Leveraging New Data Features for Enhanced Accuracy of Fake News Detection

Anonymous EMNLP submission

Abstract

The prominence of social media poses a significant threat to information integrity as the spread of fake news increases. It becomes imperative for online media outlets to develop effective strategies to mitigate the spread of fake news. This paper introduces an enhanced version of existing benchmark datasets for fake news classification. We enhance an existing benchmark dataset for fake news classification by extracting additional linguistic and psychological features, along with user comment data. In this work, we perform a qualitative analysis of linguistic, psychological and user comments trends present in fake news. We demonstrate the efficacy of our curated dataset through rigorous evaluation, showcasing their performance against state-of-the-art benchmark datasets. The findings of this work demonstrate a significant performance improvement, up to 10%, when compared to state-of-the-art datasets.

1 Introduction

In an era where social media has popularized the digital dissemination of information, the spread of false information presents a substantial threat to maintaining the trust and integrity of information in modern-day society. As fake news continues to become increasingly more prominent, it becomes crucial for social media platforms to develop preventative measures to effectively recognize and mitigate fake news on their platforms.

The focus of our dataset is to improve existing fake news classification with the use of the textual components of a news article and its related user engagement. Previous studies show that user engagement and news propagation provide valuable information for the classification of fake news. However, identifying the propagation graph or news cascade can be time-consuming, and this type of information may not be available if we want to detect fake news in the early stages. In this work, we choose the easier and more efficient method to include the propagation data - by including user comments. This paper presents a novel fake news detection dataset leveraging a comprehensive range of linguistic, psychological, and user comment features for robust classification. By incorporating linguistic and psychological patterns in the writing of both news articles and user comments, we aim to enhance the accuracy and effectiveness of existing state-of-the-art fake news detection models.

Psychological characteristics have noticeable effects when distinguishing between real and fake news. In this research, we classify these traits into three categories: emotions, swear words, and social behaviour. An in-depth analysis of these feature groups indicates that social behavioural traits, such as politeness, interpersonal conflict, moralization, pro-social behaviour and communication (Boyd et al., 2022), have elevated effects in fake and real news. Sentiment analysis is a growing field in fake news research. Studies (Zhang et al., 2021; Guo et al., 2019) have shown that fake news often exhibits emotional bias and profanity, while truthful articles are neutrally just. Postive and negative emotions, as well as tones of anxiety, sadness, and anger, are common traits found in fake news and are beneficial when incorporated with fake news classification. The inclusion of these features in fake news detection allows models to be trained to identify these trends for the accurate classification of fake news. Our research results indicate that including these psychological feature groups improves the performance of fake news detection, surpassing the performance of baseline models.

Additionally, recent research (Liu and Wu, 2018, 2020; Shu et al., 2019) in fake news mitigation has explored the inclusion of user propagation and comment data as part of the models' classification. Studies by Shu et al. (Zhang et al., 2021; Guo et al., 2019) have determined that various sentiment-based characteristics can be extracted from user comments to mitigate fake news. As part of our research, we explore the differences between user comments in fake and real news and incorporate user comment data into our proposed dataset.

The main contribution of this paper is the development and validation of a fake news dataset, which involves the addition of new features into existing dataset(s), significantly improving the accuracy and reliability of identifying misinformation by 10% in comparison to existing datasets. Additionally, we conduct a comprehensive analysis of the differences in linguistic and psychological features between fake and real news.

In this paper, we discuss related work in fake news research in Section two. We provide a detailed description of the data in our dataset and processing steps in Section three. Section four addresses the new features incorporated into the proposed dataset. To determine the efficacy of our dataset, we conduct a series of experiments which are explained and analyzed in Section five. Finally, we provide concluding remarks and list this research's limitations and future directions in Section six.

2 Related work

The reliance on robust fake news datasets becomes imperative as fake news research continues to develop. Without a detailed dataset to train the classification model, the process of model training becomes challenging. Therefore, the availability of reliable fake news datasets becomes crucial to improve the development of fake news detection. In this section, we conduct a detailed literature review of current fake news datasets and examine each dataset's various characteristics. As fake news mitigation methods develop, new datasets continue to emerge to accommodate these new methods. We classify these datasets into three categories: text-based datasets, image-based datasets, and propagation-based datasets.

2.1 Text-based datasets

Most common fake news detection models (Raza and Ding, 2022; Horne et al., 2018; Przybyla, 2020; Silva et al., 2020; Potthast et al., 2017; Karimi et al., 2018; Zellers et al., 2020; Kaliyar et al., 2020) focus on a text-based approach for classification. To train such models, several text-based datasets have emerged to streamline the training process. Additionally, these datasets contain the largest sample size compared to other forms of datasets. Datasets such as NELA-GT (Gruppi et al., 2020), Credbank (Mitra and Gilbert, 2015), FakeNewsNet (Shu et al., 2020a), and Liar (Wang, 2017) are considered state-of-the-art datasets for text-based classification as they contain a large sample of articles collected from several sources of news media. Datasets such as FakeNewsNet provide highly accurate labelling methods (Hu et al., 2021; Baly et al., 2018), as they leverage manual fact-checking platforms such as PolitiFact and GossipCop to validate the veracity of each news article. Previous research, such as Verma et al. (Verma et al., 2021), have explored incorporating pre-processed feature sets in their WELfake dataset to enhance classification accuracy and runtime. The WELfake dataset combines news articles from existing datasets to create a diverse and unbiased dataset for fake news classification. Their approach involves analyzing linguistic patterns like sentence length, readability, and subjectivity to assign a binary label to each article. Additionally, they provide a comprehensive list of linguistic features, such as readability, writing patterns, and psycho-linguistic features, to improve the effectiveness of existing fake news classification models. For our dataset, we utilize news articles and labelling inherited from the NELA-GT-2019 dataset.

2.2 Image-based datasets

Recent studies in fake news detection (Zhou et al., 2020b; Wu et al., 2021) have experimented with multimodal classification and have found promising results in the models' performance. Research has shown that images in news articles provide important details about the veracity of news articles and can be used to generate misinformation through techniques such as clickbait. Thus, research in multimodal datasets has become more prevalent in recent years. Datasets such as the Image Manipulation dataset (Christlein et al., 2012) and PS-Battles dataset (Heller et al., 2018) provide an expansive sample of manually fetched articles which contain manipulative images used for fake news. Classification models can leverage these datasets to train models on image-based classification for effective fake news mitigation.

2.3 Propagation-based datasets

More recent work (Huang et al., 2020; Jiang et al., 2019; Horne et al., 2019; Jwa et al., 2019; Nguyen

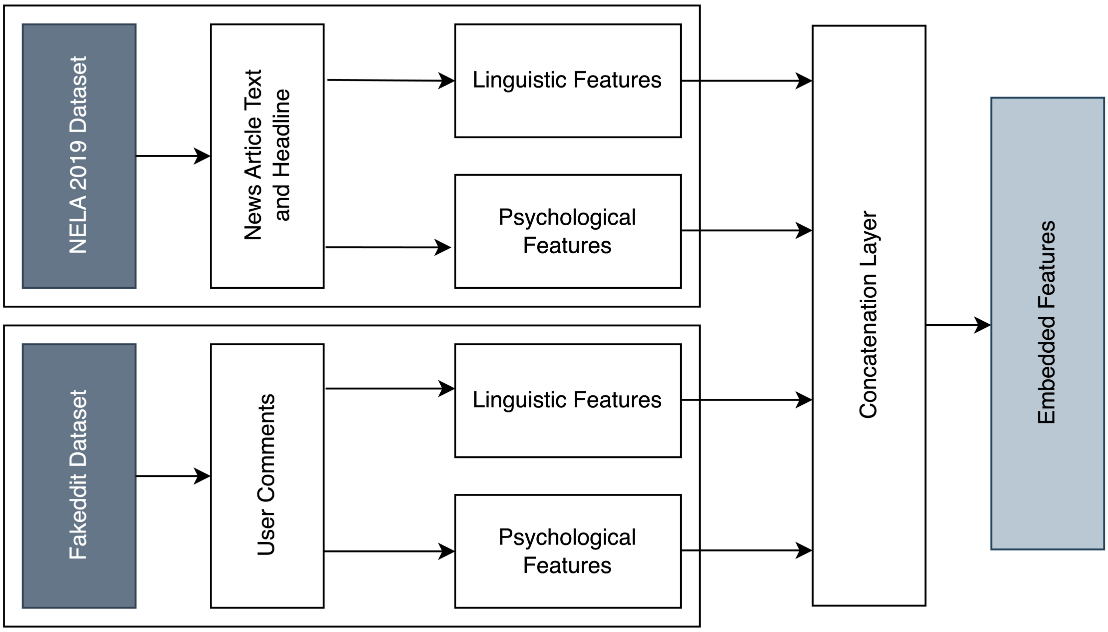


Figure 1: Overview of proposed dataset

et al., 2020; Qian et al., 2018; Cheng et al., 2021; Nakamura et al., 2019) in fake news has explored methods of fake news detection outside of the traditional text and image-based classification methods. Datasets such as Credbank and FakeNewsNet are becoming increasingly popular as they provide a comprehensive list of propagation-related features for fake news detection. User propagation data provides valuable insights into the likelihood for which different users spread fake news. This information is vital in social media-based classification, as studies (Schmitt and Spinosa, 2022) have shown that individuals with a high risk of spreading or interacting with fake news tend to propagate fake news more frequently. By limiting the spread of news by high-risk users early, the spread of fake news can be greatly reduced.

2.4 Multi-labeling

Modern research in fake news (Horne and Adali, 2017) has expanded from the binary classification problem of fake news detection to explain the type of fake news each article imposes. Datasets such as Liar, BuzzFeed, FEVER (Thorne et al., 2018), Fakeddit (Nakamura et al., 2019), and Credbank have developed their datasets to incorporate three-, four-, or six-way labelling methods. Accurate categorization of fake news into types such as satire, manipulation, and false image connection offers significant advantages in psychological-based fake news classification. This form of categorization greatly assists detection models to better understand psychological cues within fake news to improve the classification accuracy.

Finally, datasets such as Fakeddit incorporate multiple categories to create a robust dataset. Unlike other datasets, Fakeddit provides details about user engagement through user comments and user IDs, which can then be expanded to incorporate user profiles. Secondly, the dataset includes images from all collected Reddit posts for effective multimodal classification. Thirdly, user comment data is extracted from the Reddit posts to aid in text-based classification. In this work, we incorporate the user comment data from the Fakeddit dataset as a component of our dataset.

3 Data description

Many state-of-the-art news datasets (Gruppi et al., 2020; Nakamura et al., 2019; Shu et al., 2020a) source only the article's text, propagation data, or user engagement; however, the combination of multiple sources of information provides valuable benefits for fake news classification.

In this work, we merge the text-based content from NELA-GT-2019 with the user comments from Fakeddit to create a unified dataset to incorporate news articles with user comments. In addition, we make significant contributions to fake news research by incorporating linguistic and psychological features extracted from the unified dataset to improve the efficacy of fake news detection.

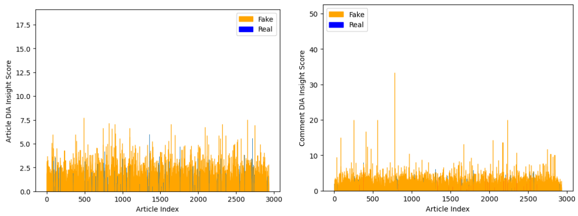
3.1 Data pre-processing

The NELA dataset consists of news articles collected from different news sources throughout 2019. Of these news sources, the dataset contains the news headline, the article's text, and metadata related to the time and source of publication. Fakeddit sources the dataset over a ten-year span of Reddit user comments from several selectively chosen subreddits. In this dataset, the user comments in the chosen subreddits, the publication timestamps and metadata related to each comment are extracted. The original news article titles and Reddit post headlines are matched to concatenate the two datasets. After concatenation, we extract the article headline, article text, user comments, date and time of publication, and the binary classification label from NELA as features to create the pre-processed dataset.

In our dataset, we leverage the news articles and user comments of the processed datasets to extract the new features prosed in this paper, which include psychological and linguistic characteristics of the textual content. Recent studies, including the work from Shu et al. (Zhou and Zafarani, 2020b; Zhang et al., 2021; Guo et al., 2019; Shu et al., 2020b, 2019, 2021), indicate linguistic and psychological features provide valuable information for the classification task. Our dataset aims to improve on existing fake news datasets by including these features as part of the dataset.

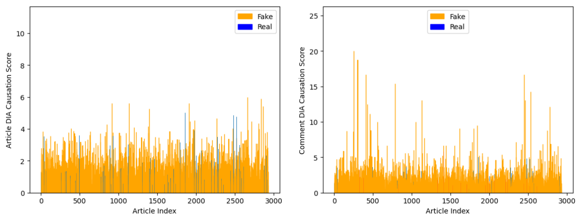
3.2 Dataset summary

After pre-processing, our collective dataset consists of 2,929 news articles and 41,867 user comments collected from the year 2019. Of these articles, 69.9% are labelled as fake news while the remaining 30.1% are real news. Our dataset contains the articles' headline, text, comments, classification label, and the metadata presented above. The articles' text and headlines are sourced directly from the NELA dataset. To obtain the full user comments, date-time stamps and upvotes, we reference the comment IDs from the Fakeddit dataset and use Reddit's API to obtain the respective metadata. For our dataset, we extract the first eight hours of user comment data after publication on Reddit. This information is then concatenated into the combined dataset.



(a) Article (b) Comments

Figure 2: DIA Insight

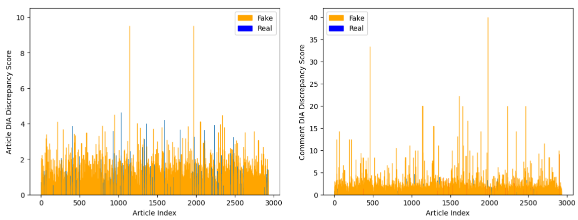


(a) Article (b) Comments

Figure 3: DIA Causation

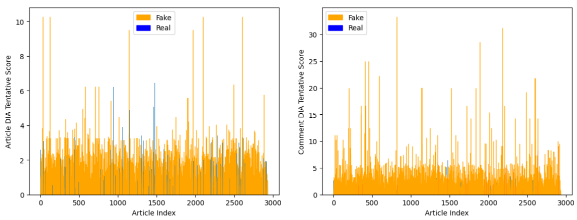
3.3 Labeling

We inherit the two-way labels from NELA dataset to classify all news articles in the dataset. Although Fakeddit provides two-, three-, and six-way labelling in their dataset, these labels are based on a full subreddit's credibility score rather than a per-comment basis. As a result, these labels are much broader compared to the news source level NELA uses for labelling. Thus, we utilize the labelling system from the NELA dataset for our dataset, which obtains labelling based on the credibility of the publishing news source. Our final dataset will contain zero to many user comments for each news article in the dataset. For each article, one binary label is provided for evaluation of the classification model.



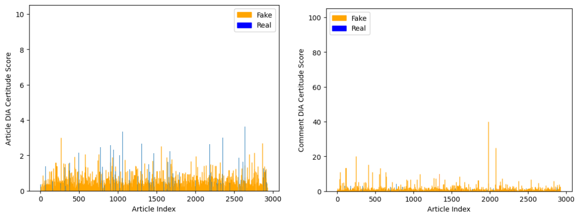
(a) Article (b) Comments

Figure 4: DIA Discrepancy



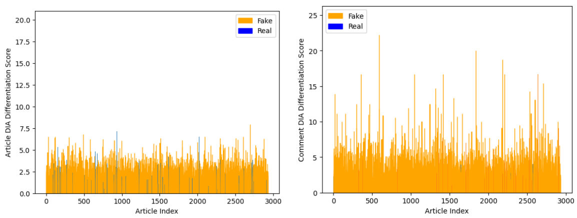
(a) Article (b) Comments

Figure 5: DIA Tentative



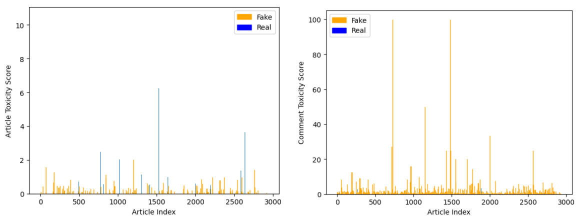
(a) Article (b) Comments

Figure 6: DIA Certitude



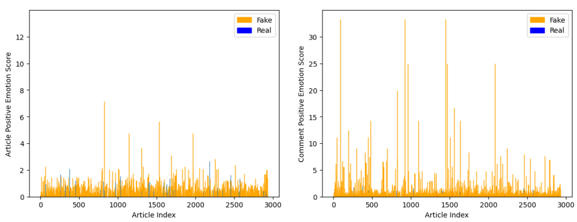
(a) Article (b) Comments

Figure 7: DIA Differentiation



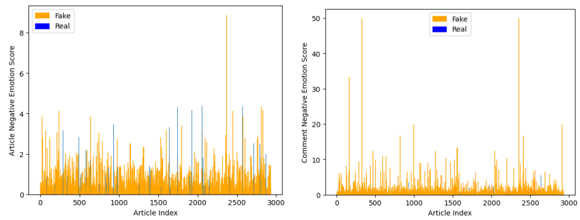
(a) Article (b) Comments

Figure 8: Toxicity



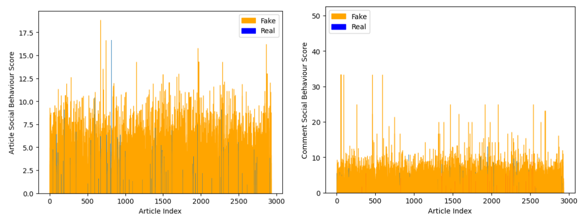
(a) Article (b) Comments

Figure 9: Positive Emotion



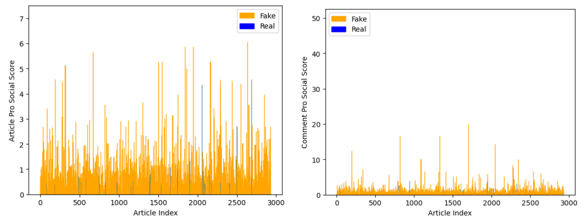
(a) Article (b) Comments

Figure 10: Negative Emotion



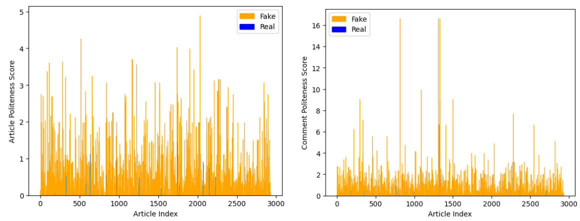
(a) Article (b) Comments

Figure 11: Social Behaviour



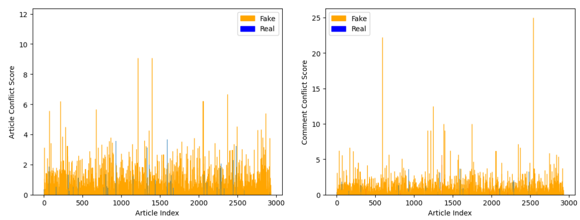
(a) Article (b) Comments

Figure 12: Pro Social



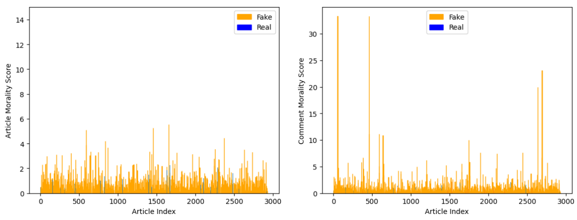
(a) Article (b) Comments

Figure 13: Politeness



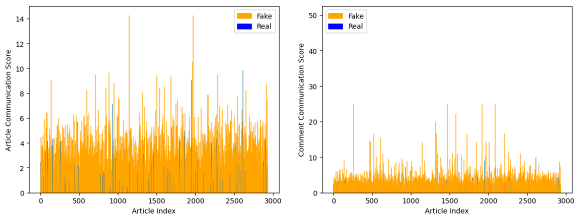
(a) Article (b) Comments

Figure 14: Conflict



(a) Article (b) Comments

Figure 15: Morality



(a) Article (b) Comments

Figure 16: Communication

4 Methodology

4.1 Problem definition

Our text-based fake news dataset is designed to aid in identifying fake news by analyzing the text within news articles and user comments. This work aims to determine the veracity of N news articles. We express the classification outcome as a binary value, $Y = (0, 1,)$, where zero indicates a truthful news article, and one signifies fake news. We define each article in N number of news articles as a set P=(P_h, P_c, P_c), where P_h represents the headline of an article, P_c represents the article's textual content, and P_u represents the related user comments.

4.2 Proposed features

Using the extracted textual features from the combined dataset, we attempt to improve existing fake news classification systems by extracting the linguistic and psychological features from input P to include in the dataset using natural language processing (NLP) and the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker et al., 2001).

4.2.1 Linguistic features

State-of-the-art fake news detection models (Zhou et al., 2020a,b) prioritize generating linguistic features from the article's text to perform the classification task. Many of these models utilize NLP tools to extract different linguistic features for classification. To improve the simplicity and runtime of these classification models, we extract different sets of linguistic features from the articles' text to be stored in our dataset. Linguistic features of our dataset include parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait-related attributes (CBA) and BERT embedding (Devlin et al., 2018). POS and CFG are commonly extracted in many classification models using NLP tools; thus, we include these features as part of our dataset. Recent classification models have explored newer embedding models for the article's text to aid in the classification task, including models such as BERT embedding (Szczepański et al., 2021; Farokhian et al., 2022; Vijjali et al., 2020; Liu et al., 2019; Devlin et al., 2018). We include the BERT embedding of the article's text, headline and user comments as part of the linguistic features in our dataset. Newer detection models (Zhou and Zafarani, 2020a) have explored the use of clickbait and disinformation-related attributes from the article's text to aid in fake news classification. Disinformation traits such as discrepancy, tentative, certitude, and differentiation as well as click-bait attributes such as "this will blow your mind" and "can change your life" have elevated values in fake news and provide valuable information when incorporated into detection models, as illustrated in figures 2 to 7. Thus, our dataset includes DIA and CBA attributes as linguistic features.

4.2.2 Psychological features

Recent studies in early fake news detection focus on different types of psychological features to improve the performance of existing classification models. A particular focus on sentiment analysis and emotions (Guo et al., 2019; Zhang et al., 2021) has shown promising results against existing state-of-the-art detection models when emotions are included in the classification task. These classification models can recognize psychological characteristics in the article's text to accurately detect fake news. Our dataset includes three psychological feature groups: emotions, swear words, and social behavioural traits. Using the articles' text content and headlines, as well as user comment data from the pre-processed dataset, we obtain the psychological features mentioned using the LIWC dictionary. LIWC utilizes TF-IDF weights to compare the frequency of words or phrases in the input text against the LIWC dictionary. LIWC then generates a numerical score out of one hundred for each psychological feature in the article's text and user comments. The resulting output indicates the percentage of terms in the input text included in the LIWC dictionary.

We perform an in-depth analysis of different psychological features' effects on the classification task against modern state-of-the-art classification models, as shown through figures 8 to 16. Social behavioural traits such as morality, politeness, interpersonal conflict, and pro-social behaviour show elevated values in fake and real news. Thus, we include these specific features as part of our dataset's social behaviour feature group. Traits such as morality and interpersonal conflict are more prevalent in fake news, as illustrated in figures 14 and 15. Traits such as pro-social behaviour, politeness and communication have elevated scores in real news, as illustrated in figures 12, 13, and 16. These trends in both fake and real news text provide valuable information for the classification model when differentiating between the two news types.

In addition, we extract the negative and positive emotions within the article's text and user comments to assist with the classification task. With negative emotional features, tones such as anxiety, anger, and sadness can be extracted from the text and used for classifying fake news. Recent work by Shu et al. (Guo et al., 2019; Zhang et al., 2021) have analyzed the trends of emotions and sentiment analysis in fake news research and identified that fake news is emotionally biased compared to real news. We see similar patterns in our analysis of the features in the figures 9 and 10. Our research in the psychological features identified interesting trends with word toxicity. As illustrated in figure 8, real news articles have elevated levels of word toxicity, such as swear words, compared to fake news. Alternatively, user comments with high levels of toxicity are often correlated with fake news. Our analysis indicates that the usage of these psychological features provides valuable information in the classification task and is a valuable contribution to improving the performance of existing detection models. By leveraging these trends, classification models can easily differentiate between fake and real news.

5 Running fake news detection models on our dataset - experiment and results

We perform in-depth analytical research to differentiate fake and real news through the use of both linguistic and psychological features.

The dataset is run against several state-of-the-art fake news classification models to determine its efficacy. Each experiment is executed on different classification models using nltk (Loper and Bird, 2002), keras (Ketkar, 2017) and scikit-learn (Pedregosa et al., 2011), on a computer with sixty-four gigabytes of RAM, a sixteen-core thirty-two thread processor and an RTX 2070 graphics card. For classification, seventy percent of the dataset is used for training, and the remaining thirty percent is used as the testing set. To reduce training loss, each model is trained over fifteen epoches and is optimized using the Adam optimizer (Loshchilov and Hutter, 2017). Each model has a dimensionality size of 768 for each linguistic and psychological embedding feature. The models use a ReLU activation layer, have a softmax output layer, and leverage the sparse categorical cross-entropy loss function.

To create a holistic analysis of our dataset's effectiveness, we chose a set of frequently used machine learning models, including multilayer perceptron models (MLP) (Ramchoun et al., 2016), XGBoost (Chen and Guestrin, 2016), convolutional neural networks (CNN) (O'Shea and Nash, 2015) and decision tree classifiers (DTC) (Swain and Hauska, 1977). Recent studies (Kaliyar et al., 2021; Szczepański et al., 2021) in fake news research have adapted BERT embedding models to improve the performance of fake news classification; thus, we include these two embedding models with the MLP classifier.

We consider the task of identifying fake news as a binary classification problem, where the resulting output is either a real (0) or fake (1) integer. To evaluate the effectiveness of our dataset against various classification models, we rely on standardized metrics (Shu et al., 2019; Vo and Lee, 2019) such as accuracy and F1 Score as evaluation metrics.

Our dataset provides the embedded values of each linguistic and psychological feature so it can be inputted into a classification model without any pre-processing steps. To train each model, we feed all embedded features as the input for the model's training from the training set and finally use the embeddings from the testing set to obtain the evaluation results from the model.

Table 1: Fake news detection accuracy on our dataset vs. original dataset

	Without Proposed Features		*With Proposed Features*		
Model	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Change in Accuracy**
BERT	0.86803	0.90779	0.97156	0.97949	+ 0.10353
DistilBERT	0.87144	0.90760	0.97156	0.97942	+ 0.10012
MLP	0.95677	0.96915	0.97497	0.98191	+ 0.01820
XGBoost	0.94539	0.96078	0.97497	0.98191	+ 0.02958
CNN	0.94880	0.96314	0.94994	0.96411	+ 0.00113
DTC	0.85552	0.89683	0.91695	0.94099	+ 0.06143

Proposed Features include User Comments, Linguistic and Psychological Features
Basic Features include Bag of Words embedding with Article Headline and Text Only

The conducted experiments provide valuable insight into the impact linguistic and psychological features have in fake news detection. Our data analysis in section four has illustrated that certain linguistic and psychological features have elevated scores in fake or real news. Classification models such as those in table 1 are trained using these elevated values to detect the trends during classification. The results indicate that our dataset improves the performance of existing models by up to ten percent when linguistic and psychological features are included as part of the model's training. With the inclusion of linguistic and particularly psychological features in our dataset, fake news classification models are able to predict fake news better than traditional text-based datasets.

6 Conclusion and future work

Due to limitations imposed by Reddit's API, our dataset only consists of the first eight hours of user comments for each news post. While sufficient for early fake news detection models, general text-based classification models are limited by the eight-hour timespan of the user comments. Future work of this dataset can entail extracting more user comment data using Reddit's API, providing a larger timespan of user comments for text-based classification. To label our dataset, we are reliant on NELA and Fakeddit's source-based labelling. Although effective, alternative labelling methods such as on a per-article basis can provide better training for classification models. Thus, leveraging manual fact-checking sources for each news article can improve the effectiveness of our dataset.

As social media becomes more prominent, it becomes imperative for social media outlets to develop effective mitigation methods to reduce the spread of fake news. Existing fake news datasets provide only the text content of news article; leaving classification models to extract various sets of linguistic and psychological features for fake news classification. However, this step increases the runtime and complexity of detection models. In this paper, we make contributions to the field of fake news research by: 1) performing an in-depth analysis of the effects of different linguistic and psychological features in both real and fake news; 2) creating a compiled dataset which includes the articles' text content with the related user comments to incorporate user propagation data with text-based classification; and 3) develop an improved fake news dataset which includes pre-embedded values of effective linguistic and psychological features to improve the efficacy of existing fake news classifiers. We execute a series of experiments to demonstrate the effectiveness of our dataset against state-of-the-art classification models and illustrate the benefits of linguistic and psychological features, and user comments in fake news research. The results of our experiments indicate the proposed dataset provides significant improvements to the performance of fake news classifiers.

7 Acknowledgments.

References

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. *arXiv preprint arXiv:1810.01765*.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*.

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Lu Cheng, Ruocheng Guo, Kai Shu, and Huan Liu. 2021. Causal understanding of fake news dissemination on social media. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 148–157.

Vincent Christlein, Christian Riess, Johannes Jordan, Corinna Riess, and Elli Angelopoulou. 2012. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on information forensics and security*, 7(6):1841–1854.

Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Mahmood Farokhian, Vahid Rafe, and Hadi Veisi. 2022. Fake news detection using parallel bert deep neural networks. *arXiv preprint arXiv:2204.04793*.

Maurício Gruppi, Benjamin D. Horne, and Sibel Adalı. 2020. Nela-gt-2019: A large multi-labelled news dataset for the study of misinformation in news articles.

Chuan Guo, Juan Cao, Xueyao Zhang, Kai Shu, and Miao Yu. 2019. Exploiting emotions for fake news detection on social media. *arXiv preprint arXiv:1903.01728*.

Silvan Heller, Luca Rossetto, and Heiko Schuldt. 2018. The ps-battles dataset-an image collection for image manipulation detection. *arXiv preprint arXiv:1804.04866*.

Benjamin D Horne and Sibel Adali. 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. In *Eleventh international AAAI conference on web and social media*.

Benjamin D Horne, William Dron, Sara Khedr, and Sibel Adali. 2018. Assessing the News Landscape: A Multi-Module Toolkit for Evaluating the Credibility of News. In *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*, pages 235–238.

Benjamin D Horne, Jeppe Nørregaard, and Sibel Adali. 2019. Robust fake news detection over time and attack. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(1):1–23.

Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763.

Qi Huang, Chuan Zhou, Jia Wu, Luchen Liu, and Bin Wang. 2020. Deep spatial–temporal structure learning for rumor detection on Twitter. *Neural Computing and Applications*, (August).

Shengyi Jiang, Xiaoting Chen, Liming Zhang, Sutong Chen, and Haonan Liu. 2019. User-characteristic enhanced model for fake news detection in social media. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 634–646. Springer.

Heejung Jwa, Dongsuk Oh, Kinam Park, Jang Mook Kang, and Heuiseok Lim. 2019. exBAKE: Automatic fake news detection model based on Bidirectional Encoder Representations from Transformers (BERT). *Applied Sciences (Switzerland)*, 9(19):4062.

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.

Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, and Soumendu Sinha. 2020. Fndnet–a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61:32–44.

Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*, pages 1546–1557.

Nikhil Ketkar. 2017. Introduction to keras. In *Deep learning with Python*, pages 97–111. Springer.

Chao Liu, Xinghua Wu, Min Yu, Gang Li, Jianguo Jiang, Weiqing Huang, and Xiang Lu. 2019. A Two-Stage Model Based on BERT for Short Fake News Detection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11776 LNAI:172–183.

Yang Liu and Yi Fang Brook Wu. 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 354–361.

Yang Liu and Yi Fang Brook Wu. 2020. FNED: A Deep Network for Fake News Early Detection on Social Media. *ACM Transactions on Information Systems*, 38(3).

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Tanushree Mitra and Eric Gilbert. 2015. Credbank: A large-scale social media corpus with associated credibility annotations. In *Proceedings of the international AAAI conference on web and social media*, volume 9, pages 258–267.

Kai Nakamura, Sharon Levy, and William Yang Wang. 2019. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. *arXiv preprint arXiv:1911.03854*.

Van Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min Yen Kan. 2020. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. *International Conference on Information and Knowledge Management, Proceedings*, pages 1165–1174.

Keiron O'Shea and Ryan Nash. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

Piotr Przybyla. 2020. Capturing the Style of Fake News. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):490–497.

Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. 2018. Neural user response generator: Fake news detection with collective user intelligence. In *IJCAI International Joint Conference on Artificial Intelligence*, volume 2018-July, pages 3834–3840.

Hassan Ramchoun, Youssef Ghanou, Mohamed Ettaouil, and Mohammed Amine Janati Idrissi. 2016. Multilayer perceptron: Architecture optimization and training.

Shaina Raza and Chen Ding. 2022. Fake news detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics*, 13(4):335–362.

Murilo FL Schmitt and Eduardo J Spinosa. 2022. Scalable stream-based recommendations with random walks on incremental graph of sequential interactions with implicit feedback. *User Modeling and User-Adapted Interaction*, pages 1–31.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020a. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020b. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data*, 8(3):171–188.

Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 9:312–320.

Kai Shu, Guoqing Zheng, Yichuan Li, Subhabrata Mukherjee, Ahmed Hassan Awadallah, Scott Ruston, and Huan Liu. 2021. Early detection of fake news with multi-source weak social supervision. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 650–666. Springer.

Renato M. Silva, Roney L.S. Santos, Tiago A. Almeida, and Thiago A.S. Pardo. 2020. Towards automatically filtering fake news in Portuguese. *Expert Systems with Applications*, 146:113199.

Philip H Swain and Hans Hauska. 1977. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3):142–147.

Mateusz Szczepański, Marek Pawlicki, Rafał Kozik, and Michał Choraś. 2021. New explainability method for bert-based model in fake news detection. *Scientific Reports*, 11(1):1–13.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. Fever: a large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355*.

Pawan Kumar Verma, Prateek Agrawal, Ivone Amorim, and Radu Prodan. 2021. Welfake: Word embedding over linguistic features for fake news detection. *IEEE Transactions on Computational Social Systems*, 8(4):881–893.

Rutvik Vijjali, Prathyush Potluri, Siddharth Kumar, and Sundeep Teki. 2020. Two stage transformer model for covid-19 fake news detection and fact checking. *arXiv preprint arXiv:2011.13253*.

Nguyen Vo and Kyumin Lee. 2019. Learning from fact-checkers: Analysis and generation of fact-checking language. *SIGIR 2019 - Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–344.

William Yang Wang. 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Yang Wu, Pengwei Zhan, Yunjian Zhang, Liming Wang, and Zhen Xu. 2021. Multimodal fusion with co-attention networks for fake news detection. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, pages 2560–2569.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2020. Defending against neural fake news. *Neurips*.

Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021. Mining dual emotion for fake news detection. In *Proceedings of the Web Conference 2021*, pages 3465–3476.

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, Jindi Wu, and Reza Zafarani. 2020b. SAFE: Similarity-Aware Multi-modal Fake News Detection. Technical report.

Xinyi Zhou and Reza Zafarani. 2020a. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. Technical Report 5.

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