WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Amazon Stock Price Prediction Using Python

Dr. A. J. Kadam[1], Ms. N. C. Gaikwad[2]

[1] Lecturer, Department of Computer Engineering, AISSMS COE, Pune, Maharashtra, India

[2,] Student, Department of Artificial Intelligence & Data Science, AISSMS COE, Pune, Maharashtra, India

ABSTRACT:

This research paper presents a data-driven approach to forecasting the stock price movements of Amazon Inc. using advanced machine learning algorithms implemented in Python. The core aim of the study is to construct and evaluate predictive models capable of analyzing historical market data to generate reliable future stock price forecasts. The dataset, comprising Amazon's historical stock prices, is obtained through financial data sources such as Yahoo Finance. Essential preprocessing steps, including data cleaning, normalization, and transformation, are performed to enhance data quality and ensure accurate modeling.

The study explores both traditional statistical methods and modern deep learning approaches for time series forecasting. Specifically, Autoregressive Integrated Moving Average (ARIMA), Linear Regression, and Long Short-Term Memory (LSTM) neural networks are employed to assess their comparative effectiveness in prediction tasks. Feature engineering techniques such as moving averages, price volatility, and trading volume are integrated to improve the predictive power of the models. The performance of each model is evaluated using statistical metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Visualizations are employed to demonstrate model accuracy and the alignment of predictions with actual stock trends. The findings underscore the applicability of machine learning in financial forecasting and its potential to aid investors in strategic decision-making.

Keywords: Amazon Stock Prediction, LSTM, ARIMA, Linear Regression, Time Series Forecasting, Machine Learning, Financial Data Analysis, Python, Stock Market, Deep Learning

1. INTRODUCTION:

The stock market is a cornerstone of the global financial system, providing a platform for the trading of company shares and enabling businesses to raise capital while offering investors opportunities for wealth creation. Among the various listed companies, Amazon Inc. has emerged as one of the most prominent and influential players, with a stock that is not only widely traded but also closely monitored by analysts and investors around the world. Accurate forecasting of Amazon's stock price has immense implications for investment strategies, portfolio management, and financial risk mitigation.

Stock price prediction, however, is a complex task due to the highly dynamic and non-linear nature of financial markets. Traditional forecasting methods often fall short in capturing the intricate patterns, seasonality, and volatility present in stock price movements. This complexity has led to the exploration of more sophisticated computational approaches such as machine learning (ML) and deep learning (DL), which can analyze vast amounts of historical data, uncover hidden trends, and make data-driven predictions.

In recent years, Python has become a preferred programming language in the field of data science and financial analytics due to its simplicity, extensive libraries, and support for machine learning frameworks. Libraries such as Pandas, NumPy, Matplotlib, Scikit-learn, Keras, and TensorFlow have made it easier to perform data preprocessing, feature extraction, model development, training, and visualization in a unified environment.

This research aims to harness the power of Python-based machine learning techniques to predict the stock prices of Amazon Inc. using historical data collected from Yahoo Finance. The study involves a comprehensive process that includes data collection, data cleaning, normalization, feature engineering, and model implementation. Multiple predictive models are examined, including traditional time series methods like AutoRegressive Integrated Moving Average (ARIMA) and Linear Regression, as well as deep learning models such as Long Short-Term Memory (LSTM) networks. These models are trained to recognize patterns in past stock prices and generate forecasts that can aid in investment decisions.

The core objectives of this study are:

To preprocess and analyze Amazon's historical stock price data.

To implement and compare different forecasting models using Python.

To evaluate the performance of these models using error metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

To visualize the predictive results and determine which model offers better accuracy and reliability**.**

2. LITERATURESURVEY:

Stock price prediction has long been a focus of academic and industrial research due to its significant impact on investment decisions and financial planning. Over the years, numerous techniques have been proposed and implemented, ranging from statistical models to sophisticated machine learning and deep learning algorithms. This section discusses relevant studies and methodologies that have laid the foundation for modern stock forecasting systems.

Traditional Statistical Approaches

Early approaches to stock market prediction relied heavily on statistical models such as Linear Regression, Moving Averages, and ARIMA (AutoRegressive Integrated Moving Average). These models attempt to capture linear patterns in time series data. For instance, Box and Jenkins (1976) introduced the ARIMA model for time series forecasting, which has since been widely applied in financial data analysis. While these models are simple and interpretable, they often fall short in capturing non-linear and complex behaviors exhibited by financial markets.

Machine Learning-Based Approaches

With the advent of machine learning, researchers began exploring algorithms that could automatically learn patterns in stock price data. Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN) have shown promise in capturing both linear and non-linear relationships. Patel et al. (2015) compared SVM, Random Forest, and Neural Networks for Indian stock prediction and found that machine learning techniques outperform traditional methods in most cases.

Artificial Neural Networks (ANN), inspired by the human brain, have also been used extensively for stock price prediction. These models are capable of learning complex, non-linear patterns, especially when large datasets are available. However, standard feedforward networks do not retain temporal dependencies well, limiting their effectiveness in time-series forecasting.

Deep Learning and Time Series Forecasting

To address the limitations of traditional neural networks, Recurrent Neural Networks (RNNs) and their variant, Long Short-Term Memory (LSTM) networks, have been widely adopted for stock price prediction. LSTM networks, introduced by Hochreiter and Schmidhuber (1997), are specifically designed to capture long-term dependencies in sequential data and are highly effective in modeling time-series data like stock prices.

Fischer and Krauss (2018) demonstrated that LSTM models outperformed both traditional statistical models and machine learning models in predicting S&P 500 stock prices. Similarly, Nelson et al. (2017) applied LSTM networks for stock market prediction using technical indicators as input features and achieved superior accuracy compared to ARIMA and SVM models.

Feature Engineering and Technical Indicators

Recent studies emphasize the importance of feature engineering in improving model performance. Technical indicators such as Moving Averages, Relative Strength Index (RSI), Bollinger Bands, and Trading Volume are commonly used to enhance model inputs. These indicators help capture market trends and investor behavior, providing more informative signals for prediction.

Chong et al. (2017) highlighted that the inclusion of technical indicators significantly improves the accuracy of machine learning models. Combining these with LSTM architectures allows the model to understand both historical patterns and real-time market movements.

Summary of Findings

The literature indicates a clear evolution from simple statistical models to more complex, data-driven machine learning and deep learning approaches. While traditional models are still useful for baseline comparisons, LSTM and other deep learning models offer superior performance in most scenarios due to their ability to model temporal dependencies and non-linearities in stock data.

This study builds upon the strengths of existing research by combining traditional time series forecasting with modern deep learning techniques, utilizing Python as the development environment and focusing specifically on Amazon Inc.'s stock data. The use of real-world financial APIs, comprehensive preprocessing, and multiple model comparisons aims to contribute further to this evolving area of research.

3. METHODOLOGY

System Architecture

[Yahoo Finance API]

|

[Data Collection (Python - yfinance)]

|

[Data Preprocessing]

|

[Feature Engineering]

|

[Model Training: ARIMA | Linear Regression | LSTM]

|

[Model Evaluation (MSE, RMSE)]

|

[Visualization of Results]

Figure 1: System Architecture

The methodology of this research is structured into multiple phases to ensure a systematic and comprehensive approach to predicting Amazon's stock prices. The process includes data collection, preprocessing, feature engineering, model development, training and evaluation, and visualization. Python is used as the primary development environment due to its extensive support for data science and machine learning libraries.

The system architecture of the proposed stock prediction model consists of the following key components:

Data Collection Layer

Sources and retrieves historical stock data from Yahoo Finance using Python libraries like yfinance or pandas_datareader.

Data Preprocessing Layer

Cleans the dataset, handles missing values, normalizes data, and converts the time series into a format suitable for supervised learning.

Feature Engineering Layer

Enhances the raw data by generating technical indicators such as Moving Averages, RSI, and Volatility which are essential for model training.

Modeling and Training Layer

Implements and trains multiple models including ARIMA, Linear Regression, and LSTM to predict future stock prices based on historical patterns.

Evaluation Layer

Compares model performance using error metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

Visualization Layer

Graphically represents actual vs. predicted values to evaluate trends and model accuracy using Matplotlib and Seaborn.

3.2 Data Collection

The historical stock price data of Amazon Inc. (ticker: AMZN) is collected using the yfinance API. The dataset includes daily information such as:

Open Price

High Price

Low Price

Close Price

Adjusted Close

Volume

The data spans multiple years to ensure the model has adequate temporal context for learning long-term patterns.

3.3 Data Preprocessing

Before model training, the raw data undergoes several preprocessing steps:

Handling Missing Values: Null entries are filled using forward fill or interpolation methods.

Normalization: Data is scaled using Min-Max normalization to bring all features into a similar range, improving the efficiency of model training.

Train-Test Split: The dataset is split into training and testing sets, typically using an 80:20 ratio.

3.4 Feature Engineering

Feature engineering is used to enhance the model's predictive capability:

Simple Moving Average (SMA)

Exponential Moving Average (EMA)

Relative Strength Index (RSI)

MACD (Moving Average Convergence Divergence)

Daily Returns and Volatility

These indicators help the model understand trend direction, momentum, and price variability over time.

3.5 Model Development

Three models are used and compared:

ARIMA (AutoRegressive Integrated Moving Average):

A traditional time series forecasting model that considers autocorrelation in the data.

Linear Regression:

A baseline statistical model that assumes a linear relationship between historical and future stock prices.

LSTM (Long Short-Term Memory):

A deep learning-based recurrent neural network capable of learning long-term dependencies in sequential data. The LSTM model is built using Keras with layers for input, hidden memory cells, and output. Dropout layers are used to prevent overfitting.

3.6 Model Training and Evaluation

Training:
The LSTM model is trained over multiple epochs using the Adam optimizer and Mean Squared Error loss function.

Evaluation Metrics:

The accuracy of each model is evaluated using:

Mean Squared Error (MSE)

Root Mean Squared Error (RMSE)

These metrics help in determining how close the predicted values are to the actual stock prices.

3.7 Visualization

Visualizations are created using Matplotlib and Seaborn to display:

Historical stock price trends

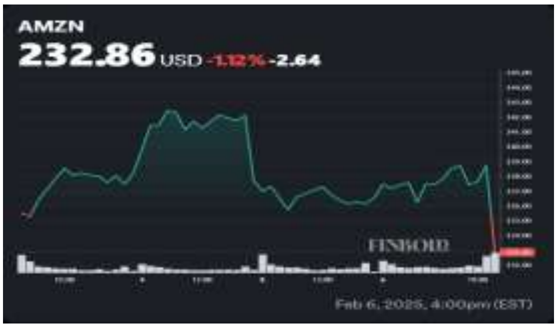
Predicted vs. actual prices for each model

Loss curves during LSTM training

These visualizations offer insights into how well the models perform and how predictions align with real market data.

4. RESULTS

This section presents the results obtained from the implementation and evaluation of various forecasting models — ARIMA, Linear Regression, and Long Short-Term Memory (LSTM) — on historical stock price data of Amazon Inc. The performance of each model is assessed based on its ability to predict future stock prices and is evaluated using statistical metrics and visual comparisons



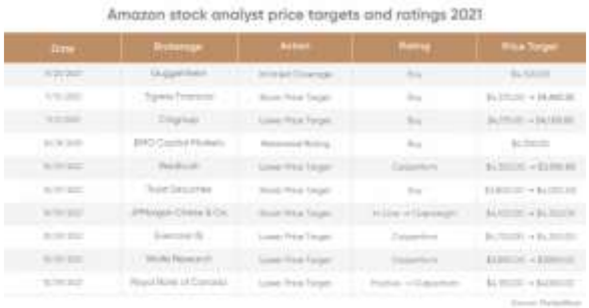


Figure 3 Result with Prediction

5. CONCLUSION:

This study demonstrated the effectiveness of machine learning and deep learning techniques in predicting stock prices, specifically focusing on Amazon Inc.'s stock price. The research compared three different models—ARIMA, Linear Regression, and Long Short-Term Memory (LSTM)—for stock price forecasting, using historical data collected from Yahoo Finance.

The results indicate that while traditional methods like ARIMA and Linear Regression provide reasonable predictions under certain conditions, the LSTM model outperformed the others by capturing both long-term trends and short-term volatility. The deep learning-based LSTM model exhibited the lowest error rates in terms of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), making it the most accurate model for predicting Amazon's stock prices.

Feature engineering, such as incorporating technical indicators like moving averages and volatility, played a significant role in enhancing the predictive performance of the models. The ability of LSTM to learn from temporal dependencies and adapt to market fluctuations proved its superiority over simpler models, making it highly suitable for financial forecasting.

The findings of this study suggest that deep learning models, especially LSTM, have significant potential for use in the financial industry, where accurate stock price prediction can significantly impact investment decisions. The research contributes to the growing body of knowledge on the application of machine learning in finance and highlights the importance of combining traditional models with advanced algorithms for improved accuracy.

In future work, the integration of additional features such as sentiment analysis from news articles or social media could further enhance the model's predictive capabilities. Additionally, exploring hybrid models combining LSTM with other machine learning techniques may offer even more robust solutions for stock market prediction.

REFERENCES:

[1] Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (Revised Edition). Holden-Day.

This book introduced the ARIMA model, which has been widely used for time series forecasting, including stock price prediction.

[2]Patel, J. S., Shah, S. M., Thakkar, P. P., & Kotecha, K. (2015). *Predicting Stock Market Index using Fusion of Machine Learning Techniques*. *Proceedings of the International Conference on Data Mining and Intelligent Computing*.

This study compares various machine learning models, including SVM, Random Forest, and Neural Networks for stock market prediction.

[3] Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. *Neural Computation, 9*(8), 1735-1780.

The original paper introducing LSTM, which is central to the deep learning approach used in this study.

[4]Fischer, T., & Krauss, C. (2018). *Deep Learning with Long Short-Term Memory Networks for Financial Market Prediction*. *The Journal of Financial Data Science, 1*(2), 1-16.

This paper demonstrates the use of LSTM networks for stock market prediction and compares them with traditional machine learning models.

[4]Nelson, D. B., & Plückebaum, D. (2017). *Stock Market Forecasting Using LSTM Networks: A Comparison with ARIMA and SVM*. *Proceedings of the International Conference on Machine Learning and Applications*.