Rajiv Gandhi University of Knowledge Technologies

R K Valley, YSR Kadapa (Dist) – 516330

A
Major Project Report
on
Stock price prediction using LSTM

Submitted By:-

J Chandu - R180634 T Guna Sekhar - R180663 C Anil Kumar - R180093



Under the guidance of

Mrs. V.Sravani (Assist. Professor)

Department of Computer Science Engineering

This project report has been submitted in fulfilment of the requirement for the Degree of Bachelor of Technology in Software Engineering.

December - 2023

Rajiv Gandhi University of Knowledge Technologies IIIT – R.K Valley, YSR Kadpa (Dist) – 516330



CERTIFICATE

This is to certify that report entitled "Stock price prediction using LSTM" Submitted by J.Chandu (R180634), T.Guna Sekhar (R180663), C.Anil Kumar (R180093) in partial fulfilment of the requirements of the award of bachelor of technology in computer science engineering is a bona fide work carried by them under the supervision and guidence.

The report has been not submitted previously in part or full to this or any other university or institute for the award of any degree or diploma.

GUIDE

HEAD OF THE DEPARTMENT

Mrs. V.Sravani	Mr.N.Satyanandaram
(Assist. Professor)	HOD OF CSE
Submitted for the practical examination on	

Internal Examiner

External Examinar

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible and who's constant guidance and encouragement crown all the efforts success.

I would like to express my sincere gratitude to **Mrs. V.Sravani**, my project guide for valuable suggestions and keen interest throughout the progress of our project.

I am grateful to **Mr. N. Satyanandaram HOD CSE**, for providing excellent computing facilities and congenial atmosphere for progressing my project.

At the outset, I would like to thank **Rajiv Gandhi University of Knowledge Technologies (RGUKT)**, for providing all the necessary resources and support for the successful completion of my course work.

DECLARATION

We hereby declare that this report entitled "Stock Price Prediction Using LSTM" submitted by me under the guidance and supervision of Mrs. V.Sravani, is a bonafied work. We also declare that it has not been of Submitted previously in part or in full to this University or other institution for the award of any degree or diploma.

Date : - 14 – 12 –2023 J. Chandu (R180634)

Place: - RK Valley T.Guna Sekhar (R180663)

C. Anil kumar (R180093)

INDEX

S.No	Title	Page No
1	Abstract	6
2	Introduction	7 - 8
3	Purpose	9
4	Overall Description	10 - 11
5	Existing System	12 -14
6	Proposed System	15 - 17
7	System Requirements	18
8	Tools and Technologies used	19 - 20
10	Data Collecting & Visualization	21 - 22
11	Choosing model & Training	23 - 24
12	Predictions	25
13	Conclusion	26

ABSTRACT

Stock price prediction using the LSTM Network on time series forecasting

This project focuses on leveraging Long Short-Term Memory (LSTM) networks for stock price prediction, applying them to the domain of Time Series Forecasting. LSTM networks are a specialized type of Neural Networks renowned for their ability to comprehend and model long-term dependencies, addressing challenges commonly encountered in handling sequential data. This project delves into the intricacies of LSTM architecture, highlighting its unique feature of utilizing gates to selectively add or remove information from the cell state.

The process of LSTM involves several key steps. Initially, the network determines the information to be retained or discarded from the cell. Subsequently, it identifies new information to be incorporated into the cell state. Finally, the LSTM network produces an output based on the state of the cells, offering predictions that take into account both short-term and long-term dependencies.

The project unfolds by applying LSTM to the task of Time Series Forecasting, specifically focusing on predicting stock prices. By understanding the LSTM process within the context of stock price prediction, this project aims to contribute valuable insights and methodologies to the realm of financial forecasting, potentially enhancing the accuracy and reliability of predictions in dynamic and complex markets.

INTRODUCTION

1. Project Context:

In the dynamic landscape of financial markets, accurate stock price predictions are imperative for informed decision-making. This project ventures into the realm of Time Series Forecasting using Long Short-Term Memory (LSTM) networks. In the volatile stock market, understanding and modeling long-term dependencies can significantly improve the precision of predictions, offering a valuable tool for investors and financial analysts.

2. Data and Dataset Description:

This project revolves around analyzing a dataset tailored for stock price prediction, encompassing historical stock prices, market indicators, and relevant financial metrics. The dataset, curated for its relevance to the financial domain, serves as the foundation for training and evaluating the LSTM network's performance in predicting stock prices.

3. Objective and Purpose:

The primary goal of this project is to harness the capabilities of LSTM networks for accurate and reliable stock price predictions. By diving into the intricacies of LSTM architecture and its unique ability to capture both short-term fluctuations and long-term trends, the project aims to advance methodologies in Time Series Forecasting within the financial sector.

4. Scope and Approach:

This analysis focuses on deploying LSTM networks to forecast stock prices, utilizing historical data and relevant market indicators. The approach involves training the network to recognize patterns,

learn from past trends, and make informed predictions. The scope encompasses understanding the LSTM process in the context of financial forecasting, with an emphasis on unveiling insights that can aid investors in navigating complex and dynamic markets.

5. Expected Outcomes:

Anticipated outcomes from this project include enhanced precision in stock price predictions, the identification of significant market trends, and a nuanced understanding of the interplay between short-term fluctuations and long-term market dynamics. By the project's conclusion, readers can expect to glean insights that contribute to refining financial forecasting models and strategies.

6. Significance:

The significance of this project lies in its potential to revolutionize financial forecasting practices. The application of LSTM networks to stock price prediction not only addresses the challenges of handling sequential financial data but also opens avenues for more accurate predictions in real-world market scenarios. Stakeholders in the financial sector stand to benefit from improved decision-making, risk management, and strategic planning based on the insights derived from this innovative

PURPOSE

The purpose of this project is to harness the capabilities of Long Short-Term Memory (LSTM) networks in the domain of Time Series Forecasting, with a specific focus on predicting stock prices. LSTM networks, known for their proficiency in capturing and modeling long-term dependencies within sequential data, present a novel approach to address challenges prevalent in handling financial time series data.

The project aims to delve into the intricacies of LSTM architecture, emphasizing its distinctive feature of employing gates to selectively include or exclude information from the cell state. By comprehensively understanding the LSTM process, the project seeks to contribute valuable insights and methodologies to the field of financial forecasting.

Through the application of LSTM networks to the task of predicting stock prices, the project endeavors to provide a nuanced understanding of how these networks can effectively handle the complexities of dynamic and intricate financial markets. The overarching purpose is to enhance the accuracy and reliability of stock price predictions by leveraging the unique capabilities of LSTM networks, thereby offering valuable contributions to the advancement of financial forecasting techniques..

OVERALL DESCRIPTION

- **1.Company Name**: Indicates the name of company (e.g., Apple,google,Netflix,amazon) chosen by users.
- 2.. **Date**: The date of a particular stock market trading day.
- 3. **Open**: The opening price of a stock on a given trading day.
- 4. **High**: The highest price reached by a stock during a trading day.
- 5. **Low**: The lowest price reached by a stock during a trading day.
- 6. **Close**: The closing price of a stock at the end of a trading day.
- 7. **Adj Close**: The adjusted closing price, accounting for corporate actions such as dividends and stock splits.
- 8.**Volume**: The total number of shares traded for a particular stock on a given trading day.

The purpose of this project is to train the LSTM network model and predict the future Stock close values, we aim to achieve the following objectives:

1. Capture Patterns Over Time:

- Leverage LSTM's ability to understand long-term trends in stock data for more accurate predictions over extended periods.

2. Handle Sequential Patterns:

- Effectively process the order and sequence of historical stock prices, recognizing the importance of temporal relationships.

3. Understand Complex Relationships:

- Model non-linear relationships within stock price movements, capturing the intricate nature of market dynamics.

4. Adapt to Relevant Information:

- Selectively use important information while filtering out noise, improving the network's ability to make informed predictions.

5. Consider Short and Long-Term Trends:

- Make predictions that factor in both short-term fluctuations and long-term patterns, providing a comprehensive view of market behavior.

6. Contribute to Financial Forecasting:

- Provide insights and methods to improve financial forecasting practices, offering practical tools for investors and analysts.

7. Improve Prediction Accuracy:

- Enhance the precision and reliability of predicting future stock prices, empowering better decision-making in the dynamic financial landscape.

EXISTING SYSTEM

Existing System: Stock price prediction using LSTM Network

1. Python Programming Language:

- Python is a versatile and widely-used programming language in data analysis and scientific computing.
- It provides a rich ecosystem of libraries (such as Pandas, NumPy, Matplotlib, Seaborn) that are essential for data manipulation, analysis, and visualization.

2. Jupyter Notebook:

- Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text.
- It provides an interactive environment where you can write and execute code in a modular and organized manner.

3. Data Loading and Preprocessing:

- Python's Pandas library is used to load, manipulate, and preprocess the dataset. You can read data from various file formats (e.g., CSV, Excel) into Pandas DataFrames for analysis.
- Data cleaning, handling missing values, and performing basic transformations are done using Pandas functions.

4. Data Visualization:

- Matplotlib and Seaborn are Python libraries commonly used for creating a variety of visualizations, including line charts, bar plots, scatter plots, histograms, and more.
- These libraries allow you to visually explore the data, identify patterns, and present insights.

5. Statistical Analysis:

- NumPy provides tools for numerical computations and statistical operations. You can calculate summary statistics, correlations, and other metrics relevant to LSTM.
- Scipy, another library, offers additional statistical functions and hypothesis testing capabilities.

6. ML Libraries:

TensorFlow, Keras, PyTorch, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn,

Statsmodels.

7. Interactive Exploration:

- Jupyter Notebook's interactive environment enables you to write and execute code cells in a step-by-step manner. This facilitates the iterative nature of LSTM.
- You can include text, explanations, and markdown cells to document your thought process and findings.

8. Collaborative Sharing:

- Jupyter Notebook documents can be easily shared with colleagues, stakeholders, or collaborators, allowing them to reproduce your analysis and understand your insights.

9. Python Ecosystem:

- Beyond the libraries mentioned, Python offers a wide range of additional libraries that you can leverage based on the specific needs of your analysis.

The combination of Python and Jupyter Notebook provides a powerful platform for conducting EDA. It allows you to load, manipulate, visualize, and analyze data in an interactive and flexible manner, making it a popular choice among data analysts and scientists.

Drawbacks of Existing System

- **1. Manual Analysis:** Current analysis is manual and time-consuming, prone to errors.
- **2. Limited Visualization :** Lack of interactive visualizations hampers effective data exploration.
- **3. Handling Large Datasets:** Challenges with processing and analyzing large datasets efficiently.
- **4. Limited Insights:** Inadequate tools for advanced insights and patterns discovery.
- **5. No Automation :** Lack of automation for ongoing data updates.
- 6. Security Risks: Potential security and data privacy concerns.
- 7. Inefficient Decision-Making: Lack of dynamic visualizations affects decision-making.
- **8.** Collaboration Issues: Difficulty in collaborative analysis and sharing insights.
- **9. Inconsistent Processes :** Absence of standardized analysis methods.
- 10. Limited Relationship Analysis: Inability to explore complex variable relationships
- 11. Everytime Training:

Need Continuous training for the predictions.

PROPOSED SYSTEM

System Overview:

The proposed system will be a run time program which runs on the jupyter-notebook[a python interpreter] to facilitate the Stock price prediction on the yahoo finance data. The people who are investing in the particular stocks are gaining the prediction price regarding stocks and make data-driven decisions.

Key Features:

1. Data Import and Preprocessing:

- User is going to download the stocks data on yahoo finance website.
- The system will handle data preprocessing tasks such as handling missing values and outliers.

2. Data Exploration and Visualization:

- Users can select attributes of interest (e.g., open,high,low,close,volume) for analysis.
- The system will generate summary statistics and visualizations (e.g., scatter plots) to provide an initial overview of the data.

3. User Stock Analysis:

- The system will offer to represent how the stocks prices are going up and down in the time series.

4. Geographical Insights:

- The system will display a map visualization to show the Stock prices on year by year.
- Users can zoom in on specific year and view stock prices.

5. Model training And testing:

- Divide the data into Training and testing and train the model using the training data.

6. Calculate mean squared Error:

- By using the testing data calculate the error value and find the error loss then fit the model.

7. Future value prediction:

- Allows users to give the input values and it will predict the the final values.

• Non-Functional Requirements :

1. Usability:

-The user interface will be intuitive, with interactive controls to customize visualizations and explore data subsets.

2. Performance:

-The system will be responsive, generating visualizations and summary statistics in realtime for a smooth user experience.

3. Security:

- User authentication and access control will ensure that only authorized users can upload data and perform analysis.

4. Scalability:

- The system will handle moderate-sized datasets and provide efficient analysis even with a large number of concurrent users.

Technology Stack:

The proposed system will be Done using Data analysis libraries (e.g., Pandas, Matplotlib, Plotly) will be utilized for processing and visualization. And machine learning model(LSTM networks) for the future stock price predictions.

Benefits:

The LSTM network project for stock price prediction provides enhanced accuracy by leveraging the network's capacity to capture both short-term fluctuations and long-term dependencies. This adaptability to time series data and non-linear relationships contributes to more reliable forecasts, aiding risk management and informed decision-making in dynamic financial markets. The project's potential for real-time predictions and innovation in financial technology further positions it to offer valuable insights for investors and financial professionals.

SYSTEM REQUIREMENTS

> SOFTWARE COMPONENTS

- Windows / Ubuntu
- Jupyter Notebook
- Terminal
- Technologies : Python
- Modules: Numpy, Pandas, Matplotlib, Datetime, Keras, tensorflow, Scikit-learn.

► HARDWARE COMPONENTS

- Processor Core i5
- Hard Disk 512 GB
- RAM 8 GB
- Internet Connections

TOOLS AND TECHNOLOGIES USED

1. Python:

- Python is a popular programming language widely used for data analysis, machine learning, and scientific computing.
- It provides a rich ecosystem of libraries and tools for data manipulation, analysis, and visualization.

2. Jupyter Notebook:

- Jupyter Notebook is an interactive web-based environment that allows you to create and share documents containing live code, equations, visualizations, and narrative text.
- It's commonly used for data analysis and exploration due to its ability to combine code execution with explanatory text.

3. NumPy:

- NumPy is a fundamental library for numerical computations in Python.
- It provides support for large, multi-dimensional arrays and matrices, along with a variety of mathematical functions to operate on these arrays efficiently.

4. Pandas:

- Pandas is a powerful library for data manipulation and analysis.
- It offers data structures like Series and DataFrame, which simplify handling and analyzing structured data, such as CSV files, SQL tables, and Excel spreadsheets.

_

5. Matplotlib:

- Matplotlib is a widely used 2D plotting library in Python.
- It provides various types of plots and charts for visualizing data, making it suitable for creating static, interactive, and animated visualizations.

6. Keras:

- An open-source high-level neural networks API written in Python. Keras serves as a user-friendly interface to TensorFlow and other deep learning libraries, making it convenient for constructing LSTM models.

7. Scikit-learn:

- A machine learning library for classical machine learning algorithms, including preprocessing, model selection, and evaluation. While Scikit-learn doesn't directly support LSTMs, it can be useful for data preprocessing and post-processing tasks.

8. Tensor Flow:

- An open-source machine learning library developed by the Google Brain team. TensorFlow provides comprehensive support for building and training deep neural networks, including LSTM networks.

DATA COLLECTING

Collecting the data from the Stocks(Eg:APPL{apple})

```
In [9]: import pandas as pd
        import yfinance as yf
        import datetime
        from datetime import date, timedelta
        today = date.today()
        d1 = today.strftime("%Y-%m-%d")
        end_date = d1
        d2 = date.today() - timedelta(days=5000)
        d2 = d2.strftime("%Y-%m-%d")
        start date = d2
        data = yf.download('AAPL',
                              start=start_date,
                             end=end_date,
                             progress=False)
        data["Date"] = data.index
        data.reset index(drop=True, inplace=True)
        data.tail(\bar{15})
Out[9]:
                                                        Close Adj Close
         3433 2023-11-22 191.490005 192.929993 190.830002 191.309998 191.309998 39617700
         3434 2023-11-24 190 869995 190 899994 189 250000 189 970001 189 970001 24048300
         3435 2023-11-27 189.919998 190.669998 188.899994 189.789993 189.789993 40552600
```

 State
 Open
 High
 Low
 Close
 Adj Close
 Volume

 3433
 2023-11-22
 191.490005
 192.929993
 190.830002
 191.309998
 191.309998
 39617700

 3434
 2023-11-24
 190.869995
 190.869998
 188.899994
 189.789993
 189.789993
 40552600

 3435
 2023-11-27
 189.919998
 190.869998
 189.399994
 190.399994
 190.399994
 38415400

 3436
 2023-11-29
 190.899994
 192.089996
 188.970001
 189.369995
 189.369995
 4301420

 3438
 2023-11-30
 189.839996
 190.320007
 188.190002
 189.949997
 189.949997
 4879400

 3449
 2023-12-01
 190.330002
 191.559998
 189.229996
 191.240005
 191.240005
 45679300

 3440
 2023-12-02
 190.210007
 194.399994
 190.179993
 193.419998
 193.419998
 6628400

 3442
 2023-12-05
 194.449997
 194.759995
 192.110001
 192.320007
 192.320007
 4108700</t

```
In [8]: data.dtypes
Out[8]: Date
                      datetime64[ns]
         Open
                             float64
         High
                              float64
         Low
                              float64
                             float64
         Close
         Adj Close
                             float64
                               int64
         Volume
         dtype: object
In [11]: data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 3448 entries, 0 to 3447
         Data columns (total 7 columns):
              Column
                         Non-Null Count Dtype
          0
              Date
                         3448 non-null
                                         datetime64[ns]
                                         float64
float64
              Open
                         3448 non-null
                         3448 non-null
              High
                         3448 non-null
                                         float64
              Low
                                         float64
                         3448 non-null
              Close
              Adj Close 3448 non-null
                                         float64
              Volume
                         3448 non-null
         dtypes: datetime64[ns](1), float64(5), int64(1)
         memory usage: 188.7 KB
```

2) Data Visualization

Apple Stock Price Analysis



3) Finding Correlation coefficient:

4) Dividing data Into training and testing sets:

```
In [8]: x = data["Open", "High", "Low", "Volume"]]
y = data["Close"]
x = x.to_numpy()
y = y.to_numpy()
y = y.reshape(-1, 1)

from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)
```

5) Choosing model:

```
In [9]: from keras.models import Sequential
    from keras.layers import Dense, LSTM
    model = Sequential()
    model.add(LSTM(128, return_sequences=True, input_shape= (xtrain.shape[1], 1)))
    model.add(LSTM(64, return_sequences=False))
    model.add(Dense(25))
    model.add(Dense(1))
    model.add(Dense(1))
```

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 4, 128)	66560
lstm_1 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 25)	1625
dense_1 (Dense)	(None, 1)	26
Total params: 117619 (:========

Non-trainable params: 0 (0.00 Byte)

```
In [10]: model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(xtrain, ytrain, batch_size=1, epochs=30)
```

```
Epoch 1/30
WARNING:tensorflow:From C:\Users\91944\AppData\Roaming\Python\Python310\site-packages\keras\src\utils\tf_utils.py:492: The name
tf. ragged. Ragged Tensor Value \ is \ deprecated. \ Please \ use \ tf. compat. v1. ragged. Ragged Tensor Value \ instead.
2758/2758 [============= ] - 32s 9ms/step - loss: 405.8429
Epoch 2/30
2758/2758 [=
        Epoch 3/30
Epoch 5/30
Epoch 6/30
2758/2758 [=========== ] - 19s 7ms/step - loss: 7.5441
Epoch 7/30
2758/2758 [==
       Epoch 8/30
2758/2758 [=
       Epoch 9/30
Epoch 10/30
2758/2758 [============== ] - 18s 7ms/step - loss: 8.3694
Epoch 11/30
2758/2758 [=
         -----] - 18s 6ms/step - loss: 5.8218
Epoch 12/30
2758/2758 [===========] - 18s 7ms/step - loss: 6.9706
Epoch 13/30
2758/2758 [============ ] - 18s 7ms/step - loss: 6.3067
Epoch 14/30
2758/2758 [============= ] - 18s 6ms/step - loss: 7.1384
Epoch 15/30
2758/2758 [=:
        ======== - loss: 4.5627
Epoch 16/30
2758/2758 [=
         Epoch 17/30
2758/2758 [===========] - 17s 6ms/step - loss: 5.1856
Epoch 18/30
2758/2758 [=
        Epoch 19/30
2758/2758 [===========] - 17s 6ms/step - loss: 5.5078
Epoch 20/30
2758/2758 [=============== ] - 18s 7ms/step - loss: 7.9913
Epoch 21/30
2758/2758 [=
         -----] - 18s 7ms/step - loss: 4.4081
Epoch 22/30
2758/2758 [===
        Epoch 23/30
2758/2758 [===========] - 19s 7ms/step - loss: 4.9961
Epoch 24/30
2758/2758 [============= ] - 19s 7ms/step - loss: 5.1487
Epoch 25/30
Epoch 26/30
2758/2758 [=
           Epoch 27/30
2758/2758 [============ ] - 18s 7ms/step - loss: 5.5437
Epoch 28/30
2758/2758 [=
          Epoch 29/30
2758/2758 [===
        Epoch 30/30
2758/2758 [============= ] - 19s 7ms/step - loss: 3.6017
```

Out[10]: <keras.src.callbacks.History at 0x25405b26230>

7)Prediction in real life

This's how it predict the Close values of Stocks based on the high, low, and Volume of the Stocks....

CONCLUSION

In conclusion, this project demonstrates the effective application of Long Short-Term Memory (LSTM) networks in the domain of Time Series Forecasting, with a specific emphasis on predicting stock prices. By delving into the intricacies of LSTM architecture and highlighting its unique features, the project establishes a foundation for understanding how LSTM networks can comprehend and model both short-term and long-term dependencies in sequential data. The outlined process of LSTM, involving the selective retention and incorporation of information through gated mechanisms, showcases the network's ability to handle complex data patterns.

Through the practical implementation of LSTM in stock price prediction, this project contributes valuable insights and methodologies to the field of financial forecasting. The utilization of LSTM networks opens avenues for enhancing the accuracy and reliability of predictions in dynamic and complex markets, providing a promising approach to addressing challenges inherent in Time Series Forecasting. Ultimately, the project underscores the potential of LSTM networks as a powerful tool for improving predictive