Project Proposal for a Fake News Classifier

Karsin Dass
University of Michigan
kdass@umich.edu

Maya Ruder
University of Michigan
mayarud@umich.edu

Abstract

The rise of fake news poses a significant challenge in the digital era, impacting societal movements and political elections. This project aims to develop a machine learning-based classifier to detect and categorize fake news articles by leveraging various supervised learning techniques, including logistic regression, Naive Bayes, K-Nearest Neighbors (KNN), and neural networks. By training these models on diverse datasets, we aim to identify distinctive patterns and linguistic features of fabricated content. The classifier's performance will be measured using cross-validation and key metrics, including accuracy, precision, recall, and F1-score, to ensure practical applicability. This research seeks to advance current baselines, emphasizing the creation of a robust, generalizable model capable of effectively detecting fake news across various domains.

1 Project Description

For our project, we will replicate and adjust Fake News Classifiers that will perform on a variety of datasets. We will use many of the methods we have done in class such as logistic regression, LDA, QDA, naive bayes, and KNN. Moreover, we will also implement methods that we have yet to learn in class such as neural networks. For this, we will need to conduct our own independent study.

1.1 Datasets

Since Fake News is a well-studied topic, we will be obtaining all of our datasets through Kaggle. We want our classifier to work on a variety of datasets, therefore, we are using American News, Soccer News, and Albanian news datasets. All of the datasets have a vairety of real and fake news postings. Using a variety of datasets, we will be able to assess the performance of our classifier for real-world problems. Some challenges we may encounter are ensuring that we develop code that works on all of the datasets.

1.2 Baselines

Some of our datasets and literature contain methods of analysis that we have seen in practice in Data Mining. According to Khanam et al., the accuracy results from their models are displayed below:

[1] 1, 2 (logistic regression)
[2] 3, 4 (discriminant analysis)
[3] 5, 6, 7 (naive Bayes)
[4] 8, 9, 10, 11, 12, 13 (neural networks as decision trees)

Preprint. Under review.

Classifier	Accuracy (%)
Logistic Regression	45.71
Naïve Bayes (G)	45.58
Naive Bayes (M)	66.39
Naïve Bayes (B)	78.17
Decision Tree	92.28
KNN	95.25
Random Forest	97.25

Table 1: Classifier Accuracies

We will attempt to replicate methods used in previous research, but we expect to have some variability due to evaluating performance on different datasets. According to the literature review, models like MLP, Naive Bayes, KNN, and Random Forest have achieved high accuracies on fake news datasets.

By replicating and modifying these approaches, we will build upon these baselines, potentially improving performance through ensemble methods and neural networks as decision trees. For example, we plan on incorporating the K-Nearest Neighbors method in our Naive Bayes method, while will be informed by the literature of which we have conducted in-depth reviews.

1.3 Evaluation Metrics

Accuracy: The percentage of correctly classified instances.

Precision: The proportion of true positives among all predicted positives.

Recall: The proportion of true positives among all actual positives.

F1-Score: The harmonic mean of precision and recall.

2 Challenges

Although we are procuring our datasets on Kaggle, we still anticipate challenges when trying to run our models. Some of the datasets have different table formats, so making sure we are thorough when checking our code is necessary for this project. Additionally, we want to run analysis on a combined dataset that will see how different models perform on news from a merged dataset from all of our datasets. Merging all of our datasets will be difficult due to differing table layouts.

A significant challenge for this project lies in the fact that many of the machine learning methods we plan to use, such as K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes, have been applied to similar classification tasks with proven accuracy. While these models have been done, our main difficulty will be fine-tuning them to achieve optimal performance for our datasets. Additionally, combining different aspects of these models, such as integrating features from multiple algorithms(KNN and Naive Bayes), may be necessary to improve out-of-sample accuracy. Balancing model complexity with the risk of overfitting will be important as our goal is to develop a solution that performs well to unseen data.

3 To-Do List

- Ensure that all datasets from Kaggle (such as the American News, Soccer News, and Albanian Fake News datasets) are properly preprocessed.
- Finalize the selection of models based on their performance and suitability for the dataset at hand.
- Develop a hybrid model that combines the K-Nearest Neighbors (KNN) algorithm with Naive Bayes to capture local patterns while utilizing Naive Bayes for modeling the probabilistic relationships between specific linguistic features.
- Create features based on word frequency, n-grams, and sentiment analysis to distinguish between real and fake news. Implement techniques like Term Frequency-Inverse Document Frequency (TF-IDF) seen in the literature to enhance feature extraction.
- Evaluate model performance using cross-validation, and previously mentioned metrics to comprehensively assess each model's effectiveness.
- Create detailed visualizations to effectively communicate the model and classifier's performance, as well as feature importance.
- Write the final report with accordance to the given rubric.

4 References

4.1 Logistic Regression

[1] Ahmed H., Traore I., and Saad S. Detection of online fake news using n-gram analysis and machine learning techniques. *Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 2017, Vancouver, Canada, Springer, pp. 127–138. https://doi.org/10.1007/978-3-319-69155-8_9.

[2] Adeyiga, Johnson Adeleke, et al. "Fake News Detection Using a Logistic Regression Model and Natural Language Processing Techniques." *Bells University of Technology*, Research Article.

4.2 Discriminant Analysis

[3] Kesarwani, Ankit, et al. "Fake News Detection on Social Media using K-Nearest Neighbor Classifier." *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2020, pp. 1-4.

[4] Ogunsuyi, Opeyemi J., and Adebola K. Ojo. "K-Nearest Neighbors Bayesian Approach to False News Detection from Text on Social Media." *I. J. Education and Management Engineering*, vol. 12, no. 4, 2022, pp. 22-32. MECS, 8 Aug. 2022, doi:10.5815/ijeme.2022.04.03.

4.3 Naive Bayes

[5] Z. Khanam et al. 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012040.

[6] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 2017, pp. 900-903, doi: 10.1109/UKRCON.2017.8100379.

[7] Moorpani, Manisha, et al. "Fake News Detection Using Naïve Bayes Algorithm." *International Research Journal of Modernization in Engineering, Technology and Science*, vol. 4, no. 4, Apr. 2022, www.irjmets.com.

4.4 Neural Networks as Decision Trees

[8] Zhou X., R.J.A., Zafarani C.S. A survey of fake news: Fundamental theories, detection methods, and opportunities. 2020;53(5):1–40.

[9] S. Patil, S. Vairagade, and D. Theng, "Machine Learning Techniques for the Classification of Fake News," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-5, doi: 10.1109/ICCICA52458.2021.9697267.

[10] Wang W. Y., *Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection*, 2017, Association for Computational Linguistics, Stroudsburg, PA, USA.

[11] Islam, M.R., Liu, S., Wang, X., et al. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, vol. 10, 82, 2020. https://doi.org/10.1007/s13278-020-00696-x.

[12] Fayaz, M., Khan, A., Bilal, M., et al. Machine learning for fake news classification with optimal feature selection. *Soft Computing*, vol. 26, pp. 7763–7771, 2022. https://doi.org/10.1007/s00500-022-06773-x.

[13] Varshney, D., Vishwakarma, D.K. Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 8961–8974, 2021. https://doi.org/10.1007/s12652-020-02698-1.

5 Reference Review

The literature surrounding fake news detection has explored various machine learning models, each with its strengths and limitations. Adeyiga et al. found that logistic regression, while interpretable, is often outperformed by more complex models such as random forests, which can capture intricate patterns in text data. Kesarwani et al. demonstrated the potential of distance-based classifiers like K-Nearest Neighbors (KNN), achieving over 90% accuracy, showing its promise for this task. Granik and Mesyura highlighted the effectiveness of Naive Bayes, particularly when used with text vectorization techniques such as Term Frequency-Inverse Document Frequency (TF-IDF). Their work was further expanded by Moorpani et al., who compared different types of Naive Bayes models, identifying Bernoulli Naive Bayes as the best performer for binary classification tasks. Finally, Islam et

al. reviewed the use of deep learning techniques and showed that neural networks, especially deep architectures, outperformed traditional models by capturing more complex patterns in textual data. Together, these studies provide a solid foundation for developing fake news detection classifiers, emphasizing the importance of choosing the right model based on the trade-off between interpretability, accuracy, precision, and computational efficiency.