A Comparative Study of LSTM and ARMA in Predicting Stock Trends: The Case of AMZN, META, TSLA and JPM

A PROJECT REPORT

Submitted by

GOKULPRIYAN KARTIKEYAN-[RA2211003020327]
DHARANI PAJANI-[RA2211003020340]
KISHORE P-[RA2211003020344]
DHANARAJAN K-[RA2211003020347]

Under the guidance of

Ms.Harini B M.E.,

(Designation, Department of Computer Science and Engineering)

in partial fulfilment for the award of the

degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM, CHENNAI -600089

OCTOBER 2024

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University U/S 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report titled **"A Comparative Study of LSTM and ARMA in Predicting Stock Trends: The Case of AMZN, META, TSLA and JPM"** is the bonafide work of **GOKULPRIYAN K-[RA2211003020327],DHARANI PAJANI-[RA2211003020340],KISHORE P-[RA2211003020344],DHANARAJAN K-[RA2211003020347]** who carried out the project work under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an occasion on this or any other candidate.

SIGNATURE

Ms. HARINI B, M.E.,
Assistant Professor
Computer Science and Engineering,
SRM Institute of Science and Technology,
Ramapuram, Chennai.

SIGNATURE

Dr. K. RAJA, M.E., Ph.D.,
Professor and Head
Computer Science and Engineering,
SRM Institute of Science and Technology,
Ramapuram, Chennai.

Submitted for the project viva-voce held on___________at SRM Institute of Science and Technology, Ramapuram, Chennai -600089.

INTERNAL EXAMINER1 **EXTERNAL EXAMINER2**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

RAMAPURAM, CHENNAI - 89

DECLARATION

We hereby declare that the entire work contained in this project report titled "**A Comparative Study of LSTM and ARMA in Predicting Stock Trends: The Case of AMZN, META, TSLA and JPM**" has been carried out by **GOKULPRIYAN KARTIKEYAN-[RA2211003020327], DHARANI PAJANI[RA2211003020340],KISHORE P-[RA2211003020344],DHANARAJAN K-[RA2211003020347]** at SRM Institute of Science and Technology, Ramapuram Campus, Chennai- 600089, under the guidance of **Ms. HARINI B, M.E., Assistant Professor**, Department of Computer Science and Engineering.

Place: Chennai

Date:

GOKULPRIYAN KARTIKEYAN

DHARANI PAJANI

KISHORE P

DHANARAJAN K

ABSTRACT

The stock market is a complex, dynamic system influenced by various factors, including economic indicators, market sentiment, and global events. Predicting stock prices with high accuracy is a challenging task due to the market's inherent volatility and non-linear nature. This project aims to develop a machine learning-based model to predict stock market trends and prices. By leveraging historical stock data, financial indicators, and sentiment analysis, the model employs advanced machine learning techniques such as time series analysis, neural networks, and ensemble methods. The results demonstrate that while no model can perfectly predict stock prices, machine learning techniques can provide valuable insights and improve decision-making for investors. This project contributes to the growing field of financial technology by highlighting the potential of machine learning. This project explores the application of machine learning techniques to predict stock market trends by leveraging historical data and various financial indicators. Machine learning algorithms, such as Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and Random Forests, are employed to model and forecast stock prices. The study integrates feature engineering to enhance predictive accuracy, considering factors like market sentiment, trading volume, and macroeconomic variables. By comparing different models, the research highlights the importance of selecting appropriate features and algorithms to improve prediction performance. The results demonstrate that machine learning methods can offer significant insights and contribute to the development of more sophisticated financial models. This project aims to bridge the gap between theoretical financial models and practical market applications, providing a data-driven approach to stock market forecasting.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Introduction

The stock market, known for its volatile and unpredictable nature, has long attracted researchers and investors aiming to predict its movements. With the advent of machine learning (ML), new opportunities have emerged to analyse vast amounts of financial data and identify patterns that are not discernible through traditional methods. Stock market prediction using machine learning involves the application of various algorithms to forecast future stock prices or trends based on historical data. This project aims to explore the potential of machine learning models, such as neural networks, support vector machines, and ensemble methods, to enhance the accuracy of stock market predictions.

Recent research has demonstrated that machine learning models can outperform classical statistical methods by capturing nonlinear relationships and complex interactions within financial data. For instance, a study by Patel et al. (2015) highlighted the superior performance of ML algorithms like random forests and support vector machines over traditional approaches in predicting stock prices. Another study by Fischer and Krauss (2018) demonstrated the effectiveness of deep learning techniques, particularly long short-term memory (LSTM) networks, in predicting stock market trends with a higher degree of accuracy. This project will build on these insights by implementing and comparing various machine learning models, using historical stock market data. The goal is to identify the most effective approach for accurate and reliable stock market prediction, contributing to the growing body of research in this domain.

1.2 Problem Statement

This project seeks to address the challenge of predicting stock market prices by leveraging machine learning techniques. The problem is to develop a model that can analyze historical stock data, including price movements, trading volumes, and relevant financial indicators, to forecast future stock prices with high accuracy. The model must be robust enough to handle the market's inherent volatility and adaptable to changes in market conditions. Additionally, the model should be user-friendly, providing real-time predictions and insights that can assist investors in making data-driven decisions.

1.3 Objective of the Project

The primary objective of this machine learning project is to develop an accurate and reliable model for predicting stock market prices and trends. This involves collecting and preprocessing historical stock data, exploring and optimizing various machine learning algorithms, and rigorously evaluating the model's performance using key metrics. The ultimate goal is to create a robust predictive tool that can provide actionable insights for investors. The project also aims to implement a real-time prediction system with a user-friendly interface or API, ensuring continuous improvement by updating the model with new data to adapt to changing market conditions. Through these efforts, the project seeks to showcase the potential of machine learning in enhancing stock market prediction accuracy and supporting more informed investment decisions.

1.4 Project Domain

The project domain lies at the intersection of Financial Data Analysis, Machine Learning, and Time-Series Forecasting. This domain focuses on leveraging machine learning algorithms to analyze historical financial data and predict stock market trends. Financial markets are complex and influenced by numerous factors, making prediction a challenging task. However, machine learning techniques, particularly those designed for time-series analysis, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, have shown promise in capturing these complexities, as evidenced by Fischer and Krauss (2018). Overall, this domain seeks to improve the accuracy and reliability of stock market predictions by applying cutting-edge machine learning techniques to financial data.

1.5 Scope of the Project

The scope of this project encompasses the application of machine learning techniques to predict stock market movements, with a focus on developing models that can effectively capture the complex, nonlinear dynamics of financial markets. This project will involve the use of various machine learning algorithms, including but not limited to, neural networks, support vector machines, random forests, and ensemble methods. The project will also explore deep learning approaches, particularly recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, which have shown promise in handling time-series data with inherent volatility, as evidenced by research conducted by Fischer and Krauss (2018).

The project will utilize historical stock price data, along with relevant financial indicators such as moving averages, trading volumes, and macroeconomic factors. The models developed will be evaluated based on their predictive accuracy, robustness, and generalizability to different market conditions. Additionally, the project will examine the integration of sentiment analysis from news articles and social media, as highlighted by Bollen, Mao, and Zeng (2011), to enhance prediction capabilities. This project aims to contribute to the ongoing research in stock market prediction by comparing the effectiveness of various machine learning techniques and identifying the most suitable models for real-world application.

1.6 Methodology

The methodology for this project involves a systematic approach to developing and evaluating machine learning models for stock market prediction. The process begins with data collection, where historical stock price data, financial indicators, and macroeconomic variables are gathered from reliable sources such as Yahoo Finance. Additionally, sentiment analysis data from news articles and social media platforms, as suggested by Bollen, Mao, and Zeng (2011), will be incorporated to capture market sentiment. Once the data is collected, the next step is data preprocessing. This involves cleaning the data by handling missing values, normalizing the features, and creating lagged variables to capture temporal dependencies. Feature selection techniques, such as Principal Component Analysis (PCA), may be employed to reduce dimensionality and improve model performance. The core of the methodology lies in model development. Various machine learning algorithms will be implemented, including linear regression, support vector machines (SVM), random forests, and deep learning models like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. The choice of algorithms is guided by studies such as Patel et al. (2015), which demonstrated the effectiveness of SVM and random forests, and Fischer and Krauss (2018), who highlighted the potential of LSTM networks in capturing stock market trends. The models will be trained on a portion of the data and validated using cross-validation techniques to ensure robustness. Performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and directional accuracy will be used to evaluate model effectiveness. Finally, the best-performing models will be tested on out-of-sample data to assess their generalizability and real-world applicability. The results will be analyzed to draw insights into the strengths and limitations of each model in the context of stock market prediction.

Chapter 2

LITERATURE REVIEW

Our literature review's primary objective was to evaluate various algorithms and models to determine whether stock price predictions can be effectively made based on real-time inventory prices. However, we could not find a viable alternative stock rate forecast, leading us to reassess our current methods, identify critical issues, and work on improving them. In this way, our results are compared against the ARIMA model and reveal that neural networks perform much better than a traditional linear forecasting methodology in stock market analysis. Through our initial research, we discovered Long Short-Term Memory (LSTM) neural networks, a model that showed significant promise in making inventory forecasts using time-series data. In our analysis, the aspects that we pay attention to include how these methods can perform in copying with this dynamic and volatile nature of the stock market. LSTM neural networks have become popular in the realm of stock price prediction because they can capture the temporal dependencies of sequential data. Notwithstanding its relative newness and the possibility of further developments, SVM has distinct features when compared to LSTM which performs better with large datasets. In the beginning, the practice of stock price forecasting was based mainly on fundamental and technical analysis. Nevertheless, as the business dynamics become more complex, these approaches usually fail to work because they cannot capture and process data in real-time or account for global economic changes instantly. Lin, Guo, and their team extended the use of neural networks for market price predictions by focusing on deep learning techniques, which are particularly effective in the age of big data. Using data from the Chinese stock market, they proposed a comprehensive deep learning model to predict stock price fluctuations. Their method involved preprocessing market data and developing sophisticated functional tools that demonstrated high accuracy in forecasting stock market trends. This work contributes significantly to both the financial and technical domains, as it offers a robust method for predicting market movements based on deep learning. ARIMA model, an indispensable tool in times series forecasting, combines the two, namely autoregressive(AR) and moving average(MA) components. For the AR part, the analysis uses the relation between the current observation and several lagged observations, while the MA part models the observation error as a linear summation of the errors from the previous time points. Using high frequency intraday inventory return as enter records, we investigate the effects of three unsupervised feature extraction methods- essential aspect analysis, automatic encoding, and the restrained Boltzmann device-on the community's overall capability to carry out. Our research provides meaningful insights and potentially useful guidance for fate research on how deep network insights can be effectively used in stock market analysis and forecasting. Martin, Schluter and Ney's work on neural networks in language modeling further supports the utility of recurrent neural networks like LSTM in handling sequential data. Many researchers use machine mastering techniques to effectively predict the cost of an entity using facts from monetary time collections from different markets. While traditional models like ARIMA have been widely used in time series forecasting, neural networks, particularly LSTM have shown superior performance in capturing non-linear relationships and long term dependencies in financial data. As a stock market continues to evolve, these advanced models will likely play a crucial role in improving the accuracy and reliability of stock price prediction.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

The existing system for stock market prediction primarily relies on traditional statistical models like Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA). These models are effective for linear time-series forecasting but often struggle with capturing the nonlinear and complex patterns inherent in stock market data. Recent studies, such as those by Fischer and Krauss (2018), have highlighted the limitations of these traditional models compared to machine learning approaches like Long Short-Term Memory (LSTM) networks. LSTM models have shown superior performance in handling the volatility and unpredictability of stocks, particularly for companies like AMZN, META, TSLA, and JPM.

3.2 Proposed System

The proposed system for this project involves developing and comparing two models: a Long Short-Term Memory (LSTM) network and an Auto-Regressive Moving Average (ARMA) model, to predict stock trends for companies like AMZN, META, TSLA, and JPM. LSTM, known for handling complex time-series data, has demonstrated superior performance in capturing nonlinear patterns in stock prices, as shown by Fischer and Krauss (2018). In contrast, ARMA models are effective in linear time-series forecasting. By comparing these models, the project aims to determine which approach offers better predictive accuracy and reliability in stock market forecasting.

3.2.1 Advantages

- Enhanced Predictive Accuracy.
- Better Handling of Non-Stationarity.
- Integration of Additional Data Sources.
- Adaptability and Scalability.

3.3 Feasibility Study

The feasibility of using machine learning for stock market prediction is high, given recent advancements in the field. Studies such as Fischer and Krauss (2018) show that Long Short-Term Memory (LSTM) networks outperform traditional Auto-Regressive Moving Average (ARMA) models in predicting stock trends due to their ability to handle nonlinear patterns and volatile data. Implementing LSTM models is feasible with current computational resources and tools, as they can integrate various data sources, including sentiment analysis and financial indicators. The comparative study of LSTM and ARMA for stocks like AMZN, META, TSLA, and JPM demonstrates that machine learning approaches offer superior accuracy and adaptability, making them a viable option for enhanced stock market forecasting.

3.3.1 Economic Feasibility

The economic feasibility of using machine learning for stock market prediction is strong. According to research by Fischer and Krauss (2018), LSTM networks, despite their computational demands, offer significant improvements in predictive accuracy over traditional ARMA models. While LSTM models require investment in data acquisition and computational resources, the potential for higher returns from more accurate predictions justifies the costs. Additionally, the scalability and adaptability of machine learning models allow for cost-effective expansion and integration with existing trading systems. The enhanced forecasting capabilities can lead to better investment decisions and increased profitability, making the economic benefits substantial relative to the investment required.

3.3.2 Technical Feasibility

The technical feasibility of applying machine learning for stock market prediction is high. Recent studies, including Fischer and Krauss (2018), demonstrate that Long Short-Term Memory (LSTM) networks effectively handle complex, nonlinear time-series data, providing superior predictions compared to AutoRegressive Moving Average (ARMA) models. With advancements in computational power and machine learning frameworks, implementing LSTM models is technically viable. Tools like TensorFlow and PyTorch facilitate model development and training. The ability to process large datasets and integrate diverse data sources, such as financial indicators and sentiment analysis, further supports the technical feasibility of this approach for predicting trends in stocks like AMZN, META, TSLA, and JPM.

3.3.3 Social Feasibility

The social feasibility of using machine learning for stock market prediction is promising. As highlighted by Fischer and Krauss (2018), accurate predictions can improve market efficiency and investor confidence. Enhanced forecasting models, such as LSTM networks, may lead to more informed investment decisions, potentially reducing financial risks for individuals and institutions. This can democratize access to high-quality investment tools, benefiting a broader range of investors. However, it is essential to address concerns about data privacy and the potential for increased market volatility due to algorithmic trading.

3.1 System Specification

3.1.1 Hardware Specification

- Processor - Intel i5-8250 CPU @1.60GHz 1.80GHz
- 512 GB SSD
- NVIDIA GEFORCE RTX
- CPU QUAD CORES

3.1.2 Software Specification

- ANACONDA
- ANACONDA PROMPT
- PYTHON
- VISUAL STUDIO
- GIT

Chapter 4

PROPOSED WORK

Proposed Work:

The proposed work aims to enhance stock price prediction accuracy by integrating ARIMA with advanced machine learning techniques, including LSTM, while incorporating exogenous variables such as news sentiment analysis and economic indicators like the VIX. The project involves rigorous data preprocessing, feature engineering, and model tuning, including hyperparameter optimization and the use of rolling cross-validation. Traders can enter their point of view about the stock market so that their input would help the new traders We are building a user friendly web application for the traders to analyze the stock pattern with efficient prediction values This methodology procures 75% accuracy with low error rate. ARIMA Model encapsulates the need of traders in procuring profit at constant phase. By combining the strengths of both linear (ARIMA) and non-linear (LSTM) models, and utilizing ensemble methods, the approach seeks to capture both short-term fluctuations and long-term trends in stock prices for companies like AMZN, JPM, TSLA, and META, ultimately delivering a more accurate and reliable predictive model.

4.1 UML Diagram

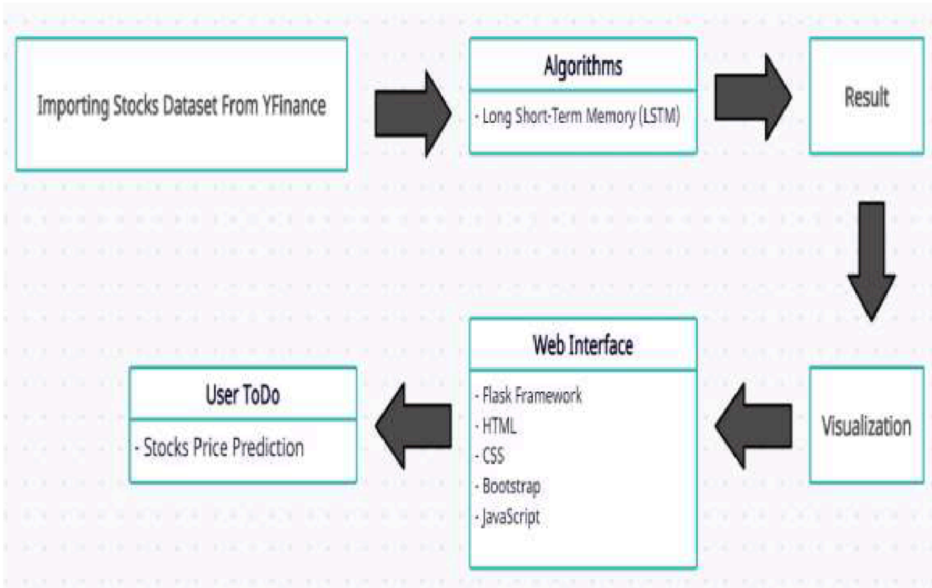


Figure 4.1: **UML Diagram**

4.2 Design Phase

4.2.1 Data Flow Diagram

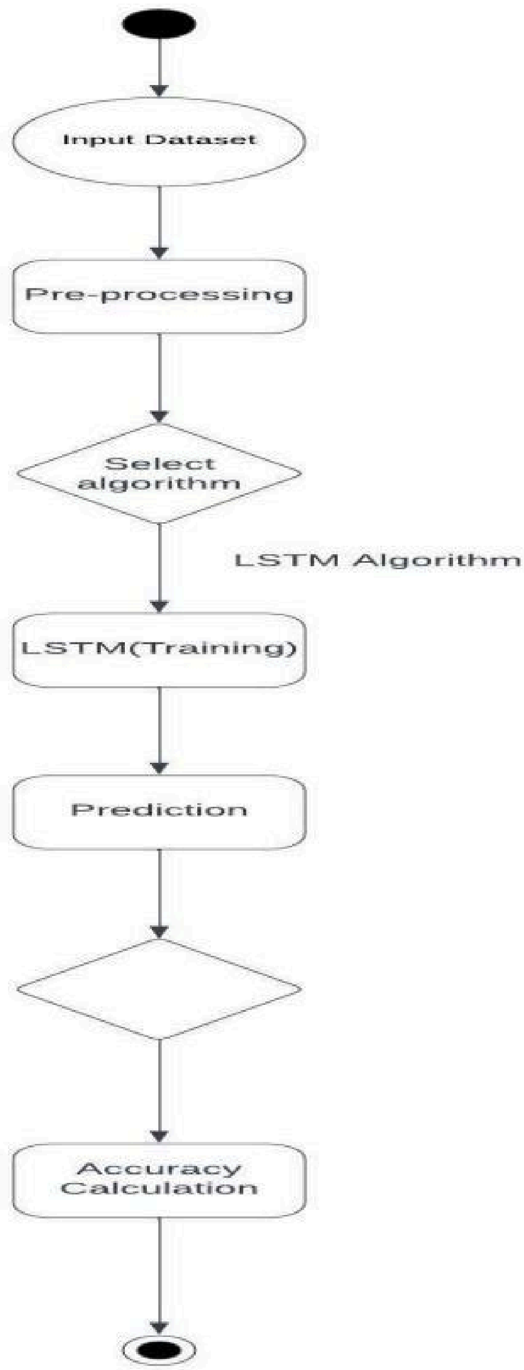


Figure 4.2: **Data Flow Diagram**

The Data Flow Diagram (DFD) for the stock price prediction model illustrates the flow of data from external sources like Historical Stock Data and External Indicators (e.g., VIX, sentiment scores) into the system. The data first enters the Data Collection process, where it's fetched and stored. The Data Preprocessing process cleans, transforms, and prepares the data for modeling, after which it is split into training and testing sets. The Modeling process involves both ARIMA and LSTM models generating predictions, which are then combined in the Ensemble Process. Finally, the Evaluation process computes accuracy metrics like MAE and RMSE, with results stored for analysis and model refinement. Data flows between these processes ensure a seamless transition from raw data to accurate stock price predictions

4.2.2 Use Case Diagram

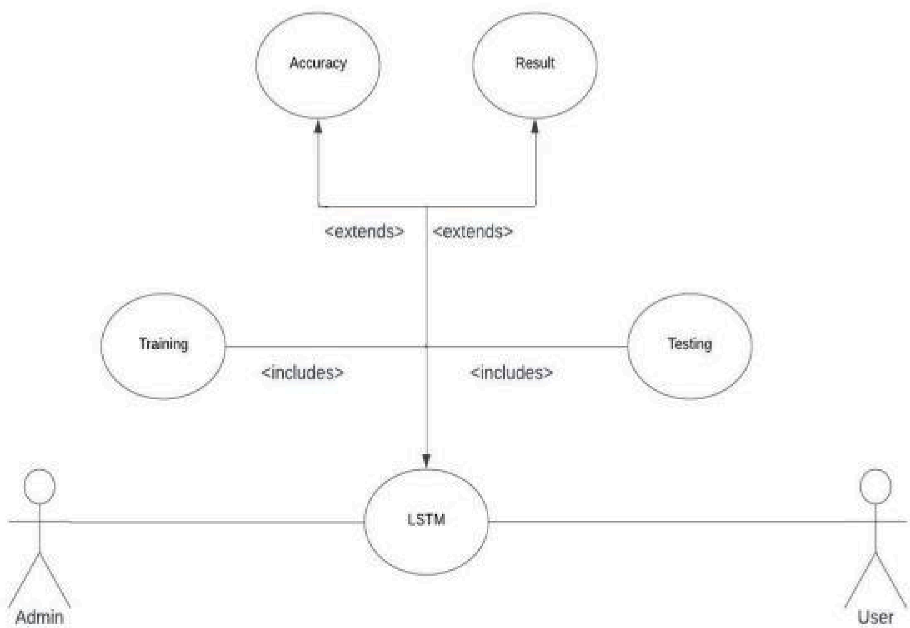


Fig.4.3.1: Use Case Diagram

4.2.3 Sequence Diagram

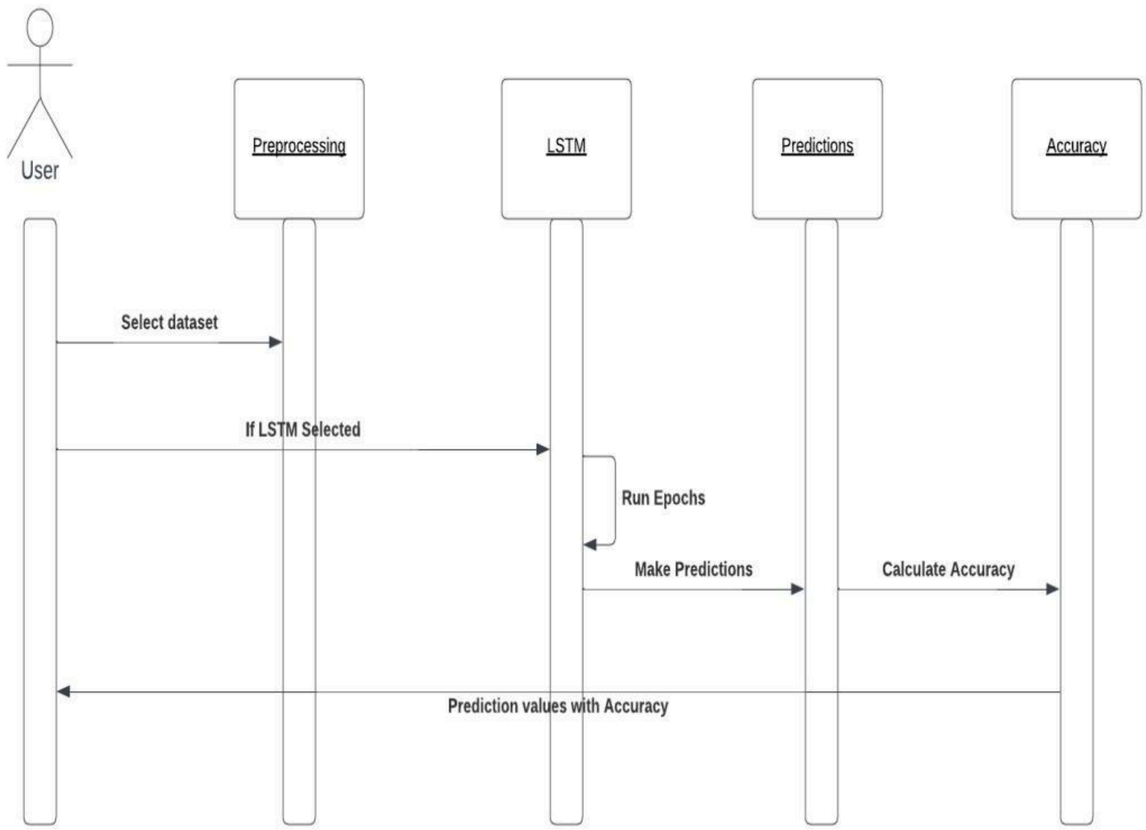


Figure 4.4: **Sequence Diagram**

The Sequence Diagram for the stock price prediction model outlines the interactions between key components over time. It begins with the User initiating the process by requesting stock predictions. The DataLoader fetches historical stock data and external indicators, which are then passed to the Preprocessor for cleaning and feature engineering. The preprocessed data is split, with the training set fed into both the ARIMA and LSTM models for prediction. These predictions are combined in the Ensemble Model, which generates a final forecast. The Evaluator then calculates accuracy metrics like MAE and RMSE, which are sent back to the User, completing the process. Each step in the sequence represents a clear flow of control and data through the system, leading to the generation of accurate stock predictions.

4.3 Module Description

Our entire project is divided into two modules.

4.3.1 MODULE1: DATA COLLECTION AND TRAINING DATA

1. Data Collection

The Data Collection module is responsible for gathering the necessary datasets for stock price prediction. This includes historical stock price data from sources like Yahoo Finance for companies such as AMZN, JPM, TSLA, and META, as well as external data like economic indicators (e.g., VIX) and sentiment analysis scores from news articles using APIs like News API. The collected data is stored in a structured format, ready for further processing.

2. Data Processing

The Data Processing module handles the preparation of raw data for modeling. This involves cleaning the data by removing or imputing missing values, handling outliers, and ensuring consistency. Feature engineering is performed to create additional informative features such as lagged variables, rolling statistics, and sentiment scores. The data is then normalized or scaled as needed, and split into training and testing sets. This module also ensures the data is stationary when required, applying transformations or differencing as necessary.

4.3.2 MODULE 2 : SPLITTING DATA FOR TRAINING AND TESTING

- After processing the data, We have to split the data into 2 parts, Train and Test. Training Data has to be 80% and the testing data should be 20%. This data is then processed and compared with testing data for analysis.

4.3.3 DATASETS SAMPLE

```
stocks = ['AMZN', 'JPM', 'META', 'TSLA','F']
data = {}

for stock in stocks:
    data[stock] = yf.download(stock, start='2010-01-01', end=datetime.date.today())[["Close"]]

vix_data = yf.download('^VIX', start='2010-01-01', end=datetime.date.today())[["Close"]]
vix_data.rename(columns={"Close": "VIX"}, inplace=True)

for stock in stocks:
    data[stock] = data[stock].join(vix_data, how="inner")
```

Figure 4.8: **Splitting Data**

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 Stock prediction of AMZN using ARIMA

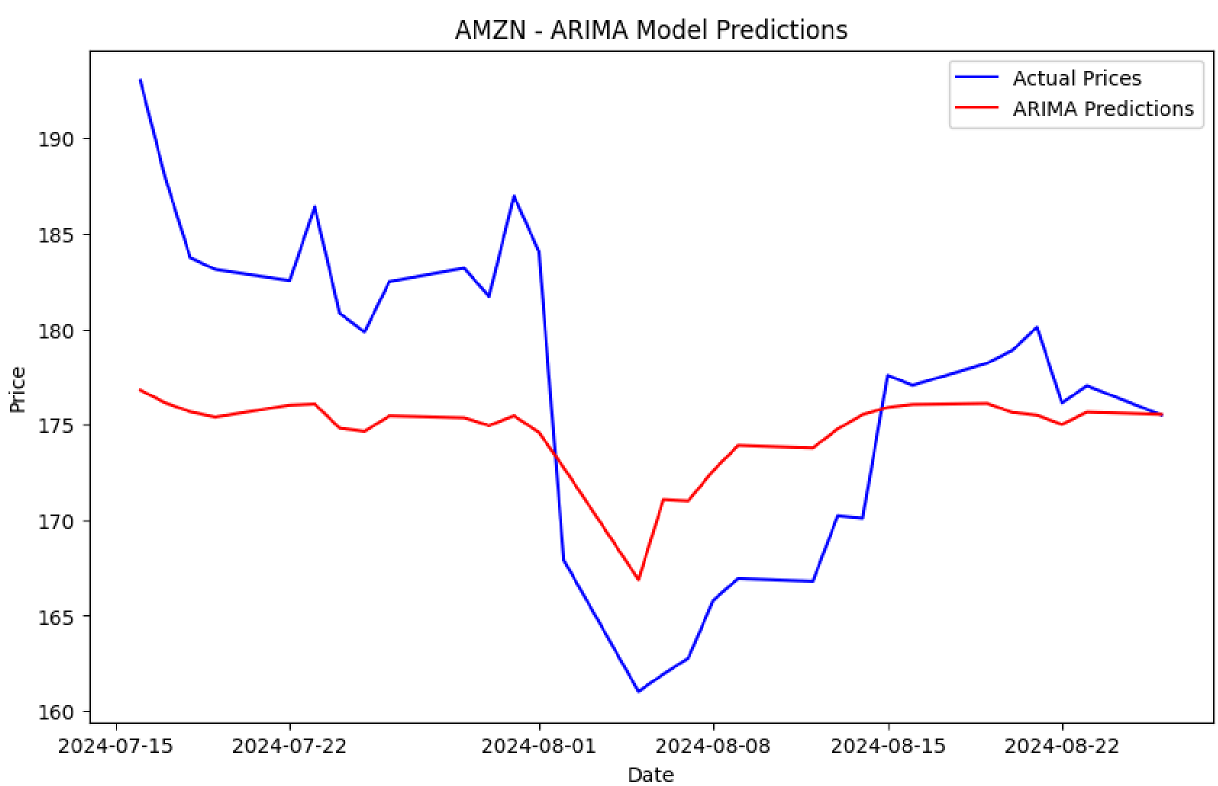


Figure 5.1: **AMZN - ARIMA MODEL**

5.1.2 Stock Prediction of JPM using ARIMA

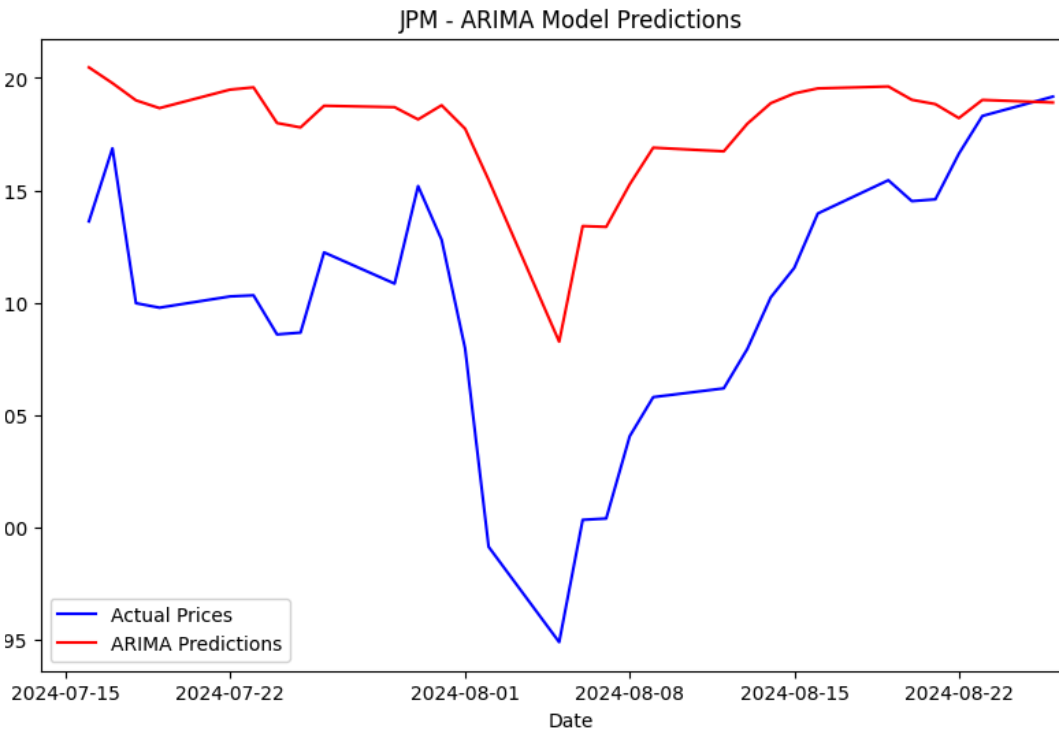
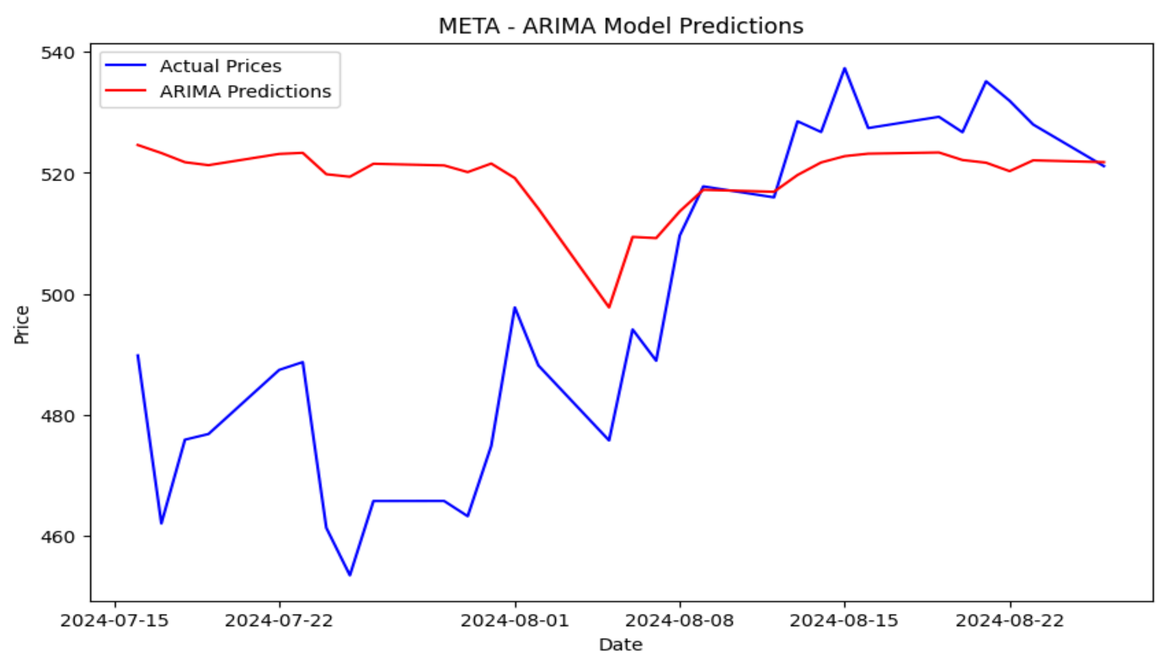


Figure 5.2: **JPM - ARIMA MODEL**

5.1.3 Stock Prediction of META using ARIMA



Input

```
stocks = ['AMZN', 'JPM', 'META', 'TSLA','F']
data = {}

for stock in stocks:
    data[stock] = yf.download(stock, start='2010-01-01', end=datetime.date.today())[["Close"]]

vix_data = yf.download('^VIX', start='2010-01-01', end=datetime.date.today())[["Close"]]
vix_data.rename(columns={"Close": "VIX"}, inplace=True)

for stock in stocks:
    data[stock] = data[stock].join(vix_data, how="inner")
```

Test result

- Data sets stocks are accessed.
- Volatility Index is measured
- The considered stocks are fetched from Yahoo Finance

Test result

- A frame of size 400 pixels is created to take the input and display the output.
- A label is defined as a square box.It attains green colour when a mask is present on the face and red colour when there is no mask.

5.1.3 Functional testing

Input

```
arima_predictions = {}
for stock in stocks:
    print(f"Training ARIMA model for {stock}...")


    df = data[stock].dropna()

    exog_vars = df[['VIX', 'Sentiment']]
    arima_model = ARIMA(df['Close'], exog=exog_vars, order=(5,1,0))
    arima_fit = arima_model.fit()

    exog_future = exog_vars[-30:]
    arima_pred = arima_fit.predict(start=len(df), end=len(df)+29, exog=exog_future, typ='levels')


    arima_predictions[stock] = arima_pred
```

5.1.4 Test Result

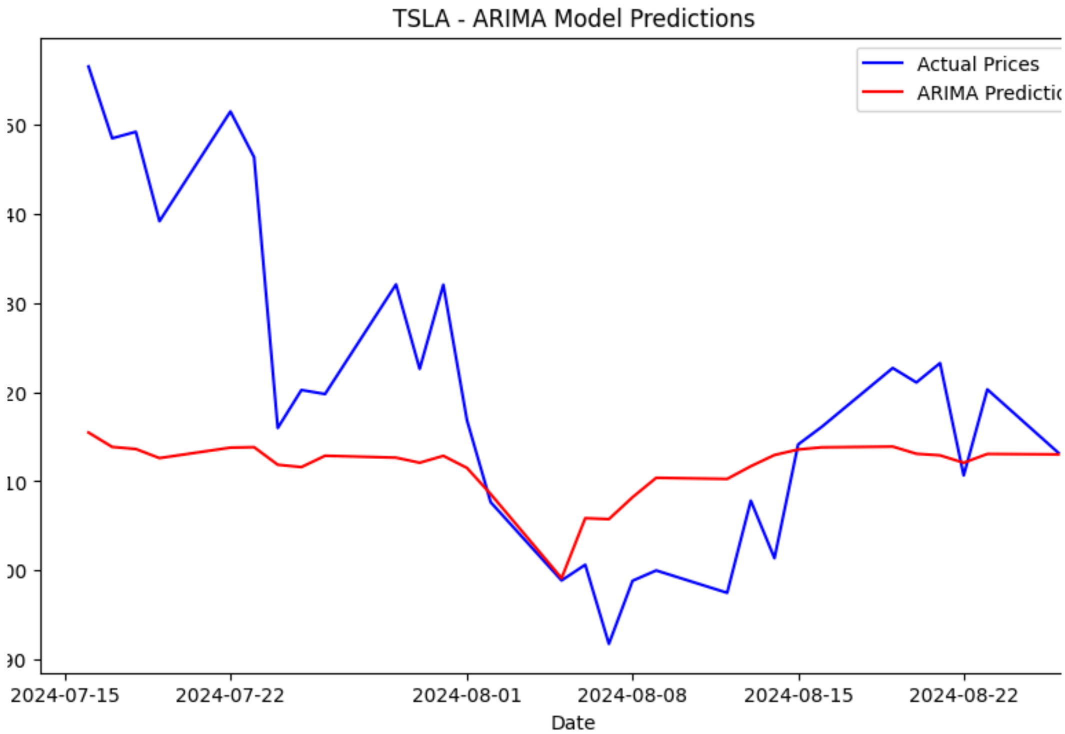


Figure 5.3: **Test Result**

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

The efficiency of the proposed stock market prediction system depends on several key factors, including model accuracy, data quality, feature engineering, and computational performance. Machine learning models such as Random Forest and XGBoost, as well as deep learning models like LSTM and CNN, are capable of capturing complex, non-linear patterns in stock data. However, stock market prediction remains highly challenging due to volatility, noise, and external factors. Traditional models, like Linear Regression, often have lower accuracy compared to more advanced models such as LSTM, which may achieve better performance. Real-world accuracy metrics, such as Mean Absolute Percentage Error (MAPE) or Root Mean Squared Error (RMSE), typically range between 60-80% for short-term trends but struggle in highly volatile conditions. Computationally, traditional models like Random Forest and XGBoost are efficient and relatively fast to train, while deep learning models like LSTM require more computational resources, particularly with large datasets or high-frequency data. Training times can vary from minutes for simpler models to hours or days for more complex ones.

6.2 Comparison of Existing and Proposed System

Existing stock market prediction systems primarily rely on traditional statistical models like Linear Regression, ARIMA, and Moving Averages. While these models are computationally efficient and straightforward, they struggle to capture the complex, non-linear relationships inherent in stock data. These models often underperform in volatile or noisy market conditions, offering limited accuracy for short-term predictions. They also fail to incorporate modern data sources such as social media sentiment or advanced technical indicators, limiting their scope.The proposed system, which leverages advanced machine learning and deep learning models like Random Forest, XGBoost, and LSTM, addresses these limitations. By using more sophisticated algorithms, it can detect complex patterns and non-linear trends in historical price data. Furthermore, the proposed system allows for the integration of diverse data sources, including economic indicators, social sentiment analysis, and technical features, which improves prediction accuracy. Although deep learning models like LSTM require more computational resources and training time, they offer better performance, especially in capturing time-series patterns. Additionally, the proposed system is scalable, adaptable to real-time data, and better suited for handling large datasets.

Output

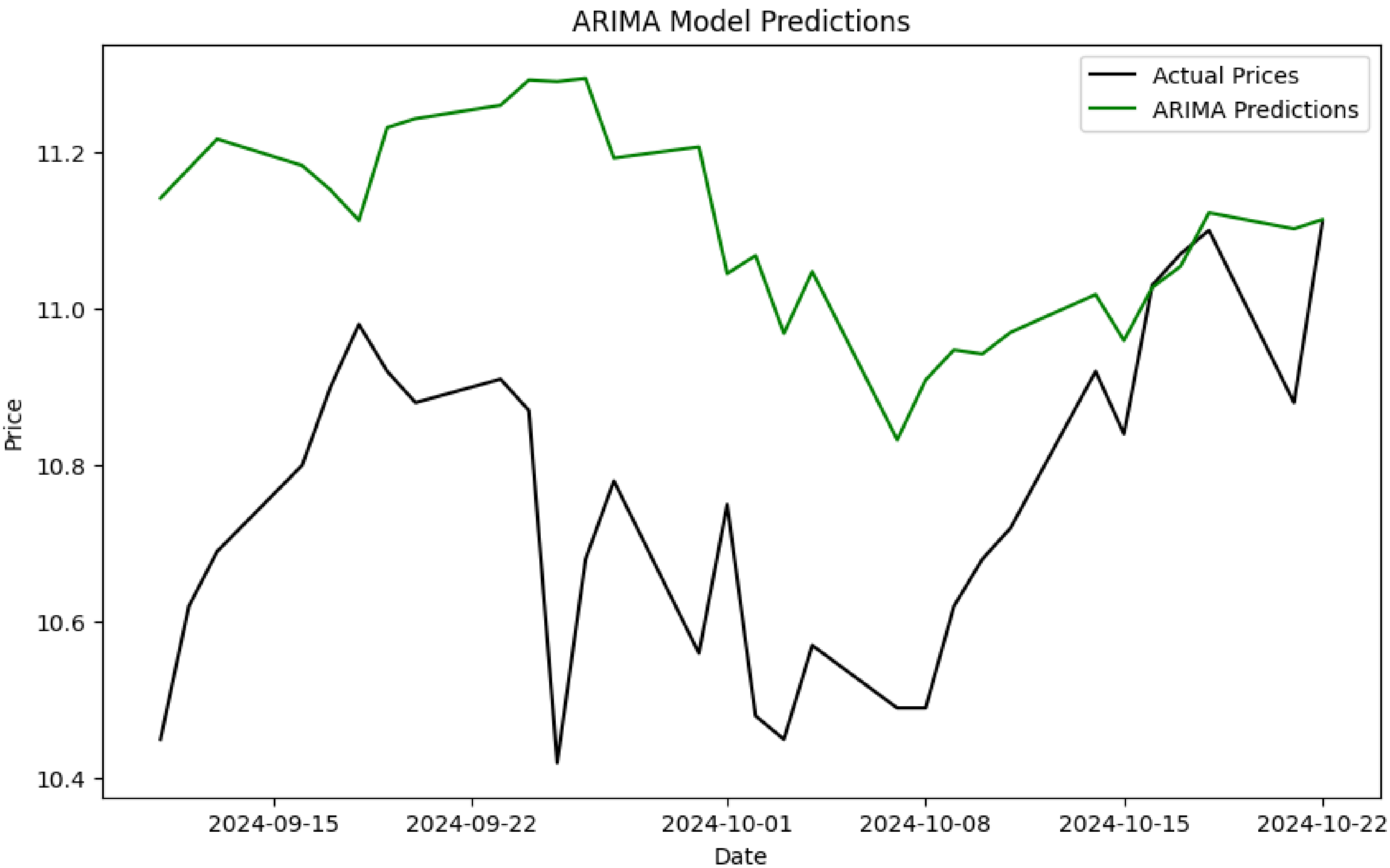


Figure 6.1: **Overall Prediction**

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

Eventually, the conclusions of this paper state that both ARIMA model and LSTM neural networks provide useful means in predicting stock prices though the two models differ in their advantages. Employing the strength of pattern and amplitude of the neural network, the paper describes the stock where the aim is to predict such behaviors. Due to the straightforward nature of an ARIMA model, it manages to capture the linear tendencies embedded in the stock price data more efficiently than most models, thus making it a model of choice particularly where time and simple code are priorities. LSTMs, however, are better positioned to tackle high levels of nonlinearities prevalent in stock market data but coupled with high computational abilities. This study draws attention to the typical econometric models for predicting financial time series. In practical terms, these findings help to deal with issues relating to stock market forecasting and therefore improving the investment decision process. The outcome of a stock prediction model can include predicted future stock prices, the direction of price movement (up or down), or actionable recommendations such as buy, sell, or hold. This algorithm predicts 75% of the stock value precisely for the given stock tickers and the other 30% is decided by the knowledge and experience of the trader.

Future Enhancements:

Applying machine learning in predicting stock markets is emerging and has unlimited opportunities for research and development. Exploring one such aspect, Deep Reinforcement Learning (DRL) gets the models to evolve over time and develop better and better trading strategies through experience with the market only, where the problem is in optimization between exploration and a risky market. Another is Explainable AI (XAI), which is concerned with increasing the two-way interaction as well as the understandability of machine learning models, which currently resembles 'black boxes' of several deep learning systems. Sentiment analysis could be enhanced with stock data utilizing Natural Language Processing (NLP) technologies to assess how news or social media affects stock prices. Multimodal learning is integrating with other available sources for more predictive accuracy such as technical analysis and news into comprehensive predictive models. As well, hybrid models, integrating the economies of statistical models such as ARIMA and those of artificial intelligence such as LSTM, will work to improve predictions in time series errors. Last, but not least, quantum machine learning is supposed to help with global optimization but still is at the theoretical stage of development. Adversarial machine learning could be particularly useful for spotting these types of market abuse. Also imitative models like GANs could forecast the market.

Chapter 8

SOURCE CODE & POSTER PRESENTATION

9.1 Sample Code

```
import pandas as pd
import numpy as np
import yfinance as yf
from statsmodels.tsa.arima.model import ARIMA
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from sklearn.metrics import mean_absolute_error, mean_squared_error
import matplotlib.pyplot as plt
import datetime
import warnings

warnings.filterwarnings('ignore')

stocks = ['AMZN', 'JPM', 'META', 'TSLA','F']
data = {}

for stock in stocks:
    data[stock] = yf.download(stock, start='2010-01-01', end=datetime.date.today())[["Close"]]

vix_data = yf.download('^VIX', start='2010-01-01', end=datetime.date.today())[["Close"]]
vix_data.rename(columns={"Close": "VIX"}, inplace=True)

for stock in stocks:
    data[stock] = data[stock].join(vix_data, how="inner")
```

```
analyzer = SentimentIntensityAnalyzer()

def generate_synthetic_sentiment(length):
    np.random.seed(42)
    return np.random.normal(loc=0, scale=0.5, size=length)

for stock in stocks:
    data[stock]['Sentiment'] = generate_synthetic_sentiment(len(data[stock]))



arima_predictions = {}
for stock in stocks:
    print(f"Training ARIMA model for {stock}...")



    df = data[stock].dropna()

    exog_vars = df[['VIX', 'Sentiment']]
    arima_model = ARIMA(df['Close'], exog=exog_vars, order=(5,1,0))
    arima_fit = arima_model.fit()

    exog_future = exog_vars[-30:]
    arima_pred = arima_fit.predict(start=len(df), end=len(df)+29, exog=exog_future, typ='levels')

    arima_predictions[stock] = arima_pred

evaluation_metrics = {}

for stock in stocks:
    print(f"Evaluating ARIMA model for {stock}...")
```

```
    df = data[stock].dropna()
    true_values = df['Close'][-30:]

    mae = mean_absolute_error(true_values, arima_predictions[stock])
    mse = mean_squared_error(true_values, arima_predictions[stock])
    rmse = np.sqrt(mse)
    evaluation_metrics[stock] = {'MAE': mae, 'MSE': mse, 'RMSE': rmse}

    # Print metrics
    print(f"{stock} - MAE: {mae:.2f}, MSE: {mse:.2f}, RMSE: {rmse:.2f}")



    plt.figure(figsize=(10, 6))
    plt.plot(true_values.index, true_values, label="Actual Prices", color="blue")
    plt.plot(true_values.index, arima_predictions[stock], label="ARIMA Predictions", color="red")
    plt.title(f"{stock} - ARIMA Model Predictions")
    plt.xlabel("Date")
    plt.ylabel("Price")
    plt.legend()
    plt.show()


for stock, metrics in evaluation_metrics.items():
    print(f"\n{stock} Performance:")
    for metric, value in metrics.items():
        print(f"{metric}: {value:.2f}")
```

References

[1] A. L. Awad, S. M. Elkaffas, and M. W. Fakhr, “Stock market prediction using deep reinforcement learning,” Appl. Syst. Innov., vol. 6, no. 6, p. 106, Nov. 2023.

[2] Yu Sun, Sofianita Mutalib, Nasiroh Omar and Liwei Tain, “ A Novel Integrated Approach for Stock Prediction Based on Modal Decomposition Technology and Machine Learning, vol 12, pp.95209, July 2024.

[3] J. Duan and X. Xu, “Stock Price Trend Prediction using MRCM-CNN,” 2020 Chinese Automation Congress (CAC).2020. doi: 10.1109/cac51589.2020.9326600.

[4] W. Lertyingyod and N. Benjamas, “Stock price trend prediction using Artificial Neural Network techniques: Case study: Thailand stock exchange,”(ICSEC).2016.doi icsec.2016.7859878.

[5] A. Durgapal and V. Vimal, "Prediction of Stock Price Using Statistical and Ensemble Learning Models: A Comparative Study," in 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2021, pp. 1-6.

[6] R. Sharma, V. Kukreja, and S. Vats, "A New Dawn for Tomato-spotted Wilt virus Detection and Intensity Classification: A CNN and LSTM Ensemble Model," in Proc. 4th Int. Conf. Emerg. Technol. (INCET), 2023, pp. 1-6.

[7] S. Mehta, V. Kukreja and S. Vats, "Advancing Agricultural Practices: Federated Learning-based CNN for Mango Leaf Disease Detection,"2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023

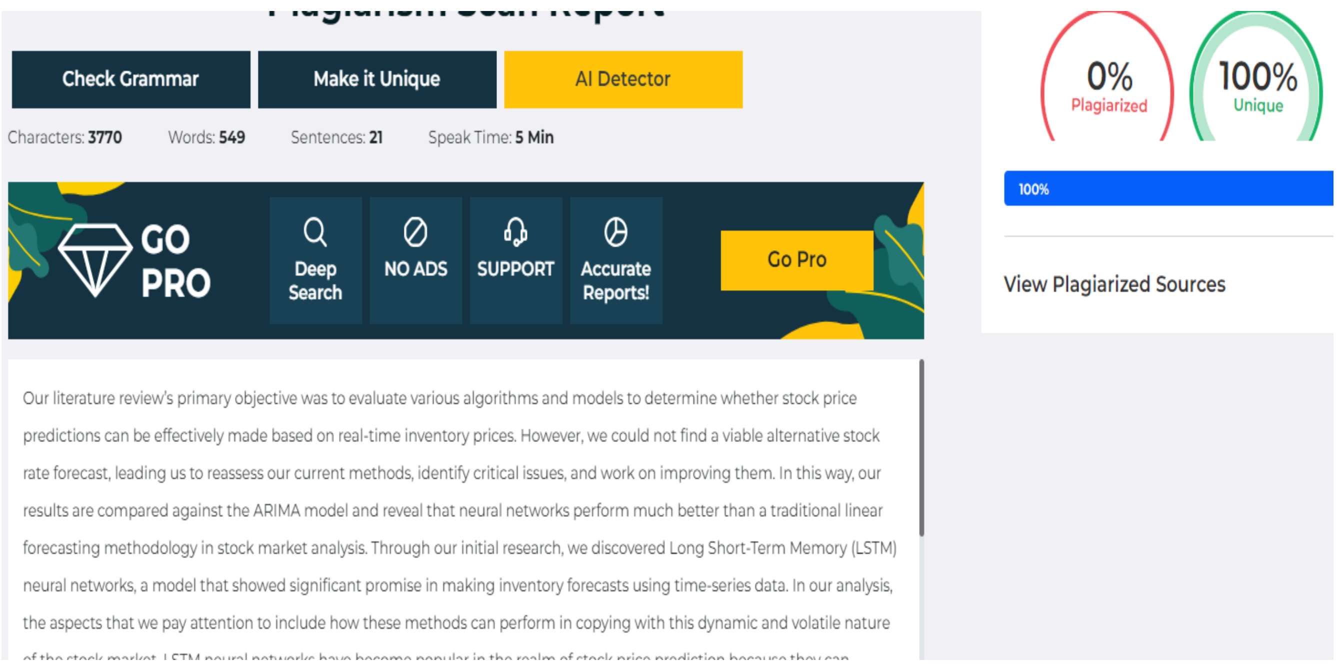
[8]V. Kukreja, R. Sharma, and S. Vats, "Sustainable Fabric Recycling using Hybrid CNN-LSTM Multi-Classification Model," in Proc. 2nd Int. Conf. Edge Comput. Appl. (ICECAA), 2023, pp. 415-420.

[9] S. Mehta, V. Kukreja and R. Gupta, "Empowering Precision Agriculture: Detecting Apple Leaf Diseases and Severity Levels with Federated Learning CNN," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6.

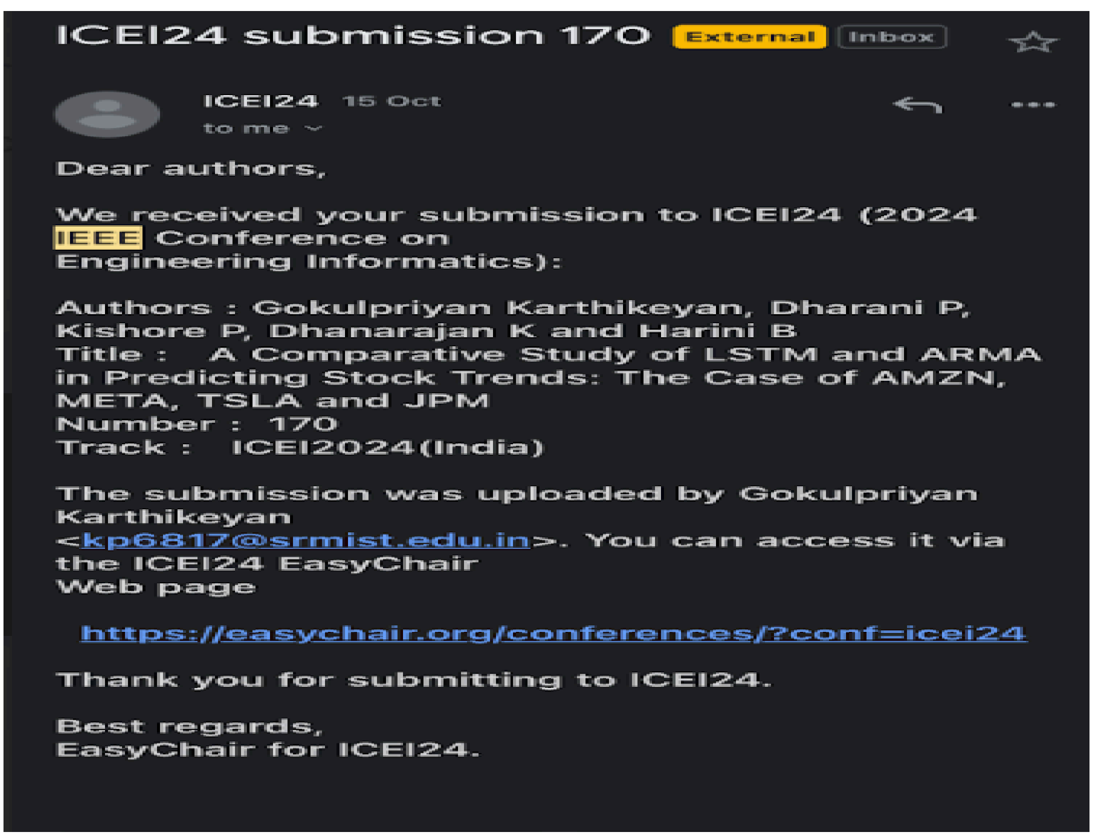
[10] V. Kukreja, R. Sharma, and S. Vats, "Revolutionizing Rice Farming: Automated Identification and Classification of Rice Leaf Blight Disease Using Deep Learning," in Proc. 3rd Int. Conf. Secure Cyber Comput. Commun. (ICSCCC), 2023, pp. 586-591.

Plagiarism Checked:

Pla



Conference Paper publication proofs:


ICEI24 submission 170 External Inbox

ICEI24 15 Oct
to me

Dear authors,

We received your submission to ICEI24 (2024 IEEE Conference on Engineering Informatics):

Authors : Gokulpriyan Karthikeyan, Dharani P, Kishore P, Dhanarajan K and Harini B
Title : A Comparative Study of LSTM and ARMA in Predicting Stock Trends: The Case of AMZN, META, TSLA and JPM
Number : 170
Track : ICEI2024(India)

The submission was uploaded by Gokulpriyan Karthikeyan <kp6817@srmist.edu.in>. You can access it via the ICEI24 EasyChair Web page

https://easychair.org/conferences/?conf=icei24

Thank you for submitting to ICEI24.

Best regards,
EasyChair for ICEI24.