iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73543

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Machine Learning Approaches for Stock Price Prediction of Amazon Company

Thatharaju Girija
Post Graduate Student, M. Tech (Embedded Systems), Jntuh University College Of Engineering, Science & Technology Hyderabad

Abstract: Forecasting stock price movements, especially for prominent companies like Amazon, remains a significant challenge and a key focus in both academic research and financial analysis. The stock market is known for its high volatility and unpredictability, making accurate prediction complex.

This study explores the application of machine learning techniques to enhance the reliability of stock price forecasting. Specifically, it focuses on deep learning models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), which are designed to capture sequential patterns in historical data. For comparative analysis, a traditional Linear Regression model is also implemented.

The experiments were conducted using real historical stock data from Amazon, employing two different feature sets: one containing fundamental price features (High, Low, Close, Volume), and another enhanced with technical indicators including the Relative Strength Index (RSI), Exponential Moving Average (EMA), and Moving Average Convergence Divergence (MACD). The effectiveness of each model was evaluated using standard performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

I. INTRODUCTION

Predicting stock prices has always been a major area of interest for investors, analysts, and researchers due to the highly volatile nature of the stock market. Traditional methods often fall short in capturing the non-linear patterns and sudden fluctuations in price movements. This project aims to address that challenge by applying deep learning models—Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)—which are designed to handle sequential data effectively. By analysing historical stock data of Amazon, the models are trained to learn past trends and make more accurate predictions of future High and Low prices. Technical indicators such as RSI, EMA, and MACD are also included to enrich the dataset and improve forecasting performance. The goal is to develop a robust system that not only predicts prices but also follows market trends, aiding in smarter investment strategies.

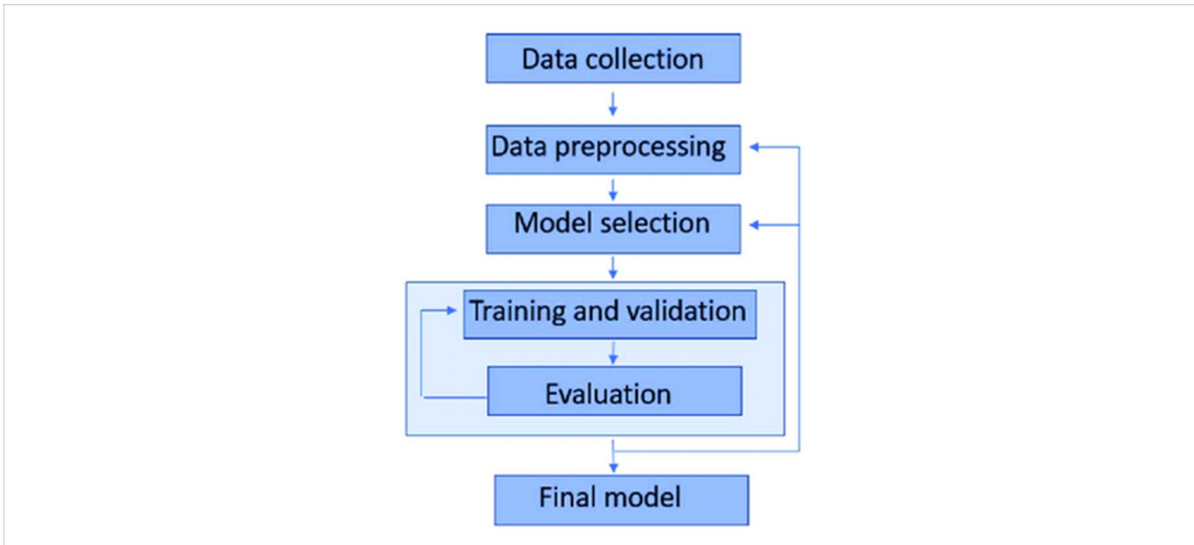
II. LITERATURE SURVEY

Several researchers have explored different machine learning approaches for stock price prediction. Rana and Kulkarni (2023) showed that advanced ensemble models like XG Boost and Light GBM outperform traditional methods in complex datasets. Mishra and Jain (2022) found polynomial regression to be more effective than linear models in capturing market volatility. Joshi and Naik (2021) used Random Forest for trend prediction, benefiting from its ability to handle noisy stock data. Patel and Shah (2020) combined ARIMA with Linear Regression to forecast short-term trends. Sarode et al. (2019) emphasized that model performance depends on dataset characteristics. Thomas and Raj (2018) demonstrated that simple models like Naive Bayes and KNN can still predict market direction when data is well-prepared. Lastly, Kumar and Reddy (2018) highlighted the strength of SVM in modelling nonlinear stock patterns.

III. OBJECTIVE

The main goal of this project is to develop a reliable system for predicting stock prices using historical data. It begins with gathering and preparing Amazon stock market data from Yahoo Finance to ensure it is clean and suitable for analysis. The project then involves building and training three different models—Linear Regression, LSTM, and GRU—to predict the following day's High and Low prices. These models are assessed using standard evaluation metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Finally, the project aims to determine how well each model can track real market trends, helping identify the most effective approach for stock price forecasting.

IV. METHODOLOGY



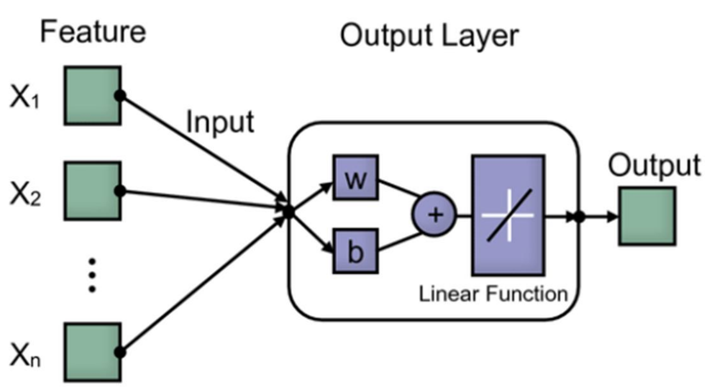
The flowchart outlines the step-by-step process followed in this stock price prediction project. It begins with loading the dataset, which involves retrieving historical stock data, typically from a source like Yahoo Finance. After loading the data, the next step is data preprocessing, where missing values are handled and features are scaled to ensure consistency and improve model performance.

Once the data is cleaned and converted into sequences, feature engineering is performed. This includes adding technical indicators such as RSI, MACD, and EMA, which help the models capture more meaningful patterns in the stock's behaviour. Following this, the process moves to model selection, where different algorithms—like Linear Regression, LSTM, and GRU—are chosen for the prediction task based on the nature of the problem.

Next, these models go through training, where they learn from the prepared data to understand the relationship between past trends and future stock prices. After training, the models are subjected to evaluation, using key performance metrics such as MAE, MSE, RMSE, and MAPE. These metrics help assess how accurate and reliable the predictions are. The final step concludes the process, indicating that the model development cycle is complete.

V. ALGORITHMS

A. Linear Regression

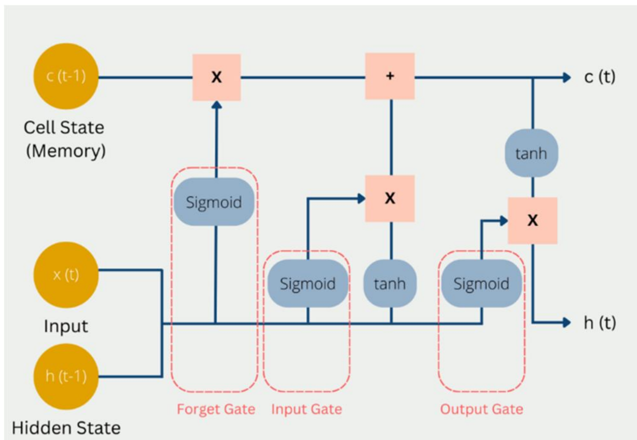


Linear Regression is a supervised machine learning algorithm used to model the relationship between one or more independent variables (inputs) and a dependent variable (output).

It assumes this relationship is linear, meaning it can be represented with a straight line.

- $\hat{y} = w_1 \cdot X_1 + w_2 \cdot X_2 + \ldots + w_n \cdot X_n + b$
- $\hat{y} = W \cdot X + b$
- W=Weights
- B=bias
- X=inputs
- Y=outputs (high price and low price)
- Input: Feature vector $X = [x1, x2, \ldots, xn]$ (e.g., High, Low, Volume, etc.)
- Prediction: $\hat{y} = w1x1 + w2x2 + \cdots + wnxn + b$
- Error: Error=$y - \hat{y}$
- Weight Calculation
- $W = (X^T . X)^{-1} X^T y$
- Minimize error to make ŷ close to actual y

B. LSTM

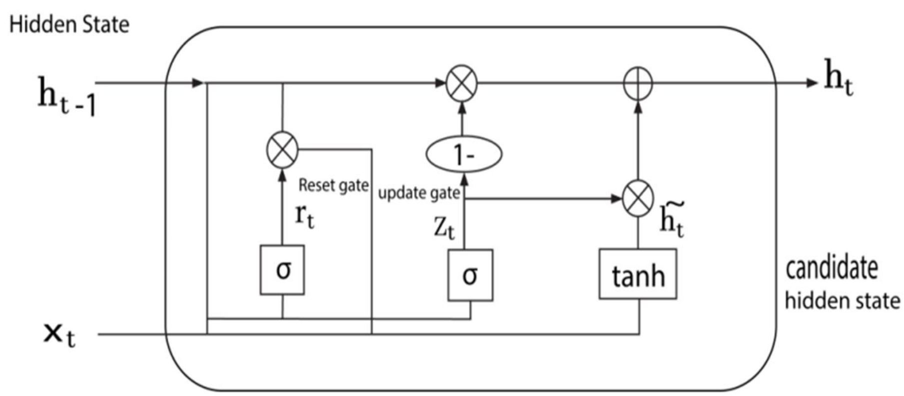


LSTM is a special type of recurrent neural network (RNN) that is widely used for processing sequential data like time series, speech, or stock prices. It is designed to remember important information over long periods and forget irrelevant parts.

- x_t = Current input (e.g., stock features of day t)
- h_{t-1} = = the output from the previous cell (Short-term output (last prediction info)
- c_{t-1} = Previous cell state (Long-term memory (important past info))
- Forget Gate
  Decides what to forget from old memory
  $f_t = \sigma (W\_f \cdot [h_{t-1}, x_t] + b\_f)$
- Input Gate
  Learns new information
  $i_t = \sigma (W\_i \cdot [h_{t-1}, x_t] + b\_i)$
  $\hat{C}_t = \tanh (W\_C \cdot [h_{t-1}, x_t] + b\_C)$
- Cell State Update
  Update memory
  $c_t = f_t \times c_{t-1} + i_t \times \hat{C}_t$
- Output Gate
  Decides what to output
  $o_t = \sigma (W\_o \cdot [h_{t-1}, x_t] + b\_o)$
  $h_t = o_t \times \tanh(c_t)$
- h_t → Output sent to next LSTM cell or Dense layer
- c_t → Updated internal memory passed to next step

- Total 32 × 60 cells run per batch (because 32 samples in a batch).
- Building the LSTM Model Architecture:
- First LSTM layer: 64 units, returns sequences for deeper learning
- Second LSTM layer: 32 units
- Dropout layers: To prevent overfitting
- To predict future prices, the model needs historical data. so created sequences of 60 and 90 days consecutive days to serve as input for the model.
- To predict future prices, the model needs historical data. so created sequences of 60 and 90 days consecutive days to serve as input for the model.
- Each sequence (X) represents 60 days and 90 days of High, Low, closing price, opening price, Volume data.
- The target (y) is the High and Low prices on the 61st day

C. GRU



GRU is a type of recurrent neural network (RNN) architecture specifically designed to handle sequential data like stock prices, speech signals, or time-series forecasting. It is a simplified version of LSTM that achieves similar performance with fewer parameters and faster training.

- Update Gate
- Controls how much past info to carry forward
  $z_t = \sigma (W\_z \cdot [x_t, h_{t-1}] + b\_z)$
- Reset Gate
- Decides how much past info to forget
  $r_t = \sigma (W\_r \cdot [x_t, h_{t-1}] + b\_r)$
- Candidate State
- New memory proposal using reset info
  $\tilde{h}_t = \tanh (W\_h \cdot [x_t, r_t * h_{t-1}] + b\_h)$
- Final Hidden State (h_t)
- Purpose: Combines old and new information
  Formula:

Building the GRU Model Architecture:

- First GRU layer: 64 units, returns sequences for deeper learning
- Second GRU layer: 32 units, without returning sequences (end of GRU layers)
- Dropout layers: To prevent overfitting
- Output layer: A dense layer with 2 units (predicts High and Low prices for the next day)

VI. RESULTS

Model	MSE (High)	MSE (Low)	MAE (High)	MAE (Low)	RMSE (High)	RMSE (Low)	MAPE (High)	MAPE (Low)	Trend Match
Linear Regression 5 days)	6.19	6.97	1.70	1.82	2.49	2.64	1.17%	1.28%	3/5 matched
Linear Regression (14 days)	6.24	7.10	1.71	1.84	2.50	2.67	1.18%	1.29%	7/14
LSTM for 14 days with technical indicators	36.40	41.26	4.74	4.83	6.03	6.42	3.33%	3.38%	9/14
LSTM for 14 days	48.30	50.05	5.11	5.19	6.95	7.07	3.37%	3.57%	9/14
Gru for 14 days	23.02	24.10	3.61	3.66	4.79	4.91	2.42%	2.58%	9/14
GRU with technical indicators	16.58	27.67	3.67	4.97	4.07	5.26	1.68%	2.27%	9/14

VII. CONCLUSION

The evaluation showed that Linear Regression gave the most accurate results in terms of error values (MSE, MAE, RMSE, MAPE) over a short 5-day prediction period. However, it was not very good at capturing the actual upward or downward movement of stock prices. Among the deep learning models, GRU performed slightly better than LSTM in terms of prediction accuracy. Linear Regression also had very low percentage errors, with MAPE close to 1.2%. On the other hand, LSTM and GRU models were better at following the correct trend of the stock, accurately predicting the direction in 9 out of 14 cases, while Linear Regression did so in only 7 out of 14. This shows that while simple models like Linear Regression are more precise, deep learning models are more effective in understanding market behaviour and trends.

REFERENCES

[1] B Shivani, Govinda Rao.Analysis of a Stock Exchange and Future Prediction Using LSTM 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST) | 978-1-6654-6633-2/22/ ©2022 IEEE

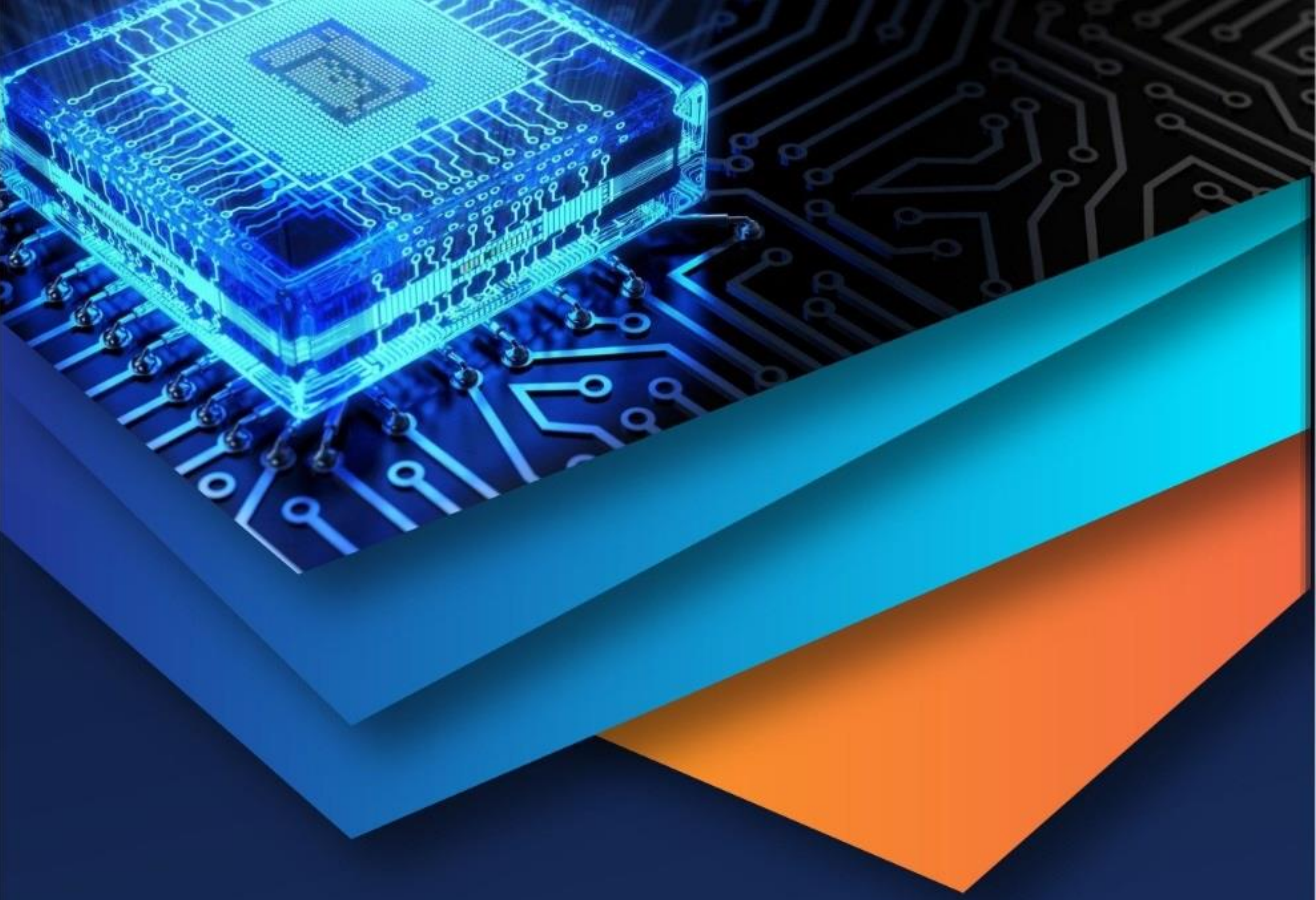
[2] Sumeet Sarode, Harsha G. Tolani, Prateek Kak, Lifna C S STOCK PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES978-1-5386-7799-5/19/ ©2019 IEEECarson Kai-Sang Leung Richard Kyle MacKinnon Yang Wang A Machine Learning Approach for Stock Price Prediction

[3] Deepika, N. & Nirupama Bhat. M. (2024). Optimized Deep Learning Model Architecture for the Feature Extraction to Predict Trend in Stock Market. International Journal of Maritime Engineering. 1. 61-74. 10.5750/ijme.v 1i1.1338.

[4] Srivastava, Shilpa & Pant, Millie & Gupta, Varuna. (2023). Analysis and prediction of Indian stock market: a machine-learning approach. International Journal of System Assurance Engineering and Management. 14. 10.1007/s13198-023-01934-z.

[5] S.M. Ajitha1, S.Karthick2, Mu.Karthikeyan3, N.T.Naveen Balaji4 AMAZON STOCK PRICE PREDICTION BY USING SMLT(2023).

[6] Vedant, N. (2024). Stock Price Prediction Research Machine Learning Model Evaluation. Open Journal of Business and Management, 12, 1251-1268.



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY