**A Rule-Based Music Recommendation System Based on Depression Levels of User Posts Extracted from Social Networks**

*A project report submitted for partial fulfilment of requirements for the degree* in **Master of Technology**

Submitted by

**Md Ashed Dewan**

Under the Supervision of

**Mr. Zafar Sarif** and **Mrs. Moumita Chatterjee**

**(**Assistant Professor Dept of CSE**)**

Department of Computer Science & Engineering

Aliah University

Kolkata, West Bengal, 700153

**DECLARATION**

I certify that the work contained in this report is original and has been done by me under the guidance of our supervisor Mr. Zafar Sarif and co-supervisor Mrs. Moumita Chatterjee, Assistant Professor, Aliah University, Kolkata.

1. The work has not been submitted to any other Institute for any degree or diploma.
2. We have followed the guidelines provided by the Institute in preparing the report.
3. We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
4. Whenever we have used materials (data, theoretical analysis, figures, and text) horn other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references.

Name and Signature of Project Member:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Enrollment No** | **Name of The Student** | **Signature of Student** |
| **1.** | **CSE212015** | **Md. Ashed Dewan** |  |

**Aliah University**

**CERTIFICATE**

It is certified that the project report entitled **“A Rule-Based Music Recommendation System Based on Depression Levels of User Posts Extracted from Social Networks”** submitted by *Md Ashed Dewan* **(***CSE212015***)** have been found satisfactory for the requirement of the degree.

**……………………………………... ……………………………………...**

Supervisor Co-Supervisor

(Mr. Zafar Sarif) (Mrs. Moumita Chatterjee)

Data : …………… **……………………………………...**

(Dr. Souvik Sen Gupta)

Head of the Department

CSE, Aliah University

**ACKNOWLEDGEMENTS**

I am deeply thankful to my advisor and Co-Advisor, Mr. Zafar Sarif and Mrs. Moumita Chatterjee for helping me throughout the course in accomplishing my final project. Their guidance, support and motivation enabled me to achieve the objectives of the project. I developed this project with the help of Faculty members of our institute, and we are extremely grateful to all of them. I also take this opportunity to thank my supervisor and co-supervisor for providing the required facilities in completing this project. I am thankful to our parents, friends, and faculty members for their motivation, guidance and help whenever needed.

Name and Signature of team Member :

Md Ashed Dewan : **……………………………………...**

# TABLE OF CONTENTS

ABSTRACT………………………………………………………...…………….09

CHAPTER 1

1. INTRODUCTION…………………………………………………………….10
   1. Motivation……………………………………………….…......….10
   2. Domain Introduction………………...…………………...…….......11
      1. Sentiment analysis……………………………………………11
      2. Lexicon-based methods………………………………………11
      3. Machine learning……………………………………………..11
      4. Deep learning…………………………………………………12
      5. Recommendation Systems……………………………………13

1.3 Thesis Outline…………………………………………………………….13

CHAPTER 2

1. LITERATURE REVIEW………………………………………………….….14

CHAPTER 3

1. BACKGROUND FUNCTIONALITY AND DESIGN………………………17

3.1. Data Acquisition…………………………………………………………18

3.2. Data Preparation…………………………………………………………19

3.3 Feature Extraction……………….….…………………………………...20

3.3.1 N-gram modeling and TF-IDF……………………………...…20

3.4 Word Embedding…………………………….……………………….….22

3.4.1 Word2vec Word Embedding………….……………………….…22

3.5 Classification…………………………………………………………….23

3.5.1 Machine Learning Classification……………………………...23

3.5.2 Deep Learning Classification……………….………………….…29

3.5.3 Recommendation Systems………………………………………32

3.5.1.1 Supervised learning………………………………………23

3.5.1.2 Unsupervised learning…………………….……………...23

3.5.1.3 Reinforcement Learning………………………………….23

3.6. Decision Tree…………………………………………………………….24

3.7. Random Forest………………………………………………...…………25

3.8. K-Nearest Neighbor (KNN)……………………………….……………26

3.9. Support Vector Machine (SVM)…………………………………………27

3.10 Nave Bayes………………………………......…………………………28

3.10 LSTM (Long Short-Term Memory)……………………….……………29

CHAPTER 4

4. EXPERIMENTAL EVALUATION...…………………………………...…….34

4.1 Machine learning Models Result…………………………………………34

4.2 LSTM Models Results….…………………………......….………………35

4.3 Recommended System Results...................................................................39

CHAPTER 5

5. CONCLUSION AND FUTURE RECOMMENDATIONS………………….42

6. REFERENCES……………………………………………………………….43

7. CODE…………………………………………………………………………44

# LIST OF TABLES

Table 1. Data Static……………………………………………….……………….18

Table 2. Parameters of the LSTM…………………...……………………………….31

Table 3. LSTM Sequential Model………………………………......…………….32

Table 4. Sentiment Level……………………..…………………………………….33

Table 5. Machine learning i) Accuracy ii) Precision iii) Recall iv) F1 Score…….34

Table 6. Confusion Matrix…………………………………….…………………….36

Table 7. LSTM : i) Accuracy ii) Precision iii) Recall iv) F1 Score…………………….36

# LIST OF FIGURES

Fig 1: Sentiment analysis approaches……………………………………………..12

Fig 2: Proposed Model……………..…………………………….……………….17

Fig 3: Non-suicide and suicide……………………………..…….……………….19

Fig 4. Nondepressed Text…………………………………………..……………….21

Fig 5. Depressed Text……………………………...……………….……………….21

Fig 6. CBOW & Skip-gram……………………………...……….……………….22

Fig 7. Decision Tree Fig………..……………………………….……………….24

Fig 8. Fandom Forest Working……………..…………………….……………….25

Fig 9. K-NN Working……………………….………………….……………….26

Fig 10. SVM Working Principle………………………………….……………….27

Fig 11. LSTM…………………………………………………….……………….30

Fig 12. Train and Validation accuracy vs No of epochs…………………………………37

Fig 13. Train and Validation accuracy vs No of epoch………………………………….37

Fig 14. ROC and AUC curve……………………………………….……………….38

# ABSTRACT

The statistics presented by the World Health Organization attribute depression to be a primary cause of concern globally, leading to suicide in most cases if left undetected. Text sentiment analysis, also known as opinion mining, is a computational technique used to determine the sentiment or emotional tone expressed in each piece of text. It involves analyzing text data to classify it into positive, negative, or neutral sentiments. Analyzing public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to employ machine learning, deep learning approaches and natural language processing techniques for training our data and evaluating the efficiency of our proposed method. With TF-IDF features machine learning models give around 90% accuracy and with Word2vec LSTM give 94% accuracy. And we develop a rule base music recommendation engine that is recommend song for depress user base on certain condition.

***Chapter 1***

**INTRODUCTION**

* 1. **Motivation**

Depression is a common mental disorder. Globally, it is estimated that 5% of adults suffer from the disorder. Along with the use of social media, depressive posts on social media are increasing at an alarming rate. February 17, 2021, Mumbai, a 21-year-old youth Planned to end his life and posted the suicide note on his Facebook account. A timely alert by a Dadar resident helped the police to find the youth and save him. We would lose a 21-year-old youth if he had not informed the police in time. In this way we have lost many people, like actor Sandeep Nahar and the Chinese girl Zoufan and Zeng. The motivation for using sentiment analysis in depression detection arises from the need to identify individuals at risk of depression, provide timely support, and improve mental health outcomes.

Also, the growing field of natural language processing (NLP) and advancements in machine learning techniques have paved the way for more accurate and sophisticated sentiment analysis models. Leveraging these technologies in depression detection can lead to improved accuracy, reliability, and scalability of sentiment analysis-based approaches.

* 1. **Domain Introduction**
     1. **Sentiment analysis:**

Sentiment analysis has gained significant attention in recent years due to the vast amount of text data available from various sources such as social media, customer reviews, surveys, news articles, and online forums. Organizations and individuals can benefit from sentiment analysis by gaining valuable insights into public opinion, customer feedback, market trends, and brand perception. The primary objective of sentiment analysis is to classify text documents or snippets into predefined sentiment categories such as positive, negative, or neutral. However, sentiment analysis can also involve more fine-grained analysis, such as identifying emotions like happiness, sadness, anger, or fear, or even analyzing the intensity or polarity of sentiments. To perform sentiment analysis, various techniques and approaches are used, including:

**1.2.2 Lexicon-based methods**: Lexicon-based approaches utilize sentiment dictionaries or lexicons that contain words or phrases associated with specific sentiment polarities. Each word or phrase is assigned a sentiment score, and sentiment analysis is performed by aggregating these scores. Lexicon-based methods are simple but may not capture the context or nuances of sentiment.

**1.2.3 Machine learning**: Machine learning techniques, such as supervised learning, are commonly used in sentiment analysis. These approaches involve training models on labeled data, where the sentiment of the text is known, and then using the trained model to classify new, unseen text. Various algorithms like Naive Bayes, Support Vector Machines (SVM), or ensemble methods like Random Forests and Gradient Boosting can be employed.

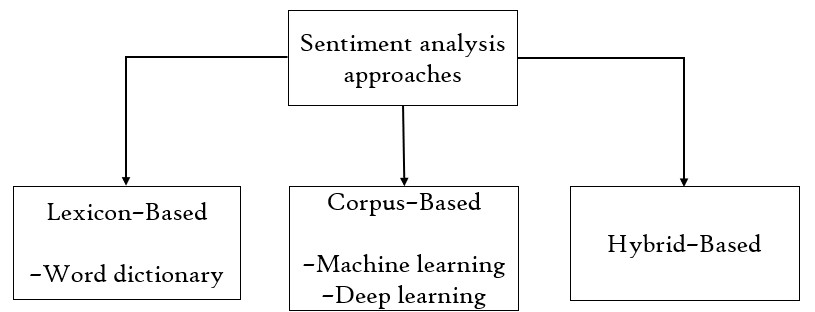
**1.2.4 Deep learning**: Deep learning models, such as neural networks, have shown promising results in sentiment analysis. Models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks can capture complex patterns and dependencies in text data, allowing for more accurate sentiment classification.

Fig 1: Sentiment analysis approaches

Applications of sentiment analysis span across various domains and industries:

* Customer feedback analysis
* Market research.
* Social media monitoring
* Financial analysis

**1.2.5 Recommendation Systems:**

There are three techniques of recommendation systems:

**Collaborative:** The collaborative method examines shared preferences across individuals by analyzing user behavior and preferences.

**Content-based:** The content-based method relies on the relationship between an item's description and the user's profile; item suggestions are made in accordance with the user's preferences [19, 20]. It is known as rule-based recommendation.

**Hybrid-based:** The hybrid strategy combines the two approaches.

**1.3 Thesis Outline**

The thesis is organized as follows:

Chapter 1 provides a general introduction to the thesis.

Chapter 2 literature review and problem Statement.

Chapter 3 introduces the necessary background functionality and designs.

Chapter 4 provides the experimental evaluation and results of the proposed method.

Chapter 5 concludes our work and the future work.

***Chapter 2***

**LITERATURE REVIEW**

In this paper [1] authors implications of hyper-parameter tuning and how it might be useful for depression research on a small Bangla social media dataset are demonstrated. The outcome demonstrates that for stratified datasets with recurrent sampling, excellent depression detection accuracy can be achieved using 5 layered LSTMs of size 128 with batch sizes 25 and learning rates of 0.0001 across 20 epochs.

In this paper [2] authors investigated word embedding models (Word2Vec, Glove) in tweets using deep learning techniques to identify sentiment polarity. Here, by including memory in a network model for prediction and visualization, we investigated how sentiment analysis using the recurrent neural network (RNN) model and long-short term memory networks (LSTMs) units can manage long-term dependencies.

The authors of [3] conducted a Chinese depression analysis. They blended psychological and machine learning expertise in their work. With the assistance of psychologists, the authors chose 90 depressed and 90 non-depressed Sina Microblog users, gathering a total of 6013 microblogs. Their model's accuracy was 80%.

In this paper [4] the authors' major focus in this study is on a combined task that combines targeted aspect-based polarity classification with target-dependent aspect detection. Using two benchmark datasets, the effectiveness of the suggested strategies is assessed for this collaborative effort. The experiment demonstrates that in two targeted aspect sentiment tasks, the suggested attention architecture and knowledge-embedded LSTM might outperform cutting-edge techniques.

In this paper [5] authors of this study investigate a novel use of recursive neural networks (RNN) coupled with deep learning for sentiment analysis of reviews. By examining various reviews and subsequently generating a score based on them, the proposed RNN-based Deep-learning Sentiment Analysis (RDSA) recommends locations close to the user's present location.

The authors of this paper [6] describe an enhanced sentiment metric (eSM), which combines a lexicon-based sentiment metric with a user profile-based correction factor, to create a music recommendation system based on sentiment intensity metrics.

In this article [7], a hybrid recommendation system for movies is proposed. It makes use of the most effective ideas from CF and CBF as well as sentiment analysis of tweets from microblogging websites. The goal of using movie tweets is to comprehend current trends, popular opinion, and user reaction to the film.

In this article [8] To show how the ensemble and hybrid techniques increase depression detection performance, the authors of conducted an experimental investigation. On each of the three datasets, numerous tests were run using various sentiment lexicons and logistic regression as the classifier. They used attention LSTM and long short-term memory (LSTM) in deep learning techniques to conduct tests for automated depression identification.

In this article [9] create a prediction model for gauging public opinion from Twitter data during the COVID-19 outbreak, the authors of this article used BOW, TF-IDF, topic analysis, document embedding, and supervised learning algorithms.

In this article [10] authors designed a novel music recommendation system based on psychotherapy. They create an emotion model and take some deference type of song and user’s latest favorite music, user’s latest playlist and develop a lstm model with word2vec word embedding. LSTM-based model can not only select the most helpful music based on users' previous mood and current emotion stimulus, but also use the care factor to adjust the results to improve users' mental status.

***Chapter 3***

**BACKGROUND FUNCTIONALITY AND DESIGN**

The process of designing a functional classifier for sentiment analysis can be broken down into five basic categories. They are as follows:

1. Data Acquisition
2. Data Prepossessing
3. Feature Extraction and Word Embedding
4. Classification
5. Recommended Music

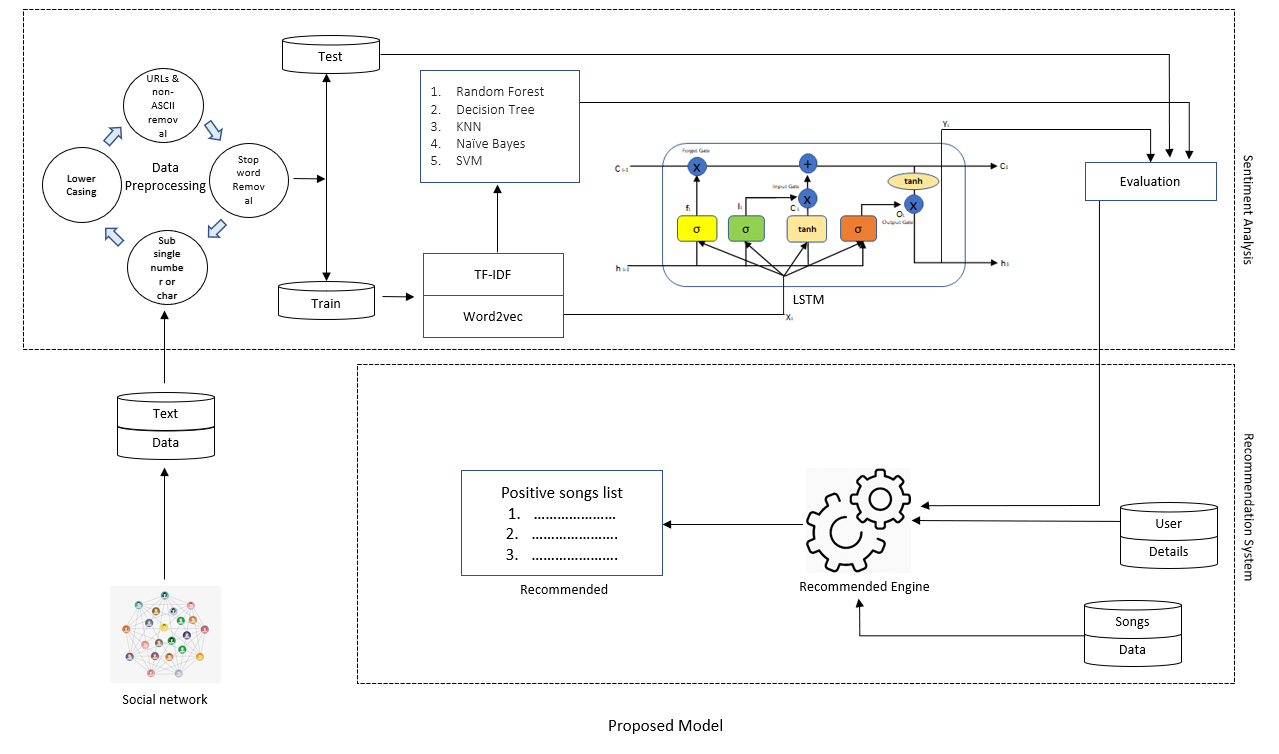


Fig 2: Proposed Model

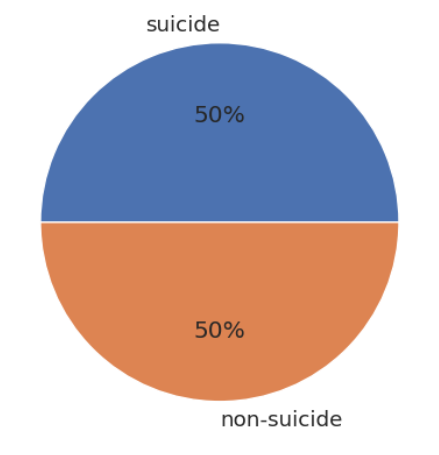
3.1. **Data Acquisition:**

|  |  |  |
| --- | --- | --- |
| Class | suicide | Non- suicide |
| Number of Text | 348096 | 348036 |
| Mean number of words | 131.33 | 131.33 |
| Min number of words | 01 | 01 |
| Max text length | 7796 | 7796 |

The proposed work uses Kaggle “Suicide and Depression Detection” data set for depression detection. The dataset is a collection of posts from the "SuicideWatch" and "depression" subreddits of the Reddit platform. The posts are collected using Pushshift API. All posts that were made to "SuicideWatch" from Dec 16, 2008(creation) till Jan 2, 2021, were collected while "depression" posts were collected from Jan 1, 2009, to Jan 2, 2021. All posts collected from SuicideWatch are labeled as suicide, while posts collected from the depression subreddit are labeled as depression. Non-suicide posts are collected from teenagers.

Table 1. Data Static

A total of 232045 English-language tweets containing these search terms were collected from 2000 users. Out of this text (46408 / 232044), text corresponding to use for testing purposes.

Fig 3: Non-suicide and suicide

Another data set are one is positive music data set which have 40 to 50 positive song, second data is user data set which have 2000 user with name, age, and gender.

**3.2. Data Preparation**

Data from online social media cannot be modelled to forecast outcomes. This causes issues with sentiment analysis and word matching. Raw social media data has the drawback of potentially including typos, misspellings, emoticons, and other offensive characters. To ensure that the computational model produces accurate predictions, the data must be preprocessed. On the analysis data, the following data preparation operations are carryout:

* URL links appearing in user posts are removed as part of preprocessing because they do not convey meaning or polarity.
* Stop words like 'a', 'an', 'the', etc.. have been removed as they are not discriminating or useful for our model.
* To improve text quality, non-ASCII characters are removed.
* Tokenizing is the process of converting sentences into collocations of a single word.
* Stemming is done to change each word to the root word.
* POS (part of speech) tagging is performed to reduce ambiguity when interpreting words.

**3.3 Feature Extraction**

Features play a significant role in text sentiment analysis. The forecast rate rises when a set of carefully chosen characteristics accurately captures each emotion. However, there is not a collection of characteristics that is universally acknowledged for accurate and unique classification. Only the TF-IDF is employed in this paper.

**3.3.1 N-gram modeling and TF-IDF:** N-gram modeling is used for examining the features that are present in posts. It is commonly used in NLP as a feature to calculate the co-occurrence probability of each text as a unigram or bigram. For n-gram modeling, TF-IDF is used as a statistical measure for highlighting the importance of a word concerning user posts. The idea behind this approach is to reduce the impact of frequently occurring less informative tokens and consider more informative tokens that occur less frequently. The tokens having a higher TF–IDF value are present in a particular post and are not present in other posts.

I removed the stop words from the dataset and restricted the term –document matrix to most frequent unigrams and bigrams. We choose the top 2000 unigrams and bigrams for each type of post. This is shown in Fig 4 and Fig 5. The words that appeared with a high frequency are depicted in both figures. Depressed text analysis of our results indicates that the lexicon that is predictive of depression includes words that contain negative emotion, feelings, self-obsession, suicidal thoughts, hostility, anger, negative words, hopelessness, meaninglessness, and present tense [24].



Fig 4. Nondepressed Text

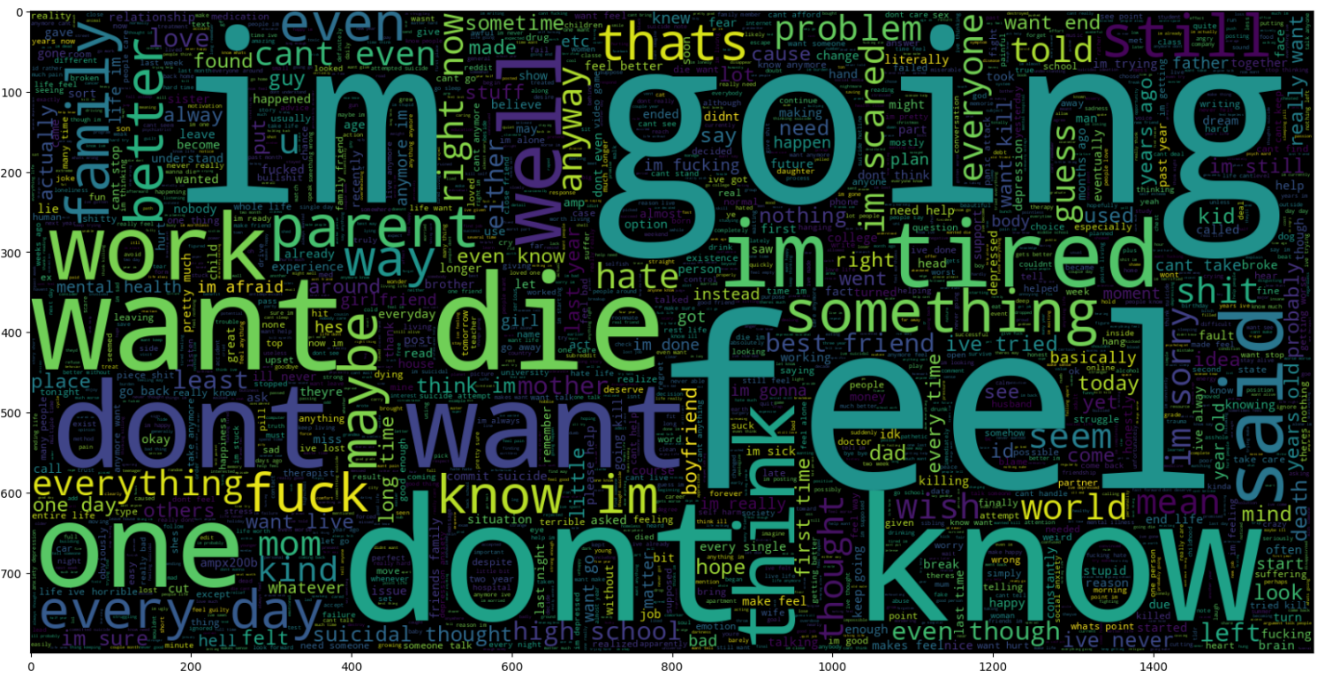


Fig 5. Depressed Text

**3.4 Word Embedding**

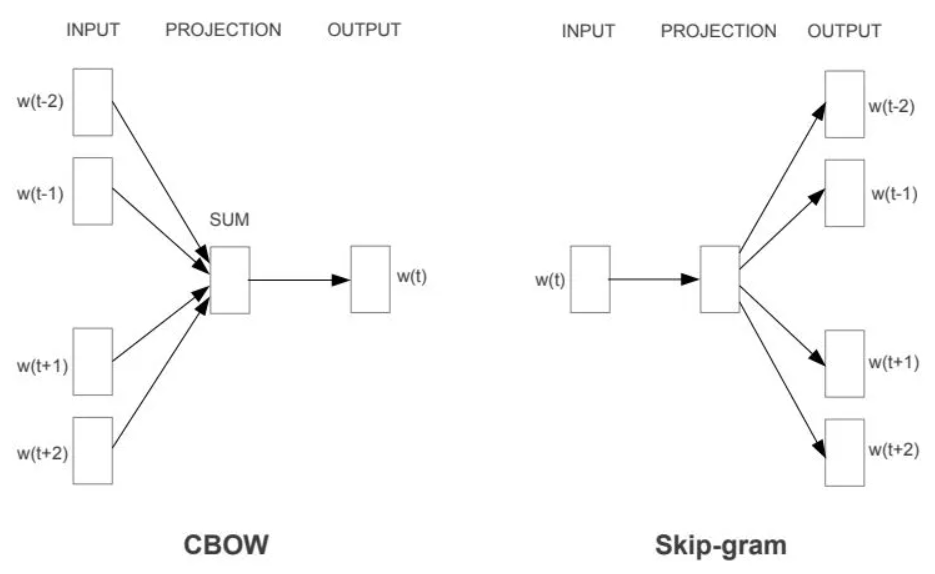
**3.4.1 Word2vec Word Embedding:** Word2Vec is a model that is frequently used in natural language processing (NPL) that can autonomously learn semantic information from a vast number of text corpora. Through an embedding space, it can reduce the physical separation between words with similar semantic properties. Word2Vec comes in two varieties: CBOW (Continuous Bag-of-Words Model) and

Fig 6. CBOW & Skip-gram

Skip-Gram (Continuous Skip-Gram Model)[21]. While CBOW can anticipate the current words using context, Skip-Gram can forecast context based on the current words. Figure 1 depicts the structure of the CBOW model and the Skip-Gram model. The input layer, projection layer, and output layer are present in each of them.

**3.5 Classification**

**3.5.1 Machine Learning Classification:**

Machine learning (ML) is essentially the area of computer science that enables computer systems to make sense of data in a manner like that of humans.Simply put, machine learning (ML) is a sort of artificial intelligence that uses an algorithm or method to extrct patterns from raw data.The main goal of ML is to make it possible for computers to learn from experience without explicit programming or human oversight.

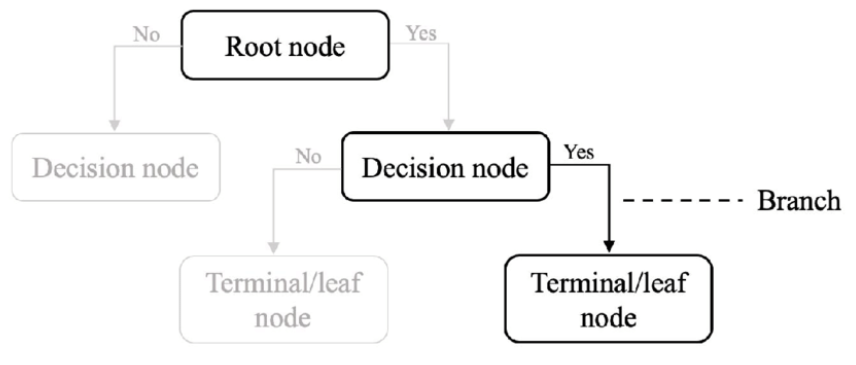
**3.5.1.1 Supervised learning:** An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans. You can use supervised learning when the output data is known. The algorithm will predict new data.

**3.5.1.2 Unsupervised learning**: Unsupervised learning is used when you do not know how to classify the data and want the algorithm to find patterns and classify the data for you. For example, an algorithm exploring customer demographic data to identify patterns would explore input data without being given an explicit output variable.

**3.5.1.3 Reinforcement Learning:** One of the most common and rising categories of machine learning algorithms is reinforcement learning. It is utilized in many autonomous systems, including automobiles and commercial robotics. This algorithm's purpose is to accomplish a task in a changing environment. Based on a few prizes that the system offers it, it can achieve this aim.

**3.6. Decision Tree:**

This kind of supervised learning technique is most frequently applied to classification issues. Unexpectedly, it functions for both continuous and categorical dependent variables. We divide the population into two or more homogenous sets using this approach. To create as many distinct groups as feasible, this is done using the 67 most important characteristics/independent variables [23].

Fig 7. Decision Tree

Since it requires minimal data preparation, the decision tree technique is quite easy to implement. A lot of people use it for data exploration. Decision trees are particularly vulnerable to noise in the dataset, though. Overfitting could become a concern because of this.

**3.7. Random Forest:**

First, N decision trees are combined to generate a random forest, and then predictions are made for each tree that was produced in the first phase.

The stages and graphic below can be used to demonstrate the working process:

Fig 8. Fandom Forest Working

Step 1: Pick K data points at random from the training set.

Step 2: Create the decision trees linked to the subsets of data that have been chosen.

Step 3: Select N for the size of the decision trees you wish to construct.

Repeat steps 1 and 2 in step 4.

Step 5: Assign new data points to the category that receives most votes by looking up each decision tree's predictions for the new data points.

**3.8. K-Nearest Neighbor (KNN):**

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbor.

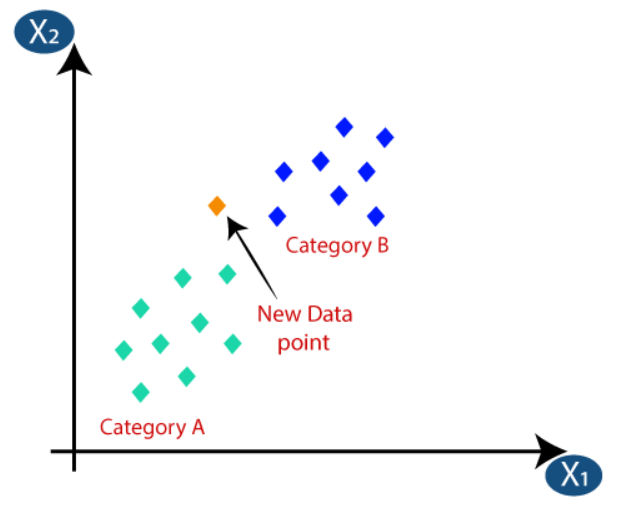
The K-NN algorithm assumes that the new case and the existing cases are comparable, and it places the new instance in the category that is most like the existing categories.

Fig 9. K-NN Working

The following algorithm can be used to describe how the K-NN works:

Step 1: Decide on the neighbors’ K-numbers.

Calculate the Euclidean distance between K neighbors in step two.

Step 3: Based on the determined Euclidean distance, select the K closest neighbors.

Step 4: Count the number of data points in each category among these k neighbors.

Step 5: Assign the fresh data points to the category where the neighbor count is highest.

Step six: Our model is complete.

**3.9. Support Vector Machine (SVM):**

SVM categorizes data points even when they are not otherwise linearly separable by mapping the data to a high-dimensional feature space. The data is changed once a separator between the categories is identified so that the separator could be drawn as a hyperplane. The group that a new record should belong to can therefore be predicted using the features of new data[22].

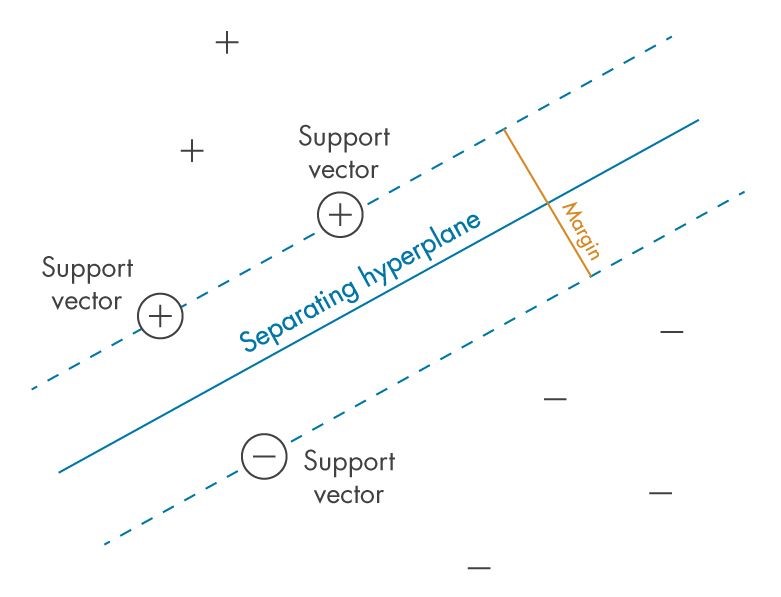
Consider the following graphic, for instance, where the data points are divided into two groups with a separating hyperplane.

Fig 10. SVM Working Principle

**3.9 Nave Bayes:**

The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem.

It is mostly employed in text categorization with a large training set.

Create frequency tables from the provided dataset.

Create a likelihood table by calculating the odds of the given attributes.

Now, determine the posterior probability using the Bayes theorem.

Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.

P(A/B) = P(B | A) P(A) / P(B)

**Where P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B. **P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true. **P(A) is Prior Probability**: Probability of hypothesis before observing the evidence. **P(B) is Marginal Probability**: Probability of Evidence.

**3.5.2 Deep Learning Classification:**

**3.10 LSTM (Long Short-Term Memory):**

Divide the dataset for each aspect and sentiment classification into 80% training data and 20% test data when the word embedding procedure finished. Sentiment analysis is the categorization of opinions in text form. It may be divided into two categories: sentiment expressed in products/movies and sentiment expressed on social networking platforms like Twitter, Facebook, and Instagram [11].

A sort of sentiment analysis known as aspect-level sentiment analysis (ALSA) examines all sentiments from all angles [12]. ALSA is also known as aspect-based sentiment analysis (ABSA). To obtain better outcomes and more accuracy, the authors of this study will conduct a deeper investigation of the attitude-based sentiment analysis (ABSA).

Features, business and sentiment polarity, and content to gather sentiments are the aspects that were employed in this study. The analysis in this study is based on application users' feedback, both favorable and negative. Model of Recurrent Neural Network with Long Short-Term Memory Deep learning models may automatically learn semantic and syntactic information, according to few studies, increasing the precision of sentiment analysis. Deep recurrent neural networks used in aspect-based sentiment analysis to extract opinions [13]. Due to its capacity to learn characteristics at an elevated level and automatically detect polarized public sentiments on certain items, deep learning has been frequently employed in sentiment analysis. RNN used to analyses explicit sentiment, model links between syntactic structures in phrases, and predict sentiment categorization [14].

One example of a supervised deep learning algorithm is the RNN. In this instance, neurons linked to one another throughout time. RNNs are designed to keep track of the information that previous neurons had so that these neurons may later transmit that information back to themselves for additional processing. Thus, data from one time instance (t1) is used as input for the subsequent time instance (t2) [15,16].

The disappearing gradient is one of the primary issues with RNNs. Any neural network's weights are adjusted during the training phase by back-propagating through the network and computing the error. But, with an RNN, it is more difficult since we must propagate these neurons across time [17] With the help of LSTM.

I can solve this issue (Long Short-Term Memory). The most recent recurrent neural network to solve the vanishing gradient issue is the LSTMA diagram of a flowchart

Description automatically generated with low confidence

Fig 11. LSTM

A cell vector value is preserved at each stage in the LSTM architecture.

With LSTM, a clear gating procedure is applied. Three binary gates—the input gate (it), forget gate (ft), and output gate—make up an LSTM (ot). The memory cell in-process update is controlled by the input gate, the memory cell's reset to zero is controlled by the forget gate, and the memory cell's information flow visibility output is controlled by the output gate [17].

𝑓𝑡 = σ (𝑊𝑓. [ℎ𝑡−1, 𝑥𝑡]) + 𝑏𝑓

A description of the parameters of the LSTM shown in down Table.

|  |  |
| --- | --- |
| Parameter | Description |
| wf | Weight vector of the forget gate layer |
| ℎ𝑡−1 | The previously hidden state vector |
| ℎ𝑡 | Output hidden state vector |
| 𝑥𝑡 | Current input vector |
| 𝑏𝑓 | Bias vector |
| 𝑖𝑡 | Current input vector |
| 𝑜𝑡 | Output vector |
| 𝑐𝑡 | Output cell memory vector |
| 𝑐𝑡-1 | Output cell memory vector |
| 𝑐𝑡 | Output cell memory vector |

Table 2. Parameters of the LSTM

The researcher concludes from the LSTM results that the word insertion strategy is ideal for aspect-based sentiment analysis. The suggested LSTM model categorized consumer product evaluating sentences as highly negative, negative, positive, and amazingly positive instead of categorizing them as positive and negative. Using training data, the LSTM model has an accuracy of 94%, while with test data, it has an accuracy of 93%.

The default LSTM parameters used as the models' parameters in this investigation. Training the size of the input embedding layer with 128 and 128 results in an analysis of the suggested model's performance. For input on three benchmark data sets, this model provides greater accuracy for the length of the 128 -word embedding vector. The memory unit is utilized to retain the words from the input used for the sentence evaluation. The suggested LSTM is built with 1192 memory units that can retain words to comprehend lengthy review paragraphs. The output layer chooses three nodes (positive, negative) to provide a sentiment score [18].

|  |  |  |
| --- | --- | --- |
| Layer(type) | Output Shape | Param |
| Embedding layer | (None, 280, 200) | 50580880 |
| Dropout layer | (None, 280, 200) | 0 |
| LSTM layer | (None, 128) | 209408 |
| Dense layer | (None, 2) | 258 |

Table 3. LSTM Sequential model

**3.4. Recommendation Systems:**

There are three techniques of recommendation systems: collaborative, content-based, and hybrid-based. The content-based method relies on the relationship between an item's description and the user's profile; item suggestions are made in accordance with the user's preferences. The collaborative method examines shared preferences across individuals by analyzing user behavior and preferences [19, 20]. The hybrid strategy combines the two approaches.

A rule-based recommendation system, also known as a content-based recommendation system, employs a set of predetermined rules to provide recommendations to users. Recommendation systems using rules can be straightforward and simple to use.

Based on well-defined rules created by the system designers. These guidelines might be founded on a variety of user or item data, including user demographics, item ratings, item genres, or other traits. For instance, a user may receive song recommendations via a rule-based recommendation system for a music streaming service based on the user's age, gender, and listening preferences.

The fact that rule-based recommendation systems are easy to comprehend, and use is one of its key advantages. Although they do not involve the use of sophisticated machine learning methods, they may also be quick. They may, however, be rigid and unable to adjust to shifting user preferences or newly added items in the suggestion database.

In general, rule-based recommendation systems can be a helpful tool for giving users recommendations, but they might not be as good at capturing the complexity of user preferences and the connections between distinct items in the recommendation database as other types of recommendation systems, like collaborative filtering or matrix factorization-based systems.

Sentiments According Sentiment values:

|  |  |
| --- | --- |
| Sentiment Range | Sentiment Level |
| -2 to -3 | Extreme negative |
| -0.5 to -2 | Negative |
| -0.5 to +0.5 | Neutral |
| +0.5 to +2 | Positive |
| +2 to +3 | Extreme positive |

Table 4. Sentiment Level

***Chapter 4***

**EXPERIMENTAL EVALUATION**

**4.1 Machine learning models results:**

The focus of this research is to detect depression by analyzing the selected user comments. I begin by running the text classification techniques on the whole dimension feature space collected from the dataset. For n-gram modeling I used IF-IDF features and Put into five different classification model Decision Tree, Random Forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Nave Bayes. Support Vector Machine (SVM) gives the best test accuracy value 93%. And lowest accuracy gives Nave Bayes 60.80%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | RF | DT | KNN | NB | SVM |
| Accuracy | 89.5 | 83.7 | 60.8 | 84.1 | 93 |
| Precision | 0.90 | 0.84 | 0.72 | 0.85 | 0.94 |
| Recall | 0.90 | 0.84 | 0.61 | 0.84 | 0.94 |
| F1-Score | 0.90 | 0.84 | 0.55 | 0.84 | 0.94 |

Table 5. Machine learning i) Accuracy ii) Precision iii) Recall iv) F1 Score

**4.2 LSTM models results:**

At this stage, I consider a variety of classification techniques to build the prediction model and determine the chance that a user would experience depression. I used 80% of the dataset for training, while the remaining 20% was used for testing. The collection is divided into sections based on social media sessions, and each section contains all the tweets from that session. Every tweet is assigned to one of two classes: depressed or non-depressive. Long-term memory and short-term memory are the classifiers utilized in the mode's development (LSTM). The LSTM model's parameters include an embedding layer, a dropout layer, an LSTM layer, and a dense layer with a SoftMax activation function Parameters using LSTM model.

For evaluating the performance of the above-mentioned approaches, we used the following evaluation metrics: i) Accuracy ii) Precision iii) Recall iv) F1 Score [12].

i) Accuracy = TP + TN / TP + TN + FP +FN

ii) Precision = TP / TP + FP

iii) Recall = TP / TP + FN

iv) F1 Score = 2\*{(recall \* Precision) / \*(recall + Precision)}

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 44768 | 1428 |
| 1 | 2460 | 44162 |

Table 6. Confusion Matrix

The experiments evaluated by using test data, we got 94% accuracy. For evaluating the performance of the above-mentioned approaches, we used the following evaluation metrics: i) Accuracy ii) Precision iii) Recall iv) F1 Score [12].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| 0 | 0.94 | 0.96 | 0.95 | 23202 |
| 1 | 0.95 | 0.94 | 0.94 | 23206 |
| Micro avg | 0.95 | 0.95 | 0.95 | 46408 |
| Macro avg | 0.95 | 0.95 | 0.95 | 46408 |
| Weighted avg | 0.95 | 0.95 | 0.95 | 46408 |
| Samples avg | 0.95 | 0.95 | 0.95 | 46408 |

Table 7. LSTM : i) Accuracy ii) Precision iii) Recall iv) F1 Score

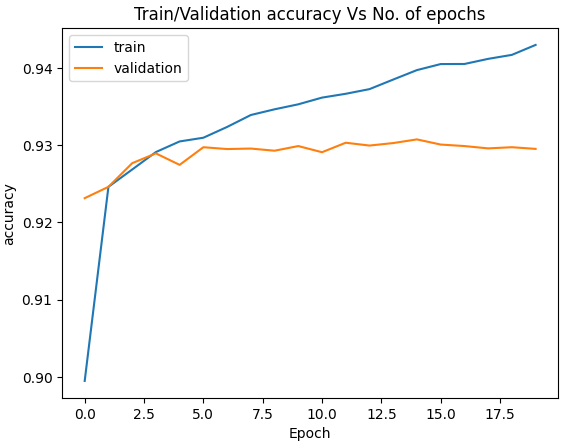
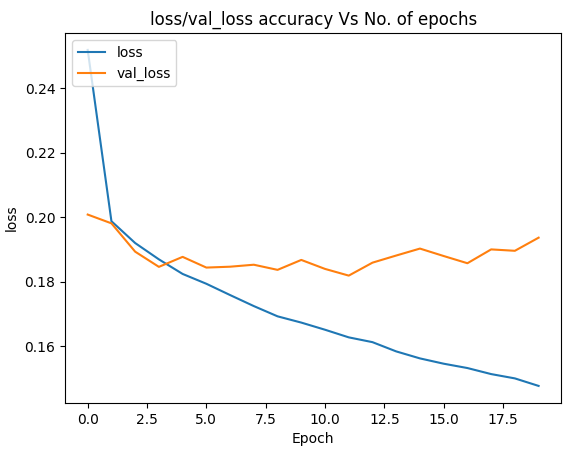


Fig 12. Train and Validation accuracy vs No of epochs

Fig 13. Train and Validation accuracy vs No of epochs

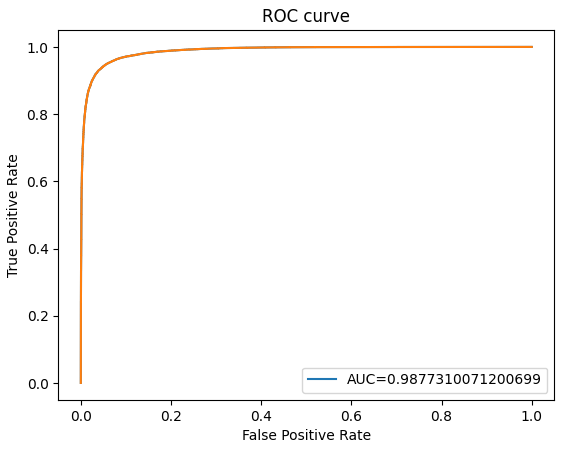
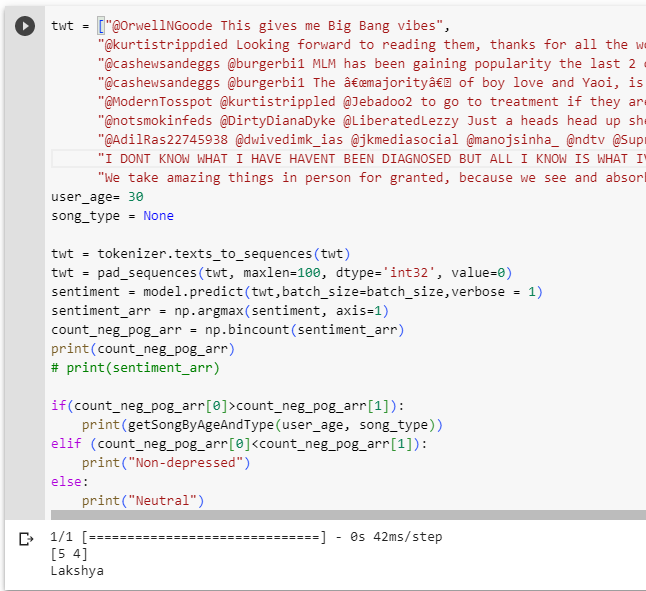
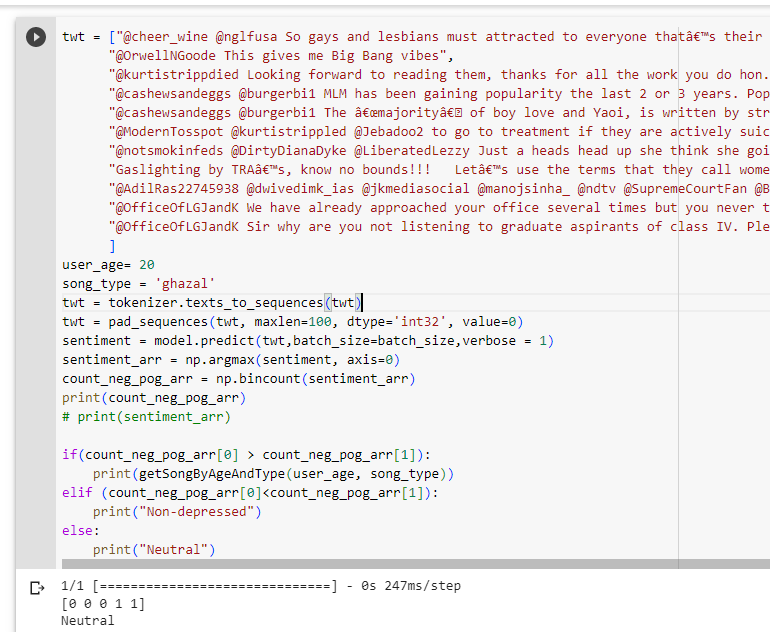
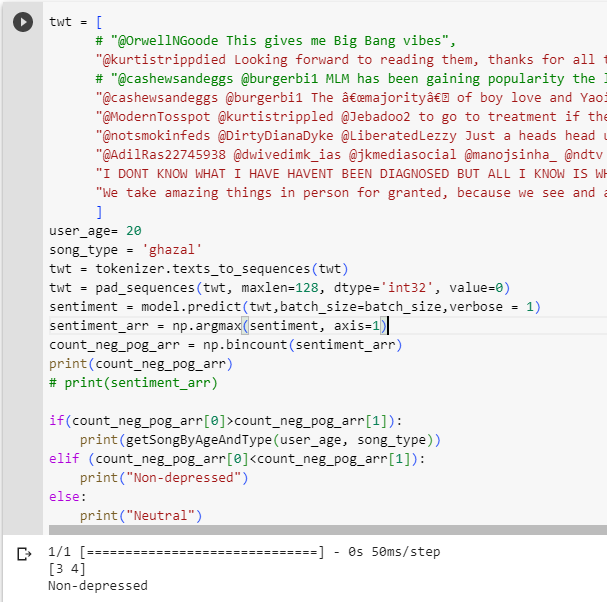


Fig 14. ROC and AUC curve

**4.3 Recommended System Results:**

****

****

***Chapter 5***

# CONCLUSION AND FUTURE RECOMMENDATIONS

In conclusion, sentiment analysis using long short-term memory (LSTM) models is a powerful and widely used technique for analyzing and understanding the sentiment of text data and user depression in individuals. And the recommended system gives a basis relaxion with music.

However, it is important to note that sentiment analysis and recommended system is not a replacement for professional diagnosis and treatment. It is recommended that individuals who may be experiencing depression seek the help of a qualified healthcare professional.

In future I develop Glove and Bart Word Embedding model also move from Content-based recommended System to Collaborative recommended system for better result.

# CODE

# Mounting google drive

from google.colab import drive

drive.mount('/content/drive')

# Setting toolkit folder as working directory

%cd /content/drive/My Drive/Project/DL/New Data

! ls

import nltk

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

#imports

from sklearn.feature\_extraction.text import CountVectorizer

from keras.preprocessing.text import Tokenizer

from keras.utils import pad\_sequences

from keras.models import Sequential

from keras.layers import Dense, Embedding, LSTM, SpatialDropout1D

from sklearn.model\_selection import train\_test\_split

from keras.utils.np\_utils import to\_categorical

from sklearn.utils import resample

from sklearn.utils import shuffle

from sklearn.metrics import confusion\_matrix,classification\_report

import re

import random

import sys

import seaborn as sns

import matplotlib.pyplot as plt

#Data

data = pd.read\_csv("Suicide\_Detection.csv")

data.head()

sns.countplot(x = 'sentiment' , data = data)

#Data prepossessing

data['tweet'][1]

import nltk

from nltk.corpus import stopwords

nltk.download('stopwords')

stop\_words = set(stopwords.words("english"))

data['tweet']=data['tweet'].apply(str)

data['tweet'] = data['tweet'].str.lower()

data['tweet'] = data['tweet'].apply(lambda i : re.sub(r'https?://[^\s<>"]+|www\.[^\s<>"]+','', i))

data['tweet'] = data['tweet'].apply((lambda x: re.sub('[^a-zA-z0-9\s]','',x)))

data['tweet'] = data['tweet'].apply(lambda x : " ".join(t for t in x.split() if t not in stop\_words))

data['tweet'][1]

print(data[ data['sentiment'] == 'non-suicide'].size)

print(data[ data['sentiment'] == 'suicide'].size)

#Tokenizer

max\_fatures = 2000

tokenizer = Tokenizer(num\_words=max\_fatures, split=' ')

tokenizer.fit\_on\_texts(data['tweet'].values)

X = tokenizer.texts\_to\_sequences(data['tweet'].values)

X = pad\_sequences(X)

X[:2]

# DATASET

TRAIN\_SIZE = 0.8

# WORD2VEC

W2V\_SIZE = 200

W2V\_WINDOW = 7

W2V\_EPOCH = 30

W2V\_MIN\_COUNT = 10

# KERAS

SEQUENCE\_LENGTH = 280

EPOCHS = 20

BATCH\_SIZE = 1024

# SENTIMENT

POSITIVE = "POSITIVE"

NEGATIVE = "NEGATIVE"

NEUTRAL = "NEUTRAL"

# EXPORT

KERAS\_MODEL = "model.h5"

WORD2VEC\_MODEL = "model.w2v"

TOKENIZER\_MODEL = "tokenizer.pkl"

ENCODER\_MODEL = "encoder.pkl"

#Word2Vec Model

%%time

documents = [\_text.split() for \_text in train.tweet]

import gensim

w2v\_model = gensim.models.word2vec.Word2Vec(vector\_size=W2V\_SIZE,

                                            window=W2V\_WINDOW,

                                            min\_count=W2V\_MIN\_COUNT,

                                            workers=8)

w2v\_model.build\_vocab(documents)

words = w2v\_model.wv.vectors

print(len(words))

print(words)

%%time

w2v\_model.train(documents, total\_examples=len(documents), epochs=W2V\_EPOCH)

w2v\_model.wv.most\_similar("love")

%%time

tokenizer = Tokenizer()

tokenizer.fit\_on\_texts(train.tweet)

vocab\_size = len(tokenizer.word\_index) + 1

print("Total words", vocab\_size)

X\_train = tokenizer.texts\_to\_sequences(data['tweet'].values)

X\_train = pad\_sequences(X\_train,maxlen=280)

Y\_train = pd.get\_dummies(data['sentiment']).values

print('x\_train shape:',X\_train.shape)

X\_test = tokenizer.texts\_to\_sequences(test['tweet'].values)

X\_test = pad\_sequences(X\_test,maxlen=280)

Y\_test = pd.get\_dummies(test['sentiment']).values

print("x\_test shape", X\_test.shape)

embedding\_matrix = np.zeros((vocab\_size, W2V\_SIZE))

for word, i in tokenizer.word\_index.items():

  if word in w2v\_model.wv:

    embedding\_matrix[i] = w2v\_model.wv[word]

print(embedding\_matrix.shape)

embedding\_layer = Embedding(vocab\_size, W2V\_SIZE, weights=[embedding\_matrix], input\_length=SEQUENCE\_LENGTH)

# embed\_dim = 280

lstm\_out = 128

model = Sequential()

model.add(embedding\_layer)

model.add(SpatialDropout1D(0.4))

model.add(LSTM(lstm\_out, dropout=0.4, recurrent\_dropout=0.4))

model.add(Dense(2,activation='softmax'))

model.compile(loss = 'categorical\_crossentropy', optimizer='adam',metrics = ['accuracy'])

print(model.summary())

batch\_size = 1024

history = model.fit(X\_train, Y\_train, epochs = 20, batch\_size=batch\_size, verbose = 1,  validation\_split=0.2)

model.evaluate(X\_test, Y\_test)

def display\_plot(history,train,validation):

    plt.plot(history.history[train])

    plt.plot(history.history[validation])

    plt.title('Train/Validation accuracy Vs No. of epochs')

    plt.ylabel(train)

    plt.xlabel('Epoch')

    plt.legend(['train', 'validation'], loc='upper left')

    plt.show()

def display\_plot(history,loss,val\_loss):

    plt.plot(history.history[loss])

    plt.plot(history.history[val\_loss])

    plt.title('loss/val\_loss accuracy Vs No. of epochs')

    plt.ylabel(loss)

    plt.xlabel('Epoch')

    plt.legend(['loss', 'val\_loss'], loc='upper left')

    plt.show()

mythreshold=0.5

from sklearn.metrics import confusion\_matrix

y\_pred = (model.predict(X\_test)>=mythreshold).astype(int)

cm=confusion\_matrix(

    Y\_test.argmax(axis=1), y\_pred.argmax(axis=1))

print(cm)

import seaborn as sns

sns.heatmap(cm, cmap="Blues")

plt.xlabel("Predicted labels")

plt.ylabel("True labels")

plt.title('HeatMap ')

plt.show()

from sklearn.metrics import roc\_curve, precision\_recall\_curve, roc\_curve

# thresholds=0.5

y\_preds = model.predict(X\_test)[::,1]

fpr, tpr, thresholdes = roc\_curve(Y\_test.argmax(axis=1), y\_preds)

plt.plot(fpr,tpr)

plt.ylabel('True Positive Rate')

plt.xlabel('False Positive Rate')

plt.title('ROC curve')

plt.show()

from sklearn.metrics import auc

import sklearn.metrics as metrics

y\_pred\_proba  = model.predict(X\_test)[::,1]

fpr, tpr, thresholdes = roc\_curve(Y\_test.argmax(axis=1), y\_preds)

auc = metrics.roc\_auc\_score(Y\_test.argmax(axis=1), y\_pred\_proba)

plt.plot(fpr,tpr,label="AUC="+str(auc))

plt.plot(fpr,tpr)

plt.ylabel('True Positive Rate')

plt.xlabel('False Positive Rate')

plt.title('ROC curve')

plt.legend(loc=4)

plt.show()

from sklearn.metrics import precision\_score, recall\_score, f1\_score, classification\_report

from sklearn.metrics import classification\_report

print (classification\_report(Y\_test, y\_pred))

song=pd.read\_csv('Song.csv')

# type(data.tweet)

song.head()

user=pd.read\_csv('suicidal-tendency-users.csv')

# type(data.tweet)

user.head()

def getPositiveSong(userId, songType=None):

    if(userId > 227): return 'Invalid user Id'

    if(songType==None):

      userAge=-1

      for index, value in enumerate(user.user\_id):

        if value == userId:

          userAge = user.age[index]

          break

      if(userAge <= 25): songType = 'bollywood'

      if(userAge > 25 & userAge <= 35): songType = 'classical'

      if(userAge > 35): songType = 'ghazal'

    isValidSongType = songType in ['bollywood', 'classical', 'ghazal']

    if(isValidSongType == False): return 'Invalid Song Type'

    songsIndexes = []

    for idx,value in enumerate(song.song\_type) :

        if(value == songType):

            songsIndexes.append(idx)

    return song.song\_name[random.choice(songsIndexes)]

getPositiveSong(4)

def getSongByAgeAndType(userAge, songType=None):

    if(userAge < 0): return 'Invalid User Age'

    if(songType==None):

      if(userAge <= 25): songType = 'bollywood'

      if(userAge > 25 & userAge <= 35): songType = 'classical'

      if(userAge > 35): songType = 'ghazal'

    isValidSongType = songType in ['bollywood', 'classical', 'ghazal']

    if(isValidSongType == False): return 'Invalid Song Type'

    songsIndexes = []

    for idx,value in enumerate(song.song\_type) :

        if(value == songType):

            songsIndexes.append(idx)

    return song.song\_name[random.choice(songsIndexes)]

getSongByAgeAndType(20, 'ghazal')

twt = ["@cheer\_wine @nglfusa So gays and lesbians must attracted to everyone thatâ€™s their sex? Why use buck anyway he has already stated quit using him as a gotcha. Everyone is just a fucking pawn to you people.",

      "@OrwellNGoode This gives me Big Bang vibes",

      "@kurtistrippdied Looking forward to reading them, thanks for all the work you do hon. I wish I could do more. I am just not the best in front of the camera.",

      "@cashewsandeggs @burgerbi1 MLM has been gaining popularity the last 2 or 3 years. Popularity with homosexuality has diminished so these other terms along side other gender woowoo speak.,"

      "@cashewsandeggs @burgerbi1 The â€œmajorityâ€ of boy love and Yaoi, is written by straight women. Of course it would be straight women who would fetishize it.",

      "@ModernTosspot @kurtistrippled @Jebadoo2 to go to treatment if they are actively suicidal it would be detrimental to treatment. Stabilization fund out why they want to be on hormones delve into the trauma. Get them to put down the weapons and find the self acceptance.",

    "@notsmokinfeds @DirtyDianaDyke @LiberatedLezzy Just a heads head up she think she going goofy hunting would keep an eye out on your page and might want to talk to her walk again about the way she being overtly homophobic",

      "Gaslighting by TRAâ€™s, know no bounds!!!   Letâ€™s use the terms that they call women  Uterus Haverâ€™s Vagina Havers Cis Women Pregnant People  Recognizing same sex attracted people for what they are isnâ€™t misogyny or bigotry.   Stop the Homophobia",

      "@AdilRas22745938 @dwivedimk\_ias @jkmediasocial @manojsinha\_ @ndtv @SupremeCourtFan @BJP4India @BJP4JnK @jwjnational @rashtrapatibhvn @VPSecretariat @INCIndia @BBCBreaking @LokSabhaSectt @LoPIndia @loksabhaspeaker @narendramodi @graduates Higher education is not a crime. We have two options left now 1) suicide 2) Class IV job",

      "@OfficeOfLGJandK We have already approached your office several times but you never tried to meet us. Please save the youth of kashmir. Graduates want to live. Graduate lives matter. We have not illegal demand, we are in merit list. Just consider graduatea for final selection. ðŸ™Please",

      "@OfficeOfLGJandK Sir why are you not listening to graduate aspirants of class IV. Please save us from suicide. We want to feed our family. Sir please help us. Please intervene. Our demand is not illegal. Higher education is not a crime. #justiceforgraduates"

      ]

user\_age= 20

song\_type = 'ghazal'

twt = tokenizer.texts\_to\_sequences(twt)

twt = pad\_sequences(twt, maxlen=100, dtype='int32', value=0)

sentiment = model.predict(twt,batch\_size=BATCH\_SIZE,verbose = 1)

sentiment\_arr = np.argmax(sentiment, axis=0)

count\_neg\_pog\_arr = np.bincount(sentiment\_arr)

print(count\_neg\_pog\_arr)

print(sentiment\_arr)

if(count\_neg\_pog\_arr[0]>count\_neg\_pog\_arr[1]):

    print(getSongByAgeAndType(user\_age, song\_type))

elif (count\_neg\_pog\_arr[0]<count\_neg\_pog\_arr[1]):

    print("Non-depressed")

else:

    print("Neutral")

twt = ["@OrwellNGoode This gives me Big Bang vibes",

      "@kurtistrippdied Looking forward to reading them, thanks for all the work you do hon. I wish I could do more. I am just not the best in front of the camera.",

      "@cashewsandeggs @burgerbi1 MLM has been gaining popularity the last 2 or 3 years. Popularity with homosexuality has diminished so these other terms along side other gender woowoo speak.",

      "@cashewsandeggs @burgerbi1 The â€œmajorityâ€ of boy love and Yaoi, is written by straight women. Of course it would be straight women who would fetishize it.",

      "@ModernTosspot @kurtistrippled @Jebadoo2 to go to treatment if they are actively suicidal it would be detrimental to treatment. Stabilization fund out why they want to be on hormones delve into the trauma. Get them to put down the weapons and find the self acceptance.",

      "@notsmokinfeds @DirtyDianaDyke @LiberatedLezzy Just a heads head up she think she going goofy hunting would keep an eye out on your page and might want to talk to her walk again about the way she being overtly homophobic",

      "@AdilRas22745938 @dwivedimk\_ias @jkmediasocial @manojsinha\_ @ndtv @SupremeCourtFan @BJP4India @BJP4JnK @jwjnational @rashtrapatibhvn @VPSecretariat @INCIndia @BBCBreaking @LokSabhaSectt @LoPIndia @loksabhaspeaker @narendramodi @graduates Higher education is not a crime. We have two options left now 1) suicide 2) Class IV job",

      "I DONT KNOW WHAT I HAVE HAVENT BEEN DIAGNOSED BUT ALL I KNOW IS WHAT IVE BEEN THROUGH. ALL I KNOW IS HOW TO HANDLE ALL MY SHIT AND LOVE OTHERS.   I'm so tired of feeling the need to verify my life, experiences, and purpose.    Do I deserve to gaslight my life?  I'm just me.",

      "We take amazing things in person for granted, because we see and absorb so much on the regular through our screens.  We seek human connection and validation through something so inhuman. Through computers.  Its no wonder why everyone's fill so disipointing. Get out! Live ya life hoe!"]

user\_age= 30

song\_type = None

twt = tokenizer.texts\_to\_sequences(twt)

twt = pad\_sequences(twt, maxlen=100, dtype='int32', value=0)

sentiment = model.predict(twt,batch\_size=BATCH\_SIZE,verbose = 1)

sentiment\_arr = np.argmax(sentiment, axis=1)

count\_neg\_pog\_arr = np.bincount(sentiment\_arr)

print(count\_neg\_pog\_arr)

# print(sentiment\_arr)

if(count\_neg\_pog\_arr[0]>count\_neg\_pog\_arr[1]):

    print(getSongByAgeAndType(user\_age, song\_type))

elif (count\_neg\_pog\_arr[0]<count\_neg\_pog\_arr[1]):

    print("Non-depressed")

else:

    print("Neutral")

twt = [

      "@kurtistrippdied Looking forward to reading them, thanks for all the work you do hon. I wish I could do more. I am just not the best in front of the camera.",

      "@cashewsandeggs @burgerbi1 The â€œmajorityâ€ of boy love and Yaoi, is written by straight women. Of course it would be straight women who would fetishize it.",

      "@ModernTosspot @kurtistrippled @Jebadoo2 to go to treatment if they are actively suicidal it would be detrimental to treatment. Stabilization fund out why they want to be on hormones delve into the trauma. Get them to put down the weapons and find the self acceptance.",

      "@notsmokinfeds @DirtyDianaDyke @LiberatedLezzy Just a heads head up she think she going goofy hunting would keep an eye out on your page and might want to talk to her walk again about the way she being overtly homophobic",

      "@AdilRas22745938 @dwivedimk\_ias @jkmediasocial @manojsinha\_ @ndtv @SupremeCourtFan @BJP4India @BJP4JnK @jwjnational @rashtrapatibhvn @VPSecretariat @INCIndia @BBCBreaking @LokSabhaSectt @LoPIndia @loksabhaspeaker @narendramodi @graduates Higher education is not a crime. We have two options left now 1) suicide 2) Class IV job",

      "I DONT KNOW WHAT I HAVE HAVENT BEEN DIAGNOSED BUT ALL I KNOW IS WHAT IVE BEEN THROUGH. ALL I KNOW IS HOW TO HANDLE ALL MY SHIT AND LOVE OTHERS.   I'm so tired of feeling the need to verify my life, experiences, and purpose.    Do I deserve to gaslight my life?  I'm just me.",

      "We take amazing things in person for granted, because we see and absorb so much on the regular through our screens.  We seek human connection and validation through something so inhuman. Through computers.  Its no wonder why everyone's fill so disipointing. Get out! Live ya life hoe!"

      ]

user\_age= 20

song\_type = 'ghazal'

twt = tokenizer.texts\_to\_sequences(twt)

twt = pad\_sequences(twt, maxlen=128, dtype='int32', value=0)

sentiment = model.predict(twt,batch\_size=1,verbose = 1)

sentiment\_arr = np.argmax(sentiment, axis=1)

count\_neg\_pog\_arr = np.bincount(sentiment\_arr)

print(count\_neg\_pog\_arr)

print(sentiment\_arr)

if(count\_neg\_pog\_arr[0]>count\_neg\_pog\_arr[1]):

    print(getSongByAgeAndType(user\_age, song\_type))

elif (count\_neg\_pog\_arr[0]<count\_neg\_pog\_arr[1]):

    print("Non-depressed")

else:

    print("Neutral")

# REFERENCES

1. Abdul Hasib Uddin, Durjoy Bapery, Abu Shamim Mohammad Arif, “Depression Analysis from Social Media Data in Bangla Language using Long Short-Term Memory (LSTM) Recurrent Neural Network Technique”, InternationalConferenceonComputer,Communication,Chemical,MaterialsandElectronicEngineering(IC4ME2),11-12July,2019
2. Ms.R.Monika, dr.S.Deivalakshmi, Dr.B.Janet, “Sentiment Analysis of US Airlines Tweets using LSTM/RNN,” 9th InternationalConference onAdvancedComputing 2019.
3. Wang, Xinyu, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. ”A depression detection model based on sentiment analysis in microblog social network.” In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 201-213. Springer, Berlin, Heidelberg, 2013.
4. YukunMa, HaiyunPeng, TahirKhan, ErikCambria1, AmirHussain, “SenticLSTM:aHybridNetworkforTargetedAspect-BasedSentiment Analysis”, Springer Science+BusinessMedia,LLC,partofSpringerNature2018.
5. Renata Lopes Rosa, Gisele Maria Schwartz, Wilson Vicente Ruggiero, and Dem´ostenes Zegarra Rodr´ıguez, “A Knowledge-Based Recommendation System That Includes Sentiment Analysis and Deep Learning,” IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 15, NO. 4, APRIL 2019.
6. Renata L. Rosa, Demóstenes Z. Rodríguez, and Graça Bressan, “Music Recommendation System Based on User’s Sentiments Extracted from Social Networks,” IEEE Transactions on Consumer Electronics, Vol. 61, No. 3, August 2015.
7. Sudhanshu Kumar, Kanjar De, and Partha Pratim Roy, “Movie Recommendation System Using Sentiment Analysis From Microblogging Data”, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 7, NO. 4, AUGUST 2020.
8. Ansari, L., Ji, S., Chen, Q., & Cambria, E. (2022). Ensemble hybrid learning methods for automated depression detection. IEEE Transactions on Computational Social Systems, 1–9. 10.1109/TCSS.2022.3154442.
9. Mahdikhani, M. (2022). Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic. International Journal of Information Management Data Insights, v2(1), Article 100053 ISSN 2667-0968. 10.1016/j.jjimei.2021.100053.
10. Zhiyuan Liu , Wei Xu, Wenping Zhang, Qiqi Jiang, “A music recommendation system based on psychotherapy,” 2023 The Authors. Published by Elsevier Ltd.
11. I. Om Prabha and G. U. Srikanth, “Survey of Sentiment Analysis Using Deep Learning Techniques.”
12. S. Cahyaningtyas, D. Hatta Fudholi, and A. Fathan Hidayatullah, “Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews,” Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, Aug. 2021, doi: 10.22219/kinetik.v6i3.1300.
13. W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, “Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis,” Mar. 2016, [Online].Available: <http://arxiv.org/abs/1603.06679>
14. L. Zhao and A. Zhao, “Sentiment analysis based requirement evolution prediction,” Future Internet, vol. 11, no. 2, 2019, doi: 10.3390/fi11020052.
15. M. Thomas and L. C.A, “Sentimental analysis using recurrent neural network,” Int. J. Eng. Technol., vol. 7, no. 2.27, p. 88, 2018.
16. P. Liu, X. Qiu, and H. Xuanjing, “Recurrent neural network for text classification with multitask learning,” IJCAI Int. Jt. Conf. Artif. Intell., vol. 2016–January, pp. 2873–2879, 2016.
17. A. Hassan and A. Mahmood, “Deep Learning approach for sentiment analysis of short texts,” 2017 3rd Int. Conf. Control. Autom. Robot. ICCAR 2017, pp. 705–710, 2017.
18. Diki Wahyudi, Yuliant Sibaroni, “Deep Learning for Multi-Aspect Sentiment Analysis of TikTok App using the RNN-LSTM Method,” Building of Informatics, Technology and Science (BITS) Volume 4, No 1, Juni 2022 Page: 169−177
19. D. Yang, and W. Lee, "Music emotion identification from lyrics," in Proc. IEEE International Symposium on Multimedia, San Diego, California, USA, pp. 624-629, Dec. 2009.
20. G. Qiu, F. Zhang, J. Bu, and C. Chen, “Domain specific opinion retrieval,” in Proc. Asia Information Retrieval Symposium on Information Retrieval Technology, Sapporo, Japan, pp. 318-329, Oct. 2009.
21. Lizhong Xiao, Guangzhong Wang, Yang Zuo, “Research on patent classification based on Word2vec and LSTM”, 2018 11th International Symposium on Computational Intelligence and Design.
22. [Sunil Ray](https://www.analyticsvidhya.com/blog/author/sunil-ray/), “Use Support Vector Machines (SVM) for Data Science**” Published On September 13, 2017**
23. [Shermain Puah](https://www.researchgate.net/profile/Shermain-Puah-2), “Predicting Students’ Academic Performance: A Comparison between Traditional MLR and Machine Learning Methods with PISA 2015” , January 2021.
24. Piyush Kumar, Poulomi Samanta, Suchandra Dutta, Moumita Chatterjee and Dhrubasish Sarkar, “Feature Based Depression Detection from Twitter Data Using Machine Learning Techniques” , Volume 66, Issue 2, 2022 Journal of Scientific Research of The Banaras Hindu University.
25. Moumita Chatterjee, Piyush Kumar, Poulomi Samanta, Dhrubasish Sarkar, “Suicide ideation detection from online social media: A multi-modal feature based technique” International Journal of Information Management Data Insights 2022.