

# Predictive Analysis of Market Movements in Stock Markets: Prediction of Trade Volumes using Stock and Interest Rate Data for the Stock Market in the US

Ragi Akshaya Reddy  
Marketing and Data Science  
Regis University  
Regis University , Denver, CO, USA  
aragi@regis.edu

## EXECUTIVE SUMMARY

Market movements represent a direct and influential aspect of decision making for traders. However, the dynamic nature of business environment imply that market movements are equally dynamic. Predictive analysis of trade volumes, as an indicator of market movements, offers an approach that would make decision making equally dynamic for traders. This project conducted predictive analysis of trade volumes using linear regression, random forest and gradient boosting models to develop best modelling approach for dynamic decision making from market movement perspective. The tuned gradient boosting model was identified as the best model, with a predictive performance at 92.42

## I. INTRODUCTION

### A. Background

The pace at which businesses are conducted presently is high. This fast-paced business environment is as a consequence of advanced technology, resulting in a dynamic business environment; consequently, market movements across stocks are as well dynamic (Zhang et al., 2023; Bindeeba et al., 2025). Although other factors are influential in trader decision making, market movements show the level of confidence on the stocks and therefore give a direct indication of market reaction (Yamani, 2023; Care & Cumming, 2024). Market movements, despite not showing the shifts in stock value, give insights into how confident the market is in the shifts (Pancić et al., 2023). This project will conduct a predictive analysis of market movements and hence allow for the forecasting of the market confidence in shifts in stock value.

### B. Problem Statement and Motivation

As an influential aspect of trader decision making, understanding of market movements gives traders an added advantage. However, the dynamic business environment and consequently dynamic market movements imply that traders cannot rely on an overview of market movement. Rather, a deeper interrogation of market movements, through data analytics is necessary to allow traders' decision making to be dynamic enough to match the present dynamic business environment. The ability of market movements in informing on the confidence of the market in the shifts in stock values implies that predictive analysis of market movements has the potential enabling traders to make anticipatory and preemptive trades that would positively impact portfolios.

## II. DATA COLLECTION AND PREPARATION METHODOLOGY

The approach taken to addressing the problem from a data point of view, for this case is summarized in Figure 1 below.



Figure 1

From the chart above, the data collection for the project involved the identification of the top 50 most valuable stocks in the US market using web scrapping in Yahoo Finance (2026). This was followed by the collection of daily data on the last 6 months for the identified stocks through Alpha Vantage (2026). The data collected from Alpha Vantage (2026) contained the following attributes; Adjusted Close. Stock Exchange Name, Regular Market Volume, Regular Market Open, Regular Market Change (%) and Market Capitalization. Following the collection of the stock market data, trade volumes data was then collected from NASDAQ Trader (2026) across stocks exchanges. Interest rate data collection formed the last data collection stage from FRED (2026) to give an external perception of market movements. Data collection, presented above in Figure 1, utilized both access to API data and web scrapping. Whereas API provides access to live data from websites, web scrapping provides access to website data that is not readily available through API or download. The Alpha Vantage API, available from Alpha Vantage (2026), formed the data source that was accessed through API among the four sources of data. Web scrapping was used to collect data from the following three remaining data sources: Yahoo Finance (2026), NASDAQ Trader (2026) and FRED (2026). Cleaning of the data was completed for each of the four data sources including filtering. The data was filtered either during the process of collection or after the collection of data from API or website. For the case of filtering during the data collection process, specifications were set to limit the scope of the data being collected from the API and web scrapping. In the case of the filtering of data after collection, filtering functions within Python were employed for the already collected data. The filtering process first targeted the scope of the data in terms of the stocks to retain, in this case the 50 most valuable stocks were retained in the data on the stock market. Secondly, filtering targeted a 6-month study period to capture data for the period covering the last 6 months. Data integration, through merging, followed

the data cleaning process. The target for the filtering of the data also formed the merging target for the data. Using date and the 50 most valuable stocks, data on the stocks, interest rate and trade volumes were merged into the final dataset for the study. After the merging of the datasets, the stock exchange name variable was converted into dummy variables prior to the modelling. The data preparation for the predictive analytics mainly involves the data partitioning into training and testing sets to allow for the model to be trained and tested using separate datasets (Kaur et al., 2024). The final dataset from the data integration process and creation of dummies for the stock exchange name variable was partitioned such that 80% of the data was allocated to the training set while 20% was allocated to the testing set.

Table 1 below gives the summary of the variables in the final dataset.

*Table 1*

| Variable                  | Description   | Nature of Variable                                  |
|---------------------------|---|---|
| Trade Volumes             | Indicator of market movements showing the amount of stocks traded for the given observation date. | Dependent variable measured on the ratio scale.     |
| Adjusted Close            | The price of the stock at close of trade for the given observation date.                          | Independent variable measured on the ratio scale.   |
| Stock Exchange            | Categorical variable with levels giving the respective stock exchange where a stock is listed.    | Independent variable measured on the nominal scale. |
| Regular Market Volumes    | Average traded volumes for the current trading session.   | Independent variable measured on the ratio scale.   |
| Regular Market Open       | The price of the stock at close of trade for the given observation date.                          | Independent variable measured on the ratio scale.   |
| Regular Market Change (%) | The percentage change in stock price from previous close price.                                   | Independent variable measured on the ratio scale.   |
| Market Capital            | Total market value of the outstanding shares for the stock.                                       | Independent variable measured on the ratio scale.   |
| FF                        | FRED (2026) interest rate indicator   | Independent variable measured on the ratio scale.   |

### III. ANALYSIS METHODS AND IMPLEMENTATION DETAILS

The solution for the challenge of dynamic market movements is based on predictive analysis of market movements to potentially enable traders to make anticipatory and preemptive trades that would positively impact portfolios. Predictive analytics offers an approach for modelling where a known combination of output and inputs trains a model for the determination of the output given a completely new set of inputs (Jamarani et al., 2024; Kaur et al., 2024). In order to implement this solution approach, the following predictive models were trained and tested for predictive performance for market movements; Linear Regression, Random Forest and Gradient Boosting. The three predictive models are distinct from each other and offer alternatives for the prediction of market movements. According to Azure (2024), an equation-based approach to predictive analytics is the linear regression; which describes the equation of

the output being a linear function of the inputs. The random forest amplifies the classification tree through employing several decision-based algorithms that split conditions for inputs in a tree like manner until all inputs are exhausted and the output predicted at the base (Ojo & Ayodele, 2025). Gradient boosting is explained in Rizkallah (2025) as an ensemble tool that trains weak predictive models sequentially with each new model learning from the previous model's residuals; and combines these models to form a predictive tool. According to Jamarani et al. (2024), the parameters for the predictive models can be tuned to improve on the predictive performance. Consequently, for each of the three predictive models, both the baseline and the tuned models will be developed and assessed for predictive performance. Presented below in Figure 2 is the summary of the methodology approach.

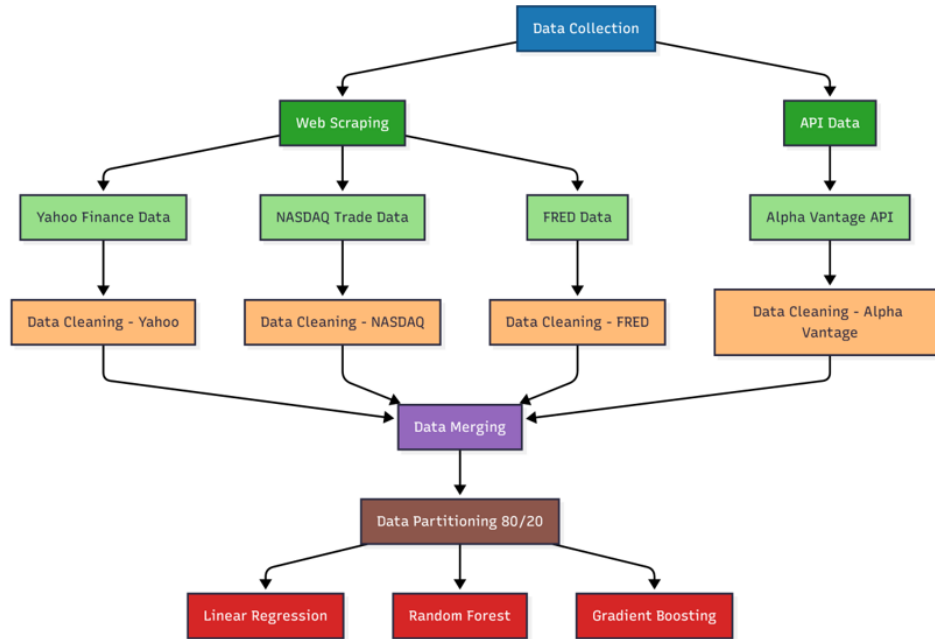


Figure 2

#### IV. RESULTS AND FINDINGS

Presented in Figure 3 below is the overview of the trade volumes across all US Stock markets. The data captures the period between December 17, 2025 and January 30, 2026 for 30 trading days across all the stock markets in the US. The line charts in Figure 3 show the trend in both the average total moving (for volumes moved) and average total (for total traded volumes). The average moving volumes show a general rising trend as the month progresses while the average total volumes generally fluctuate about the 2 billion shares mark with a notable spike (day 19) and drop (day 24) as the month progresses.

Trend in Total Trade Volumes

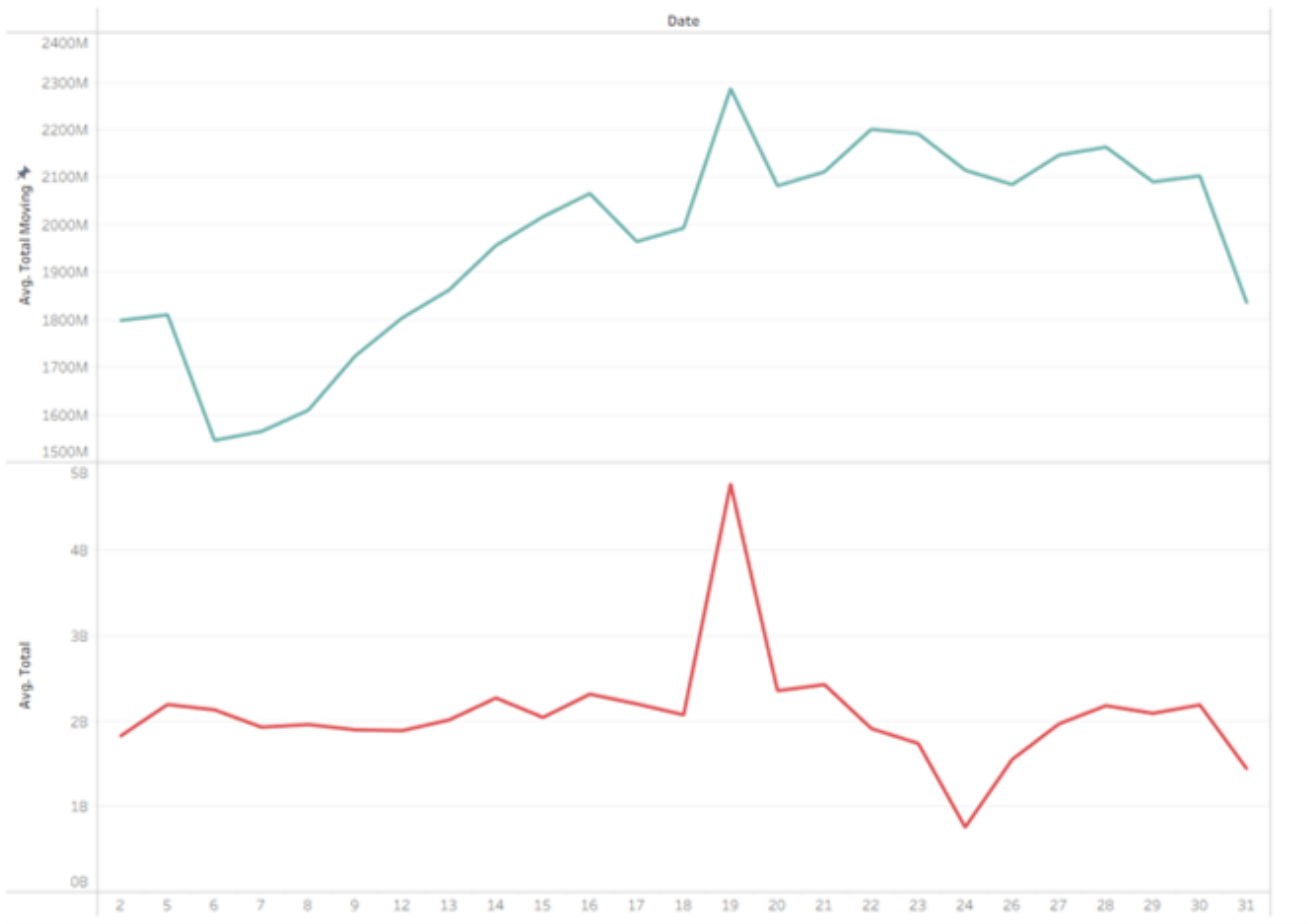


Figure 3

In Figure 4 below is the comparison of the total moving volumes across the different stock exchanges in the US over the 30-day period. We note that majority of the volumes traded were in FINRA/NASDAQ TRF Exchange, followed by the NASDAQ, ARCA and NYSE Exchanges.

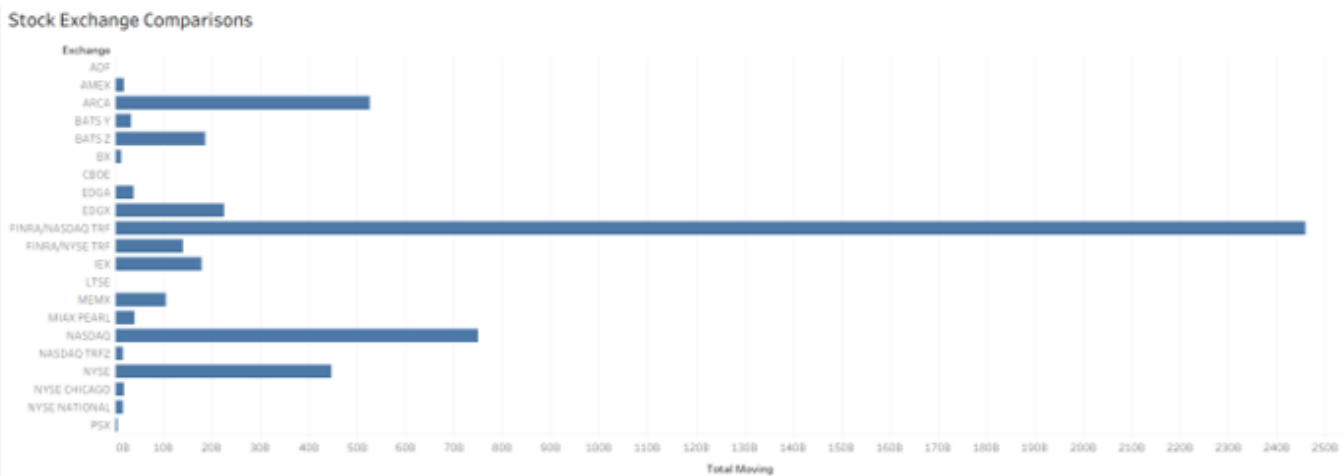


Figure 4

The results of the predictive analytics are summarized in Table 2 below. The metrics, MSE and R-

Squared are measured for the testing set and used for the comparison of the predictive performance across the three models.

Table 2

|                       | Linear Regression |           | Random Forest     |           | Gradient Boosting  |           |
|-----------------------|-------------------|-----------|-------------------|-----------|--------------------|-----------|
|                       | MSE               | R-Squared | MSE               | R-Squared | MSE                | R-Squared |
| <b>Baseline Model</b> | 1509392215684.88  | 0.8527    | 1093542355545.121 | 0.8933    | 1171378827227.7117 | 0.8857    |
| <b>Tuned Model</b>    | 1520763596104.184 | 0.8516    | 789289944668.3167 | 0.9230    | 776845239846.1954  | 0.9242    |

For the linear regression model, the baseline model had both the lowest MSE and the highest R-Squared, making the baseline model the best linear regression model. In the random forest modelling, the tuned model had both the lowest MSE and the highest R-Squared, making the tuned model the best random forest model. Gradient boosting modelling showed that the tuned model had the lowest MSE and the highest R-Squared, making the tuned gradient boosting model the best gradient boosting model. With an R-Squared value of 0.9242, the tuned gradient boosting model had a predictive performance of 92.42%, and hence the highest predictive performance across the three models.

## V. CONCLUSION

The results from the descriptive analysis indicated that the average moving volumes showed a general rising trend as the month progressed with the total moving volumes being highest for the FINRA/NASDAQ TRF Exchange, followed by the NASDAQ, ARCA and NYSE Exchanges. R-Squared for the baseline and the tuned models across all three models showed that the predictive performances exceeded 85%; hence Adjusted Close, Stock Exchange, FF (Representing Interest Rates), Regular Market Volumes, Regular Market Open, Regular Market Change Percentage and Market Capital can sufficiently predict the trade volumes for the market movements. The best model for the prediction of market movements would be the tuned Gradient Boosting Model with a prediction performance of 92.42%.

## REFERENCES

- Alpha Vantage. (2026). *API documentation*. Alpha Vantage. <https://www.alphavantage.co/documentation/>
- Azure, I. (2024). *Predictive modeling for industrial productivity: Evaluating linear regression and decision tree regressor approaches*. *Journal of AppliedMath*, 2(4), Article 1435. <https://doi.org/10.59400/jam.v2i4.1435>
- Bindeeba, D. S., Tukamushaba, E. K., & Bakashaba, R. (2025). *Digital transformation and its multi-dimensional impact on sustainable business performance: evidence from a meta analytic review*. *Future Business Journal*, 11, Article 90. <https://doi.org/10.1186/s43093-025-00511-z>
- Carè, R., & Cumming, D. J. (2024). *Technology and automation in financial trading: A bibliometric review*. *Research in International Business and Finance*, 71, Article 102471. <https://doi.org/10.1016/j.ribaf.2024.102471>

- Federal Reserve Bank of Cleveland. (2026). *Inflation nowcasting*. FRED. <https://www.clevelandfed.org/indicators-and-data/inflation-nowcasting>
- Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). *Big data and predictive analytics: A systematic review of applications*. *Artificial Intelligence Review*, 57, Article 176. <https://doi.org/10.1007/s10462-024-10811-5>
- Kaur, B., Sidhu, B. K., & Bhathal, G. S. (2024). *Predictive analytics models for commodity market: A literature review*. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3), 4525–4535. <https://doi.org/10.17762/ijisae.v12i3.7958>
- NASDAQ. (2026). *NASDAQ Trader*. NASDAQ. <https://www.nasdaqtrader.com/>
- Ojo, G. O., & Ayodele, E. (2025). *Evaluating the predictive power of random forest regression on economic growth*. *Federal Polytechnic Ilaro Journal of Pure and Applied Sciences*, 7(1). <https://fepi-jopas.federalpolyilaro.edu.ng/index.php/journal/article/view/165>
- Pancić, M., Ćucić, D., & Serdarušić, H. (2023). *Business Intelligence (BI) in Firm Performance: Role of Big Data Analytics and Blockchain Technology*. *Economies*, 11(3), 99. <https://doi.org/10.3390/economies11030099>
- Rizkallah, L. W. (2025). *Enhancing the performance of gradient boosting trees on regression problems*. *Journal of Big Data*, 12, Article 35. <https://doi.org/10.1186/s40537-025-01071-3>
- Yahoo Finance. (2026). *Most expensive stocks screener*. Yahoo Finance. <https://finance.yahoo.com/research-hub/screener/most-expensive-stocks>
- Yamani, E. (2023). *Return–volume nexus in financial markets: A survey of research*. *Research in International Business and Finance*, 65, Article 101910. <https://doi.org/10.1016/j.ribaf.2023.101910>
- Zhang, F., Yang, B., & Zhu, L. (2023). *Digital technology usage, strategic flexibility, and business model innovation in traditional manufacturing firms: The moderating role of the institutional environment*. *Technological Forecasting and Social Change*, 194, Article 122726. <https://doi.org/10.1016/j.techfore.2023.122726>