My NLP Project LUL

Braunegg, Florian[1] and Second Author[2]

[1]Graz University of Technology, Graz, Austria
[2]University of Graz, Graz, Austria

June 19, 2025

Abstract

Your abstract goes here. It should be a single paragraph of 150–250 words summarizing the report's context, aims, methods, results, and conclusions. The entire paper should not exceed 6 pages (excluding limitations, ethical considerations, references, appendix). Please do not forget to fill out the contributions for each team member in the table.

1 Introduction

Fake news have become a major concern in contemporary political and social discourse. They contribute to the polarization and fragmentation of society (Au et al., 2022) and, in extreme cases, may even pose a threat to public safety. This is exemplified by incidents such as Pizzagate, in which an armed man stormed a pizzeria in Washington, D.C., attempting to rescue nonexistent children based on fabricated information (Lipton, 2016). Due to their wide-ranging impact, addressing fake news effectively requires an approach that integrates both social and technical perspectives. While disciplines like psychology and sociology ofte explore how such misinformation spreads and shapes public opinion, computational fields, especially Natural Language Processing (NLP), provide tools to systematically analyze and detect such content. NLP enables the identification of linguistic patterns and contextual signals that are typical of deceptive or false information. Consequently, the reliable detection and classification of fake news remains a core challenge that NLP is uniquely equipped to address.

As part of the course "Natural Language Processing" at the Graz University of Technology, we analyzed and classified fake news in three consecutive stages that illustrate NLP techniques for fake news detection.

2 Related Work

Fake news can be broadly understood as information presented in the form of news that is intentionally false and designed to mislead its audience. While the definition is not universally agreed upon, broader interpretations such as the one adopted in this paper treat fake news as an umbrella term that includes both misinformation (false information shared without intent to deceive) and disinformation (false information shared deliberately). In this broader perspective, any misleading or incorrect content presented as news, regardless of intent, may be classified as fake news (de Oliveira et al. 2021; for a discussion of the terminology see Tandoc et al. 2017). Although instances of fake news can be traced as far back as 2100–1200 BC, as seen in the Babylonian Epic of Gilgamesh (Roozenbeek and van der Linden, 2024), scholarly interest in the phenomenon has only gained significant momentum in recent years (Tătaru et al., 2024). Despite definitional differences, fake news have far-reaching consequences for social cohesion, public opinion, institutional trust, and political development. Prominent examples of these consequences include the election of U.S. President Donald Trump (Allcott and Gentzkow, 2017), the Brexit referendum (Orlando, 2023), and the COVID-19 crisis (Ferreira et al., 2022), all of which are closely linked to the digitization of society.

Interest in fake news detection within the field of NLP began to grow significantly in the second decade of the 21st century. Early research on fake news detection primarily relied on traditional machine learning algorithms and, in some cases, rule-based systems. These systems operate by defining sets of rules or linguistic heuristics to classify articles. Rule-based systems generally exhibit lower accuracy compared to more advanced approaches, primarily due to their limited ability to capture contextual dependencies and adapt to dynamic linguistic variations. Consequently, they are prone to generating a high rate of false positives, especially when applied to sophisticated fake narratives or texts that deviate from their predefined rule configurations (Polu, 2024; Repede and Brad, 2023). Applications of such rule-based approaches can be found, for example, in Alotaibi and Alhammad (2022), who examined the spread of Arabic fake news during the COVID-19 pandemic, and in Yuliani et al. (2019), who developed a rule-based framework for hoax detection. When applied to larger or more complex text corpora, traditional machine learning algorithms such as Logistic Regression or Random Forrest often outperform rule-based systems. Traditional machine learning algorithms rely on statistical modeling techniques and are capable of learning patterns directly from data. Although traditional machine learning algorithms are more flexible than rule-based systems, they still lack the ability to capture contextual information within text and depend heavily on manually engineered features and expert domain knowledge. Moreover, large amounts of data are often required to effectively train and fine-tune these algorithms for specific tasks. Consequently, the performance and applicability of such models are highly dependent on how the problem is defined and on the quality of the manually engineered features they utilize (Pittman, 2025; Polu, 2024). Despite these limitations, comparative studies such as that by Sudhakar and Kaliyamurthie (2022) demonstrate the considerable potential of traditional machine learning methods. The advent of deep learning marked a significant advancement in computational classification tasks, with models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) capable of learning hierarchical and sequential patterns directly from data. However, the introduction of the transformer architecture and the subsequent development of models such as BERT, RoBERTa, and GPT represented a major breakthrough in natural language processing. Unlike earlier architectures, transformers leverage contextual embeddings and attention mechanisms to capture subtle dependencies and nuanced patterns in text that traditional machine learning models cannot handle. However, deep learning approaches still require substantial amounts of data, annotated datasets, and significant computational resources in order to achieve strong performance. Another major limitation of deep learning models is their lack of interpretability, as they often function as black-box systems whose internal decision-making processes are difficult to trace or explain (Pittman, 2025; Polu, 2024). In response to the limitations of individual approaches, recent research has proposed hybrid models for fake news detection that aim to combine the strengths of different techniques. Nasir et al. (2021) for example proposed a CNN-RNN hybrid model and Albahar (2021) an SVM-RNN-BI-GT hybrid model for fake news detection.

3 Materials and Methods

Our dataset consists of 150 articles of both fake and real news, collected and provided by students and faculty at Graz University of Technology. For stages two and three, we implemented all code using Python. In the first stage, we manually evaluated all articles to create a labeled dataset for the subsequent stages and to gain an initial understanding of how fake news is written. In doing so, we followed the definition of fake news outlined earlier in the related work section.

In the second stage, we imported the dataset using pandas (McKinney, 2010) and employed the Random Forest classifier from scikit-learn (Pedregosa et al., 2011) to classify the 150 articles as either fake or real. As features for our classifier, we extracted part-of-speech (PoS) tags, named entities (used in combination with a knowledge base), emotional content in the text, readability and difficulty scores, as well as grammatical and spelling errors. Each feature was extracted separately for both the headline and the body text of each article. We selected these features for their straightforward and interpretable nature, in contrast to statistical representations like n-grams or TF-IDF.

PoS-tags were extracted using spaCy (Honnibal et al., 2020), which also handled the tokenization. We restricted the PoS-tags to the 17 universal part-of-speech categories defined in the Universal Dependencies framework (universaldependencies.org, n.d.) to extract easily interpretable logical units. The frequency of each tag was calculated relative to the total number of tokens in the article.

The knowledge base, based on fake claims that were extracted with a one-shot prompt from GPT-4o mini to avoid bias towards our articles and due to time constraints, consists of two elements: entities (with aliases) and concise fake claims. Both elements were tokenized with spaCy and lowercased to improve matching. The claims were additionally lemmatized and filtered using spaCy's stopword list. Entities and aliases were searched in the text using spaCy's PhraseMatcher. For every match, we looked for the corresponding claims in a list of all sentences from the article that contained the entity. These sentences were lemmatized, lowercased, and filtered for alphanumeric characters and stopwords. If multiple entities were found in a sentence, each was checked separately. Since it was expected that not every claim would be worded exactly as in the knowledge base, we used WordNet provided by NLTK (Bird et al., 2009) to extract synonyms. These synonyms were also lowercased and lemmatized to improve matching accuracy. A match was recorded as soon as two words, representing the smallest logical unit, were found in the same sentence. As a feature, we extracted the number of mismatches relative to the number of sentences in the articles.

To detect emotions in the text, we compared all tokens extracted with spaCy against an extended version of the NRC Word–Emotion Association Lexicon (Mohammad and Turney, 2013), as provided by Bailey (n.d.). We calculated the relative frequency of each emotion in proportion to the total number of emotional terms in the article. Based on our findings during the manual analysis in stage one, we focused on four prevalent emotions: anger, anticipation, fear, and sadness.

To further assess article readability, we used the Flesch Reading Ease (Flesch, 1948) and the Automated Readability Index (Senter and Smith, 1967). The Flesch Reading Ease score ranges from 100, indicating very easy to read, to 0, indicating very difficult to read. The score decreases with longer average sentence lengths and more syllables per word, as longer constructions are considered harder to process. The Automated Readability Index usually ranges from 1 to 14 or higher, corresponding to U.S. school grade levels, and is based on the number of words per sentence and characters per word. To estimate word difficulty, we computed the proportion of difficult words relative to all words in the article, using the word list from Chall and Dale (1995). Both the readability score and the difficulty measure were calculated using textstat (Bansal, n.d.).

Grammatical and spelling errors were detected using LanguageTool for Python (Morris, n.d.). We calculated the number of detected issues relative to the total word count as an additional feature.

To determine the optimal combination of features and hyperparameters and to evaluate model performance, we split the dataset into an 80/20 train-test split. Hyperparameter optimization was performed using GridSearchCV from scikit-learn, with accuracy as the target metric and RepeatedStratifiedKFold with five folds and three repeats as the cross-validation strategy. This approach systematically tests various parameter combinations while controlling for overfitting. To refine the feature set, we applied the Boruta algorithm (Kursa and Rudnicki, 2010), implemented in Python by Homola (n.d.). This method evaluates feature importance by comparing actual features to randomly permuted shadow features. We selected features that ranked above the 90th percentile in importance scores, with statistical significance assessed at the 5% level. As a final step, model performance was evaluated using repeated hold-out validation over 100 random splits, reporting accuracy, macro-averaged precision, recall, and F1-score.

4 Results

Present the evaluation results you gained (objective results). Consider using tables and charts for the report. Of course, baselines are helpful for the reader to assess the performance of the method.

5 Discussion

Interpret and discuss your results (could be a more subjective results). Describe your findings here, what can we learn from the work.

6 Conclusion

Summarize the key findings and implications (design your report that one get the main insights from reading abstract/introduction/conclusions and glancing at the illustrations). Suggest future research (very briefly).

Limitations

List potential limitations of your work, e.g., only English language.

Ethical Considerations

Please consider how your work could potentially be used and may cause some harm. Also, sustainability aspects can be reported here, e.g., how much CO2e does you approach require.

Acknowledgments

Briefly acknowledge people, funding sources, or institutions.

Contributions

Table 1: List of contributions per team member.

Team Member	Contribution
Florian Braunegg	Stage0, Stage1, Stage2, Introduction, Relate Work, Methods & Materials (Materials, Stage1, Stage2), Results(Stage1, Stage2)
Another Name	Did not even care to show up.

References

Marwan Albahar. A hybrid model for fake news detection: Leveraging news content and user comments in fake news. *IET Information Security*, 15(2):169–177, 2021. doi: 10.1049/ise2.12021.

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

Fatimah L. Alotaibi and Muna M. Alhammad. Using a rule-based model to detect arabic fake news propagation during covid-19. *International Journal of Advanced Computer Science and Applications*, 13(1), 2022. doi: 10.14569/IJACSA.2022.0130114.

Cheuk Hang Au, Kevin K. W. Ho, and Dickson K. W. Chiu. The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implications. *Information Systems Frontiers*, 24(4):1331–1354, 2022. doi: 10.1007/s10796-021-10133-9.

Mark Bailey. Nrc lexicon, n.d. https://github.com/DemetersSon83/NRCLex/blob/master/nrc_en.json. Accessed: 2025-06-17.

Shivam Bansal. textstat, n.d. https://github.com/textstat/textstat. Accessed: 2025-06-17.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc., 1st edition, 2009. ISBN 0596516495.

Jeanne S. Chall and Edgar Dale. *Readability Revisited: The New Dale–Chall Readability Formula*. Brookline Books, Cambridge, MA, 1995. ISBN 1571290087.

Nicollas R. de Oliveira, Pedro S. Pisa, Martin Andreoni Lopez, Dianne Scherly V. de Medeiros, and Diogo M. F. Mattos. Identifying fake news on social networks based on natural language processing: Trends and challenges. *Information*, 12(1), 2021. doi: 10.3390/info12010038.

Carlos Ferreira, Vivian Ferreira, Rebeca Bandeira, and Luisa Ferreira. The impact of misinformation on the covid-19 pandemic. *AIMS Public Health*, 9(2):262–277, 2022. doi: 10.3934/publichealth.2022018.

Rudolf Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233, 1948. doi: 10.1037/h0057532.

Daniel Homola. Borutapy: An all relevant feature selection method for python, n.d. https://github.com/scikit-learn-contrib/boruta_py. Accessed: 2024-06-17.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spacy: Industrial-strength natural language processing in python, 2020.

Miron B. Kursa and Witold R. Rudnicki. Feature selection with the boruta package. *Journal of Statistical Software*, 36(11):1–13, 2010. doi: 10.18637/jss.v036.i11.

Eric Lipton. Man motivated by 'pizzagate' conspiracy theory arrested in washington gunfire. *The New York Times*, December 2016. https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html?_r=0. Accessed: 2025-06-16.

Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010. doi: 10.25080/Majora-92bf1922-00a.

Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013. doi: 10.1111/j.1467-8640.2012.00460.x.

Jack Morris. language-tool-python, n.d. https://github.com/jxmorris12/language_tool_python. Accessed: 2025-06-17.

Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlamis. Fake news detection: A hybrid cnn-rnn based deep learning approach. *International Journal of Information Management Data Insights*, 1(1):100007, 2021. doi: 10.1016/j.jjimei.2020.100007.

Vittorio Orlando. *Post-Truth Politics, Brexit, and European Disintegration*, pages 103–127. Springer International Publishing, Cham, 2023. doi: 10.1007/978-3-031-13694-8_6.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake

Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Jason M. Pittman. Truth in text: A meta-analysis of ml-based cyber information influence detection approaches, 2025. https://arxiv.org/abs/2503.22686. Accessed: 2025-06-17.

Omkar Reddy Polu. Ai-based fake news detection using nlp. *International Journal of Artificial Intelligence & Machine Learning*, 3(2): 231–239, 2024. doi: 10.34218/IJAIML_03_02_019.

Ştefan Emil Repede and Remus Brad. A comparison of artificial intelligence models used for fake news detection. *Bulletin of "Carol I" National Defence University*, 12(1):114–131, 2023. doi: 10.53477/2284-9378-23-10.

Jon Roozenbeek and Sander van der Linden. *The Psychology of Misinformation*. Contemporary Social Issues Series. Cambridge University Press, 2024.

Ross J. Senter and Edward A. Smith. Automated readability index. Technical Report AMRL-TR-6620, Aerospace Medical Research Laboratories, U.S. Air Force, Wright-Patterson Air Force Base, OH, November 1967.

M. Sudhakar and K.P. Kaliyamurthie. Effective prediction of fake news using two machine learning algorithms. *Measurement: Sensors*, 24:100495, 2022. doi: 10.1016/j.measen.2022.100495.

Edson Tandoc, Zheng Lim, and Rich Ling. Defining "fake news": A typology of scholarly definitions. *Digital Journalism*, 6:1–17, 08 2017. doi: 10.1080/21670811.2017.1360143.

George-Cristian Tătaru, Adrian Domenteanu, Camelia Delcea, Margareta Stela Florescu, Mihai Orzan, and Liviu-Adrian Cotfas. Navigating the disinformation maze: A bibliometric analysis of scholarly efforts. *Information*, 15(12), 2024. doi: 10.3390/info15120742.

universaldependencies.org. Universal pos tags, n.d. https://universaldependencies.org/u/pos/. Accessed: 2025-06-16.

S.Y. Yuliani, Mohd Faizal Bin Abdollah, Shahrin Sahib, and Yunus Supriadi Wijaya. A framework for hoax news detection and analyzer used rule-based methods. *International Journal of Advanced Computer Science and Applications*, 10(10), 2019. doi: 10.14569/IJACSA.2019.0101055.

A Appendix - Overview

The appendix can be used to add details, especially implementation aspects, or added evaluations. There is no page limit on the appendix. You may also report approaches that you tried, but did not work out. Additional examples can be reported, or prompts being used for generative AI.

B Appendix - Usage of AI

More details on the usage of AI: Zitiervorschlaege AI

C Appendix - Figure and Table Examples

Examples how to use figures, see Figure 1 and tables, see Table 1. In double, just let Latex layout the illustrations for you, or position them at the top or bottom of a page. Is is common to capitalise nouns that are followed by numbers, as they are considered names (proper noun), e.g., Page 4.

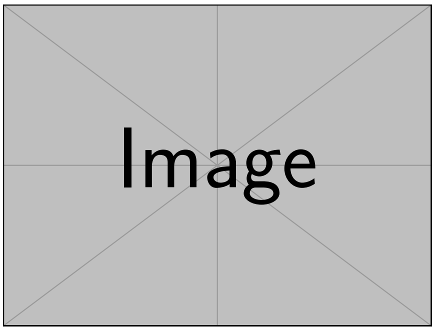

Figure 1: Example figure caption. Please consider to explain to the reader, what is depicted