Covid-19 Vaccine Misinformation Detection.

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:

DINESH

2K20/ISY/07

Under the supervision of
Dr. PRIYANKA MEEL
ASSISTANT PROFESSOR

DEPARTMENT OF INFORMATION TECHNOLOGY DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

MAY, 2022

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Dinesh Kharat, Roll No. 2K20/ISY/07 student of M. Tech., Information Systems, hereby declare that the major project titled "Covid-19 Vaccine Misinformation Detection" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of thedegree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

DKharat

Place: Delhi
Date: May 26, 2022

Mr. Dinesh Kharat
(2K20/ISY/07)

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering) Bawana
Road, Delhi-110042

CERTIFICATE

I hereby certify that the Major Project titled "Covid-19 vaccine Misinformation Detection" which is submitted by Dinesh Kharat, Roll No. 2K20/ISY/07 Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi
Date: May 26, 2022

Dr. Priyanka Meel
SUPERVISOR
ASSISTANT PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY

ACKNOWLEDGEMENT

I express my gratitude to my major project guide Ms. Priyanka Meel, Assistant Professor, lT Dept., Delhi Technological University, for the valuable support and guidance she provided in making this major project. It is my pleasure to record my sincere thanks to my respected guide for his constructive criticism and insight without which the project would not have shaped as it has.

I humbly extend my words of gratitude to other faculty members of this department for providing their valuable help and time whenever it was required.

Dinesh Kharat
Roll No. 2K20/ISY/07

M.Tech (Information Systems)

ABSTRACT

In the situation of pandemic there has been a rise in the act of spreading rumors and myths on online social media platforms and Internet blogs. This has caused harm to the society and has caused people to believe in this fake news. In this study we will be working on classifying the news as fake or real and helping people to stay safe from these myths. We are working on classifying the misinformation related to vaccines. There have been rumors since the development of the vaccine started, news about the deaths during the vaccine trials, cost of vaccine, news regarding shortage of the vaccine has been circulating on various platforms. We have referred a dataset which contains tweets regarding vaccine which has keywords such as vaccine, covid-19, covaxin, sputnik etc. These has been trained using various NLP techniques like BERT, combined BERT with CNN and BiLSTM, RoBERTa, ALBERT.

At last, we have compared the results obtained from these models.

Keywords: NLP (Natural Language Processing), BERT, RoBERTa, ALBERT.

CONTENTS

CERTIFICATE III

ABSTRACT v

CONTENTS VI

LIST OF FIGURES VIII

LIST OF TABLES IX

CHAPTER 1 INTRODUCTION 1

1.1 MOTIVATION 2

1.2 OBJECTIVE 4

1.3 ORGANIZATION OF DISSERTATION 4

CHAPTER 2 BACKGROUND AND RELATED WORK 5

2.1 LITERATURE REVIEW 5

2.2 OVERVIEW 6

2.2.1 NATURAL LANGUAGE PROCESSING 6

2.2.2 RECURRENT NEURAL NETWORK 7

2.2.3 LONG SHORT TERM MEMORY (LSTM) 8

2.2.4 CONVOLUTIONAL LSTM 8

2.2.5 Bi-DIRECTIONAL LSTM 8

2.2.6 ABERT NEURAL NETWORK 9

2.2.7 RoBERTA 16

2.2.8 ALBERT 16

2.2.9 ACTIVATION FUNCTION 17

CHAPTER 3 PROPOSED MODEL 21

3.1 METHODOLOGY 21

3.2 DATA COLLECTION 21

3.3 DATA PREPROCESSING 21

3.4 ARCHITECTURE AND IMPLEMENTATION 21

CHAPTER 4 THE EXPERIMENTAL APPROACH 25

4.1 DATA ANALYSIS 25

4.2 DATASET 25

4.3 MODEL TRAINING & VALIDATION 25

CHAPTER 5 EXPERIMENTAL RESULTS 26

5.1 MODEL EVALUATION 26

5.1.1 F1-SCORE 26
5.1.2 ACCURACY 26
5.1.3 ROC-AUC 26
5.2 MODEL PREDICTION 27
5.3 PERFORMANCE ANALYSIS 27
CHAPTER 6 CONCLUSION 28
REFERENCES..........29

LIST OF FIGURES

Figure 1: Fact check by the Wire on misreporting in the year 2020 1
Figure 2: Tweet claiming alcohol can cure virus 3
Figure 3: Claim as coronavirus as an anagram of carnivorous 3
Figure 4: Patient zero to get infected by virus 3
Figure 5: A political tweet in India 3
Figure 6: NLP 7
Figure 7: A hidden state RNN that uses one input item in the sequence to deliver relevant information to others. 7
Figure 8: An LSTM's repeating model containing four interacting layers. 8
Figure 9: Bi - Directional LSTM 9
Figure 10: The Transformer-model architecture 9
Figure 11: LSTM Network 10
Figure 12: Bi-Directional LSTM Network 10
Figure 13: Transformer model architecture 11
Figure 14: Stack of Decoders-GPT 11
Figure 15: Stack of encoders - BERT 11
Figure 16: Pre-training (Pass 1) 12
Figure 17: Fine Tuning (Pass 1) 13
Figure 18: Pre training (Pass 2) 13
Figure 19: Fine tune (Pass 2) 14
Figure 20: Pre training (Pass 3) 14
Figure 21: Initial Embeddings 15
Figure 22: output 15
Figure 23: Linear Function 17
Figure 24: Sigmoid Activation Function 18
Figure 25: Tanh Activation Function 19
Figure 26: ReLU Activation Function 19
Figure 27: SoftMax Activation Function 20
Figure 28: BERT + CNN 22
Figure 29: BERT + BiLSTM 23

LIST OF TABLES

Table 1: Details regarding existing work and it's accuracy ... 6
Table 2: Train and Validation split ... 25
Table 3: Performance analysis ... 27

CHAPTER 1

INTRODUCTION

Covid19 is an infection caused by SARS-CoV-2, a novel corona virus. The World Health Organization (WHO) initially informed the people about this new virus on 31 December 2019, after receiving a report of a cluster of patients of 'Viral Pneumonia' in Wuhan, China. On 30 January 2020, the World Health Organization (WHO) declared it a pandemic [1]. On April 12th, 2022, there were around 500 million diagnoses and 6.18 million deaths reported.

As the number of fraudulent websites on the internet grows, so does the number of fake news, necessitating the development of a classifier that can distinguish between false and true information. Compared to the past, the number of articles is steadily increasing. Because of the worrisome speed with which fake news travels on social media, there is a growing interest in research centered on automated false news identification and fact-checking. Fake news has grown into a global issue that even major internet giants like Facebook and Google are attempting to address. Without further context and human judgment, determining if a sentence is true can be challenging.

Fake news is content that is fabricated and cannot be confirmed from any source, whether it is textual, image, video, or any other form. This information is often fabricated in order to manipulate individuals for political gain or to facilitate business transitions. Fake news is any news that is not factually correct and has been labeled as false. There are numerous examples of fake news, such as claims of a cure for Covid-19, symptoms, source, and the fact that it is a bioweapon. Fig 1. Shows the facts checked done by wire in the year 2020.

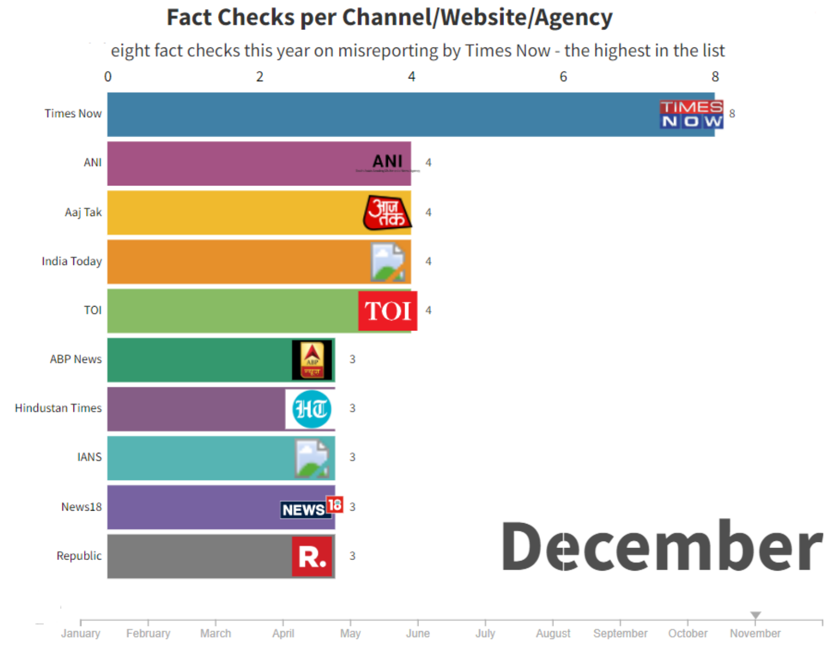


Figure 1: Fact check by the Wire on misreporting in the year 2020

Types of fake news

1. **Misleading:** Information is used to frame a situation or a person.
2. **Imposter:** When legitimate sources are imitated.
3. **False connection:** When the information isn't supported by the headlines, images, or captions.
4. **Satire or Parody:** There is no purpose to harm, but it has the ability to mislead.
5. **False context:** When legitimate content is accompanied by fake contextual data.
6. **Fabricated content:** The news information is completely misleading and intended to deceive and damage people.
7. **Manipulated content:** When real information or images is modified for the purpose of deception.

Reasons for spreading fake news

Fake news is shared by social media users for monetary, political, or amusement reasons. The goal of disseminating fake news is determined by the users who use various platforms to do it. We've gone through a couple of the reasons for disseminating fake news.

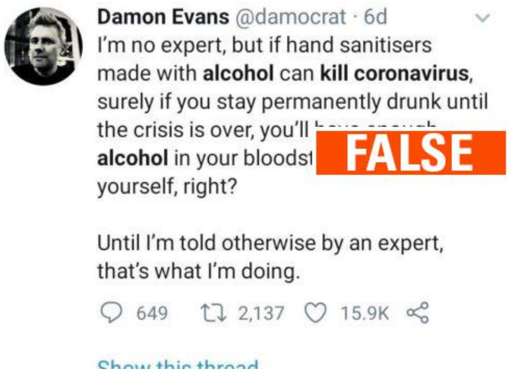
1. **Political propaganda:** Some news is distributed to persuade people to believe a certain point of view and modify their beliefs, as well as to foster hate among religions and communities, all of which have an impact on election results.
2. **Financial benefit:** To increase financial gains by manipulating the price of a stock.
3. **Defaming:** Disseminating false information in order to slander a person and harm their reputation.
4. **Fun:** Fake news can be disseminated in the form of a meme, which is an image or text that is intended to amuse and entertain.

Consequences of Fake News

1. **Political Gain:** Might change the election results, causes disbelief in democracy.
2. **Violence:** Community hatred can be spread and will result in mob violence.
3. **Financial loss:** The news that may cause drop in the sales of goods and services.
4. **Financial gain:** The news that may cause increase in the price of goods and services.
5. **Defaming:** The news will have an impact on the status of the famous personalities and business firms.

1.1. Motivation

Fake news is spread over social media in the form of photographs, text, and videos. The majority of news is built on words and images. As the epidemic spread, many facts, precautions, and medical-related news were shared via social media. As the vaccine was being developed, many assumptions were made about the testing, production, and price, causing public concern about the vaccination. The Fig 2. depicts a tweet from a user stating that alcohol is a cure for coronavirus. This tweet was a hoax since it stated that if hand sanitizer can eliminate the virus, then drinking it will do the same. This tweet had a negative impact, as 44 individuals died as a result of drinking too much alcohol and becoming poisoned. Alcohol is not a cure for coronavirus, according to the WHO. Similarly, in Fig 3. PETA, an animal rights organisation, claimed that consuming dead animals can spread the infection. According to the World Health Organization, consuming meat does not trigger any virus. According to Fig 4. a scientist working at the Wuhan Institute of Virology was the first to get infected with Covid-19. The Institute reported that the woman was fine and that she was living elsewhere, but this was false information. In Fig 5. An MP (Member of Parliament) reported that 70 people attended an event in Delhi, India, of whom 8 were positive for covid-19. She further said that the patients were spitting around and acting inappropriately. The hospital authorities corrected this with a photo of the wards, stating that 33 people attended the event and only three had tested positive for covid.



Damon Evans @damocrat · 6d

I'm no expert, but if hand sanitisers made with **alcohol** can **kill coronavirus**, surely if you stay permanently drunk until the crisis is over, you'll

alcohol in your bloodst

FALSE

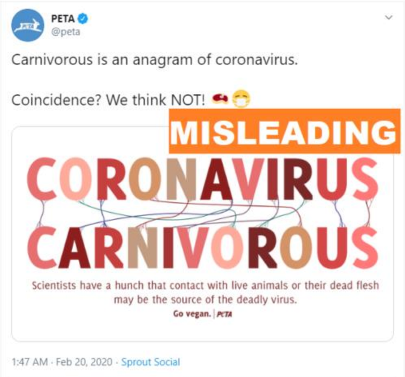
yourself, right?

Until I'm told otherwise by an expert, that's what I'm doing.

649 2,137 15.9K

Show this thread

Figure 2: Tweet claiming alcohol can cure virus



PETA @peta

Carnivorous is an anagram of coronavirus.

Coincidence? We think NOT!

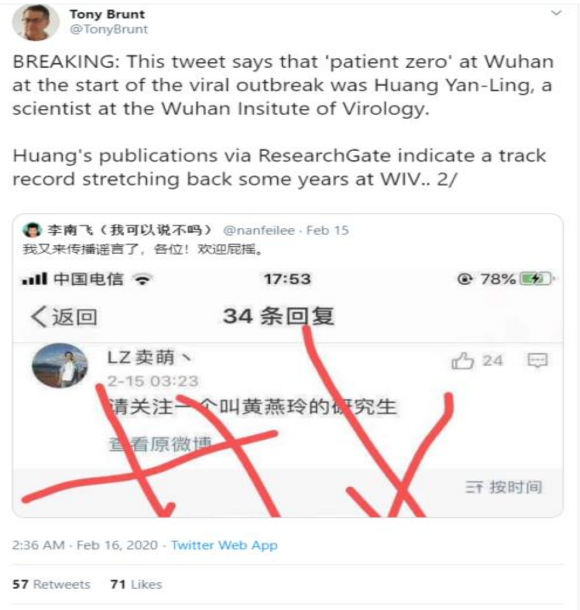
MISLEADING

Scientists have a hunch that contact with live animals or their dead flesh may be the source of the deadly virus.

Go vegan. | PETA

1:47 AM · Feb 20, 2020 · Sprout Social

Figure 3: Claim as coronavirus as an anagram of carnivorous



Tony Brunt @TonyBrunt

BREAKING: This tweet says that 'patient zero' at Wuhan at the start of the viral outbreak was Huang Yan-Ling, a scientist at the Wuhan Insitute of Virology.

Huang's publications via ResearchGate indicate a track record stretching back some years at WIV.. 2/

李南飞（我可以说不吗） @nanfeilee · Feb 15

我又来传播谣言了，各位！欢迎屁摇。

中国电信 17:53 78%

LZ卖萌丶

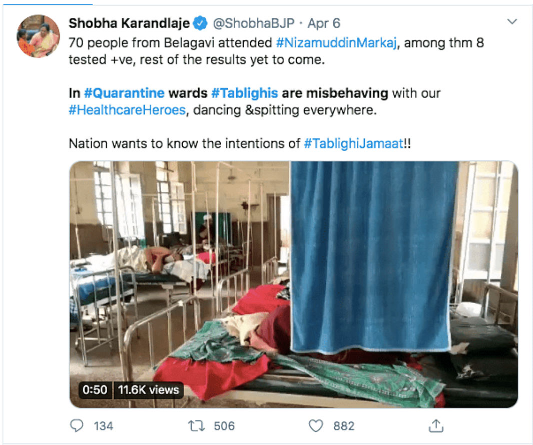
2-15 03:23

请关注一个叫黄燕玲的研究生

2:36 AM · Feb 16, 2020 · Twitter Web App

57 Retweets 71 Likes

Figure 4: Patient zero to get infected by virus



Shobha Karandlaje @ShobhaBJP · Apr 6

70 people from Belagavi attended #NizamuddinMarkaj, among thm 8 tested +ve, rest of the results yet to come.

In #Quarantine wards #Tablighis are misbehaving with our #HealthcareHeroes, dancing &spitting everywhere.

Nation wants to know the intentions of #TablighiJamaat!!

134 506 882

Figure 5: A political tweet in India

1.2. Objective

The Covid-19 related fake news have been examined in precise details in this research. The major motive of this research is to classify the text as fake and true. As we see that text based news are circulated very frequently so our focus should me mainly on classifying such text. As the spread of covid is vast there is a growing spread of related fake news as well. We have seen the impact of spreading the fake news how people and business have suffered due to spread of the fake news so our first goal should be to analyze such text messages and to classify them on the basis of their content whether they are true or false.

1.3 Organization of Dissertation:

In the 2nd chapter, we will see the topic in brief and how it develops to be essential also emphasizing how beneficial it could have been it also provides an analysis of relevant literature on Covid19 and vaccine misinformation. The second chapter will go into the fundamentals of the LSTM, BERT BiLSTM, ROBERTA.

In the third chapter the proposed models for the Covid-19 vaccine misinformation detection is discussed.

The chapter 4 presents the outcomes of our approaches on the dataset and validates the outcomes by demonstrating the likelihood of a statement being true.

The model's experimental approach will be discussed in the chapter 5.

The chapter 6 of the thesis is aimed toward a conclusion, with additional further research opportunities.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 LITERATURE REVIEW

Since the spread of covid-19 there is a situation of infodemic, as there are large number of users of the internet there are way more space for the spread of news. It is very important to see if the content that is spread around has credibility and does not have any bad impact on any individual or business. Alcohol can cue coronavirus an WhatsApp message was circulated where WHO cleared that it is false [2]. Such message shave cost loss of lives and has also caused financial loss as well. There has been various work done on fake news detection we have analyzed the previous work. The authors [3] used the Support Vector Machine to analyze the sentiments expressed in the tweets. The goal was to see how effective the models are at analyzing the sentiment classification based on the data. The data used by the researchers consisted of tweets with the term corona. Various data processing techniques such as stemming, and lemmatization were utilized to simplify the data. The usage of a feature extraction technique like TF-IDF [4] was employed to assign weights to the most frequently used words in the dataset. After that, the model is trained using classification models such as logistic-regression, decision trees, and K Nearest Neighbors, among others, and the results are compared using vectorized techniques to convert the data into vectors and transmit it to classification models. It has been completed. The method has been used by a number of researchers. The work done so far has centered on detecting fake news about the virus; the data contains numerous details about the number of deaths, false claims about the cure, and news about people breaking the rules of social distancing; these datasets have been collected from tweets, Facebook posts, and other social media posts; and works related to sentiment analysis of people by their replies. The biggest fake news and facts that began circulating through social platforms in the mid-2020s were related to vaccines; as vaccine development progressed, numerous facts were modified on these platforms, necessitating the classification of these news. The datasets utilized were from [5] they have a labeled dataset comprising of tweets related to vaccine the related work regarding covid-19 vaccine misinformation is compared in the below table 1.

Literature (Author)	Dataset	Accuracy
Cui L., Lee D [8]	Data collected from blogs websites and social media regarding covid-19 vaccine misinformation.	F1-score: 0.58
Zhou, Mulay A, Ferrara, Zafarani [9]	Data collected form news articles and tweets from twitter	F1-score: 0.83
Abdul-Mageed, Elmadany, Verma K., Lin R [10]	Tweets in different languages from different countries	F1-score: 0.92
Ismail F.H., Taha M., Abdelminaam D.S., Taha A., Nabil A., Houssein E.H.[11]	Data collected and created using various existing dataset related to covid-19.	F1-score: 0.985
Haouari, Hasanain, Suwaileh, Elsayed [12]	Tweets in Arabic regarding covid-19	F1-score: 0.74

Table 1: Details regarding existing work and it's accuracy.

2.2 OVERVIEW

This section will describe the planned approach for classifying Covid-19 fake tweets, which is implemented using NLP based BERT along with CNN and BiLSTM model.

2.2.1 NATURAL LANGUAGE PROCESSING

NLP is a branch of Human language, Computer science and Artificial Intelligence. It deals with understanding and interpreting human language. It has a wide use in performing tasks such as sentiment analysis, classification problems, speech recognition, question answering. It can used to get direct response to the question being asked.

There are 2 components of NLP:

1) **Natural Language Understanding (NLU)**

   This unit mainly focuses on understanding the human language and extracting the knowledge by analyzing the metadata of the language.

2) **Natural Language Generation (NLG)**

   It works as a translator which mainly works by taking computerized data as input and producing as output a natural language.

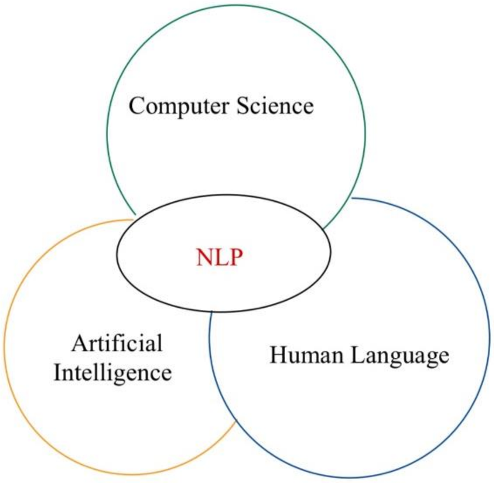


Figure 6: NLP

2.2.2 Recurrent Neural Network:

Simple neural networks are given a fascinating twist by recurrent neural networks (RNNs). The application of a vanilla neural network in circumstances requiring a series type input with no preset size is limited because it requires a fixed size vector as input. RNNs are designed to process a vast number of inputs and have no size limit by default. The Recurrent Neural Network remembers the past and bases its decisions on what it has learnt from it.

Although RNNs learn in a similar way during training, they frequently recollect information from past inputs when generating output (s). It's a part of the system. RNNs can accept one or more input vectors and generate one or more output vectors, with the output(s) influenced not just by the weights applied to the inputs, as in a conventional NN, but also by a "hidden" state vector indicating the context based on prior input(s)/output(s). As a result, depending on the preceding inputs in the series, the same input could produce a different output.

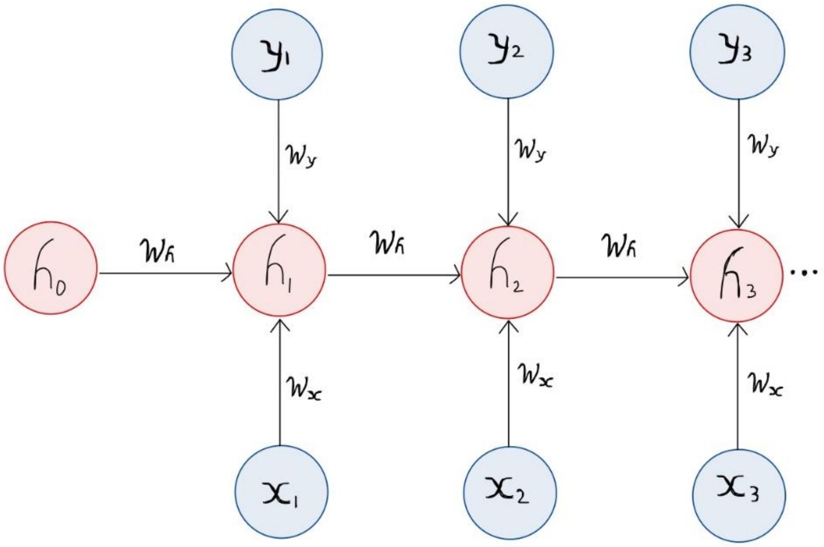


Figure 7: A hidden state RNN that uses one input item in the sequence to deliver relevant information to others.

2.2.3 Long Short Term Memory (LSTM):

LSTMs, or long-term memory networks, are a form of RNN that can learn long-term dependencies. LSTMs were generated expressly to avoid the issue of long-term dependency. It's not something people try to understand, but it's their default mode of operation to recall information for long periods of time.

Both recurring neural networks have a chain of repeating neural network modules as their shape. This recurring module in traditional RNNs would have a noncomplicated structure, such as a single tanh-layer. Although LSTMs have a structure similar to this chain, the repeating module is distinct. Instead of a single neural network layer, there are four layers, each with its own set of interactions.

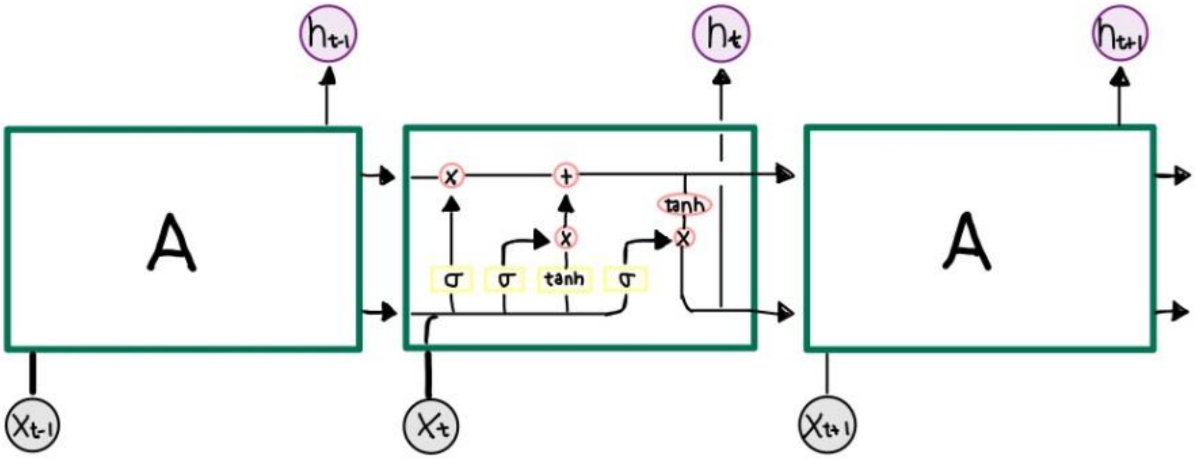


Figure 8: An LSTM's repeating model containing four interacting layers.

2.2.4 Convolutional LSTM:

The input x vector, cell output vector y, hidden state vector h, and gates are all 3D tensors with the last two dimensions as spatial dimensions (it, ft, ot). The inputs to the convolution LSTM determine the future state of each grid cell as well as the past connecting cell states.

2.2.5 Bi-Directional LSTM:

The principle behind Bidirectional Recurring Neural Networks (RNNs) is very basic. This involves replicating the network's first recurrent layer and then supplying the input sequence as it is input to the first layer and providing the replicated layer with a reversed copy of the input sequence [6]. It is distinct from the unidirectional in that the reverse operating LSTM retains future data and is capable of protecting past and future data at any point using the two secret states together [5].

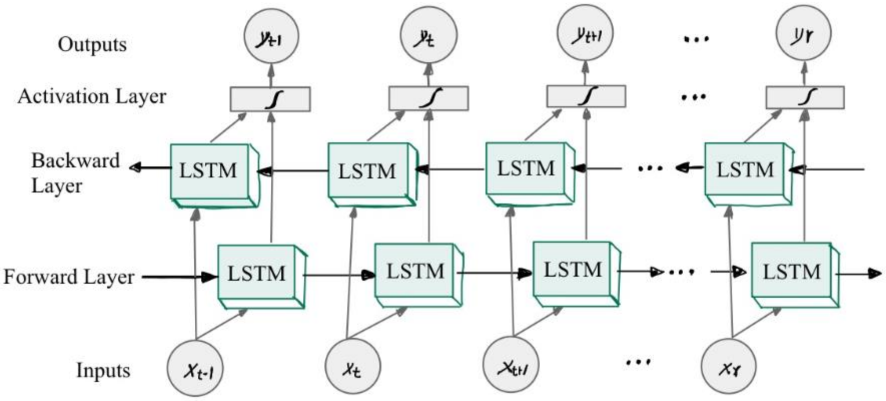


Figure 9: Bi - Directional LSTM

2.2.6 BERT Neural Network:

LSTM vs Transformers

To resolve the problem of language Translation the transformer neural network architecture was initially made. This was very well received until this point LSTM has been used to solve this problem but they have few problem themselves.

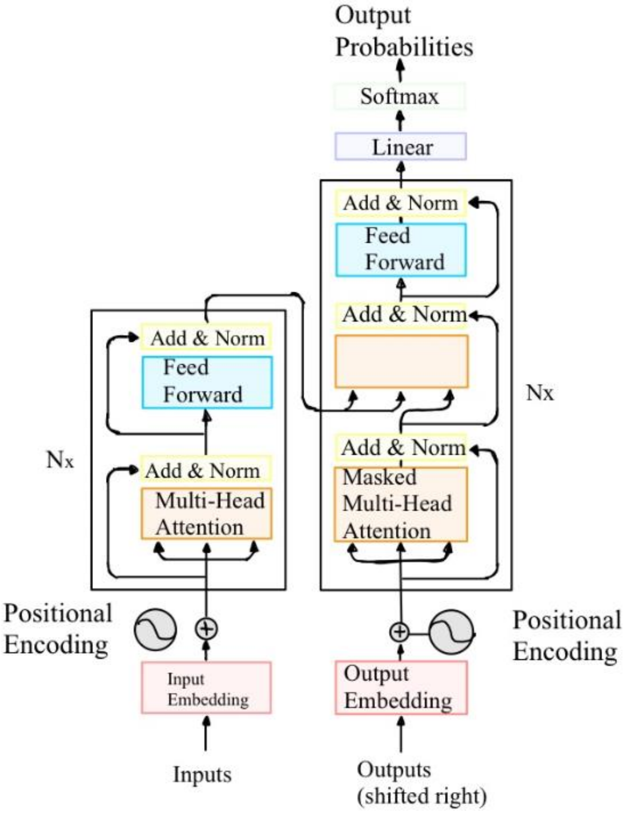


Figure 10: The Transformer-model architecture

LSTM Networks:

1) **Slow:** As the words were passed in sequentially and are generated in the same fashion, the LSTM networks are slow in terms of training. For these neural networks to learn it can take a number of time steps.

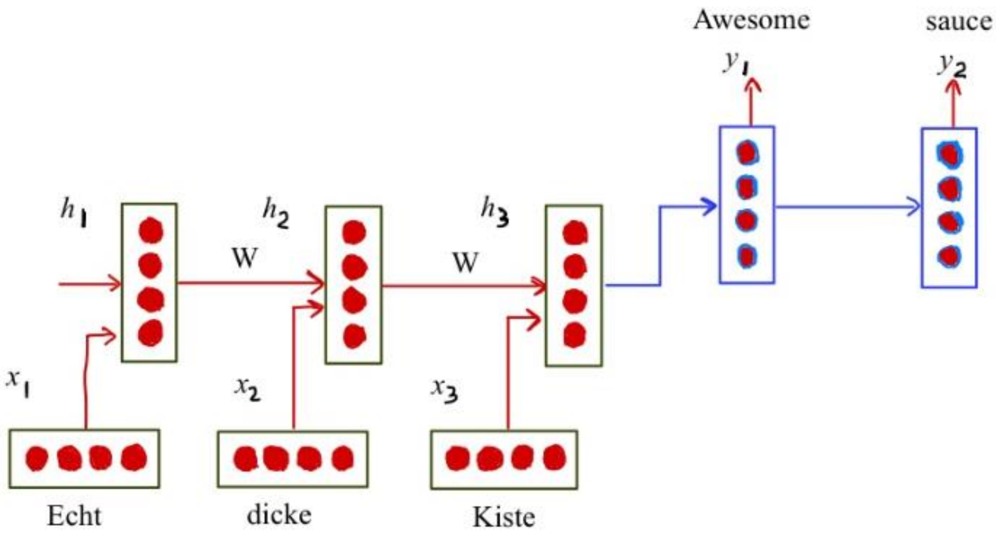


Figure 11: LSTM Network

2) **Not truly Bi-directional:** Even Bi-directional LSTM isn't great at capturing the true meaning of words because it still learns right to left and left to right context independently before concatenating it. As a result, the full context is obscured.

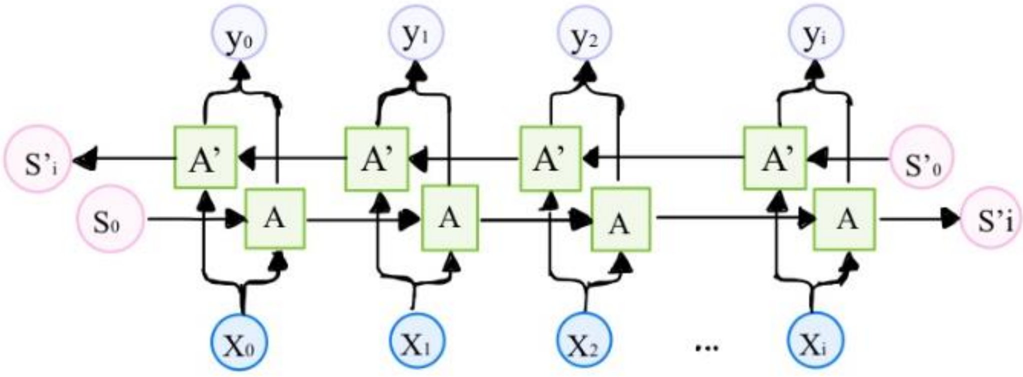


Figure 12: Bi-Directional LSTM Network

Transformer:

The transformer architecture addresses some of these concerns.

1) **Faster:** For starters they are faster since multiple words may be processed at the same time.

2) **Deeply Bi-Directional:** Word context is easier to learn because it may be learned in both directions at the same time.

Transformer Flow:

Let's have a look at the transformer in operation. Assume we wish to teach this architecture to translate from English to French. The encoder and decoder are the two main components of the transformers.

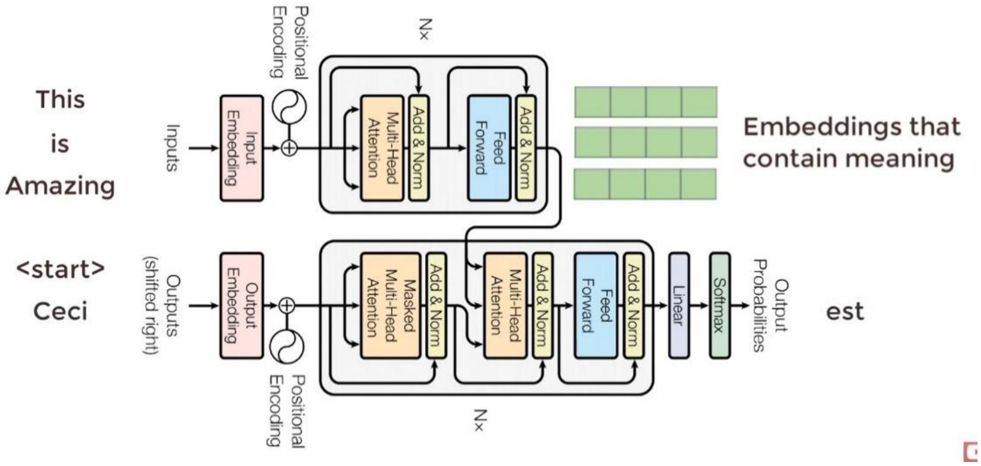


Figure 13: Transformer model architecture

Create embeddings for each one at the same time using encoder that takes the English words. These embeddings are vectors that contain the word's meaning. The decoder combines these embeddings from the encoder and the previously generated words of the translated French sentence to construct the next French word, and we continue to generate the French translation one word at a time until we reach the conclusion of the sentence.

The GPT transformer architecture is created by stacking the decoders. Stacking simply the encoders, on the other hand, yields BERT, a Bi Directional encoder representations from the transformer.

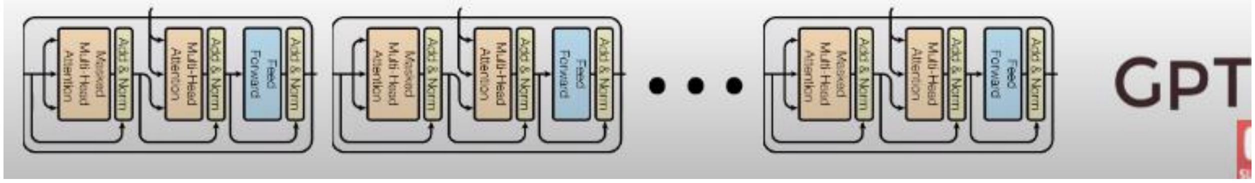


Figure 14: Stack of Decoders-GPT

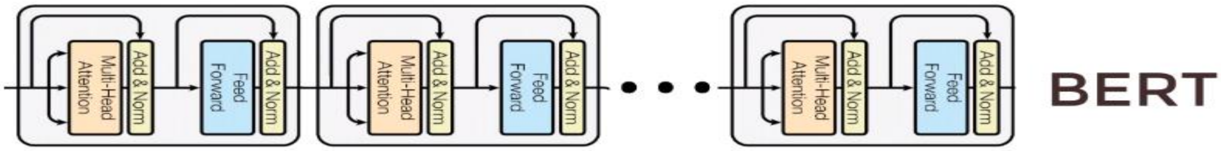


Figure 15: Stack of encoders - BERT

BiDirectional Encoder Representation from Transformer

Although the BERT transformer has a lock for conversion of language, we may use BERT to learn language conversion, answering the question, analysis of sentiments, summarizing texts, and many more tasks.

As a result, we can train BERT to understand English and then fine-tune him based on the problem we're trying to address.

How to solve problems? (BERT Training)

- Pre train BERT to recognize language.
- Fine tune BERT to learn specific task.

As a result, BERT training is separated into two phases: the first is pre-training, in which the model learns what language and context are, and the second is fine tuning, in which the model understands language but does not know how to address the problem.

1. Pre Training (Pass 1):

The purpose behind pre training is for BERT to understand what language and context are. BERT learns language by doing two unsupervised tasks at the same time: masked-language modelling and sentence prediction.

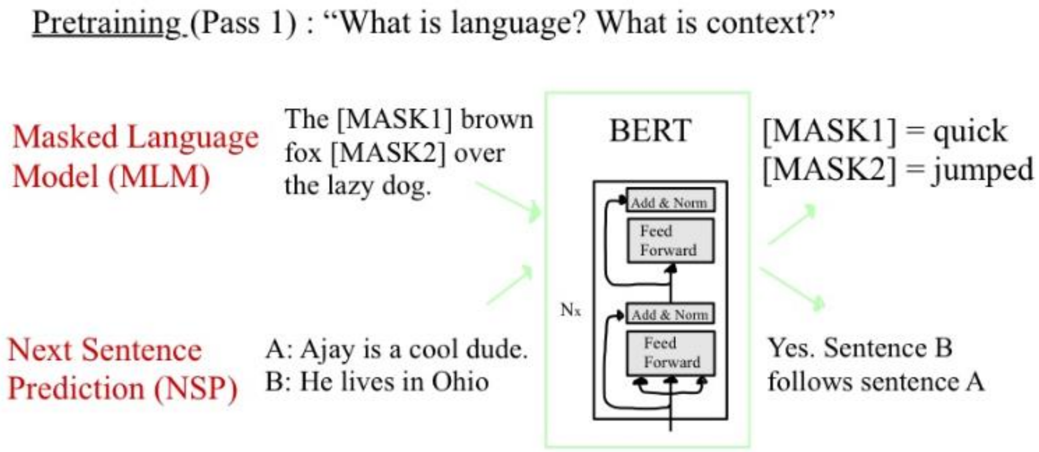

Pretraining (Pass 1) : "What is language? What is context?"
Masked Language Model (MLM)
The [MASK1] brown fox [MASK2] over the lazy dog.
BERT
Add & Norm
Feed Forward
Nx
Add & Norm
Feed Forward
[MASK1] = quick
[MASK2] = jumped
Next Sentence Prediction (NSP)
A: Ajay is a cool dude.
B: He lives in Ohio
Yes. Sentence B follows sentence A

Figure 16: Pre-training (Pass 1)

1.1. Masked Language Model (MLM)

For masked language modelling, BERT accepts a sentence and fills in the blanks with masks. The purpose is to output these mask tokens, which enables BERT grasp a bidirectional context within a sentence

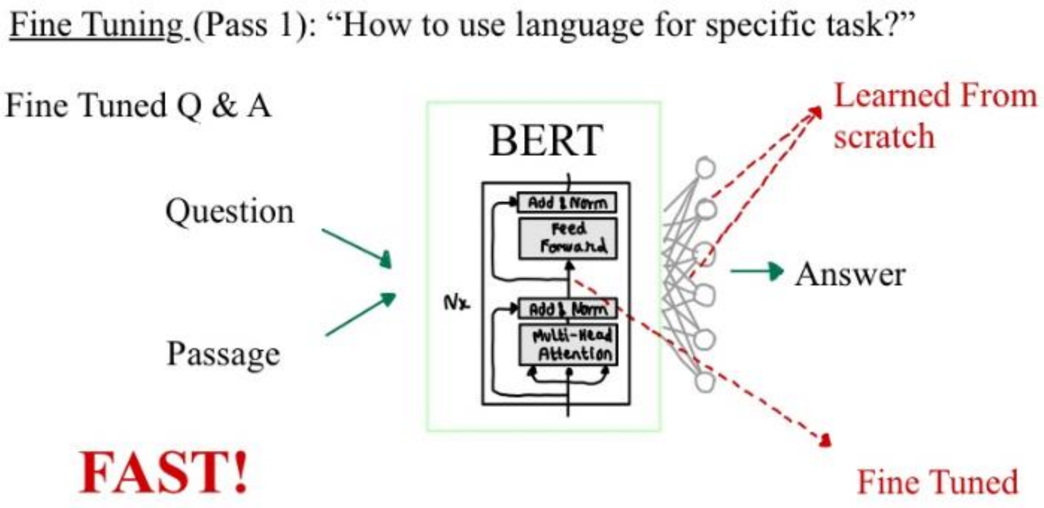
.

1.2. Next Sentence Prediction (NSP)

In order to determine whether the second sentence follows the first in next sentence prediction, BERT uses two sentences, the first in a binary classification problem. This helps, BERT understand context across different sentences, and by combining both of these, BERT gains a good understanding of language. So that’s the warmup.

Now let us see the fine tuning phase

2. Fine Tuning (Pass 1)

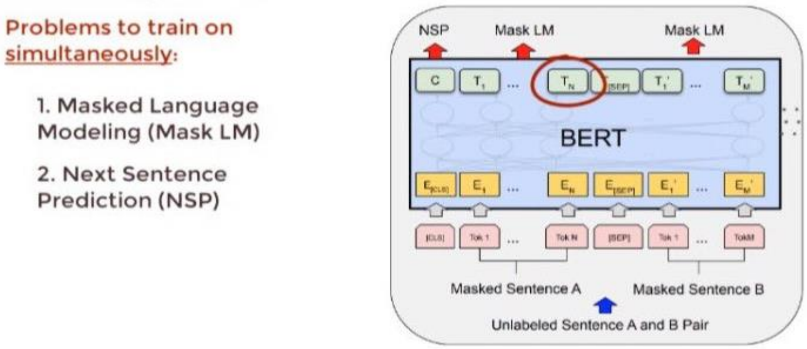


Fine Tuning (Pass 1): “How to use language for specific task?”

Figure 17: Fine Tuning (Pass 1)

So, for example, if we want to train BERT for performing the answering of question, all we have to do is replace the completely connected layers of output of the network with a new set of output layers that can basically output the expected result to the question that we want, and then we can perform supervised training using a question response dataset, the rest of the parameters of the model are just taught from scratch. So now that’s pass 1 of the explanation on pre training and fine tuning and let’s go on pass two with some more details.

3. Pre training (Pass 2)



Problems to train on simultaneously:

1. Masked Language Modeling (Mask LM)
2. Next Sentence Prediction (NSP)

Figure 18. Pre training (Pass 2)

We practiced masked language modelling and next phrase prediction during BERT pre-training. The input is a pair of two phrases, some of which are masked. Each token is a word, and we turn each of these words into an embedding using pre-learned embeddings. On the output side, this provides a decent starting point for BERT to work with. The binary output for the next sentence prediction would be 'C,' so it would be output. The following T's here are vectors of word that corelates to the output for the masked language model problem, with 1 if sentence B follows sentence A in context and 0 if sentence B does not. As a result, the number of word vectors we enter equals the number of word vectors we produce.

4. Fine tuning (Pass 2)

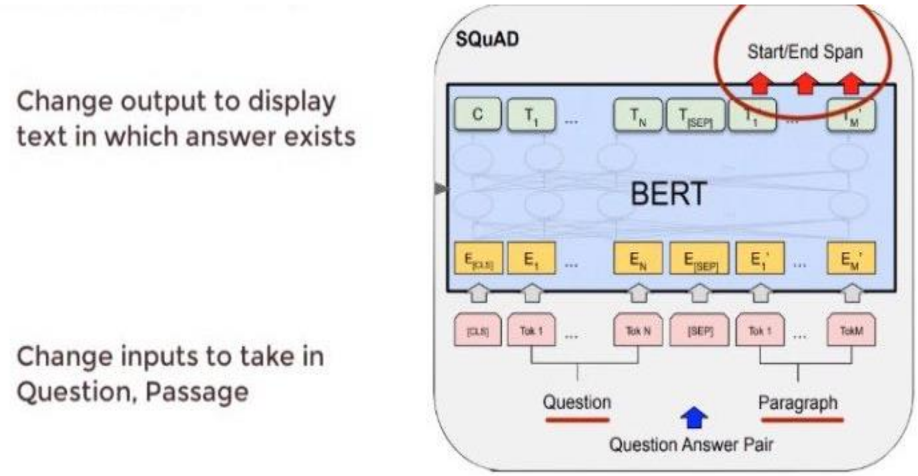


Figure 19: Fine tune (Pass 2)

However, if we want to perform some other task, by modifying the output and inputs layers the training of the model can done.

5. Pre training (Pass 3)

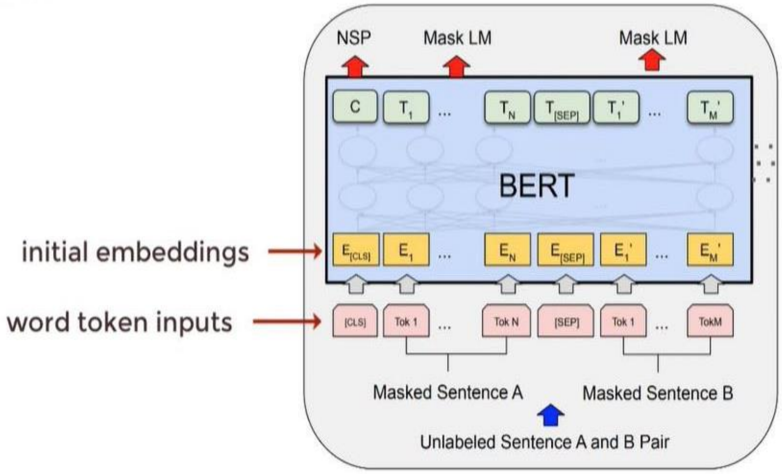


Figure 20. Pre training (Pass 3)

Now, for pass 3, we'll go over the details of how we're going to produce the embeddings

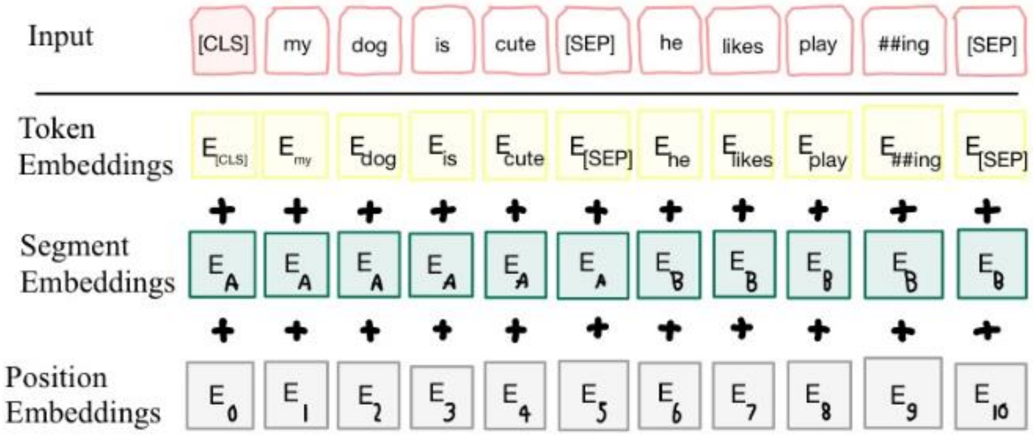

Input
[CLS]
my
dog
is
cute
[SEP]
he
likes
play
##ing
[SEP]
Token Embeddings
$E_{[CLS]}$ E_{my} E_{dog} E_{is} E_{cute} $E_{[SEP]}$ E_{he} E_{likes} E_{play} $E_{\#\#ing}$ $E_{[SEP]}$
Segment Embeddings
E_A E_A E_A E_A E_A E_A E_B E_B E_B E_B E_B
Position Embeddings
E_0 E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8 E_9 E_{10}

Figure 21. Initial Embeddings

The pre-trained embeddings are token embeddings, whereas the main study employs word piece embeddings with 30,000 tokens in the vocabulary.

The sentence number is basically encoded into a vector by the segment embeddings.

The position embedding is the vector representation of a word's position within a sentence. We acquire an embedding vector by adding all three vectors together, which we use as input to BERT. The segment and position embedding are required for temporal ordering because all of these vectors are fed into BERT at the same time, and language models require this ordering to be kept. The information is starting to come together.

Let us see the output side now.

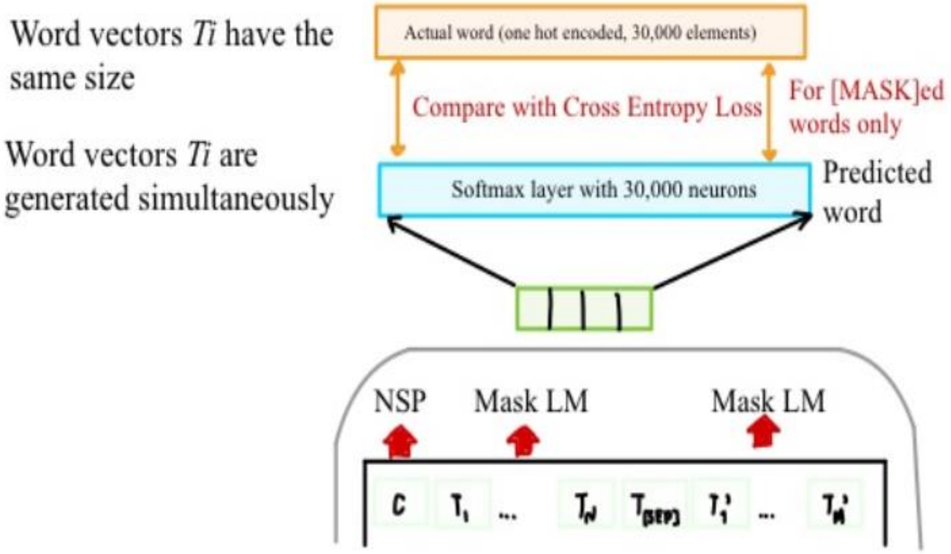

Word vectors Ti have the same size
Word vectors Ti are generated simultaneously
Actual word (one hot encoded, 30,000 elements)
Compare with Cross Entropy Loss
For [MASK]ed words only
Softmax layer with 30,000 neurons
Predicted word
NSP
Mask LM
Mask LM
C
T_1
...
T_N
$T_{[SEP]}$
T_1'
...
T_M'

Figure 22. output

The output is a binary value C and a bunch of word vectors, but we want to minimize the loss, so two things to keep in mind: all of these word vectors are the same size, and they're all generated at the same time. In this case, we'd use a activation function (softmax) to get distribution from a word vector, and the actual label of this distribution would be a single hot encoded vector for the actual word.

So that was a three passes of explaining the pre-training and fine tuning of BERT so let us put this all together

2.2.7. RoBERTA: Robustly Optimized BERT Pretraining

This is probably the one that has least updates or new stuff. It came out not that long after BERT and so, what they showed was that BERT was really under trained. So, basically they took even the same amount of data which was even 40 epochs on the data, if you do it for like 200 epochs we get even better results significantly. So they trained more epochs on the same data and they also showed that more data helps.They improved the masking and pre training using a couple of tweaks and they were able to get state of the art results.

2.2.8. ALBERT: A Lite BERT:

ALBERT which is called a Lite BERT for self-supervised learning. It has a couple of inventions. The idea here is really massive parameter sharing with the idea being that, if you share parameters you are not going to get a better language model, but we will get better sample efficiency. We will get less over-fitting when we fine tune because if we have a billion parameters and we fine tune on them on a dataset that is like 1000 label examples we are still going to overfit, but if we have a much small number of parameters we are going to get less over-fitting. So, if we get a similarly powerful model with fewer parameters we are going to get less over-fitting.

Innovation 1: Factorized embedding

So the two major innovations were instead of using the word embeddings because of it;s large table because it is the size of our vocabulary the number of word pieces times the hidden size. So it is going to be much bigger than the hidden layer. So first thing is that they used a factorized embedding table. So if they have a hidden size of 1000 they only use like 128 dimensional input embedding and they projected that to 1000 using a matrix. So instead of having 1024 by 100000 they have 128 by 100000 plus 1024 times 128 and we multiply these together and multiply the two matrices together and then effectively we have a 1024 by 100000 embedding matrix but we have a much fewer parameters i.e we are doing parameter tying, this isn’t actually parameter tying but we are doing parameter reduction in a clever way.

Innovation 2: Cross Layer parameter sharing
The other innovation is cross layer parameter sharing, it is similar and has also been done in the previous papers especially universal transformers. The idea is that we have a bunch of transformer layers let us say we have 12 layers all of them should share the same parameters. so now we can have a much bigger model that has fewer parameters than BERT has so we get less over-fitting.
One important thing to keep in mind is that ALBERT is light in terms of parameters not in terms of speed. So the model that's actually comparable to BERT. So the BERT and ALBERT are about the same but this one was actually slower. So it's only when they started making models that were much bigger in terms of compute than BERT, but doing more parameter tying, then they started getting good results. The implications of this is that we can reduce the number of parameters but still nobody has figured out how to reduce the amount of pre-training compute that is required

2.2.9 Activation Function :

To evaluate whether a neuron should be triggered or not, the activation function produces a weighted sum and then incorporates bias to it. We apply any non-linear activation function in this layer to incorporate non-linearity into our model, which speeds up training and computation. It is used to learn and comprehend complicated patterns in our data, as well as to prevent the numbers from aggregating to zero. The most common activation function is RELU (rectified linear units).

- **Linear activation function**

The linear function's equation is y= m*x.

Equation : f(x)= x

Range: -∞ to ∞

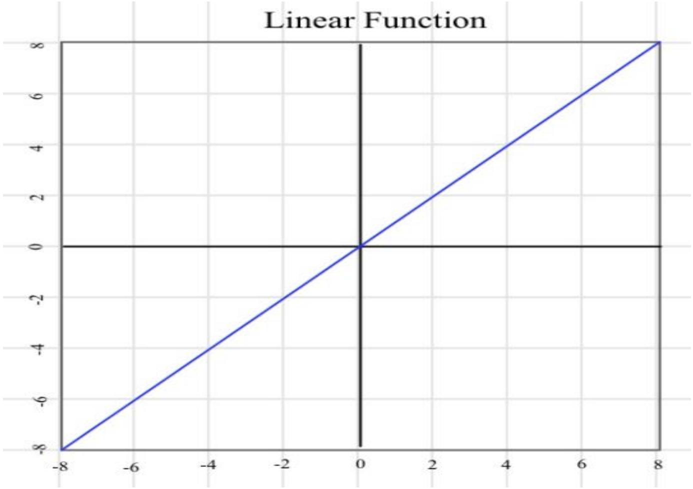


Figure 23: Linear Function

- **Sigmoid Function:**

The sigmoid or sigmoid activation function's curve resembles an 'S' shaped curve. Between zero and one is the range of the logistic activation function. Because value of the sigmoid function is limited between zero and one, the outcome is likely to be one if the value is greater than 0.5 & zero else.

Equation: $f(x) = \frac{1}{(1+e^{-x})}$

Range : [0 , 1]

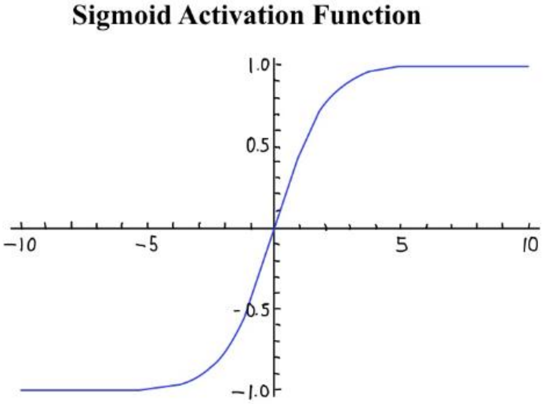


Figure 24: Sigmoid Activation Function

- **Tanh-activation function**

Tanh is a hyperbolic tangent function, similar to the logistic sigmoid. The curves of the Tanh and sigmoid activation functions are quite similar, as illustrated in figure 18, however Tanh is preferable since the whole function is zero centric.

Equation: $f(x) = \tanh(x) = \frac{2}{(1+e^{-2x})} - 1$

Range : [-1, 1]

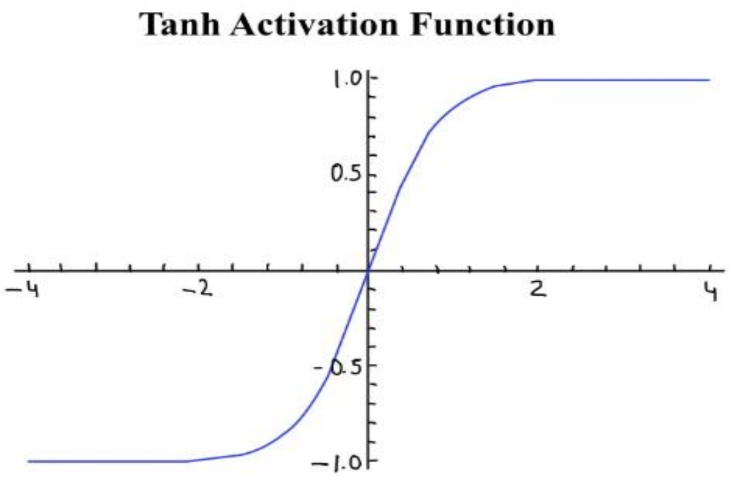


Figure 25: Tanh Activation Function

- **ReLU activation function**

It's most often used activation technique in hidden layers of a deep neural networks. It's the most used activation method in DNN hidden layers. Because the ReLU function is non-linear, we may quickly back transmit errors and trigger multiple layers of neurons. ReLU is less expensive than hyperbolic tangent and sigmoid because it uses fewer complex computations. Because just a few perceptrons are engaged at any given moment, the cnn is sparse and quick to process.

Equation => f(x)= max(0,x)

Range => [0 to ∞)

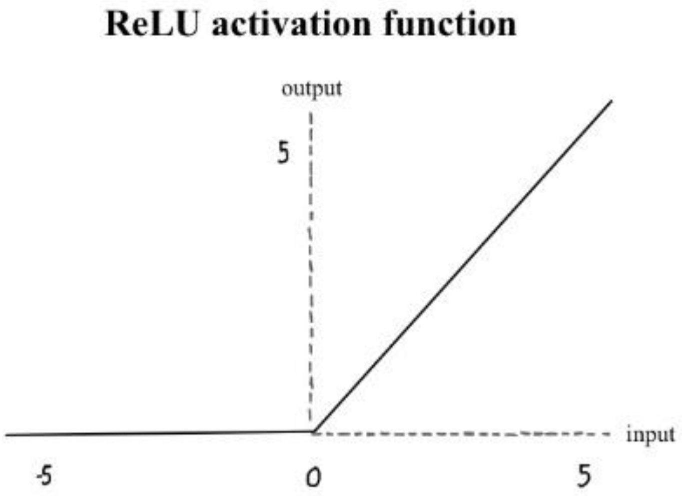


Figure 26: ReLU Activation Function

- **Softmax activation function**

The softmax function is a function that deals with classification tasks at fully connected layer. When dealing with many classes, this is commonly employed. The softmax function has range from zero to one. The softmax function is best used at the output layer of the deep neural network, where we want to use probability to characterize the classification from each input.

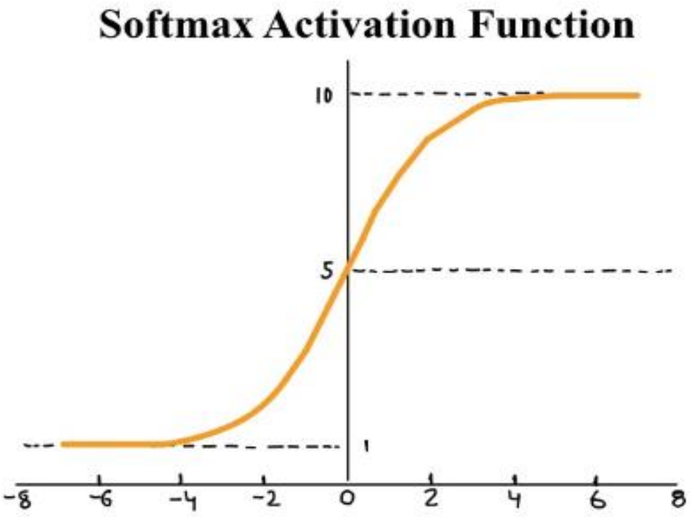


Figure 27: SoftMax Activation Function

CHAPTER 3

PROPOSED MODEL

3.1. Methodology

Text-Classification is a common Natural Language Processing methodology utilized in many studies and business development. The goal of this classification is to sort the news into one of two categories: fake or real. As Covid-19 vaccine development started recently there was no major dataset available we have taken the dataset from [5]. The below subsections discuss the details of the data and how it was collected and also we discuss the training on the data.

3.2. Data Collection

Social media platforms have been widely used to spread fake news. There is no way to check whether the tweets that have been posted have verified content. So for this study the data used is collected from twitter. Twitter provides us with a Twitter API for collection of data. Using this API we can write python code and extract tweets which contain the text posted by the users, we can fetch the tweets which contain certain keywords. The keywords that have been considered while collecting the data are 'pfizer', 'vaccine', 'sputnik', 'astrazeneca', 'sinopharm' and 'moderna'. We have taken the dataset from [5] where the researchers have taken the tweets from Dec 2020 till June 2021. We have not considered the retweets and the replies to the tweets, we have considered the tweets which are in English Language. We have the dataset which contains 10751 data entries with tweet id, tweet and label. We have split the dataset into test (10%) and train (90%).

3.3. Data Pre processing:

The BERT is made up of a stack of encoders from the transformer architecture. The initial stage of BERT training is pre-training, during which the model learns about language and context, and the second stage is fine tuning. After loading the dataset we perform Encoding and tokenization on the input data. We load the BERT Tokenizer and use it to tokenize the data, BERT uses Word Piece tokenizer. It works by splitting the word into tokens. We then preprocess the text by removing the links, punctuation, special characters, numbers. The tokenizer is then applied to these phrases, yielding an array of tokenized words, after which token IDs are assigned. Two special tokens are used, [CLS] used for classification at the beginning of every sentence and [SEP] placed at the end of the sentence to indicate the end of sentence. As BERT required fixed length of sentences we have applied padding and truncation. Lastly we use Attention mask it is useful to differentiate between padding and non padding.

3.4. Architecture and Implementation:

For comparative analysis we have used five different models, the standard BERT fine tune model, then we have added additional layers, including the BiDirectional LSTM and CNN. We have used the CNN and BiLSTM by freezing and without freezing the parameters in the BERT fine tune model.

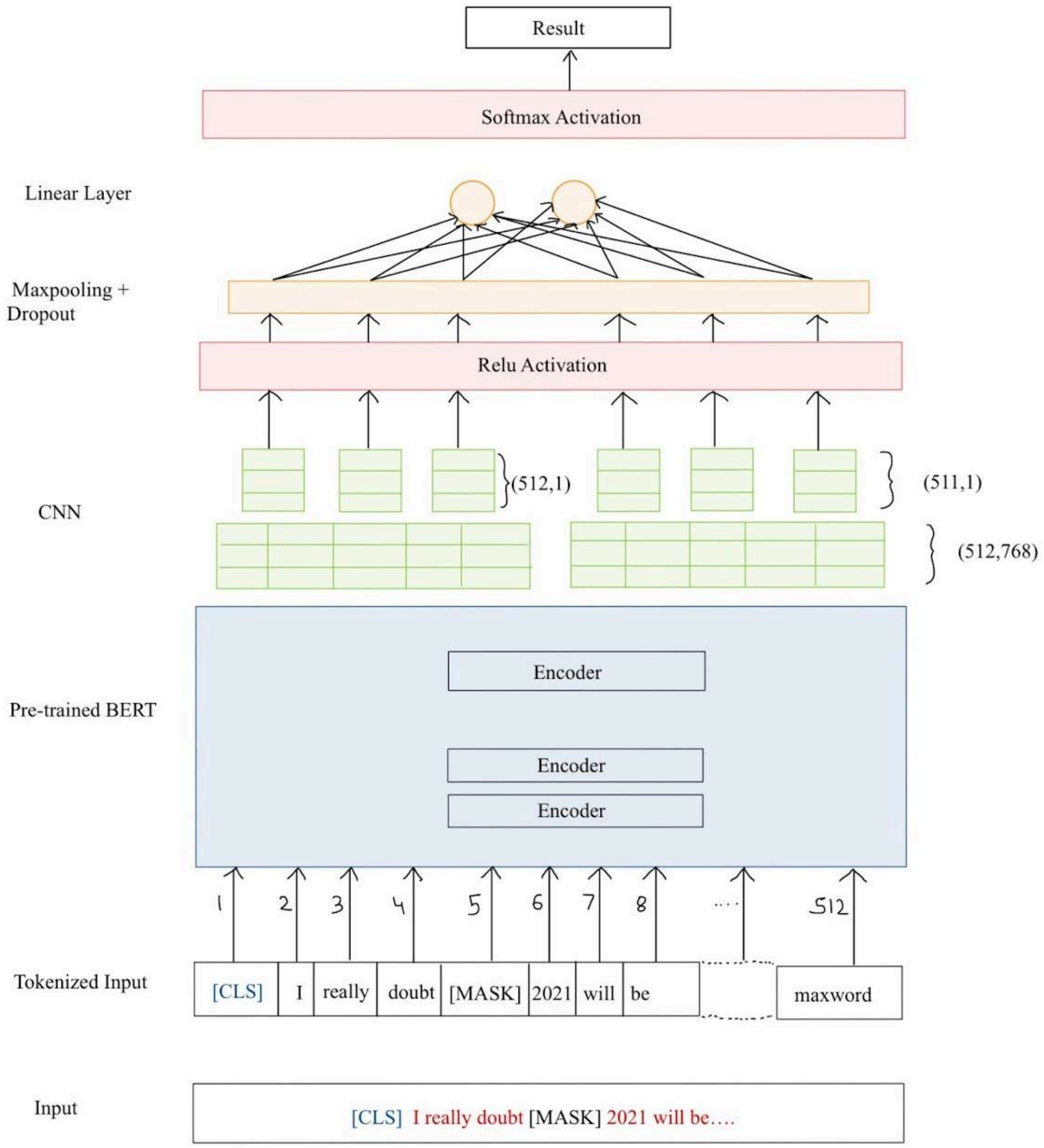


Figure 28: BERT + CNN

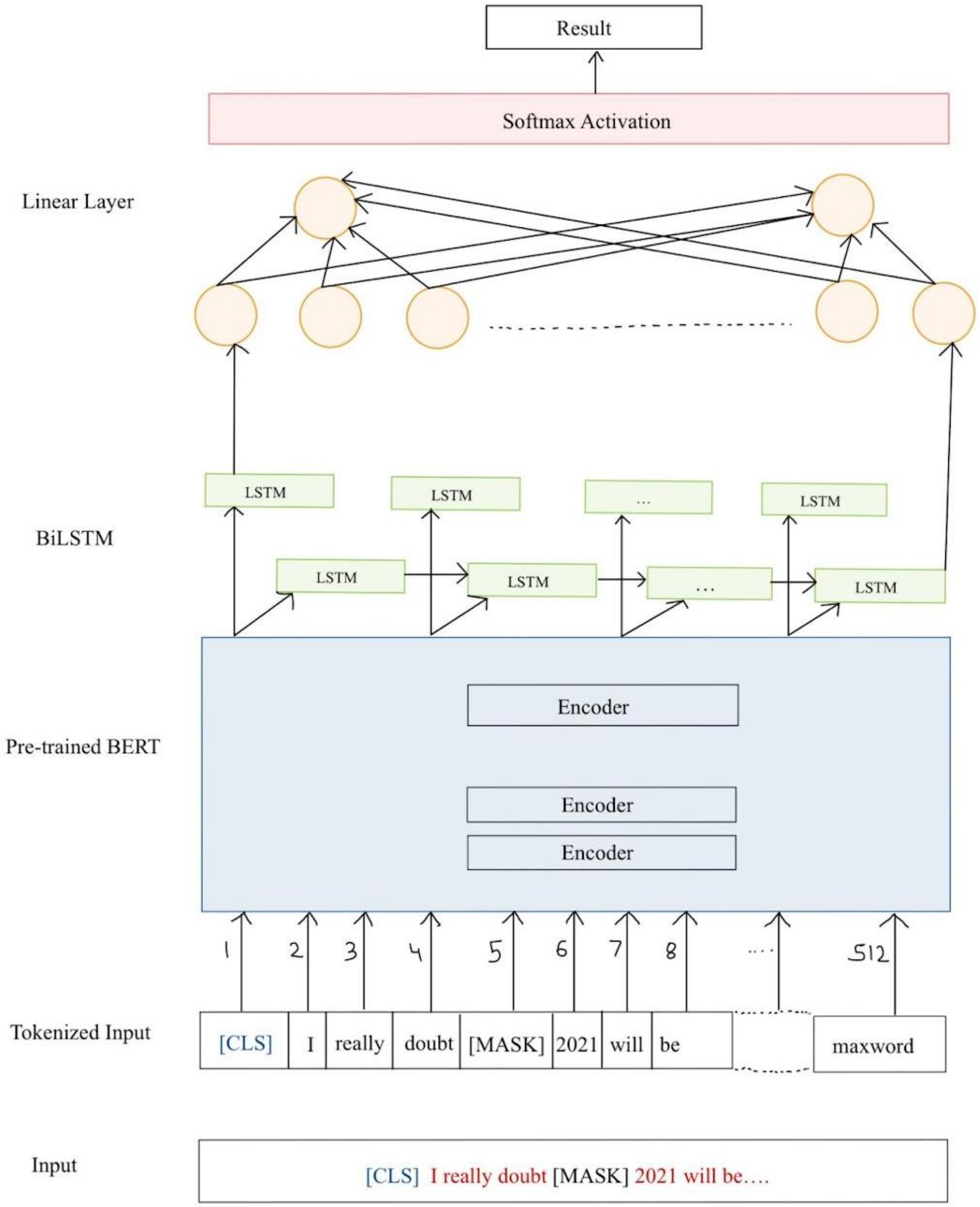


Figure 29: BERT + BiLSTM

We divide the data into training and validation sets before applying the BertForSequenceClassification model. The BERT model is then fine-tuned using a learning rate of 2e-5 and 4 epochs, with 1 addition layer. The second and third models are then enhanced with two convolution layers and the ReLU activation function, with kernel size of (1,768) and (2,768), respectively. A max-pooling layer is added, and a dropout layer with a rate of 0.1. A linear layer and a softmax-function are used. We add two BiLSTM layers and one BiLSTM layer to the fourth and fifth models, which are fine tuned BERT with BiLSTM. The model's learning rates have been set at 5e-5. The fourth model was trained for a total of ten epochs, whereas the fifth model was trained for six. The model can be seen in Fig. 21 and Fig. 22

We also used the versions of BERT, RoBERTA and ALBERT. When RoBERTA came, it was determined that BERT was undertrained. So the researchers trained more epochs on the same data, enhanced the masking and pre training with a few adjustments, and obtained better results. ALBERT, also known as a Lite BERT, is used for self supervised learning. It has two inventions: factorised embedding and cross layer parameter sharing.

Chapter 4
THE EXPERIMENTAL APPROACH

In this chapter, we'll look at our proposed model's experimental approach. This experiment was carried out with the following system configuration:

- Processor: Intel Core i5 (6th Gen)
- Main Memory: 8GB
- Secondary Memory: 1 TB
- Tools Used: Jupyter Notebook, Google Colaboratory, Anaconda.

4.1 Data analysis

Data analysis is the first stage with the purpose of conserving the best out of trash since it is the sequence of the process of analyzing, converting, cleaning and modeling data with the goal of identifying important data, reporting conclusions, and supporting decision making. One of the main factors for data analysis is to determine the data's complexity and appearance, and to ensure that the data is legitimate as well as contains the required fields.

4.2 Dataset

Social media platforms have been widely used to spread fake news. There is no way to check whether the tweets that have been posted have verified content. So for this study the data used is collected from twitter. Twitter provides us with a Twitter API for collection of data. Using this API we can write python code and extract tweets which contain the text posted by the users, we can fetch the tweets which contain certain keywords. The keywords that have been considered while collecting the data are ‘pfizer’, ‘vaccine’, ‘sputnik’, ‘astrazeneca’, ‘sinopharm’ and ‘moderna’. We have taken the dataset from [5] where the researchers have taken the tweets from Dec 2020 till June 2021. We have not considered the retweets and the replies to the tweets, we have considered the tweets which are in English Language. We have the dataset which contains 10751 data entries with tweet id, tweet and label. We have split the dataset into test (10%) and train (90%).

4.3 Model Training and Validation

Slicing the train dataset from our dataset is used for training. Accordingly, for testing purposes, a portion of the dataset is taken from the dataset. We have the dataset which contains 10751 data entries with tweet id, tweet and label. We have split the dataset into test (10%) and train (90%). There are 9675 training samples and 1076 validation samples.

Column1	True	False	Total
Train	5805	3870	9675
Validation	713	363	1076

Table 2: Train and Validation split

Chapter 5
Experimental Results

5.1 Model Evaluation

We use the following measurements based on confusion matrices conclusions for prediction evaluation.

5.1.1 F1-score

A significant number of True Negatives (TN), which in most business situations do not rely on much, contribute to the accuracy, although False Negatives (FN) and False Positives (FP) frequently have business consequences. If we need to find the right balance between Recall and Precision There is an unequal class distribution, F1-Score would be an appropriate statistic to utilize

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

5.1.2 Accuracy

The data that is correctly categorized divided by the whole dataset evaluated is how accuracy is calculated. It can also be calculated as a 1-error.
The following equation can be used to calculate the accuracy:

Accuracy = $\frac{TruePositive(TP) + TrueNegative(TN)}{TruePositive(TP)+ TrueNegative(TN) + FalsePositive(FP)+FalseNegative(FN)}$

5.1.3 ROC-AUC

The ROC AUC is used in the classification problem for the performance measurement, its value ranges from 0 to 1, if the value is 0 then the model performance is not good and if the value is 1 it means that the model performance is good.

5.2 Model Prediction

The BERT fine tune which is the basic model has test accuracy of 0.96 with a training loss of 0.0287 it has a ROC AUC and F1 score as 0.965 and 0.961. The Model 2 and Model 3 which is the BERT with CNN and parameters not freezed and with freezed has an training accuracy of 0.98 and 0.97 respectively. The Model 4 and Model 5 which is the BERT with BiLSTM and parameters not freezed and with freezed has testing accuracy of 0.945 and 0.944 respectively.
RoBERTA has a test accuracy of 0.98 and F1 score of 0.97. ALBERT has a test accuracy of 0.969 and F1 score of 0.962. The details of the metrics can be seen in the below Table 1.

Model	Test Accuracy	Train Loss	ROC AUC	F1 Score
BERT Fine Tune	0.9674	0.0287	0.9650	0.9610
CNN + BERT (Freeze Parameters)	0.9600	0.0883	0.9568	0.9519
CNN + BERT (without Freeze Parameters)	0.9600	0.0883	0.9568	0.9519
BiLSTM + BERT (Freeze Parameters)	0.9451	0.0920	0.9386	0.9325
BiLSTM + BERT (without Freeze Parameters)	0.9442	0.0922	0.9378	0.9315
ROBERTA	0.9814	0.0837	0.9967	0.9772
ALBERT	0.9693	0.0121	0.9905	0.9624

Table 3: Performance analysis

5.3 Performance Analysis

After comparing all the results we can see that the Model 6 which is the RoBERTA give the highest testing accuracy that is the 0.9814 and highest F1 score as 0.9772. So as RoBERTA was trained by the researchers for more epochs compared to the BERT so the model gets trained better and gives good results.

CHAPTER 6
CONCLUSION

We have successfully implemented the models BERT along with CNN and BiLSTM networks. We got a high-level accuracy of 98.14 on the ROBERTA model with an F1 score of 0.9772. In the future study we aim to investigate more efficient results by using different dataset which comprises of images, videos or other data format. We will try some advance models or Hybrid models to find fake news related to covid-19 vaccine. Since India has never experienced a pandemic like this in the past century, the lockdown, the policies of government, and so on should all be investigated in order to gain a better knowledge of the condition, the lockdown, government policies, and so on should all be explored. The findings may be useful for researchers and common people in recognizing and classifying fake news allowing them been more safe from the losses they may incur in terms of business and personal health.

References

[1] “WHO.” https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19.

[2] https://www.newindianexpress.com/world/2020/mar/09/believing-fake-news-iranians-turn-to-alcohol-to-prevent-covid-19-bootleg-booze-kills-27-2114502.html

[3] Samuel, Jim, G. G. Ali, Md Rahman, Ek Esawi, and Yana Samuel. "Covid-19 public sentiment insights and machine learning for tweets classification." Information 11, no. 6 (2020): 314.

[4] Elhadad, Mohamed K., Kin Fun Li, and Fayez Gebali. "Detecting Misleading Information on COVID-19." Ieee Access 8 (2020): 165201-165215.

[5] Arora, Parul, Himanshu Kumar, and Bijaya Ketan Panigrahi. "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India." Chaos, Solitons & Fractals 139 (2020): 110017.

[6] https://www.i2tutorials.com/deep-dive-into-bidirectional-lstm/

[7] Hayawi, Kadhim, Sakib Shahriar, Mohamed Adel Serhani, Ikbal Taleb, and Sujith Samuel Mathew. "ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection." Public health 203 (2022): 23-30.

[8] Cui L., Lee D. ArXiv Prepr; 2020. Coaid: covid-19 healthcare misinformation dataset. ArXiv200600885.

[9] Zhou X., Mulay A., Ferrara E., Zafarani R. Proceedings of the 29th ACM international conference on information & knowledge management [Internet] Association for Computing Machinery; New York, NY, USA: 2020. ReCOVery: a multimodal repository for COVID-19 news credibility research; p. 3205. 12.

[10] Abdul-Mageed M., Elmadany A., Nagoudi E.M.B., Pabbi D., Verma K., Lin R. ArXiv Prepr; 2020. Mega-cov: a billion-scale dataset of 100+ languages for covid-19. ArXiv200506012.

[11] Abdelminaam D.S., Ismail F.H., Taha M., Taha A., Houssein E.H., Nabil A. CoAID-DEEP: an optimized intelligent framework for automated detecting COVID-19 misleading information on Twitter. IEEE Access. 2021;9:27840–27867.

[12] Haouari F., Hasanain M., Suwaileh R., Elsayed T. ArXiv Prepr; 2020. ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection. ArXiv201008768.

[13] Hayawi, Kadhim, Sakib Shahriar, Mohamed Adel Serhani, Ikbal Taleb, and Sujith Samuel Mathew. "ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection." *Public health* 203 (2022): 23-30.

PAPER NAME

Thesis_2k20ISY07.pdf

WORD COUNT	CHARACTER COUNT
6079 Words	29974 Characters
PAGE COUNT	FILE SIZE
28 Pages	1.4MB
SUBMISSION DATE	REPORT DATE
May 24, 2022 8:18 PM GMT+5:30	May 24, 2022 8:19 PM GMT+5:30

● 15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 0% Publications database
- Crossref Posted Content database
- 14% Submitted Works database

● Excluded from Similarity Report

- Crossref database
- Bibliographic material