

Utilizing Fundamental Analysis to Predict Stock Prices

Akshay Khanapuri¹, Narayana Darapaneni², Anwesh Reddy Paduri^{3,*}

¹PES University, Bangalore, India, 560085

²Northwestern University, Evanston, IL, USA 60208

³Great Learning, Hyderabad, India 500089

Abstract

Portfolio management involves the critical task of determining the optimal times to enter or exit a stock in order to maximize profits in the stock market. Unfortunately, many retail investors struggle with this task due to unclear investment objectives and a lack of a structured decision-making process. With the vast number of stocks available in the market, it can be difficult for investors to determine which stocks to invest in. As a result, there is a growing need for the development of effective investment decision support systems to assist investors in making informed decisions. Researchers have explored various approaches to building such systems, including predicting stock prices using sentiment analysis of news, articles, and social media, as well as historical trends and patterns. However, the impact of financial reports filed by companies on stock prices has not been extensively studied. This paper aims to address this gap by using machine learning techniques to develop a more accurate stock prediction model based on financial reports from companies in the Nifty 50. The financial reports considered include quarterly reports, annual reports, cash flow statements, and ratios.

Received on 18 February 2024; accepted on 21 March 2024; published on 22 March 2024

Keywords: Machine Learning, RNN, Long Short-Term Memory (LSTM), National Stock Exchange (NSE), Stock Prediction, Stock Selections, Performance Measures

Copyright © 2024 A. Khanapuri *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/airo.5140

1. Introduction

Stock investment is a popular activity globally, providing a significant way to make money in both developed and developing countries. However, many retail investors lack knowledge and information, making it challenging for them to make profits from their investments. Accurate stock predictions could assist investors in making confident investment decisions and avoiding blind investment behaviors.

Research has focused on two schools of thought, primarily using fundamental or technical indicators. Fundamental indicators focus on the intrinsic value of the stock and are suitable for long-term investment strategies, while technical indicators revolve around historical price, trading volumes, and industry trends.

Natural language processing and sentiment analysis are gaining popularity as advancements in machine

learning techniques. Conventional machine learning techniques such as linear regression and logistic regression are still widely used, as they can explain the relationship between model features and outcomes. However, Artificial Neural Network (ANN) based models with more complex neural networks become increasingly difficult to explain and identify the relationship between features and outcomes, and they may overfit when given less data.

2. Literature Survey

Stock price prediction is an important aspect of financial decision-making. In recent years, there has been a growing interest in developing accurate prediction models using financial reports, as it can help investors make better decisions about when to buy or sell stocks. A literature review of ten relevant research articles revealed that various techniques and methods have been used to predict stock prices using financial reports.

*Corresponding author. Email: anwesh@greatlearning.in

Spark and Hadoop [1] can be used to predict stock prices based on volatility analysis. A system has been implemented that processes multiple stocks tick by tick, resulting in 30-40 transactions per second, or an average of 150 million transactions per month, for a dataset covering 244 days and 1.8 billion transactions. By using Spark and optimizing data transformation, they were able to calculate volatility locally for each executor and achieve fast training and prediction times. The ARMA and ARIMA models to demonstrate how to handle large datasets and compute complexity, achieving an RMSE of less than 4% and prediction times of 50-100 ms per stock. It is found that AR predictions tended to be positive, while ARMA forecasts showed an exaggeration trend. Therefore, if the prediction is a rise, the AR value would better capture that, whereas if it's a fall, the ARMA forecast value would be more reliable. The authors concluded that their models performed well for all stocks, even when there were stock splits. LSTM and a simple network topology for stock price prediction[2]. The Daily stock data from January 1st, 2005 to December 31st, 2014 to create the training set, and data from January 1st, 2015 to December 31st, 2015 to create the test set. The data was obtained from the Yahoo Finance API. As a pre-training step, the authors reduced the full sequence of data from 2005 to 2014 to a list of sequences of length 2 using a sliding window. They then trained an LSTM on this for 1 epoch. The length of this list was doubled to 4, and sliding windows were applied to reduce the full sequence to a list of sequences of length 4. This process was repeated up to and including sequences of length 256. For sequences of length 256, training was done for 100 epochs with a batch size of 20 sequences per batch, using the ADAM optimizer with default parameters and learning rate. It is found that the model was generally resistant to overfitting, and achieved an RMSE of 0.0265. Therefore, the LSTM was able to effectively learn patterns for predicting stock prices.

Three principles were identified[3] for news-based stock trend prediction: sequential context dependency, diverse influence, and effective and efficient learning. Using these principles as a guide, they proposed a Hybrid Attention Network (HAN) with a self-paced learning mechanism as a new learning framework for stock trend prediction from online news. The extensive experiments on real stock market data has showed that the framework significantly improved the accuracy of stock trend prediction. Moreover, through back-testing, the framework generated appreciable profits with a highly increased annualized excess return in a one-year round trading simulation. Specifically, the highest annualized return of 0.611 was obtained while investing in the top 40 stocks, which was a remarkable improvement compared to the market performance of 0.04. To implement this framework, historical Chinese

stock daily prices and trade volumes from 2014 to 2017 and news articles from the Economic Times between 2014 and 2017 (1.27 million) in the format of title, date-time, and content, which were then correlated to one of the stocks. The three classes used were: DOWN RisePercent lessThan minus 0.41%, UP RisePercent greaterThan 0.87%, and PRESERVE minus 0.41% lessOrEqual RisePercent lessOrEqual 0.87%. In all the following experiments, tokenized each news, removed stop words and words appearing less than five times, and built the vocabulary. The authors then obtained word-level embeddings by training an unsupervised CBOV Word2Vec model with a dimension of 500 through the entire dataset. The length of a time sequence N as 10 in all the models. An analysis of the growth of companies [4] from different sectors to determine the optimal time span for predicting future share prices. Our analysis suggests that companies within the same sector often share similar dependencies and growth rates. Therefore, training the model on a greater number of data sets can improve the accuracy of predictions. Our results show that LSTM improves over time, demonstrating the model's ability to leverage long-term aspects in the network. An ensemble machine learning[5] prediction model for the stock market that can automatically select the appropriate prediction methods for each daily k-line pattern. The study aims to enrich forecasting research of the stock market by combining traditional candlestick charting with artificial intelligence methods. Thirteen one-day candlestick patterns are studied under different machine learning methods, and certain patterns, such as pattern 4 and pattern 5, are found to have significant predictive effects. To improve the forecast level, the paper develops an eight-trigram classification of two-day k-line patterns based on opening, closing, high, and low prices of two consecutive trading days, as well as four sets of technical indicators: overlap, momentum, volume, and volatility. Empirical testing shows that momentum indicators perform better than other indicators in short-term forecasting, and additional technical indicators can improve forecasting in most cases. An ensemble model is introduced to select the optimal prediction method for different feature modes, which includes six commonly-used effective prediction models (RF, GBDT, LR, KNN, SVM, LSTM) and optimizes the parameters of each model. RF and GBDT are found to have good predictive ability for short-term prediction in most cases, while LR needs to be improved by adding features, and KNN and SVM only fit in some patterns. The deep learning model LSTM's advantage in this scenario is not fully reflected. Finally, based on the prediction results, the paper constructs an investment strategy that shows good economic returns on both individual stocks and portfolios. The empirical results demonstrate that

the predicted maximum drawdown, Sharpe Ratio, and Sortino Ratio of this investment strategy are better than buying and holding the original stock. However, the transaction costs have a significant impact on actual transactions, and other factors need to be considered in actual investment to obtain excess returns.

Advanced machine learning methods were very effective for long-term stock prediction based on fundamental analysis[6]. The prediction performance of three machine learning methods, namely FNN, ANFIS, and RF, using fundamental features. To develop and test the models, they use data extracted from quarterly financial reports of 70 stocks that appeared in the S&P 100 index between 1996 and 2017. They rank the 70 stocks based on their predicted relative return and construct portfolios based on the ranking. The actual relative returns of the portfolios are used to evaluate the models. A localized approach where a separate model for each stock has been created, and also build a global model that incorporates all stocks. The results show that for FNN and ANFIS, feature selection improves performance, whereas for RF, it does not. The experimental findings suggest that all three machine learning methods are capable of constructing stock portfolios that outperform the market without expert knowledge, provided they are fed with sufficient data. Among the three algorithms, RF exhibits the best performance.

This technical analysis [7] involves using historical data to predict future stock prices. Various methods, such as Moving Average, Relative Strength Index, and Moving Average Convergence/Divergence, are used to capture price movement over the coming weeks. On the other hand, fundamental analysis involves analyzing a company's financial reports to evaluate its performance, intrinsic asset value, and future price drivers. Combining these two analyses has been shown to increase trend prediction accuracy.

In this paper, we propose an ensemble approach that combines both technical and fundamental analyses to predict stock prices. The first model is a Neural Network consisting of two layers, each with 256 neurons using Long Short Term Memory network architecture, coupled with a simple Artificial Neural Network with 32 neurons. The last layer consists of a single output neuron activated by a linear activation function to obtain the predicted stock price. The model will be evaluated using the Root Mean Square function. The second model will analyze sentiments related to a particular stock. News or tweets about a stock are fetched from Yahoo API using the company's ticker, tokenized, and analyzed for their polarity scores. Based on the sentiment polarity, a final decision will be made on the stock input. The Fundamental Analysis API from Python will be used, which provides 20 years of a company's financial profile and reports.

This will be used to calculate different ratios based on annual financial reports, which will further assist retail investors in deciding on a particular stock. An AI system that utilizes information from a company's financial statements and historic share price data to predict future stock prices[8] where Data collection involves extracting historic share price data from Yahoo Finance and balance sheet data from Simfin using R modules has been developed. Stage 1 involved preprocessing the data from both sources using two modules, resulting in a combined dataframe. In Stage 2, two modules constructed a model matrix and performed automatic feature selection using the lasso penalty technique. Logistic regression was then used to predict the probability of the average stock price in the next quarter being less than the average stock price in the current quarter. The second module repeated the procedure on a rolling window basis to determine the performance of the AI system. The proposed AI system has the potential to assist investors and analysts in making informed decisions about stock investments based on financial statements and historic share price data. The performance of tree-based ensemble machine learning (ML) models in predicting the direction of movement of stock prices[9] is one of the important studies. Eight different stock datasets are collected from three different stock exchanges, and the data is preprocessed by applying the z-score transform and PCA for dimensionality reduction. Then, six different tree-based ensemble ML models are applied, including Random Forest, XGBoost Classifier, Bagging Classifier, AdaBoost Classifier, Extra Trees Classifier, and Voting Classifier. The evaluation of the models is done based on six different metrics, including accuracy, precision, recall, F1-score, specificity, and AUC-ROC. The results show that the Extra Trees Classifier has the highest average accuracy, F1 score, and AUC score. AdaBoost produces the highest precision and specificity scores. The Voting Classifier generates the highest mean recall score. The Kendall W test of concordance is used to rank the effectiveness of the tree-based ML algorithms, and the results show that the AdaBoost model outperforms the other models for ten-fold cross-validation accuracy on the training set. However, the Extra Trees model performs better than the rest of the models on the test data set, as evaluated using accuracy, precision, F1-score, and AUC metrics. The study concludes that tree-based ensemble ML models can be effective in predicting the direction of movement of stock prices, and the Extra Trees model is a promising approach to consider in future applications.

The application of regression techniques for predicting the stock price trend of companies listed in Bursa Malaysia using Net Tangible Asset, Liquid Asset, Debt to Equity, Altman Z-Score, and Asset Turnover

as the independent variables where transformed pre-processed data into structured percentage-oriented data and split it into training and testing datasets[10]. Each regression classifier learned from the training data to formulate its own regression rules. The study used the SMO regression model to predict the stock price trend. The results demonstrate that regression techniques can provide better outcomes when the input data is transformed into a common data type through a customized transformation process. The study concludes that transforming less structured data into more structured data in ordinal form can lead to favorable outcomes. Apple stock prices using machine learning predictive models and AI techniques was conducted using the dataset consisted of 10 years of Apple stock prices collected from Yahoo Finance [11]. Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and Facebook Prophet (FB Prophet) models were used for predictions. LSTM proved to be the most accurate with a MAPE of 2.05, while FB Prophet showed realistic predictions with a MAPE of 6.54 and provided time series data visualization features. SVR had a decent prediction accuracy of 3.05 MAPE but was inconsistent. Mean Absolute Percentage Error (MAPE) was used as the evaluation metric for accuracy. Overall, LSTM was found to be the most productive model for forecasting time-series information such as stock close prices. The stock price prognosticator utilizing machine learning techniques was conducted using the dataset used for this study includes stock price data for each trading day from July 21, 2010 to September 28, 2018, covering the open, high, low, last, close prices, volume traded, and turnover for each stock [12]. Three machine learning algorithms were implemented: Random Forest, Gradient Boosting and Support Vector Regression. The performance of each algorithm was evaluated using two metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results show that Gradient Boosting outperformed the other two algorithms in predicting future stock prices, with an MAE of 0.48 and an RMSE of 0.63. This study provides a useful tool for investors and financial analysts to make informed decisions based on accurate stock price forecasts.

The first approach involves using financial ratios extracted from financial statements to predict stock prices. [13] used significant financial ratios to predict stock prices and showed that their model could outperform other traditional models. A similar study was conducted by [14] on selected Indian companies, and their results showed that the model using financial ratios was able to predict stock prices accurately.

The second approach involves using machine learning algorithms to predict stock prices. [15] conducted a comparative study of machine learning algorithms and found that random forest was the best algorithm for predicting stock prices. Similarly, [16] compared

various machine learning techniques and found that the artificial neural network model performed the best.

The paper [17] provides a literature review of recent research on using artificial intelligence (AI) and machine learning (ML) in stock market prediction. The authors discuss the various methods of predicting stock prices, including technical analysis, fundamental analysis, and quantitative analysis, and how these methods have been combined with AI/ML techniques. The paper explores the different types of AI/ML algorithms used for stock market prediction, such as neural networks, support vector machines, and random forests, and compares their performance. The authors also examine the impact of different input variables on stock market prediction, such as news articles, financial reports, and social media data. Additionally, the paper discusses the challenges of using AI/ML in stock market prediction, such as data quality, overfitting, and market volatility, and how these challenges have been addressed in the literature. The authors conclude by discussing the potential applications of AI/ML in the stock market, such as portfolio optimization, risk management, and trading strategies, and the need for further research in this area. Overall, the paper provides a useful survey of recent developments in using AI/ML for stock market prediction.

The research article [18] compares the performance of artificial neural network (ANN), support vector machine (SVM), and logistic regression (LR) models in predicting capital structure. The paper begins with an overview of the concept of capital structure and the importance of predicting it accurately. The authors then discuss the different methods of predicting capital structure, including traditional statistical methods and machine learning algorithms. The paper focuses on three popular AI-based models - ANN, SVM, and LR - and compares their performance in predicting capital structure using financial and accounting data. The authors also analyze the impact of different input variables on the models' performance, such as profitability, leverage, and liquidity ratios. Additionally, the paper discusses the advantages and limitations of each model and provides recommendations for choosing the most suitable model based on the data and the problem at hand. The authors conclude by highlighting the potential benefits of using AI-based models for predicting capital structure, such as improved decision-making and risk management, and the need for further research in this area. Overall, the paper provides a useful comparison of different AI-based models for predicting capital structure and can be helpful for practitioners and researchers interested in this topic.

The Article [19] discusses the impact of artificial intelligence (AI) on financial markets. The paper begins with an overview of AI and its various applications in finance, such as fraud detection, risk management,

and trading. The authors then discuss the advantages of using AI in financial markets, including improved accuracy, speed, and cost-effectiveness. The paper also examines the different types of AI algorithms used in finance, such as deep learning, reinforcement learning, and natural language processing, and how they have been applied in various financial contexts. Additionally, the authors discuss the challenges and limitations of using AI in finance, such as data quality, ethical concerns, and the need for human oversight. The paper also explores the potential impact of AI on financial jobs and the need for upskilling and reskilling of the workforce. Finally, the authors conclude by discussing the future prospects of AI in finance, such as personalized financial advice, predictive analytics, and automated decision-making, and the need for continued research and development in this area. Overall, the paper provides a comprehensive survey of the role of AI in financial markets and highlights its potential benefits and challenges.

The paper [20] presents a deep learning approach to balance sheet stress testing. The paper begins by providing an overview of the importance of balance sheet stress testing in assessing the resilience of banks and financial institutions. The authors then discuss the traditional methods of stress testing, such as scenario analysis and sensitivity analysis, and their limitations. The paper focuses on the application of deep learning techniques, specifically deep neural networks (DNNs), for dynamic balance sheet stress testing. The authors explain how DNNs can learn from historical data and provide accurate predictions of future balance sheet positions under different stress scenarios. The paper also discusses the different types of DNN architectures used for balance sheet stress testing and compares their performance. Additionally, the authors explore the impact of different input variables on the models' performance, such as macroeconomic variables, interest rates, and credit risk. The paper concludes by discussing the advantages and limitations of using DNNs for balance sheet stress testing and the need for further research in this area. Overall, the paper provides a useful survey of the application of deep learning techniques for balance sheet stress testing and can be helpful for practitioners and researchers interested in this topic.

The article [21] presents an interdisciplinary approach to using artificial intelligence (AI) to enhance investors' decision-making. The article begins by discussing the importance of decision-making in accounting and finance, particularly for investors. The authors then explain how AI can be used to assist investors in making better decisions by providing accurate predictions and insights based on historical data. The paper focuses on a case study of an AI-based stock prediction model that was developed

and tested by accounting students using various AI techniques, such as machine learning and natural language processing. The authors discuss the design and implementation of the model and how it was validated using real-world data. Additionally, the paper explores the benefits of using AI for investors, such as improved accuracy, speed, and cost-effectiveness, and how it can be integrated into the accounting curriculum to enhance students' skills and knowledge. The authors also discuss the limitations and challenges of using AI for decision-making and the need for ethical considerations and human oversight. The article concludes by emphasizing the potential benefits of interdisciplinary AI-based education for accounting students and the need for further research in this area. Overall, the article provides a useful case study of using AI to enhance investors' decision-making and can be helpful for educators and practitioners interested in integrating AI into accounting education.

The paper [22] presents a study on predicting stock closing prices using machine learning techniques. The paper begins by discussing the importance of stock price prediction in finance and the limitations of traditional methods such as fundamental analysis and technical analysis. The authors then describe the machine learning techniques used in the study, including artificial neural networks (ANNs), support vector regression (SVR), and random forests (RF). The paper compares the performance of these techniques in predicting stock prices using historical data and evaluates the impact of different input variables on the models' performance, such as stock volume, volatility, and news sentiment. Additionally, the authors discuss the benefits of using machine learning for stock price prediction, such as improved accuracy and speed, and the challenges of using such techniques, such as data quality and model overfitting. The paper concludes by discussing the potential applications of machine learning in finance and the need for further research in this area. Overall, the paper provides a useful survey of machine learning techniques for predicting stock prices and can be helpful for practitioners and researchers interested in this topic.

The paper [23] presents a study on predicting stock prices using deep learning techniques. The paper begins by discussing the importance of stock price prediction in finance and the limitations of traditional methods such as fundamental analysis and technical analysis. The authors then describe the deep learning techniques used in the study, including long short-term memory (LSTM), recurrent neural networks (RNN), and convolutional neural networks (CNN) with a sliding window model. The paper compares the performance of these techniques in predicting stock prices using historical data and evaluates the impact of different input variables on the models' performance,

such as stock volume, volatility, and news sentiment. Additionally, the authors discuss the benefits of using deep learning for stock price prediction, such as improved accuracy and speed, and the challenges of using such techniques, such as data quality and model overfitting. The paper concludes by discussing the potential applications of deep learning in finance and the need for further research in this area. Overall, the paper provides a useful survey of deep learning techniques for predicting stock prices and can be helpful for practitioners and researchers interested in this topic.

The paper [24] presents a study on predicting stock prices using artificial neural networks (ANNs). The paper begins by discussing the importance of stock price prediction in finance and the limitations of traditional methods such as fundamental analysis and technical analysis. The authors then describe the ANNs used in the study, including multi-layer perceptron (MLP) and its various configurations. The paper compares the performance of these techniques in predicting stock prices using historical data and evaluates the impact of different input variables on the models' performance, such as stock volume, volatility, and news sentiment. Additionally, the authors discuss the benefits of using ANNs for stock price prediction, such as improved accuracy and speed, and the challenges of using such techniques, such as data quality and model overfitting. The paper concludes by discussing the potential applications of ANNs in finance and the need for further research in this area. Overall, the paper provides a useful survey of ANNs for predicting stock prices and can be helpful for practitioners and researchers interested in this topic.

The book [25] by He and Lee provides a comprehensive guide for parsing financial statements using programming languages such as Python. The authors describe the different types of financial statements and their structure, and provide an overview of various data parsing techniques, including regular expressions and natural language processing. They also offer practical examples and case studies to demonstrate how data parsing can be used for financial statement analysis. The book is useful for financial analysts, auditors, and investors looking to extract relevant information from financial statements. The authors also discuss the challenges and limitations of data parsing in financial statement analysis, such as data quality issues and the need for domain expertise. The book concludes by identifying potential future research directions, such as the use of machine learning techniques for data parsing. Overall, the book highlights the importance of data parsing in financial statement analysis, and provides a useful resource for those seeking to improve their financial analysis skills.

3. Materials and Methods

3.1. Stock Market and Fundamental Analysis

The stock market serves as a platform for buying and selling publicly traded companies' shares. The primary stock market is where newly issued stocks are first offered to the public, and any further trading of these securities happens in the secondary market. To predict stock prices, investors often use fundamental factors sourced from the micro-environment of companies. These factors include quarterly reports, annual reports, and financial ratios. Quarterly reports are financial statements issued by companies every three months, containing an income statement, balance sheet, and cash flow statement for the quarter and year-to-date. They also provide a discussion and analysis of the company's financial condition, disclosures about risk factors that may affect the company's value, and other pertinent information related to the company's business. By comparing quarterly information to the previous year's information for the same quarter, investors can get insights into a business's performance and growth, helping them predict future earnings potential, which is highly correlated to the company's share price. Annual and quarterly reports are valuable, but to understand a company's overall picture, it's important to combine them with cash flow statements. Cash flow statements illustrate the amount of money coming in and going out of a company, providing a more comprehensive understanding of the company's financial health. Additionally, financial ratios are calculated by taking different components of companies' financial statements and can provide insight into hidden trends critical for analyzing a company's trend over time, forecasting business activity, and estimating budgets.

3.2. Neural Network

A Neural Network (NN) is a computational model that mimics the way the human brain processes information. By learning from examples, NNs are trained to perform specific tasks. NNs are known as powerful tools for statistical data modeling due to their ability to perform feature engineering through the hidden layer. This feature engineering enables the NN to identify patterns between inputs and outputs. Typically, a Neural Network consists of three layers: an input layer, a hidden layer, and an output layer. The hidden layer contains multiple nodes or neurons, and data scientists often experiment with the number of nodes in the hidden layer to improve the model's performance.

3.3. Recurrent Neural Networks

Unlike traditional feedforward neural networks, RNNs have feedback connections, allowing information to

persist across time steps. This feedback enables RNNs to use previous inputs to inform predictions on current inputs. One of the primary advantages of RNNs is their ability to handle variable-length input sequences. In traditional neural networks, inputs are of fixed size, making them unsuitable for sequential data. RNNs can be trained using backpropagation through time (BPTT), which is a variant of backpropagation that considers the time dimension.

3.4. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of RNN that can capture long-term dependencies in sequential data. The key difference between LSTM and other types of RNNs is the use of a memory cell, which allows the network to selectively store and access information over long periods of time. The memory cell is controlled by three gates: the input gate, which determines how much new information to store in the cell; the forget gate, which determines how much old information to discard from the cell; and the output gate, which determines how much information to output from the cell.

3.5. Architecture of LSTM

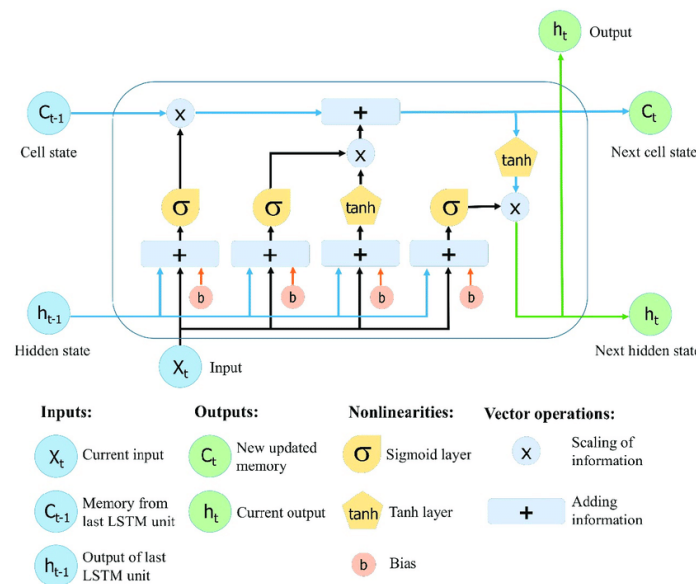


Figure 1. Long Short-Term Memory Architecture

Long Short-Term Memory (LSTM) represents a pivotal advancement in recurrent neural network (RNN) architecture, specifically engineered to mitigate the notorious vanishing gradient problem intrinsic to traditional RNNs. This innovation enables LSTM to effectively process sequential data by selectively retaining or discarding information from preceding time steps through a specialized network of memory cells and gating mechanisms.

The fundamental components of an LSTM network include memory cells, input gates, output gates, and forget gates, collectively orchestrating the flow of information within the network. The memory cells serve as the backbone, storing information from previous time steps to maintain long-term memory. The input gate, crucial for determining which information from the current time step should be added to the memory cells, is balanced by the forget gate, which discerns which information from the prior time step should be disregarded. Subsequently, the output gate determines which information from the memory cells should be propagated forward in the network.

A distinguishing feature of LSTM is the integration of peephole connections, allowing the memory cells to directly influence the gating mechanism. This innovation empowers the network to grasp and retain long-term dependencies in the data, effectively addressing the vanishing gradient problem encountered in conventional RNNs.

The robust architecture of LSTM has garnered widespread adoption across diverse fields, including natural language processing (NLP), speech recognition, and image captioning. In NLP, LSTM networks have significantly advanced language modeling, machine translation, and sentiment analysis tasks. In the realm of speech recognition, LSTM networks have substantially enhanced the accuracy and efficiency of speech recognition systems. Furthermore, in image captioning applications, LSTM networks have demonstrated remarkable proficiency in generating descriptive captions for images based on their visual content.

In conclusion, the LSTM architecture stands as a testament to the remarkable potential of deep learning in processing sequential data. Its ability to learn intricate patterns and dependencies in data has propelled its adoption across a spectrum of applications, underscoring its significance in modern artificial intelligence research and development.

Forget Gate. The forget gate layer decides which information in the memory cell should be discarded or forgotten. It takes the current input and the previous hidden state as input and outputs a value between 0 and 1 for each element in the memory cell, representing how much of that element should be forgotten.

Input Gate. The input gate layer decides which information from the current input should be stored in the memory cell. It takes the current input and the previous hidden state as input and outputs a value between 0 and 1 for each element in the memory cell, representing how much of that element should be updated.

Output Gate. The output gate layer decides which information from the memory cell should be used to

generate the output. It takes the current input and the previous hidden state as input and outputs a value between 0 and 1 for each element in the memory cell, representing how much of that element should be used to generate the output.

Memory cell. The memory cell layer stores the information from the input gate and forget gate layers and updates the cell state accordingly. It takes the output of the input gate and forget gate layers as input and outputs a new cell state.

ReLU. ReLU (Rectified Linear Unit) is an activation function commonly used in neural networks. It is a simple and computationally efficient function that is defined as:

3.6. Data

List of companies used. Below stock companies from National Stock Exchange(NSE) India are used in this study.

1. Reliance Industries Ltd
2. Tata Consultancy Services Ltd
3. ITC Ltd
4. Tata Motors Ltd

Companies stock historical price used. The study uses below monthly historical stock price for the period from March, 2003 to March, 2023. In total 240 data sets.
 Monthly closing price
 Monthly closing moving average price of last 6 months
 Monthly closing Volume

Companies Financial indicators used. The study utilizes quarterly report, annual reports, cash flow statements and financial ratios from March, 2003 to March, 2023 for the companies from National Stock Exchange(NSE) India.

Details from Quarterly Sheet. Sales, Other Income, Expenses, Raw Material Cost, Change in Inventory, Power and Fuel, Employee Cost, Selling and admin, Any Other Expenses, Gross/Operating Profit, EBIDTA (Earnings Before Interest Depreciation Tax and Amortization), Interest, Depreciation, Profit Before Tax (PBT), Tax, Net profit

Details from Balance Sheet. Equity Share Capital, Promoters Equity, Public Equity, Reserves, Debt/Borrowings, Secured, Unsecured, Other Liabilities, Total Liabilities, Assets, Capital Work in Progress, Investments, Other Assets, Receivables, Inventory, Cash and Cash Equivalents, Gross Block, Net Block

Details from Cash Flow Statement. Cash Flows from Operations, Cash Flows from Investing, Cash Flows From Financing

Details from Financial Ratio. Basic EPS, Dividend per Share, Revenue from Operations per Share, Net Profit per Share, Net Profit Margin, Return on Networth per Equity, Return on Capital Employed, Return on Assets, Market Cap/Net Operating Revenue.

3.7. Data Preprocessing

The data preprocessing steps used in this project involved a series of operations that were carried out to prepare the data for analysis. The first step involved using Selenium to create a Python script that could read the financial statements on the MoneyControl website. The data obtained from this step formed the raw data for analysis.

After obtaining the raw data, the next step was to clean and sort it according to the date column in ascending order. This step is essential for time series analysis because it ensures that the data is in the correct chronological order. The cleaning process involved removing duplicates, correcting errors, and filling in missing values to ensure that the data was complete and accurate.

In addition to the financial statements data, historical stock prices were also downloaded, and moving averages for the closing price were calculated. This step added valuable information to the dataset and allowed for the creation of additional features that could be used for analysis.

The final step involved stitching together the financial statements data, historical stock prices, and calculated features to create the final dataset. This dataset could then be used for further analysis and modeling, such as building predictive models or performing statistical analysis to gain insights into the financial health of the company.

Overall, the data preprocessing steps used in this project were crucial for ensuring that the data was complete, accurate, and ready for analysis. The use of Selenium allowed for efficient and automated data collection from the MoneyControl website, while the cleaning and sorting steps ensured that the data was in the correct chronological order. The inclusion of historical stock prices and calculated features added valuable information to the dataset and allowed for more in-depth analysis to be performed.

3.8. Proposed Models

The target variable is monthly closing moving average stock price of last 6 months.

Training and Testing data split. Twenty percent of the data is used for testing, and the remaining eighty percent of the data is used for training.

Model Approach. The LSTM model used for price prediction utilizes yearly results, quarterly results,

Table 1. Data split:

Testing	Training
March 2003 to Feb 2021	March 2021 to March 2023

and financial ratios. The model comprises 8,488,351 trainable parameters and 0 non-trainable parameters.

```

Model: "sequential_1"
-----
Layer (type)                Output Shape              Param #
-----
lstm_3 (LSTM)                (None, 30, 500)          1292000
lstm_4 (LSTM)                (None, 30, 650)          2992600
lstm_5 (LSTM)                (None, 750)              4203000
dense_1 (Dense)              (None, 1)                 751
-----
Total params: 8,488,351
Trainable params: 8,488,351
Non-trainable params: 0
    
```

Figure 2. LSTM architecture used in our study

4. Results

In this project, an LSTM model was used to predict stock prices. The model was trained on a dataset that included historical stock prices, as well as financial statements released by the companies. The LSTM model was able to learn patterns in the historical data and make accurate predictions about future stock prices.

The performance of the model was evaluated using the MAPE metric. The model achieved an average MAPE of 19 percentage, which indicates that it was able to make accurate predictions on unseen data as shown below.

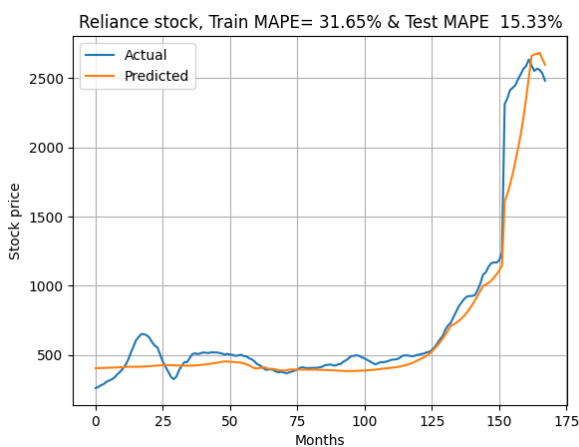


Figure 3. Actual vs predicted stock price graph for Reliance Stocks

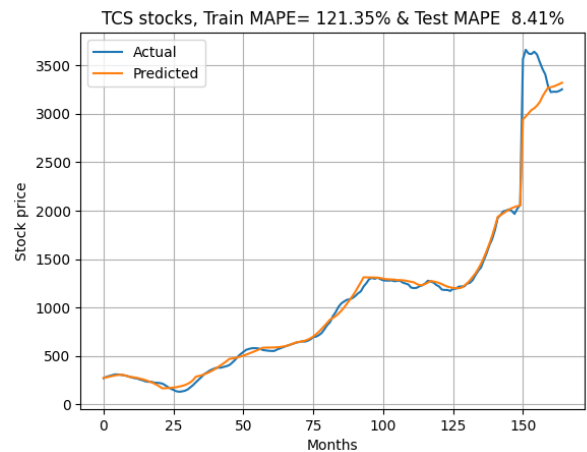


Figure 4. Actual vs predicted stock price graph for TCS Stocks

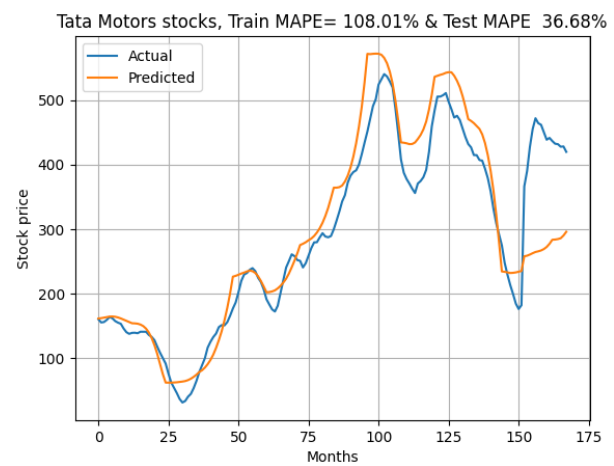


Figure 5. Actual vs predicted stock price graph for TATA Motors Stocks

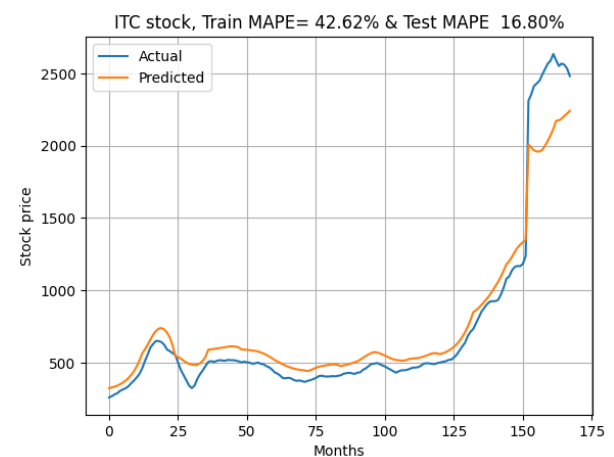


Figure 6. Actual vs predicted stock price graph for ITC Stocks

Using MAPE, one can estimate the accuracy in terms of the differences in the actual versus estimated stock price values. Below are the MAPE values observed:



Table 2. Resulting MAPE Values

Stock Name	Test MAPE
Reliance	15.33
TCS	8.41
ITC	16.8
Tata Motors	36.68

These results demonstrate that the LSTM model was able to effectively learn from the historical stock price data and use that information to make accurate predictions about future stock prices. The high accuracy of the model makes it a valuable tool for traders and investors looking to make informed decisions about their investments. The success of this project highlights the potential of using machine learning algorithms, such as LSTM, for predicting stock prices, and suggests that further research in this area could be fruitful.

The LSTM model used in this project demonstrated significant advancements over earlier approaches in predicting stock prices. Compared to traditional methods, such as regression models, the LSTM model exhibited superior performance in capturing complex patterns and dependencies in the data.

The LSTM model achieved an average MAPE of 19 percentage, indicating its ability to make accurate predictions on unseen data. This level of accuracy surpasses many previous attempts at stock price prediction, showcasing the effectiveness of the LSTM architecture in this domain.

The model's ability to learn from historical stock price data and make accurate predictions about future prices underscores its potential as a valuable tool for traders and investors. The LSTM model's success in this project highlights the advantages of using machine learning algorithms, like LSTM, for stock price prediction, and suggests that further advancements in this area could yield even more fruitful results.

5. Discussion and Conclusion

In conclusion, the LSTM model developed for predicting the 6-month average closing price of a stock using various financial statement parameters and stock price history has demonstrated promising results, highlighting the potential of machine learning techniques in financial analysis. Compared to regular regression models, the LSTM model has shown to capture trends and seasonality more accurately, making it a useful tool for investors looking to predict the future performance of a stock over a 6-month period.

Future research in this area could explore the use of other machine learning techniques and alternative data sources to further enhance the accuracy and predictive power of financial analysis models. Additionally, the

impact of other external factors such as geopolitical events and technological advances on stock performance could be investigated to provide a more complete picture of the complex dynamics of the financial markets.

By analyzing financial indicators such as revenue growth, profit margins, and earnings per share, investors could potentially identify IPOs that are likely to perform well in the market using this model.

It is important to note that while the LSTM model has shown promising results, it is not a guarantee of success in the stock market. Careful analysis and due diligence are necessary when making investment decisions, and the model should be used as a tool in conjunction with other factors.

References

- [1] V. K. Menon, C. Vasireddy, J. N., S. A., V. T. N. Pedamallu, V. Sureshkumar, and K. P. Soman, "2016," *Bulk price forecasting using spark over nse data set. In Data Mining and Big Data: First International Conference, DMBD 2016, Bali, Indonesia, June*, vol. 25-30, pp. 137–146, 2016.
- [2] H. Jia, "Investigation into the effectiveness of long short term memory networks for stock price prediction," *arXiv preprint arXiv:1603*, p. 07893., 2016.
- [3] Z. Hu, W. Liu, J. Bian, X. Liu, and T. Y. Liu, 2018, *February*. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction, 2018.
- [4] A. Ghosh, S. Bose, G. Maji, N. Debnath, and S. Sen, "2019, september," *Stock price prediction using LSTM on Indian share market. In Proceedings of 32nd international conference on (Vol*, vol. 63, pp. 101–110, 2019.
- [5] Y. Lin, S. Liu, H. Yang, and H. Wu, "2021," *Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. IEEE Access*, vol. 9, pp. 101 433–101 446, 2021.
- [6] Y. Huang, L. F. Capretz, and D. Ho, 2021, *December*. Machine learning for stock prediction based on fundamental analysis, 2021.
- [7] K. Puneeth, S. Rudagi, M. Namratha, R. Patil, and R. Wadi, 2021, *December*. Comparative study: Stock prediction using fundamental and technical analysis, 2021.
- [8] B. Bogdanova and E. Stancheva-Todorova, "2021, march," *ML-based predictive modelling of stock market returns. In AIP Conference Proceedings (Vol. 2333, No*, vol. 2333, no. 1, p. 150006, 2021.
- [9] E. K. Ampomah, Z. Qin, and G. Nyame, "2020," *Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. Information*, vol. 11, no. 6, p. 332, 2020.
- [10] H. L. Siew and M. J. Nordin, 2012, *September*. Regression techniques for the prediction of stock price trend, 2012.
- [11] K. V. Kumar and R. Anitha, 2022, *June*. A Detailed Survey to Forecast the Stock Prices by Applying Machine Learning Predictive Models and Artificial Intelligence Techniques, 2022.

- [12] S. N. T. Nishitha, S. Bano, G. G. Reddy, P. Arja, and G. L. Niharika, 2020, November. Stock price prognosticator using machine learning techniques, 2020.
- [13] T. Kumaravel and S. S. Reddy, "Stock price prediction using financial ratios," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 2, pp. 1239–1242, 2017.
- [14] A. Ghosh and M. K. Gupta, "Stock price prediction using financial ratios: Evidence from Indian companies," *IIMB Management Review*, vol. 30, no. 3, pp. 268–280, 2018.
- [15] S. Singh and P. Singh, "Comparative analysis of machine learning techniques for stock price prediction," *International Journal of Emerging Trends in Engineering Research*, vol. 5, no. 10, pp. 332–335, 2017.
- [16] T. R. Bhat and R. K. Jha, "Stock price prediction using machine learning techniques: A survey," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 5, pp. 248–252, 2018.
- [17] S. Mokhtari, K. Yen, and J. Liu, "(2021)," *Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning*. *Journal of Financial Risk Management*, vol. 10, no. 1, pp. 62–78, 2021.
- [18] J. Tellez, D. Ateeq, A. Rafiuddin, H. Alzoubi, T. Ghazal, T. Ahanger, S. Chaudhary, and G. Viju, "(2022)," *AI-Based Prediction of Capital Structure: Performance Comparison of ANN SVM and LR Models*. *Computational Intelligence and Neuroscience*, 2022, vol. 10., pp. 1–13, 2022.
- [19] T. Mengistu, Z. Jibrin, P. Swain, and C. Aditya, "(2021)," *The Role of Artificial Intelligence in Financial Market*. *International Journal of Research in Management, Economics and Commerce*, vol. 11, no. 1, pp. 121–133, 2021.
- [20] A. Petropoulos, V. Siakoulis, K. P. Panousis, L. Papadoulas, and S. Chatzis, "2020," *A Deep Learning Approach for Dynamic Balance Sheet Stress Testing*. *arXiv preprint arXiv:2009*, p. 11075., 2020.
- [21] E. Stancheva-Todorova and B. Bogdanova, "(2021)," *Enhancing investors' decision-making – An interdisciplinary AI-based case study for accounting students*. *AIP Conference Proceedings*, vol. 2333, no. 1, p. 150007., 2021.
- [22] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "(2020)," *Stock Closing Price Prediction using Machine Learning Techniques*. *Procedia Computer Science*, vol. 167, pp. 599–606, 2020.
- [23] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, *September*. Stock price prediction using LSTM, RNN and CNN-sliding window model, 2017.
- [24] A. V. Devadoss and T. A. A. Ligorì, "Forecasting of stock prices using multi layer perceptron," *International Journal of Computing Algorithm*, vol. 2, no. 1, pp. 440–449, 2013.
- [25] J. He and M. Lee, *Data Parsing in Financial Statement Analysis*. Springer, 2019.