**TEXT SENTIMENT ANALYSIS**

**MINOR 1 - PROJECT REPORT**

**SUBMITTED BY:**

**SAKSHAM SINGH (20103257) -B9**

**SRIJAN GUPTA (20103286) -B10**

**UTKARSH PATHAK (20103289) -B10**



**December - 2022**

**Submitted in partial fulfillment of the Degree of**

**Bachelor Of Technology**

**in**

**Computer Science and Engineering**

**DEPARTMENT OF CSE/IT**

**JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY UNIVERSITY, NOIDA**

**Table of contents**

|  |  |
| --- | --- |
| *Declaration*  *Certificate*  *Acknowledgment*  *Summary*  *List of figures* | *i*  *ii*  *iii*  *iv*  *v* |
| **Chapter-1**  **Introduction**  ○ 1.1 General Introduction  ○ 1.2 Problem Statement  ○ 1.3 Motivation  ○ 1.4 Brief description of the solution approach | 9-10 |
| **Chapter-2**  **Literature Survey**  ○ 2.1 Paper-I  ○ 2.2 Paper-II  ○ 2.3 Paper-III  ○ 2.4 Paper-IV  ○ 2.5 Paper-V  ○ 2.6 Paper-VI  ○ 2.7 Paper-VII  ○ 2.8 Paper-VIII | 11-12 |
| **Chapter 3**  **Requirement Analysis and Solution Approach**  ○ 3.1 Requirement Analysis  ○ 3.2 Solution Approach | 17-26 |
| **Chapter-4**  **Workflow Diagrams**  ○ 4.1 Workflow Diagram | 27-29 |
| **Chapter-5**  **Implementation**  ○ 5.1 Loading Dataset  ○ 5.2 Data Preprocessing  ○ 5.3 Splitting of dataset  ○ 5.4 Applying Models  ○ 5.5 Applying BERT transformer for fine tuning | 30-35 |
| **Chapter-6**  **Findings,Conclusion and Future Work**  ○ 5.1 Findings  ○ 5.2 Conclusion  ○ 5.3 Future Work | 36 |
| **References** | 39 |

**DECLARATION**

We hereby declare that this submission is our own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

**Place:**

**Date:**

**Signature:**

**Name: Saksham Singh**

**Enrolment No.: 20103257**

**Signature:**

**Name: Srijan Gupta**

**Enrolment No.:20103286**

**Signature:**

**Name: Utkarsh Pathak**

**Enrolment No.:20103289**

**CERTIFICATE**

This is to certify that the work titled “**TEXT SENTIMENT ANALYSIS**” submitted by “**Saksham Singh, Srijan Gupta** and **Utkarsh Pathak**” of B. Tech of Jaypee Institute of Information Technology University, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Signature of Supervisor:

Name of Supervisor: **Dr. K Vimal Kumar/Dr. Ankit Vidyarthi**

Designation: **Assistant Professor (Senior Grade)**

Date:

**ACKNOWLEDGEMENT**

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We take this opportunity to express our profound sense of gratitude and appreciation to all those who helped us throughout the duration of this project. We would like to express our special thanks of gratitude to our mentor **Dr. Ankit Vidyarthi & Mr. K Vimal Kumar** who gave us the golden opportunity to be a part of this wonderful project in the domain of Computer Vision and for providing guidance and expert supervision for this project. She has set a spectacular example on our young impressionable minds in the duration of creating this project. We are really thankful to her. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities We are making this project not only for fulfilling a college requirement but to also increase our knowledge. We express our sincere gratitude towards each and every one who helped us towards the completion of the project.

**Signature:**

**Name: Saksham Singh**

**Enrolment No.: 20103257**

**Date:**

**Signature:**

**Name: Srijan Gupta**

**Enrolment No.:20103286**

**Date:**

**Signature:**

**Name: Utkarsh Pathak**

**Enrolment No.:20103289**

**Date:**

**SUMMARY:**

Sentiment analysis or opinion mining is the computational study of people’s opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons.

First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others’ opinions.

Second, it presents many challenging research problems, which had never been attempted before the year 2000. Part of the reason for the lack of study before was that there was little opinionated text in digital forms. It is thus no surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web. In fact, the research has also spread outside of computer science to management sciences and social sciences due to its importance to business and society as a whole.

Social media is a powerful way to reach new customers and engage with existing ones. Good customer reviews and posts on social media encourage other customers to buy from your company. But the reverse is also true. Negative social media posts or reviews can be very costly to your business.One memorable example is [Elon Musk’s 2020 tweet](https://twitter.com/elonmusk/status/1256239815256797184?lang=en) which claimed the Tesla stock price was too high.The viral tweet wiped $14 billion off Tesla’s valuation in a matter of hours

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure** | **Title** | **Page No.** |
| **1** | **Recurrent Neural Network** | **1** |
| **2** | **RNN Unfolding** | **8** |
| **3** | **LSTM Cell** | **9** |
| **4** | **LSTM Network** | **10** |
| **5** | **Representation of Word2Vec** | **14** |
| **6** | **Transformer** | **15** |
| **7**  **8** | **Encoder**  **BERT** | **16**  **30** |

**Chapter 1: Introduction**

**1.1 General Introduction**

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze customer sentiment. The best businesses understand the sentiment of their customers—what people are saying, how they’re saying it, and what they mean. Customer sentiment can be found in tweets, comments, reviews, or other places where people mention your brand. Sentiment Analysis is the domain of understanding these emotions with software, and it’s a must-understand for developers and business leaders in a modern workplace.

As with many other fields, advances in deep learning have brought sentiment analysis into the foreground of cutting-edge algorithms. Today we use [natural language processing](https://www.datarobot.com/wiki/natural-language-processing/), statistics, and text analysis to extract, and identify the sentiment of words into positive, negative, or neutral categories.

## Uses of Sentiment Analysis-

## Sentiment analysis for brand monitoring

### Sentiment analysis for customer service

### Sentiment analysis for market research and analysis

**1.2 Problem Statement**

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive or negative.

**1.3 Motivation**

The Growing field of Artificial Intelligence fascinates a lot of tech freaks including us. We wanted to know how chatbots classify emotions of humans and then recommend things according to their current state of mind. So, in this project we have tried to classify emotions of human beings using texts and images.In daily life, with the emergence of various social media, more and more users like to express their feelings on Instagram, Twitter and other platforms. The content of expression is also varied, from text and emoticons at the beginning to voice and video messages later. The complexity of emotional information is increasing, and the valuable information contained in the information is also increasing. Mining this emotional characteristic information can not only improve services for the platform but also carry out customized recommendation services according to the audience’s liking degree. Likewise, it can monitor and manage public opinion for the government. Therefore, how to pro- cess and analyse multi-modal emotional information and design a better multi-modal information processing model is very important.

**1.4 Brief description of the solution approach**

Given a text , our goal is to predict whether it conveys a positive or negative emotion. Hence we want to build a text classifier for our data. There are various approaches to perform this task but for our project we pick the approach used in most state-of-the-art textual analysis systems i.e. deep learning.To construct a deep learning model which is very accurate we require huge amounts of data and compute resources. But luckily for us models like [BERT](https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html) are pre trained on large amounts of data and made publicly available. Therefore we can fine tune our model with already pre trained models like BERT on our own data to leverage what the model has already learnt .This process is called [transfer learning](https://en.wikipedia.org/wiki/Transfer_learning).

**Chapter 2: LITERATURE SURVEY**

**RESEARCH PAPER 1**

**TITLE**: **Sentiment Analysis of Public Opinion on The Go-Jek Indonesia Through Twitter Using Algorithm Support Vector Machine**

*"Sentiment analysis of public opinion on the go-jek indonesia through twitter using algorithm support vector machine." In Journal of Physics: Conference Series, vol. 1462, no. 1, p. 012063. IOP Publishing, 2020.*

They have used Support Vector Machine algorithm. SVM can be used to classify text sentiments for 3 classes. The highest accuracy generated for Multiclass tweet classification with Support Vector Machine is 91.8% with 1977 training data features. Document classification error is caused because in a class there are words that are the same as other classes and the weight of words in the other categories is greater than the class they should be, so tweets classified are more likely to approach other classes.

**RESEARCH PAPER 2**

**TITLE**: **A Study on Sentiment Analysis using Text Summarization**

*“A Study on Sentiment Analysis using Text Summarization”**2nd International Conference on the Emerging Technologies in Computing.*

The paper proposed a feature extraction scheme based on text summarization to categorize the Bangla blog texts into their respective sentiment polarities and obtained a maximum accuracy of 98.33%.In the present work, handcrafted features and rule-based algorithm have been used for sentiment analysis of Bangla texts. This study will be useful for various researchers across different domains for understanding and gaining knowledge so that advanced researches can be carried out in sentiment analysis problem.

**RESEARCH PAPER 3**

**TITLE**: **Weibo Text Sentiment Analysis Based on BERT and Deep Learning**

***“****Weibo text sentiment analysis based on bert and deep learning." Applied Sciences 11, no. 22 (2021): 10774.*

At present, deep learning models are popular in the field of sentiment analysis, but the existing traditional models can be further increased in accuracy. This paper proposes a new model based on BERT and deep learning algorithms for sentiment analysis. The model uses the BERT to convert the words in the text into corresponding word vectors, and also introduces a sentiment dictionary to enhance the sentiment intensity of the word vector, and then uses a BiLSTM network to extract the forward and reverse contextual information. In order to more or less emphasize different words in the text, the attention mechanism is added to the output of the BiLSTM network.We conducted experiments on the collected COVID-19 Weibo text dataset. Compared with the comparison model, the proposed model achieves the best experimental results. We also set up ablation experiments to explore the effects of CNN and the sentiment dictionary on the proposed model.

The research in this article points out two directions for our future work. One direction is that the BERT model consumes huge hardware resources and requires a huge corpus for training. How to train a high-quality and light weight BERT model and apply the trained BERT model to downstream tasks is our future research direction. On the other hand, we noticed that the accuracy of the experiment could be further improved by introducing an external sentiment dictionary. Therefore, our next goal is to introduce more external knowledge, such as part-of-speech information.

**RESEARCH PAPER 4**

**TITLE**: **Sentiment Analysis in English Texts**

*Advances in Science, Technology and Engineering Systems Journal* 5, no. 6 (2020): 1683-1689.

Sentiment Analysis is a Natural Language Processing field that increasingly attracted researchers, government authorities, business owners, services providers, and companies to improve products, services, and research. In this research paper, the authors aimed to survey sentiment analysis approaches. Therefore, 16 research papers that studied Twitter's text classification and analysis were surveyed. The authors also aimed to evaluate different machine learning algorithms used to classify sentiment to either positive or negative, or neutral. This experiment aims to compare the efficiency and performance of different classifiers that have been used in the sixteen papers that are surveyed. These classifiers are (Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree). Besides, the authors investigated the balanced dataset factor by applying the same classifiers twice on the dataset, one on the unbalanced and the other, after balancing the dataset. The targeted dataset included six datasets about six American airline companies (United, Delta, Southwest, Virgin America, US Airways, and American); it consists of about 14000 tweets. The authors reported that the classifier's accuracy results were very high in some datasets while low in others. The authors indicated that the dataset size was the reason for that. On the balanced dataset, the Naïve Bayes classifier, Decision Tree, and ID3 were mostly better than other classifiers and have given the almost same level of accuracy. The classifiers with Virgin America dataset reported the lowest level of accuracy due to its small size. On the unbalanced dataset, results show that the Naive Byes and ID3 gave a better level of accuracy than other classifiers when it’s applied on the balanced datasets. While (K- NN, Decision Tree, Random Forest, and Random Tree) gave a better understanding of the unbalanced datasets.

**RESEARCH PAPER 5**

**TITLE**: Sentiment Analysis: A Survey of Current Research and Techniques

*International Journal of Innovative Research in Computer sand Communication Engineering* 3, no. 5 (2015).

Opinion Mining (OM) is natural language processing dealing with tracking the mood of people regarding a product or topic. OM combines information retrieval and computational linguistic techniques handling a document’s opinions. The field’s main goal is solving problems related to opinions on products, politics in newsgroup posts and review sites. It provides automatic opinion extraction, emotions and sentiments in text tracking attitudes and feelings on the web. People express views by writing blog posts, comments, reviews and tweets on various topics. Tracking products and brands and determining if they are viewed positively or negatively is done on the web. OM has slightly different tasks and many names like opinion extraction, sentiment analysis, sentiment mining, affect analysis, subjectivity analysis, emotion analysis and review mining. But, they all come under sentiment analysis or OM.This study defined the concept of opinion in a sentiment analysis context, the main tasks being a framework of OM. Sentiment analysis deals with evaluation opinions or opinions type implying positive or negative sentiments. Reviews reveal that different features and classification algorithms combine efficiently to overcome individual drawbacks and benefit from each other’s merits. Finally they enhance sentiment classification performance. More work in future is needed to improve performance measures. The main challenge is in using other languages, dealing with negation expressions and producing an opinions summary based on product features/attributes, handling of implicit product features, complexity of sentence/ document, etc. Future research could be dedicated for the challenges.

**RESEARCH PAPER 6**

**TITLE**: Emotion and sentiment analysis of tweets using BERT

*"Emotion and sentiment analysis of tweets using BERT." In EDBT/ICDT Workshops. 2021.*

The goal of this work was the evaluation of the use of Bidirectional Encoder Representations from Transformers (BERT) models for both sentiment analysis and emotion recognition of Twitter data. We defined an architecture composed of BERT-Base followed by a final classification stage and we fine-tuned the model for the above-mentioned tasks. We measured the performance of our classifiers by considering two datasets of tweets and we obtained a remarkable 92% accuracy for sentiment analysis and a 90% accuracy for emotion analysis, from which it was possible to deduce that BERT’s language modeling power significantly contributes to achieve a good text classification.

In future work, we plan to improve the performance of our classifiers by determining the best number of layers and neurons in the final classification layers (i.e., fully connected layers) . We also intend to extend the experimentation by considering larger datasets, such as the SemEval 2017 Task 4 [23] dataset for sentiment analysis and the EmoBank [4] dataset for emotion analysis. This is particular important for the sentiment analysis task, in which we observed a repentine increment of the validation loss after the first epoch, probably due to overfitting. Although the models reach high accuracy and the approach seems promising, a comparison with other state-of-the-art classifiers will be useful to thoroughly evaluate the performance of our approach. We also intend to investi-gate the impact of BERT-Base by replacing it with other BERT distributions (e.g., BERT-Large) or traditional word embeddings, such as Word2Vec [17] or GloVe [21].

**RESEARCH PAPER 7**

**TITLE**: **Aspect-Based Sentiment Analysis Using BERT**

*"Aspect-based sentiment analysis using bert." In Proceedings of the 22nd nordic conference on computational linguistics, pp. 187-196. 2019.*

In this paper, we proposed an ABSA model that can predict the aspect related to a text for in- domain and out-of-domain. We achieve this by using the pre-trained language model BERT and fine-tuning it to a sentence pair classification model for the ABSA task. Moreover, we train the aspect classifier model with data that we generate, which consist of ’related’ and ’unrelated’ labels.

We further experimented with this approach for the sentiment classifier, by fine-tuning the model to find a relation between an aspect and a text and to make the model learn when the contextual representation showed a sentiment context. Further- more, we proposed a combined model that can classify both aspect and sentiment using only one sentence pair classification model. Experimental results show that the combined model outperforms previous state-of-the-art results for aspect based sentiment classification

**RESEARCH PAPER 8**

**TITLE**:**Sentiment analysis using product review data**

*"Sentiment analysis using product review data." Journal of Big Data 2, no. 1 (2015): 1-14.*

Sentiment analysis or opinion mining is a field of study that analyzes people’s sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process has been proposed along with detailed descriptions of each step. Exper- iments for both sentence-level categorization and review-level categorization have been performed.

**Chapter 3: Requirement Analysis and Solution Approach**

**3.1 Requirement Analysis**

**Language Used:**

· Python 3

**Libraries to be Used:**

· pandas

· numpy

· tensorflow/pytorch

· keras

· matplotlib

· sklearn

· nltk

. seaborn

. imblearn

. transformers

**Dataset:**

· Twitter data (training.1600000.processed.noemoticon.csv)

**3.2 Solution Approach**

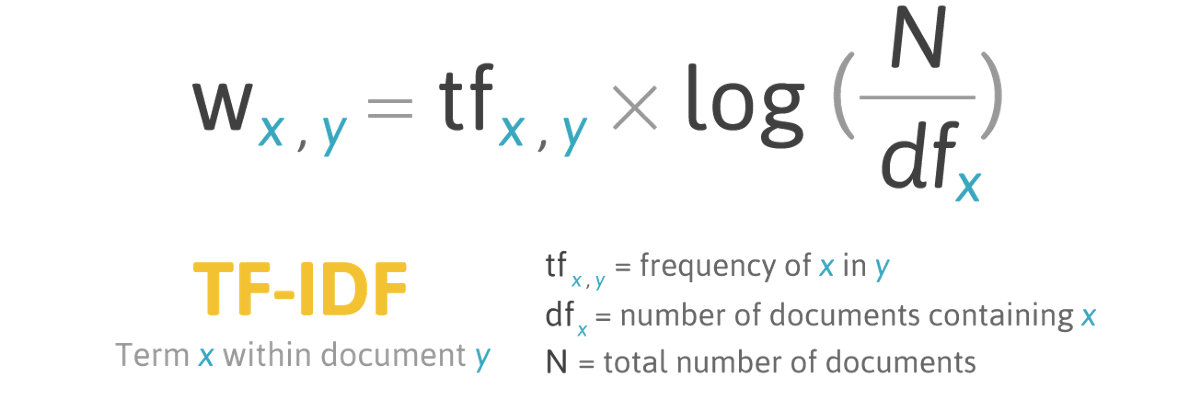
# **TF-IDF**

TF-IDF stands for*term frequency-inverse document frequency* and it is a measure, used in the fields of [information retrieval (IR)](https://en.wikipedia.org/wiki/Information_retrieval) and machine learning, that can quantify the importance or relevance of string representations (words, phrases, lemmas, etc)  in a document amongst a collection of documents (also known as a corpus).

Term frequency works by looking at the frequency of a *particular term* you are concerned with relative to the document. There are several ways, of defining frequency:

* Number of times the word appears in a document (raw count).
* Term frequency adjusted for the length of the document (raw count of occurences divided by number of words in the document).

Inverse document frequency looks at how common (or uncommon) a word is amongst the corpus.

 - (1)

Taking inverse document frequency, we can minimize the weighting of frequent terms while making infrequent terms have a higher impact.

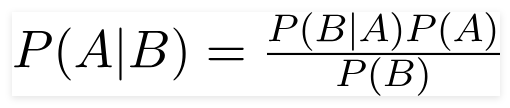
The key intuition motivating TF-IDF is the importance of a term is inversely related to its frequency across documents.TF gives us information on how often a term appears in a document and IDF gives us information about the relative rarity of a term in the collection of documents. By multiplying these values together we can get our final TF-IDF value.

The higher the TF-IDF score the more important or relevant the term is; as a term gets less relevant, its TF-IDF score will approach 0.

# **Naive Bayes Classifier**

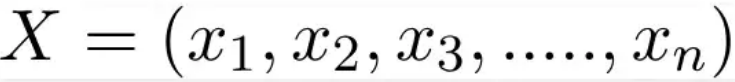
A Naive Bayes classifier is a probabilistic machine learning model that’s used for classification task. The crux of the classifier is based on the Bayes theorem.

Bayes Theorem:

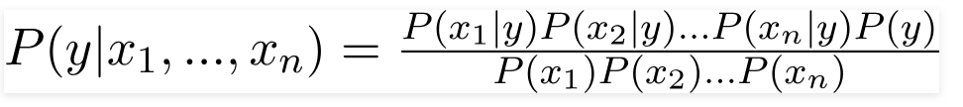
- (2)

Using Bayes theorem, we can find the probability of **A** happening, given that **B** has occurred. Here, **B** is the evidence and **A** is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

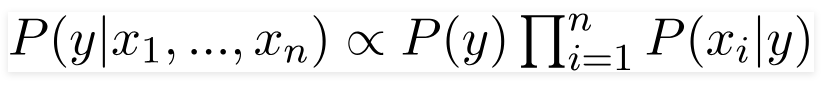
**X** is given as,

 - (3)

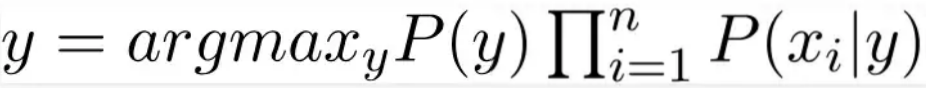
Here x\_1,x\_2….x\_n represent the features, i.e they can be mapped to outlook, temperature, humidity and windy. By substituting for **X**and expanding using the chain rule we get,

- (4)

Now, you can obtain the values for each by looking at the dataset and substitute them into the equation. For all entries in the dataset, the denominator does not change, it remain static. Therefore, the denominator can be removed and a proportionality can be introduced.

- (5)

There could be cases where the classification could be multivariate. Therefore, we need to find the class **y** with maximum probability.

 - (6)

## Recurrent Neural Network (RNN)

RNN works on the principle of saving the output of a particular layer and feeding this back to the input in order to predict the output of the layer.

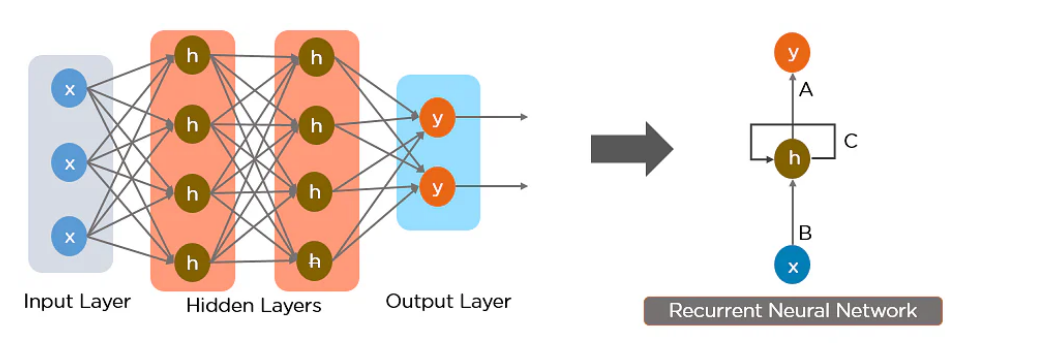


Fig 1

Recurrent Neural Network

If we have data in a sequence such that one data point depends upon the previous data point, we need to modify the neural network to incorporate the dependencies between these data points. RNNs have the concept of “memory” that helps them store the states or information of previous inputs to generate the next output of the sequence.

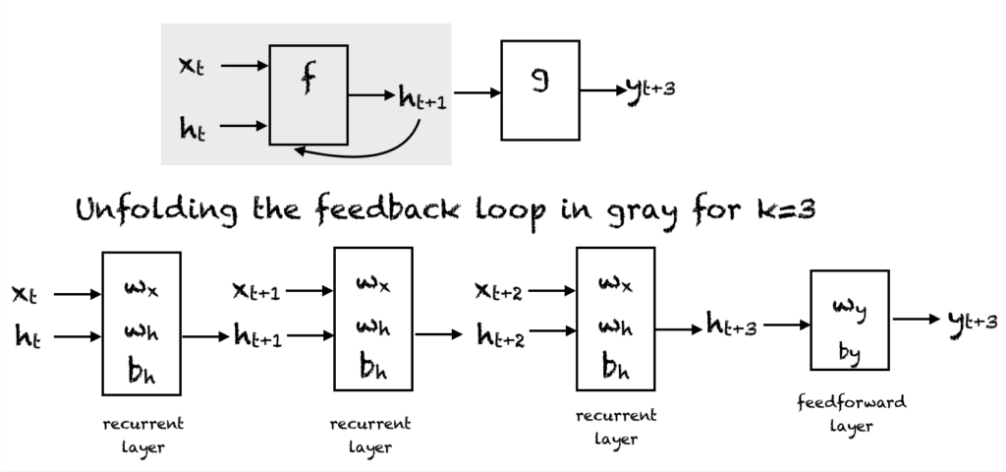


Fig 2

RNN Unfolding

RNN were created because there were a few issues in the feed-forward neural network:

* Cannot handle sequential data
* Considers only the current input
* Cannot memorize previous inputs

The solution to these issues is the RNN. An RNN can handle sequential data, accepting the current input data, and previously received inputs. RNNs can memorize previous inputs due to their internal memory.

# **Long Short Term Memory (LSTM)**

Long Short Term Memory Network is an advanced RNN, a sequential network, that allows information to persist.They remember the previous information and use it for processing the current input. The shortcoming of RNN is, they can not remember Long term dependencies due to vanishing gradient. LSTMs are explicitly designed to avoid long-term dependency problems.

The central role of an LSTM model is held by a memory cell known as a ‘cell state’ that maintains its state over time. The cell state is the horizontal line that runs through the top of the below diagram. It can be visualized as a conveyor belt through which information just flows, unchanged.

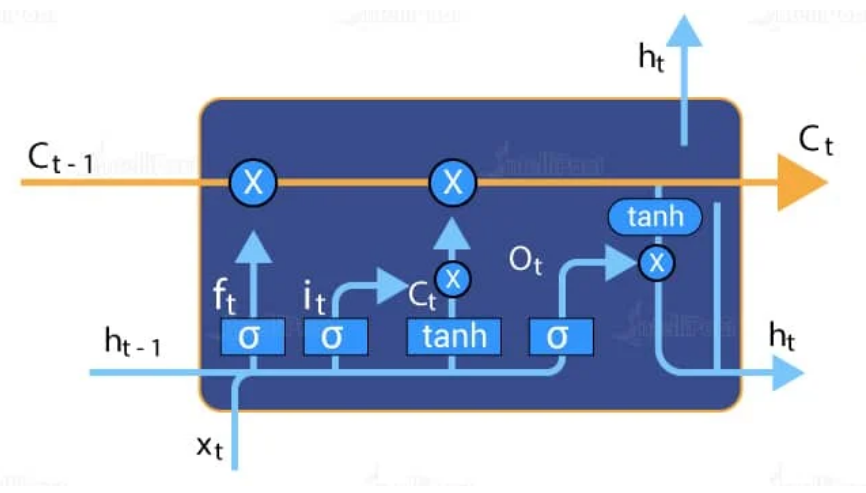


Fig 3

LSTM Cell

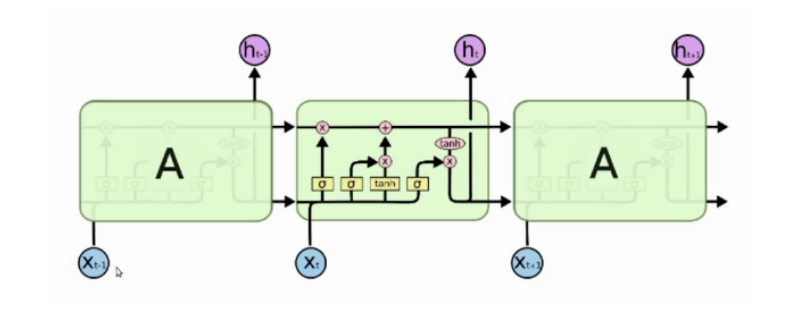


Fig 4

LSTM Network

In a cell of the LSTM network, the first step is to decide whether we should keep the information from the previous timestamp or forget it, this is done by the forget gate. Input gate is used to quantify the importance of the new information carried by the input.The output gate determines the value of the next hidden state. This state contains information on previous inputs.

## 

## Word2Vec

Word2Vec creates vectors of the words that are distributed numerical representations of word features – these word features could comprise of words that represent the context of the individual words present in our vocabulary. Word embeddings eventually help in establishing the association of a word with another similar meaning word through the created vectors.

As seen in the image below where word embeddings are plotted, similar meaning words are closer in space, indicating their semantic similarity.

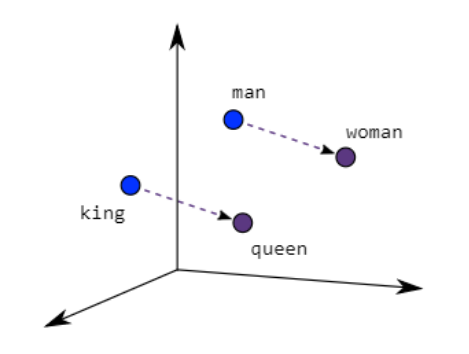


Fig 5

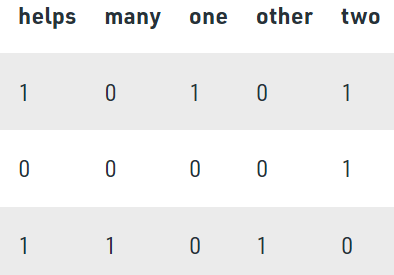
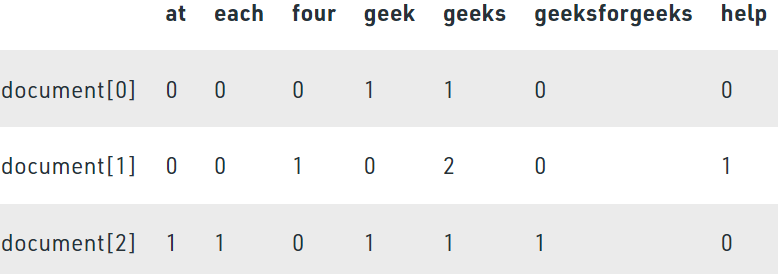
Representation of Word2Vec

**Countvectorizer**

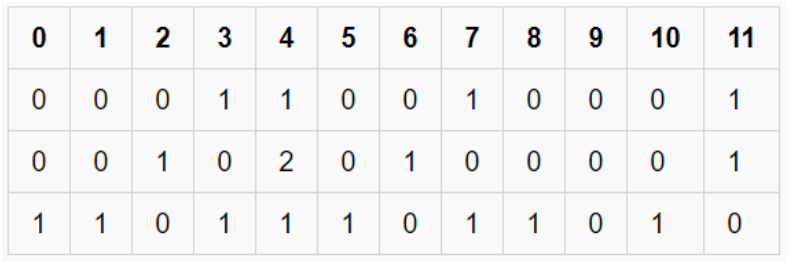
CountVectorizer is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text. This is helpful when we have multiple such texts, and we wish to convert each word in each text into vectors (for using in further text analysis).

CountVectorizer creates a matrix in which each unique word is represented by a column of the matrix, and each text sample from the document is a row in the matrix. The value of each cell is nothing but the count of the word in that particular text sample. This can be visualized as follows –

*document = [ “One Geek helps Two Geeks”, “Two Geeks help Four Geeks”, “Each Geek helps many other Geeks at GeeksforGeeks.”]*

**

CountVectoriser

******

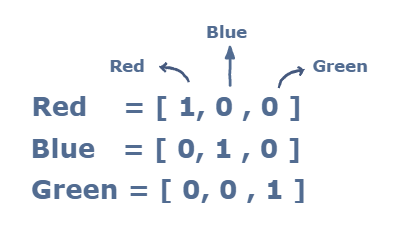
*This way of representation is known as a* ***Sparse Matrix***

## One Hot Encoding

A one hot encoding is a representation of categorical variables as binary vectors.

This first requires that the categorical values be mapped to integer values. Then, each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1.

Ex:



**Transformers**

Transformers are semi-supervised machine learning models that are primarily used with text data and have replaced recurrent neural networks in natural language processing tasks.

**How Transformers Works?**

Transformers were originally introduced by researchers at Google in the 2017 NIPS paper Attention is All You Need. Transformers are designed to work on sequence data and will take an input sequence and use it to generate an output sequence one element at a time.

For example, a transformer could be used to translate a sentence in English into a sentence in French. In this case, a sentence is basically treated as a sequence of words. A transformer has two main segments, the first is an encoder that operates primarily on the input sequence and the second is a decoder that operates on the target output sequence during training and predicts the next item in the sequence. In a machine translation problem, for example, the transformer may take a sequence of words in English and iteratively predict the next French word in the proper translation until the sentence has been completely translated. The diagram below demonstrates how a transformer is put together, with the encoder on the left and the decoder on the right.

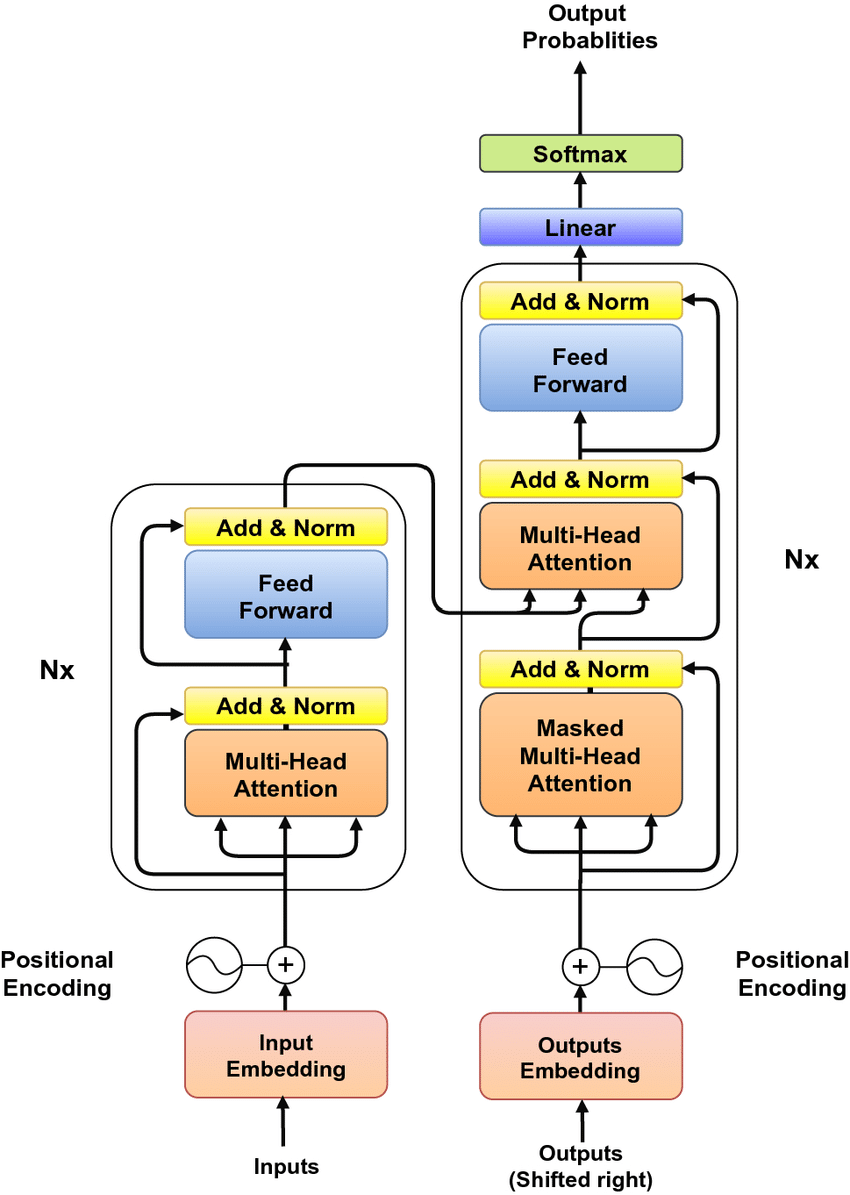


Fig 6

Transformers

It looks like there’s a lot going on in the diagram above, so let’s take a look at each component separately. The parts of a transformer that are particularly important are the embeddings, the positional encoding block, and the multi-head attention blocks.

**The Encoder**

The encoder is the part of the transformer that chooses what parts of the input to focus on. The encoder can take a sentence such as “the quick brown fox jumped”, computes the embedding matrix, and then converts it into a series of attention vectors. The multi-head attention block initially produces these attention vectors, which are then added and normalized, passed into a fully-connected layer (Feed Forward in the diagram above), and normalized again before being passed over to the decoder.

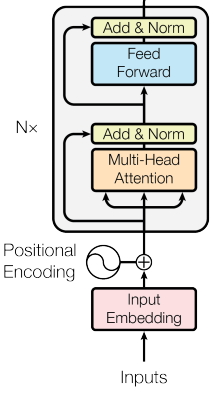


Fig 7

Encoder

**The Decoder**

During training, the decoder operates directly on the target output sequence. As per our example, let’s assume the target output is the French translation of the English sentence “the quick brown fox jumped”, which translates to “le renard brun rapide a sauté” in French. In the decoder, separate embedding vectors are computed for each French word in the sentence, and the positional encoding is also applied in the form of sine and cosine functions.

However, a masked attention block is used, meaning that only the previous word in the French sentence is used and the other words are masked. This allows the transformer to learn to predict the next French word. The outputs of this masked attention block are added and normalized before being passed to another attention block that also receives the attention vectors produced by the encoder. A feed-forward network receives the final attention vectors and uses them to produce a single vector with a dimension equal to the number of unique words in the model’s vocabulary. Applying the softmax activation function to this vector produces a set of probabilities corresponding to each word. In the context of our example, these probabilities predict the likelihood of each French word appearing next in the translation.

This is how a transformer performs tasks such as machine translation and text generation. Just as demonstrated in the figure below, a transformer iteratively predicts the next word in a translated sentence when performing translation tasks.

**BERT**

BERT, which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection.

Historically, language models could only read text input sequentially -- either left-to-right or right-to-left -- but couldn't do both at the same time. BERT is different because it is designed to read in both directions at once. This capability, enabled by the introduction of Transformers, is known as bidirectionality. The BERT framework was pre-trained using text from Wikipedia and can be fine-tuned with question and answer datasets.

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

BERT is made possible by Google's research on Transformers. The transformer is the part of the model that gives BERT its increased capacity for understanding context and ambiguity in language. The transformer does this by processing any given word in relation to all other words in a sentence, rather than processing them one at a time. By looking at all surrounding words, the Transformer allows the BERT model to understand the full context of the word, and therefore better understand searcher intent.

These word embedding models require large datasets of labeled data. While they are adept at many general NLP tasks, they fail at the context-heavy, predictive nature of question answering, because all words are in some sense fixed to a vector or meaning. BERT uses a method of masked language modeling to keep the word in focus from "seeing itself" -- that is, having a fixed meaning independent of its context. BERT is then forced to identify the masked word based on context alone. In BERT words are defined by their surroundings, not by a pre-fixed identity.

BERT is also the first NLP technique to rely solely on self-attention mechanism, which is made possible by the bidirectional Transformers at the center of BERT's design. This is significant because often, a word may change meaning as a sentence develops. Each word added augments the overall meaning of the word being focused on by the NLP algorithm. The more words that are present in total in each sentence or phrase, the more ambiguous the word in focus becomes. BERT accounts for the augmented meaning by reading bidirectionally, accounting for the effect of all other words in a sentence on the focus word and eliminating the left-to-right momentum that biases words towards a certain meaning as a sentence progresses.

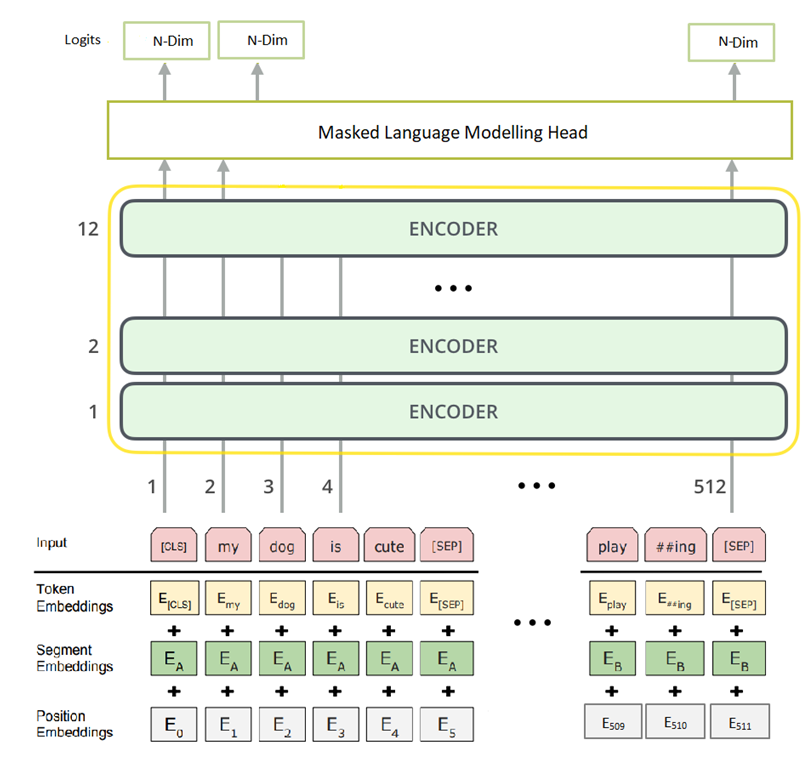
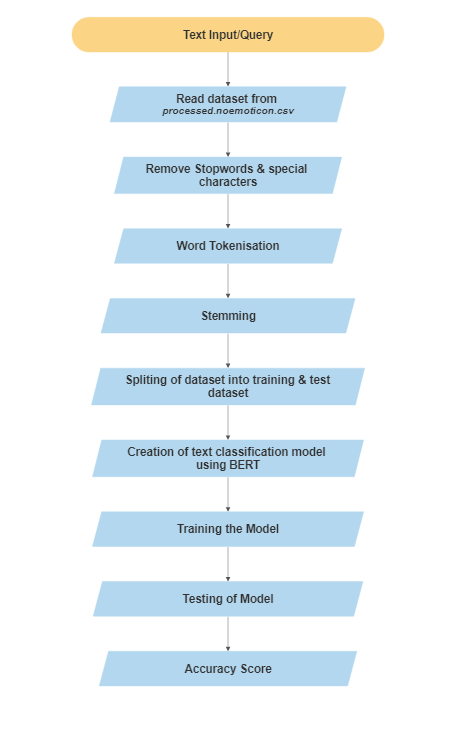


Fig 8

BERT

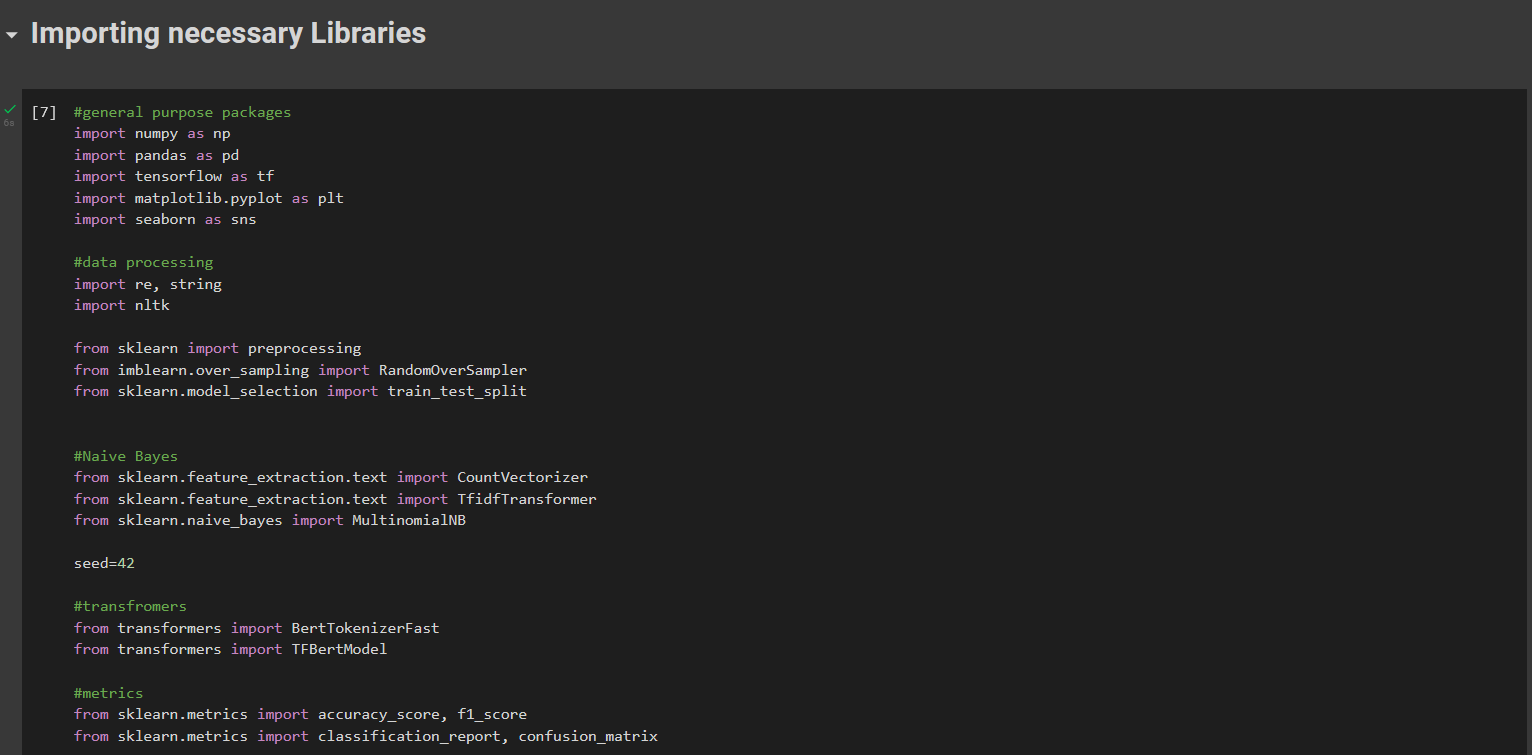
Based on the depth of the model architecture, two types of BERT models are introduced namely BERTBase and BERTLarge. The BERTBase model uses 12 layers of transformers block with a hidden size of 768 and number of self-attention heads as 12 and has around 110M trainable parameters. On the other hand, BERTLarge uses 24 layers of transformers block with a hidden size of 1024 and number of self-attention heads as 16 and has around 340M trainable parameters.

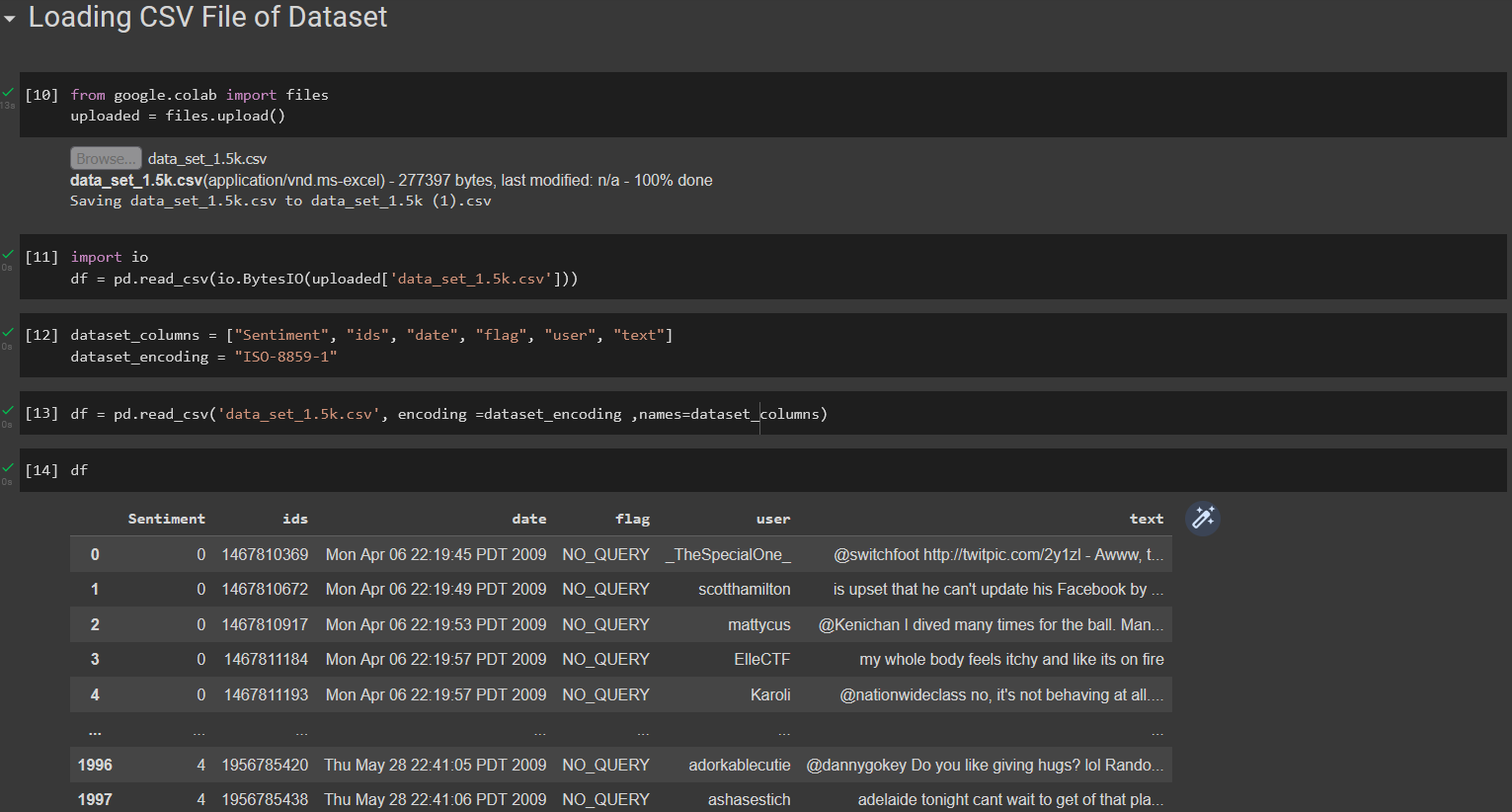
**Chapter 4: Workflow Diagram**



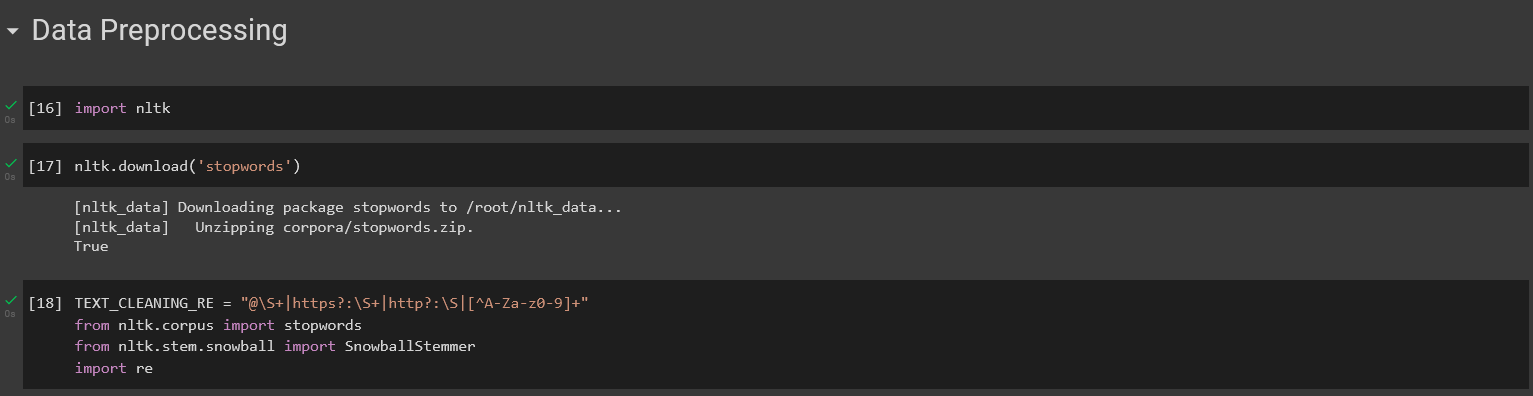
**Chapter 5:Implementation**

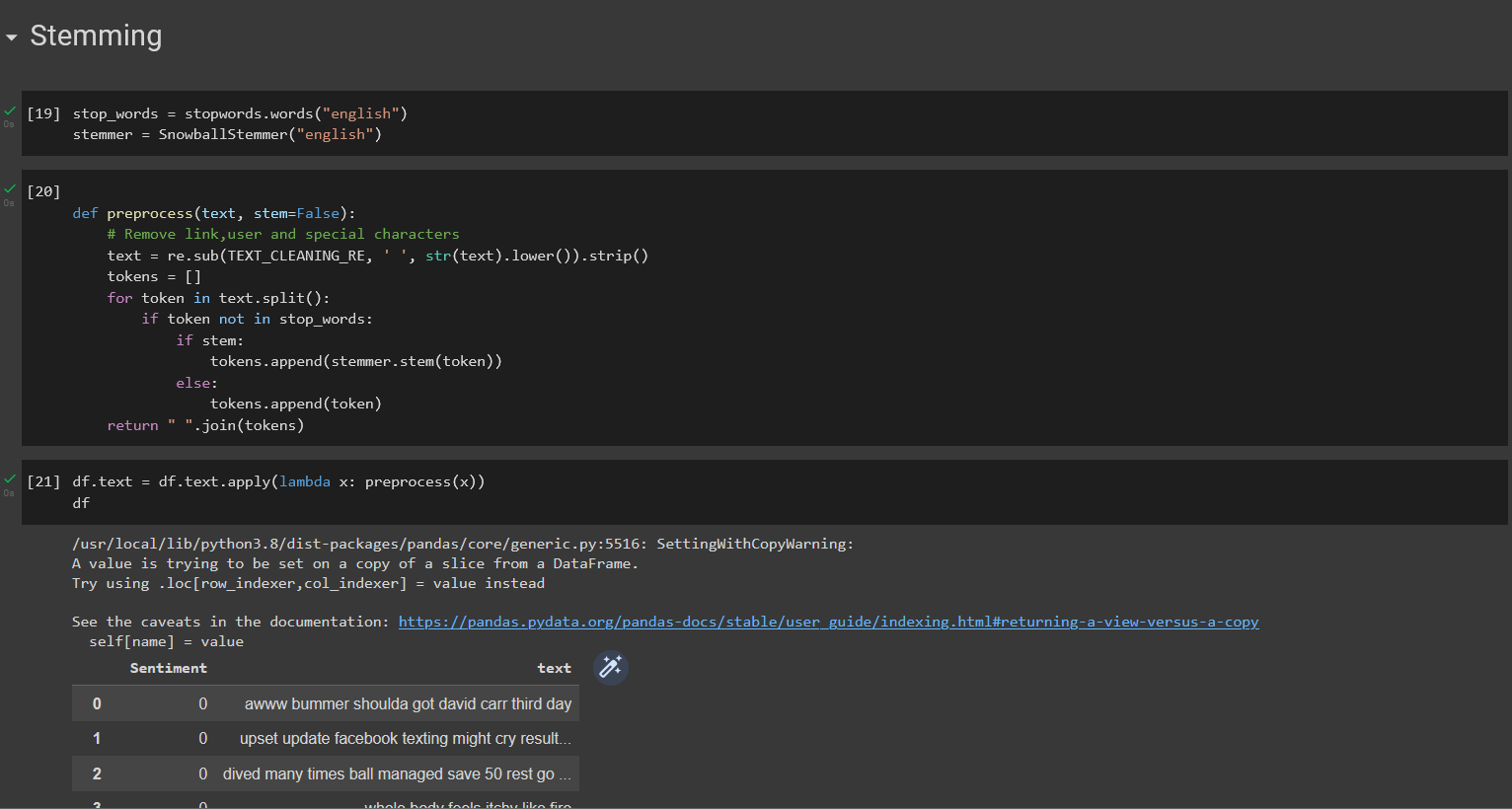
**5.1 Loading dataset**

****

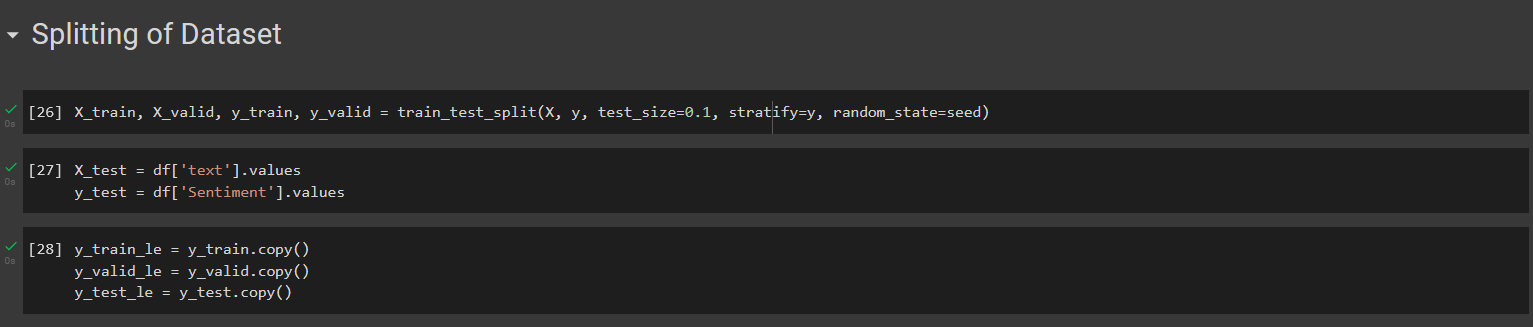
****

**5.2 Data Preprocessing**

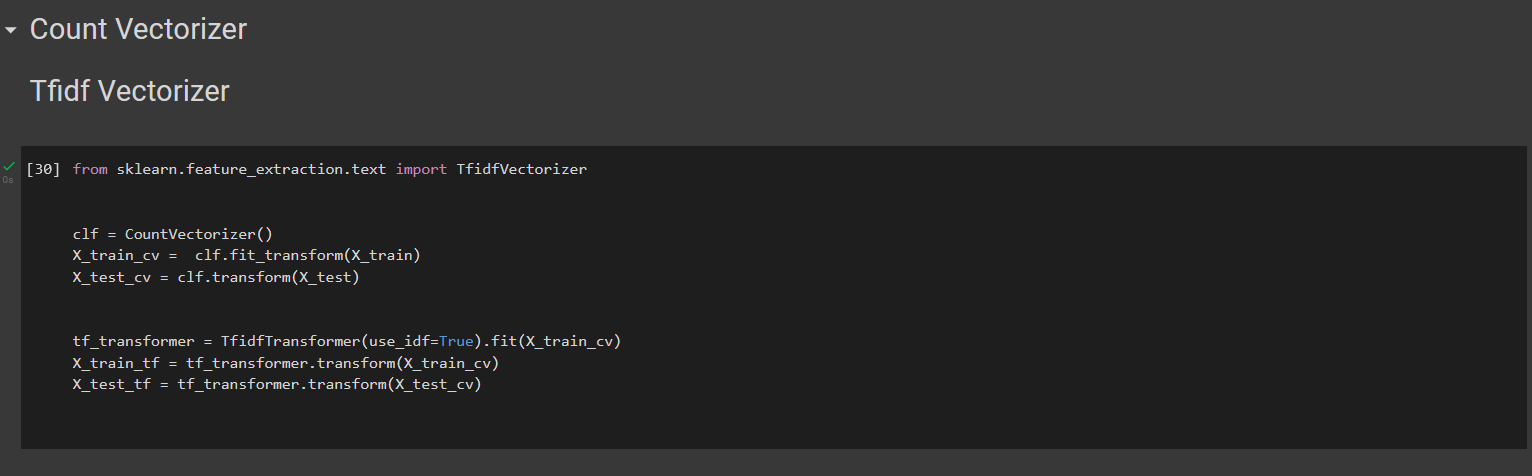
****

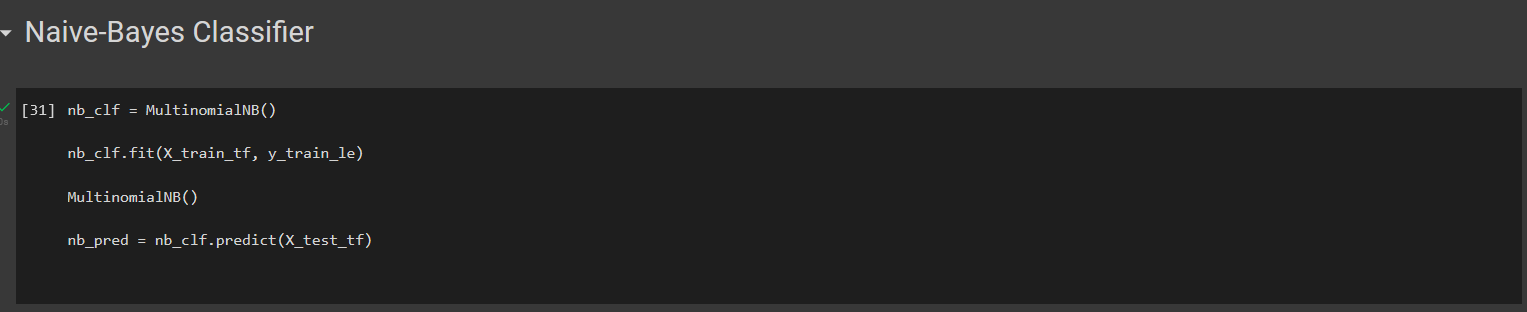
****

**5.3 Splitting Of Dataset**

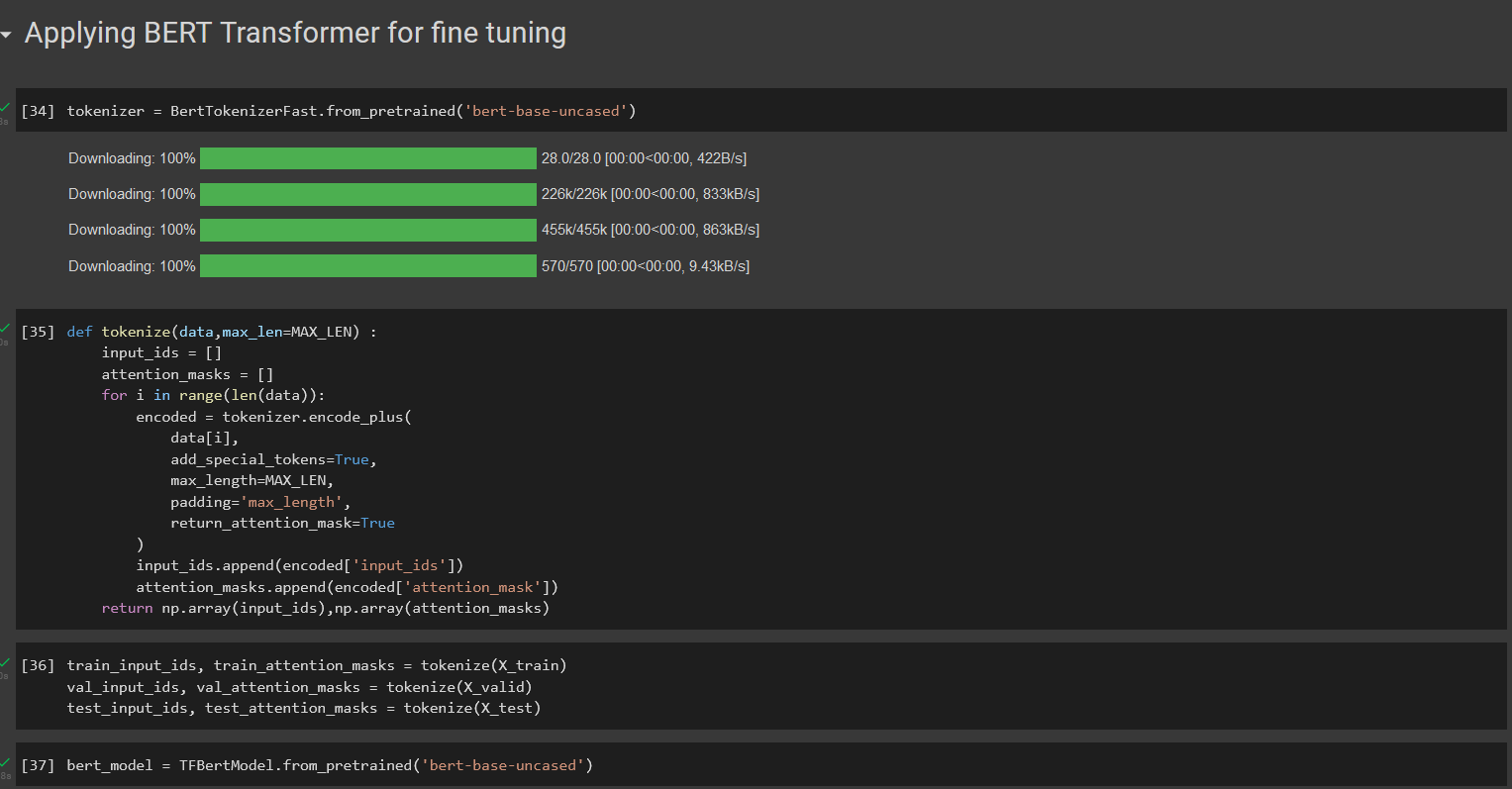
****

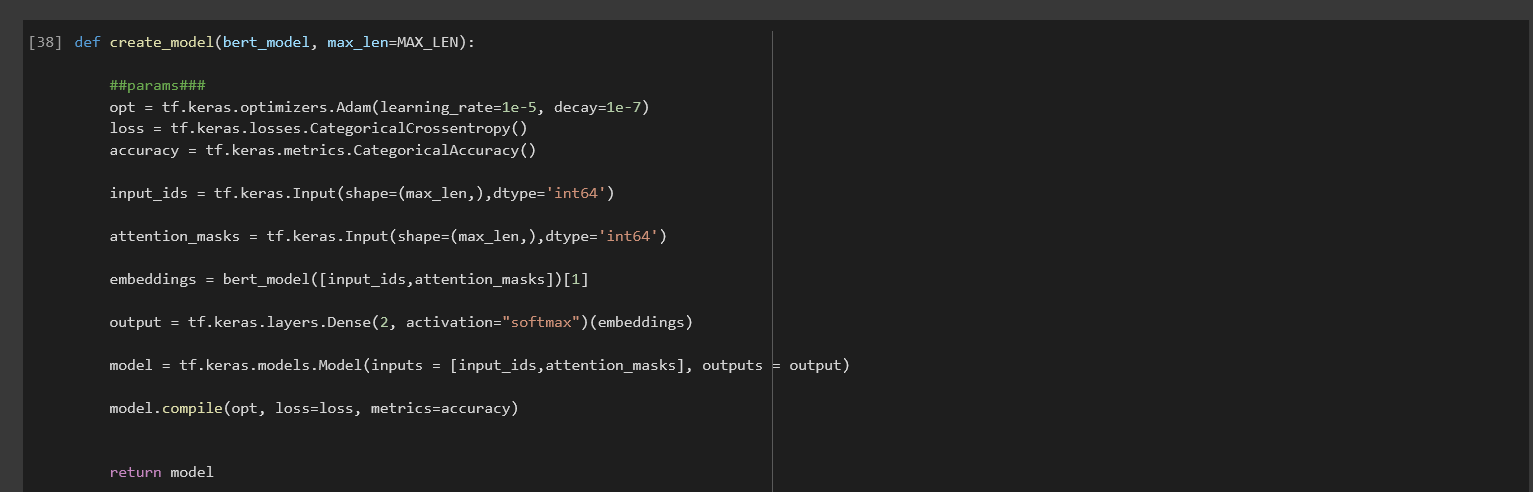
**5.4 Applying models**

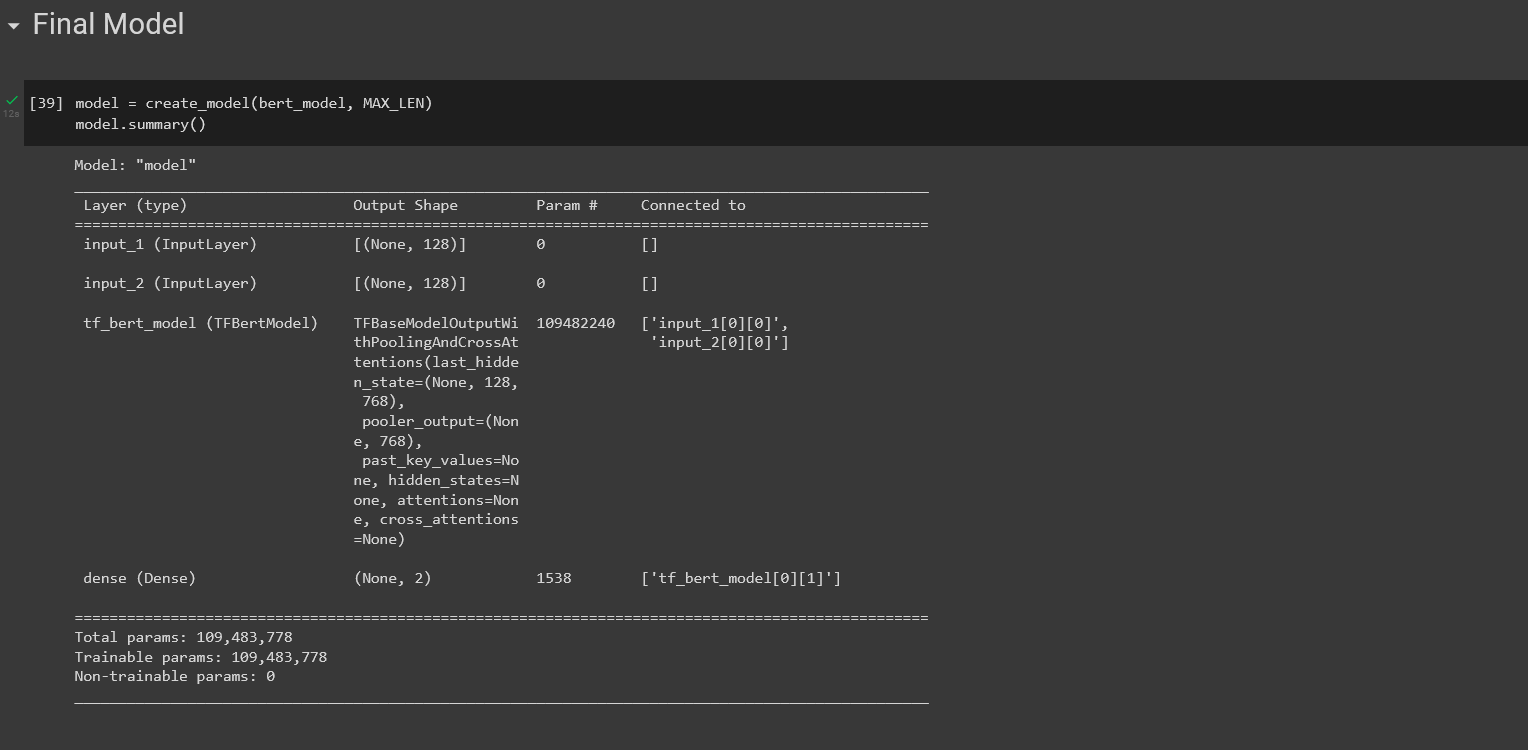
****

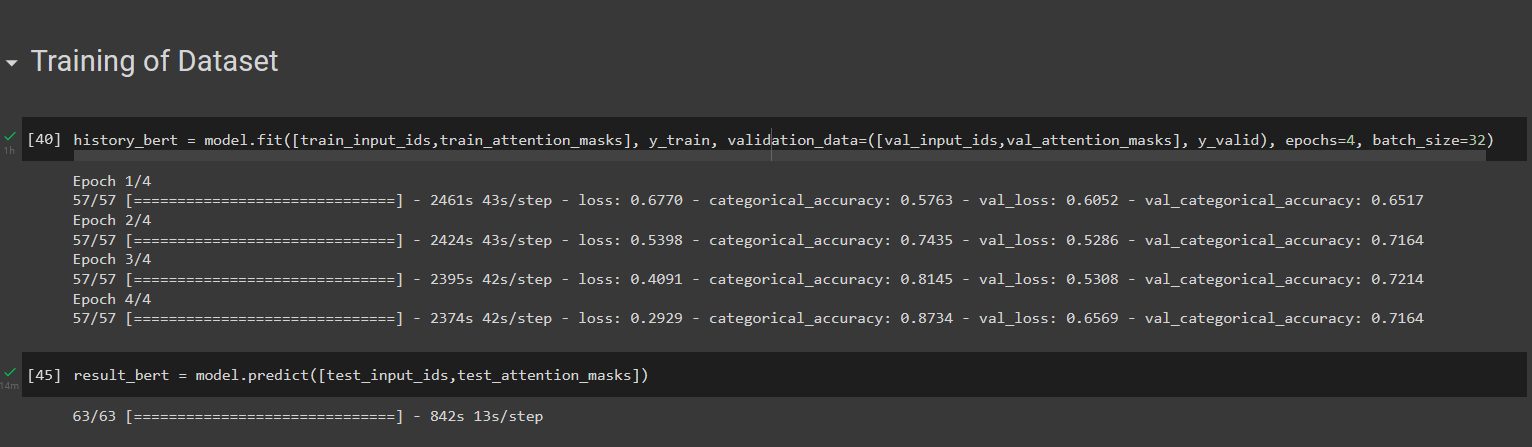
****

**5.5 Applying BERT Transformer for fine tuning**

****

****

****

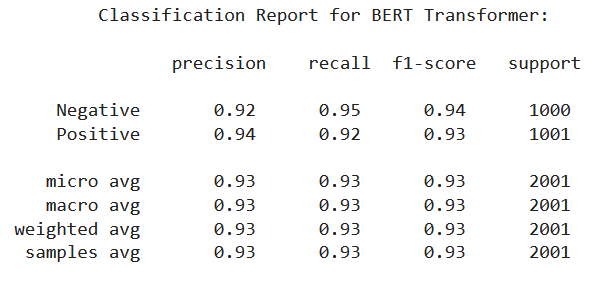
****

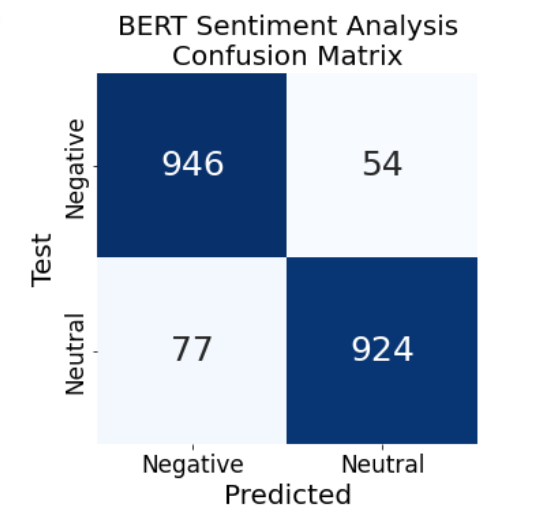
**Chapter 6: Findings,Conclusion and Future Work**

**6.1 Findings**

Accuracy for positive sentiments is 94%

Accuracy for negative sentiments is 92%



****

**6.2 Conclusion**

Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings and emotions regarding their brand. To date, the majority of sentiment analysis projects have been conducted almost exclusively by companies and brands through the use of social media data, survey responses and other hubs of user-generated content. By investigating and analyzing customer sentiments, these brands are able to get an inside look at consumer behaviors and, ultimately, better serve their audiences with the products, services and experiences they offer.

Our project presents a new way to give the ability to machine to determine the emotion with the help of the text. Emotional characteristic information can improve services for the platform and hence the audience.

It can monitor and manage public opinion for any product or scheme. It can also help to monitor public opinion for the government.

**6.3 FUTURE POSSIBILITIES**

We will be analysing different techniques to perform sentiment analysis to find efficient algorithms in order to perform this task. We will be further merging different datasets to produce more accurate results.

As future work, we will extend our experiments on the proposed method with more datasets, different combination of features, and with deep learning and other machine learning techniques.

Our project aims to determine the emotion with the text of a user. Our project can be extended to integrate with the robot to help it to have a better understanding of the mood the corresponding human is in, which will help it to have a better conversation. Any e-commerce sites which have an AI based chat bot which recommends the customer to have a good experience our project will determine the customer/s mood and accordingly it can help the chat bot to have a better recommendation.

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens. This forecast also predicts broader applications for sentiment analysis – brands will continue to leverage this tool, but so will individuals in the public eye, governments, nonprofits, education centers and many other organizations.

**REFERENCES**

**Research Papers:**

1. Syahputra, H., Basyar, L. K., & Tamba, A. A. S. (2020, February). Sentiment analysis of public opinion on the go-jek indonesia through twitter using algorithm support vector machine. In Journal of Physics: Conference Series (Vol. 1462, No. 1, p. 012063). IOP Publishing.

2. Dhar, Ankita & Mukherjee, Himadri & Bose, Rajesh & Roy, Sandip & Roy, Kaushik. (2022). A Study on Sentiment Analysis using Text Summarization.

3. Li, H., Ma, Y., Ma, Z., & Zhu, H. (2021). Weibo text sentiment analysis based on bert and deep learning. Applied Sciences, 11(22), 10774.

4. Alshamsi, A., Bayari, R., & Salloum, S. (2020). Sentiment analysis in English texts. Advances in Science, Technology and Engineering Systems Journal, 5(6), 1683-1689.

5. Jeevanandam Jotheeswaran, D. S. K. (2015). Sentiment Analysis: A Survey of Current Research and Techniques. International Journal of Innovative Research in Computer sand Communication Engineering, 3(5).

6. Chiorrini, A., Diamantini, C., Mircoli, A., & Potena, D. (2021, March). Emotion and sentiment analysis of tweets using BERT. In EDBT/ICDT Workshops.

7. Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-based sentiment analysis using bert. In Proceedings of the 22nd nordic conference on computational linguistics (pp. 187-196).

8.Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 1-14.

**Online Site**

1. Dhruvil Karani, Introduction to Word Embedding and Word2Vec, Towards Data Science, Sep 1 2018

<https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>

1. Enes Gokce, Sentiment Analysis on Amazon Reviews, Towards Data Science, May 26 2020

<https://towardsdatascience.com/sentiment-analysis-on-amazon-reviews-45cd169447ac>

1. Vijay Anandan, Sentiment Analysis on Voice Data, Analytics Vidhya, Jan 15 2021

<https://medium.com/analytics-vidhya/sentiment-analysis-of-voice-data-64533a952617>

1. Tracyrenee, RAVDESS Emotional Speech and Song dataset: an analysis, CodeX, March 20 2020

<https://medium.com/codex/ravdess-emotional-speech-and-song-dataset-an-analysis-eabe687817df>

1. SAVEE Dataset, Kaggle, July 01 2019

<https://www.kaggle.com/barelydedicated/savee-database>

1. RAVDESS Dataset. Kaggle. Jan 19 2019

<https://www.kaggle.com/uwrfkaggler/ravdess-emotional-speech-audio>

------------ THANK YOU -------------