آنالیز و بررسی ویژگی‌های ساختاری در تشخیص مکالمه‌های شایعهٔ توییتر

سروه لطفی[۱]، میترا میرزارضایی*[۲]، مهدی حسین‌زاده[۳] و وحید صیدی[۴]

[۱و۲] گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

[۳]مرکز تحقیقات بهداشت روان، پژوهشکده پیشگیری از آسیب‌های اجتماعی، دانشگاه علوم پزشکی ایران، تهران، ایران

[۴]گروه مهندسی کامپیوتر، هوش مصنوعی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

چکیده

توییتر یکی از محبوب‌ترین و مشهورترین شبکه‌های اجتماعی برخط برای گسترش اطلاعات است که در عین قابل اعتمادبودن، می‌تواند به‌عنوان منبعی برای گسترش شایعات باشد. شایعاتی غیرواقعی و فریبنده که می‌تواند تأثیرات جبران‌ناپذیری برروی افراد و جامعه به‌وجود بیاورد. در این پژوهش مجموعه کاملی از ویژگی‌های جدید ساختاری مربوط به درخت پاسخ و گراف کاربران در تشخیص مکالمه‌های شایعه توییتر استخراج شدند. این ویژگی‌ها با توجه به معیارهای سنتی گراف‌ها و معیارهای مخصوص انتشار شایعه، در بازه‌های زمانی مختلف به مدت ۲۴ ساعت از زمان شروع مکالمه‌ها درخصوص رویدادهای بحرانی در توییتر استخراج شده‌اند. نتایج حاصل از بررسی ویژگی‌های جدید، دیدگاه عمیقی از ساختار انتشار اطلاعات در مکالمه‌ها را فراهم می‌کند. براساس نتایج به‌دست‌آمده، ویژگی‌های جدید ساختاری در تشخیص مکالمه‌های شایعه در رویدادهای توییتر مؤثر هستند؛ ازاین‌رو، الگوریتم دسته‌بند شایعه مبتنی بر ویژگی‌های جدید ساختاری، زبانی و کاربران در تشخیص مکالمه‌های شایعه زبان انگلیسی توییتر ، پیشنهاد داده شد. روش پیشنهادی در مقایسه با روش‌های پایه، عملکرد بهتری دارد. همچنین، با توجه به اهمیت کاربر توییت منبع در مکالمه‌ها، این کاربر از جنبه‌های مختلفی موردبررسی و آنالیز قرار گرفت.

واژگان کلیدی: مکالمه، تشخیص شایعه، توییتر، درخت پاسخ، گراف کاربران.

Analysis of Structural Features in Rumor Conversations Detection in Twitter

Serveh lotfi[1], Mitra Mirzarezaee[2*], Mehdi Hosseinzadeh[3] & Vahid Seydi[4]

[1, 2*]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

[3]Mental Health Research Center, Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran.

[4]Department of Computer Engineering, Faculty of Technical and Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran.

Abstract

Today, online social media with numerous users from ordinary citizens to top government officials, organizations, artists and celebrities, etc. is one of the most important platforms for sharing information and communication. These media provide users with quick and easy access to information so that the content of shared posts has the potential to reach millions of users in a matter of seconds. Twitter is one of the most popular and practical/used online social networks for spreading information, which, while being reliable, can also, be a source for spreading unrealistic and deceptive rumors as a result can have irreversible effects on individuals and society.

Recently, several studies have been conducted in the field of rumor detection and verify using models based on deep learning and machine learning methods. Previous research into rumor detection

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات

●تاریخ ارسال مقاله: ۱۳۹۹/۱/۱۵ ●تاریخ پذیرش: ۱۳۹۹/۱۲/۱۱ ●تاریخ انتشار: ۱۴۰۰/۱۰/۲۷ ●نوع مطالعه: پژوهشی

سال ۱۴۰۰ شمارهٔ ۳ پیاپی ۴۹

[DOI: 10.52547/jsdp.18.3.45]

has focused more on linguistic, user, and structural features. Concerning structural features, they examined the retweet propagation graph. However, in this study, unlike the previous studies, new structural features of the reply tree and user graph in extracting rumored conversations were extracted and analyzed from different aspects.

In this study, the effectiveness of new structural features related to reply tree and user graph in detecting rumored conversations in Twitter events were evaluated from different aspects. First, the structural features of the reply tree and user graph were extracted at different time intervals, and important features in these intervals were identified using the Sequential Forward Selection approach. To evaluate the usefulness of valuable new structural features, these features have been compared with consideration of linguistic and user-specific features. Experiments have shown that combining new structural features with linguistic and user-specific features increases the accuracy of the rumor detection classification. Therefore, a rumor classification algorithm based on new structural, linguistic, and user-specific features in rumor conversation detection was proposed. This algorithm performs better than the basic methods and detects rumored conversations with greater accuracy. In addition, due to the importance of the source tweet user in conversations, this user was examined and analyzed from different aspects. The results showed that most rumored conversations were started by a small number of users. Rumors can be prevented by early identification of these users on Twitter events.

Keywords: Conversion, Rumor detection, Twitter, Reply tree, User graph.

۱- مقدمه

شبکه‌های اجتماعی به‌دلیل سهولت و سرعت اشتراک‌گذاری اطلاعات به‌ویژه در زمان وقوع رویدادهای واقعی به رسانه‌ای موثق برای ارتباطات مشترک تبدیل شده‌اند؛ علاوه‌براین، توییتر یکی از سرویس‌های اجتماعی محبوب است که با حجم زیادی از اطلاعات به‌عنوان یک منبع خبری برجسته، اغلب اطلاعات را سریع‌تر از رسانه‌های سنتی منتشر می‌کند [1]. در طی چند سال گذشته در کنار ترویج اطلاعات موثق و قابل اعتماد، از توییتر برای انتشار اطلاعات نادرست در قالب شایعات نیز سوء استفاده شده است. گسترش شایعات در رسانه‌های اجتماعی سبب ایجاد زیان‌های مالی در زیرساخت‌ها شده و زندگی انسان را در دنیای واقعی با تهدید مواجه کرده است؛ به‌خصوص در شرایط بحرانی و بروز حوادث اورژانسی به‌دلیل افزایش نگرانی‌ها و آسیب‌پذیری عاطفی، مردم بیشتر به دام شایعات و محتوی‌های جعلی می‌افتند [2]؛ بنابراین، شایعات امنیت عمومی مردم را به خطر می‌اندازد و مردم را گمراه می‌کنند.

تعاریف متعددی در منابع مختلف برای شایعه وجود دارد. برای مثال کای و همکارانش [3] شایعه را اطلاعات نادرست تعریف کرده‌اند؛ ولی در دیکشنری آکسفورد[1]،شایعه، داستان و گزارشی است که صحت[2] آن هنوز بررسی نشده است. در این پژوهش،شایعه در زمان پست آن توسط کاربر، یک ادعا در توییتر است که هنوز صحت آن تأیید نشده است. در تشخیص شایعه، برای مجموعه توییت‌های ورودی، مشخص می‌شود شایعه هستند یا خیر؟ در بررسی صحت شایعه، نتیجه شایعه به درست، نادرست و اثبات‌نشده ختم می‌شود. همچنین در بررسی تشخیص شایعات، ما با دو دسته شایعات مواجه هستیم. دسته نخست شایعاتی که درقبل شناسایی شده‌اند و در بازه زمانی طولانی در مورد آن‌ها صحبت شده و صحت آن‌ها مورد تأیید همه نیست و دسته دوم شایعاتی هستند که در حین اخبار فوری و رویدادهای اضطراری به‌وجود می‌آیند. این شایعات شناخته‌شده نیستند و شناسایی زودهنگام آن‌ها در رویدادها و شرایط اضطراری ضروری است [4]. مکالمه‌ها نقش مهمی را در توییتر ایفا می‌کنند. به‌طورتقریبی یک چهارم کاربران توییتر از طریق مکالمه‌ها با یکدیگر ارتباط برقرار می‌کنند. به‌طوری‌که بخش زیادی از توییت‌های توییتر مربوط به مکالمه‌ها است [5]. یک مکالمه در توییتر شامل یک توییت منبع (که شروع‌کننده مکالمه است) و توییت‌های پاسخ است. اگر در یک مکالمه، توییت منبع شایعه باشد، آن مکالمه را شایعه و اگر توییت منبع غیر شایعه باشد، آن مکالمه را غیر شایعه می‌نامیم. با توجه به اهمیت شناسایی مکالمه‌های شایعه که از انتشار شایعات در شرایط بحرانی و اضطراری جلوگیری می‌کند، در این پژوهش روشی برای تشخیص مکالمه‌های شایعه زبان انگلیسی توییتر، ارائه شده است.

در مطالعات قبلی تشخیص شایعات، از سه جنبه انتشار اطلاعات شایعه شامل ویژگی‌های زبانی، ویژگی‌های کاربران شرکت‌کننده در شایعه و ویژگی‌های گراف انتشار بازتوییت[3] استفاده کرده‌اند [10-6]. کاستیلیو و همکاران [11] روشی خودکار را برای ارزیابی اعتبار موضوع‌های خبری ارائه دادند. آن‌ها بر روی توییت‌های موضوع‌های

[1] https://en.oxforddictionaries.com/definition/rumour
[2] Verification
[3] retweet

[DOI: 10.52547/jsdp.18.3.45]

داغ توییتر متمرکز شدند و قابلیت اعتبار آن‌ها را بر اساس استخراج ویژگی‌های مختلف توییتر ارزیابی کردند. مجموع ویژگی‌های مورد استفاده، شامل ویژگی‌های محتوایی توییت، کاربران، موضوع و گراف انتشار بازتوییت است. کوان و همکاران [12] مجموعه جدیدی از ویژگی‌های زبانی، ساختاری مبتنی بر گراف انتشار بازتوییت و زمانی را ارائه دادند. آن‌ها در ویژگی زمانی، انتشار شایعه‌ها در بازه زمانی را مورد بررسی قرار دادند. یانگ و همکارانش [13] از ویژگی‌های ساختار مبتنی بر گراف انتشار بازتوییت و محتوای پست‌ها استفاده کردند؛ علاوه‌براین ویژگی‌ها، آن‌ها ویژگی‌های جدیدی مبتنی بر برنامه‌ای که کاربر برای ارسال پست استفاده کرده و مکانی را که رویداد اتفاق افتاده است شناسایی کردند. در این پژوهش با این فرضیه که استخراج ویژگی‌های جدید در گراف‌های مکالمه موجب افزایش دقت تشخیص شایعه‌ها در توییتر می‌شود، متفاوت از پژوهش‌های قبلی، ویژگی‌های جدید ساختاری مربوط به درخت پاسخ و گراف کاربران در تشخیص مکالمه‌های شایعه استخراج شدند. متغیرهای مستقل در این فرضیه، ویژگی‌های جدید ساختاری و متغیر وابسته دقت تشخیص شایعه است. نتایج آزمایش‌ها و ارزیابی‌ها، مؤثربودن ویژگی‌های جدید ساختاری را در تشخیص شایعه‌های توییتر نشان می دهد.

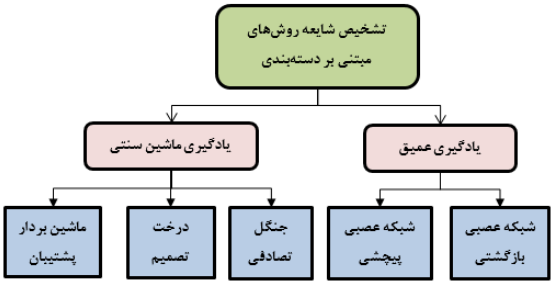
دستاوردهای علمی این پژوهش به شرح زیر است:

- در این پژوهش برای نخستین‌بار، با استفاده از الگوریتم‌های تئوری گراف ویژگی‌های جدید ساختاری درخت وزن‌دار پاسخ و گراف وزن‌دار کاربران برای مکالمه‌های توییتر در تشخیص شایعات استخراج شدند.
- فرایند انتشار مکالمه‌های شایعه در بازه‌های زمانی مختلف به مدت ۲۴ ساعت با استفاده از الگوریتم دسته‌بند مورد بررسی قرار گرفته و ویژگی‌های مهم ساختاری مکالمه‌ها شناسایی شدند.
- نتایج آزمایش‌ها مؤثربودن ویژگی‌های ساختاری در تشخیص مکالمه‌های شایعه را نشان می‌دهد به‌طوری‌که در مقایسه با روش‌های پایه مقدار F1 را ۲۰ درصد افزایش داده است.

در ادامه، ساختار مقاله بدین شکل است؛ بخش دوم پیشینه پژوهش‌ها مربوط به بررسی شایعه در شبکه اجتماعی توییتر را بیان می‌کند؛ بخش سوم روش پیشنهادی پژوهش شامل استخراج ویژگی‌ها را تشریح می‌کند، بخش چهارم آزمایش‌ها و نتایج از جنبه‌های مختلف مورد بررسی و ارزیابی قرار می‌گیرد و در بخش نهایی نتیجه‌گیری و پیشنهادهایی برای پژوهش‌های آینده مطرح می‌شود.

۲- مروری بر کارهای پیشین

درهمین‌اواخر پژوهش‌های متعددی در زمینه تشخیص و صحت شایعات با استفاده مدل‌های مبتنی بر روش‌های یادگیری عمیق و یادگیری ماشین انجام شده است [6-9,14]. شکل (۱) خلاصه رویکردهای را که در تشخیص شایعه‌ها استفاده شده است، نشان می دهد.



(شکل-۱): رویکردهای مختلف در تشخیص شایعات در کارهای پیشین را نشان می دهد.

(Figure-1): Shows different approaches to rumors detection in previous works.

در تشخیص شایعه‌ها روش‌های مبتنی بر یادگیری ماشین با استفاده از گراف انتشار اطلاعات، پژوهش‌های زیادی انجام شده است [15,16]، از جمله این پژوهش‌ها می‌توان به گراف انتشار همه‌گیری [17,18] اشاره کرد. به‌علاوه پژوهش‌هایی در شناسایی کاربران بانفوذ و منبع انتشار از طریق گراف انتشار اطلاعات انجام شده است [19-22]. در گراف‌های مکالمه توییتر، کوگان و همکارانش روشی را برای بازسازی کامل مکالمه‌های توییتر و مقایسه ساختارهای گرافیکی آن‌ها ارائه دادند [23]. در پژوهشی دیگر، نیشی و همکارانش مدل‌های مختلفی از درخت پاسخ ایجاد کردند. نتایج به‌دست‌آمده نشان داد که درجه ورودی توییت منبع، بر نحوه فرم‌دهی درخت پاسخ تأثیر مؤثری دارد [24].

در بررسی ویژگی‌های ساختاری شایعات، مندوزا و همکارانش رفتارهای کاربران در تبلیغ شایعات دروغین و اخبار واقعی مربوط به توییتر را از جنبه‌های مختلف بررسی کردند [25]؛ علاوه‌براین جین و همکارانش [26] ساختار آبشاری انتشار شایعات در شبکه‌های اجتماعی را مورد تجزیه و تحلیل قرار دادند. نتایج به‌دست‌آمده بر روی شایعات پیرامون اپیدمی ابولا نشان داد، که شایعات به‌طوردقیق همانند اخبار واقعی گسترش می‌یابند. کاستی

[DOI: 10.52547/jsdp.18.3.45]

لو و همکارانش [11] روشی خودکار برای ارزیابی اعتبار موضوع‌های خبری ارائه دادند. آن‌ها بر روی توییت‌های ۷۴۷ موضوع داغ توییتر متمرکز شدند و قابلیت اعتبار آن‌ها را بر اساس استخراج ویژگی‌های مختلف ارزیابی کردند. مجموع ویژگی‌های مورداستفاده آن‌ها در این پژوهش، شامل ویژگی‌های محتوای توییت، کاربران، موضوع و انتشار بازتوییت است. مطالعه بالا به‌عنوان منبع اصلی در بیشتر پژوهش‌های مربوط به تشخیص و صحت شایعات مورد ارزیابی قرار می‌گیرد. وثوقی و همکارانش [27] صحت شایعات را با استفاده از سه ویژگی زبانی، کاربران و ویژگی‌های انتشار زمانی مورد بررسی قراردادند. آن‌ها هفده ویژگی را انتخاب کرده و با استفاده از مدل مارکوف مخفی، در صحت شایعات دقت ۷۵ درصد را به‌دست آوردند. نتایج آن‌ها نشان داد که ویژگی‌های انتشار زمانی، کارایی بیشتری در مقایسه با ویژگی‌های زبانی و کاربران دارد. در پژوهشی دیگر وثوقی و همکارانش [28] تفاوت شیوه‌های انتشار اخبار درست و نادرست ۱۲۶۰۰۰ داستان توییتر را مورد بررسی قرار دادند. نتایج آن‌ها نشان داد که اخبار نادرست سریع‌تر از اخبار موثق پخش و اخبار کذب و دروغین سیاسی در مقایسه با اخبار ناموثق مربوط به بلایای طبیعی، داستان‌های علمی، افسانه‌های شهری و اطلاعات مالی سریع‌تر پخش می‌شوند. کووان و همکارانش [12] مجموعه جدیدی را از ویژگی‌های زبانی،کاربران، ساختاری و زمانی ارائه دادند. آن‌ها ویژگی‌های زمانی، انتشاربازتوییت شایعات را موردبررسی قرار دادند. و ویژگی‌های زبانی را با استفاده از دیکشنری LIWC[1] استخراج کردند. نتایج به‌دست‌آمده از بررسی ویژگی‌های استخراج شده در۱۳۰ موضوع شایعه و غیر شایعه نشان داد که ویژگی‌های زمانی پیشنهادشده، در تشخیص شایعات مؤثر هستند.کووان و همکارانش [29] در پژوهشی دیگر به بررسی ویژگی‌های زبانی، کاربران، ساختاری مبتنی بر گراف انتشار بازتوییت و زمانی در بازه‌های زمانی مختلف با استفاده از روش‌های آماری و یادگیری ماشین پرداختند. آن‌ها الگوریتم‌های دسته‌بندی را برای هر دو دوره کوتاه و بلند مدت تعریف کردند. یانگ و همکارانش [13] در پژوهشی مشابه کاستی لو، روشی برای تشخیص شایعات با استفاده از **ماشین بردار پشتیبان** [2] با کرنل **تابع پایه شعاعی**[3] انجام دادند. آن‌ها ویژگی‌های جدیدی در گراف انتشار بازتوییت، کاربران و محتوای پیام‌ها شناسایی کردند. جایاسمیدیس و همکارانش [17] نتایج آزمایش خود در ۷۲ شایعه مختلف شامل صد میلیون توییت ارائه دادند. آن‌ها سه نوع ویژگی مبتنی بر زبان، کاربر و گراف انتشار بازتوییت را استخراج کردند. با استفاده از درخت تصمیم نتایج قابل قبولی را به‌دست آوردند. گاپتا و همکاران [30] اعتبار اطلاعات توییتر را در چهارده رویداد خبری مهم دنیا مورد تجزیه‌وتحلیل قراردادند. آن‌ها ویژگی‌های مهمی را در پیش‌بینی اعتبار اطلاعات شناسایی کردند. این ویژگی‌ها در محتوای توییت شامل تعداد نویسه‌های منحصربه‌فرد، ضمایر و شکلک‌ها بود. در ویژگی‌های کاربران، تعداد دنبال‌کننده‌ها و طول نام کاربری به‌عنوان ویژگی‌های مهم شناسایی شدند. آن‌ها با استفاده از روش‌های یادگیری ماشین نظارتی و ویژگی‌های مهم، توییت‌ها را رتبه‌بندی کردند.

در تشخیص شایعه‌ها روش‌های یادگیری عمیق، ما و همکاران [31] مدل مبتنی بر شبکه‌های مولد تخاصمی[4]را در تشخیص شایعات ارائه کردند. در این مدل تفکیک‌کننده نقش دسته‌بند را دارد و تولیدکننده مربوطه سعی دارد با حذف نوفه‌های ناسازگار، عملکرد تفکیک‌کننده را بهبود ببخشد. تولیدکننده در شناسایی ویژگی‌های مهم‌تر شایعات از مدل توالی دنباله‌دار مبتنی بر واحد برگشتی دروازه‌دار[5] استفاده کردند. آلسیدی و همکاران [32] با استفاده از یک لایه شبکه عصبی پیچشی[6]و لایه‌های به‌طورکامل متصل[7] شایعه‌ها را تشخیص دادند. آن‌ها آزمایش‌های مختلفی را برای یافتن بهترین پارامترها در بهبود عملکرد مدل انجام دادند. آن‌ها از مجموعه‌داده‌های PHEME در آموزش مدل استفاده کردند. سانتوشکومار و همکاران [33] روشی مبتنی بر شبکه عصبی پیچشی را ارائه دادند و مدلشان را با دو شبکه عصبی پیچشی پیاده‌سازی کردند. در ورودی نخستین شبکه عصبی پیچشی همه پست‌های مربوط به یک رویداد وارد می‌شود. این ورودی‌ها در شبکه به یک بردار با طول متغیر تبدیل می‌شود. ورودی دومین شبکه عصبی پیچشی، اطلاعات زمانی و ساختاری مرتبط به پست‌های ورودی نخستین شبکه را در بر می‌گیرد که شبکه دوم آن را به برداری تبدیل می‌کند. ترکیب بردارهای خروجی شبکه‌های عصبی پیچشی به‌عنوان ورودی در درخت تصمیم برای دسته‌بندی شایعه‌ها استفاده می‌شود. زبیر اصغر و همکاران [34] مدل مبتنی بر واحدهای حافظهٔ کوتاه‌مدت طولانی دوطرفه[8] و شبکه عصبی پیچشی را ارائه کردند. در لایه واحدهای حافظهٔ کوتاه‌مدت طولانی دوطرفه، وابستگی‌های توییت براساس

[1] Linguistic Inquiry and Word Count
[2] Support Vector Machines
[3] Radial Basis Function

[4] Generative Adversarial Networks
[5] Gated recurrent unit
[6] Convolutional neural network
[7] Dense layer
[8] Bidirectional Long short-term memory

[DOI: 10.52547/jsdp.18.3.45]

محتوی اطلاعات گذشته و آینده به‌دست می‌آید و در لایه شبکه عصبی پیچشی ویژگی‌های توییت از طریق پردازش اطلاعات به‌صورت سلسله‌مراتبی استخراج می‌شود، براساس این خروجی‌ها دسته‌بندی شایعه و غیر شایعه مشخص می‌شود. لی و همکاران [35] مدلی مبتنی بر واحد برگشتی دروازه‌دار دوطرفه در شناسایی رویدادهای شایعه را ارائه دادند. ارزیابی و نتایج آن‌ها نشان داد با استفاده از چندلایه واحد برگشتی دروازه‌دار دوطرفه می‌توان اطلاعات معنایی و احساسی مربوط به مجموعه توییت‌های یک رویداد را استخراج کرد و دقت مدل را افزایش داد. جدول (۱) خلاصه‌ای پژوهش‌های مهم در زمینه تشخیص و صحت شایعه‌ها را نشان می‌دهد. همان‌طور که مشاهده می‌شود، پژوهش‌های قبلی در تشخیص شایعات، بیشتر بر روی ویژگی‌های زبانی، کاربران و ساختاری متمرکز بودند. در ویژگی‌های ساختاری گراف انتشار بازتوییت را مورد بررسی قرار دادند. اما در این پژوهش، متفاوت از پژوهش‌های قبلی، ویژگی‌های جدید ساختاری گراف درخت پاسخ و گراف کاربران در تشخیص مکالمه‌های شایعه استخراج شدند. ما متعقد هستیم که این ویژگی‌های جدید ساختاری در تشخیص شایعه‌ها موثر هستند.

(جدول-۱): خلاصه‌ای از پژوهش‌های مهم در زمینه تشخیص و صحت شایعه‌ها
(Table-1): A summary of important research in the field of rumor detection and verification

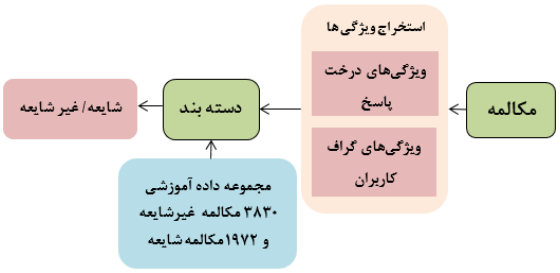
نام نویسنده	ویژگی‌ها و روش مورداستفاده
مندوزا و همکاران [25]	ویژگی‌های کاربران و گراف انتشار باز توییت
جین و همکارانش [26]	گراف انتشار بازتوییت
کاستیلیو و همکاران [36]	ویژگی‌های کاربران، موضوع، محتوای توییت و گراف انتشار بازتوییت
وثوقی و همکاران [27]	ویژگی‌های زبانی، کاربران و گراف انتشار بازتوییت
کووان و همکارانش [29]	ویژگی‌های زبانی، ساختاری مبتنی بر گراف انتشار بازتوییت و زمانی
کووان و همکارانش [37]	بررسی ویژگی‌های زبانی، ساختاری مبتنی بر گراف انتشار بازتوییت و زمانی در بازه‌های زمانی مختلف
یانگ و همکارانش [13]	ویژگی‌های ساختار مبتنی بر گراف انتشار بازتوییت، محتوای پست‌ها و مکان رویداد
جایاسمیدیس و همکارانش [38]	ویژگی‌های زبانی و گراف انتشار باز توییت
گاپتا و همکاران [30]	ویژگی‌های مبتنی بر محتوای توییت و کاربران
ما و همکاران [31]	استفاده از شبکه‌های عصبی مولد تخاصمی و واحد برگشتی دروازه‌دار
آلسیدی و همکاران [32]	استفاده از شبکه عصبی پیچشی و لایه‌های متراکم
سانتوشکومار وهمکاران [33]	استفاده از دو شبکه عصبی پیچشی
زبیر اصغر و همکاران [34]	استفاده از دو شبکه عصبی پیچشی و واحدهای حافظهٔ کوتاه‌مدت طولانی دوطرفه
لی و همکاران [35]	استفاده از شبکه عصبی واحد برگشتی دروازه‌دار دوطرفه

۳- روش پیشنهادی (user-reply-feature)

در این بخش، جزئیات مدل پیشنهادی بیان شده است. ساختار کلی مدل پیشنهادی برای تشخیص مکالمه‌های شایعه در شکل (۲) نمایش داده شده است. در این روش برای هر مکالمه، ویژگی‌های جدید درخت پاسخ و گراف کاربران استخراج شده، سپس الگوریتم یادگیری دسته‌بند بر اساس ویژگی‌های ساختاری انجام می‌شود. در مرحله بعد به‌وسیلهٔ دسته‌بند، مکالمه شایعه از غیر شایعه تشخیص داده می‌شود. مدل پیشنهادی مبتنی بر استخراج ویژگی‌های جدید درخت پاسخ و گراف کاربران user-reply-feature نامیده می‌شود. استخراج ویژگی‌های یادشده در شکل با بررسی ساختار گراف مکالمه‌ها و انتشار اطلاعات در [23,24] انجام شده است. از ویژگی‌های استخراج‌شده، ویژگی‌های مهم با استفاده از روش پوشه‌ای[1] جستجوی مستقیم ترتیبی[2] شناسایی شدند؛ سپس برای بررسی مؤثربودن ویژگی‌های انتخابی ساختاری، این ویژگی‌ها با ویژگی‌های زبانی و کاربران در تشخیص مکالمه‌های شایعه مقایسه شدند؛ علاوه‌براین در گراف کاربران با توجه به اهمیت کاربر توییت منبع، از جنبه‌های مختلفی مورد بررسی قرار گرفت. در ادامه، به شرح جزئیات مدل پیشنهادی پرداخته می‌شود.

[1] wrapper
[2] Sequential Forward Selection

[DOI: 10.52547/jsdp.18.3.45]

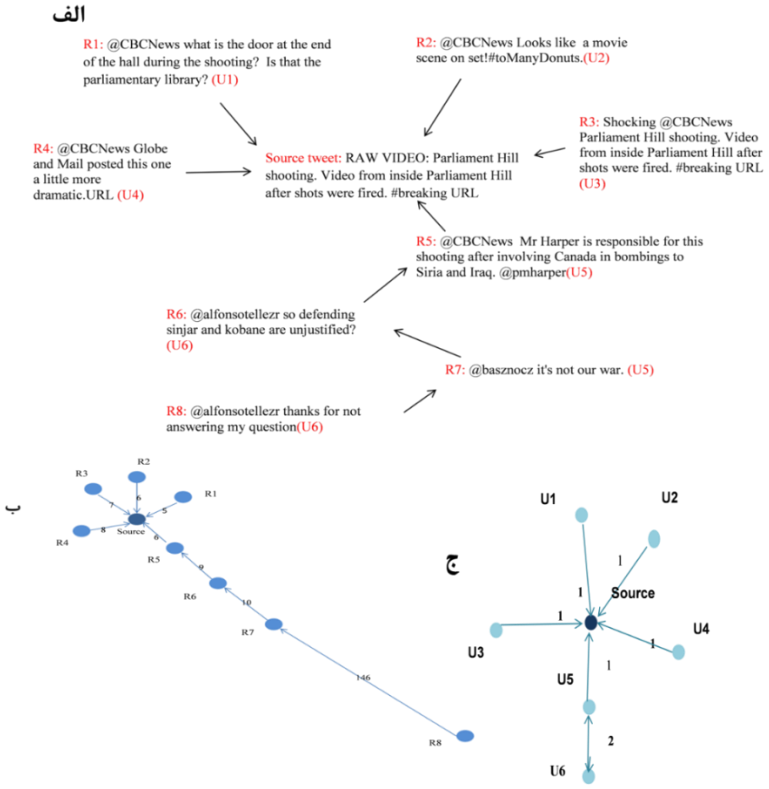


(شکل-۲): ساختار کلی مدل پیشنهادی(user-reply-feature) برای تشخیص مکالمه‌های شایعه

(Figure-2): The general structure of the proposed model (user-reply-feature)for rumor conversation detection.

۱-۳- درخت پاسخ

یک مکالمه در توییتر شامل یک توییت منبع و توییت‌های پاسخ است که در آن توییت منبع شروع‌کننده مکالمه است و توییت‌های پاسخ در جواب به توییت منبع و توییت‌های پاسخ دیگر ایجاد شده‌اند. اگر دریک مکالمه توییت منبع شایعه باشد، آن مکالمه را شایعه و اگر توییت منبع غیر شایعه باشد، مکالمه را غیر شایعه می‌نامیم. شکل (۳) بخش الف یک نمونه مکالمه شایعه در تیراندازی اوتاوا را نشان می‌دهد که شامل یک توییت منبع و هشت توییت پاسخ است. همه توییت‌های موجود در شکل با همدیگر یک مکالمه را تشکیل می‌دهند. شکل (۲) بخش ب، گراف وزن‌دار درخت پاسخ مکالمه شایعه، مربوط به بخش الف شکل را نشان می‌دهد. درخت از گره‌ها و یال‌های جهت‌دار تشکیل شده است. توییت‌های مکالمه گره‌های درخت هستند و یال‌های درخت، از سمت توییت پاسخ‌دهنده به توییتی است که به آن پاسخ داده است.



(شکل-۳): الف) یک مکالمه شایعه در تیراندازی اوتاوا را نشان می‌دهد که شامل توییت منبع و توییت‌های پاسخ است. ب) درخت پاسخ مربوط به مکالمه شایعه بخش الف را نشان می‌دهد. ج) گراف کاربران مربوط به مکالمه شایعه بخش الف را نشان می‌دهد.

(Figure-3): a) Shows a rumor conversation in the Ottawa shooting including the source tweet and the reply tweets. (b) The reply tree represents the rumored conversation in section a. (c) The user graph of the users represents the rumored conversation in section a.

یکی از ویژگی‌های مهم شایعه، جنبه زمانی آن است. انتشار شایعات در بازه‌های زمانی مختلف، متفاوت از غیر شایعات است [27,39,40]. با توجه به اهمیت ویژگی زمانی، در درخت پاسخ، مقدار وزن یال‌های بین گره پاسخ و گره توییت منبع اصلی بر اساس فاصله زمانی انتشار (برحسب دقیقه) محاسبه می‌شود. در شکل ۳-(ب) درخت پاسخ یک مکالمه نمایش داده شده است. در این درخت، وزن یال بین گره R7 (توییت منبع) و گره R8 (توییت پاسخ) برابر با ۱۴۶ است که فاصله زمانی انتشار برحسب دقیقه بین توییت پاسخ R_8 و توییت منبع اصلی را نشان می‌دهد. همان‌طورکه در شکل (۳) نشان داده شده است، در درخت پاسخ، پنج توییت به‌طور مستقیم به توییت

[DOI: 10.52547/jsdp.18.3.45]

منبع اصلی و سه توییت به توییت‌های دیگر پاسخ داده‌اند. درشکل تعداد گره‌های درخت برابر ۹ است که با تعداد توییت‌های مکالمه برابر است. شبه‌کد ایجاد درخت پاسخ به شرح زیر است.

Algorithm 1: create reply tree
Input: conversation ‘s tweets
Output: reply tree
for each tweet in conversation:
  if reply _tweet_id is null:
    Create tree
    Add tweet as root node
  else:
    Add tweet as node to reply tree
    Add edge between this node and source node

در این پژوهش ویژگی‌های ساختاری جدید درخت پاسخ برای تشخیص مکالمه‌های شایعه استخراج شدند؛ که شامل معیارهای سنتی مانند تعداد یال‌های درخت و معیارهای مخصوص انتشار شایعه از قبیل وزن یال‌های درخت پاسخ، برحسب فاصله زمانی انتشار توییت پاسخ به توییت منبع اصلی است [31]. جدول (۲) ویژگی‌های استخراج شده درخت پاسخ را نشان می‌دهد.

۲-۳- گراف کاربران

گراف کاربران پیش‌نمایشی از درخت پاسخ را نشان می‌دهد. در این گراف، گره‌ها، کاربران هستند و یال‌های آن، از طرف کاربر پاسخ‌دهنده به کاربری است که به آن پاسخ می‌دهد. شکل۳ –(ج)، گراف جهت‌دار کاربران مربوط به مکالمه شکل ۳-(الف) را نشان می‌دهد. همان‌طورکه مشاهده می‌شود، یال‌های این گراف دارای وزن هستند که تعداد توییت‌های ارتباطی کاربران با همدیگر را نشان می‌دهد. در این شکل پنج کاربر به‌طور مستقیم به کاربر توییت منبع پاسخ داده‌اند. درمجموع هفت کاربر شرکت‌کننده در مکالمه، از طریق نه توییت باهم ارتباط دارند. شبه‌کد ایجاد گراف مکالمه کاربران در بخش زیر نشان داده شده است. جدول (۲) ویژگی‌های استخراج‌شده گراف کاربران را نشان می‌دهد.

Algorithm2: create user graph
Input: conversation ‘s tweets
Output: user graph
for each tweet in conversation:
  if reply _tweet_id is null:
    Create graph
    Add user as root node
  else :
    if has_edge in graph:
      edge ['weight'] += 1
    else:
      Add user as node to user graph
      Add edge between this node and source node

(جدول-۲): ویژگی‌های جدید ساختاری درخت پاسخ و گراف کاربران را نشان می‌دهد.

(Table-2): Shows the new structural features in the reply tree and user graph

نوع گراف	توضیحات ویژگی‌ها و نحوه محاسبه آن‌ها	نماد ویژگی‌ها
درخت پاسخ	تعداد توییت‌های مکالمه (تعداد گره‌ها درخت پاسخ) $Num\_node\_tree = \sum_{i=0}^{n} node_i$	Num_node_tree
	تعداد توییت‌های پاسخی که مستقیماً به توییت منبع پاسخ می‌دهند. (درجه ورودی توییت منبع) Source_tweet_indegree = indegree(root)	Source_tweet_indegree
	طولانی‌ترین مسیر بین برگ درخت پاسخ و توییت منبع اصلی Len_ path_leaf_root_tree = max(Len_ path_leaf_root(root, $leaf_i$))	Len_ path_leaf_root_tree
	متوسط پاسخ به هر توییت مکالمه (متوسط درجه ورودی گره‌ها) $Ave\_indegree\_tweet = \frac{\sum_{i=0}^{n} indegree(node_i)}{n}$	Ave_indegree_tweet
	توییت‌های مکالمه بدون پاسخ (تعداد برگ‌ها) $\text{Num\_leaf\_tree} = \sum_{i=0}^{n} indegree(node_i) = 0$	Num_leaf_tree
	تراکم بین لینک‌های گراف (چگالی) $\text{Density\_tree}=\frac{\lvert number\ edge\rvert}{\lvert number\ node\rvert(\lvert number\ node\rvert-1)}$	Density_tree
	کمترین، بیشترین و متوسط فاصله زمانی انتشار بین توییت‌های پاسخ با توییت منبع (کمترین، بیشترین و متوسط وزن درخت) $Ave\_weigh\_tree = \frac{\sum_{i=0}^{n} weight(edge_i)}{n}$	Max,Min,Ave_weigh_tree
	تعداد یال‌های درخت پاسخ $\text{Num\_edge\_tree} = \sum_{i=0}^{m} edge_i$	Num_edge_tree
	تعداد کاربران شرکت‌کننده در مکالمه (تعداد گره‌ها) $Num\_node\_user = \sum_{i=0}^{n} node_i$	Num_node_user

[DOI: 10.52547/jsdp.18.3.45]

گراف کاربران	تعداد یال‌ها گراف کاربران $\text{Num\_edge\_user} = \sum_{i=0}^{m} edge_i$	Num_edge_user
	تعداد کاربران پاسخ‌دهنده به کاربر توییت منبع اصلی (درجه ورودی کاربر توییت منبع) Source_user_indegree = indegree(root)	Source_user_indegree
	تعامل کاربر توییت منبع با کاربرهای شرکت‌کننده در مکالمه (درجه خروجی کاربر توییت منبع) Source_user_outdegree = outdegree(root)	Source_user_outdegree
	طولانی‌ترین مسیر کوتاه ارتباط دو کاربر (قطر) $\text{Diameter} = max_{ij}(shortest\_path\_length(i,j))$	Diameter
	متوسط تعامل کاربران با یکدیگر (میانگین درجه ورودی و خروجی) $Ave\_user\_outdegree = \frac{\sum_{i=0}^{n} outdegree(node_i)}{n}$ $Ave\_user\_indegree = \frac{\sum_{i=0}^{n} indegree(node_i)}{n}$	Ave_user_ indegree or Ave_user_ outdegree
	کاربرانی که به توییت‌های آن‌ها پاسخ داده نشده است. (تعداد برگ‌ها) $\text{Num\_leaf\_user} = \sum_{i=0}^{n} indegree(node_i) = 0$	Num_leaf_user
	تعداد زیر گراف‌ها در گراف کاربران که طول مسیر کاربران شرکت‌کننده با کاربر توییت منبع برابر با ۴،۳،۲ و ۵ است.(ارتباط زنجیره‌ای کاربران شرکت کننده در مکالمه را نشان می‌دهد.) K=2,3,4,5 $\text{pattern\_chain} = \text{Len\_ path}(\text{root}, node_i) = k$	pattern_chain2,3,4,5
	بیشترین درجه ورودی و خروجی در گراف کاربران $\text{max\_user\_outdegree} = \max(\text{outdegree}(\text{node}_i))$ $\text{max\_user\_indegree} = \max(\text{indegree}(\text{node}_i))$ i=1,2,..n	max_user_ indegree or max_user_ outdegree
	متوسط تعداد توییت‌های ارتباطی بین کاربران (متوسط وزن یال‌ها) $Ave\_weigh\_user = \frac{\sum_{i=0}^{n} weight(edge_i)}{n}$	Ave_weight_user
	بیشترین و کمترین وزن یال‌ها $min\_weigh\_user = \min(weight(edge_i))$ $max\_weigh\_user = \max(weight(edge_i))$	Min,Max_weight_user
	طولانی‌ترین مسیر بین برگ و کاربر توییت منبع اصلی $\text{Len\_ path\_leaf\_root\_user} = \max(\text{Len\_ path\_leaf\_root}(\text{root}, leaf_i))$	Len_ path_leaf_root _user

۴- آزمایش‌ها و نتایج

در این بخش کارایی ویژگی‌های جدید ساختاری (جدول ۲) در فرایند تشخیص مکالمه‌های شایعه، در رویدادهای توییتر ارزیابی می‌شود. در ابتدا جزئیات مجموعه‌دادگان بررسی می‌شود، سپس ویژگی‌های مهم ساختاری در بازه‌های زمانی مختلف شناسایی می‌شوند. برای ارزیابی مفیدبودن ویژگی‌های ساختاری مهم، این ویژگی‌ها با ویژگی‌های زبانی و کاربران مقایسه شده، علاوه‌براین، روش پیشنهادی مبتنی بر ویژگی‌های ساختاری، زبانی و کاربران با روش‌های پایه تشخیص شایعه مقایسه شدند. علاوه بر این روش پیشنهادی شامل ویژگی‌های ساختاری، زبانی و کاربران در مطالعه موردی در رویدادهای توییتر مورد ارزیابی قرار گرفت. در آخر با توجه به اهمیت کاربر توییت منبع در تشخیص شایعات، این کاربر از جنبه‌های مختلف در رویدادهای توییتر بررسی شد. جدول (۳) خلاصه آزمایش‌ها و اهداف آزمایش را نشان می‌دهد. پیاده‌سازی درخت پاسخ و گراف کاربران با استفاده از کتابخانه networkx[1] در پایتون انجام شده و پیاده‌سازی و نمودارهای این پژوهش در پایتون و با استفاده از کتابخانه‌های آن انجام شده است.

[1] https://networkx.github.io/

(جدول-۳): خلاصه آزمایش‌ها و اهداف آن‌ها

(Table 3): Summary of experiments and their goals

آزمایش‌ها	اهداف
انتخاب ویژگی‌های مهم ساختاری	شناسایی ویژگی‌های مهم ساختاری در درخت پاسخ و گراف کاربران در بازه‌های زمانی مختلف ۶،۱۲،۱۸و۲۴ ساعته و انتخاب دسته بند مناسب
مقایسه ویژگی‌های ساختاری با ویژگی‌های زبانی و کاربران	ارزیابی ویژگی‌های جدید ساختاری در مقایسه با ویژگی‌های زبانی و کاربران با استفاده الگوریتم دسته بند
مقایسه با روش‌های پایه	ارزیابی روش پیشنهادی در مقایسه با کارهای پایه
مطالعه موردی رویدادهای توییتر	ارزیابی روش پیشنهادی در هر رویداد بطور مستقل
بررسی کاربر توییت منبع	مشخص کردن الگوهای پرتکرار در گراف کاربران

۴-۱- مجموعه‌دادگان

مجموعه‌دادگان جمع‌آوری شده، مکالمه‌های افراد در پنج رویداد مختلف جهان در توییتر شامل تیراندازی در شارلی هبدو، ناآرامی در فرگوسن، تیراندازی در اوتاوا، گروگان‌گیری در سیدنی، سقوط هواپیمایی ژرمن وینگیز است؛ اندکی پس از آغاز رویداد، فرایند جمع‌آوری

[DOI: 10.52547/jsdp.18.3.45]

توییت‌های هر رویداد از طریق هشتک و لغت‌های کلیدی مربوط به آن شروع شده است. سپس روزنامه‌نگاران توییت‌ها را بررسی کرده و به‌طور گزینشی توییت‌هایی را انتخاب کرده‌اند که دست‌کم صدبار بازتوییت شده باشند. برای توییت‌های انتخاب شده نیز مشخص شده است که آیا توییت‌ها شایعه هستند یا خیر. این مجموعه‌دادگان بخشی از پروژه pheme است که برای عموم[1] در دسترس و حاشیه‌نویسی مکالمه‌ها در مجموعه‌دادگان توسط روزنامه‌نگاران انجام شده است [41].

حاشیه‌نویسی مکالمه‌ها، برای همه رویدادها منجر به مجموعه‌ای شامل ۵۸۰۲ مکالمه شد که در آن ۱۹۷۲ مکالمه مرتبط به مکالمه‌های شایعه و ۳۸۳۰ مکالمه مرتبط به مکالمه‌های غیر شایعه است. همان‌طور که در جدول (۴) نشان داده‌شده است، توزیع‌های مختلفی برای این پنج رویداد وجود دارد. برای مثال در سقوط هواپیمایی ژرمن وینگیز بیشتر از پنجاه درصد مکالمه‌ها درباره شایعه صحبت کرده‌اند؛ ولی تیراندازی در شارلی هبدو درصد کمی از مکالمات درباره شایعات صحبت می‌کنند. در جدول (۴) برای هر رویداد تعداد مکالمه‌های شایعه و غیر شایعه مشخص شده است.

(جدول-۴): تعداد مکالمه‌های شایعه و غیر شایعه برای هر رویداد مشخص شده است.

(Table-4): Number of Rumor and Non-rumor Conversations of Every Event.

رویدادها	شایعات	غیر شایعه
تیراندازی در شارلی هبدو	۴۵۸	۱,۶۲۱
ناآرامی در فرگوسن	۲۸۴	۸۵۹
سقوط هواپیمایی ژرمن وینگیز	۲۳۸	۲۳۱
تیراندازی در اوتاوا	۴۷۰	۴۲۰
گروگان‌گیری در سیدنی	۵۲۲	۶۹۹
مجموع	۱,۹۷۲	۳,۸۳۰

۴-۲- انتخاب ویژگی‌های مهم ساختاری

با توجه به اهمیت ایجاد مدلی که توانایی تشخیص شایعات در گام‌های اولیه شروع مکالمات را داشته باشد، ویژگی‌های ساختاری گراف کاربران و درخت پاسخ در شش ساعت اولیه شروع مکالمه‌ها در رویدادها، استخراج شدند. در این فرایند توییت‌های پاسخی از مکالمه انتخاب شدند که در آن‌ها فاصله زمانی بین ارسال توییت منبع و توییت پاسخ حداکثر شش ساعت باشد. برای هر مکالمه ۲۸ ویژگی ساختاری جدول ۲ مربوط به درخت پاسخ و گراف کاربران از ۵۸۰۲ مکالمه استخراج شد. به‌منظور شناسایی ویژگی‌های مهم ساختاری، فرایند انتخاب ویژگی بر روی ویژگی‌های استخراج‌شده، در سه مرحله انجام شد. در گام نخست، از روش پوشه‌ای جستجوی مستقیم ترتیبی برای انتخاب ویژگی استفاده شد. این روش با مجموعه خالی از ویژگی‌ها شروع می‌شود و در هر مرحله یک ویژگی انتخاب می‌گردد که منجر به افزایش کارایی شود [42]. در گام دوم، از آنجا که در این مجموعه‌دادگان، تعداد مکالمه‌های غیر شایعه بیشتر از مکالمه‌های شایعه است؛ بنابراین مجموعه‌دادگان ما نامتوزان بوده (به‌طورتقریبی تعداد مکالمه‌های غیر شایعه دو برابر مکالمه‌های شایعه است.) و به همین منظور از دسته‌بندهای نظارتی جنگل تصادفی متوازن[2] وEasy Ensemble [43,44] و XGBoost استفاده شد. هایپر پارامتر scale_pos_weight با توجه به نامتعادل‌بودن مجموعه‌دادگان برای XGBoost تنظیم شد. پارامترهای مرتبط به دسته‌بندهای جنگل تصادفی متوزان و Easy Ensemble بر اساس پیش فرض imbalanced-learn[3] تنظیم شدند. در گام سوم، در فرایند آموزش و ارزیابی از اعتبارسنجی متقابل ۵-بخشی استفاده شد. معیار ارزیابی در انتخاب ویژگی‌ها F1-score مربوط به کلاس شایعه است. شکل (۴) نتایج انتخاب ویژگی ساختاری را در بازه زمانی شش ساعت نشان می‌دهد، همان‌طور که در شکل مشاهده می‌شود، دسته‌بند XGBoost در مقایسه با دسته‌بندهای دیگر کارایی بهتری دارد. و در ۲۸ مدل ایجاد شده توسط این دسته‌بند، مدلی با ده ویژگی بهترین کارایی را در مقایسه با مدل‌های دیگر دارد. در جدول (۵) ده ویژگی انتخابی برای دسته‌بند XGBoost در بازه زمانی شش ساعت با علامت * نشان داده شده است. به‌منظور شناسایی الگوهای مهم و متمایز بین مکالمه‌های شایعه و غیر شایعه، علاوه‌بر بازه زمانی شش ساعت، ویژگی‌های ساختاری مکالمه‌های دیتاست در بازه‌های زمانی ۱۲، ۱۸ و ۲۴ ساعت نیز استخراج شدند. در هر بازه زمانی عملیات انتخاب ویژگی مشابه بازه زمانی شش ساعت اجرا شد. نتایج انتخاب ویژگی نشان داد که در همه بازه‌های زمانی دسته‌بند XGBoost کارایی بهتری در مقایسه با دسته‌بندهای دیگر دارد. جدول (۵) نتایج انتخاب ویژگی‌ها در بازه‌های زمانی ۱۲، ۱۸ و ۲۴ ساعت با استفاده دسته‌بند XGBoost را نشان می‌دهد.

در این پژوهش برای بررسی میزان کارایی ویژگی‌های انتخاب شده در بازه‌های زمانی مختلف، نتایج میانگین وزنی F1-score، صحت، **بازخوانی** و AUC[4] (حجم زیر نمودار ROC) با استفاده از دسته‌بند XGBoost برای هر بازه زمانی جداگانه محاسبه شده است. در مجموعه‌دادگان نامتوازن معیار دقت نمی‌تواند معیار مناسبی برای اندازه‌گیری عملکرد مدل باشد. معیار AUC

[1] https://figshare.com/articles/PHEME_dataset_of_rumours_and_non-rumours/4010619

[2] Balanced Random Forest

[3] https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.ensemble.BalancedRandomForestClassifier.html

[4] Area under the ROC Curve

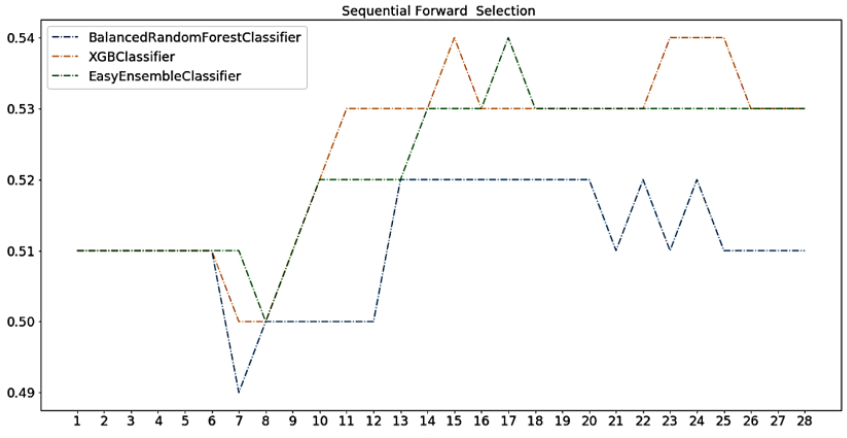
[DOI: 10.52547/jsdp.18.3.45]

در این مجموعه‌دادگان می‌تواند معیار مناسبی برای اندازه‌گیری عملکرد مدل باشد به‌خصوص در رده‌ای که تعداد نمونه‌های آن کمتر است [45].

(جدول-۵): ویژگی‌های انتخابی در بازه‌های زمانی ۶، ۱۲، ۱۸ و ۲۴ ساعت را نشان می‌دهد علامت * نشان می‌دهد که ویژگی در آن بازه زمانی انتخاب شده است.

(Table-5): Show the selected features at 6, 12, 18, and 24-hour intervals. Marks * indicates that the feature is selected at that interval.

نماد ویژگی‌های انتخاب شده	بازه‌های زمانی			
	۶	۱۲	۱۸	۲۴
Num_node_tree	*			
Source_tweet_indegree	*			
Num_leaf_tree		*	*	*
Ave_weigh_tree	*	*		
Len_ path_leaf_root _tree		*		
Num_edge_tree	*	*		
Min_weight_tree	*	*	*	*
Num_node_user	*	*	*	*
Source_user_outdegree	*	*	*	*
max_user_ outdegree		*	*	
Ave_user_outdegree				*
Diameter	*		*	
Ave_weight_user		*		
pattern_chain2			*	
pattern_chain3			*	*
pattern_chain4		*	*	*
pattern_chain5	*	*	*	*
Len_ path_leaf_root _user	*			



(شکل-۴): نتایج انتخاب ویژگی جستجوی مستقیم ترتیبی در بازه زمانی ۶ ساعته را نشان می‌دهد.

(Figure-4): Feature Selection Using the Sequential Forward Selection Method in period 6 hour.

همان‌طور که در جدول (۶) مشاهده می‌شود، معیارهای ارزیابی دسته‌بند XGBoost در بازه‌های زمانی مختلف کارایی نزدیکی دارند و این مؤثربودن ویژگی‌های ساختاری را نشان می‌دهد. این مدل امکان تشخیص زودهنگام شایعه‌ها را در گام‌های اولیه انتشار فراهم می‌کند.

[DOI: 10.52547/jsdp.18.3.45]

[DOI: 10.52547/jsdp.18.3.45]

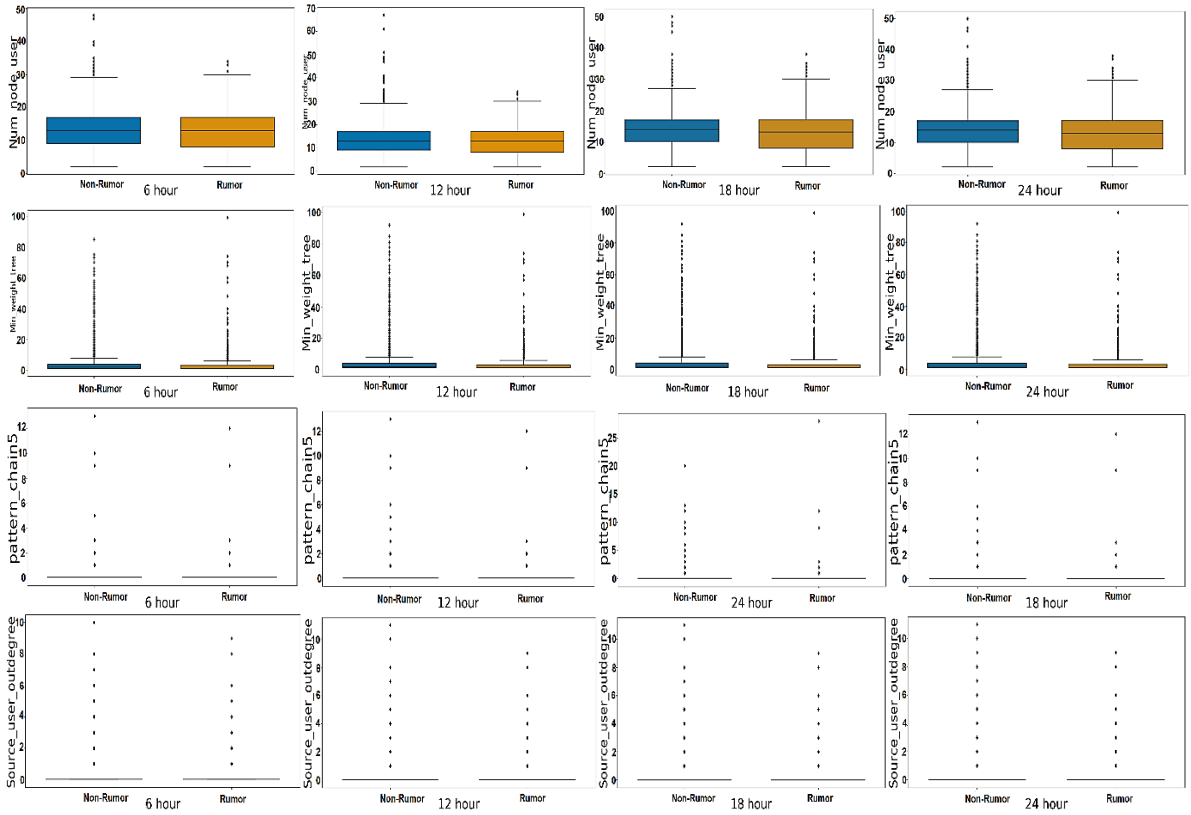
(جدول-۶): معیار ارزیابی F1-score، صحت، بازخوانی و AUC دسته‌بند XGBoost را در بازه‌های زمانی ۶، ۱۲، ۱۸ و ۲۴ ساعت را نشان می‌دهد.

(Table-6): Shows the F1-score, accuracy, recall and AUC of the XGBoost classifier at 6, 12, 18, and 24-hour intervals.

معیار ارزیابی	بازه‌های زمانی			
	۶	۱۲	۱۸	۲۴
صحت	۰/۶۶	۰/۶۷	۰/۶۶	۰/۶۷
بازخوانی	۰/۶۲	۰/۶۳	۰/۶۲	۰/۶۳
f1-score	۰/۶۲	۰/۶۴	۰/۶۳	۰/۶۴
AUC	۰/۶۲	۰/۶۳	۰/۶۳	۰/۶۳

۴-۲-۱- بررسی تفاوت ویژگی‌های مهم ساختاری

در این پژوهش برای نمایش تفاوت‌های ویژگی‌های انتخاب‌شده در جدول (۵) از نمودار جعبه‌ای استفاده شده است. از ویژگی‌های انتخابی، چهار ویژگی کوتاه‌ترین زمان پاسخ، تعداد کاربران، درجه خروجی کاربر توییت منبع و زیرالگوی که در آن فاصله مسیر ارتباطی کاربران با کاربر توییت منبع، برابر با پنج است، بررسی شدند. این ویژگی‌ها تقریباً در همه بازه‌های زمانی ۶، ۱۲، ۱۸ و ۲۴ساعته انتخاب شدند و قدرت بیشتری در تشخیص مکالمه‌های شایعه دارند. شکل (۵) نمودار جعبه‌ای چهار ویژگی انتخابی را نشان می‌دهد. همان‌طوری که مشاهده می‌کنید، کاربران شرکت‌کننده در مکالمه شایعه، واکنش سریع‌تری به توییت منبع اصلی دارند و در بازه زمانی کوتاه‌تری به شایعات پاسخ می‌دهند. همچنین در مقایسه با مکالمه‌های غیر شایعه به‌طور میانگین تعداد کاربران کمتری در مکالمه‌ها شرکت می‌کنند. با بررسی الگوهای زنجیره‌ای ارتباطی کاربران با کاربر توییت منبع، مشخص شد که در مکالمه‌ها شایعه، ارتباط کاربران شرکت کننده در مکالمه بیشتر حول و حوش کاربر توییت منبع است و در مقایسه با مکالمه‌های غیر شایعه، در آن ارتباط بین کاربران بیشتر به‌صورت زنجیره‌ای است. درجه خروجی کاربر توییت منبع در مکالمه‌های شایعه نشان می‌دهد که این کاربر با کاربرهای دیگر تعامل کمتری دارد.



(شکل-۵): نمودار جعبه‌ای چهار ویژگی که هم‌زمان در چهار بازه زمانی ۶، ۱۲، ۱۸ و ۲۴ساعته وجود دارند را نشان می‌دهد. این ویژگی‌ها شامل کوتاه‌ترین زمان پاسخ، تعداد کاربران، درجه خروجی کاربر توییت منبع و زیر الگوی که در آن مسیر ارتباطی کاربران پاسخ‌دهنده به کاربر توییت منبع برابر با ۵ باشد.

(Figure-4): The box diagram shows four features that exist at the same interval. These features include the shortest reply time, number of users, the source Tweet user's output degree, and the sub pattern that links the user to the source tweet user is 5.

۴-۳- مقایسه ویژگی‌های ساختاری با ویژگی‌های زبانی و کاربران

برای بررسی مؤثربودن ویژگی‌های جدید ساختاری، این ویژگی‌ها با ویژگی‌های زبانی و کاربران در تشخیص مکالمه‌های شایعه مقایسه شدند. با بررسی پژوهش‌های قبلی [4,5,20,22,30,33-41] ویژگی‌های زبانی و کاربران از مکالمه‌ها استخراج شدند. ویژگی‌های زبانی از خصوصیات متن توییت‌های مکالمه‌ها استخراج شدند. در این پژوهش با توجه به اهمیت توییت منبع در مکالمه‌ها و نقش متفاوت آن، ویژگی‌های زبانی برای توییت منبع و توییت‌های پاسخ به‌صورت جداگانه محاسبه شده است. در نمایش مقدار ویژگی‌های زبانی توییت منبع، از دو روش دودویی و عددی استفاده شده است. به‌طور نمونه نمایش ویژگی هشتک در توییت منبع با عدد باینری (صفر و یک) نشان داده شده، ولی در نمایش ویژگی طول توییت منبع، از یک عدد استفاده شده است. در نمایش مقدار ویژگی‌های زبانی توییت‌های پاسخ، این ویژگی‌ها در سطح مکالمه جمع شده و به‌صورت بخشی از توییت‌های پاسخ که در مکالمه دارای هشتک هستند، محاسبه (ویژگی‌های دسته‌ای)، ولی در ویژگی‌های پیوسته به‌صورت متوسط، طول توییت‌های پاسخ در مکالمه محاسبه می‌شود. به‌طورکلی در این پژوهش، ویژگی‌های زبانی در دودسته ویژگی‌های وابسته و مستقل از توییتر به‌طور جداگانه برای توییت منبع و توییت‌های پاسخ مکالمه استخراج شده‌اند. در ویژگی‌های وابسته به توییتر، علامت هشتک، منشن و نشانی‌های اینترنتی در توییت‌ها بررسی شده‌اند. در ویژگی‌های مستقل از توییتر، علامت سؤال، علامت تعجب، اختصارات، متوسط پیچیدگی کلمات، طول توییت، ایموجی‌های شاد و ناراحت، استفاده از حروف بزرگ در توییت‌ها، ضمایر اول شخص، دوم شخص و سوم شخص در توییت‌های مکالمه بررسی شده‌اند؛ این علائم نقش مؤثری در پیش‌بینی شایعات دارند [11].

علاوه‌بر استخراج ویژگی‌های زبانی بالا، پژوهش‌های قبلی نشان دادند که شایعات تحت تأثیر احساسات خاص به‌وجود می‌آیند و مردم برای بیان جلب توجه، خشم، خستگی، یا ابراز یک حس خوب، شایعات را در گروه‌ها پخش می‌کنند [46,47]. به همین منظور در این پژوهش توییت‌های منبع و پاسخ ازلحاظ سبک‌های تفکر، نگرانی‌های اجتماعی و مباحث روانشناسی و احساسات مثبت و منفی [40] بررسی شدند. در یکی از موارد بررسی شد که آیا در توییت‌های مکالمه‌ها از کلمات اضطراب (مثل نگرانی و عصبی‌بودن) استفاده شده است یا خیر؟ با توجه به پیچیدگی‌های انجام این کار، از ابزار پرس‌وجوی لغات (LIWC) نسخه ۲۰۱۵ استفاده شد. این دیکشنری شامل ۶۴۰۰ کلمه است [48,49] این کلمات در گروه‌های **اضطراب**، خشم، تفکر[1]، آزمایش‌ها، اطمینان، الفاظ نادرست[2] و ناراحتی دسته‌بندی شده‌اند. بر اساس کلمات متن توییت‌ها، دسته‌بندی هر توییت در یک یا چند گروه بالا مشخص شد. به‌طور نمونه کلماتی مانند فکرکردن، دانستن و سنجیدن در گروه تفکر قرار دارند. اگر کلمات گروه تفکر در توییت منبع باشد، مقدار ویژگی تفکر برای توییت منبع برابر با یک، در غیر این صورت صفر در نظر گرفته شده است؛ ولی محاسبه این ویژگی در توییت‌های پاسخ مکالمه، برابر است با تعداد توییت‌هایی که کلمات گروه تفکر در آن‌ها قرار دارند،بر مجموع تعداد توییت‌های پاسخ مکالمه. قبل از به‌کارگیری این ابزار، عملیات پاک‌سازی به‌وسیله حذف نشانی‌های اینترنتی، علامت ایموجی‌ها، منشن‌ها، اعداد، حذف حروف تکراری در کلمات و مواردی که توسط ابزار قابل خواندن نیست بر روی توییت‌ها انجام شده است. توییت‌های منبع و پاسخ، به‌طور جداگانه در گروه‌های مختلف از طریق دیکشنری LIWC مورد آنالیز قرار گرفتند. درمجموع ۶۶ ویژگی زبانی برای توییت منبع و توییت‌های پاسخ استخراج شده است.

در این پژوهش، ویژگی‌های مرتبط با تاریخچه، اعتبار و نقش کاربران شرکت کننده در مکالمه‌های شایعه از جنبه‌های مختلف بررسی شده است. ویژگی‌های کاربران در دو گروه کاربر توییت منبع و کاربران توییت‌های پاسخ استخراج شده و محاسبه ویژگی‌های کاربران، مشابه ویژگی زبانی مکالمه‌ها است. در یکی از موارد محاسبه ویژگی‌های کاربران مثل ویژگی نفوذ در کاربر توییت منبع، این ویژگی برابر با تعداد دنبال‌کننده‌ها[3] است؛ ولی این ویژگی برای کاربران توییت‌های پاسخ، برابر با میانگین تعداد دنبال‌کننده‌های کاربران است. نحوه محاسبه ویژگی‌های کاربران با ذکر جزئیات به شرح زیر است:

مورد قبول بودن کاربر[4]: معیاری است که نشان می‌دهد توییت‌های یک کاربر تا چه اندازه از طرف دنبال‌کننده‌ها و کاربرهای دیگر قابل قبول و به‌وسیله نرخ تعداد بارهایی که

[1] Insight
[2] Swear words
[3] followers
[4] Likability

[DOI: 10.52547/jsdp.18.3.45]

توییت‌های کاربر دارای محبوبیت[1] است، محاسبه می‌شود [27]:

$$\text{Likability} = \frac{n(favorited)}{n(status)} \quad (۱)$$

نفوذ: نفوذ به وسیله تعداد دنبال کننده‌های یک کاربر محاسبه شده و تعداد دنبال‌کننده‌های بیشتر، نفوذ بیشتر یک کاربر را نشان می‌دهد:

$$\text{Influence} = \#\text{Followers} \quad (۲)$$

اعتبار: اعتبار یک کاربر بر این اساس که حساب کاربری فرد توسط توییتر تأیید شود اندازه‌گیری می‌شود:

$$\text{Credibility} = \begin{cases} 1 & if\ verified \\ 0 & otherwise \end{cases} \quad (۳)$$

نقش کاربر: با استفاده از نسبت نرخ دنبال‌کننده‌ها به تعداد دوستان به دست می‌آید. همان‌طور که در فرمول زیر مشاهده می‌شود، این ویژگی مشخص می‌کند آیا کاربر در توییتر بیشتر ارسال‌کننده اطلاعات است یا دریافت‌کننده و تکرارکننده توییت‌های دیگران [27]:

$$\text{Role} = \frac{\#Followers}{\#Followees} \quad (۴)$$

زمان ثبت نام: به‌وسیله زمان ایجاد حساب کاربری در توییتر محاسبه می‌شود. بعضی از کاربران در شرایط اضطراری برای اشاعه شایعه اقدام به ایجاد حساب کاربری می‌کنند.

تعداد فهرست‌ها: این ویژگی به‌وسیله تعداد فهرست‌هایی که کاربر به آن متعلق است محاسبه می‌شود.

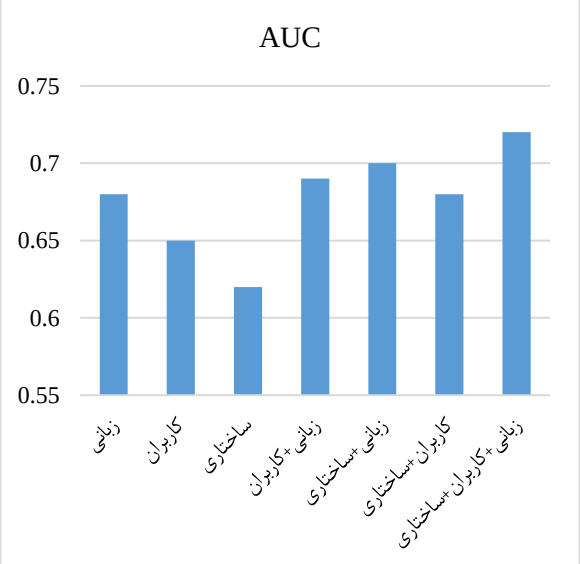
$$\text{numberlist} = \#\text{list} \quad (۵)$$

تعداد دوستان: به‌سادگی توسط تعداد دوستان محاسبه می‌شود.

$$\text{numberfollowees} = \#\text{followees} \quad (۶)$$

علاوه بر ویژگی‌های بالا از پروفایل کاربر ویژگی‌های توضیحات، مکان و نشانی اینترنتی صفحه اصلی استخراج شده‌اند. درمجموع هجده ویژگی برای کاربران توییت منبع و پاسخ در هر مکالمه استخراج شده است.

برای مقایسه ویژگی‌های انتخاب‌شده ساختاری، با ویژگی‌های زبانی و کاربران، این ویژگی‌ها از ۵۸۰۲ مکالمه استخراج شدند. با توجه به اهمیت تشخیص زودهنگام مکالمه‌های شایعه، ویژگی‌ها در شش ساعت اولیه شروع مکالمات استخراج شدند. این ویژگی‌ها شامل ده ویژگی انتخابی از مجموعه ویژگی‌های معرفی شده در جدول (۷)، به همراه ۶۶ ویژگی زبانی و هجده ویژگی کاربران است. فرایند آموزش و ارزیابی مدل بر اساس اعتبارسنجی متقابل ۵- بخشی با استفاده از دسته‌بند XGBoost انجام شد. جدول (۷) نتایج F1-score، صحت و بازخوانی رده‌های شایعه، غیر شایعه برای ویژگی‌های یادشده را نشان می‌دهد. همچنین با توجه به نامتوازن‌بودن مجموعه‌دادگان، میانگین وزنی F1-score و AUC برای رده‌ها محاسبه شدند. همان‌طوری که در جدول (۷) مشاهده می‌شود، ترکیب سه ویژگی، بیشترین کارایی را در مقایسه با ترکیب‌های دیگر دارد. با اضافه‌شدن ویژگی ساختاری به دو ویژگی کاربران و زبانی، مقدار F1-score در دو رده شایعه و غیر شایعه به بیشتر از دو درصد افزایش می‌یابد و این نشان از مؤثربودن ویژگی ساختاری در تشخیص مکالمه‌های شایعه است. در ترکیب‌های دوتایی، دو ویژگی ساختاری و زبانی، کارایی بالاتری در مقایسه با ترکیب‌های دوتایی دیگر دارد؛ ولی در ویژگی‌های انفرادی، ویژگی زبانی در مقایسه با ویژگی‌های دیگر کارایی بالاتری دارد. شکل (۶) مقدار AUC در ویژگی‌های مختلف را نشان می‌دهد. نتایج آن مؤثربودن ویژگی‌های ساختاری در ترکیب‌های مختلف ویژگی‌ها را نشان می‌دهد.



(شکل-۶): نتایج مقایسه معیار ارزیابیAUC ویژگی زبانی، کاربران و ساختاری با استفاده از دسته‌بند XGBoost را نشان می‌دهد.

(Figure-6): Show results Comparison the linguistic, user, and structure features using the XGBoost classifier by the AUCevaluation

[1] favorite

[DOI: 10.52547/jsdp.18.3.45]

(جدول-۷): نتایج مقایسه ویژگی زبانی، کاربران و ساختاری با استفاده از دسته‌بندیXGBoost را نشان می‌دهد؛ که شامل F1-score، صحت، بازخوانی کلاس‌های شایعه، غیر شایعه است و همچنین میانگین وزنی F1-score کلاس‌ها را نشان می‌دهد.

(Table-7): Show results Comparison the linguistic, user, and structure features using the XGBoost classifier. Includes F1-score, precision, recall rumor and non-rumors classes and also shows the weighted average of F1-score and AUC .

ویژگی‌ها	کلاس غیر شایعه			کلاس شایعه			میانگین وزنی دو کلاس
	بازخوانی	صحت	F1-score	بازخوانی	صحت	F1-score	F1-score
زبانی	۰/۶۵	۰/۸۱	۰/۷۲	**۰/۷۱**	۰/۵۲	۰/۶۰	۰/۶۸
کاربران	۰/۶۶	۰/۷۱	۰/۷۱	۰/۶۴	۰/۵۰	۰/۵۶	۰/۶۵
ساختاری	۰/۶۲	۰/۷۵	۰/۶۸	۰/۶۱	۰/۴۶	۰/۵۳	۰/۶۲
زبانی+کاربران	۰/۶۹	۰/۸۱	۰/۷۴	۰/۷۰	۰/۵۴	۰/۶۱	۰/۷۰
زبانی+ساختاری	۰/۷۰	۰/۸۲	۰/۷۵	**۰/۷۱**	۰/۵۵	۰/۶۲	۰/۷۱
کاربران+ساختاری	۰/۶۹	۰/۸۰	۰/۷۴	۰/۶۷	۰/۵۳	۰/۵۹	۰/۶۹
زبانی+کاربران+ساختاری	**۰/۷۲**	**۰/۸۲**	**۰/۷۷**	**۰/۷۱**	**۰/۵۷**	**۰/۶۳**	**۰/۷۱**

۴-۴- مقایسه با روش‌های پایه

در این مرحله مدل پیشنهادی که ترکیبی از ویژگی‌های زبانی، ساختاری و کاربران است با چند روش پایه مبتنی بر مهندسی ویژگی‌ها مورد ارزیابی قرار گرفت.

DT: کاستی لو و همکارانش [11] از دسته‌بند درخت تصمیم برای اعتبارسنجی اطلاعات توییتر استفاده کرده‌اند. مجموع ویژگی‌های مورد استفاده، شامل ویژگی‌های محتوای توییت، کاربر، موضوع و درخت انتشار بازتوییت است.

RF: کووان و همکارانش [50] ویژگی‌های زمانی، ساختاری و متنی مربوط به شایعات را استخراج کردند. با استفاده از الگوریتم جنگل تصادفی شایعات را شناسایی کردند.

SVM: یانگ و همکارانش [13] مشابه روش کاستی لو با استفاده ازماشین بردار پشتیبان با کرنل تابع پایه شعاعی شایعات را شناسایی کردند. آن‌ها ویژگی‌های جدیدی در ویژگی‌های ساختاری، کاربران و محتوای پیام‌ها استخراج کردند.

(جدول-۸): نتایج میانگین وزنی F1-score، صحت، بازخوانی و AUC متد پیشنهادی در مقایسه با روش‌های پایه را نشان می‌دهد.

(Table-8): Show the results weighted average of F1-score, precision and recall of the proposed method in comparison with the basic methods.

روش‌ها / معیارهای ارزیابی	بازخوانی	صحت	F1-score	AUC
DT	۰/۵۶	۰/۵۷	۰/۵۶	۰/۵۲
RF	۰/۶۱	۰/۵۷	۰/۵۸	۰/۵۲
SVM-RBF	۰/۶۵	۰/۴۳	۰/۵۲	۰/۵۰
(زبانی+کاربران+ user-reply-feature)	**۰/۷۲**	**۰/۷۴**	**۰/۷۲**	**۰/۷۱**

فرایند آموزش و ارزیابی در همه مدل‌ها بر اساس اعتبارسنجی متقابل ۵- بخشی انجام شد. ویژگی‌ها در بازه زمانی شش ساعته اولیه از شروع رویدادها از همه مکالمه‌های دیتاست استخراج شد. جدول (۸) نتایج روش‌های پایه با روش پیشنهادی را نشان می‌دهد. روش پیشنهادی، ترکیبی از ویژگی‌های زبانی، کاربران و ساختاری با استفاده از دسته‌بند XGBoost است. مقادیر پررنگ در جدول (۶)، بیشترین معیارهای ارزیابی را نشان می‌دهند. همان‌طورکه مشاهده می‌شود روش پیشنهادی در میانگین وزنی، بازخوانی، صحت، F1-score و AUC مقادیر بیشتری در مقایسه با روش‌های پایه دارد. از دلایل این امر می‌توان به استفاده از ویژگی‌های ساختاری جدید در درخت پاسخ و گراف کاربران اشاره کرد. ویژگی‌های ساختاری در DT، RF و SVM از گراف انتشار بازتوییت استخراج شده است که در مکالمه‌های توییتر کاربردی ندارد. علاوه‌براین در روش پیشنهادی در مقایسه با روش پایه، مجموعه کاملی از ویژگی‌ها زبانی و کاربران استخراج شدند که کارایی دسته‌بند را افزایش می‌دهد. همچنین دسته‌بند XGBoost در مقایسه با دسته‌بندهای درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبان در مجموعه‌دادگان نامتوزان کارایی بهتری دارد.

۴-۵- مطالعه موردی: تشخیص مکالمه‌های شایعه به تفکیک رویدادهای توییتر

با هدف ارزیابی مدل پیشنهادی که ترکیبی از ویژگی‌های زبانی، ساختاری و کاربران است، هر یک از پنج رویداد مربوط به توییتر در جدول (۴) به طور جداگانه در تشخیص مکالمه‌های شایعه ارزیابی شدند. از این‌رو برای

[DOI: 10.52547/jsdp.18.3.45]

هر رویداد، مدلی شامل ویژگی‌های زبانی، ساختاری و کاربران ایجاد شد. همان‌طورکه در جدول (۴) نشان داده شده است، توزیع های مختلفی برای مکالمه‌های شایعه و غیر شایعه در رویدادهای مختلف وجود دارد. در تیراندازی در شارلی هبدو، تعداد مکالمه‌های غیر شایعه به‌طورتقریبی ۳/۵ برابر مکالمه‌های شایعه است، اما در سقوط هواپیمایی ژرمن وینگیز تعداد آنها به‌طورتقریبی برابر است. در تیراندازی در اوتاوا، تعداد مکالمه‌های شایعه بیشتر از مکالمه‌های غیر شایعه است. جدول (۹) نتایج مربوط به ارزیابی مدل پیشنهادی در رویدادها با استفاده از دسته‌بند XGBoost را نشان می دهد. فرایند آموزش و ارزیابی مدل بر اساس اعتبارسنجی متقابل ۵- بخشی انجام شده است. همان‌طور که در جدول (۹) مشاهده می‌شود روش پیشنهادی در میانگین وزنی، بازخوانی، صحت، F1-score در سقوط هواپیمایی ژرمن وینگیز دارای مقدار ۹۷ درصد است؛ و عملکرد خوبی را نشان می‌دهد. عملکرد مدل پیشنهادی در تیراندازی در اوتاوا و گروگان‌گیری در سیدنی دارای کمترین مقدار ۷۲ درصد است. از دلایل متفاوت‌بودن نتایج می توان به توزیع‌های مختلف مکالمه‌های شایعه در رویدادها اشاره کرد و همچنین توییت‌های مکالمه‌های رویدادها در نقاط مختلف جهان درباره موضوع‌های مختلفی نوشته شده است. نتایج در جدول (۹) نشان می‌دهد به‌طورکلی، روش پیشنهادی در تشخیص مکالمه‌های شایعه عملکرد قابل قبولی در رویدادهای مختلف را نشان می‌دهد.

(جدول-۹): نتایج میانگین وزنی F1-score، صحت، بازخوانی متد پیشنهادی برای هر رویداد را نشان می دهد.

(Table-9): Show the results weighted average of F1-score, precision and recall of the proposed method for each event.

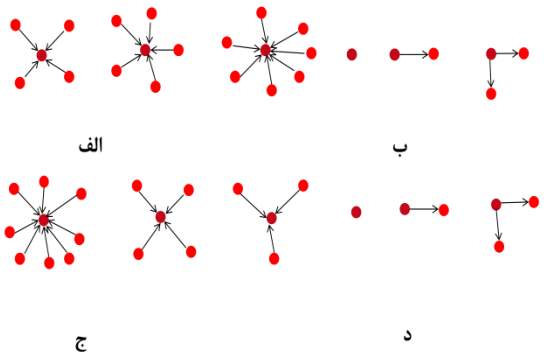
معیارهای ارزیابی / رویدادها	بازخوانی	صحت	F1-score
تیراندازی در شارلی هبدو	۰/۸۰	۰/۸۰	۰/۷۴
ناآرامی در فرگوسن	۰/۷۷	۰/۷۶	۰/۷۱
سقوط هواپیمایی ژرمن وینگیز	**۰/۹۷**	**۰/۹۷**	**۰/۹۷**
تیراندازی در اوتاوا	۰/۷۲	۰/۷۲	۰/۷۲
گروگان‌گیری در سیدنی	۰/۷۲	۰/۷۲	۰/۷۲

۶-۴- بررسی کاربر توییت منبع

با توجه به اهمیت کاربر توییت منبع، الگوهای پرتکرار گراف کاربران مرتبط با کاربر توییت منبع، از مکالمه‌های ۵ رویداد دیتاست استخراج شده است. در شکل (۷)، گره‌ها، کاربران شرکت‌کننده در مکالمه‌ها هستند و یال‌ها، ارتباط بین کاربران را نشان می‌دهد. الگوها و توزیع آن‌ها با استفاده از روش [51,52] استخراج شده‌اند. الگوهای پرتکرار در گراف کاربران بر اساس پاسخ کاربران به کاربر توییت منبع و برعکس بررسی شدند. شکل (۷) سه الگوی پرتکرار در گراف کاربران در مکالمه‌های شایعه و غیر شایعه را نشان می‌دهد که به دو دسته تقسیم می‌شوند.

شبکه‌های ستاره‌ای: در این شبکه‌ها کاربران به‌طور مستقیم با کاربر توییت منبع ارتباط دارند و گفتگوی آن‌ها به‌صورت متمرکز با کاربر توییت منبع است.

پاسخ کاربر توییت منبع: الگوهای شبکه‌های ارتباطی پاسخ کاربر توییت منبع با سایر کاربران را مورد ارزیابی قرار داده است.



(شکل-۷): بخش الف و ب سه الگوی پرتکرار شبکه‌ای ستاره‌ای و پاسخ کاربر توییت منبع به کاربران در مکالمه‌های غیر شایعه را نشان می‌دهد. بخش ج و د سه الگوی پرتکرار شبکه ستاره‌ای و پاسخ کاربر توییت منبع به کاربران در مکالمه‌های شایعه را نمایش می‌دهد.

(Figure-7): Section a and b illustrate the three most frequent star network patterns and the source tweet user's response to other users in non-rumored conversations. Section c and d illustrate the three most frequent patterns of star network and the source tweet user's response to other users in rumored conversations

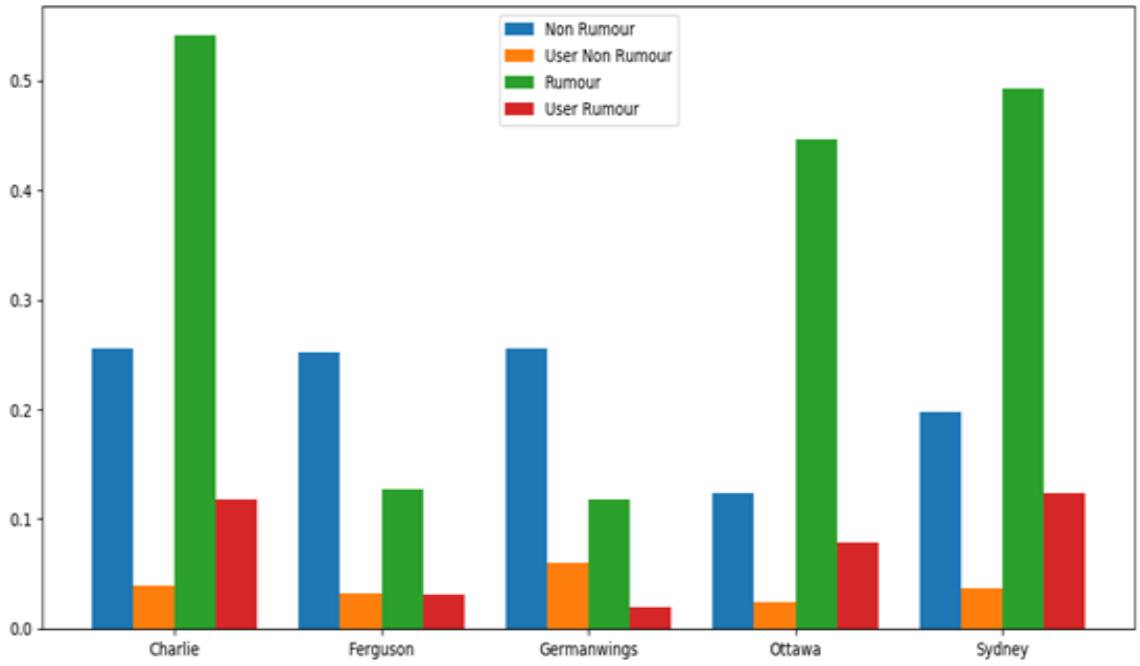
شکل (۷) نشان می‌دهد که تفاوت خاصی بین الگوهای پرتکرار مکالمه‌های شایعه در مقایسه با مکالمه‌های غیر شایعه وجود ندارد و به‌طورتقریبی از الگوهای یکسانی پیروی می‌کنند.

در ادامه بررسی شده است که آیا بیشتر مکالمه‌های شایعه توسط تعداد محدودی کاربر در رویدادهای توییتر نوشته می‌شود یا خیر؟ در شکل (۸)

[DOI: 10.52547/jsdp.18.3.45]

مشاهده می‌شود در چارلی هبدو ۵۴ درصد مکالمه‌های شایعه توسط یازده درصد کاربران نوشته شده است که در آن هر کاربر حداقل شروع‌کننده ۵ مکالمه است. همچنین در مکالمه‌های غیر شایعه در چارلی هبدو ۲۵ درصد مکالمه‌ها توسط سه‌صدم درصد کاربران نوشته شده است.

در بقیه رویدادهای شکل (۸) مشاهده می‌شود، تعداد کمی از کاربران بیشترین تعداد مکالمه‌های شایعه را شروع کرده‌اند. با شناسایی زود هنگام این کاربران می‌توان از انتشار شایعات در رویدادهای توییتر جلوگیری کرد.



(شکل-۸): تعداد مکالمه‌های شایعه و غیر شایعه هر رویداد که توسط کاربران نوشته شده است.

(Figure-8): The number of rumor and non-rumor conversations written by users for each event.

۴-۷- خلاصه آزمایش‌ها

برای بررسی رد و پذیرش فرضیه پژوهش که استخراج ویژگی‌های جدید در گراف‌های مکالمه موجب افزایش دقت تشخیص شایعه‌ها در توییتر می شود، آزمایش‌های زیادی انجام شد. در انتخاب ویژگی‌های جدید ساختاری با استفاده از الگوریتم دسته‌بند نشان داده شد که این ویژگی‌ها در همان گام‌های اولیه در تشخیص مکالمه‌های شایعه مؤثر هستند. با مقایسه ویژگی‌های جدید ساختاری با ویژگی‌های زبانی و کاربران مشخص شد که ترکیب ویژگی‌های جدید با این ویژگی‌ها، عملکرد بهتری در مقایسه با کارهای پایه دارد. همچنین مدل پیشنهادی در مطالعه موردی رویدادهای توییتر عملکرد خوبی دارد. نتایج آزمایش‌ها فرض پژوهش مبتنی بر مؤثربودن ویژگی‌های ساختاری در تشخیص شایعات را تأیید می‌کند. علاوه‌براین آزمایش‌هایی برای شناسایی الگوهای پرتکرار در گراف کاربران انجام گرفت. نتایج نشان داد مکالمه‌های شایعه در رویدادها توسط تعداد محدودی کاربر نوشته می‌شود که با شناسایی زودهنگام این کاربران می‌توان از انتشار شایعات در گام‌های بعدی جلوگیری کرد.

۵- نتیجه‌گیری و کارهای آینده

در مطالعه حاضر ویژگی‌های کامل و جدید ساختاری مربوط به درخت پاسخ و گراف کاربران، در بازه‌های زمانی مختلف در تشخیص مکالمه‌های شایعه مورد مطالعه قرار گرفت. نتایج انتخاب ویژگی‌های ساختاری نشان داد که برخی از ویژگی‌های مکالمه‌های شایعه در بازه‌های زمانی مختلف همیشگی بوده و قدرت بالایی در پیش‌بینی شایعات دارند. هر چند در بازه‌های زمانی مختلف، به‌علت متفاوت‌بودن سرعت انتشار، مجموعه ویژگی‌های انتخابی از همدیگر متفاوت هستند، با این وجود می‌توان تعیین کرد کدام ویژگی‌ها در گام‌های اولیه تشخیص شایعه مؤثر هستند. برای ارزیابی مفیدبودن ویژگی‌های انتخابی ساختاری، آن‌ها با ویژگی‌های زبانی و کاربران مقایسه شده‌اند که نتایج به‌دست‌آمده مؤثربودن آن‌ها را نشان داد. در این پژوهش دسته‌بند مبتنی بر ترکیب ویژگی‌های ساختاری، زبانی و گراف کاربران در مقایسه با روش‌های پایه، کارایی بهتری دارد و در تشخیص شایعه‌های رویدادهای توییتر مؤثر است؛ علاوه‌براین کاربر توییت منبع از جنبه‌های مختلف بررسی شده است، نتایج نشان

[DOI: 10.52547/jsdp.18.3.45]

داد که بیشتر مکالمه‌های شایعه در رویدادها، توسط تعداد محدودی کاربر نوشته می‌شود که با شناسایی زودهنگام این کاربران می‌توان از انتشار شایعات در گام‌های بعدی جلوگیری کرد.

در کارهای آینده، علاوه‌بر بررسی ویژگی‌های درخت پاسخ و گراف کاربران، می‌توان ویژگی‌های گراف منشن مکالمه‌ها را نیز استخراج کرد و آن را از جنبه‌های مختلف مورد بررسی و آنالیز قرار داد به‌طوری‌که موجب افزایش دقت در تشخیص شایعات شود. همچنین می‌توان روش پیشنهادی در این پژوهش را برای تشخیص مکالمه‌های زبان فارسی در توییتر استفاده کرد و روش تشخیص شایعات زبان فارسی [53] را توسعه داد.

۶- مراجع 6- References

[1] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," *in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56-65.

[2] K. Starbird ,L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on the red: what hazards threat reveals about the social life of microblogged information," *in Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 2010, pp. 241-250.

[3] G. Cai, H .Wu, and R. Lv, "Rumors detection in chinese via crowd responses," *in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, 2014, pp. 912-917.

[4] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, pp. 1-36, 2018.

[5] A. Ritter, C. Cherry, and B. Dolan, "Unsupervised modeling of twitter conversations," *in Human Language Technologies: The 201 ٠Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010, pp. 172-180.

[6] J. A. Reshi and R. Ali, "Rumor proliferation and detection in Social Media: A Review," *in 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2019, pp. 1156-1160.

[7] S. M. Alzanin and A. M. Azmi, "Detecting rumors in social media: A survey," *Procedia computer science*, vol. 142, pp. 294-300, 2018.

[8] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38-55, 2019.

[9] J. Cao, J. Guo, X. Li, Z. Jin, H. Guo, and J. Li, "Automatic rumor detection on microblogs: A survey," arXiv preprint arXiv:1807.03505, 2018.

[10] A. Zubiaga ,A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," arXiv preprint arXiv:1704.00656, 2017.

[11] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," *in Proceedings of the 20th international conference on World wide web*, 2011, pp. 675-684.

[12] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in Data Mining (ICDM), 2013 IEEE 13th International Conference on, 2013, 2013, pp. 1103-1108.

[13] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," *in Proceedings of the ACM SIGKDD workshop on mining data semantics*, 2012, pp. 1-7.

[14] G. Cai, H. Wu, and R. Lv, "Rumors detection in Chinese via crowd responses," *in Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, 2014, pp. 912-917.

[15] S. Goel, D. J. Watts, and D. G. Goldstein, "The structure of online diffusion networks," *in Proceedings of the 13th ACM conference on electronic commerce*, 2012, pp. 623-638.

[16] R. Cowan and N. Jonard, "Network structure and the diffusion of knowledge," *Journal of economic Dynamics and Control*, vol. 28, pp. 1557-1575, 2004.

[17] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," *in Proceedings IEEE 24th Annual Joint Conference of the IEEE*

[DOI: 10.52547/jsdp.18.3.45]

[28] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, pp. 1146-1151, 2018.

[29] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PloS one*, vol. 12, p. e0168344, 2017.

[30] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," *in Proceedings of the 1st workshop on privacy and security in online social media*, 2012, pp. 2.

[31] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors on Twitter by promoting information campaigns with generative adversarial learning," *in The World Wide Web Conference*, 2019, pp. 3049-3055.

[32] A. Alsaeedi and M. Al-Sarem, "Detecting Rumors on Social Media Based on a CNN Deep Learning Technique," *Arabian Journal for Science and Engineering*, pp. 1-32, 2020.

[33] S. Santhoshkumar and L. D .Babu, "Earlier detection of rumors in online social networks using certainty-factor-based convolutional neural networks," *Social Network Analysis and Mining*, vol. 10, pp. 1-17, 2020.

[34] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, "Exploring deep neural networks for rumor detection," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-19, 2019.

[35] L. Li, G. Cai, and N. Chen, "A rumor events detection method based on deep bidirectional GRU neural network," *in 2018 IEEE۳ rd International Conference on Image, Vision and Computing (ICIVC)*, 2018, pp. 755-759.

[36] C. Castillo, M. Mendoza, and B. Poblete, "Predicting information credibility in time-sensitive social media," *Internet Research*, vol. 23, pp. 560-588, 2013.

[37] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," PloS one, vol. 12, 2017.

[38] G. Giasemidis, C. Singleton, I. Agrafiotis, J. R. Nurse, A. Pilgrim, C. Willis, et al., "Determining the veracity of rumours on Twitter," *in International Conference on Social Informatics*, 2016, pp. 185-205.

[39] S. Kwon and M. Cha, "Modeling Bursty Temporal Pattern of Rumors," *in ICWSM*, 2014.

Computer and Communications Societies., 2005, pp. 1455-1466.

[18] V. Lampos, T. De Bie, and N. Cristianini, "Flu detector-tracking epidemics on Twitter," *in Joint European conference on machine learning and knowledge discovery in databases*, 2010, pp. 599-602.

[19] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," *Icwsm*, vol. 10, p. 30, 2010.

[20] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, pp. 131-156, 1997.

[21] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?," *IEEE Transactions on information theory*, vol. 57, pp. 5163-5181, 2011.

[22] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of consumer research*, vol. 34, pp. 441-458, 2007.

[23] P. Cogan, M. Andrews, M. Bradonjic, W. S. Kennedy, A .Sala, and G. Tucci, "Reconstruction and analysis of twitter conversation graphs," *in Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, 2012, pp. 25-31.

[24] R. Nishi, T. Takaguchi, K. Oka, T .Maehara, M. Toyoda, K.-i. Kawarabayashi, et al., "Reply trees in twitter: data analysis and branching process models," *Social Network Analysis and Mining*, vol. 6, p. 26, 2016.

[25] M. Mendoza, B. Poblete, and C. Castillo, "Twitter Under Crisis: Can we trust what we RT?," *in Proceedings of the first workshop on social media analytics*, 2010, pp. 71-79.

[26] F. Jin, W. Wang, L. Zhao, E. R. Dougherty, Y. Cao, C.-T. Lu, et al., "Misinformation propagation in the age of twitter," *IEEE Computer*, vol. 47, pp. 90.۲۰۱۴ ,۹۴-

[27] S. Vosoughi, M. N. Mohsenvand, and D. Roy, "Rumor gauge: predicting the veracity of rumors on twitter," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, pp. 50, 2017.

[DOI: 10.52547/jsdp.18.3.45]

[40] A. Zubiaga, M. Liakata, and R. Procter, "Learning reporting dynamics during breaking news for rumour detection in social media," arXiv preprint arXiv:1610.07363, 2016.

[41] A. Zubiaga, M. Liakata, R. Procter, K. Bontcheva, and P. Tolmie, "Crowdsourcing the annotation of rumourous conversations in social media," *in Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 347-353.

[42] F. Ferri, P. Pudil, M. Hatef, and J. Kittler, "Comparative study of techniques for large-scale feature selection," *in Machine Intelligence and Pattern Recognition*. vol. 16, ed: Elsevier, 1994, pp. 403-413.

[43] C. Chen, Andy Liaw, and Leo Breiman, "Using random forest to learn imbalanced data," *University of California, Berkeley 110* pp. 1-12., 2004.

[44] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, pp. 185-197, 2009.

[45] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J Inf Eng Appl*, vol. 3, 2013.

[46] C. R. Sunstein, On rumors: How falsehoods spread, why we believe them, and what can be done: Princeton University Press, 2014.

[47] A. Poulsen, "Why People Gossip and How to Avoid it," 2013.

[48] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," *Annual review of psychology*, vol. 54, pp. 547-577, 2003.

[49] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," 2015.

[50] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," *in 2013 IEEE 13th International Conference on Data Mining*, 2013, pp. 1103-1108.

[51] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Patterns of cascading behavior in large blog graphs," *in Proceedings of the 2007 SIAM international conference on data mining*, 2007, pp. 551-556.

[52] L. Tamine, L. Soulier, L. Ben Jabeur, F. Amblard, C. Hanachi, G. Hubert, et al., "Social media-based collaborative information access: Analysis of online crisis-related twitter conversations," *in Proceedings of the 27th ACM Conference on Hypertext and Social Media*, 2016, pp. 159-168.

[53] Z.Jahanbakhsh-Nagadeh, M.Feizi-Derakhshi, A.Sharifi, "A Model for Detecting of Persian Rumors based on the Analysis of Contextual Features in the Content of Social Networks", *JSDP*, 2021, pp.50-29.

[53] زلیخا جهانبخش نقده ، محمد رضا فیضی درخشی ، آرش شریفی ،" ارائه مدلی برای تشخیص شایعات فارسی مبتنی بر تحلیل ویژگی‌های محتوایی در متن شبکه‌های اجتماعی"، فصلنامه علمی پردازش علائم و داده‌ها، ۱۴۰۰، صفحه ۵۰-۲۹.

سروه لطفی در حال حاضر در مقطع دکترای تخصصی رشته مهندسی کامپیوتر در دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران مشغول به تحصیل است. ایشان از سال ۱۳۸۸ به‌عنوان عضو هیات علمی دانشگاه آزاد اسلامی واحد مریوان هستند. زمینه‌های پژوهشی مورد علاقه ایشان، یادگیری ماشین، یادگیری عمیق و تحلیل شبکه‌های اجتماعی است.

نشانی رایانامه ایشان عبارت است از:

serveh.lotfi@srbiau.ac.ir

میترا میرزارضایی استادیار دانشکده فنی و مهندسی گروه کامپیوتر دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران هستند. زمینه‌های پژوهشی مورد علاقه ایشان، بیوانفورماتیک، شبکه‌های عصبی مصنوعی، یادگیری ماشین، شناسایی الگوها و تجزیه و تحلیل شبکه‌های اجتماعی است.

نشانی رایانامه ایشان عبارت است از:

mirzarezaee@srbiau.ac.ir

مهدی حسین‌زاده دانشیار دانشکده مدیریت و اطلاع‌رسانی پزشکی، دانشگاه علوم پزشکی ایران هستند. زمینه‌های

[DOI: 10.52547/jsdp.18.3.45]

پژوهشی مورد علاقه ایشان، تجارت الکترونیک، آنالیز داده‌های عظیم، مدیریت اطلاعات و تجزیه و تحلیل شبکه‌های اجتماعی است.
نشانی رایانامه ایشان عبارت است از:
hosseinzadeh.m@iums.ac.ir

وحید صیدی استادیار دانشکده فنی و مهندسی گروه کامپیوتر دانشگاه آزاد اسلامی واحد تهران جنوب هستند. زمینه‌های پژوهشی مورد علاقه ایشان، یادگیری عمیق، هوش مصنوعی، یادگیری ماشین و تجزیه و تحلیل شبکه‌های اجتماعی است.
نشانی رایانامه ایشان عبارت است از:
v_seydi@azad.ac.ir

[DOI: 10.52547/jsdp.18.3.45]