KEC_HAWKS@DravidianLangTech 2024 : Detecting Malayalam Fake News using Machine Learning Models

Malliga Subramanian, Jayanth J R, Muthu Karuppan P, KeerthiBala A T, Kogilavani Shanmugavadivel
Kongu Engineering College
Tamil Nadu
India
{mallisenthil.cse, jayanthjr.21cse, muthukaruppanp.21cse, keerthibalaat.21cse, kogilavani.cse}@kongu.edu

Abstract

The proliferation of fake news in the Malayalam language across digital platforms has emerged as a pressing issue. By employing Recurrent Neural Networks (RNNs), a type of machine learning model, we aim to distinguish between Original and Fake News in Malayalam and achieved 9th rank in Task 1. RNNs are chosen for their ability to understand the sequence of words in a sentence, which is important in languages like Malayalam. Our main goal is to develop better models that can spot fake news effectively. We analyze various features to understand what contributes most to this accuracy. By doing so, we hope to provide a reliable method for identifying and combating fake news in the Malayalam language.

1 Introduction

In today's digital age, the proliferation of fake news has emerged as a significant challenge, disrupting the flow of accurate information and eroding public trust in media sources. In this study, the shared task is to determine the authenticity of news articles in the Malayalam language, distinguishing between fake and original news. Leveraging deep learning techniques, particularly Recurrent Neural Networks (RNNs), we preprocess the data using tokenization and padding techniques. Our model, built with RNNs known for capturing sequential dependencies in language, will be compiled, trained on the dataset shared as part of Fake News Detection in Dravidian Languages-DravidianLangTech@EACL 2024[1], and evaluated using metrics like accuracy score and classification report. Visualization techniques, including confusion matrix heatmaps, will provide insights into model performance. Ultimately, we aim to develop an effective tool for distinguishing between authentic and fake news articles, contributing to combating misinformation in the digital landscape. The following sections will delve into a comprehensive literature survey to understand existing approaches to fake news detection, followed by detailing the methods employed in our study. We will present the results of our experiments, analyze the performance metrics of the model, and finally, conclusions regarding the effectiveness of our approach and its implications for addressing the challenge of fake news in the Malayalam language.

The rest of the article is organized as follows: Section 2 provides a brief overview of existing works on fake news detection. The dataset used and the proposed model are discussed in Section 3. The results are presented in Section 4. Finally, we conclude our work in Section 5.

2 Literature Survey

In the modern digital era, combating the rapid spread of misinformation has become a critical societal challenge, necessitating innovative strategies for detection and mitigation. This literature review critically assesses the landscape of fake news detection. Mykhailo and Volodymyr (2017) presents a straightforward method using a naive Bayes classifier, achieving a respectable accuracy of approximately 0.74 on Facebook news posts. Akshay and Amey (2018) utilizes a Naive Bayes classification model for predicting the authenticity of Facebook posts, with potential improvements discussed in the paper. Bijimol and Anit Sara (2023) focuses on detecting Malayalam fake news using a Passive Aggressive Classifier, while Murari et al. (2021) provides a comprehensive review of machine learning algorithms for fake news detection across var-

[1]https://codalab.lisn.upsaclay.fr/competitions/16055

Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 266–270

ious social media platforms. Sumit and Jyoti (2022) combines SVM and Naive Bayes for detecting fake news with an accuracy of 0.84, while Rizwana Kallooravi and Mohamed (2021) emphasizes the importance of including more search engines for improved accuracy in fake news detection. Jasmine and Rupali (2021) specifically addresses the challenge of fake news detection in the Malayalam language, employing Recurrent Neural Networks (RNNs) for their ability to capture sequential dependencies and achieve high accuracy. Nihel Fatima and Abdelhamid (2021) proposes a machine learning-based system utilizing TF-IDF and SVM for effective fake news detection. Further, the results of various approaches proposed for detecting fake news have been reviewed and presented in Subramanian et al. (2024). Also, the performance of several methods proposed for the fake news shared task released during 2023 presented in Subramanian et al. (2023). In the quest to combat misinformation, the contribution of RNNs underscores their interdisciplinary nature in addressing linguistic and contextual complexities, aiming to deepen understanding and enhance countermeasures against fake news proliferation in the digital age.

3 Materials and Methods

3.1 Taskset Description

For this model we have taken taskset with different labels.Taskset are labeled as either 'Original' for presenting genuine and factual information or 'Fake' if they contain intentionally deceptive or fabricated content. This binary classification serves as a robust foundation for training models to discern between legitimate and misleading news in the Malayalam language. These refined labels contribute to the datasets' richness, fostering the development of sophisticated machine learning models capable of adeptly handling the intricacies of news classification and detection. Whether through nuanced truthfulness labels in task set or binary authenticity labels,these distinctions collectively enhance the depth and accuracy of the analyses conducted in their respective domains, while the sentence structure is adjusted for originality without altering the meaning.

3.2 Pre-processing and Feature Extraction

The model pre-processes textual data for training a Bidirectional Long Short-Term Memory (LSTM) model for binary classification. It starts with label encoding using the LabelEncoder to convert categorical labels ('Original' and 'Fake') into numerical values (0 and 1). Then, the Tokenizer class from Keras tokenizes the text data, limiting the vocabulary size based on word frequency. Sequences are padded or truncated to ensure uniform sequence length. An embedding layer is added to learn dense word representations, followed by two Bidirectional LSTM layers to capture information from past and future contexts. Dense layers with ReLU activation functions and a final sigmoid activation function for binary classification are included. A dropout layer mitigates overfitting, and the model is compiled using the Adam optimizer and binary cross-entropy loss function.

3.3 Methodology

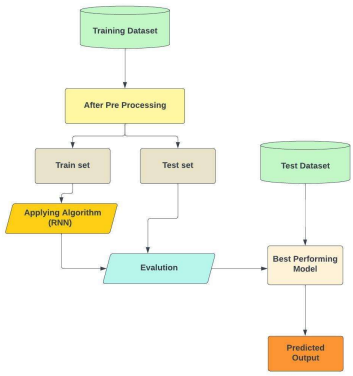

Figure 1: Flow diagram

3.4 Proposed Classifiers

For fake News detection, we propose utilizing a Recurrent Neural Network (RNN). RNNs are specifically designed for sequential data processing and are highly effective in natural language tasks. They have the ability to capture intricate dependencies within textual content by retaining memory from previous inputs. The proposed RNN model incorporates recurrent layers such as LSTM or GRU to effectively capture sequential patterns and identify offensive language patterns. This approach aims to

S.NO	NEWS	LABEL
1	Poda polayadi monea thooooo. Madar choot thooooo..	FAKE
2	e pennugalk vera valla panikum poikoode	ORIGINAL
3	Ee prandhane oke anu .....aadhyam naadu kadathande	FAKE
4	Cpm raja baranam pole aayi	ORIGINAL

Table 1: Task set

leverage the contextual understanding power of RNNs for robust offensive text detection. RNNs are selected for their strengths in handling the challenges of identifying offensive content within textual data.

4 Results and Discussion

The implementation of recurrent neural network (RNN) models in Task 1 run 1 and Task 1 run 2 for classifying Malayalam news articles as fake or original revealed critical limitations in accurately distinguishing between the two categories and achieved the overall accuracy of around 0.5833 and 0.5882. Despite training and evaluating the models on separate datasets, both runs exhibited a systemic issue where all instances were erroneously classified as original news, yielding zero true positives for fake news. This misclassification highlights fundamental flaws in the model's ability to discern meaningful patterns from the data. To address this, future improvements could involve exploring alternative model architectures incorporating attention mechanisms or more complex recurrent units, enhancing feature engineering techniques such as utilizing word embeddings or capturing semantic relationships, and mitigating data imbalance issues through oversampling or augmentation methods. Overall, these findings underscore the necessity for iterative refinement and experimentation to develop a more robust and effective classification system for combatting misinformation within the digital landscape of Malayalam news.

4.1 Performance Metrics

The evaluation of various classification models for detecting Malayalam fake news employed key metrics, including Accuracy, Precision, Recall, and F1-Score . These metrics are fundamental for assessing the classifiers' effectiveness in distinguishing between Original and Fake Malayalam NEWS.

Accuracy is calculated using the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall, also known as Sensitivity or True Positive Rate, is defined by the formula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision, or Positive Predictive Value, is given by the formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1-Score is the harmonic average of Precision and Recall, calculated as in the following formula:

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

providing a balanced assessment of the model's performance.The performance metrics for each classifier on the test dataset, along with the confusion matrix. The confusion matrix visually illustrates correct and incorrect classifications, with the X-axis representing predicted classes and the Y-axis representing actual classes. Notably, the proposed models utilizing RNN accurately classifying Original and Fake Malayalam NEWS.

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score	Weighted Precision	Weighted Recall	Weighted F1-Score
RNN	0.59	0.29	0.50	0.37	0.35	0.59	0.44

Table 2: Classification report for RNN

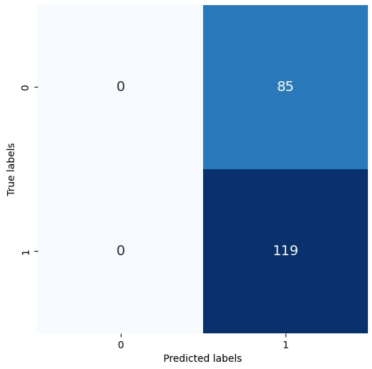

Figure 2: Confusion matrix for Malayalam fake NEWS detection Run1

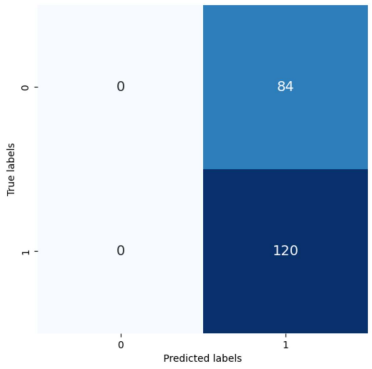

Figure 3: Confusion matrix for Malayalam fake NEWS detection Run2

Upon examining the confusion matrices in Figures 2 and 3, specific patterns emerge in the classification performance. For Taskset 1 Run 1 and Taskset 1 run 2 the RNN model exhibits a balanced performance, effectively identifying both Original and fake news.

5 Conclusion

In conclusion, the RNN model's performance in both Task 1 Run 1 and Task 1 Run 2 is characterized by an overall accuracy of around 0.5833 and 0.5882, respectively. While proficient in classifying original news instances, the model exhibits significant limitations in identifying potentially fake news. The precision for fake news is strikingly low, indicating a high number of false positives, and the recall is particularly deficient, especially in Task 1 Run 2 where it is zero.These findings underscore the imperative need for refinement, specifically in enhancing the model's ability to discern and accurately classify instances of potentially deceptive news. The observed patterns provide clear directions for future iterations, emphasizing the importance of addressing these limitations to elevate the model's effectiveness in distinguishing between original and fake news in Malayalam text.

References

Jain Akshay and Kasbe Amey. 2018. Fake News Detection. *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS).*

T K Bijimol and Santhosh Anit Sara. 2023. Malayalam Fake News Detection using Machine Learning. *National Conference on Emerging Computer Applications, 4(1)*, 4(1).

Shaikh Jasmine and Patil Rupali. 2021. Fake News Detection using Machine Learning. *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC).*

Choudhary Murari, Prashant Deepika Saxena Shashank, Jha, and Singh Ashutosh Kumar. 2021. A Review of Fake News Detection using Machine Learning. *2021 2nd International Conference for Emerging Technology (INCET).*

Granik Mykhailo and Mesyura Volodymyr. 2017. Fake news detection using naive Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)).*

Baarir Nihel Fatima and Djeffal Abdelhamid. 2021. Fake News detection Using Machine Learning. *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH).*

Thandil Rizwana Kallooravi and Muneer Mohamed, Basheer V K. 2021. Natural Language Processing of Malayalam Text for predicting its Authenticity. *Proceedings of the Yukthi 2021- The International Conference on Emerging Trends in Engineering – GEC Kozhikode, Kerala, India.*

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 ComputationalLinguistics.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, and Muskaan Singh. 2023. Overview of the Shared Task on Fake News Detection in Dravidian Languages. In *Proceedings of the Third Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, Bulgaria. European Chapter of the Association for ComputationalLinguistics.

kumar Sumit and Tiwari Jyoti. 2022. Implementing a hybrid method for fake news detection. *International Journal of Emerging Trends in Engineering Research*.