

کاربرد هوش مصنوعی در تشخیص اخبار جعلی

احمدرضا ترسلی^{۱*}

تاریخ دریافت: ۱۴۰۳/۰۲/۲۷

پذیرش مقاله: ۱۴۰۳/۰۳/۲۴

چکیده

همزمان با توسعه اینترنت، ظهور و پذیرش گسترده مفهوم رسانه‌های اجتماعی بر نحوه شکل‌گیری و انتشار اخبار تاثیر گذاشته و آن‌ها را تغییر داده است. اخبار؛ سریع‌تر، کم‌هزینه‌تر و به‌راحتی با رسانه‌های اجتماعی قابل دسترسی هستند. البته این تغییر با معایبی همراه بوده است، تا جایی که محتوای فریبنده مانند اخبار جعلی ساخته شده توسط کاربران رسانه‌های اجتماعی به‌طور فزاینده‌ای خطرناک می‌شود. هدف اصلی مقاله، ارائه راه‌حلی برای چالش تشخیص تفاوت بین اخبار واقعی و جعلی است. این مقاله از نظر هدف کاربردی است و از نظر شیوه اجرا، گردآوری و تحلیل داده‌ها پژوهشی کمی است که با انتخاب دو مجموعه داده منبع باز از وب‌گاه کگل در سال ۲۰۲۴م. (به‌عنوان جامعه آماری) و پیش‌پردازش این مجموعه داده با استفاده از سامانه رپیدماینر و بکارگیری ماتریس درهم‌ریختگی مبتنی بر پنج مدل یادگیری ماشین نظارت شده، مدل‌سازی و اجرا شده است. در این مقاله پس از تجزیه و تحلیل داده‌ها و محاسبه معیارهای ارزیابی صحت، یادآوری، دقت و امتیاز F1 (میانگین همساز) و مقایسه نتایج مشخص گردید که مدل یادگیری ماشین جنگل تصادفی در معیارهای صحت و دقت به ترتیب با ۹۸,۳ درصد و ۹۷,۹ درصد و مدل یادگیری تقویت گرادیان در محاسبه معیارهای امتیاز F1 و یادآوری به ترتیب با ۹۷,۷ درصد و ۹۸,۷ درصد بهترین نتیجه در تشخیص اخبار جعلی دارند.

واژگان کلیدی: هوش مصنوعی، یادگیری ماشین، اخبار جعلی، اطلاعات نادرست، ماتریس درهم‌ریختگی.

۱. دانش آموخته دکتری مدیریت راهبردی فضای سایبر، گروه مدیریت سایبری، دانشکده امنیت ملی، دانشگاه عالی دفاع ملی، تهران، ایران (نویسنده مسئول) a.tarrasoli@sndu.ac.ir

۱. مقدمه

با ظهور رسانه‌های اجتماعی، هر فردی می‌تواند در تولید محتوا مشارکت کند و انتشار اخبار جعلی در اینترنت را با سرعتی سرسام‌آور ساده کرده است. این امر، زمینه انتشار اطلاعات نادرست را تسهیل کرده است که اغلب یا برای افزایش ترافیک شبکه، عضوگیری و کسب سودهای مالی از ارائه تبلیغات آنلاین مانند «طعمه کلیک»، تأثیرگذاری بر باورهای افراد و در نهایت، تأثیرگذاری بر رویدادهای مهم مانند انتخابات ساخته می‌شود (ینکیکار و همکاران، ۲۰۲۳: ۱). اخبار جعلی به دلیل تهدیدی که برای جامعه ایجاد می‌کند، از بحث برانگیزترین موضوعات امروز هستند.

امروزه، افراد دارای شهرت ممکن است به دلیل ناآگاهی محض و یا به تبعیت از سیاست‌های دشمن، اطلاعات گمراه‌کننده را منتشر کنند و این یک مشکل دائمی است؛ به‌ویژه زمانی که به مداخله خارجی مرتبط باشد. کاربران رسانه‌های اجتماعی اغلب نمی‌دانند تنها هدف پست‌ها، توییت‌ها، مقاله‌ها یا سایر مطالب نوشتاری که در فضای مجازی وجود دارند، شکل‌دهی به نظرات دیگران به منظور تأثیرگذاری بر قضاوت‌های آنها است. دقیقاً، «دستکاری اطلاعات» به دلیل دانش ضعیف کاربران رسانه‌های اجتماعی از موضوعات رخ می‌دهد.

برخی از کاربران جوان، غرق «اخبار فوری» هستند و اطلاعات نادرست همیشه بخشی از برنامه‌ریزی‌های اطلاعاتی و نظامی بوده است. این رویکرد، مطمئناً در نتیجه استفاده از رسانه‌های اجتماعی و فناوری هوشمند بدتر می‌شود. دلیل آن مسئله را باید چنین توضیح داد که فن‌آوری‌های ارتباطی مدرن ابزاری عموماً ارزان و کم‌مانع برای انتشار اطلاعات به‌ویژه انتشار اطلاعات نادرست ارائه می‌کنند.

شناسایی اخبار جعلی که توسط برخی مراجع به‌صورت دستی انجام می‌گردد، همواره با مسائل عدیده‌ای مواجه است. چراکه؛

(۱) تشخیص تفاوت بین اطلاعات واقعی و نادرست، ردیابی و کنترل آنها بسیار دشوار است.

(۲) با ظهور کلان داده‌ها، ارزیابی دستی کاری زمان‌بر است و افشای اطلاعات نادرست برای تأثیرگذاری بر مخاطبین خیلی دیر اتفاق می‌افتد.

(۳) پائین بودن دقت بررسی واقعیت‌ها است که کاربران عموماً به آن اعتماد ندارند و این شرایط در متقاعد کردن افرادی که از قبل به اطلاعات نادرست اعتقاد دارند، بی‌اثر است.

(۴) اخبار جعلی آینده که ممکن است با الگوریتمی که امروز ایجاد شده قابل تشخیص نباشند. به‌عنوان مثال، جعل‌های عمیق که اصلاحات صوتی یا تصویری رقمی هستند، «بسیار واقع‌گرایانه و به سختی قابل تشخیص هستند»، و احتمالاً در آینده در جنگ اطلاعاتی رایج‌تر خواهند بود.

(۵) رمزگذاری انتها به انتها که توسط برنامه‌های پیام‌رسانی مانند واتساپ^۱ و سیگنال^۲ استفاده می‌شود، نظارت و رهگیری انتقال اطلاعات نادرست را چالش برانگیزتر می‌کند.

مهمترین گام برای جلوگیری از انتشار اخبار جعلی و نجات جان مردم، «شناسایی زودهنگام اخبار جعلی» است. افراد، ناآگاهانه اطلاعات نادرست را منتشر می‌کنند و در آن مشارکت می‌کنند. تشخیص اخبار جعلی و الگوی انتشار آن برای جامعه و دولت اهمیت فوق‌العاده‌ای داشته تا از این رشته حوادث جلوگیری کنند، انگیزه‌های پشت اخبار اینترنتی را تجزیه و تحلیل و توانایی تشخیص درست از آنچه جعلی است را توسعه دهند.

استفاده از هوش مصنوعی برای تشخیص دقیق، صحیح و زودهنگام اطلاعات نادرست، ارزشمند است و در این مقاله تلاش گردید تا با بررسی چگونگی استفاده از مدل‌های یادگیری ماشینی در پیش‌بینی و تشخیص اخبار جعلی، بر میزان دانش و درک مدیران حوزه امنیتی و دفاعی کشور بیافزاید. جامعه اطلاعاتی در عصر کوانتوم، هوش مصنوعی، اینترنت

و... نیازمند استفاده از ابزارهای متناسب با زمان برای پردازش کارآمد کلان داده‌های رایانه سپهر؛ تشخیص اطلاعات جعلی، منبع و الگوی انتشار برای پیشگیری و مقابله موثر با فعالیت‌های مجرمانه و بدخواهانه علیه امنیت و اعتماد جامعه خواهد بود.

۲. مبانی نظری و پیشینه‌شناسی تحقیق

۲-۱. اهمیت و ضرورت

هدف اصلی این پژوهش، بررسی کاربرد و چگونگی تاثیر مدل‌های طبقه‌بندی‌کننده یادگیری ماشین در تشخیص اخبار نادرست است. اهمیت و دلیل انجام مطالعه حاضر این است که الگوریتم‌های هوش مصنوعی می‌توانند حجم عظیمی از داده‌ها را به سرعت و با دقت بالا پردازش کنند و اخبار جعلی را با بررسی عواملی مانند سبک نگارش، منبع انتشار، و تناقض با اطلاعات موثق شناسایی کنند. این امر، به‌ویژه در مواقعی که انتشار اخبار جعلی می‌تواند به سرعت آسیب‌های جدی به بار آورد، بسیار حائز اهمیت است. همچنین هوش مصنوعی به‌مرور زمان و با جمع‌آوری داده‌های بیشتر می‌تواند در تشخیص اخبار جعلی دقیق‌تر عمل کند. الگوریتم‌ها می‌توانند با شناسایی الگوهای جدید در اخبار جعلی و به‌روز رسانی مداوم دانش خود، با جدیدترین روش‌های فریب‌کاری سازگار شوند. از سوی دیگر، استفاده از هوش مصنوعی برای تشخیص اخبار نادرست، نیاز به بررسی دستی تک‌تک اخبار توسط انسان را کاهش می‌دهد. این امر به کارشناسان و خبرنگاران اجازه می‌دهد تا زمان و تمرکز خود را صرف بررسی عمیق‌تر اخبار مشکوک و تأیید صحت اخبار موثق کنند.

در مورد ضرورت انجام تحقیق حاضر، همین نکته کفایت دارد که انتشار اخبار جعلی به‌خصوص در شرایط بحرانی می‌تواند باعث ایجاد ترس و وحشت در میان مردم شوند و همچنین در مورد دولت، رسانه‌ها و سایر نهادها می‌تواند اعتماد مردم به آنها را از بین ببرد. اخبار جعلی می‌تواند برای تحریک تنش و درگیری بین گروه‌های مختلف جامعه استفاده

شوند و همچنین در مواردی افراد ممکن است بر اساس اخبار نادرست تصمیماتی اشتباه در زندگی شخصی، شغلی یا حتی سیاسی خود بگیرند.

۲-۲. پیشینه‌شناسی تحقیق

اخگری و ممتازی در تحقیقی مشترک با عنوان «کاربرد هوش مصنوعی در راستی‌آزمایی اخبار: تشخیص اخبار جعلی با استفاده از متن خبر و اطلاعات منابع منتشرکننده خبر» (۱۴۰۲) با استفاده از مجموعه داده «تات» که برای زبان فارسی مناسب بوده و شامل ۱۰۸۱ خبر جعلی و ۱۰۸۱ خبر با برجسب غیرجعلی در حوزه‌های مختلف خبری از ۳۸ کانال تلگرامی است، نشان دادند استفاده از شبکه عصبی پیچشی، شناسایی اخبار فارسی منتشرشده در تلگرام و استفاده از متن خبرهای منتشرشده به همراه شناسه کانال ارسال‌کننده خبر به‌عنوان ورودی شبکه توانسته است به صحت ۹۰,۴۶ درصد در تشخیص اخبار جعلی دست یابد.

ابراهیمی و همکاران در تحقیقی با عنوان «بررسی تطبیقی سیاست کیفری انتشار اطلاعات نادرست در رسانه‌های مجازی» (۱۴۰۱) نشان دادند طی سال‌های اخیر، بسیاری از کشورها با وضع مقررات کیفری در قبال انتشار اطلاعات نادرست در شبکه‌های اجتماعی و رسانه‌های مجازی، سیاست کیفری واحدی را دنبال می‌کنند.

هاشمی و همکاران در تحقیق مشترکی با عنوان «پیشرفت تشخیص اخبار جعلی: یادگیری عمیق ترکیبی با FastText و هوش مصنوعی قابل توضیح» (۲۰۲۴) به این نتیجه رسیدند که یک مدل ترکیبی که CNN را با لایه‌های LSTM ترکیب می‌کند و با تعبیه‌های FastText تقویت شده است، در طبقه‌بندی دقیق مقالات خبری از مدل‌های دیگر بهتر عمل می‌کند. علاوه بر این، استفاده از مدل‌های مبتنی بر ترانسفورماتور، قابلیت‌های این معماری‌ها را در رمزگشایی ساختارهای نحوی پیچیده برای درک معنایی پیشرفته روشن کرده است. استفاده از هوش مصنوعی قابل توضیح از طریق LIME و LDA نه تنها

شفافیت فرآیند تشخیص را توجیه کرده است، بلکه بینش تفسیری ارزشمندی را نیز ارائه کرده است.

«پاتیل»^۱ و همکارانش در پژوهش «رویکردی جدید برای تشخیص اخبار جعلی با استفاده از هوش مصنوعی» (۲۰۲۴) نشان دادند سکوی طراحی شده توسط هوش مصنوعی مولد برای تشخیص اخبار نادرست که از LamaIndex به عنوان یک منبع قابل اعتماد استفاده می‌کند؛ با ادغام تحلیل‌های بی‌درنگ توییتر، حضور رسانه‌های اجتماعی و بر داده‌های مرتبط نیز به کاربران اطلاعات روشن‌تری در مورد محبوبیت و نفوذ محتوای خبری می‌دهد و با تجزیه و تحلیل احساسات، تجزیه و تحلیل متن، شناسایی تبلیغات، تشخیص سخنان مشوق عداوت و تنفر، و نظارت بر فعالیت ربات‌ها نیز در تجزیه و تحلیل مقالات و شناسایی منابع جعلی به خوبی عمل می‌کند.

«هو»^۲ و همکارانش در تحقیقی با عنوان «بازیگر بد، مشاور خوب: بررسی نقش مدل‌های زبان بزرگ در تشخیص اخبار جعلی» (۲۰۲۴) نشان دادند مدل‌های زبانی بزرگ^۳ فعلی ممکن است جایگزین مدل‌های زبانی کوچک^۴ تنظیم شده دقیق در تشخیص اخبار جعلی نشوند، اما می‌توانند با ارائه دلایل آموزنده چند منظری، مشاور خوبی برای SLMها باشند.

ترویکا^۵ و همکاران در مقاله «معماری گروهی شبکه عصبی عمیق برای تشخیص اخبار جعلی از محتوای اجتماعی و متنی» (۲۰۲۴) اشاره کردند معماری شبکه‌های عصبی عمیق برای تشخیص اخبار جعلی مبتنی بر متن اجتماعی و متن مشتمل بر یک شاخه متن برای زمینه مبتنی بر محتوای متنی و یک شاخه اجتماعی برای زمینه اجتماعی که این دو شاخه

1. Patil

2. Hu

3. large language models (LLMs)

4. small language models (SLMs)

5. Truică

برای ایجاد یک جاسازی شبکه جدید^۱ استفاده می‌شود، نتایج خوبی را در مجموعه داده‌های آموزشی کوچک به دست می‌آورد و از مدل‌های پیشرفته فعلی بهتر عمل می‌کند.

۲-۳. مبانی نظری و مفهوم‌شناسی

۲-۳-۱. اخبار جعلی

بر اساس تعریف لیزر^۲ و همکاران (۲۰۱۸)، اخبار جعلی به عنوان اطلاعات نادرست یا گمراه‌کننده تعریف می‌شوند که نوعی از تقلید از اخبار واقعی را به نمایش می‌گذارند (سوئتنکو و آنجلوپولوس^۳، ۲۰۲۴: ۴۶۰).

۲-۳-۲. متن کاوی

در یادگیری ماشین، نمایش داده‌ها تا حد زیادی بر صحت نتایج تأثیر می‌گذارد. به طور خاص، مسائل تجزیه و تحلیل متن باید به نمایشی تبدیل شوند که برای روش مورد استفاده مناسب باشد. داده‌های مبتنی بر متن که توسط کاربران در رسانه‌های اجتماعی به اشتراک گذاشته می‌شوند، عموماً به شکل‌های بدون ساختار هستند. به همین دلیل، داده‌های بدون ساختار استخراج‌شده از رسانه‌های اجتماعی باید با روش‌های متن‌کاوی به شکل ساختاریافته تبدیل شوند. مشکل متن‌کاوی را می‌توان به عنوان استخراج اطلاعات معنی‌دار، مفید و ناشناخته قبلی از داده‌های متنی تعریف کرد (پچیان و همکاران^۴، ۲۰۲۳: ۷). روش متن‌کاوی با پیش‌پردازش داده‌ها آغاز می‌شود که شامل سه مرحله است (مک‌کار و همکاران^۵، ۲۰۲۳: ۱).

۲-۳-۲-۱. پیش‌پردازش داده‌ها و بن‌واژه‌سازی متن

در هر زبانی، کلمات با توجه به نقشی که در جملات ایفا می‌کنند، به شکل‌های ظاهری متفاوتی خواهند بود. اما با توجه به این‌که تمامی آن‌ها از یک ریشه (بُن) ساخته می‌شوند، از نظر معنا و مفهوم در گام‌های بعدی تشخیصی به ما کمک شایانی خواهند نمود. از

1. Novel Network Embedding

2. Lazer & et.al.

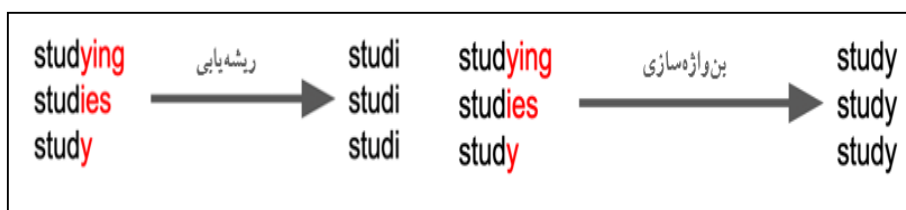
3. Setekouw, L., & Angelopoulos, S.

4. Bhiyan, V., Muthulingam, S., Sathar, G., Nalajala, S., Ch, A., & Das, M. N.

5. Makkar, K., Kumar, P., Poriye, M., & Aggarwal, S.

همین‌رو، در بسیاری از روش‌های مبتنی بر پردازش زبان‌های طبیعی (NLP)، ابتدا می‌بایست ریشه کلمات را پیدا کنیم. برای استخراج ریشه کلمات معمولاً از دو روش ریشه‌یابی^۱ و بُن‌واژه‌سازی^۲ در روش پردازش زبان‌های طبیعی (NLP) برای استخراج ریشه کلمات (کلمات مشتق شده از کلمات دیگر) استفاده می‌شوند.

بُن‌واژه‌سازی در رویکردی که برای تولید اشکال ریشه کلمات استفاده می‌شود، متفاوت از ریشه‌یابی است. تفاوت بین این دو رویکرد این است که بن‌واژه‌سازی، ریشه واقعی کلمات زبان متن را تولید می‌کند در حالی که ریشه‌یابی که به فرایند حذف پسوند کلمات گفته می‌شود؛ ممکن است ریشه واقعی کلمات را تولید نکند. بنابراین، در تحقیق حاضر از رویکرد بن‌واژه‌سازی استفاده شد، زیرا کلمات عطف شده را به درستی کاهش می‌دهد و اطمینان حاصل می‌کند که بن‌واژه، ریشه کلمه و متعلق به زبان است. نمونه‌ای از بن‌واژه‌سازی در شکل (۱) نشان داده شده است (چای،^۳ ۲۰۲۳: ۱۹).



شکل شماره ۱. بن‌واژه‌سازی از کلمات خارج شده از متن

۲-۲-۳-۲. حذف کلمات توقف

«حذف کلمات توقف»^۴ باید قبل از آموزش طبقه‌بندی‌کننده‌های مبتنی بر یادگیری ماشین از متن انجام شوند؛ زیرا کلمات توقف معنای بسیار کمی دارند و اغلب برای تکمیل ساختار جمله و اتصال عبارات استفاده می‌شوند. حروف ربط، ضمایر و حروف اضافه نیز کلمات توقف هستند. برای حذف کلمات توقف از متن خبر، هر جمله در متن نشانه‌گذاری شده و به کلمات/نشانه‌ها

1. Stemming
2. Lemmatization
3. Chai
4. Stop word removal

تقسیم می‌شود. الگوریتم در تمام نشانه‌ها تکرار می‌شود و بررسی می‌کند: آیا کلمه در لیست کلمات توقف وجود دارد یا خیر؟، بنابراین کلمه حذف می‌شود. بنابراین، تمام کلماتی مانند «به»، «او»، «است»، «an» و «the» از متن حذف می‌شوند. حدود ۴۰۰ تا ۵۰۰ کلمه توقف در زبان انگلیسی وجود دارد. برخی از کلمات توقف عبارتند از: a, an, about, by, but, that, does, when, where, what, all, am, any, against, on, above. نمونه‌ای از حذف کلمات توقف در شکل (۲) نشان داده شده است (مک‌کار و همکاران، ۲۰۲۳: ۴).

جمله با کلمات توقف	She is buying the newspaper and one bottle of milk.
واحدهای کلمات (Tokens)	She Is buying the newspaper And One Bottle Of Milk
جمله بدون کلمات توقف	buying newspaper milk

شکل شماره ۲. حذف کلمات توقف

۳-۲-۳. واحدسازی (استخراج ویژگی)

فرآیند «واحدسازی» نیز متن داده‌شده را به قسمت‌های کوچکتری تقسیم می‌کند که به آنها «واحدهای کلمات»^۲ می‌گویند و تمام علائم نگارشی را از داده‌های متنی حذف می‌کند (مک‌کار، ۲۰۲۳: ۷). فیلتر اعداد برای حذف عباراتی که حاوی اعداد هستند، اعمال شده است. مبدل حروف برای تبدیل داده‌های متنی به حروف کوچک یا بزرگ استفاده شده است. در این مقاله، تمامی عبارت‌ها به حروف کوچک تبدیل شده‌اند. در نهایت در این مرحله از فیلتر N-chars برای حذف کلماتی که کمتر از N کاراکتر تشکیل شده‌اند، استفاده شده است (مهسواری و سودها،^۳ ۲۰۲۴: ۴).

بزرگترین مشکل در متن‌کاوی، کلان‌داده‌ها هستند. بنابراین، حذف ویژگی‌های غیرمرتبط و اضافی برای بهبود دقت مدل ضروری است. در مراحل پیش‌پردازش داده‌ها، ویژگی‌ها از کلان‌داده‌های بدون ساختار استخراج می‌شوند (داس و آلفونس،^۴ ۲۰۲۳: ۳).

1. Tokenization
2. Tokens
3. Maheswari & Sudha
4. Das & Alphonse

ورودی بسیاری از روش‌های یادگیری، رقم و اعداد هستند و باید با روش‌هایی اسناد متنی (متن‌های خبری) به صورت عدد درآیند که به آن «استخراج ویژگی» گفته می‌شود. یکی از روش‌های استخراج ویژگی، ایجاد «ماتریس اصطلاح-سند»^۲ است. در این ماتریس (نک: شکل شماره ۳) تکرار هر اصطلاح در هر سند شمارش می‌شود. در استخراج ویژگی که با استفاده از مدل «فضای برداری»^۳ (VSM) شباهت بین اسناد را بر اساس بازنمایی برداری آنها مقایسه و اندازه‌گیری می‌گردد. هر سند به‌عنوان یک بردار نشان داده می‌شود، که در آن هر بعد مربوط به یک اصطلاح منحصر به فرد در مجموعه اسناد است. مقدار هر بعد نشان‌دهنده اهمیت یا فراوانی عبارت مربوطه در سند است.

اصطلاح سند(خبر)	پرتاب	موشک	ماهوره‌بر	سیمرخ
سند ۱	۱		۱	
سند ۲		۱		
سند ۳	۲	۱	۳	
سند ۴	۲		۲	۱

شکل ۳. ماتریس اصطلاح-سند مبتنی بر تکرار اصطلاح

برای وزن‌دهی و شمارش ویژگی‌های مختلف از روش‌های مختلفی می‌توان استفاده کرد: روش‌های همچون (TF-IDF) یا تکرار کلمه-برعکس تکرار سند،^۴ برعکس تکرار سند^۵ (IDF)، تکرار کلمه^۶ (TF) یا نمایش باینری^۷ برای محاسبه این وزن‌ها توسعه داده شده است (داس و آلفونس، ۲۰۲۳: ۵).

1. Feature extraction
2. Document-term matrix
3. Vector Space Model (VSM)
4. Term Frequency-Inverse Document Frequency (TF-IDF)
5. Inverse Document Frequency (IDF)
6. Term Frequency (TF)
7. Binary representation

دو روش «TF» و «TF-IDF» شناخته شده‌ترین این روش‌ها هستند. بردار TF-IDF برداری است که از روی کلمات می‌تواند ویژگی‌های مختلف را برای یک متن بسازد (در واقع متن را تبدیل به اعداد قابل فهم برای الگوریتم کند). در این مطالعه از TF برای محاسبه وزن هر کلمه در هر سند استفاده شده است. TF تعداد دفعاتی که یک کلمه در یک سند مشاهده می‌شود را نشان می‌دهد. TF با استفاده از معادله زیر محاسبه می‌شود.

$$TF = \frac{n_{ij}}{|d_i|} \quad (1)$$

d_i مجموع تمام اصطلاحات سند i است. n_{ij} تعداد کلمه j ام در سند i است.

پس از محاسبه مقدار TF برای هر کلمه در سند، ماتریس اصطلاح-سند (DTM) با توجه به وزن کلمات ایجاد می‌شود. DTM مانند شکل (۴) تعریف شده است و یک ماتریس $m \times n$ است. در ماتریس، هر ردیف نشان‌دهنده اسناد؛ ستون‌ها نشان‌دهنده عبارت‌ها؛ و سلول‌ها نشان‌دهنده اعداد واقعی هستند که وزن عبارات را در سند نشان می‌دهند (المراشی^۱ و همکاران، ۲۰۲۳: ۵).

$$\begin{array}{l} \text{سند ۱} \\ \vdots \\ \text{سند } m \end{array} \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{m1} & \cdots & W_{mn} \end{bmatrix}$$

شکل شماره ۴. ماتریس اصطلاح-سند مبتنی بر وزن اصطلاح در سند

۳-۲-۳. الگوریتم‌های طبقه‌بندی

الگوریتم‌های مختلف برای تشخیص اخبار جعلی مورد بررسی قرار گرفت. ویژگی‌های استخراج شده به پنج طبقه‌بندی‌کننده مختلف وارد شدند. در این پژوهش از الگوریتم‌های *Logistic Naïve Bayes*، *Support Vector Machine*، *Random Forest*، *Regression* و *Gradient Boosting* برای ساخت طبقه‌بندی‌کننده‌ها و بررسی استفاده از الگوریتم‌های یادگیری ماشین استفاده گردید.

۱-۳-۲. جنگل‌های تصادفی

«جنگل تصادفی»^۱ یک الگوریتم بسیار مؤثر از نظر پیش‌بینی در زمینه یادگیری ماشینی تحت نظارت، یادگیری عمیق و هوش مصنوعی است. یک جنگل تصادفی بر اساس اصل «کیسه‌گذاری»^۲ کار می‌کند که اولین گام در آن، تقسیم یک مجموعه داده به زیرمجموعه‌ها برای تولید مجموعه‌ای از درخت‌های تصمیم‌گیری^۳ است؛ پس از آموزش درختان، نتایج تولید شده برای به‌دست آوردن قوی‌ترین پیش‌بینی ترکیب می‌شوند. روش کیسه‌گذاری استفاده از چندین نمونه داده به‌جای یک نمونه است و خروجی‌های تولید شده توسط درختان تصمیم رتبه‌بندی می‌شوند و بالاترین مقدار به عنوان خروجی نهایی انتخاب می‌شود (دودجا و همکاران،^۴ ۲۰۲۳: ۲). دو روش برای تعیین نتیجه نهایی وجود دارد.

«رگرسیون جنگل تصادفی»^۵ شامل محاسبه میانگین پیش‌بینی‌های به‌دست‌آمده با در نظر گرفتن تمام پیش‌بینی‌های حاصل از درخت‌های تصمیم است. طبقه‌بندی جنگل تصادفی نیز بر اساس روش بسته‌بندی است، اما تخمین نهایی با انتخاب متداول‌ترین دسته پاسخ به جای استفاده از همه نتایج به‌دست‌آمده انجام می‌شود (گونیمات و همکاران،^۶ ۲۰۲۳: ۶).

درخت‌های تصمیم ایجاد شده توسط طبقه‌بندی‌کننده جنگل تصادفی^۷ با استفاده از ناخالصی جینی^۸ مورد استفاده برای شکافتن شاخه‌ها و انتخاب گره‌هایی که عدم قطعیت در درخت‌های تصمیم را کاهش می‌دهند، آموزش داده می‌شوند. بنابراین بهترین تقسیم با به‌حداقل رساندن ناخالصی جینی هنگام تقسیم هر گره انتخاب می‌شود. ناخالصی جینی یک گره نشان‌دهنده این احتمال است که یک نمونه بر اساس توزیع نمونه‌ها در یک گره

-
1. Random Forests (RF)
 2. Bagging
 3. Decision Trees
 4. Dudeja, D., Noonia, A., Lavanya, S., Sharma, V., Kumar, V., Rehan, S., & Ramkumar, R.
 5. Random forest regression
 6. Ghunimat, D., Alzoubi, A. E., Alzboon, A., & Hanandeh, S.
 7. Classification Random Forest
 8. Gini Impurity

به طور تصادفی انتخاب شده و به اشتباه برچسب گذاری شده است. ناخالصی جینی یک گره n به عنوان فرمول داده شده تعریف می شود.

$$I_G(n) = 1 - \sum P_i^2 \quad (2)$$

که در آن P_i احتمال متعلق به کلاس i در یک گره معین است. مقدار کم ناخالصی جینی به این معنی است که گره ها خالص هستند و هیچ شانسی وجود ندارد نمونه ای که به طور تصادفی از آن گره انتخاب شده است، به اشتباه طبقه بندی شود (سی و همکاران؛ ۲۰۲۴: ۳).

۲-۳-۳-۲. ماشین بردار پشتیبان

«ماشین های بردار پشتیبان»^۲ مجموعه ای از روش های یادگیری تحت نظارت^۳ هستند که برای حل مسائل طبقه بندی و رگرسیون طراحی شده اند و تعمیم طبقه بندی کننده های خطی هستند. این الگوریتم در دهه ۱۹۹۰ توسط «کورتس و واپنیک»^۴ توسعه یافت و با دلایل زیر به سرعت مورد استفاده قرار گرفت:

(۱) توانایی آنها در کار با کلان داده ها،

(۲) تعداد کم فرایامترها؛

(۳) تضمین های نظری،

(۴) نتایج خوب در عمل.

برخلاف سایر الگوریتم های یادگیری، این الگوریتم سعی می کند مشابه ترین مثال ها را بین کلاس ها بیاموزد تا مجموعه ای از بردارهای پشتیبان بسازد و براساس آن، با کشیدن یک مرز تصمیم به نام «ابرفصحه»^۵ داده ها را از هم جدا می کند و در نهایت با محاسبه بهترین حاشیه ابرفصحه، ابر فصحی بهینه را که تقسیم کننده کلاس ها است را بررسی می کند.

1. Xie, X., Yuan, M. J., Bai, X., Gao, W., & Zhou, Z. H.
2. Support Vector Machine (SVM)
3. Supervised Learning
4. Cortes and Vapnik
5. Hyperparameters
6. Hyperplane

SVM می‌تواند برای حل مسائل طبقه‌بندی با تصمیم‌گیری در مورد این‌که یک نمونه متعلق به کدام کلاس است، یا مسائل رگرسیون با پیش‌بینی مقدار عددی یک متغیر نیز استفاده شود. حل این دو نوع مسئله شامل ساخت تابع f است که دارای بردار ورودی X و مطابق با خروجی است (گیدو و همکاران، ۲۰۲۳: ۳).

$$Y = f(X) \quad (۳)$$

توابع کرنل^۲ توسط الگوریتم‌های SVM استفاده می‌شوند. در این مطالعه نیز از کرنل خطی^۳ که معمولاً برای مسائل طبقه‌بندی متن توصیه می‌شود، استفاده گردید. تابع کرنل خطی از پارامترهای کمتری استفاده می‌کند و سریع‌تر از سایر توابع هسته مانند توابع چند جمله‌ای و شعاعی^۴ است. مرز تصمیمی که SVM برمی‌گرداند توسط تابع کرنل خطی ارائه شده است که در فرمول زیر تعریف می‌شود.

$$f(X) = w^T x + b \quad (۴)$$

که در آن w بردار وزنی است که باید به حداقل برسد، x داده‌ای است که باید طبقه‌بندی شود، و b ضریب خطی تخمین زده شده است. از دو پارامتر w و b برای تعریف ابرصفحه استفاده می‌شود (آزه و همکاران، ۲۰۲۳: ۴).

۳-۳-۲. بیز ساده

«بیز ساده»^۵ یک نوع طبقه‌بندی بیز احتمالی ساده است که بر اساس قضیه بیز در فرمول زیر ارائه شده است. این یک طبقه‌بندی‌کننده ساده بیز را پیاده‌سازی می‌کند که متعلق به خانواده طبقه‌بندی‌کننده‌های خطی با استقلال قوی از مفروضات است. یک طبقه‌بندی‌کننده بیزی

-
1. Guido, R., Groccia, M. C., & Conforti, D.
 2. Kernel functions
 3. linear kernel
 4. polynomial and radial functions
 5. Azzeh
 6. Naive Bayes (NB)

ساده فرض می‌کند که وجود یک مشخصه برای یک کلاس مستقل از وجود ویژگی‌های دیگر است. حتی اگر این ویژگی‌ها به هم مرتبط باشند.

$$P(A/B) = (P(B/A)P(A)) / P(B) \quad (5)$$

مزیت طبقه‌بندی‌کننده NB این است که به داده‌های آموزشی نسبتاً کمی برای تخمین پارامترهای لازم برای طبقه‌بندی نیاز دارد. الگوریتم ماتریس اصطلاح-سند را برای هر کلاس (جعلی، واقعی) محاسبه می‌کند. این ماتریس شامل فهرستی از بسامدهای کلمه موجود در مجموعه‌ای از اسناد است. ورودی (m, n) ماتریس سند اصطلاحی از فراوانی کلمه "m" در سند "n" تشکیل شده است. تعداد به مفهوم تعداد دفعاتی که هر واژه/کلمه در تمام اسناد وجود دارد محاسبه می‌شود. در این پژوهش، از بیز ساده چندجمله‌ای^۱ استفاده گردید که معمولاً برای مسئله طبقه‌بندی متن استفاده می‌شود و داده‌ها به‌عنوان تعداد بردار کلمات نشان داده می‌شوند (وزیروگلو و همکاران،^۲ ۲۰۲۴: ۷).

۲-۳-۳-۴. رگرسیون پشتیبان

«رگرسیون پشتیبان»^۳ یکی از الگوریتم‌های یادگیری تحت نظارت و تجزیه و تحلیل پیش‌بینی‌کننده بر اساس مفهوم احتمال است. این الگوریتم عمدتاً برای مسائل طبقه‌بندی برای تخصیص نمونه‌ها به مجموعه‌ای از کلاس‌ها استفاده می‌شود و هدف آن، انجام طبقه‌بندی باینری است. تابع پشتیبان خروجی یک مدل خطی را به فرصتی تبدیل می‌کند که می‌تواند برای وظایف طبقه‌بندی اعمال شود.

تابع پشتیبان که در بسیاری از کاربردهای دیگر مورد استفاده قرار می‌گیرد، ابزار مؤثری برای درک و پیش‌بینی داده‌ها است. خروجی با استفاده از تابع «سیگموئید پشتیبان»^۴ به یک «تابع هزینه» تبدیل می‌شود که مقدار احتمالی را بین ۰ و ۱ برمی‌گرداند.

1. Multinomial Naive Bayes
2. Veziroglu, M., Eziroglu, E., & Bucak, I. O.
3. Logistic Regression (LR)
4. logistic Sigmoid
5. Cost Function

فرضیه رگرسیون پشتیبان به سمت حداقل کردن تابع هزینه متمایل می‌شود. بنابراین استفاده از توابع خطی به‌عنوان تابع هزینه مناسب نیست؛ زیرا می‌تواند مقداری بزرگتر از ۱ تولید کند که طبق فرضیه رگرسیون پشتیبان قابل قبول نیست. تابع سیگموئید برای ترسیم هر مقدار پیش‌بینی شده که ممکن است هر مقدار واقعی با X نشان داده شود به احتمال آن‌که باید بین ۰ و ۱ باشد استفاده شده است (زیدی و ال لوهیب، ۲۰۲۳: ۱). تابع سیگموئید به صورت زیر تعریف می‌شود (همان: ۲):

$$f(X) = \frac{1}{(1+e^{-x})} \quad (۶)$$

برای کاهش خطا در احتمالات پیش‌بینی شده توسط مدل و تولید پیش‌بینی‌های دقیق، بهینه‌سازی تابع هزینه با استفاده از گرادیان کاهشی^۲ انجام می‌گردد (همان: ۴). ضرایب بتا^۳ برای معادله رگرسیون پشتیبان از داده‌های آموزشی با استفاده از تخمین حداکثر احتمال، تخمین زده شد؛ به‌طوری‌که، مدل قادر خواهد بود مقدار بسیار نزدیک به (۱) را برای کلاس اخبار جعلی و مقدار بسیار نزدیک به (۰) را برای کلاس اخبار واقعی پیش‌بینی کند (سولومون و همکاران، ۲۰۲۳: ۷).

۵-۳-۲. تقویت گرادیان

الگوریتم تقویت گرادیان یکی از قدرتمندترین الگوریتم‌ها در زمینه یادگیری ماشین است. تقویت گرادیان^۵ اساساً فرآیند تبدیل مدل‌های یادگیری ضعیف به مدل‌های یادگیری قوی است. یکی از اجزای کلیدی تقویت گرادیان، «تابع ضرر»^۶ است که برای آموزش این مدل‌ها استفاده می‌شود (امامی و مارتینز-مونیوز، ۲۰۲۳: ۳).

1. Zaidi, A., & Al Luhayb, A. S. M.

2. Gradient Descent

3. The Beta coefficients

4. Solomon, F. A. M., Sathianesan, G. W., & Ramesh, R.

5. Gradient Boosting (GB)

6. loss function

7. Emami, S., & Martínez-Muñoz, G.

برای یادگیرندگان ضعیف، از درخت تصمیم در تقویت گرادیان استفاده می‌شود؛ جایی که خروجی‌های آن‌ها برای تصحیح باقی‌مانده‌ها در پیش‌بینی‌ها اضافه می‌شود. در تقویت گرادیان، یادگیرنده ضعیف درخت رگرسیون است و در هر تکرار، مدل با استفاده از گرادیان کاهشی تصادفی برای کاهش خطا آموزش داده می‌شود؛ یک. اولین یادگیرنده ضعیف (درخت اول) در الگوریتم GB یاد می‌گیرد تا خطا را در تکرار اول کاهش دهد.

دو. سپس، درخت دوم (یعنی دومین یادگیرنده ضعیف) به‌طور مداوم آموزش داده می‌شود و یاد می‌گیرد که خطا را در تکرار دوم درخت دوم به حداقل برساند. این فرآیند به‌طور مکرر انجام می‌شود تا زمانی که خطای مورد قبول به‌دست آید (هو و ترن، ۲۰۲۴: ۷).

۳. روش‌شناسی تحقیق

این مقاله از نظر هدف، کاربردی است و از نظر شیوه اجرا، گردآوری و تحلیل داده‌ها نیز پژوهشی کمی است. مجموعه داده مطالعه حاضر بر اساس دو مجموعه داده منبع باز است که به صورت آنلاین از وب‌گاه کگل آدر دسترس هستند.

اولین مجموعه داده شامل ۶۰۶۰ خبر است که ۵۰ درصد خبر جعلی و ۵۰ درصد خبر واقعی است. اخبار مجموعه داده اول با عناوین زیر مشخص می‌شود (کومار، ۲۰۲۴):

(۱) شناسه،

(۲) عنوان،

(۳) متن،

(۴) برچسب (جعلی و صحیح).

این داده‌ها و قالب آن در جدول (۱) ارائه شده است.

جدول شماره ۱. معیارها و زیرمعیارهای تاب‌آوری زنجیره تأمین (طهماسبی و حامی، ۱۳۹۹)			
شناسه	عنوان	متن	برچسب
8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a New York writer focusing...	Fake

مجموعه داده دوم شامل ۲۱۷۲۴ مقاله خبری و اطلاعاتی درباره آن است. مجموعه داده دوم از ۱۷۴۴۱ خبر موثق و ۵۷۵۵ خبر غیرموثق تشکیل شده است. هر مقاله خبری با موارد زیر توصیف می‌شود (گولوین، ۲۰۲۴):

- (۱) یک عنوان،
- (۲) آدرس اینترنتی،
- (۳) نام دامنه،
- (۴) تعداد بازنشر،
- (۵) برچسب (جعلی و صحیح).

شاخص‌های فوق در جدول (۲) نشان داده شده است.

جدول شماره ۲. اخبار مجموعه داده دوم				
عنوان	آدرس اینترنتی	نام دامنه	تعداد بازنشر	برچسب
Kandi Burruss Explodes Over Rape Accusation on Real Housewives of Atlanta' Reunion	http://toofab.com/2017/05/08/real-housewives-atlanta-kandi-burruss-rape-phaedra-parks-porsha-william...	toofab.com	42	Fake

دو مجموعه داده مورد استفاده برای جلوگیری از ایجاد هرگونه اختلاف در فرآیند طبقه‌بندی، با استفاده از سامانه رپیدمایتر^۱ ابتدا وارد مرحله پیش‌پردازش می‌گردند. فرآیند

1. Golovin
2. Retweet

پاکسازی داده‌ها برای حذف نویزهای درون متن خبر مورد نیاز است. دور انداختن کلیه کلمات و حروف غیرانگلیسی یا بی‌معنا در اخبار، جایگزینی کلمات عامیانه با اشکال استاندارد آنها و تبدیل متون به حروف کوچک، برای یکنواختی کار یک گام اساسی برای نگاه داشتن متن قابل استفاده در خبر است. از روش‌های پردازش زبان طبیعی (NLP) می‌توان برای تحلیل متون خبری و استخراج اطلاعات با کیفیت بالا در متن استفاده کرد. همان‌گونه که پیشتر ذکر گردید، چندین روش پردازش زبان طبیعی برای مرحله پیش پردازش مجموعه‌های داده، قبل از آموزش مدل‌های یادگیری ماشین به کار گرفته شدند. روش‌های مورد استفاده عبارتند از؛ «واحدسازی»، «بُن‌واژه‌سازی» و «حذف کلمات توقف» که بخش مهمی در متن‌کاوی می‌باشند.

پس از پیش پردازش و آماده‌سازی مجموعه داده، آن‌ها به دو قسمت تقسیم گردیدند:

(۱) «مجموعه داده آموزشی»،

(۲) «مجموعه داده تست».

بررسی مدل‌های طبقه‌بندی‌کننده یادگیری ماشین برای شناسایی اخبار جعلی و طبقه‌بندی دقیق اخبار به جعلی یا واقعی، و دستیابی به معیارهای مشخص و معتبر در ارزیابی و مقایسه هر یک از مدل‌ها یک گام مهم در مطالعه این پژوهش است. با استفاده از سامانه رپیدماینر، پنج مدل طبقه‌بندی‌کننده یادگیری ماشین که همان الگوریتم‌های رگرسیون پشتیبان، تقویت گرادیان، جنگل تصادفی، ماشین بردار پشتیبان و بیز ساده به صورت جداگانه بر روی مجموعه داده‌های تست بررسی و برای سنجش قدرت پیش‌بینی هر یک از مدل‌های یادگیری ارزیابی گردید.

۴. تجزیه و تحلیل یافته‌ها

معیارهای ارزیابی مختلف برای مقایسه عملکرد الگوریتم‌های یادگیری ماشین تحت نظارت برای تشخیص اخبار جعلی استفاده شده است. معیارهای ارزیابی اغلب به ما امکان می‌دهند اثربخشی الگوریتم را آزمایش کنیم. در مطالعه حاضر، همانطور که در جدول (۳) نشان داده

شده است، از روش ماتریس درهم‌ریختگی^۱ برای ارزیابی عملکرد تشخیص اخبار جعلی استفاده شده است. در این ماتریس، نمونه‌ها به عنوان جعلی یا واقعی طبقه‌بندی می‌شوند. هنگام آزمایش طبقه‌بندی‌کننده‌های باینری (اخبار جعلی، اخبار واقعی)، هر رکورد در یکی از چهار احتمال زیر قرار می‌گیرد:

❖ **مثبت صادق**^۲ (TP): اگر اخبار جعلی پیش‌بینی شده در واقع اخبار جعلی

باشد، پیش‌بینی TP است.

❖ **مثبت کاذب**^۳ (FP): اگر اخبار جعلی پیش‌بینی شده در واقع خبر واقعی باشد،

پیش‌بینی FP است.

❖ **منفی صادق**^۴ (TN): اگر خبر واقعی پیش‌بینی شده در واقع خبر واقعی باشد،

پیش‌بینی TN است.

❖ **منفی کاذب**^۵ (FN): اگر خبر واقعی پیش‌بینی شده در واقع اخبار جعلی باشد،

پیش‌بینی FN است.

واقعیت		جایزه آماری	پیش‌بینی
اخبار واقعی (واقعیت منفی)	اخبار جعلی (واقعیت مثبت)		
مثبت کاذب	مثبت صادق	پیش‌بینی مثبت	
منفی صادق	منفی کاذب	پیش‌بینی منفی	

جدول شماره ۳: ماتریس درهم‌ریختگی برای اخبار جعلی

پیش‌بینی‌های صحیح شامل «مثبت صادق» و «منفی صادق» هستند، در حالی که مثبت کاذب و منفی کاذب پیش‌بینی‌های نادرستی هستند که توسط طبقه‌بندی‌کننده‌ها انجام

1. Confusion matrix
2. True-Positive
3. False-Positives
4. True-Negatives
5. False-Negatives

می‌شوند. برای سنجش عملکرد مدل‌های یادگیری ماشین در این پژوهش، معیارهای ارزیابی مختلفی به شرح زیر استفاده شد:

(۱) صحت؛

(۲) امتیاز F1؛

(۳) یادآوری،

(۴) دقت.^۳

«صحت» به عنوان نسبت بین تعداد نمونه‌های طبقه‌بندی شده صحیح و تعداد کلی نمونه‌ها تعریف می‌شود (گودو و همکاران، ۲۰۲۳: ۵).

$$\frac{|TP|+|TN|}{|TP|+|FP|+|TN|+|FN|} = Accuracy(\text{صحت}) \quad (7)$$

مقدار یادآوری نسبت اخبار جعلی را نشان می‌دهد که به درستی در تعداد کل اخبار جعلی پیش‌بینی شده است.

$$\frac{|TP|}{|TP|+|FN|} = Recall(\text{یادآوری}) \quad (8)$$

معیار دقت، دقت اخبار جعلی را اندازه‌گیری می‌کند، که به درستی از کل اخبار پیش‌بینی شده در کلاس جعلی پیش‌بینی می‌شود.

$$\frac{|TP|}{|TP|+|FP|} = precision(\text{دقت}) \quad (9)$$

امتیاز F1 یا مقدار میانگین همساز؛ مقدار یادآوری و مقدار دقت به دست آمده برای تشخیص اخبار جعلی است که برای ارزیابی دقت مدل و نرخ‌های یادآوری به‌طور جمعی استفاده شده تا درک بهتری از رکوردهای طبقه‌بندی اشتباه ارائه شود.

$$2 * \frac{Recall * Precision}{Recall + Precision} = F1 - Score(\text{میانگین همساز}) \quad (10)$$

1. Accuracy
2. F1-Score
3. The Recall and Rhe Precision
4. Harmonic Mean

امتیاز F1 بالا نشان می‌دهد که مدل‌ها موارد منفی کاذب و مثبت کاذب پایینی دارند که به نوبه خود به معنای عملکرد خوب است. امتیاز F1 تعادل بین دقت و یادآوری است و بالاترین مقدار آن ۱ است که نشان‌دهنده دقت و یادآوری کامل است (همان: ۵). همانگونه که نتایج تشخیص اخبار جعلی با استفاده از معیارهای چهارگانه ارزیابی (صحت، امتیاز F1، یادآوری و دقت) در جدول (۴) گزارش شده است، در معیارهای امتیاز F1 و یادآوری، الگوریتم تقویت گرادیان بالاترین امتیازات را به ترتیب ۹۷,۷ و ۹۸,۷ به دست آورد. این الگوریتم تعادل بسیار خوبی بین معیارهای دقت و فراخوانی دارد.

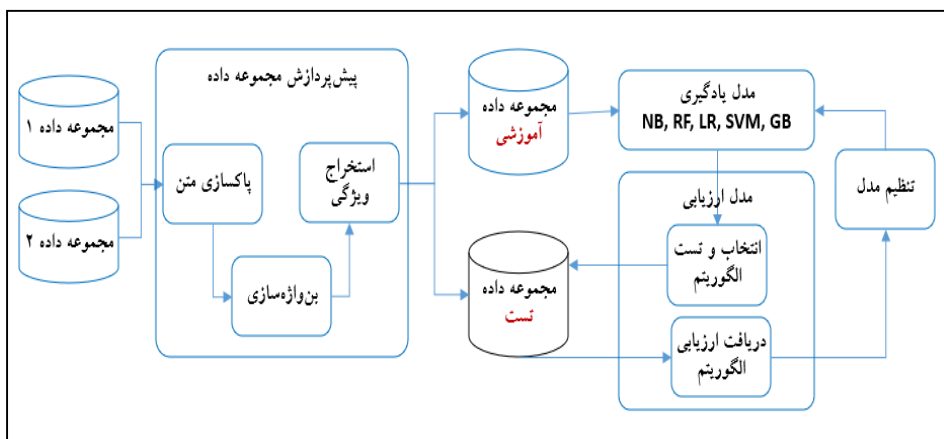
جدول شماره ۴: نتایج ارزیابی عملکرد مدل‌های یادگیری ماشین				
مدل یادگیری ماشین	% صحت	% دقت	% امتیاز F1	% یادآوری
رگرسیون پشتیبان	۹۶,۶	۹۷,۵	۹۷,۳	۹۶,۴
تقویت گرادیان	۹۷,۶	۹۶,۴	۹۷,۷	۹۸,۷
جنگل تصادفی	۹۸,۳	۹۷,۹	۹۷	۹۷,۶
ماشین بردار پشتیبان	۹۵,۲	۹۵,۳	۹۳,۶	۹۳,۴
بیز ساده	۸۹,۳	۸۹,۳	۹۱,۴	۹۱,۳

الگوریتم جنگل تصادفی قدرت تشخیص بسیار خوبی در معیارهای صحت و دقت با امتیازات به ترتیب ۹۸,۳ و ۹۷,۹ دارد.

۵. نتیجه‌گیری و پیشنهادات

۵-۱. جمع‌بندی

از آنجایی که تشخیص اخبار به صورت دستی یک کار پرهزینه است، به تخصص بالا و دانش عمیق در این حوزه نیاز دارد تا بتوان ناهنجاری‌ها را در متن خبر تشخیص داد. پژوهش حاضر، با هدف بررسی کاربرد هوش مصنوعی در تشخیص اخبار جعلی انجام شد. نتایج حاصل از روش تحقیق و تحلیل انجام شده در این پژوهش در قالب مدل مفهومی و به شرح فرآیند پیشنهادی تشخیص اخبار جعلی در شکل (۵) ترسیم و ارائه شده است.



شکل شماره ۵. فرآیند پیشنهادی تشخیص اخبار جعلی

همانطور که پیش‌تر ذکر شد، تشخیص نادرست اخبار جعلی از اخبار واقعی، به دلیل تاثیر منفی اخبار جعلی بر کاربران و عموم جامعه منجر به بروز مسائل اخلاقی رفتاری مهمی می‌گردد. بروز منفی کاذب و مثبت کاذب پیش‌بینی‌های نادرستی هستند که چنین پیامدهای زیان‌باری را ایجاد خواهند کرد. موارد منفی کاذب به‌عنوان طبقه‌بندی نادرست اطلاعات نادرست که به‌عنوان اطلاعات صحیح ظاهر می‌شوند و اجازه می‌دهند محتوای مضر از تشخیص فرار کند و در فضای برخط بدون کنترل پخش شود، می‌تواند باعث ایجاد سردرگمی و بی‌اعتمادی در میان کاربران شود و به‌طور مستقیم به‌عنوان سلاح موثر جنگ‌های ترکیبی^۱ بر انتخابات و فرآیندهایی مردم‌سالارانه تاثیر بگذارد، زیرا افراد برای تصمیم‌گیری آگاهانه به اطلاعات با کیفیت بالا نیاز دارند.

از سوی دیگر، موارد مثبت کاذب به شناسایی نادرست محتوای صحیح که به‌عنوان اطلاعات نادرست معرفی می‌شوند و در نتیجه، اطلاعات معتبر منتشر شده برخط مانند گزارش‌های خبری، به دلیل طبقه‌بندی نادرست حذف می‌گردند. چنین اشتباهاتی می‌توانند تهدید جدی برای گفتگوی آزاد باشند و به نقض آزادی‌های مدنی و موارد سانسور منجر شوند.

بنابراین استفاده از معیارهای ارزیابی که بر سنجش میزان پیش‌بینی‌های صحیح شامل مثبت صادق و منفی صادق متمرکز بوده و گزاره‌های صحت، امتیاز F1، یادآوری و دقت را بررسی می‌نمایند موجب افزایش اطمینان و اعتماد به مدل‌های یادگیری ماشین در تشخیص صحیح اخبار جعلی می‌گردند.

همانگونه که نتایج تشخیص اخبار جعلی با استفاده از معیارهای چهارگانه ارزیابی در جدول (۴) ذکر شده است، الگوریتم تقویت‌گرادیان که در معیارهای امتیاز F1 و یادآوری، بالاترین امتیازات را به ترتیب ۹۷,۷ و ۹۸,۷ به دست آورد و تعادل بسیار خوبی بین معیارهای دقت و فراخوانی ارائه نمود، از روش تقویت استفاده می‌کند و درخت‌ها را یکی یکی می‌سازد، به گونه‌ای که هر درخت از درختی که قبلاً آموزش داده شده بهره می‌برد و خطاهای خود را تصحیح و باعث ایجاد یک یادگیرنده قوی می‌شود.

الگوریتم جنگل تصادفی که قدرت تشخیص بسیار خوبی در معیارهای صحت و دقت با امتیازات به ترتیب ۹۸,۳ و ۹۷,۹ ارائه داد از راهبرد کیسه‌گذاری و از نمونه تصادفی داده‌ها استفاده می‌کند و هر درخت را جداگانه آموزش می‌دهد و نیاز به رای‌گیری برای تجمیع مدل دارد. به بیان ساده، جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند.

ترکیب روش‌های متن‌کاوی به این مدل اجازه داد تا قابلیت‌های تعمیم و عملکرد طبقه‌بندی‌کننده اخبار جعلی را بهبود بخشد. بنابراین، این مدل پتانسیل ایجاد یک محیط غنی از دانش را دارد که می‌تواند به‌طور معناداری با طبقه‌بندی دقیق اخبار به‌حداقل رساندن انتشار اخبار جعلی در سکوه‌های رسانه‌های اجتماعی کمک کند.

۲-۵. پیشنهادات

با عنایت به کاربردی بودن این پژوهش، پیشنهاد می‌گردد:

- (۱) به منظور مبارزه با اطلاعات نادرست، باید برنامه و بودجه‌ای مشخص برای دانشگاه‌ها، موسسات تحقیقاتی و سایر محققان و نهادهای دانش‌بنیان برای

«طراحی»، «تولید» و «توسعه» سامانه‌های تخصصی این حوزه تامین و اختصاص داده شود.

(۲) سازمان‌های مرتبط با حوزه پژوهش ضمن تلاش برای به حداقل رساندن کارهای تکراری، رویکردهای رایج برای توسعه، آزمایش و ارزیابی، استقرار قابلیت‌های هوش مصنوعی در تشخیص اخبار جعلی را به حداکثر برسانند. آنها باید با یکدیگر همکاری نمایند تا منابع و ابزارهای قابل استفاده و قابل اشتراک‌گذاری را ایجاد کنند. باید فرهنگ به اشتراک‌گذاری همه قابلیت‌های فعال شده با هوش مصنوعی را در صورت امکان ایجاد کنند.

منابع

الف - فارسی

- ابراهیمی، سید نصراله؛ محمودی، امیررضا؛ و میری بالاجورشری، سیده مهشید (۱۴۰۱). «بررسی تطبیقی سیاست کیفی انتشار اطلاعات نادرست در رسانه‌های مجازی»، *فصلنامه آموزه‌های فقه و حقوق جزاء*، ۱ (۱)، ۲۰-۱.
- اخگری، محمدرضا؛ ممتازی، سعیده (۱۴۰۲). «کاربرد هوش مصنوعی در راستی‌آزمایی اخبار: تشخیص اخبار جعلی با استفاده از متن خبر و اطلاعات منابع منتشرکننده خبر»، *پژوهش‌های رسانه و ارتباطات*، ۱ (۱)، ۲۶۸-۲۴۳.

ب - انگلیسی

- Almarashy, A. H. J., Feizi-Derakhshi, M. R., & Salehpour, P. (2023). "Enhancing Fake News Detection by Multi-Feature Classification". IEEE Access. Almarashy, A. H. J., Feizi-Derakhshi, M. R., & Salehpour, P. (2023). *Enhancing Fake News Detection by Multi-Feature Classification*. IEEE Access.
- Azzeh, M., Elsheikh, Y., Nassif, A. B., & Angelis, L. (2023). "Examining the performance of kernel methods for software defect prediction based on support vector machine", *Science of Computer Programming*, (226), 102916.
- Chai, C. P. (2023). "Comparison of text preprocessing methods", *Natural Language Engineering*, 29 (3), 509-553.
- Das, M., & Alphonse, P. J. A. (2023). "A comparative study on tf-idf feature weighting method and its analysis using unstructured dataset", *arXiv preprint arXiv*, 2308.04037.
- Dudeja, D., Noonina, A., Lavanya, S., Sharma, V., Kumar, V., Rehan, S., & Ramkumar, R. (2023). "Breast Cancer Diagnosis Using Bagging Decision Trees with Improved Feature Selection", *Engineering Proceedings*, 59 (1), 17; <https://doi.org/10.3390/engproc2023059017>
- Emami, S., & Martínez-Muñoz, G. (2023). *Sequential training of neural networks with gradient boosting*. IEEE Access.
- Ghunimat, D., Alzoubi, A. E., Alzboon, A., & Hanandeh, S. (2023). "Prediction of concrete compressive strength with GGBFS and fly ash using multilayer perceptron algorithm, random forest regression and k-nearest neighbor regression", *Asian Journal of Civil Engineering*, 24 (1), 169-177.
- Golovin, A. (2024). "Fake News. Fake News dataset based on FakeNewsNet". In: **Kaggle**, Available at: <https://www.kaggle.com/datasets/algord/fake-news>.

- Guido, R., Groccia, M. C., & Conforti, D. (2023). "A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers", *Soft Computing*, 27 (18), 12863-12881.
- Ho, L. S., & Tran, V. Q. (2024). "Evaluation and estimation of compressive strength of concrete masonry prism using gradient boosting algorithm", *Plos one*, 19 (3), e0297364.
- Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., & Qi, P. (2024, March). "Bad actor, good advisor: Exploring the role of large language models in fake news detection". In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 38 (20), 22105-22113.
- Kumar, R. (2024). "Fake News Prediction Dataset. Dataset is having records of real and fake news which can be used for prediction", Available at: <https://www.kaggle.com/datasets/rajatkumar30/fake-news?resource>.
- Maheswari, R. U., & Sudha, N. (2024). "An efficient fake news classification model based on ensemble deep learning techniques", *Salud, Ciencia y Tecnología-Serie de Conferencias*, (3), 649-649.
- Makkar, K., Kumar, P., Poriye, M., & Aggarwal, S. (2023). "Improvisation in opinion mining using data preprocessing techniques based on consumer's review", *International Journal of Advanced Technology and Engineering Exploration*, 10 (99), 257-277; DOI:10.19101/IJATEE.2021.875886.
- Patil, M., Yadav, H., Gawali, M., Suryawanshi, J., Patil, J., Yeole, A., ... & Potlabattini, J. (2024). "A Novel Approach to Fake News Detection Using Generative AI", *International Journal of Intelligent Systems and Applications in Engineering*, 12 (4s), 343-354.
- Pichiyar, V., Muthulingam, S., Sathar, G., Nalajala, S., Ch, A., & Das, M. N. (2023). "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis", *Procedia Computer Science*, (230), 193-202; DOI: 10.1016/j.procs.2023.12.074.
- Soetekouw, L., & Angelopoulos, S. (2024). "Digital resilience through training protocols: Learning to identify fake news on social media", *Information Systems Frontiers*, 26 (2), 459-475.
- Solomon, F. A. M., Sathianesan, G. W., & Ramesh, R. (2023). "Logistic Regression Trust-A Trust Model for Internet-of-Things Using Regression Analysis", *Computer Systems Science & Engineering*, 44 (2), 1125-1142.
- Truică, C. O., Apostol, E. S., & Karras, P. (2024). "DANES: Deep neural network ensemble architecture for social and textual context-aware fake news detection", *Knowledge-Based Systems*, (294), 111715; <https://doi.org/10.1016/j.knosys.2024.111715>.
- Veziroğlu, M., Eziroğlu, E., & Bucak, İ. Ö. (2024). "Performance Comparison between Naive Bayes and Machine Learning Algorithms for News Classification". In: *Bayesian Inference-Recent Trends*. IntechOpen.
- Xie, X., Yuan, M. J., Bai, X., Gao, W., & Zhou, Z. H. (2024). "On the Gini-impurity preservation for privacy random forests", *Advances in Neural Information Processing Systems*, (36), 45055-45082.

- Yenikar, A., Sultanpure, K., & Bali, M. (2023). "Machine learning-based algorithmic comparison for fake news identification". (pp. 67-83). In: ***AI-Based Metaheuristics for Information Security and Digital Media***. Chapman and Hall/CRC.
- Zaidi, A., & Al Luhayb, A. S. M. (2023). "Two statistical approaches to justify the use of the logistic function in binary logistic regression". ***Mathematical Problems in Engineering***.
<https://doi.org/10.1155/2023/5525675>.