Automate Emotion Analysis of Textual comments

TCS iON Digital Learning

TNSDC RIO – 125

TCS iON Remote Internships

Artificial Intelligence and Real world Application

Submitted by

PANNEERSELVAM R

BSc Computer science

Government Arts College

Udumelpet-642 120

May-2023

CERTIFICATE

        This is to certife that the project work entitled "**AUTOMATE EMOTION ANALYSIS OF TEXTUAL COMMENTS**" is submitted in partial fulfilment of the requirement for the award of the degree of Bachelor of Computer Science is the record of original work done by **R.PANNEERSELVAM** under my supervision and guidance during the period of April 2023 to May 2023 at Government Arts College , Udumalpet and that this project has no formed the basis for the award of any Degree/Diploma/Associate ship/Fellowship similar title to any candidate of any university.

SIGNATURE OF THE GUIDE   **PROJECT MENTOR**

Submitted to the Tcs iONProject viva-voce Examination held on _____

DECLARATION

I hereby declare that this project work entitled "**AUTOMATE EMOTION ANALYSIS OF TEXTUAL COMMENTS**" submitted to the TSC iON in partial fulfilment of the requirement for the award of the degree of Bachelor of Computer Science is a record of original work done by me under the supervision and guidance of Mrs. Dr. PRIYA MCA, M.Phil, ASSISTANT PROFESSOR in Computer Science, Government Arts College, Udumalpet.

TABLE OF CONTENT

- Abstract
- Introduction
- Acknowledgements
- Objective
- Introduction / Description of Internship
- Internship Activities
- Approach / Methodology
- Assumptions
- Exceptions / Exclusions
- Charts, Table, Diagrams
- Algorithms
- Challenges & Opportunities
- Risk Vs Reward
- Reflections on the Internship
- Recommendations
- Outcome / Conclusion
- Enhancement Scope
- Link to code and executable file

- Research questions and responses

ABSTRACT

Emotion Analysis is a process of identifying and analyzing the underlying emotion expressed in the textual data in the form of sentences. Sentiment analysis is a superset of Emotion Analysis. In sentiment analysis, there are only three results such as positive, negative, and neutral, but in Emotion analysis, we can make up to n number of results. In our Emotion Analysis, we are detecting ten Emotions.Sentiment Analytics tries to understand the general feeling and emotion experienced by a viewer or customer. Whereas emotional analysis uses a complex system to understand consumer responses. Unlike sentiment analysis, emotion analysis includes the subtleties of human emotions. So, for example with the positive spectrum, we may find emotions like happiness, contentment, love, excitement, etc. Software products used in emotion analysis are primarily open source and free. Most of the items are purchased online from stores like Amazon and Flipkart. Customer reviews are available for each product, and the people usually read them before purchasing. Emotion analysis can be helpful in determining a customer's emotion in long reviews. The main algorithm is a recurrent neural network (RNN) that has been shown to have a better prediction rate than other algorithms. In comparison to previous models with increased classes, this model provides better result
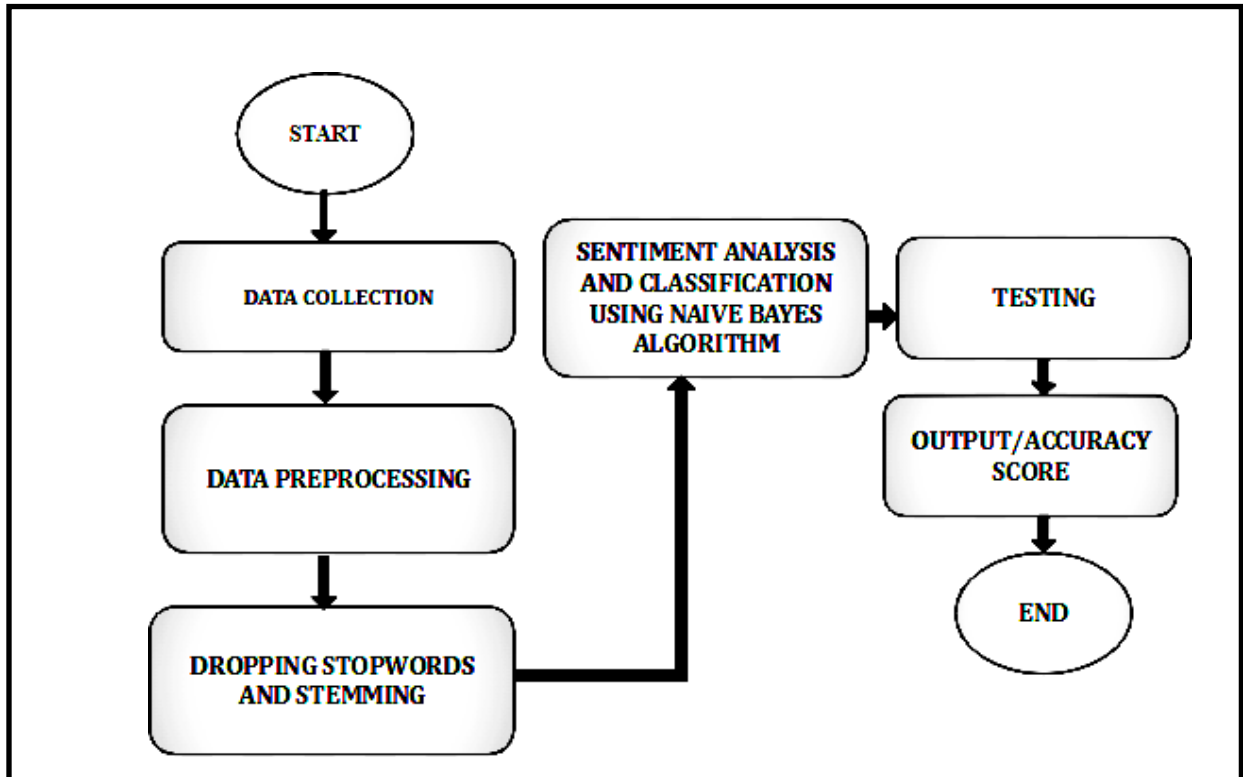
INRTODUCTION

It can be difficult, but it's also important, to identify someone's emotions by analyzing text or data that has been written or expressed by them. This is because matter expressions frequently result not only from the direct use of emotional words, but also from the interpretation of the meaning of ideas and interactions of ideas that are represented within the text of knowledge. Understanding the text's tone is essential for human-computer interactions. To provide the required and accurate result, associate degree RNN LSTM technology may recursively investigate the text knowledge. Feelings can be expressed through anger, surprise, hate, fear, happiness, and a variety of other emotions. because there is no recognised word order for emotions.Despite the considerable effort analysts have made to identify emotions from facial and audio data, the study of emotion recognition through material knowledge is still a vibrant and cutting- edge fiel

DEEP LEARING

Deep learning is a machine learning technique that teaches the computer to do what humans are more likely intended to do. It learns from examples and Deep learning is a key technology for almost all modern problems and inventions such as self-driving cars, where each car sends the information about the road structure everyDeep learning is a machine learning method that instructs computers to perform actions that people are more inclined to execute. Deep learning is a fundamental technique for the majority of contemporary issues and innovations, including self-driving automobiles, in which each vehicle transmits information about the road network every mile, which is subsequently used as training data for further development. known as picking up knowledge via examples. network, deep learning employs a multi-layer strategy. The highlights are either physically described and eliminated in conventional machine learning approaches or are chosen utilizing salient selection techniques. mile, and then the information acts as a learning data for further improvements which is called as learning from examples.Deep learning models, on the other hand, automatically learn and extract the highlights, improving performance and accuracy. In general, spontaneous estimation was also used to estimate the categorization models' superscripts. Figure 1 compares the descriptions of opinion extremes produced by deep learning and traditional machine learning (Support Vector Machine (SVM), Bayesian network, or choice tree). Deep learning and artificial neural networks are currently offering the greatest solutions to specific issues in the fields of speech and picture identification, as well as natural language processing.
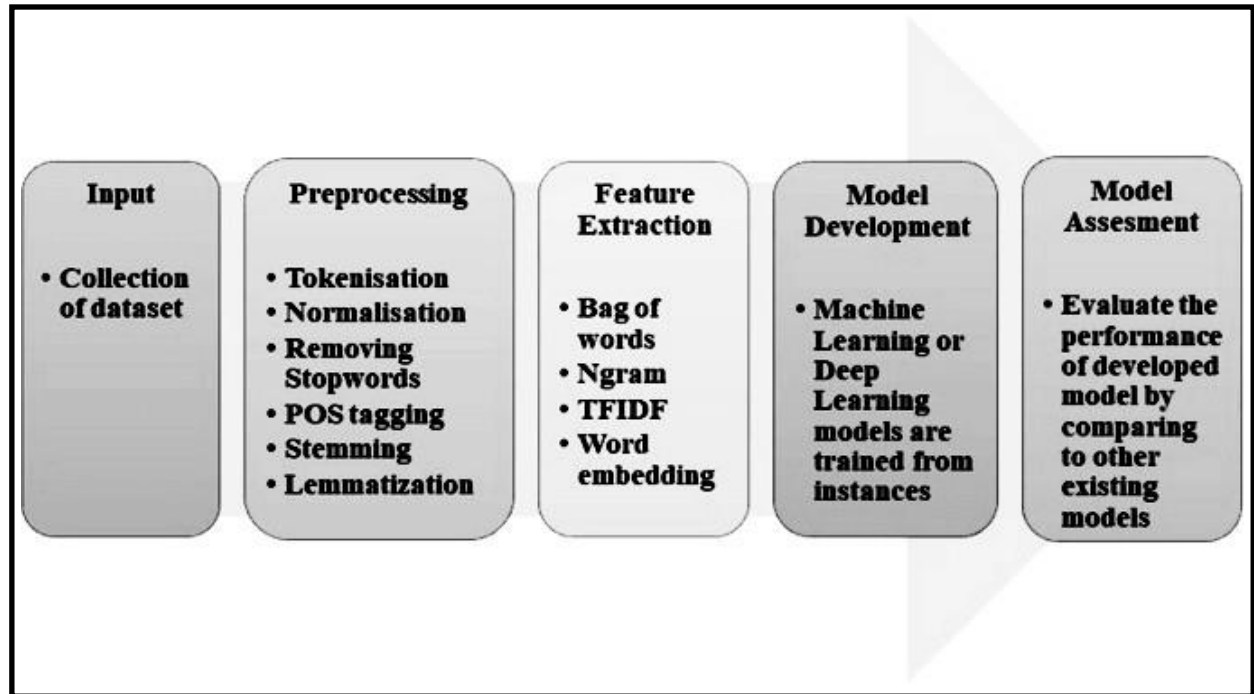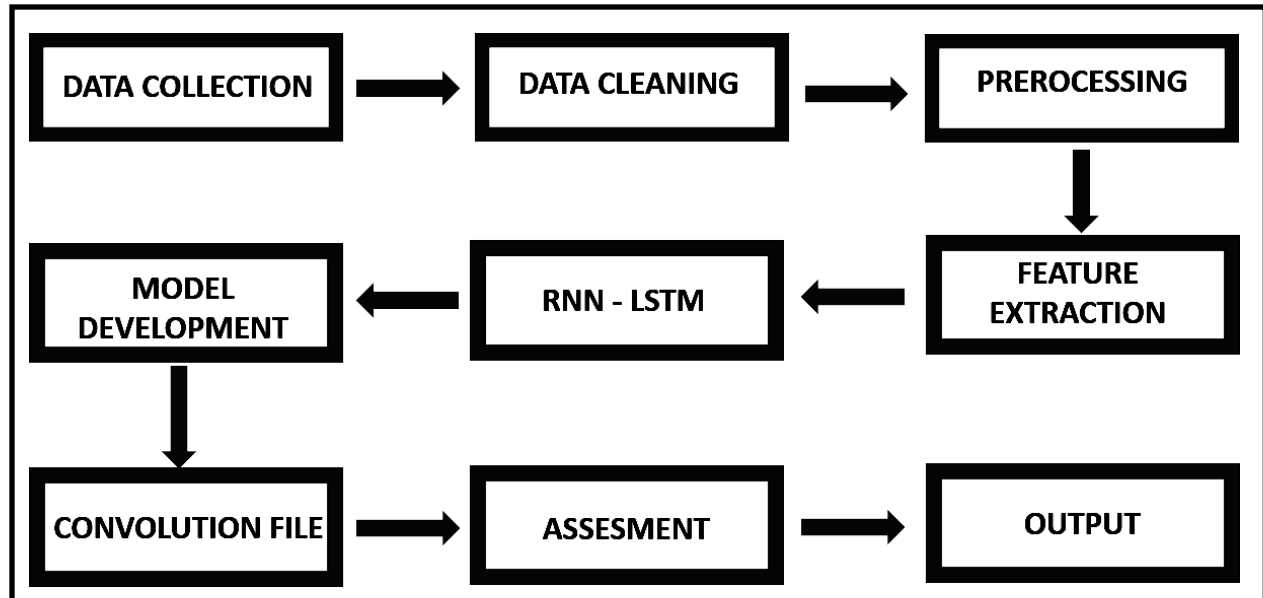
FLOW DIAGRAM

```
                    ┌─────────┐
                    │  START  │
                    └────┬────┘
                         ▼
                 ┌──────────────┐     ┌──────────────────┐     ┌──────────┐
                 │     DATA     │     │ SENTIMENT ANALYSIS│     │          │
                 │  COLLECTION  │     │ AND CLASSIFICATION│     │ TESTING  │
                 │              │     │ USING NAIVE BAYES │     │          │
                 └──────┬───────┘     │    ALGORITHM     │     └────┬─────┘
                        ▼             └─────────▲────────┘          ▼
                 ┌──────────────┐               │           ┌──────────────┐
                 │     DATA     │               │           │OUTPUT/ACCURACY│
                 │ PREPROCESSING│               │           │    SCORE     │
                 │              │               │           └──────┬───────┘
                 └──────┬───────┘               │                  ▼
                        ▼                        │            ┌─────────┐
                 ┌──────────────┐                │            │   END   │
                 │   DROPPING   │                │            └─────────┘
                 │   STOPWORDS  ├────────────────┘
                 │ AND STEMMING │
                 └──────────────┘
```

PROPOSED METHODOLOGY

          The above diagram represents the block diagram of the proposed system. It shows how the input dataset is processed and how the preprocessing works. In feature Extraction, the Word embedding is done. In model development, we deploy our RNN algorithm to get the desired result, and then we assess our model with the existing model to evaluate the performance and calculate the performance improvement.

Emotion detection system

Input	Preprocessing	Feature Extraction	Model Development	Model Assesment
• Collection of dataset	• Tokenisation • Normalisation • Removing Stopwords • POS tagging • Stemming • Lemmatization	• Bag of words • Ngram • TFIDF • Word embedding	• Machine Learning or Deep Learning models are trained from instances	• Evaluate the performance of developed model by comparing to other existing models

Algorithm Emotion detection

```
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│ DATA COLLECTION │ ───► │  DATA CLEANING  │ ───► │  PREROCESSING   │
└─────────────────┘      └─────────────────┘      └─────────────────┘
                                                           │
                                                           ▼
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│     MODEL       │ ◄─── │    RNN - LSTM   │ ◄─── │    FEATURE      │
│  DEVELOPMENT    │      │                 │      │   EXTRACTION    │
└─────────────────┘      └─────────────────┘      └─────────────────┘
        │
        ▼
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│ CONVOLUTION FILE│ ───► │    ASSESMENT    │ ───► │     OUTPUT      │
└─────────────────┘      └─────────────────┘      └─────────────────┘
```

shows the Emotion detection flow design that comprises Data collection and cleaning and preprocessing by cleaning stopwords, symbols, and user handles. Then Feature extraction using TF and IDF and then using Recurrent neural network (Long Short term Memory) to create our convolution file and then assess our project and optimize the system.

The steps to be followed in design flow of emotion detection,

i.   Emotion analysis is a Natural Language Processing(NLP) technique used to determine a person's state of things

ii.   Collected dataset and preprocessed by cleaning user handles, symbols, and stop words using neattext functions

iii.	Tokenize the data into words and sentences and then Feature extraction like word embedding, Bag of words, and TF-IDF is done.

iv.	Used count vectorizer for counting the tokens/words

v.	Used RNN and CNN techniques to predict the Emotion recursively

vi.	Used LSTM (Long Short Term Memory) from RNN

vii.	It is a recurrent model and can remember and regulate values over arbitrary time intervals. It is more useful for predictions.

viii.	Created a convolutional file and calculated the accuracy

ix.	Used the convolutional file for the detection process.

Recurrent Neural Networks (RNNs) are a type of artificial neural network layer that have unlimited impulse responses and dynamic activity. Using neural network technology, RNN offers quick and precise identification, which is necessary for this topic. RNN will recursively evaluate the text of this article to determine the frequency of words and the poles of positive and negative statements. cyclic neural networks now provide more accurate predictions than the previously employed logistic regression and Naive Bayes binomial.

An RNN's primary function is to process sequential data using internal memory that is captured by coordinate cycles.RNNs, in contrast to conventional neural networks, can remember the computation of data that has already been performed and reuse it by applying it to the following element in the layout of the information source. The long-term short-term memory (LSTM), which is skilled at relating to long-term memory as a contributor to hidden layer initialization, is a special sort of RNN. A representation of the LSTM design can be seen in Figure 6. To reformat the information for the linked network, the informational data is preprocessed (interaction similar to that described for CNN). The

LSTM, which is the subsequent layer, has 200 cells. With 128 cells to structure the message, the final layer is a completely linked layer.The final layer, assuming there are two classes to predict, employs the sigmoid-issuing capability to decrease the level 128 vector to the resulting vector of one (positive, negative).

RESULTS AND DISCUSSION

Confusion matrix, accuracy, precision, and recall metrics were used to gauge the model's performance. RNN models were used to carry out various tasks with the various datasets shown above, and feature extraction was used to assess how those computations were

presented. It also displays recall and precision for reference since F-score is derived from these two metrics.After training of 50000 sentence review. The accuracy range of the system ranges above 85 percent and the convolutional weight file's size ranges around 250 MB as shown in

```
confusion_matrix(y_test,y_pred_for_nv)

array([[ 649,    0,   64,    0,  374,    7,    0,  178,    0,   11],
       [  33,    9,   24,    0,  135,    4,    1,   75,    0,   11],
       [  79,    2, 1001,    0,  350,    4,    0,  148,    0,   61],
       [   5,    0,    5,    1,  124,    0,    0,    1,    0,    0],
       [  80,    1,  111,   29, 2674,    9,    0,  199,    0,   72],
       [  24,    0,   18,    0,  550,   17,    0,   62,    0,    4],
       [   9,    0,    9,    0,    6,    1,    2,  159,    0,    0],
       [  98,    1,   83,    0,  741,    7,   23,  841,    0,   35],
       [   9,    0,    7,    0,   13,    0,    0,    5,    1,    1],
       [  56,    5,   50,    0,  622,    4,    0,  125,    0,  319]])
```

There are nine classes in Emotion detection system,

- Happy

- Surprise

- Fear

- Neutral

- Anger

- Sad

- Joy

- Shame

- Not satisfied

First start the prediction with Multinomial Naive Bayes (NB) and accuracy of 0.528 is obtained as depicted in Figure 9. Next started the prediction with probability regression and got an accuracy of 0.682 as shown in Figure 10. Here, the accuracy increased from 0.528 to 0.682 with an average of 29% increase.This Accuracy is very low and it cannot be persistent, So RNN-LSTM is proposed.

```
nv_model = MultinomialNB()
nv_model.fit(X_train,y_train)

MultinomialNB()

# Accuracy
# method 1
nv_model.score(X_test,y_test)

0.5282621191799195
```

```
np.max(nv_model.predict_proba(vect))

0.6820992786191742
```

```
test_loss, test_acc = model.evaluate(test_dataset)

print('Test Loss: {}'.format(test_loss))
print('Test Accuracy: {}'.format(test_acc))

391/391 [==============================] - 337s 863ms/step - loss: 0.5312 - accuracy: 0.8524
Test Loss: 0.5312148928642273
Test Accuracy: 0.852400004863739
```
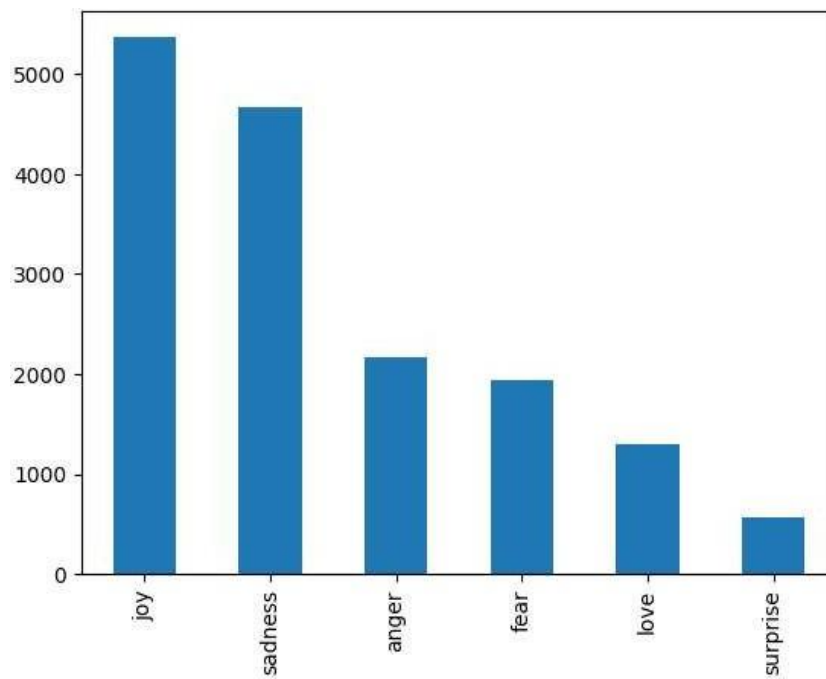
CODING: IPython.py   in Google Clob

OUTPUT:
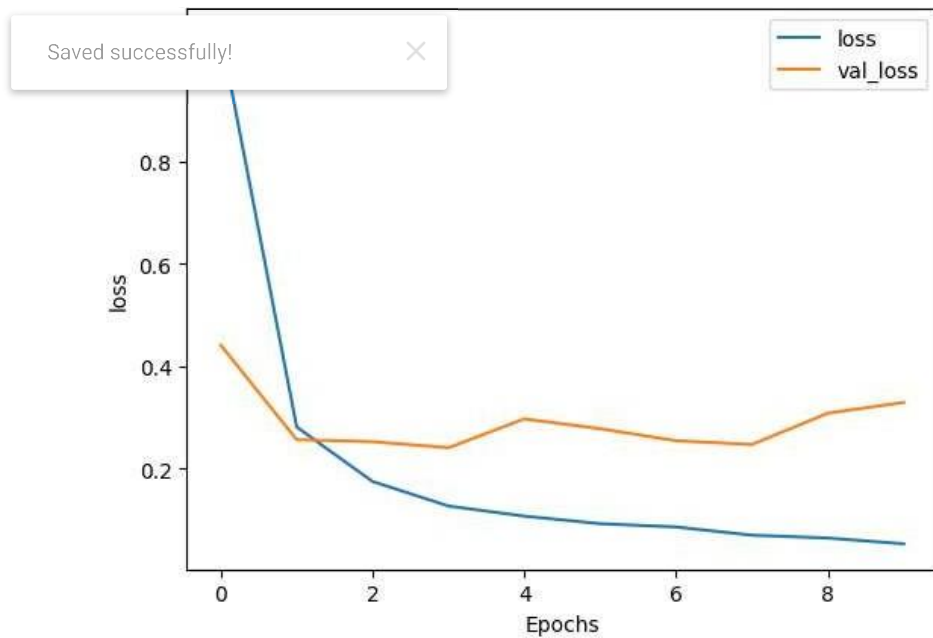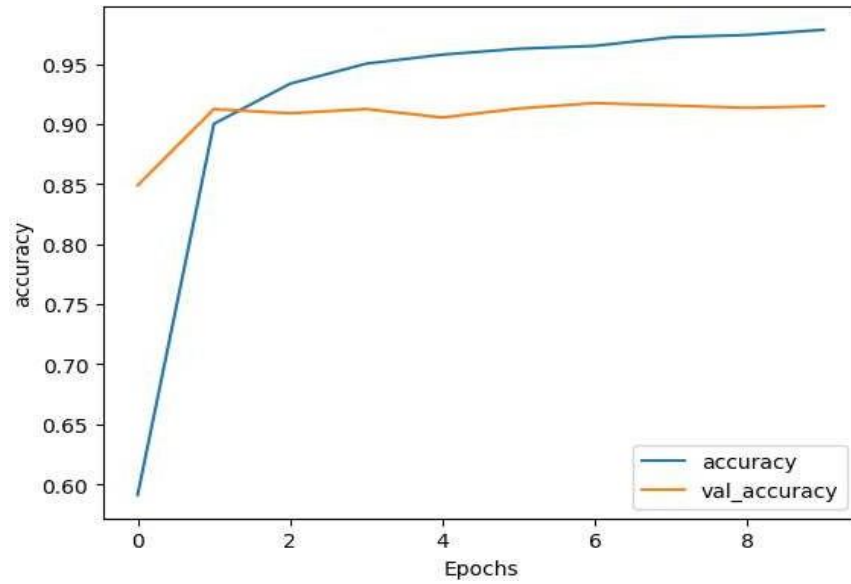
Emotios

ACCURACY

OUTPUT:


```
1/1 [==============================] - 1s 894ms/step
I am scared of hippos The prediction is ([4], [(4, 'fear')])
1/1 [==============================] - 0s 69ms/step
['im feeling rather rotten so im not very ambitious right now'
 'im updating my blog because i feel shitty'
 'i never make her separate from me because i don t ever want her to feel like i   ashamed with
 'i left with my bouquet of red and yellow tulips under my arm feeling slightly more optimistic her'
                                                                              than when i arrive
 'i was feeling a little vain when i did this one']
                        1, 0], [(0, 'sadness'), (0, 'sadness'), (0, 'sadness'), (0, 'sadness'),
                                                                              (1, 'joy')]])
```

```
y_pred =  np.argmax(model.predict(X_test_pad),
y_true = np.argmax(y_test, axis=1)              axis=1)
from sklearn import metrics
print(metrics.classification_report(y_pred, y_true))
```

```
                63/63                        164ms/step
                                             support

          0     0.96    0.91    0.93    615
                        [==============================] - 10s
          1     0.90    0.94    0.92precision 668recall  f1-score
          2     0.70    0.77    0.74    145
          3     0.84    0.86    0.85    270
          4     0.86    0.85    0.85    227
          5     0.79    0.69    0.74    75


   accuracy                     0.89    2000
  macro avg     0.84    0.84    0.84    2000
weighted avg
```

CONCLUSION

The main aim of Automated Detection of Emotion is to create an Emotion-based review system, predict Emotions from textual data and generate a product review system.The performance and accuracy of machine learning algorithms and deep learning algorithms depend on the pre-processing and size of the dataset. Many of the algorithms are not persistent but the RNN LSTM model is prevalent in emotion analysis. But also, lexicon and ML- based approaches are also evolving and getting better.