A Machine Learning Approach to Combat the Spread of Misinformation

BT5153 Applied Machine Learning in Business Analytics – Group 01
Amanda Chia Wan Ying (A0280523L), Ang Kang Jie (A0226232R), Danice Angelee C. Parel (A0280367Y), Nattaya Silprakong (A0280554A)
GitHub Link: https://github.com/angkj1995/BT5153_Group01_2024

Abstract

This paper aims to leverage machine learning (ML) techniques to combat misinformation propagation. We evaluate the effectiveness of various Natural Language Processing (NLP) techniques, including traditional approaches (bag-of-words, TF-IDF) and sophisticated deep learning methods (word and sentence embeddings), to develop a high-accuracy fake news classifier. Additionally, we explore interpretable ML techniques to enhance model transparency.

1. Introduction

1.1 Background

The proliferation of fake news poses significant political and social implications. Social media, featuring decreased cost and increase accessibility, has become a preferred platform for news consumption over traditional sources like newspapers and television. In fact, social media has surpassed television as the primary news source (Shu et al., 2017), offering timelier updates and facilitating effortless sharing and discussion. This shift has lowered dissemination barriers, allowing news to spread easier and faster, consequently increasing fake news prevalence.

Fake news is characterized as 'news articles that are intentionally and verifiably false and could mislead readers' (Shu et al., 2017). Such outlets neglect rigorous information gathering and editorial standards, thereby compromising accuracy and credibility (Lazer et al., 2018).

The impact of fake news cannot be underestimated. Notable examples include claims that 5G mobile devices causes COVID-19 (Best, 2020) and vaccine-related misinformation that eroded trust and fanned speculations during the pandemic. In the 2016 US Presidential election, fake election stories garnered 8,711,000 engagements on Facebook, exceeding the combined 7,367,000 engagements for the top 20 most-discussed election stories from 19 major news websites (Silverman, 2016). Research further suggests that fake news on Twitter is retweeted by a significantly larger user base and spreads faster than factual information, especially for political news (Vosoughi et al., 2018).

To mitigate the impact of fake news on society, several efforts revolving around fake news detection have been developed. They include manual fact-checking which relies on domain-experts or regular individuals as fact-checkers to check the veracity of news content (Zhou & Zafarani, 2020). However, this method is time-consuming and lacks scalability due to the large volume of fake news.

Automatic fact-checking offers a promising solution, leveraging Natural Language Processing (NLP) and Machine Learning (ML) to achieve scalability and cost-effectiveness (Chorąś et al., 2021). Proposed methods include NLP techniques like Term Frequency-Inverse Document Frequency (TF-IDF), network data analysis based on Network theory (Zhou & Zafarani, 2019) and reputation analysis employing neural network trained on Domain Name System (DNS) databases to detect malicious IP addresses (Lison & Mavroeidis, 2017). For the project, we will use NLP techniques and apply ML models to detect fake statements from the LIAR dataset.

1.2 Objective

With the rampant spread of misinformation and its detrimental effects, this paper aims to develop an effective fake news classifier to combat this issue. We contribute to the field by conducting a comprehensive survey of traditional and modern ML methods. Each technique is assessed using key evaluation metrics: Precision, Recall, F1-Score and Accuracy.

Secondly, this study applies interpretable ML to provide an additional layer of understanding into the decision-making process of how fake news is classified. This helps to foster greater transparency and trust in the system.

2. Dataset Description

2.1 Data Source

The LIAR dataset from Wang (2017) is a publicly available dataset for fake news detection tasks. It contains 12,800 human-labelled short statements across various contexts collected from POLITIFACT.COM, a fact-checking website. Statements are collected over a decade from 2007 – 2016 and are 17.9 tokens long on average.

2.2 Data Description

The data is available as train, validation and test sets through an 80-10-10 split. There are 6 classes describing the severity of the lie. From descending intensity, they are *pants-fire*, *false*, *barely-true*, *half-true*, *mostly-true* and *true*. The class distributions on the train set are as follows:

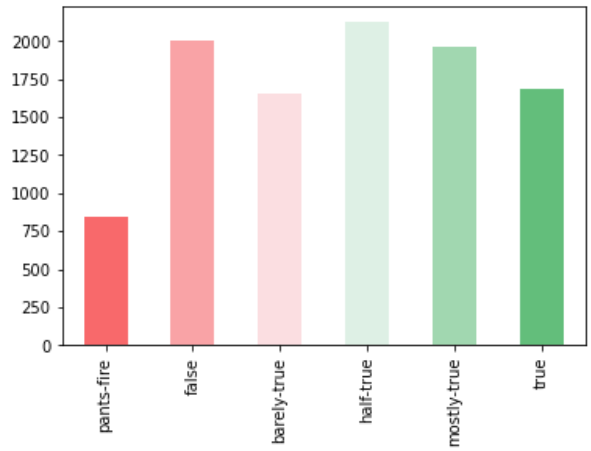

Figure 1: Class Distribution of Train Data

2.3 Data Preprocessing

The most blatant lie, *pants-fire*, is the least represented in the data. We aggregate the classes together to increase class balance and reduce the granularity of the classes for easier interpretation

Table 1: Class Aggregation Rule

Original Class	Grouped Class
pants-fire false	pants-fire
barely-true half-true	barely-true
mostly-true true	true

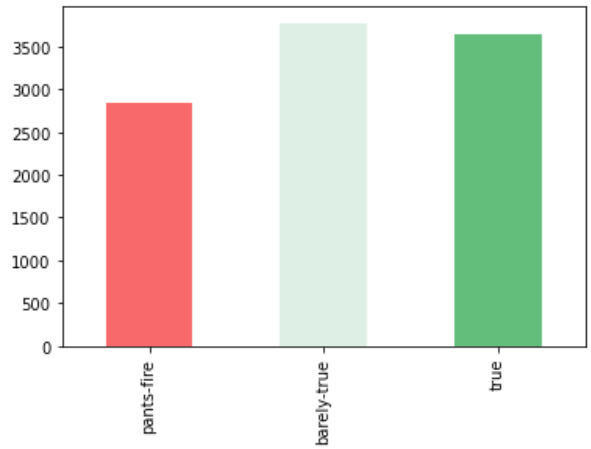

Figure 2: Class Distribution of Aggregated Train Data

Post-aggregation, the class balance has improved. Fake news now accounts for 27.7%, barely-true news accounts 36.8% while true news accounts 35.5% of the train data.

Some random snippets of the data are shown below:

Table 2: Snippets from the LIAR dataset

Statement	Label
Before World War II, very few people actually had health insurance.	True
The United States has a low voter turnout rate.	
Taxpayers subsidize 80 percent of each MARTA trip	Barely-true
Hillary Clinton said gun confiscation would be worth considering.	
Birth control pioneer Margaret Sanger was an active participant in the Ku Klux Klan.	Pants-fire
Many of the founding fathers were very actively involved in cockfighting.	

The test set was found to have lower character counts. A comparison of word counts between the train and test dataset is shown below:

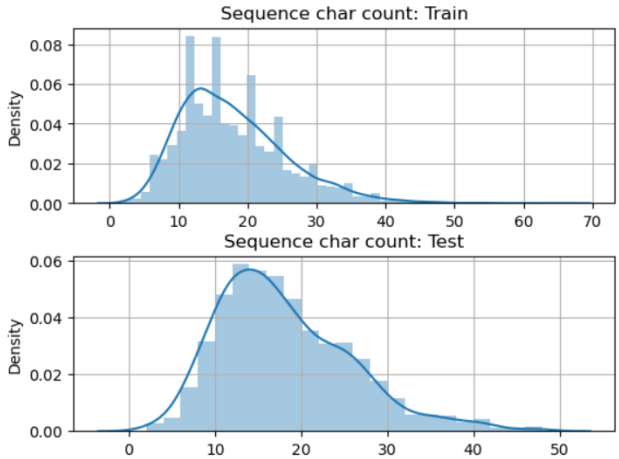

Figure 3: Distribution of Character Counts

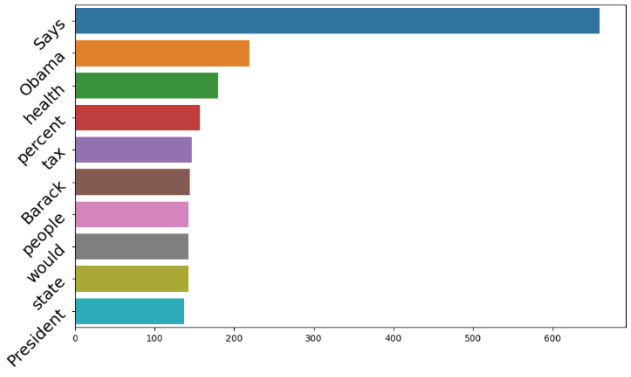

Figure 4: Plot of the Top 20 Most Frequent Words for *Pants-fire*

From Figure 4, the word 'Says' appears the most frequent. This will be handled during the text preprocessing step via stop word removal.

3. Machine Learning Methods

3.1 Baseline Bag-of-Words (BoW) and TF-IDF

The bag-of-words (BOW) model introduced by Luhn (1957) is a text representation technique used in machine learning, transforming textual data into fixed-length numerical vectors. It disregards word order and structure, focusing solely on the occurrence of known words in a document. This approach creates a vocabulary of words and measures their presence within the document, enabling algorithms to process and analyze textual data efficiently.

TF-IDF, introduced by Sparck (1972), is a numerical measure used in information retrieval to quantify a word's significance within a document relative to a corpus. It computes word frequency in a document against its frequency across the corpus, supporting tasks in search, text mining, and user modeling. The value increases with the word's frequency in the document but is dampened by its prevalence across the corpus, compensating for frequently occurring words.

The two traditional NLP techniques are implemented as a baseline comparison against more sophisticated models subsequently. We used Python's Natural Language Toolkit (NLTK) and Punkt tokenizer for sentence splitting and English stopwords. Our preprocessing involved converting texts to lowercase for standardization, removing numbers and punctuation to focus on linguistic elements, removing whitespace to ensure text consistency, and eliminating common English stopwords to emphasize meaningful words. Stemming was done using PorterStemmer to reduce words to their base form.

We configured BoW and TF-IDF with a maximum document frequency of 0.95, filtering out extremely common words that offer limited classification value. Tokenisation included both unigrams and bigrams to capture some degree of context in the text data. The *max_feature* parameter was set to 1,000 to reduce dimensionality of the text representation and lower the risk of overfitting.

3.1.1 LASSO

Logistic Regression with L1 regularization (LASSO) was used to deal with high dimensionality given the vast vocabulary inherent in the textual data. Introducing a penalty term allows the model to capture language patterns while ensuring that only the most relevant features are retained in the model, effectively reducing overfitting.

3.2 Word Embeddings (GloVe)

Word embeddings are compact numerical representations of words or phrases, aimed at mapping them from raw text to a low-dimensional space (Li & Yang, 2018). They facilitate various natural language processing tasks by capturing contextual meaning and structural roles, as introduced by Hinton (1986) and based on the semantic similarity hypothesis by Harris (1954).

We utilized Global Vectors (GloVe), an unsupervised algorithm that constructs embeddings by analyzing global word co-occurrence statistics from a text corpus (Pennington et al., 2014). This enables our models to capture more nuanced word relationships and be effective for tasks like fake news detection.

The chosen GloVe model was `glove.6B.100d`, trained on 6 billion words and outputs 100-dimensional vectors. This dimensionality strikes a balance between capturing detailed semantics and ensuring computational efficiency, making it ideal for sentence-level analysis.

To optimize the performance of our GloVe embeddings, we performed the text preprocessing steps discussed earlier. However, we opted not to use stemming in our preprocessing pipeline to maintain the original form of words, as stemming can often reduce words to their base forms, potentially stripping away nuances in meaning that GloVe embeddings are designed to capture. This approach ensures that the semantic richness and variability of language are preserved, enhancing the contextual relevance of our embeddings.

3.2.1 LASSO

Similar to Section 3.1.1, Logistic Regression with L1 regularization (LASSO) was used to make classifications. To prepare the data, sentence embeddings were created from GloVe by using sentence averaging, where the mean of all word vectors in the sentence forms a single representative vector. This approach preserves essential semantic content in a simplified form.

3.2.2 Convolutional neural networks (CNN)

GLOVE embeddings were integrated with Convolutional Neural Networks (CNNs) to harness local contextual features efficiently. CNNs excel in identifying local patterns within data, crucial for interpreting sentence structures in NLP. Unlike Recurrent Neural Networks (RNNs), CNNs can be parallelized, leading to faster processing times and less susceptibility to the vanishing gradient problem (Kim, 2014).

Initial data analysis involved exploring text length distributions to determine a uniform input size. The texts were tokenized and padded to ensure consistency across input data, preparing them for model ingestion. The tokenizer identified unique words to set a vocabulary limit, and an embedding matrix was created where each word index was mapped to a 100-dimensional GloVe vector. Words absent in GloVe were assigned zero vectors.

The CNN architecture comprised an embedding layer initialized with the GloVe matrix, fixed to preserve pre-trained semantic properties, followed by alternating convolutional and max pooling layers to extract text features efficiently. The model included dense layers with dropout for regularization, along with early stopping to prevent overfitting. Training involved 15 epochs, the Adam optimizer and categorical cross entropy to enhance training efficiency and model generalizability.

3.3 Transformers (DistilBERT)

Transformers, developed by Vaswani et al. (2017) are a novel network architecture based on attention mechanisms that achieved state-of-the-art performance upon release. Attention mechanisms allow transformers to weigh the importance of words in sequences and capture intricate linguistic dependencies effectively. They are also parallelizable, reducing training time significantly.

Subsequently, Devlin et al. (2018) introduced Bidirectional Encoder Representations from Transformers (BERT), improving fine-tuning through masked language modelling. This allowed BERT to consider both preceding and following words around an anchor word, rather than solely left-to-right (unidirectional).

We fine-tuned the *distilbert-base-uncased* model from Sanh et al. (2019), which has 40% less parameters while preserving 95% of BERT's performance. The model tokenizer has a vocabulary size of 30,522 tokens and a maximum input size of 512 tokens. We tokenize and pad all input text to a length of 512 – if the original text is less than 512, it is padded with zero. Further text preprocessing (stop-word removal, stemming, lowercase, punctuation etc.) is omitted as transformer models require full sentence context for accurate embeddings.

During fine-tuning, only the last two layers (*pre-classifier* and *classifier*) are trained, while all other parameters are frozen. To expedite fine-tuning, we only train on 1,284 randomly sampled instances from the training data. The model undergoes 20 epochs of training at a low learning rate (5e-5), and the epoch with the best validation Macro F1-Score is selected.

3.4 Sentence Transformers (all-MiniLM-L6-v2)

Sentence Transformers are a framework to compute semantically meaningful embeddings at the sentence level – similar sentences will lie close together on the embedding vector space. Reimers & Gurevych (2019) laid the initial work by introducing Sentence-BERT (SBERT), a modified version of BERT using siamese and triplet networks which could perform semantic textual similarity (STS) without computational overhead compared to BERT.

We leverage a pre-trained sentence transformer, *all-MiniLM-L6-v2*, to compute 384-dimensional embeddings for each sentence. Notably, *all-MiniLM-L6-v2* is a finetuned variant of *MiniLM* from the works of Wang et al. (2020). The vectors are then fed into a LASSO model to deal with high-dimensionality and make predictions. Similar to the DistilBERT model, text preprocessing is omitted to pass full sentence context into the model.

4. Model Evaluation and Discussion

Our evaluation of multiple machine learning models for fake news detection covers a range of traditional NLP methods and advanced deep learning techniques. The evaluation metrics focus on Macro Precision, Macro Recall, Macro F1Score, and Accuracy for a multi-class classification task across all classes of news severity. Additionally, we specifically analyzed the performance on the "pants-fire" category, which represents the most egregious form of misinformation.

In Table 3, the all-MiniLM-L6-v2 model outperformed others in all metrics of multi-class fake news detection across varying severities of misinformation. Traditional approaches (i.e. Bag of Words and TF-IDF) underperformed, while word embedding-based models (i.e. GloVe with LASSO regularization) showed considerable improvement. However, the GloVe-CNN model surprisingly underperformed even traditional methods, likely due to the shallow CNN architecture used, which may not have effectively captured the sequential patterns in the text data.

Transformer-based architectures (i.e. DistilBERT and sentence—transformers) exhibited superior ability to discern nuanced and deceptive content, crucial for identifying severe misinformation categories like "pants-fire." This highlights the potential for sentence transformers in complex NLP tasks like fake news detection, underscoring the need for sophisticated, context-aware models in operational settings.

Table 3: Multi-class Evaluation Metrics on Test Set

Model	Macro Precision (%)	Macro Recall (%)	Macro F1Score (%)	Accuracy (%)
BoW	40.4	40.2	40.1	41.3
TF-IDF	40.6	40.4	39.9	42.2
GloVe + LASSO	43.5	42.2	41.9	44.0
GloVe + CNN	40.8	39.1	38.2	40.4
DistilBERT	42.5	41.9	42.1	42.7
all-MiniLM-L6-v2	**43.6**	**43.1**	**43.0**	**44.4**

Next, we evaluated the models' performance on identifying "pants-fire" fake news, detailed in Table 4. Achieving high recall is essential for detecting the majority of fake news articles, reducing the damage of unchecked misinformation spread and preserving public trust and safety.

DistilBERT demonstrates exceptional performance, achieving both the best recall rate and F1 score. Its effectiveness in detecting severe misinformation instances makes it invaluable in high-stakes contexts (e.g. COVID-19) where a single undetected fake news article could have detrimental effects on public discourse. The high F1 score optimizes the trade-off between false positives and false negatives, thereby minimizing the resources required for manual verification of false positives – a less critical but still significant operational consideration.

Of particular interest is the high precision of the GloVE-LASSO model, which means that a news article classified as "pants-fire" is more likely to be fake, even compared to the DistilBERT model. However, in fake news detection, prioritizing recall is critical since failing to identify and control the spread of false information can lead to significant societal repercussions. High recall ensures that most fake news is flagged, reducing risks such as distorted public perceptions or manipulated election outcomes. Although low precision necessitates additional manual labor verification which is resource-intensive, the cost is outweighed by the benefits of preventing unchecked proliferation of fake news. Businesses optimizing for high recall in their detection systems not only bolster their credibility but also play a pivotal role in safeguarding public discourse, marking it as a crucial strategy in the digital information age.

Table 4: Evaluation Metrics for pants-fire on Test Set

Model	Precision (%)	Recall (%)	F1Score (%)
BoW	35.0	28.7	31.5
TF-IDF	34.3	21.6	26.5
GloVe + LASSO	**41.6**	24.0	30.4
GloVe + CNN	36.5	29.2	32.5
DistilBERT	39.1	**33.9**	**36.3**
all-MiniLM-L6-v2	39.6	28.9	33.4

5. Model Explainability

Model explainability is crucial for ensuring trust and accountability. Deep learning models, known for their opacity, can generate predictions that are difficult to interpret. This opacity may erode user trust, as individuals are the ultimate arbiters of truth in the news they consume. Without understanding the rationale behind news article classifications, users may hesitate to trust the model's output, regardless of its accuracy.

Model explanations can also reveal valuable insights to scrutinize the fit and fairness of classifiers. Certain data artifacts can sometimes induce an undesirable correlation that classifiers use for prediction (Ribeiro et al. 2016). Caliskan et al. (2017) demonstrated that models trained on word associations can inherit human prejudices and cultural associations. Specifically, if our classifier assigns undue importance to a person's name or political affiliation instead of the veracity of news claims, this may undermine model credibility.

Mishima & Yamana (2022) present a categorization of methods to explain a misinformation detection system

- Explanations by Social Features, which exploits user comments on social media
- Explanations by Feature Importance, which includes intrinsic interpretability or post-hoc methods
- Explanations by News Content, which can be applied to models having an attention mechanism
- Explanations by Knowledge Bases, which relies on fact-checking against known truths and lies

We focused on Feature Importance (BoW and TF-IDF) where the LASSO model has intrinsic explainability on word tokens, and News Content (DistilBERT) where the model uses attention mechanisms to make predictions. GloVe and Sentence Transformer embeddings are not explained as they are complex representations of linguistics more suitable for STS and clustering purposes.

5.1 Baseline Bag-of-Words (BoW) and TF-IDF

Table 5: Top 10 Feature Importance using BoW technique

Feature	Coefficient
tax break	1.13
muslim	1.11
protest	0.93
everybodi	0.90
statist	0.88
income tax	0.88
realli	0.84
rep	0.84
tax increas	0.83
sell	0.82

Table 6: Top 10 Feature Importance using TF-IDF technique

Feature	Coefficient
muslim	2.18
rep	1.86
obama	1.65
tax increas	1.61
protest	1.57
illeg	1.52
scott walker	1.46
doctor	1.44
care law	1.37
talk	1.35

The two tables above indicate the most important word tokens or n-grams most strongly associated with statements labeled as Pants-fire. Common word tokens or n-grams

such as *muslim*, *rep*, *protest* and *tax increase* were found across the two tables and suggest that statements containing these words are more likely to be classified as *Pants-fire* by the model. This provides transparency into how the model decides which statements are *Pants-fire and* aid in interpretability.

5.2 Transformers (DistilBERT)

We utilise Integrated Gradients (IG) from Sundararajan et al. (2017) to attribute deep network predictions to their input features. IG linearly interpolates the word embeddings between a zero-paddings baseline (absence of text) and the input text (presence of text). A forward pass is done for each interpolation and the gradients of the output probability distribution are computed with respect to input embeddings. Lastly, the gradients are integrated over the full linear interpolation path, creating attribution scores highlighting the specific input text components driving classification decisions.

We compute attribution scores for samples of correctly predicted fake news and real news. Unlike BoW and TF-IDF, IG provides local explainability individual news pieces, offering a flexible framework to highlight veracious content for any piece of news.

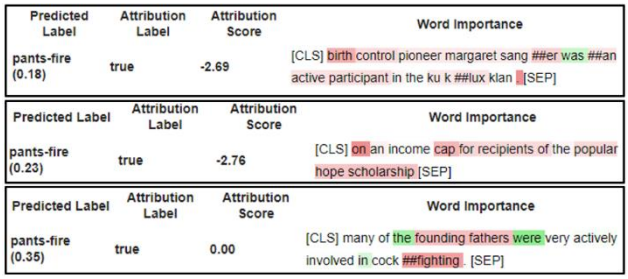

Figure 5: Attribution of words in fake news

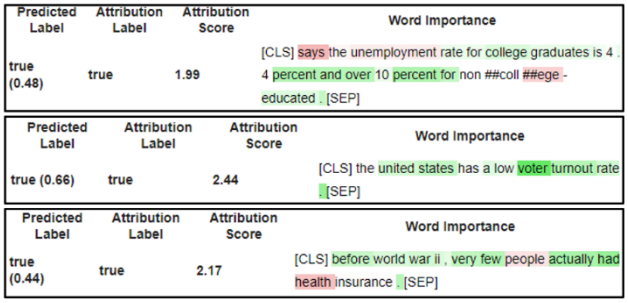

Figure 6: Attribution of words in real news

6. Conclusion

This study presents a comprehensive exploration of machine learning (ML) methodologies aimed at detecting and mitigating misinformation. Through rigorous evaluation, we have demonstrated the capabilities and limitations of various Natural Language Processing (NLP) techniques in identifying misinformation across multiple severities, with a particular focus on the most deceptive 'pants-fire' category.

We demonstrated that while traditional ML methods such as Bag-of-Words and TF-IDF, provide a solid baseline, embedding and transformer-based models like DistilBERT and all-MiniLM-L6-v2, excel in capturing the subtleties of deceptive information. This suggests a shift towards more sophisticated, context-aware techniques in operational environments where the detection of misinformation is crucial.

Additionally, we explored various methodologies for model interpretability to shed light on the inner workings of our classifiers, fostering trust and transparency in their outputs. Token importance for traditional methods (BoW, TF-IDF) provided insights on global model importance, while the application of Integrated Gradients on DistilBERT directly revealed the attention on news content contributing toward a local fake or real news classification.

For future work, we recommend exploring the integration of multimodal data sources, such as combining text with metadata or user engagement metrics, to enhance the detection capabilities further. Additionally, investigating the resilience of these models against adversarial attacks or evolving misinformation tactics will be critical in maintaining their effectiveness.

Our research contributes to the critical discourse on digital information integrity, highlighting the potential of ML to combat misinformation effectively. By advancing these technologies, we can better safeguard public discourse and contribute to a more informed and discerning society.

References

Best, S. (2020, March 3). Coronavirus hoax claims '5G causes virus' by 'sucking oxygen out of your lungs'. Mirror. https://www.mirror.co.uk/tech/coronavirus-hoax-claims-5g-causes-21620766

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186.

Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., ... & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing, 101, 107050.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162

Hinton, G. E. (1990). Preface to the special issue on connectionist symbol processing. Artificial Intelligence, 46(1-2), 1-6.

Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751. https://www.aclweb.org/anthology/D14-1181/

Lazer, D. M., 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.

Li, Y., & Yang, T. (2018). Word embedding for understanding natural language: a survey. Guide to big data applications, 83-104.

Lison, P., & Mavroeidis, V. (2017). Neural reputation models learned from passive DNS data. 2017 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2017.8258361

Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. IBM Journal of Research and Development, 1(4), 309–317.

Mishima, K., & Yamana, H. (2022). A survey on explainable fake news detection. IEICE TRANSACTIONS on Information and Systems, 105(7), 1249-1257.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532-1543. https://nlp.stanford.edu/pubs/glove.pdf

Reimers, N., & Gurevych, I. (2019, November). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 3982-3992).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

Silverman, C. (2016, November 16). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Sparck Jones, K. (1972). A Statistical Interpretation Of Term Specificity and Its Application in Retrieval. Journal of Documentation, 28(1), 11–21. https://doi.org/10.1108/eb026526

Sundararajan, M., Taly, A., & Yan, Q. (2017, July). Axiomatic attribution for deep networks. In International conference on machine learning (pp. 3319-3328). PMLR.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. science, 359(6380), 1146-1151.

Wang, W. Y. (2017, July). “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 422-426).

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Advances in Neural Information Processing Systems, 33, 5776-5788.

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.