VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum-590018

Internship Report On

Stock Market Prediction Using Twitter Sentiment Analysis

SUBMITTED IN PARTIAL FULFILMENT FOR 6TH SEMESTER

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

SUBMITTED BY

Spoorthy R 1JS19EC139

Conducted at :

VARCONS Technologies Pvt Ltd

JSS ACADEMY OF TECHNICAL EDUCATION

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

JSS Campus , Dr.Vishnuvardhan Road,Bengaluru-560060

2022-23

J.S.S ACADEMY OF TECHNICAL EDUCATION

JSS Campus, Dr Vishnuvardhan Road, Bengaluru-560060.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that the Internship titled **"Stockport Predictive Sentiment Analysis"** carried out by **Miss Spoorthy R,** a bonafide student of JSS Academy of Technical Education, in partial fulfilment for the award of **Bachelor of Engineering**, in **Electronics and Communication 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 HOD

ıator

Dr. P M Shivakumaraswamy

HOD and Professor

Dept of ECE

JSSATEB

Signature of Principal

Dr.Bhimasen Soragaon

Principal

JSSATEB

EXTERNAL VIVA

Name of the examiners:	Signature with Date
1. _____	_____
2. _____	_____

D E C L A R A T I O N

I, **Spoorthy R**, final year student of Electronics and Communication Engineering, JSS Academy of Technical Education - 560060, declare that the Internship has been successfully completed, in **VARCONS Technologies Pvt Ltd**. This report is submitted in partial fulfilment of the requirements for award of Bachelor Degree in Electronics and Communication Engineering, during the academic year 2022-2023.

Date : 25-09-2022 :

Place : Bangalore

USN : 1JS19EC139

NAME : Spoorthy R

OFFER LETTER

Internship Offer Letter

Date: **23rd August, 2022**

Name: **Spoorthy R**
USN: **1JS19EC139**

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. **Dr.P M Shivakumaraswamy**, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

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.

NAME: Spoorthy R

USN: 1JS19EC139

Abstract

Stock markets prediction is considered a considerably demanding task due to its notable returns as well as due to the high randomness within the stock market. It is difficult to estimate due to their high volatility, which is influenced by a variety of political and economic influences, changes in politics, market sentiment, and a variety of other factors. Predicting asset markets solely on historical evidence or textual records has proved inadequate.

Moreover, stock price alternations are primarily related to the capital circumstances and hot occasions/events. Nowadays, researchers have sufficiently improved prediction accuracy by taking into consideration news and social media. However, the existing strategies do not employ the different impacts that events may pose. Streaming data proves to be a perpetual real-time source of data analysis as information from different web sources can be carried.

Machine learning algorithms have been used to design new methods for developing simulation models that can forecast stock markets and tell whether they will rise or fall. Here we explore whether estimations, in terms of sentiment analysis derived from Twitter posts, can be correlated to the stock market prices.

Twitter has been the key source for providing the sentiments of the people in the real time, which are helpful for analysing the stock market. But after certain event happened the sentiments of people became more predominant on the stock price. This paper describes about the stock prices changes over the company hashtag tweet sentiment score. We have over 557 tweets over the period of 10 days. We have analysed the stock market values collected from yahoo finance. We have done the keyword analysis of tweets to find the frequent keywords used during the period about the company.

Table of Contents

CHAPTER 1

COMPANY PROFILE

VARCONS TECHNOLOGIES PRIVATE LIMITED is a Private Limited Company registration held business entity registered under The Companies Act, 2013. It's registered for pre-defined objects or activity Computer and related activities as per Activity Code mentioned under Company CIN Number U72900KA2022PTC163646. Company has paid up capital Rs. 10000 and Authorised Capital Rs. 1000000. It is set up directly by registering the company with RoC, Ministry of Corporate Affairs. Its registered office address is 8/9 5th Main 3rd Cross road Beside Sachidananada Nagar RR Nagar Bangalore Bangalore KA 560098 IN. Private Limited Company registration is registered in 2022-07-11, India and Registrar of Company is RoC - Bangalore. it is Non-govt company. Private Limited Company registration Company has paid up capital Rs. 10000 and Authorised Capital Rs. 1000000. .

Status of Company is ACTIVE. VARCONS TECHNOLOGIES PRIVATE LIMITED registered state is Karnataka, India and company category is Company limited by Shares.

Varcons offers a wide range of IT Professional Services, providing innovative, customized, and cost-effective End-to-end Product support to the Commercial and Public sector clients. Our certified experts have decades of experience delivering IT solutions to the Public Sector and enterprise. We are not only the cloud-native solution provider; we are successfully stabilizing and modernizing the legacy applications. We provide a proven cloud journey from governance, security, and risk to automated deployments. We focus on results, as we help our customers to uncover new possibilities to put them on the leading edge of cloud innovation in their industry.

To leverage the benefit of cloud technologies, you need Cloud-experts who are successfully delivered these solutions. Varcons's *CloudThrottle* framework accelerates your cloud adoption without worrying about security or governance. Varcons CloudThrottle framework with self-service blueprints enables bundled cloud-services to provision automated governance, security, advanced analytics, and billing analytics. No matter where you are in your transformation journey, Varcons experts will help you to create a unique realistic technical solution to IT modernization strategy

CHAPTER 2

ABOUT THE COMPANY

VARCONS TECHNOLOGIES PRIVATE LIMITED is **2 months & 11 days** old Private company incorporated with MCA on **11th July, 2022**. VARCONS TECHNOLOGIES PRIVATE LIMITED is listed in the class of Private company and classified as Non-govt company.

The company has **2** directors/key management personal CHIKAEGOWDANADODDI KARIYAPPA SOMALATHA, HARALAHALLI CHANDRAIAH SPOORTHI. **VARCONS TECHNOLOGIES PVT LTD** company registration number is 163646 and its Corporate Identification Number(CIN) provided from MCA is **U72900KA2022PTC163646**.

Varcons Technologies Private Limited is involved in activities such as (Computer and related activities) Other computer related activities [for example maintenance of they websites of other firms/ creation of multimedia presentations for other firms etc.] It is a digital service provider that aims to provide software, designing and marketing solutions to individuals and businesses.

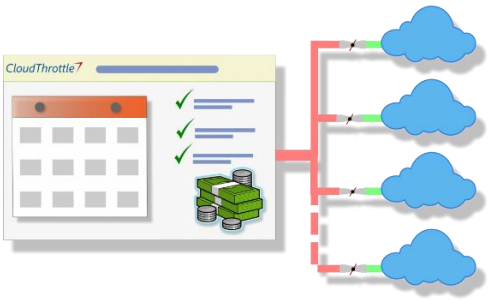
Varcons Technologies Private Limited's Annual General Meeting (AGM) was last held on N/A and as per records from Ministry of Corporate Affairs (MCA), its balance sheet was last filed on N/A. Directors of Varcons Technologies Private Limited are Haralahalli Chandraiah Spoorthi and Chikaegowdanadoddi Kariyappa Somalatha.

The Technologies is a leading provider of cutting-edge technologies and services, offering scalable solutions for businesses of all sizes. Founded by a group of friends who started by scribbling their ideas on a piece of paper, today they offer smart, innovative services to dozens of clients. They develop SaaS products, provide Corporate Seminars, Industrial trainings and much more

Smart solutions are at the core of all that they do at VCT. Our main goal is to find smart ways of using technology that will help build a better tomorrow for everyone, everywhere. SaaS offers a variety of advantages over traditional software licensing models and They here at VCT tend to include the key features of SaaS in everything they build.

2.1. Varcons' Case Study from Gov Loop in collaboration with AWS

(a)Budget Control and Management



Varcons' ***CloudThrottle*** framework enables cloud cost modeling, budget control, and management to give customers the tool they need for cloud financial management. Such as *comprehensive cost analytics*, *Cloud budget planning*, and a *cloud-cost billing model* that captures data across multiple cloud environments. It enables the customer to right-size resources and uses advanced analytics to break down costs by service, storage type, resources, and projects, putting that data into a single dashboard.

(b)AI-based budget monitoring and control engine

Maintaining budget control on 100's of accounts across multiple clouds is an operational overhead. To stay within the allocated budget, we need to track and control the Cloud usage costs at every project stage. Cloud budget control is a complicated process to manage and maintain across multiple projects, accounts, and multiple cloud vendors. Varcons' ***CloudThrottle*** *Patent-pending* AI-based budget monitoring and control engine continuously monitors and controls the allocated budget across multiple clouds to stay within ***Predictable Cloud Spending***.

2.2. Core Services

(a) Cloud Adoption and Migration

- Multi Cloud Infrastructure Design and Implementation using AWS, Azure & GCP
- Build CI/CD Pipeline and Deployment Automation
- Implement FedRAMP, NIST, internal enterprise compliance and guidelines
- Build Policies, Guardrails and Golden Images
- Cloud Operation and Maintenance (O&M)

(b) Cloud Cost Modelling

- Cloud Service costs breakdown report
- Cost trend across accounts, and portfolio
- Cost by storage, resources, service, etc.
- Cost burn-rate by portfolio, project, and accounts
- Cost on tagged resources vs. untagged resources
- Visualize historical usage data with multi-level drill down
- Anomaly Detection across multiple-accounts

(c) Application Modernization and Cloud Migration

- Application Portfolio assessment
- Application Stabilization
- Legacy Application Migration Services
- Architect Cloud Infrastructure and Integration
- Cloud-Native Development
- Architect and build disaster recovery on the cloud

(d) System and Application Monitoring

- Cloud Monitoring and Alerting
- Cloud and on-premises monitor Integration
- Heterogeneous monitoring system integration
- Synthetic Monitoring and alerting
- JVM Instrumentation and Analysis
- Design and develop custom Dashboards

- Design and develop centralized Log aggregator and repositories

(e) End-to-End Product Support

- Assess and analyse Legacy Applications
- Application Stabilization
- Application Modernization
- Legacy Application Migration to Cloud

2.3. Services provided by VACRONS technology

- **Analytics and Research**

  Let us analyze the way your users/customers interact with you/your business by gathering, studying, and understanding the consumer voice and their perception of the product/service to generate a report to help you make better market decisions
- **Comprehensive Customer Support**

  With a comprehensive range of services, they can guarantee your technology needs are not just met, but exceeded. They shall work with your Customers/users closely to understand the way your users/customers use/make use of Products/Services
- **Smart Automation Tools**

  They create API's and tools that help you automate any process with a host of features.

CHAPTER 3

INTRODUCTION

3.1. Introduction to ML

Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets. These procedures operate by construction of a model from example inputs in order to make data-driven predictions or choices rather than following firm static program instructions. "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." -- Tom Mitchell, Carnegie Mellon University. So, if they want our program to foresee, for example, traffic forms at a busy node (task T), they can run it through a machine learning process with data about previous traffic patterns (experience E) and, if it has successfully "learned", it will then do better at predicting upcoming traffic patterns.

Machine learning involves two types of tasks:

- Supervised Learning-The program is "trained" on a pre-defined set of: training examples ", which then facilitate its ability to reach an accurate conclusion when given new data.
- Unsupervised Learning-the program is given a bunch of data and must find patterns and relationships therein.

3.2. RANDOM FOREST ALGORITHM

Random forest is a Supervised Learning algorithm which uses ensemble learning methods for classification and regression. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Every decision tree has high variance, but when they combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that

particular sample data and hence the output doesn't depend on one decision tree but multiple decision trees . In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is Aggregation. A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. Random Forest has multiple decision trees as base learning models . They randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap. + Random forest is an ensemble of decision trees. This is to say that many trees, constructed in a certain "random" way form a Random Forest. + Each tree is created from a different sample of rows and at each node, a different sample of features is selected for splitting. + Each of the trees makes its own individual prediction. These predictions are then averaged to produce a single result.

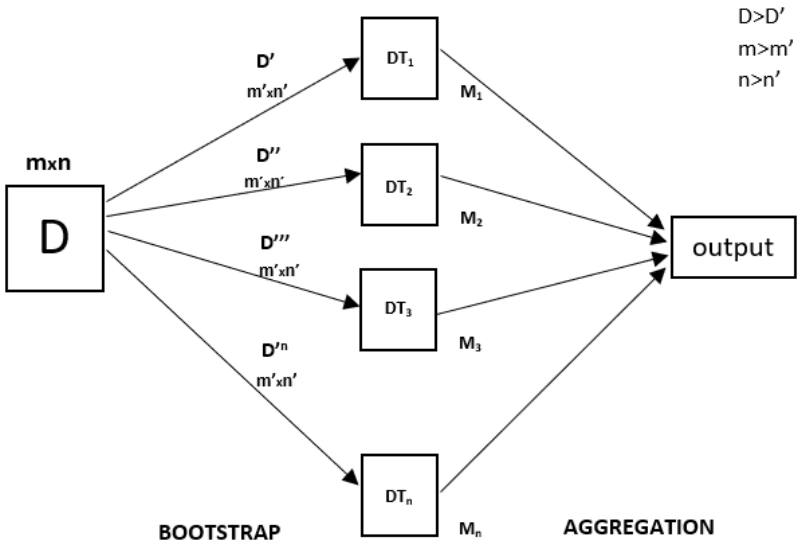


Fig 1 : Random Forest Algorithm

Problem Statement:

Stock Market Prediction Using Twitter Sentiment Analysis

Machine learning algorithms have been used to design new methods for developing simulation models that can forecast stock markets and tell whether they will rise or fall. Various sentiment analysis studies were carried out using algorithms such as support vector machines, Naive Bayes regression, etc. on various stages. The precision of machine learning algorithms depends on the amount of training data offered. In our paper, they try to maximize the prediction of stocks by collecting and reviewing data with the help of Twitter API using Random Forest Regressor.

In this paper, the Twitter Application Programming Interface (Twitter API), which offers a streaming API, has been taken into account in the study of financial data and continually returns the data. Each data collected reflects the user's status or attitude with respect to a specific subject. This is available through a basic HTTP authentication and a twitter account [8]. After all data is collected for every tick, an interpretation is initiated of the feelings relevant to each tweets and then a mood is predicted which has a direct impact on the stock status. Sentiment analysis is basically a problem of classification in which the data content is categorized with a positive or negative opinion [1]. Various models are developed based on various learning algorithms used for the training results. The streaming data is collected through the streaming API after such a model is prepared.

CHAPTER 4

SYSTEM ANALYSIS

4.1. Existing System

Twitter, a micro-blogging site, is a huge repository of public opinions expressed towards various people, services, organizations, products etc. Sentiment analysis is the process of analysing those public opinions. Sentiment analysis when combined with twitter gives useful insights into what is expressed on Twitter. It has applications in a number of domains like in stock market prediction, election results prediction, movie revenues, product reviews. Stock markets prediction is considered a considerably demanding task due to its notable returns as well as due to the high randomness within the stock market. Moreover, stock price alternations are primarily related to the capital circumstances and hot occasions/events. Nowadays, researchers have sufficiently improved prediction accuracy by taking into consideration news and social media. However, the existing strategies do not employ the different impacts that events may pose. Streaming data proves to be a perpetual real-time source of data analysis as information from different web sources can be carried.

In , authors built a system for stock prices prediction and proposed an approach that represents numerical price information by technical indicators by means of technical analysis. A deep learning model is developed to learn the sequential information within the market snapshot series.

The Granger causality test is used in where the relations between financial markets and Twitter for a 15 months period are investigated; concretely, the Twitter volume along with sentiments of the 30 large stock companies that form the Dow Jones Industrial Average (DJIA) index. Results depicted a relatively low Pearson correlation and Granger causality between the corresponding time series over the entire time period. However, the experiments showed that sentiment polarity of Twitter peaks suggests the direction of cumulative abnormal returns. Similarly, an investor sentiment proxy extracted from Twitter to investigate whether investor sentiment, as expressed in daily happiness, has predictive power for stock returns in 10 international stock markets, is utilized in . To account for complex correlations between stock returns and sentiment, a Granger non-causality test in quantiles is employed.

Twitter posts for a 6 months period were collected, and only a randomized sub-sample of the total number of tweets has been used in .

The emotions of hope and fear were measured on a daily basis with the aim of analysing the relationship between these indices and stock market indicators. Authors showed that sentiment tweet percentage was negatively correlated with stock markets, but shown a noteworthy positive association to VIX. In addition, in , approximately 250, 000 tweets related to stock prices were analysed and the results introduced a relationship among stock returns and tweets sentiment, trading and message volume, and finally volatility and disagreement. Also, authors in proposed a model for forecasting the movement of stock prices by incorporating the sentiments of the company's specific topics, derived from social media, into the stock prediction model. Comparing the average accuracy of 18 stock companies in transactions of a one year period, when only comparing the strategies for the stocks that are complicated to predict, the proposed method accomplished 9.83% better accuracy in contrast with the historical price method and 3.03% better accuracy than the human sentiment method.

Twitter's capacity of predicting consumer purchases, by noticing the association among societal Twitter trends and hourly stock prices of the top gainers and top losers of 10 companies concerning the technology sector. Experimental results depicted that the movements of stock prices are more rapidly predictive for Twitter sentiment movements. In addition, there is no significant prescient control of trending negative sentiment scores on stock markets concerning a particular subject. Moreover, introduced a novel methodology to determine investor sentiment retrieved from social media messages. More to the point, the connection between real time investor sentiment and intra-day stock returns was investigated. A words lexicon was initialized by employing a dataset with messages posted on the microblogging platform, namely Stock Twits; there, the terms are utilized by investors when they share opinions about the bearishness or the bullishness of the stock market.

Authors in identify directions for future ML stock market predictions based upon a review of current literature. A similar to the current work is the one explored in , where authors explore the effectiveness of social network analysis as well as sentiment analysis in predicting trends by mining publicly available online data sources. In our previous works, we have utilized cloud-based architectures aiming at creating sentiment analysis tools for Twitter data, based on Apache Spark framework .

4.2. Proposed System

The proposed system focuses on extraction of live tweets from the twitter and the extraction of the stocks of the particular company and continuing with sentimental analysis .

The live tweets have been extracted using the tweepy module from a particular date to present date. The dataset for stock of the company has been extracted from the Yahoo website from a particular date to present date.

The tweets along with the prices have been combined to a single data frame . The data is pre-processed . Comp, Negative, Neutral and Positive. Comp tells whether the sentence or the tweet is overall negative or positive. If the value of Comp is negative then, the sentence is negative and if the value of Comp is positive then, the sentence is positive. Vader (Valence Aware Dictionary and Sentiment Reasoner) a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media

4.3. Objective of the system:

- This paper attempts to design and implement a predictive system for guiding stock market investment.
- Predicting the movement of the stock market or particular stock of a company according to the live tweets from the Twitter using sentimental analysis.
- Our claim is that the sentiment analysis of live tweets has an impact on stock market values.

Chapter 5

REQUIREMENT ANALYSIS

5.1. Software Requirement Specification:

Minimum Software specification

- Operating System : Windows 10/10 pro/11
- Google Collab
- Sklearn Library

5.2. Hardware Requirement Specification:

Minimum hardware specification

- Processor : AMD/ I3 and above CORE PROCESSOR
- Main memory : 4GB and above
- Hard Disk : 1 Gb of hard disk space required

CHAPTER 6

DESIGN ANALYSIS

An API between Twitter and the system is established for the retrieval of the data. API stands for Application Programming Interface. In the context of APIs, the word Application refers to any software with a distinct function. Interface can be thought of as a contract of service between two applications. This contract defines how the two communicate with each other using requests and responses.

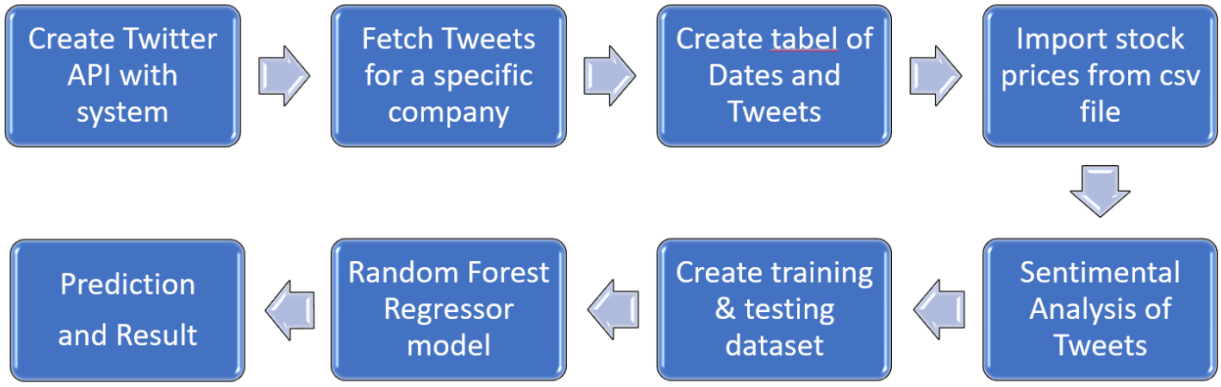


Fig.2: Flowchart of the proposed system

- Tweets are extracted from the data obtained through the API and a table is produced along with the corresponding dates of the tweets.
- The stock prices are accessed by the system from the CSV file and are correlated with the existing table of tweets data. Four new columns are added in the data frame. Comp, Negative, Neutral and Positive. Comp tells whether the sentence or the tweet is overall negative or positive. If the value of Comp is negative then, the sentence is negative and if the value of Comp is positive then, the sentence is positive.
- The data is pre-processed for the sentimental analysis and the training and testing data is split from the dataset.
- The datasets are trained using Random Forest Regressor model and the prediction is made.

CHAPTER 7

IMPLEMENTATION

7.1. Tweet Extraction

The first process is to extract the tweets from twitter. This takes place after setting up the consumer key and access token. After the tweets are fetched from twitter, special characters are removed from those tweets. The tweets are then displayed with their corresponding dates in the form of a data frame.

7.2. Dataset

After the extraction of tweets, historical data of that particular company or commodity is downloaded from the Yahoo Finance website. Yahoo Finance is a website that provides live stock prices of the company or commodity and also provides downloadable csv files of the historical data

7.3. Processing of Data

Price's column is then added to the data frame after the historical data is downloaded. Close Price of the company or the commodity is added to the Price column of the data frame. Some dates would not include any price due to some reasons like holidays. To fill in the values of the empty rows of the "Prices" column, mean of the available prices is determined and the empty rows are filled with this mean value.

7.4. Sentiment Analysis

Four new columns are added in the data frame. Comp, Negative, Neutral and Positive. Comp tells whether the sentence or the tweet is overall negative or positive. If the value of Comp is negative then, the sentence is negative and if the value of Comp is positive then, the sentence is positive. Vader (Valence Aware Dictionary and Sentiment Reasoner) a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media [7]. Approximately 88.88% positive tweets and 11.11% negative tweets were acquired by performing sentiment analysis using Vader Lexicon.

CHAPTER 8

SNAPSHOTS

```
% of positive tweets=  88.88888888888889
% of negative tweets=  11.11111111111111
[]
```

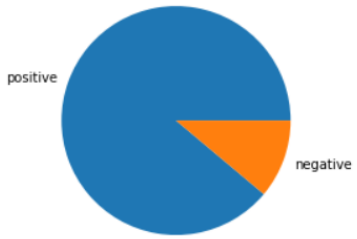


Chart - 1: Sentiment Analysis Pie Chart

```
m3=KNeighborsClassifier(n_neighbors=5)
apply_model(m3,numpy_df_train,numpy_df_test,y_train,y_test)
m3.fit(numpy_df_train,y_train)
ypred = m3.predict(numpy_df_test)
print("The Prediction's generated are\n",ypred)
print("The Training Score Generated are  ",m3.score(numpy_df_train,y_train))
print("The Testing Score Generated are  ",m3.score(numpy_df_test,y_test))
```

```
The Prediction's generated are
 [37 37 37]
The Training Score Generated are   0.6666666666666666
The Testing Score Generated are   0.3333333333333333
The Confusion Matrix is
 [[0 1 0]
 [0 1 0]
 [0 1 0]]
The Classification Report is
              precision    recall  f1-score   support

          36       0.00      0.00      0.00         1
          37       0.33      1.00      0.50         1
          38       0.00      0.00      0.00         1

    accuracy                           0.33         3
   macro avg       0.11      0.33      0.17         3
weighted avg       0.11      0.33      0.17         3
```

Chart – 2: Implementing using KNN Classifier

```
m2=RandomForestClassifier(n_estimators=100,criterion="gini",max_depth=20,min_samples_split=30)
apply_model(m2,X_train,X_test,y_train,y_test)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: DataConversionWarning: A column-vector

The Prediction's generated are
 [0 0 0 1 1 1 1 0 1 1 0 0 0 0 0 0 1 1 0 0 0 0 1 1 0 1 0 1 1 1 0 0 1 1 1 0 1
 0 0 1 1 1 0 1 0 0 1 0 1 0 0 1 1 0 1 1 1 1 0 0 1 1 0 1 1 1 1 1 0 1 0 0 0 1
 1 1 1 1 0 1 1 1 1 0 1 1 1 0 1 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 1
 1 1 1 1 1 0 0 0 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 0 0 1 1 0 0 1 1 1 1 1 1 0 1 0 1 1]
The Training Score Generated are   0.813858695652174
The Testing Score Generated are   0.5760869565217391
The Confusion Matrix is
 [[32 50]
 [28 74]]
The Classification Report is
              precision    recall  f1-score   support

           0       0.53      0.39      0.45        82
           1       0.60      0.73      0.65       102

    accuracy                           0.58       184
   macro avg       0.57      0.56      0.55       184
weighted avg       0.57      0.58      0.56       184
```

Chart – 3: Implementation using Random Forest Classifier

```
model = RandomForestRegressor(n_estimators=500, random_state=42, min_samples_split=2, min_samples_leaf=1, max_depth=10, bootstrap=True)
model.fit(numpy_df_train, y_train)
predict = model.predict(numpy_df_test)
print(predict)
print(predict.shape)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: DataConversionWarning: A column-vector y was passed when a 1d array was

[37.222 38.168 37.066]
(3,)
```

```
[ ] print("ACCURACY= ",(rf.score(numpy_df_train,y_train))*100,"%")

ACCURACY=  92.64285714285718 %
```

Chart – 4 : Final Results using Random Forest Regressor

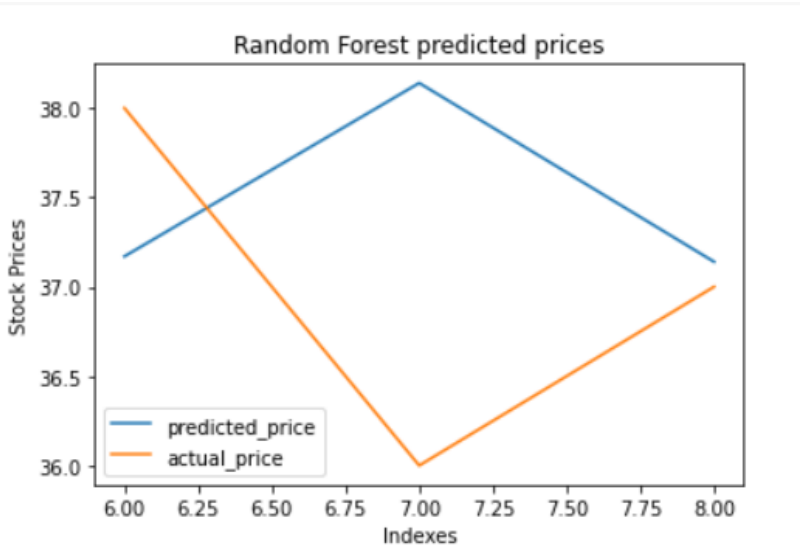


Chart – 5: Random Forest Prediction Graph

Chart -6: Final Graph

Chapter 9

Results and Conclusion

9.1. RESULT AND DISCUSSIONS

United Airlines current share price was obtained from Yahoo Finance page, which is treated as a basis for success measurements. The predicted stock price for the testing data is compared with the actual price. The data collected includes 557 tweets, ranging from 4/09/2022 up to 13/09/2022, over ten days of Twitter data, and is saved in a csv format. The collection includes a set of positively and negatively graded ratings of the company. Twitter data collected over the first five days is taken as the training dataset. And for testing, the remaining days are used. Due to the limitation of data available from Twitter API, we have to jump to historical dataset. We have acquired an accuracy of **92.64%** using historical dataset.

9.2. CONCLUSION

Various machine learning algorithms have been used to design new methods for developing simulation models that can forecast stock markets and tell whether they will rise or fall. Here we explore whether estimations, in terms of sentiment analysis derived from Twitter posts, we have correlated to the stock market prices using K Neighbor Classifier, Randon Forest Classifier and Random Forest Regressor models whose obtained results were 66.66% ,81.38% and 92.64% respectively.

The **Random Forest Regressor** out performed in predicting future data prices. Dataset is built on the basis of live results. With the assistance of Yahoo Finance, the live data and historic data is retrieved. Also, with support of twitter APIs, required for sentiment analysis is obtained. This algorithm provided an accuracy of **92.64%**. The model was not effective in situations of low or high volatile stock values. There are already various approaches to design stock models, which they leave as work to be done in the future. Some of them involve developing a business model by grouping firms based on their business, taking account of adverse impact on a company's stock price due to news about other similar businesses, and examining more general industry and global news that could suggest general stabilization of the market

Chapter 10

References

- A. Balahur, MijailKabadjov, J. Steinberger, R. Steinberger, A. Montoyo, "Challenges and Solutions in the Opinion Summarization", Journal of Intelligent Information Systems, Springer, Vol.39, No.2, pp. 375-398, 2012.
- Q. Li, Y. Chen, L. L. Jiang and P. Li and H. Chen, "A Tensor-Based Information Framework for Predicting the Stock Market", ACM Transactions on Information Systems, Vol.34,No.2, pp.1-30, 2016
- D. Nguyen and M. Le, "A Two-Stage Architecture for Stock Price forecasting by combining SOM and fuzzy-SVM", International Journal of Computer Science and Information Security, Vol. 12, No. 8, pp. 1-6, 2014. [4] B. Johan, H. Mao and X. Zeng, " Twitter Mood Predicts The Stock Market", Journal of Computational Science, Vol.2, No.1, pp.1-8, 2011
- A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks, ser. Studies in Computational Intelligence. Springer, 2012, vol. 385.
- Y. Singh and A. S. Chauhan, "Neural Networks in Data Mining," Journal of Theoretical and Applied Information Technology, pp. 37-42, 2009.
- Predicting Stock Price using RNN by Lilian Weng, Unpublished.
- A. V. Devadas and T. A. A. Ligori, "Forecasting of stock prices using multi layer perceptron," Int J Comput Algorithm, vol. 2, pp. 440–449, 2013.
- Hiroshi, K., Tetsuya, N. and Hideo, W., 2004. "Deeper sentiment analysis using machine translation technology", In Proceedings of the 20th International Conference on Computational Linguistics, pp. 1-7.
- [4] Boiy, E. and Moens, M.F., 2009. "A machine learning approach to sentiment analysis in multilingual Web texts", Information retrieval, Vol. 12, No. 5, pp.526-558.
- [5] Schrauwen, S., 2010. "Machine learning approaches to sentiment analysis using the Dutch Netlog Corpus", Computational Linguistics and Psycholinguistics, Technical Report Series, CTRS-001.
- Blinov P. D., Klekovkina M. V., Kotelnikov E. V., and Pestov O. A., 2013. "Research of lexical approach and machine learning methods for sentiment analysis", Computational Linguistics and Intellectual Technologies, Vol. 2, No. 12, pp. 48–58.
- Almatrafi, O., Parack, S. and Chavan, B., 2015. "Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014", In

Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, ACM, Bali, Indonesia, pp. 1–5

- W. Antweiler and M. Z. Frank. Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance, 59(3):1259–1294, 2004.
- A. Baltas, A. Kanavos, and A. Tsakalidis. An apache spark implementa tion for sentiment analysis on twitter data. In 1st International Workshop on Algorithmic Aspects of Cloud Computing (ALGOCLOUD), pages 15–25, 2016.
- D. Harper, Forces that Move Stock Prices. (2010) [Online] Available at: http://www.investopedia.com/articles/basics/04/100804.asp [Accessed on 27th October 2016]
- B. Liu, Sentiment analysis and opinion mining, Morgan and Claypool publishers, 2012