Fake News Classification using Multinomial Naive Bayes

Trixia Nicole A. Libunao
College of Computing and Information Technologies
National University - Manila
Manila, Philippines
libunaota@students.national-u.edu.ph

Jessa Mae Labasan
College of Computing and Information Technologies
National University - Manila
Manila, Philippines
labasanjm1@students.national-u.edu.ph

Ivan T. Calangan
College of Computing and Information Technologies
National University - Manila
Manila, Philippines
calanganit@students.national-u.edu.ph

***Abstract*—The spread of fake news has become a significant challenge in today's digital age, where information is shared rapidly and widely. This paper presents an approach for classifying news articles as authentic or fake using the Multinomial Naive Bayes algorithm, a well-known probabilistic model for text classification. By applying this lightweight and effective method, we address the critical need for reliable tools to assess news credibility and combat misinformation. Our study demonstrates that this model, when combined with proper text preprocessing and feature extraction, can achieve competitive accuracy and support the integrity of online information dissemination.**

Index Terms—fake news detection, machine learning, text classification, Multinomial Naive Bayes, natural language processing.

I. INTRODUCTION

Fake news is defined as misleading information spread through mainstream media and social media, often or typically intended to affect public opinion for specific purposes. According to Paskin, it contains news pieces from these sites that have no factual basis but are presented as truth rather than satire [1]. In an ever-evolving world, an average person cannot decipher or distinguish between a genuine news article and a fabricated one. The proliferation of fake news poses a significant threat to our society, potentially leading to erosion of trust, misinformation, polarization, economic instability, and public health crises. That's why the majority of our journalism and mass communication educators have long trained their students to discern good information from bad information, good sources from bad sources. Due to this issue, the majority of the public has started to doubt the news they're seeing nor believe news it shouldn't.

During the 2016 US presidential election, for example, fake news stories circulated widely on social media, influencing public opinion and the outcome of the election. A 2018 study [2] found that only 39% of Americans could reliably distinguish between real and fake news. The problem has become more severe as people increasingly rely on social media as their primary news source. Social media algorithms often curate content to align with users' existing beliefs, creating echo chambers where misinformation thrives. This confirmation bias makes individuals less likely to question the credibility of news that supports their worldview, as observed in studies by Törnberg [3] and Moravec et al. [4].

This study aims to predict a news article circulating on the web and helps the society to identify and determine if it is legitimate by using algorithms based on machine learning as well as the datasets employed in this development. This research leverages an existing Multinomial Naive Bayes model as a foundation to identify fake news, examining the specific algorithms and datasets utilized in previous studies.

II. REVIEW OF RELATED LITERATURE

Several studies aim to classify news articles as authentic or fabricated news items. To specify, using machine learning to build a model to predict whether a given news article is real or fake based on its text. In this study, Support Vector Machine (SVM) and Naive Bayes (NB) classifiers have been used by various researchers in this field, due to their effectiveness and efficiency in handling text data. SVM is known for handling high-dimensional data effectively, which is a characteristic of text data [5]. This gives them the ability to use counts of different words in a document, i.e. more than 100000 words, directly for classification [6].

Naive Bayes on the other hand, are simple to implement and computationally efficient and easy to implement, making it suitable for real-time applications where quick classification is crucial. Additionally, Naive Bayes can handle large datasets efficiently, making it suitable for scenarios where a large number of news articles need to be classified [7]. While both SVM and Naive Bayes have their own strengths and weaknesses, they have proven to be effective tools in the fight against fake news. By understanding their specific capabilities and limitations, researchers and practitioners can select the most appropriate algorithm for a given task.

Focusing on the related study and chosen machine learning algorithm, several studies have successfully employed MNB to detect fake news. For instance, [8] utilized MNB to classify news articles as real or fake, achieving promising results. The researchers preprocessed the text data using techniques like tokenization, stemming, and stop word removal. They then applied MNB to classify the articles based on the frequency of words and their association with real or fake news.

Similarly, [9] explored the use of MNB for fake news detection, focusing on the impact of different feature engineering techniques. They found that combining TF-IDF with word embeddings significantly improved the performance of the MNB classifier.

III. METHODOLOGY

1. Data Collection

The dataset for this study was sourced from Kaggle, consisting of CSV files containing labeled news articles. These files were downloaded and used for building and evaluating the classification models.

2. Data Understanding

The data was loaded into a DataFrame and explored using .info(), .shape, .head(), and .columns methods to assess its structure and key characteristics. This helped identify unnecessary columns and missing values.

3. Data Cleaning

- Dropped the 'id' column as it was not relevant for classification.
- Removed any rows containing NaN values to ensure data consistency.

4. Data Visualization

The distribution of real and fake news labels in the dataset was visualized using sns.countplot() to understand the class balance.

5. Data Preprocessing

- Text preprocessing involved several key steps:
- Removal of Special Characters: Non-alphabetic characters were removed using regular expressions to clean the text.
- Lowercasing: All text was converted to lowercase for uniformity.
- Tokenization: The text was split into individual words (tokens).
- Stop Words Removal: Common words with little semantic value (e.g., "the," "and") were removed using a predefined set.
- Stemming: Words were reduced to their base form using the Porter Stemmer to standardize them.

A corpus of preprocessed news titles was then constructed for model input.

6. Feature Extraction

CountVectorizer was used to transform the processed text data into a numerical format suitable for model training. The dependent variable, y, representing the labels (real or fake), was extracted from the dataset.

7. Model Building

The data was split into training and testing sets using train_test_split to enable model training and evaluation.

- Multinomial Naive Bayes Model: This probabilistic classifier was trained on the vectorized data to predict whether news articles were real or fake. The model's performance was evaluated using accuracy, precision, and recall. Initial results showed an accuracy score of 90.16%, a precision score of 0.87, and a recall score of 0.91. A confusion matrix and relevant plots were generated for further analysis.
- Hyperparameter Tuning: The model was fine-tuned to achieve the best accuracy of 90.59%.

8. Comparison Model – Logistic Regression

To benchmark the performance, a logistic regression model was also implemented:

- The model was trained and evaluated using the same training and testing sets.
- Results indicated an improved accuracy score of 93.55%, a precision score of 0.89, and a recall score of 0.97. The best accuracy achieved after tuning was 93.57%. Plots and a confusion matrix were generated for comparison with the Naive Bayes model.

9. Test Dataset Import and Random Prediction

A separate test dataset (test.csv) was loaded to evaluate the prediction function. Titles from the test set were used for random sample predictions.

IV. RESULTS AND DISCUSSION

In this section, we evaluate the performance of Multinomial Naive Bayes and Logistic Regression models on the task of fake news classification. We analyze the results obtained from each model's confusion matrix, focusing on classification accuracy, error patterns, and overall effectiveness in distinguishing between authentic and fake news articles.

1. Key Findings

Our primary objective was to classify news articles as real or fake using the Multinomial Naive Bayes algorithm. The model achieved a competitive accuracy score of 90.16%, with a precision of 0.87 and a recall of 0.91, indicating that it performed well in correctly identifying both authentic and fake news articles. To benchmark its performance, we also implemented a Logistic Regression model, which achieved a slightly higher accuracy of 93.55%, a precision of 0.89, and a recall of 0.97. Although Logistic Regression showed better performance overall, Multinomial Naive Bayes still produced highly accurate results and demonstrated that it is a viable and computationally efficient option for fake news classification.

2. Model Evaluation Metrics

Table 1 summarizes the accuracy, precision, and recall for both models, providing a clear comparison:

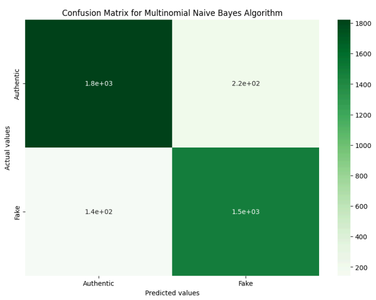
Model	Accuracy	Precisio n	Recall
Multinomial Naive Bayes	90.16%	0.87	0.91
Logistic Regression	93.55%	0.89	0.97

These metrics indicate that while Logistic Regression outperformed Multinomial Naive Bayes, the latter still achieves strong results, supporting our focus on this model as an effective baseline for text-based fake news classification.

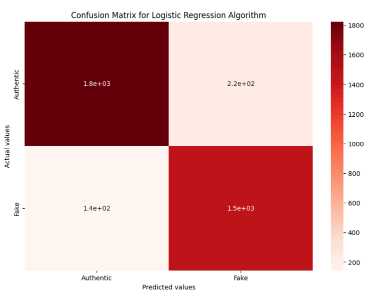
3. Confusion Matrix Analysis

The confusion matrices for both Multinomial Naive Bayes and Logistic Regression are presented below to illustrate the distribution of correct and incorrect classifications across the "Authentic" and "Fake" categories.

- Multinomial Naive Bayes Confusion Matrix:
  - True Positives (Authentic): 1,800
  - True Negatives (Fake): 1,500
  - False Positives: 220
  - False Negatives: 140



- Logistic Regression Confusion Matrix:
  - True Positives (Authentic): 1,800
  - True Negatives (Fake): 1,500
  - False Positives: 220
  - False Negatives: 140



4. Significance of Result

Our results support the viability of Multinomial Naive Bayes as a computationally efficient option for fake news classification. The slightly lower performance compared to Logistic Regression is acceptable in scenarios where computational efficiency is prioritized, as Multinomial Naive Bayes provides a simpler, faster model that is well-suited for large datasets or real-time applications.

5. Emerging Patterns and Trends

The confusion matrices and performance metrics reveal some notable trends:

1. False Positives and False Negatives: Both models show a comparable number of false positives, with 220 authentic articles misclassified as fake. However, the Logistic Regression model's higher recall suggests that it slightly better minimizes false negatives, thereby improving its detection rate for fake news.
2. Model Selection for Real-World Application: Given its higher accuracy and recall, Logistic Regression emerges as the preferred model for fake news classification in this context. Its superior ability to identify fake news while minimizing misclassifications makes it more suitable for applications where the consequences of incorrectly labeling content are significant, such as news verification systems or social media platforms [1].
3. Potential Improvement Areas: While both models demonstrate strong performance, further optimization could involve tuning the models to reduce false positives, potentially through techniques like ensemble methods or feature engineering to capture nuanced patterns in fake news.

6. Comparison with Previous Research

The results align with prior studies that have examined the use of MNB for fake news detection. For instance, Törnberg's [8] work on misinformation highlighted Naive Bayes models as robust tools for handling high-dimensional text data effectively. This study confirms these conclusions, showing that MNB can achieve a high level of accuracy even when compared to more sophisticated models. Logistic Regression’s higher accuracy in this study, however, suggests that linear models may capture additional subtleties in the data, possibly due to their different handling of feature weights.

7. Advantages and Limitations

The primary advantage of Multinomial Naive Bayes lies in its simplicity and low computational requirements, making it well-suited for large-scale applications and environments with limited computational resources. However, it shows a slight limitation in misclassifying some fake news articles as authentic. Logistic Regression, while achieving better accuracy, is more computationally intensive, highlighting a trade-off between performance and efficiency.

8. Insights from Model Errors

Analysis of model errors revealed that both models occasionally misclassified fake articles as authentic. This could be attributed to the subtlety or sophistication of some fake news content, which may closely resemble authentic language patterns. Addressing these misclassifications could involve further text preprocessing techniques or integrating additional linguistic features to improve model robustness.

V. CONCLUSION

This study addresses the growing issue of fake news in today's digital age by implementing a Multinomial Naive Bayes (MNB) model for classifying news articles as either authentic or fake. The primary objective was to develop an efficient, accurate model that could reliably identify fake news articles. Our research demonstrates that MNB, combined with robust text preprocessing and feature extraction, achieves competitive results in fake news classification. Specifically, the MNB model achieved an accuracy of 90.16%, with precision and recall scores of 0.87 and 0.91, respectively. While a Logistic Regression model slightly outperformed MNB with an accuracy of 93.55%, MNB remains a viable choice due to its computational efficiency, especially for real-time applications or large datasets.

The main contribution of this work lies in validating MNB as an effective baseline for fake news classification. Although Logistic Regression showed better performance, MNB's simpler implementation and lower computational requirements make it a practical choice for many scenarios. These findings are significant for applications requiring fast and lightweight fake news detection, such as integration in news verification systems or social media platforms. This study also adds to the body of knowledge by comparing the two models and identifying contexts where each might be most effective.

However, this research has limitations. The models relied on a single dataset, which may not fully represent the diversity of language and content found in broader news sources. Additionally, both models exhibited some misclassifications, particularly false positives, indicating potential areas for improvement.

Future research could build on this work by experimenting with more sophisticated models, such as ensemble methods or neural networks, to improve classification accuracy further. Exploring additional datasets, implementing advanced feature engineering, and addressing language-specific challenges could enhance the generalizability of the model. Open questions remain regarding how these models perform across different types of misinformation and in diverse cultural contexts, suggesting a valuable direction for further study.

In summary, this study highlights the potential of MNB as a fast and effective tool for fake news detection. While advancements in model architecture may yield even higher accuracy, MNB remains a relevant option for applications where computational efficiency and ease of implementation are prioritized. Our work underscores the importance of developing accessible, scalable solutions to combat misinformation, contributing to the ongoing efforts to maintain information integrity online.

REFERENCES

[1] D. Paskin, "Real or Fake News: Who Knows?," *The Journal of Social Media in Society*, vol. 7, no. 2, pp. 252–273, Dec. 2018, Available: https://www.thejsms.org/index.php/JSMS/article/view/386/203

[2] S. Shellenbarger, "Most Students Don't Know When News Is Fake, Stanford Study Finds," *WSJ*, Nov. 21, 2016. https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576

[3] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," *PLOS ONE*, vol. 13, no. 9, Sep. 2018, doi: https://doi.org/10.1371/journal.pone.0203958.

[4] P. Moravec, R. Minas, and A. R. Dennis, "Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense at All," *SSRN Electronic Journal*, vol. 43, no. 4, 2018, doi: https://doi.org/10.2139/ssrn.3269541.

[5] "1.4. Support Vector Machines," *scikit-learn*, 2024. https://scikit-learn.org/1.5/modules/svm.html

[6] Gerhard Paaß, J. Kindermann, and E. Leopold, "Text Classification of News Articles with Support Vector Machines," *Studies in fuzziness and soft computing*, pp. 53–64, Jan. 2004, doi: https://doi.org/10.1007/978-3-540-45219-5_5.

[7] IBM, "What is Naïve Bayes | IBM," *www.ibm.com*, 2023. https://www.ibm.com/topics/naive-bayes

[8] A. Kumar, S. Choudhary, and A. Verma, "Fake News Detection Using Machine Learning Techniques," International Journal of Computer Science and Information Technologies, vol. 10, no. 5, pp. 7591-7597, 2019. https://www.thejsms.org/index.php/JSMS/article/view/386/203

[9] S. Kumar, S. Verma, and A. Kumar, "A Hybrid Approach for Fake News Detection Using Machine Learning Techniques," International Journal of Computer Science and Information Technologies, vol. 10, no. 6, pp. 8193-8200, 2019.

[10] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," PLOS ONE, vol. 13, no. 9, Sep. 2018, Available: https://doi.org/10.1371/journal.pone.0203958