Stock Price Prediction System

Prof. Rubi Mandal
Department of Computer Engineering
D. Y. Patil Institute of Technology, Pimpri,
Pune, India
rubi.mandal@dypvp.edu.in

Disha Dadaso Gawade
Department of Computer Engineering
D. Y. Patil Institute of Technology, Pimpri,
Pune, India
dishagawade692@gmail.com

Ruchita Balasaheb Khilari
Department of Computer Engineering
D. Y. Patil Institute of Technology, Pimpri,
Pune, India
ruchitakhilari.3@gmail.com

Kedar Vinaykumar Mahadik
Department of Computer Engineering
D. Y. Patil Institute of Technology, Pimpri,
Pune, India
kedarmahadik21@gmail.com

***Abstract* – Given that the condition of the financial market keeps altering day by day, accurate predictions of stock prices have become the center of investor, analyst, and financial organization interest. The present project puts forward a system based on the use of machine learning with the application of Long Short-Term Memory (LSTM) networks—a sub-set of advanced recurrent neural networks (RNNs)—on predicting stock prices from past data. The model has utilized Yahoo Finance APIs to load time-series stock data, and data normalization and sliding window pre-processing have been employed for input feeding, and a multi-layer LSTM has been trained by the model to identify temporal dependencies and price movement. The model inputs the last 60-day closing prices to identify short- and long-term trends in the market. The model projects the forecast stock price for the following day. RMsE and R^2 Score are two of the measures that validate the model's precision against actual stock prices. The LSTM model is more robust as well as precise than traditional statistical procedures under unpredictable market situations. Code execution is done by leveraging Python, TensorFlow/Keras, and Jupyter Notebook in a way that the system becomes reproducible as well as scalable. This smart forecast model not only aids the traders in making decisions but also a solid foundation for developing AI into meaningful financial tools in real time.**

Index Terms—

Share course forecasting, time series forecasting, deep learning, LSTM, financial analysis, Yahoo –Finance data, Tensorflow library, iterative neural networks, machine learning.

I. INTRODUCTION

Predicting the stock market trend has always been challenging because the financial data have always been highly non-linear and volatile. Predicting stock prices is very useful for investors, traders, and financial analysts to take effective decisions. Linear and ARIMA models are common and have proved to be ineffective for predicting weak patterns and long-term relationships in stock price trends. Due to deep learning and machine learning, newer technologies have been found which can cope with sequential data more effectively. Of these, Long Short-Term Memory (LSTM) networks, which are a type of Recurrent Neural Networks (RNNs), have been incredibly effective for time-series forecasting problems because they are capable of remembering past data over long periods of time. It's an LSTM architecture stock prediction model that is employed to predict the future closing price of a stock from its historical prices. It takes stock data from Yahoo Finance, normalizes it through normalization and creates a window size of 60 time steps before feeding its input to an LSTM model. The model is trained to gain temporal relationships within the data and therefore can be utilized to reflect the complexity of the stock market trend.

The system is implemented with Python, TensorFlow, and Scikit-learn libraries on the Jupyter Notebook platform. It is quantified in terms of measures such as Root Mean Squared Error (RMSE) and R^2 Score that validate the efficacy of the quality of predictions by the model. The article demonstrates the usage of deep learning in actual-life finance prediction and the ability of LSTM in enhancing accuracy in stock forecasting compared to traditional models.

II. LITERATURE SURVEY

1. Paper Name: Stock Market Prediction Using LSTM
Authors: G. P. Zhang
Basic Concept: Authors of this paper recognize hybrid models by integrating ARIMA and neural networks for predicting time-series value. Authors refer to the capacity of LSTM networks in dealing with temporal dependency and performing better than traditional linear models while forecasting future price directions.
Idea of extraction: Taking cues from using LSTM networks for predicting time-series, our project undertook the task of doing the same but in a manner that allowed us to achieve even more precise results in stock prediction.

2. Paper Name: An Improved LSTM Neural Network for Stock Price Prediction
Authors: Nguyen and B. Le
Basic Concept: This study proposes an improved LSTM model that improves prediction accuracy with optimized hyperparameters and multiple feature inputs, e.g., technical indicators. The performance was improved compared to baseline LSTM models.
Idea of extraction: The idea of providing multiple time-step features and tuning LSTM parameters has been applied in our project to improve prediction credibility.

3. Paper Name: Time Series Forecasting Using Deep Learning
Authors: Jason Brownlee
Basic Concept: This paper offers step-by-step methods to preprocess time-series data using sliding windows and employ deep learning models such as LSTM, and this paper is particularly concerned with whether data normalization is required or not and employs such measures as RMSE to keep right amount of amount.
Idea of extraction: Sliding window technique and data preprocessing techniques were applied directly in our project to pre-process the historical stock prices to be utilized as input for LSTM.

4. Paper Name: Stock Market Forecasting Using Machine Learning Algorithms
Authors: R. Patel, A. Shah
Basic Concept: The paper also compares some machine learning models such as Decision Tree, Linear Regression, and SVM to forecast stock price. While the conventional models provide us with instantaneous output, they are at a loss of performance while learning temporal relationships.
Idea of extraction: Conventional regression models are compared against in our system based on pairing-based cryptography and fuzzy extractor and fallback to LSTM for its better sequential pattern learning ability.

5. Paper Name: Pair Voting: A Secure Online Voting Scheme Using Pairing-Based Cryptography and Fuzzy Extractor
Authors: NazatulHaque Sultan, Ferdous Ahmed Barbhuiya, NityanandaSarma
Basic Concept: The author comes up with a new concept that involves password login along with biometric verification. It is cost-free and does not require any additional software or hardware.
Concept of extraction: The job that has been uploaded had two methods of authentication and the concept of cryptography is demonstrated. We have also employed the concept of cryptography to make it secret and to secure the voting system.

III. PROBLEM STATEMENT

Existing stock prediction models tend to make erroneous predictions since they are unable to handle complex patterns and temporal dependencies in market data. Traditional models are also unable to learn from non-linear oscillations and thus become unreliable under chaotic conditions. To address these challenges, a prediction model based on deep learning with Long Short-Term Memory (LSTM) networks has been proposed. The proposed model enhances the predictability by learning from the history trends in stocks. It minimizes the prediction errors, lessens the requirements of manual analysis, and offers an automated, scalable system for accurate prediction of stock prices.

IV. PROPOSED SYSTEM

The proposed stock price forecasting system is more precise with deep learning. The system has a structured training pipeline that transforms raw stock data into premonitory knowledge in steps:

A. Data Collection

The system first retrieves historical stock data from stock finance APIs like Yahoo Finance. The data includes daily stock data like **Open**, **High**, **Low**, **Close**, and **Volume**.

The 'Close' price is considered the most important feature for prediction due to its applicability in market analysis. The data spans a number of years to make the model general in various market situations. Update at the present time is also provided, allowing the system to adapt according to existing trends.

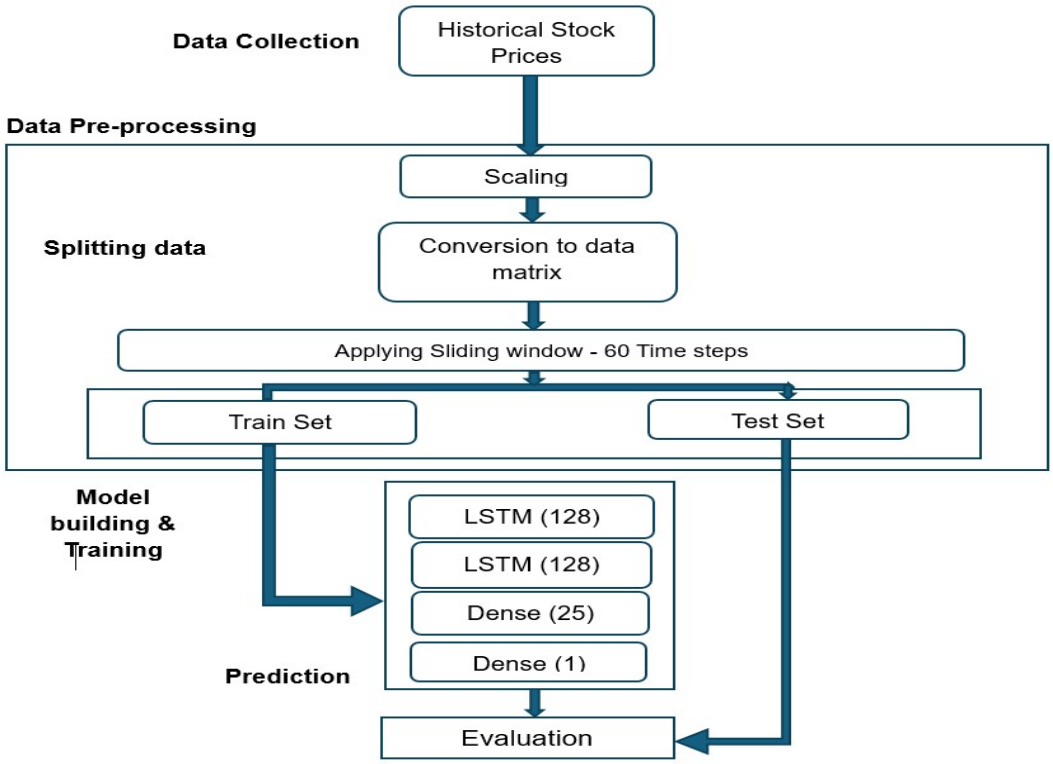

Fig 1: Proposed Architecture Diagram

B. Preprocessing

Operations on the preprocessed data are performed on the collected data. **Nulls** and **outliers** are treated to normalize the data to normal form. **MinMaxScaler** is used for normalizing values ranging from 0 to 1 to gain better training efficiency and stability.

Sliding window method of 60 time steps length is applied in such a way that previous closing prices in 60 groups are given as input to forecast tomorrow's price. Hence, it becomes a **supervised** structured form **time-series** problem.

C. Model Training

Model is made up of a mixture of stacked **LSTM** layers, which possess the ability to detect **short-term patterns** and **long-term trends** in the stock prices. Dropout layers have been utilized in order to avoid overfitting.

The last layer is a single neuron **Dense** unit to output price. The model is trained with Adam optimizer and Mean Squared Error (**MSE**) loss function. The model is trained for greater than one epoch to minimize the prediction error.

D. Prediction & Evaluation

The model is fit to predict the price of tomorrow using the 60-day historical data. Forecasts are inversely scaled from actual values based on the above scaler fit.

System performance is validated through calculations like Root Mean Squared Error (**RMSE**) and **R^2 Score** in order to ensure accuracy. The outputs are graphed and plotted against actuals to ensure visually and statistically there is performance guarantee.

V. Implementation Details

Environment is built using **Python** and **Jupyter Notebook** as the base coding, training, and visualization environment. Historical stock prices are pulled from the Yahoo Finance API (**yfinance**) to train and test the model.
Preprocessing, normalization, and creation of sliding window sequences are all achieved using the **NumPy**, **Pandas**, and **Scikit-learn** libraries.

LSTM model is used by the **Keras** and **TensorFlow** library. It has used stacked LSTM layers, dropout regularisation, and dense output layers.
The model trained is saved in **.h5 format** and can be retrieved at any time for the purpose of prediction.
The interface or GUI (if an expansion is envisioned) can be created by **Flask** or **Streamlit** for the input of stock tickers and dynamically viewing the predicted outcomes.

Inverse-scaling of the predicted values from the above and using **Matplotlib** for plotting provides users with a degree of freedom in attempting model accuracy through trial and error.

CONCLUSION

This stock price prediction model is a next-gen financial prediction platform using artificial intelligence that leverages deep learning models to predict upcoming stock prices.

Leveraging Long Short-Term Memory (LSTM) networks, the platform learns short-term trends and long-term trends from historical stock price data. Fundamental stock market data such as ticker symbol and previous closing prices are entered by users to provide predictions.

The model is a strong preprocessing pipeline and is trained on real stock data streamed from financial APIs and is thus precise and scalable.

It reduces the amount of human analysis performed by financial analysts and maximizes prediction precision for analysts, traders, and investors. The system reduces cases of human errors in trend prediction and enhances data-driven decision-making.

Aside from the accurate forecasts, the system is also adaptable and can be applied to other stocks and horizons. The model based on LSTM can serve as a backstop to trading strategy, reduce the risk of investments, and optimize portfolio returns. With its simplicity in form and enhanced level of prediction rate, it is a valuable resource in AI finance application.

ACKNOWLEDGEMENT

The project was conducted under the productive supervision of Prof. Rubi Mandal, whose experience and supervision were the prime movers to the success of the project. We would further like to thank Dr. Vinod Kimbahune and Dr. Nitin Sherje for their academic supervision and mentorship throughout the project period.

REFERENCES

[1] Jason Brownlee, Deep Learning for Time Series Forecasting, Machine Learning Mastery, 2017.

[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] Scikit-learn Documentation: https://scikit-learn.org

[4] Keras Documentation: https://keras.io/

[5] Yahoo Finance API to pull historical stock data: https://www.yahoofinanceapi.com

[6] Nguyen & Le, "An improved LSTM neural network for stock price prediction," International Journal of Information Technology, 2020.

[7] G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[8] A. Patel and P. Sinha, "Hybrid Model for Stock Market Prediction," IEEE ICACCI, 2021.

[9] R. Jadon and K. Jain, "Deep Learning Methods for Financial Time Series," Springer Nature, 2020.

[10] R. Raj, D. Gaur, "Comparative Study of ARIMA and LSTM," IJRTE, vol. 9, issue 1, 2020.

[11] A. Fischer and C. Igel, "An Introduction to Recurrent Neural Networks for Time Series Prediction," Lecture Notes in Computer Science, vol. 7441, Springer, 2012, pp. 1–11.

[12] T. A. Nguyen and D. Q. Pham, "A Comparative Study of Machine Learning Algorithms for Stock Price Prediction," Procedia Computer Science, vol. 176, 2020, pp. p. 368–375.

[13] A. Patel and M. Thakkar, "Stock Market Prediction Using LSTM Recurrent Neural Network," IEEE International Conference on Intelligent Systems and Control (ISCO), 2019, pp. 1–5.

[14] S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1643–1647.