Stock Forecasting Accuracy with a Prediction Model

Vyri Junhan Yang*
894431631

Yuhong Wang†
896967720

ABSTRACT

This paper explores the utilization of Long Short-Term Memory (LSTM) networks for forecasting stock prices, focusing on a decade of U.S. stock market data from 2010 to 2020. By leveraging LSTM models, known for their proficiency in handling time-series data, we aim to enhance predictive accuracy and provide a robust tool for strategic investment planning. Our methodology integrates data normalization, feature extraction, and iterative model optimization to effectively predict stock movements. This research not only highlights the potential of LSTM networks in capturing complex market dynamics but also discusses avenues for further enhancement through more comprehensive data inputs and advanced feature engineering. For a comprehensive review of the project, visit our GitHub repository at GitHub Repository.

1 INTRODUCTION

The domain of financial forecasting remains a significant challenge due to the complex and dynamic nature of stock markets. The task of predicting stock prices involves analyzing various patterns and trends in historical data and applying these insights to forecast future movements. The motivation behind this paper is to leverage machine learning, particularly LSTM models, to enhance prediction accuracy in the financial sector, offering a robust tool for strategic investment planning.

Stock market prediction is an area of significant interest to both academic researchers and financial practitioners. It epitomizes a critical component of the global economy where accurate predictions can lead to substantial financial gains or losses. Historically, stock market analysis has relied heavily on quantitative methods and fundamental analysis to predict price movements. However, with the advent of big data and advanced computational technologies, the focus has shifted increasingly towards sophisticated algorithms that can analyze vast datasets more efficiently and accurately.

This paper seeks to explore the application of Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Network (RNN) that is particularly suited to sequences and time-series data. LSTMs are designed to overcome the limitations of traditional RNNs by effectively managing the flow of information, enabling them to remember inputs over long periods, which is ideal for the sequence prediction necessary in stock price forecasting.

Furthermore, the paper is driven by the potential of deep learning techniques to uncover nonlinear relationships and patterns in data that are not readily apparent through traditional statistical methods. By employing these advanced machine learning techniques, the paper aims to provide a more accurate, robust, and scalable approach to stock market forecasting. This could significantly enhance the ability of investors and financial analysts to make informed decisions based on reliable predictive analytics, ultimately contributing to more stable and efficient financial markets.

*e-mail: jyang44@lsu.edu
†e-mail:ywan248@lsu.edu

2 BACKGROUND

Financial forecasting, particularly stock price prediction, has long been a challenging yet essential activity within the financial sector due to the complex and volatile nature of markets. The traditional approach to stock market analysis primarily utilized quantitative methods and fundamental analysis to predict price movements [6]. However, with the growing availability of vast datasets and the development of computational technologies, the focus has increasingly shifted towards utilizing sophisticated algorithms capable of efficiently and accurately analyzing large datasets.

2.1 Evolution from Traditional Methods to Advanced Computational Techniques

Traditionally, stock price predictions were dominated by methods that focused on linear projections of future prices based on historical data. These methods often failed to capture the complex interactions and nonlinear relationships inherent in market data. As identified in recent studies, the adoption of deep learning models, particularly those based on neural networks, has revolutionized financial forecasting by capturing these complex patterns that traditional models often miss [2].

2.2 Advancements in Deep Learning

The advent of deep learning has further revolutionized this field by facilitating the analysis of large datasets with complex, nonlinear relationships that were previously intractable. Deep learning models, particularly deep neural networks (DNNs), have demonstrated superior performance in capturing these relationships due to their deep architecture and multiple layers of abstraction, which allow them to learn from vast amounts of historical data with high variance and intricacy [3].

2.3 Challenges in Stock Price Prediction

Despite these advancements, stock price prediction remains a formidable challenge due to the chaotic nature of the financial markets, influenced by countless unpredictable factors such as economic indicators, political events, and market sentiment. This unpredictability makes the financial markets a complex system where price movements are not only based on past prices but also on an amalgam of diverse, dynamic factors that can change rapidly and without warning.

2.4 Role of LSTM Networks

LSTMs are particularly well-suited for this task as they can capture temporal dependencies and patterns over long periods, which is crucial for accurate prediction in time-series data like stock prices. Unlike traditional RNNs, LSTMs are designed to avoid the long-term dependency problem by incorporating mechanisms like forget gates, which allow the model to retain or discard information dynamically. [5] This capability enables LSTMs to learn from the long sequences necessary for predicting stock market trends, providing a more refined analysis compared to earlier models [4].

2.5 Objectives of the Paper

This paper aims to harness the potential of LSTM networks to enhance predictive accuracy in the stock market [8]. By integrating deep learning techniques with traditional financial analysis methods,

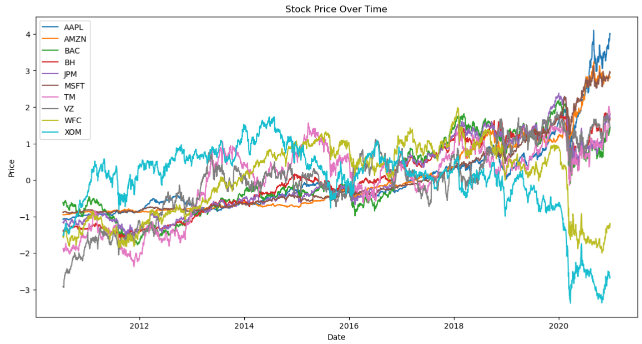

Figure 1: Standardizing the closing prices of stocks to have a mean of zero and a standard deviation of one. This process helps mitigate the effect of scale differences across various stocks and makes the training process more stable.

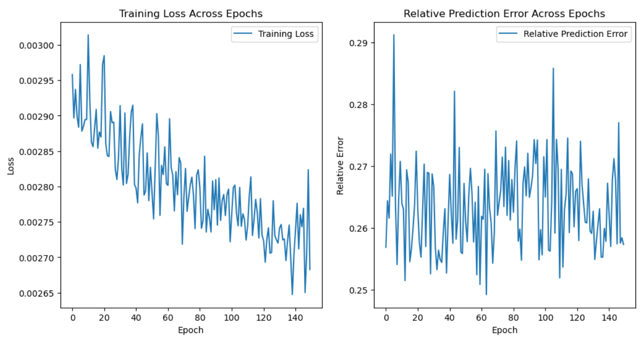

Figure 2: Comparison of Training Dynamics Over Epochs. The left panel depicts the training loss across 150 epochs, illustrating the model's gradual improvement and fluctuation in performance. The right panel shows the relative prediction error over the same epochs, highlighting variability in prediction accuracy. These metrics are crucial for understanding model convergence and behavior under training conditions.

the paper seeks to develop a robust tool for investors and financial analysts, facilitating more informed and strategic investment decisions. The ultimate goal is to contribute to more stable and efficient financial markets through improved predictive analytics, thereby supporting economic growth and financial stability.

In conclusion, the transition from traditional econometric models to sophisticated deep learning approaches represents a significant shift in the methodology of financial forecasting. By exploring and implementing LSTM networks, this paper not only addresses the complex nature of financial time series data but also enhances the analytical capabilities necessary for effective prediction and strategic decision-making in the financial sector.

3 Methodology

Our methodology for stock price forecasting involved a detailed analysis and preparation of historical stock data, emphasizing normalization and feature engineering for LSTM model training. We focused on U.S. stock data from 2010 to 2020, implementing rigorous data cleaning and feature selection to train the LSTM model effectively. This model was intricately designed to process and predict stock movements accurately, with continuous refinements through iterative training and optimization to enhance its predictive performance.

3.1 Approach

Our approach to stock price forecasting involved a methodical examination of stock price data over a defined period. We emphasized the performance of various stocks by tracking their price changes and applying a standardized method to normalize the data. The Long Short-Term Memory (LSTM) model was chosen for its proficiency in handling time series data, a characteristic that is crucial for accurately predicting stock price movements.

The methodology adopted for this paper goes beyond simple data analysis and enters the realm of predictive modeling. By integrating data normalization, feature extraction, and LSTM-based modeling, we aimed to create a robust predictive system capable of understanding and anticipating market dynamics.

3.2 Data Collection and Preparation

The data for this paper was sourced from publicly available historical stock prices listed in the Forbes Global 2000 list [1], focusing specifically on companies headquartered in the United States to maintain consistency in market closure times. The timeframe spanned from July 2010 to December 2020, providing a comprehensive dataset that includes various market conditions such as bull and bear markets, market corrections, and periods of economic stability and turmoil.

Data preparation involved several crucial steps: **Data Cleaning**: Ensuring the dataset was free from missing values and anomalies, which could lead to inaccuracies in prediction. **Data Normalization**: Standardizing the closing prices of stocks to have a mean of zero and a standard deviation of one. This process helps mitigate the effect of scale differences across various stocks and makes the training process more stable. **Feature Engineering**: Extracting relevant features that could potentially influence stock prices. Using the sliding window approach, we analyzed the closing prices of stocks for the past 20 days as the primary features [7]. Additional features, such as moving averages and percentage change, were tested to enhance the model's learning capability.

3.3 Model Development and Training

The LSTM model was developed to capture the temporal dependencies in stock price movements. The architecture of the LSTM included several layers designed to process sequences of data effectively. The model was structured to include: **Input Layer**: Accepting normalized data as input. **LSTM Layers**: One or more LSTM layers to process the input data over time. **Dense Layers**: One or more fully connected layers that interpret the LSTM outputs before making a prediction. **Output Layer**: Producing the final prediction of the next day's stock price.

Training the LSTM model involved splitting the dataset into training and testing sets, using approximately 80% of the data for training and 20% for testing. This split was crucial for evaluating the model's performance on unseen data, simulating a real-world scenario where the model would be used to predict future stock prices.

The LSTM model was intricately designed to address the complexities of stock price time series. The development process was centered around a robust feature engineering strategy and the effective utilization of PyTorch for deep learning tasks.

Feature Engineering: Primary features were derived from the closing prices of stocks for the previous 20 days, acknowledging the importance of short-term price movements in stock predictions. Additionally, data was pre-processed to fit the model's requirements, including: **Conversion to PyTorch Tensors**: Data was converted into tensors, with *unsqueeze(1)* used to add an extra dimension necessary for LSTM input. **Creating Data Loaders**: Employed PyTorch's DataLoader with a batch size of 64 and enabled shuffling to prevent overfitting, facilitating efficient and effective data handling during model training.

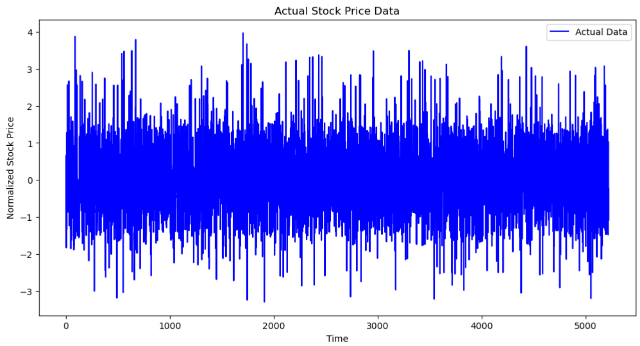

Figure 3: Actual Stock Price Data.

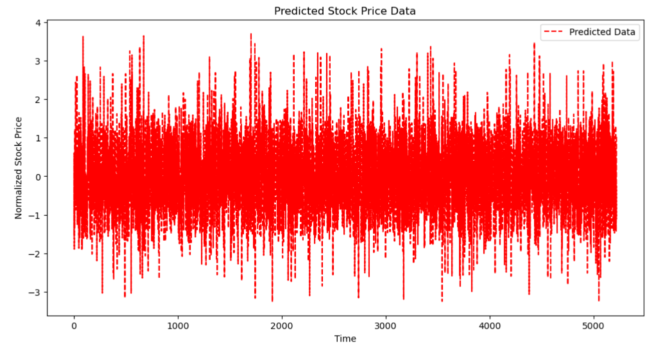

Figure 4: Predicted Stock Price Data.

LSTM Model Configuration: The architecture consisted of LSTM layers coupled with fully connected layers that process and predict from the sequential data: **Initialization**: The model was initialized to set appropriate sizes for input, hidden layers, and output, ensuring that the network structure was optimally configured for the complexity of the data. **Training and Optimization**: The training involved forward propagation to generate predictions, loss calculation using metrics such as Mean Squared Error (MSE), and backward propagation for updating model weights. Parameters such as learning rates, hidden layer sizes, and layer counts were tuned to enhance performance. The model used a range of *learning rates* [0.01, 0.001, 0.0001] and explored *hidden sizes* [50, 100, 150], with the best parameters being a *learning rate* of 0.001, a *hidden size* of 100, and two layers.

3.4 Evaluation and Iteration

After initial training, the model's performance was assessed using the test dataset. Key performance indicators included prediction accuracy, loss metrics, and the ability to generalize to new data. The iterative process involved refining the model through multiple training cycles, adjusting hyperparameters, and incorporating additional features as needed based on performance feedback.

The LSTM model underwent rigorous testing and evaluation over 150 epochs to optimize its performance and ensure its predictive accuracy. **Performance Metrics**: The system was evaluated based on its predictive accuracy and the ability to generalize on unseen data. Metrics such as average loss and relative prediction error were monitored throughout the training period. **Iterative Optimization**: The model was iteratively refined, adjusting hyperparameters and re-evaluating the training strategy based on ongoing performance metrics. This iterative approach helped in minimizing loss and enhancing the model's ability to forecast future stock prices accurately.

4 Data Analysis and Results

This section delves into the evaluation and refinement of a LSTM model for stock price prediction. It begins with a direct comparison between actual and predicted data, shedding light on the model's performance and identifying areas for improvement. The subsequent analysis assesses the model's overall performance and binary outcome accuracy, providing a baseline for its practical applicability. The chapter progresses to discuss the iterative optimization process undergone by the model, showcasing its ability to adapt and improve over time. Finally, it explores the potential for further enhancement, suggesting strategies such as incorporating more diverse data, refining feature engineering techniques, and adjusting the model's architecture to better capture complex price patterns.

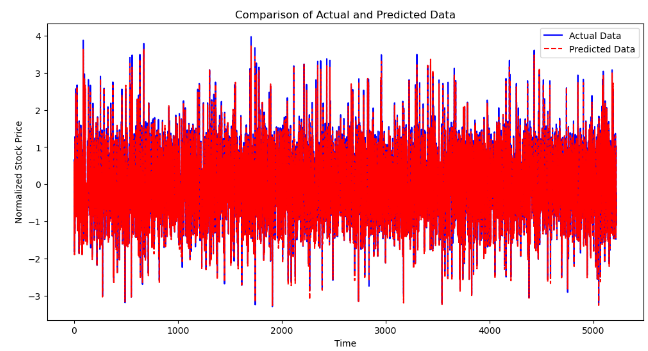

Figure 5: This graph illustrates the normalized stock prices over time, showing significant volatility and fluctuation. This visualization serves as a basis for evaluating the LSTM model's performance in predicting stock price trends, as discussed in the comparative analysis.

4.1 Comparison of Actual and Predicted Data

The analysis focused on the direct comparison between the actual stock prices (Fig. 3) and those predicted (Fig. 4) by the LSTM model over the test period. The visual representations provided clear insights into the model's ability to trace the real price trends closely, albeit with some deviations(see Fig. 5). This comparative study was crucial for validating the model's effectiveness and identifying areas where the predictive performance could be further enhanced.

4.2 Model Performance and Binary Outcomes

The model's overall performance was quantified through standard metrics. The LSTM model achieved a test loss of 0.6998 and a binary outcome accuracy of 50.67%. This accuracy rate, reflecting the model's ability to correctly predict the direction of stock price movements (up or down) on the next day, serves as a baseline for the effectiveness of the predictive system in practical trading scenarios.

4.3 Progressive Optimization and Results Analysis

Throughout the training process, the model underwent multiple iterations to optimize performance, focusing on minimizing average loss and reducing relative prediction error. These efforts resulted in progressive improvements, demonstrating the model's capacity to adapt and refine its predictions with continued training.

4.4 Potential and Improvement

The project highlighted the substantial potential of deep learning, particularly LSTM models, in predicting stock market trends. The results also suggested several avenues for further improvement:

Incorporating More Data: Including more diverse datasets, such as higher-frequency trading data or additional financial indicators, could improve the model's understanding and forecasting accuracy. **Refining Feature Engineering**: Enhancing the feature set by exploring more sophisticated financial indicators and incorporating macroeconomic factors could provide the model with more robust predictive signals. **Adjusting Model's Architecture**: Modifying the LSTM architecture, such as increasing the number of layers or experimenting with different types of recurrent neural networks, might enhance the model's ability to capture complex patterns in stock price movements.

5 Conclusion

The research project demonstrated the effective application of LSTM models in the domain of financial forecasting, specifically for predicting stock price movements. Through rigorous data preparation, model training, and iterative optimization, the LSTM model achieved a foundational level of predictive accuracy, illustrated by its performance on historical stock price data. While the model provided valuable insights, its predictive accuracy suggests substantial room for improvement. Future work will focus on expanding the dataset, refining the feature engineering process, and optimizing the LSTM architecture to enhance the model's ability to discern and predict complex stock price patterns more accurately. This ongoing refinement and expansion of the LSTM application underscore its potential to significantly impact financial analytics and investment strategies.

Acknowledgments

This work was supported in part by a grant from School of Electrical Engineering and Computer Science, Louisiana State University.

References

[1] Forbes global 2000 list. https://www.forbes.com/sites/forbesstaff/2022/05/12/forbes-global-2000-list-2022-the-top-200/?sh=3c75e7873290. Accessed: 2024-04-16.

[2] J. Heaton, N. G. Polson, and J. H. Witte. Deep learning in finance. *arXiv preprint arXiv:1602.06561*, 2016.

[3] Z. Hu, Y. Zhao, and M. Khushi. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1):9, 2021.

[4] S. Kumar, N. Pal, and A. M. Tripathi. Improving long short-term memory (lstm)-based stock market price predictions in the machine learning era. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, vol. 5, pp. 923–928. IEEE, 2024.

[5] N. Nguyen and M. M. Islam. Comparison of financial models for stock price prediction. In *2021 Joint Mathematics Meetings (JMM)*. AMS, 2021.

[6] P. Paul and R. Mukherjee. Current trends in financial forecasting. In *2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, pp. 1–5. IEEE, 2023.

[7] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman. 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, 2017.

[8] P. Vispute, J. Sujata, and N. Natraj. Predicting stock prices using machine learning techniques: An analysis of historical market data. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 725–733. IEEE, 2023.