News Science
Online ISSN: 2980-9274 Print ISSN: 2382-9788
Homepage: www.mjourcom.ir

New Strategies of Fake News Detection

Nastaran Zanjani[1]✉ | Fatemeh Pouramini [2] | Zahra Askarinejadamiri[3]

1. Corresponding author, Assistant professor, computer engineering department, Refah University College, Tehran, Iran. E-mail: zanjani@refah.ac.ir
2. Bachelor of computer engineering, computer engineering department, Refah University College, Tehran, Iran. E-mail: f.pour67@gmail.com
3. Assistant professor, computer engineering department, Refah University College, Tehran, Iran. E-mail: askarinejad@refah.ac.ir

Article Info

Article type:
Research paper

Article history:
Received: October 4, 2024
Received in revised form: October 17, 2024
Accepted: January 9, 2025
Published online: January 17, 2025

Keywords:
Social Networks,
Fake News,
Fake News Detection,
Artificial Intelligence,

ABSTRACT

Objective: This paper aims to provide a detailed overview of the most recent methods and strategies used for detecting fake news, especially in the context of rapid advancements in artificial intelligence and machine learning. With the widespread reach of fake news across social media and other digital platforms, this review focuses on identifying and evaluating effective approaches that can help tackle this growing problem.

Methods: Given the importance of detecting fake news, this paper reviews and compares various approaches utilized in this field. To this end, by studying articles published in online libraries and document repositories such as IEEE, Scopus, Elsevier, and others, we first explore different methods for detecting fake news. Then, we compare the various approaches of human-based detection with those of automated detection.

Results: The review shows that while conventional techniques like feature extraction and rule-based systems offer a good starting point, they often fall short when dealing with the complexity of modern disinformation. Deep learning models trained on large datasets have demonstrated promising results in detecting fake news, yet they still struggle with the subtlety of human-generated content and real-time applications. This highlights the need for more comprehensive solutions that can address these challenges.

Conclusions: The findings suggest that an integrated approach—one that combines language analysis, machine learning, and network-based methods—is essential for building effective fake news detection systems. As the field progresses, future research should focus on improving hybrid models, refining data quality, and incorporating user-centric insights to combat the spread of disinformation better. Combining large language models (LLMs) with context-aware systems offers a promising path for achieving higher precision in detecting both machine-generated and human-created fake news.

Cite this article : Zanjani, N., Pouramini, F., & Askarinejadamiri, Z. (2025). New Strategies of Fake News Detection. *News Science*, 13 (4), 21-45. DOI: http//doi.org/ 10.22034/lrsi.2024.481772.1261

DOI : http//doi.org/ 10.22034/lrsi.2024.481772.1261

Extended Abstract

New Strategies of Fake News Detection

Nastaran Zanjani[1]✉ | Fatemeh Pouramini [2] | Zahra Askarinejadamiri[3]

1. Corresponding author, Assistant professor, computer engineering department, Refah University College, Tehran, Iran. E-mail: zanjani@refah.ac.ir
2. Bachelor of computer engineering, computer engineering department, Refah University College, Tehran, Iran. E-mail: f.pour67@gmail.com
3. Assistant professor, computer engineering department, Refah University College, Tehran, Iran. E-mail: askarinejad@refah.ac.ir

Introduction

The expansion of the World Wide Web and online social media has transformed human communication and information sharing. Features such as ease of use and high speed have turned social media into a primary platform for online interactions. However, the growing popularity of these spaces has led to the spread of fake news, including misleading information, rumors, and false advertisements, which pose a threat to trust in authorities and businesses. These fake news stories undermine users' trust in online information and affect online purchases and user reviews. Despite the existence of verification systems like Snopes.com, their time delays and limitations have led to extensive research into developing automated systems to detect fake news. These efforts focus on elements such as the content of the news, the credibility of the author, and dissemination patterns to detect fake news in real time and with greater accuracy.This paper aims to provide a detailed overview of the most recent methods and strategies for detecting fake news, especially in the context of rapid advancements in artificial intelligence and machine learning. With the widespread reach of fake news across social media and other digital platforms, this review focuses on identifying and evaluating effective approaches that can help tackle this growing problem.

Methods

Given the importance of detecting fake news, this paper reviews and compares various approaches utilized in this field. To this end, by studying articles published in online libraries and document repositories such as IEEE, Scopus, Elsevier, and others, we first explore different methods for detecting fake news. Then, we compare the various approaches of human-based detection with those of automated detection.

Results

This article refers to two main approaches for identifying fake news: human-based detection and automated methods. In the human-based method, websites like PolitiFact and Snopes identify fake news through detailed content review by experts. Additionally, platforms like Fiskkit allow ordinary users to evaluate news and opinions, although this method may not always be reliable due to user bias. Despite the advantages of human methods, they are often time-consuming and labor-intensive, highlighting the growing need for automated systems to identify fake news. Systems like Classify.news and Factmata use machine learning and artificial intelligence to analyze content and

identify fake news with high accuracy, leveraging natural language processing (NLP) and complex algorithms.
In the second part, the article discusses automated verification methods and their applications in real-time fake news detection. These approaches utilize machine learning and deep learning algorithms to analyze content, writing patterns, and information dissemination on social networks. Real-time systems like MWPBERT, combining BERT and natural language processing, can detect fake news instantaneously. These systems rely on predictive analysis and distribution patterns on social networks to improve accuracy in identifying false information. Recent research shows that using hybrid models, such as CNN-RNN and FakeBERT, significantly enhances both the accuracy and speed of fake news detection.

Conclusions

It is concluded from the article that fake news is considered one of the biggest threats to social networks. These false reports can negatively influence decision-making in various areas, including politics, financial markets, online shopping, and even elections. Automatic detection of online fake news is a highly complex challenge, as it is closely tied to the rapid spread and dynamic, heterogeneous nature of information on social media. The article examines different aspects of fake news, including the producer, content, distribution platform, and target audience, to provide a comprehensive understanding of how fake news spreads and impacts users.
This study compares existing detection approaches and emphasizes that utilizing linguistic features, social network analysis, and dissemination patterns is critical for identifying fake news more effectively and quickly. In conclusion, the article suggests that combining machine learning with social analysis can enhance the accuracy and speed of detection, contributing to the prevention of misinformation spread.
.

Data Availability Statement
Data available on request from the authors.
Acknowledgements
The authors would like to thank anonymous reviewers.
Ethical considerations
Not applicable.
Funding
Not applicable.
Conflict of interest
The authors declare no conflict of interest.

علوم خبری
شاپا چاپی: ۲۳۸۳-۹۷۸۸ شاپا الکترونیکی: ۹۲۷۴-۲۹۸۰

استراتژی های جدید در اخبار جعلی

نسترن زنجانی[۱] ✉ | فاطمه پورامینی[۲] | زهرا عسکری نژاد امیری[۳]

۱. استادیار گروه کامپیوتر دانشکده غیرانتفاعی رفاه، تهران. ایران، (نویسنده مسئول)، رایانامه: zanjani@refah.ac.ir
۲. کارشناس مهندسی کامپیوتر، گروه کامپیوتر دانشکده غیرانتفاعی رفاه، تهران. ایران، رایانامه: f.pour67@gmail.com
۳. زهرا عسکری نژاد امیری، استادیار، گروه کامپیوتر دانشکده غیرانتفاعی رفاه، تهران. ایران، رایانامه: askarinejad@refah.ac.ir

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۷/۱۳
تاریخ بازنگری: ۱۴۰۳/۷/۲۶
تاریخ پذیرش: ۱۴۰۳/۸/۲۷
تاریخ انتشار: ۱۴۰۳/۱۲/۲۷

کلیدواژه‌ها:
شبکه‌های اجتماعی،
اخبار جعلی،
اخبار جعلی آنلاین،
هوش مصنوعی.

چکیده

هدف: این مقاله به بررسی دقیق روش‌ها و استراتژی‌های جدید در شناسایی اخبار جعلی می‌پردازد، به‌ویژه در زمینه پیشرفت‌های سریع هوش مصنوعی و یادگیری ماشین. با توجه به گسترش اخبار جعلی در رسانه‌های اجتماعی و سایر پلتفرم‌های دیجیتال، این مرور بر شناسایی و ارزیابی رویکردهای مؤثری تمرکز دارد که می‌توانند به مقابله با این مشکل فزاینده کمک کنند.

روش‌ها: با توجه به اهمیت شناسایی اخبار جعلی، این مقاله به بررسی و مقایسه رویکردهای مختلفی که در این حوزه مورد استفاده قرار گرفته‌اند می‌پردازد. بدین منظور، با مطالعه مقالات منتشر شده در کتابخانه‌های آنلاین و مراکز اسناد مانند IEEE، Scopus، Elsevier و سایر منابع، ابتدا روش‌های مختلف شناسایی اخبار جعلی را بررسی کرده‌ایم. سپس، رویکردهای شناسایی به کمک نیروهای انسانی را با روش‌های شناسایی خودکار مقایسه می‌کنیم.

نتایج: نتایج نشان می‌دهند که در حالی که تکنیک‌های سنتی مانند استخراج ویژگی‌ها و سیستم‌های مبتنی بر قوانین، نقطه شروع خوبی برای شناسایی اخبار جعلی هستند، اما در مواجهه با پیچیدگی‌های اطلاعات نادرست مدرن اغلب ناکارآمد عمل می‌کنند. مدل‌های یادگیری عمیق که بر روی مجموعه داده‌های بزرگ آموزش دیده‌اند، در تشخیص اخبار جعلی عملکرد امیدوارکننده‌ای نشان داده‌اند، اما هنوز در تشخیص محتوای تولیدشده توسط انسان و کاربردهای در لحظه (Real-time) دچار مشکل هستند. این یافته‌ها نشان می‌دهند که نیاز به راه‌حل‌های جامع‌تری وجود دارد که بتوانند این چالش‌ها را برطرف کنند.

نتیجه‌گیری: نتایج این مطالعه نشان می‌دهند که یک رویکرد یکپارچه که شامل تحلیل زبانی، یادگیری ماشین و روش‌های مبتنی بر شبکه باشد، برای توسعه سیستم‌های مؤثر شناسایی اخبار جعلی ضروری است. با پیشرفت این حوزه، تحقیقات آینده باید بر بهبود مدل‌های ترکیبی، ارتقای کیفیت داده‌ها و در نظر گرفتن ویژگی‌های کاربرمحور تمرکز داشته باشند تا بتوانند بهتر با انتشار اطلاعات نادرست مقابله کنند. ادغام مدل‌های زبانی بزرگ (LLMs) با سیستم‌های آگاه به زمینه (Context-aware) می‌تواند مسیر امیدوارکننده‌ای برای دستیابی به دقت بالاتر در شناسایی هر دو نوع اخبار جعلی تولیدشده توسط ماشین و انسان باشد.

استناد: زنجانی، نسترن؛ پورامینی، فاطمه؛ و عسکری‌نژاد امیری، زهرا (۱۴۰۳). استراتژی های جدید در اخبار جعلی. *علوم خبری*، ۱۳ (۴)، ۴۵-۲۱.
DOI : http//doi.org/ 10.22034/lrsi.2024.481772.1261

مقدمه

توسعه شبکه جهانی وب و رشد رسانه‌های اجتماعی آنلاین ارتباط مردم با یکدیگر و به اشتراک‌گذاری اطلاعات را به طرز چشمگیری متحول کرده است. ویژگی‌هایی از قبیل سهولت استفاده، کم هزینه بودن و سرعت بالای رسانه‌های اجتماعی، آن‌ها را به بستر اصلی تعاملات اجتماعی آنلاین و انتقال اطلاعات تبدیل کرده است (Bigne et al., 2018).

با این حال، به دلیل محبوبیت روزافزون رسانه‌های اجتماعی آنلاین، اینترنت به فضایی ایده‌آل برای تولید و انتشار اخبار جعلی از قبیل اطلاعات گمراه کننده، آگهی‌های جعلی، شایعات و اظهارات سیاسی جعلی، تبدیل شده است. حجم انبوهی از اطلاعات باورنکردنی و گمراه کننده در این شبکه‌ها تولید و منتشر می‌شوند که به عنوان یک تهدید بالقوه برای شبکه‌های اجتماعی آنلاین به حساب می‌آید و تاثیر منفی عمیقی بر فعالیت‌های اینترنتی مانند خرید آنلاین می‌گذارد. با انتقال اطلاعات مغرضانه و دروغین، اخبار جعلی می‌توانند اعتماد و باور مردم به مقامات، متخصصان و دولت را از بین ببرند. به عنوان مثال ۸۸٪ از مشتریان بر اساس نظرات کاربران خرید می‌کنند و ۷۲٪ از آنها به کسب و کاری که دارای بازخورد مثبت از سوی مشتریان است اعتماد دارند (Ahmed, 2017). گسترش اخبار جعلی می‌تواند این روند را خدشه دار کند. اخبار جعلی این روزها همچنان حاکم بر اینترنت است و می‌تواند عواقب سرنوشت سازی برای جامعه داشته باشد. بنابراین نیازی جدی برای تولید یک سیستم دقیق جهت شناسایی اخبار جعلی آنلاین وجود دارد.

بسیاری از سیستم‌های راستی‌آزمایی آنلاین مانند Snopes.com و PolitiFact.com مبتنی بر رویکردهای تشخیص انسانی توسط متخصصان هستند. تأخیر زمانی و محدود بودن این سیستمها به صحت‌سنجی اخبار سیاسی باعث می‌شود تا در راستی‌آزمایی سایر موضوعات متنوعی که در حجم زیاد و با سرعت در شبکه‌های اجتماعی منتشر می‌شوند کاربردی نداشته باشند.

در سال‌های اخیر، برای کمک به کاربران آنلاین در شناسایی اطلاعات واقعی، تحقیقات گسترده‌ای در جهت ایجاد یک چهارچوب موثر و خودکار برای کشف اخبار جعلی آنلاین صورت گرفته است. با این وجود، شناسایی اطلاعات معتبر از بین میلیون‌ها پیام، به دلیل ماهیت پویا و ناهمگن ارتباطات اجتماعی چالش برانگیز است. از آنجایی که داده‌های اجتماعی آنلاین، به صورت بلادرنگ ایجاد می‌شوند، برای کشف، کاوش و تفسیر اطلاعات جعلی در رسانه‌های اجتماعی آنلاین، به یک سیستم شناسایی آنلاین بلادرنگ نیاز است.

به‌علاوه تنها به کارگیری ویژگی‌های زبانی محتوای اخبار برای آشکار ساختن الگوهای اخبار جعلی کافی نبوده و ویژگی‌هایی دیگری از جمله اعتبار نویسنده خبر و الگوهای انتشار خبر نیز باید در شناسایی اخبار جعلی آنلاین به کار گرفته شوند. به این منظور اخبار جعلی را می‌توان از چهار منظر مختلف شامل: تولید کننده خبر، محتوای خبر، بستر اجتماعی و گروه هدف بررسی کرد (Zhang & Ghorbani, 2020). این تقسیم بندی درک بهتری از منابع خبری جعلی، اهداف آنها، سبک‌های نوشتاری غالب در اخبار جعلی، چگونگی توزیع اخبار جعلی از طریق اینترنت و نحوه تاثیرگذاری آنها بر خوانندگان آنلاین ایجاد می‌کند. در این مقاله هدف بررسی روشهای مختلفی است که برای شناسایی اخبار جعلی به کار گرفته می‌شود.

روش پژوهش

با توجه به اهمیت شناسایی اخبار جعلی، این مقاله به بررسی و مقایسه رویکردهای گوناگونی که در این حوزه مورد استفاده قرار گرفته است می‌پردازد. بدین منظور با مطالعه مقالات منتشر شده در کتابخانه‌های آنلاین و مراکز اسناد همچون IEEE، Scopus، Elsevier و ... ، ابتدا رویکردهای مختلف شناسایی اخبار جعلی را بررسی کرده، .سپس به مقایسه رویکردهای مختلف شناسایی به کمک نیروهای انسانی و شناسایی خودکار پرداخته‌ایم.

پیشینه پژوهش

در شناسایی اخبار جعلی میتوان به دو رویکرد کلی اشاره کرد. یکی شناسایی به کمک نیروهای انسانی که در آن متخصص به بررسی محتویات اینترنتی در حوزه‌های مختلف پرداخته و جعلی یا واقعی بودن آنها را تشخیص می‌دهند. از این قبیل می‌توان به موارد زیر اشاره کرد:

PolitiFact.com: صحت ادعاها یا اظهارات مقامات منتخب، اندیشمندان، ستون نویسان، وبلاگ نویسان، تحلیلگران سیاسی و سایر اعضای رسانه را ارزیابی می‌کند. ویراستاران یک کلمه خاص و متن کامل یک ادعا را با دقت بررسی می‌کنند و سپس صحت ادعاها و اظهارات را تایید می‌کنند. این وب سایت رابط کاربری را در اختیار کاربران قرار می‌دهد تا به متن کامل بیانیه ها، داستان ها، وعده ها و به روز رسانی‌های بررسی شده دسترسی پیدا کنند.

Snopes.com: این صفحه وب، طیف گسترده‌ای از حوزه‌ها از جمله خودرو، تجارت، رایانه، جرم، فریب و کلاهبرداری، تاریخ و... را در بر میگیرد. Snopes.com به کمک دانش افراد و سازمان‌های حرفه‌ای می‌تواند ارزیابی‌های جامعی را برای انواع مختلف منابع چاپی ارائه دهد و آنها را از نظر راستی‌آزمایی رتبه‌بندی کند.

روش دیگری در شناسایی اخبار جعلی توسط انسان وجود دارد که عبارت است از صحت‌سنجی اطلاعات از طریق جمعیت. این روش به تعداد زیادی از افراد معمولی متکی است که به عنوان یک حقیقت سنج (مانند هوش جمعی) عمل می‌کند (Tschiatschek et al., 2017). به عنوان مثال میتوان به Fiskkit11 اشاره کرد، جایی که کاربران می توانند مقالات را بارگذاری کنند، جملات را رتبه‌بندی کنند و برچسب‌هایی را انتخاب کنند که به بهترین وجه آن را توصیف می‌کنند. با این حال این روش نوپاست و نتایج حاصل از آن در مقایسه با صحت‌سنجی مبتنی بر خبرگان، به دلیل سوگیری سیاسی افراد و حاشیه نویسی متناقض آنها کمتر قابل اعتماد است، از این رو، در این روش اغلب نیاز به فیلتر کردن کاربران غیر معتبر و تحلیل نتایج متناقض صحت‌سنجی وجود دارد (Di et al., 2018; Zhou & Zafarani, 2020).

باید توجه داشت که روش تشخیص انسانی فرایندی زمان بر و پرهزینه بوده، نیازمند دخالت گسترده انسان می‌باشد. بنابراین ایجاد بسترهای شناسایی اخبار جعلی خودکار ضروری به نظر می‌رسد. در این راستا تعدادی مراجع آنلاین صحت‌سنجی اخبار وجود دارند که به تعدادی از آنها شاره می‌کنیم.

Classify.news: فضایی برای شناسایی آنلاین مقالات جعلی است. هدف آن ساختن مدلی برای تشخیص اعتبار یک مقاله صرفاً بر اساس محتوای متنی آن با استفاده از الگوریتم‌های یادگیری ماشین است. روش کار به این صورت است که مقالات خبری برچسب زده شده را جمع آوری می‌کنند و بر اساس نمونه‌های معتبر و غیرمعتبر سیستم را آموزش می‌دهند. از دو نوع مدل پیش بینی "صرفا مبتنی بر متن" به روش دسته بندی کننده بیز چند جمله‌ای و "صرفا مبتنی بر مفهوم" به روش دسته بندی کننده بوستینگ تطبیقی در این سیستم استفاده می‌شود.

Factmata.com: یک پروژه سرمایه‌گذاری شده توسط گوگل می‌باشد که به صحت‌سنجی آماری ادعاها به کمک الگوریتم‌های هوش مصنوعی و یادگیری ماشین می‌پردازد. Factmata.com بر اساس تکنیک‌های پیشرفته پردازش زبان طبیعی (NLP)، می‌تواند ادعاهای آماری را با استفاده از روابط عددی شناسایی و بررسی کند.

این وب‌سایتها الگوریتم‌های هوش مصنوعی و یادگیری ماشین را در صحت‌سنجی اخبار جعلی به کار می‌گیرند. با این حال Factmata.com فقط می‌تواند ادعاها یا اظهاراتی را که حاوی اطلاعات آماری است بررسی کند. Classify.news عمدتاً از تکنیک‌های با نظارت یادگیری ماشینی که به مجموعه داده‌ها و برچسب‌های آموزشی با کیفیت بالا نیاز دارند استفاده می‌کند. لذا مسائل و مشکلات کشف اخبار جعلی، محققان و دست اندرکاران را برای ایجاد مدل‌های سیستماتیک و رویکردهای پیش‌بینی سازگار، خودکار و جامع ترغیب می‌کند.

۱. صحت سنجی خودکار اخبار جعلی

نحوه اجرای سیستم تشخیص جعلی اخبار را می‌توان به صورت زمان واقعی و آفلاین طبقه‌بندی کرد. در سیستم تشخیص آفلاین، معمولاً از مدل‌های یادگیری ماشین به صورت دسته‌ای برای شناسایی اخبار جعلی استفاده می‌شود. سیستم تشخیص آفلاین برای طبقه‌بندی اخبار جعلی آنلاین مهم است، زیرا می‌تواند اطلاعات غیر عادی را به صورت توصیفی مورد تجزیه تحلیل قرار دهد، از جمله انتخاب تاثیر گذارترین ویژگی‌ها برای تمایز اطلاعات دروغین در بین مقادیر زیادی پیام‌های اجتماعی. دسته‌بندی

آفلاین می‌تواند بر اساس انواع اطلاعات آنلاین، به تشخیص نظرات جعلی، هجویات، فریبکاری‌ها، اخبار سیاسی و برخی از زمینه‌های مرتبط دیگر، مانند تشخیص شایعه، تشخیص تله‌کلیک، شناسایی اسپمر و تشخیص ربات بپردازد (Shu et al., 2019).

با این حال، سیستم آفلاین محدود است. مجموعه داده‌های مورد استفاده ممکن است ویژگی‌های اساسی اخبار جعلی آنلاین را نشان ندهند و مدل‌های یادگیری آموزش دیده در یک سیستم آفلاین، در شرایط دیگر پاسخگو نباشد. در سیستم تشخیص بلادرنگ، تکنیک‌های مختلف تجزیه تحلیل بلادرنگ برای تعیین جعلی بودن یا نبودن اطلاعات اجتماعی استفاده می‌شود. با استفاده از روش‌های تحلیلی پیش‌بینی کننده بر روی اطلاعات در زمان واقعی، تجزیه تحلیل بلادرنگ می‌تواند کاربرد روش‌های آفلاین را بهبود بخشد. به دلیل سرعت انتشار اخبار جعلی در شبکه‌های اجتماعی، سیستم‌های بلادرنگی توسعه یافته‌اند که با تحلیل تعاملات کاربران و نحوه انتشار اخبار، امکان شناسایی سریع اطلاعات نادرست را فراهم می‌کنند. استفاده از تحلیل توزیع اطلاعات در شبکه‌های اجتماعی و ترکیب آن با الگوریتم‌های یادگیری ماشین منجر به بهبود دقت در شناسایی اخبار جعلی می‌شود (Farokhian et al., 2022). مطالعات کمی در زمینه کشف اخبار جعلی بلادرنگ وجود دارد.به عنوان مثال، شائو و همکارانش یک سیستم مصورسازی بلادرنگ برای تجزیه تحلیل اطلاعات غیر عادی در توییتر ایجاد کرد (Shao et al., 2016). با توجه به سرعت انتشار اخبار جعلی در شبکه‌های اجتماعی، مدل‌هایی مانند MWPBERT از دو شبکه BERT به صورت موازی استفاده می‌کنند. یکی از شبکه‌ها عنوان و دیگری بدنه خبر را پردازش می‌کند، که این روش به مدل اجازه می‌دهد تا محتوای کامل اخبار طولانی را تحلیل کند (Farokhian et al., 2022). ایجاد یک سیستم آنلاین موثر، چالش برانگیز است، زیرا داده‌های ارتباطات آنلاین حساس به زمان، مداوم و ناهمگن هستند. با وجود این، سیستم تشخیص آنلاین ابزاری قدرتمند برای گرفتن ماهیت پویای اطلاعات آنلاین و مبارزه با اخبار جعلی آنلاین است و باید در آینده مورد توجه بیشتری قرار بگیرد.

در صحت‌سنجی خودکار اخبار به صورت آفلاین دو نوع رویکرد مبتنی بر اجزاء خبر و مبتنی بر روشهای داده‌کاوی وجود دارد (Oshikawa et al., 2018). در رویکردهای مبتنی بر اجزاء، چهار مولفه اصلی برای خبر درنظرگرفته می‌شود که عبارتند از: تولید کننده اخبار جعلی، مصرف کننده اخبار جعلی، محتوای اخبار جعلی و بستر اجتماعی خبر جعلی.

۱-۱. رویکرد مبتنی بر اجزاء خبر

بر اساس تجزیه تحلیل اجزای مختلف خبر، رویکردهای شناسایی اخبار جعلی را می‌توان به صورت زیر تقسیم کرد: تجزیه تحلیل تولید کننده و مصرف کننده اخبار، تجزیه تحلیل محتوای اخبار، تجزیه تحلیل بستر اجتماعی.

۱-۱-۱. تحلیل تولید کننده و مصرف کننده اخبار

تولید کنندگان اخبار جعلی آنلاین می‌توانند انسان یا ماشین باشند که هم شامل نویسندگان خبر می‌شود و هم کاربرانی که اخبار جعلی را ناخواسته منتشر می‌کنند. مصرف کنندگان خبر هم می‌توانند دانش‌آموزان، رای‌دهندگان، والدین، افراد سالخورده و... باشند. ماشین‌ها، ربات‌های اجتماعی و سایبورگ‌ها رایج‌ترین تولید کنندگان اخبار جعلی ماشینی هستند. ربات‌های اجتماعی الگوریتم‌های رایانه‌ای هستند که برای نمایش رفتارهای شبیه به انسان طراحی شده‌اند و بطور خودکار محتوا تولید می‌کنند و در رسانه‌های اجتماعی با انسان تعامل دارند (Chu et al., 2012). بسیاری از ربات‌ها به طور خاص برای توزیع شایعات، اسپم و بدافزارها، طراحی شده‌اند. سایبورگ‌ها به ربات‌های کمکی انسان اشاره دارد. پس از ثبت نام توسط یک شخص، سایبورگ می‌تواند توییت ارسال کرده و در شبکه‌های اجتماعی شرکت کند. مشابه ربات‌های اجتماعی، حساب‌های مخرب سایبورگ با انتشار اطلاعات و پیام‌های جعلی، کاربران اجتماعی آنلاین را گمراه کرده، از آنها سوءاستفاده می‌کنند، که ممکن است منجر به آسیب رساندن به اعتماد اجتماعی شود.

کوشش‌ها و تلاش‌های گسترده‌ای برای تجزیه تحلیل حساب‌های مخرب در رسانه‌های اجتماعی صورت گرفته است. در این بخش، روش‌های اصلی مورد استفاده برای تجزیه تحلیل تولیدکننده و مصرف کننده اخبار را مرور و بحث می‌کنیم.

حساب‌های مخرب، رفتاری متفاوت از کاربران قانونی دارند. بنابراین در آنالیز تولیدکننده و مصرف کننده خبر با توجه به پروفایل کاربران، رفتار پستی و زمانی آنها و نیز اعتبار حساب می‌توان به اطلاعات ارزشمندی از آنان دست یافت. اطلاعات اولیه پروفایل

کاربر شامل زبان حساب کاربری، مکان‌های جغرافیایی حساب کاربری، زمان ایجاد حساب کاربری، تعداد پست / توئیت، رفتار پستی و زمانی یک حساب، میانگین زمان بین دو پست متوالی، فراوانی پاسخ، اشتراک گذاری، و اشاره کردن[1] پارامترهایی هستند که به تشخیص صحت خبر منتشر شده توسط کاربر کمک می‌کنند. به کمک زمانسنج‌ها و برنامه‌های خودکار، دیده شده که حساب‌های مشکوکی مانند ربات‌های اجتماعی و سایبورگ‌ها در یک بازه زمانی خاص فعال هستند. در مقابل کاربران انسانی قانونی از نظر زمانی رفتارهای پیچیده‌ای دارند. شدت بالای پاسخ‌ها و رفتارهای ذکر شده ممکن است نشانگر سطح بالای مشکوک بودن یک حساب اجتماعی باشد. همچنین می‌توان اعتبار حساب را با توجه به تعداد دوستان و دنبال کنندگان[2] بررسی کرد. تعداد دنبال کنندگان یک کاربر قانونی شبکه‌های اجتماعی اغلب به دوستان خود نزدیک است. با این حال، ربات‌های اجتماعی معمولاً دوستان بیشتری نسبت به دنبال کنندگان دارند. چو و همکارانش معادله‌ای برای محاسبه اعتبار حساب با تعداد دنبال کنندگان و دوستان پیشنهاد می‌کنند. فرمول (۱) را ببینید:

$$Account\_Reputation = follower / (follower + friend) \quad (۱)$$

آن‌ها مشاهده کردند که یک فرد مشهور همیشه دارای امتیاز شهرت نزدیک به ۱ است، در حالی که این امتیاز برای یک ربات اجتماعی مشکوک نزدیک به صفر است (Chu et al., 2012).

۱-۱-۲. تحلیل محتوای اخبار

یک خبر حاوی محتوای فیزیکی (مانند عنوان، متن، شکل یا ویدیو) و محتوای غیر فیزیکی (مانند هدف، عقیده) است. با بهره‌گیری از تجزیه تحلیل عمیق محتوای اخبار، می‌توانیم الگوهای زبانی و سبک‌های نوشتن را برای اخبار واقعی و جعلی مورد تجزیه تحلیل قرار دهیم، و سپس از مهم ترین ویژگی‌ها برای تشخیص اخبار جعلی آنلاین استفاده کنیم. مطالعات فعلی در مورد تجزیه تحلیل محتوای اخبار، یا زبانی و معنایی هستند یا مبتنی بر سبک (Chu et al., 2012).

هر دو تحلیل زبانی و معنایی مبتنی بر مطالعات کلاسیک و علمی زبان طبیعی هستند. با استخراج اطلاعات مفید از محتوای اخبار، تجزیه تحلیل زبانی و معنایی می‌تواند الگوهای زبان انگلیسی، ساختارها و معانی اخبار را تحلیل کند. بیشتر سازندگان اخبار جعلی برای جلوگیری از شناسایی، از استراتژی‌های مخصوص نوشتن استفاده می‌کنند. هدف اولیه تحلیل زبانی تطبیق صلاحیت زبانی نویسنده خبر با مشاهده قالب‌های زبانی و کشف الگوهای نوشتن است. "Bag-of-words" و "n-grams" متداول ترین روش‌ها برای نمایش متون خبری خام است. در "Bag-of-words" با در نظر گرفتن هر کلمه به عنوان واحد منفرد و یکسان، متن خبر خام می‌تواند به عنوان مجموعه کلمات آن، صرف نظر از دستور زبان و دستور کلمه ارائه شود. در "n-gram"، متن خبری با دنباله‌ای از n آیتم نمایش داده می‌شود، این آیتم‌ها می‌توانند واج‌ها، هجاها، حروف یا کلمات باشند. با این حال سادگی این دو رویکرد منجر به برخی کاستی‌های واضح در پردازش متن خام می‌شود، برای نمونه، مدل "n-gram" دارای پراکندگی زیاد است و نمی‌تواند نمونه‌های خبری را که حاوی نشانه‌های ناشناخته است تفسیر کند. "bag-of-words" ممکن است با نادیده گرفتن بستر و معنای کلمات، اطلاعات قابل توجهی را از دست بدهد. از این رو از رویکردهای آنالیز معنایی استفاده می‌شود که به فرایند توصیف ساختارهای نحوی اخبار از سطح عبارات گرفته تا سطح معنایی اشاره دارد. با ترکیب کردن مدل "n-gram" با مدل " deep syntax"، تجزیه تحلیل معنایی می‌تواند میزان سازگاری بین تجربه شخصی تولیدکننده اخبار و محتوای آن را کشف کند. به عنوان مثال سازندگان اخبار جعلی غالباً برای جلب توجه خوانندگان از عناوین مبالغه‌آمیز استفاده می‌کنند، بنابراین عنوان اخبار جعلی معمولاً نامربوط یا مغایر با محتوای اخبار است درحالیکه عنوان یک خبر واقعی غالبا با محتوای بدنه خبر مطابقت دارد. مثال دیگر نظرات کاربران آنلاین است که اگر جعلی باشد ممکن است حاوی تضاد یا اشتباه باشد، زیرا نظردهندگان فریبکار تجربه‌ای در مورد ویژگی‌ها و خدمات مربوط به محصولات ندارند. محققان می‌توانند با استفاده ترکیبی از تجزیه تحلیل تولیدکننده اخبار و تجزیه

[1] Mentioning
[2] Followers

تحلیل معنایی، سازگاری بین پیشینه کاربر و محتوای اخبار را بررسی کنند، که این امر پیشرفت خوبی را برای طبقه‌بندی و کشف اطلاعات نادرست نشان داده است.

از سوی دیگر با تلاش برای به‌دست آوردن ویژگی‌های متمایز سبک نوشتن بین کابران قانونی و حساب‌های غیر عادی، تجزیه تحلیل مبتنی بر سبک نقش مهمی در شناسایی اخبار جعلی آنلاین دارد. کاربران قانونی آنلاین نظرات، هیجانات و احساسات خود را نسبت به برخی محصولات، رویدادها و خدمات از طریق رسانه‌های اجتماعی ابراز می‌کنند. در حالی که، حساب‌های آنلاین مشکوک با تغییر دادن سبک نوشتن خود یا تلاش برای تقلید از کاربران دیگر، اطلاعات فریبنده را ابراز می‌کنند. تجزیه تحلیل سبک محور را می‌توان به تجزیه تحلیل سبک فیزیکی و تحلیل سبک غیر فیزیکی تقسیم کرد. منظور از تجزیه تحلیل سبک فیزیکی فرایند استخراج ویژگی‌های فیزیکی تاثیرگذار برای تمایز اخبار جعلی از اخبار صادقانه است. این ویژگی‌ها می‌تواند شیوه نوشتن، اسلوب متن و گرایش خاص خبر را نشان دهد. تعداد فعل‌ها و اسم‌ها، تعداد کلمات عاطفی و کلمات اتفاقی. وجود نشانه‌های مشکوک مانند تعداد URL ها، هشتگ‌ها، اشاره کردن به دیگران و کلمات با حروف بزرگ در داده‌های ارتباطات اجتماعی از ویژگی‌های خوبی برای شناسایی تحلیل سبک نوشتن است (Horne & Adali, 2017). این در حالی است که تجزیه تحلیل سبک غیر فیزیکی به بررسی جنبه‌های غیر فیزیکی اخبار، مانند پیچیدگی و خوانایی متن خبر می‌پردازد. سازندگان خبر جعلی معمولاً زمان بیشتری برای نوشتن جملات نیاز دارند و در هنگام نوشتن اشتباه می‌کنند. بنابراین برخی از الگوهای خاص استفاده از کلیدها برای تحلیل سبک نوشتن قابل ردیابی هستند (Banerjee et al., 2014). به عنوان مثال کلید backspace و delete بیشتر وقتی استفاده می‌شود که یک تولیدکننده اخبار جعلی می‌خواهد پیغام غیر واقعی بنویسند.

مدل‌هایی مانند FakeBERT و RoBERTa به طور گسترده برای تشخیص اخبار جعلی استفاده می‌شوند. این مدل‌ها از قدرت تحلیل زبان طبیعی برای بررسی و تشخیص ویژگی‌های متنی که به طور معمول در اخبار جعلی یافت می‌شود، بهره می‌گیرند. به عنوان مثال، FakeBERT با تمرکز بر الگوهای زبانی و نوشتاری، توانسته است دقت بالایی در تشخیص اخبار جعلی به دست آورد (Kaliyar et al., 2021).

۱-۱-۳. تحلیل بستر اجتماعی

تحلیل بستر اجتماعی، مطالعه چگونگی توزیع سریع و گسترده اطلاعات اجتماعی و نحوه تعامل کاربران آنلاین با یکدیگر است. با این حال، بسیاری از رویکردهای اخیر برای کشف اخبار جعلی آنلاین مربوط به تجزیه تحلیل محتوای مستقیم اخبار است. معدود مطالعاتی از تجزیه تحلیل بستر اجتماعی برای پیش بینی اطلاعات آنلاین غیر عادی استفاده می‌کنند. با الگوبرداری کمی از الگوی توزیع داده‌های ارتباطات اجتماعی، تجزیه تحلیل بسترهای اجتماعی قادر به پیش بینی‌های اولیه از سرعت انتشار، مقیاس و تاثیر انتشار اطلاعات و ارائه دیدگاه‌های متنوع و عمیق برای نمایش اخبار جعلی آنلاین است (Zhang & Ghorbani, 2020). در این راستا دو نوع تجزیه تحلیل بسترهای اجتماعی پیشنهاد می شود که می‌توانند کاندیدهای بالقوه‌ای برای افزایش عملکرد روش‌های موجود برای کشف اخبار جعلی باشند که عبارتند از تجزیه تحلیل شبکه کاربر و تجزیه تحلیل الگوی توزیع هستند.

صحت بخشی از اخبار آنلاین را می‌توان در شبکه تولید کننده اخبار شناسایی کرد. کاربران مختلف آنلاین، سوابق تحصیلی، تجربیات کاری و علایق مختلفی دارند، بنابراین کاربران در رسانه‌های اجتماعی تمایل دارند گروه‌هایی را تشکیل دهند که شامل کاربران همفکر باشد. بنابراین این فرضیه منطقی است که از کاربران آنلاینی که با تولیدکننده خبر، در تعامل بسیار هستند، می توان برای پیش بینی سطح صحت اخبار استفاده کرد. در واقع تجزیه تحلیل شبکه کاربر، مطالعه‌ای برای آشکار کردن تعامل بین کاربران آنلاین است.به عنوان مثال، اگر بسیاری از حسابهای غیر عادی یا غیر قابل اعتماد، یک قطعه از خبر را "like" کرده باشند یا در مورد آن اظهار نظر[3] کرده باشند، این خبر به احتمال زیاد حاوی اطلاعات غلط و گمراه کننده است.

از سوی دیگر تجزیه تحلیل الگوی توزیع، مطالعه‌ای برای تجزیه تحلیل ویژگی‌های انتشار اطلاعات است (Russell, 2013). الگوهای انتشار اخبار آنلاین می‌تواند اطلاعات ارزشمند و آموزنده‌ای را برای پیشنهاد وقایع و پیام‌های غیر عادی فراهم کند. در

[3] Comment

سال‌های اخیر، بسیاری از محققان در حال کار بر روی تشخیص الگوی غیر عادی برای اطلاعات اجتماعی هستند، که می‌تواند به کاربران آنلاین کمک کند داده‌های متفاوت، غیر معمول یا غیر منتظره را کشف کنند. اما، کشف الگوی توزیع مشکوک برای اخبار جعلی، به دلیل ماهیت ناهمگن و پویای رفتارهای اجتماعی آنلاین پیچیده و چالش برانگیز است (Yang et al., 2012). در نتیجه، سیستم‌های مصورسازی[۴] پیشرفته همیشه به منظور پاسخگویی به چالش‌های فوق، با الگوریتم‌های یادگیری ماشین کلاسیک همراه هستند. به طور خلاصه، اطلاعات فراوانی هر روز از طریق رسانه‌های اجتماعی پخش می‌شود و بیشتر اوقات، تجزیه تحلیل محتوای اخبار برای ساختن یک سیستم موثر برای کشف اطلاعات کاذب کافی نیست. با الگوبرداری کمی از الگوی توزیع داده‌های ارتباطات اجتماعی، تجزیه تحلیل بسترهای اجتماعی قادر به پیش بینی‌های اولیه از سرعت انتشار، مقیاس و تاثیر انتشار اطلاعات و ارائه دیدگاه‌های متنوع و عمیق برای نمایش اخبار جعلی آنلاین است.

تحلیل رفتار کاربران در شبکه‌های اجتماعی و بررسی گراف‌های اجتماعی نشان داده است که الگوهای انتشار اطلاعات جعلی می‌توانند با استفاده از مدل‌های آنلاین، شناسایی و به‌صورت آنی تحلیل شوند. مدل‌های آنلاین مبتنی بر یادگیری ماشین، مانند الگوریتم‌های حاشیه بزرگ تقریبی و الگوریتم‌های غیرفعال–تهاجمی، قابلیت شناسایی اخبار جعلی را به صورت بلادرنگ (Real-time) فراهم کرده‌اند.

تحقیقات اخیر نشان داده است که استفاده از مدل‌های ترکیبی می‌تواند به بهبود دقت و سرعت شناسایی کمک کند (Zhu et al., 2024). به عنوان نمونه در پژوهشی سیستمی طراحی شده است که از یک سیستم مبتنی بر ادغام ویژگی‌های شبکه و کاربر با بهره‌گیری از شبکه مولد متخاصم شرطی برای تنظیم مجموعه داده‌ها ساخته شده است. همچنین، این سیستم با مدل‌سازی شبکه اجتماعی توئیتر به عنوان نمونه‌ای از تعاملات کاربران و تبدیل گره‌ها به بردارهای ویژگی، قادر به شناسایی کاربران منتشرکننده اخبار جعلی است. بر طبق نتایج به‌دست آمده در این پژوهش، سیستم پیشنهادی در معیار دقت ۱۱ درصد، در فراخوانی ۱۳ درصد، در معیار اف ۱۲ درصد و در صحت نیز ۱۲ درصد، بهبود یافته و توانسته است دقت تقریبی ۹۹ درصد را در شناسایی کاربران منتشرکننده اخبار جعلی به دست آورد (فرضی، ۱۴۰۰).

۱–۲. رویکرد مبتنی بر روشهای داده‌کاوی

روش دیگر شناسایی اخبار جعلی به صورت خودکار رویکردهای مبتنی بر داده‌کاوی است که در ادامه به بررسی آن می‌پردازیم. بر اساس دسته‌های مختلف تجزیه تحلیل مولفه‌ها، انواع مختلفی از ویژگی‌ها را می‌توان از داده‌های ارتباطات آنلاین استخراج کرد و برای یادگیری ساخت مدل استفاده کرد. مدل‌های یادگیری ماشین موجود را می‌توان به سه رویکرد مدل‌های نظارت شده، نیمه نظارتی و بدون نظارت تقسیم‌بندی کرد (Afroz et al., 2012). در این مقاله، مدل‌های یادگیری نظارت شده و مدل‌های یادگیری بدون نظارت عمدتا برای کشف اخبار جعلی آنلاین مورد بحث قرار می‌گیرند.

یادگیری نظارت شده: الگوریتم‌های یادگیری ماشین نظارت شده مانند درخت تصمیم‌گیری[۵]، جنگل تصادفی[۶]، ماشین بردار پشتیبانی[۷]، رگرسیون لجستیک[۸] و الگوریتم‌های یادگیری ماشین نظارت شده مانند درخت تصمیم‌گیری ، جنگل تصادفی ، ماشین بردار پشتیبانی ، رگرسیون لجستیک و K نزدیکترین همسایه به طور گسترده برای شناسایی کلاهبرداری‌های آنلاین و طبقه‌بندی اطلاعات فریبنده استفاده شده است. ایزی و رضایی با پیاده‌سازی روش K نزدیکترین همسایه به ۹۸ درصد دقت دست یافتند (ایزی و رضایی،۱۴۰۲). ترکیب این روشها نیز در شناسایی اخبار جعلی استفاده شده است به عنوان مثال در پژوهشی با توجه به ساختار کلمات انگلیسی، مدل tf-idf و الگوریتم‌های درخت تصمیم، k نزدیک‌ترین همسایه و رگرسیون لجستیک، اخبار جعلی را از اخبار واقعی با استفاده از تکنیک رأی‌گیری اکثریت شناسایی می‌کند. در این پژوهش، با به‌کارگیری این سه الگوریتم یادگیری،

[4] Visualization
[5] Decision Tree
[6] Random Forest
[7] Support Vector Machine (SVM)
[8] Logistic Regression

مدلی ترکیبی برای تفکیک اخبار جعلی از واقعی ایجاد شده است. بر اساس نتایج و با توجه به معیارهای مختلف ارزیابی دقت، مدل پیشنهادی تقریباً ۸۸ درصد دقت را نشان داد (وظیفه آبان، حسنی آهنگر،۱۴۰۲).

طی سال‌های اخیر، الگوریتم‌های یادگیری عمیق موفقیت‌های بزرگی را در حوزه‌های تشخیص گفتار و شناخت شیء بصری کسب کرده‌اند. متفاوت از تکنیک‌های معمول یادگیری ماشین که به استخراج ویژگی‌های دستی نیاز دارد، الگوریتم‌های یادگیری عمیق را می‌توان با داده‌های خام تغذیه کرد و بازنمایی‌ها را به صورت خودکار کشف کرد. به طور خاص، الگوریتم‌های یادگیری عمیق مانند شبکه عصبی بازگشتی[9]، حافظه طولانی کوتاه–مدت[10]،LSTM.دوطرفه[11]،GRU[12] در آشکار ساختن ساختار توالی در داده‌های ابعاد بالا، به خوبی عمل می‌کنند و پتانسیل چشمگیری را در پردازش زبان طبیعی، مانند طبقه‌بندی موضوعی، تجزیه تحلیل احساسات، پاسخ دادن به سوال و ترجمه زبان نشان داده‌اند. بنابراین، روش‌های مبتنی بر یادگیری عمیق راه حل‌های خوبی برای نمایش و کشف اخبار جعلی آنلاین هستند (شفیعی‌وشاره، علی‌عسگری‌رنانی و محمدی، ۱۴۰۳).

روش‌های کشف اخبار جعلی با به‌کارگیری الگوریتم‌های یادگیری عمیق به ویژگی‌های متنی دست ساز متکی نیستند و می‌توانند مفاهیم پنهان اطلاعات متنی اخبار و نویسنده را درطی زمان به دست آورند. علاوه بر این، عملکرد طبقه‌بندی یادگیری عمیق می‌تواند از طریق واحدهای محاسبات پیشرفته و لایه‌های پنهان اضافی بیشتر، بهبود یابد. تحقیقات نشان می‌دهد که LSTM با دقت ۹۲ درصد برای تشخیص اخبار جعلی عملکرد بهتری نسبت به روش‌های قدیمی مانند Naive Bayes دارد (Farokhian et al., 2022). ایزی و حسن پور (۱۴۰۲) با به‌کارگیری الگوریتم‌های یادگیری عمیق به دقت ۰.۹۶ دست یافتند که نسبت به روش‌های یادگیری ماشین نتیجه بهتری است. این مدل‌ها با تحلیل ساختار و ویژگی‌های زبانی، تفاوت‌های میان اخبار جعلی و واقعی را تشخیص می‌دهند. در پژوهشی، سه مدل ترکیبی شامل CNN+ RNN ساده، CNN + GRU و CNN + BiLSTM در قالب معماری رمزگشا–رمزگذار برای پیش‌بینی اخبار جعلی ارائه شده است (انصاری، مومن زاده و ارفعی‌نیا،۱۴۰۳). ترکیب CNN-RNNدر زمینه‌های یادگیری عمیق به عنوان یک روش کارآمد شناخته شده، زیرا این مدل‌ها قادر به ضبط ویژگی‌های ترتیبی و محلی در داده‌های ورودی هستند. این مدل‌ها با موفقیت در دو مجموعه داده خبری جعلی با طبقه‌بندی باینری (ISOT) و چندکلاسه (FNC-1) آموزش و ارزیابی شدند. نتایج نشان داد که مدل CNN + BiLSTM عملکرد بهتری نسبت به دو مدل هیبریدی دیگر در زمینه طبقه‌بندی باینری و چندکلاسه برای شناسایی اخبار جعلی دارد و به دقت ۹۶.۲۵ درصد در مجموعه داده آزمایشی دست یافته است.

عملکرد یک مدل یادگیری نظارت شده به شدت به کیفیت مجموعه داده دارای برچسب بستگی دارد. ولی تولید یک مجموعه داده گسترده و با کیفیت خوب برای کشف اخبار جعلی دشوار است، چرا که اولا مجموعه داده آنلاین در دنیای واقعی معمولاً بزرگ، ناقص، بدون ساختار، بدون برچسب و دارای عوامل مزاحم است. ثانیا هر روز مقدار زیادی از اطلاعات نادرست با اهداف متنوع و ویژگی‌های زبانی مختلف از طریق رسانه‌های اجتماعی ایجاد می‌شود که در نتیجه به‌دست آوردن برچسب درستی مستدل برای داده ها را دشوار می‌کند.

یادگیری بدون نظارت: با توجه به محدودیتهای مدل یادگیری نظارت شده که در بخش قبل به آن اشاره شد، مدل یادگیری بدون نظارت برای حل مسئله در دنیای واقعی کاربردی‌تر و عملی‌تر است. با این حال، تنها مطالعات کمی وجود دارد که به طور مستقیم بر روی کشف اخبار جعلی آنلاین از طریق رویکردهای بدون نظارت کار کند. بیشتر آنها بر تحلیل شباهت معنایی یا عقیده‌کاوی متمرکز شده‌اند و اندازه گیری تشابه بدون نظارت را برای تشخیص نظرات کاربران جعلی به‌کار برده‌اند (Ahmed, 2017). با ترکیبی از شباهت کلمه‌ای و شباهت ترتیب کلمات، رویکرد پیشنهادی احمد، می‌تواند نظرات تقریباً تکراری کاربران را با دقت بالا شناسایی کند. ونگ و همکارانش یک چهارچوب عقیده‌کاوی بدون نظارت برای تصاویر در رسانه‌های

[9] Recurrent Neural Network (RNN)
[10] Long Short Trem Memory (LSTM)
[11] Bidirectional LSTM
[12] Gated Recurrent Unit

اجتماعی ارائه دادند (Wang & Li, 2015). با بهره گیری از روابط بین اطلاعات بصری و اطلاعات متنی مرتبط، روش آنها می‌تواند عقیده مستتر در تصاویر اجتماعی را از دو مجموعه داده بزرگ پیش‌بینی کند. یک مدل بدون نظارت مولد بیزین برای تشخیص نظرات کاربران جعلی نیز در تحقیقات در این حوزه استفاده شده است (Mukherjee et al., 2012). این روش بر اساس اندازه‌گیری شباهت کسینوسی، و با استفاده از امتیازات و ویژگی‌های زمانی، به طور خودکار ویژگی‌های تفکیک کننده راویان راستگو و متقلب را تشخیص می‌دهند. ژانگ و همکارانش معتقد هستند که الگوریتم‌های یادگیری بدون نظارت مسیرهای عملی و ضروری برای کشف اخبار جعلی آنلاین هستند و باید در تحقیقات آینده اولویت اصلی قرار گیرند (Zhang & Ghorbani, 2020). آنها انواع مدل یادگیری بدون نظارت را برای تشخیص اخبار جعلی پیشنهاد می‌کنند که عبارتند از: خوشه‌بندی[۱۳]، تحلیل داده دورافتاده[۱۴]، تحلیل تشابه معنایی[۱۵]، و نهفتگی اخبار بدون نظارت[۱۶].

همانطور که اشاره شد در رویکردهای مبتنی بر یادگیری نظارت شده مجموعه‌ای از ویژگی‌های اخبار مورد تجزیه و تحلیل قرار می‌گیرند. در ادامه ویژگی‌های مهم و رایج مورد استفاده برای کشف اخبار جعلی آنلاین را مورد بحث قرار می‌دهیم. بر اساس مولفه‌های اخبار جعلی که در جدول ۱ مورد بحث قرار گرفته است، سه نوع اصلی از مجموعه ویژگی‌ها وجود دارد: ویژگیهای مبتنی بر تولیدکننده/کاربر، ویژگیهای مبتنی بر محتوای خبر، ویژگهای مبتنی بر بستر اجتماعی.

جدول (۱) دسته‌بندی ویژگیهای اخبار جعلی

ویژگیهای اخبار جعلی								
ویژگیهای مبتنی بر بستر اجتماعی			ویژگیهای مبتنی بر محتوای اخبار			ویژگیهای مبتنی بر سازنده/کاربر		
ویژگیهای مبتنی بر زمان	ویژگیهای مبتنی بر شبکه	ویژگیهای مبتنی بر توزیع	ویژگیهای بصری	ویژگیهای مبتنی بر سبک	ویژگیهای زبانی و نحوی	ویژگیهای رفتار کاربر	ویژگیهای اعتبار کاربر	ویژگیهای پروفایل کاربر

۱-۲-۱. ویژگی‌های مبتنی بر تولیدکننده/کاربر

ویژگی‌های مبتنی بر تولیدکننده/کاربر به طور گسترده برای شناسایی حسابهای آنلاین مشکوک مورد استفاده قرار می‌گیرند. هدف از این ویژگی‌ها به‌دست آوردن ویژگی‌های منحصر به فرد حساب‌های مشکوک یا حساب‌های ماشینی است. از این دسته ویژگیها می‌توان به موارد زیر اشاره کرد. مشخصات پروفایل کاربر شامل اطلاعات اصلی کاربر مانند نام حساب، اطلاعات جغرافیایی، داده‌های ثبت نام کاربر، تایید شده یا خیر، دارای توضیحی است یا خیر. مولفه اعتبار کاربر شامل امتیاز اعتبار کاربر، تعداد دوستان و دنبال‌کنندگان کاربر، نسبت بین دوستان کاربر و دنبال‌کنندگان، تعداد کل توییت ها/پست‌های کاربر و همچنین ویژگی‌های رفتار کاربر که به منظور دستیابی به الگوی رفتار کاربر هم برای کاربران فریبکار و هم کاربران مشروع مورد استفاده قرار می‌گیرند. یک ویژگی معمولی رفتار کاربر، امتیاز ناهنجاری کاربر است که با استفاده از تعداد تعامل کاربر در یک پنجره زمانی، تقسیم بر میانگین تعامل ماهانه کاربر، برای تشخیص اطلاعات غیر عادی به کار برده می‌شود (Kaliyar et al., 2021).

[13] Cluster analysis
[14] Outlier analysis
[15] Semantic similarity analysis
[16] Unsupervised news embedding

۱-۲-۲. ویژگی‌های مبتنی بر محتوای اخبار

ویژگی‌های مبتنی بر محتوای اخبار از سر نخ‌های آشکار برای کشف اخبار جعلی است. این مشخصه‌ها را می‌توان به ویژگی‌های‌زبانی و نحوی، ویژگی‌های مبتنی بر سبک نوشتن و ویژگی‌های بصری طبقه بندی کرد.

ویژگی‌های زبانی و نحوی به مولفه اساسی، ساختاری و معناشناسی زبان طبیعی اشاره دارد. اگر چه محتوای خبری جعلی همیشه به عمد برای گمراه کردن کاربران آنلاین تولید می‌شود، اما ویژگی‌های مبتنی بر زبان و نحو هنوز منابع ارزشمندی برای تجزیه تحلیل اخبار مشکوک هستند و آنها را می‌توان از این نظر به سه سطح کلمه، جمله و محتوا طبقه بندی کرد.

سطح کلمه: بسته کلمات[۱۷]، n-gram، فراوانی کلمه، فروانی وزنی IDF[۱۸]بسته کلمات ، n-gram رایج ترین ویژگی‌های زبانی برای پردازش زبان طبیعی هستند. بعلاوه وجود نشانه‌های خاص و مشکوک در محتوای خبر می‌تواند برای شناسایی اطلاعات جعلی استفاده شود (Castillo et al., 2011). علائم خاص و مشکوک شامل علامت تعجب، علامت سوال، چندین علامت تعجب، اشاره به کاربر، هشتگ، علامت لبخند، علامت اخم، حروف بزرگ، کلمات برجسته، می‌باشد. مشابه علائم مشکوک، وجود کلمات مربوط به سبک همچنین می‌تواند برای کشف اخبار جعلی آنلاین استفاده شود. کلمات مربوط به سبک عبارتند از: کلمات توقف، علائم نگارشی، نقل قول، کلمات منفی (مانند نه، هرگز، نچ، علیرغم، تردید، جعل، تخریب، انکار، تقلب و نادرست و...)، کلمات محاوره‌ای/سوگند، عبارات پرسشی چرا، چه، چه زمان، چه کسی، اسامی، ضمایر شخصی، ضمایر ملکی، قیود، حروف ندا، افعال، کلمات کمی‌ساز، کلمات مقایسه‌ای، علامت‌های تعجب، عبارات عامیانه.

LIWC [۱۹]یک برنامه تحلیل متن است که کلمات را در مقوله‌های معنادار از لحاظ روانشناختی شمارش می‌کند. LIWC شامل پنج دسته اصلی و زیر شاخه‌های مختلف مانند اجتماعی، عاطفی، شناختی، ادراکی است. بر اساس LIWC، محققان می‌توانند تعداد کلمات احساسی در محتوای اخبار را حساب کنند، که ممکن است به تعیین سطح کلی احساسات موجود درخبر کمک کند. کلمات احساسی شامل کلمات تحلیلی، کلمات حسی، کلمات بیان کننده علت، کلمات عدم توافق، کلمات موقت، کلمات یقین، واژه‌های متمایز، کلمات وابستگی، کلمات قدرت، کلمات با ارزش، کلمات خطرناک، کلمات دغدغه‌های شخصی (شغل، اوقات فراغت، دین، سرمایه)، کلمات با لحن احساسی، کلمات هیجانی (عصبانیت، غم) و.... هستند. سایر ویژگی‌های زبانی کلمات، مانند خوانایی کلمات (امتیاز خواندن بر اساس تعداد هجا در کلمات) و نسبت نوع-نشانه (تعداد کلمات منحصر به فرد تقسیم بر تعداد کل کلمات در اسناد/پست‌ها/توییت‌ها) می‌تواند سادگی و تنوع واژگان در اخبار را نشان دهد (Pennebaker et al., 2003)، و همچنین می‌تواند برای تجزیه تحلیل محتوای جعلی اخبار استفاده شود.

سطح جمله: ویژگی‌های جمله به تمام ویژگی‌های مهم در مقیاس جمله اشاره دارد و شامل برچسب گذاری پاره‌های گفتار [۲۰] ، طول متوسط جمله، میانگین طول یک توییت/پست، فراوانی علائم نگارشی، کلمات دستوری، عبارت در یک جمله، تقارن متوسط جمله(مثبت، خنثی یا منفی)، پیچیدگی جمله و.... . به طور خاص محققان پیچیدگی جمله، عمق درخت نحوی هر جمله، عمق درخت نحوی عبارات اسمی و عمق درخت نحوی عبارات فعلی را با استفاده از Stanford parser محاسبه می‌کنند که برای کشف اخبار جعلی آنلاین استفاده شده است (Horne & Adali, 2017).

سطح محتوا: ویژگی‌های محتوا به اطلاعات خام محتوای متا اخبار اشاره دارد. امتیاز کلی احساسات محتوای خبری به عنوان ابزاری قدرتمند برای اخبار جعلی و تحلیل نویسنده مشکوک تایید شده است. در مطالعات قبلی، از SentiStrength برای شناسایی شدت احساسات مثبت و منفی در سند خام استفاده شده است. به طور مشابه، ویژگیهای موضع محور[۲۱] رفتار حمایتی یا انکاری

[17] Bag-of-words
[18] Term frequency-inverted document frequency (TF-IDF)
[19] Linguistic Inquiry and Word Count
[20] Parts of speech
[21] Stance-based

یک توییت/پست را نشان می‌دهند و می‌توانند برای ارزیابی وضعیت احساسی اخبار استفاده شود. سایر ویژگی‌ها در سطح محتوای خبر، از قبیل عنوان خبری پررنگ، موضوعات خبری (زندگی اجتماعی، سیاست، فناوری و امنیت سایبری، تجاری و مالی و...)، قطعیت خبر، تعداد نشانه‌های خاص یا نمادها در کل اخبار، وجود لینک‌های خارجی یا URL ها نیز سرنخ‌های مهمی برای شناسایی آنلاین اخبار جعلی هستند.

همانطور که اشاره شد علاوه بر ویژگیهای زبانی و نحوی خبر مولفه‌های مبتنی بر سبک نیز می‌توانند برای بررسی محتوای خبر مورد استفاده قرار گیرند. این مولفه‌ها جهت نشان دادن ویژگی‌های مختلف سبک نوشتن نویسندگان اخبار جعلی به کار می‌روند. اگر چه بیشتر اوقات، نویسندگان اخبار جعلی سعی در تقلید از شیوه نوشتن یک نویسنده اخبار عادی دارند تا خوانندگان آنلاین را فریب دهند، اما هنوز هم تفاوت‌هایی وجود دارد که می‌تواند باعث ایجاد تمایز بین سازندگان اخبار جعلی و اخبار واقعی شود. به منظور شناسایی نظرات جعلی، در پژوهشی ویژگی‌های استفاده از کلیدها برای هر دوی محتوا سازان جعلی و نظردهندگان واقعی آنلاین مطالعه شده است. با تمرکز بر ویژگی‌های الگوی ویرایش (مانند تعداد پاک‌کردن‌ها، کلیک موس و استفاده از arrow key ها) و ویژگی‌های بازه زمانی (مانند مدت زمان کل گزارش، فاصله متوسط بین کلمات) محققان دریافتند که سازندگان محتوای جعلی برای اتمام نوشتن به زمان بیشتری نیاز دارند و اشتباهات بیشتری انجام می‌دهند (Ahmed, 2017) علاوه بر ویژگی‌های استفاده از کلیدها از بسیاری از ویژگی‌های مبتنی بر الگو نیز استفاده شده است (Castillo et al., 2011; Zhao et al., 2015) که می‌تواند برای تعیین الگوی منحصر به فرد محتوای اخبار جعلی استفاده شود. این ویژگی‌های مبتنی بر الگو، شامل کسری از توییت‌ها/پست‌ها است که حاوی لینک‌های خارجی، اشاره به کاربر، هشتگ ها در یک بازه زمانی، اینکه لینک‌های خارجی از نام دامنه مشهور استفاده می‌کند یا خیر و....می‌باشد.

علاوه بر متون موجود در خبر، تصاویر یا فیلم‌های موجود در یک محتوای خبری نیز نشانه‌های اساسی برای شناسایی اطلاعات مشکوک یا فریبنده هستند. مطالعات اخیر ویژگی‌های بصری را برای شناسایی اطلاعات غلط آنلاین بررسی می‌کنند. این ویژگی‌های بصری شامل تعداد تصاویر یا فیلم‌ها، امتیاز وضوح، امتیاز انسجام، نمودار توزیع شباهت، امتیاز تنوع، امتیاز خوشه‌بندی، نسبت تصویر، نسبت چند تصویر، نسبت تصویر داغ، نسبت تصویر طولانی، و... هستند.

۱-۲-۳. ویژگی‌های مبتنی بر بستر اجتماعی

ویژگی‌های مبتنی بر بستر اجتماعی برای انعکاس الگوی توزیع اخبار آنلاین و تعامل بین کاربران آنلاین طراحی شده‌اند و می‌توان آنها را در سه نوع زیر خلاصه کرد: ویژگی‌های مبتنی بر شبکه، ویژگی‌های مبتنی بر توزیع و ویژگی‌های مبتنی بر زمان.

تجزیه تحلیل مبتنی بر شبکه بر روی گروهی از کاربران آنلاین که از جنبه‌های مختلف، از قبیل مکان، پیشینه تحصیلات و عادات، مشابه هستند تمرکز دارد. ویژگی‌های مبتنی بر شبکه بر اساس شبکه‌های خاص، انتخاب و استخراج می‌شوند و می‌توانند برای مطالعه ویژگی‌های منحصر به فرد شبکه‌های خاص و شباهت و عدم تشابه حساب‌های آنلاین مختلف استفاده شوند. شو و همکارانش یک جمع بندی مختصر از تجزیه تحلیل مبتنی بر شبکه‌های مختلف، مانند شبکه وضعیت، شبکه همزمان، شبکه دوستی و شبکه انتشار ارائه می‌دهد (Shu et al., 2019).

ویژگی‌های مبتنی بر توزیع می‌توانند به کشف الگوی متمایز انتشار اخبار آنلاین کمک کنند. معمولاً می‌توان یک درخت تکثیر را برای سهولت در توصیف ماهیت توزیع یک خبر ساخت. ویژگی‌های مربوط به درخت تکثیر شامل درجه ریشه در یک درخت تکثیر، حداکثر تعداد زیر درخت، حداکثر/متوسط درجه و عمق درخت و غیره می باشد. علاوه بر این، برخی از ویژگی‌های دیگر مانند تعداد بازتوییت ها/بازنشرها برای توییت/پست اصلی و بخشی از توییت ها/پست‌هایی که برای یک حساب آنلاین مجدد توییت می‌شوند، برای بررسی تاثیر، محبوبیت و سطح مشکوک بودن اخبار جعلی آنلاین به‌کار برده می‌شوند.

از ویژگی‌های مبتنی بر زمان نیز می‌توان برای توصیف رفتار ارسال اخبار آنلاین به صورت سری زمانی استفاده کرد. این ویژگی‌ها برای شناسایی فعالیت‌های ارسال مشکوک مناسب هستند و می‌توانند برای نشان دادن سطح نادرستی اخبار آنلاین استفاده

شوند. از ویژگی‌های متداول زمانی که معمولاً استفاده می‌شود، عبارتند از فاصله بین دو پست، دفعات ارسال، پاسخ دادن و نظر دادن برای یک حساب خاص، زمان و روزی که اطلاعات اصلی ارسال می شود/به اشتراک گذاشته می‌شود یا نظر داده می‌شود.

بحث و نتیجه‌گیری

با توجه به دشواری و پیچیدگی تشخیص آنلاین اخبار جعلی، یک مدل دسته‌بندی دودویی برای تشخیص ویژگی‌های اطلاعات جعلی آنلاین بسیار ناکافی است. با توجه به بلادرنگ و ناهمگن بودن داده‌های ارتباطات اجتماعی، استفاده از سیستم‌های بلادرنگ برای تشخیص اطلاعات نادرست یا رفتارهای غیر عادی کاربر در لحظه وقوع، ضروی است. با استفاده از سیستم‌های بلادرنگ، کاربران اینترنت یا متخصصان فضای مجازی می‌توانند یک گام جلوتر از توزیع کامل اطلاعات نادرست آنلاین باشند تا اثرات چنین حملات اطلاعاتی را کاهش دهند. یک سیستم بلادرنگ که اغلب با الگوریتم‌های یادگیری بدون نظارت و تکنیک‌های مصورسازی همراه است، می‌تواند با روند جدید اخبار جعلی مطابقت داشته، در سناریوهای بیشتری استفاده شود. با این حال، حجم گسترده داده ها، ابعاد بالای داده‌های بلادرنگ و ماهیت ناهمگن جریان داده‌های آنلاین چالش‌های زیادی را برای ایجاد یک سیستم تشخیص بلادرنگ، از نظر ذخیره داده ها و محاسبه داده ها ایجاد می‌کند.

در حال حاضر اکثر مطالعات سعی در ارزیابی درست یا نادرست بودن اطلاعات آنلاین دارند، اما شناسایی هر گونه اخبار دروغ پرطرفدار یا بالقوه در اسرع وقت بسیار مهم است. به همین دلیل در کنار سیستم‌های تشخیص اخبار جعلی، پیش بینی زودهنگام خبر جعلی و مداخله در انتشار اخبار جعلی نیز مورد توجه محققان قرار گرفته است. پیش‌بینی زودهنگام اخبار جعلی با یادگیری از داده‌های گذشته، تلاش می‌کند که اخبار جعلی یا شایعات نوظهور را حتی قبل از وقوع آنها کشف کند. برخی پژوهشگران معتقدند که به کمک اطلاعات مربوط به قطبش[۲۲] و تایید کاربر، قادر به شناسایی موضوعاتی هستند که مستعد سوء اطلاعات هستند (Vicario et al., 2019). برخی پژوهشگران نیز مسئله تشخیص شایعات زود هنگام در رسانه‌های اجتماعی را بررسی کرده‌اند (Vicario et al., 2019; Zhao et al., 2015).

پیش بینی زود هنگام اخبار جعلی می‌تواند قبل از وجود اخبار جعلی، اطلاعات نادرست بالقوه را به کاربران آنلاین یادآوری کند. مداخله در انتشار اخبار جعلی می‌تواند به کاربران آنلاین کمک کند تا تاثیرات منفی اخبار جعلی را پس از انتشار اخبار جعلی پاک کنند. فرج تبار و همکارانش معتقدند که ترکیب یادگیری تقویتی با مدل شبکه فرایند نقطه‌ای (Farajtabar et al.,)، تاثیر اخبار جعلی در رسانه‌های اجتماعی را کاهش می‌دهند.

اخیرا مدل‌هایی مانند ChatGPT در تولید داده‌های مصنوعی برای ایجاد توازن در مجموعه داده‌ها استفاده شده‌اند. این داده‌ها به مدل‌های یادگیری ماشین کمک می‌کنند تا داده‌های نادرست را با دقت بیشتری طبقه‌بندی کنند. این روش به خصوص در مواردی که داده‌های نامتوازن (مانند تعداد کمی از نمونه‌های اخبار جعلی) موجود است، کاربرد دارد (Shushkevich et al., 2023) .

روش‌های نوین مبتنی بر یادگیری عمیق، شامل مدل‌های زبانی بزرگ مانند GPT-3 و BERT، در سال‌های اخیر پیشرفت‌های قابل توجهی در شناسایی اخبار جعلی داشته‌اند. این مدل‌ها به دلیل توانایی‌شان در درک زبان و شناسایی الگوهای پیچیده متنی، به عنوان ابزارهایی مؤثر در شناسایی اخبار نادرست مورد استفاده قرار گرفته‌اند. مطالعه‌ای جدید در سال ۲۰۲۴ نشان می‌دهد که مدل‌های زبانی بزرگ (LLMs) می‌توانند به طور مؤثری اخبار جعلی تولیدشده توسط ماشین را شناسایی کنند، اما تشخیص اخبار جعلی تولیدشده توسط انسان همچنان یک چالش بزرگ باقی‌مانده است. به این ترتیب، پژوهشگران پیشنهاد می‌کنند که مدل‌های شناسایی اخبار جعلی باید به گونه‌ای آموزش داده شوند که توانایی تشخیص هر دو نوع اخبار جعلی را داشته باشند (Su et al., 2023).

1. Polarization

تحقیقات جدید نشان می‌دهند که اخبار جعلی تنها به یک دسته خاص محدود نمی‌شوند و می‌توانند شامل محتوای "نیمه‌حقیقی" نیز باشند. به همین دلیل، مدل‌هایی مانند RoBERTa برای تشخیص اخبار جعلی در چندین دسته به کار می‌روند که شامل اخبار "حقیقی"، "جعلی"، "نیمه‌حقیقی" و سایر دسته‌بندی‌ها است (Shushkevich et al., 2023).

استفاده از مدل‌های چندحالته که شامل ترکیب داده‌های متنی، تصویری و صوتی هستند، می‌تواند دقت شناسایی اخبار جعلی را افزایش دهد. این رویکردها به دلیل توانایی‌شان در تحلیل اطلاعات از چند منبع مختلف، به‌خصوص در شرایطی که اطلاعات نادرست به صورت چندرسانه‌ای منتشر می‌شود، بسیار مؤثر هستند. تحقیق جدیدی در سال ۲۰۲۴ نشان داده است که رویکردهای چندحالته توانایی شناسایی اخبار جعلی با دقت بالاتری نسبت به رویکردهای تک‌حالته دارند (Vyas et al., 2024).

با توجه به ظهور مدل‌های زبانی بزرگ و افزایش تولید اخبار جعلی توسط هوش مصنوعی، نیاز به رویکردهای ترکیبی و مدل‌های چندحالته بیشتر از گذشته احساس می‌شود. این مدل‌ها می‌توانند با تحلیل چندمنظوره (Multi-modal) و استفاده از داده‌های گوناگون، دقت شناسایی را افزایش داده و چالش‌های مربوط به اطلاعات نادرست را بهتر مدیریت کنند. پژوهش‌های اخیر نشان داده‌اند که کیفیت و تنوع داده‌ها نقش کلیدی در موفقیت این مدل‌ها دارند (Kuntur et al., 2024).

نتیجه‌گیری

با گسترش روزافزون اخبار جعلی در رسانه‌های اجتماعی و پلتفرم‌های دیجیتال، اهمیت توسعه روش‌های مؤثر برای شناسایی و مقابله با این اطلاعات نادرست بیش از پیش نمایان شده است. این مقاله با بررسی جامع رویکردهای مختلف شناسایی اخبار جعلی، از روش‌های سنتی مبتنی بر قوانین و استخراج ویژگی‌ها تا تکنیک‌های پیشرفته مانند یادگیری عمیق و پردازش زبان طبیعی(NLP) ، نشان می‌دهد که هر یک از این روش‌ها مزایا و محدودیت‌های خاص خود را دارند.

نتایج بررسی‌ها نشان می‌دهد که در حالی که تکنیک‌های سنتی نقطه شروع مناسبی برای شناسایی اخبار جعلی محسوب می‌شوند، اما در مواجهه با پیچیدگی‌ها و ظرافت‌های اطلاعات نادرست تولیدشده توسط انسان‌ها و ماشین‌ها، کارایی لازم را ندارند. در مقابل، مدل‌های یادگیری عمیق و چندحالته که از داده‌های متنی، تصویری و حتی رفتاری کاربران استفاده می‌کنند، توانسته‌اند دقت بالاتری را در تشخیص اخبار جعلی به دست آورند. با این حال، این مدل‌ها نیز همچنان در تشخیص محتوای چندمنظوره و تولیدشده توسط مدل‌های زبانی بزرگ (LLMs) با چالش‌هایی روبه‌رو هستند.

به منظور بهبود کارایی سیستم‌های شناسایی اخبار جعلی، پیشنهاد می‌شود که از رویکردهای ترکیبی و یکپارچه‌ای استفاده شود که تحلیل‌های زبانی، شبکه‌های اجتماعی و رفتار کاربران را با هم ادغام می‌کنند. همچنین، افزایش کیفیت و تنوع داده‌های آموزشی و توجه به رفتار و ترجیحات کاربران می‌تواند در بهبود عملکرد این سیستم‌ها مؤثر باشد.

در نهایت، برای مقابله مؤثر با انتشار گسترده اخبار جعلی در فضای دیجیتال، همکاری میان پژوهشگران، توسعه‌دهندگان، و پلتفرم‌های اجتماعی ضروری است. این همکاری‌ها می‌توانند به توسعه ابزارهایی منجر شوند که نه تنها از دقت بالایی برخوردار هستند، بلکه قابلیت تطبیق‌پذیری و کاربرد در زمان واقعی (Real-time) را نیز دارا می‌باشند. در نتیجه، با ترکیب روش‌های پیشرفته یادگیری ماشین و تحلیل‌های چندحالته، می‌توان به راهکارهای مؤثرتری برای مقابله با این معضل رو به رشد دست یافت.

ملاحظات اخلاقی

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

سپاسگزاری

از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

منابع

انصاری، وحید؛ مومن‌زاده، حسین و ارفعی‌نیا، حسن؛ (۱۴۰۳). تشخیص اخبار جعلی با استفاده از شبکه عصبی عمیقCNN ، *هفتمین کنفرانس بین المللی مهندسی برق، کامپیوتر، مکانیک و هوش مصنوعی*، https://civilica.com/doc/2046572

ایزی، جلال و حسن پور، حسام؛ (۱۴۰۲). تشخیص و طبقه بندی اخبار جعلی به کمک پردازش زبان طبیعی و یادگیری عمیق، *ششمین کنفرانس ملی فناوری های نوین در مهندسی برق و کامپیوتر*،https://civilica.com/doc/1876620

ایزی، جلال و رضایی، پوریا؛ (۱۴۰۲). بهبود تشخیص در طبقه بندی اخبار جعلی در رسانه های اجتماعی با استفاده از الگوریتم کا – نزدیکترین همسایه، *نهمین کنفرانس بین المللی تحقیقات بین رشته ای در مهندسی برق، کامپیوتر، مکانیک و مکاترونیک در ایران و جهان اسلام*،https://civilica.com/doc/1994927

شفیعی وشاره، زهرا؛ علی عسگری رنانی، فاطمه و محمدی، شهرام؛ (۱۴۰۳). تشخیص اخبار جعلی به کمک الگوریتم های هوش مصنوعی، *هشتمین کنفرانس ملی پژوهشهای کاربردی در مهندسی برق، مکانیک و مکاترونیک*، https://civilica.com/doc/2024063

فرضی، سعید؛ (۱۴۰۰). استفاده از شبکه مولد متخاصم شرطی برای تولید داده با هدف بهبود کلاس بندی کاربران منتشرکننده اخبار جعلی، *دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره: ۱۳، شماره: ۴۷* .https://civilica.com/doc/1858925

وظیفه آبان، هادی و حسنی آهنگر، محمدرضا؛ (۱۴۰۲). استفاده از تکنیک رای‌گیری اکثریت در طبقه‌بندی اخبار جعلی از واقعی با الگوریتم‌های درخت تصمیم، رگرسیون لجستیک و کا–نزدیک‌ترین همسایگی، *ششمین همایش ملی فناوریهای نوین در مهندسی برق، کامپیوتر و مکانیک ایران*، https://civilica.com/doc/1744364

References

Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. 2012 *IEEE Symposium on security and Privacy* .DOI: 10.1109/SP.2012.34

Ahmed, H. (2017). *Detecting opinion spam and fake news using n-gram analysis and semantic similarity*

Ansari, V., Moomenzadeh, H., & Arfaeinia, H. (1403). Fake News Detection Using Deep Neural Network CNN, *7th International Conference on Electrical Engineering, Computer Science, Mechanics, and Artificial Intelligence*, (in Persian) https://civilica.com/doc/2046572.

Banerjee, R., Feng, S., Kang, J. S., & Choi, Y. (2014). Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*

Bigne, E., Andreu, L., Hernandez, B., & Ruiz, C. (2018). The impact of social media and offline influences on consumer behavior. An analysis of the low-cost airline industry. *Current Issues in Tourism*, *21*(9), 1014-1032 .https://doi.org/10.1080/13683500.2015.1126236

Castillo, C., Mendoza, M & ,.Poblete, B. (2011). Information credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web*. https://doi.org/10.1145/1963405.1963500

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on dependable and secure computing*, *9*(6), 811-824 . DOI: 10.1109/TDSC.2012.75

Di, R., Wang, H., Fang, Y., & Zhou, Y. (2018). Fake comment detection based on time series and density peaks clustering. Algorithms and Architectures for Parallel Processing: ICA3PP 2018 I*nternational Workshops, Guangzhou, China,* November 15-17, 2018, Proceedings 18

Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., & Zha, H. (2017). Fake news mitigation via point process based intervention. *International conference on machine learning*

Farokhian, M., Rafe, V., & Veisi, H. (2022). Fake news detection using parallel BERT deep neural networks. *arXiv preprint arXiv:2204.04793* .

Farzi, S. (2021). Utilizing Conditional Generative Adversarial Networks for Data Generation to Improve the Classification of Users Spreading Fake News, *Iranian Journal of Information and Communication Technology, Volume: 13, Issue: 47*. (in Persian). https://civilica.com/doc/1858925

Horne, B., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *Proceedings of the international AAAI conference on web and social media*. https://doi.org/10.1609/icwsm.v11i1.14976

Izi, J., & Hassanpour, H. (2023). Detection and Classification of Fake News Using Natural Language Processing and Deep Learning, *Sixth National Conference on New Technologies in Electrical Engineering and Computer* (in Persian) Science, https://civilica.com/doc/1876620.

Izi, J., Rezaei, P. (1402). Improving Detection in Fake News Classification on Social Media Using K-Nearest Neighbors Algorithm, *9th International Conference on Interdisciplinary Research in Electrical Engineering, Computer Science, Mechanics, and Mechatronics in Iran and the Islamic World*, (in Persian), https://civilica.com/doc/1994927.

Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, *80*(8), 11765-11788 .

Kuntur, S., Wróblewska, A., Paprzycki, M., & Ganzha, M. (2024). Fake News Detection: It's All in the Data! *arXiv preprint arXiv:2407.02122* .

Mukherjee ,A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. *Proceedings of the 21st International Conference on World Wide Web*. https://doi.org/10.1145/2187836.2187863

Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770* .

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, *54*(1), 547-577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Russell, M. A. (2013). *Mining the social web: data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more*. " O'Reilly Media, Inc .".

Shafiei-Shara, Z., Ali-Asgari-Renani, F., & Mohammadi, S. (1403). Fake News Detection Using Artificial Intelligence Algorithms. *8th National Conference on Applied Research in Electrical Engineering, Mechanics, and Mechatronics, Tehran*, (in Persian) https://civilica.com/doc/2024063.

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web*. https://doi.org/10.1145/2872518.2890098

Shu, K., Bernard, H. R., & Liu, H. (2019). Studying fake news via network analysis: detection and mitigation. *Emerging research challenges and opportunities in computational social network analysis and mining*, 43-65 .

Shushkevich, E., Alexandrov, M., & Cardiff, J. (2023). Improving multiclass classification of fake news using Bert-based models and CHATGPT-augmented data. *Inventions*, *8*(5), 112. https://doi.org/10.3390/inventions8050112

Su, J., Cardie, C., & Nakov, P. (2023). Adapting fake news detection to the era of large language models. *arXiv preprint arXiv:2311.04917* .

Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. (2017). Detecting fake news in social networks via crowdsourcing. *arXiv preprint arXiv:1711.09025*.

Vazifeh Aban, H., & Hasani Ahangar, M. R.; (1402). Using Majority Voting Technique in Classifying Fake News from Real News with Decision Tree, Logistic Regression, and K-Nearest Neighbors Algorithms, *6th National Conference on New Technologies in Electrical Engineering, Computer Science, and Mechanics of Iran*, Tehran, (in Persian) https://civilica.com/doc/1744364.

Vicario, M. D., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, *13*(2), 1-22. https://doi.org/10.1145/3316809

Vyas, P., Liu, J., & Xu, S. (2024). *Real-Time Fake News Detection on the X (Twitter): An Online Machine Learning Approach* .

Wang, Y., & Li, B. (2015). Sentiment analysis for social media images. 2015 *IEEE International Conference on Data Mining Workshop* (ICDMW). DOI: 10.1109/ICDMW.2015.142

Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on sina weibo. *Proceedings of the ACM SIGKDD workshop on mining data semantics*. https://doi.org/10.1145/2350190.2350203

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025 .

Zhao, Z., Resnick, P., & Mei, Q. (2015). Enquiring minds: Early detection of rumors in social media from enquiry posts. *Proceedings of the 24th International Conference on world wide web*.

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1-40. https://doi.org/10.1145/3395046

Zhu, Y., Li, Y., Wang, J., Gao, M., & Wei, J. (2024). FaKnow: A Unified Library for Fake News Detection. *arXiv preprint arXiv:2401.16441* .