Title

Comparing Linguistic Patterns in Fake and Real News Articles

1. Introduction

Online news has become a primary way that people learn about politics and current events. At the same time, misinformation and “fake news” spread quickly through social media platforms, making it harder for readers to evaluate what they see. Rather than debating the truth of any specific claim, this project asks whether fake and real news articles use language in systematically different ways. If consistent stylistic differences exist, they may help both human readers and automated tools assess the credibility of online news.

Using a labeled Kaggle dataset of political news articles, this paper compares fake and real news along several dimensions: document and title length, subject labels, common vocabulary, sentiment, and latent topics. We then train a simple logistic regression classifier to test whether bag-of-words features can distinguish fake from real articles based solely on textual information.

The analysis is guided by three research questions:

1. RQ1 (Structure): Are there basic structural differences between fake and real news articles, such as class balance, subject distribution, or document length?
2. RQ2 (Vocabulary): Do fake and real news articles rely on different vocabularies, even after simple preprocessing?
3. RQ3 (Sentiment and topics): Are there differences in sentiment and topics that separate fake and real news, and can these differences be captured by a simple text classifier?

2. Background and Related Work

Fake news detection has become a major area of research in natural language processing (NLP). Many studies treat the problem as a supervised text classification task and rely on labeled datasets of misinformation. For example, Patwa et al. (2021) introduce a large-scale COVID-19 fake news dataset that highlights the linguistic diversity and subtlety of online misinformation. Their work underscores the need for models that can capture more than just surface-level keyword patterns.

Recent research has explored deep learning architectures that combine multiple NLP components. Sharma, Gupta, and Rai (2024) propose a framework that integrates BiLSTM, CNN, and transformer-based modules for fake news detection, sentiment analysis, and summarization on social media. Their hybrid system achieves over 90% accuracy and outperforms traditional machine learning baselines such as Naive Bayes and Random Forest, suggesting that deep sequence models can effectively capture contextual cues in misinformation.

Sahoo and Gupta (2021) adopt a complementary approach by combining lexical, syntactic, and metadata-based features in a multi-feature deep learning model, showing that richer feature sets substantially improve detection performance on social networks.

Hybrid convolutional–recurrent architectures have also been widely explored. Javed Awan et al. (2020) propose a bimodal CNN–LSTM model that uses convolutional layers to capture local n-gram patterns and LSTM layers to model longer-range dependencies in short social media posts. Their results demonstrate that hybrid models outperform pure CNN or LSTM architectures for fake news classification. Similarly, Hannah Nithya and Sahayadhas (2022) develop an LSTM-based approach with optimal feature selection and report strong performance on fake news detection, reinforcing the value of sequence-aware models.

A parallel line of work investigates improved feature engineering and ensemble methods using classical classifiers. Dev and Bhatnagar (2024) introduce Hybrid RFSVM, an ensemble that blends Random Forest and SVM to increase robustness on imbalanced fake news datasets. Matemilola and Aliyu (2024) design an enhanced Naive Bayes algorithm with modified probability estimates to better handle noisy online news content. These studies show that even relatively simple models can benefit from careful feature design and ensemble strategies, and they remain competitive with deep learning approaches in some settings.

While most of this literature focuses on building highly accurate detectors, fewer studies emphasize interpretable linguistic differences between fake and real news. Many deep models operate as black boxes, making it difficult to explain why a particular article is labeled as fake. In contrast, the present project focuses on exploratory analysis of structural, lexical, sentiment, and topic-based patterns, using a simple linear classifier primarily as an interpretability tool. By examining length distributions, vocabulary differences, sentiment scores, topics, and interpretable logistic regression weights, this study aims to provide a transparent view of how fake and real news differ in a well-defined corpus of political articles.

3. Data and Methods

3.1 Dataset

The analysis uses the “Fake and Real News” dataset from Kaggle, which contains two CSV files: *Fake.csv* and *True.csv*. Each file includes four primary fields: title, text, subject, and date, where *subject* indicates broad topical categories such as “politics,” “government-news,” or “worldnews.” We add a new label column, set to “fake” for rows from *Fake.csv* and “real” for rows from *True.csv*, and then concatenate the two files into a single DataFrame.

After removing duplicate articles (based on identical pairs of title and text) and dropping rows with missing title or text, the final dataset contains just under forty thousand articles. The classes are roughly balanced, with slightly more fake than real articles, which is helpful for both evaluation and visualization because it avoids severe class imbalance.

3.2 Basic cleaning and feature engineering

Before any modeling, We perform a small amount of cleaning that is standard in introductory text analysis:

- Remove exact duplicate articles based on the combination of title and text.
- Drop rows with missing title or text.
- Convert the date column to a datetime object and derive a coarse year_month field.
- Ensure that title and text are stored as strings with a consistent encoding.
- Create two simple length features:
  - title_len: number of whitespace-separated tokens in the title.
  - text_len: number of whitespace-separated tokens in the body text.

We deliberately avoid heavy normalization at this stage because one goal is to keep the exploratory analysis interpretable. For example, We retain capitalization and original punctuation when examining raw token frequencies so that plots more directly reflect how the articles actually appear.

3.3 Text preprocessing for vocabulary-based analyses

For vocabulary, topic modeling, and classification, We create a cleaned version of each article's text using the following pipeline:

1. Convert all characters to lowercase.
2. Remove punctuation and non-alphabetic characters with a regular expression.
3. Split text into tokens on whitespace.
4. Remove English stopwords using the NLTK stopword list.
5. Drop very short tokens with fewer than three characters.
6. Join the remaining tokens back into a space-separated string stored in a new column, clean_text.

This representation is used as input for word frequency counts, word clouds, the topic model, and the logistic regression classifier.

3.4 Sentiment analysis

To approximate sentiment, We apply the TextBlob library to each article's body text. For every article, we compute the polarity score, which ranges from −1 (very negative) to +1 (very positive), and store it in a new column sentiment. While this measure is relatively coarse, it provides a simple way to compare how emotionally positive or negative fake and real news articles tend to be.

3.5 Topic modeling

To explore latent themes in the corpus, We fit a Latent Dirichlet Allocation (LDA) model using the cleaned text. Articles are vectorized with a CountVectorizer limited to the 5,000 most frequent tokens in the corpus. We set the number of topics to ten for interpretability and consistency with similar exploratory studies.

For each topic, we inspect the top-weighted words to assign an informal label such as “elections,” “foreign policy,” or “media and culture.” We also compute, for every article, the topic with the highest posterior probability and use this to examine how often fake and real articles fall into each topic.

3.6 Classification model

Finally, We train a simple linear classifier to test whether bag-of-words features can distinguish between fake and real news:

1. Transform the cleaned text with a TfidfVectorizer, capping the vocabulary size for efficiency.
2. Split the data into training and test sets using an 80/20 split while stratifying on the label to maintain class balance.
3. Fit a logistic regression classifier on the training set, using a sufficiently high max_iter value to ensure convergence.
4. Evaluate accuracy on the held-out test set.
5. Inspect the largest positive and negative coefficients to identify words that strongly push predictions toward the fake or real class.

Because logistic regression is a linear model, these coefficients can be interpreted as approximate feature importances, connecting the classifier back to the exploratory vocabulary analysis.

4. Exploratory Data Analysis

4.1 Class balance and subject distribution

A simple bar chart of label counts shows that the cleaned dataset contains slightly more fake than real articles, but the difference is modest. This near balance is beneficial because it prevents evaluation metrics like accuracy from being dominated by a single majority class and allows fairer comparison between fake and real news.

Next, we examine the distribution of the subject field by label. A stacked bar chart focusing on the most common subjects reveals that both fake and real articles heavily cover politics and government-related news, but the proportions differ. Some subjects, such as general “politics,” are represented in both classes, while others, such as “government-news,” are more strongly associated with real articles. This suggests that fake and real news may emphasize different subdomains even when they cover similar overall topics.

4.2 Article and title length

Histograms of the body text length show that real articles tend to cluster between roughly 300 and 800 words, with a fairly peaked distribution. Fake articles exhibit a broader distribution with a longer right tail, indicating that extremely long pieces are more common in the fake news subset. Title lengths show an even clearer pattern: real headlines are shorter and more tightly clustered around ten to twelve words, whereas fake headlines are noticeably longer on average and have a heavier right tail. This pattern is consistent with the idea that fake news often relies on longer, attention-grabbing titles to attract readers.

4.3 Vocabulary differences

To understand lexical patterns, we first look at the most frequent tokens in the raw text before preprocessing. As expected, both fake and real articles heavily use function words such as “the,” “to,” and “and,” which are common across all English prose. Because these tokens carry little semantic information, they are removed in later analyses.

After applying the cleaning pipeline and removing stopwords, more meaningful vocabulary differences emerge. A comparison of the top twenty cleaned words in each class shows that real news is dominated by institutional and event-focused terms like “said,” “government,” “president,” “reuters,” “washington,” and “states.” These words reflect conventional political reporting and reference established organizations, locations, and actors.

In contrast, fake news articles use a broader mix of emotionally loaded and conversational terms such as “people,” “one,” “image,” “via,” and “know,” alongside political names and entities. Overall, fake news appears to blend political subject matter with more informal, attention-grabbing language. Word clouds derived from the cleaned text reinforce this contrast: in the real news cloud, phrases like “white house,” “united state,” and the names of specific officials stand out, while in the fake news cloud, generic narrative terms such as “said,” “one,” “people,” and “featured image” appear relatively larger.

4.4 Sentiment patterns

Using TextBlob polarity scores, We compare the sentiment distributions of fake and real articles. Both classes are centered near neutral sentiment, which is unsurprising for political news reporting. However, the shapes of the distributions differ. Real news shows a narrow, symmetric distribution tightly clustered around zero, suggesting relatively neutral, factual language. Fake news exhibits a slightly wider distribution with heavier tails on both the positive and negative sides. This indicates that fake news may rely more frequently on emotionally charged language, whether highly negative (e.g., scandal or outrage framing) or highly positive (e.g., praise and celebration).

A small heatmap of average sentiment by subject and label further suggests that, within certain subjects, fake news tends to be slightly more polarized than real news. For example, in topics related to conspiracies or media criticism, fake articles often have more extreme sentiment scores, whereas real articles on similar subjects remain closer to neutral.

4.5 Topic modeling

The LDA model identifies ten broad topics across the corpus. Inspecting the most probable words in each topic reveals interpretable themes. For instance:

- Topic 0 is dominated by “party,” “election,” “vote,” and “campaign,” corresponding to electoral politics.
- Topic 3 features “russia,” “intelligence,” “investigation,” and “hacking,” reflecting foreign policy and security concerns.
- Topic 7 mixes “media,” “story,” “people,” and “social,” and appears to be related to media, culture, and public reaction.

By assigning each article to its most probable topic and plotting topic counts by label, we observe that some topics are relatively balanced between fake and real news, while others are skewed. Topics that revolve around mainstream policy or government processes tend to have similar numbers of fake and real articles. In contrast, topics involving conspiracies, scandals, or media criticism display a higher proportion of fake news. This supports the idea that fake and real news do not simply report the same events in different tones; they emphasize somewhat different subsets of the political agenda.

4.6 Classification results

Using TF–IDF features on the cleaned text, the logistic regression classifier achieves an accuracy of 98.4% on the held-out test set. This performance is well above random guessing and indicates that simple lexical patterns already contain substantial information about whether an article in this dataset is labeled fake or real.

To interpret the model, we examine the words with the largest positive and negative coefficients. Tokens with large positive coefficients push the prediction toward the fake class, whereas tokens with large negative coefficients push it toward the real class. Many of the fake-associated words are emotionally charged or sensational, or are terms that frequently appear in misleading or conspiratorial headlines. In contrast, real-associated words tend to be names of institutions, locations, and procedural terms typical of conventional political reporting. The fact that these patterns mirror the earlier vocabulary analysis increases confidence that the classifier is capturing meaningful stylistic differences rather than spurious artifacts.

5. Initial Inferences and Discussion

Bringing the analyses together, several consistent patterns emerge.

For RQ1, the results show clear structural differences between fake and real news. Fake news tends to have longer, more variable article bodies and noticeably longer headlines. These differences in length and subject distribution suggest that fake news in this dataset is not just a random subset of political reporting but a distinct style of content that often uses attention-grabbing titles and less standardized formats.

For RQ2, vocabulary-based analyses indicate that fake and real news rely on different lexical choices even after simple preprocessing. Real news frequently references recognizable institutions and locations, reflecting its grounding in conventional political journalism. Fake news blends political entities with more generic, conversational terms, which may make stories feel more personal and engaging while reducing ties to verifiable sources.

For RQ3, the sentiment and topic analyses reveal that fake news is more sentimentally polarized and often concentrates on controversial topics such as conspiracies, scandals, or media conflicts. Real news remains more neutral and focuses more evenly on a range of policy and government issues. The logistic regression classifier's strong performance using only TF–IDF features demonstrates that these stylistic and topical differences are systematic enough to be exploited by even a simple linear model.

At the same time, it is important not to reify any single feature, such as title length, sentiment, or specific keywords, as a definitive "fake news detector." For example, some legitimate investigative pieces also use emotionally charged language, and some fake articles may mimic the style of professional journalism. The patterns observed here should therefore be interpreted as aggregate signals: they describe how fake and real news tend to differ on average in this dataset, not hard rules for judging individual stories.

6. Limitations and Future Work

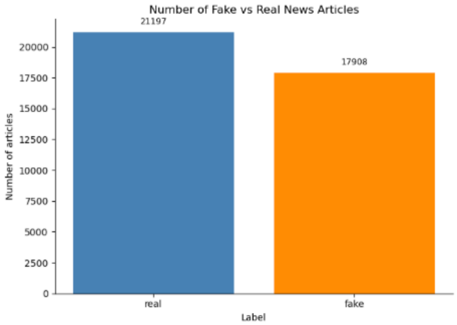
This analysis has several limitations. First, the dataset focuses on U.S. political news from a particular period. The findings may not generalize to other domains (such as health misinformation or celebrity gossip), to different countries, or to newer forms of online news. Second, the labels in the Kaggle dataset are treated as ground truth, but the process used to assign "fake" and "real" categories is not fully transparent. Any labeling errors or biases could affect the observed patterns and the classifier's performance.

Third, the text preprocessing pipeline is deliberately simple. More sophisticated approaches, such as lemmatization, phrase detection, contextual embeddings, or syntax-aware representations, could reveal additional nuances in how fake and real news differ linguistically. Fourth, the sentiment measure from TextBlob is coarse and does not capture complex emotions such as sarcasm, moral outrage, or irony, which are often present in political discourse.

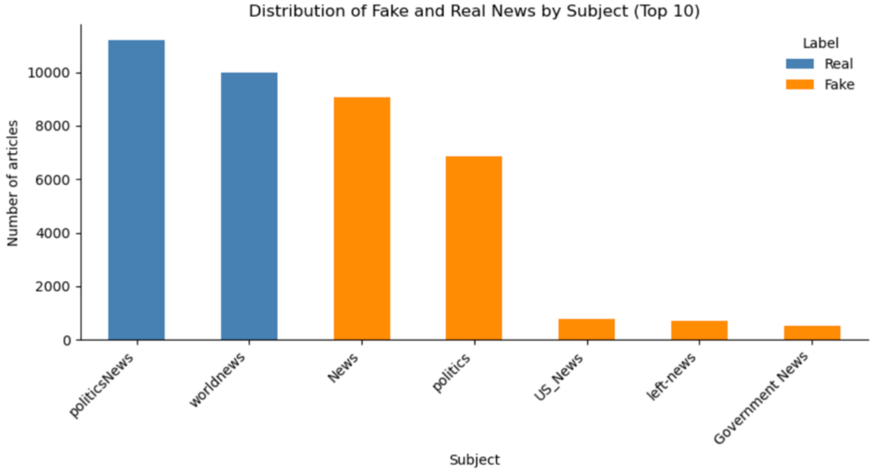
Future work could extend this project in several ways. One direction is to apply modern transformer-based language models, such as BERT or RoBERTa, and compare their performance to the simple logistic regression baseline while still maintaining some degree of interpretability (Hannah Nithya & Sahayadhas, 2022; Sharma et al., 2024). Another direction is to combine multiple datasets from different time periods and domains, for example, COVID-19 misinformation corpora such as the one introduced by Patwa et al. (2021) to assess how stable these stylistic patterns are across topics and platforms. Incorporating metadata such as source outlet, publication date, and social media engagement could also help disentangle linguistic patterns from source credibility and audience behavior. Finally, it would be valuable to connect text analysis with summarization techniques, for instance by exploring how abstractive models like BART and T5 summarize fake versus real content (Saxena & El-Haj, 2023), and to study how readers respond to these summaries in terms of sharing, commenting, or belief change.

7. Figure Captions

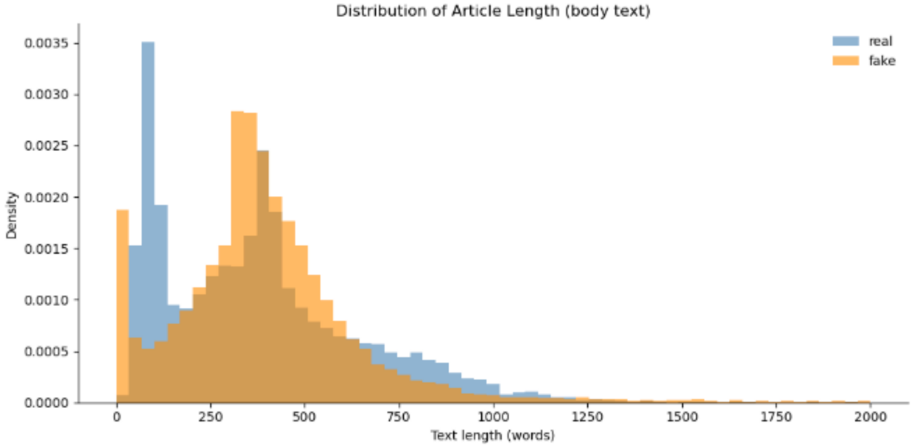
1. **Figure 1.** Number of fake and real news articles in the combined dataset.



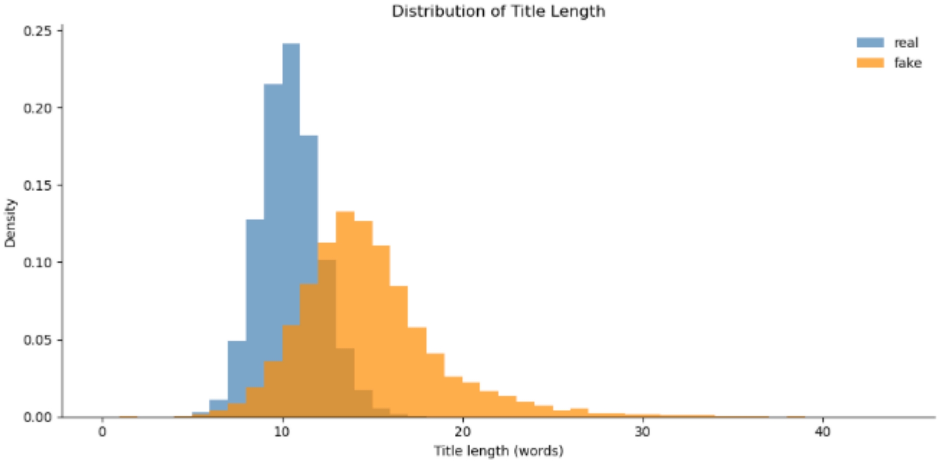
2. **Figure 2.** Distribution of fake and real news articles by subject for the top 10 subjects (stacked bar chart).



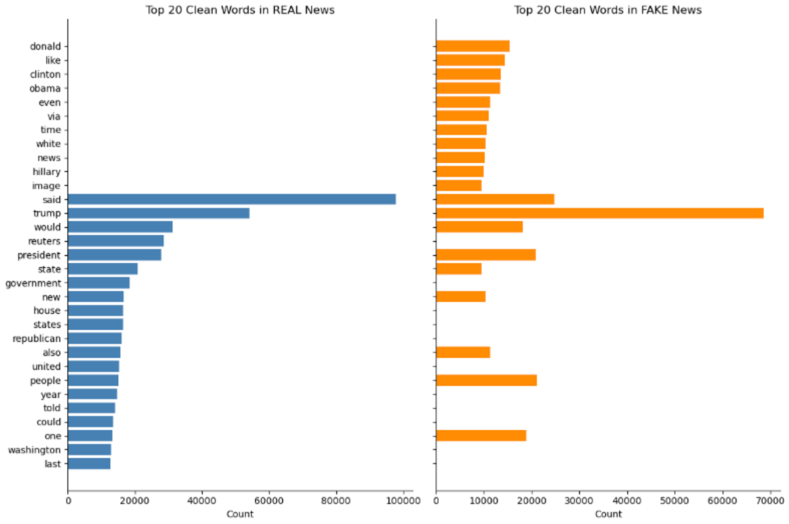
3. **Figure 3.** Distribution of article body length in words for fake and real news articles.



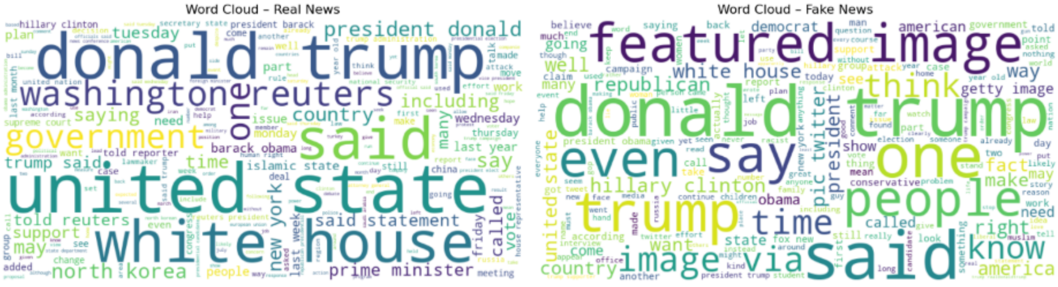
4. **Figure 4.** Distribution of title length in words for fake and real news headlines.



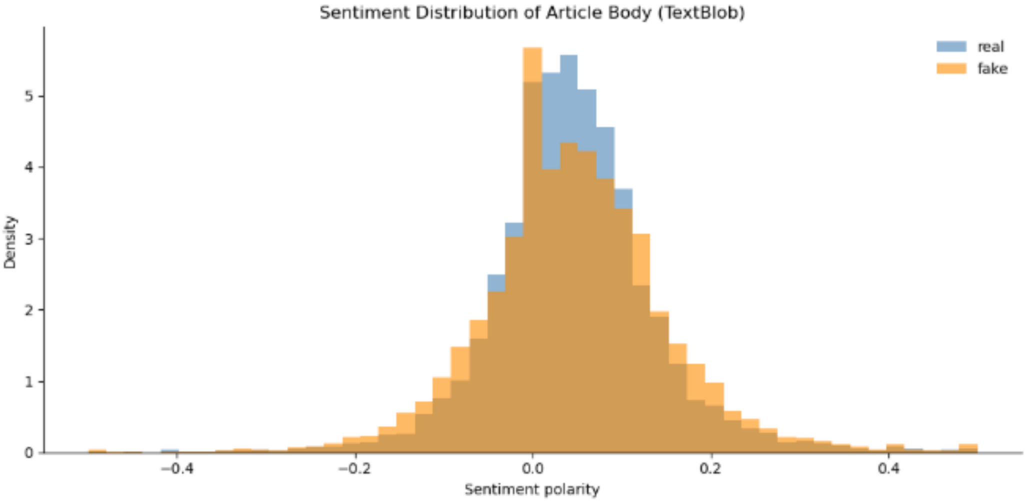
5. **Figure 5.** Top 20 most frequent cleaned content words in real and fake news articles.



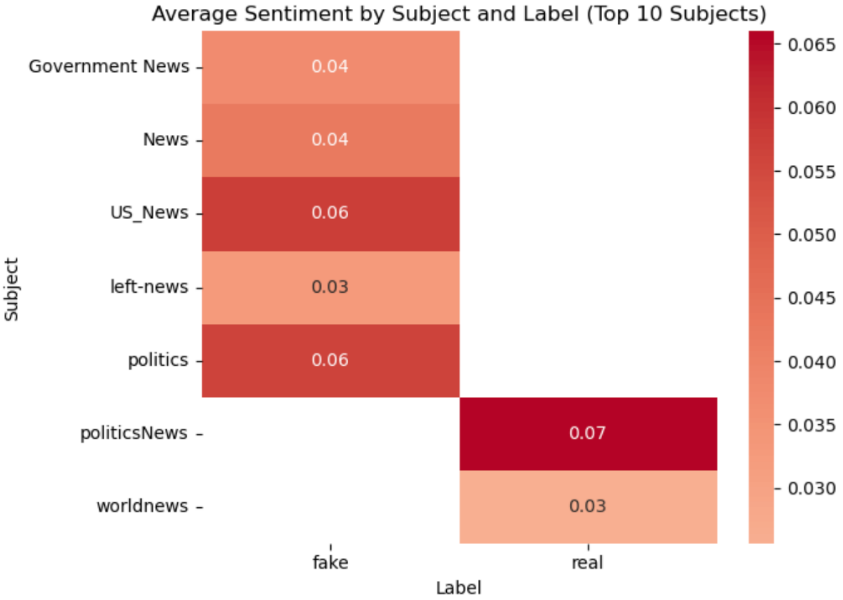
6. **Figure 6.** Word clouds of cleaned text for real and fake news articles, highlighting prominent vocabulary in each class.



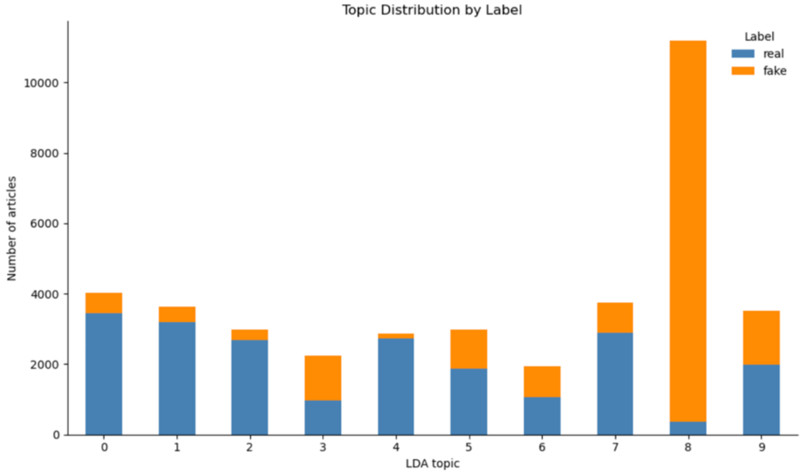
7. **Figure 7.** Sentiment distribution of article body text for fake and real news based on TextBlob polarity scores.



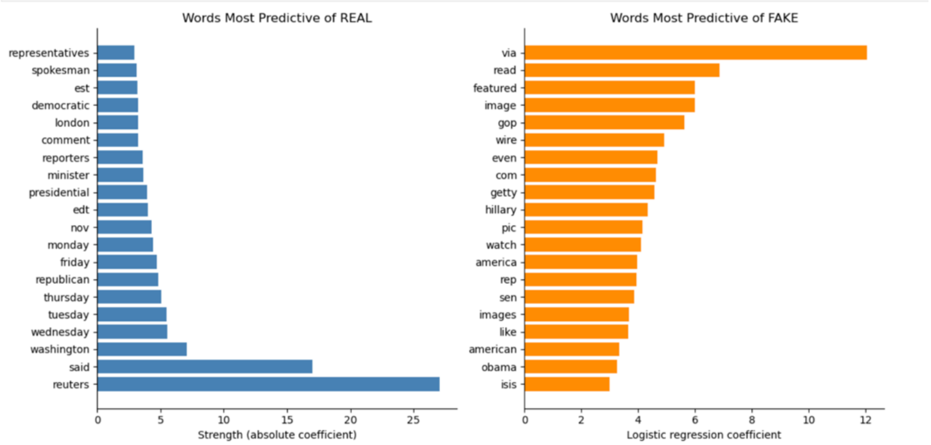
8. **Figure 8.** Average sentiment by subject and label for the top subjects, showing differences between fake and real news.



9. **Figure 9.** Topic distribution by label, showing the number of fake and real news articles assigned to each LDA topic.



10. **Figure 10.** Words most predictive of real and fake news according to the largest logistic regression coefficients on TF–IDF features.



8. References

Dev, D. G., & Bhatnagar, V . (2024). Hybrid RFSVM: Hybridization of SVM and Random Forest models for detection of fake news. *Algorithms, 17(10), 459. 10.3390/a17100459*

Hannah Nithya, S., & Sahayadhas, A. (2022). Automated fake news detection by LSTM enabled with optimal feature selection. *Journal of Information & Knowledge Management, 21(3). 10.1142/S0219649222500368*

Javed Awan, M., Shehzad, F., Muhammad, H., & Ashraf, M. (2020). Fake news classification bimodal using convolutional neural network and long short-term memory. *International Journal of Emerging Technologies, 11(5), 197–204. 10.51519/journalisi.v5i3.548*

Kaggle. (n.d.). *Fake and real news dataset*. Kaggle.
https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/data

Matemilola, A. S., & Aliyu, S. (2024). Development of an enhanced Naïve Bayes algorithm for fake news classification. *Science World Journal, 19(2), 512–517. 10.4314/swj.v19i2.28*

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). *Fighting an infodemic: COVID-19 fake news dataset*. arXiv.
https://arxiv.org/pdf/2011.03327

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features-based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing, 100, 106983. https://doi.org/10.1016/j.asoc.2020.106983*

Saxena, P., & El-Haj, M. (2023). Exploring abstractive text summarisation for podcasts: A comparative study of BART and T5 models. *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, 1023–1033.10.26615/978-954-452-092-2_110*

Sharma, R., Gupta, M., & Rai, S. (2024). Text mining and natural language processing frameworks for enhanced fake news detection, sentiment analysis, and automated summarization in social media. International Journal of Basic and Applied Sciences, 14(2), 107–112.
https://www.researchgate.net/publication/392572072_Text_Mining_and_Natural_Language_Processing_Frameworks_for_Enhanced_Fake_News_Detection_Sentiment_Analysis_and_Automated_Summarization_in_Social_Media

Ahmed, H., Traore, I., & Saad, S. (2018). *Detecting opinion spams and fake news using text classification*. **Security and Privacy, 1(1), e9.** https://doi.org/10.1002/spy2.9

Aimeur, E., Amri, S., & Brassard, G. (2023). *Fake news, disinformation and misinformation in social media: A review*. **Social Network Analysis and Mining, 13, 30.** https://doi.org/10.1007/s13278-023-01028-5

Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election*. **Journal of Economic Perspectives, 31(2), 211–236.** https://doi.org/10.1257/jep.31.2.211

Awan, M. J., Shehzad, F., Muhammad, H., & Ashraf, M. (2020). *Fake news classification bimodal using convolutional neural network and long short-term memory*. **International Journal of Emerging Technologies, 11(5), 197–204.** https://doi.org/10.51519/journalisi.v5i3.548

Bessi, A., & Ferrara, E. (2016). *Social bots distort the 2016 U.S. Presidential election online discussion*. **First Monday, 21(11).** https://firstmonday.org/ojs/index.php/fm/article/view/7090

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). *Automatic deception detection: Methods for finding fake news*. **ASIS&T Annual Meeting.** https://doi.org/10.1002/pra2.2015.145052010082

D’Ulizia, A., Caschera, M. C., Ferri, F., & Grifoni, P. (2021). *Fake news detection: A survey of evaluation datasets*. **PeerJ Computer Science, 7, e518.** https://doi.org/10.7717/peerj-cs.518

Dev, D. G., & Bhatnagar, V. (2024). *Hybrid RFSVM: Hybridization of SVM and Random Forest models for detection of fake news*. **Algorithms, 17(10), 459.** https://doi.org/10.3390/a17100459

Galli, A., Masciari, E., Moscato, V., & Sperlí, G. (2022). *A comprehensive benchmark for fake news detection*. **Journal of Intelligent Information Systems, 59(1), 237–261.** https://doi.org/10.1007/s10844-021-00646-9

Helmstetter, S., & Paulheim, H. (2018). *Weakly supervised learning for fake news detection on Twitter*. **ASONAM 2018.** https://data.dws.informatik.uni-mannheim.de/fakenews/asonam_short.pdf

Horne, B. D., & Adalı, S. (2017). *This just in: Fake news packs a lot in title, uses simpler, repetitive content…* **FEVER Workshop.** https://arxiv.org/abs/1703.09398

Hu, M., Mao, Y., & Zhang, C. (2024). *An overview of fake news detection: From a new perspective*. **Fundamental Research, 4(5), 849–869.** https://doi.org/10.1016/j.fmre.2024.01.017

Kaggle. (n.d.). *Fake and real news dataset*. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/data

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). *The science of fake news*. **Science, 359(6380), 1094–1096.** https://doi.org/10.1126/science.aao2998

Matemilola, A. S., & Aliyu, S. (2024). *Development of an enhanced Naïve Bayes algorithm for fake news classification*. **Science World Journal, 19(2), 512–517.** https://doi.org/10.4314/swj.v19i2.28

Nithya, S. H., & Sahayadhas, A. (2022). *Automated fake news detection by LSTM enabled with optimal feature selection*. **Journal of Information & Knowledge Management, 21(3).** https://doi.org/10.1142/S0219649222500368

Oshikawa, R., Qian, J., & Wang, W. Y. (2020). *A survey on NLP for fake news detection*. **LREC 2020.** https://aclanthology.org/2020.lrec-1.747/

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). *Finding deceptive opinion spam*. **ACL 2011.** https://aclanthology.org/P11-1032/

Paschalides, D., Mermigas, D., Andreou, A., Tymvios, F., & Andreou, P. (2021). *Check-It: A plugin for detecting fake news on the web*. **Online Social Networks and Media, 24, 100156.** https://doi.org/10.1016/j.osnem.2021.100156

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). *Fighting an infodemic: COVID-19 fake news dataset*. arXiv. https://arxiv.org/abs/2011.03327

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). *Automatic detection of fake news*. **COLING 2018.** https://aclanthology.org/C18-1287/

Petrou, A., Aggarwal, C. C., Panagiotopoulos, D., & Stavrakantonakis, I. (2023). *A multiple change-point detection framework on linguistic features of real and fake news*. **Scientific Reports, 13, 5953.** https://doi.org/10.1038/s41598-023-32952-3

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018). *A stylometric inquiry into hyperpartisan and fake news*. **ACL 2018.** https://aclanthology.org/P18-1022/

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). *Truth of varying shades*. **EMNLP 2017.** https://aclanthology.org/D17-1317/

Sahoo, S. R., & Gupta, B. B. (2021). *Multiple features-based approach for automatic fake news detection*. **Applied Soft Computing, 100, 106983.** https://doi.org/10.1016/j.asoc.2020.106983

Saxena, P., & El-Haj, M. (2023). *Exploring abstractive text summarisation for podcasts: A comparative study of BART and T5 models*. **RANLP 2023**, 1023–1033. https://doi.org/10.26615/978-954-452-092-2_110

Sharma, R., Gupta, M., & Rai, S. (2024). *Text mining and NLP frameworks for enhanced fake news detection...* **International Journal of Basic and Applied Sciences, 14(2), 107–112.** https://www.researchgate.net/publication/392572072

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake news detection on social media: A data mining perspective*. **SIGKDD Explorations, 19(1), 22–36.** https://doi.org/10.1145/3137597.3137600

Shu, K., Wang, S., & Liu, H. (2020). *Combating disinformation in a social media age*. **WIREs Data Mining and Knowledge Discovery, 10(6), e1385.** https://doi.org/10.1002/widm.1385

Vosoughi, S., Roy, D., & Aral, S. (2018). *The spread of true and false news online*. **Science, 359(6380), 1146–1151.** https://doi.org/10.1126/science.aap9559