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:

Ronak Uchariya

0901CS181085

Faculty Mentor:

Mr ARUN KUMAR

Assistant Professor

Submitted to:

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)

Final Year Internship Report

on

Neural Network Developer at Praedico Global Research Pvt. Ltd.

Submitted By:

Ronak Uchariya

0901CS181085

Mentor:

Mr Priyank Gupta

Project Head at PGR Pvt. Ltd.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

Internship Certificate

PRAEDICO GLOBAL RESEARCH PVT. LTD.
PGR
Worlds Finest Robotic Stock Researcher

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

Ref.: PGR-2022/I-515 **Date:** 11th – May - 2022

CERTIFICATE OF INTERNSHIP

This certificate is awarded to

Mr./Miss. RONAK UCHARIYA

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.

PRAEDICO GLOBAL RESEARCH PVT. LTD. GWALIOR (M.P.) INDIA

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 **Ronak Uchariya** (0901CS181085) 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 ARUN KUMAR**, 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.

27/05/22

Mr Arun Kumar
Assistant Professor
Computer Science and Engineering

27/05/22

Dr. Manish Dixit
Professor and Head
Computer Science and Engineering

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 **Mr Arun Kumar, 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.

Ronak Uchariya

0901CS181085

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 Arun Kumar**, 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.

Ronak Uchariya

0901CS181085

IV Year

Computer Science and Engineering

ABSTRACT

I carried out my internship at Praedico Global Research Pvt. Ltd. is a private company which is founded on 09 April 2018 it acts as a one-stop center for researchers and developers and it also offers internship opportunities to the students. The internship is an opportunity to relate what has been covered in class. The purpose of the program is to fulfill the core equipment for the award of a Bachelor's Degree in Computer Science and Engineering to get a practical aspect of the theoretical work studied at the university and to understand the operations in the Technology and Science sector and to enable students to gain experience in different Technology.

During my internship period, several approaches and exposure methods were used which included hands-on, reading relevant materials, and also questions and answer approaches.

I was assigned to the stock predictions tasks which included the Machine learning, neural network, and Django backend. The stock market 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. This model proposes a method or algorithm that can successfully fetch news from various online financial websites and along with this it labels the news i.e. how this news affect the share price of particular company as positive, negative or neutral and also predict the stock price on the basis of that sentiments and stock trends

In conclusion, this was an opportunity to develop and enhance skills and competencies in my career field which I achieved.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter I: Introduction

1.1 Overview

Stock prediction is a challenging problem in the field of finance as well as engineering and mathematics. Due to its financial gain, it has attracted much attention both from academic side and business side. Stock price prediction has always been a subject of interest for most investors and financial analysts. Nevertheless, finding the best time to buy or sell has remained a very difficult task for investors because there are other numerous factors that may influence stock prices.

Sentiment classification has become a very popular task in natural language processing area, which tries to predict sentiment (opinion, emotion, etc.) from texts. Today people are placing their comments and opinions on social media which can be shared by others also. Sentiment classification could be done in word/phrase level, sentence level and document level. Sentiment analysis has now become the dominant approach used for extracting sentiment and appraisals from online sources. Subjectivity analysis focuses on dividing language units into two categories: objective and subjective, whereas sentiment analysis attempts to divide the language units into three categories; negative, positive, and neutral.

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 performance 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 predicting news sentiment with the help of numerical data. For the model building we have used the deep learning.

Deep learning is based on the branch of machine learning, which is a subset of artificial intelligence. Since neural networks imitate the human brain and so deep learning will do. In deep learning, nothing is programmed explicitly. Basically, it is a machine learning class that makes use of numerous nonlinear processing units so as to perform feature extraction as well as transformation. The output from each preceding layer is taken as input by each one of the successive layers.

1.2 About the Company

Praedico Global Research Private Limited is a private company which is founded on 09 April, 2018. We 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.

Sunny Ralli and Priyank Gupta are the Chief Research Officer and Chief Technical Officer of the company respectively. The Registered Number of the company is 045496 and the Registered Address is 2nd Floor 204, Garima Arcade Shinde Ki Chawani, Gwalior, MP. The working office of the company is situated in Malviya Nagar, South Delhi. Its authorized capital is Rs. 1,000,000 and paid up capital is 1,00,000. Their industry code is 74999.

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.

Forecasting of stock trend has been limited to how well a trader or broker can consider the market forces or the economic and company information available. The risk associated with the stock market, implies reasoning beyond human ability has to be employed. It is due to this reason that, Praedico Global Research has been able to use Neural Networks which is a form of artificial intelligence to predict the stock market with the success rate of over 80%

With Praedico Global Research, robust softwares are built to be sensitive and identify trend on the stock market and make a high probable prediction with at least 80% confidence level.

The Company also conducts financial related workshops all over the globe. They offer their services to clients trading on stock markets all around the world. They use Financial Research and Neural Network Programming to develop hybrid products which can be used by traders and investors for better prediction of their investments.

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 proposed method, below are the software and hardware Requirements

Software Requirements

- Operating System: Windows
- Coding Language: Python
- Platform: Jupyter Notebook and Google Colab
- IDE: Visual Studio Code
- Modules: Pandas, NumPy, sklearn, keras, time, requests, json, csv, joblib, TensorFlow, BeautifulSoup, system, name

Minimum Hardware Requirements

- Processor: Intel(R) Core (TM) i5
- Memory: **4 GB**
- Hard Disk: **250 GB**

1.5 Outcome

In this internship we proposed a method which is capable of fetching Realtime news and label its sentiments with prediction of stock price. The dataset generated by this algorithm 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. Which will be useful for Forecasting of stock trend consider the market forces or the economic and company information available

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 performance 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 systems primarily focused on improving 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 its accuracy will itself improve.

Chapter III: Work Done During Internship

3.1 Proposed System

1. Scrape the HTML of the webpage and extract useful information/data from it. This technique is called web scraping or web harvesting or web data extraction.
2. In this system for the purpose of data collection, we use web scraping Python package called Beautiful Soup. There are some steps involved in web scraping using the implementation of a Web Scraping framework of Python called Beautiful Soup.
3. Send an HTTP request to the URL of the webpage you want to access. The server responds to the request by returning the HTML content of the webpage.
4. Once we have accessed the HTML content, we are left with the task of parsing the data. Since most of the HTML data is nested, we cannot extract data simply through string processing. One needs a parser which can create a nested/tree structure of the HTML data. There are many HTML parser libraries available but the most advanced one is html5lib
5. Now, we need to do is navigating and searching the parse tree that we created, i.e. tree traversal. For this task, we will be using another third-party python library, Beautiful Soup. It is a Python library for pulling data out of HTML and XML files.
6. After scraping the news, the real task come in to label it and fetching the real time stock price, and compare it with predict result

3.2 Modules Used

- Beautiful Soup
- Pandas
- NumPy
- Seaborn
- Matplotlib

3.3 TECHNOLOGY

- Python 3.8
- Machine Learning
- Natural Language Processing
- Neural Network
- Scikit-Learn
- Keras
- TensorFlow

3.4 Flowchart

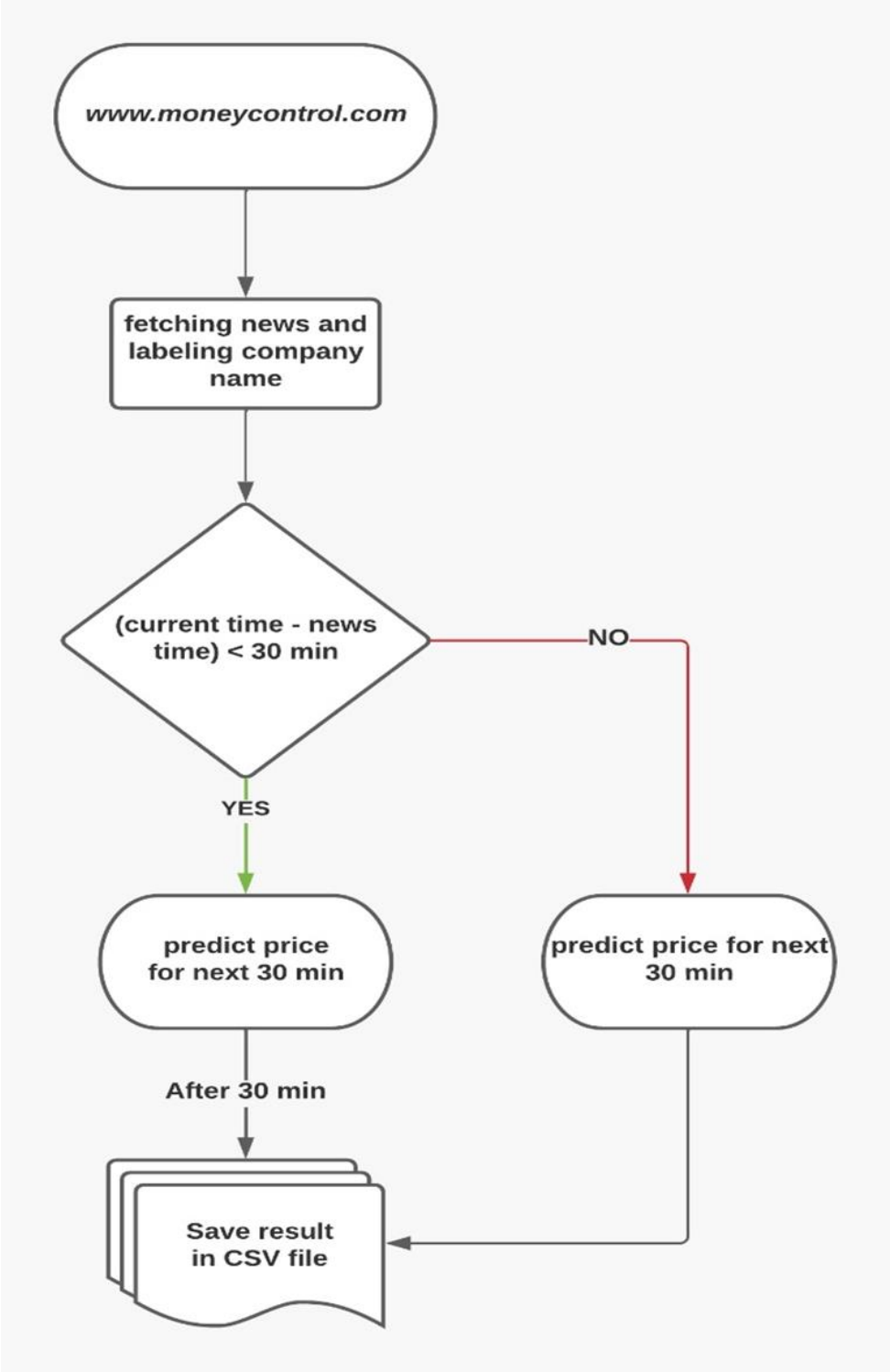

Fig 3.3.1 flow chart

Chapter IV: Coding and Results

4.1 CODING

```
import time,requests,json,csv,joblib,pandas as pd,numpy as np,tensorflow
as tf
from bs4 import BeautifulSoup
from os import system, name
from sklearn.preprocessing import MinMaxScaler
from os import system ,name
from keras.models import load_model

def   json_load_file():
        return {"lable_key_word":[ ["adani"],["asian paints",
"paint"],["axis bank", "bank"],["bajaj auto", "bajaj"],["bajaj finance",
"bajaj"],["bajaj finserv", "bajaj"],["bpcl", "bharat petroleum",
"petrol"],["bharti airtel", "airtel"],["britannia"],["cipla"],["coal
india", "coal"],["divis laboratories", "divi's lab", "divis
labs"],["reddy's laboratories", "reddys lab",
"reddy's"],["eicher"],["grasim"],["hcl tech"],["hdfc","bank","hdfc
bank"],["hdfc","bank","hdfc life insurance"],["hero motocorp",
"hero"],["hindalco"],["hindustan unilever", "hul"],["hdfc"],["icici",
"bank"],["indian oil", "ioc"],["indusind",
"bank"],["infosys"],["itc"],["jsw"],["kotak", "bank",
"mahindra"],["larsen", "lti", "l&t"],["mahindra", "m&m"],["maruti suzuki",
"maruti"],["nestle", "nestlé"],["ntpc"],["oil and natural gas
corporation", "oil & natural gas corporation", "ongc"],["power grid
corporation", "power grid"],["reliance"],["sbi", "bank"],["shree cements",
"cement"],["state bank of india", "sbi", "bank"],["sun pharma"],["tata
consultancy services", "tcs", "tata"],["tata consumer products",
"tata"],["tata motors", "tata"],["tata steel", "tata"],["tech mahindra",
"tech m", "mahindra"],["titan"],["ultratech cement", "ultratech",
"cement"],["upl"],["wipro"]]}

#utility function for update_csv()
def check_nan(close_times,date):
    t = time.mktime(time.strptime(date+' 15:29',"%Y-%m-%d %H:%M"))
    return False if t in close_times else True and True if
len(close_times) < 375 else False


# find_nan_and_fill_nan(df_2) take panda Dataframe if there is any NaN
values it fill by itself
def find_nan_and_fill_nan(df_2):
    nan_indx = [index for index, row in df_2.iterrows() if
row.isnull().any()]
    columns_with_nan = [
df_2.loc[[i]].columns[df_2.loc[[i]].isna().any()].tolist() for i in
nan_indx ]
    if nan_indx:
        df = df_2.T.sort_index().T.copy(deep=True)
        for idx,col_start in zip(nan_indx,columns_with_nan):
```

```
            for i in range(0,len(col_start),4):
                col_prev = df.loc[[idx]].loc[:,:col_start[i]].iloc[:,-5:-
4].columns[0][0]
                df.loc[idx][col_start[i+0]], df.loc[idx][col_start[i+1]],
df.loc[idx][col_start[i+2]], df.loc[idx][col_start[i+3]] =
df.loc[idx][col_prev].iloc[0], df.loc[idx][col_prev].iloc[1],
df.loc[idx][col_prev].iloc[2], df.loc[idx][col_prev].iloc[3]
        return df
    return df_2


# return a list of numbers in range(50) respectively dictionary define in
temp.json file
# label_news(news_str,news_str_2) take news heading and paragraph and
search for nifty company
def label_news(news_str,news_str_2):
    news_str = news_str +" " + news_str_2
    news_str = news_str.replace('ICICI Direct','')
    news_str,label_dict,labels = news_str.lower(), json_load_file(),list()
    for i in label_dict['lable_key_word']:
        for j in i:
            if j in news_str:
                labels.append(label_dict['lable_key_word'].index(i))
                break
    if labels:
        return labels
    return [-1]


# scrape_news() return the moneycontrol news from page define in range(),
return type dict
def scrape_news():
    table_dict = {'time':list(),'heading':list(),'paragraph':list(),'news
href':list()}
    for i in range(1,10):
        url = f"https://www.moneycontrol.com/news/business/stocks/page-
{i}/"
        response_ = requests.get(url)
        if response_.ok:
            response_ = response_.text
            soup = BeautifulSoup(response_, "lxml")
            soup = soup.find('ul', id='cagetory')
            table_rows = soup.find_all('li', class_='clearfix')
            for row in table_rows:
                table_dict['time'].append(row.find('span').text)
                table_dict['heading'].append(row.find('h2').text)
                table_dict['paragraph'].append(row.find('p').text)
                table_dict['news href'].append(row.find('a')['href'])
    return table_dict
```

```
# news_update() call the scrape_news() and compare results to (nifty 50
all news.csv) and return the new news in Pandas format
def news_update():
    new_news = pd.DataFrame(scrape_news())
    old_news = pd.read_csv(r"nifty news prediction\nifty news\nifty 50 all
news.csv")
    indx=-1
    for i in range(0,new_news.shape[0]):
        if(new_news.loc[i]['heading']==old_news.loc[0]['heading']):
            indx = i
            break
        else:
            print(f"\nSno.{i+1} {new_news.loc[i]['time'][:-4]} -->
{new_news.loc[i]['heading']}",end="")
            time.sleep(0.5)
            pass
    if(indx==0):
        print(f"\n________NO New News Updated________",end="")
        return pd.DataFrame()
    elif(indx !=-1):
        new_news.drop(labels=range(indx,new_news.shape[0]),axis=0,inplace=
True)
    print(f"\n________ {new_news.shape[0]} New News
Updated________",end="\n")
    return new_news


# get_latest_date(file_name,nifty_stock_name) it is a utility function it
check the price csv files and return date (Y-m-d) which is not present in
csv file
def get_latest_date(file_name,nifty_stock_name):
    df1 = pd.read_csv(f"nifty news
prediction\\price_csv\\{file_name}.csv",header=[0,1],index_col=0)
    t1,t2 = time.mktime(time.strptime(df1.index[2],"%Y-%m-
%d")),time.time()
    header,url = {"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/98.0.4758.102 Safari/537.36
Edg/98.0.1108.62"},
f"https://priceapi.moneycontrol.com/techCharts/techChartController/history
?symbol={nifty_stock_name}&resolution=1D&from={int(t1)}&to={int(t2)}"
    chart_data = requests.get(url,headers=header).text
    day_data =
json.loads(BeautifulSoup(chart_data,'lxml').find('body').text)
    h =  [time.strftime("%Y-%m-%d",time.gmtime(t)) for t in day_data['t']]
    for date in h:
        if date not in df1.index:
            return date
    return -1


# get_price_details(nifty_stock_name,date) return the latest price of
stock if available
def get_price_details(nifty_stock_name,t1):
    t1,t2 = time.mktime(time.strptime(t1+" 09:15 AM","%Y-%m-%d %I:%M %p")
```

```
) , time.mktime(time.strptime(t1+" 04:30 PM","%Y-%m-%d %I:%M %p") )
    header = {"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/98.0.4758.102 Safari/537.36
Edg/98.0.1108.62"}
    url =
f"https://priceapi.moneycontrol.com/techCharts/techChartController/history
?symbol={nifty_stock_name}&resolution=1&from={int(t1)}&to={int(t2)}"
    chart_data = requests.get(url,headers=header).text
    soup = BeautifulSoup(chart_data,'lxml')
    data = soup.find('body').text
    data = json.loads(data)
    return data


# update_csv() it is utility function to keep the stock price csv files
upto date
def update_csv():
    file_names = ['nifty','Adani Ports','Asian Paints','Axis Bank','Bajaj
Auto','Bajaj Finance','Bajaj Finserv','BPCL','Bharti
Airtel','Britannia','Cipla','Coal India','Divis Labs','Dr Reddys
Labs','Eicher Motors','Grasim','HCL Tech','HDFC Bank','HDFC Life','Hero
Motocorp','Hindalco','HUL','HDFC','ICICI Bank','IOC','IndusInd
Bank','Infosys','ITC','JSW Steel','Kotak Mahindra','Larsen','M&M','Maruti
Suzuki','Nestle','NTPC','ONGC','Power Grid Corp','Reliance','SBI Life
Insura','Shree Cements','SBI','Sun Pharma','TCS','TATA Cons. Prod','Tata
Motors','Tata Steel','Tech Mahindra','Titan
Company','UltraTechCement','UPL','Wipro']
    nifty_stock_name =  ["9","ADANIPORTS", "ASIANPAINT", "AXISBANK",
"BAJAJ-AUTO", "BAJFINANCE", "BAJAJFINSV", "BPCL", "BHARTIARTL",
"BRITANNIA", "CIPLA", "COALINDIA", "DIVISLAB", "DRREDDY", "EICHERMOT",
"GRASIM", "HCLTECH", "HDFCBANK", "HDFCLIFE", "HEROMOTOCO", "HINDALCO",
"HINDUNILVR", "HDFC", "ICICIBANK", "IOC", "INDUSINDBK", "INFY", "ITC",
"JSWSTEEL", "KOTAKBANK", "LT", "M%26M", "MARUTI", "NESTLEIND", "NTPC",
"ONGC", "POWERGRID", "RELIANCE", "SBILIFE", "SHREECEM", "SBIN",
"SUNPHARMA", "TCS", "TATACONSUM", "TATAMOTORS", "TATASTEEL", "TECHM",
"TITAN", "ULTRACEMCO", "UPL", "WIPRO"]
    print('\nUpdating Price csv files',end="")
    for i in range(51):
        while(True):
            date = get_latest_date(file_names[i],nifty_stock_name[i])
            if date==-1:
                print(f'\n{file_names[i]}.csv files is upto date',end="")
                break
            data = get_price_details(nifty_stock_name[i],date)
            if(data['s']=="no_data" or data['s']=="error" or
check_nan(data['t'],date)):
                print(f'\nthe market is open now, wait till {date} 03:30
PM to update {file_names[i]}.csv',end="")
                break
            time_lis,row = [ time.strftime("%H:%M",j)  for j in
[time.gmtime(i+19800) for i in data['t']]] ,dict()
            time_lis, open_lis, high_lis, low_lis, close_lis =
time_lis[:375], data['o'], data['h'], data['l'],data['c']
            for o,h,l,c,t_stamp in
```

```
zip(open_lis,high_lis,low_lis,close_lis,time_lis):
                row.update({(t_stamp,'close'):[c],(t_stamp,
'high'):[h],(t_stamp, 'low'):[l],(t_stamp, 'open'):[o]})
            row = pd.DataFrame(row,index=[ time.strftime("%Y-%m-
%d",time.gmtime(19800+data['t'][0]))])
            print(f'\n{row.iloc[[0]]}\n',end='')
            row = row.append(pd.read_csv(f"nifty news
prediction\\price_csv\\{file_names[int(i)]}.csv",header=[0,1],index_col=0)
)
            row = find_nan_and_fill_nan(row)
            row.to_csv(f"nifty news
prediction\\price_csv\\{file_names[int(i)]}.csv",index=True,index_label='d
ate')
     time.sleep(1.5)
     _ = system('cls') if (name=='nt')  else system('clear')


# get_old_price(file_name,nifty_stock_name) return the stock price csv in
Pandas format
def get_old_price(file_name,nifty_stock_name):
      df1 = pd.read_csv(f"nifty news
prediction\\price_csv\\{file_name}.csv",header=[0,1],index_col=0).iloc[:,0
::4]
    df1.columns = df1.columns.droplevel(1)
    df = pd.DataFrame()
    df['t'],df['c'],df['time'] = [int(time.mktime(time.strptime(i+'
'+j,"%Y-%m-%d %H:%M"))) for i in df1[::-1].index for j in
df1.loc[i].index],[j for i in df1[::-1].index for j in df1.loc[i]] , [i+'
'+j for i in df1[::-1].index for j in df1.loc[i].index]
    latest_date = get_latest_date(file_name,nifty_stock_name)
    if (latest_date != -1):
        latest_data = get_price_details(nifty_stock_name,latest_date)
        new_df =
pd.DataFrame({'t':latest_data['t'],'c':latest_data['c'],'time':[time.strft
ime("%Y-%m-%d %H:%M",time.gmtime(i+19800))for i in
latest_data['t']]})[:375]
        df = df.append(new_df,ignore_index=True)
    return df


# actual_price_in_next_30_min(file_name,nifty_stock_name,news_time) return
actual price for next 30 min
def actual_price_in_next_30_min(file_name,nifty_stock_name,news_time):
    old_df =
get_old_price(file_name,nifty_stock_name).query(f"t>={news_time}")
    (op,cl) = (old_df.iloc[0]['t'],old_df.iloc[-1]['t']) if
old_df.shape[0]!=0 else (0,0)
    actual_price = [old_df.query(f"t>={news_time+300*i}").iloc[0]['c'] if
(news_time+300*i<=cl and old_df.shape[0]!=0) else 'waiting' for i in
range(7)]
    actual_price_time =
[old_df.query(f"t>={news_time+300*i}").iloc[0]['time'] if
(news_time+300*i<=cl and old_df.shape[0]!=0) else 'waiting' for i in
range(7)]
```

```
    return {"price":actual_price,"time":actual_price_time}


# predict_next_30_min(file_name,nifty_stock_name,news_time) return
prediction price for next 30 min
def predict_next_30_min(file_name,nifty_stock_name,news_time):
    sclr,modl,time_step = joblib.load(f'nifty news prediction\\nifty
scalers\\{file_name} scaler'),load_model(f'nifty news prediction\\nifty
models\\{file_name} models.h5'),100
    old_df =
get_old_price(file_name,nifty_stock_name).query(f"t<={news_time}")
    inp = old_df[::-1].iloc[::10][:time_step][::-1]['c'].to_numpy()
    inp = sclr.transform(inp.reshape(-1,1))
    pridct_result_for_nxt_30min=list()
    for i in range(6):
        temp =
np.append(np.delete(inp,range(i)),pridct_result_for_nxt_30min)
        pridct_result_for_nxt_30min.append(modl.predict(temp.reshape(-
1,time_step,1)))
    pridct_result_for_nxt_30min =
np.array(pridct_result_for_nxt_30min).reshape(-1,1)
    return {"predict price":[round(i,2) for i in
sclr.inverse_transform(pridct_result_for_nxt_30min).reshape(-1).tolist()]}
def print_result(ac,pr,file_name):
    d = pd.DataFrame(ac)
    d.columns= ['actual price',file_name]
    d['predict price'] = [ac['price'][0]]+pr['predict price']
    print("\n"f"{d.set_index(file_name).T.to_string()}\n",end="")


# Save_predict_price_vs_Actual_price() save the prediction price and
actual price in nifty 50 news file
def Save_predict_price_vs_Actual_price():
    file_names = ['Adani Ports','Asian Paints','Axis Bank','Bajaj
Auto','Bajaj Finance','Bajaj Finserv','BPCL','Bharti
Airtel','Britannia','Cipla','Coal India','Divis Labs','Dr Reddys
Labs','Eicher Motors','Grasim','HCL Tech','HDFC Bank','HDFC Life','Hero
Motocorp','Hindalco','HUL','HDFC','ICICI Bank','IOC','IndusInd
Bank','Infosys','ITC','JSW Steel','Kotak Mahindra','Larsen','M&M','Maruti
Suzuki','Nestle','NTPC','ONGC','Power Grid Corp','Reliance','SBI Life
Insura','Shree Cements','SBI','Sun Pharma','TCS','TATA Cons. Prod','Tata
Motors','Tata Steel','Tech Mahindra','Titan
Company','UltraTechCement','UPL','Wipro']
    nifty_stock_name =  ["ADANIPORTS", "ASIANPAINT", "AXISBANK", "BAJAJ-
AUTO", "BAJFINANCE", "BAJAJFINSV", "BPCL", "BHARTIARTL", "BRITANNIA",
"CIPLA", "COALINDIA", "DIVISLAB", "DRREDDY", "EICHERMOT", "GRASIM",
"HCLTECH", "HDFCBANK", "HDFCLIFE", "HEROMOTOCO", "HINDALCO", "HINDUNILVR",
"HDFC", "ICICIBANK", "IOC", "INDUSINDBK", "INFY", "ITC", "JSWSTEEL",
"KOTAKBANK", "LT", "M%26M", "MARUTI", "NESTLEIND", "NTPC", "ONGC",
"POWERGRID", "RELIANCE", "SBILIFE", "SHREECEM", "SBIN", "SUNPHARMA",
"TCS", "TATACONSUM", "TATAMOTORS", "TATASTEEL", "TECHM", "TITAN",
"ULTRACEMCO", "UPL", "WIPRO"]
    full_name = ['Adani Ports and Special Economic Zone Ltd.','Asian
Paints Ltd.','Axis Bank Ltd.','Bajaj Auto Ltd.','Bajaj Finance
```

```
Ltd.','Bajaj Finserv Ltd.','Bharat Petroleum Corporation Ltd.','Bharti
Airtel Ltd.','Britannia Industries Ltd.','Cipla Ltd.','Coal India
Ltd.','Divis Laboratories Ltd.','Dr Reddys Laboratories Ltd.','Eicher
Motors Ltd.','Grasim Industries Ltd.','HCL Technologies Ltd.','HDFC Bank
Ltd.','HDFC Life Insurance Company Ltd.','Hero Motocorp Ltd.','Hindalco
Industries Ltd.','Hindustan Unilever Ltd.','Housing Development Finance
Corporation Ltd.','ICICI Bank Ltd.','Indian Oil Corporation
Ltd.','IndusInd Bank Ltd.','Infosys Ltd.','ITC Ltd.','JSW Steel
Ltd.','Kotak Mahindra Bank Ltd.','Larsen & Toubro Ltd.','Mahindra and
Mahindra Ltd.','Maruti Suzuki India Ltd.','Nestle India Ltd.','NTPC
Ltd.','Oil and Natural Gas Corporation Ltd.','Power Grid Corporation of
India Ltd.','Reliance Industries Ltd.','SBI Life Insurance Company
Ltd.','Shree Cements Ltd.','State Bank of India','Sun Pharmaceutical
Industries Ltd.','Tata Consultancy Services Ltd.','TATA Consumer Products
Ltd.','Tata Motors Ltd.','Tata Steel Ltd.','Tech Mahindra Ltd.','Titan
Company Ltd.','UltraTech Cement Ltd.','UPL Ltd.','Wipro Ltd.']
    sector = ['Infrastructure - General','Paints & Varnishes','Banks -
Private Sector','Auto - 2 & 3 Wheelers','Finance - Leasing & Hire
Purchase','Finance - Investments','Refineries','Telecommunications -
Service','Food Processing','Pharmaceuticals','Mining &
Minerals','Pharmaceuticals','Pharmaceuticals','Auto - LCVs &
HCVs','Diversified','Computers - Software','Banks - Private Sector','Life
& Health Insurance','Auto - 2 & 3 Wheelers','Iron & Steel','Personal
Care','Finance - Housing','Banks - Private Sector','Refineries','Banks -
Private Sector','Computers - Software','Cigarettes','Steel - Large','Banks
- Private Sector','Infrastructure - General','Auto - Cars & Jeeps','Auto -
Cars & Jeeps','Food Processing','Power - Generation & Distribution','Oil
Drilling And Exploration','Power - Generation &
Distribution','Refineries','Diversified','Cement - Major','Banks - Public
Sector','Pharmaceuticals','Computers - Software','Plantations - Tea &
Coffee','Auto - LCVs & HCVs','Iron & Steel','Computers -
Software','Miscellaneous','Cement - Major','Chemicals','Computers -
Software']
    new_news,wait = news_update(),0
    for i in new_news[::-1].index:
        print("\n\n"f"Sno.{i+1}       {new_news.loc[i][0]}\n{new_news.loc[
i][1:].to_string()}\n",end="")
        labels =
label_news(new_news.iloc[i]['heading'],new_news.iloc[i]['paragraph'])
        if labels[0] == -1:
            print('\n_____________________________other
news_____________________________',end="")
            news_row = new_news.iloc[[i]].copy(deep=True)
            news_row[["nifty name","company","sector","price 0
min","predict 5 min","predict 10 min","predict 15 min","predict 20
min","predict 25 min","predict 30 min","price 5 min","price 10 min","price
15 min","price 20 min","price 25 min","price 30 min"]] = ["out of nifty
50","other","other",0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0]
            news_row = news_row.append(pd.read_csv(r'nifty news
prediction\nifty news\nifty 50 all news.csv'),ignore_index=True)
            news_row.to_csv(r'nifty news prediction\nifty news\nifty 50
all news.csv',index=False) if wait==0 else wait==1
        else:
            for label in labels:
```

```
                news_time =
time.mktime(time.strptime(new_news.iloc[i]['time'][:-4],"%B %d, %Y %I:%M
%p"))
                ac =
actual_price_in_next_30_min(file_names[label],nifty_stock_name[label],news
_time)
                pr =
predict_next_30_min(file_names[label],nifty_stock_name[label],news_time)
                news_row = new_news.iloc[[i]].copy(deep=True)
                news_row[["nifty name","company","sector","price 0
min","predict 5 min","predict 10 min","predict 15 min","predict 20
min","predict 25 min","predict 30 min","price 5 min","price 10 min","price
15 min","price 20 min","price 25 min","price 30 min"]] =
[file_names[label],full_name[label],sector[label]] + [ac['price'][0]] +
pr['predict price'] +ac['price'][1:]
                print_result(ac,pr,file_names[label])
                if (ac['price'][-1]!='waiting'and wait==0):
                    news_row = news_row.append(pd.read_csv(r'nifty news
prediction\nifty news\nifty 50 all news.csv'),ignore_index=True)
                    news_row.to_csv(r'nifty news prediction\nifty
news\nifty 50 all news.csv',index=False)
                else:
                    wait = 1

#utility function
def main ():
    tf.autograph.experimental.do_not_convert(func=None)
    update_csv()
    Save_predict_price_vs_Actual_price()
```

4.2 Results:

For a website moneycontrol the following data scrape and analyse

index	time	heading	paragraph	news href
0	May 12, 2022 05:46 PM IST	Buy Bajaj Auto; target of Rs 4350: Emkay Globa...	Emkay Global Financial is bullish on Bajaj Aut...	https://www.moneycontrol.com/news/business/sto...
1	May 12, 2022 05:46 PM IST	Buy UPL; target of Rs 930: Emkay Global Financial	Emkay Global Financial is bullish on UPL has r...	https://www.moneycontrol.com/news/business/sto...
2	May 12, 2022 05:46 PM IST	Buy Gujarat Gas; target of Rs 630: Emkay Globa...	Emkay Global Financial is bullish on Gujarat G...	https://www.moneycontrol.com/news/business/sto...
3	May 12, 2022 05:45 PM IST	Buy Max Financial; target of Rs 1030: Emkay Gl...	Emkay Global Financial is bullish on Max Finan...	https://www.moneycontrol.com/news/business/sto...
4	May 12, 2022 05:45 PM IST	Buy KEI Industries; target of Rs 1350: Emkay G...	Emkay Global Financial is bullish on KEI Indus...	https://www.moneycontrol.com/news/business/sto...
5	May 12, 2022 05:45 PM IST	Hold Dalmia Bharat; target of Rs 1615: Emkay G...	Emkay Global Financial recommended hold rating...	https://www.moneycontrol.com/news/business/sto...
6	May 12, 2022 05:45 PM IST	Hold Voltas; target of Rs 1000: Emkay Global F...	Emkay Global Financial recommended hold rating...	https://www.moneycontrol.com/news/business/sto...
7	May 12, 2022 05:45 PM IST	Hold PVR; target of Rs 2165: Emkay Global Fina...	Emkay Global Financial is bullish on PVR has r...	https://www.moneycontrol.com/news/business/sto...
8	May 12, 2022 05:45 PM IST	Buy Canara Bank; target of Rs 282: Emkay Globa...	Emkay Global Financial is bullish on Canara Ba...	https://www.moneycontrol.com/news/business/sto...
9	May 12, 2022 05:19 PM IST	Bajar Gupshup | Markets tank for fifth consecu...	Markets tanked for the fifth consecutive trade...	https://www.moneycontrol.com/news/business/mar...

Table 4.2.1: Table shows news scraping from website.

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

news

time	May 12, 2022 02:46 PM IST
heading	Buy Bajaj Auto; target of Rs 4350: Emkay Globa...
paragraph	Emkay Global Financial is bullish on Bajaj Aut...
news href	https://www.moneycontrol.com/news/business/sto...

Labels

[Bajaj Auto, Bajaj Financ, Bajaj Finserv]

Predictions

	price 0 min	price 5 min	price 10 min	price 15 min	price 20 min	price 25 min	price 30 min
Bajaj Auto	3611.95	3607.35	3594.7	3594.65	3605	3604.49	3604.49
	price 0 min	price 5 min	price 10 min	price 15 min	price 20 min	price 25 min	price 30 min
Bajaj Financ	5981.55	5976.55	5978.05	5974.65	5970.7	5987.65	5978.6
	price 0 min	price 5 min	price 10 min	price 15 min	price 20 min	price 25 min	price 30 min
Bajaj Finserv	13729.25	13721.75	13720	13737.75	13716.95	13767.35	13737

Table 4.2.2: Table shows stock price prediction after fetching actual price from website.

4.3 Prediction Graphs

Here are some graphs on the basis of prediction and actual result

Bajaj Auto

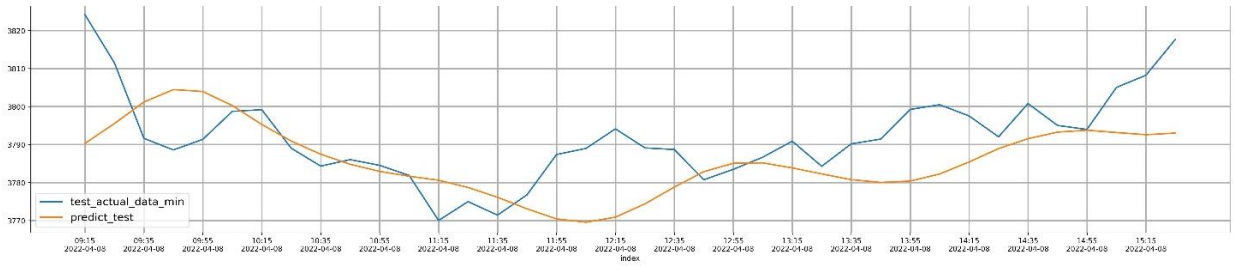

Fig 4.3.1 prediction vs actual price

Bajaj Finance

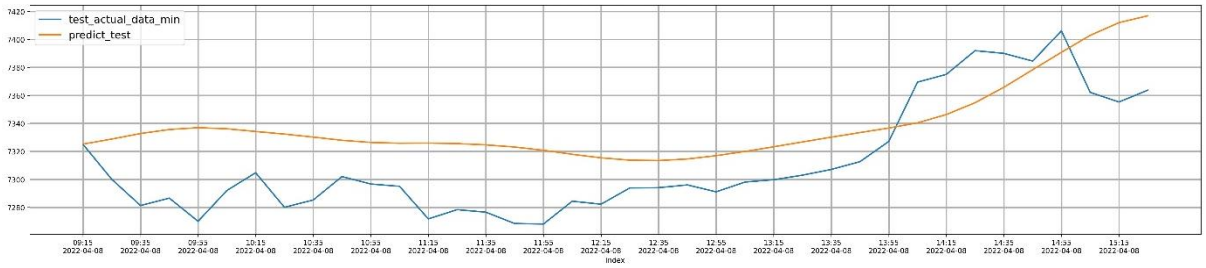

Fig 4.3.2 prediction vs actual price

Bajaj Finserv

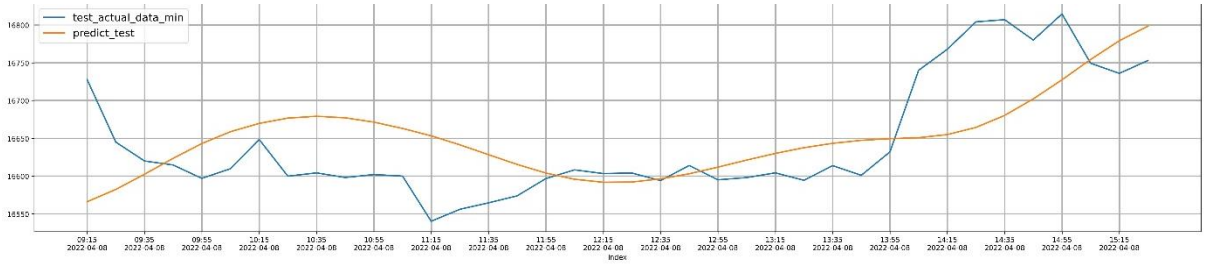

Fig 4.3.3 prediction vs actual price

prediction line Actual real time price

Chapter V: Conclusion and Future Work

5.1 Conclusion

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 algorithm 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 methodology by incorporating more collection sources so that the dataset generated become more versatile. Also, to use 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.

References

1. Stock Market Today | Bombay Stock Exchange Updates | Global Stock Market News (moneycontrol.com)
2. Yahoo Finance - Stock Market Live, Quotes, Business & Finance News
3. Google Finance - Stock Market Prices, Real-time Quotes & Business News

FPR Report 1

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchaniya		Department	CSE	
Industry/Organization	PRAEDICO GLOBAL RESEARCH PVT. LTD.		Date/Duration	10-1-2022	10-2-2022
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		**Name of Faculty Mentor**		**Sign**	

FPR Report 2

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchariya		Department	CSE	
Industry/Organization	PRAEDICO GLOBAL RESEARCH PVT. LTD.		Date/Duration	10-2-2022 / 02-3-2022	
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		**Name of Faculty Mentor**		**Sign**	

FPR Report 3

MADHAV INSTITUTE OF TECHNOLOGY AND SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Date:

FORMAT

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchariya		Department	CSE	
Industry/Organization	PRAEDICO GLOBAL RESEARCH PVT. LTD.		Date/Duration	2-3-2022 / 16-3-2022	
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	Priyanka Gupta				
Signature of Industry Mentor					

Receiving Date		**Name of Faculty Mentor**		**Sign**	

FPR Report 4

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchamiya		Department	CSE	
Industry/Organization	PRAEDICO GLOBAL RESEARCH PVT. LTD.		Date/Duration	16-3-2022 / 1-4-2022	
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	Priyank				

Receiving Date		**Name of Faculty Mentor**		**Sign**	

FPR Report 5

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchaniya		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		**Name of Faculty Mentor**		**Sign**	

FPR Report 6

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Ronak Uchaniya		Department	CSE	
Industry/Organization	PRADICO GLOBAL RESEARCH PVT. LTD		Date/Duration	15-4-2022 / 10-5-2022	
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		**Name of Faculty Mentor**		**Sign**	