Tamil Fake News Detection using a Transformer Based Approach

Supraja S[1], Debika Priya D[2], Divyasri K[3], and KS Gayathri[4]

[1] Sri Sivasubramaniya Nadar (SSN) College of Engineering, Chennai, India[1]supraja2010341@ssn.edu.in
[2] debika2010561@ssn.edu.in
[3] divya2011037@ssn.edu.in
[4] gayathriks@ssn.edu.in

Abstract. The widespread use of social media has led to the spread of false information, particularly in this day where the internet is pervasive. Low-resource languages like Tamil, where there are little resources for detecting fake news, face an even greater challenge in this regard. To tackle this, we are concentrating on creating a strong system that uses transformer-based model efficiency to classify fake news in Tamil. A transformer-based strategy in this endeavor, employing BERT based models like TamilBERT,Labse and XLNet , operate though models trained on a large corpus of Tamil data. Transformers provide a quick fix because of their exceptional ability to do parallel calculations. The dataset holds equal proporations of true and fraudulent Tamil news samples, so better accuracy for classification can be expected . TamilBERT is a transformer-based language model designed specifically for Tamil. hence easy to grasp the nuances of the language. LaBSE on the other hand, is a versatile language-agnostic model that is good at finding similarities in news material across languages since it provides sentence embeddings for effective analysis. XLNet, Because of its bidirectional autoregressive and autoencoding techniques, excellent tool for identifying subtle contextual links in language, which is useful when identifying linguistic clues that may be suggestive of misinformation in fake news detection tasks.

Keywords: Fake news detection · BERT · XLNet · Tamil

1 Introduction

In today's world, individuals have a vast array of information at their disposal. As an unexpected consequence, this has led to the proliferation and dissemination of fake news. The ease with which such misinformation spread has raised concerns worldwide. Plenty of fake news detection systems for languages like English have arise to tackle this issue. However, there is a major gap when it comes to low resource languages like Tamil. Tamil is of particular interest due to its vast usage in several countries such as India, Sri Lanka, Singapore, Malaysia to name a few. Due to its widespread usage, it is important to ensure that for primary communication in Tamil a model to check the authenticity of information. As fake news often relies on the cultural nuances of a language, it is crucial to develop an approach specific to Tamil language.

Various machine learning and deep learning techniques have demonstrated exceptional abilities when it comes to natural language processing tasks. Particularly for fake news detection, where understanding the language is primary, deep learning has proved to be better than other approaches. In the context of Tamil fake news detection, deep learning can better understand the complexity and nuances of Tamil to provide more accurate information.

Transformers are neural network based architecture used widely for natural language processing models[10].Transformer based approaches in particular can be favored for the task of fake news identification. Their self-attention mechanism and ability to capture long range dependencies in text, make it suitable to understand the context and semantics behind the text. Developing such a transformer based approach to Tamil fake news detection, ensures that an optimal system is developed.

This research paper, aims to address the pressing issue of fake news detection in Tamil, using transformer based language models, renowned for understanding and identifying complex patterns in such ancient sophisticated language. Starting off, the texts are primarily tokenized and are fed them into language models .In this paper, three such models used are: TamilBERT[10][2], LaBSE and XLNet[8], in an effort to identify the most suitable model for real-time detection of fake news in Tamil. TamilBERT was identified as the optimal approach, due to its high accuracy and lower overhead.

The paper is organised as follows: section 2 presents the literature review on the works done in Tamil News Fake detection and possible deep learning approaches used for the same. Section 3 presents the proposed approach used in the design of Tamil fake news detection. Section 4 discusses on the dataset [5]used in modeling and the various deep learning models used in the experimental study. Section 5 presents the conclusion and future work.

2 Literature Review

Fake news detection has been commonly done in the English language. BERT based models were combined with explainable Artificial Intelligence (xAI) techniques, Local Interpretable Model-Agnostic Explanations (LIME) methods to obtain an accuracy of 98% for classification of fake news [10]. These were further tested with various data sources such as NELA-GT-2019, Fakeddit ets and several transformer models to compare the accuracies obtained, to determine the suitable real-time model [9]. An improved approach was proposed, with a fine-tuned exBAKE model by using pre-training based on a BERT model to accurately understand the contents of articles to classify fake news in English [5] .

Additionally, fake news detection has been explored in many low resource languages. The paper 'Automatic Fake News Detection for Romanian Online News' compared the results obtained using CNN, Bert, RoBERTa and classified the Romanian corpus using Naive Bayes and SVM. They achieved an accuracy of 98.2% using CNN the model [2]. In the paper [11] various techniques like Hierarchical Networks, Attention network, DistilBERT, BERT, CNN, LSTM were tried out of which Pretrained LM with bert-cased had the highest accuracy with 98.4% on the Contraint@AAAI 2021 Covid-19 Fake news detection english dataset.

'Fake-news detection system using machine learning algorithms for Arabic-language content' [1] collected and manually tagged 206,080 Arabic tweets from twitter. The features were then selected and extracted using TF-IDF and ANOVA then further classified using naive bayes, Random Forest, Logistic regression, random committee. RC and RF achieved the highest accuracy with about 97.3%.

Fake news detection was attempted in low resource languages like Slovak too in the paper 'Improvement of Misleading and Fake News Classification for elective Languages by Morphological Group Analysis' [6]. They attempted to process the 160 collected articles based on morphological analysis which was done using Parts of Speech tagging and then further classified using decision tree. They achieved a 75% accuracy.

The paper 'Automatic Online Fake News Detection Combining Content and Social Signals' explained fake news detection in Italian from corpus collected from FacebookData, PolitiFact, and BuzzFeed. It combined social content and news content features to propose a HC-CB-3 detection system with logistic regression to achieve a 99% accuracy. Machine learning models like SVM, Naive Bayes, Logistic Regression were used to classify fake news in Indian media The dataset used in this, contained headlines from Tamil articles translated to English. Amongst the algorithms used, Naive Bayes gave the highest accuracy of 97.6%.

When it comes to Tamil fake news detection, a transformer based approach was suggested and implemented using the XLM-RoBERTa, mBERT and MuRIL [3] obtaining an accuracy of 84% for tamil. mBERT and XLM-RoBERTa were also used to perform fake news detection in four Dravidian languages: Telugu, Tamil, Kannada, and Malayalam, which reported an accuracy of 93.31% [8].

3 Proposed approach

The main objective is to identify the most suitable transformer based model for real time detection of Tamil fake news from headlines. For this, we have considered three models, TamilBERT, LaBSE and XLNet and have compared their accuracies and architecture to obtain the most suitable model.

TamilBERT is a pretrained model that is suitable for detecting fake news as it is trained on a large and diverse dataset of Tamil text from a variety of sources. It is developed based on the fine-tuning of existing models on a monolingual corpus [4]. This makes the model understand the nuances of the Tamil language and to identify the patterns that are often used in fake news. Moreover, TamilBERT is not just a statistical model that memorizes patterns. It can understand the meaning of the tamil text it is processing, making it suitable for fake news detection.

Additionally, the TamilBERT approach is more specific to Tamil compared to other transformer models that exist, which cover a range of Indic languages. This specificity, provides it an edge over other transformer approaches. While other approaches are trained with lesser data in several languages, this approach is trained with more data only in Tamil. This makes it ideal to help the model understand more clearly, the one language it is trained for, Tamil.

Furthermore, the TamilBERT approach, takes the entire Tamil headline as input, allowing it to be used in real-time application, where plenty of preprocessing for each headline is not feasible. This ensures the system has plenty of learning parameters while at the same time, has lead overhead and is not very computationally intensive.

3.1 Architecture

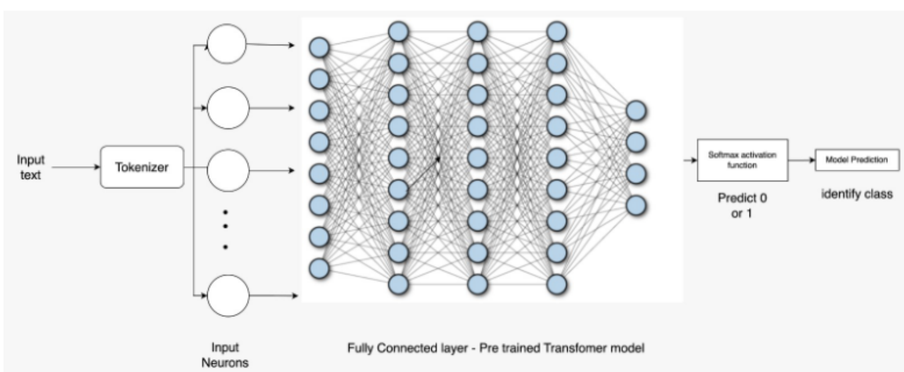

Fig. 1. Systematic architecture diagram for pre-trained language model text classification

This is the systematic architecture diagram of using pre-trained language models on text classification. The input text, which is tamil headlines, is tokenized and passed into fully connected dense layers during training phase and the obtained value is fed into softmax function 0 or 1 and find if the news is fake or not.

4 Experimental Setup

4.1 Dataset

For the purpose of this paper, the dataset described in [7] is considered. The dataset contains Tamil news headlines from various sources, under 5 major domains, namely: politics with 1674 records, miscellaneous with 1521 records, technology with 966 records, entertainment with 589 records and sport with 476 records each. Furthermore, the dataset contains an English version of each of the Tamil headlines. The dataset also specifies the author of the headlines. It contains a total of 5273 records with 2949 fake headlines, and 2324 real ones, making it a balanced dataset.

Another dataset used in this paper is specified in [8] which consits of a number of dravidian languages. Amongst those we have used the tamil dravidian fake segment of the dataset. This contains a total of 6277 records with 3195 real news and 3082 fake news in tamil. The dataset consists of news over the period of January 2021 to March 2022 in a wide range of news categories namely, sports, political, celebrity, COVID-19, educational and business news. The performance of the approaches in both datasets are compared

4.2 TamilBERT

The Tamil headlines given in the dataset were preprocessed to remove all digits, links and other unnecessary punctuation to retain only vital information. The next step was to invoke the tokenizer defined as a part of TamilBERT [12], to tokenize the headlines in the training dataset.

The model is designed with two input layers, one for tokenized text sequences and another for an attention mask. These inputs are essential for handling tokenized text data and ensuring that the model focuses on relevant tokens during processing. The tokenized sequences and attention mask are combined to a dictionary, which is passed into the TamilBERT model to get a pooled output, which captures the crucial contextual information in the input text.

For classification, additional layers are introduced. A dropout layer with a rate of 0.2 is applied to mitigate overfitting. Subsequently, a dense layer with 64 units and a hyperbolic tangent (tanh) activation function is added to introduce non-linearity and feature transformation. Another dropout layer with a rate of 0.2 follows. Finally, a dense layer with a single unit and a sigmoid activation function is included. This layer produces a binary classification output, where the output value represents the likelihood of the input text being fake or real.

The mentioned model trains a total of 237605505 parameters and achieves an accuracy of 99.19% in training and 97.37% in testing.Furthermore, based on k-fold cross validation, the model on an average contains an accuracy of 98.9%. Fig 2 depicts a comparison between the accuracy obtained by the model in the training and testing phase of each epoch.

Fig 3 shows a representation of the loss in the model for each of the epochs run, in the training as well as testing phase.

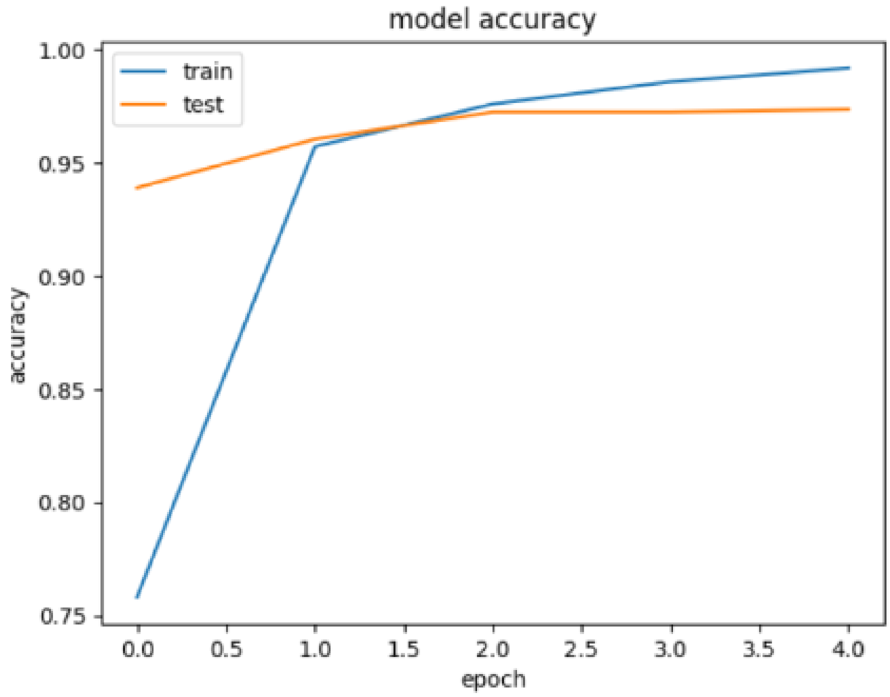

Fig. 2. TamilBERT model accuracy curve

4.3 LaBSE

As a part of preprocessing, punctuation, emojis and stop words were removed from the headlines. Subsequently, label encoding is performed, assigning numerical values to facilitate binary classification.

Two methods of feature extraction were performed. Firstly, sentence transformers in the Language-agnostic BERT Sentence Embedding (LaBSE) model are utilized to generate pre-trained embeddings to capture the semantic context of the text. Secondly, TF-IDF features were extracted to capture information regarding word frequency. The two feature sets are combined to enhance the model's ability to comprehend and classify text accurately.

Models such as SVM with a RBF kernel are trained on the fused feature set, capitalizing its versatility in classifying fake news from real ones. Additionally, a random forest model was also used, aimed at improving the accuracy of the application.

4.4 XLNet

In this model, after the train test split, the encoding is done ,where input_ids and attention_masks are defined .where input_ids are a list of token ids that

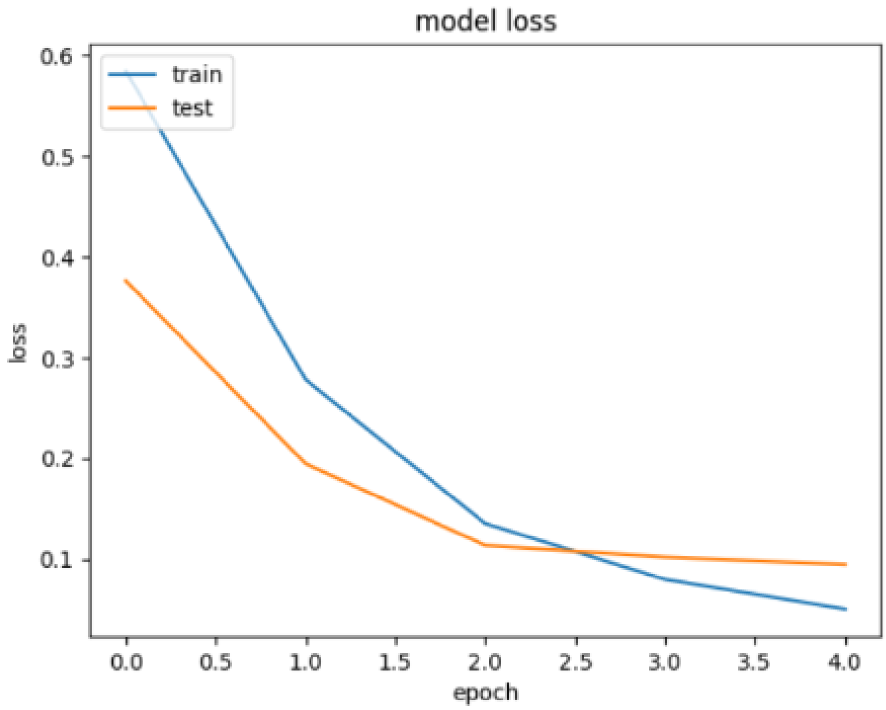

Fig. 3. TamilBERT model loss curve

the model will be supplied and attention_masks, list of indices defining which tokens the model should pay attention to . Now pretrained XLnet model is and after adding layers, it is compiled with Adam optimizer . The created model is compiled and to be fit in 4 epochs. The train and validation accuracies were found to be 95% and 85% respectively. Fig 4 represents the Binary Entropy Loss on each epoch for training and validation dataset

Fig 5 represents the accuracy of the prediction vs actual value for training and validation dataset. The accuracy and Loss for Training dataset are 93% and 0.2 respectively while that of validation dataset are 85% and 0.4 respectively, Overall the results proves it is a generalized model neither overfitting or underfitting for data values.This model has over 116000000 trainable parameters trained on dense fully connected hidden layers

Table 1 shows the results obtained by the models used on the dataset considered [7]. We observe that for this dataset, Bert based language models prove to be efficient in terms of computation intensity and accuracy.

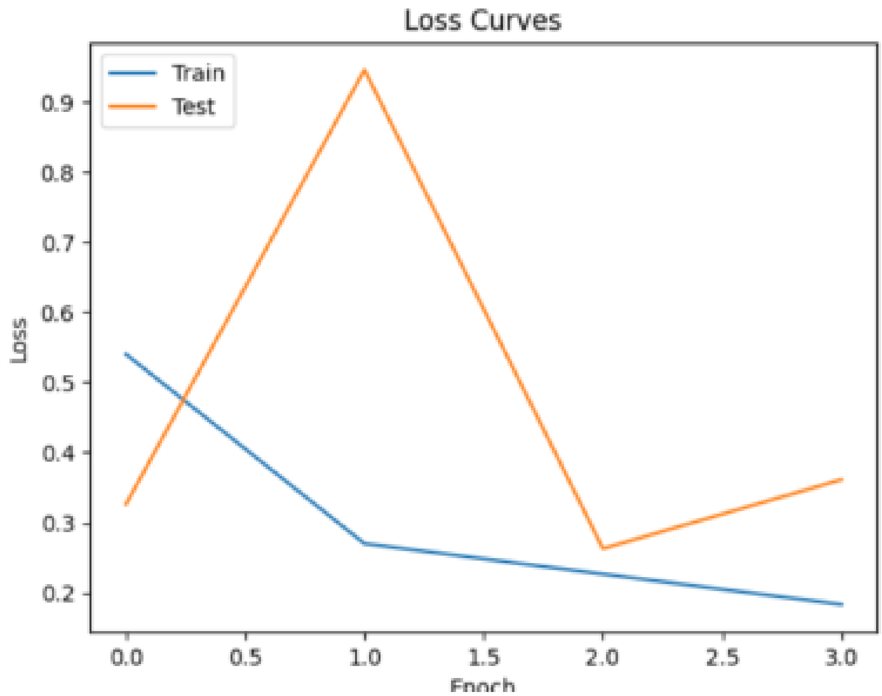

Fig. 4. XLNet model loss curve

Table 1. Performance obtained in each approach

Name	TamilBERT	LaBSE
Accuracy	0.989	0.99
Precision	0.97	0.99
Recall	0.97	0.99
F1 Score	0.97	0.99

5 Conclusion and Future Work

The results achieved enable us to draw the conclusion that the BERT variation models (TamilBERT, LaBSE) have better performance compared to XLNet model . The LaBSE model is much more computationally intensive than TamilBERT. This is due to the fact that two separate features need to be extracted for any Tamil headline and combined for the LaBSE model. The TamilBERT model however, requires no such extraction, and the tamil headline can be directly fed to the model to obtain the classification. This makes TamilBERT more appealing and suitable for real time applications, as it requires fewer computations while giving results similar to LaBSE. There are many other transformer based language models like Robert-a, featMLP that can be used for this classsification. The scope of this research can be extened to other low resource languages as well

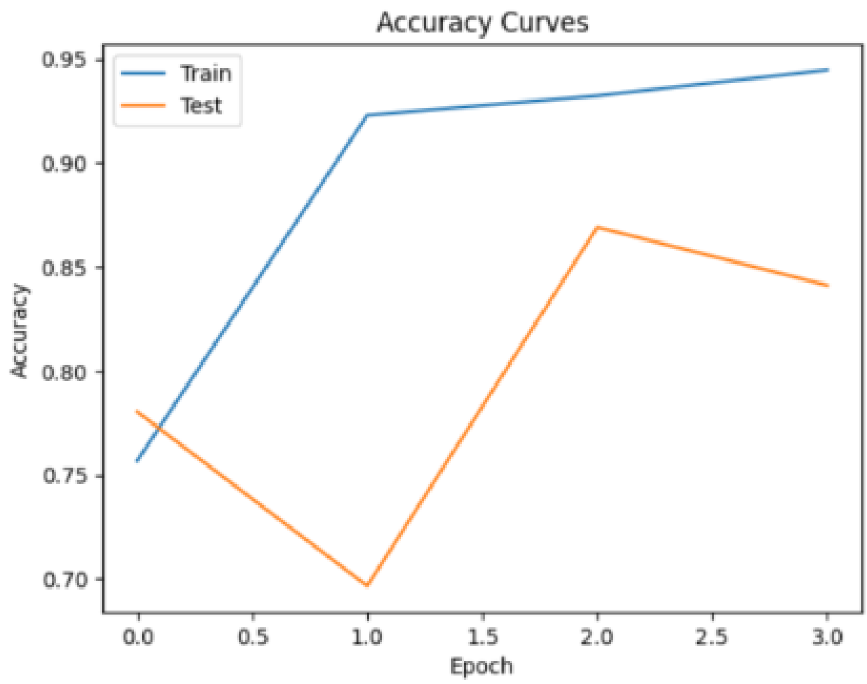

Fig. 5. XLNet model loss curve

with additional preprocessing of textlike including category of news, removing emojis, links etc.

References

1. Moutaz Alazab, Albara Awajan, Ammar Alazab, Ansam Khraisat, Abeer Alhyari, and Reem Saadeh. Fake-news detection system using machine-learning algorithms for arabic-language content. *Journal of Theoretical and Applied Information Technology*, 100(16):14, 2022.
2. Marius Cristian Buzea, Stefan Trausan-Matu, and Traian Rebedea. Automatic fake news detection for romanian online news. *Information*, 13(3), 2022.
3. RamakrishnaIyer LekshmiAmmal Hariharan and Madasamy Anand Kumar. Impact of transformers on multilingual fake news detection for tamil and malayalam. In Anand Kumar M, Bharathi Raja Chakravarthi, Bharathi B, Colm O'Riordan, Hema Murthy, Thenmozhi Durairaj, and Thomas Mandl, editors, *Speech and Language Technologies for Low-Resource Languages*, pages 196–208, Cham, 2023. Springer International Publishing.
4. Raviraj Joshi. L3cube-hindbert and devbert: Pre-trained bert transformer models for devanagari based hindi and marathi languages, 2023.

5. Heejung Jwa, Dongsuk Oh, Kinam Park, Jang Mook Kang, and Heuiseok Lim. exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). *Applied Sciences*, 9(19), 2019.
6. Jozef Kapusta and Juraj Obonya. Improvement of misleading and fake news classification for flective languages by morphological group analysis. *Informatics*, 7(1), 2020.
7. T. T. Mirnalinee, Bhuvana Jayaraman, A. Anirudh, R. Jagadish, and A. Karthik Raja. A novel dataset for fake news detection in tamil regional language. In Anand Kumar M, Bharathi Raja Chakravarthi, Bharathi B, Colm O'Riordan, Hema Murthy, Thenmozhi Durairaj, and Thomas Mandl, editors, *Speech and Language Technologies for Low-Resource Languages*, pages 311–323, Cham, 2023. Springer International Publishing.
8. Eduri Raja, Badal Soni, and Samir Kumar Borgohain. Fake news detection in dravidian languages using transfer learning with adaptive finetuning. *Engineering Applications of Artificial Intelligence*, 126:106877, 2023.
9. Shaina Raza and Chen Ding. Fake news detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics*, 13:1–28, 05 2022.
10. Mateusz Szczepański, Marek Pawlicki, Rafał Kozik, and Michał Choraś. New explainability method for bert-based model in fake news detection. *Scientific Reports*, 11, 12 2021.
11. Apurva Wani, Isha Joshi, Snehal Khandve, Vedangi Wagh, and Raviraj Joshi. pages 153–163. Springer International Publishing, 2021.
12. Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc Le. Xlnet: Generalized autoregressive pretraining for language understanding, 06 2019.