KNOX COLLEGE

FAKE NEWS DETECTION

BY

Tram Ngo, Anh Phan, Akbota Serikkyzy

March 2025

Table of Content

Chapter 1

INTRODUCTION

In this day and age, especially with the rapid rise of AI-generated content, misinformation has become more pervasive and sophisticated than ever before. The sheer volume of fake news articles circulating online makes it increasingly difficult to distinguish fact from fiction. Traditional methods of fact-checking are no longer sufficient to keep pace with the speed at which false information spreads. In this paper, we took a data-driven approach to address this challenge by applying machine learning models to detect fake news.

Our goal was to identify the most effective models for this task, optimize their parameters, and evaluate their performance on real-world data. We experimented with a variety of models, including logistic regression, support vector machines (SVM), decision trees, ensemble methods and deep learning architectures such as neural networks.

Once the models were trained, we tested them using a diverse set of sample articles collected from various online sources, including both human-written content and AI-generated text.

Our results were promising: several models achieved accuracy rates of $\geq 90\%$ on the test set, indicating strong potential for real-world application. The high accuracy suggests that the models were able to effectively identify linguistic patterns, contextual inconsistencies, and stylistic markers that are characteristic of fake news.

These findings highlight the potential of machine learning as a powerful tool for combating misinformation. However, we also recognize the limitations and challenges that remain, such as bias within our dataset and the ambiguity of ”fake” articles.

Chapter 2

LITERATURE REVIEW

We reviewed several existing studies that employed a similar statistical approach to fake news detection. Our goal, within the constraints of the allotted time, was to replicate and extend these methods to find room for improvement. Most prior approaches followed a structured pipeline consisting of data cleaning $\rightarrow$ vectorizing words $\rightarrow$ training 3–5 models $\rightarrow$ drawing conclusions.(1) (2)

Following this established framework, we adopted a similar process while introducing variations in the choice of models. The data cleaning phase involved removing stop words, lemmatizing, and handling punctuation, URLs, hashtags, and so on, to standardize the text. For vectorization, we experimented with techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings to convert textual data into numerical form.

In the model training phase, we expanded on prior work by testing a broader range of models, including logistic regression, support vector machines (SVM), random forests, Naive Bayes, and so on. We also experimented with deep learning models and ensemble methods. This allowed us to evaluate not only the predictive accuracy of each model but also their robustness to different types of misinformation, including AI-generated content. We also experimented with the number of models to use, as previous papers tended to work with 3-5. Ultimately, we settled on 5 of the best performing models.

Chapter 3

DATA AND ANALYSIS

One of the most important steps in Data analysis is choosing the right data resource to analyze. Choosing the right data for analysis is critical since it ensures that your findings are correct and relevant. Good data enables precise decisions, saves time, and allows for meaningful comparisons. It also assures compliance with legal and ethical requirements. The dataset we choose to start our project is “Fake and Real News Dataset” credited by Clément Bisaillon on Kaggle (3). This dataset contains 23,502 labeled fake news and 21,417 labeled true news with parameters like title, text, subject, and the date it was published. The next step we took was conducting Exploratory Data Analysis, which helps to explore the patterns, trends, missing values, and duplicate values. Additionally, we explored patterns and insights such as the top common words in each category. We conducted word frequency analysis and analyzed common terms in both fake and real news articles to uncover key differences.

3.1 Data Cleaning

Data cleaning helps to make the data consistent and ready to analyze. In these two datasets, we handled duplicate values to ensure data integrity (there were no missing values). Next, we cleaned the text by removing URLs, mentions, hashtags, and punctuation to reduce noise. We also removed stop words such as “the”, “is”, and “and” to focus on meaningful terms. Additionally, we applied ntlk’s lemmatization package to simplify words by reducing them to their base forms (e.g., “running” to “run”). Finally, the cleaned data was saved for further analysis. Here is the text before and after cleaning.

- Text before cleaning: `Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that.`

- Text after cleaning: `donald trump wish american happy new year leave it at that`

3.2 Data Visualization

Figure 3.1. shows the top common words likely to appear in fake or real news. For example, the word "said" appears a lot in real news, indicating that if the articles contain this word, it's likely to be real. Furthermore, one of the interesting sights is that the word 'trump' appears the most in fake news. The reason could be because he was a high-profile political figure, especially during the 2016 U.S. election. Fake news stories about him spread widely since they grabbed attention, sparked reactions, and generated clicks. Moreover, he just tended to make untruthful statements.

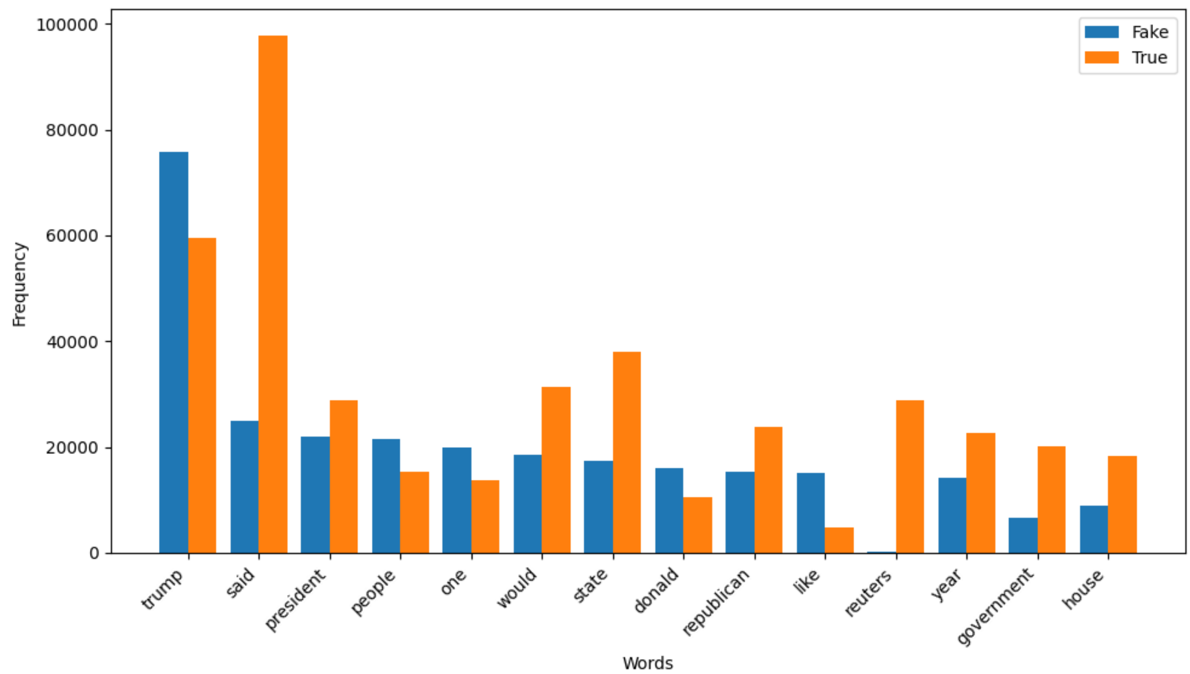

Figure 3.1: Most common words in dataset

Figure 3.2 shows the relationship between fake and true news articles. There are a total of 38,647 data points (red and blue) in this graph. Red points represent Fake news and blue points represent True news, and each point is originally 100,896 dimensional. Each point has been applied dimensional reduction using t-Distributed Stochastic Neighbor Embedding (t-SNE).

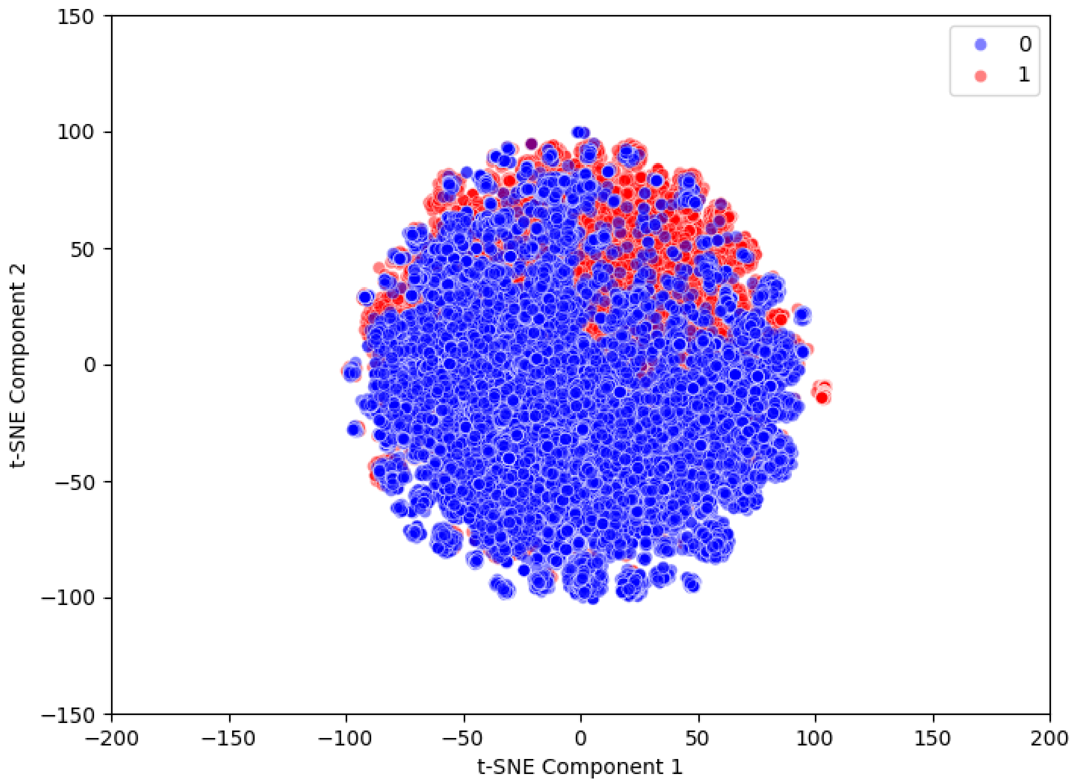

Figure 3.2: Relationship between fake and true news

3.3 Feature Importance and Evaluation

When the dataset is ready, it's time to analyze it. Since this dataset contains only texts, we decided to use TF-IDF (one of the Feature Engineering techniques). It is a technique that calculates the relevance of a term in a text based on its occurrence in numerous documents. Words that appear frequently in one text but seldom in others are given higher weights, improving the model's capacity to distinguish between common and unique phrases. This strategy effectively reduced noise by removing overly common words and stressing crucial textual elements.

TF-IDF Formula:

The TF-IDF score of a word t in a document d is calculated as:

$$TF-IDF(t,d) = TF(t,d) \times IDF(t)$$

Term Frequency (TF) measures how often a word appears in a document compared to the total number of words in that document:

$$TF(t,d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$

Inverse Document Frequency (IDF) reduces the weight of common words that appear in many documents and highlights rare but important words:

$$IDF(t) = \log\left(\frac{N}{1+df}\right)$$

TF-IDF will give out a (38,647x100,896) matrix. The rows represent the total number of fake and true news through 100,896 columns. There are with 6,278,505 non-zero values, indicating this is a sparse matrix where 83% of the values of the data are zero. This means that, among 100,896 unique words, there are a lot of words that are absent in some document, and there are some words that are too common in such documents.

3.4 Train/Test Split

We decided to use a 70/30 split which helps the model learn better because it has more training data. This is especially important in fake news detection, where text data is complex and requires more examples for the model to recognize patterns. A larger training set reduces errors and helps improve accuracy. In addition, since fake and real news articles may not be evenly distributed, having more training data helps balance classes. With this 70/30 split, we ensure that the model has enough data to learn while keeping a sufficient test set to check its performance.

3.5 Model Preparation

For our fake news detection analysis, we used several machine learning models that are well suited for text classification. Below, we provide a brief explanation of each model and why it is a good fit for our analysis.

Logistic Regression is a simple and effective model for binary classification problems such as fake news detection. It uses a linear decision boundary to separate real and fake news, making it easy to interpret and efficient for large datasets. Despite its simplicity, it achieves high accuracy by learning the most relevant features in text data. In our case, we used the L2 regularization (penalty = 'l2') to prevent overfitting and C=0.1 to control the strength of regularization, ensuring a good balance between bias and variance.

Support Vector Machine is a powerful algorithm that works by finding the optimal hyperplane to separate different classes. It performs exceptionally well in text classification tasks due to its ability to handle high-dimensional data. SVM helps detect fake news by maximizing the margin between real and fake news samples, making it a robust choice for this task. In our case, we used the RBF kernel, which helps capture complex decision boundaries, and set class_weight='balanced' to adjust for potential class imbalances, ensuring fair classification across different categories.

Neural Networks are designed to capture complex patterns in data by using multiple layers of neurons. They are particularly useful in text classification because they can learn intricate relationships between words and phrases. For fake news detection, a neural network can identify subtle differences in writing styles and language patterns between real and fake news articles. In our case, we used three fully connected layers (64, 32, and 1 neurons), ReLU activation functions, and a dropout rate of 0.2 to prevent overfitting. We optimized the model using the Adam optimizer and trained it using binary cross-entropy loss, which is well suited for classification problems.

Bagging (Bootstrap Aggregating) is an ensemble learning technique that combines multiple weak classifiers to improve accuracy. By averaging predictions from different models, Bagging reduces overfitting and enhances generalization. This approach ensures more stable and reliable classification results. In our case, we implemented a Bagging classifier with 50 estimators, setting max_samples=0.8 and max_features=0.8 to ensure diversity in the training samples and reduce overfitting.

Decision Trees classify data by learning simple rules based on input features. They work well for the detection of fake news because they create a structured decision-making process to differentiate between real and fake news. In our case, we restricted the tree depth to max_depth=15, set min_samples_split=5 to ensure meaningful splits, and used min_samples_leaf=2 to prevent overly complex branches.

3.6 Model Evaluation

To evaluate the effectiveness of different machine learning models in detecting fake news, we used four key metrics: accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model, while precision indicates the proportion of true fake news predictions out of all instances predicted as fake. Recall captures how well the model identifies actual fake news instances, and the F1-score provides a balance between precision and recall. These metrics allow for a comprehensive understanding of each model's performance, particularly in handling imbalanced datasets, where the number of real and fake news articles may not be equal.

We trained and tested several models to determine which performed best in detecting fake news. The ensemble model using bagging achieved the highest accuracy of 99.66%, followed closely by the decision tree at 99.58%. The support vector machine and neural network models also demonstrated strong performance, with accuracies above 98.9%. Logistic regression, while still effective, performed slightly lower than the other models, with an accuracy of 96.82%. These results are shown in Table 3.1.

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**ROC-AUC score**
Logistic Regression	0.9682	0.9804	0.9489	0.9644	0.9944
SVM	0.9909	0.9926	0.9875	0.9900	0.9988
Neural Network	0.9906	0.9907	0.9886	0.9896	0.9987
Ensemble (Bagging)	**0.9966**	0.9990	0.9935	**0.9963**	**0.9997**
Decision Tree	0.9958	**0.9958**	**0.9949**	0.9953	0.9943

Table 3.1: Evaluation Metrics

Figure 3.3 shows that most of our models perform well, with their ROC curves being close to the top left corner. This indicates a high true positive rate and a low false positive rate between different classifiers. The minimal gap between the curves suggests that all models effectively distinguish between real and fake news, with only slight variations in performance.

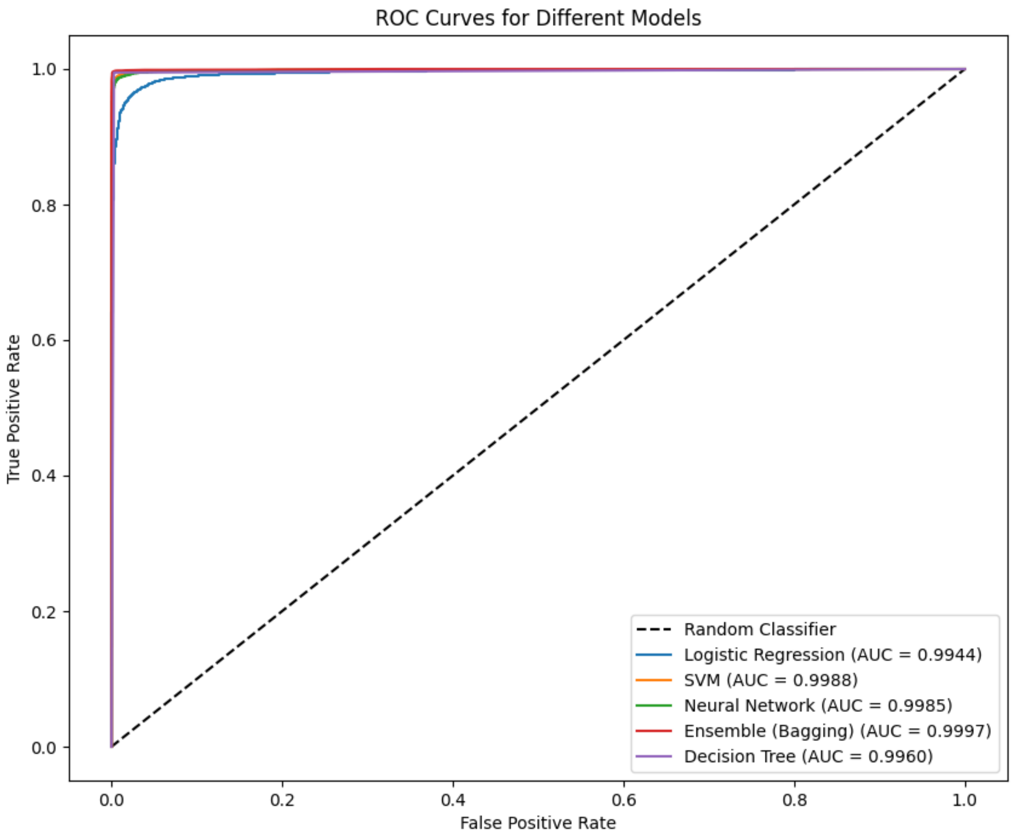

Figure 3.3: ROC Curves for Different Models

Chapter 4

RESULTS AND DISCUSSION

4.1 Limitations

Going into our project with our dataset, we knew it would show more bias towards articles on US politics. From our exploratory data analysis, we figured that with "Trump" being the most commonly occurring word, our model would perform better with certain topics than others. Future improvements would come from a more inclusive and extensive dataset.

A limitation we encountered was the notion of "fakeness". It is vague what determines whether an article is fake or true. If based on factuality, then we are not sure whether a statistical approach would suffice for how misinformation can take form. It could be a half-truth, an incomplete truth, or a truth with implications within the given context. The most solid way to detect fake news is still via fact-checking. We do not believe that any statistical approach can outdo that.

Another limitation we had was collecting information from actual news websites. With the rise of data scraping to use as AI training material, we had difficulty testing our model for several news media outlets. This raises the question of ethical data collection.

4.2 Application

We created a Jupyter Notebook that contains all the necessary information, takes user input in text or URL form, and runs it through our models. The models will then take a majority vote to determine whether the given article is real or fake.

Article	Verdict
Trump Administration Moves to End New York's Congestion Pricing Tolls - New York Times (4)	Real
Wave of Russian strikes kill at least 20 and injures dozens, Ukraine says - BBC (5)	Real
Trump Organization to develop $1.5 billion golf course and hotel project in Vietnam - Reuters (6)	Real
Tampa Bay Rays withdraw from planned $1.3 billion ballpark in St. Petersburg, citing storms, delays - AP News (7)	Real
AI Generated Article on Politics (8)	Fake
House Democrats Vow To Hold President Accountable With Agriculture Bill Where First Letter Of Every Line Spells Out 'Impeach Trump' - The Onion (9)	Fake
President Trump is Remaking America into a Manufacturing Superpower - The White House (10)	Fake

Table 4.1: Categorization of Articles from Various Outlets

Table 4.1 shows the result of our model as we tested in on articles from various news media outlets. We noticed that it labeled articles from major outlets as Real. Upon fact-checking these articles ourselves, we found this to be accurate. We also experimented on AI-generated articles and satirical ones, which our model successfully categorized as false.

Chapter 5

CONCLUSION

Ultimately, with everything discussed above, particularly the limitations, we recognize that statistically approaching misinformation detection remains a challenge. The dynamic and evolving nature of misinformation, combined with the increasing sophistication of AI-generated content, presents obstacles that even the most advanced models may struggle to overcome. While our model demonstrated strong performance under the test conditions, achieving high accuracy rates and promising predictive power, it is important to acknowledge that real-world application introduces additional complexities.

Misinformation often adapts to detection mechanisms, and subtle shifts in language, tone, and context could reduce the effectiveness of any static model over time. Moreover, factors such as dataset bias, sample diversity, and the ability to generalize to unseen data remain potential vulnerabilities. While our results suggest that the model performs well under controlled settings, the complexity and variability of real-world information mean that outcomes may differ when faced with novel misinformation tactics.

Despite these limitations, we believe that our model provides a solid foundation for future work in fake news detection. However, our findings should be interpreted with caution. Continuous retraining with updated data, further refinement of feature selection, and the integration of ensemble models may be necessary to maintain and improve performance in the face of evolving disinformation strategies.

Bibliography

[1] Jouhar, J., Pratap, A., Tijo, N., & Mony, M. "Fake news detection using python and machine learning." Procedia Computer Science, 233, 763–771. 5th International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2024). [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S1877050924006252

[2] Prachi, N. , Habibullah, Md., Rafi, Md., Alam, E., & Khan, R. "Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms." Journal of Advances in Information Technology Vol. 13, No. 6, December 2022. [Online]. Available at: https://www.jait.us/issues/JAIT-V13N6-652.pdf

[3] Bisaillon, C. "Fake and real news dataset." Kaggle, 2017 [Online]. Available at: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

[4] Oreskes, B., LeyStefanos, A., Hu, C. & Oreskes, B. "Trump Administration Moves to End New York's Congestion Pricing Tolls." New York Times, 2025. [Online]. Available at: https://www.nytimes.com/2025/02/19/nyregion/trump-congestion-pricing-nyc.html?smid=url-share

[5] Wilson, C. "Trump threatens 200% tariff on alcohol from EU countries." BBC, 2025. [Online]. Available at: https://www.bbc.com/news/live/cx2gprz84rlt

[6] Nguyen, P., Petty, M. "Trump Organization to develop $1.5 billion golf course and hotel project in Vietnam." Reuters, 2024 [Online]. Available at: https://www.reuters.com/business/trump-organization-develop-15-bln-golf-course-hotel-project-vietnam-2024-10-08/

[7] Anderson, C. "Tampa Bay Rays withdraw from planned $1.3 billion ballpark in St. Peters-

burg, citing storms, delays." AP News, 2025. [Online]. Available at: https://apnews.com/article/tampa-bay-rays-ballpark-cae72812c5f9d04804c139e06764a048

[8] "Write me a fake news article.". ChatGPT, 2025. [Online]. Available at: https://chatgpt.com/share/67d38853-b528-8002-9c81-1f261fb8efa4

[9] "House Democrats Vow To Hold President Accountable With Agriculture Bill Where First Letter Of Every Line Spells Out 'Impeach Trump.'" The Onion, 2019. [Online]. Available at: https://theonion.com/house-democrats-vow-to-hold-president-accountable-with-1838020720/

[10] "President Trump is Remaking America into a Manufacturing Superpower." The White House, 2025. [Online]. Available at: https://www.whitehouse.gov/articles/2025/03/president-trump-is-remaking-america-into-a-manufacturing-superpower/