Enhanced BERT Model for Identifying Fake News

Report submitted to the SASTRA Deemed to be University as the requirement for the course

MAT499: PROJECT PHASE - I

Submitted by

NAME: NANDHINI DEVI S
Reg. No: 125150033

November 2024

SCHOOL OF ARTS, SCIENCES, HUMANITIES AND EDUCATION
THANJAVUR, TAMIL NADU, INDIA – 613 401

SCHOOL OF ARTS, SCIENCES, HUMANITIES AND EDUCATION

THANJAVUR – 613 401

Bonafide Certificate

This is to certify that the report titled "**Enhanced BERT Model for Identifying Fake News"** submitted as a requirement for the course **MAT499: PROJECT PHASE - I** for M.Sc. Data Science programme, is a bonafide record of the work done by (**Ms. NANDHINI DEVI S, Reg. No:125150033**) during the academic year 2024-2025, in the School of Arts, Sciences, Humanities and Education, under my supervision.

Signature of Project Supervisor :

Name with Affiliation : Dr.SwaminathanVenkataraman M.Sc., M. Phil., Ph.D.,

Date : 18.11.2024

Project *Viva voc*e held on ________________

Examiner 1 **Examiner 2**

SCHOOL OF ARTS, SCIENCES, HUMANITIES AND EDUCATION

THANJAVUR – 613 401

Declaration

I declare that the report titled "**Enhanced BERT Model for Identifying Fake News"** submitted by me is an original work done by me under the guidance of **Dr. Swaminathan Venkataraman M.Sc., M.Phil., Ph.D.,** during the third semester of the academic year 2024-2025, in the **School of Arts, Sciences, Humanities & Education**. The work is original and wherever I have used materials from other sources, I have given due credit and cited them in the text of the report. This report has not formed the basis for the award of any degree, diploma, associate-ship, fellowship or other similar title to any candidate of any University.

Signature of the candidate(s) : S. Nandhinidevi

Name of the candidate(s) : Nandhini Devi S

Date : 18.11.2024

Acknowledgements

My sincere thanks to Prof **R. Sethuraman**, Chancellor, SASTRA Deemed to be University for facilitating us to do this project.

I am grateful to our Vice Chancellor **Dr. S. Vaidhyasubramaniam**, SASTRA Deemed to be University) for being a source of inspiration.

I thank our Registrar **Dr. R. Chandramoulli**, SASTRA Deemed to be University for encouraging and supporting me for this project.

I sincerely thank our Dean **Dr. K. Uma Maheswari**, Dept. of SASHE, SASTRA Deemed to be University forencouraging our endeavours for this project.

I am grateful to my project guide **Dr. Swaminathan Venkataraman**, M.Sc., M.Phil., Ph.D., SASTRA Deemed to be University for his valuable suggestions, guidance, constant supervision and supporting me in all stages for the successful completion of this project.

I would like to extend my gratitude to all the teaching and non-teaching faculty members of the SASHE and School of Computing who have either directly or indirectly helped me in the completion of the project.

TABLE OF CONTENTS

List Of Figures

List of Tables

ABSTRACT

Fake news on social media is a growing problem, often spread by unreliable sources. Detecting fake news requires advanced tools, or NLP models like BERT are proving to be very effective. When tested against standard models, pretrained BERT performs better and learns faster. Using a large dataset from Kaggle with 21,417 real and 23,502 fake news, BERT accuracy has improved, achieving a high precision and F1 score of 99.96%. BERT models help reduce errors and the study also focuses on Tamil news, showing BERT's potential in regional languages. Using a balanced Tamil news dataset from GitHub with 5,273 labeled articles (2,949 fake and 2,324 real), the fine-tuned model reached an accuracy and F1 score of 83.43%. The other transformer models like DistilBERT, RoBERTa, ALBERT, and DeBERTa are also implemented, but BERT achieved better accuracy for the Tamil news dataset. Overall, this study shows that a transformer model can help identify fake news across different languages.

Keywords*: Bidirectional encoder representations from transformers (BERT), fake news, pretrained model, social networks.*

Chapter 1

1. INTRODUCTION

In recent years, the issue of fake news on social media has become a serious problem as numerous unqualified sources continue feeding the mills of misinformation. Thus, proper identification and management of this misinformation become an important concern in the instance of trustworthy information online. In this scenario, deep learning models, among which BERT (Bidirectional Encoder Representations from Transformers) have been found to be very effective in detecting fake news. Based on the concept of transfer learning, BERT showed a tremendous improvement compared to traditional baseline models with excellent accuracy, not to mention its relatively less training time.

Training BERT on an enormous dataset of 23,502 fake and 21,417 real news articles from Kaggle resulted in high accuracy with very high F1 values up to 99.96%. This model performed better in terms of accuracy while minimizing the false positives and negatives rate. Another important thing is regional implementation of identifying false news. The area that shows importance is in Tamil media, which shares common issues with misinformation. Fine-tuning of a deep learning model was applied on the balanced dataset of 5,273 Tamil news articles retrieved from GitHub, out of which 2,949 were labeled as fake and 2,324 as real. The model attained an accuracy and F1 score of 83.43%; in evidence that, in fact, machine learning is effective in supporting the spread of information across languages and regional contexts.[1]

1.1 LITERATURE SURVEY

S.No	Title	Author(s)	Year	Journal Name	Remark
1.	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.	J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova.	2018	Arxiv	BERT Bidirectional Encoder Representations from Transformers can be a very powerful language model that learns through context in both ways. It can achieve top performance on various kinds of language tasks like question answering, without needing any major changes for every task.
2.	Combat COVID-19 infodemic using explainable natural language processing models.	Jackie Ayoub a, X. Jessie Yang b, Feng Zhou a.	2021	Information Processing & Management	Misinformation about COVID-19 on social media is successfully detected using a model that uses DistilBERT and SHAP. It helps increase public confidence in COVID-19 information since it is accurate and provides an explanation for its forecasts.
3.	Fake news detection on social media: A data mining perspective.	K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu.	2017	Arxiv	fake news is spreading on social media, but it is quite difficult to detect due to decoy content and

					intricate user interactions. This literature review discusses the methods and future directions to improve the detection of fake news on social media.
4.	Fake News Detection Using Enhanced BERT.	Shadi A. Aljawarneh and Safa Ahmad Swedat.	2022	IEEE Transactions on Computational Social Systems	Fake news is a growing issue on social media. It would thus be interesting to see if pretrained models like BERT still outperform others and train faster for detecting fake news in the following study, reporting an accuracy and F1 score of 99.96% on a dataset of real and fake news articles.
5.	Fake news detection in Arabic tweets during the COVID-19 pandemic.	A. B. Nassif, A. Elnagar, O. Elgendy, and Y. Afadar.	2020	International Journal of Advanced Computer Science and Applications	The paper focuses on detecting fake news in Arabic tweets related to COVID-19 using advanced analysis techniques.
6.	Influence of fake news on Twitter during the 2016 U.S. Presidential election.	A. Bovet and H. A. Makse.	2019	Nature Communication	The paper analyzes the impact of fake news on Twitter during the 2016 U.S. Presidential election.

7.	Hybrid real-time protection system for online social networks.	M. B. Yassein, S. Aljawarneh, and Y. Wahsheh.	2019	Foundations of Science	The paper presents a hybrid real-time protection system designed to safeguard online social networks from various threats.
8.	Context-based fake news detection model relying on deep learning models.	E. Amer, K.-S. Kwak, and S. El-Sappagh.	2022	Electronics	Electronics involves the design and use of circuits and devices for controlling and processing electrical signals.
9.	Influence of fake news on Twitter during the 2016 U.S. Presidential election.	A. Bovet and H. A. Makse	2019	Nature Communications	The study examines how fake news on Twitter influenced public perception during the 2016 U.S. Presidential election.

Table 1.1 Literature survey

1.2 PROBLEM STATEMENT

1. To use transformer models to detect fake news in English.
2. To use transformer models to detect fake news in Tamil.

1.3 OBJECTIVES

1. To Implement text preprocessing for Tamil fake news detection, in Tamil language dataset.
2. To implement text preprocessing for English fake news detection, in the English language.
3. To Develop an enhanced BERT Model for fake news detection.

1.3 EXISTING SYSTEM

Most of the present fake news detection models rely primarily on traditional machine learning or early deep learning models. Traditional methods to categorize false information include the Naive Bayes, Support Vector Machines-SVM, and Random Forest that are mostly based on such basic features as word frequencies or specific keywords relating to fake news. However, these models suffer at large dataset scales and fail in grasping the deeper language and events context required for the accurate detection of fake news. Another area of methods used is deep learning techniques in the shape of CNNs and RNNs, which can capture much more patterns in text than might be possible to obtain from hand-crafted features. Although these models have provided some betterment, they still lack insight into capturing complex structures in sentences and overlook vital contextual information. Initially, some pretrained models, such as Word2Vec and GloVe, found their way to capture relationships among words. These, however, make words understandable only in isolation rather than full capture of context at the sentence level. Because of these, traditional and early deep learning systems often require extensive manual feature engineering and are not able to easily handle the complexity that fake news detection poses. For this reason, they are less accurate on real-world datasets where nuances of language and context play a major role in distinguishing between false information and real news.[1]

1.4 PROPOSED SYSTEM

The proposed system aims to implement Tamil identification of fake news using advanced transformer models, specifically enhancing the BERT model. Key elements of the system include text preprocessing for Tamil-language data using the Natural Language Toolkit (NLTK) and creating an improved BERT model. The balanced dataset, which contains 5,273 manually labeled Tamil news articles (both fake and real), was sourced from GitHub. The enhanced BERT model was fine-tuned, and unfreezing the final layer allowed for better text representation and context understanding. This system achieved an accuracy and F1 score of 83.42%, surpassing the performance of other models in Tamil fake news detection.

Chapter 2

2. METHODOLOGY

2.1 WORKFLOW

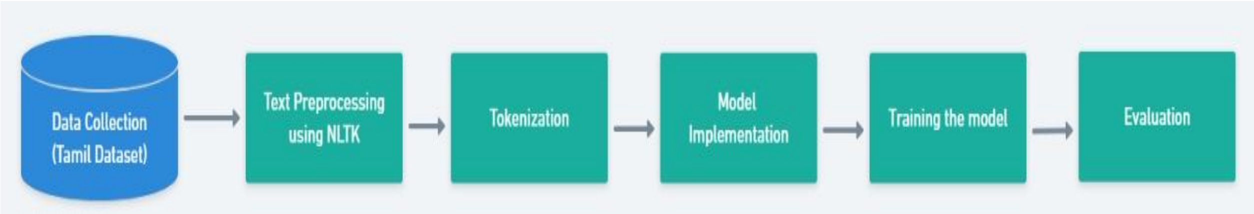

Fig 2.1 Flow Chart

2.2 ENHANCED BERT MODEL

BERT is an opensource framework for machine learning in NLP. BERT was designed with the purpose of teaching computers What unclear language means within the text using surrounding text for context. The BERT framework was pre-trained from Wikipedia texts and can be adjusted to both question-and-answer data sets.

BERT stands for Bidirectional Encoder Representations from Transformers. As it is a deep learning model, transformers are the basis of it. Each element in the output is connected to all the input elements. Their weightings depend on their connections.

Unlike before, where language models only read text one way: either from left to the right or from right to left, with no exception of both being done in one line, BERT works by reading it both ways in one line. This is because the transformer models made this possible and the capability is called bidirectionality. What BERT uses bidirectionality for are two related tasks in the area of NLP: masked language modeling (MLM) and next sentence prediction (NSP).

The goal of MLM training is to hide a sentence's words. Then, the program tries to guess what the hidden word is by looking at the context around it. The goal of NSP training is to see if the program can tell if two sentences are connected in a logical way or if they are just random.[10]

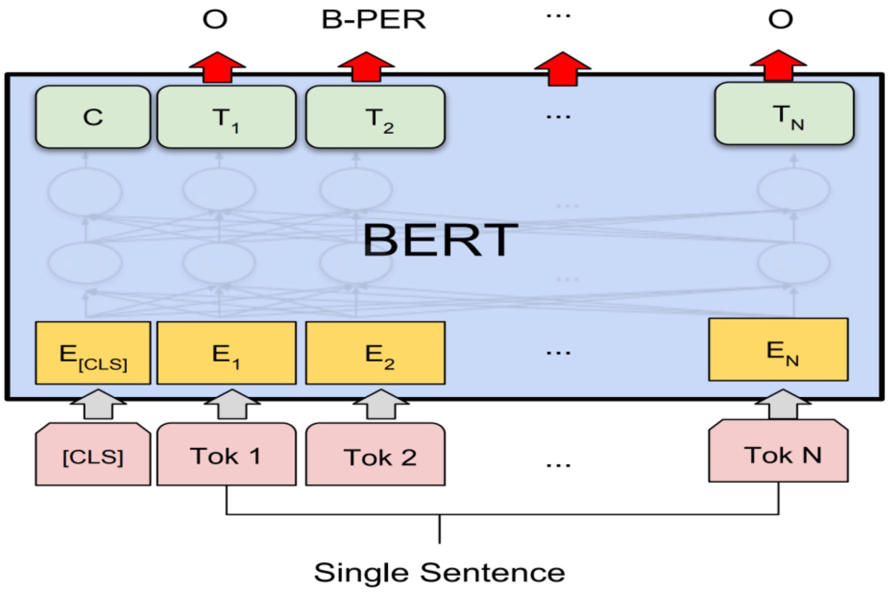

Fig 2.2 Bert Architecture [11]

2.2.1 ENHANCED BERT WORKS

Any NLP technique is intended to comprehend human language as people speak. With BERT, the goal could be a missing word that has to be guessed. For this to happen, models typically learn from a very large set of specific, labeled training data. In practice, heavy workup on hand-labeling the data by linguists is necessary.BERT was pre-trained on simply only a corpus of unpadded, unlabelled plain text, all of English Wikipedia and the Brown Corpus. It continues to learn through unsupervised learning from unlabelled text as its quality improves even as it is being used in real-world applications like Google search.

BERT's pretraining provides it with a basic understanding which will serve it in making answers. From that base, BERT can be adjusted to the continually evolving searchable content and questions and become tailored to a user's wants. That is referred to as transfer learning.[11]

2.2.2 TRANSFORMERS

The transformer is that component of the model which further enables BERT in better understanding the contextual and ambiguous meaning involved while communicating. While processing a sentence, the transformer not only sees each word in isolation but with regard to each other word as well. Thus, with consideration of all phrases around it, the transformer helps BERT understand the complete meaning of the word and, hence, the searcher's requirement also.

This is very different from the traditional method of language processing, which was known as word embedding. This approach has previously been used in models such as GloVe and word2vec. It would link each word to a vector that presented only one part of that word's meaning.[11]

2.2.3 ENHANCED BERT TOKENIZER

The BERT tokenizer utilizes a method called subword-based tokenization. Subword-tokenization breaks unknown words into smaller words or letters so that the model can get some meaning from the tokens. For example, 'boys' is split into 'boy' and 's'. BERT uses a workpiece algorithm in order to create its vocabulary. The workpiece algorithm makes subwords based on how likely characters are to appear together.[12]

2.2.4 MASKED LANGUAGE MODELING

Word embedding models need large amounts of structured data. They are really good at many typical NLP tasks; however, they do worst regarding question-answering, that relies in huge measure on the context and prediction because all the words are tied to a specific vector or meaning.MLM is applied to BERT such that it prevents the term under consideration from both seeing itself and possessing a set meaning that doesn't change with its context. In BERT, words derive meanings strictly from the words surrounding them rather than a fixed identity.[12]

2.2.5 SELF-ATTENTION MECHANISMS

BERT applies a self-attention system that finds and understands the relationships between the words in a sentence. This is due to the key components of BERT's design, namely the bidirectional transformers. This is crucial because sometimes a word can have a different meaning as the sentence continues to unfold. Each new word adds to The NLP algorithm's overall interpretation of the term it is concentrating on. In general, the more words used in every phrase or sentence, the more vague the focus word is.BERT takes into consideration the extra meaning by reading in both directions, taking into account every additional word in a phrase might affect the key term and removing that which would push left-to-right flow in making words gather as a statement develops toward a particular meaning.

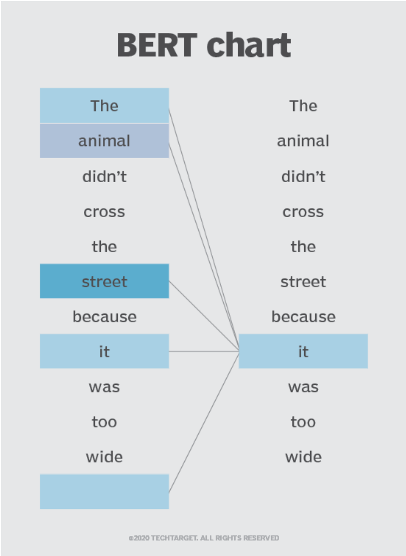

Fig 2.3 Self-Attention Mechanisms [12]

In the example above, BERT is trying to determine what the word "it" is referring to in the sentence: it's an animal not a street. Using a self-attention mechanism, the model then compares the options. The option with the largest score is considered the best fit. Here The word "it" means "animal" and not "street." Had this phrase been a search query the results would have demonstrated such better and more precise comprehension BERT reached.[12]

2.2.6 NEXT SENTENCE PREDICTION

NSP is a training technique that enables BERT to predict whether one sentence could follow another. This tests how well BERT understands the relationships between sentences. BERT is exposed to pairs of sentences that are appropriately matched and others that are not, where it learns to be able to distinguish between them. With time, BERT becomes more effective at making accurate predictions of subsequent sentences. Generally, both of these NSP and MLM techniques are applied simultaneously.[12]

2.3 RoBERTa MODEL

RoBERTa (short for "Robustly Optimized BERT Approach") is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which was developed by researchers at Facebook AI. Like BERT, RoBERTa is a transformer-based language model that uses self-attention to process input sequences and generate contextualized representations of words in a sentence. One key difference between RoBERTa and BERT is that RoBERTa was trained on a much larger dataset and using a more effective training procedure.Specifically, 160GB of text—more than ten times the size of the dataset used to train BERT—was used to train RoBERTa. RoBERTa also employs a dynamic masking strategy during training, which aids in the model's acquisition of stronger and more broadly applicable word representations.

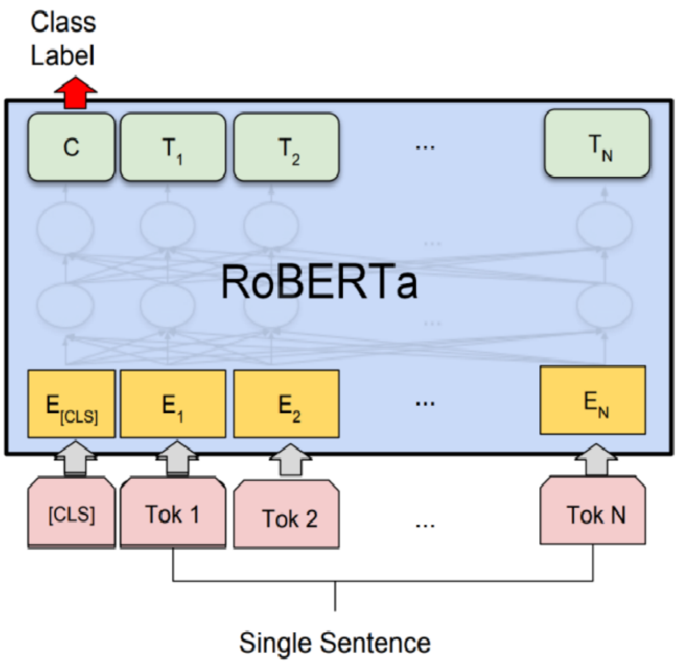

Fig 2.4 RoBERT Architecture[13]

It has been demonstrated that RoBERTa performs better than BERT and other cutting-edge models on a range of natural language processing tasks, including as question answering, text classification, and language translation. It is now a common option for both academic and commercial applications, and it has served as a foundation model for numerous other effective

NLP models. All things considered, RoBERTa is a strong and efficient language model that has advanced a variety of applications and significantly advanced the field of natural language processing. Robustly Optimized BERT Pre-training Approach is what RoBERTa stands for. Researchers from Washington University and Facebook presented it. In order to reduce pre-training time, this work aimed to optimize the BERT architecture's training.[13]

2.4 DISTILBERT MODEL

The DistilBERT model appeared on the blog Smaller, faster, cheaper, lighter: Introducing A distilled form of BERT is called DistilBERT. Together with the paper DistilBERT is a distilled BERT that is lighter, quicker, less expensive, and smaller. TheDistilBERT is a compact, quick, affordable, and lightweight Transformer model that was trained using BERT base distillation. Compared with the google-bert/bert-base-uncased model, DistilBERT has 40% fewer parameters than the former, runs 60% faster and preserves more than 95% of BERT results as determined by the GLUE language comprehension test.[14]

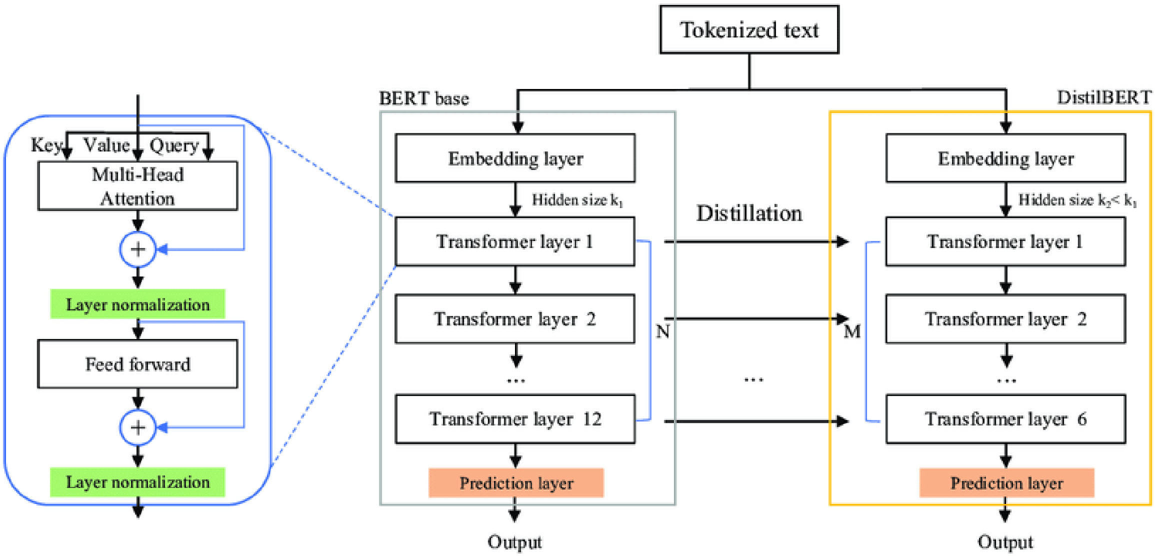

Fig 2.5 DistilBERT Architecture[14]

2.5 ALBERT MODEL

In 2018, Google AI researchers proposed the BERT. Similar to the 2012 AlexNet revolution in computer vision, BERT has produced a similar change in NLP. It makes it possible to use the vast volumes of text data that are available for self-supervised model training.In 2019, Google Research researchers came up with the idea for ALBERT. This paper's objective is to enhance the BERT architecture's training and performance through the use of several methods, such as parameter sharing, embedding matrix factorization, and inter-sentence coherence loss.Utilized dataset Like BERT, ALBERT is pre-trained using the 16 GB of uncompressed data from the English Wikipedia and Book CORPUS datasets.[15]

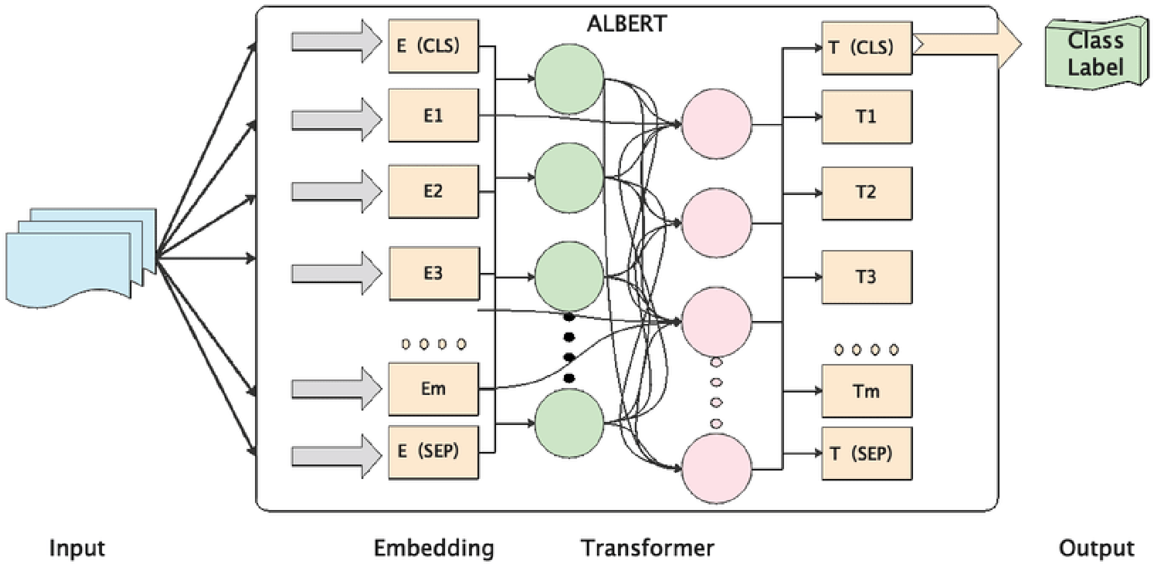

Fig 2.6 ALBERT Architecture [15]

2.6 DEBERTA MODEL

DeBERTa is a Transformer-based neural language model that uses an improved mask decoder and a disentangled attention mechanism to outperform the BERT and RoBERTa models. The mechanism of disentangled attention, the attention weights between words are calculated using disentangle matrices on their relative positions and contents, but each word is represented unmodified using two vectors that encode its position and content, respectively. To anticipate the masked tokens for model pre-training, the output softmax layer is swapped out for the enhanced mask decoder. Furthermore, the model's generalization on downstream tasks is enhanced through fine-tuning using a novel virtual adversarial training technique.[16]

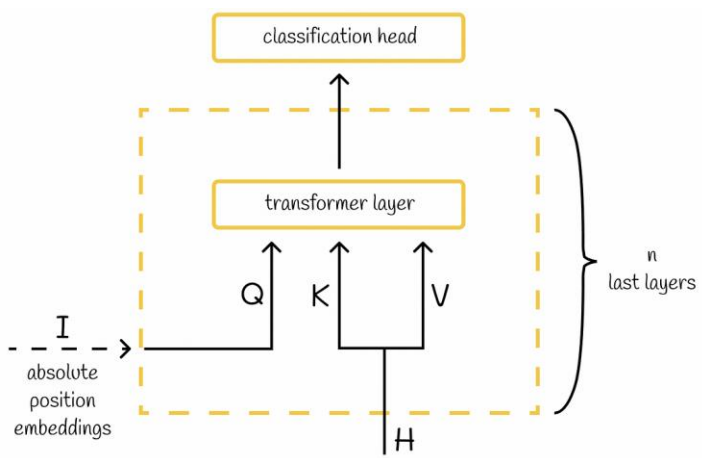

Fig 2.7 Debert Architecture [16]

Chapter 3

3. IMPLEMENTATION

3.1 DATASET

The dataset was downloaded from GitHub
Fake-News-Headlines-In-Tamil/Tamil-News-Headlines.csv at main · AAnirudh07/Fake-News-Headlines-In-Tamil · GitHub

The Tamil fake news dataset is balanced with 5,273 manually labeled news articles (fake and real) Fake news: 2,949 , True news: 2,324
Columns: English Version, Label, News, Author, Date, Authenticity

Column	Description
English Version	English translation of the news article
Label	Categories like Politics, Business, Entertainment, Sports
News	Translated Tamil news article
Author	News writers
Date	Publication date
Authenticity	Whether the news is labeled as fake or true

Table 3.1 Dataset

3.2 TEXT PREPROCESSING

Text Preprocessing Using NLTK (Natural Language Toolkit) : In the proposed system for Tamil fake news detection, text preprocessing plays a crucial role in preparing the data for analysis. Using the Natural Language Toolkit (NLTK), various preprocessing steps ensure that the text is clean and standardized before it's fed into the model. Tokenization is the first step, where text is split into individual words, allowing the model to understand each word separately. Lowercasing is then applied to convert all text to lowercase, ensuring consistency and helping the model treat words like “News” and “news” as the same. Punctuation Removal is performed next, eliminating symbols and marks that do not contribute to the meaning of the text, reducing noise in the data. Stopword Removal follows, where common but insignificant words (such as "and," "the," "is") are removed, so the model can focus on meaningful content. Finally, Stemming or Lemmatization is used to reduce words to their base form. For example, words like “running” and “ran” are reduced to “run,” helping the model recognize different forms of the same word. These preprocessing steps help create a clean and standardized dataset, improving the model's ability to detect fake news accurately.

News	Authenticity	Processed_news
பாஸ்வேர்டை பகிரும் பயனர்களிடம் கூடுதல் கட்டணம...	0	பாஸ்வேர்டை பகிரும் பயனர்களிடம் கூடுதல் கட்டணம்...
இந்தியாவில் நடப்பு ஆண்டின் இறுதிக்குள் ‘ஏகே-2...	0	இந்தியாவில் நடப்பு ஆண்டின் இறுதிக்குள் ‘ஏகே203...
பட்ஜெட் விலையில் மோட்டோ E22s ஸ்மார்ட்போன் இந்...	0	பட்ஜெட் விலையில் மோட்டோ e22s ஸ்மார்ட்போன் இந்த...
கலாம் கண்ட கடைசி கனவை நனவாக்குவோம்	0	கலாம் கண்ட கடைசி கனவை நனவாக்குவோம்
பட்ஜெட் விலையில் ரெட்மி ஏ1+ ஸ்மார்ட்போன் இந்த...	0	பட்ஜெட் விலையில் ரெட்மி ஏ1 ஸ்மார்ட்போன் இந்திய...

Fig 3.1 Text Preprocessing

3.3 PRE-TRAINED BERT TOKENIZER

Tokenization is transforming raw text into a format for processing by a model. BERT tokenizer itself, already pre-trained and optimized to be specific to BERT, provides a subword or token representation of the text that is compatible with the original vocabulary of BERT. Therefore, it can break an out-of-vocabulary word into known subwords, appropriately capture subtle nuances of language, and be coherent with the training of BERT. Tokenization often involves adding special tokens to it, such as [CLS] for classification and [SEP] to separate sentences with added text context to the model.

3.4 TRAINING THE ENHANCED BERT MODEL

In order to identify the patterns that differentiate one class from another, the model must be trained by optimizing its parameters on a labeled dataset. Important training actions consist of Data Splitting: Training and validation (or test) sets are created from the dataset. In order to make sure the model generalizes properly; this division enables it to learn from the training data while also occasionally assessing its performance on unseen data.

Optimization: A loss function computes the discrepancy between the model's predicted and actual labels during training. In order to reduce this loss and steer the model toward improved performance, the optimizer subsequently modifies the model's parameters. Computation is made more efficient by training in phases using methods like batch processing.

Epochs & Iterations: The model iterates across the complete dataset several times during the training process, which spans several epochs. The model can improve its comprehension and lower mistakes with each epoch.

```
Epoch 1/10, Loss: 0.5910
Epoch 2/10, Loss: 0.4753
Epoch 3/10, Loss: 0.4166
Epoch 4/10, Loss: 0.3421
Epoch 5/10, Loss: 0.3080
Epoch 6/10, Loss: 0.2564
Epoch 7/10, Loss: 0.2063
Epoch 8/10, Loss: 0.1738
Epoch 9/10, Loss: 0.1708
Epoch 10/10, Loss: 0.1209
```

Fig 3.2 Epochs and loss

3.5 EVALUATION

The Enhanced BERT model's performance in the actual world can be assessed after training on a different test dataset. Among the important evaluation metrics are:

Accuracy: Indicates how accurately the model predicts outcomes for each class overall. It is defined as the ratio of accurate predictions to all predictions.

Precision: Indicates the percentage of true positive predictions among all of the model's positive predictions, focusing on how accurate positive predictions are.

Recall: Shows the percentage of true positives found out of all actual positives, indicating how well the model catches all pertinent instances of the positive class.

F1 Score: Indicates the harmonic mean by combining precision and recall into a single metric. When working with unbalanced data, the F1 score is very helpful because it balances

		Predicted Class Label	
		Positive	Negative
Actual Class Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{\sum (TP + TN)}{\sum (TP + TN + FP + FN)}$$

$$Precision = \frac{\sum TP}{\sum (TP + FP)}$$

$$Recall = \frac{\sum TP}{\sum (TP + FN)}$$

$$F1\ score = \frac{\sum 2TP}{\sum (2TP + FP + FN)}$$

Fig 3.3 Model Performance Metrics

Chapter 4

4. RESULTS & DISCUSSION

4.1 RESULTS

4.1.1 TAMIL FAKE NEWS

Criteria 1: 70% training, 15% validation, 15% testing.

	Model	Accuracy	Precision	Recall	F1-Score
0	BERT	0.8265	0.8276	0.8265	0.8267
1	RoBERTa	0.6263	0.7130	0.6263	0.5651
2	DistilBERT	0.8036	0.8088	0.8036	0.8013
3	ALBERT	0.6939	0.7314	0.6939	0.6725
4	DeBERTa	0.8163	0.8163	0.8163	0.8160

Table 4.1 Criteria 1 Comparison Table

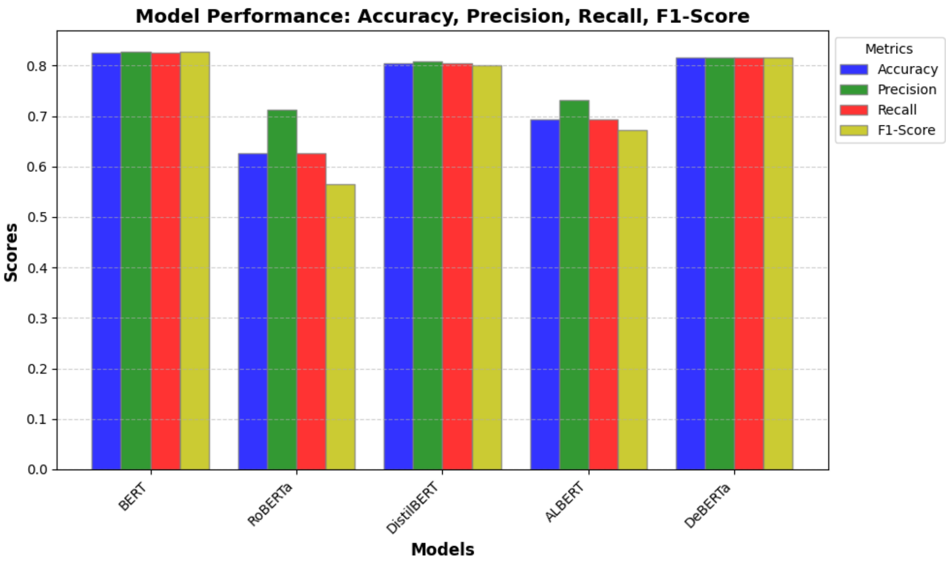

Fig 4.1 Criteria 1 Comparison Chart

Criteria 2: 70% training, 30% testing.

	Model	Accuracy	Precision	Recall	F1-Score
0	BERT	0.8342	0.8362	0.8342	0.8330
1	RoBERTa	0.7455	0.7652	0.7455	0.7353
2	DistilBERT	0.8144	0.8294	0.8144	0.8143
3	ALBERT	0.7315	0.7836	0.7315	0.7254
4	DeBERTa	0.7398	0.7650	0.7398	0.7272

Table 4.2 Criteria 2 Comparison Table

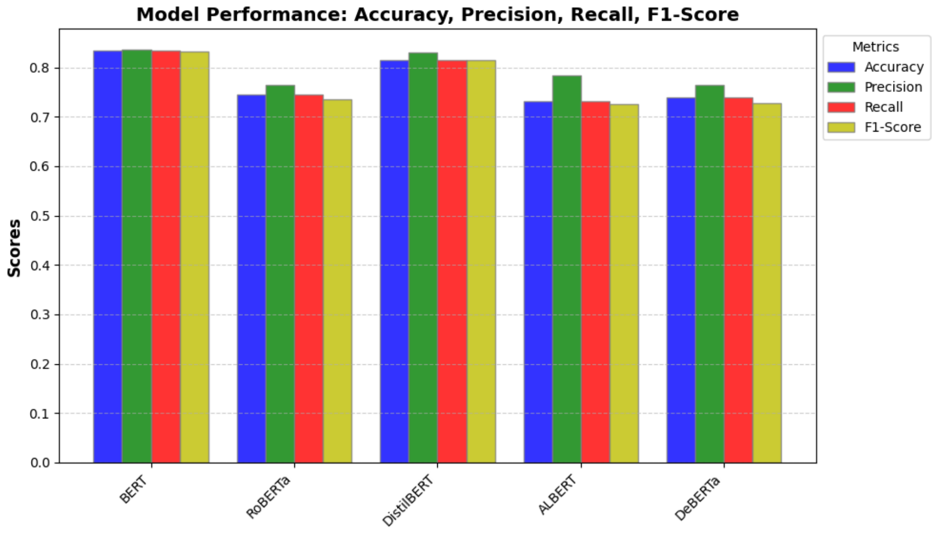

Fig 4.2 Criteria 2 Comparison Chart

4.1.2 ENGLISH FAKE NEWS

Splitting Criteria	Loss	Accuracy	Precision	Recall	F1 Score
First criteria	0.0023	0.9996	0.9996	0.9996	0.9996
Second criteria	0.0030	0.9993	0.9993	0.9993	0.9993

Table 4.3 English Fake News Criteria 1 and Criteria 2

Chapter 5

5. CONCLUSION & FUTURE WORK

5.1 CONCLUSION

The Enhanced BERT model was specifically designed with the intention of detecting fake news among English data. It bases improvements on the foundational capabilities of BERT, adding upgrades that identify more accurate and efficient performances of fake news.In the experiment on Tamil fake news detection, numerous transformer models were experimented, including Enhanced BERT. Among those, Enhanced BERT showed the best score in terms of accuracy, precision, recall, and F1 score. This implies that Enhanced BERT made a better distinction between fake news with fewer mistakes and has captured relevant patterns in both the English and Tamil datasets.

5.2 CHALLENGES & LIMITATION

This Enhanced BERT model obtained a fantastic accuracy and F1 score of 83.42%, which is enough to make this model different from other models towards accurate detection of fake news. Fine-tuning adjustment like freezing the last layer was highly essential for capturing better text representations that improved performance. The process helped the model distinguish relevant characteristics to detect fraudulent news.Using the pre-trained Enhanced BERT also proved to be useful in the task of significantly reducing the time taken for training. The model could take advantage of prior knowledge by the application of the pre-trained layers, which resulted in the model both training much faster and performing very well overall.

5.3 FUTURE WORK

In the future, plans involve exploring pretrained models to identify false information on larger, multilingual datasets by combining datasets in various languages.

Chapter 6

6. REFERENCE

1. S. A. Aljawarneh and S. A. Swedat, "Fake News Detection Using Enhanced BERT," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 4, pp. 4843-4850, **DOI:** 10.1109/TCSS.2022.3223786, 2024.

2. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

3. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.

4. M. B. Yassein, S. Aljawarneh, and Y. Wahsheh, "Hybrid real-time protection system for online social networks," *Foundations of Science*, vol. 25, no. 4, pp. 1095–1124, 2019.

5. X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Information Processing & Management, vol. 57, no. 2, Art. no. 102025, 2020.

6. S. Raza and C. Ding, "Fake news detection based on news content and social contexts: A transformer-based approach," International Journal of Data Science and Analytics, vol. 13, pp. 335–362, 2022.

7. E. Amer, K.-S. Kwak, and S. El-Sappagh, "Context-based fake news detection model relying on deep learning models," Electronics, vol. 11, no. 8, pp. 1255, 2022.

8. A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 U.S. Presidential election," Nature Communications, vol. 10, no. 1, pp. 1–14, 2019.

9. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.

10. Hashemi-Pour, C., & Lutkevich, B. (2024b, February 15). BERT language model. Search Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model

11. Walidamamou. (2024, April 8). Master Named Entity Recognition with BERT in 2024. UBIAI. https://ubiai.tools/mastering-named-entity-recognition-with-bert/

12. Kohli, A. (2023, January 25). Notes on BERT tokenizer and model - Anmol Kohli - Medium. Medium. https://medium.com/@anmolkohli/my-notes-on-bert-tokenizer-and-model-98dc22d0b64
13. GeeksforGeeks. (2023, January 10). Overview of ROBERTa model. GeeksforGeeks. https://www.geeksforgeeks.org/overview-of-roberta-model/

14. DistilBERT. (n.d.). https://huggingface.co/docs/transformers/en/model_doc/distilbert

15. GeeksforGeeks. (2022, January 27). ALBERT A Light BERT for Supervised Learning.

GeeksforGeeks. https://www.geeksforgeeks.org/albert-a-light-bert-for-supervised-learning/

16. Papers with Code - DeBERTa Explained. (n.d.). https://paperswithcode.com/method/deberta

6.1 DATA AVAILABILITY

17. The dataset was downloaded from GitHub. Fake-News-Headlines-In-Tamil/Tamil-News-Headlines.csv at main · AAnirudh07/Fake-News-Headlines-In-Tamil · GitHub