Stock Price Prediction Using Machine Learning

Ammar Khan, Swati Pandit, Tushar Gaur, Jay Gupta, Ritik Gupta, Sanchit Sharma, Swati Pandit

Department of Computer Science & Engineering, Vivekananda College of Technology and Management Aligarh

Abstract-Price prediction of stocks is a fundamental problem in the financial domain, as it provides valuable information to investors, traders, and analysts making informed decisions. Predicting stock prices is an issue that has intrigued both investors and researchers for decades. ML algorithms have produced effective tools for predicting stock prices in the last few years. This paper reviews the major ML algorithms for predicting stock prices, including long short-term memory (LSTM) and SVM networks. It also explores applications and some of the caveats of working with stock data using ML price prediction.

The main focus is to evaluate their prediction capabilities and prevision skills and to select the most promising models for the real world. This paper provides a comprehensive review of machine learning methods used for stock price prediction. Although factors such as market volatility can complicate predictions, the study highlights how machine learning techniques can enhance stock price forecasting, ultimately assisting investors in making more informed decisions.

Keywords: Stock price prediction, Machine Learning, LSTM, SVM, Streamlit.

INTRODUCTION

Forecasting stock prices has long been a key challenge for investors, analysts, and financial institutions. In a classical sense, stock price prediction involves numerous economic indicators, stock performance, and market sentiment. But technology has advanced so rapidly has become an important and useful tool to perform stock price forecasting and prediction. By analyzing, You can give it large datasets and let it find complex patterns, so machine learning algorithms can offer insights and predictions about stock price movement. This study aims to examine the performance of SVM and LSTM models for stock prediction. data, utilizing historical stock data from Google (GOOG). It is a clean time series dataset extracted from a CSV file, and the Average Magnitude of Relative Error (AMRE), tests the finance and mean squared-Error, Root Mean Squared Error, (R^2).

Machine learning models are primarily used to predict stock prices, evaluate the accuracy of these predictions, and assess the models' practical applicability in real-world financial markets. This project will focus on data sampling, cleaning, feature selection, modelling, and performance evaluation to explore the effectiveness of machine learning in predicting stock prices. Machine learning algorithms analyse historical data to forecast financial markets, with researchers typically working with data up to 2023. stock market trends. Implemented by: Umer Ghani, M Awais, and Muhammad Muzammul displaying a stock market prediction using data from CSV file 5-year data in (GOOGL) [3]

In this support Vector Machines (Random forest-based algorithm model Junwei Chen [4] selected regression and LSTM to SVM) which is utilized in the section.

This will introduce long short-term memory (LSTM) and SVM. Researchers often use them for stock market analysis.

During analysis, in different For instance, the model features such as historical financial data, the accuracy of the model, predictions and must be considered to achieve the correct and aligned return goal of stock prediction. So, detailed research and Evaluation are crucial to carry out, so that you choose ML algorithms only after mastering stock data. There is no guarantee that the ML algorithm will give us the right prediction in each case. Specific challenges and issues will be addressed.

2.1 Moving Average (MA):
The most popular method of time-series forecasting is the Autoregressive moving average.
It relies on three components:

It relies on past values to forecast future values. MA (Moving Average) terms, which incorporate past forecast errors.
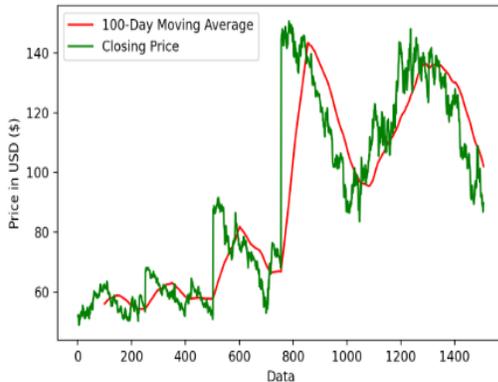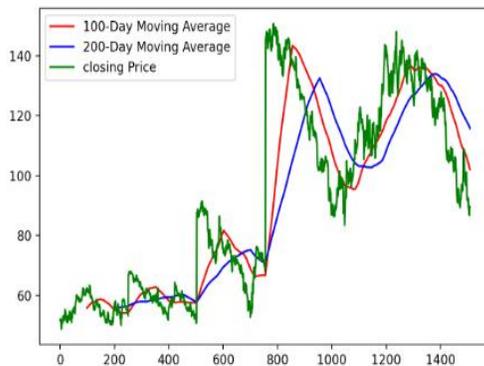

Fig1.100-day MA

Fig2.200-day MA

2.2. Long Short-Term Memory (LSTM): In financial domain, ML algorithms like Long-Short-term Memory (LSTM) are considered some of the best algorithms. LSTM is considered superior in predicting stock prices due to its capacity to process whole sequences of information (Long Short-Term Memory) is a type of Recurrent Neural Network (RNN) designed to capture and extract valuable information from long-term sequential data, making it particularly effective at identifying long-term trends in time-series data. The vanishing gradient is the main reason for the inefficiency of RNNs for longer sequences. Perhaps the most cited common algorithm is the (Long Short-Term Memory) LSTM introduced by Hochreiter & Schmid Huber (1997) which added memory cells and gating mechanisms. Memory cells

and gating mechanisms. allow the network to decide if it should remember a pattern.

LSTM Methodology Recent studies have indicated that the LSTM model is superior to traditional methods while predicting stock price:

LSTM models yielded higher prediction accuracy than logistic regression and gradient boosting models (Fischer & Krauss, 2018).

Selvin et al. demonstrated that LSTM-based models exceeded standard RNNs in financial time-series forecasting.

Some key strengths of LSTM networks concerning time series volatility prediction include:

Chances of covering time-based correlations in the financial docs.

Better handling of non-linearity than traditional statistical approaches.

More accurate forecasting of stock trends than conventional ML models.

But LSTMs also have some disadvantages:

So, it should be trained with big data in consideration, but compared to SVM less CPU intensive.

2.3. Support_Vector_Machines (SVM):
The SVM is utilized to articulate analysis for linear and nonlinear processes, such as stock price prediction and classification of stock spouts and peaks. One or more 1-dimensional points in 2-dimensional space represent a hyperplane that separates two classes for a certain classification using an SVM-dimensional hyperplane.

The optimized hyperplane is the one that has the maximum distance from both classes. There are various SVM Kernel functions that can help change data dimension and allow researchers to transform linearly inseparable data into a separate dataset. It can create the most complex and curve-fitting dataset and identify nonlinear relationships between stock prices and input variables. Several research studies show that SVM can be powerful and effective for stock predictions Jingdong Yang's 2023 analysis shows that Kernel functions in SVM yield high accuracy in price forecasting and ensure results achieve a global minimum.
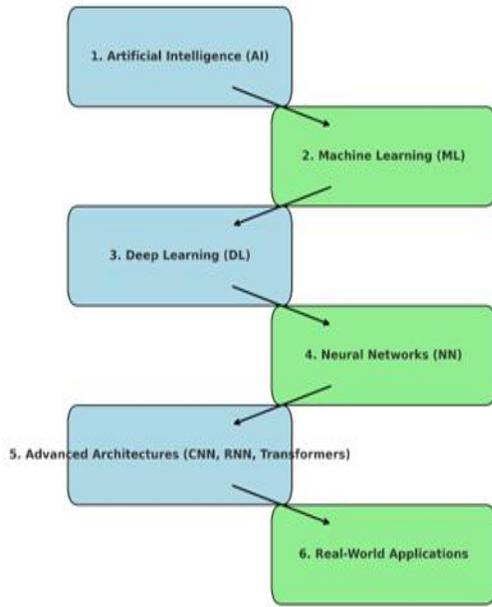
Fig3.machine Learning

LITERATURE VIEW

In Price prediction of stock, ML techniques are often used.

• Neural Network commonly used for stock price prediction. ANN consists of multiple layers of neurons that are interconnected, wherein the input data is processed to learn and provide predictions. ANNs can model non-linear relationships between stock prices and features.

Applications: Paoet al. found (ANNs) outperform traditional models in stock market predictions. prediction (2006) due to the capacity to model complex non-linear relationships.

 Deep-Learning Models

LSTM networks are specialized RNNs, making them ideally suited for time-series forecasting, as RNNs are designed to learn long-term dependencies from sequential data.

Applications: Zhang employed LSTMs for predicting stock market trends and demonstrated the superiority of LSTMs to machine learning approaches in modelling sequential patterns and dependencies in stocks. Essential Data Sources for Stock Price Prediction Many different data sources exist, and even within a single data source, many different possible features can be generated. Some of the most popular data sources where you can get your data are: Historical Stock Data

Price data and trading volume are the primary inputs for most machine-learning models. These data points form the foundation of many prediction models, such as time series analysis.

METHODOLOGY

1. Data Collection and Pre-processing

• Data Source: The primary data source will be historical stock prices from financial platforms in CSV format.

2.Pre-processing:

Remove unnecessary strings: Eliminate strings that are not needed. Missing values can b filled using interpolation or statistical methods like forward or backward filling.

Normalization: Data has to be normalized for different features across a common (e.g., e.g., Min-Max Scaling).

3. Feature selection — find the features that have the highest influence on stock price prediction This will involve:

Historical Stock Data: Previous day close price, Volume, etc

4.Model selection:

Long-Short-Term Memory (LSTM): RNN is this type will be use to manage time-series data. Stock price prediction a beneficial using LSTMS due to their capability of managing based dependencies and long-time patterns.

5. Model Training and Validation

Training: Implement the training of the model on the historical stock data (with features that may include past prices, trading volume, and sentiment data) from a subset of this dataset. The model will be trained on existing prices as features and future stock prices or labels.

6. Prediction and Evaluation

• Prediction: Use the trained model for predicting future stock prices or market trends on the unseen test data.

• Evaluation: Compare stock prices predicted against stock prices in the test data to evaluate model performance [We will assess performance according to the evaluation metrics we introduced earlier (MAE, RMSE, R^2).

7. Deployment & real-time application

• Model Deployment: After reaching an acceptable performance level, the model is ready for production on a streamlet application.
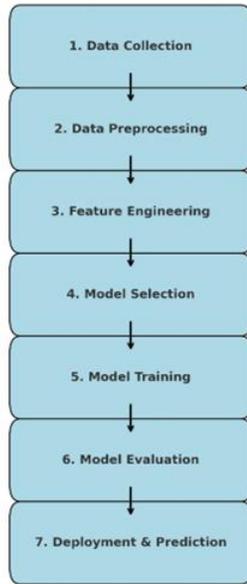

Fig4.phases of development

CONCLUSION

In summary, ML in stock price prediction has drawn the huge attention of financial and AI researchers as a result of significant benefits in different fields of applications. In addition to the above two popular ML algorithms (SVM and LSTM), various types of ML-based models have been developed and work in the stock market successfully. Despite the significant benefits, several challenges need to be addressed while using ML. The price prediction of stocks project using machine learning and Streamlit aims to provide an accessible, interactive tool for forecasting stock prices. By utilizing advanced algorithms, such as deep learning models like Long Short-Term Memory (LSTM), this project introduces innovative approaches for predicting stock market trends. Additionally, the integration of sentiment analysis enhances the predictive capabilities.

Streamlit allows for the design of a simple, intuitive web application where users can interact with the model and access predictions in real-time, making the tool user-friendly and effective for both beginners and seasoned investors. This project not only helps in stock price prediction but also illustrates the significance of integrating multiple data sources—historical stock data, in making accurate forecasts.
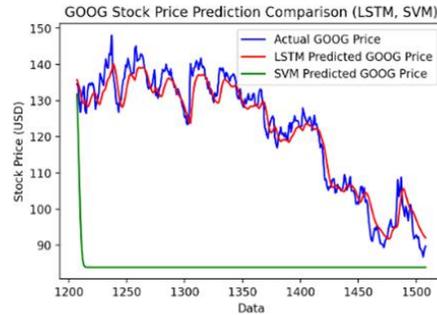
REFERENCE

[1]   Obthong,M.Tantisantiwog,   N. Jeamwatthanachai, W., and Wills, G. (2020) A survey on machine learning for stock price prediction: algorithms and technique.International Conference on Finance, Management and IT Business, 63-71.

[2]   Umer, M. Awais, M. and Muzammul, M. (2019) Stock market prediction using machine learning (ML) algorithms. Distributed Computing and Artificial Intelligence Journal, 8(4).

[3]   Chen, J. (2023) Analysis of Price Prediction Using Machine Learning.

[4]   Panwar, B., Dhuriya, G., Johri, P., Yadav, S. S., and Gaur, N.(2021)Stock market prediction using SVM. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 629-631.

[5]   Emioma, C. C. and Edeki, S. O. (2021) Stock price prediction using machine learning on least-squares basis. In Journal of Physics: Conference Series, 1734 (1): 012058.

[6]   Yang, J. (2023) Support Vector Machine-based Stock Prediction Analysis. Highlights in Business, Economics and Management, 3: 12-18.

[7]   Kumar, M. and Thenmozhi, M. (2006) Forecasting stock index movement: A comparison of support vector machines and random forest.

[8]   Di Persio, L. and Honchar, O. (2016) Artificial neural networks architectures for stock price prediction: Comparisons and applications. International Journal of circuits, systems and signal processing, 10(2016): 403-413.

[9]   Shahvaroughi Farahani, M., and Razavi Hajiagha, S. H. (2021) Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. Soft computing, 25(13): 8483-8513.

[10]   Selvin et al. (2017) demonstrated that LSTM-based models outperformed standard RNNs in financial time-series forecasting.

[11]   Fischer & Krauss (2018) found that LSTM models achieved higher prediction accuracy than logistic regression and gradient boosting models.