu Sinha, FNDNet – A deep convolutional neural network for fake news detection, Cognitive Systems Research,Volume 61,,2020 Pages 32-44,

[21] Mohammadreza Mohammadrezaei, Mohammad Ebrahim Shiri, Amir Masoud Rahmani, "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms", Security and Communication Networks, vol. 2018, Article ID 5923156, 8 pages, 2018. https://doi.org/10.1155/2018/5923156.

[22] Vergeer, M., Hermans, L., & Sams, S., "Online Social Networks and Micro-blogging in Political Campaigning:The Exploration of A New Campaign Tool and A New Campaign Style", Party Politics, 19(3), 477- 501, 2013.

[23] https://github.com/NikhilCodes/Fake-Twitter-Account-Detection-Keras/tree/master/DATASET.

[24] X. Zhang, S. Zhu, and W. Liang, "Detecting spam and promoting campaigns in the Twitter social network," Proc. - IEEE Int. Conf. Data Mining, ICDM, pp. 1194–1199, 2012.

[25] Fazil, M., & Abulaish, M. (2018). A Hybrid Approach for Detecting Automated Spammers in Twitter. IEEE Transactions on Information Forensics and Security,13 (11), pp. 2719 - 2707. https://doi.org/10.1109/TIFS.2018..2825958