ارائهٔ مدلی برای تشخیص شایعات فارسی مبتنی بر تحلیل ویژگی‌های محتوایی در متن شبکه‌های اجتماعی

زلیخا جهانبخش نقده[۱]، محمدرضا فیضی درخشی*[۲] و آرش شریفی[۳]

[۱ و۳] گروه مهندسی رایانه، واحد علوم و تحقیقات تهران، دانشگاه آزاد اسلامی، تهران، ایران

[*۲] گروه مهندسی رایانه، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز، ایران

چکیده

شایعه یک تلاش جمعی است که در آن از قدرت واژگان برای تفسیر یک موقعیت مبهم ولی جذاب استفاده می‌شود؛ بنابراین، شناسایی زبان شایعه می‌تواند در تشخیص شایعات کمک‌کننده باشد. پژوهش‌های پیشین برای حل مسأله تشخیص شایعه بیشتر بر روی اطلاعات متنی موجود در ریتوییت و توییت پاسخ کاربران و کمتر بر روی متن اصلی شایعه متمرکز شده‌اند. اغلب این پژوهش‌ها بر روی زبان انگلیسی بوده و کارهای محدودی در زبان فارسی انجام شده است؛ از این‌رو، این مقاله 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 توییتر و تلگرام ارائه می‌کند. مدل پیشنهادی شایعات فارسی مجموعه‌داده توییتر را با معیار F- ۰/۸۴۸، شایعات مجموعه‌داده زلزله کرمانشاه را با معیار F- ۰/۹۵۲ و شایعات تلگرامی را با معیار F- ۰/۸۶۷ شناسایی کرده است؛ که نشان‌دهنده توانمندی مدل پیشنهادی برای شناسایی شایعات تنها با تمرکز بر ویژگی‌های محتوایی متن شایعه منبع است.

واژگان کلیدی: تشخیص شایعات فارسی، تحلیل محتوی، ویژگی‌های محتوایی فیزیکی و غیرفیزیکی، پردازش متن.

A Model for Detecting of Persian Rumors based on the Analysis of Contextual Features in the Content of Social Networks

Zoleikha Jahanbakhsh-Nagadeh[1], Mohammad-Reza Feizi-Derakhshi [2*] & Arash Sharifi[3]

[1,3]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

[2*]Department of Computer Engineering University of Tabriz, Tabriz, Iran.

Abstract

The rumor is a collective attempt to interpret a vague but attractive situation by using the power of words. Therefore, identifying the rumor language can be helpful in identifying it. The previous research has focused more on the contextual information to reply tweets and less on the content features of the original rumor to address the rumor detection problem. Most of the studies have been in the English language, but more limited work has been done in the Persian language to detect rumors. This study analyzed the content of the original rumor and introduced informative content features to early identify Persian rumors (i.e., when it is published on news media but has not yet spread on social media) on Twitter and Telegram. Therefore, the proposed model is based on physical and non-physical content features in three categories including, lexical, syntactic, and pragmatic. These features are a combination of the common content features along with the proposed new content-based features. Since no social context information is available at the time of posting rumors, the proposed model is independent of propagation-based features and relies on the content-based information of the original rumor. Although in the proposed model, much information (including user information, the user's

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات

•تاریخ ارسال مقاله: ۱۳۹۸/۰۳/۲۳ •تاریخ پذیرش: ۱۳۹۹/۰۷/۰۲ •تاریخ انتشار: ۱۴۰۰/۰۳/۰۱ •نوع مطالعه: پژوهشی

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

reaction to the rumor, and propagation structures) are ignored, but helpful content information can be obtained for classification by content analysis of the original rumor.

Several experiments have been performed on the various combinations of feature sets (i.e., common and proposed content features) to explore the capability of features in distinguishing rumors and non-rumors separately and jointly. To this end, three machine learning algorithms including, Random Forest (RF), AdaBoost, and Support Vector Machine (SVM) have been used as strong classifications to evaluate the accuracy of the proposed model. To achieve the best performance of classification algorithms on the training dataset, it is necessary to use feature selection techniques. In this study, the Sequential Forward Floating Search (SFFS) approach has been used to select valuable features. Also, the statistical results of the t-test on the P-value (<=0.05) demonstrate that most of the new features proposed in this study reveal statistically significant differences between rumor and non-rumor documents. The experimental results are shown the performance of new proposed features to improve the accuracy of the rumor detection. The F-measure of the proposed model to detect Persian rumors on the Twitter dataset was 0.848, on the Kermanshah earthquake dataset was 0.952 and on the Telegram dataset was 0.867, which indicated the ability of the proposed method to identify rumors only by focusing on the content features of the original rumor text. The results of evaluating the proposed model on Twitter rumors show that, despite the short length of Twitter tweets and the extraction of limited content information from tweets, the proposed model can detect Twitter rumors with acceptable accuracy. Hence, the ability of content features to distinguish rumors from non-rumors is proven.

Keywords: Persian rumors detection, Content analysis, Physical and non-physical content features, Text processing.

۱- مقدمه

پیشرفت بسیار سریع رسانه‌های اجتماعی، نقش کلیدی در انتشار اخبار در جامعه مدرن دارد. اغلب کاربران رسانه‌های اجتماعی نه‌تنها اخبار را می‌خوانند، بلکه آن را نیز انتشار می‌دهند. به این ترتیب، انتشار اخبار به‌راحتی صورت می‌گیرد و در این بین ممکن است، اخباری انتشار یابند که به تعریف و تبیین وضعیت‌های مهمی کمک کنند که به‌صورت دوپهلو ظاهر می‌شوند. این اخبار شایعاتی هستند که بر اثر تدابیر از پیش اندیشیده‌شده، و براساس مطالعات نظام‌مند در زمینه جامعه هدف، تهیه و انتشار می‌یابند و باعث بروز مشکلات متعددی در حوزه‌های اجتماعی، اقتصادی، سیاسی و ... می‌شوند؛ از این‌رو، یک شایعه، تکه‌ای از اطلاعات در حال انتشار است که صحت آن در زمان انتشار توسط منابع معتبر و موثق خبری تأیید نشده است [1]. بر اساس تعاریف گذشته، ابتدا لازم است، تعریف روشنی از شایعه داشته باشیم. در این پژوهش، شایعه به‌عنوان "قطعه‌ای از اطلاعات تأئیدنشده در حال انتشار است که با استفاده از قدرت واژگان و مؤلفه‌های محتوایی قابل مشاهده در متن، ایده‌ها، احساسات و دیدگاه‌های مهم نویسنده را به مخاطب منتقل می‌کند و باعث تحریک هیجانات، احساسات و جلب‌توجه مخاطب می‌شود." معرفی شده است.

در این تعریف به ویژگی‌های خاص و مرکزی شایعه تأکید شده است؛ یعنی ویژگی‌های محتوایی فیزیکی که می‌تواند برای هر خواننده جلب‌توجه‌کننده باشد و ویژگی‌های محتوایی غیرفیزیکی که بیان‌گر احساسات و عقاید درونی نویسنده است. نویسندهٔ شایعه با استفاده از این ویژگی‌های محتوایی گاهی ترس را فریاد می‌کشد، گاهی امید واهی و آینده‌ای روش را نوید می‌دهد؛ از این‌رو، این ویژگی‌های محتوایی می‌توانند وجه تمایز شایعات از غیرشایعات بوده و در فرآیند تشخیص و تعیین اعتبار شایعات نقش به‌سزایی داشته باشند.

در چند سال اخیر، پژوهش‌های متعددی در زمینه شناسایی خودکار شایعات در زبان نگلیسی و چینی –[2] [10] انجام شده که برای این منظور از مجموعه ویژگی‌هایی در سه سطحِ کاربر، محتوی و شبکه انتشار استفاده شده است. برخلاف پژوهش‌های انجام‌شده در زمینه شایعه در زبان انگلیسی که سابقه‌ای بیش از نیم قرن دارد، پژوهش‌های چندانی بر روی شایعات فارسی انجام نشده است و عمر چندانی هم ندارد؛ به‌طوری‌که پژوهش‌هایی که در زبان فارسی بر روی شایعات انجام شده، بیش‌تر در حوزه علوم انسانی [11] در زمینه تأثیر روانی شایعه در سازمان‌ها و جامعه بوده است. بر اساس آخرین پژوهش‌ها تنها دو کار پژوهشی [12] و [13] در حوزه سیستمی و رایانه‌ای بر روی تشخیص شایعات در زبان فارسی انجام شده است.

در طول دهه اخیر، به‌کارگیری فناوری در ایران نیز سبب تسهیل و تسریع فرآیند انتشار اطلاعات و تأثیرگذاری بر دیگران شده و از طرفی این موضوع استعداد نهانی انتشار شایعات را به همین نسبت فراهم می‌کند؛ بنابراین، با توجه به اثرات سوء انتشار شایعه در بین عموم و ایجاد نفرت و ترس و عصبانیت و ترغیب مردم به انتشار

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

شایعات، اعتبارسنجی، تشخیص و کنترل شایعات به مسائلی خطیر در شبکه‌های اجتماعی برخط قرار گرفته است. بنابراین طراحی و پیاده‌سازی سامانه‌هایی خودکار برای شناسایی شایعات فارسی از اهمیت بسیار زیادی برخوردار است.

این پژوهش قصد دارد، برای حل این مسأله، به‌جای تحلیل اطلاعات در سه سطحِ کاربر، محتوا و شبکه انتشار، مدلی مبتنی بر اطلاعات متنی شایعه منبع و بدون وابستگی به اطلاعات مربوط به پیشینه انتشار شایعه ارائه کند؛ به‌طوری‌که مدل پیشنهادی، تنها با تمرکز بر روی ویژگی‌های محتوایی شایعه منبع و بدون درنظرگرفتن اطلاعات متنی مربوط به ریتوییت و توییت‌های پاسخ، قادر به شناسایی شایعات باشد؛ بنابراین، مدل پیشنهادی برای شناسایی شایعات در متون فارسی، بر روی مجموعه ویژگی‌های مبتنی بر محتوای پیام از جمله، بافت متنی، کنش‌گفتار، واژگان رایج در شایعات، تحلیل سطحی و عمیق ویژگی‌های لغوی، نحوی و عملگرا متمرکز شده است. به‌منظور استخراج این ویژگی‌ها، حجم بالایی از پیام‌های رد و بدل‌شده در شبکه پیام‌رسان تلگرام و شبکه اجتماعی توییتر مورد تجزیه و تحلیل قرار گرفته است. این پژوهش، با به‌دست‌آوردن این ویژگی‌ها به‌دنبال پاسخ‌گویی به پرسش‌های زیر است:

۱) آیا تفاوتی در ساختار لغوی و گرامری متون شایعه و غیرشایعه وجود دارد؟

۲) آیا تنها با استفاده از ویژگی‌های محتوایی متن شایعه منبع و بدون در نظر گرفتن اطلاعات ریتوییت و پاسخ کاربران می‌توان شایعات را شناسایی کرد؟

۳) شایعه‌پردازان از چه ویژگی‌های محتوایی فیزیکی و غیرفیزیکی برای افزایش جذابیت متن استفاده می‌کنند؟

تعدادی از پژوهش‌های پیشین [16]–[14] نیز به‌منظور حل مسأله تشخیص شایعه بر روی ویژگی‌های محتوایی متمرکز شده‌اند، اما این درحالی‌ست که آنها اغلب ویژگی‌های مدنظرشان را از اطلاعات ریتوییت و محتوای توییت پاسخ کاربران استخراج می‌کنند. به‌عبارتی، آنها برای انجام فرآیند تشخیص شایعه نیاز به تاریخچه انتشار شایعه و اطلاعاتی دارند که بعد از انتشار شایعه منبع توسط کاربران در فضای اجتماعی ردوبدل می‌شود. از این‌رو، نوآوری‌های اصلی این پژوهش برای شناسایی شایعات فارسی عبارت است از:

- **عدم وابستگی به پیشینه انتشار شایعه:** در این پژوهش، یک مدل مبتنی بر ویژگی‌های محتوایی پست منبع برای تشخیص شایعات ارائه شده است، که برخلاف پژوهش‌های پیشین نیازی به انتظار برای پخش شایعه و جمع‌آوری اطلاعات کمکی همچون ریتوییت و توییت پاسخ کاربران نیست، بلکه تنها با تمرکز بر ویژگی‌های محتوایی متن منبع، شایعه‌بودن یا نبودن آن متن تعیین می‌شود.
- **ارائه ویژگی‌های محتوایی جدید:** به‌منظور تحلیل محتوایی عمیق‌تر متون شایعه، مجموعه‌ای از ویژگیهای جدید شامل چهار ویژگی لغوی، دو ویژگی نحوی و شش ویژگی عمل‌گرا به‌عنوان ویژگی‌های متمایزکننده شایعات از غیرشایعات معرفی شده است. این مجموعه ویژگی‌ها به‌همراه سایر ویژگی‌های محتوایی رایج معرفی‌شده در پژوهش‌های پیشین برای شناسایی شایعات به‌کار گرفته شده‌اند.

در ادامه، ساختار مقاله به این شکل است: در بخش ۲ پژوهش‌های پیشین در زمینه تشخیص شایعات بحث شده و در بخش ۳ مدل پیشنهادی برای شناسایی شایعات فارسی شرح داده شده است. بخش ۴ مجموعه‌دادهٔ به‌کاررفته در مقاله را معرفی می‌کند. در بخش ۵ نتایج تجربی و ارزیابی‌های پژوهش نمایش داده شده و بخش ۶ به بحث و نتیجه‌گیری پژوهش و پیشنهاد کار آینده اختصاص داده شده است.

۲- پیشینه پژوهش

مسألهٔ تشخیص و تعیین اعتبار خودکار شایعات از چندین سال قبل مورد توجه قرار گرفته و پژوهش‌های وسیعی به‌طور پیوسته در این زمینه انجام و نیز پیشرفت‌هایی حاصل شده است. در جدول (۱)، حوزه مجموعه ویژگی‌های به‌کاررفته در روش‌های پیشین و روش پیشنهادی ما نمایش داده شده است. برخی از کارهای انجام شده در زمینه تشخیص و تعیین اعتبار شایعات در سال‌های اخیر به شرح زیر است:

در سال ۲۰۱۱ کاستیلو و همکاران [2] از روش‌های دسته‌بندی باناظر دودویی برای ارزیابی صحت یا عدم صحت شایعات موجود بر روی توییتر انگلیسی استفاده کردند و برای انجام تحلیل‌هایشان چهار نوع ویژگی را مورد بررسی قرار دادند: (۱) ویژگی‌های مبتنی بر پیام (تعداد واژگان و نویسه‌ها، تعداد واژگان مجزا، متوسط طول واژه و ...)، (۲) ویژگی‌های مبتنی بر شبکه (شامل اطلاعات کاربر، تعداد دوستان کاربر و ...)، (۳) ویژگی‌های برجسته توییتر (شامل، URL، هشتگ و ...) و (۴) ویژگی‌های مبتنی بر شبکه انتشار. کاستیلو و همکارانش تنها به

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

تخمین درک شخصی کاربران از معتبربودن توییت‌ها پرداختند و بر روی بررسی میزان اعتبار پیام‌ها متمرکز نشدند. در تحلیل‌هایی که بر روی زلزله سال ۲۰۱۰ در شیلی انجام دادند، آنها به این نتیجه رسیدند که نسبت بین توییت‌های حمایت‌کننده[1]و نقض‌کننده[2] شایعات یک به یک است.

ژائو و همکاران [10] دو نوع الگوی زبانی اصلاحی[3] و درخواستی[4] برای شایعات پیشنهاد دادند؛ بنابراین، پرسش و پاسخ‌های کاربران در مورد وضعیت رویداد (سؤالات تأیید/تصدیق) و اصلاحات/مشاجرات در مورد آن رویداد را مورد تجزیه و تحلیل قرار دادند. به‌منظور استخراج این الگوها، واحدهای لغوی یک‌گرام، دوگرام و سه‌گرام از محتوای مجموعه‌ای از پیام‌های برچسب‌گذاری‌شده استخراج شده و فرکانس هر یک از این ویژگی‌های در کل مجموعه‌داده بررسی، سپس، الگوهای استخراج‌شده با استفاده از روش‌های خی۲[5] و سود اطلاعاتی (IG)[6] رتبه‌بندی می‌شوند تا ویژگی‌های مهمی که قابلیت متمایزکردن توییت‌های شایعه از غیرشایعه را دارند، شناسایی شوند.

در سال ۲۰۱۴، کووان و همکاران [5] تعیین اعتبار شایعات در توییتر را بر اساس سه دسته ویژگی انجام دادند: (۱) الگوهای انتشار شایعه (۱۱ ویژگی بر اساس سری‌های زمانی به‌دست‌آمده از وضعیت شبکه اجتماعی در لحظه iام)، (۲) شبکه انتشار و شبکه دوستی[7] (پانزده ویژگی ساختاری بر اساس سه ساختار شامل شبکه دوستی، LCC شبکه دوستی، شبکه انتشار)، (۳) استخراج ۶۵ ویژگی زبانی با استفاده از ابزار معنایی LIWC[8]. آنها برای تجزیه و تحلیل متن از ابزار LIWC استفاده کردند که اجزای ساختاری، احساسی و شناختی یک متن را با استفاده از فرهنگ واژگان ارزیابی می‌کند. این پژوهش‌گران، واژگان احساسی مثبت، واژگانی با کنش تجربی و ضمیر اول شخص را به‌عنوان ویژگی‌هایی متمایزکننده شایعات از غیرشایعات معرفی کردند.

در سال ۲۰۱۱، قزوینیان و همکاران [3] سه مجموعه ویژگی برای تعیین اعتبار شایعات بر روی توییت‌های انگلیسی مورد بررسی قرار دادند. این ویژگی‌ها عبارتند از: (۱) ویژگی های مبتنی بر محتوا (اجزای گفتار(POS)[9]، تعریف برچسب برای هشتگ‌ها و برچسب URL برای url و استخراج N گرام‌ها)، (۲) ویژگی‌های مبتنی بر شبکه (تمرکز بر روی رفتار کاربر: ویژگی‌های ریتوییت و توییت پاسخ) و (۳) ویژگی‌های خاص توییتر (هشتگ و URL).

در سال ۲۰۱۵، حمیدیان و دیاب [6] برای تعیین اعتبار شایعات انگلیسی دو سری آزمایش شامل، دسته‌بندی تک‌مرحله‌ای (SRDC)[10] و دسته‌بندی دومرحله‌ای (TRDC)[11] با استفاده از روش نظارتی انجام دادند. مراحل دسته‌بندی دومرحله‌ای عبارتند از: (۱) تشخیص شایعه(معتبر یا نامعتبر)، (۲) دسته‌بندی نوع شایعه (تائیدی، تکذیبی، سؤالی یا بی‌طرف). آنها سه مجموعه ویژگی برای انجام آزمایش‌هایشان به‌کار گرفتند: (۱) ویژگی‌های مبتنی بر محتوای توییت، (۲)، ویژگی‌های عمل‌گرا[12] (احساسات، شکلک‌های اینترنتی، موجودیت-های نامدار)، (۳) ویژگی‌های مخصوص توییتر (تعداد توییت‌های پست‌شده در یک روز، شناسه کاربری،ریتوییت، پاسخ‌ها، هشتگ‌ها و URL).

در سال ۲۰۱۵، وثوقی [17] از میان ویژگی‌های تعریف‌شده در مقاله [2]، هفده ویژگی را به‌عنوان مهم‌ترین و موثرترین ویژگی‌ها برای شناسایی شایعات معرفی و مورد تحلیل قرار داد. وثوقی صحت توییت‌های شایعه انگلیسی را در دوره‌های زمانی مختلفی با استفاده از دسته‌بندهای DTW و مدل پنهان مارکوف (HMM)[13] مدل‌سازی کرد. بهترین نتیجه به‌دست‌آمده در دوره زمانی از نیمه دوم زمان شروع شایعه تا زمان تأیید صحت/عدم‌صحت شایعه بود که از دقت ۰/۶ تجاوز نکرد. این دوره زمانی، به‌طور میانگین در حدود سه ساعت و بیست دقیقه بعد از شیوع شایعه بود.

در سال ۲۰۱۵، وویو و همکاران [8] از ساختار انتشار پیام‌ها برای شناسایی شایعات بر روی شبکه اجتماعی چینی Sina Weib استفاده کردند. آنها معتقد بودند، ساختار انتشار شایعات متفاوت از غیرشایعات است. به‌طوری‌که، یک شایعه ابتدا توسط یک کاربر معمولی ارسال، سپس توسط تعدادی رهبر نظرریتوییت و پشتیبانی و درنهایت توسط تعداد زیادی کاربر معمولیریتوییت می‌شود. برعکس، یک پست غیرشایعه، توسط یک رهبر نظر پست می‌شود و به‌طورمستقیم توسط کاربران معمولی زیادی ریتوییت می‌شود.

در سال ۲۰۱۶، جاسمیدیس و همکاران [7] در دوره زمانی مشابه روش وثوقی (ده دوره زمانی) به دقت دست‌کم ۷۶٪ دست یافتند که دقتی بالاتر از روش وثوقی [17] است. این دقت ۷۶٪ در تعیین اعتبار شایعه به‌طور میانگین در حدود یک ساعت و پنجاه دقیقه بعد از شیوع شایعه به‌دست آمده است. نویسندگان برای هر شایعه چهار مجموعه داده شامل هشتاد معیار برای تعیین قابلیت اعتماد جمع‌آوری کردند: (۱) اطلاعات توییت‌ها (محتوا،

[1] Supporting
[2] Debunking
[3] Correction
[4] Enquiry
[5] Chi-Squared
[6] Information Gain
[7] Friendship
[8] Linguistic Inquiry and Word Count
[9] Part of Speech

[10] Single-step Rumor Detection and Classification
[11] Two-step Rumor Detection and Classification
[12] Pragmatic
[13] Hidden Markov Models

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

مهرزمانی،ریتوییت و غیره)، (۲) اطلاعات کاربران (شماره کاربری، تعداد پست‌ها، دوستان و فالوورهای کاربر و غیره)، (۳) اتصالات شبکه اجتماعی (شبکه دوستان) و (۴) ۴۰۰ توییت آخر کاربران قبل از شروع شایعه.

در سال ۲۰۱۶، حمیدیان و دیاب [9] در کار دیگری، از مجموعه ویژگی‌های مطرح شده در کار قزوینیان [3] به‌همراه پانزده ویژگی پیشنهادی پیشین خود [6] استفاده کردند. نویسندگان به بازیابی شایعه (RR)[1] و بررسی اعتماد[2] بر روی توییتر انگلیسی پرداختند. با توجه به طول کوتاه توییت‌ها، نویسندگان برای تعیین موضوع شایعه در فرآیند RR، بردار نهان توییت (TLV)[3] را به‌کار گرفتند. ویژگیTLVیک نمایش صد بُعدی از هر توییت ایجاد می‌کند تا اطلاعات بیشتری از محتوای توییت برای تعیین شباهت معنایی بین توییت‌ها در دسترس باشد. نویسندگان همچنین معیار اعتماد را معرفی و تغییرات میزان اعتماد منتشرکنندگان شایعه را مابین سال‌های ۲۰۱۰ تا ۲۰۱۶ بررسی کردند.

در سال ۲۰۱۶، چووا و همکاران [18] شش مجموعه ویژگی را برای بررسی نقش زبان در تعیین اعتبار شایعات پیشنهاد دادند که عبارتند از: قابلیت درک[4](تعداد واژگان با طول بیشتر از شش کاراکتر)، احساسات (تعداد واژگان احساسی مثبت، منفی و خنثی)، زمان (تعداد واژگان در زمان‌های گذشته، حال و آینده)، کمیت(۳ ویژگی)، استایل نوشتاری (۶ ویژگی)، و عنوان (نوع شایعه). آنها از ابزار LIWC برای اندازه گیری ویژگی‌های زبانی شایعات استفاده کردند.

در سال ۲۰۱۷، زوبیاگا و همکاران [19] از دو نوع ویژگی برای دسته‌بندی شایعات استفاده کردند. (۱) ویژگی‌های مبتنی بر محتوی (هفت ویژگی شامل، POS، علامت سؤال و تعجب و ...). (۲) ویژگی‌های اجتماعی (پنج ویژگی شامل، تعداد توییت، مدت زمان عضویت کاربر، اعتبار کاربر و ...). اما، اغلب این ویژگی‌ها در سناریوی زمان واقعی به‌دست می‌آیند.

در سال ۲۰۱۸، تاکور و همکاران [21] یک چارچوب یادگیری ماشین باناظر برای تشخیص شایعات منتشرشده در توییتر پیشنهاد دادند. این چارچوب از دو مرحله (۱) دسته‌بندی توییت‌ها در دو رده شخصی و غیرشخصی، (۲) تشخیص موضوع توییت‌های خروجی مرحلهٔ نخست تشکیل شده است.

وثوقی و همکاران در سال ۲۰۱۸ [22] به بررسی انتشار اخبار واقعی، نادرست و مختلط (تا حدودی درست، تا حدودی نادرست) بر روی مجموعه داده‌ای از اخبار و شایعات منتشرشده از سال ۲۰۰۶ تا ۲۰۱۷ در توییتر پرداختند. آنها دریافتند که داستان‌های دروغین باعث ترس، انزجار و تعجب می‌شوند، و داستان‌های واقعی الهام‌بخش از پیش‌بینی، غم، شادی و اعتماد بودند. نتایج آزمایش‌های آنها نشان داد که سرعت انتشار اخبار واقعی و نادرست توسط ربات‌ها یکسان است، اما درحقیقت خبرهای نادرست بیشتر از اخبار واقعی پخش می‌شوند، زیرا انسان‌ها با احتمال بیشتری اقدام به انتشار می‌کنند.

لی و همکاران در سال ۲۰۱۹ [15] تعیین اعتبار شایعات را در دو مرحله انجام دادند: مرحله نخست، تعیین موضع[5] کاربر نسبت به شایعه، مرحله دوم: تعیین اعتبار شایعه (صحیح، نادرست یا تأییدنشده). آنها برای دسته‌بندی موضع، یک مدل شبکه عصبی با ویژگی‌های زبان را به‌کار بردند؛ و برای تعیین اعتبار شایعه، از اطلاعات محتوای شایعه، اعتبار منبع و کاربر، موضع کاربر، و مسیر انتشار رویداد استفاده کردند.

خینگ و همکاران در سال ۲۰۱۹ [16] یک چارچوب یادگیری جدید به‌وسیله شبکه عصبی بازگشتی مبتنی بر ادراک بصری عمیق[6](ViP-RNN) برای تشخیص شایعه پیشنهاد کردند. آنها ویژگی‌های سطح بالا و اطلاعات متنی را مورد بررسی قرار دادند. در این مدل از RNN برای به‌دست‌آوردن وابستگی‌های موقت اطلاعات متنی پست‌های مرتبط با شایعه و ترکیب ویژگی‌های لغوی سطح پایین با تعاملات معنایی سطح بالا با روش سلسله‌مراتبی با درک بصری شبکه عصبی کانولوشن (CNN)[7] به‌کار گرفته شده است. برای ادغام اطلاعات آموخته‌شده به‌وسیله RNN وCNN، آنها لایه‌های کانولوشن و بازگشتی را در یک مدل ترکیب کردند.

بر اساس مطالعات ما، کارهای انجام‌شده در حوزه سیستمی و رایانه‌ای بر روی تشخیص شایعات فارسی در سال ۲۰۱۷ توسط زمانی و همکاران [12] و در سال ۲۰۱۸ توسط محمودآباد و همکاران [20] است که بر روی توییت‌های فارسی انجام شده است. در هر دو کار، مجموعه‌ای از ویژگی‌های ساختاری (شامل، بررسی نوع توییت (ریتوییت یا توییت پاسخ)، تعداد ریتوییت، و ...)، محتوایی و ویژگی‌های خاص توییتر (شامل پنجاه‌هزار یونیگرام از شایعات توییتر، تعداد نویسه‌های توییت، URL، هشتک و ...) و ویژگی‌های کاربر (مانند، تعداد دوستان، دنبال‌کننده‌گان[8]، اعتبار کاربر و ...) مورد ارزیابی قرار گرفته است. زمانی و همکاران [12] ارزیابی‌های خود را بر روی مجموعه‌ای از توییت‌های شایعه و غیرشایعه فارسی با موضوعات مختلف انجام دادند. نتایج آزمایشگاهی آنها نشان داد که ویژگی‌های کاربران نسبت به سایر ویژگی‌ها

[1] Retrieval Rumor
[2] Belief
[3] Tweet Latent Vector
[4] Comprehensibility
[5] Stance
[6] Deep Visual perception based Recurrent Neural Network
[7] Convolutional Neural Network
[8] Follower

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

در شناسایی شایعات مؤثرتر عمل می‌کنند. محمودآباد و همکاران [20] ارزیابی‌های خود را بر روی رویداد "زلزله کرمانشاه" در توییتر با استفاده از دسته‌بندهای مختلف انجام دادند. نتایج ارزیابی‌های آنها، جنگل تصادفی و meta.RandomSubSpace را به‌عنوان بهترین دسته‌بند در شناسایی شایعات نشان می‌دهد.

این پژوهش، برای شناسایی شایعات تنها بر روی ویژگی‌های محتوایی متن منبع متمرکز شده است. وجه تمایز این پژوهش با پژوهش‌های پیشین این است که، پژوهش‌هایی [19] ,[18] ,[16] که برای حل مسأله تشخیص شایعه بر روی ویژگی‌های محتوایی متمرکز شده‌اند، از اطلاعات استخراج‌شده از واکنش کاربران (ریتوییت یا توییت‌های پاسخ) به خبر منتشرشده کمک می‌گیرند. برای استخراج این اطلاعات، لازم است، مدت‌زمانی از انتشار شایعه سپری شود؛ اما، این پژوهش قصد دارد بدون صرف زمان برای انتشار شایعه و بدون درنظرگرفتن واکنش کاربران، تنها با تمرکز بر روی اطلاعات محتوایی متن منبع شایعات فارسی را شناسایی کند. مزیت مدل پیشنهادی نسبت به کارهای پیشین این است که، این مدل به‌دلیل عدم وابستگی به اطلاعات پیشینه انتشار، قادر است در همان لحظات اولیه انتشار که اطلاعاتی از واکنش کاربران در دسترس نیست، شایعات را شناسایی کند.

(جدول-۱): فهرست ویژگی‌های معرفی‌شده در پژوهش‌های پیشین مبتنی بر یادگیری باناظر برای تشخیص شایعه در مقایسه با روش پیشنهادی.

(Table-1): The list of features introduced by supervised learning- based previose works for rumor detection in compration with the proposed method.

مرجع	ویژگی‌ها														
	محتوی											کاربر	شبکه انتشار	منحصر به توییتر	پیشینه انتشار	
	لغوی	نحوی	عملگرا												ریتوییت	توییت پاسخ
			احساسات	موجودیت‌های نامدار	واژه‌های عاطفی	هیجان	کنش گفتار	نفی	ارزش خبری							
[2]	√		√					√				√	√	√		
[10]	√														√	√
[5]	√		√		√			√					√			
[3]	√	√			√								√	√	√	√
[6]	√		√	√	√							√			√	√
[17]	√												√		√	√
[8]												√	√		√	
[7]	√											√	√	√	√	√
[9]	√		√	√	√							√			√	√
[18]	√											√	√	√	√	√
[19]	√	√										√		√		
[12]	√											√	√	√	√	√
[20]	√											√	√	√	√	√
روش پیشنهادی	√	√	√	√	√	√	√	√	√							

۳- روش پیشنهادی

در این پژوهش، با استفاده از روش‌های متن‌کاوی و تحلیل محتوا، مدلی مبتنی بر ویژگی‌های محتوایی به‌منظور شناسایی شایعات در زبان فارسی ارائه شده است. هر شایعه شامل اطلاعات محتوایی فیزیکی و غیرفیزیکی است که می‌تواند در شناسایی شایعات نقش مهمی ایفا کنند. در روش پیشنهادی با تحلیل این دو دسته ویژگی، مجموعه اطلاعات ارزشمندی از محتوای شایعه برای شناسایی آن استخراج می‌شود. مراحل مدل پیشنهادی برای تشخیص شایعات عبارتند از:

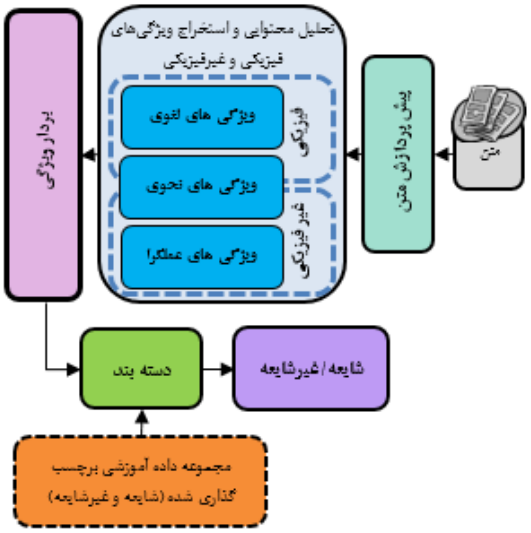
- مرحله نخست، پیش‌پردازش متن ورودی؛
- مرحله دوم، تحلیل محتوا و استخراج ویژگی‌های فیزیکی و غیرفیزیکی؛

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

- مرحله سوم، نمایش متن به‌شکل بردار ویژگی بر اساس ویژگی‌های محتوایی استخراج‌شده از مرحله قبل؛
- مرحله چهارم، آموزش دسته‌بند بر اساس بردارهای ویژگی ایجادشده برای مجموعه‌داده آموزشی؛
- مرحله پنجم، ایجاد بردار ویژگی برای متن ورودی و تشخیص شایعه‌بودن یا نبودن آن متن به‌وسیله دسته‌بند.

در ادامه، به شرح جزئیات مربوط به هر یک از مراحل پرداخته می‌شود. همچنین، ساختار کلی مدل پیشنهادی برای تشخیص شایعات فارسی در شکل (۱) نمایش داده شده است.



(شکل-۱): ساختار کلی مدل پیشنهادی برای تشخیص شایعات.

(Figure-1): The general structure of the proposed model to identify rumors.

۳-۱- پیش‌پردازش

متون برخط از جمله پست‌های منتشرشده در شبکه‌های اجتماعی و پیام‌رسان به‌طورمعمول حاوی نوفه هستند. از آنجایی‌که مدل پیشنهادی مبتنی بر تحلیل ویژگی‌های محتوایی متن شایعه است، بنابراین، انجام پیش‌پردازش داده‌ها به‌عنوان نخستین گام در فرایند تحلیل محتوای متن به‌منظور نرمالیزه‌کردن یک متن غیراستاندارد به یک حالت متعارف ضروری است. در این پژوهش، شش مرحله پیش‌پردازش با استفاده از کتابخانه هضم (Hazm)[1] بر روی متون انجام شده است. هضم یک کتابخانه منبع باز برای انجام پردازش‌های لازم بر روی زبان فارسی در پایتون است که با استفاده از کتابخانه NLTK منتشر شده است. مراحل انجام‌شده برای پیش‌پردازش متون شامل موارد زیر است:

[1] http://www.sobhe.ir/hazm/

- نرمال‌سازی[2]: یکسان‌سازی نویسه‌ها با جایگزین‌کردن نویسه‌های استاندارد در متن ورودی.
- تقطیع[3] جملات و واژه‌ها: جداسازی جملات و واژه‌ها.
- ریشه‌یابی لغوی[4]: استانداردسازی واژگان به‌صورت ریشه واژه، مثال: "کتاب‌ها"←"کتاب".
- ریشه‌یابی معنایی[5]: این فرایند تا حدود زیادی شبیه ریشه‌یابی لغوی است، اما با این تفاوت که واژه ریشه همواره از نظر فرهنگ واژگانی یک واژه صحیح است (یعنی در لغت‌نامه وجود دارد)؛ اما در ریشه‌یابی لغوی ممکن است چنین نباشد مثال: "می‌روم"←"رفت#رو".
- تگ‌گذاری اجزای گفتار (POS)[6]: عمل انتساب برچسب‌های نحوی (از قبیل اسم، صفت، قید، فعل، حروف و ...) به واژه‌ها و نشانه‌های تشکیل‌دهنده یک متن است.
- درخت تجزیه وابستگی[7]: درخت وابستگی رابطه بین واژگان را در یک جمله نمایش می‌دهد. بنابراین، پیچیدگی گرامری متن برمبنای عمق درخت وابستگی آن متن محاسبه می‌شود.

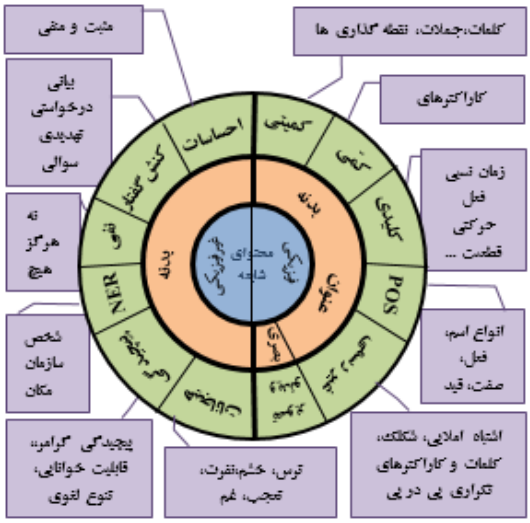
۳-۲- تحلیل محتوا و استخراج ویژگی‌های محتوایی

تحلیل محتوا، فنی است که به‌وسیله آن مشخصات خاص پیام به‌طور روشمند و دقیق جهت استنباط علمی شناسایی می‌شوند. در تحلیل محتوا، پژوهش‌گران به‌دنبال شناخت اثر پیام بر روی مخاطبان هستند. محتوای یکی متن دارای دو مجموعه ویژگی فیزیکی و غیرفیزیکی است. دسته نخست ویژگی‌هایی هستند که در متن قابل مشاهده‌اند، دسته دوم، بیان‌گر احساسات و دیدگاه‌های نویسنده متن است که قصد انتقال آن به خوانندگان را دارد؛ از این‌رو، در مدل پیشنهادی برای تحلیل محتوای شایعات، با بهره‌گیری از روش‌های پردازش زبان طبیعی (NLP)[8] دو مجموعه‌ویژگی فیزیکی و غیرفیزیکی معرفی شده‌اند که قابلیت متمایزسازی شایعات و غیرشایعات را دارند (شکل ۲). در ادامه، جدول‌های (۲ و ۳) مجموعه‌ویژگی‌های به‌کاررفته را برای تشخیص شایعات در زبان فارسی به‌ترتیب در دو دسته فیزیکی و غیرفیزیکی

[2] Normalization
[3] Tokenization
[4] Stemming
[5] Lemmatization
[6] Part of Speech tagging
[7] Dependency tree
[8] Natural Language Processing

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

نشان می‌دهند. ویژگی‌های پیشنهادی به‌صورت پررنگ و با علامت ستاره (*) نمایش داده شده است.



(شکل-۲): شایعات و ویژگی‌های محتوایی مرتبط با آن
(Figure-2): Rumors and content features related to it

۱-۲-۳- محتوای فیزیکی شایعات

محتویات فیزیکی یک متن حامل‌ها و قالب‌های به‌کاررفته در آن متن است. همان‌طورکه در شکل (۲) نشان داده شده است، محتوای فیزیکی شایعات شامل عنوان خبر، متن اصلی و عناصر دیگری مانند تصاویر یا فیلم‌ها است؛ بنابراین، هر مؤلفه موجود قابل مشاهده در متن داده‌های اجتماعی برخط، مانند URL، هشتگ، ایموجی، تصویر، و فیلم به‌عنوان یک ویژگی فیزیکی در نظر گرفته می‌شوند. در این پژوهش با توجه به نبود وجود اطلاعات بصری (شامل، تصویر و ویدئو) در مجموعه‌دادهٔ گردآوری‌شده، تنها عناصر موجود در عنوان و بدنه متن به‌عنوان سبک نوشتاری پردازش شده است.

از آنجایی‌که، اخبار موثق به‌طورمعمول با نوشتار رسمی و با الگوی گرامری استاندارد بیان می‌شوند، بنابراین دارای سبک نوشتاری یکسانی هستند؛ اما شایعات با سبک‌های نوشتاری متفاوتی بیان می‌شوند [23]. این تفاوت در الگوی نوشتاری به‌دلیل تفاوت در نوع شایعه است؛ زیرا، شایعات براساس معیارهای گوناگونی، نظیر معیار زمان، موضوع، انگیزه انتشار و نظایر آن می‌توانند دسته‌بندی شوند. برای مثال، تقسیم‌بندی شایعات براساس انگیزش‌های روان‌شناختی شایعه‌پردازان شامل شایعات هیجانی، ترس‌آور و خسمانه است که شایعات ایجادشده در هر یک از این حوزه‌ها با ویژگی‌های محتوایی متفاوتی بیان می‌شوند تا مورد توجه مخاطب قرار گیرند؛ از این‌رو، با پردازش حجم زیادی از اخبار و شایعات فارسی، مجموعه‌ای از ویژگی‌های محتوایی فیزیکی در دو دسته لغوی[1]، نحوی[2] برای تجزیه و تحلیل سبک نوشتاری معرفی و مورد ارزیابی قرار گرفته است. جدول (۲) ترکیبی از ویژگی‌های محتوایی فیزیکی به‌کاررفته در پژوهش‌های پیشین و ویژگی‌های پیشنهادی جدید در پژوهش حاضر را نمایش می‌دهد. ویژگی‌های جدید با علامت ستاره (*) مشخص شده‌اند.

(جدول-۲): فهرست ویژگی‌های محتوایی فیزیکی برای تشخیص شایعات. ویژگی‌های پیشنهادی جدید با علامت * مشخص شده‌اند.

(Table-2): The list of physical features for rumor detection. The proposed new features are marked with the "*" mark.

دسته	زیردسته	ویژگی (توضیحات)	اختصار
لغوی	کمیتی	#واژگان	W
		#جملات	S
		نقطه‌گذاری (؟، !، : و ...)	P
		نویسه‌های خاص(~ , @, #, $, %,)	SC
	کمی	نویسه‌های عددی (۱۰۰، ۸۰٪ و ...)	NLC
		واژگان شمارشی (دومین، صد و ...)	NW
		واژگان کمیّت‌سنج (چند، بسیار و ...)	Qty
	کلیدی	واژگان قطعی (حتما، باید و ...)	Cer
		واژگان غیرقطعی(شاید، احتمالاو ...)	UCer
		واژگان سؤالی (چرا، کجا و ...)	QW
		واژگان استنتاجی (نتیجتاً، ...)	IW
		افعال حسی (دیدن، شنیدن و ...)	SV
		زمان نسبی (فردا، امروز، دیشب و ...)	RT*
		افعال حرکتی (نوسان، افتاد و ...)	MV*
		عبارات آغازین(خبر مهم، فوری و ...)	SP*
		عبارات پایانی(کپی‌اجباری، اطلاع‌رسانی کنید و ...	EP*
نحوی	POS	اسم	N
		فعل	V
		صفت	Adj
		قید	Adv
		ضمیر (سوم شخص و اول شخص جمع)	Pro
	غیررسمی	اشتباه املایی	SM
		واژگان بد	BW
		واژگان عامیانه	SW
		شکلک ☺	Emj
		واژگان پی در پی در یک جمله	CW*
		نویسه‌های پی در پی در یک کلمه	CC*

[1] Lexical
[2] Syntactic

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

۱-۱-۲-۳- ویژگی‌های لغوی

ویژگی‌های لغوی، معانی یا احساسات خاصی را بیان می‌کنند و راهنمای خیلی مهمی برای توصیف متن هستند. این ویژگی‌ها از سطح واژگان متن استخراج می‌شوند. ویژگی‌های لغوی به‌کاررفته در این پژوهش عبارتند از:

- **ویژگی‌های کمیتی**[1]: ویژگی‌هایی از قبیل تعداد واژگان، طول واژگان، طول جمله، تعداد نشانه‌ها و ... مورد بررسی قرار گرفته است. نتایج تحلیل محتوایی نشان داد که ویژگی‌های کمیتی در شایعات فارسی حدود سیزده درصد بیشتر از غیرشایعات است.
- **ویژگی‌های کمّی**[2]: شایعه‌پردازن برای واقعی جلوه‌دادن موضوع شایعه سعی می‌کنند از اعداد و ارقام بیشتری به‌عنوان اطلاعات آماری استفاده کنند. برای مثال در بخشی از شایعه مربوط به زلزله تهران آمده است *"لطفًا اطلاع‌رسانی کنید: رییس شورای شهر تهران همین الان زنده تو شبکه خبر اعلام کرد امشب زلزله بزرگ پایتخت به احتمال هشتاددرصد میاد."* در این شایعه، فرد شایعه‌ساز با اشاره به درصد رخداد زلزله و سایر ویژگی‌های متنی سعی در افزایش ارزش خبری متن دارد تا اعتماد مخاطب را جلب کند؛ بنابراین، سه دسته ویژگی کمّی شامل، نویسه‌های عددی(NLC)[3]، واژگان شمارشی (NW)[4] و کمیّت‌سنج(Qty)[5] مورد تحلیل قرار گرفته است. میانگین رخداد هر یک از سه ویژگی کمّی $F = \{NLC.NW.Qty\}$ در متن T به‌طور مجزا با استفاده از فرمول (۱) محاسبه می‌شود:

$$Score_F(T) = \frac{\sum_{i=1}^{|S(T)|} |F(S_i(T))|}{|S(T)|} \qquad (۱)$$

در فرمول (۱)، $|S(T)|$ تعداد جملات متن T و $|F(S_i(T))|$ یک مقدار بولین است که رخداد یا عدم رخداد ویژگی F در جمله i ام از متن T را بیان می‌کند.

- **ویژگی‌های کلیدی**: بر اساس تئوری آلپورت و پستمن [24]، "اهمیت" و "ابهام" دو فاکتور اساسی در افزایش قدرت انتشار شایعه هستند؛ بنابراین، لازم است ویژگی‌هایی مدنظر قرار گیرد که بیان‌گر اهمیت و ابهام موجود در متن هستند. از این‌رو، در این پژوهش، دو مجموعه ویژگی برای ارزیابی این دو فاکتور معرفی شده است: (۱) ویژگی‌های تعیین‌کننده اهمیت: واژگان مرتبط با قطعیت(Cer)[6]، افعال حسی(SV)[7]، زمان نسبی(RT)[8]، افعال حرکتی(MV)[9]، عبارت آغازین(SP)[10]، عبارات پایانی(EP)[11]. (۲) ویژگی‌های تعیین‌کننده ابهام: واژگان مرتبط با عدم‌قطعیت (UCer)[12]، واژگان سؤالی (QW)[13]، واژگان استنتاجی (IW)[14]. از میان ویژگی‌های معرفی‌شده در این دو دسته، ، چهار ویژگی "زمان نسبی"، "افعال حرکتی"، "عبارات آغازین" و "عبارات پایانی" ویژگی‌های پیشنهادی پژوهش حاضر است. دلایل معرفی این ویژگی‌ها این است که، (۱)شایعه‌پردازان، برای ایجاد یک جلوه داغ خبری از زمان‌های نسبی در بیان شایعه استفاده می‌کنند. (۲) آنها با به‌کارگیری افعال حرکتی همچون "اوج‌گرفتن"، "جهش"، "افتادن"، "تکان دادن"، "نوسان" و غیره سعی در افزایش هیجان موضوع دارند. (۳) با استفاده از قدرت واژگان در دو موقعیت بالقوه، یعنی ابتدا و انتهای متن، با استفاده از عبارات هیجانی همچون "فوری فوری"، "خبر مهم" و ...؛ و همچنین عبارات درخواستی در عبارات پایانی متن شایعه همچون "کپی اجباری" و "لطفا اطلاع‌رسانی کنید" و ...، احساسات مخاطبان را نسبت به انتشار و اطلاع‌رسانی شایعه تحریک می‌کنند.

میانگین رخداد هر یک از هفت ویژگی کلیدی نخست $F = \{Cer.UCer.QW.IW.SV.RT.MV\}$ در متن T نیز به‌طور مجزا با استفاده از فرمول (۱) محاسبه می‌شود، و برای دو ویژگی SP و EP، به‌ترتیب عنوان متن و جمله آخر از بدنه شایعه مورد پردازش قرار می‌گیرد.

۲-۱-۲-۳- ویژگی‌های نحوی

این نوع ویژگی‌ها اجزای متن را در سطح جمله نمایش می‌دهند. ویژگی‌های نحوی بررسی‌شده بر روی شایعات در این پژوهش شامل برچسب‌گذاری اجزای واژگانی کلام (POS)، پیچیدگی متن و رسمی یا غیررسمی بودن متن است.

---

[1] Quantity
[2] Quantitative
[3] Numeral
[4] Number
[5] Quantifier
[6] Certainty
[7] Sensory Verbs
[8] Relative Time
[9] Motion Verbs
[10] Start Phrase
[11] End Phrase
[12] Uncertainty
[13] Questuion Words
[14] Inferring Words

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

• **اجزای واژگانی کلام**: براساس مطالعات انجام‌شده بر روی چندصد شایعه فارسی، مشخص شد که شایعه‌پردازان از صفت و قید بیشتری برای توصیف واژگان، تغییر ماهیت واژگان دیگر، افزایش هیجان موضوع و تغییر درک مخاطب نسبت به موضوع استفاده می‌کنند.

• **غیر رسمی**[1]: نتایج تحلیل محتوایی متون شایعه و غیرشایعه نشان داد که متون شایعه نسبت به متون غیرشایعه دارای سبک نوشتاری غیررسمی‌تری هستند. در این پژوهش، شش نوع ویژگی معرفی شده است که وجود این ویژگی‌ها در متن منجر به غیررسمی‌شدن محتوا شده و از ارزش خبری آن می‌کاهد. این ویژگی‌ها عبارتند از: (۱) واژگان بد(BW)[2] (مجموعه‌ای شامل ۹۶۵ واژه عامیانه و مبتذل فارسی[3])، (۲) واژگان عامیانه(SW)[4]، (۳) شکلک(Emj)[5]، (یک نمایش تصویری از حالت چهره با استفاده از علائم نقطه‌گذاری، اعداد و حروف الفبا که برای بیان احساسات شخصی به‌کار می‌روند، مانند، :-) لبخند)[6]. (۴) اشتباه املایی(SM)[7]، برای شناسایی اشتباه املایی موجود در متون فارسی، از پیاده‌سازی موجود بر روی گیت‌هاب[8] استفاده شده است؛ سپس، میزان اشتباه املایی موجود در متن T با استفاده از فرمول (۲) محاسبه می‌شود:

$$Score_{SM}(T) = \frac{|SM(T)|}{|W(T)|} \qquad (۲)$$

در فرمول (۲)، $|SM(T)|$ تعداد اشتباه املایی در متن T است و $|W(T)|$ تعداد واژگان موجود در متن T می‌باشد.

(۵)واژگان متوالی(CW)[9]، (۶) نویسه‌های متوالی(CC)[10]. تکرار متوالی واژگان و نویسه‌ها باعث غیررسمی‌شدن ساختار نحوی متن می‌شود. چنین ساختاری اغلب در متون شایعه ظاهر و به‌ندرت در غیرشایعات دیده می‌شود؛ از این‌رو، در این مقاله، دو ویژگی CW و CC به‌عنوان ویژگی‌های نحوی جدید پیشنهاد داده شد. به‌منظور استخراج ویژگی CW، هر واژه از متن با واژه همجوار خود در متن مقایسه می‌شود. در صورت تشابه واژه‌های مجاور (قبل یا بعد) با آن واژه (مثال: هشدار هشدار)، به ویژگی CW مقدار دودویی ۱ اختصاص داده می‌شود. به همین ترتیب، برای استخراج ویژگی CC، به‌ازای هر واژه موجود در متن، نویسه‌های آن واژه مورد بررسی قرار می‌گیرند، اگر نویسه‌ای بیش از دو مرتبه به‌صورت متوالی تکرار شده باشد (مثال: فوررری) به ویژگی CC مقدار دودویی ۱ اختصاص داده می‌شود.

(جدول-۳): فهرست ویژگی‌های محتوایی غیرفیزیکی برای تشخیص شایعات. ویژگی‌های پیشنهادی جدید با علامت * مشخص شده‌اند.

(Table-3): The list of non-physical contextual features for rumor detection. The proposed new features are marked with the "*" mark.

دسته	زیردسته	ویژگی (توضیحات)	اختصار
لغوی	پیچیدگی	میانگین طول کلمات	WL
		میانگین طول جملات	SL
		پیچیدگی گرامری(عمق‌درخت وابستگی)	DDT
		میانگین مکث در جمله	Pous
		قابلیت خوانایی	Rdbl
		تنوع لغوی	LD
	کنش گفتاری	بیانی	SA-D*
		درخواستی	SA-R*
		تهدیدی	SA-T*
		سوالی	SA-Q*
	معیار عاطفی	خشم	AN
		نفرت	D
		ترس	F
		پیش‌بینی	AC
		اعتماد	T
		غم	S
		تعجب	SU
	احساسات	قطبیت مثبت	PS
		قطبیت منفی	NS
	NER	موجودیت‌های نامدار (نام شخص، سازمان، مکان)	NE
	نهی	نفی (نه، هرگز، هیچ‌کس، هیچ و ...)	Neg
	هیجان	هیجان (نسبت مجموع صفت، قید و افعال حسی و حرکتی به کل کلمات)	Emtv*
	ارزش خبری	ارزش خبری (مؤلفه‌های ارتقاءدهنده ارزش‌خبری متن)	News*

[1] Informal
[2] Bad words
[3] http://www.farsishahri.com
[4] Slang words
[5] Emoticon
[6] https://unicode.org/emoji/charts/full-emoji-list.html, Available in 1 April 2019.
[7] Spelling Mistake
[8] https://github.com/eteamin/async_faspell
[9] Consecutive Words
[10] Consecutive Chars

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

۲-۲-۳- محتوای غیرفیزیکی شایعات

محتویات غیرفیزیکی عبارتند از نظرات[1]، هیجانات[2]، کنش‌گفتار[3] و احساساتی[4] که سازندگان شایعه می‌خواهند آن را بیان کنند. محتوای غیرفیزیکی هسته اصلی شایعات است، زیرا شامل همه ایده‌ها ، احساسات و دیدگاه‌های مهمی است که نویسندگان می‌خواهند به خوانندگان منتقل کنند. قطبیت احساسات یکی از ویژگی‌های مهم محتوای غیرفیزیکی برای شایعات است. سازندگان شایعه به‌منظور ایجاد اخبار قانع‌کننده خود، اغلب ابراز احساسات مثبت یا منفی را در متن با قدرت بیشتری بیان می‌کنند. جدول (۳)، مجموعه ویژگی‌های محتوایی غیرفیزیکی به‌کار رفته در پژوهش را نمایش می‌دهد.

۱-۲-۲-۳- پیچیدگی متن

پیچیدگی متن از جمله ویژگی‌های محتوایی غیرفیزیکی است که می‌تواند اطلاعات مفیدی از آن به‌دست آید. پنج فاکتور ارزیابی‌شده برای تعیین پیچیدگی اجزای متن عبارتند از: (۱) پیچیدگی واژگان (WL)[5] (میانگین نویسه‌های موجود در واژگان)، (۲) پیچیدگی جملات (SL)[6] (میانگین تعداد واژگان موجود در هر جمله)، (۳) پیچیدگی گرامری (عمق درخت وابستگی (DDT)[7])، (۴) تنوع لغوی(LD)[8]: شایعه‌پردازان به‌منظور جلب‌توجه مخاطب به موضوع شایعه واژگان مهم و عاطفی مرتبط با موضوع شایعه را در متن داستان تکرار می‌کنند، از این‌رو، تنوع لغوی (LD) متن کاهش می‌یابد. برای محاسبه معیار LD از فرمول (۳) استفاده شده است:

$$Score_{LD}(T) = \frac{|V_{uniq}(T)|}{|V(T)|} \qquad (۳)$$

در فرمول (۳)، $|V_{uniq}(T)|$ تعداد لغات یکتای موجود در متن T است و $|V(T)|$ به تعداد کل لغات موجود در متن اشاره می‌کند.

(۵) قابلیت خوانایی (Rdbl)[9] متن: در این پژوهش، از یک رویکرد خودکار مبتنی بر روش‌های یادگیری ماشین [25] برای ارزیابی قابلیت خوانایی متون فارسی استفاده شده است. نتایج ارزیابی این مدل بر روی متون شایعه و غیرشایعه فارسی نشان داد که قابلیت خوانایی شایعات نسبت به متون غیرشایعه پایین‌تر است. قابلیت خوانایی پایین می‌تواند به‌شدت تحت تاثیر جملاتی با طول کوتاه باشد. از آنجائیکه شایعه‌پردازان از جملات کوتاه برای بیان داستان‌شان استفاده می‌کنند، آن می‌تواند یکی از دلایل پایین‌بودن قابلیت خوانایی شایعات باشد.

۲-۲-۲-۳- ویژگی‌های عمل‌گرا

عمل‌گرایی سطح بالایی از پردازش داده‌های متنی و گفتاری است که دربارهٔ معنای متفاوت یک جمله در متون مختلف قضاوت می‌کند. به‌عبارتی دیگر، عملگرایی به کاربردهای زبان برای رساندن یک مطلب به مخاطب یا مخاطبان، در حالت عملی و یا در نوشتار و گفتار طبیعی می‌پردازد. ویژگی‌های مرتبط با عمل‌گرایی که در این پژوهش بر روی متون شایعه و غیرشایعه بررسی شده است، عبارتند از: کنش‌گفتار (SA)[10]، تحلیل احساسات[11] در دو ردهٔ مثبت و منفی، موجودیت‌های نامدار(NE)[12]، واژه‌های عاطفی[13] در هفت دسته، نفی (Neg)[14]، هیجان‌انگیزی (Emtv)[15] و ارزش خبری (News)[16]. جدول (۳) مجموعه‌ویژگی‌های محتوایی غیرفیزیکی به‌کاررفته در این پژوهش را نمایش می‌دهد. چهار رده SA، معیار هیجان(Emtv) و ارزش خبری(News) به‌عنوان ویژگی‌های عمل‌گرای پیشنهادی جدید هستند که با علامت ستاره (*) مشخص شده‌اند.

- **کنش‌گفتار:** عملی است که به‌وسیله گوینده با نطقی خاص انجام می‌شود، مانند SA عذرخواهی، تعریف و تمجید، قسم‌خوردن، تهدید، درخواست و غیره [26]. درک SA یک متن می‌تواند به بهبود تجزیه و تحلیل آن متن و همچنین درک بهتر از وضعیت ذهن نویسنده کمک کند. برای تشخیص SA شایعات، از یک دسته‌بند SA باناظر [27] استفاده شده است که SA متون فارسی را بر اساس ویژگی‌های محتوایی تعیین می‌کند. نتایج پژوهش قبلی ما [27] نشان می‌دهد که شایعات به‌طورمعمول با SA بیانی (SA-D)[17] (انتقال اطلاعات به شنونده.)، تهدیدی (SA-T)[18] (تهدید شنونده برای انجام کاری خاص که نمی‌خواهد انجام

[1] Opinions
[2] Eemotions
[3] Speech Act
[4] Sentiment
[5] Word Length
[6] Sentence Length
[7] Dependency parse tree
[8] Lexical Diversity
[9] Readability

[10] Speech Act
[11] Sentiment
[12] Named Entity
[13] Emoticons
[14] Negation
[15] Emotiveness
[16] Newsworthy
[17] Speech Ac- Declaration
[18] Speech Act-Threat

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

دهد.)، سؤالی (SA-Q)[1] (جملات سؤالی جهت پرسش، یا جهت تصدیق) و در برخی موارد با SA درخواستی (SA-R)[2] (درخواست مودبانه از کسی برای انجام‌دادن یا متوقف‌کردن چیزی) بیان می‌شوند؛ ازاین‌رو، این چهار رده SA به‌عنوان چهار ویژگی با یک مقدار عددی مابین صفر و یک در فرآیند تشخیص شایعات درنظر گرفته شده است.

- **معیار عاطفی:** شایعه‌پردازان از قدرت واژه‌های عاطفی برای ایجاد ترس، نگرانی، خشم و نفرت در مخاطب و تحریک او برای انتقال احساساتش به دیگران استفاده می‌کنند؛ از این‌رو، وجود واژه‌های هیجانی در متن باعث افزایش سرعت انتشار آن می‌شود. برای تشخیص نوع عاطفی لغات موجود در متن، از لغت‌نامه احساسی[3] شورای پژوهش‌های ملی کانادا (NRC)[4] استفاده شده که توسط محمد و تورنی [28] فراهم شده است. NRC شامل ۱۴۱۸۳ واژه انگلیسی و ارتباطات آنها با هشت احساس اولیه (خشم (AN)[5]، ترس (F)[6]، پیش بینی (AC)[7]، اعتماد(T)[8]، تعجب (SU)[9]، غم(S)[10]، لذت(J)[11] و نفرت (D)[12]) و دو نوع احساسی (منفی و مثبت) است. محمد و تورنی [28] علاوه‌بر زبان انگلیسی نسخه‌های مختلفی از این لغت‌نامه را در بیش از صد زبان دیگر از جمله زبان فارسی با استفاده از ترجمه گوگل فراهم کرده‌اند. ازآنجایی‌که، این ترجمه‌های ماشینی حاوی اشتباهات زیادی بود، از سه فرد خبره برای تصحیح این ترجمه‌ها کمک گرفته شد. به این ترتیب، نوع عاطفی لغات موجود در متن با استفاده از NRC به‌صورت یک عدد باینری (صفر یا یک) استخراج شده و معیار عاطفی متن در هفت دسته با استفاده از فرمول (۱) محاسبه شده است.

- **تحلیل احساسات:** هدف از تحلیل احساسات، پی‌بردن به این واقعیت است که اطلاعات موجود در متن دارای چه نوع احساسی است (مثبت یا منفی). در این پژوهش، از روش مبتنی بر لغت‌نامه[13] NRC برای تحلیل احساسات متن استفاده شده است. به‌این صورت که، ابتدا نوع احساسی لغات موجود در متن تعیین می‌شوند؛ سپس با استفاده از فرمول (۴ و ۵) به‌ترتیب معیار احساسی مثبت ($Score_{PSent}(T)$) و منفی ($Score_{NSent}(T)$) متن T محاسبه می‌شود:

$$Score_{PSent}(T) = \frac{|PSntm(S)|}{|PSntm(S)| + |NSntm(S)|} \quad (۴)$$

$$Score_{NSent}(T) = \frac{|NSntm(S)|}{|PSntm(S)| + |NSntm(S)|} \quad (۵)$$

در فرمول‌های (۴ و ۵)، $|PSntm(S)|$ تعداد عبارات مثبت در جمله S است و $|NSntm(S)|$ تعداد عبارات منفی در جمله S است.

- **موجودیت‌های نامدار (NER)**[14]: ذکر اسامی افراد معروف، دانشمندان، فیلسوفان، سازمان‌ها یا مؤسسات به‌عنوان یک موضوع یا یک منبع خبری باعث افزایش ارزش خبری پیام می‌شود و مخاطب بر اساس محبوبیت یا نفرت از آن موجودیت نامدار، پیام را منتشر می‌کند. موجودیت‌های نامدار موجود در متن با استفاده از روشی مبتنی بر HMM [29] در سه دستهٔ نام فرد، سازمان و مکان استخراج شده است. همچنین، برای محاسبه میانگین رخداد موجودیت‌های نامدار در یک متن از فرمول (۱) استفاده شده است.

- **هیجان:** در این پژوهش، چهار فاکتور صفت، قید، افعال حسی و حرکتی به‌عنوان عوامل تأثیرگذار هیجانی معرفی شده‌اند که قابلیت افزایش حس هیجانی را در خواننده و تغییر نگرش او نسبت به موضوع دارند؛ بنابراین، ویژگی "هیجان" (Emtv)[15] به‌عنوان یک ویژگی جدید پیشنهاد شده است که با استفاده از فرمول (۶) محاسبه می‌شود:

$$Score_{Emtv}(T) = \frac{|Adj(T)| + |Adv(T)| + |SV(T)| + |MV(T)|}{|W(T)|} \quad (۶)$$

در فرمول (۶)، نسبت مجموع تعداد صفت (Adj)[16]، قید (Adv)[17]، افعال حسی (SV) و حرکتی (MV) به تعداد کل واژگان W موجود در متن T محاسبه می‌شود.

[1] Speech Act-Question
[2] Speech Act-Request
[3] http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
[4] National Research Council Canada
[5] Anger
[6] Fear
[7] Anticipation
[8] Trust
[9] Surprise
[10] Sadness
[11] Joy
[12] Disgust
[13] Lexicon Based
[14] Named Entities Recognition
[15] Emotiveness
[16] Adjective
[17] Adverb

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

• **ارزش خبری**[1]: یکی از فاکتورهای تأثیرگذار در پذیرش شایعه توسط مخاطب، ارزش خبری آن شایعه است. یک خبر زمانی ارزش خبری بالایی دارد که (۱) جدید باشد، (۲) جزو علاقه‌مندی‌های مخاطب باشد، (۳) دارای اطلاعات آماری در ارتباط با موضوع باشد، (۴) با قطعیت بیان شود و (۵) در مورد رویدادهای غیر عادی، استثنایی، عجیب یا نادر باشد. در این پژوهش، معیار "ارزش خبری" به‌عنوان یک ویژگی محتوایی جدید در تحلیل شایعات پیشنهاد شده است. این پژوهش، مجموعه متغیرهای محتوایی، شامل زمان نسبی (RT)، موجودیت‌های نامدار (NE)، اطلاعات آماری (SI)[2] (نویسه‌های عددی و نمادهای ٪ $ و ...) و واژگان مرتبط با قطعیت (Cer) را معرفی می‌کند که باعث افزایش ارزش‌خبری یک متن می‌شوند. از طرفی، وجود اشتباهات املایی در متن از ارزش‌خبری آن می‌کاهد. به‌این ترتیب، ارزش‌خبری یک متن با استفاده از فرمول (۷) محاسبه می‌شود:

$$Score_{Emtv}(T) = \frac{|RT(T)| + |NE(T)| + |SI(T)| + |Cer(T)|}{|V(T)| + |SM(T)|} \qquad (۷)$$

در فرمول (۷)، مجموع چهار فاکتور متنی افزاینده ارزش خبری (RT، NE، SI و Cer) نسبت به مجموع تعداد واژگان (V) موجود در متن T و تعداد اشتباهات املایی محاسبه می‌شود.

۳-۳- نمایش متن

نمایش متن یکی از روش‌های پیش‌پردازش است که پیچیدگی متن را کاهش می‌دهد و باعث می‌شود که تحلیل متن برای دسته‌بند ساده‌تر شود. یکی از روش‌های اساسی برای نماش برداری متن، مدل فضایی برداری (VSM)[3] است که باعث ایجاد بردارهایی با ابعاد بالا می‌شود که دارای عناصری غیرمفید نیز است [30]؛ زیرا در این روش تمامی عناصر موجود در متن به‌عنوان ویژگی در نظر گرفته می‌شوند؛ اما در این پژوهش، تنها به ویژگی‌های لغوی اکتفا نشده است، بلکه مجموعه‌ای شامل پنجاه ویژگی محتوایی فیزیکی و غیرفیزیکی از متن استخراج می‌شود؛ بنابراین، به‌ازای هر متن ورودی یک بردار پنجاه بعدی ایجاد می‌شود که هر بعد بیانگر یکی از ویژگی‌های متمایزکننده بین شایعات و غیرشایعات است.

[1] Newsworthy
[2] Statistical information
[3] Vector Space Model

$$\begin{bmatrix} & F_1 & F_2 & \dots & F_n & C \\ T_1 & w_{11} & w_{12} & \dots & w_{1n} & C_j \\ T_2 & w_{21} & w_{22} & \dots & w_{2n} & C_j \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ T_n & w_{n1} & w_{n2} & \dots & w_{nn} & C_j \end{bmatrix}$$

(شکل-۳): نمایش برداری متون آموزشی به‌صورت ماتریس متن-ویژگی

(Figure-3): The vector representation of training texts as text-feature matrix.

به این ترتیب، مجموعه‌ای از n متن در مدل فضای برداری با ماتریس متن-ویژگی[4] (شکل ۳) نمایش داده می‌شود. فرض می‌شود $T = \{T_1^j, T_2^j, T_3^j, \dots, T_n^j\}$ مجموعه‌ای از n متن آموزشی در دو رده $(C_j: j = 0,1)$ شایعات $(j = 0)$ و غیرشایعات $(j = 1)$ باشد. متن T_i با یک بردار وزن‌دار $T_i = \{w_{i,1}, w_{i,2}, w_{i,3}, \dots, w_{i,m}\}$ نمایش داده می‌شود، که $w_{i,j}$ یک مقدار عددی در محدوده صفر و یک است که نشان‌دهندهٔ وزن مجموعه‌ویژگی‌های f_j در متن T_i است. این بردارهای ویژگی به‌عنوان ورودی دسته‌بند در نظر گرفته می‌شوند.

۴- مجموعه‌داده

در این پژوهش، مدل پیشنهادی برای دسته‌بندی متون شبکه‌های اجتماعی در دو دسته شایعه و غیرشایعه بر روی دو مجموعه‌داده توییتر و تلگرام ارزیابی شده است؛ ازاین‌رو، مجموعه‌ای از پست‌های منتشرشده بر روی تلگرام و توییتر به‌منظور ارزیابی مدل پیشنهادی جمع‌آوری شده‌اند.

۴-۱- تلگرام[5]

سرویس پیام‌رسان فوری تلگرام بیش از پنجاه میلیون کاربر در ایران دارد و شصت درصد از پهنای باند اینترنت برای استفاده از این برنامه صرف شده است؛ بنابراین، با توجه به محبوبیت تلگرام در بین مردم ایران، این پژوهش، شایعات منتشرشده در تلگرام را مورد تجزیه و تحلیل قرار داده است. برای این منظور، با استفاده از API ارائه‌شده توسط آزمایشگاه سامانه‌های هوشمند رایانه‌ای (ComInSys)[6] دانشگاه تبریز، چندین هزار پست تلگرام فارسی از یکم می ۲۰۱۷ تا سی‌ام مارس ۲۰۱۸ از کانال‌های عمومی مختلف جمع‌آوری شده است که تحت عنوان Sepehr_RumTel01 [31] در دسترس است. این

[4] Feature-Text
[5] Telegram
[6] www.cominsys.ir

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

API به‌ازای هر پیام تلگرامی قابلیت استخراج داده‌هایی را دارد که عبارتند از: شناسه پیام(MessageId)، تاریخ انتشار پیام (Date)، شناسه فرستنده(Sender)، شناسه گیرنده (Reciever)، متن پیام (Text) که حاوی شکلک‌های اینترنتی نیز است، نام اطلاعات محتوایی موجود در متن (ContentTypeName) که شامل سه مقدار "فیلم"، "تصویر" یا "تهی" است و مقدار ویژگی محتوایی (Content) که شامل یک پیوند به فیلم یا تصویر موجود در پست تلگرامی است. مجموعه‌داده Sepehr_RumTel01 از کانال‌های تلگرامی سه تارنمای ایرانی، گمانه (Gomaneh.com)، ویکی‌هوآکس (wikihoax.org) و آنتی‌شایعه (Shayeaat.ir) استخراج شده است که این تارنماها شایعات فارسی را تحلیل و اعتبارسنجی می‌کنند. با کمک سه فرد خبره، متن کامل ۸۸۲ شایعه تلگرامی با موضوعات متفاوت از این مجموعه‌داده استخراج، بررسی و با مقدار '۰' برچسب‌گذاری شده است؛ سپس، به همان تعداد پست غیرشایعه تلگرامی با موضوعات مختلف از کانال تلگرامی سایت‌های خبری معتبر به مجموعه‌داده اضافه و با مقدار '۱' برچسب‌گذاری شده است. برای انتخاب این سایت‌های خبری از نتایج تقوی‌فرد و همکاران [32] کمک گرفته شده است. آنها چندین سایت خبری را بر اساس ابعاد مختلف کیفیتی مورد بررسی قرار داده و درنهایت ۶۱ سایت خبری را به‌عنوان سایت‌های خبری با کیفیت معرفی کرده‌اند.

۲-۴- توییتر

توییتر یک سرویس شبکه اجتماعی میکرو وبلاگ‌نویسی است که کاربران می‌توانند پیام‌های کوتاه (توییت) تا ۱۴۰ نویسه را منتشر و تبادل کنند. قابلیت دسترسی، سرعت و سهولت استفاده توییتر، آن را به یک رسانه ارزشمند اجتماعی برای مقاصد مختلف تبدیل کرده است. این سرویس به‌عنوان یکی از رسانه‌های اصلی برای پژوهش در حوزه NLP عمل می‌کند. برای ارزیابی مدل پیشنهادی برای تشخیص شایعات بر روی توییت‌های توییتر، از مجموعه‌داده‌های برچسب‌گذاری‌شده توسط زمانی و همکاران [12] استفاده شده است. این مجموعه‌داده شامل ۷۸۳ توییت شایعه فارسی با موضوعات مختلف و ۷۸۳ توییت تصادفی فارسی است.

همچنین، مدل پیشنهادی بر روی مجموعه‌داده KNTUPT[1] [13] مورد ارزیابی قرار گرفته است. در این مجموعه‌داده تمام توییت‌های فارسی را از ۲۴ نوامبر ۲۰۱۷ تا ۸ دسامبر ۲۰۱۷ در مورد زلزله کرمانشاه جمع‌آوری شده است، که شامل ۳۵۹۸۰۴۹ توییت و ۴۳۴۵ توییت شایعه است. درمجموع، شصت شایعه مختلف در مورد زلزله کرمانشاه در این مجموعه وجود دارد که ۵۸ مورد مربوط به شایعات نادرست و دو مورد شایعه درست است.

۵- آزمایش‌ها و ارزیابی‌ها

در این پژوهش، آزمایش‌های متعددی با ترکیب‌های مختلفی از مجموعه ویژگی‌های محتوایی رایج و پیشنهادی جدید انجام شده است تا کارایی ویژگی‌ها به‌صورت جداگانه و مشترک در فرآیند تشخیص شایعات مشخص شود. برای این منظور، از الگوریتم‌های یادگیری ماشین باناظر برای انجام عملیات یادگیری و دسته‌بندی استفاده شده است. هر الگوریتم دسته‌بند معایب و مزایایی دارد. بر اساس نتایج و بحث‌های انجام‌شده بر روی الگوریتم‌های دسته‌بندی در پژوهش‌های پیشین [35]–[33]، سه الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF)[2] [36]، AdaBoost [37] و ماشین بردار پشتیبان (SVM)[3] [38] به‌عنوان دسته‌بندی‌های قوی برای ارزیابی دقت مدل پیشنهادی به‌کار گرفته شده است. برای پیاده‌سازی این الگوریتم‌های یادگیری ماشین، کتابخانه scikit-learn در پایتون مورد استفاده قرار گرفته است.

اعتبارسنجی متقابل 10-fold. به‌منظور ارزیابی عملکرد مدل پیشنهادی در تعیین دو رده شایعه و غیر شایعه، و نشان‌دادن اینکه نتایج این مدل تا چه اندازه قابل تعمیم و مستقل از داده‌های آموزشی است، از روش اعتبارسنجی متقابل 10-fold استفاده شده است. نتایج این ارزیابی‌ها بر اساس معیارهای ارزیابی دقت[4]، بازخوانی[5]، معیار F[6] و معیار سطح زیر منحنی ROC[7] نمایش داده شده است.

(جدول-۴): ارزیابی مدل پیشنهادی بر روی مجموعه‌دادهٔ توییتر با استفاده از سه دسته‌بند RF، AdaBoost و SVM.

(Table-4): Evaluation of the proposed model on dataset of Twitter using three classifier RF, AdaBoost and SVM.

RF				
	دقت	بازخوانی	Fمعیار	ROC
شایعات	۰/۸۳۴	۰/۸۱۴	۰/۸۲۴	۰/۸۲۶
غیرشایعات	۰/۸۱۸	۰/۸۳۸	۰/۸۲۸	۰/۸۲۶
میانگین	۰/۸۲۶	۰/۸۲۶	۰/۸۲۶	۰/۸۲۶
AdaBoost				
شایعات	۰/۸۱۱	۰/۸۰۴	۰/۸۰۸	۰/۸۳۳

[1] The dataset is available on https://trlab.ir/res.php?resourceid=3978-1-5386-8274-6/18/$31.00 2018 IEEE

[2] Random Forest
[3] Support Vector Machine
[4] Perision
[5] Recall
[6] F-measure
[7] Receiver Operating Characteristic

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

غیرشایعات	۰/۸۱۶	۰/۸۲۳	۰/۸۲۰	۰/۸۳۳
میانگین	۰/۸۱۴	۰/۸۱۴	۰/۸۱۴	۰/۸۳۳
SVM				
شایعات	۰/۷۴۷	۰/۸۱۶	۰/۷۸۰	۰/۸۸۰
غیرشایعات	۰/۸۵۱	۰/۷۹۲	۰/۸۲۱	۰/۸۸۰
میانگین	۰/۸۰۶	۰/۸۰۲	۰/۸۰۳	۰/۸۸۰

(جدول-۵): ارزیابی مدل پیشنهادی بر روی مجموعه‌دادهٔ تلگرام با استفاده از سه دسته‌بند RF، AdaBoost و SVM.

(Table-5): Evaluation of the proposed model on dataset of Telegram using three classifier RF, AdaBoost and SVM.

	دقت	بازخوانی	Fمعیار	ROC
RF				
شایعات	۰/۸۶۲	۰/۸۷۵	۰/۸۶۸	۰/۹۵۰
غیرشایعات	۰/۸۷۳	۰/۸۶۰	۰/۸۶۶	۰/۹۵۰
میانگین	۰/۸۶۷	۰/۸۶۷	۰/۸۶۷	۰/۹۵۰
AdaBoost				
شایعات	۰/۸۳۷	۰/۸۱۷	۰/۸۲۷	۰/۹۰۱
غیرشایعات	۰/۸۲۲	۰/۸۴۱	۰/۸۳۱	۰/۹۰۱
میانگین	۰/۸۲۹	۰/۸۲۹	۰/۸۲۹	۰/۹۰۱
SVM				
شایعات	۰/۸۱۴	۰/۷۸۵	۰/۷۹۹	۰/۸۰۳
غیرشایعات	۰/۷۹۳	۰/۸۲۱	۰/۸۰۷	۰/۸۰۳
میانگین	۰/۸۰۳	۰/۸۰۳	۰/۸۰۳	۰/۸۰۳

در جداول (۴ و۵) نتایج ارزیابی مدل پیشنهادی را به‌ترتیب بر روی دو مجموعه‌داده توییتر و تلگرام با استفاده از سه دسته‌بند RF، AdaBoost و SVM نمایش داده شده است. نتایج ارزیابی به‌وسیله سه دسته‌بند نشان می‌دهد که دسته‌بند RF بهترین نتایج را ارائه می‌کند. از این‌رو، در آزمایش‌های بعدی، RF به‌عنوان بهترین دسته‌بند به‌کار گرفته شده است.

تجزیه و تحلیل آماری ویژگی‌های محتوایی. در این بخش، توزیع آماری مجموعه ویژگی‌های محتوای معرفی‌شده در این پژوهش برای دسته‌بندی شایعات و غیرشایعات مورد ارزیابی قرار گرفته است. به‌منظور تشخیص ویژگی‌های متنی که از لحاظ آماری مابین شایعات و غیرشایعات اختلاف معناداری دارند، آزمون تی[1] به‌کار گرفته شده است. با این فرض که هر یک از ویژگی‌های محتوایی که با فرکانس‌های مختلفی در شایعات و غیرشایعات ظاهر می شود، آن ویژگی می‌تواند شایعات و غیرشایعات را از یکدیگر متمایز کند.

نتایج (P-value (<= 0.05 نشان می‌دهد که بیشتر این ویژگی‌ها دارای تفاوت‌های آماری معناداری بین شایعات و غیر شایعات هستند. برای مثال، بر اساس نتایج آزمون تی، میانگین معیار خوانایی شایعات بیشتر از غیرشایعات است، این نتیجه نشان می‌دهد که شایعه‌پردازان اغلب از جملات کوتاه‌تر با ساختار زبانی به‌ندرت پیچیده استفاده می‌کنند. همچنین، میانگین فراوانی ویژگی‌های مبتنی بر قطعیت در شایعات بیشتر است. این نتیجه از آنجا ناشی می‌شود که شایعه‌پردازان با استفاده از واژگان مرتبط با قطعیت قصد جلب اعتماد مخاطب را دارند؛ علاوه‌براین، شایعه‌پردازان از ضمایر اول شخص جمع و سوم شخص برای بیان شایعه استفاده می‌کند تا اینکه یا میزان قطعیت داستان را بالا ببرند و یا نقل شایعه را به دیگران ارجاع دهند؛ علاوه‌براین، بر اساس نتایج تی تست برای ویژگی‌هایی همچون واژگان عامیانه و اشتباه املایی اثبات شده است که شایعات دارای استایل نوشتاری غیررسمی هستند. جدول (۶) وزن متوسط هر یک از ویژگی‌ها را در دو مجموعه‌داده شایعه و غیرشایعه نشان می‌دهد. بر اساس نتایج ارزش P، ویژگی‌هایی که از لحاظ آماری در دو دسته شایعه و غیرشایعه قابل‌توجه نیستند عبارتند از: عدم قطعیت، مکث، نفی و نفرت و کنش‌گفتار درخواستی.

همچنین، نتایج آماری آزمون تی در جدول (۶) ثابت کرد که ویژگی‌های پیشنهادی جدید در این پژوهش شامل CC، CW، SP، EP، MV، RT، Emtv، News و سه SAرایج (SA_Dec، SA_Thrt و SA_Ques) دارای مقدار P کمتر از ۰/۰۵ هستند، و این نشان‌دهنده وجود تفاوت معنادار این ویژگی‌ها در دو دسته شایعه و غیرشایعه است؛ بنابراین، این ویژگی‌ها به‌عنوان ویژگی‌های متمایزکننده بین شایعات و غیرشایعات می‌توانند در فرآیند تشخیص شایعات مفید باشند.

انتخاب ویژگی. یک مؤلفهٔ بسیار مهم برای کاهش اندازه بردار ویژگی و بهبود عملکرد دسته‌بند است، که به شناخت ویژگی‌ها و اهمیت آنها کمک می‌کند [39]؛ بنابراین، برای رسیدن به بهترین عملکرد ممکن با الگوریتم‌های دسته‌بند بر روی مجموعه آموزشی، لازم است از روش‌های انتخاب ویژگی استفاده شود. در این پژوهش، از رویکرد جستجوی شناور رو به جلوی متوالی (SFFS)[2] به‌منظور انتخاب ویژگی‌هایی با ارزش اطلاعاتی بالا استفاده شده است. SFFS یکی از روش‌های انتخاب ویژگی از نوع بسته‌بندی[3] است که زیرمجموعه‌ای از ویژگی‌های ارزشمند را بر اساس ارزیابی عملکرد دسته‌بند (برای مثال، RF یا SVM) از مجموعه کل ویژگی‌ها انتخاب می‌کند. این روش می‌تواند دقت دسته‌بندی را بهبود بخشد، اما در مواقعی که تعداد ویژگی‌ها زیاد باشد، زمان انتخاب ویژگی به‌طور قابل توجهی افزایش می‌یابد.

[1] T-test

[2] Sequential Forward Floating Search

[3] Wrapper

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

(جدول-۶): میانگین فرکانس ۵۰ ویژگی همراه با مقدار P آزمون تی تست برای مقایسه فرکانس ویژگی‌های متنی بین اسناد متنی شایعه و غیرشایعه در پست‌های تلگرام. (ویژگی‌های پیشنهادی جدید با علامت * مشخص شده‌اند)

(Table-6): The average frequency of 50 features, along with the p-value of a student t-test for comparing frequencies in Rumor versus Non-Rumor groups in Telegram posts.

	W	S	P	SC	NLC	NW	Qty	Cer	UCer	QW	IW	SV	RT*	MV*	SP*	EP*	Emtv*
میانگین فرکانس در شایعات	۰/۰۷۰	۰/۰۹۷	۰/۰۹۰	۰/۰۰۸	۰/۰۵۶	۰/۰۵۸	۰/۱۹۲	۰/۱۵۰	۰/۶۰۶	۰/۱۰۱	۰/۲۰۹	۰/۰۲۴	۰/۳۸۰	۰/۴۸۶	۰/۵۰۲	۰/۴۴۱	۰/۵۴۵
میانگین فرکانس در غیرشایعات	۰/۰۳۸	۰/۰۲۳	۰/۰۳۴	۰/۰۰۲	۰/۰۳۸	۰/۰۳۲	۰/۲۲۷	۰/۰۸۸	۰/۶۱۱	۰/۰۴۳	۰/۱۵۴	۰/۰۰۵	۰/۳۱۰	۰/۶۰۹	۰/۲۰۰	۰/۱۴۰	۰/۳۴۸
P-value	۳/۸۴E-۱۹	۳/۰۷E-۳۷	۱/۱۲E-۳۰	۰/۰۸۳	۰/۰۳۹	۰/۰۰۵	۲/۷۶E-۰۹	۵/۴۵E-۱۲	۰/۳۴۵	۲/۱۵E-۰۷	۰/۰۰۱	۳/۳۰E-۰۵	۰/۰۰۱	۳/۶۷E-۰۸	۳/۴۵E-۲۹	۷/۸۳E-۳۲	۲/۳۰E-۱۳

	N	V	Adj	Adv	Pro	WL	SL	DDT	Pous	Rdbl	LD	SM	BW	SW	Emj	CW	CC
میانگین فرکانس در شایعات	۰/۰۸۲	۰/۰۵۷	۰/۱۰۰	۰/۰۹۲	۲/۵۸۱	۰/۳۵۵	۰/۸۰۰	۰/۱۹۲	۰/۰۹۳	۰/۵۹۳	۰/۷۶۱	۰/۰۵۲	۰/۰۸۲	۰/۴۴۱	۰/۶۹۴	۰/۰۲۵	۰/۰۵۰
میانگین فرکانس در غیرشایعات	۰/۰۴۸	۰/۰۳۰	۰/۰۶۶	۰/۰۴۱	۱/۹۶۶	۰/۴۴۴	۰/۱۰۷	۰/۲۲۷	۰/۰۹۴	۰/۵۴۳	۰/۸۱۳	۰/۰۴۰	۰/۰۰۳	۰/۱۴۰	۰/۳۷۴	۰/۰۰۷	۰/۰۰۸
P-value	۹/۲۳E-۱۸	۱/۸۱E-۱۳	۵/۸۷E-۱۱	۱/۲۹E-۲۰	۱/۷۱E-۲۸	۲/۳۲E-۳۰	۳/۹۷E-۱۰	۲/۲۵E-۰۹	۰/۴۲۶	۰/۰۱۷	۶/۲۸E-۱۲	۰/۰۰۲	۰/۰۰۱	۷/۸۳E-۳۲	۱/۱۱E-۴۴	۰/۰۰۱	۲/۲۳E-۰۸

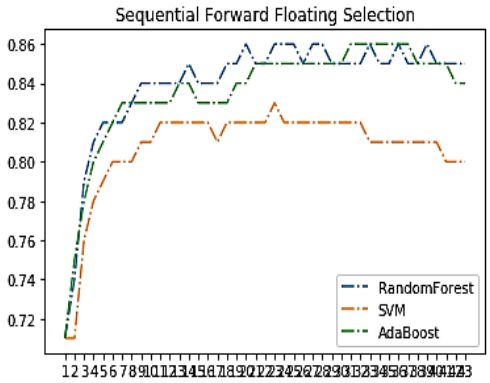
	News	SA-D	SA-R	SA-T	SA-Q	PS	NS	NE	AN	D	F	AC	T	S	SU	Neg
میانگین فرکانس در شایعات	۰/۷۷۹	۰/۳۱۸	۰/۰۰۲	۰/۰۳۸	۰/۰۷۰	۰/۳۹۲	۰/۳۰۱	۰/۵۰۹	۰/۴۷۶	۰/۲۸۴	۰/۶۱۰	۰/۵۷۷	۰/۹۰۹	۰/۵۴۳	۰/۳۵۰	۰/۰۸۱
میانگین فرکانس در غیرشایعات	۰/۵۰۹	۰/۵۷۲	۳/۲۷E-۰۵	۰/۰۱۴	۰/۰۲۴	۰/۴۵۷	۰/۲۳۴	۰/۷۷۸	۰/۳۲۴	۰/۲۵۰	۰/۴۳۳	۰/۴۳۵	۰/۹۳۵	۰/۴۱۱	۰/۲۳۰	۰/۰۷۰
P-value	۱/۴۵E-۳۶	۹/۹۲E-۲۱	۰/۱۵۹	۰/۰۰۳	۲/۵۳E-۰۵	۰/۰۰۲	۰/۰۰۱	۳/۷۴E-۳۶	۴/۵۸E-۱۱	۰/۰۵۷	۳/۵۰E-۱۴	۵/۱E-۱۰	۹/۱۹E-۱۶	۱/۵۷E-۰۸	۳/۱۴E-۰۸	۰/۱۸۲

برای شناسایی ویژگی‌های مؤثر در شناسایی شایعات، روش SFFS یک‌بار بر روی مجموعه‌بردارهای ویژگی‌های استخراج‌شده از توییتر و بار دیگر بر روی تلگرام به‌کار گرفته شده است. در شکل (۴)، نتایج روش SFFS با استفاده از سه دسته‌بندRF ، AdaBoost و SVM نمایش داده شده است. همان‌طورکه نتایج نشان می‌دهد، بهنرین عملکرد دسته‌بند بر روی مجموعه‌داده تلگرام با انتخاب ۳۲ ویژگی از ۵۰ ویژگی دقتی در حدود ۸۶/۳ درصد و بر روی مجموعه داده توییتر با انتخاب ۴۰ ویژگی دقتی در حدود ۸۴/۲ درصد دارد.

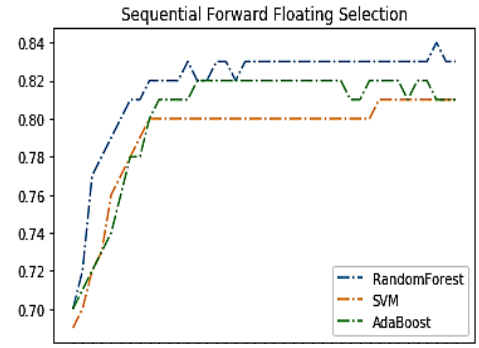
بر اساس نتایج SFFS، مجموعه ویژگی‌های موثر در تشخیص شایعات عبارتند از: Rdbl، SA-Q، SA-T، SA-D، Adj، Adv، SV، NW، NLC، QW، SM، RT، CC،CW، MV، LD، SL، WL، Emj، Pro، EP، SP، AD، D، S، T، UNcer، Emtv، News، Neg، NS، BW، AC،SW. این نتایج، قابلیت اغلب ویژگی‌های جدید در متمایزسازی شایعات از غیرشایعات را نشان می‌دهد.

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]



(الف) تلگرام



(شکل-۴): نرخ تشخیص روش SFFS براساس میانگین معیار-F با استفاده از ویژگی‌های مبتنی بر محتوی برای تشخیص شایعات فارسی منتشرشده در توییتر و تلگرام.

(Figure-4): Recognition rate of SFFS method based on average F-measure using contextual-based features to detect Persian rumors published in Twitter and Telegram.

ارزیابی تأثیر ویژگی‌های پیشنهادی جدید. مدل پیشنهادی بر اساس سه آزمایش مجزا با استفاده از دسته‌بند RF ارزیابی شده است. نتایج ارزیابی بر اساس معیارهای ارزیابی دقت، فراخوانی، و معیار F در جدول (۷) نمایش داده شده است. در نخستین آزمایش، تشخیص شایعه تنها بر اساس ویژگی‌های پیشنهادی در مطالعات پیشین انجام شد که کارایی دسته‌بند بر اساس معیار F، ۰/۷۶۵ است. در آزمایش دوم، ویژگی‌های پیشنهادی جدید شامل چهار ویژگی لغوی (RT، MV، SP و EP)، دو ویژگی نحوی (CC و CW) و شش ویژگی عمل‌گرا (Emtv، News، SA-Dec، SA-Req، SA-Thrt و SA-Ques)، به مجموعه‌ویژگی‌های محتوایی پیشین اضافه شد که عملکرد دسته‌بند از ۰/۷۶۵ به ۰/۸۵۰ بهبود یافت. این نتیجه نشان‌دهنده تأثیر مثبت دوازده ویژگی جدید پیشنهادی در بهبود دقت دسته‌بندی است. در آزمایش سوم، مجموعه‌ویژگی‌های انتخاب‌شده به‌وسیلهٔ روش SFFS از کل ویژگی‌های به‌کار گرفته‌شده در آزمایش دوم برای شناسایی شایعات به‌کار گرفته شده است. در این آزمایش، بهترین نتیجه به‌دست‌آمده بر اساس معیار F در حدود ۰/۸۶۷ است.

(جدول-۷): تأثیر متفاوت ویژگی‌های پیشین و پیشنهادی بر روی دقت دسته‌بندی پست‌های تلگرام در دو رده شایعه و غیر شایعه با استفاده از دسته‌بندRF در سه آزمایش مجزا.

(Table-7): Different impact of previuse and proposed features on the classification accuracy of Telegram posts in two rumor and non-rumor classes by RF classifier in three separate experimrnts.

	دقت	بازخوانی	Fمعیار-	ROC
ویژگی‌های متنی رایج در مطالعات پیشین				
شایعه	۰/۷۸۵	۰/۷۲۹	۰/۷۵۶	۰/۸۳۳
غیرشایعه	۰/۷۴۸	۰/۸۰۱	۰/۷۷۳	۰/۸۳۳
میانگین	۰/۷۶۶	۰/۷۶۵	۰/۷۶۵	۰/۸۳۳
ویژگی‌های رایج + ویژگی‌های پیشنهادی				
شایعه	۰/۸۴۵	۰/۸۵۶	۰/۸۵۰	۰/۹۲۴
غیرشایعه	۰/۸۵۴	۰/۸۴۳	۰/۸۴۹	۰/۹۲۴
میانگین	۰/۸۵۰	۰/۸۵۰	۰/۸۵۰	۰/۹۲۴
۳۲ ویژگی برتر				
شایعه	۰/۸۶۲	۰/۸۷۵	۰/۸۶۸	۰/۹۵۰
غیرشایعه	۰/۸۷۳	۰/۸۶۰	۰/۸۶۶	۰/۹۵۰
میانگین	۰/۸۶۷	۰/۸۶۷	۰/۸۶۷	۰/۹۵۰

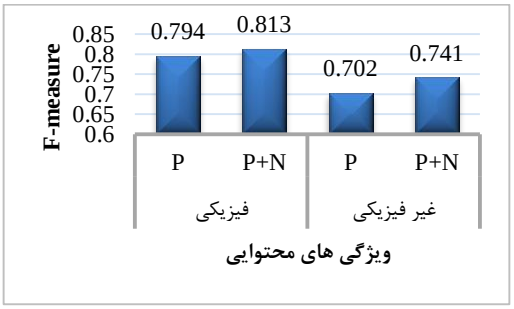
ارزیابی ویژگی‌های محتوایی فیزیکی و غیرفیزیکی. در این بخش، ۸ آزمایش مختلف (چهار ارزیابی بر روی مجموعه‌داده توییتر و چهار ارزیابی بر روی مجموعه‌داده تلگرام) به‌منظور ارزیابی تأثیر ویژگی‌های محتوایی فیزیکی و غیرفیزیکی در کارایی دسته‌بند برای تشخیص شایعات انجام شده است. برای این منظور، به‌ازای هر دسته ویژگی، در آزمایش نخست ویژگی‌های محتوایی پیشین (P) به‌کارگرفته شده‌اند، سپس در آزمایش بعدی ترکیبی از ویژگی‌های پیشین (P) و ویژگی‌های پیشنهادی جدید (N) به‌کار رفته است.

شکل (۵)، کارایی مؤثر ویژگی‌های پیشنهادی جدید را در دو مجموعه‌ویژگی‌های محتوایی فیزیکی و غیرفیزیکی نشان می‌دهد. همچنین، نتایج ارزیابی‌ها بیان‌گر این است که در مجموعه‌داده توییتر ویژگی‌های فیزیکی و در مجموعه‌داده تلگرام ویژگی‌های غیرفیزیکی تأثیر بیشتری در متمایزسازی شایعات از غیرشایعات داشته‌اند.

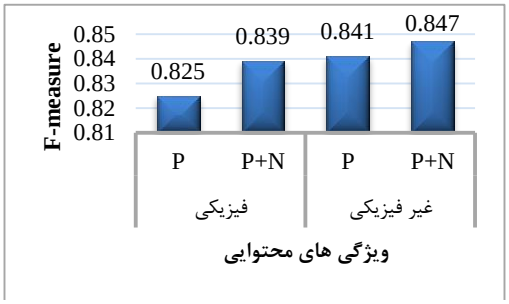
منحنی یادگیری. به‌منظور ارزیابی اینکه آیا افزایش داده آموزشی می‌تواند کارایی شناسایی شایعات را بهبود بخشد، فرآیند یادگیری RF به‌عنوان بهترین دسته‌بند مورد تحلیل قرار گرفته است. بنابراین، منحنی یادگیری بر اساس مجموعه ویژگی‌های محتوایی پیشین(P)، پیشنهادی(N) و ترکیب این دو مجموعه(P+N) با استفاده از مقادیر افزایشی داده‌ها رسم شده است که برای دو مجموعه‌داده توییتر و تلگرام به ترتیب در شکل (۶-الف و ۶-ب) نمایش داده شده است. به‌طور کلی، روند یادگیری

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

در هر دو مجموعه‌داده، بهبود مستمری را نشان می‌دهد، به این معنی که مقادیر بیشتر داده‌های آموزشی می‌تواند عملکرد دسته‌بندی را بهبود بخشد.



(الف) توییتر



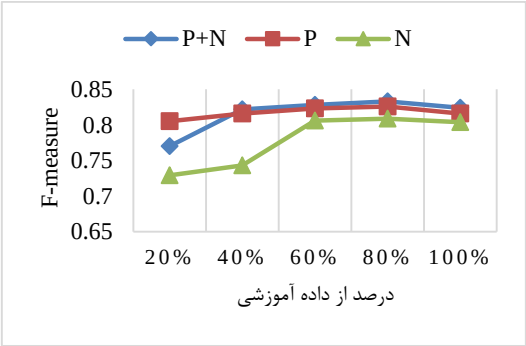
(شکل-۵): نمایش تأثیر دو دسته ویژگی‌های محتوایی فیزیکی و غیرفیزیکی پیشین (P) و پیشنهادی جدید (N) برای تشخیص شایعات.

(Figure-5): A demonstration of the impact of two categories of physical abd non-physical content features of privouse and new proposed to detect rumors.

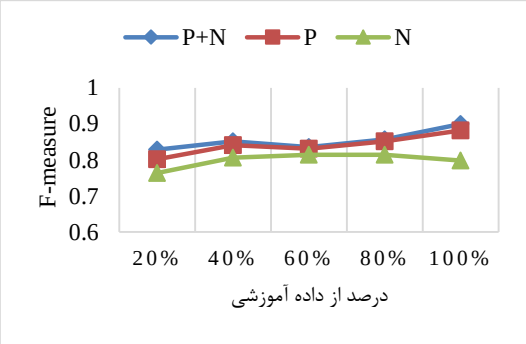
ارزیابی روش پیشنهادی در مقایسه با کارهای پیشین.

به‌منظور ارزیابی تأثیر ویژگی‌های پیشنهادی در مسأله دسته‌بندی شایعات و غیرشایعات و مقایسه با روش‌های پیشین، در این بخش، روش پیشنهادی با سه روش پیشین مقایسه شده است:

- DT[1]: کاستیلو و همکاران [2] لیستی از ویژگی‌های محتوایی را به‌همراه مجموعه ویژگی‌های ساختاری و کاربر برای اعتبارسنجی شایعات پیشنهاد دادند، و برای ارزیابی ویژگی‌های پیشنهادی از دسته‌بند درخت تصمیم (DT) استفاده کردند. در این آزمایش، تأثیر ویژگی‌های محتوایی معرفی‌شده در این پژوهش در مقایسه با ویژگی‌های محتوایی ارائه‌شده در [2] بر روی مجموعه‌داده تلگرام مورد ارزیابی قرار گرفته است.
- RDPT[2]: در این آزمایش، مدل پیشنهادی بر اساس ویژگی‌های انتخاب‌شده توسط SFFS با کار زمانی و همکاران [12] بر روی مجموعه‌داده توییتر آنها مورد ارزیابی قرار گرفته است.
- KNTUPT: به‌منظور مقایسه کارایی مدل پیشنهادی نسبت به کار محمودآباد و همکاران [20]، از مجموعه‌داده به‌کار رفته در [20] و [28] استفاده شده که این مجموعه‌داده شامل توییت‌های شایعه و غیرشایعه در مورد رویداد زلزله کرمانشاه است.



(الف) توییتر



(ب) تلگرام

(شکل-۶): منحنی‌های یادگیری و دقت دسته‌بندی بر اساس درصد افزایشی داده آموزشی و سه آزمایش مختلف.

(Figure-6): Learning curves and classification accuracy based on % training data and three different experiments.

(جدول-۸): مقایسه مدل پیشنهادی با مدل پایه کاستیلو [5] و دو کار [4] و [25] بر روی شایعات فارسی.

(Table-8): Comparison of the proposed method with Castilo's baseline model [5] and two works [4] and [25] on Persian rumors

مجموعه داده	روش	دقت	بازخوانی	Fمعیار-
DT	کاستیلو و همکاران [2]	۰/۶۹۹	۰/۶۹۶	۰/۶۹۵
	روش پیشنهادی	۰/۷۸۲	۰/۷۸۲	۰/۷۸۲
RDPT	زمانی و همکاران [12]	۰/۸۱۹	۰/۸۱۶	۰/۸۱۷
	روش پیشنهادی	۰/۸۴۸	۰/۸۴۸	۰/۸۴۸
KNTUPT	محمودآباد و همکاران [20]	./۹۵۹	۰/۹۶	۰/۹۶
	روش پیشنهادی	۰/۹۵۲	۰/۹۵۱	۰/۹۵۲

در جدول (۸) نتایج مقایسه روش پیشنهادی با کار پایهٔ انجام‌شده توسط کاستیلو و همکاران [2] و دو کار [12] و [20] که بر روی شایعات فارسی انجام شده نمایش داده شده است. همان‌طور که نتایج نشان می‌دهند، روش

[1] Decision Tree
[2] Rumor Detection for Persian Tweets

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]

پیشنهادی برای تشخیص شایعات تنها با تمرکز بر روی ویژگی‌های محتوایی شایعه منبع، به‌ترتیب با میانگین معیار F ۰/۸۴۸ و ۰/۹۵۲ عملکرد بهتری نسبت به کار زمانی و همکاران [12] و محمودآباد و همکاران [20] در شناسایی شایعات دارد. دلیل بالابودن دقت روش پیشنهادی در ارزیابی بر روی مجموعه‌داده KNTUPT نسبت به دو آزمایش دیگر این است‌که، این داده شامل مجموعه‌ای از توییت‌ها و ریتوییت‌ها و توییت‌های پاسخ کاربران در رابطه با رویداد زلزله کرمانشاه است. از آنجایی‌که، مجموعه‌داده آموزشی و آزمون بر روی یک رویداد می‌باشد؛ بنابراین دسته‌بند عملیات آزمون را بر روی مجموعه‌داده‌هایی انجام می‌دهد که درقبل آموزش دیده است. این در حالی است که مجموعه‌داده تلگرام [31] به‌کاررفته در آزمایش نخست (DT) و مجموعه‌داده توییتر [12] در آزمایش دوم (RDPT) شامل مجموعه‌ای از شایعات و غیرشایعات با موضوعات به‌طورکامل متفاوت است. ازاین‌رو، آموزش و آزمون دسته‌بند بر روی مجموعه شایعات و غیرشایعات با موضوعات متفاوت صورت می‌گیرد.

۶- نتیجه‌گیری و کارهای آینده

در این مقاله، مسأله تشخیص شایعات در مورد شایعات منتشرشده در پیام‌رسان تلگرام و شبکه اجتماعی توییتر مورد بررسی قرار گرفته است. برای حل این مسأله، بر روی اطلاعات متنی شایعات تمرکز شده و پست‌های مرتبط با شایعه از جمله Retweet و توییت پاسخ در نظر گرفته نشده است. به‌عبارتی هیچ پیشینه‌ای از شایعه مدنظر قرار نگرفته است. تشخیص شایعه بر اساس اطلاعات محتوایی شایعه منبع می‌تواند سامانه را از بسیاری از وابستگی‌ها جدا کند، زیرا در این مورد نیاز به اطلاعاتی در مورد تاریخچه انتشار شایعه همچون اطلاعات کاربر و ساختار شبکه انتشار وجود ندارد. براساس تحلیل محتوایی صدها متن شایعه و غیرشایعه، مجموعه‌ای از ویژگی‌های محتوایی فیزیکی و غیرفیزیکی که تأثیر به‌سزایی در شناسایی شایعات دارند معرفی شده است. این ویژگی‌ها ترکیبی از ویژگی‌های محتوایی رایج در مطالعات پیشین به‌همراه ویژگی‌های پیشنهادی جدید متنی است. نتایج تجربی نشان می‌دهد که ویژگی‌های محتوایی پیشنهادی در کنار ویژگی‌های پیشین می‌تواند کارایی دسته‌بند را در شناسایی شایعات بهبود ببخشد؛ به‌طوری‌که، شناسایی شایعات تلگرامی بر مبنای اطلاعات محتوایی با معیار F ۰/۸۷۶، شایعات توییتر [12] با ۰/۸۴۸ و رویداد زلزله کرمانشاه (KNTUPT) با معیار F ۰/۹۵۲ انجام شده است؛ بنابراین، نتایج ارزیابی مدل پیشنهادی بر روی شایعات توییتر نشان دادند که، با وجود طول کوتاه توییت‌های توییتر و استخراج اطلاعات محتوایی محدود از توییت‌ها، مدل پیشنهادی توانایی تشخیص شایعات توییتر را با دقت قابل قبولی دارد. این امر نشان‌دهنده توانمندی ویژگی‌های محتوایی در متمایزکردن شایعات از غیرشایعات است.

با توجه به تفاوت در شیوهٔ نوشتاری شایعات در حوزه سیاست، فناوری یا آموزش، ممکن است نیاز به توجه بیشتری به جنبه‌های زبانی سبک نوشتاری باشد؛ درحالی‌که برای شناسایی شایعات در قلمرو افراد مشهور، ممکن است، نیاز به تمرکز بر روی تفاوت‌های نوشتاری مرتبط با احساسات و درک مخاطب باشد؛ بنابراین، با تحلیل عمیق‌تر ویژگی‌های محتوایی در حوزه‌های متفاوت می‌توان به نتایج مؤثری در فرآیند شناسایی شایعات دست یافت.

۷- مراجع 7- References

[1] A. Zubiaga et al., “Detection and resolution of rumours in social media: A survey,” *ACM Comput. Surv.*, vol. 51, no. 2, p. 32, 2018.

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

[3] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, “Rumor has it: Identifying misinformation in microblogs,” *in Proceedings of the conference on empirical methods in natural language processing*, 2011, pp. 1589–1599.

[4] R. Dayani, N. Chhabra, T. Kadian, and R. Kaushal, “Rumor: Detecting Misinformation in Twitter,” in 3rd Security and Privacy Symposium, 2015.

[5] 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.

[6] S. Hamidian and M. T. Diab, “Rumor Detection and Classification for Twitter Data,” in the Fifth International Conference on Social Media Technologies, Communication, and Informatics, 2015.

[7] G. Giasemidis et al., “Determining the veracity of rumours on Twitter,” *in International Conference on Social Informatics*, 2016, pp.

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

185–205.

[8] K. Wu, S. Yang, and K. Q. Zhu, “False rumors detection on Sina Weibo by propagation structures,” *in 2015 IEEE 31st International Conference on Data Engineering*, 2015, vol. 2015-May, pp. 651–662.

[9] S. Hamidian and M. Diab, “Rumor Identification and Belief Investigation on Twitter,” *in Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2016, pp. 3–8.

[10] Z. Zhao, P. Resnick, and Q. Mei, “Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts,” *in Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1395–1405.

[۱۱]سمانه سلطانی‌پور، احسان رحیم‌زاده، "تحلیل محتوای اخبار و شایعات تکذیب شده در رسانه‌های مکتوب و آنلاین ایران در آستانه برگزاری انتخابات دهمین دوره مجلس شورای اسلامی"، دومین همایش ملی رسانه، ارتباطات و آموزش‌های شهروندی، تهران، ۱۱ اسفند ۱۳۹۵.

[11] E. Rahim zade, S. Soltanipour, “Analysis of the content of denied news and rumors in the print and online media of Iran on the eve of the 10th term of the Islamic Consultative Assembly,” in the second national conference on media, communications and education Citizenship, Tehran, 2017, (In Persian).

[12] S. Zamani, M. Asadpour, and D. Moazzami, “Rumor detection for Persian Tweets,” *in 2017 Iranian Conference on Electrical Engineering (ICEE)*, 2017, pp. 1532–1536.

[13] M. Seifikar, S. Farzi, and S. D. Mahmoodabad, “Kermanshah Earthquake Event Tracking Through Persian Tweets,” *in 2018 9th International Symposium on Telecommunications (IST)*, 2018, pp. 424–428.

[14] A. Y. K. Chua and S. Banerjee, “Linguistic predictors of rumor veracity on the Internet,” *in Lecture Notes in Engineering and Computer Science*, 2016, vol. 1, pp. 387–391.

[15] Q. Li, Q. Zhang, and L. Si, “eventai at semeval-2019 task 7: Rumor detection on social media by exploiting content, user credibility and propagation information,” *in Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, pp. 855–859.

[16] F. Xing and C. Guo, “Mining Semantic Information in Rumor Detection via a Deep Visual Perception Based Recurrent Neural Networks,” *in 2019 IEEE International Congress on Big Data (BigDataCongress)*, 2019, pp. 17–23.

[17] S. Vosoughi and D. Roy, “A human-machine collaborative system for identifying rumors on twitter,” *in 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 2015, pp. 47–50.

[18] A. Y. K. Chua and S. Banerjee, “Linguistic predictors of rumor veracity on the Internet,” pp. 387-391, 2016.

[19] A. Zubiaga, M. Liakata, and R. Procter, “Exploiting context for rumour detection in social media,” *in International Conference on Social Informatics*, 2017, pp. 109–123.

[20] S. Mahmoodabad, … S. F.-2018 9th I., and U. 2018, “Persian Rumor Detection on Twitter,” ieeexplore.ieee.org.

[21] H. K. Thakur, A. Gupta, A. Bhardwaj, and D. Verma, “Rumor Detection on Twitter Using a Supervised Machine Learning Framework,” Int. J. Inf. Retr. Res., vol. 8, no. 3, pp. 1–13, Jul. 2018.

[22] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science* (80-.)., vol. 359, no. 6380, pp. 1146–1151, 2018.

[23] A. Bondielli and F. Marcelloni, “A survey on fake news and rumour detection techniques,” *Inf. Sci. (Ny)*., vol. 497, pp. 38–55, 2019.

[24] G. W. Allport and L. Postman, The psychology of rumor. Henry Holt, 1947.

[25] H. Mohammadi and S. H. Khasteh, “A Machine Learning Approach to Persian Text Readability Assessment Using a Crowdsourced Dataset,” Oct. 2018.

[26] M. M. Homayounpour and A. S. Panah, “Speech Acts Classification of Persian Language Texts Using Three Machine Leaming Methods,” Int. J. Inf. Commun. Technol. Res., vol. 2, no. 1, pp. 65–71, 2010.

[27] Z. Jahanbakhsh-Nagadeh, M.-R. Feizi-Derakhshi, and A. Sharifi, “A Speech Act Classifier for Persian Texts and its Application in Identifying Rumors,” *J. Soft Comput. Inf. Technol. (JSCIT) Vol*, vol. 9, no. 1, 2020.

[28] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word-emotion association lexicon,” *in Computational Intelligence*, 2013, vol. 29, no. 3, pp. 436–465.

[29] H. Moradi, F. Ahmadi, and M.-R. Feizi-Derakhshi, “A Hybrid Approach for Persian Named Entity Recognition,” *Iran. J. Sci. Technol. Trans. A Sci.*, vol. 41, no. 1, pp. 215–222, 2017.

[30] V. Korde and C. N. Mahender, “Text classification and classifiers: A survey,” Int. J. Artif. Intell. Appl., vol. 3, no. 2, p. 85, 2012.

[31] A.-R. Feizi-Derakhshi et al., “Sepehr_RumTel01,” 2019.

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

[۳۲]محمدتقی تقوی فرد، پیام حنفی زاده، ابوالفضل کزازی، حمید جعفری، "مدل بومی ارزیابی کیفیت سایت‌های خبری (NEWSQUAL)"، مجله علمی-پژوهشی رایانش نرم و فن‌آوری اطلاعات، دوره ۷، شماره ۱، بهار و تابستان ۱۳۹۷، صفحه ۵۶-۷۱.

[32] H. Jafary, M.-T. Taghavifard, P. Hanafizadeh, and A. Kazazi, “Native Quality Assessment Model of news sites (NEWSQUAL),” *J. Soft Comput. Inf. Technol.*, vol. 7, no. 1, pp. 56–71, 2018, (In Persian).

[33] M. Wainberg, B. Alipanahi, and B. J. Frey, “Are random forests truly the best classifiers?,” *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 3837–3841, 2016.

[34] J. Wainer, “Comparison of 14 different families of classification algorithms on 115 binary datasets,” arXiv Prepr. arXiv1606.00930, 2016.

[35] A. J. Wyner, M. Olson, J. Bleich, and D. Mease, “Explaining the success of adaboost and random forests as interpolating classifiers,” *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1558–1590, 2017.

[36] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[37] Y. Freund and R. E. Schapire, “A desicion-theoretic generalization of on-line learning and an application to boosting,” *in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 904, 1995, pp. 23–37.

[38] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995.

[39] N. El Aboudi and L. Benhlima, “Review on wrapper feature selection approaches,” in Proceedings - 2016 International Conference on Engineering and MIS, ICEMIS 2016, 2016.

زلیخا جهانبخش نقده فارغ‌التحصیل دکترای گروه مهندسی رایانه، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران است. ایشان هم اکنون عضو هیئت علمی دانشگاه آزاد اسلامی واحد نقده است. علایق پژوهشی ایشان پردازش زبان طبیعی، داده‌کاوی و تحلیل شبکه‌های اجتماعی است. نشانی رایانامه ایشان عبارت است از:

zoleikha.jahanbakhsh@srbiau.ac.ir

محمدرضا فیضی درخشی دانشیار گروه مهندسی رایانه در دانشگاه تبریز است. او همچنین در حال حاضر به‌عنوان سرپرست آزمایشگاه سامانه‌های پردازش هوشمند رایانه‌ای (سپهر) در همان دانشگاه خدمت می‌کند. وی دکترای خود را در رشته مهندسی رایانه از دانشگاه علم و صنعت ایران دریافت کرده است. زمینه‌های پژوهشی ایشان، پردازش زبان طبیعی (شامل خلاصه‌سازی متن، پردازش معنایی وب و اسناد و ...)، تحلیل شبکه اجتماعی، الگوریتم‌های بهینه‌سازی و پایگاه داده‌ها است.

نشانی رایانامه ایشان عبارت است از:

mfeizi@tabrizu.ac.ir

آرش شریفی استادیار گروه مهندسی رایانه، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران است. وی دکترای خود را در رشته مهندسی رایانه از همان دانشگاه دریافت کرده است. زمینه‌های پژوهشی ایشان یادگیری ماشین، محاسبات نرم، و یادگیری عمیق است.

نشانی رایانامه ایشان عبارت است از:

a.sharifi@srbiau.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]

[DOI: 10.52547/jsdp.18.1.50]

[Downloaded from jsdp.rcisp.ac.ir on 2025-12-14]
[DOI: 10.52547/jsdp.18.1.50]