VISVESVARAYA TECHNOLOGICAL IVERSITY

BELAGAVI – 590018,Karnataka

INTERNSHIP REPORT

ON

"Stockprice prediction using twitter sentiment analysis"

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

BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE ENGINEERING

Submitted by:

Vilas Hegde

4SU19CS111

Conducted at
Varcons Tech Pvt.Ltd

SDM INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering

Accredited by NBA, New Delhi

Ujire

SDM INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Accredited by NBA, New Delhi
Ujire

CERTIFICATE

This is to certify that the Internship titled **"Stockprice prediction using twitter sentiment analysis"** carried out by **Mr. Vilas Hegde,** a bonafide student of SDM Institute of Technology, 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, **Vilas Hegde**, final year student of Computer Science and Engineering, SDM Institute of Technology - 574240, declare that the Internship has been successfully completed, in Varcons Tech Pvt Ltd. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Branch name, during the academic year 2022-2023.

Date : 25-09-2022 :

Place : Ujire

USN : 4SU19CS111

NAME : Vilas Hegde

OFFER LETTER

Internship Offer Letter

Date: **23rd August, 2022**

Name: **Vilas R Hegde**
USN: **4SU19CS111**

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, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – branch code, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate Prof, 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, forhelping 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.

Vilas Hegde
4SU19CS111

ABSTRACT

In contemporary era, the use of social media has reached unprecedented levels. Among all social media, Twitter is such a popular micro-blogging service, which enables users to share short messages in real time about events or express own opinion. In this paper, we examine the effectiveness of various machine learning techniques on retrieved tweet corpus. We apply machine learning model to predict tweet sentiment as well as find the correlation between twitter sentiment and stock prices. We accomplish this by mining tweets using Twitter's search API and process it for further analysis. To determine tweet sentiment, we test the effective two machine learning techniques: Naïve Bayes classification and Support vector machines. By evaluating each model, we discovered that support vector machine gives higher accuracy though cross validation. After predicting tweet sentiment, we have mined stock historical data using Yahoo finance API. We have designed feature matrix for stock market prediction using positive, negative, neutral and total sentiment score and stock price for each day. We have applied same machine learning algorithm to determine correlation between tweet sentiments and stock market prices and analyzed how tweet sentiments directly correlates with stock market prices.

Table of Contents

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of Compsoft Technologies

Compsoft Technologies, 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 clients requirements.

Compsoft Technologies, 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. Compsoft Technologies work with their clients and help them to defiine 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

2. ABOUT THE COMPANY

Products of Varcons Tech Pvt Ltd.

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 zutorials. 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 softwaredevelopment is possible by using specialized Android applications.

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 be considered as a specifific variant of client–server software where the client software is downloaded 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, and acts 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 specifified 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 allow 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.

Web design

It is encompasses 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 mark up 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.

Departments and services offered

Compsoft Technologies plays an essential role as an institute, the level of education, development of student's skills are 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 Compsoft Technologies 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 Compsoft Technologies 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 skill 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.

Services provided by Compsoft Technologies.

- 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

3. INTRODUCTION

Introduction to ML

Stock market prediction has been an active area of research for a long time. The Efficient Market Hypothesis (EMH) states that stock market prices are largely driven by new information and follow a random walk pattern. Though this hypothesis is widely accepted by the research community as a central paradigm governing the markets in general, several people have attempted to extract patterns in the way stock markets behave and respond to external stimuli. In this paper, we test a hypothesis based on the premise of behavioral economics, that the emotions and moods of individuals affect their decison making process, thus, leading to a direct correlation between ”public sentiment” and ”market sentiment”. We perform sentiment analysis on publicly available Twitter data to find the public mood and the degree of membership into 4 classes - Calm, Happy, Alert and Kind (somewhat like fuzzy membership).

Problem Statement

- How to utilize social media to assess market sentiment and predict the behavior of stock of certain company, market and stock indexes to identify an opportunity for trading?
- In order to design feasible system for identifying an opportunity for trading based on user sentiment. We have used twitter as our social media and analyze it to determine behavior of following stocks: Google, Yahoo, Apple, Microsoft and Goldman Sachs.
- We have chosen twitter as our social media because of its real time facet and importance of real time decision in trading. We have preferred to mine finance historical data from yahoo finance.
- The five stocks were determined based on opinion and mindshare on twitter, ensuring that we will have enough data model how their social data affect their stock.
- Last is to learn and investigate how machine learning techniques can be used to identify trends.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. Existing System

2. Proposed System

3. Objective of the System

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- Processor: Minimum 1 GHz; Recommended 2GHz or more
- Ethernet connection (LAN) OR a wireless adapter (Wi-Fi)
- Hard Drive: Minimum 32 GB; Recommended 64 GB or more
- Memory (RAM): Minimum 1 GB; Recommended 4 GB or above

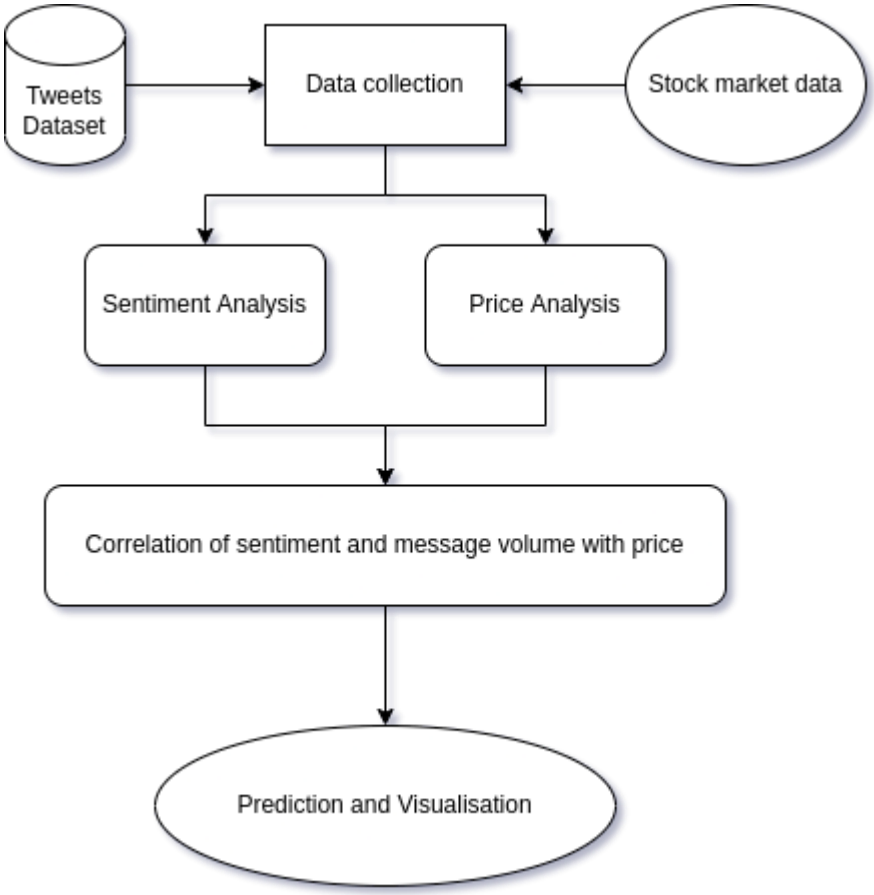
Software Requirement Specification

- Code editor (IDE)
- Operating System (Windows/Mac/Linux)
- Anaconda

CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS



This is a flowchart of Stock price prediction using Twitter Sentiment Analysis. By using Machine Learning Model of random forest regressor.

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

The application is clubbed in single file. Hence user do not need to run different file to run different tasks. Get twitter data and yahoo finance data from kaggle and yahoo finance for given company keyword and collect tweet for given company stock and store it in text file then process tweets, remote stop words and create feature set for training data.

Classifying training data and test classified model on test data for tweet sentiment analysis. Calculate total number of tweets, positive tweet and negative tweets and representing them in piechart. Classify stock prediction dataset and predict upcoming stock direction

TESTING

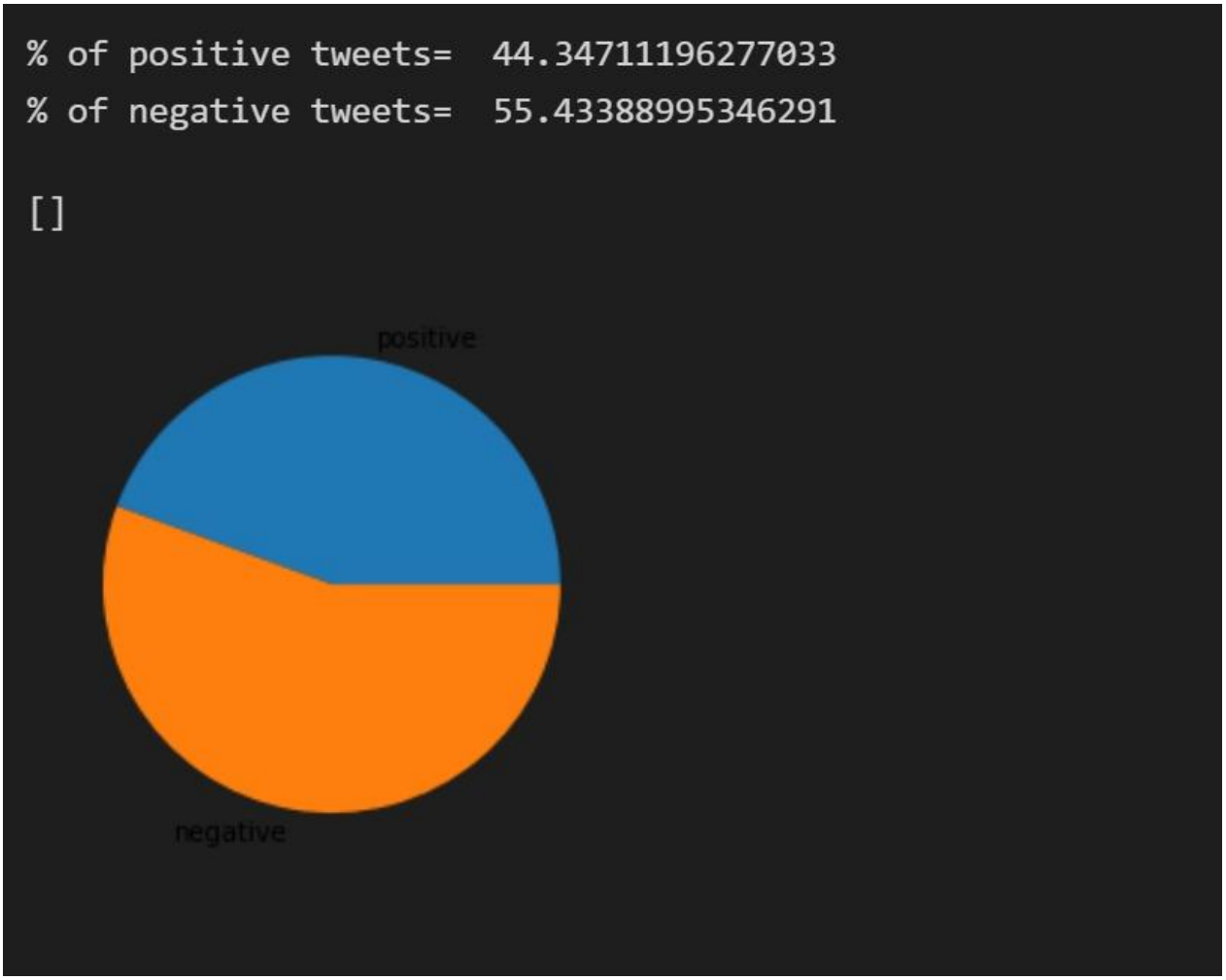
The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing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

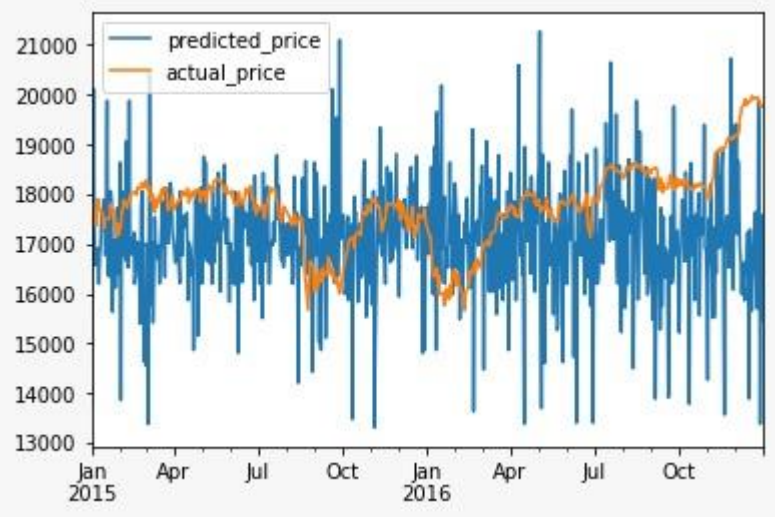
CHAPTER 8

SNAPSHOTS

8. SNAPSHOTS



This piechart shows the polarity of positive and negative tweets volume in the dataset.



This is the graph for total predictions vs the actual stock price for every month.

This is the graph which shows the accuracy of random forest regressot for training and testing data.

CHAPTER 9

CONCLUSION

9. 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.

10. REFERENCE

1. Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis.
   http://ieeexplore.ieee.org.ezproxy.gl.iit.edu/stamp/stamp.jsp?tp=&arnumber=6753954&tag=1
2. Stock Prediction Using Twitter Sentiment Analysis
   http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf
3. Large-scale Multiclass Support Vector Machine Training via Euclidean Projection onto the Simplex
   http://www.mblondel.org/publications/mblondel-icpr2014.pdf
4. Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation
   http://apps.cs.utexas.edu/tech_reports/reports/tr/TR-2124.pdf
5. https://github.com/ravikiranj/twitter-sentiment-analyzer