ABSTRACT

Humans have such a wide range of emotions and they can express their emotions through various methods. Writing is one of them, however by the reign of social media now people like to express their emotions on social media. In this research effort we used 6 emotions Sadness, Joy, Anger, Fear, and Surprise. We used multiple models and trained each model with 16000 training data and tested with 2000 test data.

Three deep learning models—Bidirectional Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)—are being used in this study. The outcomes show how well the models that were put into use work for identifying tweet emotions. The accuracy of the Bidirectional GRU model is 91.85%, whereas the accuracy of the CNN model is somewhat higher at 91.95%. With an accuracy of 88.50%, the LSTM model performs admirably. To learn more about the classification skills of the models, evaluation measures including confusion matrix, f-1 scores, precision, recall, and ROC AUC scores are used.

By comparing the Bidirectional GRU, CNN, and LSTM models on the Hugging Face Emotion dataset, this effort makes a contribution to the field of twitter emotion identification. In order to choose the best models for their applications, researchers and practitioners can benefit greatly from the findings. As a foundation for future study and advancements in sentiment analysis and social media analytics, the assessment metrics and analysis offered here.

Overall, the research effort shows how well Bidirectional GRU, CNN, and LSTM models recognize tweet emotions. The results expand sentiment analysis and offer helpful recommendations for social networking analytics applications.

Keywords: Tweet, Emotion Recognition, Sentiment Analysis, GRU, LSTM, CNN comparative analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Understanding user sentiment and behavior is largely dependent on the ability to recognise emotions, especially in the setting of social media. Different types of machine learning methods, including supervised and unsupervised learning, can be used to tackle the problem of emotion identification in tweets. Because there were labeled data in the dataset, we used supervised machine learning in this case. We were able to train models using supervised learning to recognize patterns in the labeled tweets and generate precise predictions about unobserved data. Using deep learning techniques, this research effort focuses on tweet emotion identification. Convolutional neural network (CNN), long short-term memory (LSTM), and bidirectional gated recurrent unit (GRU) performance is compared. By offering information on the efficiency of these deep learning models, the findings advance the area of social media emotion identification. This can also serve as a guide for academics and professionals when choosing the best models for sentiment analysis on tweet data. Such simulations may help with brand management or marketing efforts by helping to understand user attitudes, spot patterns, and make data-driven decisions. Additionally, they can support the study of public opinion, assisting decision-makers in determining how the public feels about a range of topics. Such models can also help detect offensive or improper content during content moderation, improving online safety and community standards enforcement.

1.2 Motivation

It is crucial to assess emotions on social media. According to a recent Pew Research Centre survey, over 70 percent of Americans use social networking sites, with Twitter being among the most popular platforms. Since people freely express their thoughts on the site and there are millions of tweets written every day, it is crucial to correctly analyze the emotion. This will enable us to understand how they are feeling [1]. We can use the power of labelled data to build models that can precisely anticipate tweet sentiments using supervised machine learning. This strategy has produced positive results in a number of trials. As an illustration, Wang et al.'s work [2] used supervised machine learning to classify tweet emotions with an accuracy of 89%. This illustrates how supervised learning may successfully capture the complex emotions conveyed in tweets. However, we created three different models to properly detect the emotions. Among the three models, two have a better accuracy rate than Wang et al.

1.3 Rationale of the Study

The number of social media users is increasing constantly. They share about their opinions and emotions about anything and everything on these platforms. Different types of online business are growing based on these users. Thus, learning about their mood is essential. The purpose of this research effort is to explore and compare different types of deep learning models for tweet emotion recognition. The results of this study can enhance sentiment analysis and offer useful information for social media analytics tools.

1.4 Research Question

- How do Bidirectional GRU, CNN, and LSTM models compare in predicting tweet emotions?
- What are the strengths and considerations associated with each deep learning model for tweet emotion recognition?
- How can evaluation metrics provide insights into the effectiveness of deep learning models for tweet emotion recognition?

- How can accurate tweet emotion recognition benefit social media marketing and customer feedback analysis?
- What are the potential applications of tweet emotion recognition in political campaigns and sentiment analysis?
- How difficult is the task?

1.5 Expected output

The goal is to compare different models that we have create to analyze the data and see how well each model works in detecting the true emotions. We want to find the best model and discover the strengths and weaknesses of each. By evaluating these models we can use them in real life situation such as social media marketing and political campaigns. Moreover, we want to look at the effectiveness of deep learning models for tweet emotion recognition in sentiment analysis.

1.6 Project Management and Finance

Google Colab, a cloud-based tool that encouraged cooperation and sped up the creation of tweet emotion detection models, was used to manage the project effectively. Google Colab does away with the requirement for pricey equipment setup because to its high-performance computing capabilities.

The project uses the Hugging Face Emotion dataset, which is open source and helps reduce data collecting expenses, was financially advantageous. Cost-effectiveness was further enhanced by Google Colab's free use of computing resources and the accessibility of open-source libraries.

1.7 Report Layout

- Background study
- Research Methodology
- Experimental Result and Discussion
- Summary, Conclusion and Future Analysis
- Reference

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

It is crucial to comprehend the algorithms and models utilized in the context of tweet emotion identification. The Bidirectional Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) are the three models that this research focuses on employing.

RNN can be classified as either bidirectional GRU or unidirectional GRU. Convolutional layers are used by CNN to extract pertinent features, whereas LSTM employs them to keep long-term relationships in sequences. In order to make the best choice we have to be aware of the advantages and limitations of each type. We can learn more about their functionality and prospective uses through examination and analysis.

This enhances the study of tweet emotion identification and offers recommendations for sentiment analysis and social media analytics by investigating these deep learning models. Several researchers who used several models in their study are mentioned in the section.

2.2 Related Works

Deep learning techniques have been used in a number of researches on tweet emotion identification. M. Alfa Riza. (2021) classified twitter emotions LSTM with an accuracy of 73% [nth]. To identify the temporal relationships and significant elements in tweets, they used a bidirectional LSTM and an Fast Text method.

Nested LSTM and LSTM were used for emotion identification in text in a different research by Haryadi (2019). Their model produced findings that were competitive and had an accuracy of 99.167% [nth]. The LSTM architecture made it possible to accurately anticipate emotions by extracting useful features from text contents.

GRU Neural Network was also investigated by Liu. (2020) for text emotion recognition. The accuracy of their GRU-based model was 84% [nth]. They used the GRU neural network with attention mechanism perfectly, which helped with accurate emotion categorization.

Santosh Kumar Bharti (2022) carried out another important study in the area of text emotion detection, in which they combined word embeddings and deep neural networks for text emotion categorization [nth]. They used Convolutional neural network (CNN) and Bi-GRU were exploited as deep learning technique with an accuracy of 80.11%.

These papers show how well deep learning models, such as CNN, LSTM, and bidirectional GRU, perform in twitter emotion identification tasks. Building on these earlier studies, our effort seeks to examine and contrast the performance of these models on the Hugging Face Emotion dataset, offering insightful information about their advantages and limitations.

2.3 Comparative Analysis and Summary

Here the results and findings of GRU, CNN, and LSTM models are presented in this section. The models we have created has performed better than many other works on this field, with the accuracy rates of 91.95% for CNN and 91.85% for Bidirectional GRU on text based emotion recognition. This shows the effectiveness of our method in correctly distinguishing between emotions. We used Google Colab's free GPU which resulted in e quicker training of the datasets and also enhanced the performance of the models. These research effort shows how quickly deep learning approaches can work for emotion recognition and also can be applicable in real life scenarios.

2.4 Scope of the Problem

The issue of tweet emotion detection is the main subject of this research effort. Our goal is to investigate and evaluate how well various deep learning models predict the emotions conveyed in tweets. We want to give insights into the models' capabilities and constraints in comprehending and categorizing tweet emotions by assessing and contrasting the models' efficacy.. The study intends to further sentiment analysis and social media analytics by concentrating primarily on the usage of these models in the context of social media.

2.5 Challenges

In order to address the challenge of varying tweet lengths, we used padding and truncating techniques into the datasets because of the uneven lengths of the different data on dataset. By padding we added zeros to the shorter tweets to make them equal in length, while truncating involves cutting off excess words from longer tweets. Padding and truncating are preprocessing techniques. We used these so that the model can process all the data effectively resulting in easier training of the model and better accuracy.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The primary focus of the research effort was to figure out how deep learning models work on text based emotion detection. We used three deep learning models such as LSTM, Bidirectional GRU and CNN. By identifying emotion expressed in the tweets presented in the dataset valuable insights can be obtained for different applications such as sentiment analysis, customer feedback and more.

- Operating system: Windows.
- Programming language: Python 3.
- Environment: Google Colaboratory
- Libraries used: TensorFlow, Keras, Matplotlib, Pandas, Seaborn etc.
- GPU: Tesla T4 GPU, based on the Turing architecture, for accelerated deep learning model inference.

3.2 Data Collection Procedure

The dataset was virtually ready for implementation when it was taken from Hugging Face. While unsplitted the dataset consisted of 20000 tweets as data. But then the dataset was splitted into 3 datasets, among them there was 16000 training data in the train-dataset, 2000 data in the validation dataset and another 2000 data in the test-dataset. The dataset had two columns namely text and labels. In labels column there are six labels representing six emotions. The emotions are 0 = Sadness, 1=Joy, 2=Love, 3 =Anger, 4 = Fear, 5 = Surprise. Here the ratio among training, validation and testing dataset were 8:1:1 which lead to a better accuracy of the model. By utilizing Hugging Face Emotion dataset we ensured the availability of high quality labeled data that represented real world emotions expressed in tweets.

text (string)	label (class label)
"im feeling rather rotten so im not very ambitious right now"	0 (sadness)
"im updating my blog because i feel shitty"	0 (sadness)
"i never make her separate from me because i don t ever want her to feel like i m ashamed with her"	0 (sadness)
"i left with my bouquet of red and yellow tulips under my arm feeling slightly more optimistic than when i arrived"	1 (joy)
"i was feeling a little vain when i did this one"	0 (sadness)
"i cant walk into a shop anywhere where i do not feel uncomfortable"	4 (fear)
"i felt anger when at the end of a telephone call"	3 (anger)
"i explain why i clung to a relationship with a boy who was in many ways immature and uncommitted despite the excitement i should have been feeling for getting accepted into the masters program at the university of virginia"	1 (joy)
"i like to have the same breathless feeling as a reader eager to see what will happen next"	1 (joy)
"i jest i feel grumpy tired and pre menstrual which i probably am but then again its only been a week and im about as fit as a walrus on vacation for the summer"	3 (anger)
"i don t feel particularly agitated"	4 (fear)

Figure 1- Training dataset of Emotion

text (string)	label (class label)
"i didnt feel humiliated"	0 (sadness)
"i can go from feeling so hopeless to so damned hopeful just from being around someone who cares and is awake"	0 (sadness)
"im grabbing a minute to post i feel greedy wrong"	3 (anger)
"i am ever feeling nostalgic about the fireplace i will know that it is still on the property"	2 (love)
"i am feeling grouchy"	3 (anger)
"ive been feeling a little burdened lately wasnt sure why that was"	0 (sadness)
"ive been taking or milligrams or times recommended amount and ive fallen asleep a lot faster but i also feel like so funny"	5 (surprise)
"i feel as confused about life as a teenager or as jaded as a year old man"	4 (fear)
"i have been with petronas for years i feel that petronas has performed well and made a huge profit"	1 (joy)
"i feel romantic too"	2 (love)
"i feel like i have to make the suffering i m seeing mean something"	0 (sadness)
"i do feel that running is a divine experience and that i can expect to have some type of spiritual encounter"	1 (joy)

Figure 2- Test dataset of Emotion

3.2.1 Data Preparation and Preprocessing

Since we are going to build deep learning models for the research effort, we prepared the dataset accordingly. For this reason several preprocessing steps were performed. At first the tweets were tokenized using the Tokenizer class from Keras. The tokenizing process converted the tweets into sequence of numerical tokens. Each of these tokens represents a unique word in the vocabulary. The tokenizer was configured to only consider the first 10000 most used words and the unknown words were marked as UNK(unknown).

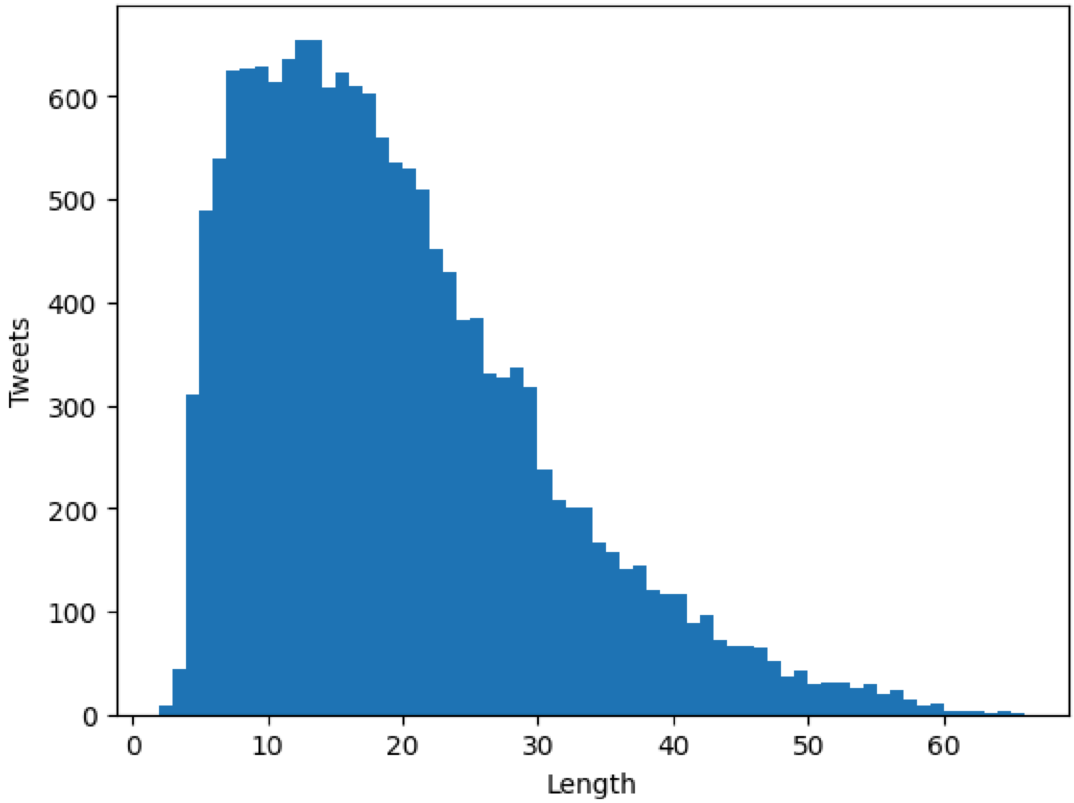


Figure 3- Distribution of Tweet Lengths

From the plot it is evident that most of the tweets have a length between 10-20 tokens and very few tweets have a length over 50 tokens and the number lessens after 60. Later, the tokenized sequences were padded and truncated. That ensured a consistent length of all input sequences. The sequences were padded to a maximum length of 50 tokens and any exceeding tokens were truncated from end of the sequences without this step long sequences could have caused memory or performance issues during training. Padding was performed by adding zeros to the end of the sequences. To prepare the twitter data for training the deep learning models, the preprocessing processes of tokenization, padding, and truncating were essential. The input sequences were uniformly lengthened as a result of these stages, allowing for efficient training and prediction.

3.3 Statistical Analysis

We created three models that we presumed will work best for emotion detection in this context. The models were created using LSTM, Bidirectional GRU and CNN. We conduct a thorough statistical study to assess the effectiveness of the trained models. To determine how well they are able to forecast tweet emotions, we use a variety of assessment indicators. We found out about the accuracy, confusion matrix, F-1 score, Precision, Recall and ROC-AUC score for each models. To assess how well the trained models are performing, we conducted a thorough statistical study.

3.4 Proposed Methodology

Flow chart:

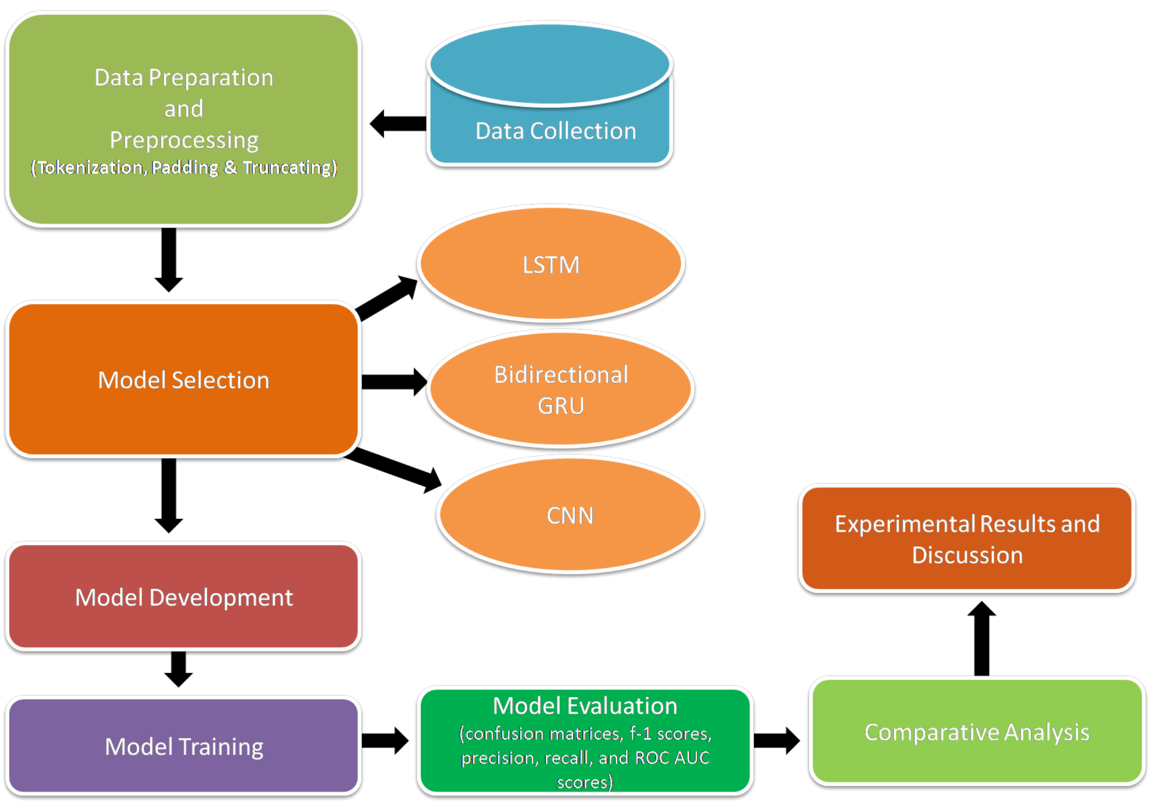


Figure 4: Methodology of Tweet Emotion Recognition

For research methodology we have gone through a few steps:

- Data Collection: We get a dataset of labeled tweets commented on with particular feeling categories. The dataset serves as the establishment for preparing and assessing the profound learning models for tweet feeling acknowledgment.

- Data Preprocessing: We preprocess the collected information by performing errands such as content normalization, tokenization, padding and truncating. This step guarantees that the information is in a reasonable arrange for assist investigation and show preparing.

- Model Selection: We select three conspicuous profound learning models for tweet feeling acknowledgment: Bidirectional Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). These models have performed well.

- Model Development: We created and arranged the chosen models utilizing fitting libraries and systems such as TensorFlow and Keras. Based on their particular architectures, the models are built with specified layers, activation functions, and input forms.

- Model Training: We prepare the models utilizing the preprocessed dataset. Amid preparing, the models learn to recognize designs and connections within the tweet information to precisely foresee the comparing feeling categories.

- Model Evaluation: We assess the prepared models utilizing different assessment measurements, accuracy, recall, ROC AUC scores, f-1 scores, confusion matrix etc. These measurements give bits of knowledge into the models' execution in terms of precision, exactness, and in general classification capacity.

- Comparative Analysis: We conduct a comparative investigation of the three models to survey their qualities, contemplations, and execution in foreseeing tweet feelings. This examination makes a difference us distinguish the demonstrate that yields the most noteworthy precision and gives profitable experiences for selecting an suitable show for tweet feeling acknowledgment assignments.

- Experimental Results and Discussion: We display the exploratory comes about, counting the precision rates accomplished by each demonstrate. We talk about the discoveries, compare them with existing investigate, and analyze the suggestions and confinements of our approach.

The proposed methodology provides a systematic framework for conducting the project, ensuring that each step is carefully executed to achieve accurate and reliable results in tweet emotion recognition.

3.5 Implementation Requirements

The dataset that we have used in this project was already splitted into three parts such as Train, Validation and Test. We used 80% of the data to train the deep learning models, which gave us a large number of labeled tweets to study correlations and trends in. 20% of the data was left over, and it was evenly divided into two parts, one for testing and one for validation. As a result, we were able to evaluate how well the trained models performed on new data and determine how well they predicted tweet emotionIt requires enough amounts of RAM and storage to handle the dataset and model training procedure and a steady internet connection to access the necessary databases, resources, and libraries throughout installation

A sufficient amount of time and computing power to build the models and conduct analyses was also needed. These implementation requirements, which include the training and assessment of the deep learning models for tweet emotion identification, are important to guarantee the project's effective execution. We also evaluated the models through various procedures such as confusion matrix, precision, recall, F-1 score, ROC AUC score. The result of these procedures shows that we were successful in terms of model selection, because all of the models performed well for emotion recognition.

CHAPTER 4

Experimental results and discussion

4.1 Experimental Setup

We used python on Google colaboratory for the experiment. Since Google Colab is a suitable environment to execute deep learning models and libraries such as tensorflow and keras. The project was executed in a machine where Windows was used as the operating system. In google colab we used the Tesla T4 GPU which enhanced the model training and inference. The specific details and methods will be further elaborated in the following discussion. We also used 3 suited models for implementing deep learning into the dataset and get the best result.

4.2 Experimental Results & Analysis

At first we labeled the emotions in case into numeric values so we can train the model at ease. Three models were created using three different approaches to see which one works best in order to recognize emotion from text data.

Model-1: We created the model using LSTM architecture. A brief description of the model is given below:

Embedding layer: Converts input that has been integer-encoded into dense vectors of a fixed size.

Bidirectional LSTM layer 1: Processes the input sequence backwards as well as forwards, gathering contextual information.

Bidirectional LSTM layer 2: Increases the model's capacity to recognize long-term relationships in the input sequence through the use of a bidirectional LSTM layer.

Dense layer: Uses a softmax activation function to carry out the final classification. The model's input shape is (batch_size, input_length), where batch_size indicates how many samples there are and input_length indicates how long each input sequence is.

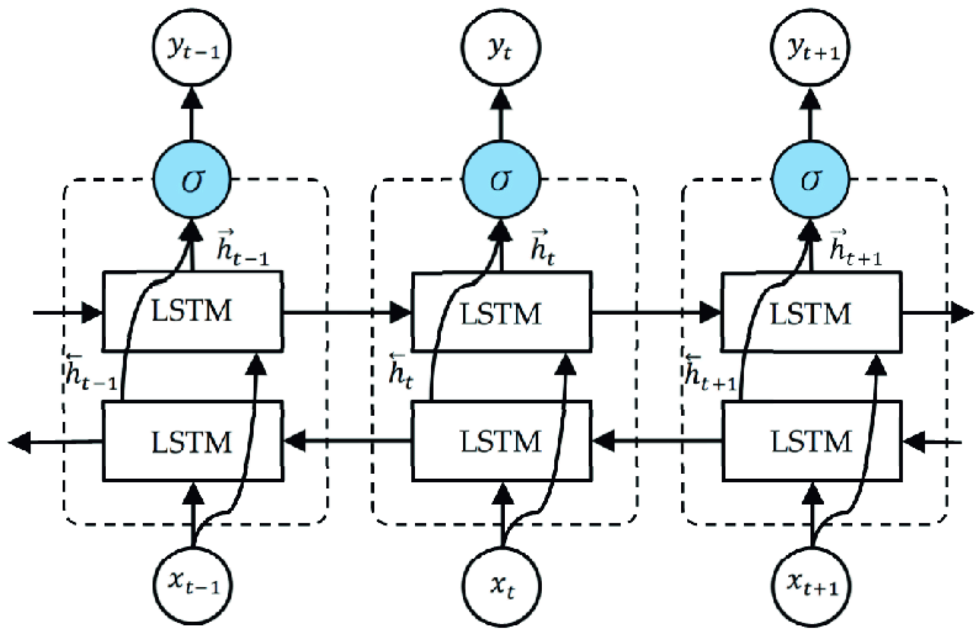


Figure- Architecture of Bidirectional LSTM

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 50, 16)            160000

 bidirectional (Bidirectiona  (None, 50, 40)           5920
 l)

 bidirectional_1 (Bidirectio  (None, 40)               9760
 nal)

 dense (Dense)               (None, 6)                 246

=================================================================
Total params: 175,926
Trainable params: 175,926
Non-trainable params: 0
```

Figure- Summery of Created Model

This model is trained using the following configuration:

Loss Function: Loss function for multi-class classification problems is sparse categorical cross-entropy.

Optimizer: Adam is an optimizer well-known for its quick convergence and flexible learning rate.

Metrics for evaluation: Accuracy, which gauges how well the algorithm performs in detecting tweet emotions.

The model uses backpropagation to change the parameters in order to minimize the loss function during training. To improve the model's predictions, the optimizer modifies the weights. The accuracy, which represents the percentage of properly categorized samples, is used to assess the training progress.

Afte training the model we evaluated it and plotted the Accuracy vs Epochs graph as well as Loss vs Epochs graph. From there we can see how well our model has performed. As the number of epochs increases the accuracy increases as well. However after a certain number of epochs the accuracy remains the same meaning that the model has peaked its accuracy. Same is shown for Loss vs Epochs too. As the number of epochs increases the Loss decreases and after a certain number of epochs it remains constant. The graph is shown later on for a better understanding. It also shows the lines of training and validation data. Since the lines are close to each other it can be said that the model is performing well.

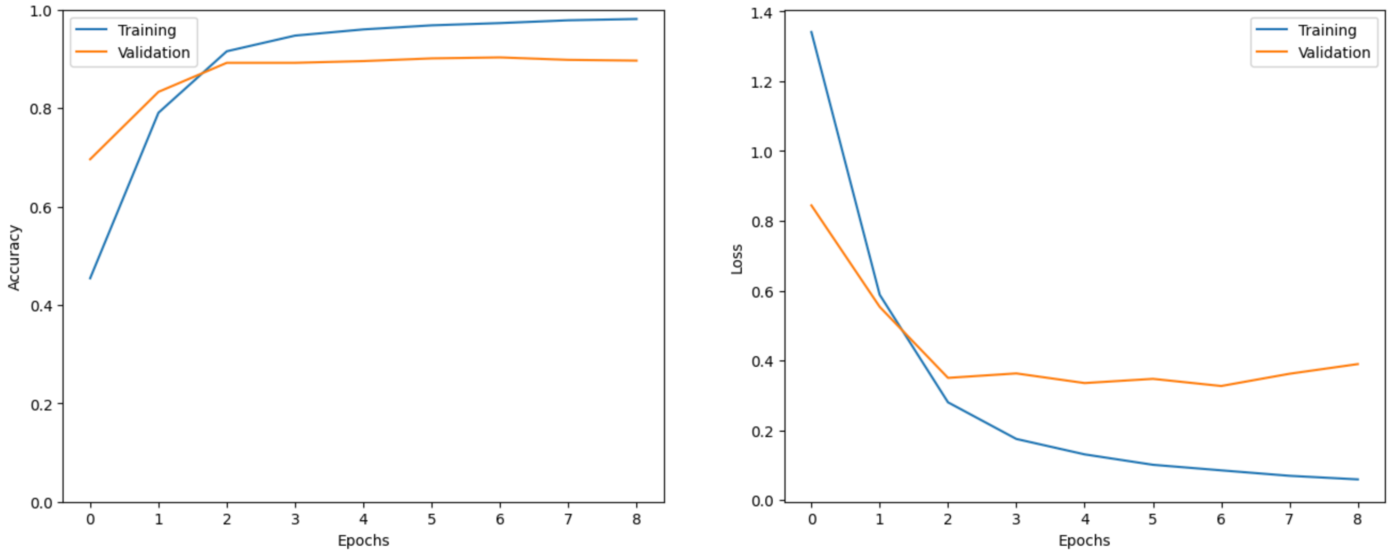


Figure – History of Training and Validation Data

For evaluating the model we found out the accuracy. The accuracy of LSTM model was 88.50%. We later on plotted the confusion matrix for the model

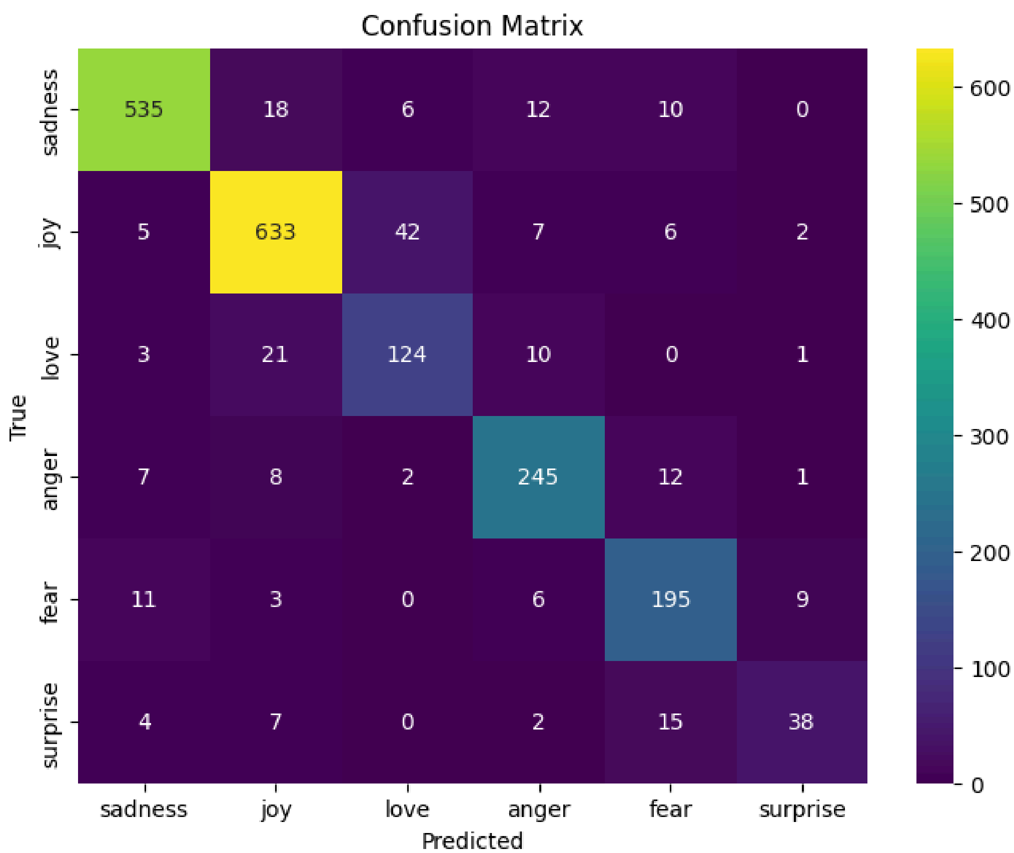


Figure- Confusion Matrix for LSTM

From the confusion matrix we can say that the model predicted the emotions Sadness and Joy with most accuracy. However, it made the most mistakes for Joy and Love.

We also found out about precision, recall, f-1 score and ROC AUC scores of the model which will be discussed further on. We plotted evaluation matrices. And this shows the same result as before.

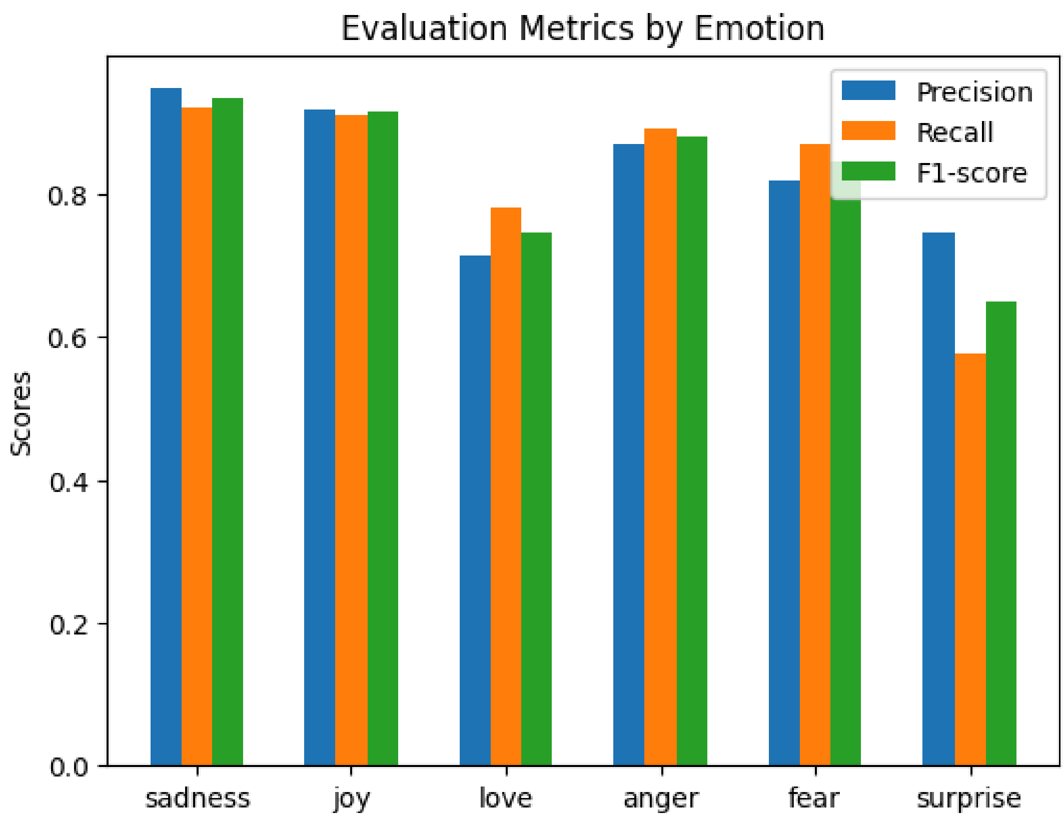


Figure – Evaluation Metrics for LSTM model

Model-2:

The GRU (Gated Recurrent Unit) architecture is used to build Model-2. The input is first transformed into dense vectors of a specified size by an embedding layer. Two Bidirectional GRU layers make up the following levels. These layers process the input sequence both forward and backward while recording contextual data. For the final classification, a thick layer with a softmax activation function is employed. The GRU model architecture is similar to the LSTM model.

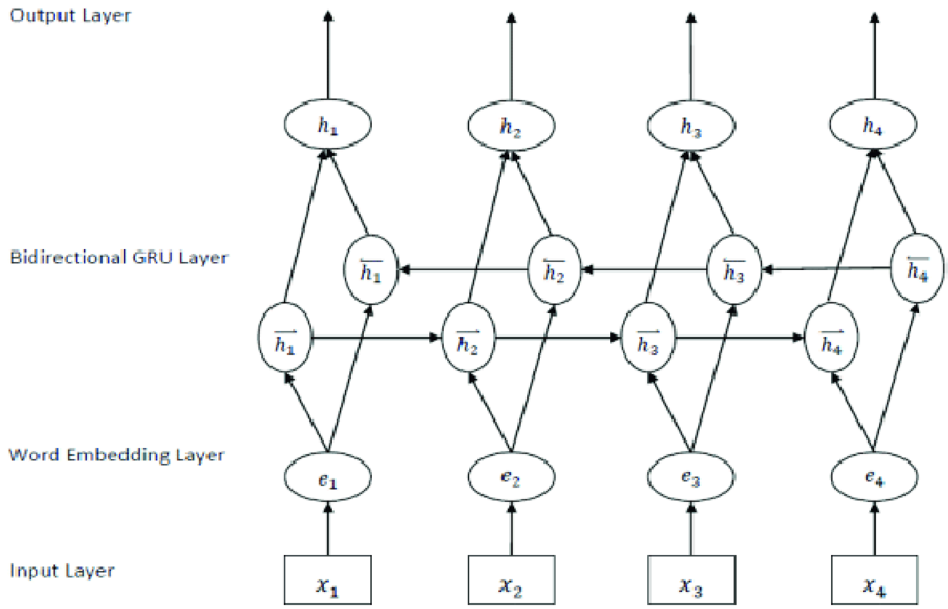


Figure- Architecture of Bidirectional GRU

The model is built using an Adam optimizer, a sparse categorical cross-entropy loss function, and accuracy as the evaluation measure. The model summary includes details on the total number of parameters in the model as well as the number of parameters in each layer.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 50, 32)            320000

 bidirectional (Bidirectiona  (None, 50, 128)          37632
 l)

 bidirectional_1 (Bidirectio  (None, 128)              74496
 nal)

 dense (Dense)               (None, 6)                 774

=================================================================
Total params: 432,902
Trainable params: 432,902
Non-trainable params: 0
```

Figure- Summery of model-2(Bidirectional GRU)

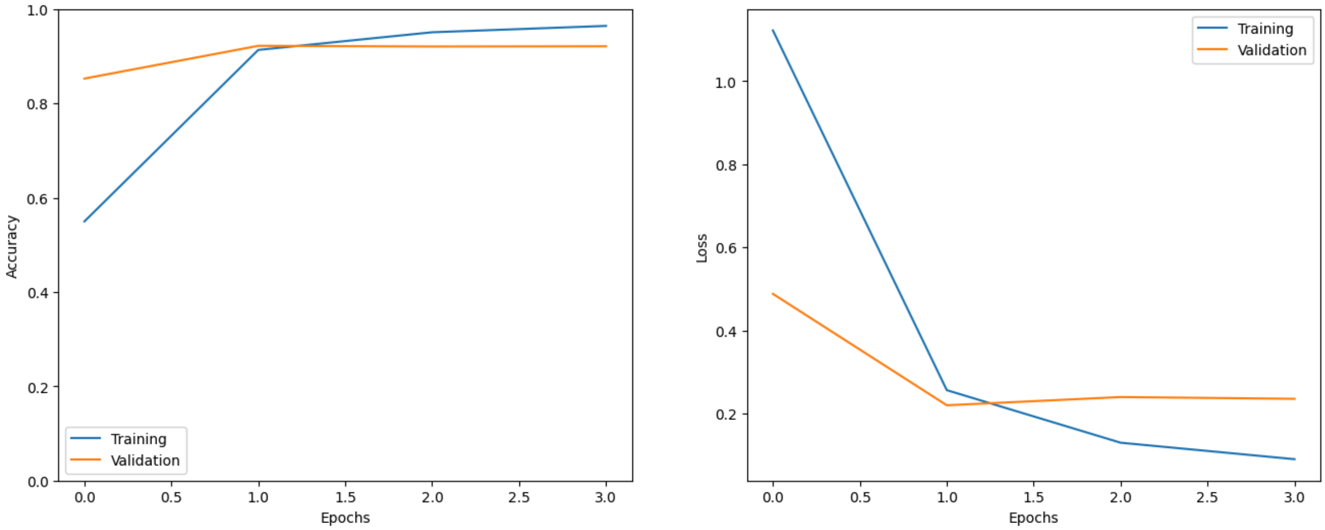


Figure – History of Training and Validation Data for model-2

To know how well the model has worked we also visualized the training history of the model. We generated two graphs. We can see that Training and validation curve is very close to one another. Even closer than the LSTM model. So we can see that this model has performed even better than LSTM. We later on found out the accuracy of the model which was 91.85%.

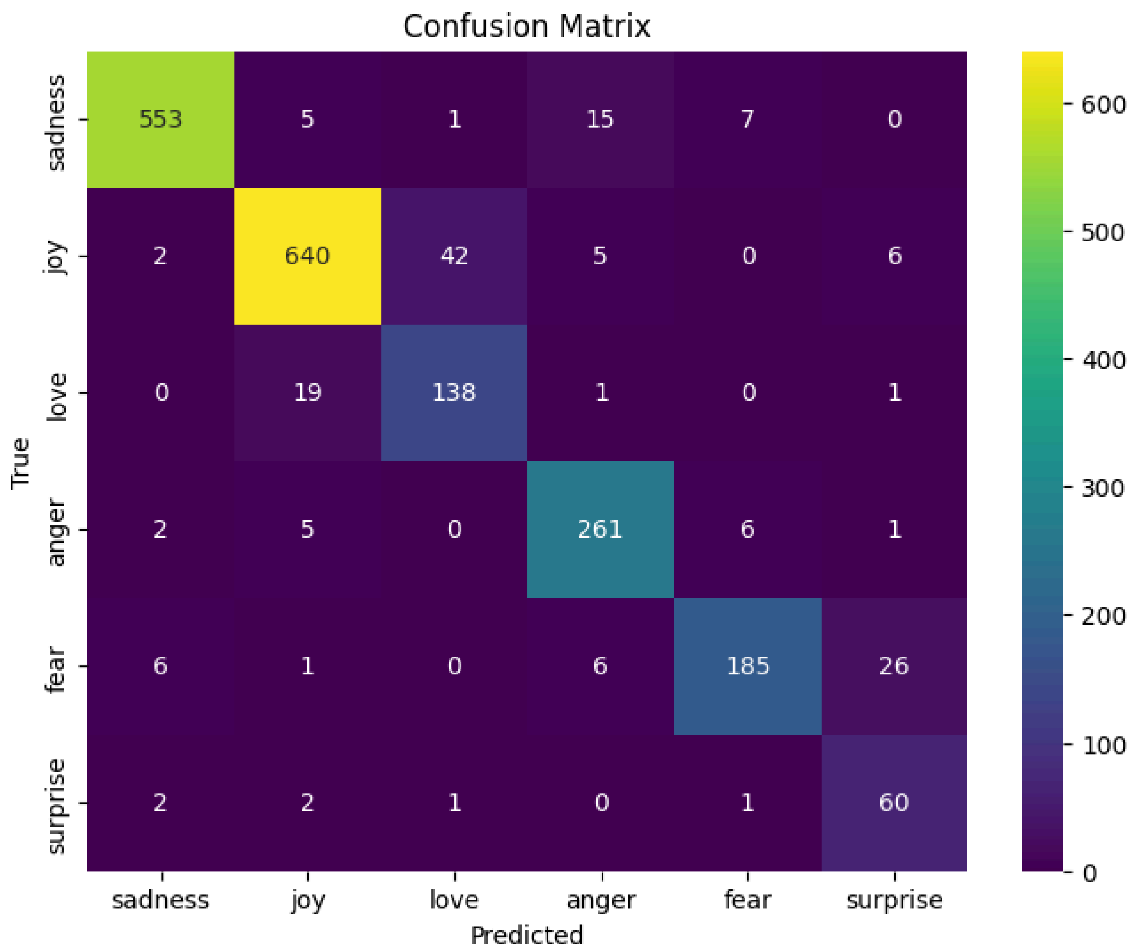


Figure- Confusion matrix for Bidirectional GRU model.

From the confusion matrix we can say that the model predicted the emotions Sadness and Joy with most accuracy. However, it made the most mistakes for Joy and Love, same as model-1.

We also used some metrics for evaluating the model. They are visualized for better understanding.

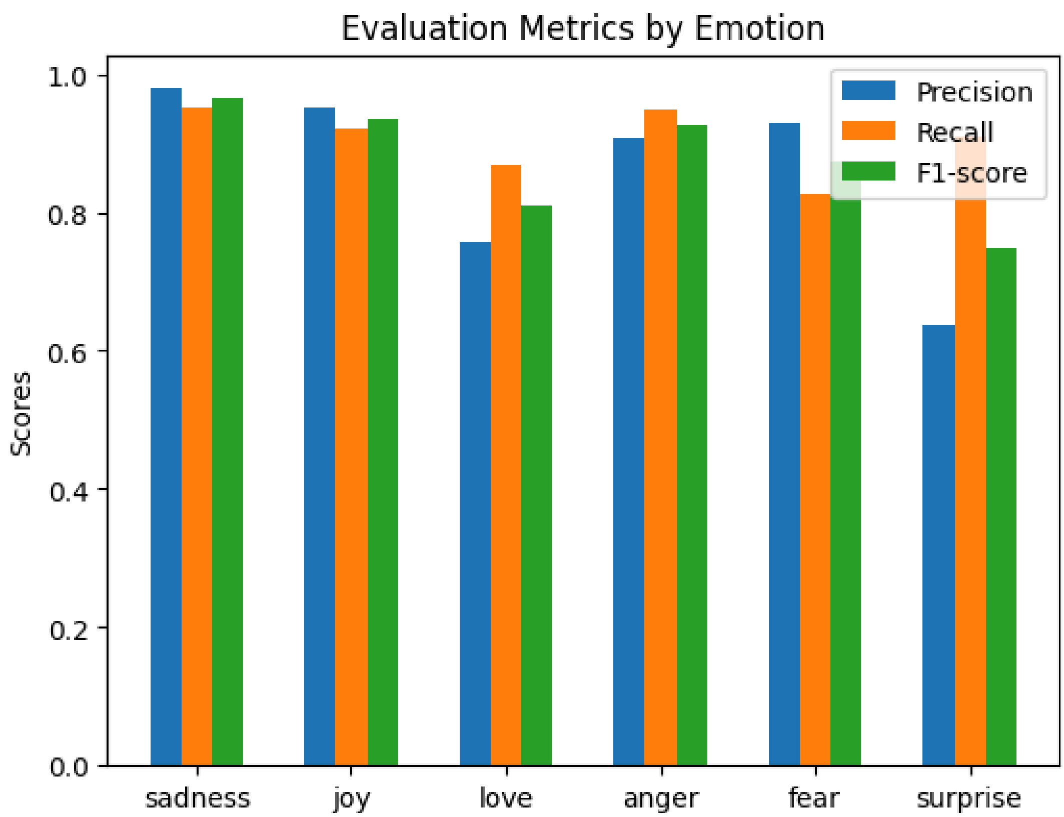


Figure- Evaluation metrics for model-2

Model-3:

Model-3 uses a CNN (Convolutional Neural Network) architecture to recognise tweet emotion, unlike model-1 and model-2 where LSTM and GRU architecture was used.. It begins with an embedding layer that turns the input into dense vectors, then moves on to a 1D convolutional layer that extracts features. The most important characteristics are captured using max pooling, and classification is performed using a dense layer with softmax activation.

This makes use of the special characteristics of CNNs to capture regional patterns and spatial linkages in the input data, as opposed to Models-1 and Models-2.. Due to its unique architectural design, the Model-3 may be able to better capture specific aspects and enhance tweet emotion recognition as a whole. While evaluating the model we can understand that even better. This model has a slightly better accuracy than other models. The accuracy of model- 3 is 91.95%.

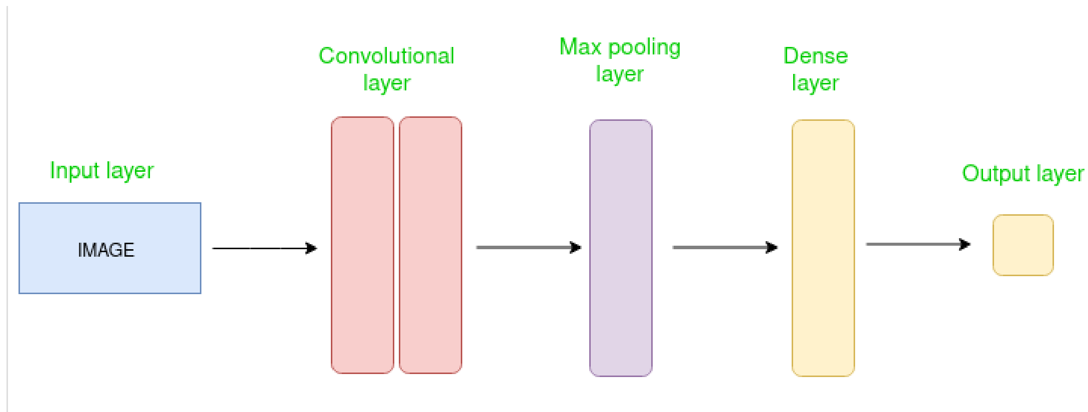


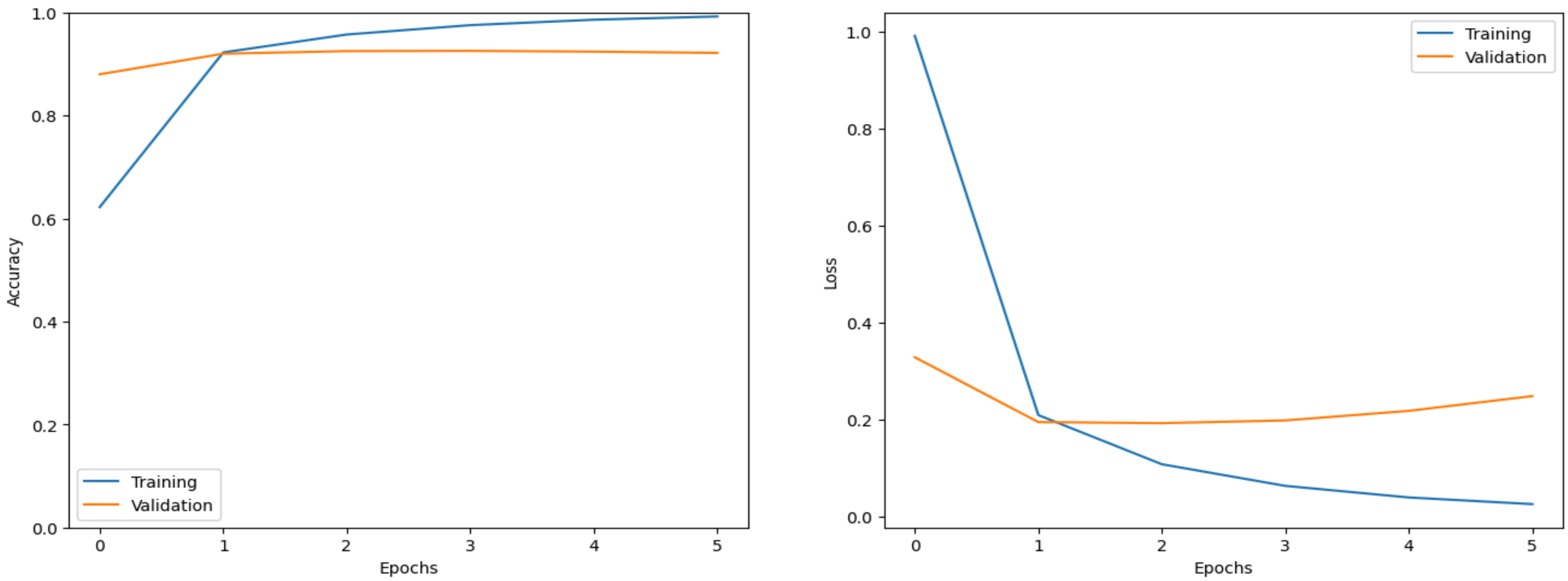
Figure- Architecture of CNN

It comprises of an embedding layer that creates dense vectors from the input data. Then, to extract features from the input, a convolutional layer with 128 filters and a kernel size of 5 is employed. The most crucial characteristics are then captured by passing the output via a global max pooling layer. There are two thick levels after that, each with 64 and 6 units. The sparse categorical cross-entropy loss function and Adam optimizer are used to train the model. The structure and quantity of parameters in each layer are displayed in the Model-3 summary.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 50, 32)            320000

 conv1d (Conv1D)             (None, 46, 128)           20608

 global_max_pooling1d (Globa  (None, 128)              0
 lMaxPooling1D)

 dense (Dense)               (None, 64)                8256

 dense_1 (Dense)             (None, 6)                 390

=================================================================
Total params: 349,254
Trainable params: 349,254
Non-trainable params: 0
```

Figure- Summery of model-3(CNN)

We also plotted the graph to see how well the model is performing. The training and validation curve is close to each other and the accuracy increases with each epochs. It shows how effectively the model is learning. It shows that the model is not suffering high



bias or high variance issues.

Figure – History of Training and Validation Data for model-2

We also found the confusion matrix we can say that the model predicted the emotions Sadness and Joy with most accuracy. However, it made the most mistakes for Joy and Love, same as model-1 and model-2. As a result we can say that all three models are being confused to differentiate between love and joy. We also used some metrics for evaluating the model. They are visualized for better understanding.

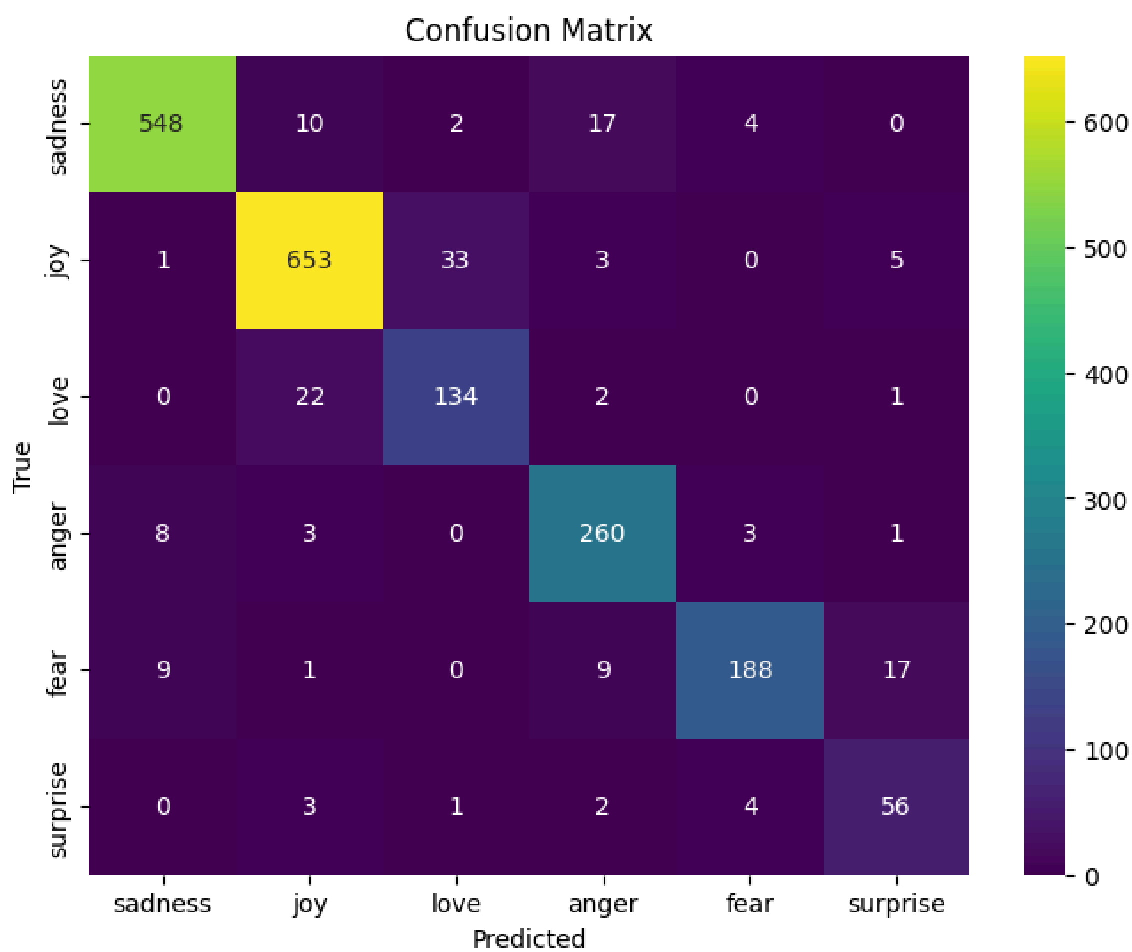


Figure- Confusion matrix for CNN model

We also used some metrics for evaluating the model. They are visualized for better understanding.

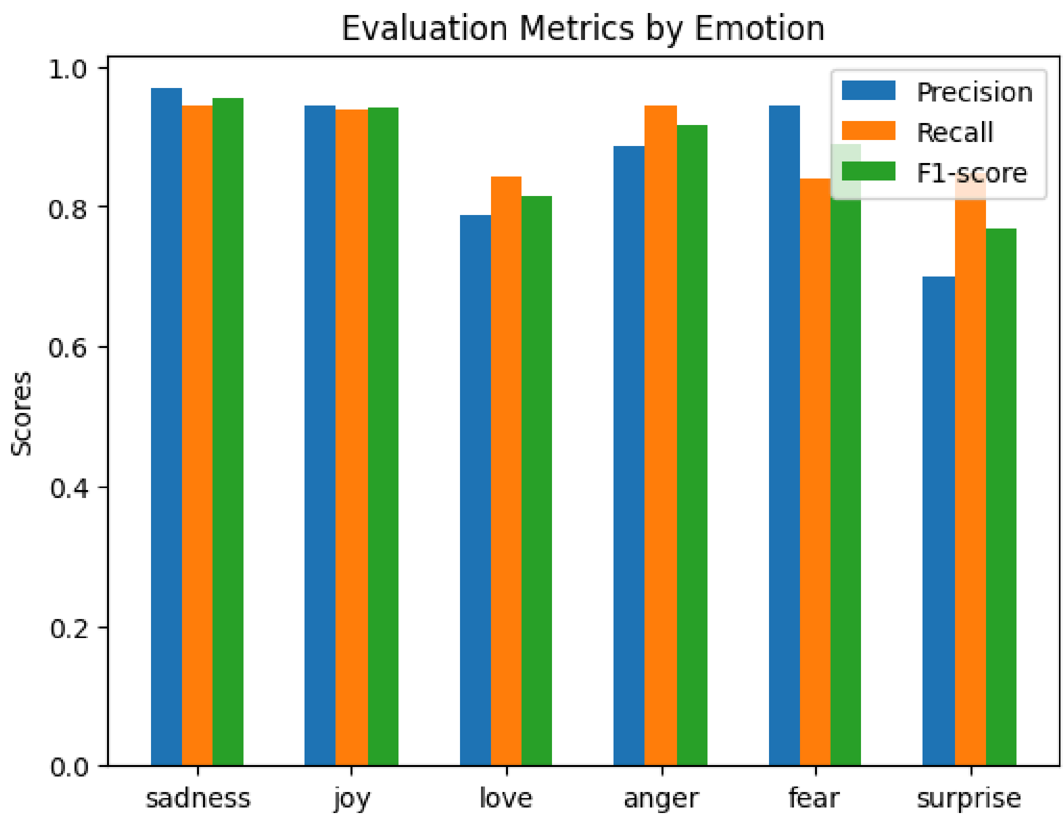


Figure- Evaluation metrics for model-3

4.3 Discussion

This In this phase, we will make the judicial structure of our suggested model clear. We have taken into account the F-1 score, recall, accuracy, and precision.

4.3.1 Accuracy

It speaks about the proportion of testing data predictions that were correct. Accuracy of model-1 was 88.50%, for model-2 it was 91.85% and for model-3 it was 91.95%. This shows that using CNN architecture resulted in a better accuracy rather than LSTM and GRU.

$$Accuracy = \frac{TruePositive+TrueNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative}$$

4.3.2 Precision

It shows the percentage of positively expected observations that really occurred. The precision values for the three models show how accurately each model predicted the correct outcomes. For Model 1, a precision score of 0.886 indicates that 88.6% of the projected positive instances are in fact genuine positive cases. Model 2 has a precision of 0.925, which means that 92.5% of the projected positives are really true positives. A high accuracy of 0.923 for Model 3 indicates that 92.3% of its positives are accurate.

To concluse, these precision values show that the models are capable of producing precise positive predictions, with Models 2 and 3 obtaining a little greater accuracy than

Model 1. Higher accuracy values represent more accurate affirmative case identification, which is essential for tweet emotion recognition.

$$Precision\ =\ \frac{TruePositive}{TruePositive+FalsePositive}$$

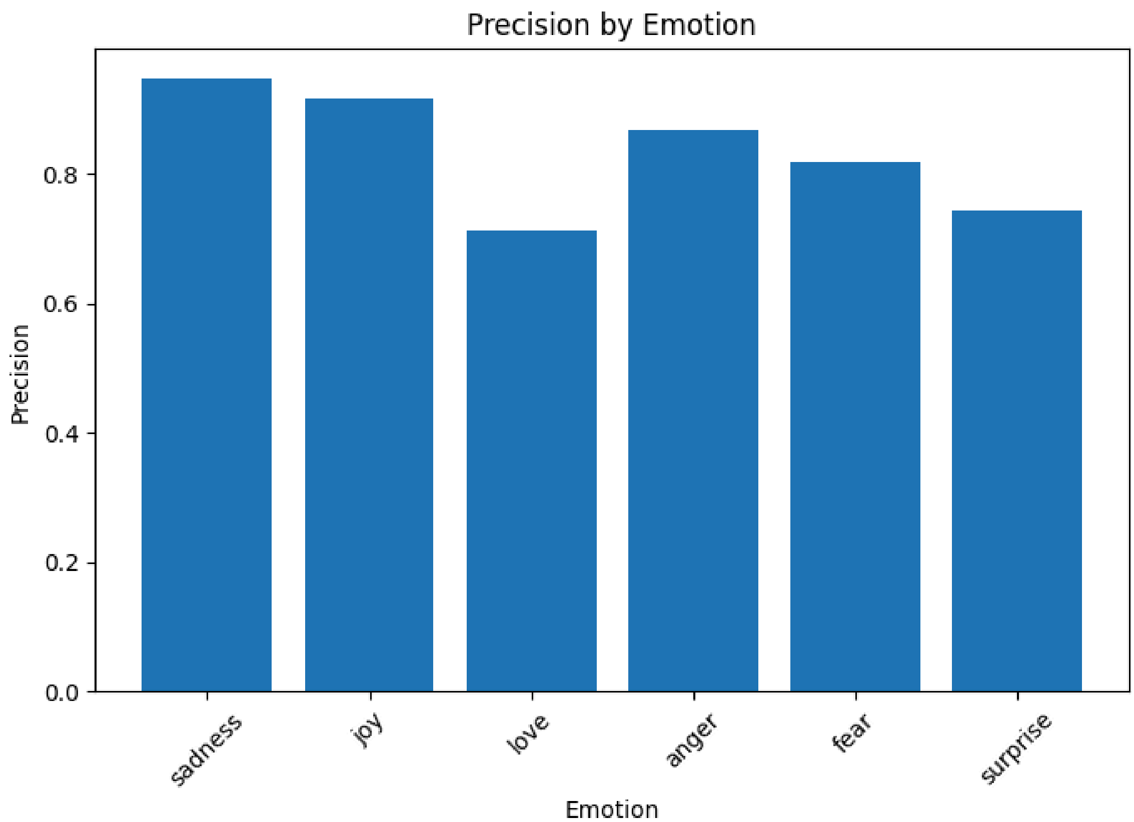


Figure-Precision By emotion for Model-1

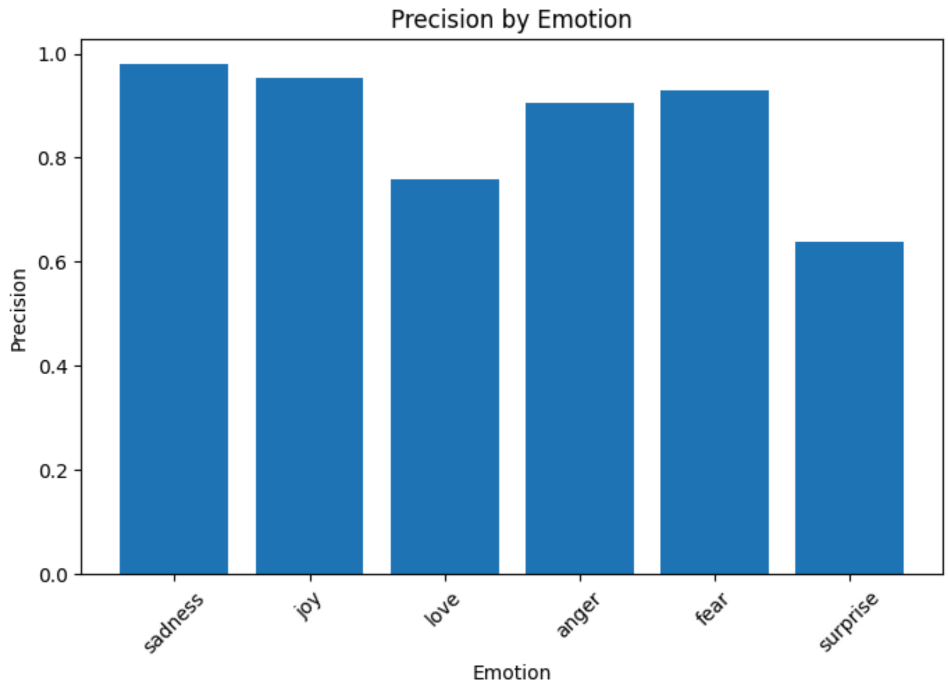


Figure-Precision By emotion for Model-2

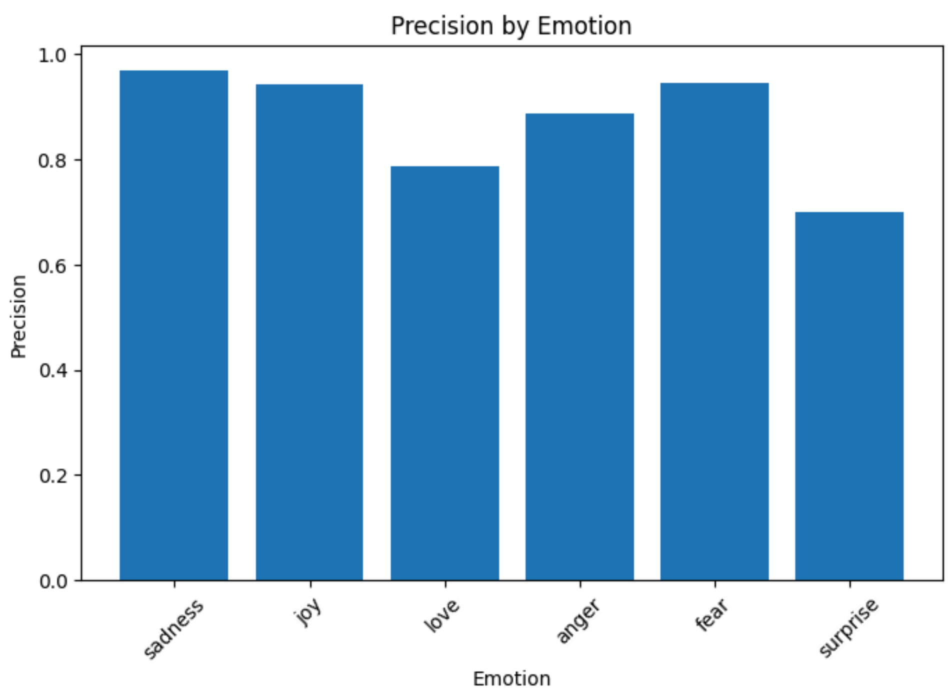


Figure-Precision By emotion for Model-3

4.3.3 Recall

The three models' recall values show how well each model was able to recognise positive instances among all of the real positive examples in the dataset. Model 1's recall rating of 0.885 indicates that it can recognise 88.5% of the real positive cases. Model 2 can properly identify 91.85% of the positive instances according to its recall score of 0.9185. Similar to Model 2, Model 3 has a recall of 0.9195, which indicates that it can correctly identify 91.95% of the positive cases.

In conclusion, these recall numbers show how well the models capture a large percentage of real positive cases. Models 2 and 3 had marginally higher recall than Model 1, demonstrating an improvement in their capacity to accurately identify positive cases in tweet emotion recognition.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

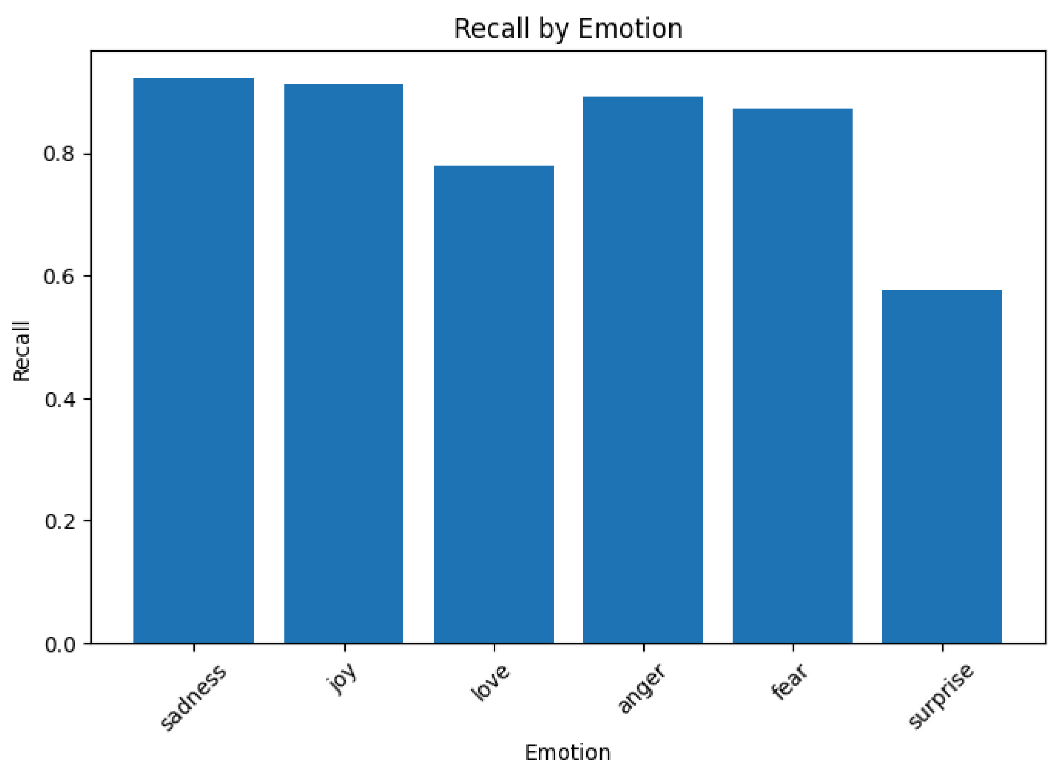


Figure- Recall by emotion for model-1

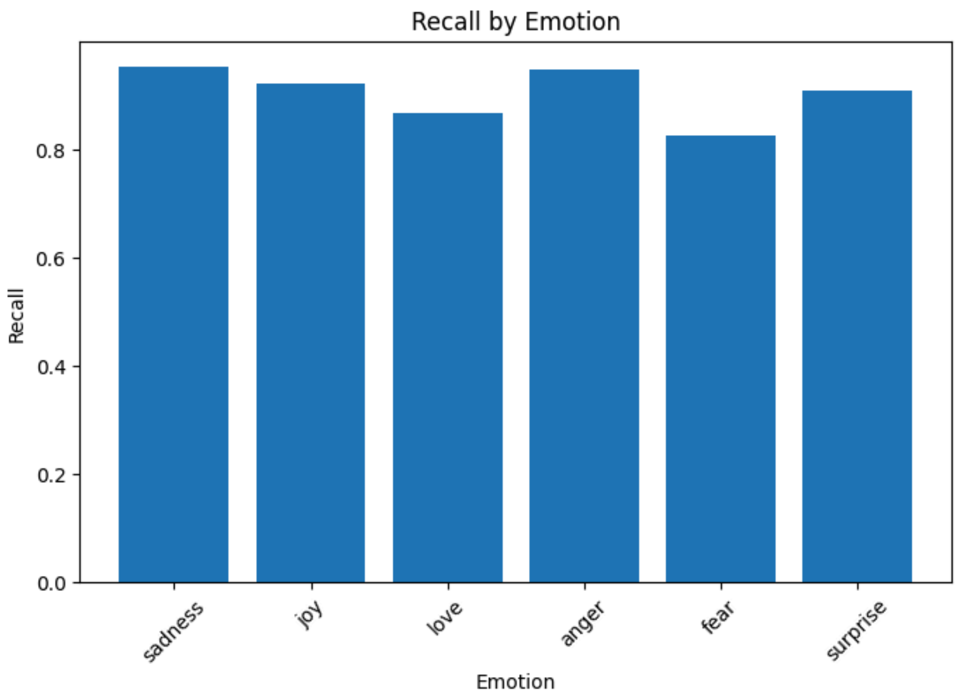


Figure- Recall by emotion for model-2

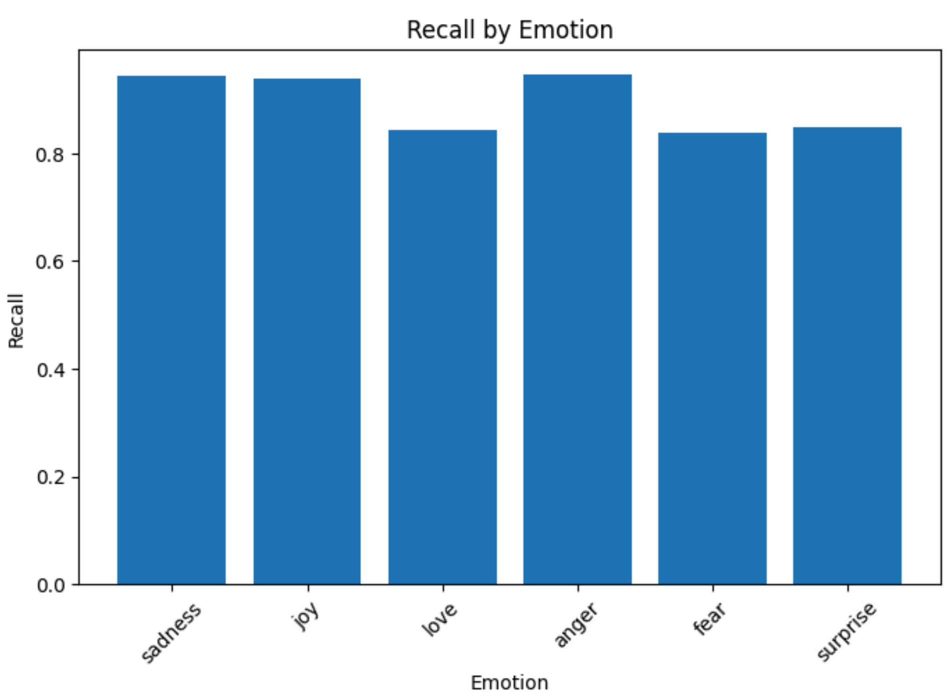


Figure- Recall by emotion for model-3

4.3.4 F-1 Score

A model's total performance is gauged by the F1 score, which considers both precision and recall. The F1 score for Model 1 is 0.885, suggesting that accuracy and recall are well-balanced. The F1 score for Model 2 is 0.920, suggesting a somewhat improved overall performance. Similar to Model 2, Model 3 likewise has an F1 score of 0.920, indicating a similar performance. Simply put, these F1 values imply that Models 2 and 3 have outperformed Model 1 in terms of overall accuracy for recognising tweet moods. The F1 score takes into account the model's accuracy (ability to categorise positive instances accurately) and recall (ability to identify all positive cases). Therefore, Models 2 and 3 demonstrate better performance in accurately identifying tweet emotions compared to Model 1.

$$F - 1\,Score = 2 * \frac{Recall * Precision}{Recall + Precision}$$

4.3.5 AUC-ROC Curve:

The effectiveness of a model in binary classification tasks is graphically represented by the ROC AUC (Receiver Operating Characteristic Area Under the Curve) curve. However, the ROC AUC curve might not be immediately applicable in this situation because the models in question are conducting multi-class categorization.

We may instead concentrate on the accuracy statistic, which shows the proportion of accurately identified samples. We can see from the accuracy figures for each model that Model 2 has a better accuracy than Models 1 and 3. This shows that Model 2 does a better job of properly categorizing tweet emotions.

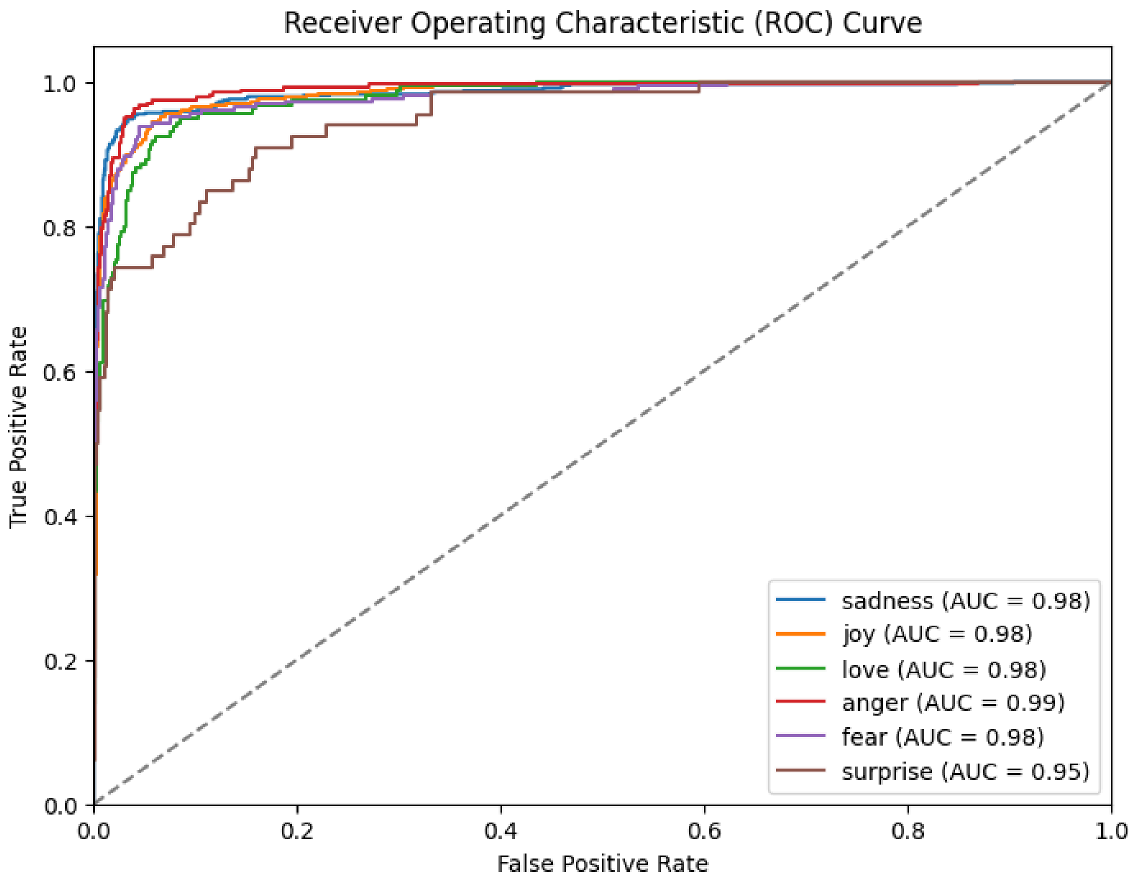


Figure- AUC-ROC Curve Analysis

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Deep learning methods for tweet emotion identification have a significant social impact. This technology provides invaluable information for organisations, enabling them to comprehend client mood and adjust their goods and services appropriately by precisely reading and analysing emotions conveyed in tweets. As a result, businesses flourish and are better able to make data-driven decisions to efficiently serve their customers' demands. Additionally, tweet emotion identification helps organizations and government quickly assess public opinion in crisis situations. By facilitating efficient response coordination, resource distribution, and prompt assistance to impacted areas, this information eventually improves crisis management and public safety.

5.2 Impact on Environment

Deep learning methods for tweet emotion identification have a little environmental effect. Although these models' training and deployment include some energy use, it is far less than in other industries. Improvements in corporate practises, crisis management, mental health care, and policy-making are just a few examples of possible good social effects that may also indirectly benefit sustainability and environmental wellbeing. Overall, the positive effects on society outweigh the little negative effects on the environment when computing resources are used responsibly.

5.3 Ethical Aspects

Significant ethical ramifications result from the creation and use of tweet emotion recognition models. It is important to protect user privacy and secure their consent before using their data. Since it can result in user profiling for malicious intentions. To stop discriminatory results, bias and fairness concerns must be resolved. Building confidence depends on the models' interpretability and transparency. Responsible deployment also calls for accountability measures like audits and adherence to ethical standards. Promoting justice, openness, and accountability in the employment of tweet emotion recognition algorithms requires taking certain ethical considerations into account.

5.4 Sustainability Plan

In conclusion, a sustainability strategy for tweet emotion recognition models include maximising computing resources, ongoing development, ethical data handling, and encouraging cooperation. To lessen the influence on the environment, this includes effective algorithms, model compression, and cloud computing. Regular improvements and updates improve performance and reduce resource use based on changing data and user input. Privacy is protected by using responsible data practises including safe storage and data reduction. Within the research community, cooperation and knowledge exchange later sustainability and effectiveness. We can guarantee the long-term viability and beneficial effects of tweet emotion recognition algorithms by taking these elements into account.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

Using deep learning approaches, the study constructed and assessed tweet emotion identification models. Three models (LSTM, GRU, and CNN) classified tweet emotions with great accuracy. Privacy, prejudice, justice, accountability, and openness were all ethical issues that were discussed. A sustainability strategy emphasising resource optimisation and assuring ongoing progress was suggested. Overall, the work made a contribution to the fields of deep learning and natural language processing.

6.2 Conclusions

Technology has a significant impact on the world we live in today, making it more accessible and linked. In light of this, we have created a simple and effective method for predicting tweet moods. Our objective is to make it easier to comprehend and analyse the emotions represented in tweets. Our initiative wants to help people effectively recognise and interpret emotions in tweets by utilising cutting-edge technologies and creative models. We are devoted to improving and broadening our models in order to guarantee their performance in a variety of real-world circumstances.

6.3 Implication for Further Study

This research effort has tried to make a small contribution in the world of deep learning. By exploring various deep learning architectures, incorporating information and conducting user studies are promising directions for further investigation of this matter. Moreover, cross culture and cross-lingual studies can also be performed. Suppose, Tweet Emotion recognition models can also be trained on Bangla language. For that a sustainability strategy emphasising resource optimisation and assuring ongoing progress is suggested.