SpiderNet Project Proposal: An AI Model to Predict Stock Market Prices

Savannah Stephenson, Janavi Bahalala, Christian Duff, Allision Adams, Alicia Mares

***Abstract*—The stock market is a complex and dynamic system, driven by a number of factors including economic indicators, investor behavior, and global events. To address the problem of the inherent volatility and unpredictability of the stock market, we will apply AI models to predict stock trends, which may potentially offer a more reliable system of prediction. In this paper, we will explore existing research in the field of stock market prediction, outline the team's approach to the problem, give an overview of the architecture used for our AI model, and present our findings on the effectiveness of our AI model in predicting the stock price of individual stocks.**

I. Introduction

THE stock market is a complex and dynamic system, driven by a number of factors including economic indicators, investor behavior, and global events. Predicting its fluctuations is a challenge; traditional methods of stock market prediction often rely on historical trends, statistical analysis, and expert opinion, which can be subject to human biases and errors. However, recent advancements in artificial intelligence (AI) present new opportunities to approach stock market forecasting in a more coherent and unbiased manner.

Our project focuses on developing an AI-based time series prediction model for Apple stock prices. We will utilize deep learning techniques to analyze historical stock price data from the online dataset repository Kaggle, use TensorFlow in Python to set up and train the model, and use Google Colab for development. The output of the model will be a single numerical value representing the low stock price prediction for that day. We will then visualize these predictions against actual stock market values, and provide an assessment of the model's accuracy.

The motivation for utilizing AI in stock market predictions stems from its ability to detect complex patterns that may not be immediately apparent to human analysts, potentially offering a more reliable prediction system. By avoiding emotional biases and increasing the consistency of predictions, AI can provide a wealth of insight to traders, investors, and financial institutions.

We will address the problem of inherent volatility and unpredictability of the stock market, where applying AI models such as LSTMs or ANNs can help improve accuracy in predictions. This approach holds significance in a world where financial markets have global impacts, and even slight improvements in prediction accuracy can turn into significant financial benefits.

In this paper, we will explore existing research in the field of stock market prediction, outline the architecture used for our AI model, and present our findings on the effectiveness of our AI model in predicting the stock price of Apple.

II. Related Work

SPIDERNET conducted a literature review in order to gain insight about Stock Market Prediction; below are the summaries of the article that the team reviewed.

A. Stock Price Prediction Using Machine Learning with Source Code [5]

Predicting stock prices is complex due to many influencing factors making high-accuracy predictions difficult. To address this challenge, we use Machine Learning algorithms like RNNs and LSTMs. Techniques like Moving Average (SMA & EMA) focus on past data points, weighing recent ones more. Short-term predictions offer probabilistic market models. RNNs and LSTMs enhance the predictive power of time-series data. The historical stock data from Yahoo Finance is analyzed using Python and its libraries (Pandas, TensorFlow and Keras). The models are trained to weigh recent vs older data for data price forecasting. This article provides a practical guide for understanding stock price forecasting using advanced machine learning algorithms like RNNs and LSTMs. It simplifies the complex of nature of predictions by integrating different techniques and using Python libraries like Pandas, TensorFlow, and Keras.

B. Stock Market Prediction base on Statistical Data Using Machine Learning Algorithms [4]

The goal of the study was to identify the most effective machine learning model for predicting stock prices with high accuracy. For example: Random Forest and Support Vector Machine (SVM). The article does two types of analysis: Technical Analysis and Fundamental Analysis. The article used historical stock price data, such as open price, high price, low price, close price from Yahoo Finance. As well as various financial indicators, such as Relative Strength Index (RSI), Moving Average Convergence/Divergence), and other were computed to train Machine Learning (ML) models.Additionally, the article sought to predict stock price trends at the end of each business day. Fundamental Analysis was used to analyze public sentiment based on Twitter data, aiming to predict stock price movement by classifying positive or negative sentiments. To measure AI's ability, the model was trained, where the data was divided into training, validation, and testing sets. Various training algorithms like Decision

Trees (DT), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) were trained and tested. Evaluation metrics for regression models predicting stock prices include the following: (i) R-squared (R^2): which measures how well predictions match actual data. (ii) Mean Absolute Percentage Error (MAPE): shows the average deviation of predictions from actual values. (iv) Root Mean Squared Error (RMSE): indicates magnitude of prediction errors. In technical analysis, the linera regression model outperformed others. In fundamental analysis, the SVM model had the best performance in classufying public sentiment. In summary, the AI looks at past stock prices tweets to predict future stock prices ot action (buy,sell, or hold). To see if AI is predicting accurately, we measure R^2 (how well AI fits the real data), MAPE (average error), and RMSE (how bug the mistakes are). This article provided a comprehensive overview of different machine learning techniques for stock price prediction backed by statistical analysis and real-world data. It equips readers with practical insights into how to apply these model effectively.

C. Stock Market Prediction Using LSTM Recurrent Neural Network [2]

This article is focused on building a model using Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM), to predict future stock market values. The article compares Long Short-Term Memory (LSTM) networks to ARIMA, showing LSTSM's superiority in predicting stock prices. LSTM is a type of Recurrent Neural Network (RNN) with the ability to memorize data sequences to forecast the adjusted closing prices of a portfolio. An Artificial Neural Network (ANN) comprises of an input layer, hidden layers, and an output layer. The dataset that are used in this article are of two stocks: Google and Nike from the New York Stick Exchange obtained from Yahoo Finance. In summary, this article evaluates how precisely machine learning can predict stock prices and how increasing the number of training epochs improves the model's accuracy. Ultimately, the article provided insights into effectiveness of LSTM networks for stock price prediction, showcasing their advantages over traditional models. It offers practical guidance on improving predictive accuracy through training techniques.

D. Effectiveness of Artificial Intelligence in Stock Market based on Machine Learning [1]

This article explored Stock Marlet Prediction (SMP) by utilizing Artificial Intelligence (AI) through two main approaches: Technical Analysis and Fundamental Analysis. In technical analysis, Machine Learning (ML) regression algorithms are applied to historical stock prices from Yahoo Finance to forecast daily trends. In fundamental analysis, Machine Learning (ML) classification are used to analyze public sentiment, primarily through Twitter data, to predict stock market movements. By combining these two approaches, the paper investigates how historical data and public sentiment can influence stock predictions. The results showed moderate success, indicating that AI has potential but is is not yet advance enough to consistently outperform stock market. This article provided a comprehensive overview of how AI and machine learning can be applied to stock market prediction through both technical and fundamental analysis, and highlighted the potential and limitations of current AI technologies in this field.

E. The Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions [3]

The rise of global digitization has transformed Stock Market Prediction, modernizing traditional trading models. As market capitalization grows, crock trading has become a key investment avenue for financial investors. Analysts have developed tools to forecast stock price movements, aiding investor in making informed decisions. Advanced trading models utilize non-traditional data from social media, while machine learning techniques like text data analytics have significantly improved prediction accuracy. However, stock market analysis remains a complex research challenge due to its dynamic and unpredictable nature. This article outlines machine learning approaches for stock prediction within a generic framework, analyzing findings from 2011 to 2021 sourced from digital libraries like ACM and Scopus. Additionally, a comprehensive comparative analysis highlights significant trends, pricing emerging researchers with a foundational understanding to pursue further research in this evolving field. Some of the keywords that are used in this article are: machine learning, Support Vector Machine (SVM), and generic review. By offering a timeline of the advancements in machine learning for stock market prediction, the article serves as a valuable resource for researchers and practitioners looking to navigate and contribute to this dynamic area of study.

F. Literature Review Conclusion

Through these articles, our team gained valuable insights into various techniques used in AI algorithms towards stock market predictions which aided in deciding on our next steps in our approach.

III. Proposed Approach

TO find the best model architecture for the stock market prediction problem our team proposed a literature search. Each member of the team found a paper or article whose topic was similar to our proposed project. The papers were gathered by the team members individually in order to obtain a wide spread of authors and paper types. Each paper was then read and summarized for the group at large so that we could gain a better understanding of the challenge we were tackling. The team member responsible for summarizing each paper took notes in a document stored in the team's shared Google Drive so that all members had access to the information even after the meeting where the papers were discussed.

Following this broad literature search the team assigned two different machine learning (ML) algorithms per team member in order to identify the best approach for our stock market prediction problem. Each team member summarized their

assigned algorithm in a few sentences, outlining the knowledge of how it worked and what kind of tasks it has been utilized for in the past. Based on these findings, the team selected five model approaches that were deemed suitable for further investigation: ANN, LSTM, Random Forest, Regression, and a combined approach that incorporated sentiment analysis.

SpiderNet decided that for these preliminary modes, we should start off attempting to train our model to predict only one stock. After some team discussion over what stock would have ample data and be useful to our stakeholders we chose Apple. Apple is a household name and is well-established with over 40 years of documented stock data. Its longevity and prominence ensure that there will be continued interest in the company's performance, both from investors and the general public. From our literature search we had identified a few different sources to collect data from, and met as a team to discuss what option was preferred by the team for our first attempt. We came to the conclusion that using datasets from Kaggle would fit the team's workflow best, partially because they had a large amount of data related to our problem but also because unlike our other option Yahoo Finance, Kaggle datasets are not behind any kind of paywall.

This first draft models of the five algorithms that the team chose to investigate will be used to evaluate feasibility of future steps including the inclusion of other stocks, visualization of our model's predictions, and accuracy evaluation.

IV. System Design

SPIDERNET will use Python to code their AI Models due to the accessibility of multiple libraries in Python that are specifically designed for creating, training, and evaluating AI, including TensorFlow Keras and Scikit-Learn. Python also offers graph-plotting libraries such as plotly and matplotlib, as well as data manipulation libraries such as numpy and pandas that will be invaluable when working with stock market data. The coding interface that the team is using is Google Colab. Google Colab offers easy collaboration across users and through some of our team members past experience, we know that it has been utilised for AI models for similar projects. The team is also utilising HTML and CSS for the project website, because they allow for high customization. Lastly, the architecture we will be using for the model will be determined after the team's investigation into the algorithms that we found the most suitable for our problem. This means that going forward we will either use ANN, LSTM, Regression, Random Forest, or a combined approach. It should be noted that all of these algorithms are suitable for either financial or time series problems, which is the categorization of our stock prediction problem. We also hope to incorporate evaluation metrics, like $R\hat{2}$, RMSE (Root Mean Square Error), and MAPE (Mean Absolute Percentage Error), to measure the accuracy of the AI.

Appendix A
Tentative Timeline

- Select Single Stock for Model Implementation and gather a dataset.
- Set up model architecture and share a colab notebook with the team for each identified algorithm candidate.
- Train and test model for single stock.
- Team decides what model they found the most easy to use.
- Visualize model predictions.
- Integrate visualization into project website.
- Assign and work on adding to the project proposal.
- Project Checkpoint I:
  - Deadline: October 27th (Sunday) at 11:59 PM.
  - Extend the project proposal to 5 pages, including a progress report for each team member.
  - Post it on the project webpage.
- Each team member adds an additional stock model.
- Integrate models into website.
- Assign and work on adding to the project proposal.
- Project Checkpoint II:
  - Deadline: November 24th (Sunday) at 11:59 PM.
  - Extend the checkpoint I report to 7 pages, including a progress report for each team member.
  - Post it on the project webpage.
- Finalize project state.
- Assign and work on final report portions.
- Assign and work on project presentation slides.
- Practice Presentation as Group.
- Final Project Report and Presentation:
  - Deadline: December 8th (Sunday) at 11:59 PM.
  - Final report (IEEE two-column style format, Times New Roman, Font 10): 9 pages (excluding references and appendix) for a 4-member team, 10 pages for a 5-member team (named: Report_GroupNumber.pdf).
  - YouTube video link.
  - Project presentation slides (named: Presentation_GroupNumber.ppt).
  - One-page project summary (named: Summary_GroupNumber.ppt).
  - Project source codes/data/readme file (named: Code_GroupNumber.zip).
  - Submit a peer evaluation form to assess team members.

Appendix B
Task Assignments

Savannah Stephenson	Research Nieve Bayes and Random Forests	Draft Model Random Forest
Janavi Bahalala	Research Artificial Neural Networks and Fuzzy Algorithms	Draft Model ANN
Christian Duff	Research Support Vector Machine and Deep Neural Networks	Draft Model Combined Approach
Allision Adams	Research Genetic Algorithm and Regression Algorithm	Draft Model Regression
Alicia Mares	Research LSTM and Hybrid Approache	Draft Model LSTM

REFERENCES

[1] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning," *International Journal of Computer Applications*, vol. 183, no. 7, pp. 1–8, Jun. 2021, doi: https://doi.org/10.5120/ijca2021921347.

[2] A. Moghar and M. Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network," *Procedia Computer Science*, vol. 170, pp. 1168–1173, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.049.

[3] N. Rouf *et al.*, "Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions," *Electronics*, vol. 10, no. 21, p. 2717, Nov. 2021, doi: https://doi.org/10.3390/electronics10212717.

[4] Md. M. Akhtar, A. S. Zamani, S. Khan, A. S. A. Shatat, S. Dilshad, and F. Samdani, "Stock market prediction based on statistical data using machine learning algorithms," *Journal of King Saud University - Science*, vol. 34, no. 4, p. 101940, Mar. 2022, doi: https://doi.org/10.1016/j.jksus.2022.101940.

[5] P. Raval, "Stock Price Prediction using Machine Learning with Source Code," *ProjectPro*, Apr. 09, 2024. https://www.projectpro.io/article/stock-price-prediction-using-machine-learning-project/571