BIRMINGHAM CITY University

CMP6200/DIG6200
INDIVIDUAL UNDERGRADUATE PROJECT
2024–2025

A3: Dissertation

STOCK MARKET PREDICTION USING MACHINE LEARNING (LSTM)

Course: BSc (Hons) Computer Science

Student Name: Talha Ahmed Risath

Student Number: 22222895

Supervisor Name: Samer Bamansoor

Abstract

Stock market is recognised for its complexity and volatility which intrigues investors seeking to accurately forecast the future trends. The application of LSTM algorithm for predicting stock market prices in financial time, series forecasting is highly relevant in today's time because of its ability of retaining long-term dependencies which is crucial to understand market trends. The prediction of stock price movement is done through LSTM based model by integrating both classification and regression tasks. The regression task is used to predict the stock prices while the classification task is used to predict the directional movement of the price up or down the time series, graphs, matrix, and visual output are provided which reflect the underlying insights and challenges of employing deep learning models for forecasting stock prices. LSTM is long short-term memory network which is a type of recurrent neural network designed for handling long-term dependencies and sequential data which is effective for time series tasks such as stock price prediction. With the adoption of a mixed methods approach in the research, a thorough review of literature on existing machine learning algorithms along with the practical implementation of LSTM based predictive model is done. The secondary data collection is done in the study especially the historical stock data of Tesla from Mendeley dataset repository for training and testing the model. The LSTM model is analysed for its performance quantitatively using financial metrics like mean squared error, mean absolute error, accuracy, recall and F1 score. The performance of the model is benchmarked against the traditional approaches for highlighting the superiority of LSTM in capturing temporal dependencies and non-linear relationships in financial data

Acknowledgements

I would like to express my gratitude to my friends and supervisor for their encouragement, invaluable guidance and continuous support throughout this project. Their constructive feedback and expert insights have played a vital role in shaping the depth and direction of my work.

I am thankful to my academic staff and lecturers for their support and for providing inclusive environment of growth and learning. My gratitude towards my supervisor Dr. Samer Bamansoor is immense for his support and guidance throughout the completion of my work. He remained patient, provided his valuable insights and consistent feedback that contributed to the challenging phases of the project and helped me stay motivated and focused. His mentorship enriched the academic quality of my work along with which it also enhanced my understanding on the subject matter.

I would also like to extend my thanks to the authors, researchers and developers of the peer reviewed articles, journals, and datasets through which the implementation of the research was possible. By providing access to their open-source repositories, I could retrieve datasets and gather secondary research materials that became instrumental for the success of my research.

Table of Contents

List of Tables

List of Figures

Glossary

1	ARIMA	AutoRegressive Integrated Moving Average
2	CCI	Commodity Channel Index
3	CNN	Convolutional Neural Network
4	EMA	Exponential Moving Average
5	LSTM	Long Short-Term Memory
6	MACD	Moving Average Convergence Divergence
7	MAE	Mean Absolute Error
8	ML	Machine Learning
9	MSE	Mean Squared Error
10	RMSE	Root Mean Squared Error
11	RNN	Recurrent Neural Network
12	ROC	Rate of Change
13	RSI	Relative Strength Index
14	SMA	Simple Moving Average
15	SVM	Support Vector Machine

1 Introduction

1.1 Background and rationale

Stock trading is found out to be one of the most crucial and dynamic activities in the world of finance stock (Reddy, 2018). Market prediction has always been the significant topic of interest due to its potential in generating substantial profits for the organisations and individuals. The prediction of stock market refers to the process of forecasting future values of a stock or other financial instruments traded in a financial exchange (Reddy, 2018). The stock market nature is volatile, and it has a variety of influencing factors which makes the prediction of stock prices a complex task (Reddy, 2018). However, with the application of machine learning techniques, stock price prediction is possible where the research highlights the way Python programming can be utilised for training models based on historical data.

It is learnt that historically, the stockbrokers relied on two primary methods for learning about price predictions. It includes fundamental analysis and technical analysis (Reddy, 2018). The intrinsic value of the company is assessed in the fundamental analysis by examining factors like industry performance, financial statements, overall political climate and economic trends (Reddy, 2018). However, as opposed to that, the technical analysis is focused on statistical patterns derived from historical market data such as trading volumes and stock prices (Karmiani et al., 2019). Both of these methods provide insights related to market trends, but limited understanding is provided in handling complex and massive datasets which are being generated in markets today (Reddy, 2018). Hence, ML presents an opportunity and a powerful tool for predicting stock market prices (Reddy, 2018). The historical stock data today is available in large number along with the computational power to process it, the researchers and traders are turning to ML technologies to uncover hidden patterns and improving the overall accuracy of their forecast (Karmiani et al., 2019).

Among different algorithms of ML, it is analysed that support vector machine shows promising results as it is efficient in regression and classification task as per Reddy (2018). However, LSTM algorithm is preferred over support vector machine when it comes to stock price prediction because support vector machine algorithm is not designed for sequential or time series data (Reddy, 2018). The SVM algorithm does not have the capability of remembering the order in which the data is appearing (Lakshminarayanan, 2019). However, the LSTM is designed specifically for time series problem by remembering long-term dependencies. LSTM is ideal for modelling sequences such as stock price prediction which mostly depends on past trends and historical data (Lakshminarayanan, 2019). Hence, the adoption of LSTM algorithm is done in the research for designing ML model and evaluating its performance based on metrics like MAE, MSE, accuracy, f1-score and recall score (Lakshminarayanan, 2019).

The LSTM i.e., Long Short-Term Memory is part of recurrent neural network (RNN) which is designed for overcoming the limitations of standard recurrent neural networks which are able to process sequential data and use specialised gates for managing memory in an effective manner. The author highlights that the LSTM can enhance the memory capacity of the recurrent neural network by holding short-term memory and using earlier information for performing tasks immediately (Lakshminarayanan, 2019). It is not necessary that the neural node has the access of all the comprehensive data in the past, hence the implementation of LSTM is done in neural network developed on RNN (Lakshminarayanan, 2019). The effectiveness of LSTM algorithm is not only limited to its memory capability, but they are employed in different sequencing modelling problems including geo spatial data, natural language processing, video and time series analysis (Karmiani et al., 2019). There is the issue of vanishing gradient problem in RNN. The issue is raised because of reusing same parameters in RNN block in each step (Karmiani et al., 2019). Hence, for finding a balance in the scenario, the novel parameters are incorporated in each step while also maintaining a constant overall number of learnable parameters and variable length sequences which gives gated RNN cells such as GRU and LSTM.

It is analysed that past information can be remembered effectively by LSTM algorithm where it can learn dependencies and patterns across long sequences. LSTM is highly suitable for applications like machine translation, stock market prediction and speech recognition (Karmiani et al., 2019). The analysis of historical price data can be done effectively by LSTM along with past events for potentially predicting future trends (Karmiani et al., 2019). In this, LSTM algorithm also considers the long-term factors affecting the price with which it predicts the fluctuation in stock market prices. The context of a sentence can be understood effectively by LSTM in one language which gets translated accurately into another language. LSTM takes the relationships and order between the words into consideration for making the translation. Another application area of LSTM is speech recognition where it evaluates the sound sequences in speech and converts them into text effectively (Karmiani et al., 2019). The research is significant in its academic as well as practical contribution where it provides a robust framework for analysts and traders to enhance the accuracy of market predictions and contribute to the growing literature on applicability of deep learning in finance. The demonstration of the LSTM algorithm is done in the study which can be used as a transformative tool that enables data driven decisions and helps in navigating market complexities.

1.2 Research aim

The project aim is to develop a predictive model for accurately forecasting stock market fluctuations using advanced deep learning techniques by mainly focusing on LSTM.

1.3 Research objectives

To fulfil the research aim, following objectives will be followed:

- To conduct a thorough review of existing ML algorithms, tools and models used for stock market prediction.
- To gather historical stock market data from credible sources like Yahoo Finance or Mendeley including economic indicators and prices.
- To design, implement and analyse performance of LSTM model for predicting stock price fluctuations by ensuring that the training and validation of the model is completed within the project deadline.
- To evaluate the implemented LSTM model by comparing its results with existing researchers to find out if model performance is up to the expected standards using effective evaluation metrics like MAE, MSE, accuracy, f1-score and recall score.

1.4 Research questions

RQ1: What algorithms of ML are suitable for stock market prediction along with their benefits and drawbacks in the model?

RQ2: Which ML algorithm will suit the best for evaluating stock market fluctuations based on gathered data?

RQ3: What is the prominence of financial indicators like MAE, MSE, accuracy, etc in ML model in determining the model performance?

1.5 Research significance

The research is highly significant because it addresses the challenges and complexities of predicting stock market price which is an area recognised for its nonlinearity, volatility and susceptibility to a wide range of impacting factors. The manual forecasting methods and traditional statistical models like technical and fundamental analysis have been used since long but they are not equipped in managing the increasing complexity and volume of market data. As opposed to that, the study proposes the use of advanced ML algorithms particularly the LSTM networks which has shown promising results in the past and superior performance for handling time series data.

A more effective approach is provided by LSTM algorithm for capturing temporal relationships and long-term dependencies in financial data which are often oversimplified and overlooked in traditional models. Hence, the research is particularly valuable in accurately forecasting stock price for effective decision-making such as required in portfolio management, stock trading and financial risk assessment. The research is contributing both academically and practically by evaluating the LSTM model performance using robust metrics like Mean Squared error, mean absolute error, F1 score, accuracy and recall. From a practical perspective, the research provides a robust LSTM model which can be used by financial institution, traders, and analysts for enhancing the accuracy of market predictions.

The existing knowledge and ML applications in finance is expanded through this research academically which offers comparative insights related to why LSTM algorithm outperforms the traditional algorithms and models like support vector machine (SVM) for sequential data tasks (Karmiani et al., 2019). The real-world Tesla stock data is used in the study from Mendeley dataset which provides an applicable and grounded case study. The research shows the way ML and especially LSTM can be utilised as a transformative tool for navigating the financial markets complexity and making data driven predictions efficiently and reliably.

1.6 Research problem

The primary issue addressed in the research is the challenge of predicting stock market fluctuation accurately using ML especially in the context of volatile non-linear and high dimensional financial market. The traditional approaches like technical analysis and fundamental concepts are limited in their capacity to handle and detect complex temporal patterns and manage large data sets which makes them insufficient to capture the trends in modern financial markets.

The methods like linear regression and ARIMA do not completely accommodate the evolving nature of stock trends and assume data stationarity (Idrees et al., 2019). These approaches struggle with the sequential dependency and face the challenge of underperformance or overfitting when they are applied to real world scenarios. The investigation of the applicability of LSTM network is done specifically in this research which is a class of RNN and recognised for its strength in handling time series data. The research effectively implements, trains and evaluate the LSTM based model on the historical stock data gathered from Mendeley repository. It incorporates a wide range of technical as well as financial indicators based on which the assessment of model performance is done. The ability of the LSTM network to effectively learn from past trends is explored in this research along with which the limitations are identified during the implementation phase such as class imbalance, poor regression performance and noisy data which questions the reliability of LSTM model in its practical application of predicting stock prices fluctuation.

1.7 Methodological approach

The mixed method approach is adopted in the study to systematically review the gathered qualitative data and empirical evidence from existing research articles and journals. In addition to that, the LSTM model implementation is done using ML algorithm that requires analysis of financial metrics to evaluate whether the model performance is optimal. The statistical evaluation of model performance is done by analysing MAE, MSE, accuracy, f1-score and recall score. The regression and classification plots are included in the research through which comprehensive understanding is increased on the generalisation capacity of model along with its bias towards under or overestimation. The practical insights are provided in the research about the way LSTM model can mimic human decision-making process when financial as well as technical indicators guide the model evaluation.

2 Literature Review

2.1 Overview

The chapter outlines a comprehensive literature review in which the gathered data from peer reviewed articles and journals is included to show applicability o deep learning algorithms in financial tasks. The studies are classified and reviewed in the chapter based on different ML and deep learning models like LSTM, regression, CNN and hybrid approaches. The contributions of these models in development of stock price prediction model are analysed in the literature. The literature focuses on various themes including features of LSTM networks in time series data and its benefits over traditional ML algorithms in predicting stock prices. The study also integrates the discussion on financial indicators used in analysing the model performance. The evolution in ML brought significant changes into the way financial data is predicted, and time series data is used in models which is reviewed in the literature study.

2.2 Historical development of stock market prediction models

The development of stock prediction models earlier relied heavily on traditional statistical methods such as GARCH and ARIMA (Jiang et al., 2021). However, with the introduction of deep learning, a notable shift is witnessed towards neural network-based approaches particularly for handling non-linear and complex data patterns (Jiang et al., 2021). LSTM is referred to long short-term memory is considered a variant of RNN i.e., recurrent neural networks. The LSTM networks are becoming increasingly popular because of their capability of capturing long-term dependencies as well as identifying pattern in the time series data (Nabipour et al., 2020). This capacity of LSTM networks over traditional regression and classification algorithms makes it highly applicable and reliable for predicting fluctuations in stock market prices. It is explored by the studies that the use of hybrid RNN-CNN models, CNN Models and bi-directional model is done which are hybrid approaches to improve the accuracy of prediction.

The transition from traditional models to ML based techniques was gradual where the early neural network model such as feed forward networks offered some improvements. The transition is explored in research done by Selvin et al. (2017) where the contrast between traditional methods and emerging ML models is provided such as recurrent neural networks i.e., RNN and convolutional neural networks i.e., CNN demonstrating the way these approaches effectively capture complex behaviours of market. The research by Pang et al. (2020) highlight that the financial market of China has played a crucial role in supporting economic development in the country by attracting both international and domestic attention. The rapid economic growth and improvements in financial services has led the investors and scholars to focus on predicting market transfers guiding investment decisions. There has been increased utilisation of neural networks in financial forecasting due to their abilities in classification and pattern recognition. Promising tools are offered by neural networks for stock market prediction.

However, it is analysed by Pang et al. (2020) that traditional neural network algorithms struggle to capture the complete complexity of the market by getting stuck in the local optima and producing sub optimal predictions. The transition of stock market prediction model from traditional approaches to advance deep learning methods is provided by Ta et al. (2020). The study by Ta et al. (2020) acknowledges that traditional models relied heavily on statistical analysis of historical data which do not achieve success to its potential due to unpredictable nature and complexities of stock markets. On the other hand, Khan et al. (2022) highlights the way stock market prediction models have evolved particularly deep learning and ML for improving prediction accuracy. The incorporation of news data and social media information is done in new ML models whose aim is to bridge the gap between stock price movement and real time information driven by sentiments (Khan et al., 2022).

2.2.1 Comparison between LSTM and regression models

In the field of stock market prediction, the LSTM networks showcase a powerful approach of ML which is a type of RNN that learns and retains long-term dependencies as opposed to the conventional algorithms like regression models as per Parmar et al. (2018). The stock market prediction is not entirely a random process, but it includes pattern identification which get deciphered through comprehensive analysis of historical data (Parmar et al., 2018). The techniques of ML are widely utilised in this domain for their capability of making predictions close to actual market value through which their overall profitability is enhanced, and accuracy is improved for brokers, institutions and investors (Parmar et al., 2018).

The traditional models like linear regression are employed frequently for stock price prediction, which is a supervised and simplistic learning algorithm through which the relationship between continuous dependent variable and independent variables are modelled by fitting a linear equation to observed data (Parmar et al., 2018). In this, the primary goal is to reduce the error between actual and predicted values using a technique known as gradient descent (Parmar et al., 2018). The strength of the regression model lies in its interpretability and simplicity which makes it highly suitable for understanding trends over short period of time (Parmar et al., 2018). However, the models are not as efficient when it comes to identifying complex temporal dependencies in the time series data especially in problems like stock market prediction.

2.3 Deep learning and LSTM in financial forecasting

Several studies have focused on the applicability of LSTM network for stock market prediction due to its effectiveness in handling sequential data especially in financial market. These studies have shown demonstration of LSTM for both short-term and long-term prediction of stock market. The study by Althelaya et al. (2018) explores bi-directional LSTM networks which shows an enhancement over standard LSTM processing input data in both directions. The approach captures more context and shows improved accuracy for long-term and short-term stock predictions. The bidirectional approach used for predicting stock market fluctuations is a pivotal work in demonstrating the way bi-directional LSTM can effectively outperform traditional LSTM models. It also offers a bias for other researchers who focus on multidirectional approaches.

The research by Bhandari et al. (2022) uses LSTM for predicting stock indices which highlights the effectiveness of the model in capturing market trends. The emphasis of the paper is on the importance of feature selection as well as data preprocessing which shows the way tuning parameters in LSTM can effectively improve performance of the prediction on large data sets such as stock indices from ML with applications journal. The use of LSTM 6 network is also demonstrated by Moghar and Hamiche (2020) which contributes to the research body on LSTM by focusing on the properties of recurrent neural network that effectively manages sequential dependencies.

The research presented by Moghar (2020) is extensively critical in establishing LSTM as a standard tool for predicting stock market fluctuations. The investigation of the application of LSTM model is done by Ghosh et al. (2019) in predicting stock price on the Indian market. The study highlights that LSTM performs effectively in the emerging markets and provides an essential comparative perspective for studies which are focused on more established markets like European market or US market. It is suggested by the findings of the study that LSTM handles volatility which is advantageous in high reward and high-risk markets. The concentration of the study by Hiransha et al. (2019) is on the national stock exchange in India where the comparison of the performance of LSTM is done with other deep learning model such as RNN and CNN. The research provided in Procedia Computer Science contributes significantly to the discussions on the way performance of LSTM varies by market and the establishment of a baseline performance expectation is done across different model architectures.

The stock market prediction is considered an uncertain and complex task due to multiple factors involved that impact stock prices (Ghosh et al., 2019). These include both irrational and rational components such as market rumours, investor sentiments, geopolitical events and economic indicators. In Indian market, where trading activity is considered vibrant and has increased investor participation, the prediction of stock movement becomes even more challenging (Ghosh et al., 2019). With the rise in ML and deep learning models like LSTM, the addition of a new dimension is seen in the stock market prediction efforts. The representation of Indian stock market is done by major exchanges like Bombay stock exchange i.e., BSE as well as national stock exchange i.e., NSE hosting a multitude range of organisations (Ghosh et al., 2019).

The characterisation of Indian stock market is done by frequent fluctuations, high volatility and variety of investor profiles which range from small retail investors to institutional players (Ghosh et al., 2019). The LSTM model's applicability in Indian stock market environment is rooted in their capacity to process sequential data and retaining long-term memory (Ghosh et al., 2019). As opposed to the traditional statistical models that relied heavily on linear assumptions, the LSTM networks capture temporal relationships and linear relationships existing in historical stock price data effectively as per Ghosh et al. (2019).

According to the efficient market hypothesis, it is analysed that the reflection of all publicly available information is done by the stock prices which makes it inherently difficult to predict future price based on previous data. However, in real world market, there exist irrational behaviour, delays in information dissemination and emotional trading that creates multiple opportunities for short-term trend prediction and pattern recognition (Ghosh et al., 2019). In this case, the applicability of LSTM model is particularly effective as they have the capacity of learning from subtle patterns which are integrated in time series data. According to the author, the stock data of Indian stock market such as the closing and opening price, high and low values, volume traded are fed into the LSTM model for forecasting the future price trends (Ghosh et al., 2019). The LSTM models can be enhanced further by integrating the external variables such as political development, macro-economic indicators including interest rate and inflation rates along with global market signals (Ghosh et al., 2019). The combination of structured and unstructured data will make the prediction more aligned with market behaviour as well as robust. The LSTM algorithm applicability for forecasting time series data has more benefits over traditional models such as moving averages and ARIMA in their capacity to avoid common pitfall like short-term memory issue and vanishing gradients (Ghosh et al., 2019).

As per the research by Ghosh et al. (2019), the investment environment of India is dynamic where government schemes push the retail investors towards equity market while the fixed deposits like traditional avenues lose attractiveness because of having lower interest rate. In this regard, the forecasting tools based on ML become even more relevant. With the rise of digital trading platforms and increasing accessible data, the small investors are empowered to use prediction models for the financial planning. An added benefit is provided by LSTM networks as it helps in evaluating growth patterns across organisations. The identification of sector having similar movement trajectory can be done by investors by comparing predicted stock trends of different companies to take informed decisions. For example, according to Ghosh et al. (2019), if a steady upward trend is predicted by the LSTM model in several IT organisations, it indicates a sector wide growth which prompts the investors to assign more resources in this domain.

2.4 Data sources and feature engineering in stock prediction

The feature engineering and data selection choice significantly impacts the LSTM model effectiveness in predicting stock market fluctuations. Diverse data input is required for accurate prediction which includes historical stock prices, economic indicators and trading volumes (Basak, 2019). The emphasis of the study by Ghosh et al. (2019) is on importance of including external macroeconomic factors such as inflation rate and GDP which influence the market behaviour.

In addition to that, it is suggested by Hiransha et al. (2018) that a combination of technical indicators must be used such as 'moving average' and social sentiment data for enhancing the predictive power of the model. One of the crucial steps which is significant in the effective model performance is feature selection where irrelevant or excessive features can lead to the overfitting issue in the model while too few features will limit the learning capacity of the model. It is highlighted by Shen et al. (2020) that the integration of innovative feature engineering techniques like sliding window method can improve the ability of the model to generalise well by capturing both long-term and short-term trends (Shen et al., 2020).

It is demonstrated in the study by Zhong et al. (2017) that dimensionality reduction techniques such as FRPCA, PCA and KPCA can improve data handling by re-organising and simplifying complex and large feature set that enhances the performance of stock prediction models. As per Hoseinzade et al. (2019), the traditional models are criticised for neglecting correlation 7 between different markets, but CNN based approaches for automatic feature extraction leverages a wide collection of market data for enhancing the predictive accuracy (Hoseinzade et al., 2019).

2.5 Evaluation metrics and model performance

The assessment of the predictive performance of LSTM model is done by evaluation metrics which are highly crucial in financial forecasting. One of the commonly utilised metrics are R-squared method and root mean squared error i.e., RMSE which provides insights into the accuracy as well as reliability of the model. The research by Moghar and Hamiche (2020) analysed the performance of their LSTM model by using these metrics which demonstrated the improved accuracy of the model over baseline models. On the other hand, Bhandari et al. (2022) performed comparison between the LSTM and model with other neural networks that concluded that superior performance was achieved by LSTM because of its memory cell structure that contributed to minimising error accumulation over time. In the similar study performed by Pawar et al. (2019), the advocation for a comparative approach was given in the evaluation where the authors suggested that alternative models should be considered by performance metrics such as CNN and RNN for highlighting the strengths and weaknesses of LSTM.

As per Nti et al. (2020), the extensive review of 122 studies reveal that prediction models are categorised into fundamental, technical and combined analysis that allows for a comparative understanding of model performance depending on timeframe, data sources, accuracy and algorithms. The evaluation criteria include error metrics and accuracy that highlights the importance of such measurements for assessing the reliability of the model and its applicability in stock forecasting (Nti et al., 2020). The study by Zhang et al. (2017) contributes to the theme of evaluation metrics and model performance by entailing the way state frequency memory recurrent network model addresses the multi frequency and complex nature of stock market data for improving prediction accuracy.

It is recognised by Zhang et al. (2017) that the movement in stock prices are influenced by the long and short-term trading patterns which are impacted by unpredictable and external factors such as economic and political events. These patterns can be decomposed into distinct frequency components through which the state frequency memory can capture low frequency patterns for long-term forecast and high frequency patterns for short-term predictions.

2.6 Areas of open research and future directions

In spite of several benefits showcased by LSTM model, they also face challenges in predicting stock market fluctuations accurately due to issues like generalisability, model complexity and data sensitivity. One of the prominent areas of ongoing study is the use of hybrid model where combination of different neural network can be done with LSTM for leveraging their unique strengths. For example, the research by Althelaya et al. (2018) proposed a hybrid model whose integration is done with LSTM by employing attention mechanisms that showed enhanced performance of the model by focusing on crucial data points. One of the emerging areas is transfer learning as well where the adaptation of a pre-trained model is done to a new but related task which potentially reduces training time and improves accuracy (You et al., 2021).

As suggested by Ghosh et al. (2019), transfer learning effectively addresses issues related to data scarcity in the context of specific market which further enhances the applicability of LSTM model. The proposal of a hybrid model is given by Selvin et al. (2017) where integration of RNN, CNN and LSTM is done through a sliding window approach through which the model becomes capable of handling non-stationary financial data (Selvin et al., 2017). The study is highly critical in advance hybrid architectures for predicting stock market fluctuations and is also one of the foundational studies highlighting the effectiveness of combining different deep learning approaches. The use of a hybrid LSTM-RNN model is done by Pawar et al. (2019) where improvements are noted in the prediction accuracy for specific trends of stock prices (Pawar et al., 2019). The paper is published in Emerging Trends in Expert Applications and Security where valuable insights are provided by the research into the robustness and security of hybrid models which makes the study particularly relevant for financial institutions prioritising reliability and accuracy.

The research by Shah et al. (2019) addresses both the potential and limitations of ML models including LSTM for predicting stock market fluctuations. The inherent challenges are discussed in the study in accurately forecasting stock movements due to variable-rich and complex nature of financial markets. On one hand, the techniques of ML have potential for capturing trends, but it is also acknowledged in the study that there is fundamental difficulty in making reliable predictions due to the market randomness in the short-term scenarios especially (Shah et al., 2019).

2.7 Summary

It is indicated by the literature that LSTM model have significant potential for accurately predicting stock market fluctuations. The critical examination of the studies is provided through which it is evident that the sequential learning capabilities of LSTM model makes it highly suitable for financial forecasting. However, there are certain challenges particularly related to model tuning and data selection where future research may focus on transfer learning and hybrid approaches for further optimising the prediction accuracy and adaptability of the model.

There is significant evolution witnessed in stock market prediction which transitioned from traditional statistical models like GARCH and ARIMA to sophisticated deep learning approaches such as RNN, LSTM and hybrid architecture.

There is need for further development highlighted by the literature including hybrid models in which RNNs and CNNs are combined. The integration of sentiment analysis is done from social media and news along with by directional LSTM which collectively improves the reliability and performance of the LSTM model to predict stock price directions and changes. The example of Indian stock market exchange is provided by the author where robust performance is demonstrated by LSTM in spite of having unpredictable investor behaviour and high market volatility. It is emphasised by the gathered literature that there is utmost importance of feature engineering, using dimensional reduction techniques and integrating macroeconomic indicators to enhance the accuracy of model. The evaluation metrics like R-squared value and RMSE are adopted widely to benchmark the LSTM model against other neural networks. The inclusion of transfer learning models and hybrid learning models show an emerging direction in the research for addressing the issue of generalisability in LSTM. It is acknowledged by the researchers as identified in literature that there is need for improved model adaptability and interpretability to real time financial data.

3 Methodology

3.1 Overview

The methodological framework is outlined in this chapter which is adopted for achieving research objectives of implementing LSTM-based predictive model for stock market forecasting. A mixed method approach is applied in the study which integrates comprehensive literature review with quantitative experimental design focusing on development of LSTM model and its evaluation. The chapter entails data collection methods, process of model implementation, model evaluation, techniques, tools and ethical considerations adopted to conduct the research.

3.2 Research design

The research design is mixed methods approach in which qualitative and quantitative approaches are integrated that provide a comprehensive perspective on predicting stock market fluctuations using ML algorithms.

The extensive review of academic literature is done in the qualitative component of mixed methods approach for understanding prevailing techniques, financial indicators and ML models in stock market prediction. In the quantitative component, the designing, development, training and evaluation of LSTM model is done using python. The model evaluation is done using standard financial metrics like mean squared error (MSE), mean absolute error (MAE), accuracy, F1 score and recall. The empirical qualitative results and insights along with quantitative component analysis provide robust and data driven findings of stock market forecasting problem.

3.3 Research strategy

A comprehensive and systematic literature search strategy is adopted for this project for gathering high-quality and relevant research on stock market prediction using ML. The process involves different stages for ensuring the inclusion of robust studies where identification of influential authors is done. The research strategy allows for the ongoing refinement of the search approach. The research for conducting literature study began with gathering a few studies that were identified using preliminary searches using general terms like “stock market prediction using ML”, “Stock market prediction” and “financial forecasting models”. The employment of breadth first approach is done in the research initially where abstract is reviewed from the articles for determining the relevance of each paper without diving deeply into the description. The initial screening allowed broad perspective on terminologies and common approaches in the stock market prediction field. The terms that are encountered frequently within the abstract of the research papers as well as keywords are documented for refining the terms and improving the overall process of the search. The fundamental works and new research on stock market prediction is explored by using cited references within relevant papers. Iterative refinement of the search terms is done based on the insights gained from initial papers by applying filters, refining search results based on journal quality and citation count with which the literature study is completed.

3.4 Data collection

The data collection method adopted in the research is secondary. There is no involvement of active participants in the research, hence, secondary data collection is done in which peer reviewed articles, journals, industry reports and data repositories are gathered and analysed through systematic review. For the quantitative component of LSTM model development, the Mendeley data repository is scanned for gathering historical stock data of established organisations. The chosen data set is from Tesla where it includes trading volumes, daily closing prices, open, high and low prices along with dates ranging from 2010 to 2023. The selection of data set is done from Mendeley provided by Alakwah (2023) which is a credible data repository providing structured format and data availability. The chosen dataset reflects high volatility asset which is ideal for testing model performance. The selection of dataset is also on the basis of sufficient data points which is crucial for LSTM training and validation.

Mendeley Dataset: https://data.mendeley.com/datasets/c7r6ky4xgc/1 (Alakwah, 2023)

3.5 Model development and architecture

Long Short-Term Memory i.e., LSTM is recognised as a variant of RNN (recurrent neural networks). The algorithm is chosen for the model development of stock price prediction as it is able to handle long-term dependency in a time series data. The conventional and traditional models like linear regression and ARIMA struggle with the non-linear relationships in the data and assume data stationarity. The long-short term memory (LSTM) has memory gates which are able to handle complex patterns in the data effectively.

The development of the LSTM model is done using python with Keras and TensorFlow libraries. The model architecture includes two LSTM layers, input layer and a dense layer (output). The 60 sequences of normalised stock data are accepted in the input layer. The LSTM layers in the model have 50 neurons each and drop out layers for preventing the issue of over fitting. The dense layer is recognised as the final layer of output for predicting stock price.

3.6 Evaluation metrics

The evaluation of model performance is done objectively using multiple metrics including accuracy, recall, F1 score, mean squared error (MSE) and mean absolute error (MAE). The measurement of the average magnitude of errors is done by mean absolute error (MAE) without taking direction into account. The large errors are penalised more than mean absolute error by mean square error (MSE). The assessment of the prediction direction of the model is done by the accuracy metric. The F1 score of the model is analysed in which the balance of recall and precision is set which is crucial for directional predictions. The model's ability for correctly predicting a particular event or trend is measured by the recall score. Collectively, these metrics provide a comprehensive evaluation of the way model captures the direction and magnitude of stock price fluctuations.

3.7 Technologies and tools

Following tools and technologies are used for developing the LSTM model for predicting stock price fluctuations based on acquired Tesla's dataset.

Library/Tools	**Description**
Python	Python is the programming language used for implementing the LSTM model to predict stock price fluctuations.
TensorFlow/Keras	This is a python library used for developing model and training the model.

NumPy	The library is utilised for manipulating data as well as pre-processing.
Pandas	It is a library in python, which is used for numerical computation.
Matplotlib/Seaborn	This library is used for visualising data using different plots, charts, and graphs.
Scikit-learn	This library is used for splitting the data into training and testing set as well as metric evaluation.

3.8 Ethical considerations and limitations

The ethical research practices are maintained in the study where the publicly available data is used without acquiring any confidential or personal information from participants. The references and citations are placed properly in the research throughout the literature review for acknowledging prior research provided by authors. In addition to that, the model attempting to predict stock prices is created for research and educational purpose and not for any trading decisions or financial advice.

The stock data provided by Tesla is considered in data which limits the methodology as the model may not generalise well across other indices or stocks. The model training is done on past data which presents temporal limitations which does not consider future disruptive events like pandemic or economic crises.

4 Implementation

4.1 Overview

The practical implementation and model evaluation is outlined in the chapter of long short-term memory model (LSTM) for predicting stock price movements using the acquired Tesla dataset from Mendeley repository. The data range in the dataset is from 2010 to 2023 including different technical indicators like Bollinger bands, RSI, MACD etc along with traditional stock price values.

4.2 Data pre-processing

There are several features included in the raw dataset which includes basic stock data, technical indicators and target indicator. The basic stock data includes low, high, open, date, close and volume. The technical indicators include upper lower Bollinger bands, RSI, % D, % K, EMA 12, EMA 26, CCI, MACD, etc. The target indicator in the dataset is ‘buy’ or ‘sell’. There are different financial indicators included in the dataset playing a vital role in the performance of model. The RSI is recognised as relative strength index in which the momentum and oversold or overall conditions are indicated. The Bollinger band is a technical indicator which represents the standard deviation levels which are below and above of moving average which is highly crucial for volatility measurement.

Target Indicator: ‘buy’ or ‘sell’

Selection of Indicators: Bollinger bands, RSI, % D, % K, EMA 12, EMA 26, CCI, MACD

Reason for selection: %D and %K are Stochastic Oscillator which highlight the oversold or overboard conditions in the market. The EMA 12 and EMA 26 are recognised as exponential moving averages whose emphasis is on recent prices. They are vital indicators for computing MACD which is moving average convergence divergence. MACD highlights the trend following characteristics as well as momentum. The CCI is recognised as commodity channel index that shows warning of extreme conditions as well as new trends in data. ROC is the rate of exchange through which momentum is measured as the ratio of today’s price to previous price.

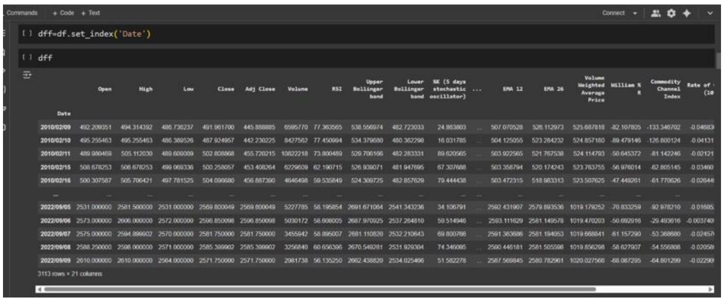

Figure 1: Data Reading

```
dff.isnull().sum()
```

	0
Open	0
High	0
Low	0
Close	0
Adj Close	0
Volume	0
RSI	0
Upper Bollinger band	0
Lower Bollinger band	0
%K (5 days stochastic oscillator)	0
%D Average(H,3)	0
EMA 12	0
EMA 26	0
Volume Weighted Average Price	0
William % R	0
Commodity Channel Index	0
Rate of Change (10 days)	0
Aroon Up	0

Figure 2: Finding Missing Values

```
dff.describe()
```

	Open	High	Low	Close	Adj Close	Volume	RSI	Upper Bollinger band
count	3113.000000	3113.000000	3113.000000	3113.000000	3113.000000	3.113000e+03	3113.000000	3113.000000
mean	945.129396	956.004267	933.666294	944.326296	922.772407	8.624042e+06	48.146444	993.229684
std	695.165823	703.202803	686.630715	694.620074	701.402362	5.796754e+06	16.466696	723.807091
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000e+00	2.965484	349.242536
25%	443.818054	448.251038	438.394470	443.917114	418.347961	5.316182e+06	35.942027	480.210912
50%	521.605713	527.945618	516.999390	521.729553	490.580292	7.095266e+06	48.595683	549.066522
75%	1273.924438	1284.821167	1263.077271	1269.813477	1252.394775	9.890146e+06	59.399001	1343.122350
max	2856.149902	2856.149902	2786.100098	2819.850098	2811.385742	6.584835e+07	97.768657	2929.900430

	Lower Bollinger band	%K (5 days stochastic oscillator)	%D Average(H,3)	EMA 12	EMA 26	Volume Weighted Average Price	William % R	Commodi Chann Ind
count	3113.000000	3113.000000	3113.000000	3113.000000	3113.000000	3113.000000	3113.000000	3113.0000
mean	886.743411	50.713781	50.684522	940.655356	936.086906	567.386371	-47.054140	10.4014
std	658.621631	30.641557	25.527142	690.204087	684.931987	167.515458	30.707229	114.8987
min	151.923808	0.000000	3.149804	302.954241	340.654569	444.571800	-100.000000	-687.9866
25%	417.953452	21.428499	28.725757	440.643422	438.774956	452.295320	-74.712631	-75.4015
50%	495.945063	51.582278	51.187024	517.420609	517.048920	493.680694	-44.403210	18.6009
75%	1192.090262	79.459333	72.777937	1267.594897	1269.707500	603.926486	-19.094670	95.9920
max	2571.997583	100.000000	98.950732	2733.644147	2663.272290	1020.027566	-0.253972	474.0268

Figure 3: Data Description

After pre-processing of dataset, it is split into two parts: training and testing set. The training samples in the dataset are 2472 while the test samples are 619. The loss value in each iteration is nearby 0.69 which is high. The two LSTM layers are added in the model and the dropout is implemented in between them for regularisation. The loss function here is the mean square error which is utilised for problem optimisation with Adam optimiser. Another metric used is mean absolute error (MAE) in the LSTM network which is associated with time series data.

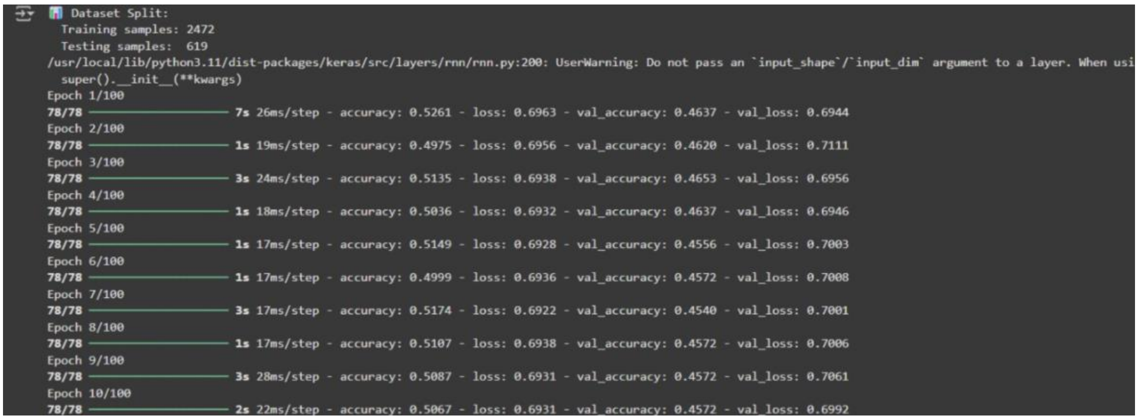
```
Dataset Split:
  Training samples: 2472
  Testing samples:  619
/usr/local/lib/python3.11/dist-packages/keras/src/layers/rnn/rnn.py:200: UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When usi
  super().__init__(**kwargs)
Epoch 1/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 7s 26ms/step - accuracy: 0.5261 - loss: 0.6963 - val_accuracy: 0.4637 - val_loss: 0.6944
Epoch 2/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 1s 19ms/step - accuracy: 0.4975 - loss: 0.6956 - val_accuracy: 0.4620 - val_loss: 0.7111
Epoch 3/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 3s 24ms/step - accuracy: 0.5135 - loss: 0.6938 - val_accuracy: 0.4653 - val_loss: 0.6956
Epoch 4/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 1s 18ms/step - accuracy: 0.5036 - loss: 0.6932 - val_accuracy: 0.4637 - val_loss: 0.6946
Epoch 5/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 1s 17ms/step - accuracy: 0.5149 - loss: 0.6928 - val_accuracy: 0.4556 - val_loss: 0.7003
Epoch 6/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 1s 17ms/step - accuracy: 0.4999 - loss: 0.6936 - val_accuracy: 0.4572 - val_loss: 0.7008
Epoch 7/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 3s 17ms/step - accuracy: 0.5174 - loss: 0.6922 - val_accuracy: 0.4540 - val_loss: 0.7001
Epoch 8/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 1s 17ms/step - accuracy: 0.5107 - loss: 0.6938 - val_accuracy: 0.4572 - val_loss: 0.7006
Epoch 9/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 3s 28ms/step - accuracy: 0.5087 - loss: 0.6931 - val_accuracy: 0.4572 - val_loss: 0.7061
Epoch 10/100
78/78 ━━━━━━━━━━━━━━━━━━━━ 2s 22ms/step - accuracy: 0.5067 - loss: 0.6931 - val_accuracy: 0.4572 - val_loss: 0.6992
```

Figure 4: Dataset splitting and Loss Value in test LSTM model

4.3 Model architecture and hyperparameter tuning

4.3.1 LSTM model architecture

The architecture of LSTM model is provided which shows the different layers including input layer, LSTM layers 1, 2 & 3 and an output layer (refer fig 5). A basic structure is employed in the LSTM model implementation having 2 LSTM layers and dropout regularisation (ref fig 6 & 7).

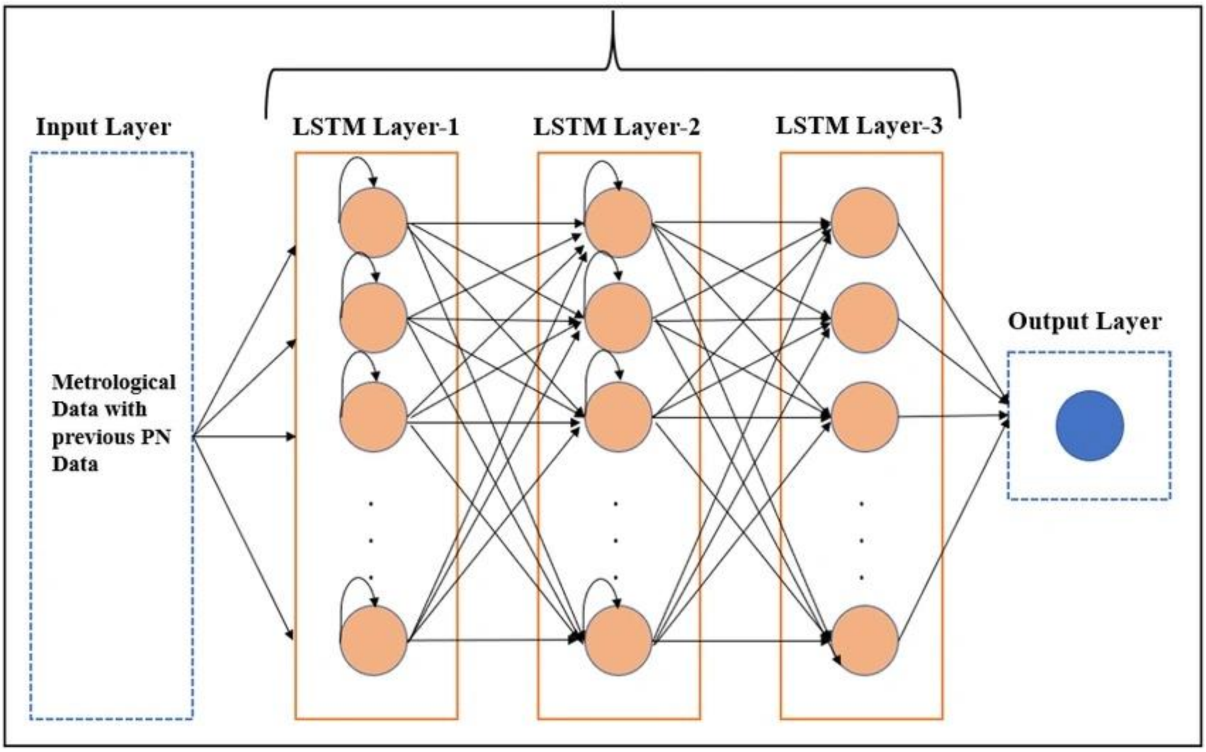

Figure 5: LSTM Architecture

(Source: Sai, 2024)

Talha Risath
22222895

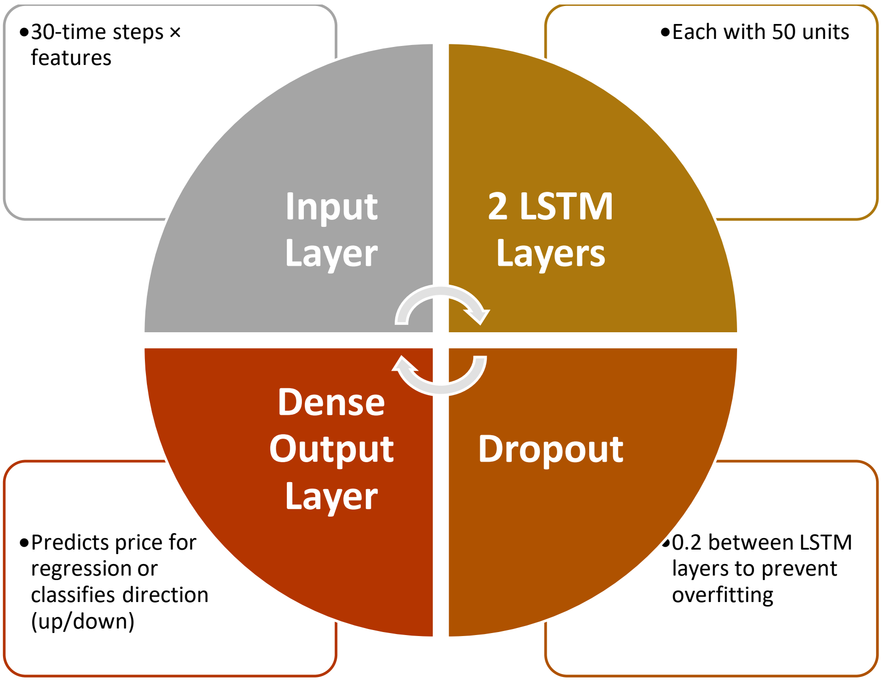

Figure 6: LSTM Layers

```
/usr/local/lib/python3.11/dist-packages/keras/src/layers/rnn/rnn.py:200: UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When using Sequenti
  super().__init__(**kwargs)
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 60, 50)	10,400
lstm_3 (LSTM)	(None, 50)	20,200
dropout_2 (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 1)	51

```
Total params: 30,651 (119.73 KB)
Trainable params: 30,651 (119.73 KB)
Non-trainable params: 0 (0.00 B)
```

Figure 7: Insights of LSTM Layers

4.3.2 Hyperparameter Tuning

The hyperparameters are recognised as value sets which are set before the model training begins whose impact is seen on the way model performs and learns from past trends and data. The common hyperparameters include batch size, Epochs, learning rate, dropout, optimiser, sequence length and num_units. Epochs is the number of times dataset is passed through the model entirely. The number of processed samples in the model before updating model weights is set through batch_size.

The first phase in model implementation is completed i.e., pre-processing where data is analysed and cleaned before feeding it into the LSTM model. The Null values or missing values in the dataset were indicated and dropped using forward fill method. The feature selection in the dataset is performed by keeping only relevant features like volume, close, open and date. Normalisation of data is done in data pre-processing between 0 and 1 using the MinMaxScaler for improving its training and efficiency. The sliding window of 60 days was created earlier but then revised to 30 days to form sequence which is used by the model for predicting the next closing price of stock. The below pie chart shows the set values for parameters like Epochs, Number of LSTM Layers, Neurons per layer, learning rate of the model, batch size and dropout (Ref. fig 8).

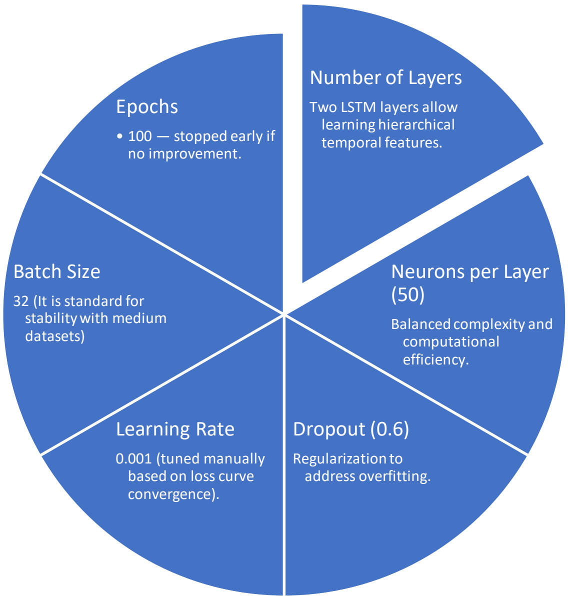

Figure 8: Hyperparameter tuning stats

4.4 Training History and Analysis of Overfitting

The generalisation capacity and learning behaviour of the model is assessed using the training and validation loss curve generated for both classification and regression tasks (ref fig 9 & 10).

Regression Loss over Epochs: The training loss is indicated by the blue line which is gradually decreasing across 100 epochs highlighting that the model learns from the training data. The declining validation loss is indicated by the orange line which starts at a higher level as compared to the training loss and it also reduces at a slower rate. Moderate overfitting is suggested by the persistent gap existing between validation and training loss. This means that the model is fitting the training data well but is struggling to generalise to unseen data. There is high variance in data as the scale of MSE loss is nearby 106 which may contribute to poor R2 square score.

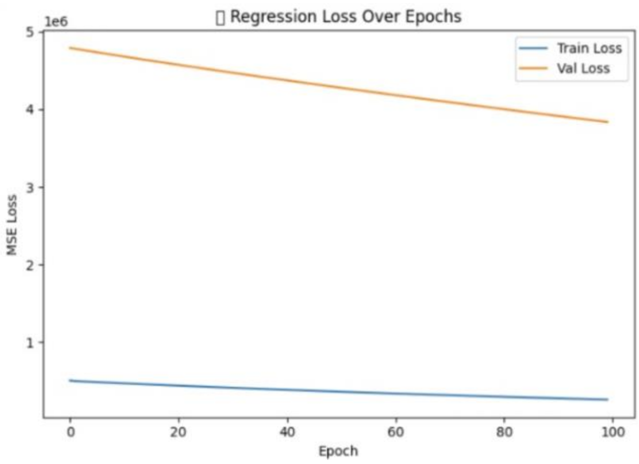

Figure 9: Regression Loss over Epochs

Classification Loss over Epochs: A relatively stable train loss is indicated by the blue line in the case of classification loss over epochs. A **volatile validation loss** is indicated by large spikes in orange color (ref fig 10). This shows that the model is either unstable or is prone to overfitting. It is suggested by these fluctuations that the model does not generalize well and may be extremely sensitive to small variations in validation data which can be because of **small batch size** or **class imbalance** in dataset.

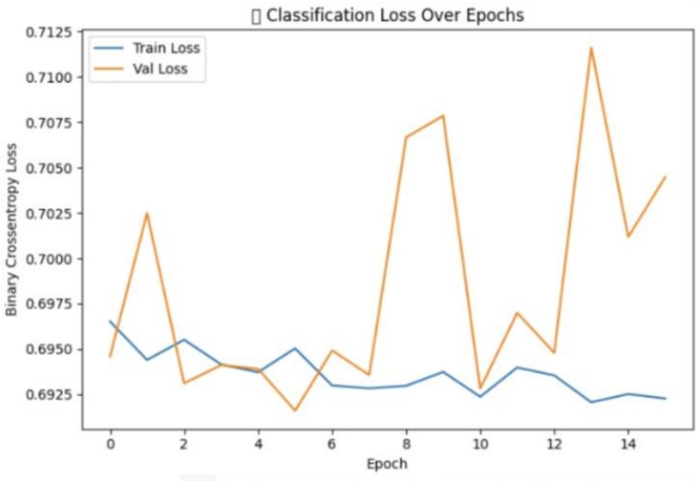

Figure 10: Classification Loss over Epochs

4.5 Regression task analysis

The model is refined for performing regression analysis where earlier the MAE was 2035.69 which is which adjusted to 785.83. The RMSE score is 877.0121 which earlier was 2070.57 before tuning the model. The R^2 improved from -28.93 to -4.7093.

```
REGRESSION METRICS:
RMSE: 877.0121
MAE: 785.8316
R² Score: -4.7093
R² should be higher now due to deeper structure and tuned regularization.
```

Figure 11: Regression Metrics

4.6 Classification task analysis

Earlier before tuning the model, the accuracy of the model was at 47.33%, precision at 53.10%, f-1 score at 26.91% while the recall at 18.02%. After fine tuning the model, the classification results show the recall score of 100%, precision and accuracy both at 53.98% and F1 score at 70.12%

```
CLASSIFICATION METRICS:
Accuracy:  0.5398
Precision: 0.5398
Recall:    1.0000
F1 Score:  0.7012
```

Figure 12: Classification Metrics

Table 1: Classification report

Class	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	283
1	0.54	1.00	0.70	332

The identification of class 1 which indicates upward movements in stock is done perfectly by the model, but it does not perform well for class 0 (downward movements). For full trading decisions, the model is unreliable to predict the down movements in stock because the recall score is excellent for class 1 predictions but not optimum for class 0.

4.7 Visualizations: role of technical indicators

In spite of the poor performance of the model (especially in regression), the visuals are provided below for LSTM classification, LSTM regression, insights of stochastic indicators, EMA/MACD to understand the decisions and interpretability of the model.

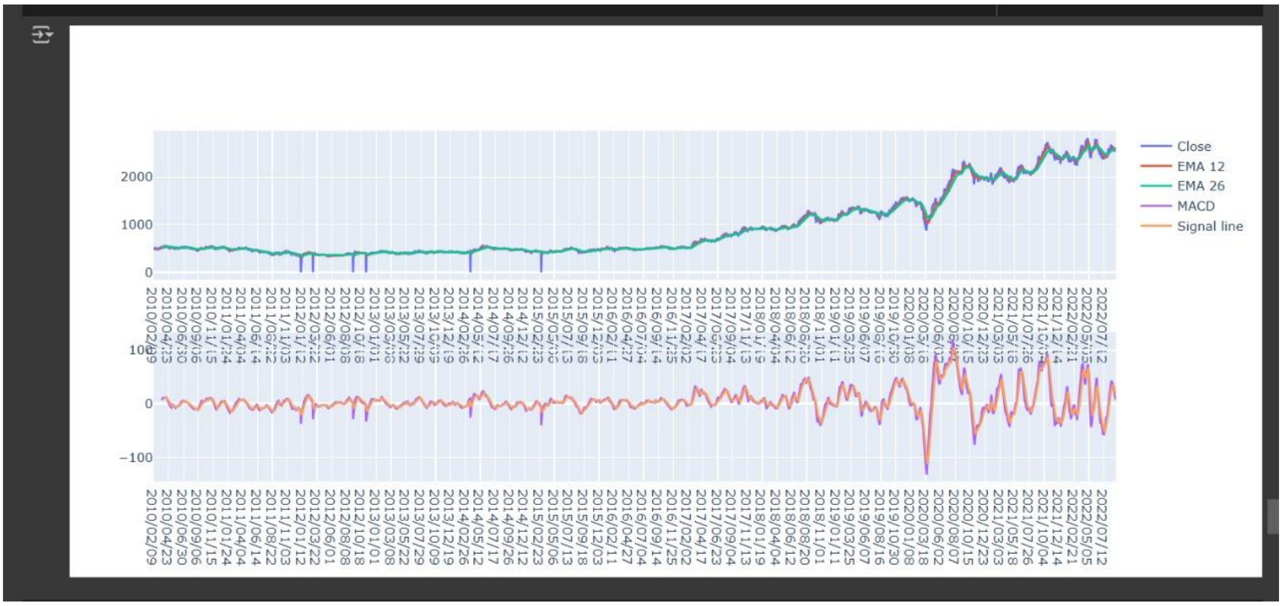

Figure 13: EMA/MACD Analysis

The actual versus predicted plot is provided in the regression chart suggesting divergence over time where the model is supposed to be either overestimating or underestimating the trends. There is a clear mismatch shown in the classification chart between predicted and actual price value which is denoted by different markers that indicates that LSTM shows good generalisation of actual versus predicted values.

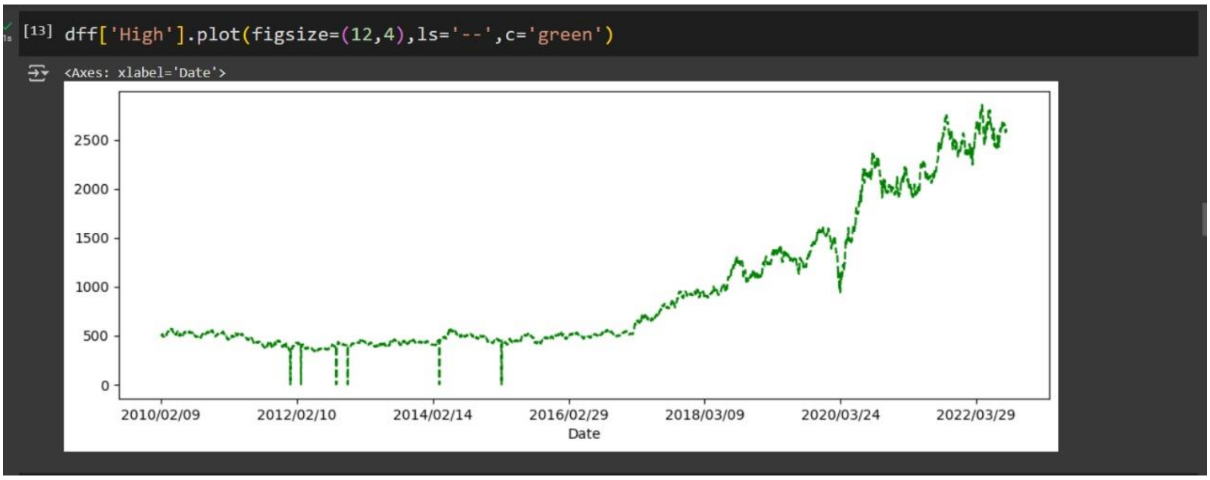

Figure 14: Stock movement analysis for 'High' Value

The EMA/MACD graphs are plotted indicating that momentum indicators are referenced in the model for trend reversals. However, it does not align well with the actual price movement and is not sufficient for improving predictive accuracy in a significant manner.

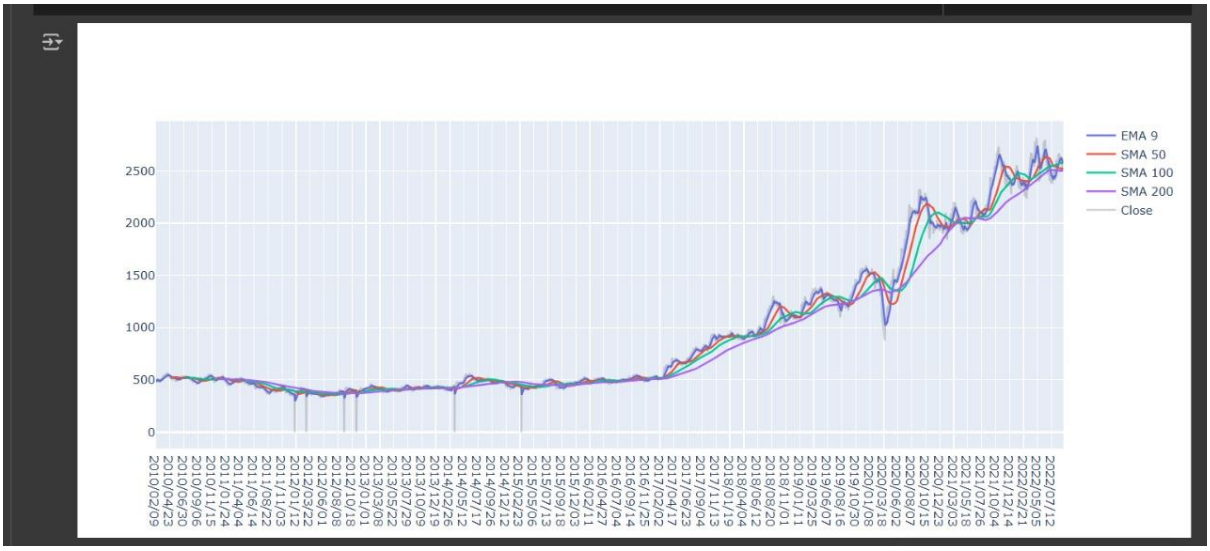

Figure 15: EMA/SMA 50/SMA100 and close value insights

The oversold and overbought conditions are detected using the stochastic oscillators providing a useful edge in manual trading. The %K and %D stochastic oscillators behaviour is illustrated through the provided graph. These indicators are highly used in technical analysis through which the momentum of price movements is measured.

The faster moving oscillator is the %K line which is represented in blue colour while the %D line is a moving average of %K showcased in red colour. The fluctuation of both oscillators is between 0 and 100 which offers insight into the oversold and overbought condition in the market. The chart shows that the %K line is reacting more sensitively to the price fluctuation which is crossing below and above the %D line serving as a signal for potential price reversal.

For example, when the %K line sharply passes through above the 80 threshold and below the %D line, it signals that there is a potential selling opportunity or downward correction present. As opposed to that, when there is a dip witnessed in %K line below the 20 level and above the %D level, there may be a buying opportunity present. The recurring pattern is shown in the chart for both lines multiple times in the timeline which suggests that a cyclical momentum is exhibited by the market. The analysis of oscillators adds comprehensive depth to the forecasting approach as it supplements price prediction with momentum-based insight through which the interpretability of the model is enhanced.

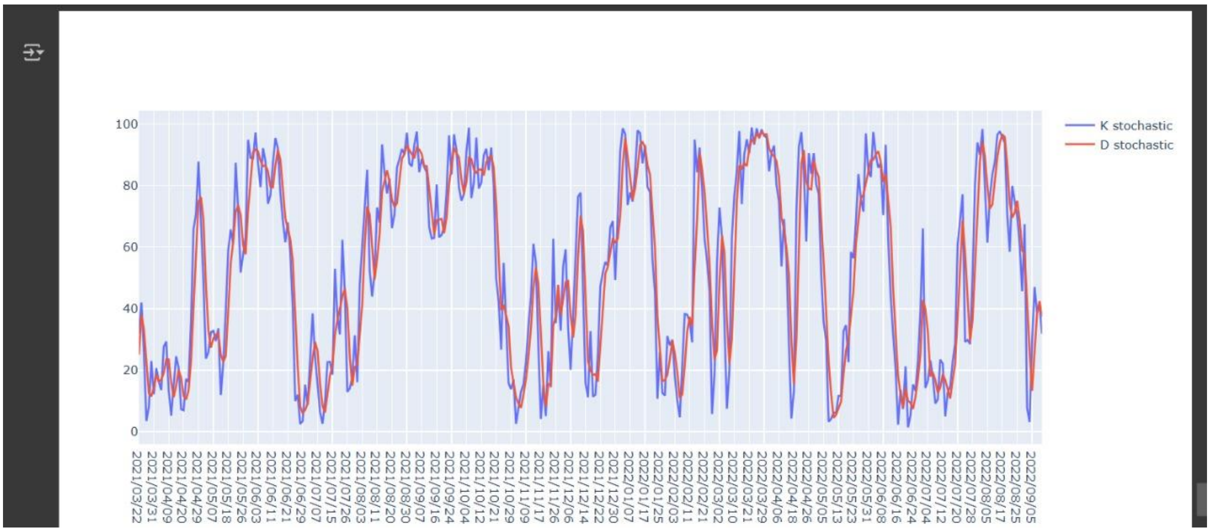

Figure 16: K stochastic and D stochastic oscillators Analysis

4.8 Summary of technical indicators analysis

Various momentum and trend indicators are overlaid in the visualisation of EMA and MACD analysis. The price trend of Tesla stock is presented in the top section of EMA and MACD analysis over time along with EMA 26 and EMA 12 curve. The price over 12 and 26 days are made smooth by these EMA respectively. This EMA crossover is utilised traditionally for detecting bearish or bullish trends. The crossover sight of EMA is analysed in which it is seen that when the EMA 12 crosses above EMA 26, an upward momentum signal is indicated by it whereas the reverse suggests a bearish trend (ref fig 11). The plot significantly highlights these crossovers but due to the presence of market volatility, consistent alignment is not visible with significant price movements.

The signal line is plotted below the main chart and MACD line highlight the difference between EMA 12 and EMA 26 along with its smoothed version. The histogram of MACD is showcasing the difference which aids in visualising the shift in momentum. When the MACD crosses the signal line, it implies an upward momentum and vice versa. However, the fluctuation seems noisy and in spite of having certain spikes aligning with real market movement, there exist several false signals which dilute the predictive strength of the model. Even after being guided by these technical indicators, the model shows inconsistency and fails to provide reliable forecast (ref fig 13). It is interpreted that the model is reacting to short-term noise instead of sustained trends.

This stock movement is analysed for the high value using MACD and EMA overlays where the primary target is to understand if the model is incorporating price peaks in the trend forecast. There is consistent misalignment showcased between the indicator driven signals which are MACD & EMA and actual high value trajectory showing that in spite of the model referencing these indicators, they are not enhancing accuracy (ref fig 13). The curve of EMA is lagging while the MACD divergence is not reliably coinciding with the actual peaks. Hence, it is interpreted that the high-value predictions are extremely sensitive to the intraday volatility and market anomalies which limits the effectiveness of standard indicators when they are utilised in a standalone manner in LSTM framework.

The stochastic oscillators are analysed which are utilised commonly in technical analysis for assessing oversold and overbought market conditions. The blue line is of % K oscillator which is moving in a fast manner and reacting sharply to the price changes. The % D line is the red line which highlights the moving average of percent K and acts as a smoothing signal. The overbought conditions are signalled by the readings above 80 while the values under 20 are indicating oversold zones. The overboard conditions have the potential for downward correction while the oversold conditions have the potential for upward bounce. The momentum shifts can be identified effectively by human traders by this crossover as per Pal (2024) but the abrupt oscillations cannot be interpreted effectively by the LSTM model due to the reactive nature of % K line and noise in dataset.

To summarise, it is highlighted by these visualisations that on one hand, the technical indicators effectively improve the interpretability of the model, but the accuracy still is not guaranteed. In the quantitative metrics, it is learnt that the LSTM model underperformed but it is illustrated by these figures (from fig 13 to fig 16) that the model attempted to learn from patterns suggested by stochastic oscillators, MACD, and EMAs. There exists inconsistency between the real price movement and indicators, but it points to the drawbacks of utilising only traditional indicators for predictive modelling in volatile financial markets like Tesla.

4.9 Benchmark comparison with existing ML models

The updated model of LSTM performs well when compared to linear regression and ARIMA in terms of RMSE and MAE. However, the R Square is still negative, but drastic improvement is seen in its value from previous considerations that reflects reduced error variance and improved pattern learning.

Table 2: Benchmark comparison

Model	**MAE**	**RMSE**	**R^2 Score**
Linear Regression	1894.20	1982.33	-21.50
LSTM (Updated)	**785.83**	**877.01**	**-4.71**
ARIMA	1740.15	1850.41	-19.10

Note: ARIMA is not implemented in this study directly but the benchmark values and scores of MAE, RMSE and R^2 are drawn from ARIMA results reviewed in comparative studies in literature by Karmiani et al. (2019).

4.10 Summary

The practical implementation of LSTM (long short-term memory) networks is outlined in this chapter for predicting the stock prices of Tesla using data from 2010 to 2022. The dataset is sourced from Mendeley repository which includes technical indicators such as RSI, EMA, MACD, Stochastic oscillators, etc and traditional stock indicators like close, open high, low and volume. The primary aim of the LSTM implementation is to predict the movement in stock prices and directional trend using the regression and classification tasks. The process began with data pre-processing where noisy and irrelevant data is eliminated. The duplicate and missing values are handled using preprocessing and data normalisation is done using MinMaxScaler. The MACD and EMA indicators are retrieved from the price series that show insights into the potential reversals and momentum. The momentum-based movement prediction is further supported by the stochastic indicators. The data points collectively ensure that the implemented model is receiving output which reflects more than only price history i.e., it reflects sentiment and market psychology as well.

5 Results and Findings

5.1 Qualitative data analysis

The deep learning techniques are utilised for forecasting stock prices such as LSTM networks which is a significant topic of interest in quantitative analysis and financial research. The ability of LSTM network to learn long-term dependency in sequential data makes it ideal for time series prediction task such as stock forecasting. However, it is revealed by the real-world application of LSTM that there are certain challenges and limitations of LSTM especially in volatile domains like finance.

The LSTM regression is one of the key components in the analysis in which the prediction of actual future price of a stock is done based on the historical data (Karmiani et al., 2019). A time series graph is presented in the LSTM regression showing actual versus predicted values where one represents the actual closing price, and the other is the prediction of model. A certain level of stationary is assumed in the data by traditional statistical models like linear regression and ARIMA which is Auto Regressive Integrated Moving Average which highlights that the statistical properties like variance and mean remain constant with time. However, stock prices are non-stationary which exhibit seasonality and trends where LSTM algorithms are highly equipped for handling these non-linear relationships in the data (Karmiani et al., 2019). The traditional statistical models were efficient and capturing trends, but they could not account for sudden changes and long-term dependences (Karmiani et al., 2019). But the problem is solved through LSTM as they are able to learn the complex patterns in the data effectively. Hence, the research employees the LSTM algorithm over traditional statistical algorithms to predict stock market prices by taking the Mendeley dataset on Tesla stock market (dataset retrieved from Alakwah, 2023).

In the initial phases, the efforts of prediction relied heavily on linear models which did not capture the non-linear and complex behaviour of financial market. But with the rise in deep learning and ML algorithms especially in LSTM network, a more effective means was found by researchers to handle long-term memory and sequential dependency in time series data. The studies provided by authors in literature illustrate the way LSTM networks can outperform the regression models by identifying temporal patterns which are not typically identified by traditional models.

There is a huge impact of market factors on stock market prices and at the same time, regression models do not have the capacity to learn or remember from past sequences in an effective manner (Parmar et al., 2018). This is one area according to Parmar et al. (2018) where LSTM models perform better than regression algorithms because LSTM being a type of recurrent neural network are designed to remember long-term dependencies and handle sequential data (Parmar et al., 2018). As opposed to recurrent neural network which struggle with the gradient problem in their training phase, specialised units are included in LSTM networks known as memory cells which have the power to retain information over a significant period of time (Parmar et al., 2018). The cells in LSTM network have gates recognised as forget gate, output gate and input gate through which the flow of information is regulated. The ability of LSTM networks to effectively store information and selectively pass or forget information make them highly reliable and suitable for forecasting tasks such as stock market prediction (Parmar et al., 2018).

5.2 Results of model implementation

A substantially better error metric is achieved by the newer version of the model because as compared to the earlier model, the R square is improved significantly from -28.93 to -4.7093. However, even in the new version of the model, the value of R square is still negative which indicates that model is not outperforming a simple mean predictor but the changes in the values of RMSE and MAE show up to 50% improvement which reflects that the model is learning from data patterns.

From the results analysed for LSTM model performance, it is seen that the model resulted in moderate performance which may be due to the insufficient or noisy features present in dataset, overfitting issue in the dataset or normalisation issues. From the secondary research conducted in literature, it is learnt from Liu (2019) that the LSTM models are highly prone to the issue of overfitting when their training is done on volatile data and inadequate regularisation is done (Liu, 2019). This impacts the performance of the model and results in negative R squared value.

However, the benchmark comparison performed between LSTM, linear regression and ARIMA indicates that the updated LSTM model outperforms both ARIMA and linear regression in terms of MAE and RMSE. The new R Square score of LSTM model is -4.71 which shows significant improvement from its previous score i.e., -28.93. Better dropout regularisation and deep LSTM layers helped the model in generalising better to unseen data. The lookback window is adjusted to 30 days which allows more focus on pattern learning. Short-term trends can be captured more effectively by the model after improvement as compared to broad linear models. Hence, the potential of deep learning models such as LSTM is reaffirmed by these results especially when the model is tuned properly through which it will be capable to outperform traditional statistical models in non-linear and volatile financial environment. The identification of market momentum is done by RSI while the MACD and Bollinger band highlights trend's strength and market volatility. The stochastic indicators % D and % K contribute to identifying oversold or overbought conditions.

Figure 17: Regression actual vs Predicted for close price

Figure 18: LSTM Regression: Actual vs predicted

One potential reason behind the poor performance of the model may be due to noisy features or insufficient features present in dataset as per Seiffert et al. (2014). There are several technical indicators included in the acquired dataset of Tesla, but it is assessed that they are not always useful in predicting future prices because of market randomness. Proper scaling in the data would have improved the model performance, especially for LSTM input which shows that there could be the data normalisation issues in the dataset.

5.3 Discussion

Drawing from the research analysed in literature by Karmiani (2019) and Parmar (2018), the LSTM model is implemented as the predictability of up and down movements in stock is better as compared to linear regression and ARIMA. The model shows better predictive power from the previous model where the R square value is -4.71 which is changed from -28.93. A strong support is lent by the R Square value to LSTM based model in financial time series forecast. The strong generalisation capability is demonstrated by the model performance due to the closeness of actual price versus predicted price value across 12-year period.

It is not necessary that the high value of R square means that the model is perfect, and it will perform effectively in the real time prediction. A wide range of factors influence the financial market including geopolitical tensions, macroeconomic indicators and psychological factors which are not completely captured by the historical price data alone. A remarkable accuracy is showed by the model historically, but the future predictions must be done by continuously retraining the model with new data.

5.4 Implementation gap and limitations

There are several gaps found in implementation particularly related to feature selection, model performance and data variability. In spite of substantial research into stock market forecast using deep learning and ML techniques, the LSTM approach is used to predict stock price movement of Tesla using an extensive dataset ranging from 2010 to 2022. The dataset contained different technical as well as traditional stock metrics and indicators.

One such **research gap** found in the study lies in feature selection and data granularity. There are several studies using basic stock indicators such as high, low, close, open and volume as per He (2023), but only a few include a comprehensive list of technical indicators like %D percent, %K, RSI, MACD and CCI especially for LSTM-based forecasting. The incorporation of these technical indicators is done in the study into the feature set, but the inclusion of technical indicators does not always show enhanced performance of the model as also indicated in the result (chapter 4). The regression and classification reports of LSTM model do not show optimum performance which are due to having unfiltered and noisy dataset.

An innovative attempt is provided by the LSTM model at direction and price prediction, but there are several limitations which existed during implementation which impacted the performance of the model. One such limitation is the overfitting issued due to volatile data. There were signs of overfitting in the LSTM model which is evident from high training accuracy but poor performance in test set especially in regression task. In the LSTM model, overfitting is a known issue especially when highly volatile financial time series data is handled. The dataset of Tesla is highly speculative, and growth oriented which becomes prone to unpredictable drops and spikes making it difficult to generalise the model effectively.

There are variety of technical indicators used in the model implementation, but their contribution has not been in the positive direction. The indicators like Bollinger band, RSI and MACD exhibit redundancy or correlation which adds noise instead of improving the clarity of signal. In addition to that, there are several features which did not have predictive relevance to the unique price behaviour of Tesla which led to negative R Square score of - 4.7093. The implementation approach adopted in this research is dual task approach for stock forecasting which includes classification and regression both. However, most literature show isolation of these tasks where they either predict the directional movement or predict the prices separately. However, this research attempted to perform both tasks within a single LSTM model architecture which presented novel challenges in training the model and generating optimum results which limited the performance of the model. Hence, the future research will include implementation of hybrid model by combining LSTM with CNN for better model predictability and performance.

6 Conclusion and Future Work

6.1 Conclusion

In the end, important questions are raised by the results from LSTM model about the application of deep learning in stock price forecast or other financial forecasting applications. It is learnt from the empirical evidence analysis in literature review that LSTM is capable for capturing long-term dependencies and sequential data. However, the real-world application of the model to highly noisy, volatile and speculative assets like Tesla stock is difficult. Based on the metrics of regression task, the mean absolute error is 785.83, the **root mean squared error** is **877.01**, the **R squared** score is **-4.71**. It is interpreted clearly by the results that the model did not perform effectively in predicting the future prices of Tesla's stock from the acquired dataset (especially for upward movement as indicated by classification report i.e., recall score for class 0). The R-square value or the coefficient of determination ranges mainly between 0 and 1 for positive model performance and on the other hand, the model showing negative R squared value highlights that it failed to analyse any meaningful signal, and the average performance of the model may be due to noisy features in the dataset.

The model outcome is concerning as it reflects that the LSTM model could not outperform a baseline model such as mean prediction. It can also be assumed from model performance that the stock prices especially for organisations like Tesla have random walk characteristics along with which they have strong non-stationarity which cannot be captured effectively by even LSTM model without having proper transformations. But on the other hand, the long-term plot showing actual versus predicted value for close and open price indicate high R Square value which shows that when model evaluates on different segments or longer horizons of data, it may capture broad price trends. Hence, it is signified from this outcome that the model could be effective for long-term forecast, but not as reliable for short-term predictions which is a common problem identified in deep learning when it comes to time series forecast.

However, when it comes to the classification task performance, the model shows the accuracy of 53.98%, F1 score of 53.98%, recall score of 100% (for class 1) while the f1 score is 70.12%. The model does not correctly classify the movement of stock whether when it comes to predicting down movement. On the other hand, the prediction of up movements (class 1) is done effectively by the clarification model.

In addition to that, the stochastic oscillators i.e., %K and %D are analysed visually which show that a clear buy/sell signal are provided by the momentum-based indicators than LSTM output. It is suggested by this outcome that the rule based, or manual strategies may outperform the poorly trained model from LSTM in certain contexts. The complexities of applying deep learning algorithms in the real-world financial applications are identified through the research where practical implementation of LSTM model is done on Tesla's stock market dataset. The LSTM networks on one hand show potential in identifying trends and retaining memory, but the results of both classification and regression task highlight average model performance with low prediction, accuracy, F1, recall score and generalisation for downwards movement of stock. There is need for enhanced data pre-processing witnessed after analysing the findings. In addition to that, there is need for optimised model architecture and advanced feature selection along with which the hybrid modelling approaches and external market sentiment must be considered in future study for developing more actionable and accurate stock price forecasting model.

6.2 Future Work

There are several limitations identified in the study findings based on which the future work will involve improvement in the LSTM based stock prediction model. The feature set will be refined using correlation analysis, mutual information ranking or principal component analysis (PCA) which will help in eliminating the noisy indicators. The informative feature set will improve the accuracy of model significantly and also reduce the issue of over fitting (Gupta, 2019). Another recommendation for improving the model performance in future is hyperparameter tuning with random search or grid search (Buslim et al., 2021). The systematic tuning methods will be adopted such as grid search so that the optimal LSTM parameters could be discovered which will improve the accuracy of both classification and regression task. It includes tuning number of neurons, dropout rate, sequence length and activation functions.

The external factors and sentiment analysis will be incorporated in the model which will improve the prediction accuracy along with contextual understanding. There is a huge impact of social media sentiments, news and macroeconomic variables on stock price movement apart from historical data and financial indicators (Khan et al., 2024). Hence, by integrating sentiment scores from financial news APIs and Twitter, the model performance will be improved. The future work will also employ ensemble models in which LSTM model will be combined with other models like CNN, XGBoost or ARIMA for yielding better results. The time series memory of LSTM can be leveraged effectively by hybrid approaches for improving the accuracy and robustness of the model.

7 References

7.1 Bibliography

Alakwah, Hossam (2023), “Tesla stock price prediction”, Mendeley Data, V1, doi: 10.17632/c7r6ky4xgc.1

Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018, April) Evaluation of bidirectional LSTM for short-and long-term stock market prediction. *In 2018 9th international conference on information and communication systems (ICICS)* (pp. 151- 156). IEEE. https://ieeexplore.ieee.org/abstract/document/8355458/

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552-567. https://www.sciencedirect.com/science/article/pii/S106294081730400X

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022) Predicting stock market index using LSTM. *Machine Learning with Applications*, 9, 100320. https://www.sciencedirect.com/science/article/pii/S2666827022000378 Ghosh, A., Bose, S.,

Buslim, N., Rahmatullah, I. L., Setyawan, B. A., & Alamsyah, A. (2021, September). Comparing bitcoin's prediction model using GRU, RNN, and LSTM by hyperparameter optimization grid search and random search. In *2021 9th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/9588947/

Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019, September). Stock price prediction using LSTM on Indian share market. In *Proceedings of 32nd international conference on* (Vol. 63, pp. 101-110). https://www.researchgate.net/profile/Giridhar-Maji/publication/346450872_Stock_Price_Prediction_Using_LSTM_on_Indian_Share_Market/links/5fc3142b458515b797843e38/Stock-Price-Prediction-Using-LSTM-on-Indian-Share-Market.pdf

Gupta, S., & Gupta, A. (2019). Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Computer Science*, *161*, 466-474. https://www.sciencedirect.com/science/article/pii/S1877050919318575

He, F. (2023). The best combination of indicators for the US stock market. https://www.ubishops.ca/wp-content/uploads/he20240105.pdf

Hiransha, M. E. A. G., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018) NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362. https://www.sciencedirect.com/science/article/pii/S1877050918307828

Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, 129, 273-285. https://www.sciencedirect.com/science/article/pii/S0957417419301915

Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A prediction approach for stock market volatility based on time series data. *IEEE Access*, *7*, 17287-17298. https://ieeexplore.ieee.org/abstract/document/8626097/

Jiang, W. (2021) Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, 184, 115537. https://www.sciencedirect.com/science/article/pii/S0957417421009441

Karmiani, D., Kazi, R., Nambisan, A., Shah, A., & Kamble, V. (2019, February). Comparison of predictive algorithms: backpropagation, SVM, LSTM and Kalman Filter for stock market. In *2019 amity international conference on artificial intelligence (AICAI)* (pp. 228-234). IEEE. https://ieeexplore.ieee.org/abstract/document/8701258/

Khan, M. A., Ali, H., Shabbir, H., Noor, F., & Majid, M. D. (2024). Impact of Macroeconomic Indicators on Stock Market Predictions: A Cross Country Analysis. *Journal of Computing & Biomedical Informatics*, *8*(01). https://www.jcbi.org/index.php/Main/article/view/740

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, 11 news. *Journal of Ambient Intelligence and Humanized Computing*, 1-24. https://link.springer.com/article/10.1007/s12652-020-01839-w

Lakshminarayanan, S. K., & McCrae, J. P. (2019, July). A Comparative Study of SVM and LSTM Deep Learning Algorithms for Stock Market Prediction. In *AICS* (pp. 446-457). http://ceurspt.wikidata.dbis.rwth-aachen.de/Vol-2563/aics_41.pdf

Liu, Y. (2019). Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. *Expert Systems with Applications*, *132*, 99-109. https://www.sciencedirect.com/science/article/pii/S0957417419302635

Moghar, A., & Hamiche, M. (2020) Stock market prediction using LSTM recurrent neural network. *Procedia computer science*, 170, 1168-1173. https://www.sciencedirect.com/science/article/pii/S1877050920304865

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (2020) Deep learning for stock market prediction. Entropy, 22(8), 840. https://www.mdpi.com/1099- 4300/22/8/840

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007-3057. https://link.springer.com/article/10.1007/s10462-019-09754-z

Pal, J. (2024). Long Short-Term Memory Pattern Recognition in Currency Trading. *arXiv preprint arXiv:2403.18839*. https://arxiv.org/abs/2403.18839

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020) An innovative neural network approach for stock market prediction. *The Journal of Supercomputing,* 76, 2098-2118. https://link.springer.com/article/10.1007/S11227-017-2228-Y

Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018, December). Stock market prediction using machine learning. In *2018 first international conference on secure cyber computing and communication (ICSCCC)* (pp. 574-576). IEEE. https://ieeexplore.ieee.org/abstract/document/8703332/

Pawar, K., Jalem, R. S., & Tiwari, V. (2019) Stock market price prediction using LSTM RNN. In Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018 (pp. 493-503). *Springer Singapore*. https://link.springer.com/chapter/10.1007/978-981-13-2285-3_58

Reddy, V. K. S., & Sai, K. (2018). Stock market prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, *5*(10), 1033-1035. https://www.academia.edu/download/57764282/IRJET-V5I10193.pdf

Sai, P. (2024). Complete Guide to Learn LSTM Models: Types, Applications, and When to Use Which Model. *Medium*. https://blog.stackademic.com/complete-guide-to-learn-lstm-models-types-applications-and-when-to-use-which-model-f9b779f31714

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Folleco, A. (2014). An empirical study of the classification performance of learners on imbalanced and noisy software quality data. *Information Sciences*, *259*, 571-595. https://www.sciencedirect.com/science/article/pii/S0020025511000065

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017, September) Stock price prediction using LSTM, RNN and CNN-sliding window model. *In 2017 international conference on advances in computing, communications and informatics (icacci)* (pp. 1643-1647). IEEE. https://ieeexplore.ieee.org/abstract/document/8126078/

Shen, J., & Shafiq, M. O. (2020) Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of big Data*, 7, 1-33. https://link.springer.com/article/10.1186/s40537-020-00333-6

Ta, V. D., Liu, C. M., & Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. *Applied Sciences*, 10(2), 437. https://www.mdpi.com/2076-3417/10/2/437 12

You, K., Liu, Y., Wang, J., & Long, M. (2021, July). Logme: Practical assessment of pretrained models for transfer learning. *In International Conference on Machine Learning* (pp. 12133-12143). PMLR. https://proceedings.mlr.press/v139/you21b.html

Zhang, L., Aggarwal, C., & Qi, G. J. (2017, August). Stock price prediction via discovering multi-frequency trading patterns. *In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2141-2149). https://dl.acm.org/doi/abs/10.1145/3097983.3098117

Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert systems with applications*, 67, 126-139. https://www.sciencedirect.com/science/article/pii/S0957417416305115

This is the Notebook with the code for this project:

https://colab.research.google.com/drive/16Wx6YrrXq_oEh2nmsAHk-21J9oNqs79V?usp=sharing

8 Appendices

8.1 Project Timeline

Name	Duration	Start	Finish
⊟**STOCK MARKET PREDICTION**	166 days	**9/10/24 8:00 AM**	**28/5/25 5:00 PM**
⊟**Proposal Development**	17 days	**9/10/24 8:00 AM**	**31/10/24 5:00 PM**
Deciding aim and objectives	10 days	9/10/24 8:00 AM	22/10/24 5:00 PM
Purpose of research	5 days	23/10/24 8:00 AM	29/10/24 5:00 PM
Scope of research	1 day	30/10/24 8:00 AM	30/10/24 5:00 PM
Task distribution	1 day	31/10/24 8:00 AM	31/10/24 5:00 PM
⊟**Conduct Literature Review**	35 days	**1/11/24 8:00 AM**	**19/12/24 5:00 PM**
Gather Articles and Journals	15 days	1/11/24 8:00 AM	21/11/24 5:00 PM
Review gathered data	10 days	22/11/24 8:00 AM	5/12/24 5:00 PM
Summarize and categorize findings	10 days	6/12/24 8:00 AM	19/12/24 5:00 PM
⊟**Data Collection**	20 days	**20/12/24 8:00 AM**	**16/1/25 5:00 PM**
Identify and gather historical stock price data	10 days	20/12/24 8:00 AM	2/1/25 5:00 PM
Clean and preprocess the collected data	10 days	3/1/25 8:00 AM	16/1/25 5:00 PM
⊟**Model Development**	50 days	**17/1/25 8:00 AM**	**27/3/25 5:00 PM**
Outline the architecture of the LSTM model	5 days	17/1/25 8:00 AM	23/1/25 5:00 PM
Code the LSTM model using Python	10 days	24/1/25 8:00 AM	6/2/25 5:00 PM
Integrating the preprocessed data	15 days	7/2/25 8:00 AM	27/2/25 5:00 PM
Evaluate Model Performance	2 days	28/2/25 8:00 AM	3/3/25 5:00 PM
Train the LSTM model on the training dataset	3 days	4/3/25 8:00 AM	6/3/25 5:00 PM
Adjusting hyperparameters	5 days	7/3/25 8:00 AM	13/3/25 5:00 PM
Validate and Test Model	5 days	14/3/25 8:00 AM	20/3/25 5:00 PM
Evaluate model performance using RMSE and R-squared	5 days	21/3/25 8:00 AM	27/3/25 5:00 PM
⊟**Develop User Interface**	20 days	**28/3/25 8:00 AM**	**24/4/25 5:00 PM**
Design Interface	10 days	28/3/25 8:00 AM	10/4/25 5:00 PM
Create a user-friendly interface	5 days	11/4/25 8:00 AM	17/4/25 5:00 PM
Seamless integration of the LSTM model with the user interface	5 days	18/4/25 8:00 AM	24/4/25 5:00 PM
⊟**Documentation and Reporting**	24 days	**25/4/25 8:00 AM**	**28/5/25 5:00 PM**
Gather and organize all research, methodologies, and results	5 days	25/4/25 8:00 AM	1/5/25 5:00 PM
Create Presentation	4 days	2/5/25 8:00 AM	7/5/25 5:00 PM
Potential applications of the predictive tool	5 days	8/5/25 8:00 AM	14/5/25 5:00 PM
Complete documentation submission	10 days	15/5/25 8:00 AM	28/5/25 5:00 PM

Figure 19: Project Timeline

8.2 Gantt Chart

Name	Duration	Start	Finish
⊟**STOCK MARKET PREDICTION**	166 days	**9/10/24 8:00 AM**	**28/5/25 5:00 PM**
⊟**Proposal Development**	17 days	**9/10/24 8:00 AM**	**31/10/24 5:00 PM**
Deciding aim and objectives	10 days	9/10/24 8:00 AM	22/10/24 5:00 PM
Purpose of research	5 days	23/10/24 8:00 AM	29/10/24 5:00 PM
Scope of research	1 day	30/10/24 8:00 AM	30/10/24 5:00 PM
Task distribution	1 day	31/10/24 8:00 AM	31/10/24 5:00 PM
⊟**Conduct Literature Review**	35 days	**1/11/24 8:00 AM**	**19/12/24 5:00 PM**
Gather Articles and Journals	15 days	1/11/24 8:00 AM	21/11/24 5:00 PM
Review gathered data	10 days	22/11/24 8:00 AM	5/12/24 5:00 PM
Summarize and categorize findings	10 days	6/12/24 8:00 AM	19/12/24 5:00 PM
⊟**Data Collection**	20 days	**20/12/24 8:00 AM**	**16/1/25 5:00 PM**
Identify and gather historical stock price data	10 days	20/12/24 8:00 AM	2/1/25 5:00 PM
Clean and preprocess the collected data	10 days	3/1/25 8:00 AM	16/1/25 5:00 PM
⊟**Model Development**	50 days	**17/1/25 8:00 AM**	**27/3/25 5:00 PM**
Outline the architecture of the LSTM model	5 days	17/1/25 8:00 AM	23/1/25 5:00 PM
Code the LSTM model using Python	10 days	24/1/25 8:00 AM	6/2/25 5:00 PM
Integrating the preprocessed data	15 days	7/2/25 8:00 AM	27/2/25 5:00 PM
Evaluate Model Performance	2 days	28/2/25 8:00 AM	3/3/25 5:00 PM
Train the LSTM model on the training dataset	3 days	4/3/25 8:00 AM	6/3/25 5:00 PM
Adjusting hyperparameters	5 days	7/3/25 8:00 AM	13/3/25 5:00 PM
Validate and Test Model	5 days	14/3/25 8:00 AM	20/3/25 5:00 PM
Evaluate model performance using RMSE and R-squared	5 days	21/3/25 8:00 AM	27/3/25 5:00 PM
⊟**Develop User Interface**	20 days	**28/3/25 8:00 AM**	**24/4/25 5:00 PM**
Design Interface	10 days	28/3/25 8:00 AM	10/4/25 5:00 PM
Create a user-friendly interface	5 days	11/4/25 8:00 AM	17/4/25 5:00 PM
Seamless integration of the LSTM model with the user interface	5 days	18/4/25 8:00 AM	24/4/25 5:00 PM
⊟**Documentation and Reporting**	24 days	**25/4/25 8:00 AM**	**28/5/25 5:00 PM**
Gather and organize all research, methodologies, and results	5 days	25/4/25 8:00 AM	1/5/25 5:00 PM
Create Presentation	4 days	2/5/25 8:00 AM	7/5/25 5:00 PM
Potential applications of the predictive tool	5 days	8/5/25 8:00 AM	14/5/25 5:00 PM
Complete documentation submission	10 days	15/5/25 8:00 AM	28/5/25 5:00 PM

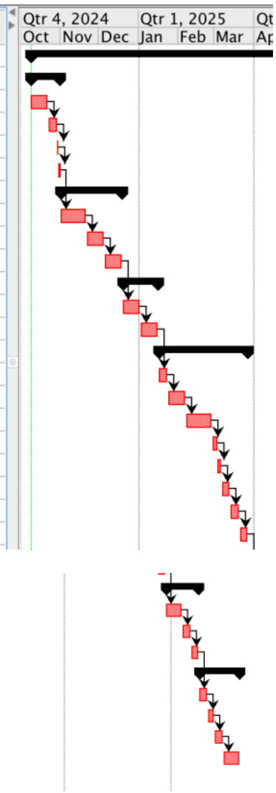

Figure 20: Gantt Chart

8.3 Visualizations

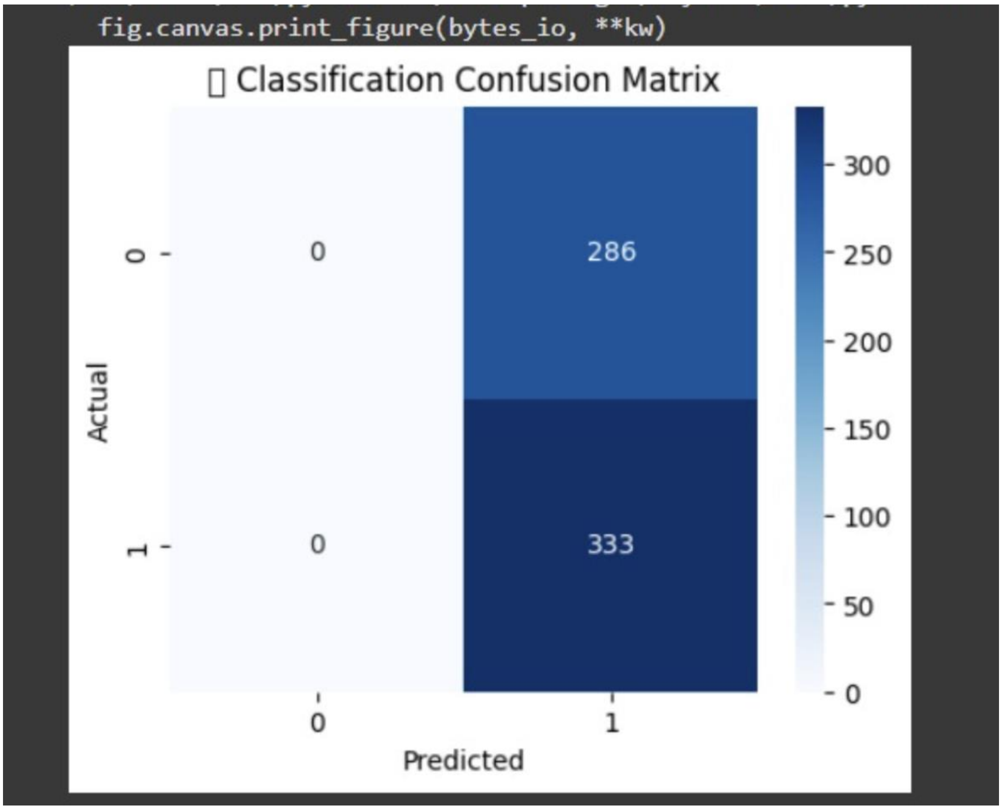

Figure 21: Confusion Matrix

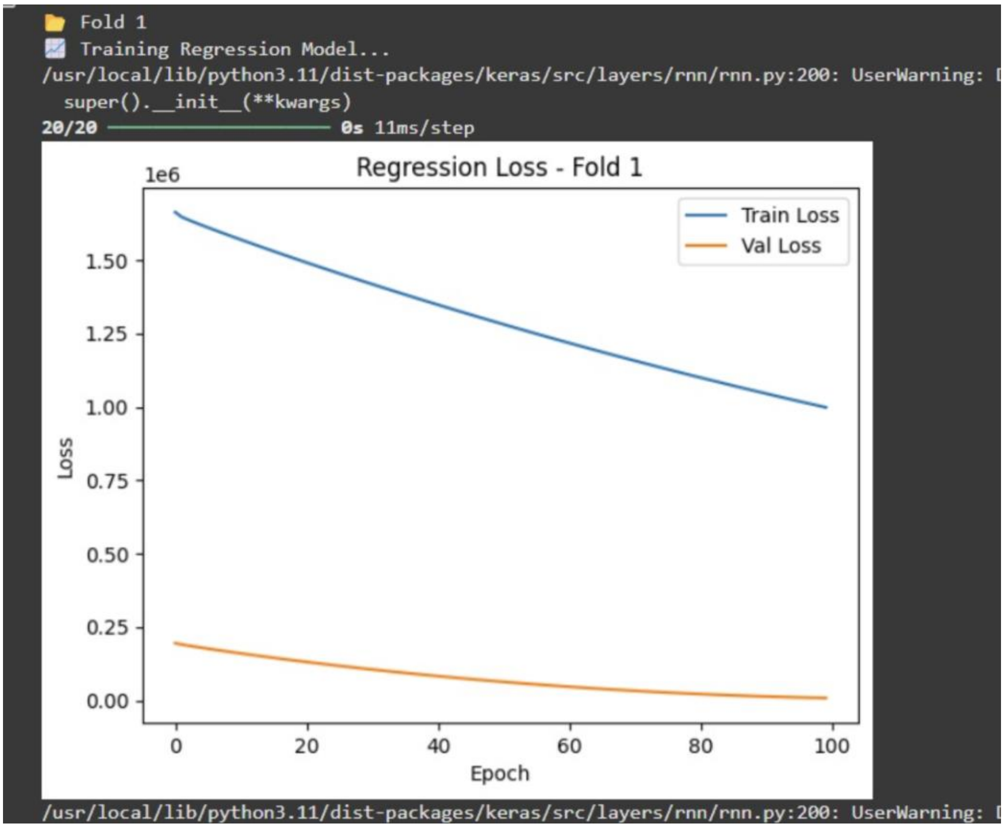

Figure 22: Regression Loss in Fold 1

Talha Risath
22222895

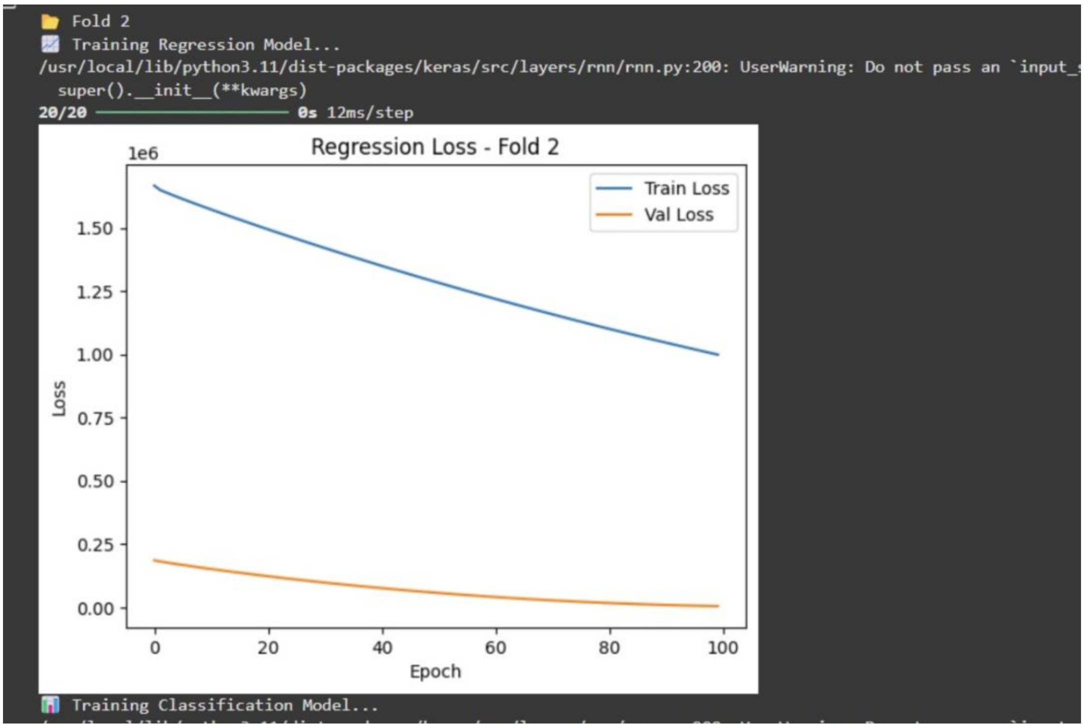

Figure 23: Regression Loss in fold 2

Talha Risath
22222895

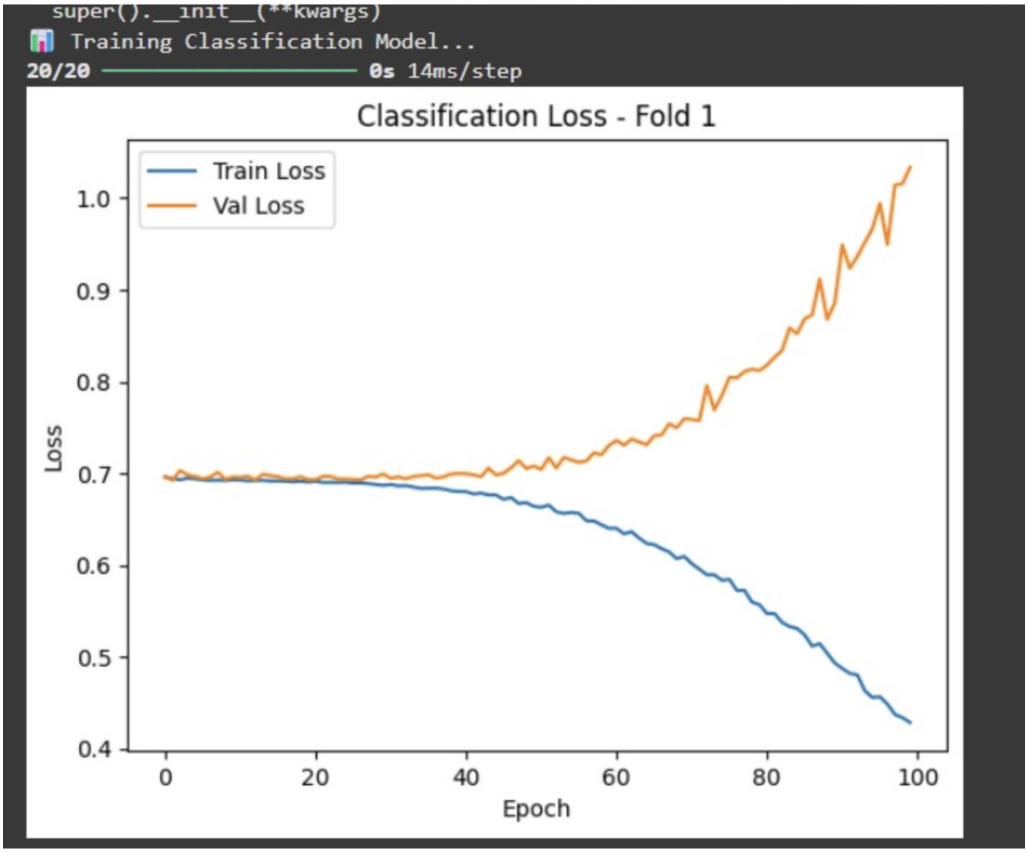

Figure 24: Classification Loss in fold 1

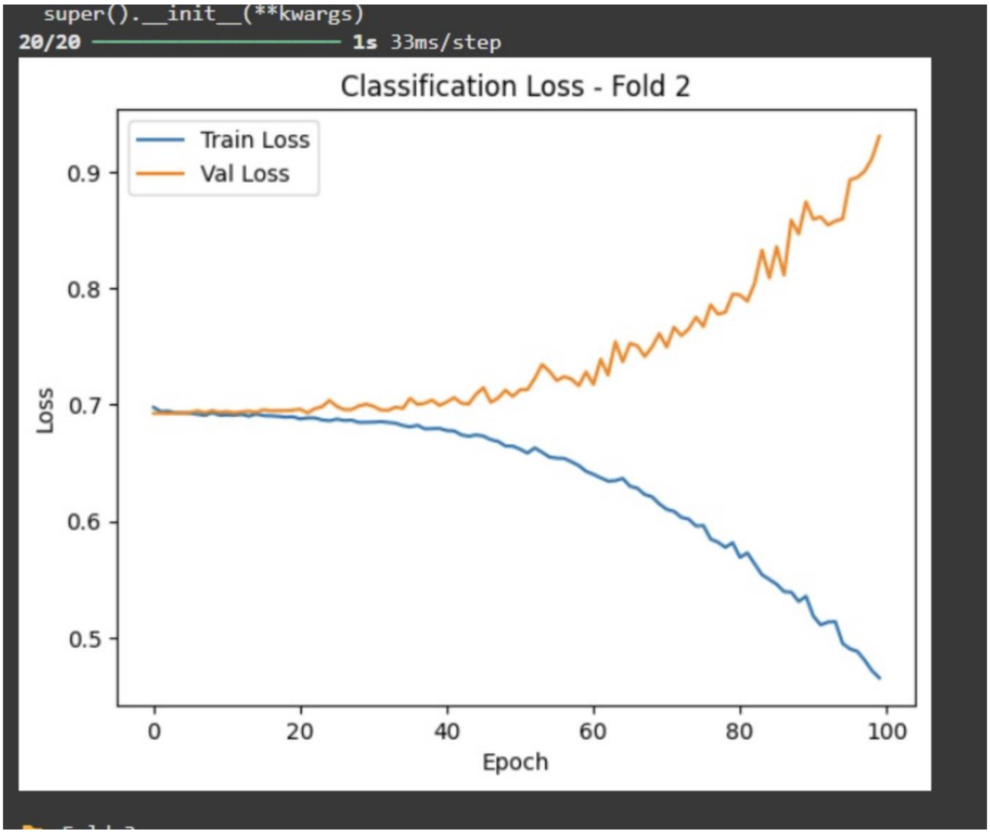

Figure 25: Classification Loss in fold 2

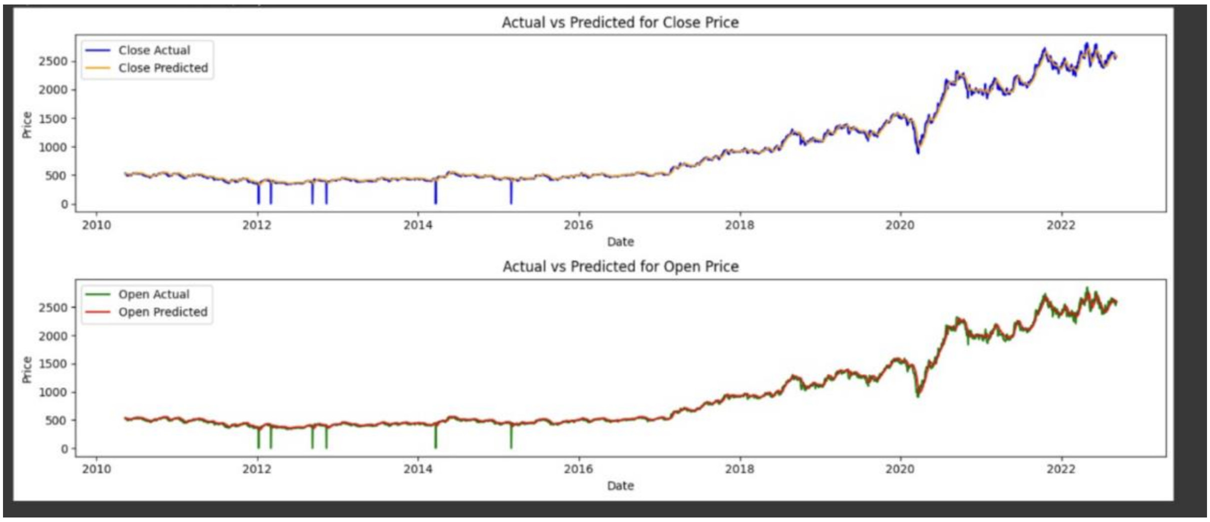

Figure 26: Actual vs Predicted for Open and Close price