VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"STOCKPORT PREDICTIVE SENTIMENT ANALYSIS"

Submitted in partial fulfilment for the award of degree(18CSI85)

**BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE AND ENGINEERING**

Submitted by:

MALLIKARJUN A G (2LG19CS019)

Conducted at
VARCONS TECHNOLOGIES Pvt Ltd

GOVERNMENT ENGINEERING COLLEGE, TALAKAL DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
Recognised by Govt. of Karnataka,Affilited to VTU,Belagavi

590018

GOVERNMENT ENGINEERING COLLEGE, TALAKAL
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
Recognised by Govt. of Karnataka,Affilited to VTU,Belagavi 590018

CERTIFICATE

This is to certify that the Internship titled **"STOCKPORT PREDICTIVE SENTIMENT ANALYSIS"** carried out by **Mr. MALLIKARJUN A G,** a bonafide student of Government Engineering College, Talakal, in partial fulfillment for the award of **Bachelor of Engineering**, in **COMPUTER SCIENCE AND ENGINEERING** under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

Signature of Guide   **Signature of HOD**   **Signature of Principal**

External Viva:

Name of the Examiner   Signature with Date

1)_____

_____

_____
2)_____

D E C L A R A T I O N

I, **MALLIKARJUN A G** final year student of Computer Science and Engineering, Government Engineering College, Talakal - 583238, declare that the Internship has been successfully completed, in **VARCONS TECHNOLOGIES Pvt Ltd**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Computer Science and Engineering, during the academic year 2022-2023.

Date : 23 SEPT 2022

Place : KOPPAL

USN : 2LG19CS019

NAME : MALLIKARJUN A G

OFFER LETTER

INTERNSHIP OFFER LETTER

Varcons Technologies Pvt Ltd

Date: **23rd August, 2022**

Name: **Mallikarjun A G**
USN: **2LG19CS019**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Varcons Technologies Pvt Ltd**, effective Start Date **23rd August, 2022**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Spoorthi H C
Director
VARCONS TECHNOLOGIES PVT LTD
213, 2st Floor,
18 M G Road, Ulsoor,
Bangalore-560001

A C K N O W L E D G E M E N T

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our Principal, **Dr. Virupaxi Bagodi** for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – CSE, **Dr. Virupaxi Bagodi** for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We express our deep and profound gratitude to our guide, Asst Prof, **Tejaswini Eshwar Achar,** for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

MALLIKARJUN A G   (2LG19CS019)

ABSTRACT

In this era of growing social media users, Twitter has significantly large number of daily users who post their opinions in the form of tweets. This presents an idea of extracting sentiments out of the tweet and an approach towards classifying a tweet into positive, negative or neutral. This approach can be in many ways useful to any organization, who gets mentioned or tagged in a tweet. Generally, the tweets being unstructured in format, first of all the tweet needs to be converted into the structured format. In this paper, tweets are resolved using pre-processing phase and access of tweets has been accomplished via libraries using Twitter API. The datasets need to be trained using algorithms in a way, such that, it becomes capable of testing the tweets and it releases the required sentiments out of the feeded tweets. The growth and advancement in social media platforms engaged a huge number of users. Social media platform like twitter where users can post their tweets in 280 characters. Because of the limited number of characters in tweets, it becomes easy for the sentiment analysis. On Twitter 550 million of tweets are posted daily. Twitter also represents all age group people and also a fair representation of gender. Therefore, the sentiment analysis of twitter data becomes somewhat general sentiments of society. In this paper, we will compare various Machine Learning methods like the Naïve Bayes Classification method, Support Vector Machine Classification Method and Maximum Entropy Classification method. We will see how sentiments analysis is done by this classification algorithm and what is the accuracy and precision in these cases.

Table of Contents

CHAPTER 1
COMPANY PROFILE

A Brief History of Varcons Technologies Pvt Ltd

Varcons Technologies Pvt Ltd, was incorporated with a goal "To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients' requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever-increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting client's requirements.

Varcons Technologies Pvt Ltd, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Varcons Technologies Pvt Ltd work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence "Technology helps you to Delight your customers" and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

Varcons Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever-increasing automation requirements, Varcons Technologies Pvt Ltd specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting client's requirements. The organization where they have a right mix of professionals as a stakeholder to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well". Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, we strive hard to achieve it.

2.1 Products of Varcons Technologies Pvt Ltd.

2.1.1 Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later.

As of March 2015, the SDK is not available on Android itself, but software development is possible by using specialized Android applications.

2.1.2 Web Application

It is a client–server computer program in which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include web mail, online retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can beconsidered as a specific variant of client–server software where the client software isdownloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, andacts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a "smart" client that performs all the work and queries a "dumb" server, or a "dumb" client that relies on a "smart" server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn't allows for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry. Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process.

This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

2.1.3 Web design

It is encompassing many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

2.2 Departments and services offered

Varcons Technologies Pvt Ltd plays an essential role as an institute, the level of education, development of student's skills is based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Varcons Technologies Pvt Ltd gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Varcons Technologies Pvt Ltd always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skills development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

2.2.1 Services provided by Varcons Technologies Pvt Ltd.

•   Core Java and Advanced Java

•   Web services and development

•   Dot Net Framework

•   Python

•   Selenium Testing

•   Conference / Event Management Service

•   Academic Project Guidance

•   On The Job Training

•   Software Training

CHAPTER 3

INTRODUCTION

Twitter is a trending micro-online journal service wherein users can justify their sentiments in the form of "tweets". These tweets sometimes express opinions about different topics. Sentiment analysis is the prediction of emotions in a word, sentences or corpus of documents. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention

3.1 Introduction to ML

Machine Learning is a branch of the broader field of artificial intelligence that makes use of statistical models to develop predictions. It is often described as a form of predictive modelling or predictive analytics and traditionally, has been defined as the ability of a computer to learning. In basic technical terms, machine learning uses algorithms that take empirical or historical data in, analyze it, and generate outputs based on that analysis. In some approaches, the algorithms work with so-called "training data" first and then they learn, predict, and find ways to improve their performance over time.

Three different Machine Learning Classifiers such as LR, NB, SVM were applied for predicting the sentiment of the tweets.

- **Naive Bayes Classifier:** NB predicts the sentiment of the test dataset as positive or negative using a Multinomial NB classifier [4]. This classification is done based on Bayes' theorem. It is regarded as one of the most suitable for word counts.

- **Support Vector Machine (SVM):** SVM is a supervised machine learning algorithm which is widely regarded as one of the best text classification algorithms was proposed by Vapnik in 1992. It finds the linear separator with maximum marginal distance using support vectors in high dimensional space [4]. Here both linear kernel function and stochastic gradient descent (SGD) learning method have been used for predicting the sentiment of tweets and the results were compared.

- **Logistic Regression (LR):** LR is a machine learning Classifier which classifies based on probability prediction. Here the probability prediction must be transformed into a binary value using the logistic function.

3.2 Problem Statement

Sentiment analysis of in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews, documents, web blogs/articles and general phrase level sentiment analysis. These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text. The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised and semi-supervised approaches, and there is a lot of room of improvement. Various researchers testing new features and classification techniques often just compare their results to base-line performance. There is a need of proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications.

CHAPTER 4

SYSTEM ANALYSIS

The process by which an individual (s) studies a system such that an information system can be analyzed, modeled, and a logical alternative can be chosen.

4.1 Existing System

The existing system, uses knowledge base approach to classify the tweets into either positive, negative or neutral. But employing this method results in less accuracy of the classification. In existing system, they have employed lexicon-based method to compute the sentiment of the data coming front twitter which resulted in low accuracy rate. There is a lot of overhead while computing the sentiment of a sentence, because for each word this method retrieves the sentiment from a predefined word dictionary.

4.2 Proposed System

To overcome the drawbacks of the methods we have reviewed above, we propose a new model for sentiment analysis. In this model we combine many techniques to reach our final goal of emotion extraction. The steps for the process are documented below.

1. Retrieval of Data: Public Twitter data is mined using the existing Twitter APIs for data extraction. Tweets would be selected based on a few chosen keywords pertaining to the domain of our concern, i.e., product reviews. We have elected to use the Twitter API due to ease of data extraction.

2. Preprocessing: In this stage, the data is put through a preprocessing stage in which we remove identifying information such as Twitter handles, timestamps of the message and embedded links and videos. Such information is largely irrelevant and may cause false results to be given by our system.

3. Tweet Correction: As tweets are written for human perusal, they often contain slang, misspellings and other irrelevant data. Thus, we correct the misspellings in the sentences and look to replace the slang in the sentences with words from standard English that may roughly relate to the slang in question. As slang itself can be used to display a wide variety of sentiment, often with greater emotional impact, this process is necessary so that slang words may be considered as part of the emotion expressed.

4. Polarity detection: In this step we begin the second phase of our proposed system, in which we try to identify the polarity of the sentence in question. If emoticons exist in the statements, they will be used as well to compute the overall polarity of the statement.

We aim to find sentences where the polarity detection is not very clear or where the expressed sentiment may be low. We also try to isolate the opinion words in the sentence in relation to a given concept in the sentence.

    a. We train the system to understand the relation between words in various contexts. Pre-existing dictionaries like Sentic Net can be used in this phase to segregate the emotion from the context it is in.

    b. Once the opinion words are identified with context, we can find the polarities of the words using NLTK-Senti WordNet.

    c. To help with detection of the concepts associated, we train our system on a large dataset that expresses a wide variety of complex and ambiguous emotions. The system is given this data in an unsupervised fashion and will proceed by clustering.

5. Emotion Extraction: Emotion models often map the core emotions to a computational scale from which we can broadly classify and detect the emotions expressed. For the purposes of our system, we consider the "Plutchik's Wheel of Emotion" which divides all emotions into an eight-point wheel which represents the intensity and complexity of human feeling as we move from the centre of the wheel to the outer rim. The central core is made of 8 basic emotions that decrease in intensity as we move away from the centre, often blending with one or more emotion to become increasingly complex. For example, the wheel may express the simple emotions "rage" and "loathing" at the centre, but the rims contain the harder to identify emotions of "contempt", "boredom" and "annoyance".

    a. Mapping: Once the emotional relation has been extracted, we map it to Plutchik's model using a neuro-fuzzy inference system. As ambiguous phrases contain a high probability of expressing two or more emotions together in order to create a complex feeling, a neuro-fuzzy system is designed so that the emotions may be computed to a membership function instead.

    b. Once the system calculates the degree of membership of the emotion or emotions expressed in the statements, we use it to determine the most significant emotions. This value is decided after comparing all the degrees of membership given by the opinion words in the statement.
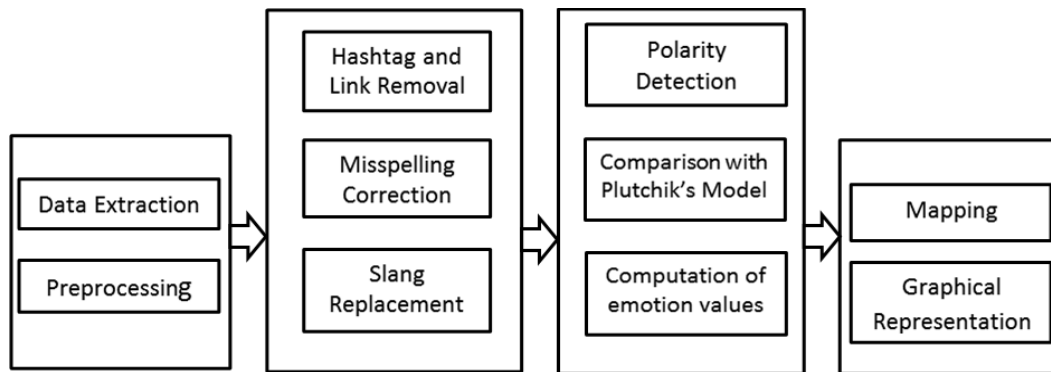
4.2.1 Model of Proposed System

Fig 4.3: Model of Proposed System

4.3 Objective of the System

The objective of feeling investigation is recognizing content assumption extremity. Assumption examination could be taken as an order issue. According to [1-2], sentiment analysis is a process that isolates the content into positive, negative or neutral conclusion. Profound neural system and the Gaussian blend model is one of the hearty models for normal preparing language.

Conviction analysis is useful for consumers who are trying to research a product or service, or marketers researching public opinion of their company/product. However, doing the analysis of tweets that express human emotions isn't an easy job. A lot of challenges are involved in terms of tonality, polarity, lexicon and grammar of the Twitter sentiment analysis using bag of words tweets. They tend to be highly unstructured and non-grammatical and therefore it gets difficult to   interrupt   their   meanings.

CHAPTER 5

REQUIREMENT SPECIFICATION

Hardware Requirement Specification

- RAM : 128MB, 256MB recommended
- Output Device : Monitor, 1366* 768 resolution
- Processor : Intel 3, 1.1 GHz or faster processor

Software Requirement Specification

- Software : Jupiter notebook
- Operating system : Windows 11
- Programming language : Python
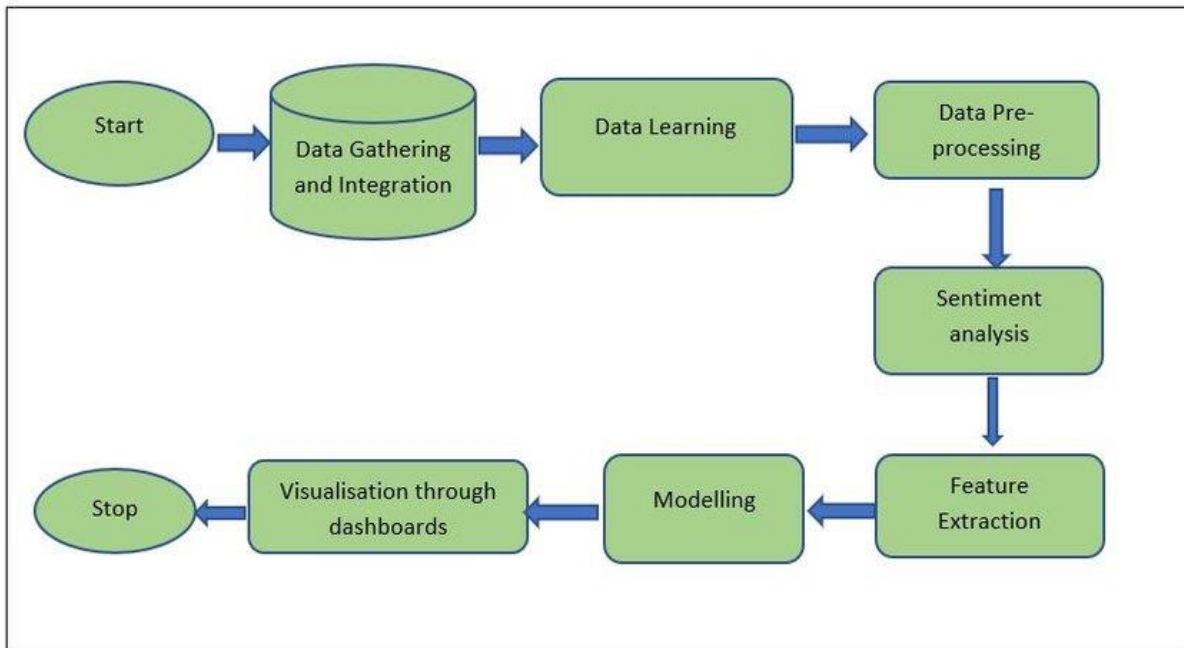
CHAPTER 6
DESIGN AND ANALYSIS

Fig 6.1: Aspect Based Sentiment Analysis Process Flow Diagram

Process starts from gathering of data from different sources. After data gathering and data understanding data pre-processing is conducted. Pre-processing of data includes the data cleaning steps like removal of numbers, removal of stop words, removal of white spaces etc. Once the data is clean python code is applied on data to fetch polarity / sentiment score from reviews. Feature extraction is done by qdap package in R. After getting final data set with aspects and sentiment score five different supervised machine learning algorithms were tested and and highest accurate model is selected. Finally, visualization is done with tableau an R.

CHAPTER 7

IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the changeover and an evaluation of change over methods a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

TESTING

The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.

2. Unit testing is the important and major part of the project. So, errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So, unit testing is conducted to individual modules.

3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

CHAPTER 8

SNAPSHOTS
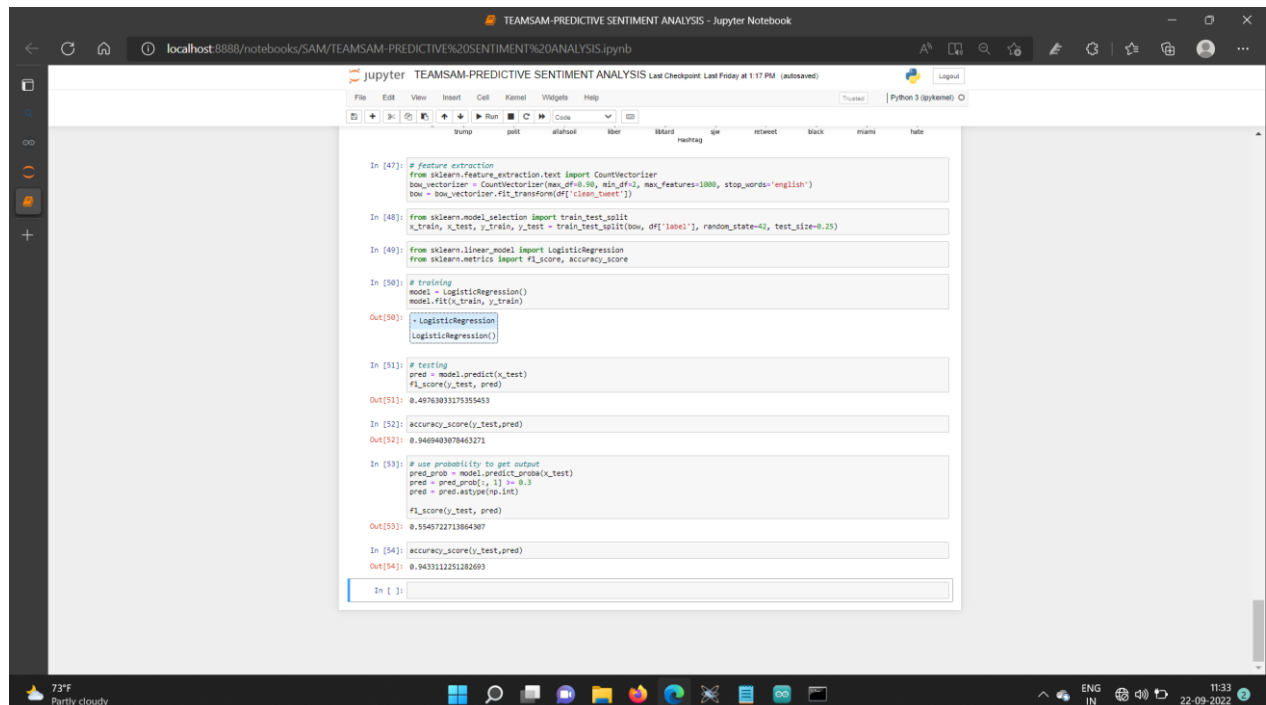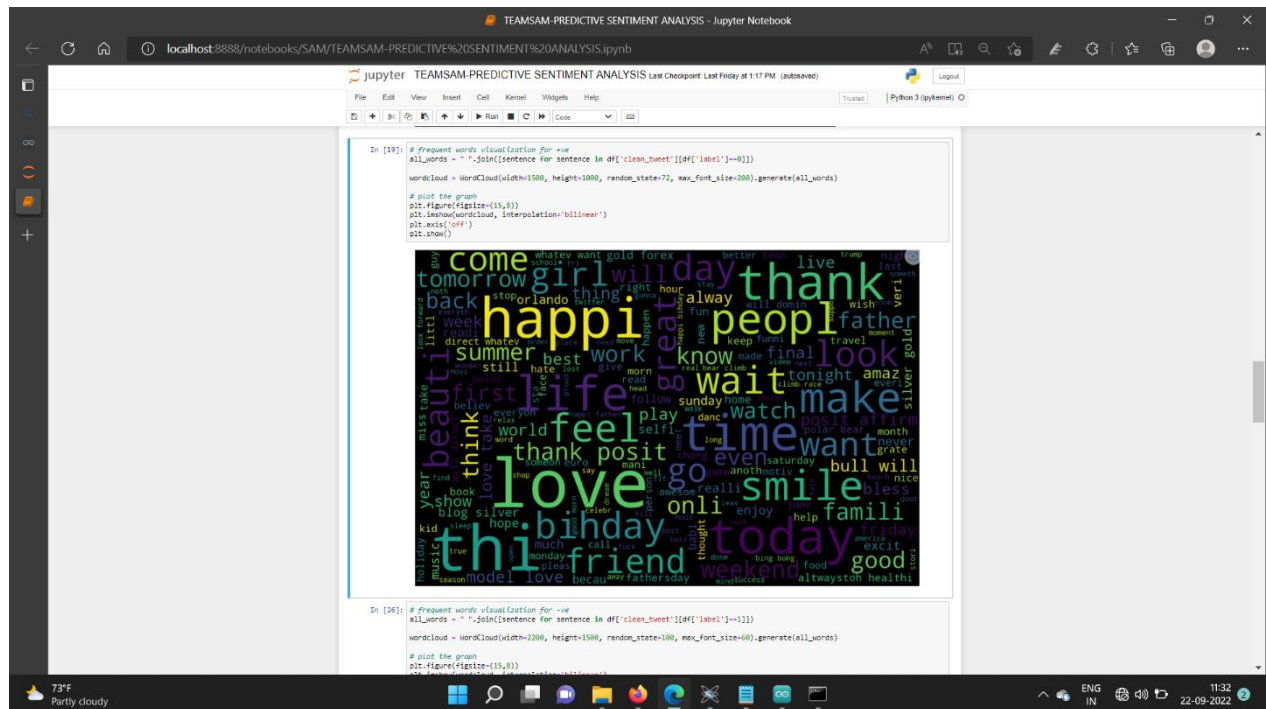
CONCLUSION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- ❖ Automation of the entire system improves the efficiency

- ❖ It provides a friendly graphical user interface which proves to be better when compared to the existing system.

- ❖ It gives appropriate access to the authorized users depending on their permissions.

- ❖ It effectively overcomes the delay in communications.

- ❖ Updating of information becomes so easier

- ❖ System security, data security and reliability are the striking features.

- ❖ The System has adequate scope for modification in future if it is necessary.

REFERENCES

[1] Ashwin Perti, Munesh Chandra Trivedi, Amit Sinha, "Development of intelligent model for twitter sentiment analysis", Materials Today proceedings, issue 1, August,2020.

[2] Gonzalo A.Ruz, Pablo A. Henríquez, Aldo Mascareno, "Sentiment analysis of Twitter data during critical events through Bayesian networks classification", Future Generation Computer Systems ,Volume 106,issue May 2020.

[3] Richard A.Plunz, Yijia Zhoua, Maria Isabel Carrasco Vintimillab, Kathleen Mckeownc, Tao Yud, Laura Uguccionia, Maria Paola Sutto, "Twitter Sentiment in New York city parks as measures of well being", Volume 189, issue April 2019.

[4] Soumya S, Pramod K.V "Sentiment Classification Malayalam tweets using machine learning techniques", ICT Express, issue 22, April 2020.

[5] Akilandeswari Ja, Jothi G, "Sentiment Classification of Tweets with Non-Language Features", ICACC-2018, Procedia Computer Science 143 (2018) 426–433.

[6] Filippo Chiarello, Andrea Bonaccorsi, Gualtiero Fantoni, "Technical Sentiment Analysis Measuring Advantages and Drawbacks of New Products Using Social Media", Computers in Industry, Volume 123, December 2020.

[7] Mohd Zeeshan Ansaria , M.B. Aziza, M.O. Siddiquib ,H. Mehraa , K.P.Singha "Analysis of Political Sentiment Orientations on Twitter ", ICCIDS 2019, Procedia Computer Science 167 (2020) 1821–1828.

[8] Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi, Wejdan Abdullah AlShehri "Sentiment Analysis of Twitter Data", International Conference on Computer & Information Science (ICCIS) ,2019.

[9] Ankita Sharma, Udayan Ghose, "Sentimental Analysis of Twitter Data with respect to General Elections in India", ICITETM 2020, Procedia Computer Science 173 (2020) 325–334.

[10] Sushree Das, Ranjan Kumar Behera, , Mukesh kumar, Santanu Kumar Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction", ICCIDS 2018, Procedia Computer Science 132 (2018) 956–964.