A

Major Project

On

SENTENCE SENTIMENTS USING MACHINE LEARNING WITH DATA ANALYSIS

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

Lakshmi Dharani Palanki(187R1A05H3)

Rayi Supriya (187R1A05H5)

V Aishwarya (187R1A05H8)

Under the Guidance of

Dr. M. Malyadri

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NIRF, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401.

2018-22.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**SENTENCE SENTIMENTS USING MACHINE LEARNING WITH DATA ANALYSIS**" being submitted by **Lakshmi Dharani Palanki(187R1A05H3), Rayi Sypriya(187R1A05H5) & V Aishwarya(187R1A05H8)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under our guidance and supervision during the year 2021-22.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. M. Malyadri
Associate Professor
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held on _____

ACKNOWLEGDEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project

We take this opportunity to express my profound gratitude and deep regard to my guide Dr. M. MALYADRI, Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing.help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark. We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) Mr. A. Uday Kiran, Mr. J. Narasimharao, Dr. T. S. Mastan Rao, Mrs. G. Latha, Mr. A. Kiran Kumar, for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to Dr. K. Srujan Raju, Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully

We are obliged to Dr. A. Raji Reddy, Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sn. Ch. Gopal Reddy, Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of CMR Technical Campus who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

LAKSHMI DHARANI PALANKI (187R1A05H3)

RAYI SUPRIYA (187R1A05H5)

V AISHWARYA (187R1A05H8)

TABLE OF CONTENTS

ABSTRACT

Social networking sites have millions of people who share their thoughts day by day as posts/sentences. To identify the best approach for sentiment of the sentence, we have to verify the precision and accuracy of various algorithms. Sentiment Analysis is an area of text data mining and NLP which has an immense impact on socio-economic activities. The research of sentiment analysis of sentence data can be performed in different aspects using various algorithms. This project shows sentiment analysis types and techniques used to perform extraction of sentiment from sentences. In this survey paper, we have taken comparative study of different machine learning techniques and approaches of sentiment analysis having twitter as data.

LIST OF FIGURES

LIST OF SCREENSHOTS

1. INTRODUCTION

1.1 PROJECT SCOPE

With the current outbreak of Covid19, there has been a huge hurling of various negative comments for a particular community. There are also posts that are forwarded without any proper supervision which can be disrespectful and violating. There are many algorithms present that can be used for NLP and Sentence Sentiment Analysis, it is important to study and analyze them to provide the best results. Our primary goal while developing the project was to apply the learnings to battle the harmful texts and create a healthy environment for everyone.

1.2 PROJECT PURPOSE

There has been a huge spike in the usage of social networking sites like Twitter, Facebook, and YouTube in the recent years. The area of sentiment analysis is known as opinion mining; it is under the umbrella of computational linguistics and data mining. Its main aim is to detect the person's mood, behavior and opinion from text documents. With the expanded use of social networking sites, sentiment analysis techniques have started to use these sites' public data to do sentiment analysis studies in different sociological areas, such as politics, sociology, economy and finance. Most of the data that is available in social networks is unstructured. Such unstructured data is almost 80% of the data all over the world. This makes it difficult to analyze and gain valuable judgment from such data. Sentiment analysis or opinion mining is the important technique, which helps in detecting opinions of people on social media data.

1.3 PROJECT FEATURES

Opinions of others can be important when it is needed to make a decision. When those decisions involve valuable resources people think about their companions' past experiences. Now a day's social media gives new tools to conveniently share ideas with peoples linked to the World Wide Web. Though sentiment analysis concentrates on polarity detection (positive, negative or neutral). Twitter is a micro blogging site which contains a large number of short length utilities for marketing, social networking. For example, political parties might be eager to know whether people constructs, aspect level straightforwardly gives attention to the opinion or sentiment. It is based on the concept that an opinion resides of an attitude and a destination of opinion. support their curriculum or not. In the present scenario the need to gather opinions from social networking sites and draw conclusions that what people like or dislike, has been the most important perspective. The objective of this review paper is to discuss concept of sentiment analysis of twitter tweet.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

The System Analysis generally consists of three phases:

- Problem definition
- Existing system that describes about the current scenario and what are its drawbacks.
- Proposed system that explains the advantages of selecting the approach outlined in the project.

2.1 PROBLEM DEFNITION

Social media sites are becoming flooded with posts from users. These posts contain all sorts of meanings and can be harmful and spread hatred to the end users. To filter out the negative impacting reviews, there is a need for an analysis on the emotions and intentions of the posts.Since it is not feasible and labor consuming to always use an individual, implementing a machine learning algorithm is the best use case to solve this issue. Sentiment analysis for the posts will lead to a healthy content generation for the users.This allows us to understand that if a chuck oof text is positive or negative or its neutral using the natural language processing and the machine learning.

2.2 EXISTING SYSTEM

The social networking sites like Twitter, Facebook, and YouTube have gained so much popularity nowadays. The area of sentiment analysis is known as opinion mining, it is under the umbrella of computational linguistics and data mining. Its main aim is to detect the person's mood, behavior and opinion from text documents. The algorithms used in Sentence Sentiment Analysis are not selected with care either due to the lack of skill or ignorance of them.

Most of the data that is available in social networks is unstructured. Such unstructured data is almost 80% of the data all over the world. This makes it difficult to analyze and gain valuable judgment from such data.

2.2.1 Disadvantages

- Unstructured data makes it hard to process and operate on the data
- Difficult to analyze data using the older algorithms
- Difficult to gain valuable judgment from data which is not fit
- Low accuracy rates due to improper study and analysis of data.

2.3 PROPOSED SYSTEM

The proposed system studies and analyzes the algorithms to help us in selecting the algorithm with the higher accuracy value. This would help in improving the results of the test dataset by helping the system choose the right approach.

The research of sentiment analysis of any data can be performed in different aspects. This paper shows sentiment analysis types and techniques used to perform extraction of sentiment from sentences. In this survey paper, we have taken comparative study of different machine learning techniques and approaches of sentiment analysis having the train and test data.

2.3.1 Advantages

- Overcome the disadvantages of the existing system
- NLP improves processing time and makes it faster to provide the results
- Usage of machine learning algorithms helps in understanding the approach
- The comparison proves to be a better approach to study all the available algorithms and derive insights regarding the accuracy.

2.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

2.4.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to
high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely.
depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2.5 HARDWARE AND SOFTWARE REQUIREMENTS

There are several basic requirements for this project to work appropriately categorized as the software requirements and the hardware requirements.

2.5.1 SOFTWARE REQUIREMENTS

For developing the application the following are the Software Requirements:

- Python

  Operating Systems supported Windows 7 Windows XP Windows 8

  Debugger and Emulator Any Browser (Particularly Chrome)

2.5.2 HARDWARE REQUIREMENTS

For developing the application the following are the Hardware Requirements:

Processor: Pentium IV or higher

RAM: 2 GB

Space on Hard Disk: minimum 512MB

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

An Architectural diagram is a graphical representation of a set of concepts, that are part representation of a set of concepts, that are part of an architecture, including their principles, elements and components.

In our system architecture diagram, we have components as Load data, Process data, Train data, Test Data and Generate results. Our system flow includes the action of the user who login and load the required data set in the software and process the raw data taken to convert it into a well structured format. Train the data set by providing the target or keywords on which the data segregation or classification would occur.

The result generated using different algorithms in the training module will be tested in the test results phrase. The accuracy of the result is compared and noted down for different algorithms when the different inputs are given. Then the actual input is given and the required outputs are observed.

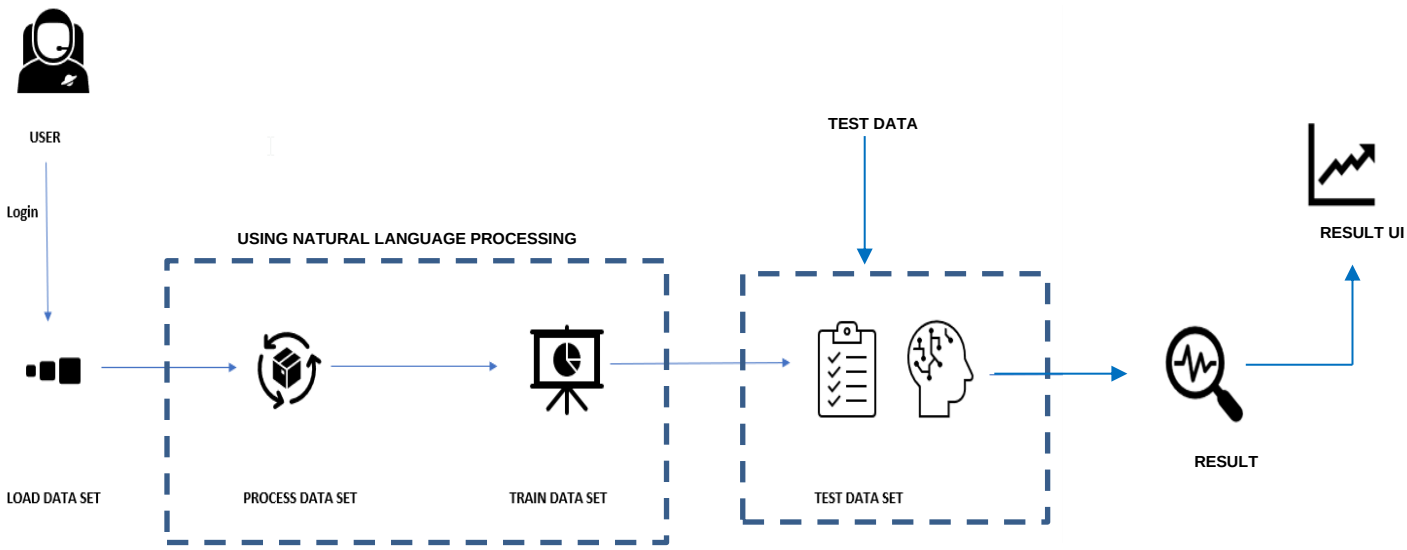

Figure 3.1 System Architecture

3.2 IMPLEMENTATION MODULES

The system has 5 main modules that are used recurrently for the different algorithms.They are:-

- **Load Data**

  The data presented to the module in the form of raw input. The input can be either a structured, semi-structured or unstructured form of data.

- **Process Data**

  The preprocessing is the main part of any Machine Learning process. The data needs to be processed and made ready into a form that can be used by the algorithm. Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

- **Training Data**

  In this phase the data sets are trained with a number of keywords to classify the sentiment of a particular sentence. The training set is the material through which the computer learns how to process information. The training data is absolutely essential to the process – it can be thought of as the "food" the system uses to operate.

- **Test Data**

  The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. If the test set does contain examples from the training set, it will be difficult to assess whether the algorithm has learned to generalize from the training set or has simply memorized it.

- **Generate Results**

  Results obtained are the sentiment classification of the sentence. We receive output in the form of a confusion matrix and the accuracy of each algorithm is displayed as output.

  The other values that are created as output of the process are precision, recall and F-score that are extremely important for an ML algorithm.

  We do the following testing for other algorithms and find the accuracy, precision, recall and FScore values that are displayed as Output.

3.3 UML DIAGRAMS

3.3.1 CLASS DIAGRAM

Class diagrams show the static structure of classifiers in a system. It provides a basic notation for other structure diagrams prescribed by UML which are helpful for developers.

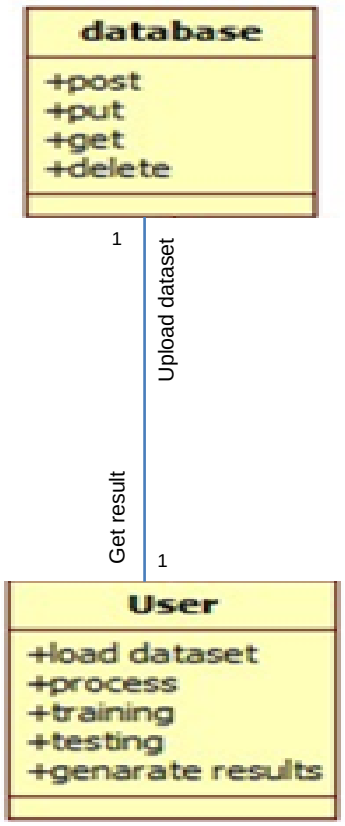

Figure 3.3.1 Class diagram

3.3.2 USE CASE DIAGRAM

Use case diagrams are typically developed in the early stage of development and people often apply use case modeling for the following purposes:

➔ Specify the context of a system

➔ Capture the requirements of a system

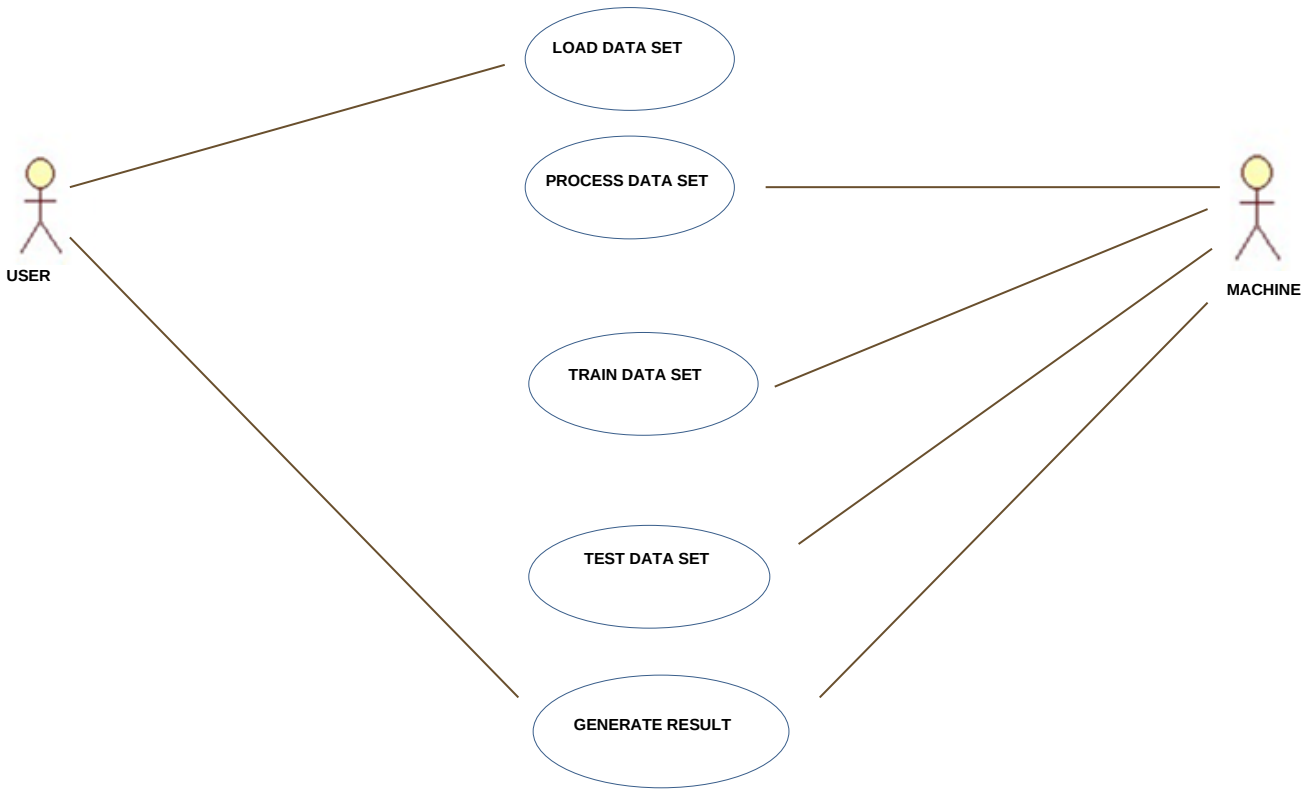

Figure 3.3.2 Use Case Diagram

3.3.3 SEQUENCE DIAGRAM

Sequence Diagrams captures the interaction that takes place in a collaboration that either realizes a use case or an operation (instance diagrams or generic diagrams)

➔ Model high-level interaction between active objects in a system

➔ Model the interaction between object instances within a collaboration that realizes a use case

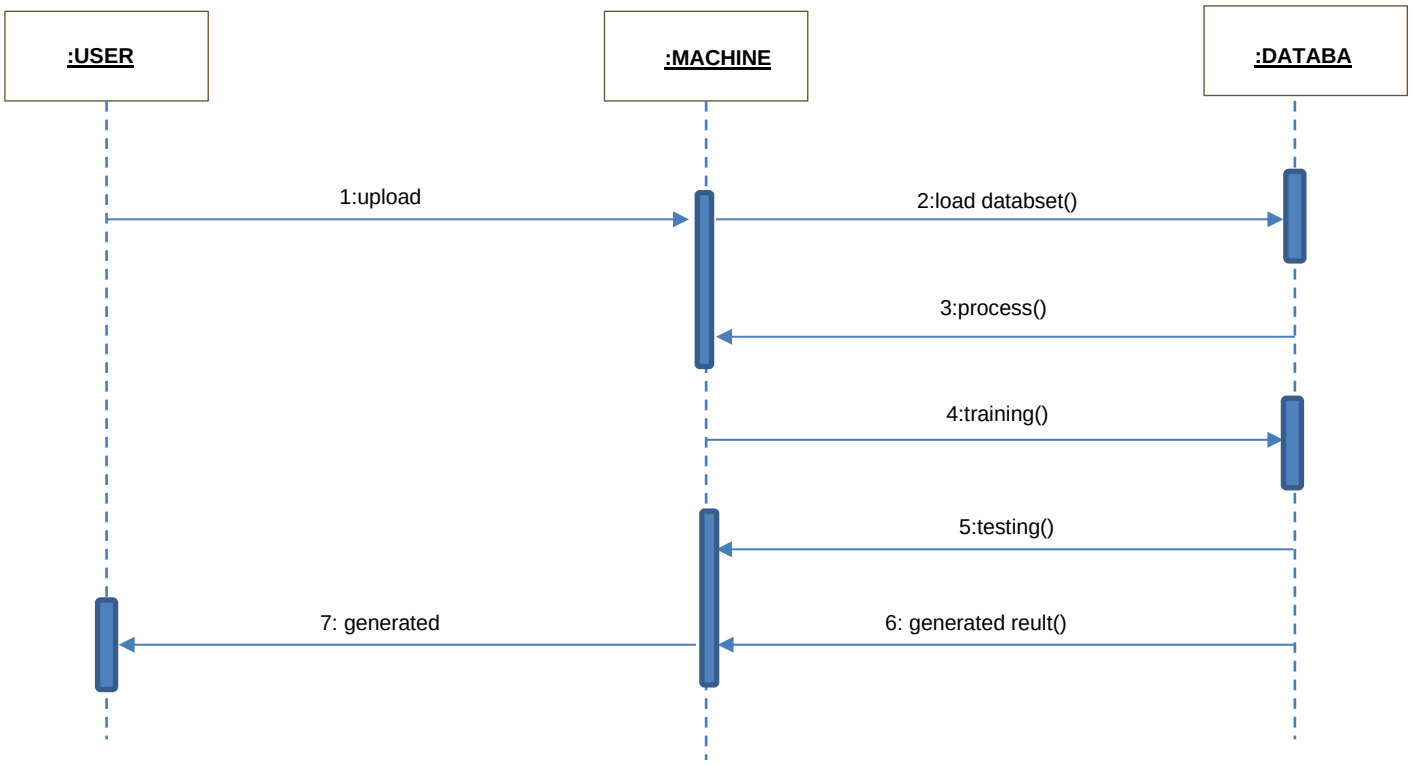

Figure 3.3.3 Sequence Diagram

3.3.4 ACTIVITY DIAGRAM

Activity Diagrams describe how activities are coordinated to provide a service which can be at different levels of abstraction.

➔ Model workflows between/within use cases

➔ Model complex workflows in operations on objects

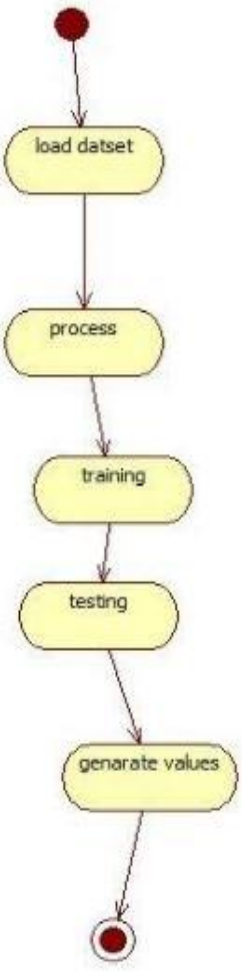

Figure 3.3.4 Activity Diagram

4. IMPLEMENTATION

4. IMPLEMENTATION

4.1 ALGORITHM

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation.

NLP is an advanced algorithm used for the sentiment understanding of documents.It includes two major steps.They are:

- **Training :** In this phase the entire data set is trained with a certain number of keywords so as to classify the sentiment or emotion behind the sentence.
- **Prediction :** In this phase the trained data sets are used for the prediction of the sentiment behind the sentence that is given as an example.

For this project, we have used the following Machine Learning Algorithms to analyze and the compare the results:

- Logistic Regression
- Support Vector Machine
- KNearestNeighbor
- Decision Tree
- Random Forest Classifier

4.1.1. LOGISTIC REGRESSION

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model.

For example,

- To predict whether an email is spam (1) or (0)
- Whether the tumor is malignant (1) or not (0)
- To predict if the sentiment of the sentence is Positive (1) or Negative (0)

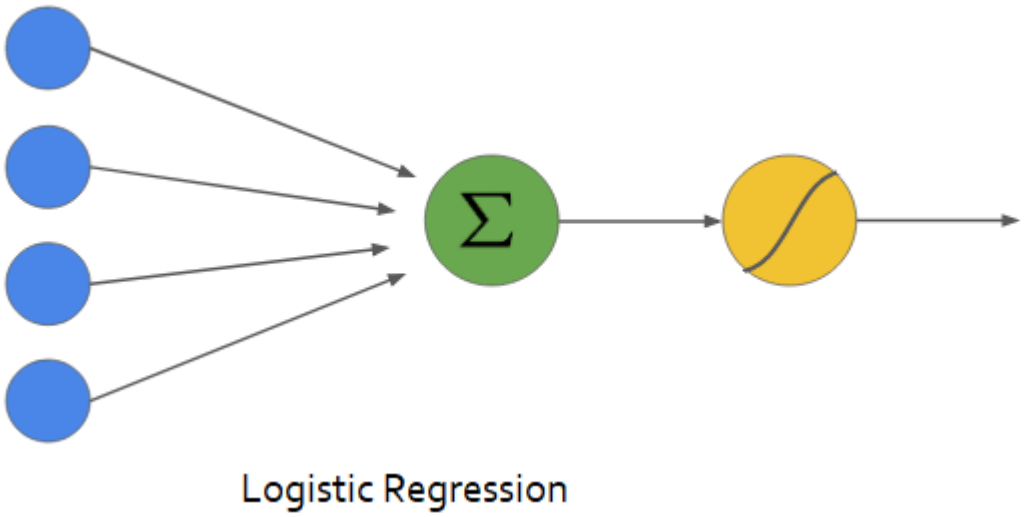

Figure 4.1.1 Logistic regression

4.1.2. SUPPORT VECTOR MACHINE

Support vector machines are highly preferred by many as it produces significant accuracy with less computation power. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks. But, it is widely used in classification objectives. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

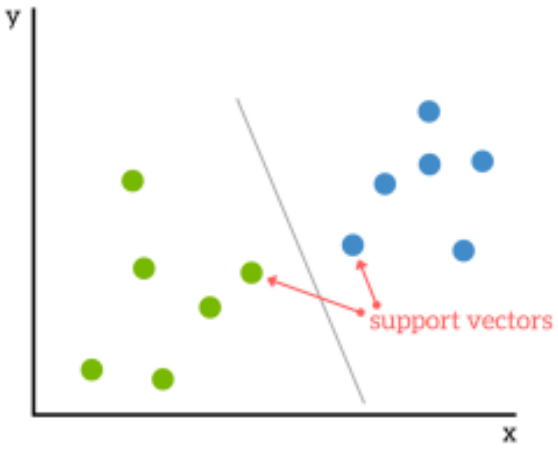

Figure 4.1.2 Support vector machines

4.1.3 K-NEAREST NEIGHBOR

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. The K in the KNN stands for the number of nearest neighbors that are to be considered for evaluation and classification.

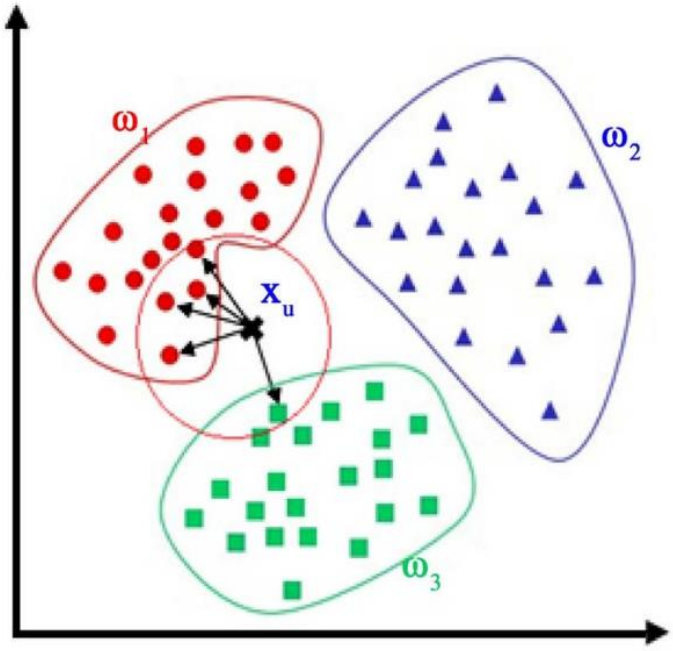

Figure 4.1.3 K-Nearest Neighbors

4.1.4. DECISION TREE

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as the learning decision tree from data and the tree is called the Classification tree as the target is to classify passengers as survived or died. Regression trees are represented in the same manner, just they predict continuous values like the price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

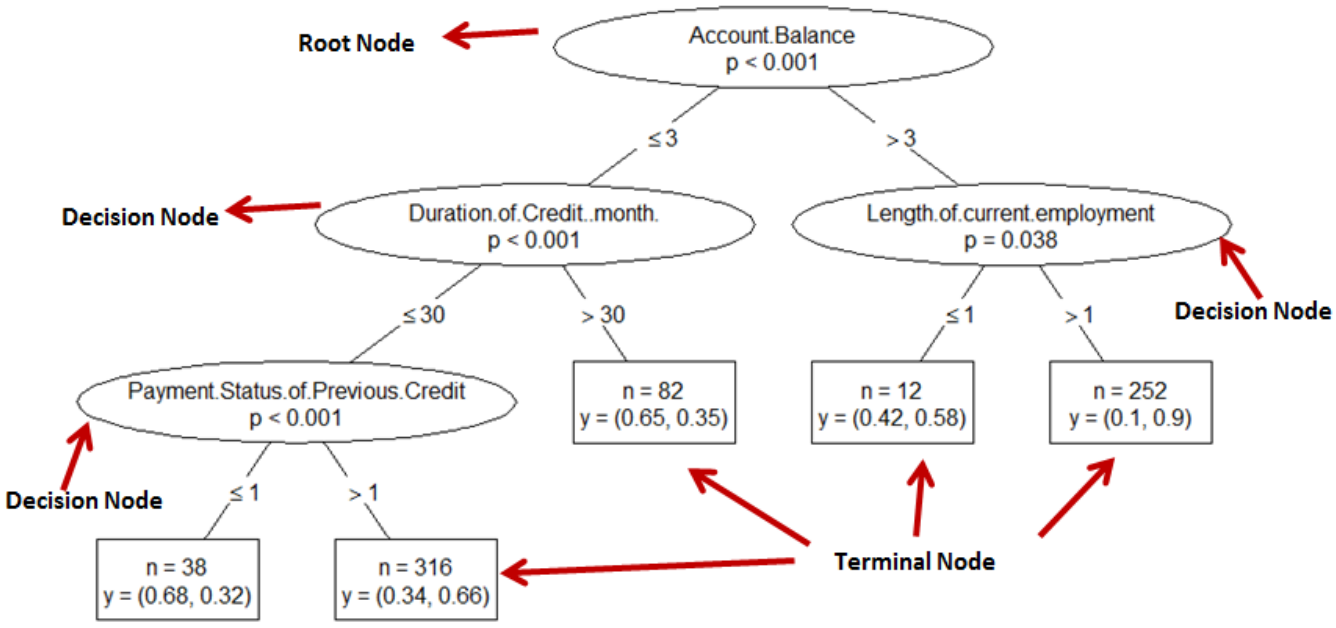

Figure 4.1.4 Decision Tree

4.1.5. RANDOM FOREST CLASSIFIER

It is an ensemble tree-based learning algorithm. The Random Forest Classifier is a set of decision trees from a randomly selected subset of training set. It aggregates the votes from different decision trees to decide the final class of the test object.

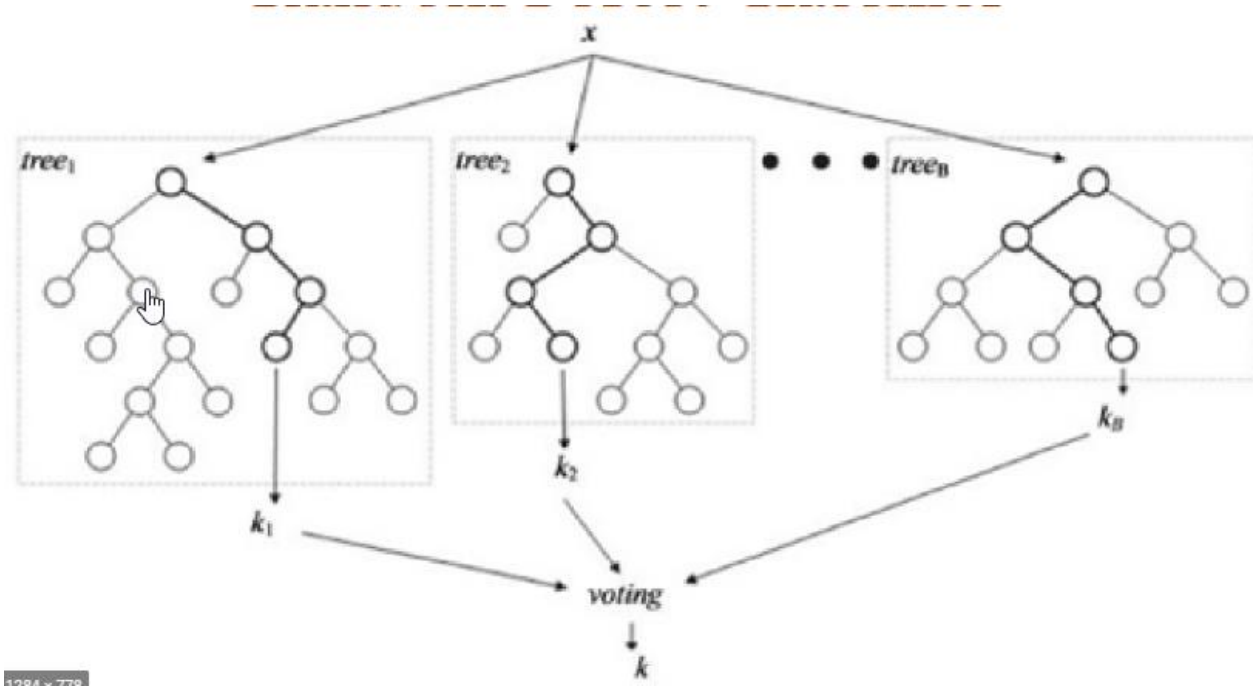

Figure 4.1.5 Random Forest Classifier

4.2 LANGUAGE USED - PYTHON

Python is a general purpose, dynamic, high level, and interpreted programming language. It supports an Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures.

Python is easy to learn yet powerful and versatile scripting language, which makes it attractive for Application Development. Python's syntax and dynamic typing with its interpreted nature make it an ideal language for scripting and rapid application development.Python supports multiple programming patterns, including object-oriented, imperative, and functional or procedural programming styles.

General purposes of the python specifications are as follows:

Web Applications : We can use Python to develop web applications. It provides libraries to handle internet protocols such as HTML and XML, JSON, Email processing, request, beautifulSoup, Feedparser etc.

It also provides Frameworks such as Django, Pyramid, Flask etc to design and develop web based applications. Some important developments are: PythonWikiEngines, Pocoo, PythonBlogSoftware etc.

Desktop GUI Applications : Python provides a Tk GUI library to develop user interfaces in python based applications. Some other useful toolkits wxWidgets, Kivy, pyqt that are usable on several platforms. The Kivy is popular for writing multitouch applications.

Software Development : Python is helpful for the software development process. It works as a support language and can be used for build control and management, testing etc.

Business Applications : Python is used to build Business applications like ERP and e-commerce systems. Tryton is a high level application platform

4.3 SAMPLE CODE

LOAD DATA

```
with open('Cell_Phones_and_Accessories_5.json') as json_data:
    d = json.load(json_data)
```

PROCESS DATA

```
#cleaning unwanted symbols
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
comment_dict = defaultdict(list)
for i in range(len(dataset)):
    sentence = re.sub('[^a-zA-Z.]',' ',dataset['reviewText'][i])
    sentence = sentence.lower()
    sentence = sentence.split('.')
    for k in range(len(sentence)):
        review = sentence[k].split()
        review = [word for word in review if not word in set(stopwords.words('english'))]
        sentence[k] =  ' '.join(review)
        comment_dict[i].append(sentence[k])
#delete unwanted " words
for j in range(len(comment_dict)):
    comment_dict[j] = [comment_dict[j][i] for i in range(len(comment_dict[j])) if comment_dict[j][i] not
in '']
for i in range(len(comment_dict)):
    reviewText[i] = ('. '.join(comment_dict[i][j] for j in range(len(comment_dict[i]))))

# spelling correction
for i in range(len(reviewText)):
    b = TextBlob(reviewText[i])
    reviewText[i] = b.correct()
dataset_corrected= DataFrame({'reviewerID': reviewerID, 'productID': productID, 'liked_and_seen':
liked_and_seen, 'reviewText': reviewText, 'rating': rating, 'summary': summary, 'unixTime': unixTime,
'date': date})
```

TRAINING DATA

```
# Part-of-speech Tagging
pos_dict = defaultdict(list)
for i in corpus_key:
    for j in range(len(corpus_list[i])):
        text = TextBlob(corpus_list[i][j])
        text = text.tags
        pos_dict[i].append(text)
pattern1 =  defaultdict(list)
for i in pos_dict_key:
    for j in range(len(pos_dict[i])):
        if(len(pos_dict[i][j]) == 2):
            if((pos_dict[i][j][0][1] == 'JJ' and pos_dict[i][j][1][1] == 'NN') or
            (pos_dict[i][j][0][1] == 'JJ' and pos_dict[i][j][1][1] == 'NNS')):
                #pattern1
      pattern1[i].append(pos_dict[i][j])
Finding Maximum scored words.
max_abs_score = defaultdict(list)
for i in range(len(tuple_word_score)):
    maxx = 0
    for j in range(len(tuple_word_score[i])):
        temp = abs(tuple_word_score[i][j][1])
        if(temp > maxx):
            maxx = abs(tuple_word_score[i][j][1])
            maxx_word = tuple_word_score[i][j][0]
    max_abs_score[i].append((maxx_word, maxx))
# Collecting important opinion words.
list_max_abs_score = []
for i in range(len(max_abs_score)):
    list_max_abs_score.append(max_abs_score[i])
```

TEST DATA

```
# Making the Confusion Matrix
y = comparision_dataframe['Calculated rating']
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(rev_rate, y)
accuracy = 0
for i in range(len(cm)):
   for j in range(len(cm)):
if(i == j):
        accuracy = accuracy + cm[i][j]
accuracy = accuracy/ len(rev_rate) * 100
# Calculating precision, recall and FScore values
precision = TP/(TP+FP)
recall = TP/(TP+FN)
FScore = 2*(recall * precision) / (recall + precision)
```

GENERATE RESULTS

```
# Fitting Logistic Regression to the Training set
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)

classifier.fit(X_train, y_train)
# Predicting the Test set results
y_pred = classifier.predict(X_test)

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
```

5. SCREENSHOTS

5. SCREENSHOTS

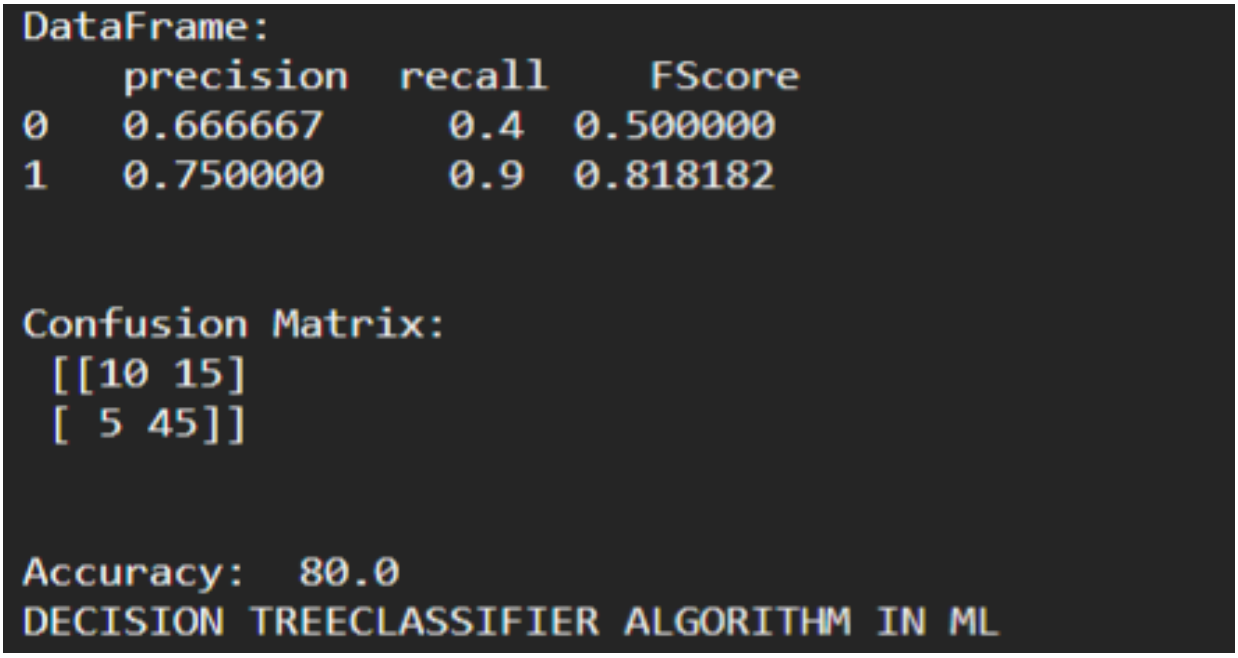

Fig 5.1 Decision Tree Classifier Output

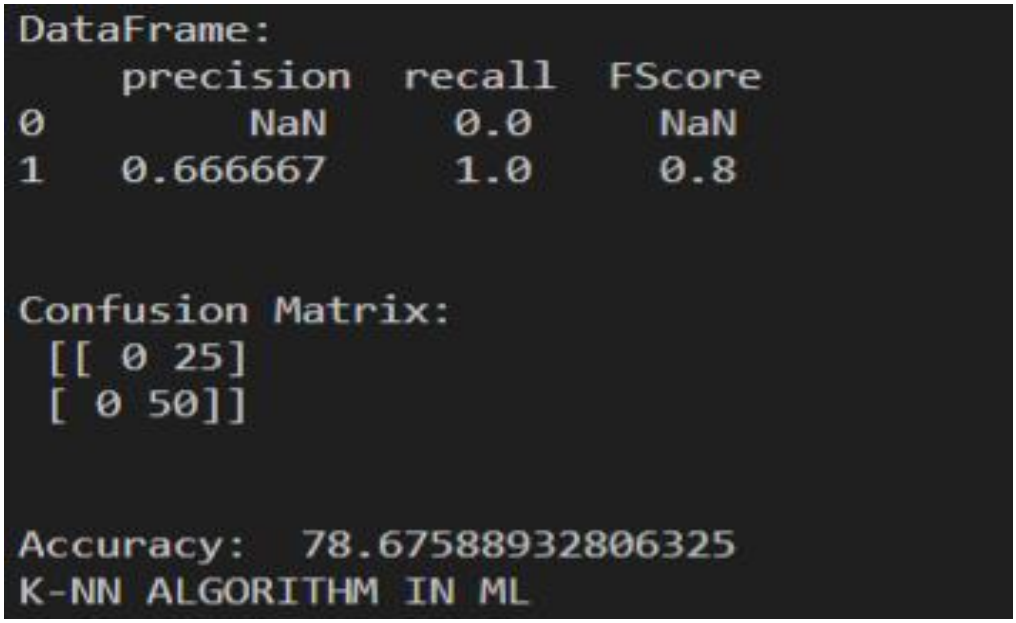

Figure 5.2 KNN Algorithm Output

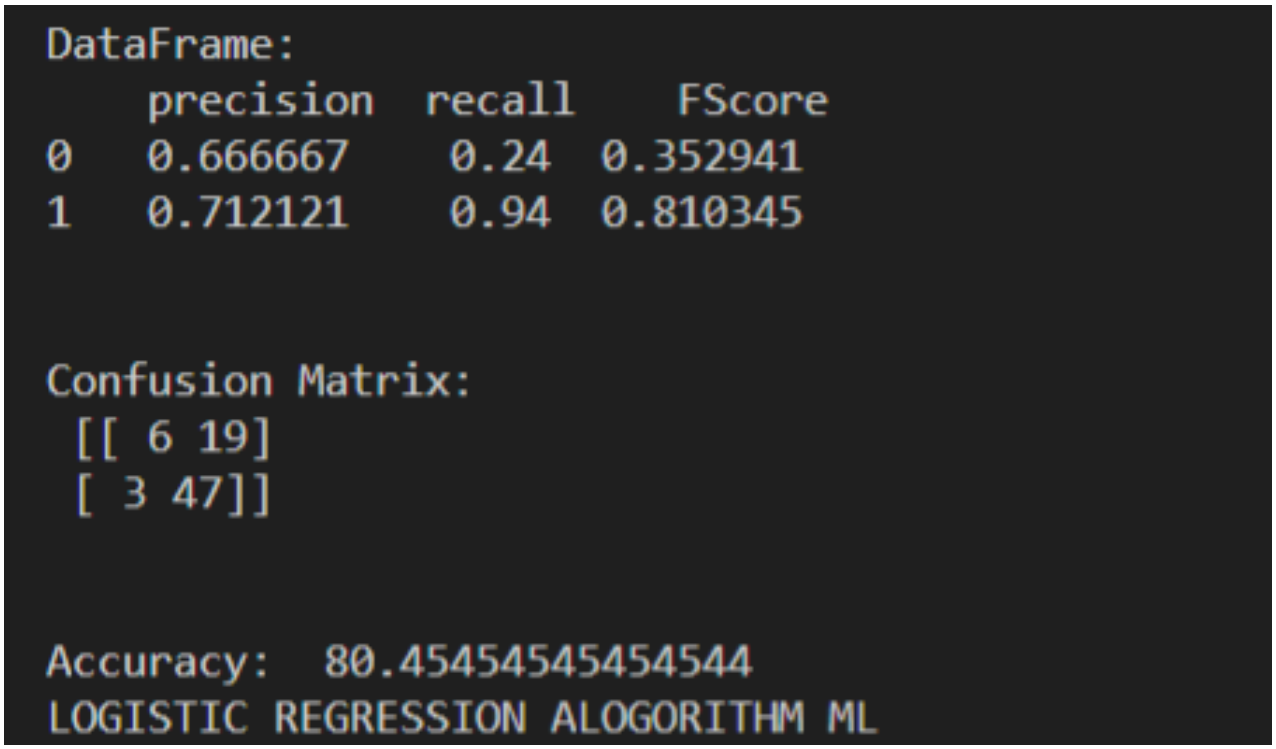

Fig 5.3 Logistic Regression

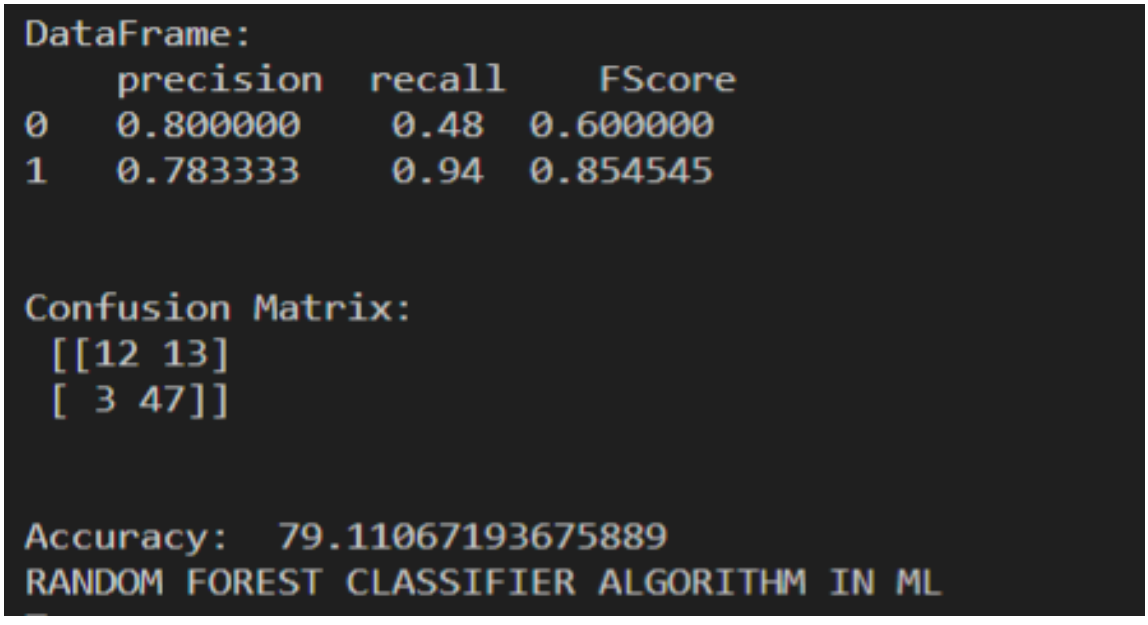

Figure 5.4 Random Forest Classifier Output

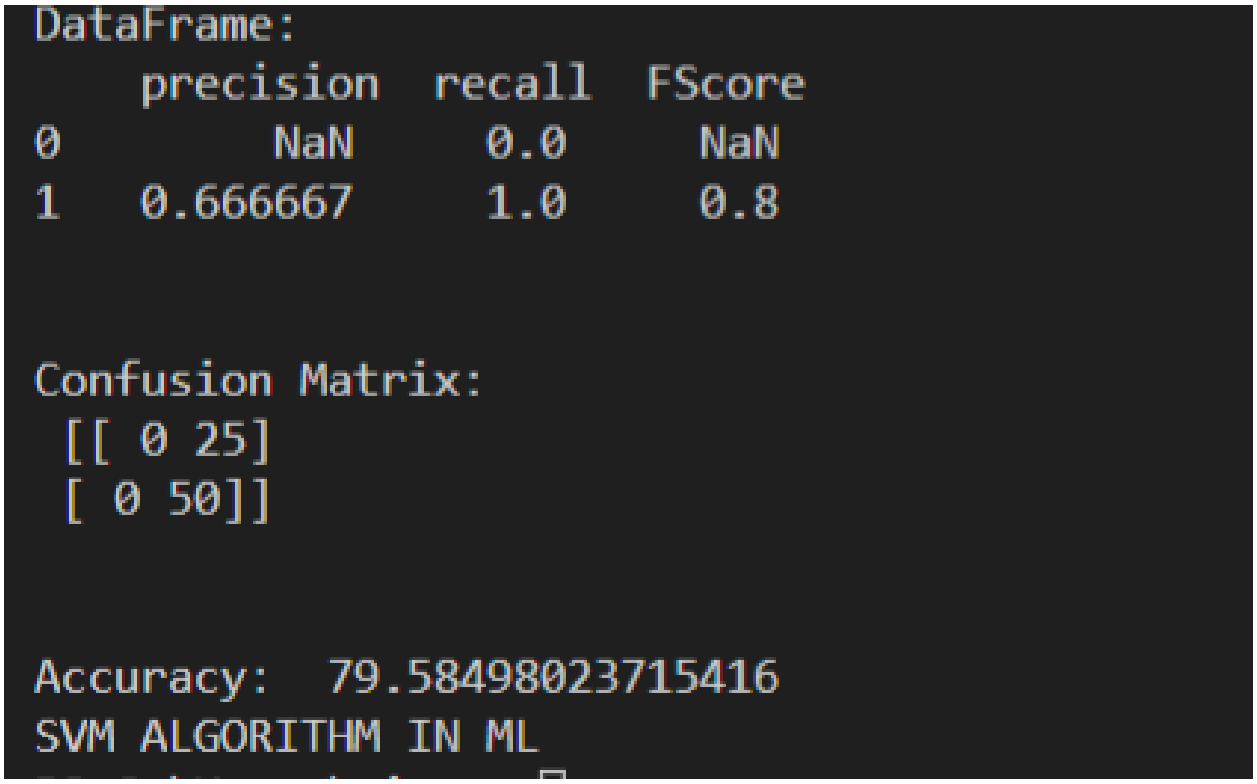

Figure 5.5 SVM Algorithm Output

6. TESTING

6. TESTING

6.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

6.2 TYPES OF TESTING

6.2.1 UNIT TEST

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.2.2 INTEGRATION TEST

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfied, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6.2.3 FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

6.2.4 SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

6.2.5 WHITE BOX TESTING

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purposeful. It is used to test areas that cannot be reached from a black box level.

6.2.6 BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated as a black box.

You cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

6.2.7 TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

6.3 TEST CASES

A test case is a set of conditions or variables under which a tester will determine whether a system under test satisfies requirements or works correctly. The process of developing test cases can also help find problems in the requirements or design of an application.

Test template followed to describe the results of this project are mentioned below:

Test Case ID - The ID of the test case.
Test Case - Test case name
Expected Result - The expected result of the test.
Actual Result - The actual result of the test; to be filled after executing the test.
Result - Any comments on the test case or test execution.

Test Case ID	Test Case	Expected Result	Actual Result	Result
6.2.1	Select an excel file for loading	The program should throw an error	The program throws an error	Pass
6.2.2	Print the value of the dataframe	The dataframe should have record of data	The dataframe has the record of the data	Pass
6.2.3	Print the first 300 records of dataframe	Prints the 300 rows	Prints the 300 rows	Pass
6.2.4	Pass a number as test value	Should throw a Value Error	Throws a ValueError	Pass
6.2.5	Verify the output of the program	The program should display the confusion matrix and accuracy values for each algorithm	The program displays the confusion matrix and accuracy values for each algorithm	Pass

7. CONCLUSION

7.1 CONCLUSION

The analysis of sentences and other forms of textual data is being done in different points of view to mine the opinion or sentiment. This project report defined the concept of sentiment analysis and opinion mining with respect to various levels of sentiment analysis. We then looked at the various Machine Learning algorithms that are compared to find the best algorithm for the test dataset. This report discussed different techniques of sentiment analysis and methodology for sentiment analysis. The idea of Sentiment Analysis is used in many areas of the modern world and has been impacting the way data is presented over the internet. Our project helped to identify a portion of the Sentence Sentiment analysis and outlined the process that needs to be followed.

There are various applications of this project which can enhance the way algorithms are selected and used. Based on the study, we can derive insights and suggestions for the analysis of Sentence Sentiments. Especially, with the rise in social media usage during the lockdown periods all over the world, the developers of the social networking sites can deploy more useful algorithms.

7.2 FUTURE SCOPE

The future of NLP is very vivid. There are innumerably large applications of the algorithm in various fields of industries and applications. This project has the potential to select the best algorithm for a Sentence Sentiment application and allows the user to view the accuracy percentage. This project can further be combined with various Speech Recognition or Image Recognition algorithms to apply this for a wider community of people. According to many market statistics, data volume is doubling every two years, but in future this time span may get further reduced. The vast portion of this data (about 79 percent) is text data. Natural Language Processing (NLP) is the sub-branch of Data Science that attempts to extract insights from “text.” Thus, NLP is assuming an important role in Data Science. Industry experts have predicted that the demand for NLP experts will grow exponentially in the near future.

With the recent improvements in the world of hardware, the GPU and TPU, there are many applications for all the Machine Learning and AI fields. With the exponential growth of multi-channel data like social or mobile data, businesses need solid technologies in place to assess and evaluate customer sentiments. So far, businesses have been happy analyzing customer actions, but in the current competitive climate, that type of customer analytics is outdated.

Now businesses need to analyze and understand customer attitudes, preferences, and even moods – all of which come under the purview of sentiment analytics. Without NLP, business owners would be seriously handicapped in conducting even the most basic sentiment analytics.

8. BIBILOGRAPHY

8.1 REFERENCE

[1] Surnar, Avinash, and Sunil Sonawane. "Review for Twitter Sentiment Analysis Using Various Methods."

[2] Eliacik, Alpaslan Burak, and Erdoğan Erdoğan. "User-weighted sentiment analysis for the financial community on Twitter." *Innovations in Information Technology (IIT), 2015 11th International Conference on*. IEEE, 2015.

[3] Ahmed, Khaled, Neamat El Tazi, and Ahmad Hany Hossny. "Sentiment Analysis over Social Networks: An Overview." *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015.

[4] Ko, Youngjoong, and Jung Yun Seo. "Automatic text categorization by unsupervised learning." *Proceedings of the 18th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 2000.

[5] Kharche, S. R., and Lokesh Bijole. "Review on Sentiment Analysis of Twitter Data." International Journal of Computer Science and Applications 8 (2015).

[6] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012): 1-167.

[7] Singh, Prabhsimran, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government." ICT Express (2017).

[8] Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data." Journal of Big Data 2.1 (2015): 5.

[9] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014.

[10] Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013.

[11] Amolik, Akshay, et al. "Twitter sentiment analysis of movie reviews using machine learning techniques." International Journal of Engineering and Technology 7.6 (2016): 1-7.

8.2 GITHUB LINK

https://github.com/dharanipalanki/SENTENCE-SENTIMENT-USING-MACHINE-LEARNING-WITH-DATA-ANALYSIS.git