Fake News Detection Using Bert Semantic Similarity

Suyash Sawant[1] Prathmesh Pawar[2] Rutva Patel[3] Suman Parui[4] Prof. Jyoti Dange[5]
[5]Assistant Professor
[1,2,3,4,5]Department of Electronics and Telecommunication Engineering
[1,2,3,4,5]Atharva College of Engineering, Mumbai, India

Abstract— The swift progress of the applications of NLP (Natural Language Processing) techniques particularly in the domain of semantic similarity are quite beneficial for the detection of 'fake news', that is, dishonest news stories that comes from the non-reputable sources. Therefore, to tackle this problem by building a model supported a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) TFIDF matrix, (word tallies relative to however usually they're employed in different articles in your dataset) which will solely get you thus so far. However, these models don't think about the necessary qualities like word ordering and context. It's completely feasible that 2 articles which are similar in their word count are going to be fully totally different in their actual definition. The information science community has responded by taking actions against the matter and also took initiatives like there's a Kaggle competition referred to as because the "Fake News Challenge" and Facebook is using AI to filter faux news stories out of users' feeds. Combatting the faux news may be a classic text classification project with an uncomplicated proposition. So, it is doable for us to create a model that may differentiate between "Real "news and "Fake" news? thus, a planned work on collecting a dataset to train the model to understand context of sentences which can in turn predict faux and real news.

Keywords: Fake News Detection, Bert Semantic Similarity

I. INTRODUCTION

These days' faux news is making totally different problems from grim articles to a made-up news and set up government information in some organizations. Fake news and lack of trust within the media are tremendously growing issues with immense ramifications in our society. Obviously, a deliberately dishonest story is "fake news "however late voluble social media's discourse is dynamical in its definition. A number of them currently use the term to dismiss the facts counter to their most popular viewpoints.

The importance of misinformation among USA political discourse was the topic of weighty attention, notably following the USA president election. The term 'fake news' became common expression for the problem, notably to explain factually incorrect and dishonest articles revealed largely for the aim of constructing cash through page views. Among this paper, we've tried to provide a model that may accurately predict the probability that a given article is faux news. Facebook has been at the geographic point of abundant critique following media attention. they need already enforced a feature to flag faux news on the location once a user see's it ; they have additionally aforesaid in the public that they're engaged on to develop a tool to differentiate these articles in an automatic means. Certainly, it's not a straightforward task. A particular algorithm formula should be politically unbiased – since faux news exists on each end of the spectrum – and additionally give equal balance to legitimate news sources on either end of the spectrum. additionally, the question of legitimacy may be a troublesome one. However, so as to resolve this downside, it's necessary to own an understanding on what Faux News is composed of particularly. Later, it in required to appear into how the techniques within the fields of machine learning, linguistic communication process facilitate general people to discover faux news.

II. OBJECTIVE

The goal of this project is given an article's headline body text, predict whether or not that articles text agrees with the opposite article headlines or disagrees with the headline, discusses a similar topics/claim as per the headline, or is unrelated to the headline. To characterize and analyse faux/fake news threat detection. the most objective is to discover the faux news, that may be a classic text classification downside with a uncomplicated proposition. Therefore, it is required to create a model that may differentiate between "Real" news and "Fake" news. For our project we've adopted and are following the subsequent

A. Methodology:

1) To differentiate between the important and fake news using Bert pretrained model along with finetuning for better accuracy.
2) The Bert model is trained to predict the news input into 3 categories.
   - Contradiction (Fake news)
   - Neutral (The Unbiased news)
   - Entailment (Similar news – Real news)
3) To look for linguistics similarity of user input we have a tendency to use the NewsAPI that may be a straightforward, easy-to-use REST API that returns JSON search results for current and historic news articles revealed by over seventy-five thousand worldwide sources together with Google News Library
4) To realize higher user expertise, we've enforced our model by employing a webapp.

III. GENERAL DESCRIPTION

A. BERT – Transformer Based Model

The abbreviation for BERT is Bidirectional Encoder Representations from Transformers. It comes from a paper published by Google AI Language in 2018. It supports the concept that fine-tuning a pretrained language model will facilitate the model achieve better results in the downstream tasks. BERT is meant to assist computers perceive the meaning of ambiguous language in text by making the use of close text to ascertain context. .It is an open-source machine learning framework derived from natural language (NLP) and from the architecture of Google's Transformer Model. The BERT framework was pre-trained from the corpora of

various data from Wikipedia and later be fine-tuned with question-and-answer datasets. BERT may be also termed as a multi-layer bidirectional Transformer encoder which also supports the functionality of fine-tuning. BERT is derived from Transformers, a deep learning model during which each output part is connected to each input part, and also the weights between them are dynamically calculated primarily based upon their association.

Language models could only read text input sequentially that is they can do processing unidirectionally but couldn't do both at the same time. BERT is different because it's designed to read in both directions directly. This capability, enabled by the introduction of Transformers, is understood as bidirectionality.

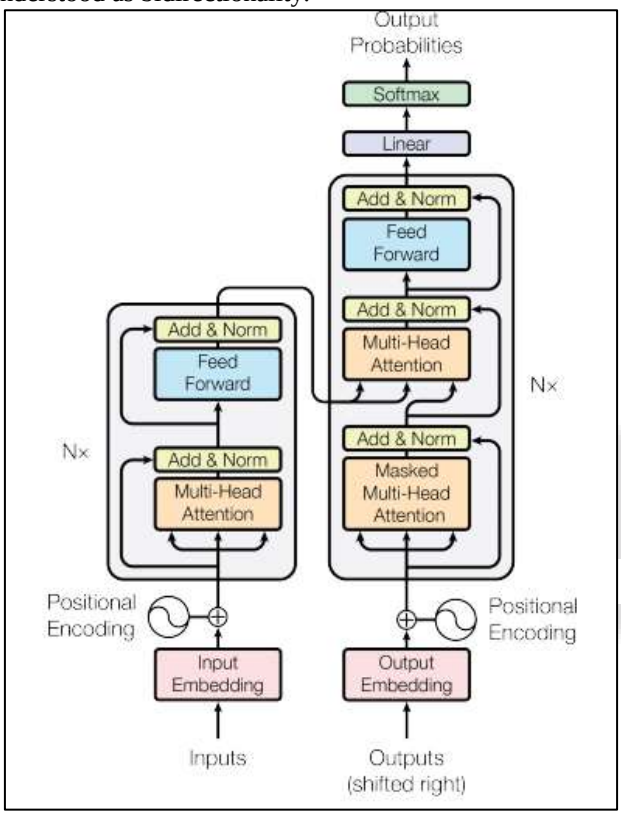


Fig. 1.a: Transformer Model architecture

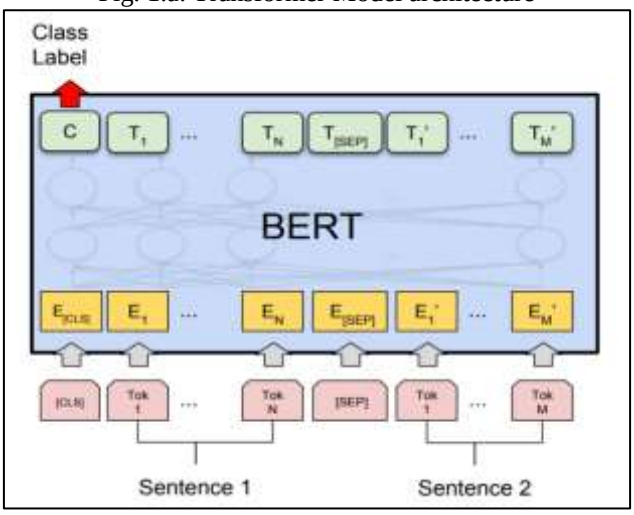


Fig. 1.b: Sentence Pair Classification model architecture

B. BERT – Deconstruction of the Model

Using this bidirectional capability, BERT is pre-trained on two different, but related, NLP tasks: Masked Language Modeling and Next Sentence Prediction.

The purpose of Masked Language Model (MLM) training is to mask a certain word in a sentence and then have the model predict what word was hidden (masked) based on the hidden word's context trained on a large corpus of data. The objective of Next Sentence Prediction training is to possess the program predict whether two given sentences have a logical, sequential connection or whether their relationship is just random.

IV. SOFTWARE IMPLEMENTATION

A. Dataset Used with Flow of Process

The computer code implementation demonstrates the employment of SNLI (Stanford language Inference) Corpus to predict sentence linguistics similarity with Transformers. The SNLI corpus consists of an assortment of 570k human written English sentence pairs manually labelled for balanced classification with the labels Entailment, Contradiction, and Neutral, supporting the task of language inference (NLI), additionally called recognizing textual entailment (RTE). we've solely trained the model for the highest layers to perform "feature extraction", which can permit the model to use the representations of the pretrained model. In this project after creating the Bert model we are going to fine-tune the model for better accuracy. The model will take 2 sentences as inputs which outputs a similarity score for these 2 sentences between the classes Contradiction, Neutral and Entailment.

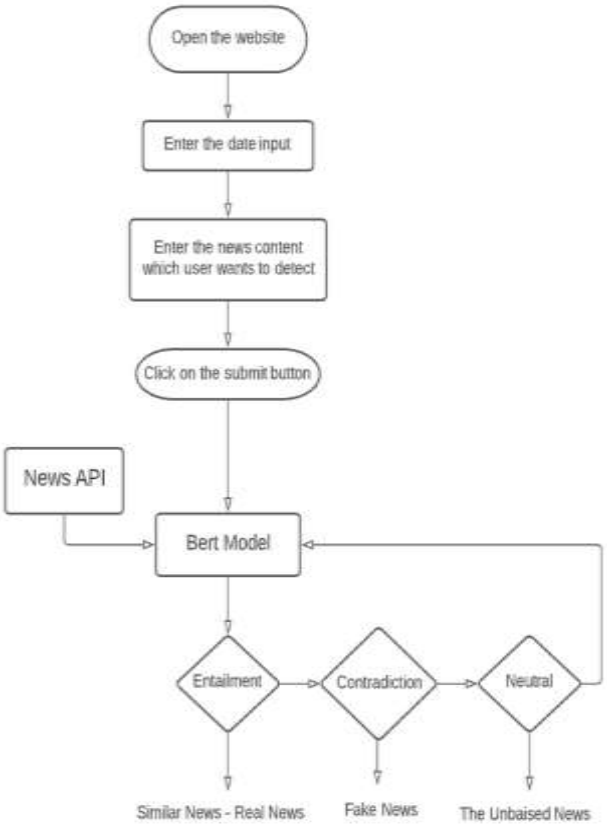


Fig. 2.a: Flow Chart of Webapp

B. Building the Model

After pre-processing the dataset by performing exploratory data analysis and filling out the nan data by taking the weighted average we find the embeddings that are the vector representations of text where word or sentences with similar meaning or context have similar representations. We later generate batches of the data and load our BERT Tokenizer to encode the text. Later we convert batch of encoded features to NumPy array. We will use base-base-uncased pretrained model. .Thus, finally we begin with creating the Bert model. We create the model under a distribution strategy scope. Attention masks indicates to the model which tokens should be attended to Token type ids are binary masks identifying different sequences within the model. We load and freeze the BERT model to reuse the pretrained features without modifying them and Add trainable layers on top of frozen layers to adapt the pretrained features on the new data. Later we apply hybrid pooling approach to bi_lstm sequence output. Training is completed just for the highest layers to perform "feature extraction", which can allow the model to use the representations of the pretrained model. After fine tuning `Bert model` is unfreezed and retrained with a very low learning rate. This can deliver meaningful improvement by incrementally adapting the pretrained features to the new data. We train the model end to end and perform evaluation on test data

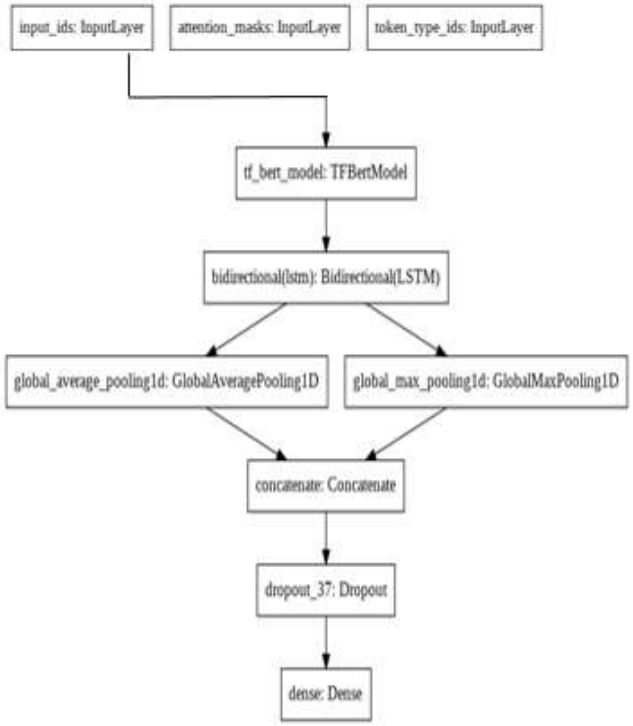


Fig. 2.b: Block Diagram of Bert Model

C. Software used:

1) Jupyter Notebook / Google Colab
2) Pycharm / Visual Studio IDE
3) Sublime

D. Libraries used:

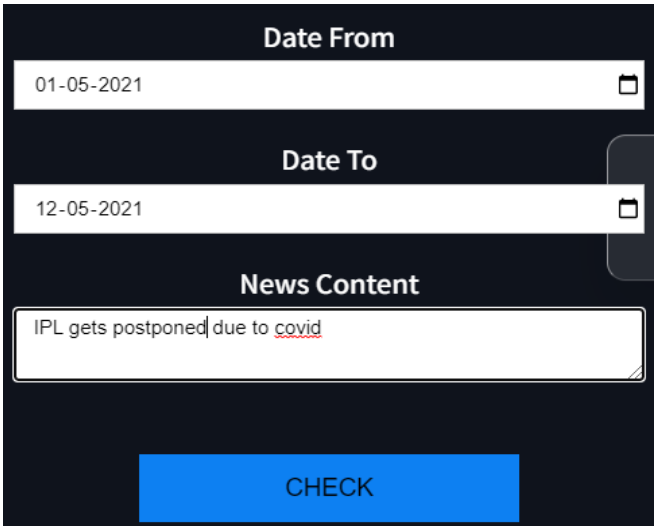
1) Tensorflow
2) Keras
3) Transformers
4) Numpy
5) Pandas
6) GoogleNews and NewsApi
7) Django Web Framework

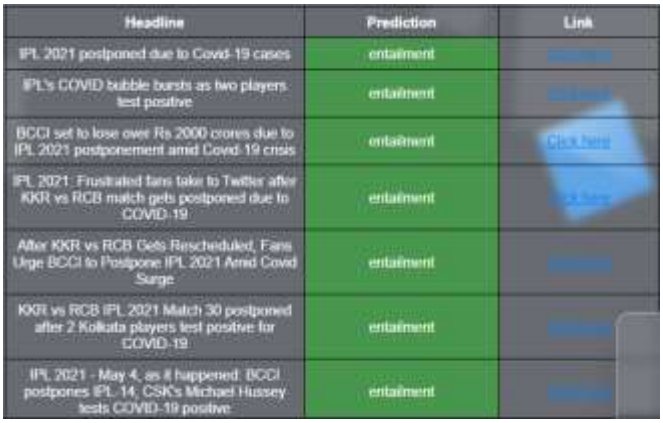
V. RESULTS & OUTPUTS

The Bert model was trained for about 3-4 hours and after finetuning the Bert Model , an accuracy of 80% was achieved on the test data.

A. Fake News

B. Real News





Headline	Prediction	Link
IPL 2021 postponed due to Covid-19 cases	entailment	
IPL's COVID bubble bursts as two players test positive	entailment	
BCCI set to lose over Rs 2000 crores due to IPL 2021 postponement amid Covid-19 crisis	entailment	Click here
IPL 2021: Frustrated fans take to Twitter after KKR vs RCB match gets postponed due to COVID-19	entailment	
After KKR vs RCB Gets Rescheduled, Fans Urge BCCI to Postpone IPL 2021 Amid Covid Surge	entailment	
KKR vs RCB IPL 2021 Match 30 postponed after 2 Kolkata players test positive for COVID-19	entailment	
IPL 2021 - May 4, as it happened: BCCI postpones IPL 14; CSK's Michael Hussey tests COVID-19 positive	entailment	

VI. Applications

The recent speedy progress of neural network-based NLP analysis and research, particularly on learning linguistics text representations, are capable to actually novel merchandise products like sensible Compose and consult with Books. It also can facilitate improve performance on a spread of NLP tasks that have restricted amounts of coaching information, like building sturdy text classifiers from as few as one hundred labelled examples. With the assistance of the linguistics similarity model, we can get alerted about the fake news that's keeps circulating through social media platforms and numerous news channel media to avoid unnecessary controversies and public stunts that are quite common these days.

VII. Conclusion

In this project we've demonstrated that the sentence similarity task in BERT pretraining stage successfully captured semantic information in sentences, and can be used to determine the similarity of two news articles positively and negatively. It can also predict if the sentences are neutral in nature.

In order to attain a more efficient way in terms of computation required in a way to measure document or sentence similarity is doing sentence embeddings (vector form). In contrast to RNN and its variants (e.g., LSTM and GRU), extracting sentence embeddings from transformer models is not as straight forward as they are attention model and are feed forward networks.

Acknowledgement

We would like to thank all the staff members of the Electronic and Telecommunication department of Atharva College of engineering who have provided us with various facilities and guided us in the development of such a sound and effective project idea. Their inputs have invariably helped the project enormously.

Reference

[1] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding by Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova.
[2] Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context by Zihang Dai, Zhilin Yang, Yiming Yang, William W. Cohen, Jaime Carbonell, Quoc V. Le and Ruslan Salakhutdinov.
[3] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification.
[4] Radford, A., & Salimans, T. (2018). Improving Language Understanding by Generative Pre-Training.
[5] Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). Attention is all you need.
[6] Z. Chen, H. Zhang, X. Zhang, and L. Zhao. 2018. Quora question pairs.

