KEC_AI_DATA_DRIFTERS@DravidianLangTech 2025: Fake News Detection in Dravidian Languages

Kogilavani Shanmugavadivel[1], Malliga Subramanian[2], Vishali K S[1], Priyanka B[1], Naveen Kumar K[1]

[1]Department of AI, Kongu Engineering College, Perundurai, Erode.
[2]Department of CSE, Kongu Engineering College, Perundurai, Erode.
{kogilavani.sv, mallinishanth72}@gmail.com
{vishaliks.22aid, priyankab.22aid, naveenkumark.22aid}@kongu.edu

Abstract

Detecting fake news in Malayalam possess significant challenges due to linguistic diversity, code-mixing, and the limited availability of structured datasets.We participated in the Fake News Detection in Dravidian Languages shared task, classifying news and social media posts into binary and multi-class categories. Our experiments used traditional ML models: Support Vector Machine (SVM), Random Forest, Logistic Regression, Naive Bayes and transfer learning models: Multilingual Bert (mBERT) and XLNet. In binary classification, SVM achieved the highest macro-F1 score of 0.97, while in multi-class classification, it also outperformed other models with a macro-F1 score of 0.98. Random Forest ranked second in both tasks. Despite their advanced capabilities, mBERT and XLNet exhibited lower precision due to data limitations. Our approach enhances fake news detection and NLP solutions for low-resource languages.

1 Introduction

Fake news detection is a critical challenge in the digital age, where misinformation spreads rapidly online, influencing public opinion and policy-making. This issue is even more pronounced in regional languages like Malayalam due to linguistic complexities, cultural nuances, and diverse online content. Developing effective detection methods is essential to ensure informed decision-making.

Detecting fake news in low-resource languages like Malayalam is challenging due to limited annotated data, linguistic diversity, and writing style variations like code-mixing and Romanization. Additionally, Malayalam's complex morphology and informal online discourse make classification difficult. Existing approaches primarily use machine learning and transfer learning techniques to improve classification accuracy.

This research develops a fake news detection system for Malayalam by evaluating traditional machine learning models and transformer-based approaches independently. We applied SVM, Random Forest, Multinomial Naive Bayes, and Logistic Regression for text classification. Additionally, transfer learning models such as mBERT and XLNet were tested to analyze multilingual text. Each model's effectiveness was assessed separately to determine the most suitable approach for fake news classification in Malayalam.

2 Literature Review

Subramanian et al. (2024a) highlighted how social networks spread misinformation, affecting public understanding and trust. The FakeDetect-Malayalam shared task addresses this with two subtasks: Task 1 classifies social media posts as genuine or fake, while Task 2 categorizes fake news into five labels, including False and Mostly True.

Subramanian et al. (2024b) highlighted the rapid spread of fake news in Malayalam online. In Task 1, they secured ninth place using RNNs to classify news as Original or Fake, leveraging their ability to capture sequential patterns. Their study aims to enhance accuracy and improve fake news detection.

Qu et al. (2024) proposed QMFND, a quantum fusion model using Quantum Convolutional Neural Networks (QCNN) to merge text and image data. Tested on Gossip and Politifact datasets, it shows robustness against quantum noise, enhances expressibility, and performs as well as or better than classical models.

K et al. (2024) emphasized the need for fake news detection in Malayalam, a low-resource language. They introduced a curated dataset categorizing news by inaccuracy levels. Baseline models, including multilingual BERT and ML classifiers, showed potential, with Logistic Regression on LaBSE achieving an F1 score of 0.3393. Addressing data imbalance is key to improving accuracy.

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 260–264
May 3, 2025

Shanmugavadivel et al. (2024) emphasized the rapid spread of false information on social media and the need for fake news detection. The study validated YouTube comments using ML models like Naive Bayes, SVM, Random Forest, and Decision Tree. Presented at DravidianLangTech@EACL 2024, it applies ML and NLP techniques to combat misinformation.

Babu et al. (2023) proposed a machine learning approach for vectorizing and tokenizing news headlines. Experimental results demonstrate that this method surpasses existing fake news detection techniques, demonstrating its effectiveness in various topics and languages.

S et al. (2023) provided detailed instructions for preparing a manuscript for the RANLP 2023 proceedings on this page. These guidelines apply to both initial submissions and final versions, including an example of the required format. The authors must follow all provided instructions to ensure compliance.

Hu et al. (2022) explored deep learning approaches for fake news detection, including supervised, weakly supervised, and unsupervised methods. The study evaluates models using news content, social context, and external data while reviewing FND datasets, identifying limitations, and proposing future research directions.

Baarir and Djeffal (2021) proposed a machine learning system for fake news detection using TF-IDF with bag of words, n-grams for feature extraction, and SVM for classification. The system was trained on a curated dataset of true and fake news, demonstrating its effectiveness in identifying misinformation. However, detecting fake news remains challenging due to limited datasets and analytical approaches.

Ahmad et al. (2020) developed a machine learning ensemble model to classify news articles based on linguistic characteristics. Evaluated on four real-world datasets, the model demonstrated superior performance compared to individual classifiers in detecting disinformation.

3 Problem and System description

The Fake News Detection from Malayalam News task aims to identify fake news in Malayalam social media posts and articles. It classifies the text into two categories: Fake or original, with the dataset containing labeled posts and articles. The task also includes categorizing fake news into five labels: False, Half True, Mostly False, Partly False, and Mostly True. Researchers will use machine learning models, embeddings, and transfer learning to distinguish between accurate and misleading information. This task contributes to creating a robust fake news detection system for Malayalam content, promoting reliable communication, and reducing misinformation. Subramanian et al. (2025) Out of 128 teams, our system secured the 13th rank in task 1 and the 5th rank in task 2.

4 Dataset description

The shared dataset consists of Malayalam news articles categorized into two tasks. In Task 1 (Binary Classification), the training dataset consists of 1,659 Original and 1,598 Fake class labels, while the test dataset consists of 1,019 rows. In Task 2 (Multiclass Classification), the training dataset consists of 1,384 False, 2 True, 162 Half True, 295 Mostly False, and 57 Partly False labels, with a test dataset containing 200 rows. In addition, a validation dataset is provided to evaluate the performance of the model prior to testing.

Dataset	No. of Comments
Train	3257
Dev	815
Test	1019

Table 1: Task 1 Dataset Description

Dataset	No. of Comments
Train	1900
Test	200

Table 2: Task 2 Dataset Description

5 Methodology

5.1 Data pre-processing

To enhance text quality for fake news classification, we applied a structured pre-processing pipeline. The text was lowercased, and URLs, HTML tags, special characters, and numbers were removed. Tokenization filtered out non-informative words using custom Malayalam stopwords. Stemming (PorterStemmer) and lemmatization (WordNetLemmatizer) normalized words, while short and duplicate words were removed to reduce redundancy. This process improves data quality for machine learning analysis. Figure 1 illustrates the workflow from data pre-processing to classification.

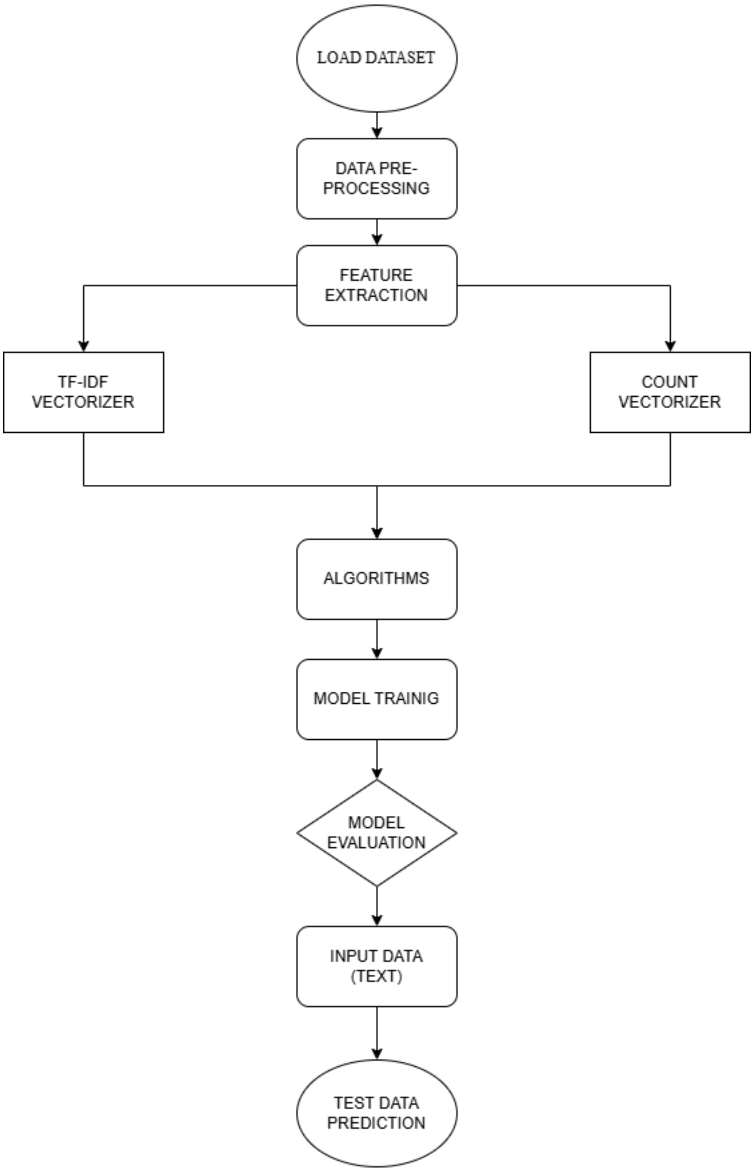

Figure 1: Proposed System Workflow

5.2 Encoding module

For our dataset, we utilized TF-IDF Vectorizer and Count Vectorizer from sklearn.feature-extraction for feature extraction. TF-IDF assigns weights to words based on their importance, reducing the influence of commonly used terms. Count Vectorizer, on the other hand, generates a matrix representing word frequencies, highlighting frequently occurring words in the text. By combining both techniques, we ensure a balanced representation of important terms and common patterns. This approach enhances the model's ability to capture textual features effectively, leading to improved classification performance. Ultimately, these methods contribute to better accuracy in fake news detection.

5.3 Model description

To classify Malayalam news as original or fake, we used SVM, Random Forest, Multinomial Naive Bayes, and Logistic Regression for text classification. Naive Bayes assigns probabilities based on Bayes' theorem, Random Forest builds multiple decision trees, SVM finds an optimal hyperplane, and Logistic Regression predicts binary outcomes. Additionally, XLNet and mBERT were applied for transfer learning, with XLNet leveraging bidirectional context and mBERT supporting multilingual analysis. Each model was tested independently, and results showed that traditional ML models outperformed transformer-based approaches for fake news detection in Malayalam.

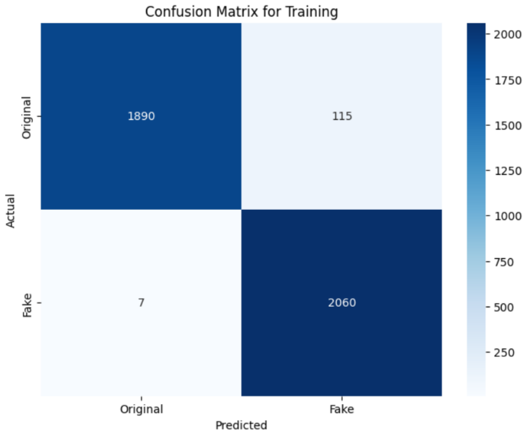

Figure 2: Confusion Matrix for Task 1

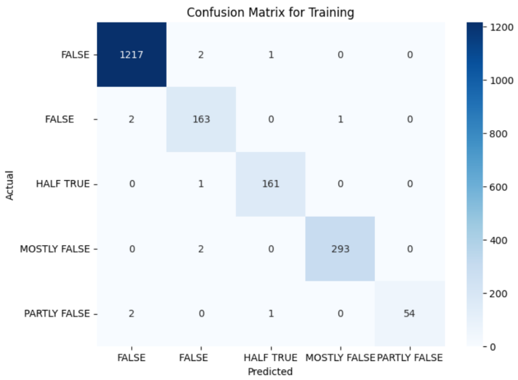

Figure 3: Confusion Matrix for Task 2

6 Experimental Analysis

In this experiment, we used the Malayalam dataset for the classification of fake news, applying four machine learning models and two transfer learning models for two tasks. In task 1 (binary classification), SVM achieved the highest accuracy (97%), followed by Random Forest (96%), Logistic Regression (94%), Multinomial Naive Bayes (90%), XLNet (80%), and mBERT (83%). SVM was the best performer, excelling in both accuracy and macro F1 score. In task 2 (multiclass classification), SVM again led with 98% accuracy, followed by Random Forest (95%), Logistic Regression (92%), and Multinomial Naïve Bayes (91%). Traditional models, especially SVM and Random Forest, outperformed deep learning models like XLNet and mBERT, proving to be the most reliable for fake news detection in Malayalam. Github Repository: Fake News Detection

7 Limitations

Our approach relies heavily on labeled datasets, which are limited to Malayalam and other Dravidian languages. The imbalance in multi-class classification affected model performance, especially for underrepresented labels. Although traditional ML models performed well, transfer learning models struggled due to data scarcity and domain-specific challenges. Furthermore, variations in code-mixed and informal text reduced the accuracy of the classification. Enhancing dataset quality, incorporating advanced preprocessing techniques, and optimizing deep learning models are crucial to improve fake news detection in Malayalam.

Model	Macro F1-Score
Support Vector Classifier	0.97
Random Forest	0.96
Logistic Regression	0.94
Naive Bayes	0.90
mBert	0.83
XLNet	0.80

Table 3: Macro F1-Score Metrics for Task 1

Model	Macro F1-Score
Support Vector Classifier	0.98
Random Forest	0.95
Logistic Regression	0.92
Naive Bayes	0.91

Table 4: Macro F1-Score Metrics for Task 2

8 Conclusion

We applied multiple machine learning and transfer learning models for fake news detection in Malayalam. SVM and Random Forest performed exceptionally well, achieving high Macro-F1 scores in both tasks, with SVM scoring 0.97 in binary and 0.98 in multi-class classification. Although mBERT and XLNet had lower accuracy due to data constraints, they demonstrated the potential of context-aware models for low-resource languages. Each model was applied independently to assess its effectiveness, and the results highlight the superiority of traditional ML models over transformer-based approaches in this context. Future work should focus on improving the quality of the data set, improving feature extraction techniques and optimizing deep learning models to further advance fake news detection in Malayalam and other Dravidian languages.

References

Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. 2020. Fake news detection using machine learning ensemble methods. *Complexity*, 2020:1–11.

Nihel Fatima Baarir and Abdelhamid Djeffal. 2021. Fake news detection using machine learning. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pages 125–130.

Tina Babu, Rekha R Nair, Adithya Challa, Rahul Srikanth, Sri Sai Aravindan, and Suhas S. 2023. Fake news detection using machine learning algorithms. In *2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*, volume 1, pages 1–7.

Linmei Hu, Siqi Wei, Ziwang Zhao, and Bin Wu. 2022. Deep learning for fake news detection: A comprehensive survey. *AI Open*, 3:133–155.

Devika K, Hariprasath .s.b, Haripriya B, Vigneshwar E, Premjith B, and Bharathi Raja Chakravarthi. 2024. From dataset to detection: A comprehensive approach to combating Malayalam fake news. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 16–23, St. Julian's, Malta. Association for Computational Linguistics.

Zhiguo Qu, Yunyi Meng, Ghulam Muhammad, and Prayag Tiwari. 2024. Qmfnd: A quantum multimodal fusion-based fake news detection model for social media. *Information Fusion*, 104:102172.

Malliga S, Bharathi Raja Chakravarthi, Kogilavani S V, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, and Muskaan Singh. 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*, pages 59–63, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Kogilavani Shanmugavadivel, Malliga Subramanian, Sanjai R, Mohammed Sameer B, and Motheeswaran K. 2024. Beyond tech@DravidianLangTech2024 : Fake news detection in Dravidian languages using machine learning. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 124–128, St. Julian's, Malta. Association for Computational Linguistics.

Malliga Subramanian, , B Premjith, Kogilavani Shanmugavadivel, Santhia Pandiyan, Balasubramanian Palani, and Bharathi Raja Chakravarthi. 2025. Overview of the Shared Task on Fake News Detection in Dravidian Languages: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, Premjith B, Vanaja K, Mithunja S, Devika K, Hariprasath S.b, Haripriya B, and Vigneshwar E. 2024a. Overview of the second shared task on fake news detection in Dravidian languages: DravidianLangTech@EACL 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 71–78, St. Julian's, Malta. Association for Computational Linguistics.

Malliga Subramanian, Jayanthjr J R, Muthu Karuppan P, Keerthibala T, and Kogilavani Shanmugavadivel. 2024b. KEC_HAWKS@DravidianLangTech 2024 : Detecting Malayalam fake news using machine learning models. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 266–270, St. Julian's, Malta. Association for Computational Linguistics.