Predicting Stock Prices with Machine Learning and Virtual Stock Market Game

Priyanshu Mavale
Department of Information Technology
Vidyalankar Institute of Technology
Mumbai, India

Mohit Sharma
Department of Information Technology
Vidyalankar Institute of Technology
Mumbai, India

Harshad Phadatare
Department of Information Technology
Vidyalankar Institute of Technology
Mumbai, India

Prof. Ajitkumar Khachane
Department of Information Technology
Vidyalankar Institute of Technology
Mumbai, India

***Abstract*- This innovative stock trading simulation platform offers users an immersive experience in the world of stock market investing, using virtual currency. Featuring a user-friendly interface, individuals can easily grasp the essentials of purchasing and selling stocks, managing their investment portfolios, and making informed decisions within a secure, risk-free setting. The virtual currency system ensures that users can refine their trading abilities without any financial jeopardy. Moreover, the simulator delivers real-time market insights, educational materials, and analytical tools to support users in reaching their investment objectives. Our primary project objective is to gamify the experience and foster collaboration among friends within dedicated rooms, enabling them to assess their comparative skills. Users can leverage predictive models, trained using LSTM algorithms, for investment purposes. Extensive testing, employing various prediction algorithms like LSTM, Gradient Boosting Regression, and Random Forest Regression, revealed LSTM's superior performance in stock price forecasting, establishing its efficiency and reliability.**

Keywords—Virtual Currency, Gamification, Prediction Models, Long Short Term Memory (LSTM)

I. INTRODUCTION

Numerous individuals lack familiarity with the intricacies of the stock market, hindering them from capitalizing on its potential opportunities. Common misunderstandings include the notion that stock market participation is solely reserved for the affluent or too convoluted for the average individual to comprehend. Additionally, apprehensions regarding potential financial losses or uncertainties about making informed investment choices may deter some from investing. By dedicating time to understand the stock market and its advantages, individuals can make well-informed decisions regarding investment, thereby progressing toward their financial objectives.

Our paper introduces a platform where users can invest and compete with friends in a stock trading simulation utilizing virtual currency. Catering to both novices and seasoned traders, this platform offers an enjoyable and interactive environment for honing investment skills. Collaborate with friends to construct portfolios, monitor progress, and execute informed trades in a simulated stock market. Equipped with real-time market data and a user-friendly interface, our simulator provides a risk-free avenue to refine trading strategies. Embark on your stock trading journey today and ascertain the top performer among your circle of friends.

Historically, predicting stock prices has posed a formidable challenge for investors and traders. However, recent advancements in artificial intelligence have enhanced the accuracy of stock price forecasting. Introducing a cutting-edge stock price prediction model harnessing Long Short-Term Memory (LSTM) neural networks. Leveraging historical stock data, our model generates forecasts for future prices. Boasting remarkable precision and the capacity to decipher intricate financial data relationships, our LSTM model equips investors and traders with invaluable insights to guide their investment decisions.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Welcome to our stock consulting blog, where seasoned investors provide expert advice, analysis, and insights on the stock market and investing. Our team is committed to furnishing valuable information and guidance to assist you in navigating the multifaceted realm of investing. From market trends and stock selections to investment strategies and risk management, our blog covers a diverse array of topics to empower individuals, regardless of experience level, to succeed in the stock market.

RELATED WORKS

The Sayavong Lounnapha et al. [1] explore the application of Convolutional Neural Networks (CNNs) in stock price prediction. They gathered historical stock data and employed it to train a CNN model for forecasting future stock prices. Data preprocessing involved normalization and partitioning into training, validation, and testing sets. The model underwent training via backpropagation, utilizing a mean squared error loss function. The study concludes by demonstrating the effectiveness of CNNs in stock price prediction, surpassing conventional time series prediction methods. It underscores the potential of deep learning techniques in finance and emphasizes the significance of hyperparameter tuning for optimal performance.

Ferdiansyah et al. [2] present a study on utilizing Long Short-Term Memory (LSTM) networks to predict Bitcoin prices. Historical Bitcoin price data from Yahoo Finance served as the basis for training and evaluating an LSTM model. Data preprocessing included normalization and division into training, validation, and testing sets. The LSTM architecture comprised multiple LSTM layers and fully connected layers. Training employed backpropagation with a mean squared error loss function. The study concludes by showcasing the effectiveness of LSTM networks in Bitcoin price prediction, outperforming traditional time series prediction methods. It highlights the potential of deep learning techniques in finance and underscores the importance of hyperparameter tuning for optimal performance.

Li X, Wu P, and Wang W [3] present a study on incorporating both stock prices and news sentiments for stock market prediction. They collected historical stock prices and news sentiments from the Hong Kong stock market for training and evaluating a machine learning model. Data preprocessing involved normalizing stock prices and computing sentiment scores for news articles. The designed machine learning model integrated both stock prices and news sentiments as inputs for predicting future stock prices. Training utilized backpropagation.

II. Our Proposed System *A. PROJECT FLOW DIAGRAM :*

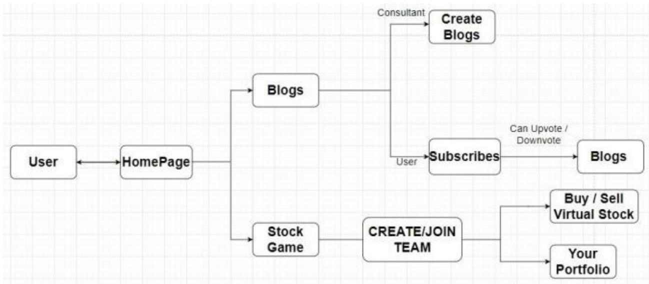

Figure (a) Project Flow

This defines the Project Flow Diagram of our entire project which overall comprises three phases: Game platform, Subscription based blogs, Machine Learning Algorithms.

B. GAME PLATFORM

Online group virtual stock games simulate the stock market environment, enabling users to form groups and trade virtual stocks using virtual currency. These games foster collaboration, idea sharing, and friendly competition among participants.

Within these virtual stock games, users have the option to create or join groups, engaging in trading activities alongside their friends. Together, they navigate the complexities of the stock market, learning from both successes and setbacks. The game monitors each individual's virtual portfolio, providing real-time updates on gains or losses based on the performance of virtual stocks. Periodic ranklists are generated for individual rooms, reflecting participants' standings relative to the current stock prices.

Such group-based virtual stock games serve educational purposes, offering students a hands-on experience in understanding the stock market and investment dynamics while fostering teamwork. Additionally, novice investors can leverage these platforms to gain practical experience and insights from peers before venturing into real-money investments.

C. SUBSCRIPTION BASED BLOGS

Subscription-based stock market blogs offer invaluable resources for individuals seeking to remain abreast of market trends and investment prospects. Managed by financial professionals and analysts, these blogs furnish expert insights and recommendations concerning stocks, bonds, and various investment instruments.

Typically, these blogs furnish timely updates on stock market happenings, encompassing breaking news, earnings disclosures, and other pertinent data influencing stock valuations. Consequently, subscription-based stock market blogs serve as a crucial resource for investors keen on staying well-informed and making prudent investment choices. It's imperative to conduct thorough research and select a blog aligning with your requirements and financial constraints. ve

D. MACHINE LEARNING ALGORITHM FOR PREDICTING STOCK PRICES:

1. Algorithmic Details :

LSTM (Long Short-Term Memory) stands as a notable variant of recurrent neural networks (RNNs), extensively employed in forecasting stock prices. Its adeptness in time-series analysis derives from its capability to retain and process historical data while incorporating present inputs. This inherent trait renders LSTM proficient in modeling and predicting time-dependent stock prices.

Several pivotal features characterize LSTM:

- Memory Cell: LSTM incorporates a memory cell adept at storing information across extended durations, facilitating the retention of long-term dependencies within input sequences.
- Gates: Comprising input, forget, and output gates, LSTM regulates the influx and efflux of information to and from the memory cell.
- Nonlinear Activation Function: Employing nonlinear activation functions like the hyperbolic tangent (tanh), LSTM transforms input and output data, enabling the capture of intricate relationships.
- Backpropagation through Time (BPTT): LSTM undergoes training via BPTT, a variant of the backpropagation algorithm tailored for RNNs. This mechanism updates the model's parameters based on the disparity between predicted and actual outputs.
- Deep LSTM: LSTM configurations can be layered to construct deep LSTM networks featuring multiple LSTM layers. This architecture enables the acquisition of hierarchical representations of input data, facilitating the discernment of complex patterns.

Consequently, LSTM was utilized to prognosticate stock prices for the subsequent 90-100 days. The fundamental principle underlying LSTM revolves around leveraging a memory cell capable of retaining information across protracted intervals.

In an LSTM framework, each gate is represented by a sigmoid function, yielding values within the range of 0 to 1. These sigmoid functions serve to modulate information flow, with values close to 0 indicative of gate closure, impeding information passage, and values near 1 signaling gate opening, facilitating information transmission.

Moreover, the memory cell undergoes modification via a tanh function, producing outputs within the -1 to 1 range. This function orchestrates the adjustment of cell states by augmenting or diminishing information within the cell. In the context of time-series prediction, LSTM harnesses past values to forecast future outcomes.

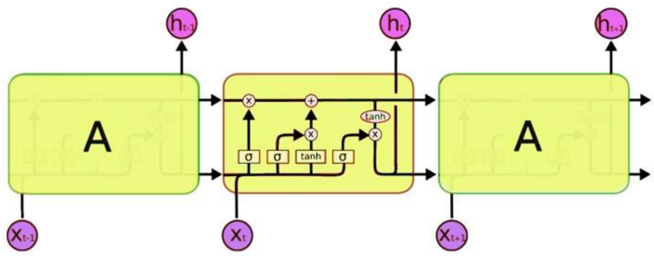

Figure (b) LSTM Algorithm

2. Architecture of Proposed Model

The architecture of the proposed work is shown in the below figure .

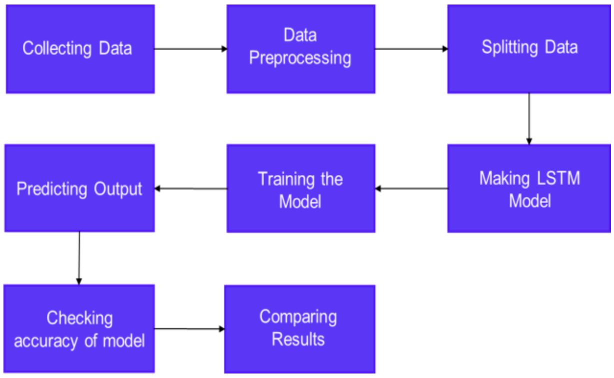

Figure (c) Data Flow Diagram

● Data Collection: Initial data acquisition involves gathering stock details such as opening and closing prices from Yahoo Finance.

● Data Pre-Processing: Subsequent to data collection, preprocessing is undertaken to eliminate redundant fields or empty columns, streamlining the dataset.

● Data Splitting: The dataset is partitioned into training and testing sets to facilitate the creation of algorithmic models

● LSTM Model Training: The training dataset is fed into an LSTM model to impart learning and develop predictive capabilities.

● Output Prediction: Following model training, predictions are generated for the testing datasets, envisioning stock prices for the forthcoming 3-6 months.

● Accuracy Evaluation: The accuracy of our algorithm is assessed based on the performance of predictions on the testing dataset.

3. DATASET DESCRIPTION

For training our machine learning model, we utilize data sourced from Yahoo Finance, comprising diverse stock price information such as open, high, low, close, and volume for any given date range. Specifically, we focus on the "Date" and "Closing Price" fields, crucial for both training and prediction tasks. The dataset is dynamically updated to reflect the latest trading information, ensuring the models remain current. This data forms the basis for training models capable of forecasting stock prices for the subsequent 90-100 days.

The dataset encompasses details pertaining to various stocks, offering insights into future stock price movements. Each dataset entry represents a distinct stock and encompasses the following attributes:

- Date: Denoting the trading day's date.

- Open: Reflecting the asset's opening price on the trading day.

- High: Signifying the peak price reached by the asset during the trading day.

- Low: Representing the lowest price attained by the asset on the trading day.

- Close: Indicating the asset's closing price on the trading day.

- Adj Close: Reflecting the adjusted closing price, accounting for factors like stock splits and dividends.

- Volume: Signifying the total volume of shares or contracts traded on the trading day.

- Symbol: A unique symbol utilized to identify the asset.

- Market Capitalization: Representing the total value of all outstanding shares of a company's stock.

- Dividend Yield: Calculated as the annual dividend payment divided by the current stock price.

- Price-to-Earnings (P/E) Ratio: Derived from the price per share divided by the earnings per share over the past 12 months.

- 52-Week High: Denoting the highest price attained by the asset in the preceding 52 weeks.

- 52-Week Low: Reflecting the lowest price reached by the asset in the preceding 52 weeks.

● weeks.

4. IMPLEMENTATION

Pre-processing of the dataset involves several essential steps:

1. Data Collection: Initially, data is sourced from Yahoo Finance, comprising various stock details such as opening and closing prices for a specified date range.

2. Data Cleaning: Following data collection, cleaning procedures are implemented to rectify errors, remove duplicates, and address outliers. Techniques like handling missing values, eliminating duplicates, and managing outliers are applied to ensure data integrity.

3. Data Transformation: Data transformation encompasses scaling and normalization processes to standardize data values, ensuring they fall within a consistent range.

4. Feature Selection: Relevant variables crucial for analysis are selected to optimize model performance. This involves identifying variables likely to have a significant impact on the predictive outcomes.

5. Data Splitting: The dataset is divided into training and testing sets to evaluate model performance accurately. The training set is utilized to train the model, while the testing set is employed to assess the model's efficacy on unseen data.

For stock price prediction:

1. Algorithm Selection: Multiple machine learning algorithms are evaluated to determine the most effective approach. Following performance assessment, the LSTM algorithm is chosen for its superior predictive capabilities.

2. Data Preprocessing for LSTM: Preprocessing steps are executed to prepare data for LSTM model training. Historical data is dynamically fetched and preprocessed, with emphasis on scaling using the Min-max scaler to standardize price values. The "Date" column along with its corresponding closing prices are utilized for training and prediction purposes.

3. Training Dataset Creation: A sliding window approach is employed to construct the training dataset for the LSTM model. Sequential iterations through the dataset generate input-output pairs, with 99 values utilized as input (x_totrain) and the 100th value serving as the output (y_totrain).

4. LSTM Model Construction: Leveraging the Keras API, a sequential LSTM model is constructed comprising multiple layers. The model architecture includes three LSTM layers followed by a Dense layer, facilitating predictions. Each LSTM layer comprises 50 units with a dropout of 0.2, while the return_sequences parameter is set to True for the first two LSTM layers to yield sequence outputs.

5. Model Compilation and Training: The model is compiled with 'mean_squared_error' loss function and 'adam' optimizer, tailored for regression problems. Subsequently, the model is trained using the fit() method, updating parameters to minimize the loss function.

6. Visualization of Predictions: Finally, the trained model is employed to visualize predicted stock price values, enabling insights into future price trends.

IV. RESULT

In our project, we conducted a comprehensive analysis of various algorithms to predict stock prices. Through extensive research, we determined that LSTM stands out as the most efficient algorithm due to its ability to store information over extended periods, making it well-suited for time-series data like stock prices.

Additionally, we developed a stock game simulator utilizing virtual currency, aimed at providing users with an engaging and interactive platform to learn about the stock market. This simulator incorporates features to enhance the trading experience, such as the creation of blogs by seasoned investors or financial experts. Furthermore, we introduced the functionality to create game rooms, allowing users to trade stocks with friends and gain hands-on experience with virtual money before venturing into real investments. These features contribute to creating an interactive space for users to engage in virtual trading while also accessing insights from experts through blogs.

Furthermore, we conducted a comparative analysis of various machine learning algorithms for stock price prediction, where LSTM emerged as the top performer based on Root Mean Square Error (RMSE) values. This rigorous testing allowed us to identify LSTM as the optimal algorithm for accurate stock price forecasting.

Overall, our project integrates advanced machine learning techniques with interactive features to provide users with an immersive learning experience in the stock market, facilitating both education and practical trading in a virtual environment.

Machine Learning Algorithms	RMSE
LSTM Algorithm	11.93
Random Forest Regression	13.02
Gradient Boosting Regression	15.08

Figure (d) Comparison of Machine Learning Algorithm

V. Future Work

Implementing a real currency system entails several key components. Firstly, a secure payment gateway must be established to facilitate safe transactions. Additionally, a robust bid tracking system is necessary to manage bids effectively and ensure transparency.

Fair transaction algorithms are crucial to maintain integrity and prevent fraud. Offering multiple payment options, including credit cards and PayPal, enhances user accessibility. Mobile compatibility is essential to accommodate users accessing the platform via smartphones or tablets.

A user-friendly interface with clear instructions simplifies navigation and enhances user experience. Regular bug fixes and prompt customer support are essential for addressing technical issues and providing assistance to users. Overall, these measures ensure the reliability, security, and user-friendliness of the real currency system.

VI. Conclusion

In conclusion, engaging in a stock price game using virtual currency offers a dynamic and educational avenue for exploring the world of investing without financial risk. Participants can cultivate valuable skills in stock analysis, decision-making, and portfolio management while enjoying a competitive and collaborative environment. Nonetheless, it's crucial to acknowledge the limitations of virtual trading, as it may not always mirror real-market conditions accurately. Strategies employed in games may not directly translate to real-life investments. Thus, while virtual trading serves as a valuable learning tool, it should complement real-world experience and guidance from professionals for a comprehensive understanding of investment practices.

VII. Acknowledgment

We would like to acknowledge our ”HOD Vipul Dalal”, for supporting us throughout the project. Thanks to ” Prof. Ajitkumar Khachane” for her guidance and efforts. The project “Stock Street” would not have been possible without them. We are Thankful to “Vidyalankar Institute of Technology” who provided us with such a stimulating atmosphere for us to work on our project. And lastly, we express our gratitude to everybody who has been the source of inspiration throughout the project.

VIII. References

[1]. Research on Stock Price Prediction Method Based on Convolutional Neural Network, IEEE 2019 - Sayavong Lounnapha et al.

[2]. An LSTM-Method for Bit-coin Price Prediction: A Case Study Yahoo Finance Stock Market, IEEE 2019- Ferdiansyah et al.

[3]. Li X, Wu P, Wang W. 2020. Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong.

[4]. Yu, H., Wang, X., Xie, Y., & Yang, X. (2020). Stock price prediction based on LSTM with multiple technical indicators. Journal of Physics: Conference Series, 1668(1), 012013.

[5]. Singh, N., Kumar, A., & Kumar, P. (2020). Predicting stock prices using artificial neural networks: A survey. Journal of King Saud University-Computer and Information Sciences, 32(4), 434-441.

[6]. Zhang, S., & Jin, Y. (2021). Stock price prediction based on LSTM with technical indicators and textual data. Applied Soft Computing, 103, 107184.

[7]. Chen, X., Sun, H., & Zhang, X. (2020). An improved stock price prediction method based on
CNN-LSTM. Journal of Physics: Conference Series, 1558(1), 012032.

[8]. Song, X., Wu, Y., & Zhang, Y. (2019). A stock price prediction method based on the LSTM model. Journal of Intelligent and Fuzzy Systems, 37(3), 3173- 3183.

[9]. Li, X., Li, Q., Li, J., & Li, J. (2020). Forecasting stock price with technical indicators and sentiment analysis using deep learning. Neural Computing and Applications, 32(7), 2691-2701.