Identifying Deceptive Content: A Study on Clickbait and Fake News Detection

Shashank Rangarajan, Chia-Yu Tung, Rishabh Ghosh, Michael Guastalla, Edwin Wang
Department of Computer Science, University of Southern California
{sr87317, ctung, ghoshris, guastall, kuanchun}@usc.edu

Abstract

The proliferation of fake news and clickbait articles on social media platforms has become a major concern for individuals, organizations, and society as a whole. In this article, we study the relationship between clickbait and fake news detection tasks by implementing state-of-the-art models for both of the tasks and studying their predictions on various datasets. Our findings suggest a positive correlation between models predicting fake news and clickbait. Training models with diverse data from multiple sources can enhance their adaptability and resilience by eliminating bias towards any specific source. Such diversity can render the model more versatile and robust.

1 Introduction

Social media has become a popular platform for sharing views and news due to the availability of digital devices, affordable internet, and free subscriptions. However, studies show that this popularity has given rise to the circulation of fake news and clickbait. As a result, detecting fake news and clickbait on social media has become an urgent and crucial research topic.

This project aims to study the relationship between fake news and clickbait detection tasks in order to see if these tasks are interrelated and if models trained on one task can be adapted to the other. After experimenting with multiple datasets for both of these tasks, we conclude that a model trained for fake news detection can aid in clickbait detection and that the predictions from the models are positively correlated for both fake news and clickbait.

2 Related work

2.1 Clickbait Detection

Clickbait detection gained popularity through datasets and challenges like the Webis Clickbait challenge (Potthast et al., 2018). The challenge winners proposed a two-level classification approach that combined 65 classifiers into a feature vector fed into a second layer classifier, and a recent study (Qiu et al., 2021) introduced the Click-BERT model that improved detection capabilities. Our models, based on BiLSTM, Bi-GRU, and BERT, were fine-tuned on multiple datasets according to the methods described in section 4, inspired by the above research.

2.2 Fake News Detection

Fake news detection gained popularity through challenges and datasets like "r/Fakeddit" (Nakamura et al., 2019), LIAR (Wang, 2017), and Fake News Corpus (Szpakowski, 2017), where the authors of (Nakamura et al., 2019) proposed multimodal models using a dataset of over 1 million samples, emphasizing the importance of novel aspects of multimodality and fine-grained classification unique to Fakeddit. The text-only model from (Nakamura et al., 2019) served as an inspiration for our fake news models, which are discussed in section 4.

2.3 Relevance between Clickbait and Fake News Detection

Although fake news and clickbait detection appear similar, few studies have explored their relationship. One related work, (Lee et al., 2021), proposed a multi-task framework that trained a misinformation detection model to detect news bias, clickbait, and rumors. We were inspired by this and plan to study the relationship between fake news and clickbait detection models, as discussed in section 3.

3 Problem Description

Our literature survey revealed that fake news detection and clickbait detection share similar characteristics, suggesting a possible overlap. However, the available datasets treat them as mutually exclusive. To explore their relationship, our project aims

to break down our high-level goals into specific objectives.

- Implement existing state-of-the-art models for both clickbait and fake news detection.
- Analyze the performance of these models on various publicly available datasets
- Examine the predictions from these models for the correlation between fake news and clickbait detection

4 Methodology

4.1 Datasets

Following were the datasets we used:

1. **Clickbait Detection:**
   (a) *Webis Clickbait 17 (Potthast et al., 2018)*
   (b) *Clickbait-Detector (Mathur, 2017)*
   (c) *Fake News Corpus (FNC) (Szpakowski, 2017)*
2. **Fake News Detection:**
   (a) *Fakeddit (Nakamura et al., 2019)*
   (b) *LIAR (Wang, 2017)*
   (c) *Fake News Corpus (FNC) (Szpakowski, 2017)*

4.2 Data preprocessing

For our LSTM and GRU-based models (section 4.3) we performed data preprocessing steps such as tokenization, lemmatization, stop-word-removal, etc. However we soon noted that pre-processing methods like these cannot be easily adapted to datasets across the clickbait and fake news detection tasks, hence we decided to perform minimal preprocessing and rely on pre-trained BERT models to provide contextual embedding in our datasets.

4.3 Model Architectures

We implemented many model architectures as follows:

- *Bi-LSTM with GloVe*: Following (Papadopoulou et al., 2017), we implemented a Bi-LSTM model with GloVe embedding as our baseline for clickbait detection in the initial phase.
- *Bi-GRU with GloVe*: Similar to the Bi-LSTM model, we tried a Bidirectional GRU model instead.
- *Siamese-BERT with FC*: Inspired from (Kolla, 2019) we implemented a Siamese-BERT for the LIAR dataset, where the branches fed content and metadata separately into RoBERTa model. The outputs from RoBERTa model were averaged, concatenated, and fed into a fully-connected layer followed by a 2-way classifier.
- *RoBERTa with FC*: Inspired from the text-only model in (Nakamura et al., 2019), our architecture fed the average of RoBERTa embeddings into a fully-connected layer of 300 relus with a dropout which fed into a 2-way classifier.

5 Experimental Results

5.1 Clickbait Detection Experiments

Initial experiments on clickbait detection involved creating LSTM and GRU models, and evaluating them on the kaggle dataset (Anand, 2020) and Clickbait-detector dataset (Mathur, 2017). To avoid the burden of data preprocessing (section 4.2), we utilized BERT-based embeddings for our model input and developed the RoBERTa + FC model. Additionally, to address dataset diversity, we merged the clickbait detection datasets in section 4.1 and trained the model on this combined dataset.
In Table 1, our models' performance on different datasets is presented. Notably, the combined model exhibits generalizability, while individual models show dataset-specific bias, despite being slightly outperformed by individual models on their respective datasets.

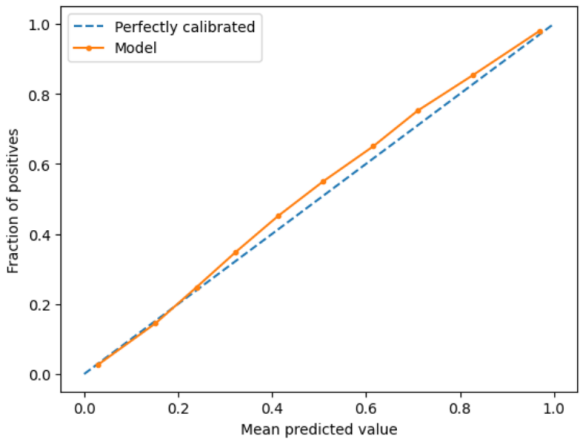


Figure 1: Calibration curve of Clickbait Model

Figure 1 displays the calibration graph of the combined clickbait model. The ECE score of 0.02, representing the average deviation from perfect calibration, indicates the model is well-calibrated. Therefore, the predicted logits align well with the true probability of a data point having a specific label.

Dataset	Mathur's Model			Webis model			FNC Model			Combined Model		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Mathur's	0.92	0.91	0.92	0.75	0.68	0.74	0.69	0.72	0.69	0.82	0.81	0.82
Webis 2017 Clickbait	0.53	0.54	0.61	0.86	0.67	0.77	0.54	0.35	0.53	0.69	0.52	0.68
FNC	0.56	0.46	0.56	0.48	0.27	0.48	0.89	0.89	0.89	0.88	0.88	0.88
Combined	0.59	0.52	0.58	0.55	0.36	0.54	0.84	0.83	0.84	0.86	0.85	0.86

Table 1: Performance of various clickbait models across multiple datasets.

Dataset	Fakeddit Model			FNC Model			Combined Model		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Fakeddit	0.86	0.82	0.85	0.54	0.36	0.50	0.84	0.79	0.83
FNC	0.52	0.52	0.52	0.86	0.86	0.86	0.85	0.84	0.85
Combined	0.68	0.65	0.68	0.71	0.66	0.70	0.83	0.82	0.83

Table 2: Performance of various fake news models across multiple datasets.

5.2 Fake News Detection Experiments

Our initial experiments with the RoBERTa + FC architecture inspired from (Nakamura et al., 2019) showed promising results across the datasets. When setting up the experiment, we combined datasets for fake news detection as we did for clickbait (section 4.1) to create a combined dataset that accounts for diversity in our data.

Table 2 shows the performance of our models on the fake news detection task. Similar to clickbait detection we see that the combined model, again, showcases good generalizability across the datasets.

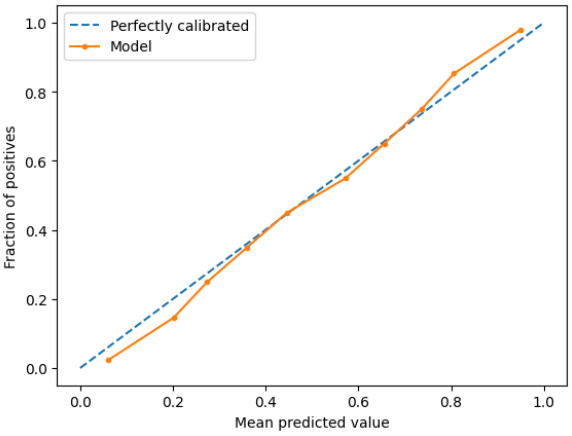


Figure 2: Calibration curve of Fake News Model

Figure 2 shows the calibration graph for the combined fake news model. We measured an ECE score of 0.025 for this graph. While this score is slightly worse than the one we obtained for clickbait, it still performs well on this metric when compared to many modern neural networks as discussed in (Guo et al., 2017).

Apart from this, we also experimented with the Siamese-BERT model on the LIAR dataset (Wang, 2017), and we saw poor performance due to the small dataset size. However, we noticed a performance boost of 10% when the RoBERTa + FC model pre-trained on fakeddit dataset was fine-tuned on LIAR dataset. It is interesting that external data could aid in the better detection of fake news for the LIAR dataset.

5.3 Exploring relationship between Clickbait and Fake News Detection

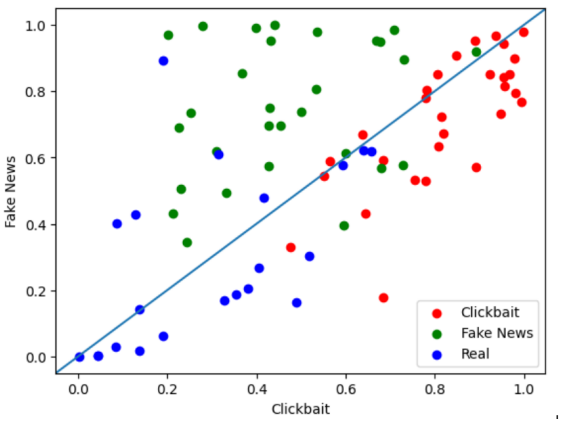


Figure 3: Relationship between Clickbait and Fake News data points.

To investigate the relationship between fake news and clickbait, we first extracted 75 samples from the Fake News Corpus (Szpakowski, 2017) (made up of 25 clickbait, 25 real, and 25 fake news

examples). Next, we fed these inputs to both of our combined models.

In Figure 3 we plot the `%fakeness` v/s `%clickbait` from the predictions of the combined models.

Our analysis of Figure 3 indicates that the `%clickbait` is positively correlated to `%fakeness`, in that both the models predict clickbait samples correctly while we see that the fake news samples are misclassified by the clickbait model and therefore predicts less `%clickbait` for fake news samples. This can also be seen in Table 3.

Task	ACC	F1	AUC
ClickBait dataset (in FakeNews Model)	0.82	0.78	0.81
FakeNews dataset (in ClickBait Model)	0.56	0.56	0.59

Table 3: Performance of clickbait and fake news models on cross datasets.

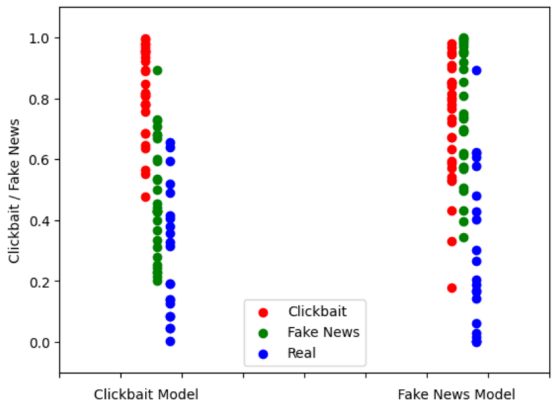
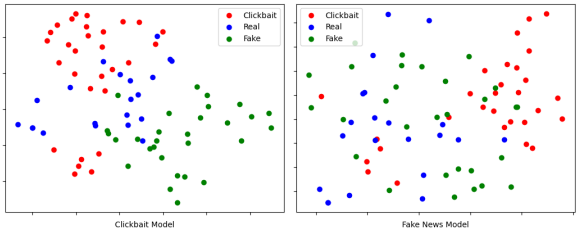


Figure 4: The correlation between Clickbait and Fake News is demonstrated by the data points on both the clickbait model and fake news model.

Both the Figure 4 and Table 3 indicate that the fake news model was successful in classifying clickbait as fake news, while the clickbait model struggled to distinguish the realness of fake news. These findings support the intuition that fake news often uses exaggerated headlines similar to those of clickbait, but clickbait typically contains truthful information.

5.4 Visualization of data points

We visualized the data points by extracting the 300 hidden dimensions from our models and applying t-SNE to plot them in a 2D space. Figure 5a and Figure 5b show the values of clickbait, fake news, and real data points from the clickbait model and fake news model respectively. We see that for the clickbait model, the fake-news samples are far away from the clickbait samples, however for the fake news model, the clickbait samples are interspersed between fake news samples, this can be a likely reason for the performance showcased in Table 3.



(a) (b)

Figure 5: t-SNE visualizations for Clickbait Model(a) and Fake News Model (b)

6 Conclusions and future work

We found that clickbait and fake news share similar characteristics. Our fake news model performs well for clickbait datasets, but the reverse is not always true, possibly due to the syntactic features that clickbait relies on, which are easier to learn than the semantic features required for fake news, which holds truthfulness. Our analysis also revealed issues with the datasets, such as untrustworthy sources reposting genuine articles labeled as fake and short titles lacking meaningful information. We also found that combining multiple datasets improves the robustness of fake news detection and clickbait detection, which highlights the importance of diversity in the training data. Our findings suggest a positive correlation between models predicting fake news and clickbait, emphasizing the need to train models with diverse data from multiple sources to enhance their adaptability and resilience by eliminating bias towards any specific source.

To improve fake news and clickbait detection, future studies can explore multimodal approaches using images and ensemble methods. While our study focused on titles/headlines across datasets, we did not pursue this approach. Additionally, GNNs can capture complex relationships and interactions in a network, including between news articles, sources, authors, and other entities, to predict the veracity of news. GNNs can be combined with other machine learning models and features, such as content analysis and source credibility, for greater accuracy. AI-generated datasets may also prove useful as misinformation evolves over time.

A Division of Labour

Task **Description**	**Member** (last name)
Implement state-of-the-art clickbait detection model	Tung, Ghosh
Implement state-of-the-art fake news detection model	Rangarajan, Guastalla
Perform analysis of models on different datasets	Tung, Rangarajan, Guastalla, Ghosh, Wang
Explore models that make fine-grained predictions	Rangarajan, Ghosh, Guastalla
Final paper presentation	Rangarajan, Ghosh, Guastalla
Github: Identifying Deceptive Content	Tung, Rangarajan, Guastalla, Ghosh, Wang
Final report	Tung, Rangarajan, Guastalla, Ghosh, Wang

Please find the code here:
https://github.com/rshashank13/identifying-deceptive-content

References

Aman Anand. 2020. Clickbait dataset on kaggle. https://www.kaggle.com/datasets/amananandrai/clickbait-dataset?datasetId=609158.

BuzzFeedNews. 2017. Buzzfeednews. https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data.

Webis Clickbait Challenge Dataset. 2017. clickbait challenge dataset. https://zenodo.org/record/6362726#.YsbdSTVBzrk.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bhanuka Gamage, Adnan Labib, Aisha Joomun, Chern Hong Lim, and KokSheik Wong. 2021a. Baitradar: A multi-model clickbait detection algorithm using deep learning. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2665–2669.

Bhanuka Gamage, Adnan Labib, Aisha Joomun, Chern Hong Lim, and KokSheik Wong. 2021b. Baitradar: A multi-model clickbait detection algorithm using deep learning.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On calibration of modern neural networks.

Manideep Kolla. 2019. Triple branch bert siamese network for fake news classification on liar-plus dataset. https://github.com/manideep2510/siamese-BERT-fake-news-detection-LIAR.

Nayeon Lee, Belinda Z. Li, Sinong Wang, Pascale Fung, Hao Ma, Wen tau Yih, and Madian Khabsa†. 2021. On unifying misinformation detection.

Saurabh Mathur. 2017. clickbait-detector. https://github.com/saurabhmathur96/clickbait-detector/tree/master/data.

Gilbert E Mitra, T. 2015. Credbank: A large-scale social media corpus with associated credibility annotations. https://github.com/compsocial/CREDBANK-data.

Kai Nakamura, Sharon Levy, and William Yang Wang. 2019. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection.

Olga Papadopoulou, Markos Zampoglou, Symeon Papadopoulos, and Ioannis Kompatsiaris. 2017. A two-level classification approach for detecting clickbait posts using text-based features.

Martin Potthast, Tim Gollub, Matthias Hagen, and Benno Stein. 2018. The clickbait challenge 2017: Towards a regression model for clickbait strength.

Changyuan Qiu, Chenshu Zhu, and Haoxuan Shan. 2021. Click-bert. https://github.com/PeterQiu0516/Click-BERT.

Shashank Rangarajan. 2023. identifying-deceptive-content. https://github.com/rshashank13/identifying-deceptive-content.

Benjamin Riedel, Isabelle Augenstein, Georgios Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the fake news challenge stance detection task.

Maciej Szpakowski. 2017. Fake news corpus. https://github.com/several27/FakeNewsCorpus.

William Yang Wang. 2017. "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)*, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

Tianyi Xie, Thai Le, and Dongwon Lee. 2021a. Checker: Detecting clickbait thumbnails with weak supervision and co-teaching. In *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pages 415–430, Germany. Springer Science and Business Media Deutschland GmbH. Funding Information: Acknowledgement. The works of Thai Le and Dongwon Lee were in part supported by NSF awards 1742702, 1820609, 1909702, 1915801, 1934782, and 2114824. Publisher Copyright: © 2021, Springer Nature Switzerland AG.; European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD 2021 ; Conference date: 13-09-2021 Through 17-09-2021.

Tianyi Xie, Thai Le, and Dongwon Lee. 2021b. Checker: Detecting clickbait thumbnails with weak supervision and co-teaching.

Savvas Zannettou, Sotirios Chatzis, Kostantinos Papadamou, and Michael Sirivianos. 2018a. The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 63–69.

Savvas Zannettou, Sotirios Chatzis, Kostantinos Papadamou, and Michael Sirivianos. 2018b. The good, the bad and the bait: Detecting and characterizing clickbait on youtube.