Project Report – 2

On

Stock Market Prediction System

Subject Code: 3IT31

Subject: Mini Project

Academic Year: 2023 – 2024

Team Members: Pratham Satani (21IT417)

Vatsal Patel (21IT414)

Deep Hirapara (21IT403)

Batch: D – 10

Project Guide: Prof. K. J. Padhiyar

Birla Vishvakarma Mahavidyalaya

Engineering College, Vallabh Vidyanagar

[An Autonomous Institution]

14. Project Implementation (up to 60%)

14.1 Data Acquisition and Management

- Aim
  - Ensure data is captured without loss or corruption from diverse sources.
  - Implement error-checking, and validation mechanisms to maintain data integrity.
  - Handle real-time and/or historical data without delays.
  - Clean, normalize, and transform raw data into usable, structured formats.

- Dataset
  - Data to be used in this project will be downloaded from finance.yahoo.com.
  - This data is free and fully open-source.
  - Format of this data will be in CSV (comma-separated values) format.
  - Timeframe of the data will be 10 years back from the current date.

Date	Open	High	Low	Close	Adj Close	Volume
01-01-2013	45.15833	45.15833	44.11667	44.44167	23.72343	1375158
02-01-2013	44.5	45.41666	44.49167	45.33333	24.19941	1936392
03-01-2013	45.65833	45.65833	44.88333	45.25833	24.15938	1146306
04-01-2013	45.35	48.13333	45.25833	46.85833	25.01348	10010520
07-01-2013	45.375	48.30833	45.375	46.875	25.02238	10881438

- Data Acquisition
  - Dataset will be downloaded using the ‘yfinance’ module of the Python Programming Language.
  - Size of the dataset will be no larger than 1 MB.
  - Following snippet shows the Data Acquisition methodology:

```
import yfinance as yf
stock = yf.Ticker(<name>).history(<period>)
```

  - Downloaded data will be temporarily stored in the working memory allocated to the program and then once the task is completed, it will be deallocated.
  -

- Data Preprocessing
  - The values in the dataset are comparatively larger in magnitude than that accepted by the Machine Learning model.
  - Thus, the values are normalized between [0, 1] using the ‘scikit-learn’ module’s data preprocessing functionality. Following code snippet shows how it is done.

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler(feature_range=(0,1))
scaled_data = scaler.fit_transform(stock)
```

Date	Open	High	Low	Close	Adj Close	Volume
01-01-2013	0.110675	0.103422	0.110829	0.102931	0.06153	0.004367
02-01-2013	0.105238	0.105544	0.113998	0.110398	0.065993	0.006149
03-01-2013	0.114805	0.107529	0.117307	0.10977	0.065618	0.00364
04-01-2013	0.112258	0.127858	0.120476	0.123168	0.073627	0.03179
07-01-2013	0.112465	0.129295	0.121462	0.123308	0.073711	0.034556

14.2 Price Prediction Module

- Aim
  - Forecast the closing prices of selected stocks for a time horizon of up to two months.
  - Utilize Long Short-Term Memory (LSTM) networks to model the sequential nature of stock price data.
  - Build a sequential model architecture as outlined in Fig 14.2.2, detailing the layers and hyperparameters of the LSTM network.
  - Specify the plotting approach used to compare actual stock prices with the model's predictions (as illustrated in Fig. 14.2.3).

- Methodology
  - Future prices of selected stocks for up to 2 months will be predicted.
  - Long Short-Term Memory (LSTM) will be used for prediction of the prices.
  - The reason behind using LSTM in time-series prediction is that LSTMs have higher memory power than RNNs for a more extended period.

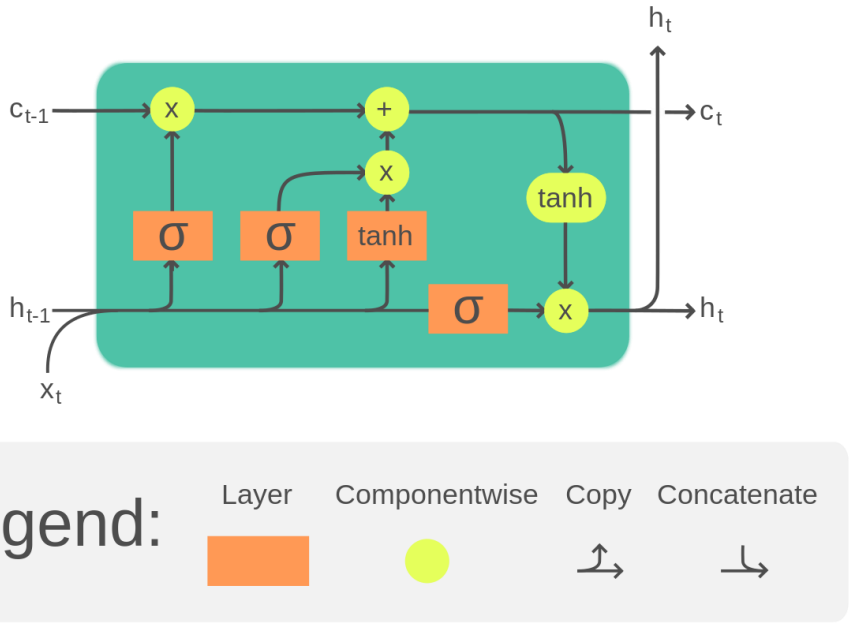

Fig 14.2.1: Architecture of LSTM cell

– Following image shows actual architecture of the model used for predicting the stock prices

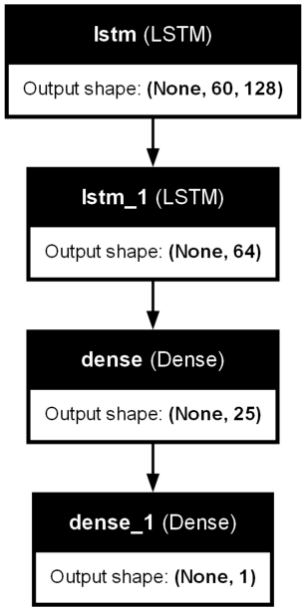

Fig 14.2.2: Architecture of the Actual ML Model

- Procedure
  – Following code snippet shows how the model is created and trained

```
model = Sequential()
model.add(LSTM(128, return_sequences=True,
input_shape=(60,1)))
model.add(LSTM(64, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))
model.compile(optimizer='adam',
loss='binary_crossentropy',
metrics=['accuracy'])
history = model.fit(x_train, y_train,
batch_size=1, epochs=1)
```

  – Following code snippet shows how predictions are made using the trained model

```
predictions = model.predict(x_test)
```

```
predictions =
scaler.inverse_transform(predictions)
```

- Plotting the results
  - Following snippet shows how the results are plotted.

```
# Plot the data
train = data[:training_data_len]
valid = data[training_data_len:]
valid['Predictions'] = predictions

# Visualize the data
plt.figure(figsize=(16,6))
plt.title('Model')
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price INR', fontsize=18)
plt.plot(train['Close'], linewidth=1)
plt.plot(valid[['Close', 'Predictions']],
linewidth=1)
plt.legend(['Train', 'Val', 'Predictions'],
loc='lower right')
plt.show()
```

Fig 14.2.3: 2-months' price prediction for TCS

14.3 News Feed Module

- Aim
  - Aggregate news headlines from several sources.
  - Provide the cleaned data to the Sentiment Analysis module.
  - The current source is marketaux.com.

- Methodology
  - The module is specifically designed to gather targeted news related to stock market movements. It filters news based on stock symbols and potentially regional preferences.
  - It forgoes the use of a dedicated news library like NewsApiClient and interacts directly with the MarketAux API using basic HTTP requests.
  - Following code snippet demonstrates the working of this module

```
conn =
http.client.HTTPConnection("api.marketaux.com"
)
params = urllib.parse.urlencode(
            {
                "api_token":
CONSTANTS.marketaux_api_token,
                "symbols": f"SENSEX",
                "countries": "in"
            }
        )
conn.request("GET","/v1/news/all?{}".format(pa
rams))
res = conn.getresponse()
data = res.read()
print(data.decode("utf-8"))
```

	text	sentiment
0	Stock Market Today Live Updates: Sensex and Nifty opened in red after making a record high on Monday. Asian indices remained largely in negative territory with Nikkei still above the 40,000 mark.	
1	Share Market Today Live Updates: Brent crude oil price is trading 0.14 per cent lower at $86.77 a barrel while US WTI is down 0.16 per cent at $82.59 a barrel.	
2	The broader Nifty 50 jumped 355.95 points, or 1.62 per cent to close at a record high of 22,338.75.	
3	The BSEâ€™s 30-share Sensex jumped 408.86 points or 0.55 per cent to close at a life time high of 74,085.99. The broader Nifty surged 0.53 per cent, or 117.75 points to a record high of 22,474.05.	

Fig 14.3: News fetched from the Internet using the News Feed Module

14.4 Sentiment Analysis Module

- Aim
  - To extract relevant sentiment indicators from various textual sources, quantifying and categorizing them to determine the overall market or public opinion about a specific stock or sector.
  - Develop or implement techniques to accurately detect positive, negative, and neutral sentiment within the gathered textual data.
  - Consider classifying sentiment into finer-grained emotions (e.g., fear, optimism, joy) for more nuanced analysis.
- Methodology
  - Combines a rule-based lexicon (the custom dictionary) with the machine-learning-trained VADER model for sentiment analysis.
  - Explicitly retrieves news related to stocks, making sentiment analysis more tailored to stock price prediction.
  - Provides simplified sentiment representation (positive, negative, neutral) for potential integration with a price prediction model.
  - Following code snippet shows the functioning of the Sentiment Analysis Module.

```
def getNews(self, query):
    self.news =
    self.News.get_everything(query)

def preprocessText(self, text):
    tokens = word_tokenize(text.lower())
    filtered_tokens = [token for token in
    tokens if token not in
    stopwords.words('english')]
    lemmatizer = WordNetLemmatizer()
        lemmatized_tokens =
     [lemmatizer.lemmatize(token) for token in
     filtered_tokens]
    processed_text = '.join(lemmatized_tokens)
    return processed_text

def getSentiments(self):
     self.getNews("Sensex")
```

```
self.news =
self.news.filter(items=["text"])
scores =
self.news["text"].apply(self.analyzer.pol
arity_scores)
scores_df =
pd.DataFrame.from_records(scores)
self.news = self.news.join(scores_df)
idx = 0
self.news["compound"] =
self.news["compound"].apply(lambda x: 1
if x > 0.2 else (-1 if x < -0.2 else 0))
```

– After the above processing, the overall sentiment is shown in the form of a Pie Chart.

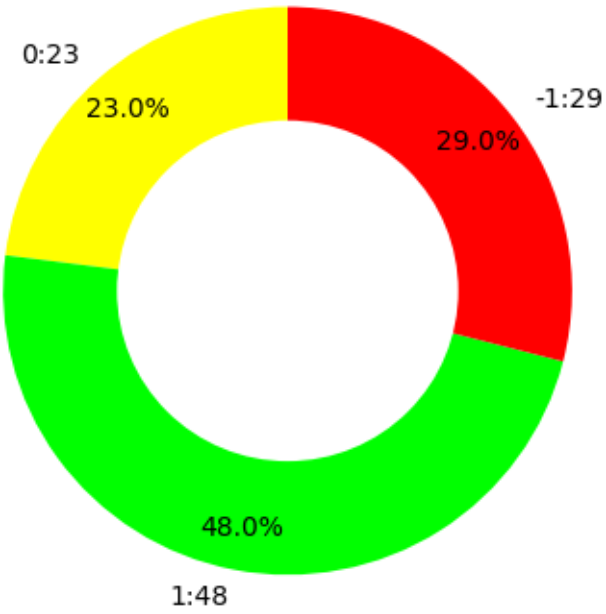

Fig 14.4: Overall market sentiment,

Green: Bullish; Yellow: Neutral; Red: Bearish

15. Test Cases

15.1 Test Case: RELIANCE.NS

- Data Acquisition and Management
  - Raw Data

Date	Open	High	Low	Close	Adj Close	Volume
02-01-2012	318.5535	324.4738	314.1419	323.0109	294.1662	9404053
03-01-2012	325.6167	332.3142	324.7481	331.2856	301.702	10244609
04-01-2012	332.5428	334.1657	326.1196	327.3997	298.1631	9270951
05-01-2012	326.8739	331.4228	317.3192	319.7879	291.231	14479600
06-01-2012	318.6678	330.9884	318.485	328.0397	298.7459	10287605

  - Pre-processed Data

Date	Open	Close	High	Low	Volume	Adj Close
02-01-2012	0.003687	0.005145	0.004542	0.002417	0.131817	0.004638
03-01-2012	0.0063	0.008203	0.007432	0.006394	0.143599	0.007395
04-01-2012	0.008863	0.006767	0.008114	0.006909	0.129951	0.0061
05-01-2012	0.006765	0.003954	0.007103	0.003609	0.202961	0.003564
06-01-2012	0.003729	0.007004	0.006943	0.004046	0.144202	0.006314

- Price Prediction Module
  - 2-month's predicted price of RELIANCE.NS

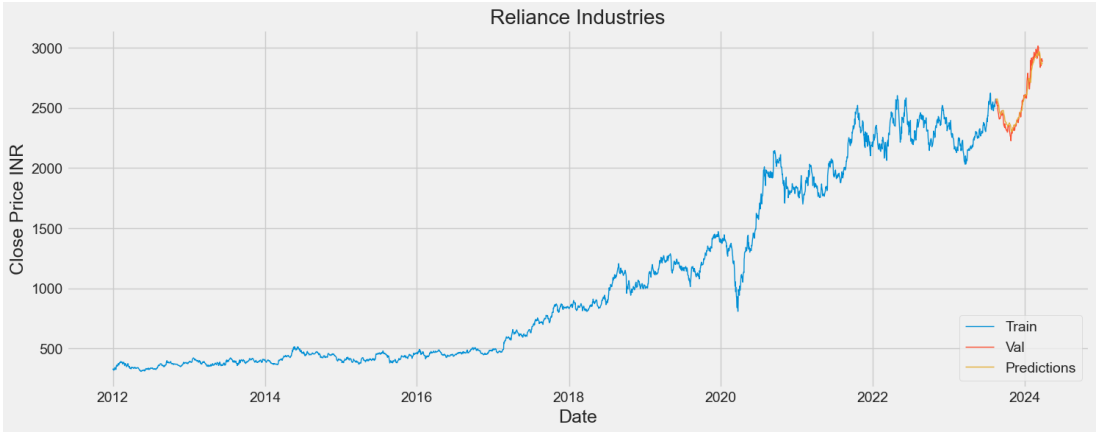

Fig 15.1.1: Predicted price movement of RELIANCE.NS

- News Feed Module
  - News related to RELIANCE INDUSTRIES

text	sentiment
Other guests expected to attend the lavish celebrations include Sundar Pichai, Bob Iger, and Ivanka Trump, reports say.	
Disney and billionaire Mukesh Ambaniâ€™s conglomerate have signed a binding pact to merge their media operations in India, creating a sector behemoth valued at $8.5 billion.	
Nita Ambani casually wore an enormous emerald necklace and a 52.58-carat diamond ring dubbed the Mirror of Paradise.	
The Ambani family hosted a three-day bash in Jamnagar to celebrate the upcoming wedding between Anant Ambani and Radhika Merchant.	
"You know, I never really wanted to get a watch. But after seeing that, I was like, watches are cool," Zuckerberg said of Ambani's watch.	

- Sentiment Analysis Module
  - Overall sentiment for RELIANCE.NS

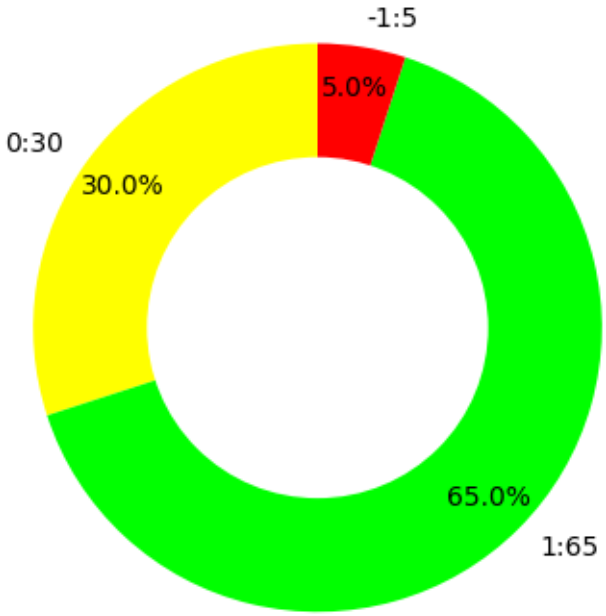

Fig 15.1.2: Overall sentiment for RELANCE.NS

15.2 Test Case: INFY.NS

- Data Acquisition and Management
  - Raw Data

Date	Open	High	Low	Close	Adj Close	Volume
02-01-2012	344.9	352.0812	342.75	351.1187	268.8162	5826176
03-01-2012	353.0125	360.8687	352.5	358.0375	274.1132	8903008
04-01-2012	355.625	359.375	355.1313	356.8	273.1657	7341424
05-01-2012	354.9	359.5	353.9312	355.0187	271.802	7125272
06-01-2012	354.75	358.45	351.3875	354.0187	271.0364	7087632

  - Pre-processed Data

Date	Open	Close	High	Low	Volume	Adj Close
02-01-2012	0.044889	0.05116	0.048462	0.050917	0.03503	0.039326
03-01-2012	0.049751	0.055293	0.053682	0.056745	0.053529	0.042546
04-01-2012	0.051317	0.054554	0.052795	0.058318	0.04414	0.04197
05-01-2012	0.050882	0.05349	0.052869	0.057601	0.042841	0.041141
06-01-2012	0.050793	0.052893	0.052245	0.05608	0.042614	0.040676

- Price Prediction Module
  - 2 months' predicted price of INFY.NS

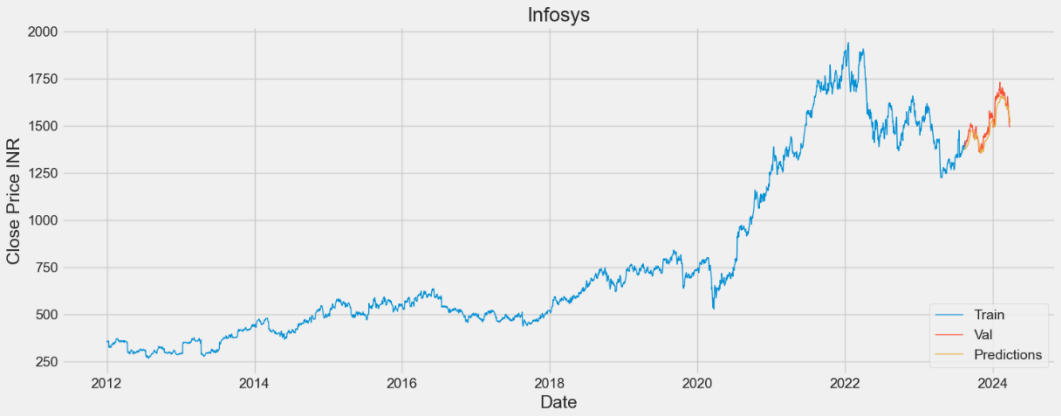

Fig 15.2.1: Predicted price movement of INFY.NS

- News Feed Module
  - News related to INFY

Responsible AI (RAI) Office will serve as the custodian of ethical use of AI and ensure solutions align with emerging guardrails for AI across geographies BEI
The CPI is expected to have risen 0.4% in February on a monthly basis, and could heavily influence the timing of the Federal Reserve's rate cuts.
Sandip Agarwal says for Accenture, the challenge is on the discretionary side or on the consulting side, where we have small presence for some of the nam
CLSA analysts say HCL and Infosys growth guidance would be a negative catalyst for TCS, HCL and Wipro
(marketsc

- Sentiment Analysis Module
  - Overall market sentiment for INFY.NS

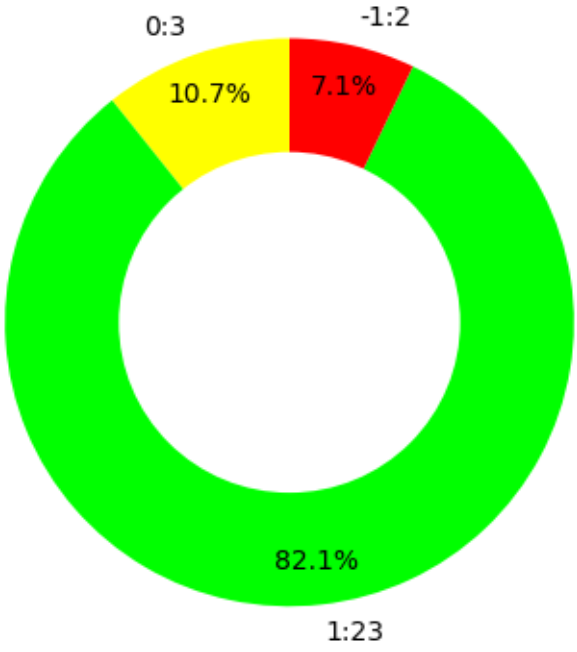

Fig 152.2: Overall sentiment for INFY.NS

Remarks/Suggestions:

Signature