Project Report

On

# Twitter Sentiment Analysis

Submitted impartial fulfilment for the requirement for the Award of degree of Master of Computer Applications

By

Shashwat Vashimkar (182120003)

**Under the Guidance of**

Dr. Sanjay Sharma



**Session 2018-2021**

**Department of Mathematics & Computer Applications**

**Maulana Azad National Institute of Technology Bhopal (MP)**

**April 2021**



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 1

**Declaration**

I, hereby declare that the work presented in this project entitled "Twitter Sentiment Analysis" presented in partial fulfilment for the award of the degree of Master of

Computer Applications submitted in the Department Of Mathematics and Computer Applications, Maulana Azad National Institute of Technology, Bhopal is an authentic work

carried out from 1st January 2021 to 5th May 2021 under the guidance of

**Dr. Sanjay Sharma,** MANIT Bhopal.

The matter embodied in this project has not been submitted by me or anybody else to any institution for award of any other degree or diploma.



Signature of Student

**Counter Signed by:**

**Dr. Sanjay Sharma**

Supervisor:

**Dr. Madhavi Shakya**

Head, Department Of Mathematics and Computer Applications MANIT, Bhopal.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 2

**Acknowledgement**

Here, I gladly present this project report on “Twitter Sentiment Analysis” as part of

the 6th semester Master in Computer Applications. I take this occasion to thank God, almighty for blessing me with his grace and taking our endeavor to a successful culmination. I extend my sincere and heartfelt thanks to me esteemed guide, Dr. Sanjay Sharma for providing me with the right guidance and advice at the crucial junctures and for showing me the right Way. I extend my sincere thanks to my respected head of the department Dr. Madhvi Shakya, for allowing me to use the facilities available.

I am highly indebted to-

**Dr. Sujoy Das**

**Dr. R S Thakur**

**Dr. G S Thakur**

**Dr. Amit Baghat**

**Dr. Vishnu Priya**

for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. Last but not the least, I would like to express my gratitude towards my parents & my friends for the support and encouragement they have given me during the course of my work.

Submitted By -

**Shashwat Vashimkar (182120003)**



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 3

**Certificate**

**2018-2021**

**DEPARTMENT OF MATHEMATICS AND COMPUTER APPLICATIONS**



**MAULANA AZAD NATION INSTITUTE OF TECHNOLOGY BHOPAL**

**(M.P.) – 462003**

This is to certify that Shashwat Vashimkar (182120003) have carried out the project work. In this report entitled **“Twitter Sentiment Analysis”** for the award of Master of Computer Application in Maulana Azad National Institute of Technology Bhopal

(M.P).

This report is the record of the candidate’s own work carried out by them under our supervision and guidance. This project work is the part of their Master in Computer Applications in Information Technology curriculum.

Their performance was excellent and we wish them good luck for their future endeavors.

**Signature of Project Guide** **Signature of Head of Department**



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 4

**Abstract**

Twitter is a popular social networking website where members create and interact with messages known as “tweets”. This serves as a mean for individuals to express their thoughts or feelings about different subjects. Various different parties such as consumers and marketers have done sentiment analysis on such tweets to gather insights into products or to conduct market analysis.

Furthermore, with the recent advancements in machine learning algorithms, we are able improve the accuracy of our sentiment analysis predictions.

Maulana Azad National Institute Of Technology Bhopal (M.P) Page 5

Maulana Azad National Institute Of Technology Bhopal (M.P) Page 6

**TABLE OF CONTENTS**

1. Introduction 09

1.1 Objective ................................................................................................................

1.2 Statement of the problem ........................................................................................

1.3 Solution to solve this problem .................................................................................

1.4 Machine learning approach in Sentiment Analysis.....................................................

2. Algorithm 10-12

2.1.1 Huge Datasets (Pandas Data Frame). ..............................................................

2.1.2 Supervised Learning ......................................................................

2.1.3 Count Vectorizer ..........................................................................

2.1.4 Tf-Idf Vectorizer .................................................................

3. Analysis Technique or Flow chart 13

4. Working Flow of Model 14

5. Technologies Used. 15

6. System Analysis 16

6.1. Requirement Analysis 17

6.1.1 Configuration 18

6.2. Let's Explore the Project 19

6.3 Data Pre- processing 24

6.3.1 Why data processing .....................................................................

6.3.2 Steps involved in Data processing .................................................

7. Software Engineering Paradigms 26

8. Analysis Document 27

8.1 Software Requirement Specification 28

9. Design 29

10. Code Efficiency 30

11. Code Optimization 31

12. Testing 32

13. Implementation 34

14. Evaluation 37

15. Maintenance 38

16. Conclusion and Recommendations. 40

17. Bibliography 42



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 7

**Introduction**



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 8

* 1. Objective

The Objective of our project is to predict the sentiment of tweets given

by the members, our model predict the sentiment of the tweets to be

Positive or Negative.

Technical Objective

The technical objectives will be implemented in Python. The system

must be able to access a list of tweets. It must predict the sentiment

of the tweets to be Positive or Negative.

Experimental Objective

Two Version of the prediction system will be implemented, Count Vectorizer

and Tf-Idf Vectorizer. The experimental objective will be to compare the Count

Vectorizer result with Tf-Idf Vectorizer result. We will test and evaluate both

the systems with the same test data to find their prediction accuracy.

Solution to this problem

We will implement the system using two different machine learning techniques

1. Count Vectorizer
2. TF-IDF(Term Frequency Inverse Document Frequency)

We will train the system using 80% of member’s tweets data and then test our

model to check the accuracy of the model using the remaining 20% of historic

data. Our solution uses a different algorithm and different techniques to perform

the prediction.

Maulana Azad National Institute Of Technology Bhopal (M.P) Page 9

1. **Introduction:**

Sentiment Analysis Model is an important way to determine the sentiments of the member’s tweets regarding certain post.

Sentiment analysis is the process of detecting positive or negative sentiment in text. It’s often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically [analyzing customer feedback](https://monkeylearn.com/blog/customer-feedback-analysis/), such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers’ needs.

For example, using sentiment analysis to automatically analyze 4,000+ reviews about your product could help you discover if customers are happy about your pricing plans and customer service.

Maybe you want to gauge [brand sentiment](https://monkeylearn.com/blog/brand-sentiment/) on social media, in real time and over time, so you can detect disgruntled customers immediately and respond as soon as possible.

Some of most important libraries is mentioned below

* **Numpy**
* **Sklearn**
* **Panda**
* **Matplotlib**

**Some of mostly used important libraries:**

**NUMPY:-**

NumPy is the fundamental package for scientific computing with Python. It containsamong other things:

* a powerful N-dimensional array object
* sophisticated (broadcasting) functions
* tools for integrating C/C++ and Fortran code
* useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the [BSD license,](https://www.numpy.org/license.html#license) enabling reuse with few restrictions.

**Sklearn:-**

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon **NumPy, SciPy** and **Matplotlib**.

**Pandas:**

It is the most popular python library that is used for data analysis. It provides highly optimized performance with back-end source code is purely written in ***C*** or Python. We can analyze Data in Panda with

1. Series- Series is one dimensional(1-D) array defined in pandas that can be used to store any data type.
2. Data Frame-Data Frame is two-dimensional(2-D) data structure defined in pandas which consists of rows and columns.

**Matplotlib :-**

Matplotlib is a python 2D plotting library which produces publication quality figures in a variety

of hard copy formats. It is for data visualization, it can be used in python scripts you can generate

plots, histograms, power spectra, bar charts, pie charts, error charts etc.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 10

**Some Important MODEL used in this project:-**

* **Huge Data Set( Pandas DataFrame)**
* **Supervised Learning**
* **Count Vectorizer**
* **TFIDF Vectorizer**

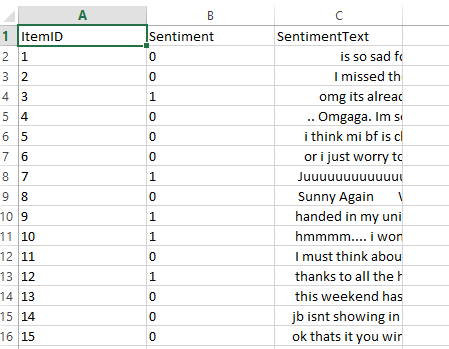
**Huge DataSet (Pandas DataFrame):-**

Pandas DataFrame is two-dimensional size- mutable, potentially heterogeneous tabular data structure with labelled axes(rows and columns). A data frame is a two-dimensional data structure, I,e.,data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data,rows, and columns.

First lets look at the dataset. It is tweets.csv.

It has 3 columns - "S.No.", “Sentiment” and "Sentiment text" and describes the sentiment for the tweets based on model.





Maulana Azad National Institute Of Technology Bhopal (M.P) Page 11

**Supervised Learning.**

**Training the system:**

Supervised learning is when the model is getting trained on a labelled dataset.

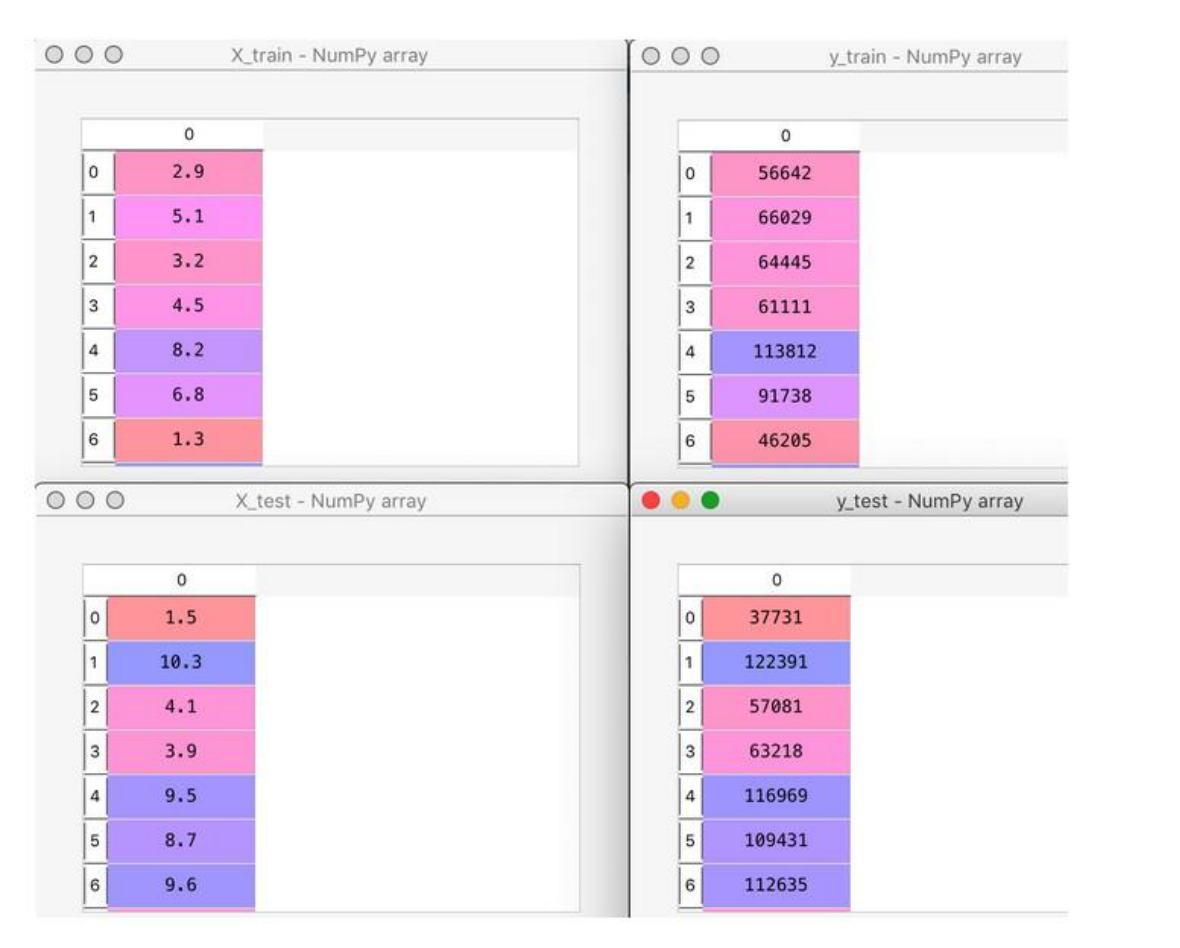
Labelled dataset is one which have both input and output parameters.

In this type of learning both training and validation datasets are labelled

While training the model, data is usually split in the ratio of 80:20 i.e. 80% as training data and rest as testing data. In training data, we feed input as well as output for 70% data. The model learns from training data only. We use different machine learning algorithms(which we will discuss in detail in next articles) to build our model. By learning, it means that the model will build some logic of its own.

Once the model is ready then it is good to be tested. At the time of testing, input is fed from remaining 20% data which the model has never seen before, the model will predict some value and we will compare it with actual output and calculate the accuracy.

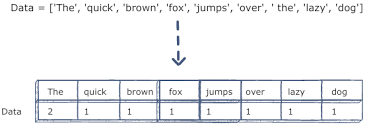
Below is the sample screenshot of X\_train, y\_train, X\_test and y\_test



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 12

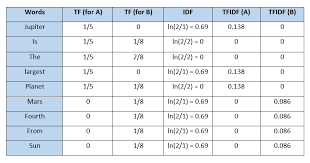
**Count Vectorizer**

Scikit-learn’s CountVectorizer is used to convert a collection of text documents to a vector of term/token counts. It also enables the ​pre-processing of text data prior to generating the vector representation. This functionality makes it a highly flexible feature representation module for text.



**Tf-Idf Vectorizer**

TF-IDF is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. This is done by multiplying two metrics: how many times a word appears in a document, and the inverse document frequency of the word across a set of documents. It has many uses, most importantly in automated [text analysis](http://www.monkeylearn.com/text-analysis/), and is very useful for scoring words in machine learning algorithms for [Natural Language Processing](https://monkeylearn.com/blog/definitive-guide-natural-language-processing/) (NLP).TF-IDF (term frequency-inverse document frequency) was invented for document search and information retrieval. It works by increasing proportionally to the number of times a word appears in a document, but is offset by the number of documents that contain the word. So, words that are common in every document, such as this, what, and if, rank low even though they may appear many times, since they don’t mean much to that document in particular.

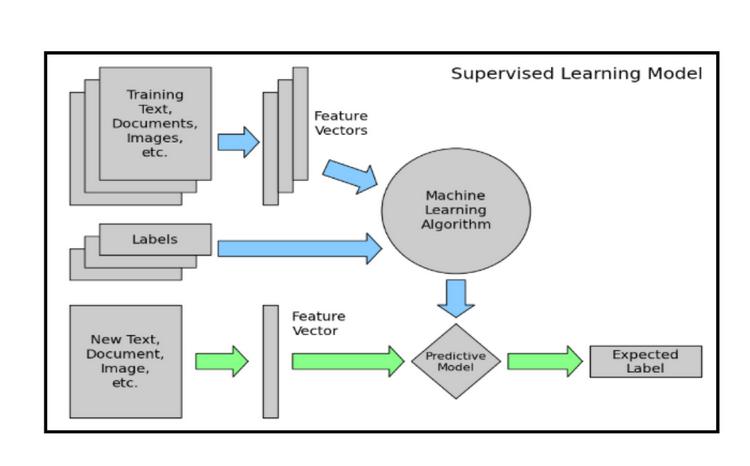


.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 13

**Working flow of Model.-**



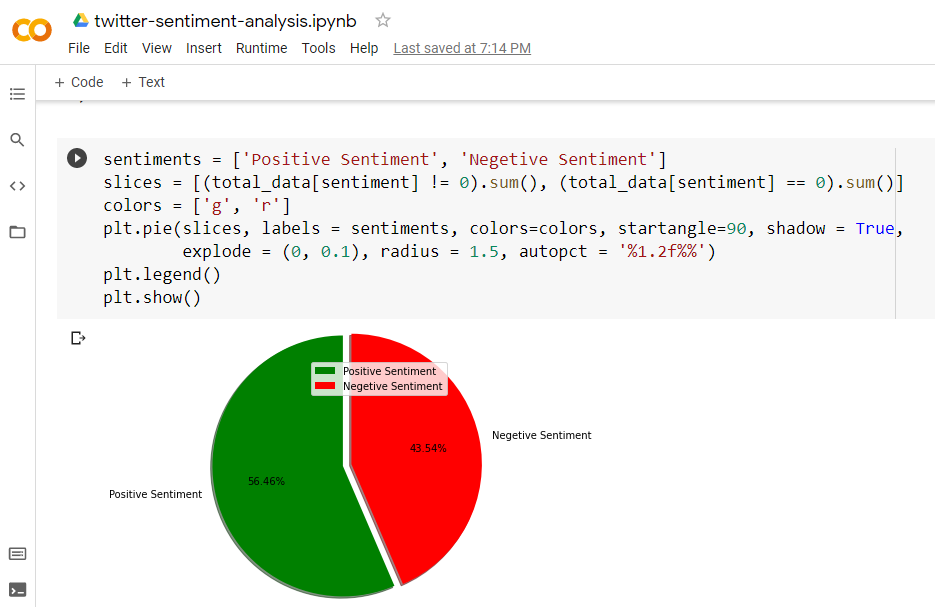
Maulana Azad National Institute Of Technology Bhopal (M.P) Page 14

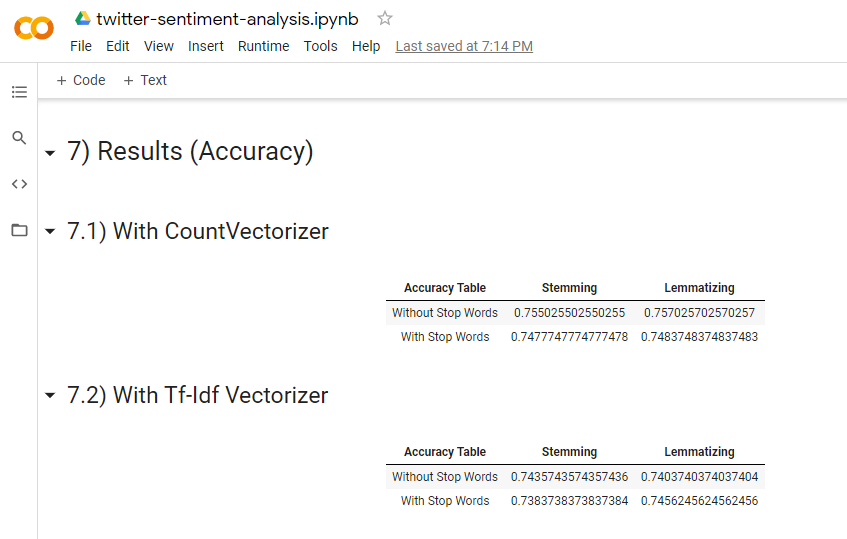
**Working of Model:**

The user can select the dataset of reviews and our code after pre-processing or cleaning

the dataset will train the model and fit the model on the algorithms used and will show the

output.





Maulana Azad National Institute Of Technology Bhopal (M.P) Page 15

**Project Scope:**

The project has a wide scope, as it is not intended to a particular organization. This project is going to develop generic software which can be applied by any education organization. Moreover it provides facility to record real time traffic data and on the base of that data manage the traffic signals dynamically.

**Proposed System:**

Establish a method of this system are easy to use and can make traffic system dynamic and real time with synchronized manner.

* 1. **Technologies Used:**
* **Languages: Python**
* **IDE: Google Colab**
* **Tools: Machine Learning Algorithms**
* **MS Excel / Csv\_files**

**ANALYSIS TOOLS:**

**Data Collection Tools:**

We have collected the huge DataSet of twitter reviews as a Training data.This collection helps us to understand the labeled data.Using this labeled data now we can predict the Sentiments of it.

**Charting Tools:**

The graphical representation of system & activities help us to understand theminor problems and the flow procedures, which can become very helpful in analysis as well as to generate the new system requirements.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 16

**Dictionary Tools**:

This tool help us to maintain & record the data & description of systemelement through Data items, processes & data stores. After the use of tools we have to do the analysis of a system through main factors.

**6.1 Requirement Analysis**

Requirement Analysis means studying or observing the current Business System to find how it works and where improvement can be made. It may include ways of capturing or processing data, producing information, or supporting management.

The First step of Twitter Sentiment Analysis is the identification of need. In this regard I have researched how these tweets can change the scenario. This was essential so as to know the processes that were being followed.

Requirements analysis is a software engineering task that bridges the gap between system level software allocation and software design.

**Analysis is a bridge between system engineering and software design**

Software requirements analysis may be divided into five areas of effort:-

1. Problem recognition.
2. Evaluation and synthesis.
3. Modelling.
4. Specification.
5. Review.

Initially, the analyst studies the system specification and the software project plan. Problem evaluation and solution synthesis is the next major area of effort for analysis.

Upon evaluating current problems and desired information (input and output) the analyst begins to synthesize one or more solutions.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 17

During the evaluation and solution synthesis activity, the analyst creates models of the system in an effort to better understand data and control flow, functional processing and behavioural operation, and information content. The model serves as a foundation for software design and as the basis for the creation of a specification for the software.

**6.1.1 Configuration:**

**Hardware Configuration:**

The section of hardware configuration is an important task related to the software development insufficient random access memory may affect adversely on the speed and efficiency of the entire system. The process should be powerful to handle the entire operations. The hard disk should have sufficient capacity to store the file and application.

Processor: I7 6th generation and above

Processor speed: 3.0 GHz Onwards

System memory: 25 GB minimum

Cache size: 512 mb

RAM: 4 GB(Minimum)

Network card: Any card can provide a 100mbps speed

Graphics card: 4 GB(Minimum)

Network connection: UTP or Coaxial cable connection

Hard disk: 100 Gb

Monitor: SVGA Color 15” 5

**Software Configuration:**

A major element in building a system is the section of compatible software since the software in the market is experiencing in geometric progression. Selected software should be acceptable by the firm and one user as well as it should be feasible for the system. This document gives a detailed description of the software requirement specification. The study of requirement specification is focused specially on the functioning of the system. It allows the developer or analyst to understand the system, function to be carried out the performance level to be obtained and corresponding interfaces to be established.

Operating system: Windows 10.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 18

**Let's Explore Our Project**

**\*\*\_\_\_Dataset\*\*\_\_\_**

**First let us look at the dataset. It is tweets.csv.**

**It has 3 columns - "ItemId",”Sentiment” and “Sentiment Text".**

**Column Description:**

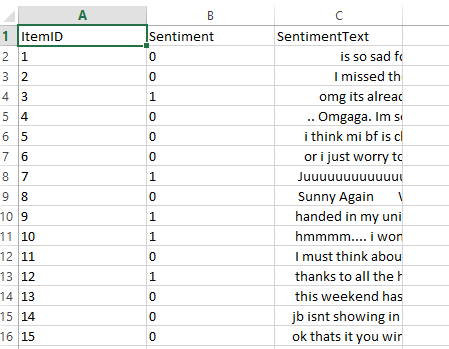
Item Id -> Numbering of each rows

Sentiment -> This column contains value either 0 or 1 (0 for Negative Sentiment and 1 for

Positive Sentiment) corresponding to the text.

Sentiment Text -> Tweets by different members.

**Below is the screenshot of the dataset**



**Maulana Azad National Institute of Technology (M.P)** **Page 19**

**\*\*\_\_\_\_\_Project Objective\*\*\_\_\_**

**We want to build a model to predict the sentiments of the tweets given by the members.**

**Let's get started.**

**Step 1: Load the Dataset**

**We need to load dataset from our project file. First we are importing the important libraries**

**and mounting the drive and then loading the dataset for using it in the code.**

**# Step 1 - Load Data**

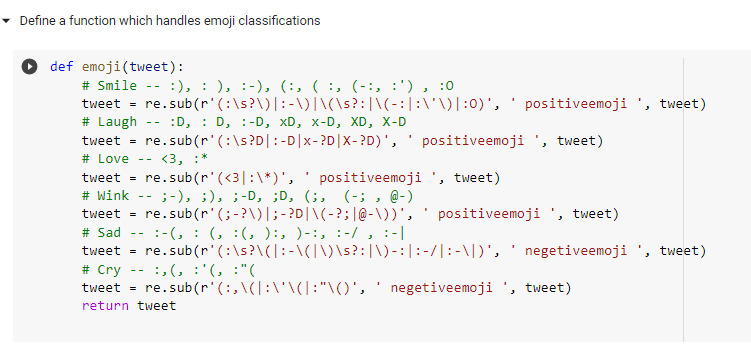


**# Step 2 – Preprocessing the Data**

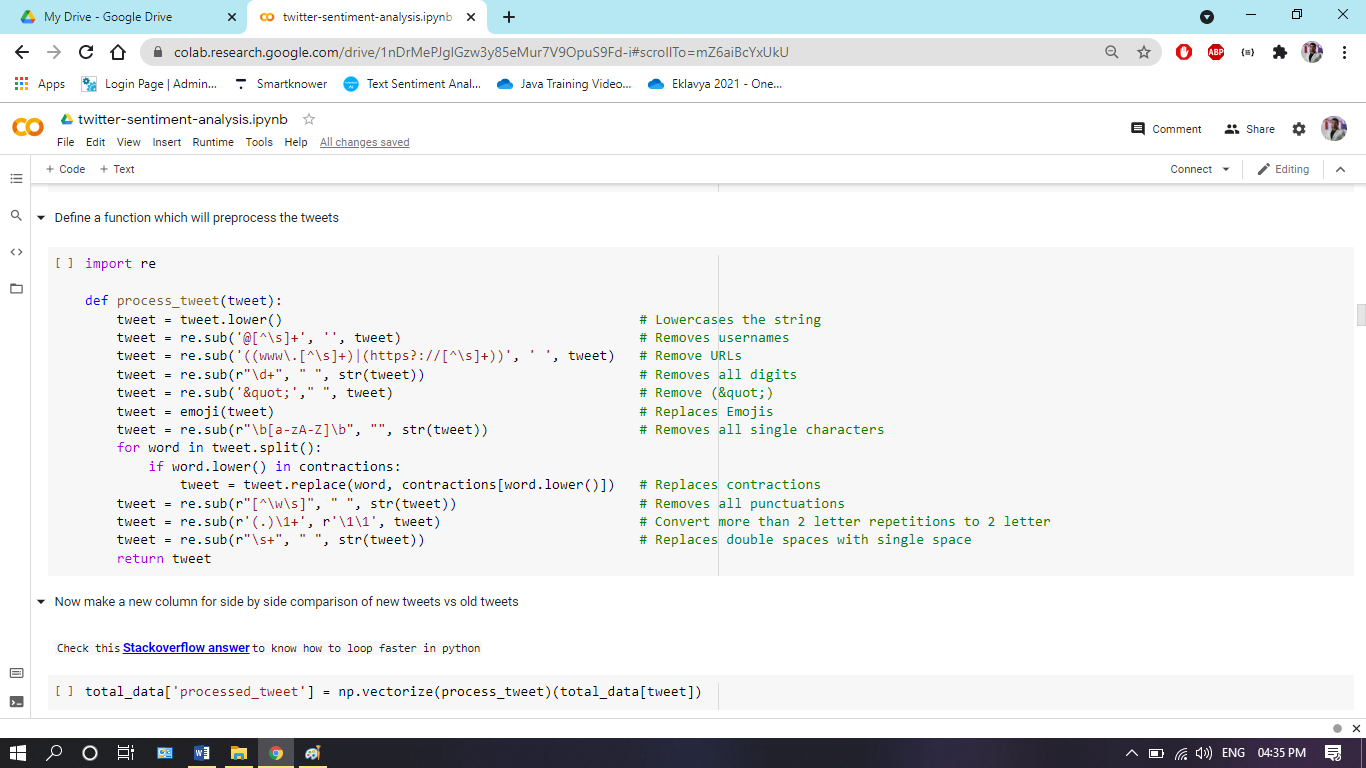
* **Convert every tweets to lower case**
* **Remove Twitter username**
* **Remove punctuations, numbers and special characters**
* **Convert more than 2 letter repetitions to 2 letter ( example (wooooooow --> wow))**
* **Remove extra spaces**
* **Remove URLs**
* **Emoji analysis**
* **Handle contractions words**
  + **" can't " >> " can not "**
  + **" won't " >> " will not "**
  + **" should't " >> " should not "**
* **Tokenization**
* **(Optional) Remove**[Stop words](https://www.geeksforgeeks.org/removing-stop-words-nltk-python/)
* **(Optional) Text Normalization (**[Stemming](https://www.geeksforgeeks.org/python-stemming-words-with-nltk/)**/**[Lemmatization](https://www.geeksforgeeks.org/python-lemmatization-with-nltk/)**)**

**Maulana Azad National Institute of Technology (M.P) Page 20**

**Handling emojis :**



**Preprocessing the tweets:**



**Maulana Azad National Institute of Technology (M.P) Page 21**

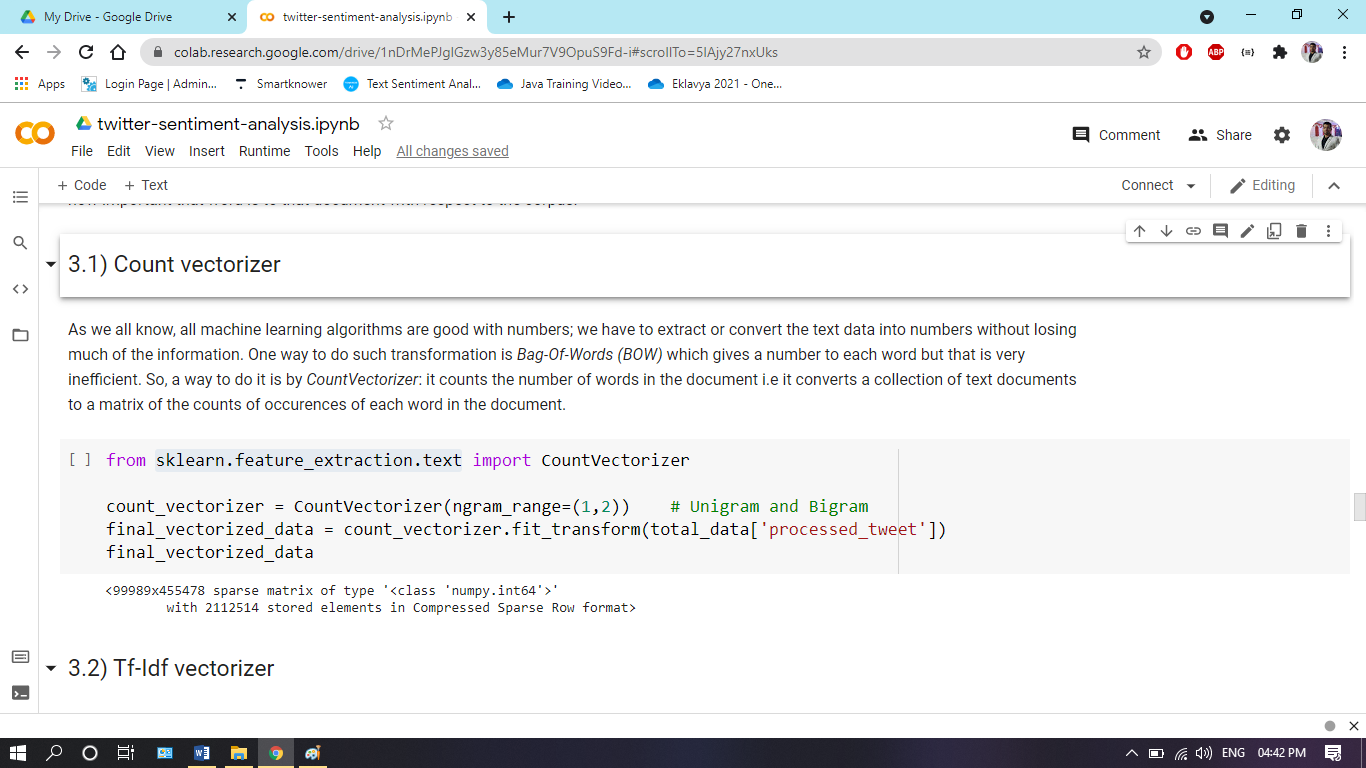
**Step 3: Fit Count Vectorizer model to datasets**

First we will build a simple count vectorizer model to see what prediction it makes and then

compare it to the prediction made by the test data to see how much accurate our model is.

We will be using the countVectorizer class from the library sklearn.feature\_extraction.text

We create an object of the count vectorizer class and call the fit method passing the X and y.

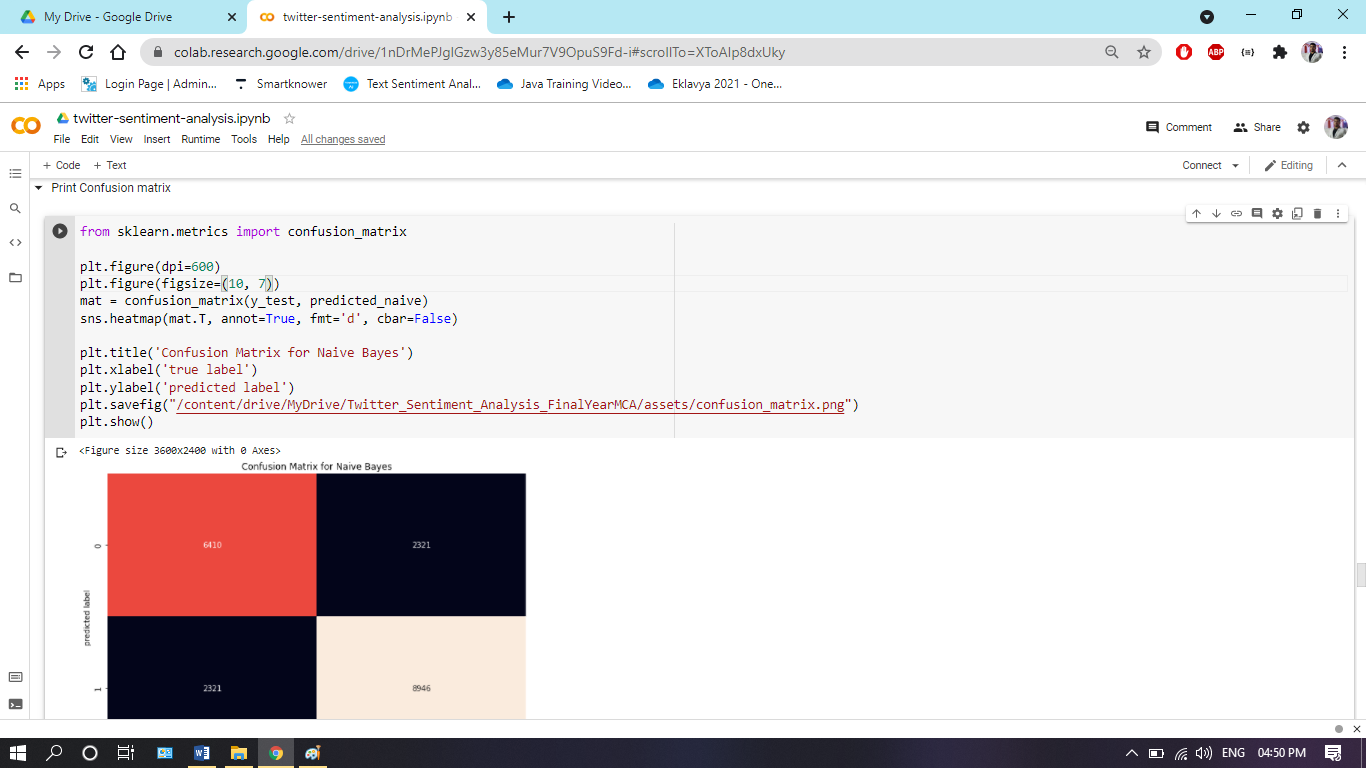


**Maulana Azad National Institute Of Technology (M.p.) Page 22**

**Step 4: Splitting the dataset in to train and test**

****

**Step 5: Printing the confusion matrix**



**Maulana Azad National Institute of Technology (M.p) Page 23**

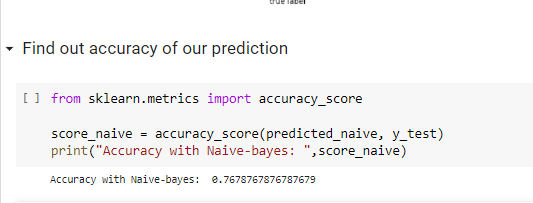
**Accuracy of our Model**

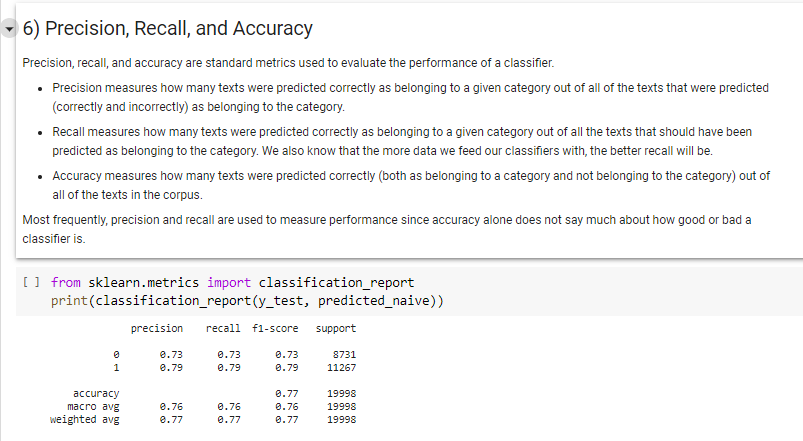
**r2=r2\_score(Y\_test,Y\_pred) # coefficent of determination**

**From testing data we can know easily know how much our model is accuracy, we have values against, testing data,but we predict values and after result came we can check with actual value of test data**

**and predict value of test data if difference is nearby 1 then we will say our model accuracy is**

**very good, this difference is called coefficient of determination.**





**So in this case by using Count Vectorizer we got 76.78% accuracy.**

**Maulana Azad National Institute of Technology (M.P)** **Page 24**

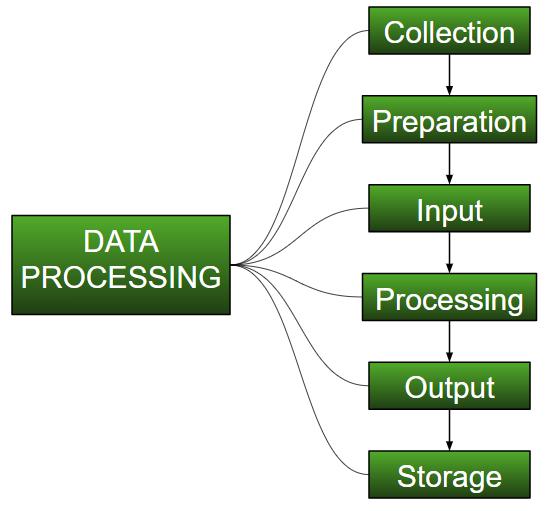
**6.2 Data Pre-Processing:**

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

**Why Data Processing:**

Data pre-processing is done for achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. And there is another aspect that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

Steps involved in Data Processing



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 25

**6.3 Data Set Description**

Pandas Data Frame is a two-dimensional size- mutable, potentially heterogeneous tabular data structure with labelled axes (rows and columns). heterogeneous tabular data structure with labelled axes (rows and columns). tabular fashion in rows and columns. Pandas Data Frame consists of three

principal components, the data, rows, and columns.

For data collection I used Csv files. Csv files is a database where we can store data in excel format, we can store as many as data.

From that training data I need 20% test data from these data we predict our sentiment if above 80% accuracy will come then we will say our model is accurate.

The data set contains members data

1. Sentiment
2. Sentiment Text



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 26

**7. Software Engineering Paradigms**

**Software Process Models**

To solve actual problems in an industry setting, a software engineer or a team of engineers must incorporate a development strategy that encompasses the process, method and tools.

A process model for software engineering is chosen based on the nature of the project and application, the methods and tools to be used, and the control and deliverables that are required.

Among the various available Software Process Models (Waterfall models, spiral model Prototyping model, etc) the best suited for this project is the Prototyping Model.

To develop the system Proto Type Modal is applied as the company wanted to see the software development process and appearance of the software so that the idea of the functionally of the system can be understood. The software developed in three phases in first phase the dry proto type is developed in which after the analysis the screens are designed and no validation are performed and also no database functionality is incorporated. After the demonstration of dry proto type suggestions from the client are noted and the development process is moved to the second phase i.e. wet prototype the actual designed is incorporated, validations are performed and the software is submitted to the user for acceptance and testing and then after final submission of the software is produced with user manual.

**Prototyping Model**

Often, a customer defines a set of general objectives for software but does not identify detailed input, processing or output requirements. In other cases the developer may be unsure of the efficiency of an algorithm, the adaptability of an operating system, or the form that human –machine interaction should take, in these and many other situations, prototyping paradigm may offer the best approach.

The prototype can serve as “the first system’’. It is true that both the customer and developers like the prototyping paradigm. Users get a feel for the actual system and developers get to build something immediately. Yet prototyping can also be problematic for the following reasons:

1. The customer sees what appears to be a working version of the software that the prototype is held together “with chewing gum and a baling wire’’ unaware that in the rush to get it working we haven`t considered overall software quality or long



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 27

term maintainability. When informed that the product must be rebuild so high level of quality can be maintained. Too often, software development management relents.

* 1. The developer often makes implementation compromises in order to get a prototype working quickly. An inappropriate operating system or programming language may be used simply because it is available and known, an efficient algorithm may become familiar with these choices and forget all the reasons why they were inappropriate. The less than ideal choice has now become an integral part of the system.

1. **Analysis Documents**

Document Analysis is a technique used to gather requirements during the requirements elicitation phase of a project. It describes the act of reviewing the existing documentation of comparable business processes or systems in order to extract pieces of information that are relevant to the current project, and therefore should be consider projects requirements.

**8.1 Software Requirement Specification:**

The software requirements specification is produced at the culmination of the analysis task. The function and performance allocated to software as a part of system engineering are refined by establishing a complete information description a detailed functional and behavioral description, a indications of performance requirements and design constraints, appropriate validation criteria and other data pertinent to requirements. The National Bureau of standard IEEE and the US Department of Defense have all proposed candidate formats for software requirements specifications. For our purpose however the simplified outline presented may be used as a framework for the specification.

**S.R.S. Of Our Project:**

**Introduction:** It states the goals and objectives of the software descriptions in the context ofthe computer – based system actually, the introduction may be nothing more than the software scope of the planning document.

**Information description**

It provides a detailed description of the problem that the software must solve information contest and relationships, flow and structure are documented. Hardware, software and human



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 28

interfaces are described for external system elements and internal software function. The Department of Administrative Reforms and Public Grievances aims to facilitate the pursuit of excellence in governance. It gives the information about the citizen with grievances against public a institution and government organization viz. central Government Ministries / departments / Organizations / State Governments /UT Administrations and services to them. It forward all the information to citizens with grievance to the concerned Government organization within 15days under intimation to the petition.

**Functional Description**

It present description of each function required to solve the problem. A processing narrative is provided for each function, design constraints are stated and justified, performance characteristics are stated, and one or more diagrams are included to graphically represent the overall structure of the software and interplay among software function and other system elements under this topic, the functional working of the various modules where give the complete picture for the data flow from one location to other. There proper linking between them. The department of Administrative Reforms and Public Grievance aims to facilitate the pursuit of excellence in governance through promotion of:-

1. Improvements in Auction structures and processes.
2. Consumer friendly initiatives including redressal of grievances.
3. Documentation, incubation and discrimination of best practices.
4. Codification and simplification of procedures.
5. Internet Based.

**Behavioural Description**

This section of the specification examiners the operation of the software as a consequences of external events and internally generated control characteristics. In this topic may project shows the event flow diagram which tells us at what step what will be the result.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 29

**9. Design**

A thorough analysis of Information System requirements helps in defining an effective design of the information system. This is essential so as to meet the requirements of the system in an efficient way. Thus, System Design can be described as the process of planning the new business system.

Following points were kept in mind while the System was designed

* That it performs the Required job
* Presents information and instructions in an acceptable fashion
* Produces accurate results
* Provides an acceptable interface and method of interaction
* Is perceived by the users as a reliable system.
* The Cost of the maintenance of the system should be minimum subject to the condition that it must satisfy all user requirements
* The System must be stable under all conditions and can be easily operated
* It Conforms to the Standards
* It should be modifiable depending upon the changing needs of the user



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 30

**10. Code Efficiency**

In software project we must keep a balance between the language used & the area of the application as well as database management system were actual data is being stores. There must be provision for those situations which are not meeting at the time of making the project for example the codes should be written keeping the aspect that software should compatible with others database or data can be migrated from outside. The information systems are designed with space & time complexity and cost saving in mind. The coding systems are methods in which conditions, words, ideas or relationship are expressed by a code. A code is an ordered collection of symbols designed to provide unique identification of an entity or attribute. It may be brief number, title or symbol. The main purpose of codes is to facilitate the identification and retrieval of times of information from the system.

There are many possible-coding structures. The main types of codes are described below:

**Classification Code**: Classification is the best described as the establishment of categories ofentities, types and attributes in a way that brings like or similar items together according to pre-determined relationships. A classification is by nature an order systematic structure. So the classification code places separate entities like events, people, or object into distinct groups called classes. A code is used to identify one class from another. Using the code the user classifies the event into one of several possible categories and records the code.

Classification codes vastly simplify the input process because only a single-digit code is required. The need for writing lengthy description or making judgments is eliminated.

**Function Code**: Function codes states the activities or work to be performed without spellingout all of the details in narrative statement.

It has been recognized that good coding style can overcome many of the deficiencies of a primitive programming language, while poor styles can defeat the intent of an excellent language. The goal of a good coding style is to provide easily understood, straight-forward and elegant code. The guidelines for coding includes:

* Use of a few standard control constants.
* Use of unconditional branching (go to) in a disciplined manner.
* Introduction of user-defined data types to model entities in the problem domain.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 31

* + Hiding of data structures behind access functions.
  + Providing standard documentation prologues for each subprogram and compilation unit.
  + Use of indentation, parenthesis, and blank spaces, blank lines, borders and comments to enhance readability.

1. **Code Optimization**

The main purpose of the optimization of code is to reduce redundancy of code, for this user has to write procedure function for repetitive. If in two situation the same type of interface needed then only one interface should be created and it may be called by passing different parameter.

The basic design model uses the analysis model as the framework for implementation. The analysis model captures the logical information about the system, while the design model must add details to support efficient information access. During design optimization we must:

* + Add redundant association to minimize access cost and maximize convenience.
  + Rearrange the computation for greater efficiency.
  + Saved derived attributes to avoid re-computation of complicated expression.
  + During report generation we used an optimization technique of rearranging execution order for greater efficiency.
  + Redundancy of data in the tables has been removed through Normalization

Technique. All the tables are designed in normalized form.

* Explicit and Implicit locks are used where needed.
  + Explicit update: In this each derived attribute is defined in terms of one or more fundamental base object. We determined that which derived attributes are affected by each change to a fundamental attribute and inserted code into the update operation on the base object to explicitly update the derived attribute that depends on it.
  + Optimization of Resource: During coding we have kept the strict vigil to use the same function again by writing the code in a module, so it keeps our coding minimize.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 32

1. **Testing**

In this section we have tested our system in different modes.

Basically the testing is done on different level like Unit testing, Integration testing , System testing and Acceptance testing. These different testing tests different types of faults.

We have tested in the following manner and whatever the fault we have got, fixed it immediately.

Clients Need  Acceptance Testing



Requirements  System Testing

Design Integration Testing



Code Unit Testing



Testing levels

**SYSTEM TESTING**

As the part of system testing we execute the program with the intent of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. The ultimate aim is quality assurance.

Tests are carried out and the results are compared with the expected document. In the case of erroneous results, debugging is done. Using detailed testing strategies a test plan is carried out on each module. The various tests performed are unit testing, integration testing and user acceptance testing.

**Unit Testing:**

The software units in the system is are modules and routines that are assembled and integrated to perform a specific function. As a part of unit testing we executed the program for individual modules independently. This enables, to detect errors in coding and logic that are contained within each of the three module. This testing includes entering data that is filling forms and ascertaining if the value matches to the type and entered into the database. The various controls are tested to ensure that each performs its action as required.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 33

**Integration Testing:**

Data can be lost across any interface, one module can have an adverse effect on another, sub functions when combined, may not produce the desired major functions. Integration testing is a systematic testing to discover errors associated within the interface. The objective is to take unit tested modules and build a program structure. All the modules are combined and tested as a whole. Here the admin module, sec module and student module options are integrated and tested. This testing provides the assurance that the application is well integrated functional unit with smooth transition of data.

**User Acceptance Testing:**

User acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keep the records of applicants and making changes to the details and password whenever required.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 34

**13. Implementation**

A crucial phase in then systems life cycle is the successful implementation of the new system design. Implementation means converting a new system design into operation. This creating computer-compatible files, training the operating staff before the system is up and running. A major factor in conversion is not disrupting the functioning of the organization.

The implementation phase of software development is also concerned with translating design specification into source code. it is necessary to write source code and internal and documentation so that conformance of the code to its specification can be easily verified, and so that debugging, testing and modification are eased. This can be achieved by making the source code as clear and straight forward as possible. The implementation team should therefore be provided with-defined set of software requirements. An architectural design specification and a detailed design description.

User training is another important area, which is responsible for minimizing resistance to change and giving the new system a chance to provide it worth. Training aids, such as user-friendly manuals, data dictionary, and job performance aids that communication information about the new system; help screens provide the user with a good start on the new system.

Following conversion, it is desirable to review the performance of the system and to evaluate it against established criteria. Software maintenance follows conversion, which includes minor enhancements or corrections to problems that surface late in the system’s operating.

**Conversion:**

The objective of conversion is to put the system into operation while holding costs, risks and personal problems to a minimum. It involves three major steps:

* Creating computer-compatible files
* Training the operating staff
* Installing hardware/software

Procedures and documents are unique to the conversion phase. These are shown below:

* Conversion begins with a review of the project plan, the system test documentation, and the implementation plan. The parties involve are the user, the project team, programmers and operators.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 35

* The conversion portion of the implementation plan is finalized and approved.
* Files are converted.
* Parallel processing between the existing and the new system is initiated.
* Results of computer runs and operators for the new system are logged on a special form.
* Assuming no problems, parallel processing is discounted. Implementation results are documented for reference.
* Conversion is completed. Plans for the post-implementation review are prepared. Following the review, the new subsystem is officially operational.

**File Conversion:**

File conversion involves capturing data and creating a computer file from existing files. There can be problems like staff shortage for loading data. Also, specialized training necessary tom prepares records in accordance with the new system specifications. In most cases, an outside agency performs this function for a flat rate. If a computerized system already exists, then copying existing files for a new system is a concern area. Program can be written to copy files intact for the new system and test programs on both systems. A file comparison program can be used for this purpose.

Many systems are prone to errors because of insufficient attention given to data entry control or protective features like audit control trails. These items must be part of the overall plan of for conversion. Before a data entry operator starts working, a data entry validation program can be written to keep track of wrong data entered.

At the time installation of the software it is required that oracle must be installed on the server of the company and it must contain the data files related to the software before implementing the developed system on all machines

Connected to the server. For this the exported data is imported to the client machine and the database is checked that all tables are properly imported to the client machine or not as the software is developed in oracle database and also going to be implemented in oracle is no other conversion is required.

A good audit control trail is the key to detecting fraud and errors in data entry.

**User Training:**

Analysis of user training focuses on two factors-user capabilities and the nature of the system being installed. The requirements of the system also range from very simple tasks like using a



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 36

pocket calculator to complex tasks like learning to program a database system. Tasks that require the user to follow a well-defined, concrete, step-by-step procedure-require limited problem solving. For this, the training level and duration is basic and brief.

**Hardware/Software Installation:**

Adequate time and resources for installation of software and hardware must be allotted in the development schedule. User can be trained on the installation procedure. The detailed instructions can be a part of the user manual.

**14. Evaluation**

Evaluation of the system is performed to identify its strength and weaknesses. The actual evaluation can occur along any of the following dimensions:

* **Operational Evaluation**: Assessment of the manner in which the system functions,including ease of use, overall reliability and level of utilization.
* **Organizational Impact**: Identification and measurement of benefit to
  + The organization in such areas as financial concerns, operational efficiency and competitive impact.
* **User Manager Assessment**: Evaluation of attitudes of senior and user managerwithin the organization as well as end users.
* **Development Performance**: Evaluation of the development process in accordance with such yardsticks as overall development time and effort, conformance to budgets and standards and other project management criteria.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 37

**15. Maintenance**

Software maintenance is used to describe the software engineering activities that occur following delivery of a software product to the customer.

The maintenance of existing software can account for over 60% of all effort expanded by a development organization, and the percentage continues to rise as more software is produced.

* Activities involved in maintenance of a software product include:
  + Analysis activities
  + Standards and guidelines
  + Design activities
  + Implementation activities
  + Supporting documents
* Configuration management is concerned with tracking and controlling of the work product that constitute a software product.
* Configuration management tools include:
  + Configuration management databases
  + Version control libraries
* Automated tools to support software maintenance include technical support tools and managerial support tools.
* Other automated tools include
  + Text editors
  + Debugging aids
  + Linkage editor

Computer Maintenance covers a wide range of activities. Many activities performed during development of Airlines Reservation System for Global Airways to enhance the maintainability of it are as follows-

1. **Analysis Activities:** The analysis phase of software development is concerned withdetermining customer requirements and constraints and establishing feasibility of the product. From maintenance view point, the most important activities that occur during analysis are establishing standards and guidelines for the project and the work products to ensure uniformity of the products; setting of milestones to ensure that the work products are produce on schedule; specifying quality assurance procedures to



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 38

ensure development of high quality documents; identifying product enhancements that will most likely occur following initial delivery of the system; and estimating the response (personnel, equipment, floor space) required to perform the maintenance activities.

1. **Standards and Guidelines:** Various types of standards and guidelines we developedto enhance the maintainability of our software.
2. **Design Activities:** Design is concerned with developing the functional components,conceptual data structures and interconnection in a software system. The most important activity for enhancing maintainability during the design.

Automated tools to support software maintenance include technical support tools and managerial support tools. Tolls to support the technical aspects of software maintenance span the spectrum from analysis and design tools to implementation to debugging the testing tools. Automated tolls include text editors debugging aids, cross-reference generators, linkage editors, comparators, complexity metric calculators, and version control system and configuration management databases. Text editors permit a rapid, efficient modification of source programs, test data and supporting documents. Text editors can be used to insert the replace segments of source code, internal comments, test data and supporting documents; to systematically change the occurrences of an identifier or other textual strings to locate all references to a given identifier or other string of text; and to save both old and new version of a routine, test files or document. A syntax-directed text editor can ensure that all cross-references in the supporting documents are correctly updated.

Debugging aid provide traps, and traces assertion checking and history file aid in locating the causes of known errors. System-level-cross-reference generators provide cross-reference listing of procedure calls, statements usages and data references. Cross-references directories provide the calling structures of who calls whom and for where, and procedures names and statements numbers where formal parameters local variables and global variables are defined, set and used.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 39

**16. Conclusion and Recommendations**

In this model, we used the Count Vectorizer and Tf-Idf Vectorizer. We saw that the Count Vectorizer result is 79% accurate compare to linear regression.

While linear regression also give 70 to 75% accuracy which is also good.

**16.1 Recommendations and future studies**

Two algorithms, Count Vectorizer and Tf-Idf Vectorizer were used in this study and a data set from was applied to train and test the models. The system can only predict the

Sentiments of the tweets.

Maulana Azad National Institute Of Technology Bhopal (M.P) Page 40



**18. Future Scope of the Project**

This software is made according to the requirement of organization but expansion of modules can be easily included in the software.

**Handling emotion ranges**: We can improve and train our models to handle a range of sentiments. Tweets don’t always have positive or negative sentiment. At times they may have no sentiment i.e. neutral. Sentiment can also have gradations like the sentence, this is good, is positive but the sentence, this is extraordinary is somewhat more positive than the first. We can therefore classify the sentiment in ranges, say from -2 to +2.

**Using symbols**: During our pre-processing, we discard most of the symbols like commas, full-stops, and exclamation mark. These symbols may be helpful in assigning sentiment to a sentence.

**Software Development:**

The total span of developing the Database Management System that I got was 3 months.

During this duration, I was involved in the different aspects of system development viz.

System analysis, System design and development.

In this limited period of time, it was impossible for me to devote as much time as I would have wanted, on each of the phases.

Despite this limitation, the project gave me an opportunity to understand and learn about the various phases of the Software development life cycle.

It was a good opportunity for me to learn about the various phases of System development in depth, by means of documents, System Study and interacting with System users and further clarifications of doubts with the guide.

Since it was an individual activity, it helped me even more to strive harder to gain invaluable experience and knowledge to handle a software development project.

**Achievements:**

Working on a live project form its initial stages gave me an advantage of understanding the various needs of the customer and devising a plan to satisfy his requirements. Working on this project has benefited me in a number of ways. It gave me an opportunity to understand the users need, interact with them on personal basis and consequently develop a plan to meet their requirements.

The project gave me an experience of working in a professional set-up enriched with quality systems. I got an opportunity to learn and discover the various powerful features .An



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 41

important aspect of developing an information system is to understand, in depth the entire functionality and features of the system before planning to develop the proposed system.



Maulana Azad National Institute Of Technology Bhopal (M.P) Page 42

**20. Bibliography**

1. <https://www.geeksforgeeks.org/machine-learning/>
2. <https://www.kaggle.com/>
3. GitHub: <https://github.com/Shashwat5497/twitter_senstiment_analysis.git>
4. <https://www.analyticsvidhya.com/>

Maulana Azad National Institute Of Technology Bhopal (M.P) Page 43