PREDICTION OF STOCK VOLUME FROM DOW JONES INDEX USING BUSINESS ANALYTICS CONCEPT

Anastacia[1], Evita Linata[2], Ivana Jocelyn[3], Veline Chensiery[4], Genesis Sembiring[5]

1)Universitas Pelita Harapan, Medan

2) Universitas Pelita Harapan, Medan

3) Universitas Pelita Harapan, Medan

4) Universitas Pelita Harapan, Medan

5) Universitas Pelita Harapan, Medan

e-mail:

anastacianass@gmail.com *)

evitalinata@gmail.com *)

ivanajocelynn@gmail.com *)

velinechensiery@gmail.com *)

ABSTRACT

Because of human constraints in computational reasoning, rational decision making is confined, which is why people experience bounded rationality. In many cases, a decision support system is required to help researchers make decision, including in predicting variables in stock market. Trading volume is one of the most important factors influencing the stock market's behavior. The higher the trading volume of a stock, the more active and liquid the market is which made it more attractive because the more liquid a stock is, the narrower its spread will be and there will be less risk of them being stuck with an undesired stock position. This study is made on the purpose of predicting trading volume from Dow Jones Index using the Business Analytics concept, including Descriptive Analytics, Predictive Analytics, and Prescriptive Analytics. A managerial contribution was also developed for assisting a trader or investors in determining which index to invest on. Even so, the weighting results showed only minor variances in value among the 16 criteria used.

Keywords: *prediction, stock volume, machine learning*

1. INTRODUCTION

The contrast between how humans and robots process information is highlighted by computational thinking. Machines can perform billions and even trillions of calculations per second, whereas people can only do so and correct errors in real time. Computational thinking is sometimes promoted as a universal approach to problem resolution since machines are able turn a mistake in an algorithm into an expensive calamity before any person has a chance to react (Denning, 2019).

The value of a valuation report's quality for the most part depends mainly on how well the expert valuer has applied basic valuation theory to the observations made, the data gathered, and the situation under consideration. But even so, however well an expert valuers may make these decisions, humans are subject to the constraints of human information processing (Amidu, 2020). As a result, data-driven methodologies combined with computational power are potential techniques for developing a decision-making system, and advanced analytics, a major disruptive breakthrough, is increasingly replacing traditional techniques of summarizing, viewing, and reporting data.

Stock trend prediction has long been a fascinating issue that has been explored extensively by experts from several professions. Investors are one profession that concerns in accurate stock market prediction; however, stock markets are driven by fluctuating elements such as microblogs and news, making it difficult to anticipate stock market indexes based just on past data. The significant stock market volatility highlights the need of accurately assessing the impact of external factors in stock forecasting. Machine learning algorithms can be used to forecast stock markets (Wasiat Khan, 2020).

The chance of a share's value falling to far owing to market movements is rare, but despite that fact, the risk still exists. These movements, which have an impact on stock price and trading volume, are not easy to forecast. Furthermore, these movements also have an influence on how people behave in terms of capital savings or investment, stock price, and the increase or decrease of risk for investors. In general, anticipating stock market behavior using various approaches and procedures is a valuable tool to aid investors in acting with better certainty, taking into account the risks and volatility of an investment, and knowing when to buy at the lowest price and sell at the highest price (Garehchopogh, Bonab, & Khaze, 2013).

Machine learning, a well-known technique with numerous uses, has been actively researched for its potential in financial market prediction. Support vector machine

(SVM) and reinforcement learning are two popular algorithms that have been claimed to be highly effective in tracing the stock market and optimizing the profit of stock option purchases while keeping the risk low (Shen).

This study was done on the purpose of predicting stock volume from Dow Jones Stock Index using the business analytics concept. The reason why we chose volume to predict is because Trading volume is one of the most important factors influencing the stock market's behavior. The number of lots bought and sold in a daily basis is referred to as stock trading volume. The larger the trading volume of a stock, the more active and liquid the market is. When the volume of stocks traded are high, it means that the stocks are liquid and is more preferred by investors because the more liquid a stock is, the narrower its spread will be and there will be less risk of them being stuck with an undesired stock position. Which is why we think this study is important to do. In technical analysis, trading volume is more essential than stock price in approving price patterns. If we can forecast the future direction of a stock's trading volume, we can be more certain in predicting how prices will change, continue, or stop their trend (Garehchopogh, Bonab, & Khaze, 2013). And by using Business Analytics, we will be able to extract and visualize strong insights from an organization's data assets depends on the ability to detect meaningful patterns and recognize signals and trends within datasets (Jones, Cournane, Sheehy, & Hederman, 2016).

2. LITERATURE REVIEW

2.1 Stock Volume

Volume in stock market is a technical indicator that defined as the number of shares traded each day, each week, and many more and is an important indicator analysis as it used to measure the worth of stock price movement (Abbondate, 2010). The stock price-volume relationship provides insight into the structure of financial markets that can describe how information is disseminated in the markets. The price-volume relationship is of great significance for event studies that use a combination of stock returns and volume data to draw inferences. It is an integral part of the empirical distribution of speculative prices. And it can provide insight into future markets (Karpoff, 1987).

The other things that also involved in the stock price and trading volume in stock, they have different characteristics of non-stationarity which cause the requirement treatment in order to induce stationarity to be differences. A series of raw trading volume has some features of each linear and non-linear time trend with gradual changes in some statistical property of it (Gallant, 2015).

A number of measures of volume have been proposed by different studies on trading activity of financial markets (Andrew W. Lo & Jiang W. Wang, 2000) such as:

- Total number of shares traded as a measure of volume (A. Ronald Gallant, Peter E. Rossi, George Tauchen, 2015)
- Aggregate turnover as a measure of volume (Andrew W. Lo & Jiang W. Wang, 2000)
- Individual share volume as a measure of volume (Lamaoureux and Lastrapes, 1990).

In stock exchange or capital market activities, the volume of stock trading activity and the frequency of stock trading is one of the elements that has become one of the ways to see the market reaction to information that enters the capital market. Investors are doubtful when small trading volume are happening in the market. On the other hand, a large trading volume indicates that the stock is in demand by investors. The increase in trading volume is an increase in the buying and selling activity of investors in the stock exchange which has an influence on the level of offer of a spread (Khoirayanti & Sulistiyo, 2020).

Stock Volume is important for an investor because it is one of the main factors that affect stock prices in the market. The higher the volume of supply and demand for a stock, the greater its effect on stock price fluctuations on the Stock Exchange and increase the volume stock trading that is coveted by the public so that it will affect the increase in stock prices (Rahayu & Masud, 2019).

2.2 The Dow Jones Index

The fluctuations of stock prices in one country are influenced by the movements of stock prices in other countries. Economic changes in other parts of the world frequently have an impact on a country's economy. The Dow Jones Index is one of the three major indexes used by the New York Stock Exchange in Unites States. Economic growth in developed countries is fundamentally has a relationship with the economy of developing countries. The United States economy has a strong influence on the economies of other countries, especially developing countries.

Cha & Cheung's (1998) states that although most of the unexpected variations in stock returns in emerging Asian markets were due to domestic shocks, the impact of the US and Japan was greater in Hong Kong and Singapore than in Korea and Taiwan. The United States is one of the export destination countries for Indonesia. If its economic growth has increased, it can encourage Indonesia's economic growth through export operations, direct investment, or investment in the Indonesian capital market.

According to Mansur (2005), "the Indonesian capital market through the Jakarta Stock Exchange is an inseparable part of global stock exchange activities, because stock exchanges located close to each other often have the same investors". It would be interesting to know whether the Dow Jones index influences the stock price fluctuations of Indonesian stock exchange issuers, particularly in the agricultural industry.

Recent studies that discuss the relationship between the volume of foreign buy and the Dow Jones index on company stock prices. Mayzan & Sulasmiyati (2018) and Sudarsana (2014) states that the partial test results show that the Dow Jones Index has a positive influence on the Composite Stock Price Index, implying that an increase in the Dow Jones Index can raise the value of the Composite Stock Price Index. Besides that, Jayanti (2014) has a similar opinion, the Dow Jones index partially affects the Composite Stock Price Index (Elfahmi, 2020).

2.3 Business Analytics Concept

Business analytics is a platform or a place to use a technological data that collected and managed in different digital platforms such as databases and data warehouses, and transform it into actionable insights. One of the fastest growing fields is business analytics, which has emerged as a potential business enabler in public and private sectors. According to Watson (2009), Business analytics can be defined as "a board category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions" (Bayrak, 2015).

2.3.1 Descriptive Analytics

Descriptive analytics means that the number of the result is cause by the summarization the data in term of describing what occurred in the sample. Descriptive statistics can be use in comparing some of the sample. Descriptive Statistics also may help researchers to detect the characteristics of sample that can impact their result (Cheryl Bagley Thompson, 2009).

Descriptive statistics are used to summarize data in organized way along with describing the relationships between variables in a sample/ population. Descriptive Statistics include types of variables such as ratio, nominal, interval and ordinal. Descriptive analytics are an important part of the initial data analysis which provide a basis for comparing variables with inferential statistical tests. Therefore, in order to have good research, it is important to report the most appropriate descriptive statistics using misleading results (Kaur, Stoltzfus, & Yellapu, 2018).

Mean

Mean or average is the common indicator of central tendency which refers to the average value of a group of numbers. (Sykes, Gani, & Vally, 2016). There are several types of mean which is viz. arithmetic mean, weighted mean, geometric mean and harmonic mean (S., 2011).

$$\text{Mean (m)} = \frac{\text{sum of the terms}}{\text{numbers of terms}}$$

Median

Medan is the number that count from the dividing the distribution into a part which half of the number are above the median and half are below it when the data are arranged in numerical method. Median is the central value that can be used in extremely high or low value situations (Sykes, Gani, & Vally, 2016).

$$\text{If n is even:} \quad \text{Median} = X\left\lfloor\frac{n}{2}\right\rfloor$$

$$\text{If n is odd:} \quad \text{Median} = \frac{X\left\lfloor\frac{n-1}{2}\right\rfloor + X\left\lfloor\frac{n+1}{2}\right\rfloor}{2}$$

Description:

X = ordered list of values in data set

n = number of values in data set

Standard Deviation

When we use descriptive statistics, we need to use measures of dispersion, one of it is by using standard deviation. Standard deviation is a measure of dispersion when describing the sample whether we are describing the sample in numbers and words or in a figure. The use of standard deviation is to tell us how far from the mean and the average number (Andrade & Chittaranjan, 2020).

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma\,(\chi i - \mu)^2}{N}}$$

Description:

σ = population standard deviation

N = the size of population

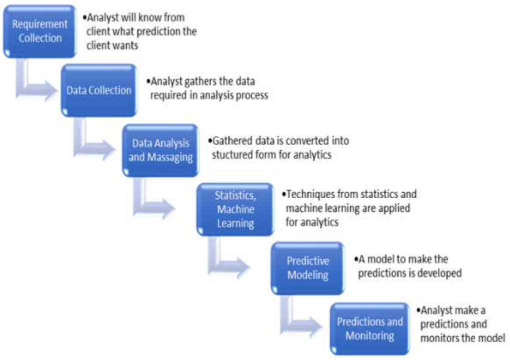
χ_I = each value from the poluation μ = the population mean

2.3.2 Predictive Analytics

Predictive analytics is a subset of advanced analytics and used in predicting the future events. Predictive analytics analyzes current and historical data to generate future predictions using the techniques from statistics, data mining, machine learning, and artificial intelligence. All predictive analytics models are divided into two categories: classification models and regression models. Classification models predict whether or not variables belong to a specific class, whereas regression models predict a number.

Predictive analytics involves several steps procedures that allow the data analyst to predict the future based on current and historical data (Kumar & Garg, Predictive Analytics: A Review of Trends and , 2018).

Figure 8 Predictive Analytics Process



Random Forest

Random forest is a relatively new Ensemble Supervised Machine Learning approach. Machine learning techniques have applications in data mining. Descriptive and predictive data mining are the two primary categories of data mining. Descriptive data mining is more concerned with describing the data, categorizing it, and summarizing it. Predictive data mining examines previous data to identify trends or draw inferences for future predictions. Predictive data mining has its roots in the traditional statistical model building procedure. Predictive model construction is based on feature analysis of predictor variables.

Random Forest generates a group of decision trees. Breiman used the randomization approach, which works well with bagging or random subspace approaches, to achieve variation among base decision trees. Research work in the area of Random Forest aims at either improving accuracy or improving performance (reducing time required for learning and classification), or both. Some work aims to experiment with Random Forest by utilizing online continuous stream data, which is important nowadays because data streams are generated by a variety of applications. Because Random Forest is an ensemble technique, experiments are carried out with its basis classifier, such as Fuzzy Decision Tree (Kulkarni & Sinha, 2013).

Deep Learning

Deep learning, a class of machine learning algorithms inspired by the biological process of neural networks, dominating in many applications and outperforming traditional machine learning algorithms. Deep learning becomes more useful when the amount of training data is increased. Deep learning models use numerous layers, which are made up of a variety of linear and non-linear transformations. With the expansion in data quantity, or breakthroughs in the field of big data, traditional machine learning algorithms have demonstrated their limitations in data analysis (Kumar & Garg, Deep Learning as a Frontier of Machine Learning: A , 2018).

2.3.3 Prescriptive Analytics

Prescriptive analytics is a type of data analytics in which the actions that must be taken in order to achieve a specific goal are determined. This is a relatively new feature of analytics that allows users to "prescribe" a variety of possible actions to help them find a solution. It advises on possible outcomes and answers using optimization and simulation algorithms. As a result, prescriptive analytics is all about making (Brodre, D.A. & Mahagaonkar S, 2019).

Prescriptive analytics can recommend the best decision options for taking advantage of the predicted future, as well as show the consequences of each decision. This is an effort to determine the impact of future decisions so that possible outcomes can be predicted before the decision is made. It goes beyond descriptive and predictive analytics to suggest one or more potential causes of action. The effectiveness of the suggestions is determined by how well these models incorporate a mix of structured and unstructured data, represent the domain under investigation, and capture the consequences of the decisions under consideration (Alexandros, B, 2018).

Prescriptive analytics focuses not only on the Why, How, When, and What, but also on how to act in order to take advantage of the situation. Prescriptive analytics is frequently used to assess a company's analytics maturity. Prescriptive analytics, according to IBM, is "the final phase" of business analytics (Rijmenam, 2013). Prescriptive analytics evaluates and determines new ways to operate, with the goal of achieving business objectives while balancing all constraints.

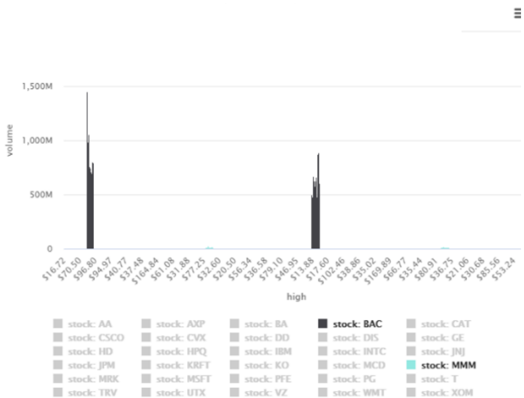
3. DATA AND METHODOLOGY

The dataset we chose was retrieved from The UCI Machine Learning Repository, which is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. This dataset contains weekly data for the Dow Jones Industrial Index and has been used in computational investing research. By Dr. Michael Brown from University of Maryland University College.

Based on the data provided in Appendix A (see endnote), there are 16 attributes in his dataset, and since we are predicting Stock Volume, the attributes volume will be the label and the rest of the independent attributes will be used to help us predict the stock volume. Using the Business Analytics concept, we will provide three analyses, which are Descriptive Analysis, Predictive Analysis, and Prescriptive Analysis.

3.1 Descriptive Analytics

Figure 9 Descriptive Analytics



The statistic that we examine is the stock volume, as we can see above, we take BAC (black candles) and MMM (blue color) stock for us to examine with BAC as the highest volume and MMM as the lowest volume from all of the stock.

The highest volume of BAC stock is 1.453.438.639 when the price reach $14.69, the lowest volume of BAC stock is 479.389.107 when the price reach $11.69. whereas for the highest volume of MMM stock is 26.110.543 then the price reach $90.50 and the lowest volume of MMM stock is 10.705.558 when the price reach $89.39. The smallest or minimal volume is 9.718.851, whereas the highest or maximal volume is 1.453.438.639. The average is 117.387.644,835 and the standard deviation is 159.232.227,996.

3.2 Predictive and Prescriptive Analytics Process

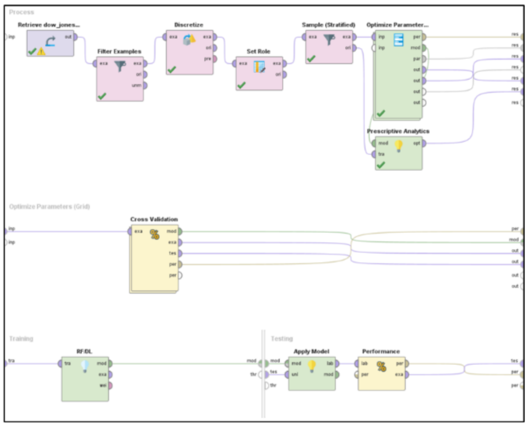


Figure 10 Business Analytics Sequences Using Rapid Miner

3.2.1 Predictive Analytics

In predictive analysis, the first step to do is **preparing the data**, by filtering out the attributes with missing value. To do so, the operator we use is Filter Example, and choose the attributes with missing values and filter them out. After the data is cleaned, use the Discretize by User Specification operator to **discretize the label** based on user's needs. The reason in doing so is because the label we chose, volume, has a numerical value but in predicting the accuracy of this attribute, nominal value for the attributes is needed.

After some discussion was made with a stock expert, we decided to categorize the attributes into three entry which are LOW with the Upper Limit of 245.000.000, MID with the Upper Limit of 398.000.000, and HIGH with Infinity as the Upper Limit.

Upon doing so, the **Set Role** operator is the placed in the model with the purpose of setting the chosen attributes volume as the label. Furthermore, to **reduce the dimension** of the overall dataset, we use only use 0.6 sample of the dataset.

Use **Cross Validation** to divide the data that will be trained and one that will be tested, in this paper, we train the data using two algorithms which are the Random Forest algorithm and Deep Learning algorithms to compare which algorithm will exhibit better results. To test the data, **Apply Model** and **Performance** operators are used. All the operators in to test and train the data and to choose the best data is processed inside the **Optimize Parameters** operator where we can choose which parameters we would like to optimize.

3.2.2 Prescriptive Analytics

In prescriptive analytics, these operators reverse that procedure, starting with a model and a desired output, and prescribing an optimized input to achieve the desired outcome. We implement the **Prescriptive Analytics** operator. This process is conducted with a relative sample, with a **ratio** of **0.6** by using the **Operator Sample (Stratified).**

In addition, Discretize by User Specification operator we use **discretize the label** based on user's needs. Due to the label that we are using is volume which mean has a numerical value in predicting the accuracy of this attribute, nominal value for the attributes is needed.

4. RESULT AND DISCUSSION

4.1 Descriptive Analytics

The descriptive statistics table can be seen Appendix B (see endnote). We can see that the average or the mean of volume is 115603991.883. After the calculation, we can conclude that volume below average is more than volume above average. For the median itself we can calculate it by taking the maximum amount divided by two which is 527207687.5. For the standard deviation we can see from the table above which is 152084379.271. We can see from the statistics that the frequency at most is 580 which is the mode with the volume among 9.718.851 until 154.090.829.8. We will also compare the **percent_change_price** and **percent_change_next_weeks_price**. The difference is that percent_change_price is the price changes for this one week, whereas percent_change_next_week_price is the estimation of the price changes for next week. The minimum and the maximum are the same in both places, whereas the average is quite different which is 0.030 and 0.193, and the standard deviation is close enough which is 2.505 and 1.664. and for the visualization itself as a whole is quite same. We can conclude that the price for this week and next week not so different and there are only slight changes.

4.2 Predictive Analytics

a. Random Forest

Random forest is a supervised learning strategy for training data that combines a decision tree and a bagging method. The following parameters were optimized in this study: number_of_trees, criterion, apply_pruning, and accuracy. The result (Figure 9) shows that the model accuracy was 95.85%. From 432 of the samples that being tested, 414 predictions are precise, and the rest was predictions errors. In prediction of HIGH volume traded, the model made showed 95.24% accuracy. This means that this model had successfully predict the HIGH volume we aim for in a great way.

Table 18 Confusion Matrix of Random Forest

	true LOW	true MID	true HIGH	class precision
pred. LOW	365	9	1	97.33%
pred. MID	3	29	4	80.56%
pred. HIGH	1	0	20	95.24%
Class recall	98.92%	76.32%	80.00%	

Accuracy: 95.85%

b. Deep Learning

We also use Deep Learning Algorithm to compare which one is better to predict stock volume. In this case, the parameters we chose to optimize was local_random_seed, L1, and accuracy. As can be seen on the result (Figure 10), the overall accuracy of this Deep Learning model was 94.92% which was a little below the Random Forest prediction Model. From 432 of the samples that being tested, 410 predictions are precise, and the rest was predictions errors. In prediction of HIGH volume traded, the model made showed 84.00% accuracy. This means that this model shows less accuracy in predicting the HIGH stock volume we aim for. The result itself is not bad but Random Forest shows to be a potential model to predict HIGH stock volume.

Table 19 Confusion Matrix of Deep Learning

	true LOW	true MID	true HIGH	class precision
pred. LOW	374	9	1	97.40%
pred. MID	5	15	3	65.22%
pred. HIGH	2	2	21	84.00%
Class recall	98.16%	57.69%	84.00%	

Accuracy: 94.92%

c. Algorithm Comparison

To summarize the performance of the two algorithm we compared, below provided the result in Figure 11. We can see that Random Forest algorithm has better or higher accuracy. This result shows prove that random forest performs better than the neural network model when used to predict stock index movement (Depari, 2020).

Table 20 Algorithm Comparison

Algorithm	Accuracy	Precision (pred. HIGH)	Recall (true HIGH)
Random Forest	95.85%	95.24%	80.00%
Deep Learning	94.92%	84.00%	84.00%

4.3 Prescriptive Analytics

Table 21 Prescriptive Analytics Result

date	Prediction (volume)	stock	quarter	confidence		
				low	mid	high
Jan 14, 2011	high	BAC	2	0.405	0.164	0.432

Changes percentage			Previous week volume	Days to next dividend	Percent return for next dividend
In price	In price for next week	In volume over last week			
-0.199	5.301	53.104	424378544	47	1.285

close	open	Next week close	Next week open	high	low
$27.49	$19.68	$83.24	$14.26	$15.50	$40.22

This is the result of prescriptive analytics by using the process that shown in the chapter 3 (figure 3) As we stated above that the lamber that we are using is volume and the stock type that we analyze is **BAC**. As we can see from the result stated above, it shows that the prediction (volume) is **HIGH** with the confidence level of high has the highest value compared to the low and mid which is **0.432.** Meanwhile, the percent change price -0.199 and the percent change price for next week will be surplus 5.301 with the open price $19.68 and close price $27.49. It is predicted that next week open price will be $14.26. The percent return to next dividend is 1.285 and 47 days to go for the next dividend.

5. CONCLUSION

This study is written to predict stock volume from Dow Jones Index by using business analytics concept. As stated previously, a higher trade volume shows that a stock or commodity has a stronger overall market interest. Stocks with higher volume are exchanged more frequently and quickly than those with lesser volume. As a result, a large transaction volume usually indicates a high amount of liquidity for a particular security or commodity in the market. The more liquid a stock is, the narrower its spread will be. Because market makers would be able to buy and sell quickly, there will be less risk of them being stuck with an undesired stock position. Which is why we choose to do the prediction of volume, since it is a great indicator to a good market. In details, we are predicting the volume of the stock for one week from the Dow Jones Index and compare it to other week to know the fluctuations of the stock volume with the assistance of a machine learning software Rapid Miner.

There were 15 input variables and one label variable. To reduce the dimensionality, we used a 60% sample of the data set. Random Forest and Deep Learning Algorithm were used to test and compare the performance of the two algorithms, and with a 95.85% accuracy, Random Forest was chosen as the most accurate algorithm. This research contributes by stressing a clear, succinct, and robust analytical framework for assisting stock traders and researchers in predicting stock's volume. It also makes a managerial contribution by assisting a trader or investors in determining which index to invest on. Even so, the weighting results showed only minor variances in value among the 16 criteria we used. More daily data could provide a more comprehensive and more complete picture. As a result, using a larger and more compound dataset in future studies may disclose other important, and hidden information in Stock.

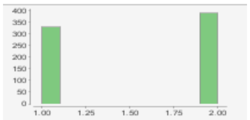
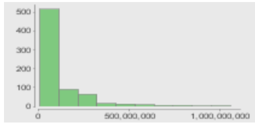
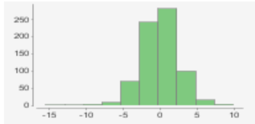
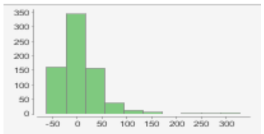
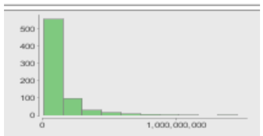
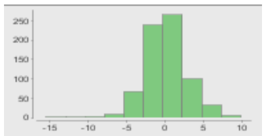
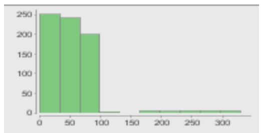
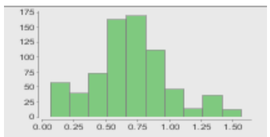
APPENDIX A

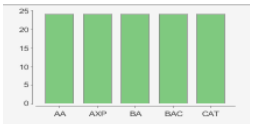
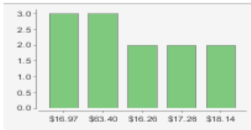
Table 22 Data Description

No	Attributes	Description	Roles
1	quarter	the yearly quarter (1 = Jan-Mar; 2 = Apr=Jun)	Independent Variable
2	stock	the stock symbol (see above)	Independent Variable
3	date	the last business day of the work (this is typically a Friday)	Independent Variable
4	open	the price of the stock at the beginning of the week	Independent Variable
5	high	the highest price of the stock during the week	Independent Variable
6	low	the lowest price of the stock during the week	Independent Variable
7	close	the price of the stock at the end of the week	Independent Variable
8	volume	the number of shares of stock that traded hands in the week	Dependent Variable (Label)
9	percent_change_price	the percentage change in price throughout the week	Independent Variable
10	percent_chagne_volume_over_last_wek	The percentage change in the number of shares of stock that traded hands for this week compared to the previous week	Independent Variable
11	previous_weeks_volume	the number of shares of stock that traded hands in the previous week	Independent Variable
12	next_weeks_open	the opening price of the stock in the following week	Independent Variable
13	next_weeks_close	the closing price of the stock in the following week	Independent Variable
14	percent_change_next_weeks_price	the percentage change in price of the stock in the following week	Independent Variable
15	following week days_to_next_dividend	the percentage change in price of the stock in the following week	Independent Variable
16	percent_return_next_dividend	the percentage of return on the next dividend	Independent Variable

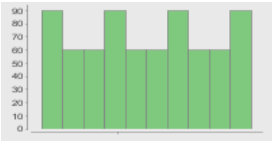
APPENDIX B

Table 23 Descriptive Statistics Table

Name	Type	Min	Max	Average	Standard Deviation	Visualization
quarter	Integer	1	2	1.542	0.499	
volume	Integer	9718851	1054415375	115603991.883	152084379.271	
percent_change_price	Real	-15.423	9.882	0.030	2.505	
percent_change_volume_over_last_wk	Real	-61.433	327.409	5.593	40.543	
previous_week_volume	Integer	9718851	1453438639	117387644.835	159232227.996	
percent_change_next_weeks_price	Real	-15.423	9.882	0.193	2.664	
day_to_next_dividend	Integer	0	329	52.260	45.882	
percent_return_next_dividend	Real	0.066	1.564	0.692	0.305	

Name	Type	Least	Most	Values	Visualization
stock	Polynominal	XOM (24)	AA (24)	AA (24), AXP (24), BA (24), BAC (24), [26 more]	
open	Polynominal	$94.38 (0)	$37.26 (3)	$37.26 (3), $44.75 (3), $19.68 (2), $20.32 (2), [718 more]	
high	Polynominal	$94.81 (0)	$12.60 (2)	$12.60 (2), $17.24 (2), $17.64 (2), $19.39 (2), [709 more]	
low	Polynominal	$92.30 (0)	$16.97 (3)	$16.97 (3), $63.40 (3), $16.26 (2), $17,28 (2), [707 more]	
close	Polynominal	$93.73 (0)	$33.07 (3)	$33.07 (3), $36.00 (3), $41.52 (3), $46.25 (3), [707 more]	
next_week_ open	Polynominal	$93.21 (0)	$44.75 (3)	$44.75 (3), $14.94 (2), $19.68 (2), $20.32 (2), [716 more]	
next_week_ close	Polynominal	$94.01 (0)	$33.07 (3)	$33.07 (9), $36.00 (3), $41.52 (3), $12.31 (2), [711 more]	

Name	Type	Earliest Date	Latest Date	Duration	Visualization

date	Date	Jan 14, 2011	Jun 24, 2011	161 days	

ACKNOWLEDGEMENT

Since the beginning until the result of this research process, the authors of this paper have received constructive advice from several parties which are very well appreciated. The author's lecturer, Mr Genesis Sembiring, deserves a special thanks for his generosity and efforts in guiding this research process. Furthermore, the authors express sincere gratitude to the academic reviewers of the ***5th" National Conference on Business, Management, and Accounting (NCBMA)*** for providing the authors a chance to participate in this event.

REFERENCES

A. Ronald Gallant, Peter E. Rossi, George Tauchen. (2015). Stock Prices and Volume. *Trading Volume and Stock Return*.

Abbondate, P. (2010). Trading Volume and Stock Indices: A Test of Technical Analysis. *Trading Volume and Stock Return: Empirical Evidence for Asian Tiger*.

Alexandros, B. (2018). *Prescriptive Analytics: Survey of Approaches Methods. Athens.*

Amidu, A.-R. (2020). Research in Valuation Decision Making Processes: Educational Insights and Perspectives. *Journal of Real Estate Practice and Education*, 19.

Andrade, & Chittaranjan. (2020). Understanding the Difference Between Standard Deviation and Standard Error of the Mean, and Knowing When to Use Which. *Sage Journal*.

Andrew W. Lo & Jiang W. Wang. (2000). Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. *Trading Volume and Stock Return*.

Bayrak, T. (2015). A Review of Business Analytics: A Business Enabler or Another Passing Fad. *Social and Behavioral Sciences 195*, 230-231.

Brodre, D.A. & Mahagaonkar S. (2019). Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Alogrithms. *Journal of Engineering Applied Sciences and Technology*, 371-376.

Cheryl Bagley Thompson, P. R. (2009). Basics of Ressearch. *Descriptive Data Analysis*, 59.

Denning, P. J. (2019). Computational Thinking. In M. Tedre, *An introduction to computational thinking that traces a genealogy beginning centuries before the digital computer* (p. Preface xiii). Monterey: MIT Press.

Depari, G. S. (2020). Forecasting Bitcoin Return: A Data Mining Approach. *Review of Integrative Business and Economics Research, Vol. 10, Issue 1*, 63.

Elfahmi, R. (2020). The Effect of Foreign Buy and the Dow Jones Index on Stocks . *Journal of Research in Business, Economics, and Education; Volume 2; Issue 6*, 1443.

Gallant, A. R. (2015). Stock Price and Volume. *Trading Volume and Stock Return*.

Garehchopogh, F. S., Bonab, T. H., & Khaze, S. R. (2013). A Linear Regression Aprroach to Prediction of Stock Market Trading Volume: A Case Study. *International Journal of Managing Value and Supply Chains (IJMVSC) Vol.4, No. 3, September 2013*, 25.

Jones, S., Cournane, S., Sheehy, N., & Hederman, L. (2016). A Business Analytics Software Tool for Monitoring and Predicting Radiology Throoughput Performace.

Karpoff, J. M. (1987). The Relations between Price Changes and Trading Volume: A Survey. *Trading Volme and Stock Return*.

Kaur, P., Stoltzfus, J., & Yellapu, V. (2018). Biostatistics. *Descriptive statistics*, 4.

Khoirayanti, R. N., & Sulistiyo, H. (2020). Pengaruh Harga Saham, Volume Perdagagan, dan Frekuensi Perdagangan terhadap Bid-Ask Spread. *Jurnal Ilmiah Akuntansi Fakultas Ekonomi*, 234.

Kulkarni, V. Y., & Sinha, D. P. (2013). Random Forest Classifiers :A Survey and Future. *International Journal of Advanced Computing, ISSN:2051-0845, Vol.36, Issue.1*, 1144 - 1146.

Kumar, V., & Garg, M. L. (2018). Deep Learning as a Frontier of Machine Learning: A . *International Journal of Computer Applications (0975 – 8887) Volume 182 – No.1*, 22.

Kumar, V., & Garg, M. L. (2018). Predictive Analytics: A Review of Trends and . *International Journal of Computer Applications (0975 – 8887) Volume 182 – No.1*, 31-33.

Lamaoureux and Lastrapes. (1990). Heteroscedasticity in Stock Return Data. *Trading Volume and Stock Return*.

Rahayu, T. N., & Masud, M. (2019). Pengaruh Tingkat Suku Bunga, Nilai Tukar Rupiah dan Volume Perdagangan Saham Terhadap Harga Saham Perusahaan Manufaktur. *Jurnal Ilmu Ekonomi*, 38.

S., M. (2011). Measures of central tendency: The mean. *PMC PubMed Central*.

Shen, S. (n.d.). Stock Market Forecasting Using achine Learning Algorithms. 1.

Sykes, L. M., Gani, F., & Vally, Z. (2016). The meaning of the MEAN, and other statistical terms commonly used in medical research. *SciELO* .

Wasiat Khan, M. A. (2020). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1.