TechWhiz@DravidianLangTech 2024: Fake News Detection Using Deep Learning Models

Madhumitha M, Kunguma Akshatra M, Tejashri J, C Jerin Mahibha
Meenakshi Sundararajan Engineering College, Chennai, India
madhumithamurthy2002@gmail.com, kungumaakshatra@gmail.com,
tejashrijayachandran@gmail.com, jerinmahibha@gmail.com

Abstract

The ever-evolving landscape of online social media has initiated a transformative phase in communication, presenting unprecedented opportunities alongside inherent challenges. The pervasive issue of false information, commonly termed fake news, has emerged as a significant concern within these dynamic platforms. This study delves into the domain of Fake News Detection, with a specific focus on Malayalam. Utilizing advanced transformer models like mBERT, ALBERT, and XMLRoBERTa, our research proficiently classifies social media text into original or fake categories. Notably, our proposed model achieved commendable results, securing a rank of 3 in Task 1 with macro F1 scores of 0.84 using mBERT, 0.56 using ALBERT, and 0.84 using XMLRoBERTa. In Task 2, the XMLRoBERTa model excelled with a rank of 12, attaining a macro F1 score of 0.21, while mBERT and BERT achieved scores of 0.16 and 0.11, respectively. This research aims to develop robust systems capable of discerning authentic from deceptive content, a crucial endeavor in maintaining information reliability on social media platforms amid the rampant spread of misinformation.

1 Introduction

Navigating the digital realm, social media emerges as a pivotal force in disseminating information, presenting both opportunities and challenges. The surge in fake news instigated the initiation of the Fake News Detection in Dravidian Languages task(Subramanian et al., 2023), slated for presentation at DravidianLangTech@EACL in 2024 [1]. Task 1 plays a crucial role, concentrating on categorizing social media text to meticulously distinguish between original and fake news. The challenge lies in discerning genuine content from misleading information. Task 2 shifts focus to detecting fake news in Malayalam-language articles, aiming to develop models classifying misinformation into five distinct types. Beyond bolstering natural language processing capabilities, the initiative actively contributes to fostering a more informed and reliable digital environment. This effort not only upholds information integrity but also reinforces the pillars of a trustworthy digital space, marking a beacon for responsible technological advancements and ethical digital communication practices.

[1]https://sites.google.com/view/dravidianlangtech-2024

2 Related Works

Different deep learning models, including CNNs, LSTMs, ensembles, and attention mechanisms had been employed for fake news detection by kum (2020). The CNN + bidirectional LSTM ensembled network with attention had achieved the highest accuracy. Jwa et al. (2019) had proposed "exbake", a model for detecting fake news that uses Bidirectional Encoder Representations from Transformers (BERT). The model had outperformed previous models in terms of F1-score by effectively analysing the relationship between news headlines and body text. Raza and Ding (2022) had addressed early detection challenges and labelled data scarcity in fake news. The proposed transformer-based framework had taken into account both news content and social contexts, resulting in greater accuracy shortly after news dissemination. They had emphasised the significance of incorporating multiple features for better classification. Schütz et al. (2021) had made a contribution to the detection of disinformation by presenting a content-based classification approach based on pre-trained transformer models. Their experiments with models such as XLNet, BERT, and RoBERTa had yielded promising results, highlighting the effectiveness of transformers in achieving high accuracy even with small datasets. The Transformer-based fake news detection framework proposed by Raza and Ding (2022) had demonstrated higher accuracy in early detec-

Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 200–204
March 22, 2024 ©2024 Association for Computational Linguistics

Category	Training Dataset	Evaluation Dataset
Fake	1,599	1,658
Original	406	409

Table 1: Data Distribution for task 1

Category	Training Dataset
Half true	141
Mosty false	239
Mostly true	1
Partly false	42
False	1,246

Table 2: Data Distribution for task 2

tion, addressing label shortage challenges with effective features and a novel labelling technique. Application of Cross Lingual model for sentiment classification on Tamil-English code mixed data had been proposed by (Jerin et al., 2021).

Qazi et al. (2020) had employed an attention-based transformer model for social media fake news detection, discovering a 15 percentage improvement with a hybrid CNN model, underscoring the significance of attention mechanisms. Kula et al. (2021) had explored transformer-based neural network models, showcasing effectiveness in fake news detection using precision, F1-score, and recall metrics. Kumar et al. (2021) proposed an XLNet fine-tuning model, outperforming existing models in multi-class and binary-class fake news detection. The KATMF framework by Song et al. (2021), integrating Knowledge augmented transformer, had excelled in multimodal fake news detection on a real-world dataset. TRANSFAKE, a multi-task transformer model introduced by Jing et al. (2021) had outperformed competitors by jointly modeling body content and comments, emphasizing the importance of considering multiple modalities for comprehensive fake news analysis.

3 Data set

The datasets utilized for implementing fake news detection in Task 1 comprised the training, evaluation, and test datasets provided by the shared task organizers (Subramanian et al., 2024). Each instance in the training dataset was labeled to specify whether the text is fake or original. For Task 2, the datasets used for fake news detection included the training and test datasets from the task organizers. Each instance in the Task 2 training dataset was labeled to indicate one of the five fake categories: False, Half True, Mostly False, Partly False, and Mostly True. Table 1 illustrates the data distribution for the training and development datasets in Task 1, while Table 2 presents the distribution for Task 2. In Task 1, the training dataset consisted of 3,257 instances, with 1,599 falling under the fake category and 1,658 under the original category. The Malayalam development dataset comprised 815 instances, including 406 in the fake category and the remainder in the original category, highlighting data imbalance. The Malayalam test dataset contained 1,019 instances for evaluating model predictions. For Task 2, the training dataset encompassed 1,669 instances across five categories: Half True (141 instances), Mostly False (239 instances), Mostly True (1 instance), Partly False (42 instances), and False (1,246 instances), providing a diverse set for model training and evaluation. The Malayalam test dataset for Task 2 included 250 instances for evaluating model predictions.

4 System Description

In Task 1, social media text was classified as original or fake using XLM-RoBERTa, ALBERT, and mBERT. The methodology involved preprocessing data by removing unwanted characters and training models on datasets from the task organizers. XLM-RoBERTa, with the highest development accuracy, was chosen for final predictions. Task 2 addressed the FakeDetect-Malayalam challenge, focusing on five fake news categories. Transformer models (XLM-RoBERTa, mBERT, and BERT) handled data encoding and tokenization. In the absence of a designated development dataset, training data was split for validation. Optimization with AdamW and cross-entropy loss during training led to mBERT achieving the highest accuracy, highlighting the importance of precise classification in streamlined preprocessing, training, and model selection. Figure 1 illustrates the proposed architecture for both tasks.

4.1 XLM-RoBERTa Model

The XLM-RoBERTa model Conneau et al. (2019), developed by Facebook AI, stands as a pioneering advancement in natural language processing (NLP). Tailored for cross-lingual tasks, it exhibits remarkable proficiency in comprehending and processing text across diverse languages. Leveraging extensive

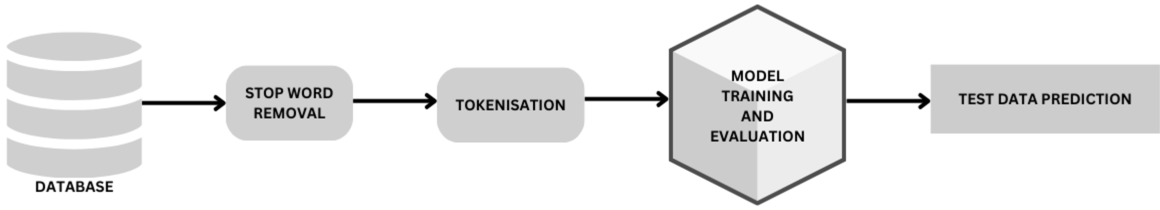

Figure 1: System Architecture

pre-training on multilingual datasets and the transformative potential of the transformer architecture, XLM-RoBERTa is optimized for Malayalam by Raja et al. (2023). This enhancement strengthens fake news detection, emphasizing the model's multilingual capabilities and performance, establishing XLM-RoBERTa as a crucial asset for identifying deceptive content in the digital communication landscape.

4.2 mBERT Model

mBERT (Multilingual BERT) Devlin et al. (2018), a transformer-based language model for multilingual natural language processing, excels in cross-lingual tasks. Inspired by BERT, its architecture, featuring multiple transformer layers and 768 hidden units, is particularly adept at capturing linguistic nuances ("c" model). In his 2022 PhD thesis, "Multi Languages Fake News Detection," Ali (2022) enhances accuracy in identifying and combating fake news across diverse languages using mBERT. This resonates with our focus on Fake News Detection Using Deep Learning Models, underscoring mBERT's multilingual capabilities and effectiveness in addressing specific linguistic nuances.

4.3 ALBERT Model

The ALBERT Model (A Lite BERT) (Lan et al., 2019), a transformative language model by Google Research, uniquely balances efficiency and performance. Employing parameter-sharing techniques, it achieves a compact yet accurate architecture. Widely adopted, Wang et al. (2022) leverage ALBERT for fake news detection, combining its advanced capabilities with multi-modal circulant fusion for robust performance in handling the complexities of identifying deceptive content in textual data.

4.4 BERT Model

The BERT Model Devlin et al. (2018), a transformative force in natural language processing, excels in understanding context and relationships. Developed by Google, BERT's bidirectional approach and attention mechanisms capture nuanced nuances. In BERT model for fake news detection based on social bot activities in the COVID-19 pandemic, Heidari et al. (2021) utilize BERT to scrutinize social bot behaviors, enhancing fake news identification during the pandemic. This emphasizes BERT's versatile application in addressing contemporary challenges related to misinformation.

Model	F1-Score	Accuracy
XLMRoBERTa	0.84	0.84
ALBERT	0.56	0.58
mBERT	0.84	0.84

Table 3: Performance Score for Task 1

Model	F1-Score	Accuracy
XLMRoberta	0.21	0.74
BERT	0.11	0.25
mBERT	0.16	0.63

Table 4: Performance Score for Task 2

5 Result

The evaluation of task performance focused on the macro-F1 score. In Task 1 of DravidianLangTech@EACL 2024, fake news detection assessment relied on macro F1-Score and Accuracy. XLMRoBERTa secured 3rd position on the leaderboard with a notable macro F1-Score of 0.84 and accuracy of 0.84. Despite similar Task 1 scores, XLMRoBERTa's overall excellence, especially in cross-lingual tasks, pre-training, and adaptability, influenced its preference over mBERT. Models like XLMRoBERTa, ALBERT, and mBERT made unique contributions, as depicted in Table 3. In Task 2, XLMRoBERTa maintained superiority, securing 11th position on the leaderboard with a macro F1-Score of 0.21 and accuracy of 0.74, while

Text	Model	Predicted Label	Actual Label
`Athippo avark evde venelm aavamloo Thabileegenedirey 100 video ittu...`	m-BERT	Fake	Original
`Ithiney kurich onnenkilum ittallo`	ALBERT	Original	Fake

Table 5: Error Analysis

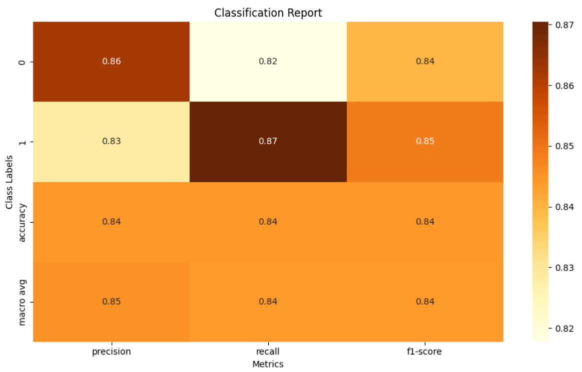

Figure 2: Classification Report - Task 1 XLMRoBERTa Model

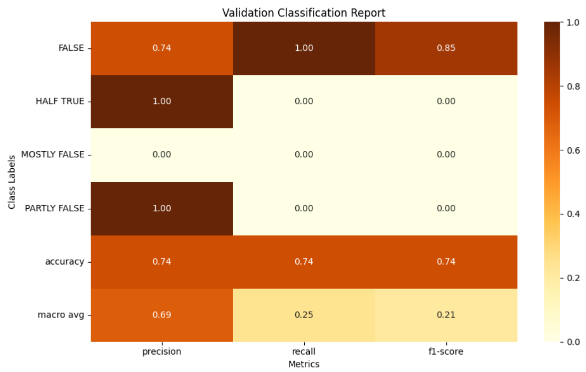

Figure 3: Classification Report - Task 2 XLMRoBERTa Model

mBERT and BERT exhibited lower scores, summarized in Table 4. Figures 2 and 3 depict the classification reports for Task 1 and Task 2, respectively. The classification reports for both tasks showcase the performance of the XLM-RoBERTa model, which outperformed all other models.

6 Error Analysis

The F1 score from the XLMRoBERTa model for both Task 1 and Task 2 exposes false positive and false negative prevalence, linked to inherent data imbalance. Class "false" with more instances, exhibits a high F1 score, precision, and recall, indicating Fake News. Lower training instances intensify misclassifications, highlighting the data imbalance impact. To address this, data augmentation is suggested for enhanced model performance. Misclassified Malayalam texts due to data imbalance are detailed in Table 5, emphasizing the need for strategies to overcome these challenges in Fake News detection.

7 Limitation

XLMRoberta, ALBERT, and mBERT excel but face hurdles: cross-lingual challenges, biased predictions due to data imbalances, uncertainties in generalizing to new datasets, resource-intensive fine-tuning for diverse languages, interpretability issues, and a textual focus overlooking multimodal fake news complexities. Recognizing and addressing these constraints is essential for refining approaches and ensuring robust fake news detection in transformer-based models.

8 Conclusion

The DravidianLangTech@EACL 2024 initiative, focused on fake news detection in Dravidian languages, signifies a critical step in countering misinformation on social media. Task 1, employing mBERT, ALBERT, and XMLRoBERTa, revealed XMLRoBERTa as the optimal model, showcasing its prowess in discerning authenticity. In Task 2, where XMLRoBERTa, mBERT, and BERT were employed, XMLRoBERTa consistently outperformed, emphasizing its effectiveness in classifying Malayalam-language articles. This collective effort not only advances natural language processing but also reinforces the importance of reliable information dissemination in the digital age. The XMLRoBERTa model's dual-task success highlights its pivotal role in navigating the challenges of fake news, setting the stage for continued innovation in Dravidian language technology.

References

2020. Fake News Detection using Deep Learning Models: A Novel Approach, author=Kumar, Sachin and Asthana, Rohan and Upadhyay, Shashwat and

Upreti, Nidhi and Akbar, Mohammad. *Transactions on Emerging Telecommunications Technologies*, 31(2):e3767.

Md Monsur Ali. 2022. *Multi Languages Fake News Detection*. Ph.D. thesis, Hochschule Rhein-Waal.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale. *CoRR*, abs/1911.02116.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, abs/1810.04805.

Maryam Heidari, Samira Zad, Parisa Hajibabaee, Masoud Malekzadeh, SeyyedPooya HekmatiAthar, Ozlem Uzuner, and James H Jones. 2021. BERT Model for Fake News Detection Based on Social Bot Activities in the COVID-19 Pandemic. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0103–0109. IEEE.

Mahibha Jerin, Kayalvizhi Sampath, and Thenmozhi Durairaj. 2021. Sentiment Analysis using Cross Lingual Word Embedding Model. In *Proceedings of Forum for Information Retrieval Evaluation-FIRE 2021*.

Quanliang Jing, Di Yao, Xinxin Fan, Baoli Wang, Haining Tan, Xiangpeng Bu, and Jingping Bi. 2021. TRANSFAKE: Multi-Task Transformer for Multimodal Enhanced Fake News Detection. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Heejung Jwa, Dongsuk Oh, Kinam Park, Jang Mook Kang, and Heuiseok Lim. 2019. exBAKE: Automatic Fake News Detection model based on Bidirectional Encoder Representations from Transformers (BERT). *Applied Sciences*, 9(19):4062.

Sebastian Kula, Rafał Kozik, Michał Choraś, and Michał Woźniak. 2021. Transformer Based Models in Fake News Detection. In *International Conference on Computational Science*, pages 28–38. Springer.

Ashok Kumar, Tina Esther Trueman, and Erik Cambria. 2021. Fake news detection using XLNet fine-tuning model. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, pages 1–4. IEEE.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Momina Qazi, Muhammad US Khan, and Mazhar Ali. 2020. Detection of Fake nNews using Transformer Model. In *2020 3rd international conference on computing, mathematics and engineering technologies (iCoMET)*, pages 1–6. IEEE.

Eduri Raja, Badal Soni, and Sami Kumar Borgohain. 2023. nlpt malayalm@ DravidianLangTech: Fake News Detection in Malayalam using Optimized XLM-RoBERTa Model. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 186–191.

Shaina Raza and Chen Ding. 2022. Fake News Detection based on News Content and Social Contexts: a Transformer-based approach. *International Journal of Data Science and Analytics*, 13(4):335–362.

Mina Schütz, Alexander Schindler, Melanie Siegel, and Kawa Nazemi. 2021. Automatic Fake News Detection with Pre-Trained Transformer Models. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII*, pages 627–641. Springer.

Chenguang Song, Nianwen Ning, Yunlei Zhang, and Bin Wu. 2021. Knowledge Augmented Transformer for Adversarial Multidomain Multiclassification Multimodal Fake News Detection. *Neurocomputing*, 462:88–100.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, Premjith B, Sandhiya Raja, Vanaja, Mithunajha S, Devika K, Hariprasath S.B, Haripriya B, and Vigneshwar E. 2024. Overview of the Second Shared Task on Fake News Detection in Dravidian Languages. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, Malta. European Chapter of the Association for Computational Linguistics.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, Muskaan Singh, Sandhiya Raja, Vanaja, and Mithunajha S. 2023. Overview of the Shared Task on Fake News Detection from Social Media Text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Xingang Wang, Xiaomin Li, Xiaoyu Liu, and Honglu Cheng. 2022. Using ALBERT and Multi-modal Circulant Fusion for Fake News Detection. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2936–2942. IEEE.