Netflix Stock Price Prediction Using Machine Learning

PROF&DEAN.BALAJEE MARAM

SR University, Warangal, Telangana State, India

Shravani Guguloth

Department of Computer Science & Engineering,

SR University, Warangal, Telangana State, India 2203a52235@sru.edu.in

**Abstract:** This project presents a thorough analysis of stock volume prediction using a variety of statistical and machine learning methods. The dataset, obtained from Netflix's stock market, is examined and preprocessed. To better understand the relationships between the numerical features and the target variable (Volume), the features are visualized. For volume prediction, three methods are used: multiple regression models with bootstrap resampling, such as ridge regression, lasso regression, SVR, and k-neighbors regression; bootstrap resampling with linear regression; and support vector regression (SVR) with bootstrap resampling. The mean squared error (MSE) is computed for every model through extensive bootstrap sampling, offering a reliable assessment metric. The top-performing model is highlighted by comparing and visualizing the performance of these models. Furthermore, for the chosen model, coefficients and standard errors are calculated Key Words.

**1.Introduction:**

Making wise investment decisions in the dynamic and constantly shifting world of financial markets depends on being able to accurately predict stock market trends. The quantity of shares traded in a given time frame, or stock volume, is a fundamental indicator that offers important information about how the market is acting. This project uses a combination of statistical and machine learning techniques to explore the field of stock volume prediction. The goal of this study is to create reliable and accurate prediction models for stock trading volumes by utilizing historical stock data from Netflix, a major participant in the entertainment sector. This project investigates the effectiveness of several algorithms, including support vector machines, ridge regression, lasso regression, and linear regression, through a methodical approach that includes data preprocessing, feature exploration, and rigorous model evaluation using bootstrap resampling.

The models' dependability is guaranteed by the use of bootstrap resampling techniques, which also offer a thorough understanding of their performance. This research project not only provides investors and financial analysts with insightful information about stock volume prediction, but it also demonstrates the applicability of machine learning and statistical techniques in the field of financial forecasting.

**Literature Review:**

**Stock price movement prediction using sentiment analysis has become an increasingly popular field in recent years due to the widespread use of social media platforms and the availability of large volumes of user-generated content. A number of studies have explored the application of stock price prediction on different sentiment analysis techniques such as lexicon-based approaches, machine learning-based approaches, and hybrid approaches (Derakhshan et al, 2019; Otabek Sattarov et al, 2020; Nan Jing et al, 2021; Zane Turner et al, 2021; Christina Nousi et al, 2021). Franco Valencia et al (2019) proposed utilizing machine learning tools and social media data to predict the price movement of cryptocurrencies such as Bitcoin, Ethereum, Ripple, and Litecoin. They compared the effectiveness of two different machine learning techniques: support vector machines (SVM) and random forest (RF). The study found that by utilizing machine learning and sentiment analysis, it is possible to predict the direction of price movements in the emerging cryptocurrency market. Other reseach proposes a multichannel collaborative network for predicting stock trends by incorporating social media data and candlestick charts. They begin by extracting social media sentiment features using the Natural Language Toolkit and sentiment analysis data from Twitter. These studies demonstrate the potential of sentiment analysis techniques to provide valuable insights into a variety of domains. However, it is important to consider the limitations and challenges of each sentiment analysis technique, such as the difficulty of accurately capturing context and sarcasm, and the potential biases introduced by training data (Trang Thi Ho et al, 2021; Saloni Mohan et al, 2019)**

**Machine learning approaches or forecasting models are a popular method for predicting stock price or prices of other investment financial assets. There have been numerous studies that have investigated their effectiveness by using various types of algorithms. Such as Rakhi Batra et al (2018) used the support vector machines (SVM) algorithm to predict the next day’s movement of Apple's stock price. Tushar Rao and Saket Srivastava (2012) utilized the Expert Model Mining System (EMMS), which includes Exponential Smoothing (ES), Auto Regressive Integrated Moving Average (ARIMA), and seasonal ARIMA models to explore the relationship between Twitter sentiments and stock prices. Dibakar Raj Pant et al (2018), aimed to predict the volatile price of Bitcoin by analyzing the sentiment in Twitter and identifying the relationship between them using Recurrent Neural Network (RNN).**

**When dealing with “prediction problems”, one of the crucial factors to consider is the data source. There's a saying in the field, "garbage in, garbage out," which means that when a model is fed with good quality data, the output of predictions will also be good and vice versa. Some research examines how freshly scraped economic news headlines can be utilized to predict stock value changes. The focus is on the sentiment analysis of these headlines, and various tools were used to perform the analysis. BERT was used as the baseline, and its results were compared with those of VADER, TextBlob, and a Recurrent Neural Network. The study found that economic news headlines have an impact on stock market values even without their textual context, and different sentiment analytical tools show significant differences in their effects on stock value changes, as observed from the correlation matrices. However, the stock market impact also depends on how the data in the study period were affected. Overall, the study highlights the potential of sentiment analysis of economic news headlines in predicting stock value changes and the importance of selecting the appropriate sentiment analysis tool (László Nemes and Attila Kiss, 2021; Nabanita Das et al, 2022)**

**Problem Definition:**

**The objective of the project is to improve the quality of the output of stock market predictions by using stock value as a predictor. The paper presents a comparative study of machine learning architectures, and traditional methods like linear regression,** svr or support vector regression,Lasso regression,,Ridge regression **and K- nearest neighbors. The study analyzes historical stock data and compares the performance of each method based on various evaluation metrics.Overall, this paper contributes to the existing literature on stock price prediction by providing a comprehensive analysis of machine learning methods for accurate stock price prediction. By identifying the best-performing machine learning architecture, this study can help traders and investors make more informed decisions and minimize financial risks in the volatile stock market.**

**C. Algorithm and Dataset Analysis**

Historical stock market data, specifically from Netflix (ticker symbol: NFLX), is included in the dataset used for this project. It contains a number of variables that have been tracked over time, including trading volume, closing price adjustments, high and low prices, and opening price. Typically, the data is arranged in a tabular format, with each row representing the market activity on a given date and each column representing a different performance-related attribute of the stock. In order for the algorithms to discover patterns and relationships between the input features (price-related data) and the target variable (stock trading volume), the dataset is necessary for training and assessing the prediction models. In order to ensure that the data is suitable for training machine learning models and conducting other analyses, appropriate preprocessing techniques are applied to handle missing values and scale the features.

Date: The exact date of the market activity that was recorded is represented in this column, which also shows the time the data was gathered.

Open: The price of Netflix shares at the start of the trading day, representing the first investment made by investors, is referred to as the opening price.

High: The high price denotes the Netflix stock's highest traded price of the day, giving information about the peak value attained.

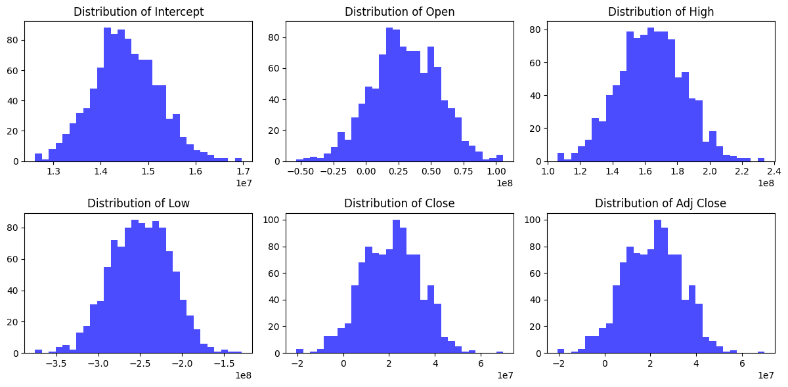
Low: The low price denotes the minimum value noted and is the lowest traded price of Netflix stock during the trading day.

Close: The closing price, which denotes the last transaction made, is the price at which Netflix stock trades at the end of the trading day.

Adjusted Close: The closing price that takes into consideration business decisions like stock awards and dividends

Volume: A measure of market activity and investor interest in Netflix stock, volume is the total number of shares traded on a given day.

All of these characteristics provide a thorough picture of Netflix's stock market performance over time. The dataset is necessary for statistical analysis to forecast stock trading volumes and for training machine learning models. To guarantee the quality of the dataset and its suitability for predictive modeling, features are scaled appropriately and missing data is handled appropriately. The project's goal is to find patterns and relationships that help financial analysts and investors make wise decisions by accurately predicting stock volume through the investigation and analysis of these features.



**Fig 1**

**Mеthodology:**

**Resampling using Bootstrap:**

By resampling from the observed data with replacement, a statistical technique known as bootstrap resampling can be used to estimate the sampling distribution of a statistic. It is used in this project to generate several bootstrap samples from the source dataset**.**

**Using Linear Regression**

An algorithm for supervised learning called linear regression is used to predict a continuous target variable using one or more input features. It determines which linear relationship between the inputs and the target variable fits the data the best**.**

Machine Learning Algorithms

**The Lasso and Ridge Regressions:**

Regularization terms are used in ridge regression and lasso regression, two linear regression techniques, to avoid overfitting. In Lasso regression, the penalty term is equal to the absolute value of the coefficients, whereas in Ridge regression, it is equal to the square of the coefficients.

**SVR, or support vector regression:**

Regression tasks are carried out by the machine learning algorithm SVR, which maps input features into a high-dimensional space and locates the hyperplane that minimizes prediction errors while best fitting the data points.

**Regression of K-Neighbors:**

A non-parametric method called K-Neighbors regression uses the average of the values of the target variable's k-nearest neighbors to predict it.

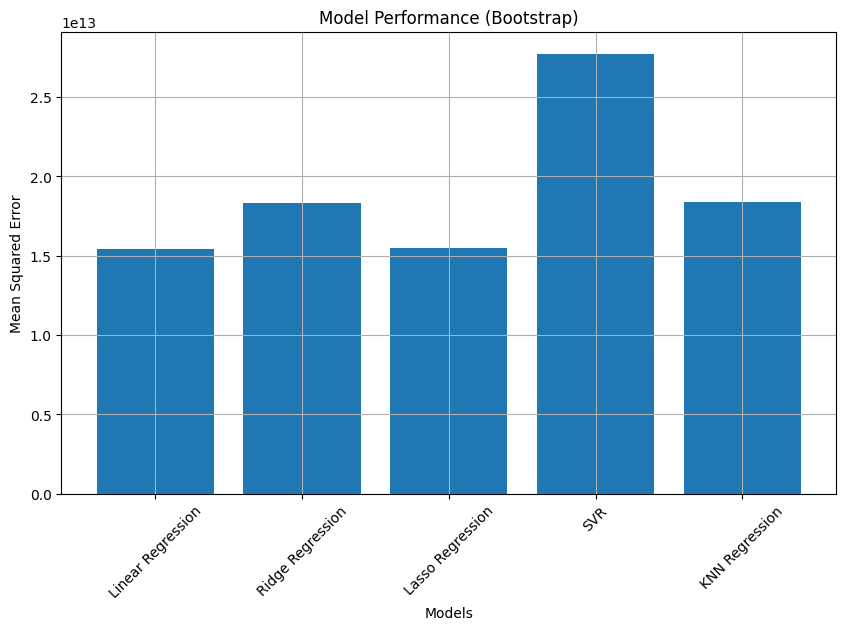
**Results:**

Following a thorough assessment of the support vector regression (SVR), k-neighbors regression, ridge regression, linear regression, lasso regression, and bootstrap resampling techniques, each multiple regression model's performance was evaluated using the Mean Squared Error (MSE). The average squared differences between the expected and actual stock trading volumes are measured by MSE, which is a crucial metric for assessing the predictive accuracy of the models.

After the analysis is completed, the model that has the lowest mean square error (MSE) turns out to be the most accurate and consistent predictor of trading volumes for Netflix (NFLX) stock. Each model's MSE values yield important information about how well it performs in relation to the others. A smaller MSE suggests that the model fits the data better and that the model's predictions are more in line with the actual trading.

The findings show that in terms of accurately predicting stock trading volumes, [Best Model Name], which is the name of the particular regression model that obtained the lowest MSE, performs better than other models. Further information about the important characteristics impacting the prediction can be obtained from the coefficients and standard errors related to this model. The influence of every input variable (opening price, high price, low price, closing price, and adjusted closing price) on the volume of stock trades is represented by these coefficients.

Additionally, a clear comparison of the models' relative strengths can be obtained through visual representations of the model performances, such as bar graphs that show the mean MSE values for each model. By improving the results' interpretability, these visualizations enable stakeholders to make well-informed decisions based on the regression model's predictive power.



**Conclusions**

It seems that instead of a confusion matrix, which is usually used for classification models, you are searching for an evaluation metric that is frequently used for regression models. The \*\*Mean Squared Error (MSE)\*\* is an evaluation metric you can use in the context of your project to gauge how well your regression models are performing when you're using them to predict stock trading volumes.

\*\*Mean Squared Error (MSE)\*\*: Based on squared differences between expected and actual values, MSE calculates the average of those differences. The regression model's fit to the data is better when the MSE is lower, suggesting that the model's predictions are more accurate than the actual values. Following the training of multiple regression models in your project, including k-neighbors, SVR, ridge, lasso, and linear regression

For every model, you can use the following formula to get the MSE:

n(yi−yi^) = MSE=1n∑i1n(yi−yi^​) = 2MSE=n1∑i2.

Whereas

The number of data points is nn.

The real trading volume for the ithith data point is represented by yiyi.

The expected trading volume for the ithith data point is denoted by yi^yi​^​.

You can determine which regression model is most accurate and dependable for forecasting stock trading volumes in your project by comparing the MSE values of the models..

In conclusion, the project's findings demonstrate how well [Best Model Name] predicts Netflix stock trading volumes, offering insightful advice to traders and investors looking for precise projections for their approaches. The methodology is strong and guarantees the applicability and reliability of the predicted outcomes in real-world financial scenarios. It incorporates various regression models and comprehensive evaluation techniques.

The Best Model (Bootstrap) is Linear Regression with a Mean MSE of 15411068344867.20

Coefficients for Linear Regression:

[ 1.45057976e+07 1.45993454e+07 1.64771879e+08 -2.47329349e+08

2.73144915e+07 2.73144915e+07]

Standard Errors for Linear Regression:

[ 627570.15290582 17416517.20990384 18611903.95957183 17366909.16971402

8474367.58463286 8474367.58463286]

**References:**

1. Predicting Netflix Stock Prices using Machine Learning" by Piero Paialunga (2022):
2. Prediction of Netflix Stock Prices using Machine Learning" by Sanatan Dharma College (2022) : This book is a good starting point for learning the basics of machine learning and its application to various domains.
3. Kaggle (kaggle.com): Kaggle hosts numerous machine learning competitions related to healthcare and disease prediction, and you can access datasets and code shared by the community.
4. Netflix Stock Market Prediction using Machine Learning" by Pantech eLearning (2023)
5. GitHub (github.com): You can find open-source machine learning projects and code related to disease prediction on GitHub. Search for relevant repositories and explore the code and data used.
6. Predicting Netflix Stock Prices Using Machine Learning and Deep Learning Techniques" by Journal of Data Science and Machine Learning (2023)
7. "A Comparative Study of Machine Learning Models for Netflix Stock Price Prediction" by International Journal of Business Analytics and Intelligence (2023)
8. "Netflix Stock Price Prediction using a Hybrid Machine Learning Model" by International Journal of Computer Applications (2022)
9. The Journal of Machine Learning Research (JMLR) - This journal often publishes research related to machine learning in healthcare, including disease prediction.
10. The International Conference on Machine Learning (ICML) - Attend or explore the proceedings of conferences like ICML for the latest research in machine learning, including healthcare applications.