Fake News Detection

Deepti Gururaj Baragi, Royal Pathak, Rohan Raut, Srabanti Guha

1 Introduction

Fake news refers to misleading or false information presented as real news to deceive people deliberately. Fake news can come in many forms, like fake user-generated content, enhanced image by photoshop, satire or parody, clickbait, etc. In this era of the internet, the increasing growth of social networks like Facebook, Twitter, etc., have become familiar sources of information. The unrestricted freedom of sharing any news and ease of publication on the Internet lead to an abundance of opportunities for cybercriminals to spread the fake word quickly. Besides that, people's age, orientation, social usage (Ma et al., 2014), perceived preference of news, self-perceptions of opinion leadership (Guess et al., 2019) etc., influence users to be an object of spreading fake news themselves. Moreover, peoples' trust in the news, which has more shares, likes, views, etc., contributes to believing fake sensational, emotional news more than real ones. The eye catchy visual representation of news, like attached audio, image, embedded content, and links of social media posts, also plays an essential role in increasing people's belief in fake news.

Fake news negatively affects many fields of our society, like public health, election, emergency management, and response. It can easily mislead people, affect their decision and change their perspectives. In times of election, fake news can harm the reputation of any individual, organization, or political party and can affect election outcomes (2016 US election). Besides that, during the pandemic, it can spread discredited health information like in covid; the Novax movement influenced people's decision to receive vaccines and hampered the fight against the global crisis. So, it's essential to have a reliable news ecosystem in our society, and that's why we need to detect fake news in its early stage and mitigate it before it reaches and spreads among people.

Detecting fake news is always challenging because fact-checking each piece of information is only sometimes possible. Fake news is purposefully fabricated with appealing emotions and sensational words to attract the attention of familiar people. Hence, extracting relevant information that distinguishes fake news from real and using them in algorithmic approaches is an effective way to detect fake news than using human intervention. Machine learning models, graph networks, and deep network approaches are already used to combat the fake news problem. Previous works use user-news-publisher relation, writing style, propagation pattern, source credibility, and various content levels (Shu et al., 2019), (Kumar and Shah, 2018), (Zhou and Zafarani, 2020), (Zhou et al., 2020a) as features for detecting fake news. In our work, we propose to use content information as our features and use them in various supervised machine learning and compare their performance. We focus on readability, writing style (average number of words, words characters, lower case characters, etc.), and emotion features (anger, anticipation, fear, etc.) of news content to identify fake news from real ones. Readability analysis considers various linguistic levels like the number of complex words, long words, and text cohesion while computing features. It helps us to distinguish between the fact that content is generated by professional journals or just a rookie to attract people. The writing style also plays an essential role in detecting fake news. For instance, while we have more exclamation marks in the content, there is a possibility that the news is fake because cyber criminals try to make that news sensational with more exclamation. On the other hand, more quotations can be a symbol of accurate, trustworthy news. Fake news can also include more uppercase URLs to gain attention. The ugly truth of fake news is they always try to attach emotion to news to attract people to consume, like, or share the news and spread them through it. That's why we propose to include readability, writing style, and emotion features from news content in our work.

1.1 Problem description

The accomplishment of our work will help in the following tasks:

- Analysis of news content and extract appropriate features from them.
- Identification of highly performed supervised learning models in the detection of fake news.

In brief, our work will help to solve the problem of detecting fake news using a variety of content features and provide better understanding of the role of those features together in the detection of fake news for different supervised learning techniques.

2 Related Work

Few studies address fake news detection. (Kumar and Shah, 2018) proposed a survey comprising various aspects of false information. The survey consisted of the mechanisms, reasons, impact, and characteristics in spreading false information. They developed algorithms based on feature, graph, and propagation modeling for effectively detecting false information.

(Shu et al., 2019) proposed a framework-'TriFN' modeling relationships among publishers, news contents, and social engagements for fake news detection. An experiment was conducted using fake news datasets, which demonstrated the proposed framework's effectiveness and the importance of tri-relationship for fake news prediction. The study by (Monti et al., 2019) suggested a fake news detection model based on geometric deep learning. The model was trained and tested on news stories verified by professional fact-checking organizations. The results indicated that the features such as social network structure and propagation showed high accuracy in fake news detection. It also pointed out that propagation-based approaches can be complementary to content-based approaches.

(Zhou et al., 2020a) proposed a theory-driven model for fake news detection. The model investigated news content considering various levels categorized based on social and forensic psychology. The detection of fake news was conducted using a supervised machine-learning framework.

(Yang et al., 2019) proposed investigating fake news by using an unsupervised learning framework. The variables used in the experiment were truths of news and the credibility of users. They used a Bayesian network model to find the conditional dependencies among the truths of the news, opinions, and credibility of users. They proposed an efficient collapsed Gibbs sampling approach to infer news truths and users' credibility without any labeled data.

Some of the research implements a bidirectional training approach for fake news detection. (Kaliyar et al., 2021) proposed FakeBert, a BERT-based (Bidirectional Encoder Representations from Transformers) deep learning approach to detect fake news. They combined parallel blocks of the single-layer deep Convolutional Neural Network (CNN) having different kernel sizes and filters with the BERT. They used various fake and real news articles during the 2016 U.S. General Presidential Election as a dataset. They evaluated the model with parameters such as Accuracy, FPR, FNR, and cross-entropy loss.

The paper by (Shrestha et al., 2020) discusses the problem of detecting fake news spreaders on Twitter. They proposed a binary classification task on the PAN at CLEF 2020 shared task on profiling fake news spreaders in two languages: English and Spanish. They used features computed from a set of tweets by users' such as writing style, word and char n-grams, BERT semantic embedding, and sentiment analysis. In addition, they investigated the role of psycho-linguistic (LIWC) and personality features in detecting fake news spreaders. The results showed that personality features do have a significant impact on user-sharing behavior.

The paper focusing on the ability of people to determine the reliability of news based on varying combinations of information by (Spezzano et al., 2021) showed that the automated techniques were more accurate than human samples. They used information such as news title, image, source bias, and excerpt that impact accuracy of users in identifying real and fake news and compared human performance to automated detection based on the information used.

3 Dataset

The ReCOVery (Zhou et al., 2020b) dataset was created to make studying and combating COVID-19 information easier. Following a search and investigation of almost 2,000 news outlets, 60 were found to have extremely high or low levels of credibility. A total of 2,029 news pieces about coronavirus, published from January to May 2020, are gathered in the archive, along with 140,820 ts that show how these news articles have spread on the Twitter social network, inheriting the authority of the medium on which they were published. The repository offers textual, visual, temporal, and network information in its collection of news items about the coronavirus. It is possible to compromise dataset scalability and label accuracy when determining news trustworthiness. The news data includes the news id, the article's URL, the publisher, the publish date, the author's name, the article's title, the article's image, the article's body content, the political bias, the nation of the news source, and the ground truth. The article's reliability provides the ground truth, where 1 indicates reliability and 0 shows unreliable. Similarly, each tweet's news id and tweet id are included in social media data. The only data we are using for this project is news text data.

4 Methodology/Feature engineering

The features we posit to add for the fake news detection is described below:

4.1 Readability

Readability measures the complexity of the textual content, which gets computed from t content written by the user (ts in our case). It also represents which stage of textual content complexity a person can recognize. In order to determine that, we used popular readability measures in our analysis:

- Flesh Reading Ease
- Flesh Kincaid Grade Level
- Coleman Liau Index
- Gunning Fog Index
- Simple Measure of Gobbledygook Index (SMOG)
- Automatic Readability Index (ARI)
- Lycee International Xavier Index (LIX)
- Dale-chall Score

The Flesch scale ranges from 0 to 100. Higher Flesch reading-ease ratings suggest that the text is easier to read, while lower scores indicate that it is more difficult to read. The Coleman Liau Index gauges the text's readability based on the word's characters. The Gunning Fog Index, the Flesh Kincaid Grade Level, the SMOG Index, the Automatic Readability Index, and the Gunning Fog Grade Level are algorithmic heuristics used to determine readability based on the number of educational years needed to comprehend the text. Lastly, the Dale-Chall reading test gauges the text's difficulty using a vocabulary list that fourth-graders are familiar with. We use this group of eight readability features to measure the complexity of a user's writing style.

4.2 Writing Style

This set of features captures the writing style of the news item authored by the same user. Specifically, we computed the average number of words, the average number of uppercased words, the the average number of characters per user news item, the percentage of stop-words, and the use of part of speech such as the number of nouns, proper nouns, personal nouns, possessive nouns, pronouns, determinants, adverbs, interjections, verbs, and adjectives.

4.3 Emotion-based features

Fake news is purposely sparked by emotionally charged statements to sway public opinion and damage feelings of a particular group by inciting their anger, fear, and mistrust in the direction of the event, individual, and agency. Thus, we calculated emotional traits such as anger, happiness, sadness, fear, disgust, anticipation, surprise, and trust. We computed emotion features including anger, pleasure, disappointment, fear, disgust, anticipation, marvel, and trust. We used the Emotion Intensity Lexicon (NRC-EIL) given in Mohammad (2018) and happy, unhappy, angry, do not care, inspired, afraid, amused, and aggravated using [1]Emolex.

Firstly, we cleaned up the news content by expanding contraction phrases, using LanguageTool 6 to fix spelling and grammar errors, swapping out negated terms for their WordNet antonym, removing forestall phrases, and lemmatizing the words.

[1] https://sites.google.com/site/emolexdata/

Following that, we calculated function vectors using the methods suggested by Milton et al. (2020) and Milton and Pera (2020). We specifically looked up each word in the two emotion dictionaries and mapped the corresponding affect values of the words that matched. To create an emotion vector, we then normalized the ranks of each emotion class using the full range of emotions that were retrieved from a news text. If two lexicons had the same emotion, such as sad in NRC-EIL and unhappiness in Emolex, we took the average of the two computed values into consideration.

4.4 BERT Embeddings

Bert Embeddings: BERT is derived from Vaswani et al. (2017). The semantic meaning of the news text is significantly preserved by the word- and sentence-level semantics. To embed the representation of each news, we employed the BERT based uncased text embedding model (Devlin et al., 2019). A neural network classifier with one feed-forward layer receives the embedding derived from BERT as input. As it produces better classification results, we fine-tuned the BERT classification model (Devlin et al., 2019).

Bert Model Training: We trained one-layer feed-forward neural network classifiers using linguistic features. We used a grid search to find the ideal parameters. To achieve the best classification results, the neural network's batch size, number of epochs, and learning rate taken are 8, 3, and 1e 5, respectively.

Finally, the BERT model successfully classifies news with high recall and precision. The model is also less dependent on the presence of specific phrases, whose popularity may change over time, thanks to the BERT classifier.

5 Experiments

We used a binary classification task to automatically identify whether a given piece of news is real or fake. Specifically, we used the features described in Section 4 which includes readability, writing style, and emotion features as input to various machine learning algorithms, namely Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, and Extra Trees Classifier, and selected the best performing classifier as our proposed model. We found that CatBoost is the best performing model.

Classifiers	AUROC	AvgP.
Logistic Regression	0.79	0.861
SVC	0.58	0.711
Random Forest	0.77	0.854
Extra Trees	0.81	0.891
CatBoost	**0.83**	**0.904**
XGBoost	0.82	0.900

Table 1: Textual Features: The performance of our textual features according to different classifiers on fake news classification and comparison with various supervised techniques. The best values are in bold.

Classifiers	AUROC	AvgP.
Logistic Regression	0.86	0.909
SVC	0.73	0.801
Random Forest	0.82	0.895
Extra Trees	0.85	0.912
CatBoost	**0.88**	**0.935**
XGBoost	0.88	0.931

Table 2: BERT embeddings: The performance of our extracted BERT embeddings according to different classifiers on fake news classification and comparison with various supervised techniques. The best values are in bold.

Using those three types of textual features separately in CatBoost model, we see that model with emotion features has AUROC (0.772) and precision (0.866) which is higher than other models with readability or writing based features. For models with readability and writing features, AUROC are 0.719 and 0.731 respectively and precision are 0.822 and 0.834 respectively.

Furthermore, We extend our work by extracting the BERT layers and combine with textual (readability, writing, emotion) features to improve our result by using it with proposed features. As our data is unbalanced, we used class weighting to deal with it and performed 5-fold cross-validation. We considered the Area Under the ROC Curve (AUROC) and Average Precision (AvgP) as evaluation metrics, which are well-suited for unbalanced data.

We also check feature importance by using Random Forest as in Figure 1. Our result illustrates average feature importance on five cross-validations. Our result also shows that the top 2 features are readability features (Dale-chall, Coleman Liau Index), one is writing features (number of verb past

Classifiers	AUROC	AvgP.
Logistic Regression	0.86	0.905
SVC	0.70	0.783
Random Forest	0.82	0.897
Extra Trees	0.85	0.911
CatBoost	**0.88**	**0.933**
XGBoost	0.88	0.932

Table 3: BERT embeddings + Textual Features: The performance of our textual features merged with BERT embeddings according to different classifiers on fake news classification and comparison with various supervised techniques. The best values are in bold.

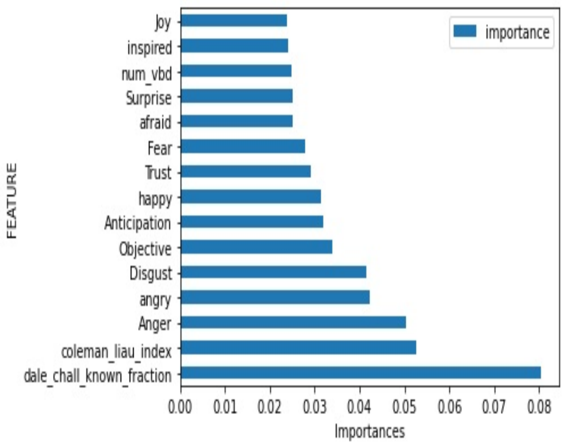

Figure 1: Top 15 Feature Importance based on average of feature importance on five cross-validations.

tense) and rest of them are emotion features.The work from (Shrestha and Spezzano, 2022) has also considered the same features to characterize and predict fake news spreaders in social networks on PAN 2020 dataset. Instead, we consider the same features, readability, emotions, and writing style to predict the fake news in the ReCOVery dataset with high AUROC and Average Precision.

6 Conclusions

According to a survey among US adults conducted by PEW research center[2], 59% of twitter users regularly get their day to day news from twitter, 54% of facebook users regularly get their day to day news from facebook. Similar statistics are seen in other social media like reddit, youtube, and tiktok. The consumption of news from social media is increasing in a rapid manner. However, social media has been used to spread fake news which creates confusion about the basic facts of day-to-day current events. The same PEW Research center reports 64% of americans believe that fake news is creating confusion about basic facts of current events.This has impacted health sector especially during the time of COVID-19, election outcomes, emergency management and response, and decline of trust in insitutions. Hence, the computational analysis of news with its readability, writing style, and emotional features are must to mitigate the fake news spread in social media.

Our classification results are reported in Table 3 according to AUROC and average precision. As the table shows, among all the considred classifiers, XGBoost achieved the best results with an AUROC of 0.8280 and average precision of 0.8982 for fake news classification. Further, Using supervised learning techniques with proposed features helps to consistently classify fake news as readability, writing style, and emotions are the important indicators for fake news which is not feasible by human expertise.

[2] https://pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

7 Acknowledgements

The authors would like to appreciate Dr.Spezzano for providing the opportunity to leverage the application of the supervised machine learning algorithms in fake news detection with a service learning initiative to address teenagers' vulnerability to web misinformation.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788.

Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*.

Long Ma, Chei Sian Lee, and Dion Hoe-Lian Goh. 2014. Understanding news sharing in social media: An explanation from the diffusion of innovations theory. *Online information review*, 38(5):598–615.

Ashlee Milton, Levesson Batista, Garrett Allen, Siqi Gao, Yiu-Kai D Ng, and Maria Soledad Pera. 2020. “don’t judge a book by its cover”: Exploring book traits children favor. In *Fourteenth ACM Conference on Recommender Systems*, pages 669–674.

Ashlee Milton and Maria Soledad Pera. 2020. What snippets feel: Depression, search, and snippets.

Saif Mohammad. 2018. Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*.

Anu Shrestha and Francesca Spezzano. 2022. Characterizing and predicting fake news spreaders in social networks. *Int. J. Data Sci. Anal.*, 13(4):385–398.

Anu Shrestha, Francesca Spezzano, and Abishai Joy. 2020. Detecting fake news spreaders in social networks via linguistic and personality features. In *CLEF*.

Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 312–320.

FRANCESCA Spezzano, A Shrestha, JERRY ALAN Fails, and BRIAN W Stone. 2021. That’s fake news! investigating how readers identify the reliability of news when provided title, image, source bias, and full articles. *Proceedings of the ACM on Human Computer Interaction journal*, 5.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5644–5651.

Xinyi Zhou, Atishay Jain, Vir V Phoha, and Reza Zafarani. 2020a. Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, 1(2):1–25.

Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020b. Recovery: A multimodal repository for covid-19 news credibility research. In *Proceedings of the 29th ACM International Conference on Information Knowledge Management*, pages 3205–3212.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.