MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Final Year Internship Report

on

Neural Network Developer at Praedico Global Research Pvt. Ltd.

Submitted By:
YASH JAIN
0901CS181123

Faculty Mentor:

Mir Shahnawaz Ahmad

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

A final year internship report submitted in partial fulfillment of the requirement for the

degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Yash Jain

0901CS181123

Internship Faculty Mentor:

Mir Shahnawaz Ahmad

Assistant Professor

Submitted to:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

Address: First Floor, Garima Arcade, Shinde ki Chhawani, Gwalior
Email: mail@praedicoglobalresearch.com
Website: www.praedicoglobalreserach.com

Ref.: PGR-2022/I-513

Date: 11th – May - 2022

CERTIFICATE OF INTERNSHIP

This certificate is awarded to

Mr./Miss. YASH JAIN

In appreciation for your accomplishments in the company as an intern

(Position titled- *"Neural Network Developer"*)

at **Praedico Global Research Pvt. Ltd.,**

from **Jan 10th, 2022 to May 10th, 2022**.

We take this opportunity to wish you a long, happy and successful career.
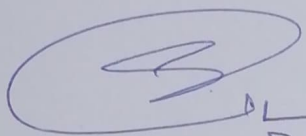
Authorized Signatory
Praedico Global Research Pvt. Ltd.

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to
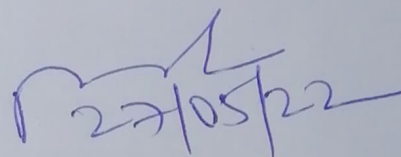RGPV, Bhopal)

CERTIFICATE

This is certified that **Yash Jain** (0901CS181123) has submitted the Internship report titled **Neural Network Developer at Praedico Global Research Pvt. Ltd.** of the work he has done under the mentorship **of Mir Shahnawaz Ahmad**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

Mir Shahnawaz Ahmad

Assistant Professor

CSE

Dr Manish Dixit

Professor and Head

CSE

Dr. Manish Dixit
Professor & HOD
Department of CSE
M.I.T.S. Gwalior

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to
RGPV, Bhopal)

DECLARATION

I hereby declare that the work being presented in this Internship report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in CSE at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Mir Shahnawaz Ahmad,Assistant Professor**, Department of CSE.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Yash Jain

0901CS181123

IV Year

Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to
RGPV, Bhopal)

ACKNOWLEDGEMENT

The full semester internship has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary internship as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering, for allowing** me to explore this internship. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Mir Shahnawaz Ahmad**, Assistant Professor, Department of Computer Science and Engineering, for his continued support and close mentoring throughout the internship. I am also very thankful to the faculty and staff of the department.

Yash Jain

0901CS181123

IV Year

Computer Science and Engineering

ABSTRACT

The stock maret is very complex and volatile. It is impacted by positive and negative sentiments which are based on media releases. Publication of financial news articles , impact the decision made by investors and , therefore , change the market state.It makes them important source of data for financial predictions. Some of the company specific factors that can affect the share price like news releases on earnings and profits , future estimated earnings , announcement of dividends etc. News articles and social media data plays very important role in financial predictions , but the most important thing in that is how to collect these news and use this for future financial predictions. In this Internship we proposed a method or algorithm that can successfully fetch news from various online financial websites and along with this it lables the news i.e. how this news affect the share price of particular company as positive, negative or neutral.

Keywords : Stock Market , Predictions , News , Financial

List of Figures

CONTENTS

Chapter I: Introduction

1.1 Overview:

In recent past years, the rise of the stock market is remarkable. With the rise in stock market, more and more people are ready to invest in stock market. With the increasing interest of people makes this topic more important for research. It is the place where buyers and sellers come for the purpose of buying and selling stocks and make profit. It is becoming the most vital part of counry's economy as it is a way to make investment and make higher capital.Also to make higher profit one must have in depth knowledge of market as sometime in lieu of making profit investors lost their money.

Prediction in stock market can be of great importance for the investors if it is made correctly. The efficient prediction in stock market may offers invertors a helpful guidance in order to take the appropriate decisions and measures whether to buy or sell shares. In recent years , many methods for predicting stocks market have been introduced , it involves **fundamental analysis** which is based on published financial statements. Second one is **technical analysis** which involves technical analysis of stocks based on its previous perfomance or charts.Third one is **ML methods** applied on huge historical data and prices. The last one is **sentiment analysis** which involves prediction with the help of published news , articles or blogs.

This paper focuses on the generation of real time dataset containing financial news along with its label (positive , negative or neutral). In recent years many

studies related to stock market prediction has been proposed which primarily focus on building machine learning and neural network based models and improving their accuracy. In all these studies the dataset taken is not real time. As it is known that for the predictions of stock movement real time data as well as huge amount of data is needed all these studies are based on old historical data which make that models less effective for real time data. In this paper two algorithms is proposed first one is responsible for fetching current rates of stock and second one is responsible for fetching realtime news from various online resources and labeled it as positive , negative or neutral. The benefits of this realtime dataset is that when machine learning models are applied to these collected realtime data it becomes more effective and accurate.

Sentiment Analysis helps us to find or understand emotion behind the text data . It also helps us to understand whether the given text has positive , negative or neutral tone. It plays important role to understand customer feedback about variety of topics , products or anything. The most important part in sentiment analysis is text preprocessing which involves variety of steps to be performed in text before it is being given to machine for learning. These steps are basically used to prepare data. These steps are : lower casing of words, punctuation removal, stopwords removal , tokenization , stemming , lemmatization etc.

1.2 About the Company

Praedico Global Research Private Limited is a private company which is founded on 09 April, 2018. It deals in Stock Market Training, Stock Marketing Predictive Softwares, Robotic Stock Trading, Global Equity Research, Portfolio Designing, Financial Literacy and Stock Market Research Using Deep Learning. Praedico Global Research Pvt. Ltd. is India's first "coordinated worldwide research cum preparing" company. They take a shot at the model of spreading financial literacy

all over the globe and have their own exploration model for India and worldwide stock trades.Praedico Global Research focuses on the use of artificial intelligence to forecast the trends in the stock markets across the globe. The company has achieved many successes within the limited period of time it has been in existence.

1.3 Learning Objectives/Internship Objectives:

- ✓ Internships are generally thought of to be reserved for college students looking to gainexperience in a particular field. However, a wide array of people can benefit from Training.
- ✓ Internships in order to receive real world experience and develop their skills.
- ✓ An objective for this position should emphasize the skills you already possess in the areaand your interest in learning more.
- ✓ Internships are utilized in a number of different career fields, including architecture, engineering, healthcare, economics, advertising and many more.
- ✓ Some internship is used to allow individuals to perform scientific research while othersare specifically designed to allow people to gain first-hand experience working.
- ✓ Utilizing internships is a great way to build your resume and develop skills that can be emphasized in your resume for future jobs. When you are applying for a Training Internship, make sure to highlight any special skills or talents that can make you stand apart from the rest of the applicants so that you have an improved chance of landing the position.

1.4 Software Requirements and Specification

For the above proposed method below are the software and hardware

Software Requirements

- Operating System : Windows
- Coding Language : Python

- Platform : Jupyter Notebook and Google Colab
- IDE : Visual Studio Code
- Modules : Pandas , Numpy , sklearn , keras and beautifulsoup

Hardware Requirements

- Processor : Intel(R) Core(TM) i5
- Memory : 8GB
- Hard Disk : 1TB

1.5 Outcome

In this internship we proposed a method which is capable of fetching realtime news and label its sentiments.The dataset generated by this algorithms helps to achieve higher accuracy in predictions.Also this algorithm can be used to make progressive machine learning and neural network model which will give prediction on real time basis.

Chapter II: Literature Review

2.1 Existing System

Prediction in stock market can be of great importance for the investors if it is made correctly. The efficient prediction in stock market may offers invertors a helpful guidance in order to take the appropriate decisions and measures whether to buy or sell shares. In recent years , many methods for predicting stocks market have been introduced , it involves **fundamental analysis** which is based on published financial statements. Second one is **technical analysis** which involves technical analysis of stocks based on its previous perfomance or charts.Third one is **ML methods** applied on huge historical data and prices. The last one is **sentiment analysis** which involves prediction with the help of published news , articles or blogs.

2.2 Critical Analysis of Existing System

In all these methods the one primary thing that is very important is Data . For the prediction in stock market the most important thing is the real and accurate data. As these prediction has to be made on real time data so there is a need to collect these data.All the existing proposed methods are based on very historical or old data , all these proposed system primarily focused on improoving machine learning algorithm or creating more complex systems . But the most important thing is that we feed most recent and real time data to our model it accuracy will itself improove .

Chapter III : Work Done During Internship

3.1 Proposed System

We focuses on the generation of real time dataset containing financial news along with its label (positive , negative or neutral). In recent years many studies related to stock market prediction has been proposed which primarily focus on building machine learning and neural network based models and improoving their accuracy. In all these studies the dataset taken is not real time. As it is known that for the predictions of stock movement real time data as well as huge amount of data is needed all these studies are based on old historical data which make that models less effective for real time data. In this paper two algorithms is proposed first one is responsible for fetching current rates of stock and second one is responsible for fetching realtime news from various online resources and labeled it as positive , negative or neutral. The benefits of this realtime dataset is that when machine learning models are applied to these collected realtime data it becomes more effective and accurate.

3.2 Modules Used

1.Data Collection Module:

The aim of this module is able to fetch realtime news form online websites then label this news as positive , negative or neutral. In this module there are two sub modules :

1. Responsible for creating csv files of companies with the following parameters open, close, dayhigh, daylow, high52, low52, etc.

2. Responsible for fetching news from online website then fetching names of the companies present in the news and label news as positive or negative.

2.Basic Feature Extraction Module:

In this modules basically we try to capture or extract some feature or underlying pattern from text data these includes:

1. Number of words

2.Number of characters

3.Average Word Length

4.Number of Stopwords

5.Number of Special Characters

6.Number of Numerics

7.Number of Uppercase Words

3.Basic Pre-Processing Module:

In this module we basically pre process our collected text data so that it become ready to feed into our machine learning algorithm . This includes basic pre-process steps:

1.Lowercase

2.Removing Punctuation

3.Removal of stop words

4.Common Word Removal

5.Rare words Removal

6.Spelling Correction

7.Tokenization

8.Stemming

9.Lemmatization

4.Text Vectorization Module

After the completion of above two modules there is a need to convert collected text data into the vector form so that machine learning algorithm can work on that . We used two main approaches here:

1.Bag of Words

2.Tf- Idf Vectorization

5.Model Building

This is our last module in this we apply machine learning algorithm to our collected text data . Here we have used LSTM algorithm for predictions.

3.3 Flowchart
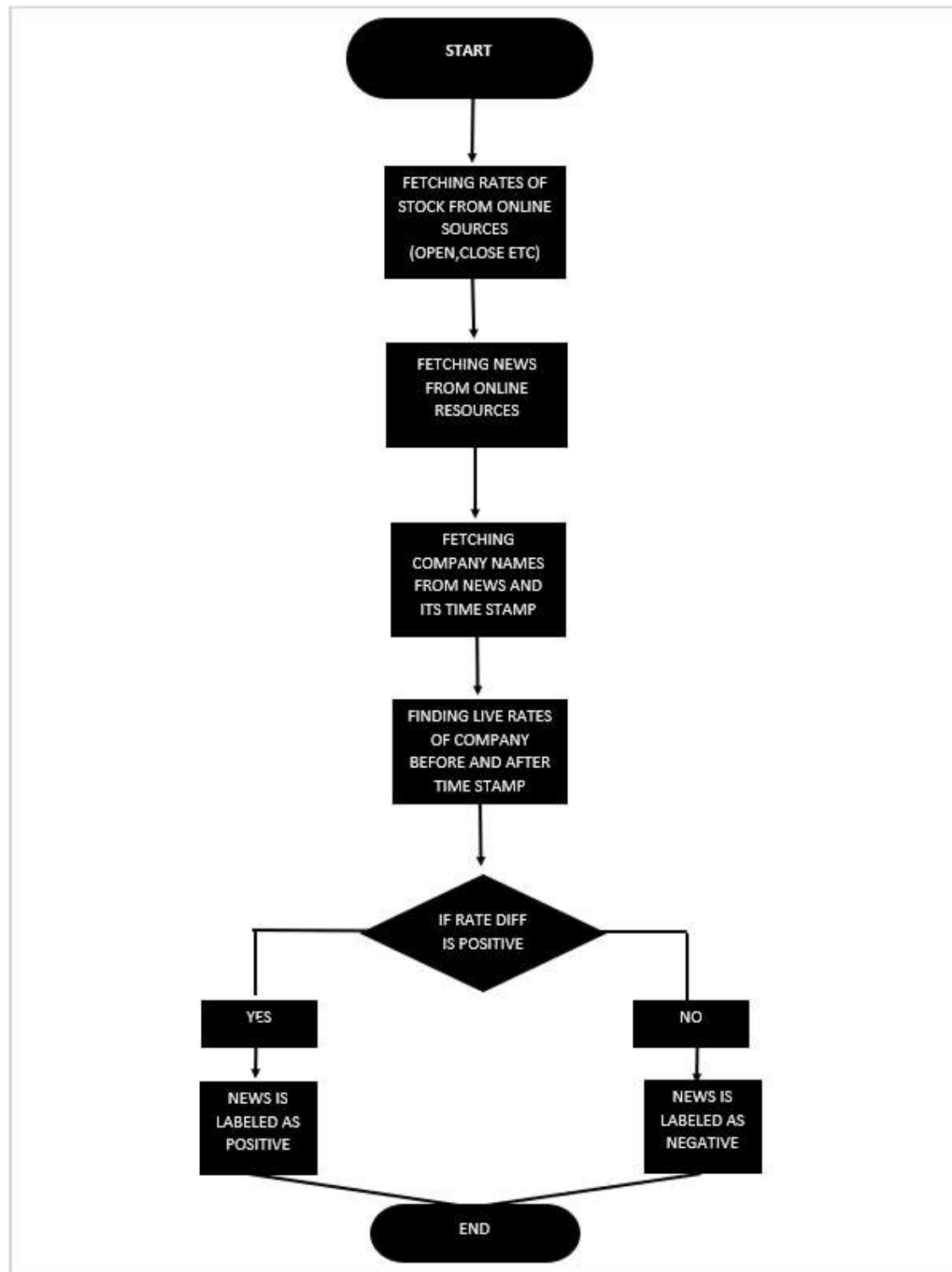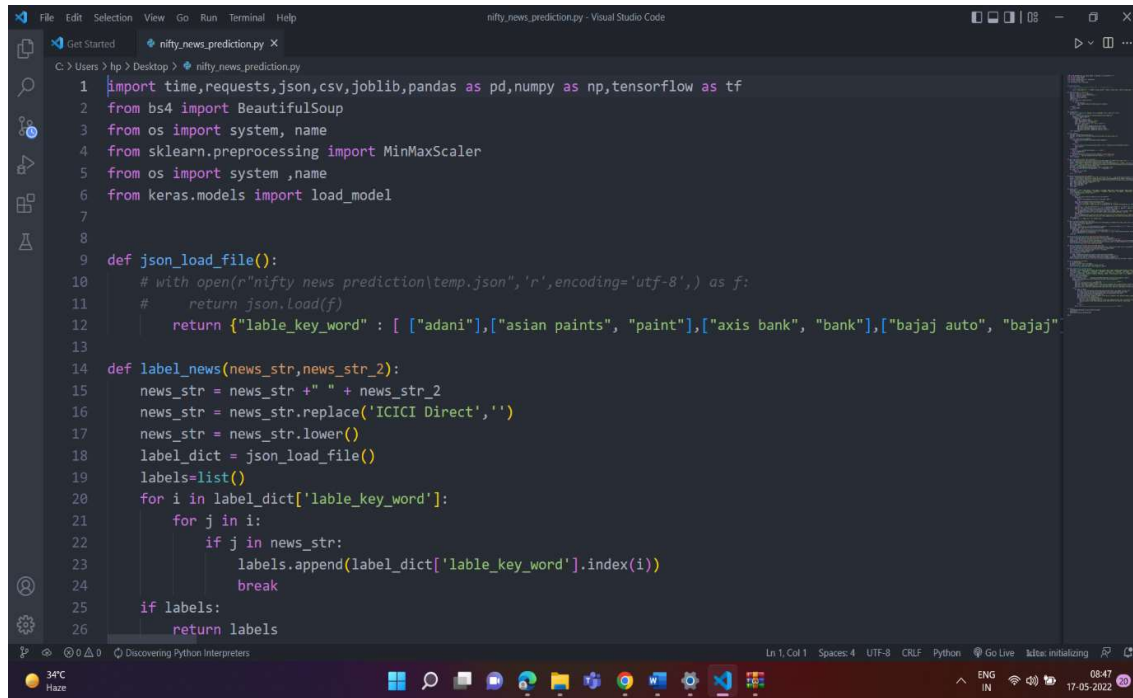
Fig 3.3.1 : Flowchart representing flow of work done

3.4 Technology

- Python 3.8
- BeautifulSoup
- requests
- Pandas
- Numpy
- Matplotlib
- Seaborn
- Sklearn
- Keras
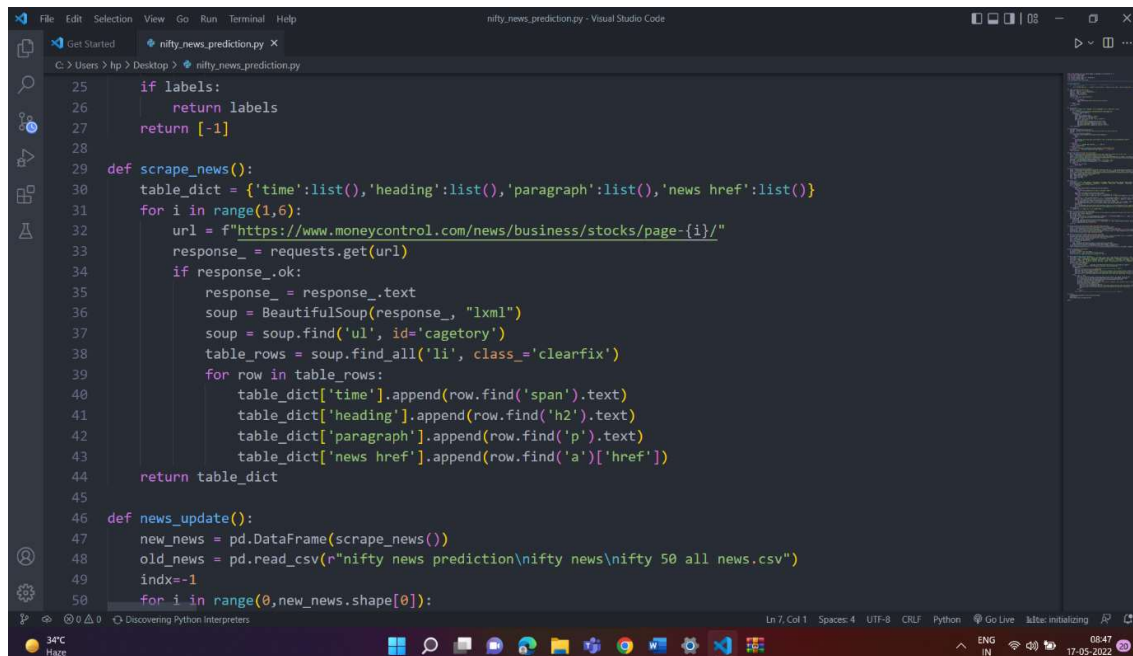
Chapter IV: Coding and Results

4.1 Coding

Fig 4.1.1 : Funnction used inside project

Fig 4.1.2 : Funnction used inside project

Fig 4.1.3 : Funnction used inside project

**Fig 4.1.4 : Funnction used inside project**

4.2 Results

For a particular news this is how information is stored and predicted

NEWS
Tata Motors share price slips 1% after CLSA cuts rating, lowers target price

Company Name	Time Stamp(TS)	Rate at TS	Actual rate after 10 min	Predicted rate after 10 min	Error	Difference (Pred - TS)	Label	Error %(abs)
Tata Motors	10:00 AM	430	433	432.70	0.70	+2.70	+ve	+- 10%

	time	heading	paragraph	news href	nifty name	company	sector	price 0 min	predict 5 min	predict 10 min	predict 15 min
0	May 05, 2022 02:07 PM IST	Buy Tanla Platform; target of Rs 1867: YES Sec...	YES Securities is bullish on Tanla Platform ha...	https://www.moneycontrol.com/news/business/sto...	out of nifty 50	other	other	0.0	0.00	0.00	0.00
1	May 05, 2022 02:03 PM IST	Buy Sonata Software; target of Rs 900: ICICI D...	ICICI Direct is bullish on Sonata Software has...	https://www.moneycontrol.com/news/business/sto...	out of nifty 50	other	other	0.0	0.00	0.00	0.00
2	May 05, 2022 02:01 PM IST	Hold Hero MotoCorp; target of Rs 2495: ICICI D...	ICICI Direct recommended hold rating on Hero M...	https://www.moneycontrol.com/news/business/sto...	Hero Motocorp	Hero Motocorp Ltd.	Auto - 2 & 3 Wheelers	2492.0	2502.42	2501.58	2501.31

Fig 4.2.1 : Table shows prediction of a price after fetching news by taking its time stamp into consideration.

4.3 Prediction Graphs
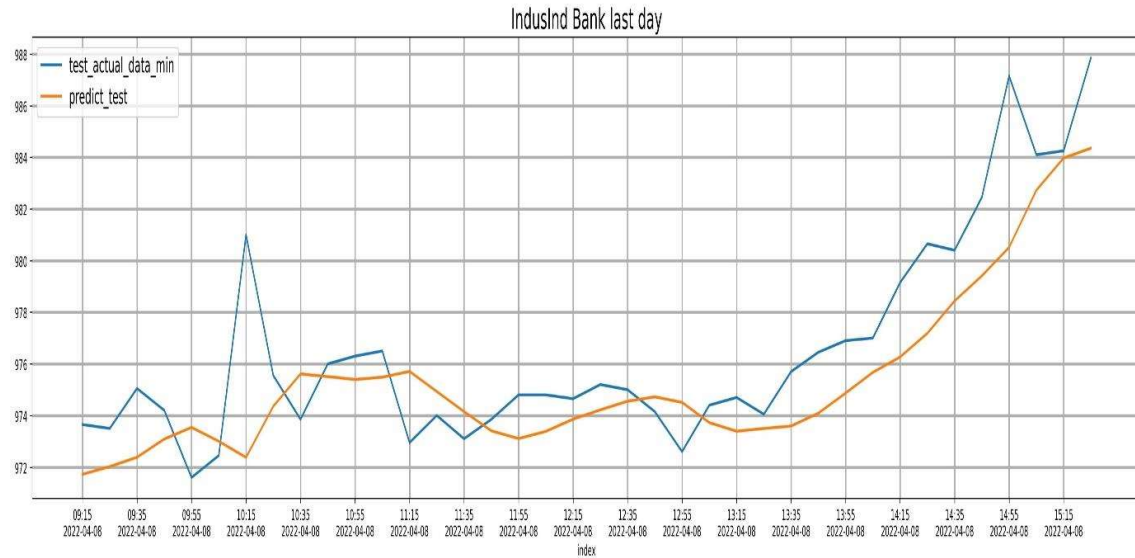
Fig 4.3.1 : This Fig represents actual and predicted price for a company (Indusland Bank). This chart actually shows price of a stock in 20min interval where the orange curve shows the prediction curve here algorithm used is LSTM.
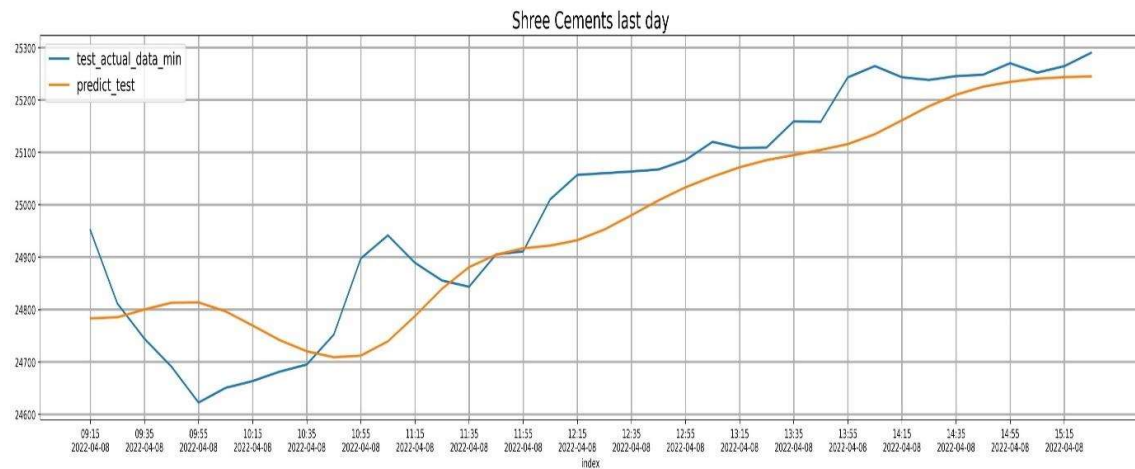
Fig 4.3.2 : This Fig represents actual and predicted price for a company (Shree Cements). This chart actually shows price of a stock in 20min interval where the orange curve shows the prediction curve here algorithm used is LSTM.
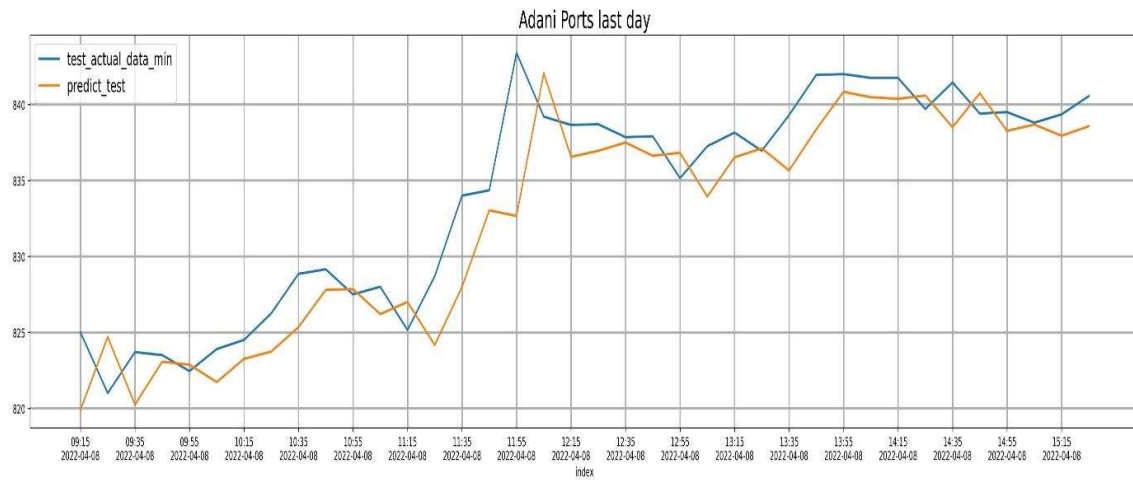
Fig 4.3.3 : This Fig represents actual and predicted price for a company (Adani Ports). This chart actually shows price of a stock in 20min interval where the orange curve shows the prediction curve here algorithm used is LSTM.
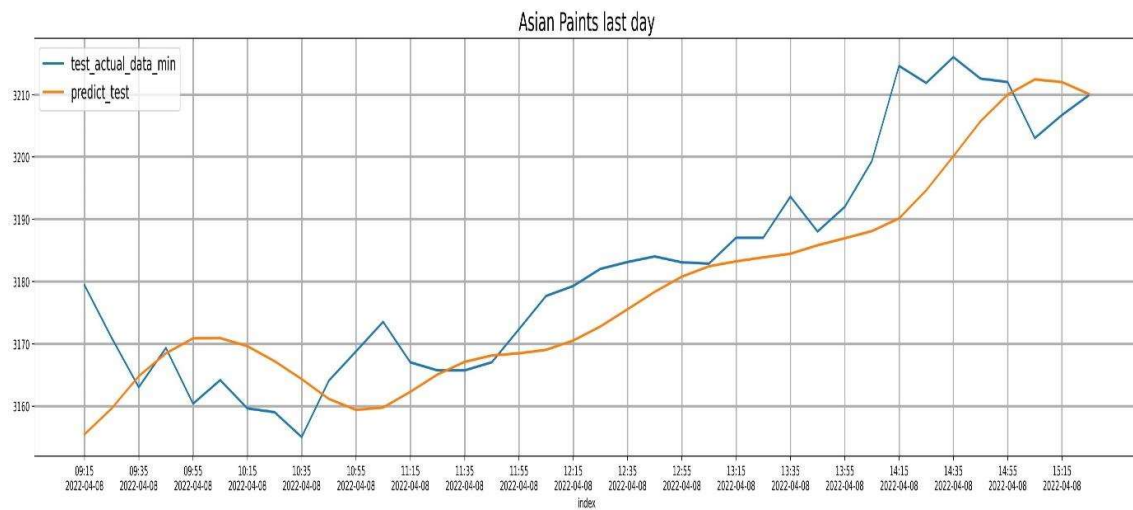

Fig 4.3.4 : This Fig represents actual and predicted price for a company (Asian Paints). This chart actually shows price of a stock in 20min interval where the orange curve shows the prediction curve here algorithm used is LSTM.

Chapter V : Conclusion and Future Work

5.1 Conclusions

In this internship the focus is on collecting data rather than applying algorithm as we know that data is the most important thing for accurate predictions. In this internship we proposed a method which is capable of fetching realtime news and label its sentiments.The dataset generated by this algorithms helps to achieve higher accuracy in predictions.Also this algorithm can be used to make progressive machine learning and neural network model which will give prediction on real time basis.
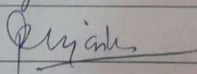
5.2 Future Work

Enhancing the above proposed methodlogy by incorporating more collection sources so that the dataset generated become more versatile. Also to used a Fuzzy Logic system for discrete prediction like to tell for a particular news whether the stock of the company mention in this news is likely to buy , sell and hold.

Refrences

[1] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using contextcapturing features," Decision Support Systems, vol. 55, no. 3, pp. 685– 697, 2013.

[2] Ayman E. Khedr, S.E.Salama, Nagwa Yaseen,"Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis", International Journal of Intelligent Systems and Applications(IJISA), Vol.9, No.7, pp.22-30, 2017. DOI: 10.5815/ijisa.2017.07.03

[3] Adusumilli, R.: Predicting stock prices using a keras LSTM model. NikolaNews (2019)

[4]. Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012).

[5] armar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock market prediction using Machine Learning. Paper presented at the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).
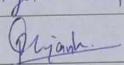
FPR Reports

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	YASH JAIN		Department	CSE	
Industry/Organization	Praeduo Global Research Pvt. std.		Date/Duration	10-Jan-2022 to 15-Feb-2022	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work				✓	
Learning capacity/Knowledge up gradation			✓		
Performance/Quality of work				✓	
Behaviour/Discipline/Team work				✓	
Sincerity/Hard work			✓		
Comment on nature of work done/Area/Topic	ML & NN NLP Stock Market.				
OVERALL GRADE (Any one)	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT ✓				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor	Priyank				
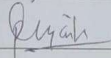
Receiving Date	12/02/2022	Name of Faculty Mentor	Mir Shahnawaz Ahmad	Sign	

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	YASH JAIN		Department	CSE	
Industry/Organization	Praedico Global Research Pvt Ltd		Date/Duration	16 - Feb - 2022 to 28 - Feb - 2022	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge up gradation					✓
Performance/Quality of work				✓	
Behaviour/Discipline/Team work				✓	
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	ML & NN Model Building Stock Market				
OVERALL GRADE (Any one)	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor					
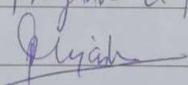
Receiving Date	02/03/22	Name of Faculty Mentor	Mir Shahnawaz Ahmad	Sign	
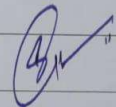
FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	YASH JAIN		Department	CSE	
Industry/Organization	Praedico Global Research Pvt Ltd		Date/Duration	1-Mar-2022 to 15-Mar-2022	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge up gradation				✓	
Performance/Quality of work				✓	
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	ML &NN Model Building & Evaluation				
OVERALL GRADE **(Any one)**	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor					

Receiving Date	16/03/22	Name of Faculty Mentor	Mir Shahnawaz Ahmad	Sign	

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	YASH JAIN		Department	CSE		
Industry/Organization	Praedico Global Research Pvt ltd		Date/Duration	15- Mar - 2022 to 31- Mar - 2022		
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**	
Punctuality/Timely completion of assigned work					✓	
Learning capacity/Knowledge up gradation					✓	
Performance/Quality of work					✓	
Behaviour/Discipline/Team work					✓	
Sincerity/Hard work					✓	
Comment on nature of work done/Area/Topic	Madel Testing & Improvement					
OVERALL GRADE (Any one)	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT					
Name of Industry Mentor	Priyank Gupta					
Signature of Industry Mentor						

Receiving Date	30/03/2022	Name of Faculty Mentor	Mr. Mir Shahnawaz Ahmad	Sign	
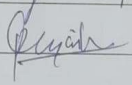
FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Yash Jain		Department	CSE	
Industry/Organization	PRAEDICO GLOBAL RESEARCH PVT. LTD.		Date/Duration	1-4-2022 / 15-4-2022	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge up gradation					✓
Performance/Quality of work					✓
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	ML & NN, NLP Stock Market				
OVERALL GRADE (Any one)	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor					

Receiving Date	15/04/2022	Name of Faculty Mentor	Mir Shahnawaz Ahmad	Sign	

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	YASH JAIN		Department	C SE	
Industry/Organization	Praedico Global Research Pvt. Ltd.		Date/Duration	15-April-2022 to 10-May-2022	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge up gradation					✓
Performance/Quality of work					✓
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	Project Completion & wrap up				
OVERALL GRADE (Any one)	POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT				
Name of Industry Mentor	Priyank Gupta				
Signature of Industry Mentor					

Receiving Date	10/05/2022	Name of Faculty Mentor	Mir Shahnawaz Ahmad	Sign	