تشخیص شایعه در شبکهٔ اجتماعی توییتر با استفاده از ویژگی‌های توییت و کاربر

مسلم صامت عمرانی*[۱]، محمّد صنیعی آباده[۲] و نصرالّله مقدّم چرکری[۳]

[۱]کارشناس ارشد هوش مصنوعی دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه تربیت مدرّس، تهران، ایران

[۲و۳]دانشیار دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه تربیت مدرّس، تهران، ایران

چکیده

با شنیدن هر خبر در شبکه‌های اجتماعی، واکنش‌ها به آن متفاوت است و از زوایای مختلف موجب برانگیخته‌شدن حس کنجکاوی می‌شود. مهم‌ترین بخش آن فهمیدن صحّت‌وسقم خبر است. شایعه، خبری نامعتبر است؛ یعنی هنوز تأیید نشده و ممکن است در صورت نداشتن اعتبار موجب خسارات جبران‌ناپذیری شود؛ ازاین‌رو، تشخیص آن بسیار مهم است. تشخیص شایعه و یا به‌عبارتی مشخص‌کردن اعتبار آن نقش اساسی در جلوگیری از خبر نادرست دارد. در این مقاله، با استفاده از ویژگی‌های جدید دستی مبتنی بر توییت، کاربر و ترکیبی از این دو و با استفاده از چهار دسته‌بند یادگیری ماشین، شایعهٔ موجود در شبکه‌های اجتماعی تشخیص داده شد؛ همچنین با توجه به نامتعادل‌بودن مجموعه‌داده از روش بیش‌نمونه‌برداری استفاده و با توجه به تفاوت ویژگی‌ها از نرمال‌سازی استفاده شده‌است. نتایج نشان داد این روش با وجود سادگی نسبت به روش‌های یادگیری ماشین و عمیق بهبود قابل توجّهی داشته‌است و مقدار صحّت به ۰/۹۹ رسید.

واژگان کلیدی: تشخیص شایعه، یادگیری ماشین، ویژگی کاربر، ویژگی توییت، ویژگی دستی.

Rumor Detection on Twitter using tweet and user features

Moslem Samet Omrani[1*], Mohammad Saniee Abadeh[2] & Nasrollah Moghaddam CHarkari[3]

MA Student of Artificial Intelligent of Faculty of Electrical & Computer Engineering,
[1]Tarbiat Modares University, Tehran, Iran
[2,3]Asscociate Professor of Faculty of Electrical & Computer Engineering,
Tarbiat Modares University, Tehran, Iran

Abstract

When every news item is posted on social media, reactions to it are different and arouse curiosity from different viewpoints. The most important part is to understand the accuracy of the news. A rumor is invalid news, meaning it has not yet been confirmed and it may cause irreparable damage if it is not valid. Therefore, it is very important to detect it. Rumor detection, or in other words, determining its validity, plays an essential role in preventing fake news. Naturally, every phenomenon of normal and anomaly is transmitted to people through social networks. Every News Reactions to that news are different. Depending on the importance of the news, it may be widely covered or it may not have a specific reaction. But if the news spreads widely, it arouses curiosity from different angles. The news is false or true, or the news is valid or invalid. In this work, an attempt was made to identify rumors on social networks by using Hand-Crafted features based on tweets, users and a combination of the two, oversampling and normalization, and by using machine learning classification. Using 4 machine

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۴۰۱/۹/۱۲ • تاریخ پذیرش: ۱۴۰۳/۵/۲۹ • تاریخ انتشار: ۱۴۰۳/۸/۸ • نوع مطالعه: کاربردی

سال ۱۴۰۳ شمارهٔ ۲ پیاپی ۶۰

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

learning classifiers, including Support vector machine, Logistic regression, K-nearest neighbors and Random forest, the two rumors on social networks were detected. Two data sets, PHEME 2017 and PHEME 2018, have been used. The results on these two datasets show that in PHEME 2017, the random forest classifier shows an accuracy of 0.988 using tweet and combination features. Also, these features show a precision of 0.987, which is better than other classifiers used in this work. This classifier has a better recall than other classifiers along with logistic regression with a value of 0.986. Also, this classifier obtained better results with the two mentioned features, with 0.987. In the PHEME 2018 dataset, it obtained the RF classifier with an accuracy of 0.969 using tweet and combination features, and it has better performance in precision, recall and F1. In addition, the user feature in the classifier of k nearest neighbors brings better results than the other two features.

Keywords: **rumor detection, machine learning, user Feature, tweet Feature, Hand-Crafted Feature**

۱-مقدمه

رسانه‌های اجتماعی به ابزار انتشاری مهمی برای اهالی رسانه [۱] و روش اصلی مصرف برای مردم که به‌دنبال آخرین اخبار هستند[۲]، تبدیل شده‌اند. اهالی رسانه ممکن است از رسانه‌های اجتماعی برای گزارش دربارهٔ افکار عمومی در مورد اخبار فوری و حتی برای کشف داستان‌های احتمالی جدید استفاده کنند؛ درحالی‌که شهروندان ممکن است، پیشرفت اخبار و وقایع را از طریق کانال‌های رسمی یا از طریق پست‌های شبکهٔ خود دنبال کنند. درحقیقت، شبکه‌های اجتماعی به‌ویژه به‌دلیل توانایی ذاتی خود در انتشار اخبار، بسیار سریع‌تر از رسانه‌های سنتی فوق‌العاده مفید هستند[۳]؛ باوجوداین، این تأثیر مثبت رسانه‌های اجتماعی با هزینه‌ای روبه‌رو می‌شود؛ برای مثال، کنترل‌نکردن و بررسی واقعیت بر روی پست‌ها، رسانه‌های اجتماعی را زمینه‌ساز باروری برای انتشار اطلاعات غیرقابل تأیید و یا دروغین می‌کند[۴].

افراد اغلب پست‌هایی را از خود یا افراد دیگر تأیید می‌کنند که منبع و اعتبار آن‌ها مورد تأیید نیست. بیشتر اوقات، یک عنوان جذّاب برخلاف یک محتوای احتمالی غیرقابل اثبات یا نادرست، ممکن است، هزاران بار به اشتراک گذاشته شود؛ در این شرایط اگر بخواهیم با استفاده از ویژگی‌های دستی عمل دسته‌بندی را انجام دهیم، باید ارتباط بین توییت اصلی یا منبع و توییت‌های واکنش را مدنظر قرار دهیم. بسیاری از ویژگی‌های مورد استفاده در یادگیری عمیق تفسیرپذیری مناسبی ندارند؛ به‌عبارت ساده‌تر عوامل شایعه‌بودن یا نبودن یک توییت به‌شکل تفسیری بیان نمی‌شوند. در برخی از کارها از تنها ویژگی سری زمانی استفاده شده‌است که به‌تنهایی قادر به دسته‌بندی و تشخیص شایعه نیست. در بسیاری از کارهایی که در حوزهٔ یادگیری ماشین صورت‌گرفته بیشتر از ویژگی‌های خام مورد استفاده در مجموعه‌داده‌ها استفاده شده‌است؛ درصورتی‌که می‌توان از ویژگی‌های مورد استفاده در مجموعه‌داده استفادهٔ بهتری کرد.

در این‌کار قرار است با استفاده از بیش‌نمونه‌برداری و استخراج ویژگی‌های کاربر و توییت، که بسیاری از آن‌ها جدیدند و ترکیبی از این دو و با استفاده از چهار دسته‌بند ماشین بردار پشتیبان، k نزدیک‌ترین همسایه، ترابری رگرسیون و جنگل تصادفی شایعه در شبکهٔ اجتماعی تشخیص داده شود.

در ادامهٔ مقاله بخش دو به مرور کارهای گذشته اختصاص دارد. بخش سه به روش پیشنهادی می‌پردازد. در بخش چهار به آزمایش و مقایسهٔ دسته‌بندهای روش پیشنهادی پرداخته می‌شود. بخش پنج به مقایسه با کارهای قبلی و در نهایت بخش شش به نتیجه‌گیری و کارهای آینده پرداخته می‌شود.

۲- پیشینهٔ پژوهش

هنگامی‌که شروع به صحبت در مورد سازوکارهای تشخیص شایعه می‌کنیم، باید از این واقعیت آگاه باشیم که با وجود اینکه یک حوزهٔ پژوهشی جدید است، راه‌های مختلفی برای تشخیص شایعه در رسانه‌های اجتماعی وجود دارد. همان‌طورکه بسیاری از پژوهش‌گران فکر می‌کنند، تصحیح شایعه می‌تواند به شناسایی یک شایعه و ایجاد آگاهی عمومی کافی برای رفع آن کمک کند، چالش در این رویکرد برای شناسایی اصلاح شایعات در شرایط اضطراری بسیار دشوار است. ویژگی‌های پیام می‌تواند نقش اساسی در تشخیص و تصحیح شایعه داشته باشد. یک ایده، مبتنی بر گراف است. در این‌کار تخمینی از اعتبار کاربر بر اساس تعاملات وی و برداشت از گراف اجتماعی وی در توییت‌هایش به‌دست می‌آید. عمدهٔ تأثیرات در ویژگی‌های این کار مربوط به دنبال‌کنندگان کاربر است. نتایج نشان داده‌است که این کار ارزیابی سریع

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

و دقیقی از اعتبار کاربر ارائه می‌دهد[۵]. برای غلبه بر مقاومت در برابر پاسخ‌های مختلف و آسیب‌پذیری در مقابل حملات گوناگون روشی پیشنهاد شد که از یک شبکهٔ گراف لبهٔ وزن‌دار و یک مولد پاسخ آگاه از موقعیت استفاده شود. با استفاده از رمزگذار مبتنی بر تبدیل‌کننده کلیهٔ نشانه‌های موجود در رشته گفت‌وگو رمزگذاری، همچنین از یک گراف کانولوشن برای جاسازی توکن استفاده و برای غلبه بر عدم‌تشخیص موقعیت، از یک مولد متخاصم آگاه از موقعیت استفاده شد تا تشخیص‌دهنده را با استفاده از اضافه‌کردن یک پاسخ خصمانه به رشته گفت‌وگو آموزش دهد[۶]. در کار دیگر یک مدل دسته‌بندی مبتنی یادگیری ماشین و تجزیه‌وتحلیل N-gram مبتنی بر واژه برای دسته‌بندی خودکار پیام‌های توییتر معتبر و غیرمعتبر معرفی شد. بهترین عملکرد با استفاده از ترکیبی از یونیگرام‌ها و بیگرام‌ها، LSVM[1] به‌عنوان دسته‌بندی‌کننده و TF-IDF[2] به‌عنوان یک راه‌کار استخراج ویژگی به‌دست آمد[۷].

در یک کار با استفاده از مدل‌های یادگیری ماشین و و ویژگی‌های محتوا و کاربر به‌دقّت‌های بسیار خوبی رسیدند؛ برای این کار از پنج ویژگی کاربر، هجده ویژگی محتوا و ۲۸۰ تعبیهٔ واژه برای هر توییت استفاده شد[۸]. در کار دیگر مدلی پیشنهاد شد که به تبدیل «اطلاعات شایعه» می‌پردازد؛ این مدل بر اساس ۳۹ ویژگی (۲۲ ویژگی محتوا، ۱۷ ویژگی کاربر) است. هفت دسته‌بند در این کار استفاده شد[۹]. کار دیگری وجود دارد که در آن از معماری BERT استفاده شد. از BERT برای بازنمایی هر توییت در یک بردار بر اساس معنای متنی جملهٔ آن استفاده و از دسته‌بندهای مختلف برای تشخیص شایعه در توییتر استفاده شد[۱۰].

در کار دیگر مدلی مبتنی بر گراف که برای تشخیص شایعهٔ ناهمگن ساخته شده‌است پیشنهاد شد. نحوهٔ شبیه‌سازی روش‌های مختلف استتار در شبکه‌های اجتماعی برای فرار از آشکارساز شایعه بررسی شد. برای پرداختن به این دو موضوع، استتار احتمالی؛ یعنی انواع حمله با درنظرگرفتن محدودیت‌های دامنه به‌دقّت تعریف و سپس یک چارچوب یادگیری متخاصم مبتنی بر گراف جدید پیشنهاد شد که حملات را قادر می‌سازد تا با اضافه‌کردن آشفتگی‌های عمدی پویا آشکارساز را فریب دهند؛ در همین‌حال، آشکارساز برای یادگیری ویژگی‌های ساختاری متمایزتر برای مقاومت در برابر چنین آشفتگی‌هایی تقویت می‌شود. مهاجم و آشکارساز مکرّر یکدیگر را تقویت می‌کنند[۱۱].

ساختار گراف با پردازش توییت‌ها براساس ساختار مکالمه و نه زمان، بر کاستی‌های معماری‌هایی مثل RNN[3] و CNN[4] غلبه می‌کند. این مدل از سه ماژول تشکیل شده‌است: ماژول بازنمایی توییت[5]، ماژول انتشار مکالمه[6] و ماژول دسته‌بند. TRM[7] اطلاعات سطح بالای یک توییت را می‌گیرد و بازنمایی می‌کند، CPM بازنمایی توییت را از طریق ساختار گراف منتشر می‌کند و CM یک شبکهٔ عصبی عمیق است.؛ از آنجاکه ماژول CPM می‌تواند اطلاعات همسایگان محلی را جمع‌آوری و بازنمایی از گره‌های دیده‌نشده ایجاد کند، این مدل می‌تواند برای کارهای بدون نظارت استفاده شود[۱۲]. کار دیگر با استفاده از RvNN است.

ورودی این مدل درخت انتشار است که ریشهٔ یک پست منبع دارد و هر گره درختی یک پست پاسخ است. معناشناسی محتوای پست و رابطهٔ پاسخ بین پست‌ها از طریق فرایند یادگیری ویژگی به‌طور بازگشتی در امتداد ساختار درختی صورت می‌گیرد. از ویژگی‌های ساختاری برای یادگیری سیگنال‌های شایعه و تقویت بازنمایی‌ها با تجمع بازگشتی سیگنال‌ها در طول شاخه‌های مختلف استفاده می‌شود؛ در این‌کار، دو مدل از پایین به بالا و بالا به پایین استفاده شد[۱۳].

رویکرد تکاملی و یادگیری عمیق اساس کار بعدی بود. این مدل راه‌حل‌هایی برای تشخیص شایعه در سطح متن با توضیحات و دسته‌بندی ژن با تشخیص جهش ارائه می‌دهد و توانایی تمایز بین نمونه‌های واقعی و تولیدشده را حفظ می‌کند، همچنین این مدل نمونه‌های جعلی را تولید می‌کند. تولیدکننده[8] برخلاف کارهای قبلی، یک روش یادگیری تقویتی را به‌کار می‌گیرد. با ترکیب GAN[9] و RL[10] می‌تواند بازنمایی‌های بافتی را با کیفیت بالاتر تولید و آموزش تضاد را متعادل کند. با انتخاب واژگان در جمله بی‌ثباتی آموزش کاهش می‌یابد. مشکل تولید خطای مرحله‌به‌مرحله در GAN با تولید و ارزیابی واژه حل شده‌است[۱۴].

[1] Linear Support Vector Machine
[2] Time Frequency-Inverse Document Frequency

[3] Recurrent Neural Network
[4] Convolutional Neural Network
[5] Tweet Representation Module
[6] Conversation Propagation Module
[7] Classification Module
[8] Generator
[9] Generative adversarial networks
[10] Reinforcement learning

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]
[DOI: 10.61186/jsdp.21.2.15]

به‌منظور استفادهٔ کامل از اطلاعات موجود در رسانه‌های اجتماعی، مدلی پیشنهاد شد که شایعات را بر اساس یک رویداد شناسایی کرد. برای استفادهٔ کامل از اطلاعات توییت و به‌دست‌آوردن یک بازنمایی در سطح بالا، یک معماری سلسله‌مراتبی در نظر گرفته شد. برای تعمیم مدل از داده‌های مخالف بیشتری برای آموزش مدل استفاده شد. چارچوب سلسله‌مراتبی به‌طور مشترک از اطلاعات سطح پست و رویداد استفاده می‌کند؛ برای این‌کار روش یادگیری تخاصمی سلسله‌مراتبی استفاده شد تا مدل پیش‌بینی قوی را تحت تعبیه‌های در سطح پست و رویداد ارائه دهد[۱۵].

مدل بعدی از دو نوع اطلاعات عینی و ذهنی استفاده کرد. اطلاعات ذهنی مربوط به توییت، نظرات و غیره و اطلاعات عینی برگرفته از ویکی‌پدیا و فرهنگ لغت بایدو است؛ همچنین در این‌کار از دو ماژول بازیابی مدارک و تشخیص شایعه استفاده شد؛ در این‌کار اطلاعات عینی به‌عنوان مدرک در نخستین ماژول مورد بازیابی قرارگرفت و سپس در اختیار ماژول دوم قرارگرفت[۱۶]. مدل دیگر به‌صورت پویا حداقل تعداد پست‌های مورد نیاز برای شناسایی یک شایعه را می‌آموزد؛ برای این منظور، یادگیری تقویتی با RNN ادغام شده‌است و پست‌های رسانه‌های اجتماعی در زمان واقعی نظارت می‌شود تا تصمیم گرفته شود چه‌زمانی شایعات دسته‌بندی شود[۱۷]. مدل دیگر براساس یادگیری عمیق با استفاده از یک‌مدل BiLSTM-CNN[1] است؛ کار پیشنهادی شامل اکتساب مجموعه‌داده، پیش‌پردازش، نمایش ویژگی، رمزگذاری ویژگی، استخراج ویژگی و دسته‌بندی است. BiLSTM که لایهٔ متوالی نیز نامیده می‌شود، اطلاعات توالی را در هر دو جهت (به جلو و عقب) حفظ می‌کند؛ درحالی‌که لایهٔ CNN بازنمایی ویژگی‌های تولیدشدهٔ لایهٔ BiLSTM را به تصویر می‌کشد و درنهایت دسته‌بندی می‌شود[۱۸]. مدل یادگیری عمیق دیگر یک مدل به روش GAN پیشنهادی است؛ در این‌کار تولیدکننده[2]های مبتنی بر شبکهٔ عصبی، نمونه‌های آموزشی تولید می‌کنند تا متمایزکنندهٔ شایعه را گیج کند و مجبور شود ویژگی‌های قدرتمندتر را از داده‌های آموزشی تقویت‌شده بیاموزد؛ در واقع، این کار برعکس مدل‌های معمول GAN که تولیدکنندهٔ قوی‌تری را ایجاد می‌کند، تأکید بر روی متمایزکننده‌است[۱۹]. از سازوکار چندتوجهی در تبدیل‌کننده برای مدل‌سازی تعامل از راه‌دور بین توییت‌ها استفاده شد. اطلاعات ساختار درختی در پژوهش مختل می‌شود و وابستگی از نظرات کاربران بین رشته‌های گفت‌وگوی مختلف به‌دست می‌آید. از سازوکار چندتوجهی تبدیل‌کننده برای به‌دست‌آوردن ویژگی‌های هر پست وبلاگ اصلی و نظر بازتوییت، محاسبهٔ هم‌بستگی بین آن‌ها و دادن وزن هم‌بستگی استفاده می‌شود[۲۰].

کار دیگر بر اساس تشخیص شایعه بر استخراج انتشار یا تعاملات داده در میان توییت‌های منبع و بازتوییت یا واکنش‌های بعدی آن‌ها تمرکز دارد؛ با این‌کار، ممکن است پست منبع هیچ جواب و واکنشی نداشته باشد یا واکنش کمی داشته باشد؛ برای ازبین‌بردن این مشکل یک چارچوب جدید تشخیص شایعه مبتنی بر موضوع پیشنهاد شده‌است؛ در ابتدا به‌طور خودکار دسته‌بندی موضوع روی میکروبلاگ‌های منبع انجام شد، سپس بردار موضوع پیش‌بینی‌شده میکروبلاگ‌های منبع با تعبیه‌های کلمهٔ میکروبلاگ‌های منبع برای تشخیص شایعه ترکیب شد[۲۱]. کار دیگر روش‌های با نظارت و بدون نظارت را با هم ترکیب می‌کند. اخبار جدید هشتگ‌های جدید دارند. مسئلهٔ واژگان خارج از واژگان[3] چالش دیگر است؛ هدف این کار آن است که تعبیهٔ واژه را بیاموزد و برای کاهش واژگان خارج از واژگان و قرارگرفتن در حالت بین موضوعی به مدل RNN آموزش دهد. مدل، بازنمایی توزیع‌شده را به موازات یادگیری عمیق یاد می‌گیرد و از ویژگی‌های محتوایی و اجتماعی در این‌کار استفاده شده‌است[۲۲]؛ در این‌کار نویسندگان مقاله از یک سامانهٔ BILSTM سلسله‌مراتبی چندضرر استفاده کردند. آن‌ها دو سطح پست و رویداد را درنظر گرفتند تا در زمان آموزش صرفه‌جویی و از ناپدیدشدن شیب[4] جلوگیری شود. برای جلوگیری از این کاهش دقّت از سازوکار توجّه استفاده شده‌است[۲۳]. در یک کار تنها با تمرکز برروی متن اصلی شایعات فارسی و معرفی ویژگی‌هایی با ارزش اطلاعات محتوایی بالا، مدلی مبتنی بر ویژگی‌های محتوایی فیزیکی و غیرفیزیکی برای تشخیص شایعات فارسی منتشرشده برروی

[1] Bidirectional Long Short Term Memory
[2] Generators
[3] Out of vocabulary
[4] Vanishing Gradient

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]
[DOI: 10.61186/jsdp.21.2.15]

توییتر و تلگرام ارائه می‌کند. مدل پیشنهادی شایعات فارسی مجموعه‌دادهٔ توییتر را با معیار F ۰/۸۴۸، شایعات مجموعه‌دادهٔ زلزله کرمانشاه را با معیار F ۹۵۲/. و شایعات تلگرامی را با معیار F ۸۶۷/. شناسایی کرده‌است[۲۴].

۳-روش پیشنهادی

در این کار سعی بر آن شده تا با استفاده از ویژگی‌های دستی که در مجموعه‌دادهٔ مورد استفاده وجود دارد دسته‌بندی شایعه را انجام دهیم. قبل از پرداختن به چگونگی انجام این کار لازم است با مجموعه‌دادهٔ مورد استفاده و شمای ساده‌ای از روش پیشنهادی آشنا شویم:

۳-۱-مجموعه‌داده

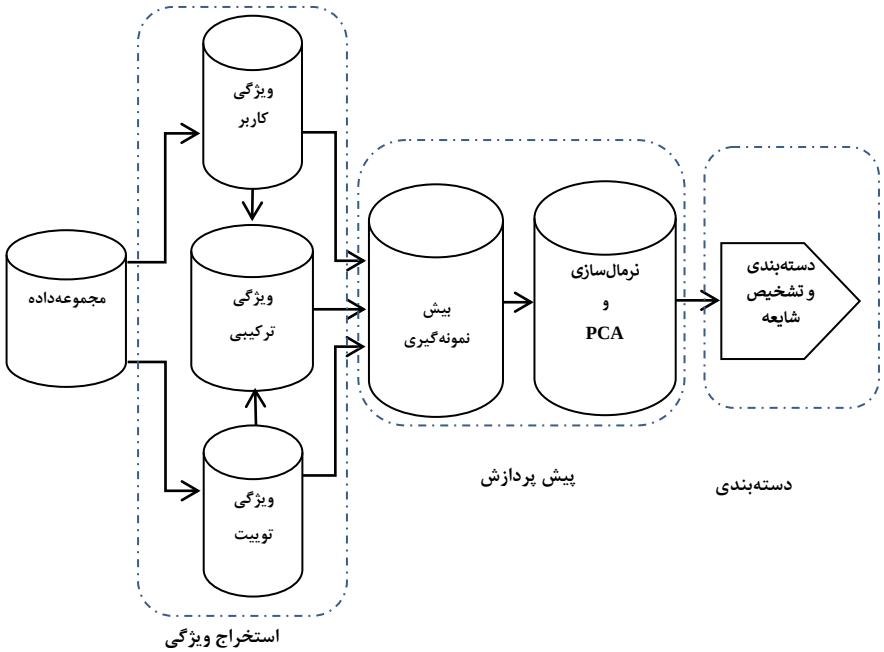
داده‌های دو مجموعه‌دادهٔ شایعهٔ مورداستفاده در این مطالعه از توییت‌هایی که در جریان اخبار فوری ارسال شده‌است مشتق شده‌اند. نام آن‌ها PHEME 2017 و PHEME 2018 [۱۲] است. جدول (۱) آمار این دو مجموعه‌داده را شرح می‌دهد؛ دو مجموعه‌داده حاوی تعداد زیادی ویژگی‌اند که می‌توانند برای مهندسی ویژگی استفاده شوند و برای تشخیص شایعات مفید است.

۳-۲-شرح مدل پیشنهادی

ما در ابتدای این کار از مجموعه‌دادهٔ موجود دو دسته ویژگی مبتنی بر کاربر و توییت را استخراج می‌کنیم؛ سپس اعمال پیش‌پردازش را روی آن انجام می‌دهیم و در آخر روی داده‌ها دسته‌بندی صورت می‌گیرد. این فرایند در شکل (۱) نشان داده شده‌است.

Statistic	PHEME2017	PHEME2018
Users	۴۹۳۴۵	۵۰۵۹۳
Posts	۱۰۳۲۱۲	۱۰۵۳۵۲
Events	۵۸۰۲	۶۴۲۵
Rumor	۱۹۷۲	۲۴۰۲
NonRumor	۳۸۳۰	۴۰۲۳
Balance Degree	٪۳۴	٪۳۷.۴

(جدول- ۱): مجموعه‌داده‌های PHEME
(Table-1): PHEME DataSets



(شکل-۱): نمای کلی از مدل پیشنهادی
(Figure-1): Scheme of Propose model

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]
[DOI: 10.61186/jsdp.21.2.15]

۳-۳-استخراج ویژگی‌ها

۳-۳-۱-ویژگی‌های مبتنی بر توییت

در این قسمت تنها ویژگی‌های جدید در این‌کار معرفی می‌شوند. تعداد واژگان یک رکورد نشان‌دهندهٔ طول آن و ملاکی برای اهمیّت توییت منبع است. اگر بزرگ باشد، بحث‌برانگیزبودن آن را نشان می‌دهد. تعداد نویسه‌های هر توییت، هم طول و هم وجود حروف اضافه و واژگان کوچک تأثیرگذار را نشان می‌دهد. بیشترین تعداد تکرار یک واژه ممکن است نشان‌دهندهٔ مطرح‌بودن آن در شبکه‌های اجتماعی باشد. نسبت بیشترین تعداد یک واژه به تعداد واژگان توییت مشابه حالت قبل است؛ اگرچه کمتر پیش می‌آید واژه در یک توییت تکرار شود. بیشترین تعداد تکرار یک‌واژه به تعداد واژگان کل توییت‌های مربوط به یک رکورد نشان‌دهندهٔ اهمیّت آن واژه در گفت‌وگوهای مربوط به یک رکورد است و در تشخیص شایعه کمک می‌کند. اگر تعداد علاقه‌مندی‌ها به بازتوییت عدد بزرگی باشد، نشانهٔ موافقت و تأیید افراد مختلف با توییت منبع است که به کمک ویژگی‌های دیگر بسیار تعیین‌کننده خواهد بود.

۳-۳-۲-ویژگی‌های مبتنی بر کاربر

یک کاربر ممکن است، در مورد یک رویداد توییت‌های زیادی بزند؛ درصورتی‌که مشخصات کاربری ثابتی دارد. وجود توییت‌های مختلف با واژگان متفاوت و با طول‌های مختلف هم به‌لحاظ معنایی و هم به‌لحاظ متنی مواضع متفاوتی را ایجاد می‌کند؛ از

طرفی ویژگی‌های کاربری به‌نوعی هویت کاربر و فعالیت‌های آن را مشخص می‌کند که ترکیب خصوصیات کاربری منجر به شناخت خصوصیات کاربر می‌شود و از طریق این شناخت می‌توان موضع کاربر را در قبال خبر تعیین کرد.

در مورد ویژگی‌های مبتنی بر توییت در ابتدا عمل پیش‌پردازش اعمال کردیم. همان‌طورکه در جدول (۲) نشان داده شد این کار با حذف عناصر غیرواژهٔ متن از قبیل حذف شکلک، نشانی اینترنتی، نشانی کاربری و غیره انجام شد.

این کار به این علّت انجام شد که بتوانیم ویژگی‌ها را تنها از واژگان متن توییت استخراج کنیم؛ دلیل دیگر اینکه بسیاری از این عناصر ممکن است، موجب تشخیص نادرست دسته‌بندها شود. بسیاری از ویژگی‌های مورد استفاده در این‌کار دستی هستند؛ به‌عبارت دیگر علاوه‌بر ویژگی‌های استخراج‌شده از مجموعه‌داده، ویژگی‌هایی نیز وجود دارند که با رابطهٔ دو یا چند ویژگی دیگر به‌وجود آمدند.

ویژگی‌های توییت جدید به‌کاررفته در این‌کار تعداد واژگان یک رکورد، بیشترین تعداد تکرار واژه، نسبت بیشترین تعداد واژه به تعداد واژگان توییت، بیشترین تعداد تکرار واژه به تعداد واژگان کل توییت‌های مربوط به یک رکورد و تعداد علاقه‌مندی‌ها به بازتوییت هستند.

یک رکورد در اینجا به‌معنی یک توییت منبع و پاسخ‌های مربوط به آن است. تعداد واژگان یک رکورد بیان‌کنندهٔ اهمیّت آن رکورد است و در تشخیص شایعه اهمیّت دارد. بیشترین تعداد تکرار واژه در یک توییت نیز اهمیّت آن واژه را نشان می‌دهد. نسبت بیشترین تعداد واژه به تعداد واژگان یک توییت نیز ویژگی قدرتمندی محسوب می‌شود؛ زیرا پویاتر و منعطف‌تر از ویژگی قبلی است و اگر عدد بزرگ (نزدیک به یک) یا کوچکی (نزدیک به صفر) باشد، اهمیّت آن نسبت به ویژگی قبلی بیشتر است. بیشترین تعداد تکرار واژه به تعداد واژگان کل توییت‌های مربوط به یک رکورد نیز اهمیّت واژه را نسبت به سایر واژگان در سطح رکورد بیان می‌کند و بدون‌شک اگر عددی بزرگ باشد، تأثیر معنایی واژه را در سطح رکورد نشان می‌دهد که این موضوع منجر به قدرت تفکیک‌پذیری بالای آن می‌شود. ویژگی دیگری که به این کار اضافه شده نسبت علاقه‌مندی‌ها به بازتوییت است؛ زمانی‌که این ویژگی عدد بزرگی باشد، نشان می‌دهد که توییت مورد نظر (توییت منبع یا پاسخ) جالب بوده‌است، اما نه در این حد که تعداد بازتوییت آن بالا باشد؛ یعنی اینکه حسّاسیت این توییت خیلی بالا نیست. درصورتی‌که در نقطهٔ مقابل، بالابودن تعداد بازتوییت‌های یک توییت نشان از جلب‌توجّه و حسّاسیت بالای آن دارد.

در طرف دیگر ویژگی‌های کاربری وجود دارند. تعداد علاقه‌مندی‌ها به تعداد دنبال‌کنندگان یک ویژگی جدید است که در این کار مورد استفاده قرارگرفت؛ بزرگ‌بودن این نرخ را می‌توان این‌گونه استنباط کرد که توییت‌های کاربر موردنظر توجّه افراد بیشتری را جلب کرده‌است. دیگر ویژگی مورد استفاده نرخ تعداد فهرست به سن حساب است.

طبیعی است که هرچه این نرخ بزرگ‌تر باشد، میزان فعالیت بیشتر کاربر را نشان می‌دهد؛ همچنین نشان می‌دهد که کاربر وقت زیادی را در شبکه‌های اجتماعی می‌گذراند و در زمینهٔ خاصی تمرکز نمی‌کند.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

(جدول-۲): پیش‌پردازش توییت و فهرست ویژگی‌های توییت و کاربر
(Table-2): preprocessing of Tweet & User features

پیش پردازش ویژگی توییت

- حذف شکلک
- حذف نشانی اینترنتی
- حذف نشانی کاربری
- حذف هشتگ
- حذف نشانه
- حذف تصویر نگاشت
- حذف حروف‌چینی

فهرست ویژگی‌های توییت

- تعداد واژگان هر توییت
- تعداد واژگان یک رکورد*
- تعداد کاراکترهای هر توییت
- بیشترین تعداد تکرار یک واژه*
- نسبت بیشترین تعداد یک واژه به تعداد واژگان توییت *
- بیشترین تعداد تکرار یک واژه به تعداد واژگان کل توییت‌های مربوط به یک رکورد*
- تعداد علاقه‌مندی‌ها
- تعداد بازتوییت
- تعداد علاقه‌مندی‌ها به بازتوییت *

فهرست ویژگی‌های کاربر

- تأیید
- تعداد علاقه‌مندی‌ها به تعداد دوستان*
- تعداد دنبال‌کنندگان به تعداد دوستان
- تعداد دنبال‌کنندگان به سن حساب
- تعداد دوستان به سن حساب
- تعداد فهرست‌ها به سن حساب *
- تعداد استاتوس به سن حساب
- تعداد دنبال‌کنندگان کاربر منبع به متوسط دنبال‌کنندگان کاربران واکنش دهنده*
- تعداد دوستان کاربر منبع به متوسط دوستان کاربران واکنش دهنده*
- تعداد علاقه‌مندی‌های کاربر منبع به متوسط علاقه‌مندی‌های کاربران واکنش دهنده*
- تأیید × تعداد دوستان*
- تأیید × تعداد دنبال‌کنندگان*
- تأیید × تعداد فهرست *
- تأیید × تعداد استاتوس*

تأیید × ویژگی
اگر تأیید =۱ آنگاه ویژگی
اگر تأیید =۰ آنگاه ۰
* به‌معنی ویژگی جدید است

ویژگی بعدی نرخ تعداد دنبال‌کنندگان کاربر توییت منبع به متوسط تعداد دنبال‌کنندگان کاربران واکنش‌دهنده است؛ این ویژگی می‌گوید اگر این نرخ نزدیک به یک باشد، نشان می‌دهد که سطح کاربر توییت منبع به‌احتمال زیاد با کاربران واکنش‌دهنده یکی است؛ در این‌صورت برای مثال اگر کاربر توییت منبع دنبال‌کنندهٔ کمی داشته باشد احتمال بروز شایعه بالاست؛ زیرا می‌خواهد به قصد افزایش دنبال‌کننده شایعه‌پراکنی کند، از طرفی نشان‌دهندهٔ موقعیت اجتماعی پایین‌تر کاربر است؛ از این‌رو، کنار هم قرارگرفتن این ویژگی‌هاست که احتمال شایعه‌بودن یا رد آن را بالا می‌برد.

با همین استدلال ویژگی‌های مشابهی را در مورد دوستان و علاقه‌مندی‌ها نیز می‌توانیم به‌کار ببندیم. در کارهای قبلی یکی از ویژگی‌های پرکاربرد در مورد این مجموعه‌داده ویژگی تأیید بود. در این‌کار ما نیز از این ویژگی استفاده می‌کنیم، امّا این ویژگی را به‌صورت ضریب به‌ویژگی‌های تعداد دوستان، تعداد دنبال‌کنندگان، تعداد فهرست و تعداد وضعیت[1] اضافه می‌کنیم که در جدول (۲) نشان داده شده‌است. چنانچه کاربر مورد تأیید در توییتر باشد در ویژگی موردنظر ضرب و حاصل همان ویژگی می‌شود و چنانچه تأییدنشده نباشد، ویژگی تأیید مقدار صفر می‌گیرد؛ درنتیجه مقدار نهایی صفر می‌شود.

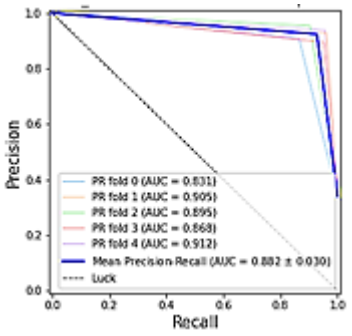
۳-۳-۳-پیش‌پردازش

برای ازبین‌بردن عدم‌تعادل در مجموعه‌داده‌های مورد استفاده در این‌کار در ابتدا از بیش‌نمونه‌گیری استفاده می‌کنیم. راه‌کار استفاده‌شده SMOTE است؛ با اعمال این راه‌کار داده‌های آموزش در هر لایه متعادل خواهد شد، امّا پیش از اینکه از این روش بیش‌نمونه‌گیری استفاده کنیم با داده‌های غیرمتعادل یک‌بار عمل دسته‌بندی را انجام می‌دهیم تا تفاوت قبل و بعد از این روش را نشان دهیم. برای این‌کار از دسته‌بند SVM

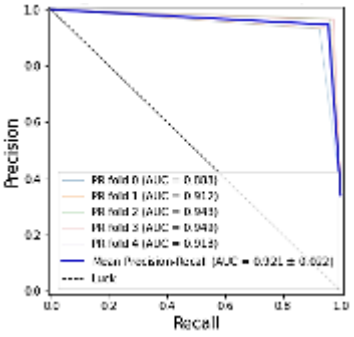
[1] Status

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]
[DOI: 10.61186/jsdp.21.2.15]

استفاده می‌کنیم؛ همان‌طورکه در شکل‌های (۲ و ۳) ملاحظه می‌کنید عمل بیش‌نمونه‌برداری موجب بهبود در Precision-Recall می‌شود. این تغییر در Precision-Recall تقریباً ۴٪ است؛ به‌عبارت‌دیگر فاصلهٔ دقّت و بازخوانی با این‌کار کمتر می‌شود.



(شکل – ۲): دقت- فراخوانی قبل از بیش‌نمونه‌برداری
(Figure-2): Precision-Recall before oversampling



(شکل – ۳) : دقت- فراخوانی بعد از بیش‌نمونه‌برداری
(Figure-3) : Precision-Recall after oversampling

بعد از این مرحله داده‌های موجود را نرمال‌سازی می‌کنیم تا داده‌ها در یک بازه قرارگیرند. در این‌کار با توجه به تنوّع داده‌ها در مجموعه‌داده، بازه‌های متفاوتی مشاهده می‌شود. از طرفی داده‌ها متنوّع‌اند و با این‌کار همه در یک بازه قرارمی‌گیرند؛ این عمل موجب می‌شود که فاصلهٔ داده‌ها از یکدیگر به‌شکل غیرمعمول زیاد نباشد. برای عمل نرمال‌سازی از روش min-max استفاده می‌کنیم تا داده‌ها در بازه صفر و یک قرار گیرند. پس از عمل نرمال‌سازی نوبت به استخراج ویژگی است؛ همان‌طور که درقبل گفته شد، برای این‌کار از PCA استفاده می‌کنیم تا بتوان از ویژگی‌های مفید برای دسته‌بندی استفاده کنیم؛ دلیل دیگر این است که از زمان اجرا کاسته می‌شود. برای این کار بعد از عمل PCA، ۹۵٪ واریانس مؤثّر آن‌را استفاده می‌کنیم.

$$\text{Normalization} = \frac{\text{X}-\text{Min}}{\text{Max}-\text{Min}} \qquad (۱)$$

۳-۳-۴-دسته‌بندی

برای آموزش و ارزیابی دسته‌بندهای ارائه‌شده در این کار مجموعه‌داده را به سه قسمت آموزش، اعتبارسنجی و آزمون تقسیم و از پنج لایه[۱] اعتبارسنجی[۲] متقابل استفاده کردیم تا تعمیم خوبی از نتایج داشته باشیم. در جدول (۳) نحوهٔ تقسیم‌بندی این داده‌ها به مجموعه‌های آموزشی[۳]، اعتبارسنجی[۴] و آزمایشی[۵] را می‌بینید.

(جدول-۳): تقسیم‌بندی داده‌ها قبل از بیش‌نمونه‌گیری و در هر لایه
(Table-3): Devide data before Oversampling in a fold

PHEME 2018	PHEME 2017		
تعداد کل داده‌ها	تعداد کل داده‌ها	درصد	دسته
۴۴۹۸	۴۱۷۸	۷۰	آموزشی
۶۴۲	۵۸۰	۱۰	اعتبارسنجی
۱۲۸۵	۱۱۶۰	۲۰	آزمایشی

۳-۳-۵-معیارهای اندازه‌گیری

معیارهای اندازه‌گیری در این کار صحّت[۶]، دقّت[۷]، فراخوانی[۸] و F1 هستند.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \qquad (۲)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \qquad (۳)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \qquad (۴)$$

$$\text{F1 score} = 2 * \frac{\text{Prec}*\text{Rec}}{\text{Prec}+\text{Rec}} \qquad (۵)$$

TP (True Positive) نمونهٔ مثبت که به‌درستی پیش‌بینی شدند. FP (False Positive) نمونهٔ مثبت که به غلط پیش‌بینی شدند. TN (True Negative) نمونهٔ منفی که به‌درستی پیش‌بینی شدند. FN (False Negative) نمونهٔ منفی که به غلط پیش‌بینی شدند.

ما در این کار از چهار دسته‌بند Support Vector Machine، K Nearest Neighbors، Logistic Regression و Random Forest استفاده کردیم. عمل دسته‌بندی بر روی سه دسته ویژگی‌های مبتنی بر کاربر، مبتنی بر توییت (در بخش‌های قبل توضیح داده شد) و ترکیب این دو صورت می‌گیرد. با توجّه به اینکه ما ترکیب توییت منبع و توییت‌های واکنش مربوط به همان توییت منبع را به‌عنوان یک سطر یا رکورد درنظر می‌گیریم، مشاهده می‌کنیم که درعمل تعداد ویژگی‌های هر دسته افزایش می‌یابد؛ زیرا به تعداد توییت‌های هر رکورد دسته‌های ویژگی وجود دارد؛ برای

[1] Fold
[2] 5 Fold Cross Validation
[3] Training
[4] Validation
[5] Testing
[6] Accuracy
[7] Precision
[8] Recall

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

مثال اگر تعداد ویژگی‌های کاربری پانزده عدد باشد، برای رکوردی با بیست توییت، تعداد مجموع ویژگی‌های کاربری برای آن رکورد به سیصد عدد می‌رسد. افزایش ویژگی می‌تواند یک مزیّت بزرگ محسوب شود. تنوّع و تعدّد ویژگی‌ها عامل بسیار مؤثری برای افزایش کارایی در دسته‌بندها محسوب می‌شود.

۴-مقایسهٔ دسته‌بندهای مقاله

همان‌طورکه در جدول (۴) ملاحظه می‌کنید ما در ابتدا برای مجموعه‌دادهٔ PHEME 2017 به‌ازای هر دسته‌بند نتایج سه ویژگی را با هم مقایسه کردیم؛ در دسته‌بند جنگل تصادفی (Random forest) دستهٔ ویژگی‌های ترکیبی برتری نسبی دارند؛ دلیل برتری ترکیب ویژگی‌ها به دو دستهٔ دیگر را می‌توان علاوه‌بر ویژگی‌ها به تجمیعی‌بودن دسته‌بند جنگل تصادفی نیز نسبت داد.

دسته‌بند بعدی ماشین بردار پشتیبان است. همان‌طورکه مشاهده می‌شود، در حالت مشابه مثل جنگل‌های تصادفی در اینجا نیز بیشترین مقدار متعلّق به ترکیب ویژگی‌هاست. به‌علاوه عمق و تعداد درختان را محدود کردیم تا از پیش‌پردازش جلوگیری شود. اگر دقّت، فراخوانی و F1 برای شایعه و غیرشایعه به‌طور جداگانه محاسبه شوند، آن‌گاه بیشترین میزان دقّت در غیرشایعه متعلّق به دستهٔ ویژگی‌های توییت با ۹۹٪ است. در مورد فراخوانی بیشترین مقدار شایعه متعلّق به ویژگی توییت با ۹۸٪ است. بیشترین مقدار هر دو مورد شایعه و غیرشایعه در F1 متعلّق به ترکیب ویژگی‌ها با ۹۶٪ و ۹۸٪ است. در مقایسهٔ بین دو ویژگی توییت و کاربر، ویژگی کاربر از برتری کوچکی برخوردار است. در مورد غیرشایعه نیز با معیار دقّت، ویژگی‌های توییت تقریباً ۹۹٪ است که بهتر از ترکیب ویژگی‌هاست؛ همچنین این آزمایش‌ها را روی مجموعه‌دادهٔ PHEME 2018 انجام دادیم؛ نتایج نشان دادند که روی این مجموعه‌داده RF با ویژگی توییت، کاربر و ویژگی ترکیبی در هر چهار معیار بهترین عملکرد را دارد.

در بین ویژگی‌ها نیز RF با ویژگی ترکیبی عملکرد بهتری نسبت به سایر ویژگی‌ها دارد؛ زیرا ترکیب ویژگی‌ها باعث افزایش قدرت تفکیک‌پذیری می‌شود. نتایج آزمایش‌ها با این مجموعه‌داده در جدول (۵) نشان داده شده‌است.

(جدول-۴): نتایج آزمایشات به ازای چهاردسته‌بند برای PHEME 2017

(Table-4): Results of experiments for 4 Classifiers for PHEME 2017

RF						LR					
Feature	Class	Acc	Pre	Re	F1	Feature	Class	Acc	Pre	Re	F1
Tweet	All	**۰/۹۸۸**	**۰/۹۸۷**	**۰/۹۸۶**	**۰/۹۸۷**	Tweet	All	۰/۹۶۶	۰/۹۵۷	۰/۹۷۱	۰/۹۶۳
	R		۰/۹۸۲	۰/۹۸۲	۰/۹۸۲		R		۰/۹۱۸	۰/۹۸۷	۰/۹۵۱
	NR		۰/۹۹۰	۰/۹۹۰	**۰/۹۹۰**		NR		۰/۹۹۳	۰/۹۵۵	۰/۹۷۲
User	All	۰/۹۸۲	۰/۹۷۹	۰/۹۸۲	۰/۹۸۰	User	All	۰/۹۷۲	۰/۹۶۶	۰/۹۷۳	۰/۹۶۹
	R		۰/۹۶۸	۰/۹۸۰	۰/۹۷۴		R		۰/۹۴۴	۰/۹۷۶	۰/۹۶۰
	NR		۰/۹۹۰	۰/۹۸۳	۰/۹۸۶		NR		۰/۹۸۷	۰/۹۷۰	۰/۹۷۸
Combine	All	**۰/۹۸۸**	**۰/۹۸۷**	**۰/۹۸۷**	**۰/۹۸۷**	Combine	All	**۰/۹۸۴**	**۰/۹۷۹**	**۰/۹۸۶**	**۰/۹۸۲**
	R		**۰/۹۸۲**	**۰/۹۸۵**	۰/۹۸۴		R		**۰/۹۶۲**	**۰/۹۹۲**	**۰/۹۷۶**
	NR		**۰/۹۹۲**	**۰/۹۸۹**	۰/۹۹۰		NR		**۰/۹۹۶**	**۰/۹۸۰**	**۰/۹۸۷**
SVM						**KNN**					
Feature	Class	Acc	Pre	Re	F1	Feature	Class	Acc	Pre	Re	F1
Tweet	All	۰/۹۰۲	۰/۹۴۲	۰/۹۶۱	۰/۹۵۰	Tweet	All	۰/۹۵۵	۰/۹۴۵	۰/۹۶۰	۰/۹۵۱
	R		۰/۸۹۱	**۰/۹۸۳**	۰/۹۳۵		R		۰/۸۸۲	۰/۹۷۵	۰/۹۲۶
	NR		**۰/۹۹۱**	۰/۹۳۸	۰/۹۶۴		NR		۰/۹۸۶	۰/۹۳۳	۰/۹۵۹
User	All	۰/۹۰۶	۰/۹۵۱	۰/۹۵۰	۰/۹۵۱	User	All	**۰/۹۸۰**	**۰/۹۷۵**	**۰/۹۸۰**	**۰/۹۷۸**
	R		۰/۹۳۵	۰/۹۳۵	۰/۹۳۵		R		**۰/۹۵۹**	**۰/۹۸۳**	**۰/۹۷۰**
	NR		۰/۹۶۶	۰/۹۶۷	۰/۹۶۷		NR		**۰/۹۹۱**	**۰/۹۷۸**	**۰/۹۸۵**
Combine	All	**۰/۹۶۲**	**۰/۹۶۴**	**۰/۹۷۳**	**۰/۹۶۹**	Combine	All	۰/۹۶۱	۰/۹۵۶	۰/۹۵۸	۰/۹۵۷
	R		**۰/۹۴۰**	۰/۹۷۹	**۰/۹۵۹**		R		۰/۹۳۷	۰/۹۵۰	۰/۹۴۴
	NR		۰/۹۸۹	**۰/۹۶۸**	**۰/۹۷۸**		NR		۰/۹۷۴	۰/۹۶۷	۰/۹۷۰

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

(جدول-۵): نتایج آزمایش‌ها به‌ازای چهاردسته‌بند برای PHEME 2018

(Table-5): Results of experiments for 4 Classifiers for PHEME 2018

Feature	Tweet				User				Combine			
Classifier	Acc	Pre	Re	F1	Acc	Pre	Re	F1	Acc	Pre	Re	F1
RF	**۰/۹۶۹**	۰/۹۶۶	۰/۹۶۸	**۰/۹۶۷**	۰/۹۶۷	۰/۹۶۳	۰/۹۶۸	۰/۹۶۵	**۰/۹۶۹**	**۰/۹۶۹**	**۰/۹۷۰**	**۰/۹۶۷**
KNN	۰/۹۰۱	۰/۸۹۱	۰/۸۹۹	۰/۸۹۵	۰/۹۵۹	۰/۹۵۵	۰/۹۵۹	۰/۹۵۷	۰/۹۳۰	۰/۹۲۹	۰/۹۲۸	۰/۹۲۸
SVM	۰/۹۳۶	۰/۹۲۶	۰/۹۴۶	۰/۹۳۳	۰/۹۳۴	۰/۹۲۶	۰/۹۳۴	۰/۹۳۰	۰/۹۴۴	۰/۹۳۵	۰/۹۵۲	۰/۹۴۱
LR	۰/۹۵۷	۰/۹۵۱	۰/۹۰۰	۰/۹۶۸	۰/۹۵۲	۰/۹۴۵	۰/۹۵۶	۰/۹۵۰	۰/۹۵۶	۰/۹۴۸	۰/۹۶۳	۰/۹۵۴

۵-مقایسه با روش‌های یادگیری ماشین

ما در ابتدا سه دسته ویژگی کار خود را با کارهای گذشته مقایسه کردیم؛ در مقایسهٔ ویژگی‌های ترکیبی همان‌طورکه مشاهده می‌شود با معیار صحّت، دسته‌بند RF در [۸] بهترین نتیجهٔ کارهای قبلی را با ۰/۹۴ به‌دست می‌آورد.

در کار پیشنهادی با معیار صحّت هر چهار دسته‌بند بهتر بوده و بیشترین مقدار مربوط به RF با ۰/۹۹ است. در معیار دقّت نیز برتری با کار پیشنهاد شده‌است. در [۸] دقّت شایعه۰/۹۲ مربوط به RF است.

در غیرشایعه ۰/۹۷ و با همین دسته‌بند است. این درحالی‌است که در کار پیشنهادی بیشترین مقدار غیرشایعه مربوط به LR، SVM و RF با ۰/۹۹ با ویژگی‌های توییت و ترکیبی و مقدار شایعه نیز در LR و RF به‌ترتیب به ۰/۹۶ و ۰/۹۸ می‌رسد.

در [۸] در شایعه، RF فراخوانی ۰/۹۶ و در غیرشایعه فراخوانی ۰/۹۲ حاصل می‌شود؛ درصورتی‌که در کار پیشنهادشده در شایعه LR، RF و SVM به‌ترتیب ۰/۹۹، ۰/۹۸ و ۰/۹۸ است. در غیرشایعه نیز RF، LR و SVM فراخوانی‌های ۰/۹۹ ، ۰/۹۸ و ۰/۹۷ را به‌دست آوردند. در مورد F1 نیز در مقالهٔ پایه مقادیر مشابه ۰/۹۴ برای شایعه و غیرشایعه به‌دست آمد. این در حالی‌است که در کار پیشنهادشده برای غیرشایعه و شایعه دسته‌بندهای RF و LR مقادیر مشابه ۰/۹۹ و ۰/۹۸ را به‌دست آوردند. اینکه این دسته‌بند بهترین عملکرد را دارد، به‌تجمیعی و تصادفی‌بودن آن برمی‌گردد و از اولویت‌بندی ویژگی‌ها استفاده و ویژگی‌هایی را استفاده می‌کند که بیشینه قدرت تفکیک‌پذیری را دارد. کاری که ما انجام دادیم تا دقّت بالاتری به‌دست‌آوریم، استفادهٔ بیشتر از ویژگی‌های کاربری است. در کارهای پیشین ویژگی‌های کاربری کمتری استفاده شدند که بیشتر آن‌ها خام‌اند؛ یعنی مستقیم از مجموعه‌داده استخراج شده‌است؛ این درحالی‌است که ما از چهارده‌ویژگی کاربری در کار خود استفاده کردیم؛ ازاین‌رو، وجود ویژگی‌های بیشتر می‌توانند مکمّل‌های خوبی باشند و قدرت تفکیک‌پذیری را افزایش دهند. علاوه‌براین ما در هر دو دسته ویژگی‌های توییت و کاربر ویژگی‌هایی ساختیم که ارتباط معناداری بین کاربر توییت منبع و کاربران توییت‌های واکنش برقرار می‌شود. در مقایسهٔ دیگر، روش پیشنهادشده را با درنظرگرفتن هر دو کلاس با کارهای گذشته مقایسه کردیم. همان‌طورکه ملاحظه می‌شود کار پیشنهادی دارای برتری نسبت به کارهای پیشین است. در کار [۷] دسته‌بند LR با کمک ویژگی TF-IDF به‌دقّت ۰/۸۶۲ دست‌یافت که در مقایسه با کار روش پیشنهادی با دقّت ۰/۹۸۷ برای ویژگی‌های توییت و ترکیبی با دسته‌بند RF بسیار پایین‌تر است. البته هر چهاردسته‌بند عملکرد بسیار خوبی دارند. در مورد فراخوانی، دسته‌بند KNN و با ویژگی TF برتری با [۷] است و مقدار ۰/۹۹۸ به‌دست آمد. این در حالی‌است‌که در کار [۹]، SVM با ویژگی کاربر، LR با ویژگی کاربر، SVM با ویژگی محتوایی و LR با ویژگی محتوایی به‌ترتیب با ۰/۹۹۶، ۰/۹۷۳، ۰/۹۶۸ و ۰/۹۶۳ عملکرد بسیار خوبی داشت. بهترین مقدار فراخوانی مربوط به روش پیشنهادی مربوط به RF با ویژگی توییت با مقدار ۰/۹۸۶، RF با ترکیب ویژگی‌ها با مقدار ۰/۹۸۷، LR با ترکیب ویژگی‌ها با مقدار ۰/۹۸۶، SVM با ترکیب ویژگی‌ها با مقدار ۰/۹۷۳ و KNN با ویژگی کاربر با مقدار ۰/۹۸۰ است که عملکرد بهتری را در مجموع به همراه داشتند. بهترین نتایج ویژگی TF در [۷] با معیار فراخوانی به‌دست آمده‌است؛ در مقابل، فراخوانی هر سه دسته ویژگی در کار پیشنهادی عملکرد بهتری دارند.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

(جدول-۶): نتایج کار یادگیری ماشین ۱ با PHEME 2017
(Table-6): Results with machine learning 1 with PHEME 2017

PHEME 2017				
Method	Acc	Pr	Re	F1
LR+TF	۰/۸۴۱	۰/۸۴۶	۰/۹۲۸	۰/۸۸۵
LR+TF-IDF	۰/۸۴۶	۰/۸۶۲	۰/۹۰۷	۰/۸۸۶
RF+TF	۰/۸۲۴	۰/۸۲۸	۰/۹۲۷	۰/۸۷۴
RF+TF-IDF	۰/۸۲۰	۰/۸۴۱	۰/۸۹۸	۰/۸۶۹
KNN+TF	۰/۶۸۵	۰/۶۷۸	۰/۹۹۸	۰/۸۰۷
KNN+TF-IDF	۰/۸۲۵	۰/۸۳۴	۰/۹۱۸	۰/۸۷۴

(جدول-۷): نتایج یادگیری ماشین ۲ با PHEME 2017
(Table-7): Results of machine learning 2 with PHEME 2017

PHEME 2017					
Method	Acc	class	Pr	Re	F1
RF (C+U+WE)	۰/۹۴	NR	۰/۹۷	۰/۹۶	۰/۹۴
		R	۰/۹۲	۰/۹۲	۰/۹۴
KNN(C+U+WE)	۰/۸۲	NR	۰/۹۶	۰/۹۷	۰/۸۴
		R	۰/۷۴	۰/۷۰	۰/۸۰
SVM (C+U+WE)	۰/۸۰	NR	۰/۸۳	۰/۸۳	۰/۸۰
		R	۰/۷۶	۰/۷۶	۰/۸۰
LR (C+U+WE)	۰/۸۱	NR	۰/۸۴	۰/۸۳	۰/۸۳
		R	۰/۷۹	۰/۸۰	۰/۷۹

البته در [۷] فراخوانی بهتر از دقّت بود و این نشان می‌دهد که هر دو ویژگی در شناسایی نمونه‌های درست عملکرد خوبی داشتند، امّا در کار پیشنهادی این دو معیار تقریباً در یک سطح قرار دارند.

(جدول-۸): نتایج یادگیری ماشین ۳ با PHEME 2017
(Table-8): Results of machine learning 3 PHEME 2017

PHEME 2017				
Method	Acc	Pr	Re	F1
RF+Content	۰/۶۷۷	۰/۷۱۰	۰/۸۲۳	۰/۷۹۶
RF+User	۰/۸۲۲	۰/۸۲۶	۰/۹۲۴	۰/۸۵۲
RF+Combine	۰/۸۳۴	۰/۸۳۱	۰/۹۵۶	۰/۸۸۶
KNN+Content	۰/۶۹۱	۰/۷۱۲	۰/۸۵۲	۰/۸۰۳
KNN+User	۰/۷۳۸	۰/۷۹۵	۰/۹۹۶	۰/۸۷۶
KNN+Combine	۰/۷۱۳	۰/۷۹۳	۰/۸۴۵	۰/۸۲۳
SVM+Content	۰/۷۰۸	۰/۷۱۱	۰/۹۶۸	۰/۸۳۰
SVM+User	۰/۷۱۴	۰/۶۹۷	۰/۹۹۶	۰/۸۰۵
SVM+Combine	۰/۷۳۵	۰/۷۱۲	۰/۹۳۴	۰/۸۱۳
LR+Content	۰/۷۳۲	۰/۷۰۸	۰/۹۶۳	۰/۸۲۴
LR+User	۰/۶۹۱	۰/۶۹۴	۰/۹۷۳	۰/۸۰۶
LR+Combine	۰/۷۲۹	۰/۷۰۳	۰/۹۳۰	۰/۸۰۰

در مورد F1 نیز بهترین نتیجهٔ کارهای قبلی مربوط به کار [۷] و دسته‌بند LR و ویژگی TF-IDF و همچنین کار [۹] و دسته‌بند RF و ترکیب ویژگی‌ها با مقدار مشابه ۰/۸۸۶ است؛ درحالی‌که در روش پیشنهادی بهترین نتایج مربوط به دسته‌بند RF با ویژگی‌های توییت و ترکیبی و با مقدار مشابه ۰/۹۸۷، LR با ترکیب ویژگی‌ها با ۰/۹۸۲ و KNN با ویژگی کاربر با ۰/۹۷۸ است که نشان‌دهندهٔ برتری کار پیشنهادی است. نتایج کارهای پیشین در زمینهٔ یادگیری ماشین کامل در جداول (۶، ۷ و ۸) آورده شده‌است.

۵-۱-مقایسه با روش‌های یادگیری عمیق

در این قسمت نیز روش پیشنهادی را در دو حالت دو کلاس جداگانه و با هم با روش‌های قبلی مقایسه می‌کنیم. در کار پیشنهادی و با مجموعه‌دادهٔ PHEME 2017 معیار صحّت هر چهاردسته‌بند بهتر بوده و بیشترین مقدار مربوط به RF با ۰/۹۹ است. در کارهای مورد مقایسه بهترین نتیجه مربوط به [۱۷] با ۰/۹۵۷ است. در دقّت و در مورد غیرشایعه [۱۱] به مقدار ۰/۹۳۱ رسید؛ همچنین [۱۰] نیز به ۰/۸۹۰ رسید، امّا در روش پیشنهادی و در مورد غیرشایعه مربوط به LR، SVM و RF با ۰/۹۹۰ و با ویژگی‌های توییت و ترکیبی است. در مورد شایعه نیز [۶] با مقدار ۰/۸۳۷ بهترین نتیجهٔ کارهای قبلی را دارد.

(جدول-۹): نتایج کار یادگیری عمیق ۱ برای PHEME 2017
(Table-9): Results with deep learning 1 with PHEME 2017

PHEME 2017					
Method	Acc	Class	Pr	Re	F1
Alkhodair Model	-----	NR	۰/۸۳۳	۰/۸۴۷	۰/۸۳۹
		R	۰/۷۲۶	۰/۷۰۶	۰/۷۱۶
TDRD (CNN_text_ predicted topics)	۰/۸۲۶	NR	۰/۸۳۱	۰/۹۲۵	۰/۸۷۵
		R	۰/۸۱۳	۰/۶۳۵	۰/۷۱۳
GAN-GRU	۰/۷۸۱	NR	۰/۷۹۱	۰/۷۶۶	۰/۷۷۸
		R	۰/۷۷۳	۰/۷۹۶	۰/۷۸۴
CGAT	۰/۸۹۳	NR	**۰/۹۳۱**	۰/۹۰۳	۰/۹۱۷
		R	۰/۸۲۳	۰/۸۷۱	۰/۸۴۶
PARG	۰/۸۴۸	NR	۰/۸۶۳	۰/۸۲۹	۰/۸۴۵
		R	۰/۸۳۷	۰/۸۶۷	۰/۸۵۱
TD-RvNN-GA	۰/۸۵۵	NR	۰/۸۶۸	۰/۹۲۲	۰/۸۹۴
		R	۰/۸۲۱	۰/۸۱۹	۰/۷۶۷
MLP+BERT	۰/۸۴۵	NR	۰/۸۹۰	۰/۸۷۶	۰/۸۸۳
		R	۰/۷۵۸	۰/۷۸۳	۰/۷۷۱

در روش پیشنهادی مقدار شایعه نیز LR، KNN و RF به‌ترتیب به ۰/۹۶۲، ۰/۹۵۹ و۰/۹۸۲ می‌رسد. در مورد فراخوانی و غیرشایعه نتیجه هستند؛ درحالی‌که در روش پیشنهادی RF با ویژگی‌های توییت و ترکیبی به‌ترتیب به مقادیر ۰/۹۹۰ و ۰/۹۸۹ دست یافت. در مورد F1 نیز در غیرشایعه بهترین نتیجه در کارهای قبلی مربوط به [۱۱] با ۰/۹۱۷ و [۱۰] با ۰/۸۹۴ است؛ درحالی‌که در روش پیشنهادی RF با ویژگی‌های توییت و ترکیبی به‌دقّت مشابه ۰/۹۹۰ رسید. در مورد شایعه نیز در معیار دقّت

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]
[DOI: 10.61186/jsdp.21.2.15]

بهترین نتیجه در کارهای قبلی به [۶] با ۰/۸۳۷ اختصاص دارد. در روش پیشنهادی نیز RFبا ویژگی‌های توییت و ترکیبی به مقادیر مشابه ۰/۹۸۲ دست یافت و عملکرد بهتری داشت. در مورد فراخوانی نیز [۱۱] با ۰/۸۷۱ بهتر عمل کرده‌است. در مقابل در روش پیشنهادی LR با ویژگی ترکیبی به ۰/۹۹۲ رسید. در مورد F1 نیز [۱۱] مقدار ۰/۸۴۶ را کسب کرد؛ درحالی‌که در روش پیشنهادی بازهم RF با ویژگی‌های توییت و ترکیبی به ۰/۹۸۲ و ۰/۹۸۴ رسید، امّا اگر با معیارهای گفته‌شده هر دو کلاس را در نظر بگیریم [۱۶] در صحّت، دقّت، فراخوانی و F1 به مقادیر ۰/۹۵۷، ۰/۹۵۰، ۰/۹۶۳ و ۰/۹۵۷ دست یافت. در این مورد در روش پیشنهادی RF با ویژگی توییت و ترکیبی با معیار صحّت به مقدار مشابه ۰/۹۸۸، با معیار دقّت به ۰/۹۸۷ و فراخوانی به مقادیر ۰/۹۸۶ و ۰/۹۸۷ و F1 به مقدار ۰/۹۸۷ دست‌یافت. در مورد مجموعه‌دادهٔ PHEME 2018 نیز [۱۵] با معیارهای صحّت، دقّت، فراخوانی و F1 به مقادیر ۰/۹۳۷، ۰/۹۳۲، ۰/۹۳۶ و ۰/۹۳۴ به بهترین نتیجه از بین کارهای مورد مقایسه دست‌یافت؛ همچنین روش [۱۶] بعد از آن بهترین نتیجه را داشت. روش پیشنهادی نیز با RF و با صحّت، دقّت، فراخوانی و F1 به‌ترتیب ۰/۹۶۹، ۰/۹۶۹، ۰/۹۷۰ و ۰/۹۶۷ عملکرد بهتری داشته‌است. نتایج کارهای پیشین در زمینهٔ یادگیری عمیق کامل در جداول (۹ و ۱۰) آورده شده‌است.

(جدول-۱۰): نتایج کار یادگیری عمیق ۲ برای PHEME 2017و2018

(Table-10): Results with deep learning 1 with PHEME 2017& 2018

PHEME 2017				
Method	Acc	Pr	Re	F1
MHA	۰/۹۲۶	۰/۸۳۴	۰/۹۵۶	۰/۸۹۱
BiLSTM-CNN+ task-specifc WE	۰/۸۶۶	۰/۸۶۰	۰/۸۶۰	۰/۸۶۰
RDM	**۰/۹۵۷**	**۰/۹۵۰**	**۰/۹۶۳**	**۰/۹۵۷**
LOSIRD	۰/۹۱۴	۰/۹۱۵	۰/۹۰۰	۰/۹۰۶
HAT-RD	۰/۹۲۵	۰/۹۲۵	۰/۹۱۱	۰/۹۱۷
GMD-CNN	۰/۸۴۷	-----	-----	۰/۸۴۸
CSRD	۰/۹۰۰	۰/۸۹۳	۰/۸۶۹	۰/۸۸۱
PHEME 2018				
Method	Acc	Pr	Re	F1
MHA	۰/۹۱۹	۰/۸۹۲	۰/۹۲۳	۰/۹۰۷
LOSIRD	۰/۹۲۵	۰/۹۲۲	۰/۹۲۴	۰/۹۲۳
HAT-RD	**۰/۹۳۷**	**۰/۹۳۲**	**۰/۹۳۶**	**۰/۹۳۴**
GMD-CNN	۰/۸۰۸	-----	-----	۰/۸۰۹
CSRD	۰/۹۱۹	۰/۸۹۲	۰/۹۲۳	۰/۹۰۷

۶- نتیجه‌گیری و کارهای آینده

در این کار ما با استفاده از یادگیری ماشین و استفاده از چهاردسته‌بند و با استفاده از سه‌دسته ویژگی توییت، کاربر و ترکیب این دو عمل تشخیص شایعه را روی دو مجموعه‌داده انجام دادیم. ویژگی‌ها در این‌کار یا مستقیم از مجموعه‌داده استخراج و یا با استفاده از روابط دستی ساخته شدند. در آینده می‌توان برای حل این مسئله از روش‌های سری زمانی به همراه یادگیری عمیق و یا شبکه‌های پیچیدهٔ پویا استفاده کرد.

۷- مراجع 7-References

[1] N. Diakopoulos, M. De Choudhury, and M. Naaman, "Finding and assessing social media information sources in the context of journalism," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 2451-2460.

[2]A. Hermida, "Twittering the news: The emergence of ambient journalism," *Journalism practice*, vol. 4, no. 3, pp. 297-308, 2010.

[3] S. Vieweg, "Microblogged contributions to the emergency arena: Discovery, interpretation and implications," *Computer Supported Collaborative Work*, pp. 515-516, 2010.

[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, no. 2, pp. 1-36, 2018.

[5] H. Slimi, I. Bounhas, and Y. Slimani, "TWITTER USERS CREDIBLITY EVALUATION BASED ON SOCIAL GRAPH IMPRESSION," in *Proceedings of the 16th International Conference on Applied Computing (AC), Cagliari, Italy*, 2019, pp. 11-18.

[6] Y.-Z. Song, Y.-S. Chen, Y.-T. Chang, S.-Y. Weng, and H.-H. Shuai, "Adversary-Aware Rumor Detection," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 1371-1382.

[7] N. Hassan, W. Gomaa, G. Khoriba, and M. Haggag, "Credibility detection in twitter using word n-gram analysis and supervised machine learning techniques," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 1, pp. 291-300, 2020.

[8] M. R. I. Nawab, K. M. Shahiduzzaman, T. Eng, and M. N. Jamal, "Rumor Detection in Social Media with User Information Protection," *European Journal of Electrical Engineering and Computer Science*, vol. 4, no. 4, pp. 1-8, 2020.

[9] M. Azer, M. Taha, H. H. Zayed, and M. Gadallah, "Credibility Detection on Twitter News Using Machine Learning Approach," *International Journal of Intelligent Systems and Applications*, vol. 13, no. 3, pp. 1-10, 2021.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]

[10] R. Anggrainingsih, G. M. Hassan, and A. Datta, "BERT based classification system for detecting rumours on Twitter," *arXiv preprint arXiv:2109.02975*, 2021.

[11] X. Yang, Y. Lyu, T. Tian, Y. Liu, Y. Liu, and X. Zhang, "Rumor detection on social media with graph structured adversarial learning," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 1417-1423.

[12] J. Li, Y. Sujana, and H.-Y. Kao, "Exploiting microblog conversation structures to detect rumors," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 5420-5429.

[13] J. Ma, W. Gao, S. Joty, and K.-F. Wong, "An attention-based rumor detection model with tree-structured recursive neural networks," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 4, pp. 1-28, 2020.

[14] M. Cheng, Y. Li, S. Nazarian, and P. Bogdan, "From rumor to genetic mutation detection with explanations: a GAN approach," *Scientific Reports*, vol. 11, no. 1, pp. 1-14, 2021.

[15] S. Ni, J. Li, and H.-Y. Kao, "Rumor Detection on Social Media with Hierarchical Adversarial Training," *arXiv preprint arXiv:2110.00425*, 2021.

[16] J. Li, S. Ni, and H.-Y. Kao, "Meet the truth: Leverage objective facts and subjective views for interpretable rumor detection," *arXiv preprint arXiv:2107.10747*, 2021.

[17] K. Zhou, C. Shu, B. Li, and J. H. Lau, "Early rumour detection," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 1614-1623.

[18] 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*, vol. 12, no. 4, pp. 4315-4333, 2021.

[19] 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.

[20] L. M. S. Khoo, H. L. Chieu, Z. Qian, and J. Jiang, "Interpretable rumor detection in microblogs by attending to user interactions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 05, pp. 8783-8790.

[21] F. Xu, V. S. Sheng, and M. Wang, "Near real-time topic-driven rumor detection in source microblogs," *Knowledge-Based Systems*, vol. 207, p. 106391, 2020.

[22] S. A. Alkhodair, S. H. Ding, B. C. Fung, and J. Liu, "Detecting breaking news rumors of emerging topics in social media," *Information Processing & Management*, vol. 57, no. 2, p. 102018, 2020.

[23] Y. Sujana, J. Li, and H.-Y. Kao, "Rumor detection on Twitter using multiloss hierarchical BiLSTM with an attenuation factor," *arXiv preprint arXiv:2011.00259*, 2020.

[۲۴] جهانبخش نقده زلیخا، فیضی درخشی محمد رضا، شریفی آرش. «ارائه مدلی برای تشخیص شایعات فارسی مبتنی بر تحلیل ویژگی‌های محتوایی در متن شبکه‌های اجتماعی»، *پردازش علائم و داده‌ها*، ۱۴۰۰؛ ۱۸ (۱) :۲۹-۵۰

[24] Z. Jahanbakhsh-Nagadeh, M.-R. Feizi-Derakhshi, and A. Sharifi, "A Model for Detecting of Persian Rumors based on the Analysis of Contextual Features in the Content of Social Networks," (in eng), *Signal and Data Processing*, Research vol. 18, no. 1, pp. 50-29, 2021, doi: 10.52547/jsdp.18.1.50.



مسلم صامت عمرانی مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر- نرم‌افزار از دانشگاه پیام نور محمودآباد در سال ۱۳۸۸ و کارشناسی ارشد را در رشتهٔ مهندسی کامپیوتر- هوش مصنوعی از دانشگاه تربیت مدرس در سال ۱۴۰۰ دریافت کرده‌است. زمینه‌های پژوهشی مورد علاقهٔ ایشان عبارت‌اند از: داده‌کاوی، متن‌کاوی و پردازش تصویر.

نشانی رایانامه ایشان عبارت است از:

Moslem.ai1983@gmail.com



محمّد صنیعی آباده مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر- نرم‌افزار از دانشگاه صنعتی اصفهان در سال ۱۳۸۰، کارشناسی ارشد را در رشتهٔ مهندسی کامپیوتر- هوش مصنوعی از دانشگاه علم‌وصنعت در سال ۱۳۸۱ و دکترا را در رشتهٔ مهندسی کامپیوتر- هوش مصنوعی از دانشگاه صنعتی شریف در سال ۱۳۸۶ دریافت کرده‌است. زمینه‌های پژوهشی مورد علاقهٔ ایشان عبارت‌اند از: نظرکاوی در وب و شبکه‌های اجتماعی، سامانه‌های هوشمند مبتنی بر عظیم داده‌کاوی، روش‌های فرامکاشفه‌ای در داده‌کاوی و داده-کاوی در پزشکی و بیوانفورماتیک.

نشانی رایانامه ایشان عبارت است از:

saniee@modares.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]



نصرالّه مقدّم چرکری مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر- نرم‌افزار از دانشگاه شهید بهشتی در سال ۱۳۶۴، مدرک کارشناسی ارشد و دکترا را در رشتهٔ مهندسی کامپیوتر از دانشگاه یاماناشی ژاپن در سال‌های ۱۳۷۱ و ۱۳۷۴ دریافت کرده‌است. زمینه‌های پژوهشی ایشان عبارت‌اند از: آنالیز و بازیابی تصویر و ویدیو، شبکه‌های حس‌گر و بی‌سیم، شناسایی رخداد در سامانه‌های ویدیویی، شبکه‌های پیچیده و بیوانفورماتیک.

نشانی رایانامه ایشان عبارت است از:

moghadam@modares.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-04]

[DOI: 10.61186/jsdp.21.2.15]