WordWizards@DravidianLangTech 2024:Fake News Detection in Dravidian Languages using Cross-lingual Sentence Embeddings

Akshatha Anbalagan[1], Priyadharshini T[1], Niranjana A[1],
Shreedevi Seluka Balaji[1], Durairaj Thenmozhi[1]
akshatha2210397@ssn.edu.in , priyadharshini2210228@ssn.edu.in ,
niranjana2210379@ssn.edu.in , shreedevi2210389@ssn.edu.in , theni_d@ssn.edu.in
[1]**Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu, India**

Abstract

The proliferation of fake news in digital media has become a significant societal concern, impacting public opinion, trust, and decision-making. This project focuses on the development of machine learning models for the detection of fake news. Leveraging a dataset containing both genuine and deceptive news articles, the proposed models employ natural language processing techniques, feature extraction and classification algorithms.

This paper provides a solution to Fake News Detection in Dravidian Languages - DravidianLangTech 2024 [1]. There are two sub tasks: Task 1 - The goal of this task is to classify a given social media text into original or fake. We propose an approach for this with the help of a supervised machine learning model – SVM (Support Vector Machine). The SVM classifier achieved a macro F1 score of 0.78 in test data and a rank 11. The Task 2 is classifying fake news articles in Malayalam language into different categories namely False, Half True, Mostly False, Partly False and Mostly True.We have used Naive Bayes which achieved macro F1-score 0.3517 in test data and a rank 6.

1 Introduction

Fake News refers to false or misleading information presented as genuine news. This misinformation is often disseminated through traditional media, social media platforms, or other online channels. The intent behind fake news is typically to deceive and manipulate public opinion, influence political processes, or generate click-throughs for financial gain. In 2017, fake news has been seen to have influenced the US elections and the British Brexit vote, and locally in South Africa Finance Minister Pravin Gordhan, newspaper editors and journalists have become targets for fake news peddlers. In other instances breaking news on social media has turned out to be false and based on hoaxes and hearsay (Shu et al., 2017).

Following this, the task of Fake News Detection in Dravidian Language DravidianLangTech 2024 (Chakravarthi et al., 2021; Subramanian et al., 2023) aims to classify a set of social media texts and news articles using principles of feature extraction and machine learning.

The subsequent sections of this paper are structured as follows: Section 2 reviews relevant literature identified through a thorough literature survey. Section 3 presents an overview of the dataset, while Section 4 elaborates on the methodology employed for the task at hand. Section 5 delves into the discussion of results, and Section 6 concludes the paper.

2 Related Works

In recent years, there has been significant research in the domain of fake news detection, with a predominant focus on examining and identifying hoaxes within their primary dissemination channel: social media. The prevalent approach involves assessing the likelihood of a particular post being false by analyzing its inherent characteristics, such as likes, followers, shares, etc. This analysis typically employs traditional machine learning methods like classification trees, SVM, and similar techniques (Rodríguez and Iglesias, 2019).

In Natural Language Processing (NLP), there are many ways to approach the subject, and researchers have documented several methods in well-known literature. Recently, with advancements in turning image and speech into text, researchers have worked on creating and testing models that are both fast and accurate.

In the cited paper (Sharma et al., 2020), authors employed Artificial Intelligence, Natural Language Processing and Machine Learning for binary classification of news articles as fake or original. They assessed website authenticity using a two-step pro-

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

Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 162–166

cess: Static Search (training Naive Bayes, Random Forest, and Logistic Regression) and Dynamic Search (with three search fields).

In this work (Adiba et al., 2020), Naive Bayes Classifier, a Bayesian approach of Machine Learning algorithm has been applied to identify the fake news. The researchers showed how the wealth of corpora can assist algorithm to improve the performance. The dataset collected from an open-source, has been used to classify whether the news is authenticated or not. Initially,they achieved classification accuracy about 87 percent which is higher than previously reported accuracy and then 92 percent by the same Naive Bayes Algorithm with enriched corpora.

In the cited paper (Khanam et al., 2021), authors utilized Python's scikit-learn library for tokenization, feature extraction, and vectorization, employing tools like Count Vectorizer and Tiff Vectorizer. They conducted feature selection methods based on confusion matrix results to determine the most fitting features for achieving the highest precision. The study revealed that many research papers favored the Naive Bayes algorithm, achieving prediction precision within the range of 70-76 percent. Qualitative analysis, relying on sentiment analysis, titles, and word frequency repetition, was commonly employed in these studies.

3 Dataset

The dataset of Task 1 is a list of social media comments in English and Malayalam with the labels either fake or original. The sources of data are various social media platforms such as Twitter and Facebook. Figure 1 describes the data distribution of this task.

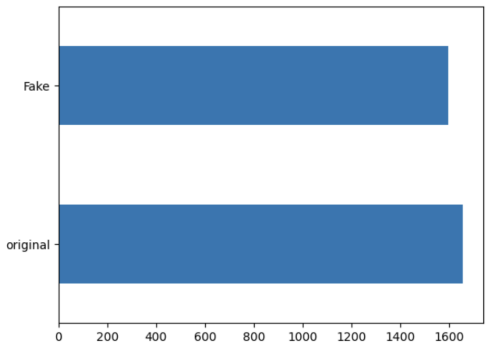

Figure 1:Task 1 dataset distribution

The dataset of Task 2 consists of news articles in Malayalam with the labels - False, Half True, Mostly False, Partly False and Mostly True. Figure 2 shows the data distribution of this task.

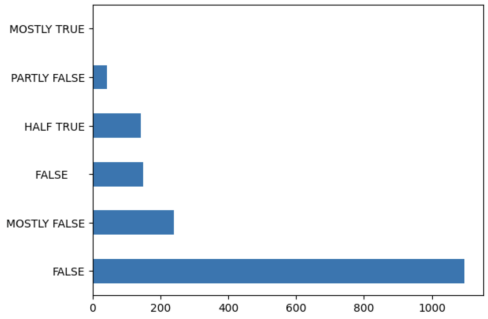

Figure 2:Task 2 dataset distribution

4 Methodology

4.1 Preprocessing

Preprocessing of given data involves cleaning, transforming, and organizing raw data into a format that is suitable for analysis or model training. The goal of data preprocessing is to improve the quality and usability of the data.

1. The elimination of punctuation marks and special characters serves to diminish noise and variations in the data, facilitating a clearer focus on the essential content making it easy for the model to learn the features.

2. Data reduction involves eliminating a significant portion of fillers or stop words present in any text which lack essential information for text analysis tasks. We have used a custom list of stop words in Malayalam and NLTK's English stop words list to remove such non-informative words from the text.

3. Vectorization is the method of representing words as vectors and is commonly referred to as word vector representations or word embeddings. In this paper, we experiment with the below mentioned word vector representations:

TF-IDF (term frequency-inverse document frequency) is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. It helps to identify the most important words in a document by taking into account both the frequency of the word in the document and the rarity of the word in the entire collection of documents. This technique was used for Task-1.

CountVectorizer converts a collection of text documents into a matrix of token counts. It is used to transform a corpus of text to a vector of term/token counts. The vocabulary of known words is formed,

which is subsequently used for encoding unseen text later. This method was utilized for Task-2.

A particular limitation faced here is that of the data set being in the Malayalam language, for which comparatively less resources are available. Thus a lot preprocessing techniques like lemmatisation and stemming could not be performed. The model may struggle to capture the subtleties of language use in different contexts.

4.2 LaBSE Feature Extraction

Feature extraction is a process that reduces the number of dimensions needed to define a large dataset by creating a smaller set of new features while discarding many existing ones. This involves transforming raw data into numerical features for further processing.

Language-agnostic BERT Sentence Embedding, or LaBSE, stands out as a multilingual language model developed by Google, building upon the BERT model.(Feng et al., 2020) In its pre-training process, LaBSE combines masked language modeling with translation language modeling. It is designed to be language-agnostic and has demonstrated superior performance compared to other existing sentence embedders.

This model proves useful for obtaining multilingual sentence embeddings and for bi-text retrieval. Recognized as the state-of-the-art in sentence embedding, LaBSE encodes sentences into a shared embedding space, ensuring that similar sentences are positioned closer to each other.

4.3 Models Used

Different researchers used different machine learning classifiers for checking the authenticity of news. According to their experiments the SVM and Naïve Bayes classifiers are best for detecting fake news. These two are better than other classifiers on the basis of accuracy they provide (Al Ayub Ahmed et al., 2021).

SVM: Support Vector Machine, is a supervised learning method that works for both classification and regression tasks. It is like a tool we use to organize information. SVM helps us find a special line, called a hyperplane, in a space with many different aspects or features (we call this space N-dimensional, where N is the number of features).

The idea behind SVM is to make decisions based on differences. Imagine some points in our data that are very important for making decisions; these are called support vectors. They are the ones closest to our decision line, or hyperplane (Patel et al., 2022).

SVM does its job by transforming our data into a space with many dimensions. This makes it better at figuring out patterns and making predictions, especially when our data is not easily separated in a straight line.

Naive Bayes: This classification technique is based on Bayes theorem, which assumes that the presence of a particular feature in a class is independent of the presence of any other feature. It provides way for calculating the posterior probability.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

P(c|x) = posterior probability of class given predictor

P(c) = prior probability of class

P(x|c) = likelihood (probability of predictor given class)

P(x) = prior probability of predictor

(Ranjan, 2018)

5 Result And Analysis

5.1 Performance Metrics

We evaluate our model using the classification report which provides a summary of the performance metrics for a machine learning model, typically used for binary or multiclass classification tasks.

Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positives.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall (Sensitivity or True Positive Rate): Recall is also known as sensitivity or true positive rate and is defined as the ratio of correctly predicted positive observations to the all observations in the actual class.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Macro Average and Weighted Average: Macro average calculates metrics independently for each class and then takes the unweighted average. It

treats all classes equally. Weighted average calculates metrics for each class but uses the support of each class as weights when averaging. It considers class imbalance.

F1-Score: F1-score is the weighted average of precision and recall, providing a balance between the two metrics.

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

In summary, precision, recall, and F1-score are key metrics to evaluate the model's performance, considering aspects of true positives, false positives, and false negatives.

5.2 Analysis

Table 1 shows the Classification Report obtained from the train data using SVM that was used for Task 1.

Table 1: Task 1 Training Report

Label	Precision	Recall	F1 score
original	0.75	0.85	0.80
fake	0.83	0.70	0.76

Table 2: Task 1 Development Report

Label	Precision	Recall	F1 score
original	0.85	0.96	0.90
fake	0.95	0.83	0.89

The development data was used to evaluate the model after the training phase.The predictions on the development data gave a macro F1 score of 0.89 as shown in Table 2. Table 3 shows the Classification Report obtained from the train data using Naive Bayes Model which was used for Task 2.

5.3 Result

The training and development data has been used to train and evaluate the models. Predictions of labels is done on the test data.

Table 4 and Table 5 describe the macro F1-score of the prediction and the rank secured in the competition.

Table 3: Task 2 Training Report

Label	Precision	Recall	F1 score
False	0.81	0.99	0.89
Half True	0.90	0.33	0.49
Mostly False	0.95	0.37	0.53
Partly False	0.67	0.10	0.17
Mostly True	0.80	0.00	0.00

Table 4: Rank Secured in Task 1

Team Name	Macro F1-Score	Rank
CUETDUO	0.88	1
PunnyPunctuators	0.87	2
WordWizard	**0.78**	**11**

6 Conclusion

In this research paper, we present a solution to the Fake News Detection in Dravidian Languages-DravidianLangTech 2024. Our approach involves classifying the text data into categories such as original and fake and False, Half True, Mostly False, Partly False and Mostly True. We leverage a Language-agnostic Sentence Embedder known as LaBSE and machine learning models, SVM (Support Vector Machine) and Naive Bayes. The models demonstrated good F1-scores and accuracy signifying their effectiveness in accurately identifying deceitful news from original news as the rankings are shown in Table 4 and Table 5.

These findings underscore the potential of Natural Language Processing (NLP) in automating fake news detection. This contribution has great implications in the society and politics since people are not influenced by fake news.

Table 5: Rank Secured in Task 2

Team Name	Macro F1-Score	Rank
CUETBinaryHackers	0.5191	1
CUETSentimentSilles	0.4964	2
WordWizard	**0.3517**	**6**

Limitations

Fake news is constantly evolving, and new tactics are regularly employed to keep a check on them. A model trained on historical data may struggle to adapt to emerging patterns of misinformation. Thus dynamic nature of fake news poses a limitation to the project.

Identifying relevant features for effective model training can be challenging, especially when dealing with a language like Malayalam which has unique linguistic features which are not well-captured by standard NLP techniques.

Ethics Statement

Ethical fake news detection is crucial for safeguarding information integrity and preserving the societal fabric. Detecting and combatting misinformation helps maintain public trust in media and democratic processes, preventing the erosion of confidence in reliable sources. Moreover, ethical efforts contribute to preventing harm by curtailing the dissemination of false information that can lead to negative consequences such as violence. By promoting informed decision-making, ethical fake news detection empowers individuals to make choices based on accurate and reliable information, ultimately fostering a more responsible and well-informed society.

References

Farzana Islam Adiba, Tahmina Islam, M Shamim Kaiser, Mufti Mahmud, and Muhammad Arifur Rahman. 2020. Effect of corpora on classification of fake news using naive bayes classifier. *International Journal of Automation, Artificial Intelligence and Machine Learning*, 1(1):80–92.

Alim Al Ayub Ahmed, Ayman Aljabouh, Praveen Kumar Donepudi, and Myung Suh Choi. 2021. Detecting fake news using machine learning: A systematic literature review. *arXiv e-prints*, pages arXiv–2102.

Bharathi Raja Chakravarthi, Gaman Mihaela, Radu Tudor Ionescu, Heidi Jauhiainen, Tommi Jauhiainen, Krister Lindén, Nikola Ljubešić, Niko Partanen, Ruba Priyadharshini, Christoph Purschke, Eswari Rajagopal, Yves Scherrer, and Marcos Zampieri. 2021. Findings of the VarDial evaluation campaign 2021. In *Proceedings of the Eighth Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 1–11, Kiyv, Ukraine. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Z Khanam, BN Alwasel, H Sirafi, and Mamoon Rashid. 2021. Fake news detection using machine learning approaches. In *IOP conference series: materials science and engineering*, volume 1099, page 012040. IOP Publishing.

Alpna Patel, Arvind Kumar Tiwari, and SS Ahmad. 2022. Fake news detection using support vector machine.

Aayush Ranjan. 2018. *Fake news detection using machine learning*. Ph.D. thesis.

Álvaro Ibrain Rodríguez and Lara Lloret Iglesias. 2019. Fake news detection using deep learning. *arXiv preprint arXiv:1910.03496*.

Uma Sharma, Sidarth Saran, and Shankar M Patil. 2020. Fake news detection using machine learning algorithms. *International Journal of Creative Research Thoughts (IJCRT)*, 8(6):509–518.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

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.