**Final Report**

**Table of Contents**

1. Abstract
2. Introduction
3. Purpose and Objectives
4. Scope
5. Data Description
6. Business Understanding/ Analytics Questions:
7. Exploratory Data Analysis
8. Data Visualization
9. Methods /Modeling
10. Model Evaluation
11. Feature Work
12. Conclusion
13. References

**Abstract:**

This project analyzes the dynamic and complicated world of stock market investments by thoroughly examining historical stock data for seven significant businesses: AT&T, JP Morgan Chase & Co., Spotify Technology, Chipotle Mexican Grill, NVIDIA Corporation, TATA ELXSI, and Spotify. The objective of this study is to evaluate the impact of various market factors on stock performance and to find patterns in daily returns and trading volumes using advanced statistical approaches and modeling techniques like ANOVA and linear regression. Including insights on market patterns, volatility, and the financial soundness of the participating companies, the research offers a thorough knowledge of how these stocks have performed over time. The study helps investors make well-informed decisions by combining comprehensive data analysis with attractive visualizations to match investment strategies to market opportunities and risk profiles. The analysis's conclusions are meant to improve investment strategies by providing a data-driven understanding of market dynamics. This project examines the correlations present among various trading behaviors and changes in stock prices, as well as the consequences of these interactions for predicting future performance. These kinds of information are important for financial feature and portfolio management since they will offer a solid platform for both short-term trading and investment decisions.

**Introduction:**

The important to collect and analyze large volumes of data could significantly improve the investments and decision-making in the fast-growing world of financial markets. With the help of this committed group of data analysts and finance enthusiasts, this project will attempt to analyze historical stock data from seven top companies like, AT&T, TATA ELXSI, Spotify Technology, JP Morgan Chase & Co., NVIDIA Corporation, Chipotle Mexican Grill, Meta, and AT&T. The goal of the project is to analyze the complexities of stock market behaviors. The project aims to give a detailed knowledge of the variables influencing stock performance by carefully examining daily returns, trading volumes, and price variations. The project is based on modeling and analyzing stock market patterns using advanced statistical techniques and tools. This study analyzes the correlation between trading volumes and stock prices by using ANOVA and linear regression to see how these variables change over time. The present study gives valuable information for investors seeking to manage their portfolios by showing not only the normal characteristics of these equities but also their volatility and related hazards. In addition, the project handles the need for comprehensive risk evaluation in portfolio management, helping investors to make more educated decisions based in quantitative analysis and empirical data. The analysis assists in developing strategies that strike a balance between current market risks and anticipated earnings by finding patterns and connections in the data. This strategy is especially important in a time when technology developments and global events are having a greater impact on financial markets, making traditional investment methods risky and more difficult.

**Purpose and Objectives:**

This project's main goal is to analyze and understand the complexities of the stock market by doing a thorough statistical analysis of historical stock data from seven eminent corporations. These businesses have been carefully selected for their market power and the variety of industries they represent. This study examines for patterns and trends in stock price, trading volume, and daily return changes that may indicate future market behavior.

The following are the project's main objectives:

**Trend analysis:** To identify long-term patterns and trends in stock price that could throw a spotlight on the overall health and prospects of the selected business.

**Predictive Modeling:** Based on historical data, the project aims to develop predictive models that can anticipate changes in stock prices using ANOVA and linear regression. This is evaluating how well a variety of market indicators, including opening, and closing prices, highs and lows, and trade volumes are able to predict future events.

**Comparative Analysis:** Based on previous performance, analyze the selected companies stocks to figure out which ones may have the most potential for dividends, growth, or stability.

**Scope:**

The goal of this project was to identify underlying patterns and trends in daily returns, volume, and price changes by analyzing stock market data from several major companies. The study focused on predicting stock behaviors and identifying important variables impacting market dynamics using innovative statistics and machine learning models. This gave stakeholders access to a dependable platform for understanding complex financial statistics as well as predicting insights that may guide investment decisions. In order to ensure a thorough approach to addressing the issues of financial data analytics, the project included significant stages of data cleaning, exploratory analysis, model creation, and validation. In order to further improve the estimated accuracy, this study aims to provide the basis for future research that may include macroeconomic features, multi-variable models, or other modeling approaches like deep learning.

**Data Description:**

**Date:** This column keeps track of the date when the stock data was entered. Every entry is associated with a certain day and is usually written as YYYY-MM-DD.

**Open:** The stock price at the start of the trading day.

**High:** The price at which the stock achieved its highest point throughout the trading day.

**Low:** The share price that was traded at its lowest point all through the trading day.

**Close:** The stock price at the end of the trading day.

**AdjClose:** The adjusted closing price takes into consideration all company events, including rights provides, dividends, and stock splits.

**Volume:** This shows the number of shares that were traded all over the duration of the day. Increasing volumes may be evidence of increased interest or significant business news.

**Company:** Identifies the company that holds the stock data. Listed within the businesses in your dataset are NVIDIA, TATA ELXSI, and more companies.

**Daily\_Return:** It displays the percentage change in the stock price between the closing prices of the day before and the day after.

**Year:** Collected from the Date column, this indicates the year that the stock data was taken, allowing for annual analysis.

A table of numbers and numbers

Description automatically generated

**Business Understanding / Analytics Questions:**

1. **What is the link between daily returns and trading volume for each company in the dataset?**

The purpose of this question is to find out if significant changes in business volume whether positive or negative precede or follow significant changes in prices. Understanding this correlation could help in accurately predicting market timing, particularly for traders and investors with short time spans.

Furthermore, it is useful with deciding whether stocks are more variable because of large trading volumes, which may help portfolio managers calculate risk.

1. **What is the reaction of stock prices to news specific to a sector or market-wide economic announcements?**

Investors can assess a stock's sensitivity to external variables by looking at how the price of the stock responds to significant events. This information is crucial for creating strategies for profiting from or protect against such events.

Additionally, by showing the mutual dependence of certain sectors with sectoral developments and shifts in the world economy, this study may offer new ideas on diversification strategies.

1. **Which company's daily high and low prices show the least number of variations, suggesting the most consistent stock price?**

Finding the most stable companies may be especially helpful for risk-averse investors who want to minimize their exposure to daily market volatility while focusing on long-term benefits.

Stability analysis may also indicate whether companies, based on their performance in relation to market and sector changes, may be undervalued or overvalued.

1. **What trends can be observed in these companies stock performance over time, and how may these trends help guide investing strategies?**

Knowing seasonal patterns and cyclical trends might make it easier to time the market and make investments at the best periods of the year. These kinds of insights are useful for building a diversified portfolio that can withstand or benefit from stock performance moves that are predictable.

**Exploratory Data Analysis:**

Below is the code we did in the first step of our project. The method of numeric conversion is used to many stock variables, such as Open, High, Low, Close, Adjusted Close, and Volume for each business, in the initial step of data preparation for the exploratory data analysis (EDA). By ensuring that all relevant columns have numerical formats, this standardization makes it easier to do accurate and efficient statistical analysis and comparisons between various datasets.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

The combined\_data dataset's structure str() shows that it has 12,912 observations distributed across 10 variables, representing the day-to-day activity of the stock market for multiple companies. Several stock attributes are included in the dataset: Date, Volume (numeric), Company (a factor with seven levels representing different companies), Daily\_Return (numeric), and Year (numeric). For evaluating stock performance and trends as well as creating projections for the future, this structure is essential.

A close up of numbers

Description automatically generated

The below output following provides descriptive data for every variable. It records the trade volume, date range, and stock price metrics (Open, High, Low, Close, and Adjusted Close) for multiple companies between 2015 and 2024. It helps identify volatility by displaying the range of daily returns and displaying the minimum and highest numbers. Understanding core tendencies and dispersion is made easier with the addition of quartiles and means for each stock indicator. This is important for examining patterns in stock performance and helping investors make well-informed selections.

A close-up of numbers

Description automatically generated

The command sum(is.na(combined\_data)) provides a total count after checking the full dataset for missing values. In this case, it shows that the dataset has 0 missing values, indicating that there is no need for imputation, or the removal of missing entries and that the data is complete.

A close-up of blue text

Description automatically generated

The code distinct(data\_clean), which indicates that duplicate items have been successfully removed from the dataset, is given as evidence of this. By making sure every data point is distinct, this stage improves the standard of the investigations that come after it. Having all original columns Date, Open, High, Low, Close, Adjusted Close, Volume, Company, Daily Return, and Year retained, the printed result shows a sample of the unique data that is crucial for comprehensive stock market analysis. Ensuring data is reliable and authentic in financial modeling is dependent upon this refreshing.

A screenshot of a computer code

Description automatically generated

After normalization, the stock data is displayed in the below output. All numerical variables, including Open, High, Low, Close, Adjusted Close, and Volume, are scaled between 0 and 1. By normalizing the data range, this transformation makes it simpler to compare results across scales and encourages more efficient analysis. Both the Daily Return and the Volume have been normalized; the volume shows a significant amount of change, which may be connected to trading activity. For this part of the dataset, the Year column has been normalized to zero, indicating that this typical has been treated as constant or single.

A table of numbers and symbols

Description automatically generated

The output data\_standardized shows stock data that has gone through standardization, which usually involves dividing each variable's value by its standard deviation and eliminating the mean. By putting the data on a common scale with a mean of zero and a standard deviation of one, this transformation normalizes the data. By considering all variables equally, this enhances the performance of many statistical models and is essential for comparison analysis across multiple metrics. Changes in trading activity and the intensity of price movement on different days are shown by variations in the volume and daily return numbers.

A table of numbers with red and black text

Description automatically generated

The result displayed is a transposed form of the summary statistics for a stock performance dataset. The data was initially shown in a regular tabular way, with each row indicating a statistic for an independent variable (Open, High, Low, Close, Adjusted Close, Volume, Daily Return) (mean, median, standard deviation, interquartile range). It is now easier to read and analyze the data because of the transposition, where each row now represents a variable, and the columns represent different statistics. As an example, the initial price's average (mean) is 2662.36, but its median is significantly lower at 200.47, suggesting a skewed distribution. The large standard deviations show significant instability in both trading volume and stock prices, especially in volume. The below output enables comparative analyses of several variables, making it easier to recognize trends and deviations in the dataset.

A screenshot of a computer

Description automatically generated

**Data Visualization**

The below graph shows daily return as well as histograms for each important stock measure, including open, high, low, close, adjusted close, and volume. Each variable's range of values is shown on the x-axis, while the frequency of observations is shown on the y-axis. With few outliers dealing at higher values, the majority of variables, including Open, High, Low, and Close, are heavily skewed to the left, indicating that most stocks move within a lower range. As is typical of stock market returns, the Daily Return histogram centers around zero, meaning that most daily returns are almost unchanged.

A screenshot of a graph

Description automatically generated

The below time series graph shows a comparative analysis of the closing stock prices of several companies from 2015 till after 2023. AT&T (red), CMG (orange), JP Morgan (green), Meta (blue), NVIDIA (light blue), Spotify (grey), and TATA ELXSI (pink) are the various companies represented by each line. A large number of companies show flat closing prices in the lower part of the graph, with just a few exceptions over time. On the other hand, JP Morgan's line, which is colored green, indicates an important and rapid rise starting in late 2022, indicating a big increase in the value of their stock. To comparing the performance of multiple businesses in a collection and assessing trends over time quickly, this type of display is important. For example, an in-depth review of the variables that caused JP Morgan's stock's rapid rise, such as market developments, business performance improvements, or industry expansion might provide information about potential opportunities for investment and dangers.

A graph with numbers and lines

Description automatically generated

The trading volume for all companies in the dataset  from 2016 to 2024 is shown in the graph titled as "Volume Traded Over Time". Since every business is represented by an individual color, it is easy to figure out over time which trade operations each one participates in. The vertical spikes indicate times of increased trading volume, which may be connected to significant business announcements, news about the market, or changes in investor mood. Active trading periods involving multiple companies are shown by the densely crowded parts of the graph, especially those with overlapping colors. Investors and analysts may more easily see trends in trade, possible market influences, and market responses to outside events with the help of this representation.

A graph with different colored lines

Description automatically generated

The "Distribution of Daily Returns" graph shows, for a number of companies, the frequency of daily return percentages over an interval of time. The histogram's assigning colors approach allows one to quickly compare the distribution of daily returns across the various companies. Most of the returns are centered around zero, indicating that most days the stocks displayed little to no return a feature common to stable economic times. All of the companies returns are most concentrated around zero, indicating that large gains or losses were fewer in number. On the other hand, the distribution includes tails on both sides, indicating days with significant positive and negative returns. Investors can better understand the risk and volatility of each company's stock by using this type of representation; for example, larger distributions indicate more volatile stocks. Such type of understandings are cruicial for the portfolio management because of they are enable the choice of resources that fit an investor's risk tolerance and help in the decision-making process about risk diversification.

A graph of a number of daily returns

Description automatically generated

The change of AT&T's stock price from January 2020 to December 2022 is shown in the below graph. The three-year timeline appears to be on the x-axis, and while the closing stock price is to be represented on the y-axis. With a sharp drop at the start, a recovery period from mid-2020 to 2021, and another decline in 2022, the chart shows significant swings. The graphic highlights the difficulties and successes the business had during this time, which may have been impacted by external economic variables, business performance, and market conditions.

A graph with numbers and lines

Description automatically generated

The frequency distribution of daily stock returns for every year from 2015 to 2024 can be seen in the graph "Distribution of Daily Returns Over Years". Every layer represents a single year, with colors changing from dark purple to outstanding yellow in subsequent years. The fact that the distributions cluster around zero indicates most days have very little variation in returns. But each year's distribution is different, with varying spreads and shapes that correspond to varying degrees of market volatility. Investors can better grasp the temporal variability in market risk by looking at distributions that are narrower to signal more stable years and larger to show years with greater price swings.

A graph of a number of numbers and a number of daily returns

Description automatically generated with medium confidence

The "Average Daily Return by Company" graph shows how seven companies performed in comparison to one another based on their average daily returns. For easy identification, each bar is color-coded to represent a specific company. Particularly, with significantly higher average daily returns than the others, JP Morgan and NVIDIA stand out and suggest better stock performance. While AT&T, CMG, and SPOT have slightly lower returns, META and TATAELXSI additionally demonstrate remarkable returns. For investors who want to figure out which companies provide longer-term average returns and direct their financial strategies toward greater financial opportunities, this representation is important.

A graph of different colored bars

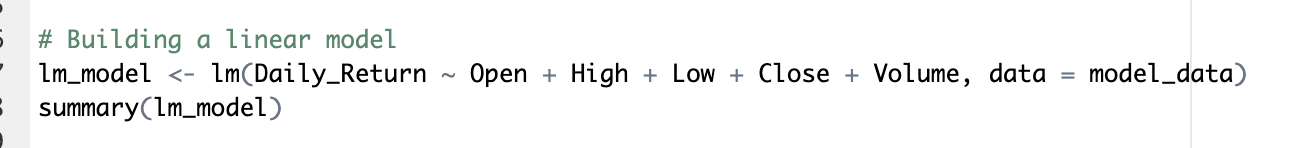
Description automatically generated

**Methods / Modeling:**

We use statistical methods to carefully examine the data in relation to our provided questions in the Methods and Modeling portion of our research. The initial part of this section explains the procedures followed in the data analysis, including the models and statistical tests that were used. We make sure that every technique is suitable for the queries that are at hand as well as the properties of the data. To guarantee the validity and reliability of our study, the reasoning behind the selection of particular approaches is discussed, setting the foundation for an in-depth analysis of the findings that follows. This methodical process makes sure that findings are well-supported by the available evidence. We used appropriate statistical methods that match with the features of our data and the goals of our research. Our models are based on relevant and reliable data inputs since every variable in the study has been carefully chosen based on its relevance to the investing questions inquired.

**Linear Regression:**

Our project's linear regression model provides an in-depth look of the variables influencing daily stock returns. To identify the underlying patterns in stock price movements and business dynamics, we incorporate data like Open, High, Low, Close, and Volume into our model. This model helps in assessing the opening, closing, and intraday changes predictive value on the day's results. The model's coefficients quantify the impact of each variable, giving insight on how daily stock profitability is impacted by various market conditions and trade volumes. Developing strategies and risk management are influenced by the statistically significant components that are projected to have consistent impact on returns, as shown by the significance levels given to each coefficient. For predicting future movements and optimizing trading decisions based on previous information, this analytical method is important.



The linear regression model's summary output provides useful data about the correlation between a dataset's daily returns and other stock market variables. According to the model, a significant number of predictors have significant effects on daily returns, prices that are "open," "high," "low," and "close" have particularly large affects (positive for low and negative for open). P-values less than 0.05 are found in almost all variables, except for "Close" and "Volume," indicating that these relationships are statistically significant at conventional levels. The model may account explain within 1.985% of the variability in daily returns, according to the Multiple R-squared value of 0.01985. Although relatively low, this proportion is normal for time series data in the financial domain, where many unmodeled elements are involved. The model is statistically significant, offering accurate projections of the effects of the included variables, as shown by the F-statistic of 52.28 on 12906 degrees of freedom.

A screenshot of a computer

Description automatically generated

**ANOVA:**

The output displayed is the result of an ANOVA (Analysis of Variance) test that used pairwise comparisons to compare the mean daily returns of several companies. The contrast between company combinations, the estimated difference in mean returns, the t-ratio, the degrees of freedom (df), the standard error (SE), and the p-values are all listed in the table. At a 95% confidence level, a non-significant p-value (higher than 0.05) suggests there is no statistically significant difference in the daily returns between the firm pairings. As a case study, the comparison of AT&T and NVIDIA's returns shows a statistically significant difference with a significant p-value of 0.0094. Other couples, however, such as AT&T and CMG, do not exhibit a statistically significant difference (p-value 0.9842). This study supports the process of making statistically informed investment decisions by identifying the companies whose returns differ considerably from those of the others.

A screenshot of a computer screen

Description automatically generated

As shown by the p-value of 0.0193, the given ANOVA results show a statistically significant difference in daily returns across the different companies under consideration. This p-value indicates that the average daily return of at least one business differs significantly from the others, as it is less than the generally accepted threshold of 0.05. In order to determine if the variations in each company's returns may be the result of chance, the research involved comparing the means of its daily returns. The F-value of 2.523 provides more evidence for the existence of return variability between companies. These findings suggest that greater research into the particular company pairings that differ from one another might be useful in order to guide financial analysis or investment approaches that are more targeted. Understanding how industry or company-specific factors may affect stock performance is an essential initial phase in developing more intelligent financial and portfolio management techniques.

A close-up of numbers

Description automatically generated

**Random Forest:**

The random forest model's (rf\_model) summary provides some important statistics on the characteristics and performance of the model:

**Predictions and Error:** The analysis indicates 500 for the mean squared error (MSE), which might be a placeholder or error as MSE values normally range from 0 to 1 or are scaled based on the context of the situation.

**Variable Importance:** Two measures, **%IncMSE** and **IncNodePurity,** are used to determine each feature's importance in predicting the target variable. The two measures show the corresponding roles of each variable to the accuracy of the model and the homogeneity of the nodes in the trees. Among the most significant indicators are the close and open prices.

A screenshot of a computer

Description automatically generated

**Logistic Regression:**

The variable Return\_Binary in this logistic regression model is designed to categorize daily stock returns as either non-positive (0) or positive (1) depending on whether the Daily\_Return is larger than zero. This method of binary classification reduces the result to a format that can be used with logistic regression, which forecasts the likelihood of a specific result (a positive return in this example). To make sure that both training and testing sets are reflective of the whole data, the dataset combined\_data is then divided into training and testing sets using an 80/20 split. The createDataPartition function from the caret package is used to accomplish the split. It offers an accurate method to generate stratified samples based on the Return\_Binary variable's distribution.

Using the glm function and a binomial family, a logistic regression model is built to indicating a logistic model. The Open, High, Low, Close, and Volume predictors common quantitative metrics used for predicting stock movements are among the numerous predictors included in the model. These variables are predicted to impact the probability of a stock generating a positive gain and offer insights into the daily trading dynamics. By model fitting to the training set, our goal is to identify correlations between these predictors and the probability of a profitable result. This method aids in forecasting future patterns based on previous data, in addition to helping to understand significant factors of stock performance. Metrics such as accuracy, precision, recall, and the AUC-ROC curve which are crucial for validating the model's performance in binary classification tasks will be looked at in order to evaluate the efficacy of this model by looking at its prediction accuracy on the test set.

A screenshot of a computer code

Description automatically generated

**Evaluation:**

For the reason of analyzing the underlying assumptions of the linear regression model that is utilized for predicting daily returns based on stock market indicators, the residual plots that are presented here are important.

**Residuals vs Fitted:** This graph looks for equal variance and non-linearity. As the spread increases along the fitted values, the non-random pattern indicates to a possible variability or non-linearity in the connection.

**Residuals Q-Q Plot:** This quantile-quantile plot analyzes the residuals' normality. The tails' departure from the straight line indicates the possibility of outliers or the chance that the residuals are not normally distributed.

**Scale-Location:** Also called a spread-location plot, this type of plot is used to determine if the residuals are homoscedastic. The upward trend suggests that the variance of the residuals is not constant.

**Residuals vs Leverage:**This graphic helps in recognizing significant situations that might significantly influence the regression line (high leverage points). Points that could be significantly impacting the model's predictions are shown by the distance lines.

A collage of graphs and charts

Description automatically generated

This residuals plot checks the assumption of consistency across fitted values in a linear regression model by showing the residuals on the vertical axis compared to the fitted values on the horizontal axis. The residuals should ideally show no pattern and form a random cloud centered on the horizontal line at zero. With no obvious trend or systematic departing, the residuals appear quite randomly in this figure, indicating that the requirement of equal variance may be somewhat met. The model might not fully account for every relevant factor or interactions, as indicated by the clustering around the center, which points to possible underfitting.

A graph with black circles and red line

Description automatically generated

The variable significance for a Random Forest model is shown on the graph, which measures the impact of each feature (Open, High, Low, Close, and Volume) on the predicted accuracy of the model. '%IncMSE' is used in the left plot to show the amount that each variable decreases the Mean Squared Error (MSE) when it is incorporated into the model; a larger number indicates more significance. The 'IncNodePurity' plot on the right shows the role of each feature to the rise in node purity when the attributes are separated inside the model. Variables that better segment the data are shown by higher values. Understanding what components in the model are the most important and predictive needs this information.

A screenshot of a graph

Description automatically generated

The logistic regression model's results, which classify daily stock returns as positive or negative, provide a variety of performance indications. With a precision of 0.595, the model is around 59.5% accurate when it forecasts a positive return. However, the model's recall of 0.914 indicates that it is quite successful, correctly detecting 91.4% of all real positive returns. There is still opportunity for development, particularly in precision, as seen by the F1 score of 0.7209, which finds a very acceptable balance between recall and accuracy. To increase the model's forecast accuracy, these measures might direct more model modification.

A computer code with numbers and symbols

Description automatically generated

**Feature Work:**

The results of the study indicate the number of interesting areas requiring more research. Adding extra variables, such as industry-specific evaluates or macroeconomic data, would be helpful to see whether it helps us better understand how stock prices change. Experimenting with different machine learning techniques or statistical models may also help us determine which strategies work best in particular economic situations. Analyzing how events across the world affect stock prices directly might be another fascinating subject. To see how rapidly the markets respond, this may include a more thorough time-based study or even real-time data processing. By taking these actions, we might greatly improve our comprehension of the financial markets.

**Conclusion:**

This study utilized innovative statistical methods and machine learning approaches to provide an in-depth examination of the stock market dynamics of various major companies. The results offer further light on market behavior by showing the large impact of daily trade volume and price variations on daily returns. Using a number of models, such as random forests and linear regression, the study showed the complexity of financial data is and how many aspects effect stock performance. Even though the models were not perfect, their ability to forecast daily returns increased as more factors were added, indicating the possibility of even more advanced models in the future. This approach is useful for investors looking to optimize their strategies based on past data trends, as well as for academic understanding. The project's results highlight how crucial it is to continuously evaluate and modify models in order to account for the always changing market conditions.

**References:**

1. Tsay, R. S. (2010). Analysis of Financial Time Series. Wiley-Interscience.
2. Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). The Econometrics of Financial Markets. Princeton University Press.
3. Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting. Springer.
4. Fabozzi, F. J., Focardi, S. M., & Kolm, P. N. (2014). Quantitative Equity Investing: Techniques and Strategies. Wiley Finance.
5. Murphy, J. J. (1999). Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance.
6. Shiller, R. J. (2015). Irrational Exuberance. Princeton University Press.
7. Wilmott, P. (2013). Paul Wilmott Introduces Quantitative Finance. Wiley.
8. Mandelbrot, B., & Hudson, R. L. (2004). The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward. Basic Books.
9. Breiman, L. (2001). "Random Forests", Machine Learning, 45(1), 5-32.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics.
11. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
12. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
13. Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine," Annals of Statistics.
14. Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics.
15. Lantz, B. (2013). Machine Learning with R. Packt Publishing.