STOCK PRICE PREDICTION AND FORECASTING USING RNN AND LSTM

Project Report submitted to
Shri Ramdeobaba College of Engineering & Management,
Nagpur in partial fulfilment of requirement for the award of
degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

Srushti Dhakate (25)
Prathamesh Gujar (52)
Prathamesh Rajbhoj (53)
Utkarsh Sathawane (68)

Guide

Prof. S. Jain

RCOEM
Shri Ramdeobaba College of
Engineering and Management, Nagpur

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering and Management, Nagpur 440013

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

Dec 2023

STOCK PRICE PREDICTION AND FORECASTING USING RNN AND LSTM

Project Report submitted to

Shri Ramdeobaba College of Engineering & Management, Nagpur in partial fulfilment of requirement for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

Srushti Dhakate (25)
Prathamesh Gujar (52)
Prathamesh Rajbhoj (53)
Utkarsh Sathawane (68)

Guide

Prof. S. Jain

RCOEM
Shri Ramdeobaba College of Engineering and Management, Nagpur

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering and Management, Nagpur 440013

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

Dec 2023

SHRI RAMDEOBABA COLLEGE OF ENGINEERING & MANAGEMENT, NAGPUR

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Thesis on **"Stock Price Prediction And Forecasting Using RNN and LSTM "** is a bonafide work of Srushti Dhakate, Prathamesh Gujar, Prathamesh Rajbhoj and Utkarsh Sathawane submitted to the Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur in partial fulfilment of the award of a Degree of Bachelor of Engineering. It has been carried out at the Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur during the academic year 2023-2024.

Date:

Place: Nagpur

Prof. Sweta Jain	Dr. Ramchand Hablani
Project Guide	H.O.D
Department of Computer Science and Engineering	Department of Computer Science and Engineering

Dr. R. S. Pande

Principal

DECLARATION

I, hereby declare the thesis “STOCK PRICE PREDICTION AND FORECASTING USING RNN AND LSTM ” submitted herein, has been carried out in the Department of Computer Science and Engineering of Shri Ramdeobaba College of Engineering & Management, Nagpur. The work is original and has not been submitted earlier as a whole or part for the award of any degree / diploma at this or any other Institute / University.

Date:

Place: Nagpur

Name of the Students	Roll No.	Signature
Srushti Dhakate	25	
Prathamesh Gujar	52	
Prathamesh Rajbhoj	53	
Utkarsh Sathawane	68	

APPROVAL SHEET

This report entitled

"STOCK PRICE PREDICTION AND FORECASTING USING RNN AND LSTM "

By

Srushti Dhakate

Prathamesh Gujar

Prathamesh Rajbhoj

Utkarsh Sathawane

is approved for the degree of Bachelor of Engineering.

Name & signature of Supervisor(s)

Name & signature of External Examiner(s)

Name & signature of HOD

Date:

Place: Nagpur

ACKNOWLEDGEMENT

The project is a combined effort of a group of individuals who synergize to contribute towards the desired objectives. Apart from the efforts by us, the success of the project shares an equal proportion on the engagement and guidance of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

I would like to extend my heartfelt thanks to Prof. S. Jain for her support and encouragement during this project. Her knowledge, expertise, and willingness to help have been instrumental in the successful completion of this work.

We would like to show our greatest appreciation towards Dr. Ramchand Hablani, Head, Department of Computer Science and Engineering, RCOEM, Nagpur for providing us with the facilities for completing this project and for his constant guidance.

We would like to express our deepest gratitude to Dr. Rajesh S. Pande, Principal, RCOEM, Nagpur for providing us the opportunity to embark on this project.

Finally, we extend our gratitude to all the faculty members of the CSE department who have always been so supportive and provided resources needed in our project development. We are very grateful to all of them who have unconditionally supported us throughout the project.

Date: — Projectees

ABSTRACT

The goal of this research is to use the capabilities of advanced machine learning techniques, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, to address the difficulties associated with stock price prediction and forecasting. Making wise investment decisions requires accurate predictions in the dynamic and volatile world of financial markets. Investors are exposed to unanticipated risks when traditional methods fail to capture the complex patterns and dependencies present in time-series financial data.

The principal aim of this study is to create and apply a predictive model that utilises RNNs and LSTMs to improve the precision of stock price predictions. These neural network architectures are especially useful for modelling the intricate and dynamic nature of stock prices because they are well-suited for capturing sequential dependencies. By delving deeply into the application of RNNs and LSTMs to stock market analysis, this study aims to offer a thorough grasp of the possible advantages and difficulties related to these cutting-edge machine learning methods.

The proposed model's performance against conventional forecasting techniques, its robustness in managing market volatility, and the economic impact of better stock price predictions on investment decision-making are all part of the research objectives. The study also intends to advance knowledge of how artificial intelligence is changing investment practices, which will benefit the larger field of financial analytics.

The results of this study should improve stock price forecasts and have an impact on risk management techniques, investment strategies, and overall financial stability. The suggested model has the ability to enable stakeholders to make wise decisions in the face of market uncertainty by giving investors more trustworthy insights into market trends. This research aims to add to the current discussion on the use of RNNs and LSTMs in stock price prediction by disseminating its findings and useful suggestions. It also provides an insightful look at the developing interface between sophisticated machine learning methods and financial analytics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Expansion
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
MARS	Multivariate Adaptive Regression Splines
ARIMA	Autoregressive Integrated Moving Average
PCA	Principal Component Analysis
TSLM	Time Series Linear Models
PIL	Python Imaging Library
DL	Deep Learning
ML	Machine Learning
GPU	Graphics Processing Unit
TL	Transfer Learning
API	Application Programming Interface
PIL	Python Imaging Library
OS	Operating Systems
AUC	Area Under the Curve
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

CHAPTER 1

INTRODUCTION

Making wise investment decisions requires being able to anticipate and forecast stock prices in the fast-paced world of financial markets. In order to improve the accuracy of stock price predictions, this paper focuses on the integration of cutting-edge machine learning techniques, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks. Proven for their ability to capture sequential dependencies in time-series data, RNNs and LSTMs show promise in modelling the complex dynamics of stock market fluctuations. This report's main goal is to examine the fundamental ideas behind stock price forecasting and prediction while evaluating the performance of LSTMs and RNNs in this particular setting. Through a thorough analysis of relevant literature and case studies, readers will acquire valuable insights into the dynamic financial market environment.

In the sections that follow, we'll explore the fundamentals of stock market analysis, the importance of precise forecasts in financial decision-making, and the inner workings of RNNs and LSTMs with regard to stock price forecasting. Readers will gain a sophisticated understanding of the developing relationship between financial modelling and artificial intelligence through this investigation.

1.1 Problem Definition

Predicting stock prices is difficult because of the financial markets' inherent volatility and complexity. Time-series data frequently contains complex patterns and dependencies that are difficult for conventional methods to capture. The problem statement focuses on creating precise forecasting models using Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN). The intention is to overcome the drawbacks of conventional methods and offer investors more trustworthy forecasts of future stock price movements so they can make wise decisions in ever-changing financial landscapes.

1.2 Motivation

The necessity for sophisticated tools to navigate the complex world of financial markets is the driving force behind researching STOCK PRICE PREDICTION AND FORECASTING USING RNN AND LSTM. Conventional approaches frequently fail to capture the complex dynamics and quick changes that are present in stock prices, making investors vulnerable to unanticipated risks. A strong solution is provided by the development of Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks. These sophisticated machine learning models are particularly good at identifying sequential patterns in time-series data, which offers a chance to improve the precision of stock price forecasts.It is crucial to use RNNs and LSTMs for stock market analysis in this era of unprecedented data availability. Giving investors access to a more advanced and reliable predictive framework will enable them to make wise decisions even in the face of market volatility. Furthermore, enhanced stock price forecasts have a significant potential economic impact that could affect risk management, investment strategies, and overall financial stability.

By utilizing these state-of-the-art tools, we hope to further the development of financial forecasting and promote a more flexible and resilient approach to investing in a world where the state of the economy is constantly shifting. The motivation behind this research stems from the belief that incorporating RNNs and LSTMs into stock price prediction models will be a major step toward developing a more precise, trustworthy, and perceptive understanding of financial markets.

1.3 Overview

A groundbreaking foray into the field of financial analytics, stock price prediction and forecasting employing Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks is motivated by the urgent need for sophisticated predictive models in a more intricate and dynamic trading environment. The goal of this study is to overcome the shortcomings of conventional stock price prediction techniques, which frequently fail to identify the complex relationships and patterns present in time-series financial data.

Investors face many difficulties as a result of the volatility and unpredictability of stock prices, which calls for a more advanced and flexible method of market analysis.

This research is primarily driven by the possibility of a paradigm shift in the way that investments are chosen. Investors are exposed to unanticipated risks and market fluctuations because conventional methods frequently fail to provide timely and accurate predictions. With the help of RNNs and LSTMs, which are excellent at simulating sequential dependencies in data, we hope to create a more sophisticated understanding of stock price fluctuations. Investors are then equipped with timely and accurate insights as a result, allowing them to trade the financial markets with more assurance and accuracy.

The incentive also encompasses the wider economic ramifications of more accurate stock price forecasts. For investment strategies, risk management, and overall financial stability, the ability to predict market trends is critical in today's globalised and interconnected financial landscape. Improved prediction models could impact resource distribution, improve portfolio management, and support a more flexible and resilient financial system.

The incorporation of RNNs and LSTMs into stock price prediction models signifies a paradigm shift in financial analytics as we begin this investigation. The goal of this research is to advance predictive modelling by providing a forward-looking viewpoint consistent with the never-before-seen accessibility of financial data. Combining cutting-edge machine learning methods with the complexities of financial markets could improve stock price forecasts and change the way that people invest in the face of a constantly shifting global economic environment. Our goal in conducting this research is to spur innovations that will enable investors and other stakeholders to more adeptly and resiliently negotiate the intricacies of the financial markets.

1.4 Objectives

- Create and put into use a stock price forecasting predictive model that makes use of long short-term memory (LSTM) and recurrent neural networks (RNN).
- Examine how well RNNs and LSTMs can capture sequential dependencies in time-series financial data to improve stock price prediction accuracy.
- To measure gains in predictive power, compare the suggested model's performance to that of conventional forecasting techniques.
- Examine how well the RNN and LSTM-based models can withstand the innate volatility and unpredictability of stock prices.
- Examine the possibility that RNNs and LSTMs could offer investors timely insights into developing market trends so they can make wise decisions in volatile financial markets.
- Contribute to the field of financial analytics' evolution by expanding knowledge about how to combine sophisticated machine learning methods with stock market analysis.

CHAPTER 2

LITERATURE SURVEY

The field of stock price prediction has undergone a significant transformation, shifting from traditional statistical and econometric models to the exploration of deep learning methodologies. Early attempts, such as autoregressive integrated moving average (ARIMA) and linear regression, fell short in capturing the intricate and non-linear dynamics of financial markets. With the advent of deep learning, particularly the utilisation of recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), researchers have found more effective tools for modelling the complex relationships within financial time series data. This paradigm shift has sparked a surge of interest, prompting a reevaluation of predictive modelling approaches in the financial domain..

The success of stock price prediction using deep learning hinges on meticulous data preprocessing and feature engineering strategies. Financial time series data, often characterised by noise and irregularities, necessitate techniques like data normalisation, time-series decomposition, and feature scaling. The aim is to empower models to extract meaningful patterns from the intricate web of financial data, thereby enhancing predictive accuracy and robustness in forecasting stock prices.

Recent strides in stock price prediction using deep learning extend beyond individual model architectures. Ensemble methods, encompassing techniques like stacking and bagging, have gained traction for their ability to amalgamate the strengths of multiple models, enhancing overall robustness and generalisation. This technique involves adapting pre-trained models from related tasks to the domain of stock price prediction, leveraging knowledge from diverse sources. These recent trends underscore the importance of combining models and harnessing pre-existing knowledge to address challenges like uncertainty and data sparsity, contributing to the dynamic landscape of deep learning applications in stock price prediction. As researchers delve into these novel approaches, they contribute to the evolving understanding and refinement of deep learning methodologies in financial forecasting.

1. Short-term stock market price trend prediction using a comprehensive deep learning system [1]:

    - Researchers collected data from the Chinese stock market for two years and presented a thorough customisation of feature engineering and deep learning-based models for predicting stock market price trends. The suggested approach is comprehensive in that it covers stock market dataset pre-processing, the use of numerous feature engineering techniques, and a customised deep learning-based system for stock market price trend prediction.
    - The study involved the collection of two years' worth of data from the Chinese stock market, focusing on devising a comprehensive customization of feature engineering and a deep learning-based model aimed at predicting the price trend of stock markets.
    - The proposed solution was structured into three distinct parts. Firstly, emphasis was placed on feature selection to ensure that the chosen features were highly effective. Following this, a thorough examination of the data was conducted, incorporating dimensionality reduction techniques. The final and primary contribution of the research involved the construction of a prediction model specifically tailored for target stocks.
    - In evaluating the effectiveness of the proposed LSTM model, notable results were achieved, with a binary accuracy rate of 93.25%. This high precision level in predicting the bi-weekly price trend underscored the efficacy of the model.
    - The preprocessing of data involved the application of Principal Component Analysis (PCA), resulting in the extraction of five principal components. Subsequently, the model underwent training for 150 epochs to further refine its predictive capabilities. These findings collectively highlight the success and robustness of the proposed methodology in predicting stock price trends in the Chinese stock market.

2. Stock Market Prediction Using Machine Learning [2] :

- This paper aimed to enhance the accuracy and reliability of predicting future stock prices through the application of machine learning techniques. The researchers' primary contribution lies in the utilisation of the novel Long Short-Term Memory (LSTM) Model as a method for determining stock prices.
- The dataset employed for analysis was sourced from Yahoo Finance, comprising approximately 900,000 records of the necessary stock prices and other relevant values. To implement the predictive model, a sequential architecture was designed, involving the stacking of two LSTM layers on top of each other with an output value of 256
- In the model configuration, a dropout value of 0.3 was set, implying that 30% of the total nodes would be frozen during the training process to mitigate the risk of overfitting. While acknowledging the myriad machine learning models available, the paper focused on two crucial ones: the Regression Based Model and the LSTM Based Model. The results revealed the LSTM model's superior efficiency in making predictions, highlighting its effectiveness in the context of stock price forecasting.
- Additionally, the research delved into the specifics of the LSTM model architecture and its parameters. By stacking two LSTM layers, the model demonstrated a capacity to capture and learn intricate patterns within the sequential stock price data. The choice of an output value of 256 contributed to the model's ability to extract and represent complex features, enhancing its predictive capabilities.
- The incorporation of a dropout value of 0.3 in the model configuration played a pivotal role in preventing overfitting during the training process, ensuring that the LSTM model maintained its generalisation ability when applied to new, unseen data. This detailed examination of the model's architecture and parameter tuning adds depth to the understanding of how the LSTM-based approach contributes to the increased accuracy and reliability observed in predicting future stock prices.

3. Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models [3]:

    - This paper provides a comprehensive exploration of stock price prediction through an extensive analysis involving a diverse array of models, covering time series, econometric, and various learning-based approaches. The study leverages datasets encompassing the time period from January 2004 to December 2019, sourced from Infosys, ICICI, and SUN PHARMA, with the overarching goal of identifying the most effective model within specific sectors.
    - The considered models span a spectrum of methodologies, encompassing Holt-Winters Exponential Smoothing for time series, ARIMA for econometrics, Random Forest and Multivariate Adaptive Regression Splines (MARS) for machine learning, and simple Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) for deep learning. By employing these diverse models, the research aims to provide a comprehensive understanding of their relative strengths and weaknesses in the context of stock price prediction.
    - Among the machine learning models, Multivariate Adaptive Regression Splines (MARS) emerges as the standout performer, demonstrating a notable edge in predictive capabilities. Simultaneously, within the domain of deep learning, Long Short-Term Memory (LSTM) proves to be the most effective for accurate stock price prediction. Importantly, across three distinct sectors—IT, Banking, and Health—MARS consistently asserts itself as the premier model for sales forecasting. This overarching observation not only highlights the adaptability and robustness of the MARS model but also underscores its efficacy in providing accurate stock price predictions across a spectrum of diverse sectors.
    - In conclusion, the findings of this study contribute valuable insights for practitioners and researchers engaged in stock market analysis. The versatility of the MARS model and the effectiveness of LSTM in specific contexts offer a nuanced understanding of the interplay between different prediction techniques, guiding future research and practical applications in the dynamic field of stock price prediction.

4. A Survey on Stock Market Prediction UsingMachine Learning Techniques [4]:

   - This paper conducts an extensive review and comparative analysis of diverse techniques employed in the prediction of stock market trends and performance. The objective is to critically evaluate the efficacy of various prediction parameters within the framework of stock market analysis, shedding light on their applicability and effectiveness in capturing the complexities of financial markets.
   - Specifically, the study focuses on the examination of the latest prediction techniques employed in the stock market domain. The techniques under consideration include Artificial Neural Network, Neuro-Fuzzy System, Time Series Linear Models (TSLM), and Recurrent Neural Network (RNN). Each of these methods is scrutinised with a keen eye on their distinct advantages and disadvantages in the context of predicting stock market behaviour.
   - Artificial Neural Network, known for its ability to model complex relationships, is assessed for its strengths and potential shortcomings in capturing the intricate patterns inherent in stock market data. The Neuro-Fuzzy System, which combines fuzzy logic and neural network principles, is explored to understand its adaptability to the dynamic nature of financial markets. Time Series Linear Models (TSLM) are evaluated for their efficacy in modelling the time-dependent nature of stock prices. Lastly, the study delves into the capabilities and limitations of Recurrent Neural Network (RNN), known for its capacity to consider sequential dependencies in data.

In conclusion, this literature survey has comprehensively examined various stock price prediction techniques, ranging from traditional methods to state-of-the-art machine learning and deep learning models. The comparative analysis of recent approaches, such as Artificial Neural Network, Neuro-Fuzzy System, Time Series Linear Models (TSLM), and Recurrent Neural Network (RNN), has illuminated their distinct strengths and limitations. Moving forward, a holistic understanding of these techniques and their contextual applicability is crucial, offering researchers and practitioners valuable insights for designing effective and adaptive stock price prediction models in the ever-evolving landscape of financial markets.

CHAPTER 3

METHODOLOGY

3.1 Proposed Methodology

Stock price prediction and forecasting involve analysing historical market data to forecast future price movements in stocks. Prediction is concerned with predicting the future value or direction of a stock's price by analysing patterns and trends in data such as past prices, volumes, and external indicators. Forecasting offers a broader perspective by providing estimates, probabilities, or ranges for future price movements while accounting for uncertainties and risks associated with predictions. Both serve as tools for investors and analysts, but it is critical to recognise the inherent complexities of financial markets and use these forecasts as decision-making aids rather than definitive future indicators.

Modern techniques for stock price prediction, such as machine learning algorithms and deep learning models like RNNs and LSTMs, outperform traditional methods by handling complex relationships and capturing non-linear patterns inherent in financial data. They offer improved accuracy, scalability to handle large datasets, and adaptive learning capabilities, continuously updating predictions with new information. Moreover, modern techniques automate feature extraction and reduce assumptions about data distribution, making them more flexible and robust to noisy or irregular data, essential traits in the dynamic landscape of financial markets.

Despite their strengths, these techniques might demand substantial computational resources, extensive data for training, and could be prone to overfitting. Interpreting their results might also pose challenges due to their inherent complexity. While modern methods offer significant advantages in predicting stock prices, their successful application often involves a balanced integration of various techniques and expert insights, acknowledging the uncertainties and risks inherent in financial forecasting.

In Figure 3.1, the system architecture is depicted for constructing a stock price prediction and forecasting model based on Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) mechanisms.

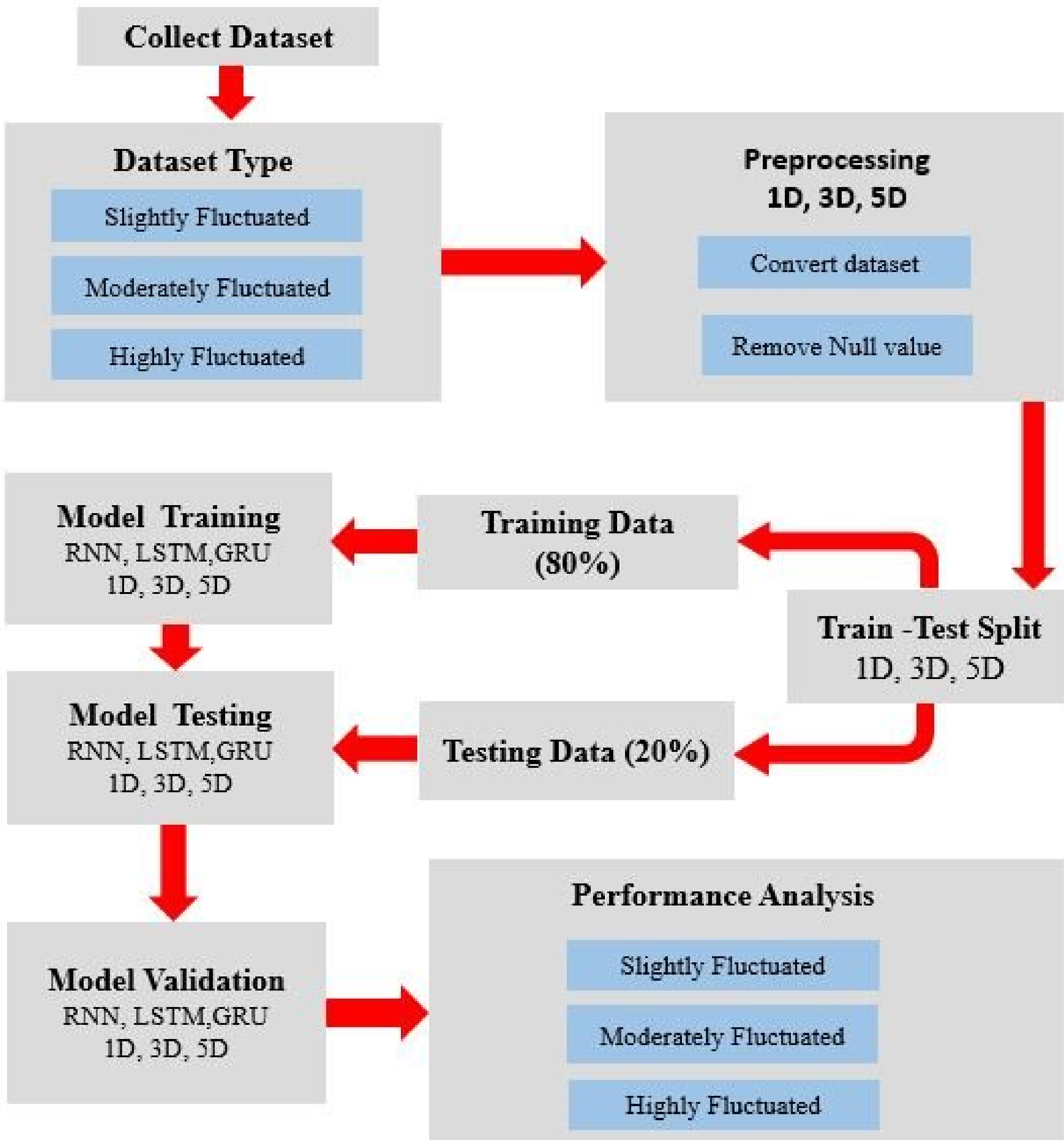

Fig. 3.1 System Architecture

3.2 Modelling Approach

A detailed approach for building a RNN and LSTM based stock price prediction and forecasting model is discussed below.

1. **Data Gathering**: Obtained a large historical stock price data from a financial database called Yahoo Finance.

2. **Data Cleaning:** Handle missing values, adjust for stock splits, and ensure consistency in the time series data.

3. **Feature Selection:** Determine relevant features such as open, high, low, close prices, volume, technical indicators, sentiment analysis data, etc.

4. **Scaling**: Scale the features to a range suitable for neural networks, typically between 0 and 1 or using techniques like Min-Max scaling or Standardization.

5. **Train-Test Split:** Divide the data into training and testing sets. A common split is 70-30 or 80-20 for training and testing respectively.

6. **Model Configuration:** Design the RNN or LSTM architecture. Define the number of layers, number of neurons in each layer, activation functions (like ReLU, sigmoid, or tanh), dropout rate (to prevent overfitting), etc.

7. **Compile the Model:** Specify the loss function (such as mean squared error for regression problems), optimizer (like Adam or RMSprop), and evaluation metrics.

8. **Training Process:** Feed the training data into the model and train it using backpropagation and gradient descent. Adjust hyperparameters iteratively to optimise performance.

9. **Validation:** Monitor the model's performance on the validation set to prevent overfitting. Use early stopping to halt training when the model stops improving.

10. **Model Evaluation:** Assess the model's performance on the unseen test data using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), etc.

11. **Future Prediction:** Utilise the trained model to forecast future stock prices beyond the available dataset by feeding the model with recent historical data.

12. **Deployment**: Once the model has demonstrated satisfactory performance, it can be deployed in real-world scenarios. This may involve integrating the model into a web or mobile application, allowing users to choose company name, years of prediction and get predicted data.

CHAPTER 4

IMPLEMENTATION

4.1 Technology Stack: Table 4.1 represents the technology stack used in the development.

Technology	Function	Logo
Python	For Deep learning model development and to build the backend	
Google Colaboratory	To run and execute deep learning models using Python	
Streamlit	To build an interactive user interface.	
FastAPI	To build APIs with Python	
Google Cloud Platform	For efficient and fast model training, and for hosting web application	

Table 4.1 Technology Stack

4.2 Code Structure:

- **Dataset:** Contains yfinance dataset downloaded from yahoo.com
- **Deployment :** Contains complete Deployment of Streamlit application.
- **Documents :** Contains documentations.
- **SavedModels:** Stores model versions with various epochs and model architecture

Actual code structure from codebase is shown in Figure 4.1

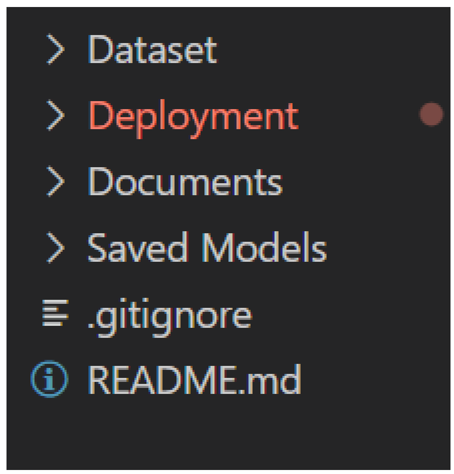

Fig. 4.1 Code Structure

4.3 Libraries Used:

- **Tensorflow**: TensorFlow is a powerful open-source library for machine learning and deep learning tasks. It provides a flexible and efficient framework for developing and deploying neural networks.

- **Keras**: Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.

- **Numpy**: Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Moreover Numpy forms the foundation of the Machine Learning stack.

- **Image**: PIL (Python Imaging Library) is an open-source library for image processing tasks that requires python programming language. PIL can perform tasks on an image such as reading, rescaling, saving in different image formats. PIL can be used for Image archives, Image processing, Image display.

- **Matplotlib**: Matplotlib is a cross-platform, data visualisation and graphical plotting library (histograms, scatter plots, bar charts, etc) for Python and its numerical extension NumPy. Here it is used for representing Accuracy vs Loss graph.

- **Shutil**: The Shutil module allows you to do high-level operations on a file, such as copy, create, and remote operations. It falls within the umbrella of Python's basic utility modules. This module aids in the automation of the copying and deleting of files and folders. It is used for setting up dataset structure

- **Os**: Python OS module provides the facility to establish the interaction between the user and the operating system. It offers many useful OS functions that are used to perform OS-based tasks and get related information about operating systems. Here OS was used to get the folder's path from the system.

- **Time**: The Python time module provides many ways of representing time in code, such as objects, numbers, and strings. It also provides functionality other than representing time, like waiting during code execution and measuring the efficiency of your code. Here we calculated the training time using this library.

- **Google-Cloud**: The Cloud Client is used to access Google Cloud APIs programmatically. It was used to access data from cloud buckets and deploy it on cloud functions.

4.4 Implementation:

- **Data Collection:** We trained our model on Yfinance Dataset. The "yahoo_finance_dataset(2018-2023)" dataset is a financial dataset containing daily stock market data for multiple assets such as equities, ETFs, and indexes. It spans from April 1, 2018 to March 31, 2023, and contains 1692 rows and 7 columns. The data was sourced from Yahoo Finance, and the purpose of the dataset is to provide researchers, analysts, and investors with a comprehensive dataset that they can use to analyse stock market trends, identify patterns, and develop investment strategies. Figure 4.2 shows a snapshot of the yahoo_finance Tesla dataset, including Date field.

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Volume	Adj Close
2	6/29/2010	19	25	17.540001	23.889999	18766300	23.889999
3	6/30/2010	25.790001	30.42	23.299999	23.83	17187100	23.83
4	07-01-10	25	25.92	20.27	21.959999	8218800	21.959999
5	07-02-10	23	23.1	18.709999	19.200001	5139800	19.200001
6	07-06-10	20	20	15.83	16.110001	6866900	16.110001
7	07-07-10	16.4	16.629999	14.98	15.8	6921700	15.8
8	07-08-10	16.139999	17.52	15.57	17.459999	7711400	17.459999
9	07-09-10	17.58	17.9	16.549999	17.4	4050600	17.4
10	07-12-10	17.950001	18.07	17	17.049999	2202500	17.049999
11	7/13/2010	17.389999	18.639999	16.9	18.139999	2680100	18.139999
12	7/14/2010	17.940001	20.15	17.76	19.84	4195200	19.84
13	7/15/2010	19.940001	21.5	19	19.889999	3739800	19.889999
14	7/16/2010	20.700001	21.299999	20.049999	20.639999	2621300	20.639999
15	7/19/2010	21.370001	22.25	20.92	21.91	2486500	21.91
16	7/20/2010	21.85	21.85	20.049999	20.299999	1825300	20.299999
17	7/21/2010	20.66	20.9	19.5	20.219999	1252500	20.219999

Tesla.csv - Tesla.csv

Ready Accessibility: Unavailable

Fig. 4.2 Data Collection

- **Google Cloud GPU Server Connection:** Google Cloud GPU processors are used for enhancing model training power and efficiency. Figure 4.3 shows a Google Collab Notebook which is connected with Cloud Engine - Virtual Machine instance.

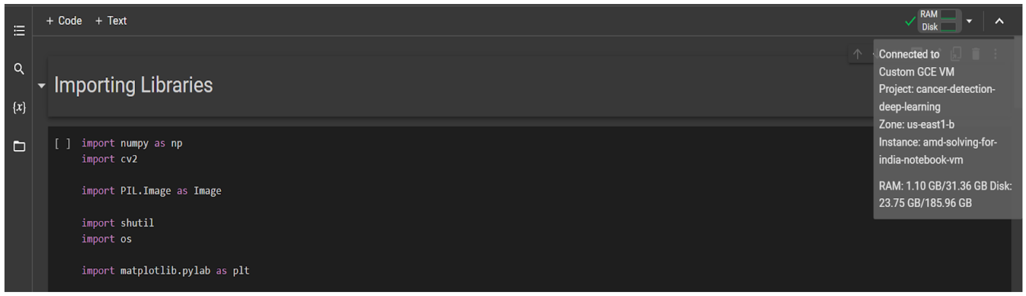

Fig. 4.3 GPU Server Connection

- **Data Visualization:** Dataset is visualised and following is the information from dataset, 7 columns of data with 1692 entries of different stocks. The column includes Date, Opening price (Open), Highest price (High), Lowest price (Low), Closing Price (Close), Volume, Adj Close.

  Figure 4.4 represents the top 5 instances from dataset and the column wise information.

```
data.head()
```

	Date	Open	High	Low	Close	Volume	Adj Close
0	6/29/2010	19.000000	25.00	17.540001	23.889999	18766300	23.889999
1	6/30/2010	25.790001	30.42	23.299999	23.830000	17187100	23.830000
2	7/1/2010	25.000000	25.92	20.270000	21.959999	8218800	21.959999
3	7/2/2010	23.000000	23.10	18.709999	19.200001	5139800	19.200001
4	7/6/2010	20.000000	20.00	15.830000	16.110001	6866900	16.110001

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1692 entries, 0 to 1691
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       1692 non-null   object
 1   Open       1692 non-null   float64
 2   High       1692 non-null   float64
 3   Low        1692 non-null   float64
 4   Close      1692 non-null   float64
 5   Volume     1692 non-null   int64
 6   Adj Close  1692 non-null   float64
dtypes: float64(5), int64(1), object(1)
memory usage: 92.7+ KB
```

Fig. 4.4 Data Visualization

- **RNN Model:** Core part of any Deep learning model is its layers. This model has 5 combined layers of Simple RNN and Dropout layers and at last a Dense Output layer for final prediction.

  Figure 4.5 shows the actual implementation of sequential RNN layers using keras.

```
# initializing the RNN
regressor = Sequential()

# adding first RNN layer and dropout regulatization
regressor.add(
    SimpleRNN(units = 50,
              activation = "tanh",
              return_sequences = True,
              input_shape = (X_train.shape[1],1))
             )

regressor.add(
    Dropout(0.2)
             )


# adding second RNN layer and dropout regulatization

regressor.add(
    SimpleRNN(units = 50,
              activation = "tanh",
              return_sequences = True)
             )

regressor.add(
    Dropout(0.2)
             )

# adding third RNN layer and dropout regulatization

regressor.add(
    SimpleRNN(units = 50,
              activation = "tanh",
              return_sequences = True)
             )

regressor.add(
    Dropout(0.2)
             )

# adding fourth RNN layer and dropout regulatization

regressor.add(
    SimpleRNN(units = 50)
             )

regressor.add(
    Dropout(0.2)
             )

# adding the output layer
regressor.add(Dense(units = 1))
```

Fig. 4.5 RNN Model

- **LSTM Model for Forecasting:** LSTM consists of 3 layers of LSTM and a final Dense layer for forecasting stock data.

  Figure 4.6 shows the actual implementation of a stacked LSTM model using keras.

Creating the Stacked LSTM model

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM


model=Sequential()
model.add(LSTM(50,return_sequences=True,input_shape=(100,1)))
model.add(LSTM(50,return_sequences=True))
model.add(LSTM(50))
model.add(Dense(1))
model.compile(loss='mean_squared_error',optimizer='adam')
```

Fig. 4.6 Stacked LSTM Model

- **Model Compilation:** Model is trained using Adam Optimizer and Cross Entropy as loss function. Training was performed for 100 epochs in a batch of 32.

  Figure 4.7 shows a snapshot from notebook to represent model hyperparameters.

```
model.compile(
    optimizer='adam',
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=False),
    metrics=['accuracy']
)


history = model.fit(
    train_ds,
    batch_size = BATCH_SIZE,
    validation_data = val_ds,
    verbose = 1,
    epochs = EPOCHS,
)
```

Fig. 4.7 Model Compilation

- **Model Evaluation on training data (RNN):** Model is evaluated on Validation data splitted before and is scaled back from 0-1 values to its original values. Figure 4.8 shows the code structure of the section where the model is evaluated.

```
# predictions with X_test data
y_pred_of_test = regressor.predict(X_test)
# scaling back from 0-1 to original
y_pred_of_test = scaler.inverse_transform(y_pred_of_test)
print("Shape of y_pred_of_test :",y_pred_of_test.shape)
```

```
9/9 [==============================] - 0s 45ms/step
Shape of y_pred_of_test : (288, 1)
```

Fig. 4.8 Model Evaluation on training data

- **UI Development:** Interactive User Interface is created using Streamlit in python. Various components are Navbar, HomePage, About pages are added in the UI to make it more user friendly.

- **Backend Development:** Backend is served using a python based framework FAST-API. Saved model is loaded from the cloud bucket, image is pre-processed and then the predict function is used for class prediction on the requested image. Following figure 4.9 is the code snippet of the backend functionality implemented for development.

```
global model
if model is None:
    download_blob(
        BUCKET_NAME,
        "models/final.h5",
        "/tmp/final.h5",
    )
    model = load_model("/tmp/final.h5")
```

Fig. 4.9 Backend Development

- **Testing:** Manual Testings were done which includes testing functions, modules, and specific features in isolation. Tools like pytest can be helpful for creating and running unit tests.

4.5 User Interface:

- Home Page: In Figure 4.10, the depicted interface corresponds to the Home Page of the suggested application.

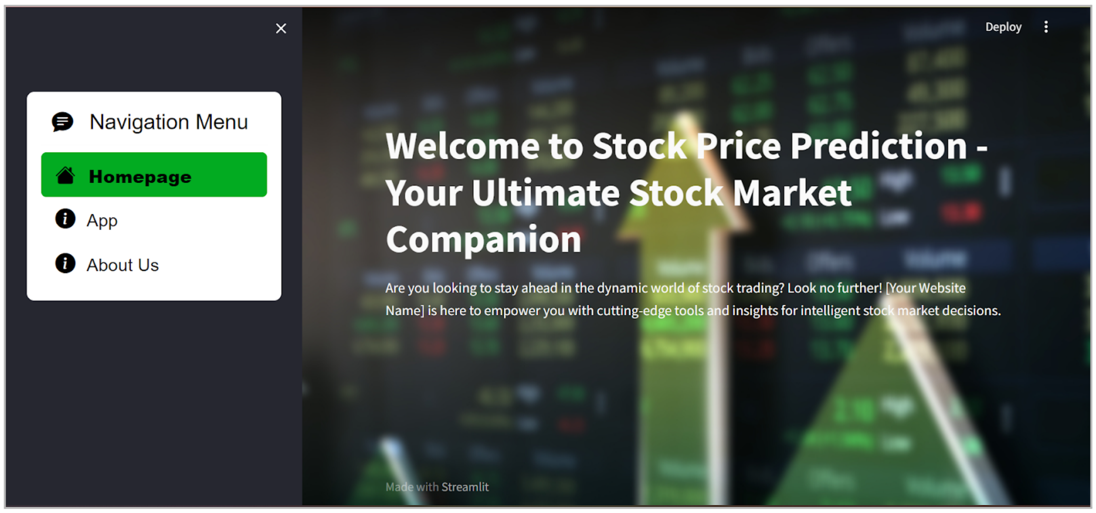

Fig. 4.10 HomePage

- About page: Figure 4.11 represents the About page of the application, presenting essential information about the application's purpose and features.

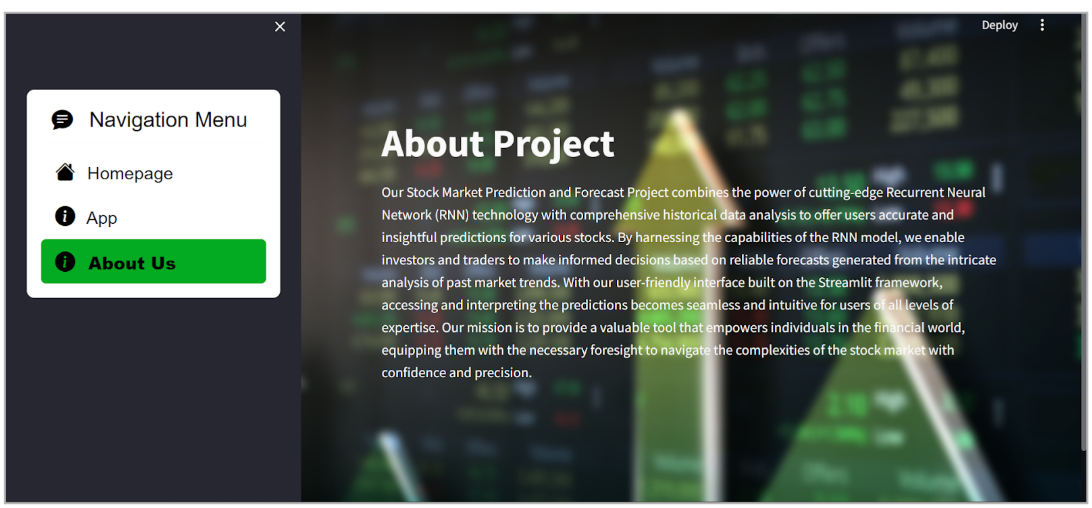

Fig. 4.11 About us

- Selecting Stocks to predict: In the following figure 4.12, users can select the stock to analyse.

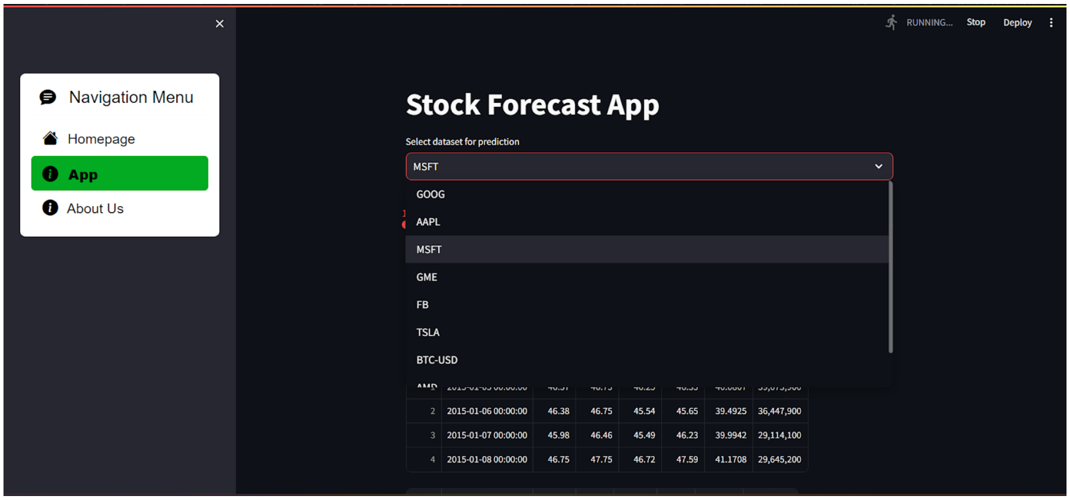

Fig. 4.12 Stock Selection Section

- Selecting years of prediction: Figure 4.13 depicts the section within the application where users can select the years for prediction.

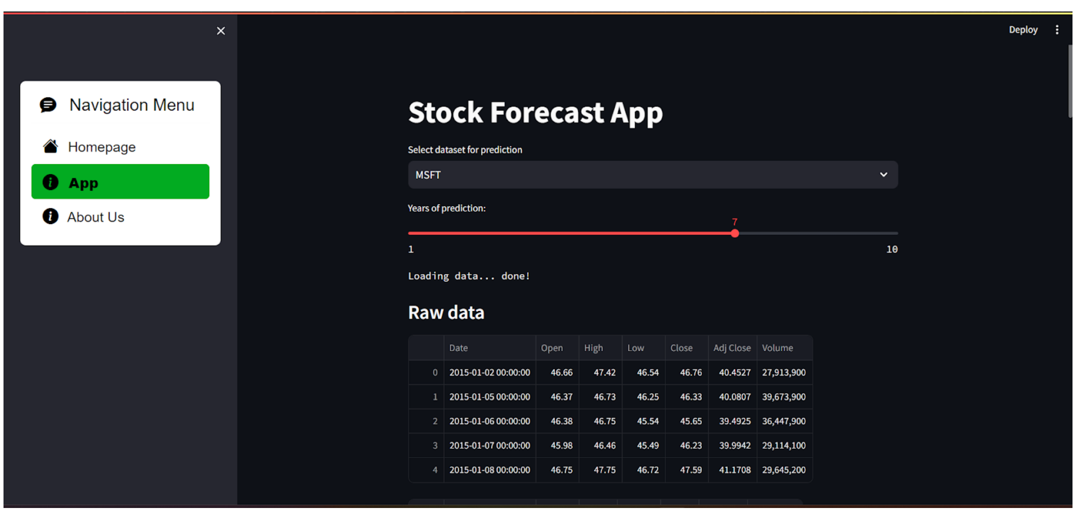

Fig. 4.13 Selecting years of Prediction

- Raw Data visualisation: Following Figure 4.15, is the section displaying raw data of the selected stock for selected years as shown.

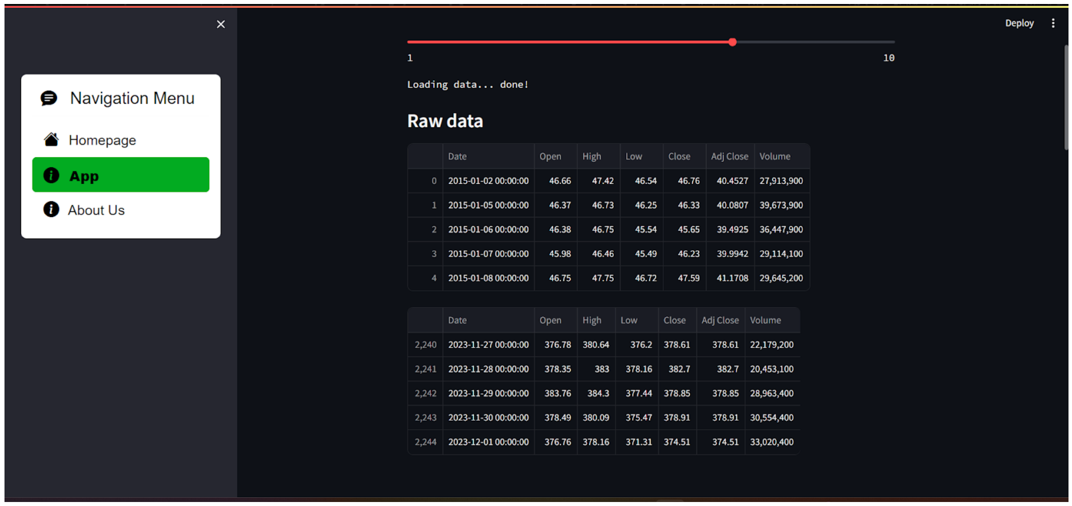

Fig. 4.14 Raw data Visualisation

- Time Series Data Section: In Figure 4.15, the section displaying time series data is accompanied by a range slider, as illustrated.

Fig. 4.15 Time series data with range slider

- Forecasting Section: Figure 4.16 is a visual representation of the forecasting section of the application which forecasts the stock data of selected years.

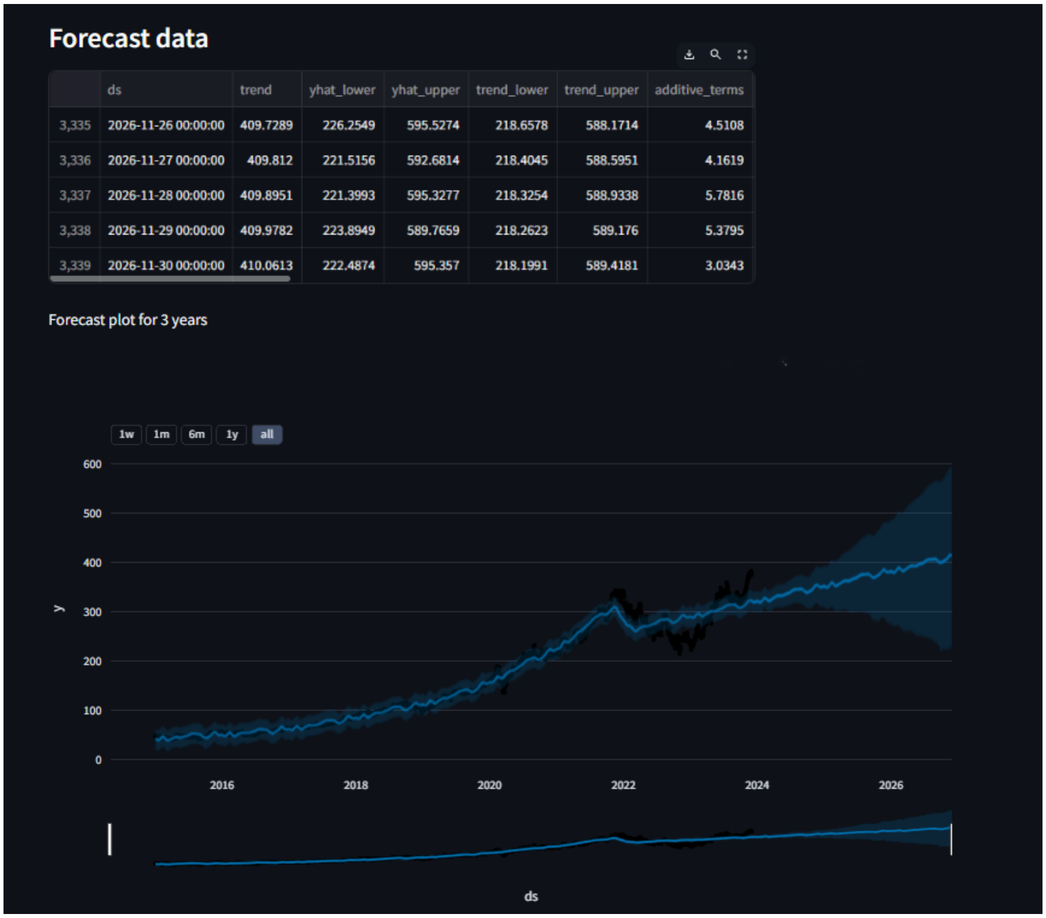

Fig. 4.16 Forecasting Page

CHAPTER 5

RESULTS AND EVALUATION

5.1 Results and Deployments:

- **Web-App Link :** https://indian-stocks.streamlit.app/

5.2 Model Analysis:

- **Graphical Analysis:** Figure 5.1 visually represents the graphical analysis of predictions on both Training and Validation data.

Fig. 5.1 Train-Validation Prediction

- **Prediction for 30 days:** In Figure 5.2, the graph illustrates the forecasting for the next 30 days.

[<matplotlib.lines.Line2D at 0x782018fc6170>]

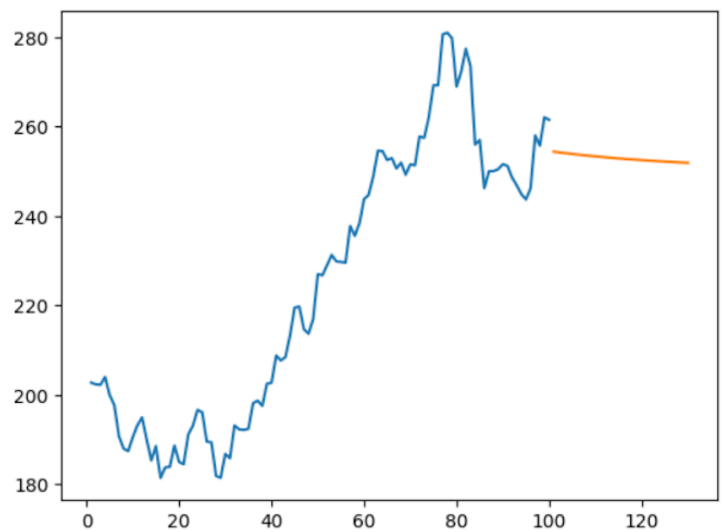

Fig. 5.2 Next 30 days Forecasting

- **Forecasting for 5 years:** Figure 5.3 presents a graphical representation of the 5-year forecasted data for the stock selected by users.

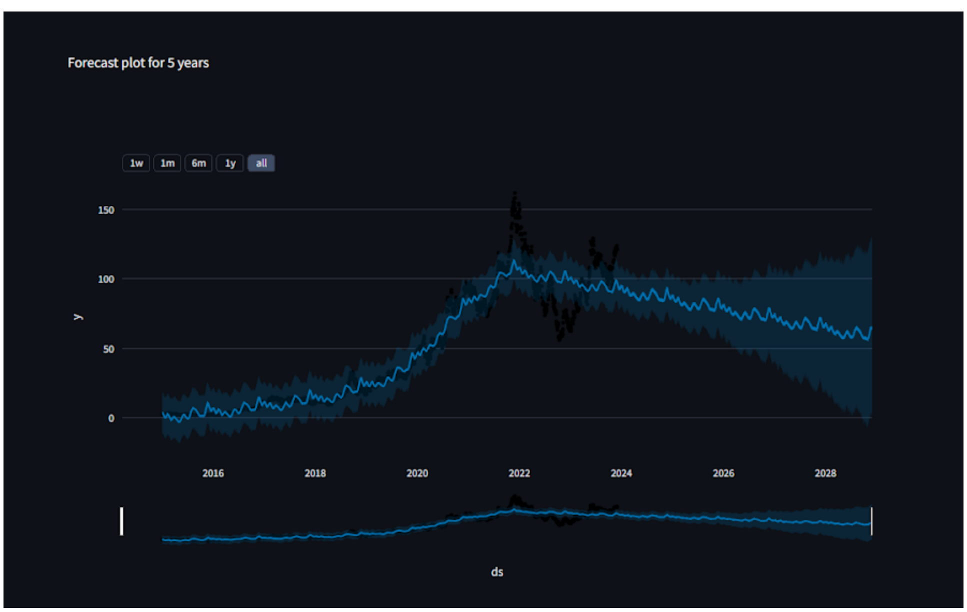

Fig. 5.3 5 years forecasting for user selected stock

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

Conclusion:

In conclusion, the project successfully leveraged advanced machine learning models, specifically LSTM and RNN, for stock price prediction using the yfinance dataset. The integration of these models into a user-friendly application built with Streamlit provides a practical and accessible platform for users interested in exploring stock market trends and making informed decisions.

The LSTM and RNN models demonstrated their capability to capture complex temporal dependencies in the financial data, allowing for more accurate predictions of stock prices. The application's interface, designed with Streamlit, provides an intuitive experience for users to interact with and visualise the model predictions.

As part of future improvements, there are various avenues to explore, including feature engineering, hyperparameter tuning, ensemble methods, real-time data integration, and backtesting functionalities. These enhancements aim to refine the application's accuracy, usability, and practicality, positioning it as a valuable tool for both novice and experienced investors.

By deploying the application on cloud platforms, the project would ensure scalability and accessibility, allowing users to benefit from its capabilities without the constraints of local installations. Furthermore, potential collaboration and social features could foster a community around the application, encouraging knowledge sharing and strategy discussions among users.

In summary, the project marks a significant stride in the domain of stock price prediction, combining sophisticated machine learning models with a user-centric application. The continuous pursuit of innovation and responsiveness to user needs will contribute to the project's longevity and relevance in the dynamic landscape of financial technology.

Future Scope:

The successful development of our Stock price prediction and forecasting application using machine learning opens up promising avenues for future advancements and enhancements. Here are some potential areas of future scope for this application:

1. Feature Engineering: Explore additional financial indicators and economic factors to enhance the input features for your models. Features such as moving averages, relative strength index (RSI), and other technical indicators can be valuable. Consider sentiment analysis on financial news or social media data to incorporate market sentiment into your models.
2. Ensemble Methods: Combine multiple models to create an ensemble for improved prediction accuracy. You can experiment with combining RNN and LSTM models or even incorporate traditional machine learning models.
3. Hyperparameter Tuning: Optimise the hyperparameters of your models to achieve better performance. Techniques such as grid search or Bayesian optimization can be employed for this purpose.
4. Real-Time Data Integration: Enhance your application to handle real-time data updates. This could involve implementing a mechanism to fetch and process streaming data, allowing users to make more timely decisions.
5. Deployment on Cloud Platforms: Deploy your application on cloud platforms like AWS, Google Cloud, or Microsoft Azure for scalability and accessibility. This would also allow users to access the application from anywhere without the need for local installations.
6. Risk Management Features: Integrate risk management tools and features into your application. This could involve incorporating stop-loss mechanisms, portfolio optimization, or risk assessment tools to help users manage their investments more effectively.

REFERENCES

1. Shen, J., Shafiq, M.O. Short-term stock market price trend prediction using a comprehensive deep learning system. J Big Data 7, 66 (2020). https://doi.org/10.1186/s40537-020-00333-6

2. I.Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 574-576, doi: 10.1109/ICSCCC.2018.8703332.

3. A. Chatterjee, H. Bhowmick and J. Sen, "Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models," 2021 IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, 2021, pp. 289-296, doi: 10.1109/MysuruCon52639.2021.9641610.

4. Polamuri, Dr & Srinivas, Kudipudi & Mohan, A.. (2020). A Survey on Stock Market Prediction Using Machine Learning Techniques. 10.1007/978-981-15-1420-3_101.

5. Sen, Jaydip, and Tamal Datta Chaudhuri. "Decomposition of time series data of stock markets and its implications for prediction: an application for the Indian auto sector." arXiv preprint arXiv:1601.02407 (2016).

6. Sen, Jaydip, and Tamal Datta Chaudhuri. "A time series analysis-based forecasting framework for the Indian healthcare sector." arXiv preprint arXiv:1705.01144 (2017).

7. Ullah, Kifayat, and Muhammad Qasim. "Google stock prices prediction using deep learning." 2020 IEEE 10th international conference on system engineering and technology (ICSET). IEEE, 2020.

8. Mehtab, Sidra, and Jaydip Sen. "Stock price prediction using convolutional neural networks on a multivariate timeseries." arXiv preprint arXiv:2001.09769 (2020).

9. Shen, Jingyi, and M. Omair Shafiq. "Short-term stock market price trend prediction using a comprehensive deep learning system." Journal of Big Data 7.1 (2020): 1-33.

10. Dos Santos Pinheiro, Leonardo, and Mark Dras. "Stock market prediction with deep learning: A character-based neural language model for event-based trading." Proceedings of the Australasian Language Technology Association Workshop 2017. 2017.