Ashish Shrestha

APPLICATION OF MACHINE LEARNING ALGORITHM TO MEASURE A FIRM'S PEFORMANCE

School of Technology and Innovations
Master's thesis in Industrial Systems Analytics

Vaasa 2021

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Ashish Shrestha
Title of the Thesis: Application of machine learning algorithm to measure a firm's performance.
Degree: Master of Science in Technology
Programme: Industrial Systems Analytics
Supervisor: Ahm Shamsuzzoha
Year: 2020 **Pages:** 117

ABSTRACT :

Machine learning techniques are an emerging field in today's world. The objective of this thesis was to use machine learning methodology to measure a company's performance by using forecasting techniques in financial statements. This information can be useful for investors, managers, and analysts. The financial statement data collected were from 250 companies from the United States of America. The methodology that was applied was Long Short-Term Memory. The forecasting method used was time-series forecasting. The software used for running the code was Juypter. The conclusion of the study shows that machine learning algorithms can be applied for forecasting firm performance. The program shows the results for the future prediction of the performance of companies.

KEYWORDS: Machine learning, forecasting, Long short-term memory

TABLE OF CONTENTS

List of Figures

ACKNOWLEDGEMENTS

This Master's thesis was undertaken to test the application of machine learning algorithm for the analysis of financial statements. I would like to thank Mr. Rasheed Alabi for guiding me with this work. I would also like to thank to the thesis supervisor, Assistant Professor, Ahm Shamsuzzoha for helping me to complete this thesis process. I am grateful for the support that my wife Nabina Shrestha has given to me throughout this period.

1. INTRODUCTION

1.1 Background and purpose of the study

The yearly publication of annual reports of companies are an important source of information for investors and analysts (Qiu, 2007). An annual report shows a company's performance for the past and present year (Qiu, 2007). It includes the reason for price and sales changes, revenue and cost changes, planned expenses and future possibilities (Qiu, 2007). These measures, with the help of mathematical financial indicators can predict the future trends of the firm (Qiu, 2007). Based on these trends and values, the investors and analysts can make decisions for a profitable investment (Qiu, 2007). A firm's financial data are available to the general public through financial statements. The financial statements include balance sheets, income statements, and cash flow statements (Qiu, 2007).

Financial analysis is performed by skilled personnel (Alicia, 2019). It refers to assessment of the viability, stability and profitability of a business (Alicia, 2019). There are two types of financial analysis. Fundamental analysis and technical analysis. Fundamental analysis uses data from the financial statements. It uses different ratios such as earnings per Share, return On Equity for calculation of business value. Similarly, technical analysis, uses trends from the moving averages. The trends are calculated from the statistical calculations. (Alicia, 2019).

Matt (2018) stated, *"Financial forecasting is an estimate of future financial outcomes for a project. The important role in financial forecasting is predicting the revenue*". The advantages of financial forecasting for business are that it provides better control over cash flow. It helps to distribute money in different sectors of a company. Likewise, it creates a benchmark which can be used to match performance, identify loopholes, and perform necessary action. Similarly, it helps to identify financial risk. Therefore, identification of financial risks helps to control the future risks. The financial forecasting method can also

predict the future cash requirements. It can project the future expenses which is necessary to make financial decisions. Likewise, it can help to get better idea about the projected expenses. This will be able to determine the money needed for business. Lastly, it is necessary tool for investing (Matt, 2018).

According to Demirbag, et al. (2006), "*Performance measurement is an important area of effective management for any firm*". The evolution of machine learning techniques and algorithms have made the measurement of the firm's performances more accurate, reliable and fast (Miyakawa et al. 2017). The need for using machine learning for measurement of the firm's performance is in providing fast and reliable information for profitable use (Miyakawa, et al. 2017).

The need for financial forecasting is that it provides better flow over cash flow. In a company, cash is an important sector to manage. Likewise, it shows the financial viability of new ventures. It acts as a benchmark for indicators such as identifying loopholes, and taking necessary actions (Matt, 2018).

Machine learning is the method of study using computer algorithms. It is seen as a subset of artificial intelligence. It is related to the technologies for making computing devices be able to learn from input data and make reasonable output. The application of machine learning is very large and includes a number of different areas including the banking sector, engineering, etc. It works with statistical data. The thesis aims to combine Machine Learning (ML) algorithms and historic data to achieve profitable result regarding measurement firm performances (Alpaydin, 2020).

1.2 Research questions of the study

The research question is based on the financial indicators, choice of prediction, evaluation criteria and experimental design. These foundations have made the research questions presented later in this section. The research study has four objectives, as follows:

a. To design a machine learning model which will be able to forecast a company's future return on equity (ROE), diluted and basic earnings per share (EPS), revenue and return on assets (ROA).

b. To use the machine learning Long Short-Term Memory (LSTM) model to predict a 5 (2020-2025) year forecast of the company's return on equity (ROE), earnings per share (EPS), revenue and return on assets (ROA). The measures are based on ROE (Return of Equity) and EPS (Equity of Share).

Mathematically,

$$\text{ROE} = \frac{Net\ income}{Average\ shareholder's equity} \quad (1)$$

$$\text{EPS} = \frac{Profit - Preferred\ dividends}{Weighted\ average\ common\ shares} \quad (2)$$

c. Classification of the firm based on the return on equity and earnings per share. We use t to refer to the year corresponding to the annual report and t +1 to the preceding year (Qiu, Padmini, Nick, 2007).

I. Better performance: If the ROE ratio in year t+1 is greater than the ROE in year t by more than 5% then the company has a better performance (Qiu, Padmini, Nick, 2007).

II. Neutral performance: If the ratio in year t+1 is within 5% of the ROE ratio in year t, then the company is classified as neutral in performance (Qiu, Padmini, Nick, 2007).

III. Worst performance: If the ROE ratio in year t+1 is less than the ROE ratio in year t by more than 5% has its worst performance (Qiu, Padmini, Nick, 2007).

d. Evaluate the model's strengths and weaknesses. The evaluation or analysis will be based on quantitative results.

The objectives of the thesis are based on the firm performance, financial indicators, machine learning algorithm and forecasting methods. The above four objectives have produced the following two research question.

- RQ 1: How does financial statement forecasting support companies in measuring performance?

- RQ 2: How can machine learning algorithm be applied successfully for forecasting financial statements in companies?

1.3 Structure of the thesis

The thesis is divided into three sections: theoretical study, empirical study and conclusion. The theoretical section consists of an introduction, a literature review on machine learning approaches to financial statements and time series forecasting methods. The empirical study consists of research methods and data collection. The final section the conclusion, consists of the results, recommendations, and analysis from the empirical study.

2 LITERATURE REVIEW

In this section, the focus is on research work related to building predictive models, firm performance, and forecasting.

Predictive analysis is performance in many ways. Machine learning is a new evolving sector in artificial intelligence (Charles, 2007) and the utilization of machine learning technique predictive analysis is possible. Its application to finance and banking sectors has shown improvements in quality standards (Charles, 2007). Machine learning techniques are widely used in various interdisciplinary contexts. Techniques such as regression, classification, clustering, etc. have shown improvements in forecasting for stock prices, crude oil, etc.

2.1 Predictive models with company annual reports

The study by Qiu (2007) applied machine learning technique to company annual reports (Qui, 2007). The study presented method for measuring firm performances based on return on equity (ROE) and earnings per share (EPS). The result of the research is assessment of the predictive potential of the company's annual reports. The research focuses on financial performance indicators like operating earnings, net income and current ratio and stock returns.

In this case, scalar vector multiplication method (SVM), ROE and EPS. Second, it uses 10K or company annual reports for ML application. Third, the research problems of the thesis are based on five dimensions. These are, financial performance indicators, choice of predictions, evaluation criteria, document representation and experimental studies. These three core measures match with the study objective of this thesis.

The core foundation of the dissertation is based on the work of three different researchers. One is Kohut and Segars (1992), whose research study implemented return on equity. The research includes studies of annual reports to measure firms´ performance using return on equity (ROE). The second work is from Zhang et al. (2004), whose research was based on comparative neural network models and a variety of linear statistical models in forecasting. The final work used is Kloptchencko et al. (2002), who implemented 7 ratios to characterize firm performance. These three-profitability ratios are liquidity ratio, two solvency ratio and one efficiency ratio.

The present study includes data collection of annual reports from 250 companies from different financial sectors. The data collected was from the Edgar database's SEC filings. The companies were selected from SIC codes. The industrial areas that the reports were taken from are food and associated products, tobacco products, textile mill products, apparel and other finished products, lumber and wood products, and furniture and fixtures.

The dissertation assesses the textual contents of the annual reports by building predictive models. The result is obtained by testing three Scalar Vector Multiple models (SVM-score, SVM-prob and SVM-multi) using different evaluation benchmarks: predictive accuracy, cost of errors, comparison with majority vote baseline and analysts` forecasts, portfolio return and robustness with different class definitions (Qiu, 2007).

A weakness of the dissertation is that it did not evaluate the firm performances using machine learning algorithm. The measures that were applied for the results were only based on two dimensions, which were ROE and EPS. The model that it used was scalar vector multiples. The results could only be used for analysts and investors for investment purposes. Likewise, the study was concerned with research on applications which were already implemented in several research studies.

2.2 Determinants of financial performance indicators

The performance of firm is a dependent variable (Selvam et al., 2016), and. According to Selvam et al, (2016), *"These are profitability performance, growth performance, market value performance of firm, customer satisfaction, employee satisfaction, environmental audit performance, corporate governance performance and social performance.*" The subjective model was developed with nine determinants/dimensions (Selvam et al., 2016).

In the research study of Selvam et al. (2016), "*Determinants of financial performance: A meta-analysis",* different meta-analysis of the results from 320 published studies are discussed. The study indicates that, the firm performance is measured based on profitability performance, market value performance and employee satisfaction. These performance measures are taken from table 1. The measurements are based according to Selvam (2016) firm performance table. Selvam (2016) stated *"the comprehensive model is constructed based on firm performance."* The figure1 below shows the list of identified determinants for firm performance.

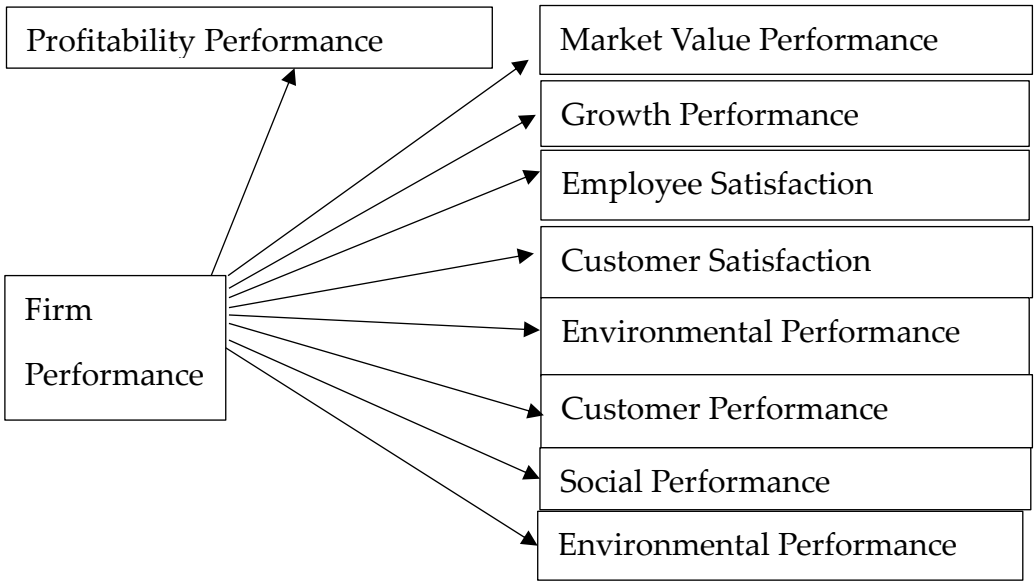


Figure 1. List of determinants of firm performance (Selvam, 2016)

2.2.1 Profitability performance

This is the method to earn profit from a business. Profit is produced after the deduction of revenue and expenses which are related to the operating business activities. Profitability measures a firm's past ability to generate (Selvam et al., 2016).

2.2.2 Market value performance

Value performance is related to prices in the market. It represents the external assessment and future performance of the firm. It is the ability to predict stock trends, based on publicly disclosed information. The information is relevant to stock returns which is good for investors and stakeholders. Diversification strategy provides risk minimization and return maximization (Selvam et al., 2016).

2.2.3 Growth performance

Growth performance refers to positive change in size. The stock indices apart from being an indicator serve as a benchmark measuring the performance of stock. Maximization of stakeholder and investors are revealed on the stock market by the indices of financial reports and other required information. These refers to positive change in size and/or maturation over a period (Selvam et al., 2016).

2.2.4 Employee satisfaction

Employees are a trained and skilled personnel group who have clearly defined work descriptions. Their jobs in an organization are defined by roles and responsibilities, the work environment, and their experiences with management. Therefore, an organization

must have human resources management to produce better outcomes (Selvam et al. 2016).

2.2.5 Customer performance

Customer performance is related to customer satisfaction. It includes the needs to be evaluated from the demands of customers. The focus of business improvement must always be the customer. The expectations of the customers must always be fulfilled by the companies (Selvam et al., 2016).

2.2.6 Social performance

According to Selvam, 2016, *social performance can be defined as "the effective translation of an institution's mission into practice in line with what is accepted as social value"*. It is a way to satisfy communities. Social performance is about making an organization's social mission a reality (Selvam et al., 2016).

2.2.7 Environmental performance

Vasanth, et al., (2015a) stated that, *"It is essential that when the company earns more profit from the operation of the business, it should spend a portion of amount towards environmental protection."* The world's environment is degrading continuously. The deterioration of the quality of the environment is a major issue. The growing number of industries and population is the reason for the pollution, which is the main factor involved in the performance of the environment (Selvam, et. al. 2016).

2.2.8 Dimensions of firm performance

The dimensions of the firm performances are the dimensions and sample indicators for firm performances. The dimensions for firm performances cover three variables: - profitability performance, growth performance, and market value performance. The strategic performance includes six dimensions, namely employee satisfaction, customer satisfaction, environmental performance, environmental audit performance, corporate governance performance and social performance (Selvam, et. al. 2016).

Table 1 shows the nine dimensions and list of indicators for each dimension for firm performance. The list shows 46 indicators revealing firm performance (Selvam et al., 2016) (Santos & Brito 2012).

S. N.	Dimension	Sample indicators	Number of sample indicators
1	Profitability performance	1. Return on assets 2. EBTIDA margin 3. Return on investment 4. Net income/revenues 5. Return on equity 6. Economic value added	6
2	Market value performance	1. Earnings per share 2. Changes in stock price 3. Dividend yield 4. Stock price volatility 5. Market value added	6

		6. Tobin's Q (market value/replacement value assets)	
3	Growth performance	1. Market-share growth 2. Asset growth 3. Net revenue growth 4. Net income growth 5. Amount of employees growth	5
4	Employee satisfaction	1. Turnover 2. Investments in employee's development and training 3. Wages and rewards policies 4. Career plans organizational climate 5. General employees' satisfaction	6
5	Customer satisfaction	1. Mix of products and services 2. Number of complaints 3. Repurchase rate 4. New customer retention 5. General customer satisfaction 6. Number of new products services launched	6

6	Customer satisfaction	1. Mix of products and services, 2. Number of complaints, repurchase rate, 3. New customer retention, general customer satisfaction, 4. Number of new products/services launched	6
7	Environmental Performance	1. Environmental policy environmental 2. Audit report 3. Environmental review	3
8	Corporate governance performance	1. Board size 2. Board independence 3. Outside directors 4. Insider ownership	4
9	Social performance	1. Employment of minorities, 2. Number of social and cultural projects, 3. Number of lawsuits filed by employees, 4. Customers and regulatory agencies	4

Table 1. The nine dimensions and list of indicators for each dimension.

3 THEORETICAL FRAMEWORK

In this section, the measures related to forecasting, financial statements, balance sheet, cash flow and industrial classification methods will be discussed. The measures are related to the forecasting for machine learning methods and the algorithms for machine learning.

3.1 Forecasting

According to Scott (2001), a forecast is the predictability of an event. It includes the following three questions.

I. How well are the factors understood?
II. How much data is available?
III. Can the forecasts affect the thing we are trying to forecast?

There are two factors which makes forecasting results. They are plans and goals. Planning a forecast is related to the objective and the goal of a project. The desired result is the true outcome (Rob & George, 2018). The future demand of organizations depends upon three aspects. These are short, medium- and long-term forecasts. Short-term forecasts are for lesser duration, medium forecasts for future resource requirements, and long-term forecasts for strategic planning (Rob & George, 2018).

The forecasting methods largely depends on available data, which can be either quantitative or qualitative (Rob & George, 2018). The qualitative data methods are guesswork. Quantitative data methods are applied when there are historic data available, and a reasonable past pattern will continue in the future (Rob & George, 2018).

Forecasting is an important financial indication to predict future trends for an institution or organization. It is based on historical data. The data are for example, sales, revenues, etc. The historical data are then implemented to mathematical indicators like simple moving average, volume weighted average etc. The results are the possible trend patterns which show future profitable possibilities. An organization invests capital-based forecasts for new products, factories, retail outlets and contracts with executives. This is in the hope of profitable earnings (Scott, 2001).

3.2 Financial statements

According to, Petrit (2019), *"financial statements are a structured financial presentation and transactions undertaken in an organization*". The objective of financial statements is to provide information on the current position and financial changes. It is a very important basis for making managerial decisions (Asllanaj, 2008 & Petrit, 2019).

The objective of financial statements is to provide information about the financial situation, financial performance and changes in an entity's financial position which is usable by a wide range of users in making their economic decisions (Lewis, & Pendrill, 2004). There are three financial statements which interprets the quantitative data of a company's performance. These are: income statements, balance sheets and statement of cash flow.

According to Petrit, 2019, financial statements and reports provide information on:

I. Assets
II. Liabilities
III. Equity
IV. Income and expenditure and
V. Cash Flow

3.3 Balance Sheets

The balance sheet is financial statement. It reports a company's assets, liabilities and shareholder equity. It shows the amount invested by shareholders and the company's ownership of the investments by shareholders (Adams, 2019).

There are two types of sub accounts in the balance sheet. The assets account includes the current and fixed assets of the company (Frank, 1989). Current assets include cash, market securities, accounts receivable, inventories, prepaid expenses, etc. (Frank, 1989) Similarly, the other sub account includes the liabilities and equity. This includes accounts payable, short term debt, accrued expenses and notes payable (Frank, 1989).

Figure2 below shows an example of a balance sheet. It shows the assets, properties, liabilities, and shareholders' equity for Teddy Fab Inc. (freshbooks, N.D.)

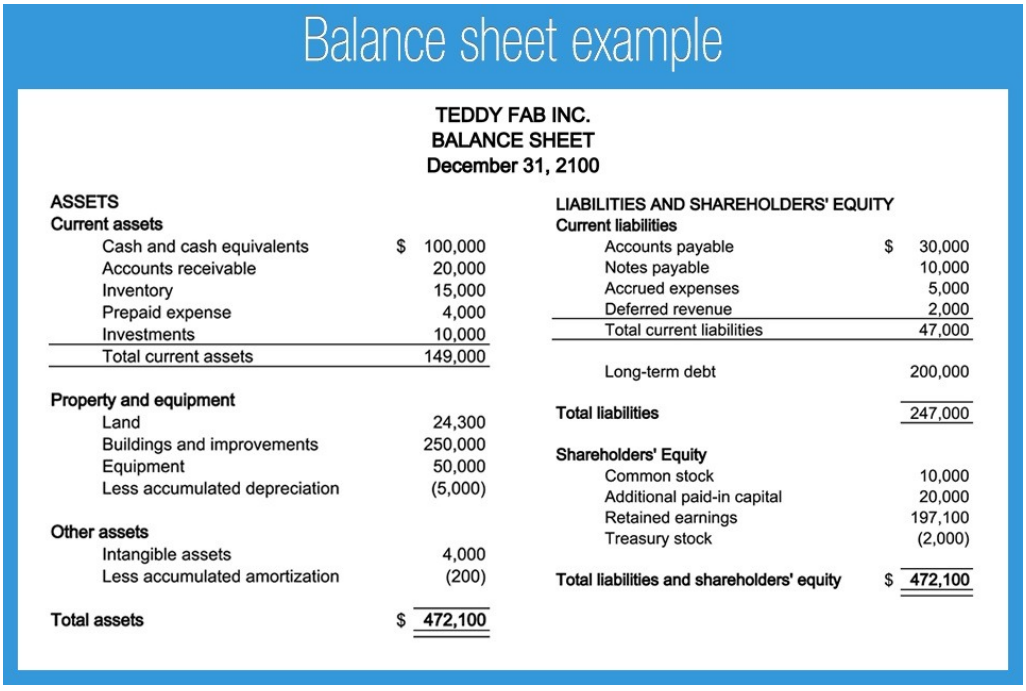


Figure 2. Balance sheet for Teddy fab Inc (Timothy and Joseph, 2003)

3.4 Cash Flow

This is a financial statement which shows an analysis of operating, investing and financial activities. It is related with the flow of cash which is in and out of business. It is useful in determining the short-term viability of a business firm. According to Patrick et al (2002), *"Cash flow helps the investors and creditors to access the ability of the firm to generate positive future cash flow, ability to meet the debt obligations and to shed light on the cash and non-cash aspect of the investing and financial transactions."*

The operating activities are net income, depreciation, the increase or decrease in marketable securities, accounts receivable, inventory, prepaid expenses, account payable, and accrued expenses (Timothy and Joseph, 2003). Figure 3 below shows cash flow statement.

	© Corporate Finance Institute. All rights reserved.	Startup year						Terminal year
		2016	2017	2018	2019	2020	2021	2022
	Balance Sheet Check	*OK*	*OK*	*OK*	*OK*	*OK*	*OK*	*OK*
	Cash Flow Statement							
	Cash from Operations							
	Net Income	(2,573,040)	(1,340,320)	42,855	925,320	2,086,560	3,706,200	4,303,920
	Depreciation	100,000	120,000	140,000	250,000	300,000	420,000	520,000
	Changes in non cash working capital							
	Accounts Receivable	52,529	23,456	45,022	68,373	72,976	105,363	68,813
	Inventory	900,493	154,849	396,740	733,068	766,356	1,084,438	755,260
	Accounts Payable	880,708	(111,228)	7,108	69,245	265,394	383,178	256,961
1	Total	(2,545,354)	(1,509,854)	(251,799)	443,124	1,812,621	3,319,576	4,256,808
	Cash Invested							
	Capital Expenditures	500,000	100,000	100,000	550,000	250,000	1,100,000	600,000
	Acquisitions	0	0	0	0	0	0	0
2	Total	500,000	100,000	100,000	550,000	250,000	1,100,000	600,000
	Cash from Financing							
	Increase (decrease) in Debt	0	0	0	0	0	0	0
	Equity Issued (repurchased)	8,000,000	0	0	0	0	0	0
	Dividends Paid	0	0	0	0	0	0	0
3	Total	8,000,000	0	0	0	0	0	0
	Total Cash	4,954,646	(1,609,854)	(351,799)	(106,876)	1,562,621	2,219,576	3,656,808
	Opening balance	0	4,954,646	3,344,792	2,992,993	2,886,117	4,448,738	6,668,315
	Increase (Decrease)	4,954,646	(1,609,854)	(351,799)	(106,876)	1,562,621	2,219,576	3,656,808
4	Closing balance	4,954,646	3,344,792	2,992,993	2,886,117	4,448,738	6,668,315	10,325,123

Figure 3. Cash Flow (Timothy and Joseph, 2003)

3.5 Industrial Classification System

This is a type of economic taxonomy which organises companies into industrial groups. The grouping is based on similar products, and financial markets. It is used by national and international statistical agencies for summarizing the economic conditions (Christin, 2005).

3.5.1 Standard Industrial Classification (SIC)

This is the four-digit industrial code which was developed by the United States Office of Management for Researchers and Practitioners. (Christin, 2005) The first digit covers 10 divisions for example, mining or manufacturing: -the first two digits cover 81 major groups like oil and gas extraction or paper and allied products, the first three digits cover industry groups like converted paper, all four industries like Envelopes. (Christin, 2005)

3.5.2 North American Industry Classification System (NAICS)

The numbering system employes a five or six-digit code. The first two digits shows the business sector. The third digit the subsector. The fourth the industry group and the fifth the NAICs industries. The system is applied in three countries ;- The United States of America, Canada and Mexico, and covers twenty sectors (Christin, 2005).

3.5.3 Global Industry Classification Standard (GICS)

The Global Industry Classification Standard is a development by Morgan Stanley Capital International (MSCI) and Standard & Poor (S&P) (Christin, 2005). The GICS system consists of 10 sectors like energy or financials, 23 industry groups like oil and gas or Insurance with 59 industries like oil and gas drilling or insurance brokers, and 122 sub-industries (Christin, 2005). The system links an eight-digit code to each company (Christin, 2005). Data are available from December 1994 for S&P 1500 companies, and from June 1999 for non-S&P companies (Christin, 2005). The classification of companies is primarily

based on revenues but also on earnings and market perception. Diverse companies are members of separate industry groups or industries (Christin, 2005).

3.5.4 Dow Jones Global Classification Standard (DJGCS)

The Dow Jones Global Classification Standard provided by Dow Jones covers approximately 45,000 securities worldwide (Christin, 2005). Companies are classified in 10 general economic sectors like financial or consumer - cyclical, 18 market sectors like banks or automobiles, 51 industry groups like auto parts and finally, 89 sub-groups like tires (Christin, 2005). The classification of individual companies is based on revenues from dominant lines of business (Christin, 2005). World scope provides current DJGCS data based on the sub-group level (Christin, 2005). This means that every company is linked to one sub-group that consists of three characters (Christin, 2005).

3.5.5 World scope industry groups (WSIG)

Thomson Financial provides a four-digit numeric code system where each company is linked to one code based on the net sales or revenues figure (Christin, 2005). The first two digits represent one of 27 major industry groups (Christin, 2005). Major groups are for instance aerospace, automotive or chemicals (Christin, 2005). The next two digits represent sub-groups that cover a more detailed industry classification within the major groups (Christin, 2005). The major group, financial contains the most sub-groups (12) (Christin, 2005). The major group beverages on the other hand have only three sub-groups (Christin, 2005). Diversified companies with no clear primary segment but several similar important segments have their own sub-groups. Companies that cannot be linked to a major group are classified in the group miscellaneous (Christin, 2005).

3.5.6 Value Line (VL)

Value Line is a comprehensive source of information and covers approximately 100 industries. (Christin, 2005) The Value Line database contains fundamental data (both current and historical) on more than 7,500 publicly traded North American, European, and

Asian firms. (Christin, 2005) It includes hundreds of items on each firm, with balance sheet and income statement data. Companies are assigned to industries by sales information. Industries are for instance tobacco or medical services. (Christin, 2005) Some industries are separated into speciality and diversiefied classes. There are no sub-categories (Christin, 2005).

3.5.7 Fama and French(FF)

Fama and French developed a classfication system that linked the existing SIC groups based on 4-digits to 48 industries. Their intention was not to develop a new classfication structure: - they were only interested in a manageable number of industries (Christin, 2005).

3.5.8 Return on equity (ROE)

ROE is a measure of financial performance. It is calculated by dividing net income and shareholder's equity. Return on net assets is also considered. It is a good method to develop sustainable growth and dividend growth (Marshall, 2020). The mathematical interpretation of ROE is given in Equation 1 in Chapter 1.

The net income is the amount of income, net expenses, and taxes that a company generates for a given period. Likewise, the average shareholder's equity is calculated by adding equity from the beginning of a period (Marshall, 2020).

The usage of ROE is that it is used for comparing the performances of companies. It is a measure of management's ability to generate income. A percentage of between 15-20% is considered good (Marshall, 2020).

3.5.9 Earnings per share (EPS)

EPS is the company's profit divided by the outstanding shares of its common stock. It shows the firm's money on each share. It can be arrived at in several forms. It excludes discontinued items or operations (David, 2020).

The calculation of earnings per share (EPS) is done by taking the net income statement and the balance sheet to find the period-end number of common shares, dividends paid and the net earnings of the income (David, 2020). Earnings per share (EPS) are interpreted mathematically.

3.5.10 Diluted earnings per share

These are calculated by using the quality of a company's earnings per share if all convertible securities are exercised. The convertible securities are outstanding convertible preferred shares, convertible debenture, stock options and warrants (Akhilesh, 2019).

3.5.11 Basic earnings per share

This is the ratio of net income deducted by the preferred dividends and the weighted average common shares. It tells investors of the firm's net income allotted to each share of common stock. It is used for businesses with simple capital structure (Akhilesh, 2019).

3.5.12 Return on Assets (ROA)

ROA is the ratio of net income to total assets. It is an indicator of how well a company utilizes the assets generating its earnings. It is best used when comparing similar companies or comparing it to its previous performance. The issue with ROA is that it cannot be used across industries (Marshall, 2020).

Mathematically,

$$Return\ on\ Assets = \frac{Net\ Income}{Total\ Assets} \quad 3$$

3.5.13 Revenue

Revenue is the income generated from business operations and includes discounts and deductions from returned merchandise. It is also known as sales on the income statement, or sales as in the price to sales ratio. There are many ways to calculate revenue (Will, 2019).

According to Will, (2019), mathematically,

$$\text{Sales Revenue} = Sales\ Price * number\ of\ Units\ Sold \quad 4$$

4 EMPIRICAL RESEARCH FRAMEWORK

4.1 Supervised Learning Model

Supervised machine learning algorithm is one of the machine learning models in which we teach or train the machine using data, which is labelled, or some data is already tagged with correct answer (Wilson, 2019). The supervised machine learning model is the input to the process and known responses which are based on observations (Aidan, 2019). An example of supervised machine learning includes distinguishing a good or bad mobile phone from a given set of data (Aidan, 2019).

Supervised machine learning is where there is an input variable (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output (Wilson, 2019).
Mathematically,

$$Y=f(x) \qquad 5$$

The goal is to approximate the mapping function so well that when there is new input data (x) the output variables (Y) for that data can be predicted. (Brownlee, 2016) Likewise, the objective of the supervised machine learning model is to correctly label the new input from the existing data (Aidan, 2019). The process of supervised machine learning starts by collecting the data set. The next process is the data preparation and data processing. The feature subset selection is the next process. In this process identifying and removing the unnecessary features are conducted. Likewise, the end process is the training and evaluation (Kotsiantis, 2007). Figure 4 shows the process of supervised machine learning. The process starts with the problem and ends with classification.

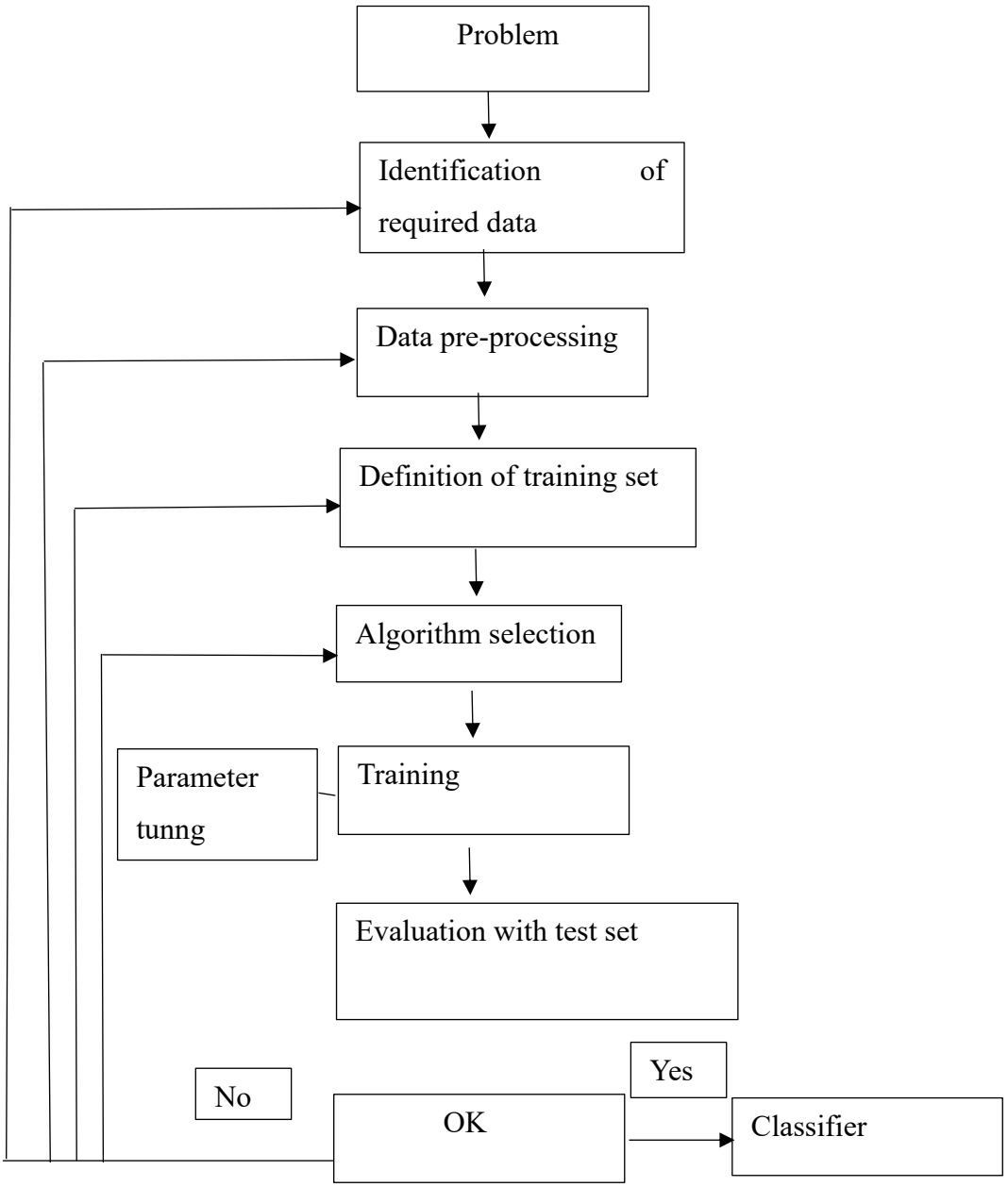


Figure 4. The process of supervised machine learning (Kotsiantis, 2007**)**

There are two main examples of supervised learning techniques: - these are regression and classification. In this thesis, forecasting and classification type of machine learning was examined.

4.2 Classification

The classification model attempts to draw conclusions from observed values (Kirill, 2017). There are several classification models: - these include logistic regression, decision tree, random forest, gradient boosted tree, multi-layer perception, one-vs-rest and naïve Bayes (Kirill, 2017).

The classification is of two types (Alpaydin, 2004 & Hastie et al., 2001), namely linear and non-linear classification. If the items are linearly classified between two attributes [X1, X2] and there are two classes A and B, then the linear equation can be computed based on the equation. (Alpaydin, 2004 & Hastie et al., 2001)

$$y = a_1 x_1 + a_2 x_2 + a_0 \quad (6)$$

$$y = a_0 + a_1 x_1 + a_1 x_1^2 + b_1 x_2 + b_2 x_2^2 \quad (7)$$

Figures 5 and 6 below show a graphical illustration of linear and non- classification.

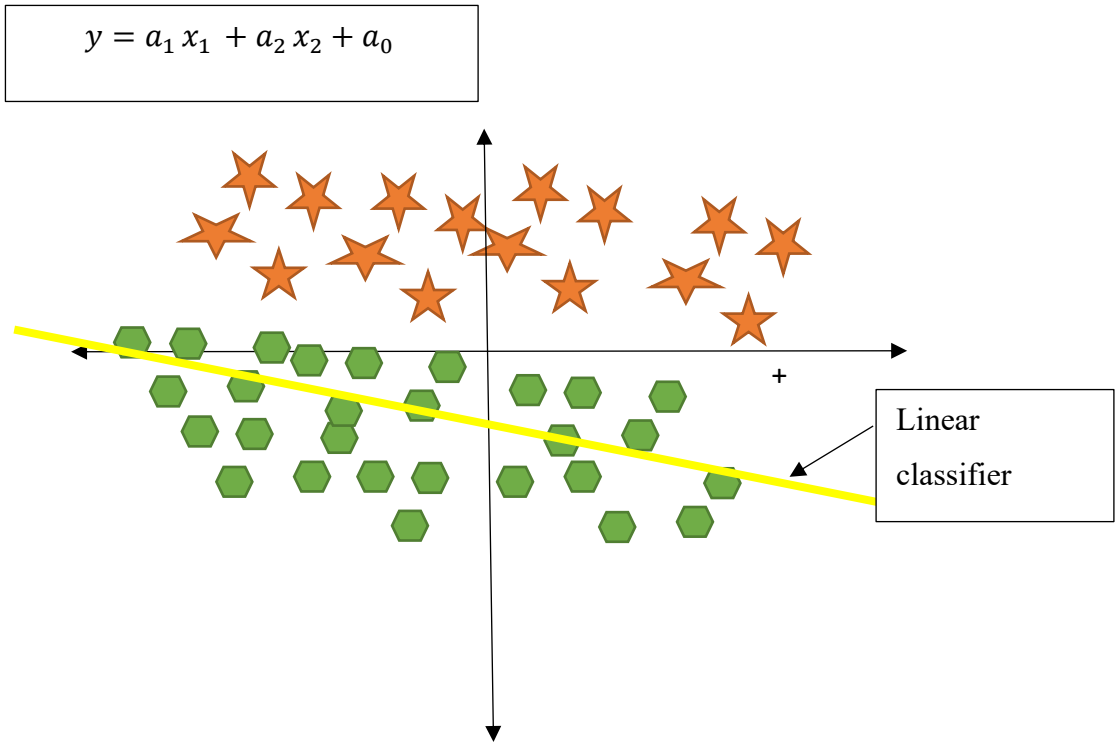


Figure 5. Linear classification

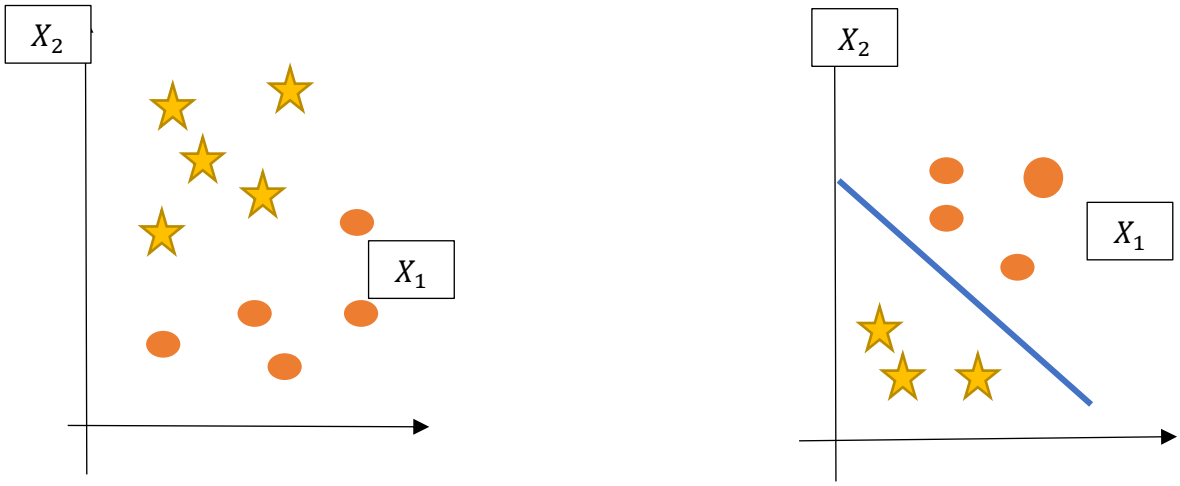


Figure 6. Graph showing non-linear mapping of attributes.

4.3 Recurrent Neural Network

Recurrent neural networks (RNN) are a type of neural network in which the output of previous step is feed as input of the current step. As we know, in traditional neural networks, all inputs and outputs are independent of each other but in cases like when the next output is dependent on the previous output, as in a sentence when the next word is dependent of the previous work, RNN solve this problem with the help of a hidden layer. The main and most important feature of RNN is hidden status, which remembers some information about the sequence.

RNN have a memory which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of the parameters, unlike other neural networks.

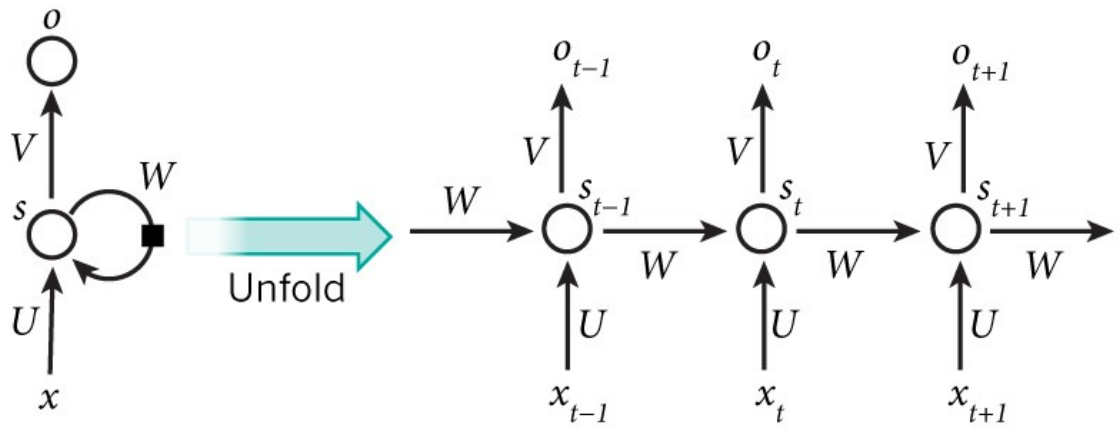


Figure 7. RNN

4.4 Time-series analysis

A time series is a sequence where a metric is recorded over regular time intervals. (Selva, 2019) It is defined as a set of random variables ordered with respect to time. It is a statistical technique that deals with trend analysis. It takes existing series of data and forecasts the data values to future time (Hanis, Curtis, Thalassinos, 2012). The goal of the forecast is to predict future unknown values.

The data available to firms are compared in time series. These are used to forecast inter-company financial performances. The trend developed from time series analysis is used to predict future earnings, sales and ratio. There are several methods for time series analysis. The methods are simple moving average, exponential moving average, and double exponential smoothing, amongst others. (Hanis, Curtis, Thalassinos, 2012).

4.5 Long Short-Term Memory

This is an artificial recurrent network architecture. It is used in the field of deep learning. It is well-suited to classifying, processing, and making predictions based on time series data (Hochreiter, 1997). It is composed of a cell, an input gate an output gate and a forget gate (Hochreiter, 1997). The advantages of long short-term memory (LSTM) are constant error backpropagation within memory cells which results in LSTM's ability to bridge a very long time (Hochreiter, 1997).

The architectures of long short-term memory (LSTM) contain special units called memory blocks in the recurrent hidden layer (Sak, Senior, Beaufays, 2014). Likewise, it also contains memory cells with self-connections storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information. The input gates control the flow of input activations. The output gates control the output flow of cell activations into the rest of the network. Figure 8 shows the memory block and LSTM architecture.

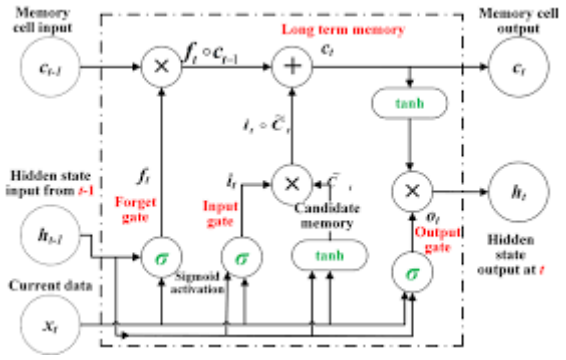


Figure 8. Long short-term memory (Hochreiter, 1997)

Mapping from an input sequence $x = (x_1, \ldots\ldots x_T)$ to an output sequence $y = (y_1, \ldots\ldots y_T)$ by calculating the network unit activations uses the following equations.

$$i_t = \sigma(W_{ixx_t} + W_{imm_t} + W_{icc_{t-1}} + b_i) \quad 8$$

$$f_t = \sigma(W_{fxx_t} + W_{fmm_{t-1}} + W_{fcc_{t-1}} + b_i) \quad 9$$

$$c_t = f_t . c_{t-1} + i_t \, \theta \, g(W_{cxx_t} + W_{cmm_t} + b_c) \quad 10$$

$$o_t = \sigma(W_{oxx_t} + W_{omm_{t-1}} + W_{occ_t} + b_o) \quad 11$$

$$m_t = o_t \, \Theta h(c_t) \quad 12$$

$$y_t = \emptyset(W_{ymm_t} + b_y) \quad 13$$

where the W terms denote weight matrices(e.g. Wix is the matrix of weights from the input gate to the input), Wic, Wfc, and Woc are diagonal weight matrices for peephole connections, the b terms denote bias vectors (bi is the input gate bias vector), σ is the logistic sigmoid function, and i, f, o and c are respectively the input gate, forget gate, output gate and cell activation vectors, all of which are the same size as the cell output activation vector m , $\theta$is the element-wise product of the vectors, g and h are the cell input and cell output activation functions, and in this paper tanh, and ϕ is the network output activation function.

4.6 Why LSTM model?

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically by their default behaviour, not something which they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer. LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way. The key to LSTMs is the cell state, the horizontal line running through the top of the diagram.

The cell state is kind of like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions. It is very easy for information to just flow along it unchanged. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates.

Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation. The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through. A value of zero means “let nothing through,” while a value of one means “let everything through!” An LSTM has three of these gates, to protect and control the cell state.

In the LSTM layer, ‘RELu’ input layer was used with using appropriate input shape. Further added 10 LSTM units to the LSTM layer. as the output layer Dense layer was used with the 10 output layers. In the dataset the predictions were set to 10 years of time. In the loss function “mean_squared_error” function was used and as the optimizer, ‘adam ’ optimizer function was used. for the batch size and the epochs, values of 1 and 100 were added, respectively.

When considering the dataset, there is not test dataset to evaluate trained model. in that sense the loss of the trained model was plotted and mentioned below.

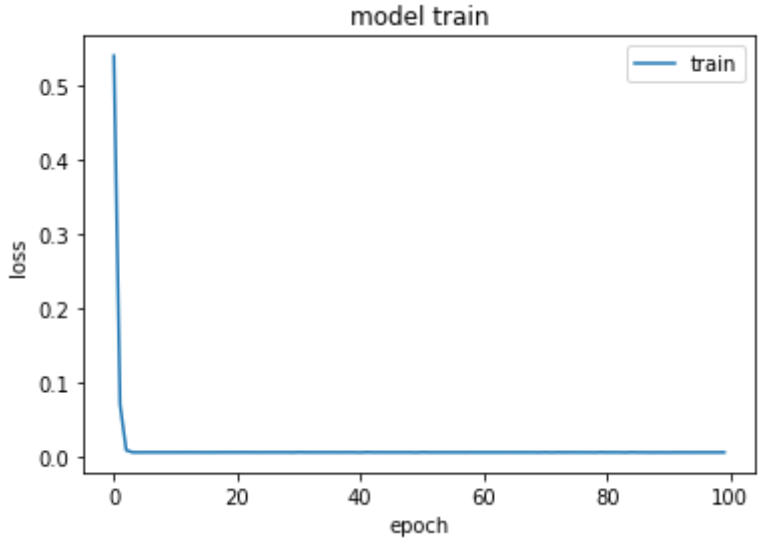


Figure 9. Model train

When model in trained, it clearly stated that the lack of data is really troubling the model. even though the loss is reduced to 10% the model is started overfitting. The reason for the overfitting is the lack of data. There is only 250 data were remaining. As the suggestion for that adding more data could resolve this problem. When predicting for the future years, using past years data is important. In this model I have used 10 years of past data to predict the future data for another 10 years. Also, it can be used to predict 5 years of future predictions by using 5 years of data. When predicting the future ROE its necessary to get all the previous data on ROE. Like mentioned above, to predict the 10 years of future values, it's better to have past data with 10 years of time. In the training dataset, the amount of the training data is too small to get predictions for more than 5 or 10 year. It is the major drawback. Since this is a 5- or 10-years' time value prediction, there is a necessity of having considerably large scale of data is important.

LSTMs are stochastic, meaning that you will get a different diagnostic plot each run. The advantages of long short-term memory (LSTM) are constant error backpropagation within memory cells which results in LSTM's ability to bridge a very long time period.

It can be useful to repeat the diagnostic run multiple times (e.g., 5, 10, or 30). The train and validation traces from each run can then be plotted to give a more robust idea of the behavior of the model over time.

Due to dataset intensity and features, decided to make 10 LSTM units, with this dataset neither get too much space into memory units or neurons also don't get too low space that data overwrite to each other. Besides using only memory units may turns all data into linear form and makes model overfit. By added LSTM features to prevent the over-fit.

In the LSTM layer, 'RELu' input layer was used with using appropriate input shape. Further added 10 LSTM units to the LSTM layer. as the output layer Dense layer was used with the 10 output layers. In the dataset the predictions were set to 10 years of time.

The reason for the overfitting is the lack of data. There is only 250+ data were remaining. As the suggestion for that adding more data could resolve this problem. Due to lack of dataset, unable to use validation data to test the model during training, to handle this problem there is a solution that is to use cross validation to dataset, In this data splits into different pieces and train on defined pieces and validate on defined pieces. Data again randomly splits pieces and again model trains that overcome the issue of lack of dataset also prevent model to overfit.

5 METHODOLOGY

Long Short-Term Memory (LSTM) was used in this study. The research used a quantitative methodology. The analysis was done using Jupyter notebook with Python version. The needed library was imported in the Jupyter notebook.

The goal of the analysis was to find the forecast and to classify the companies based on their return oneEquity (ROE), earnings per share (EPS), return on assets (ROA) and revenue. Likewise, the result of the forecast of the company was used for the classification. The expected result was multi-class classification. The result was based on the 5% rule (Qiu, Padmini, Nick, 2007). If the company's 5-year average forecast is better than the preceding year the result will show better performance (Qiu, Padmini, Nick, 2007). Likewise, if the result shows no change, it will show neutral performance (Qiu, Padmini, Nick, 2007). Similarly, if the company's average performance was less than 5% it shows worst performance. The forecast and classification years were from 2020 to 2025.

The data preparation of the company was selected based on return on equity (ROE), return on assets (ROA), diluted and basic earnings per share (EPS) and revenue with respect to year. The selected years were from 2010 to 2019. The data were collected from the Orbis database. The data were arranged in columns. The first column consisted of the name of the company. The second column is the ROE, ROA, EPS and revenue with respect to the years from 2010 to 2019. The diluted earnings per share was arranged with respect to the years from 2015 to 2019.

The extracted data was pre-processed. These pre-processed data were loaded into Jupyter notebook. The data was divided into a training and testing set. The data used for training was 80%, and for testing 20% data were used. The input was the return on equity (ROE), earnings Per share (EPS), return on assets (ROA) and revenue. The output was the forecast result of the company for that particular year and classification was based on better performance, worst performance, and neutral performance.

After, completion of the data pre-processing, the data was ready for training a model. The algorithm that was used for the model was long short-term model. The trained model was saved to hard disk with separate folder name.

The name and year of the company to be forecast can be written in the blank space at the bottom where the name of the company is written. This is then followed by the year of the company. The code is then run by pressing the run button. The result is the forecast value of the particular year. The results are the ROE, ROA and the revenue and the classification, which is good, bad or neutral performance. The code for the forecasting can be found in Appendix 1.

The same process can be done with the classification. The code for the classification can be found in Appendix 2. The companies are classified in three categories. These are good, bad and neutral. The code for classification is attached in Appendix 2.

6 RESULTS AND ANALYSIS

In this section, the forecast results for return on asset (ROA), return on equity (ROE), revenue and diluted and basic earnings per share (EPS) will be discussed.

6.1 Forecast for Return on Assets

The forecast for return on assets is shown in Figure 9 below. The number in the x-axis represents the companies while the Y-axis depicts the variables. There are 210 under consideration. The highest results are companies with number between 120-129 and 134-138. These companies are Brooks International, Deere & Co., Digi International Inc, Kennametal Inc, Netapp Inc and Nordson Corp.

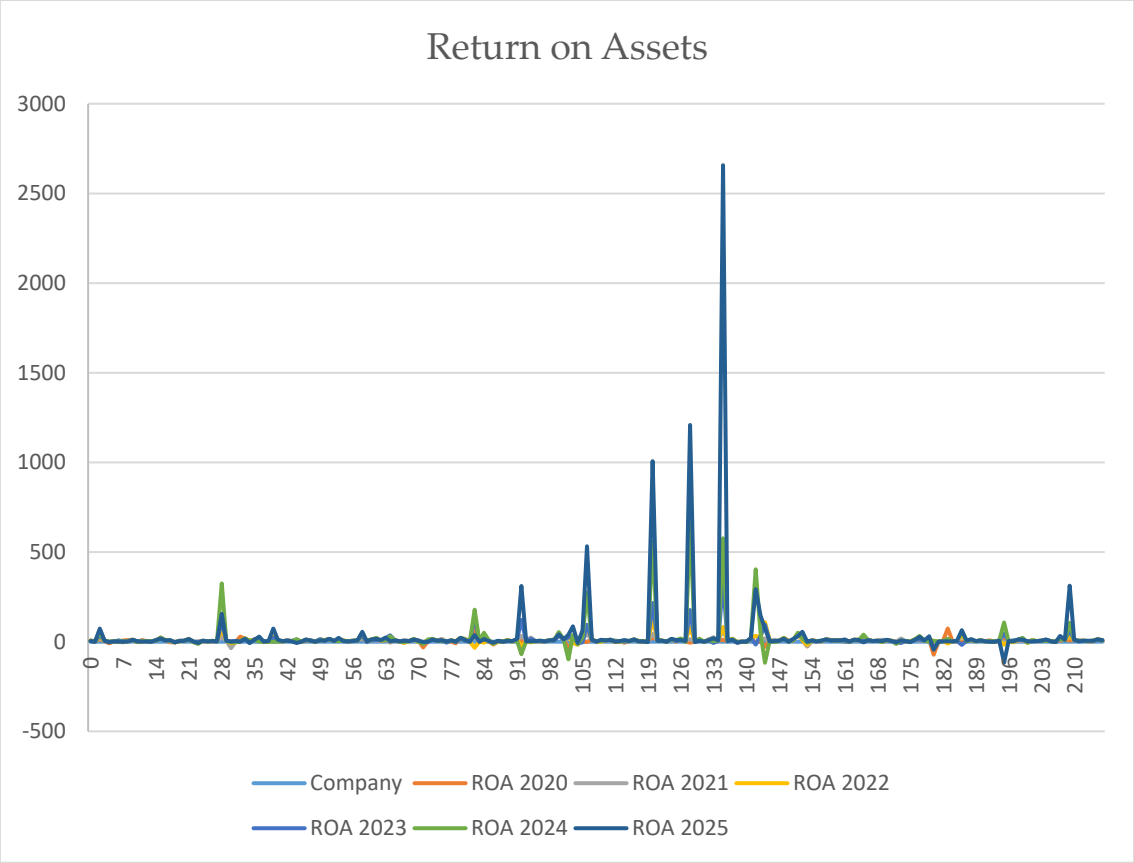


Figure 10. Return on assets.

The x-axis shows the companies for the year 2020. The y-axis shows the companies which are in the return on assets values. The return on assets for the year 2020 are shown in the figure 11 below. The forecast shows that the highest return on assets value are the companies 80 (light path technologies inc), 100 (Arrowhead pharmaceuticals inc), 150 (Tapestry Inc) and 185(Suburban propane partners). The lowest are in 70-72 (Cree inc, Extreme networks).

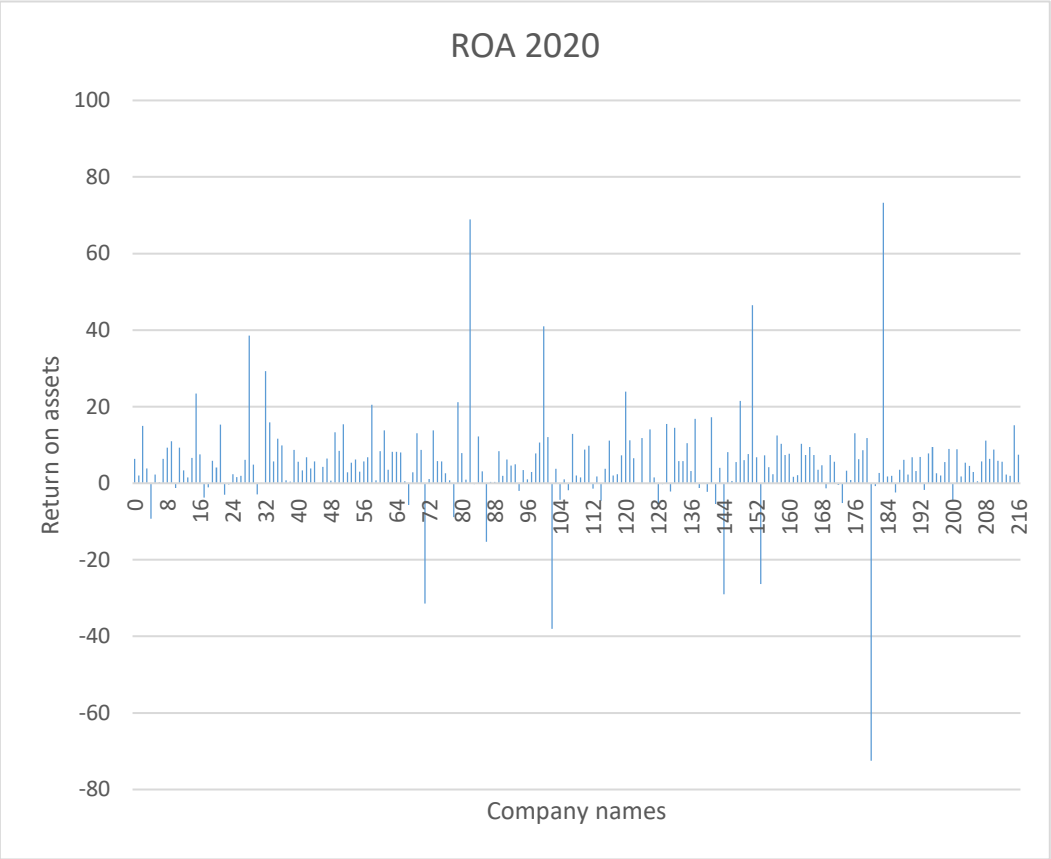


Figure 11. Return on assets for the year 2020.

The figure 12 below shows the return on assets for 2021. The x-axis shows the company names. The y-axis shows the return on assets value. The highest values are the companies 106(Farmers Bro), 120 (Lazboy inc) and 134(Brook's automation Inc). The lowest values are the 30 (Charter communications), 100 (Arrowhead inc) and 153 (Commercial Metal).

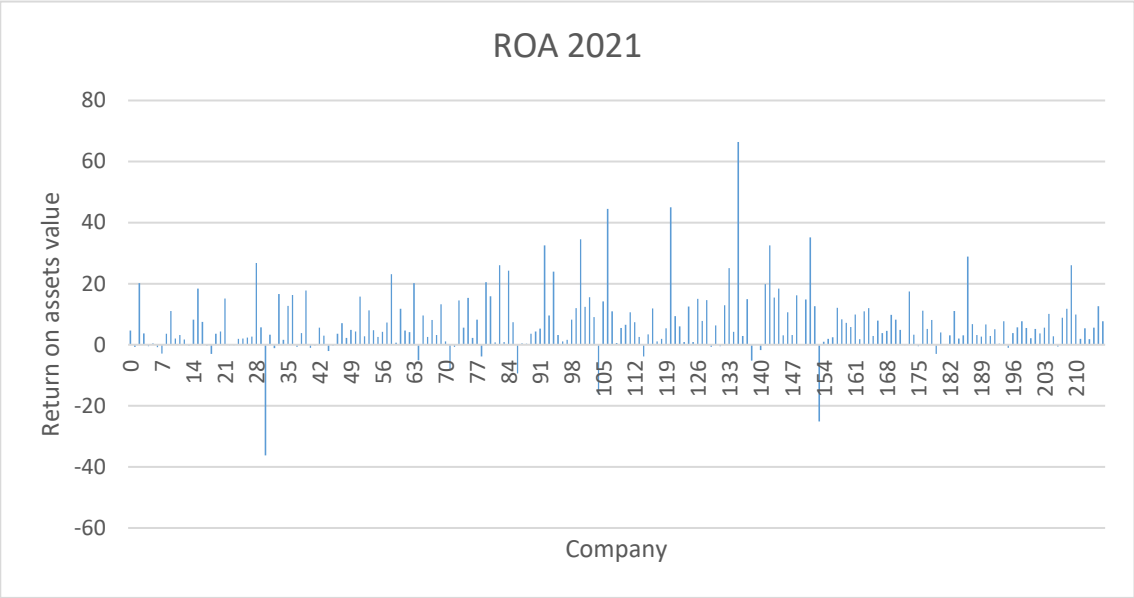


Figure 12. Return on assets for the year 2021.

The figure 13 below shows the return on assets for the year 2022. The x-axis shows the companies. Likewise, the y-axis shows the return on assets values. The highest returns are the companies with values 108 (Hain celestial group inc), 120 (La-z boy Inc) and 145(Boot barn holding inc). The lowest are the companies with 85 (Micron technology inc), 153(Commercial Metal) and 195(General motors company).

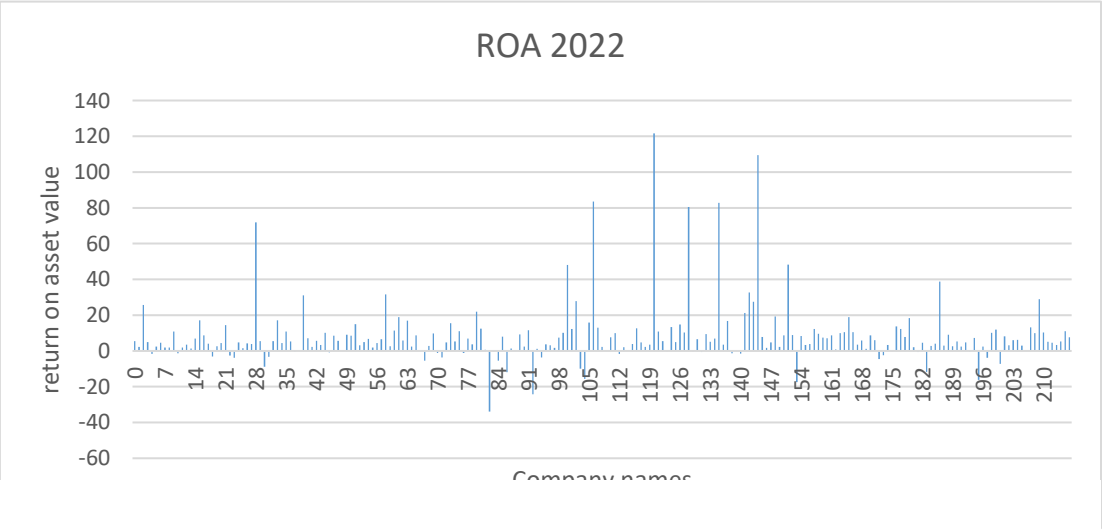


Figure 13. Return on assets for the year 2022.

The figure 14 below shows the return on assets for the year 2023. The X-axis shows the company names, and the Y-axis shows the values. The highest return on assets is 28 (Cnx resources corporation), 120 (Laz boy inc) and 135 (Deere and co). The lowest return on assets is the 34(T-mobile US inc), 85 (Micron technology inc)and 180 (Amazon Inc.).

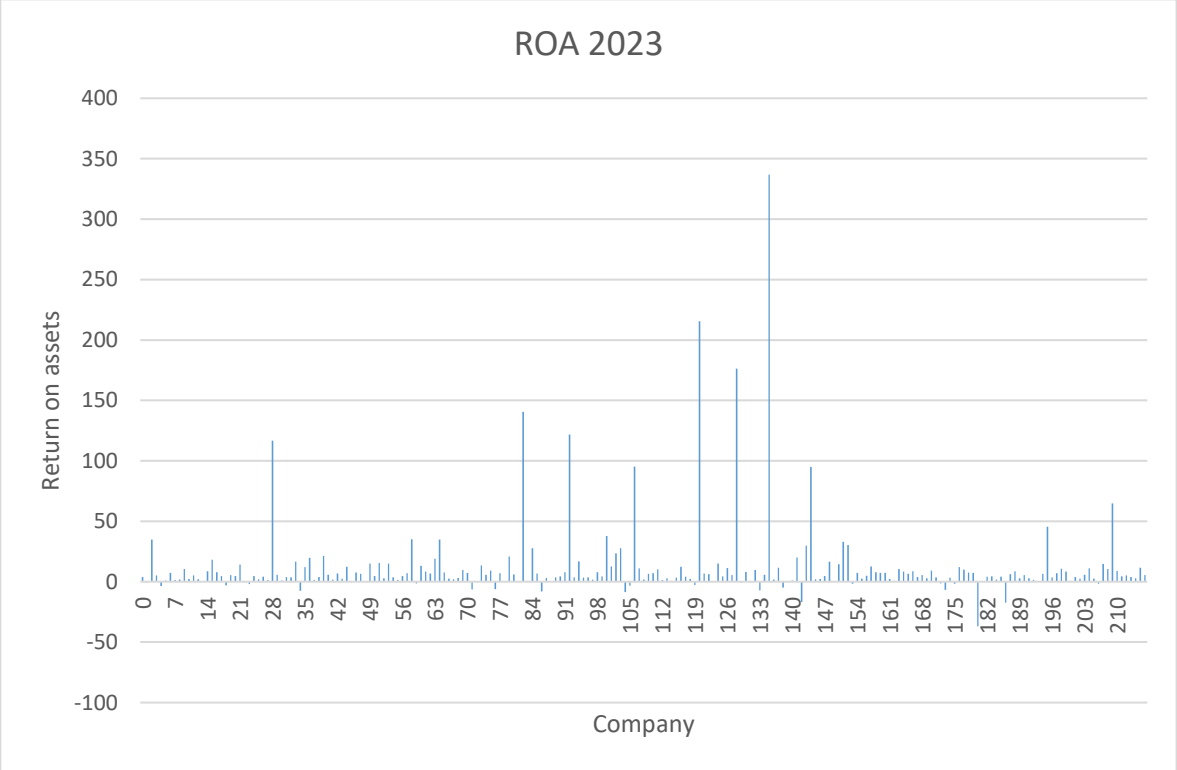


Figure 14. Return on assets for the year 2023.

The figure 15 is shown below. The x-axis are the company names. The y-axis is the return on assets value. The highest return on assets is 28 (Cnx resources corporation), 120 (Laz-boy Inc), 128(Mercury systems inc) and 135(Deere and co). The lowest company are, 92(Aviat networks), 100 (Arrowhead pharmaceuticals) and 145 (Boot barn holdings inc).

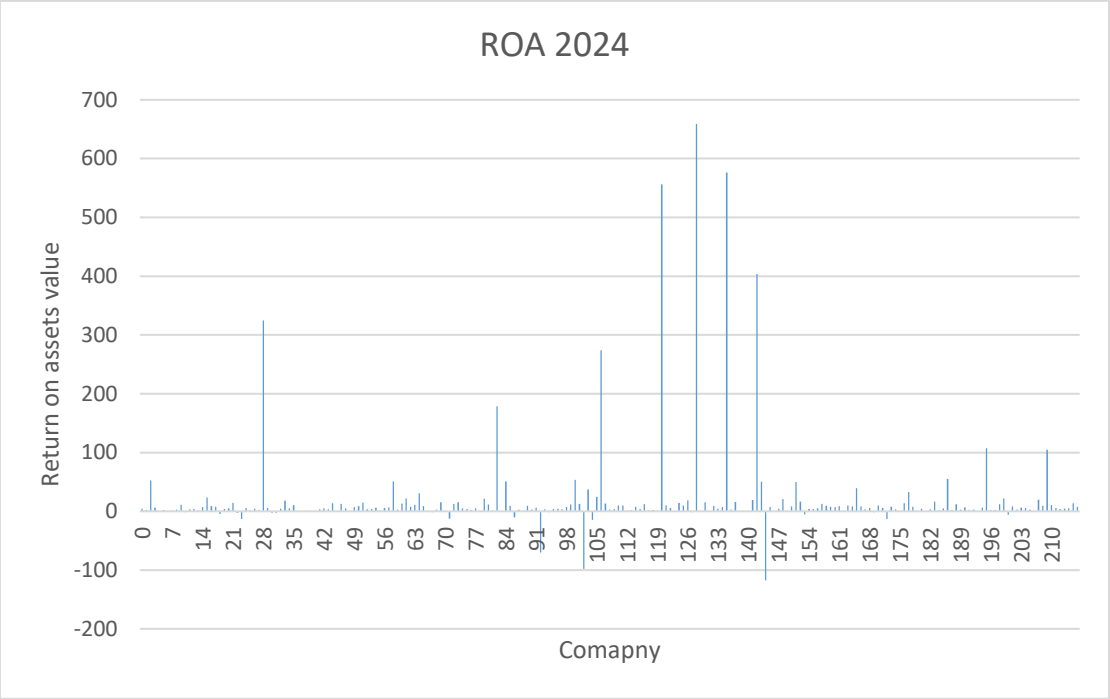


Figure 15. Return on assets for the year 2024.

The figure 16 below shows the return on assets of 2025. The highest return on assets is 120(Laz boy inc), 128 (Mercury systems inc) and 135(Deere and co). The lowest company is 195 (General motors co).

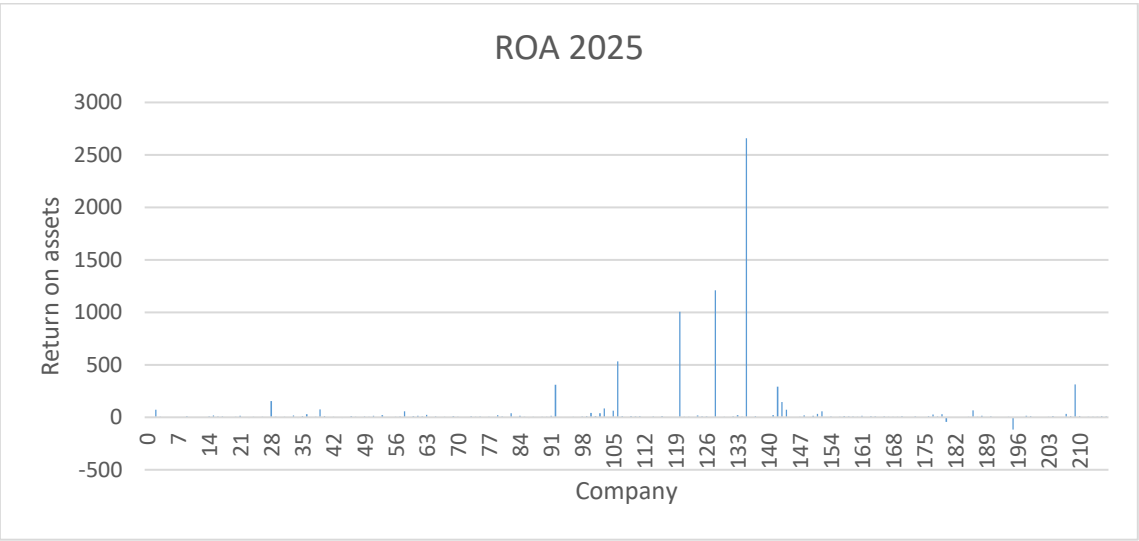


Figure 16. Return on assets for the year 2025.

6.2 Forecast for Return on Equity

The figure 17 below shows the return on equity for the year 2020. The companies with the highest values are 10 (Avid bioservies inc), 15 (Cardinal ethanol inc) and 143(Western digital corp). The lowest return on equity for the companies are 30(Charter communications), 55(RGC resources inc), 191(Aat corp) and 207(Applied industrial technologies).

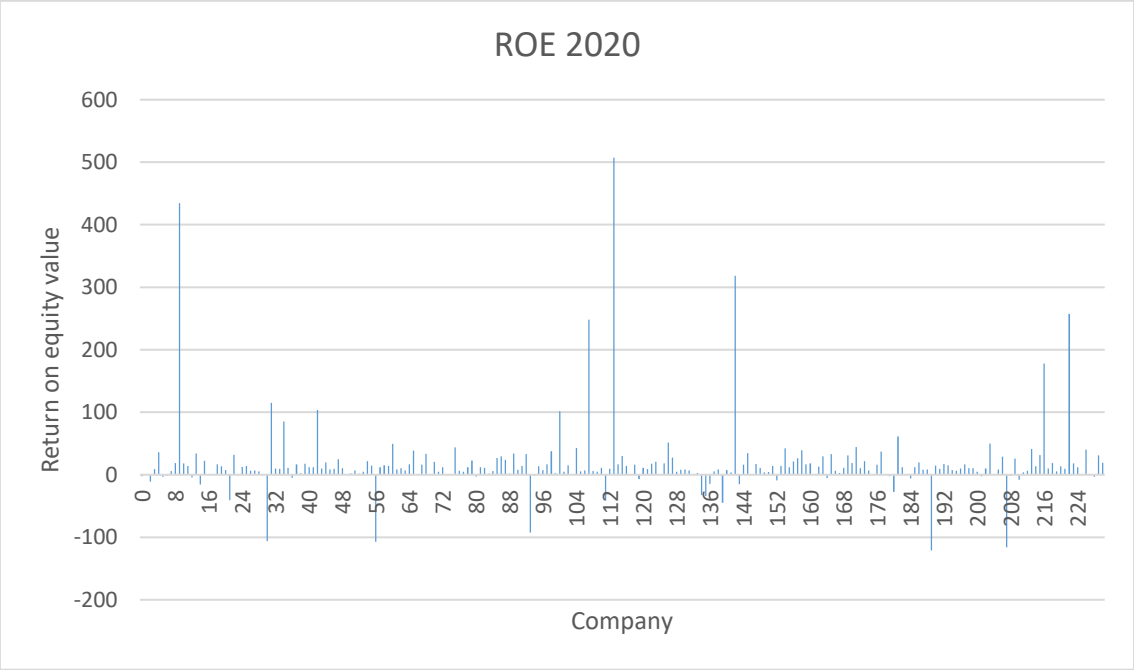


Figure 17. Return on equity for the year 2020.

The figure 18 below shows the return on equity for the year 2021. The highest return on equity for the year 2021 are the companies labels 145(boot barn holdings inc), 215(Patterson companies) and 223 (Cvs health corporations). The lowest ones are the companies with the labels 30 (Charter communications), 100 (Arrowhead pharmaceuticals inc) and 188 (Eaton Vance corp).

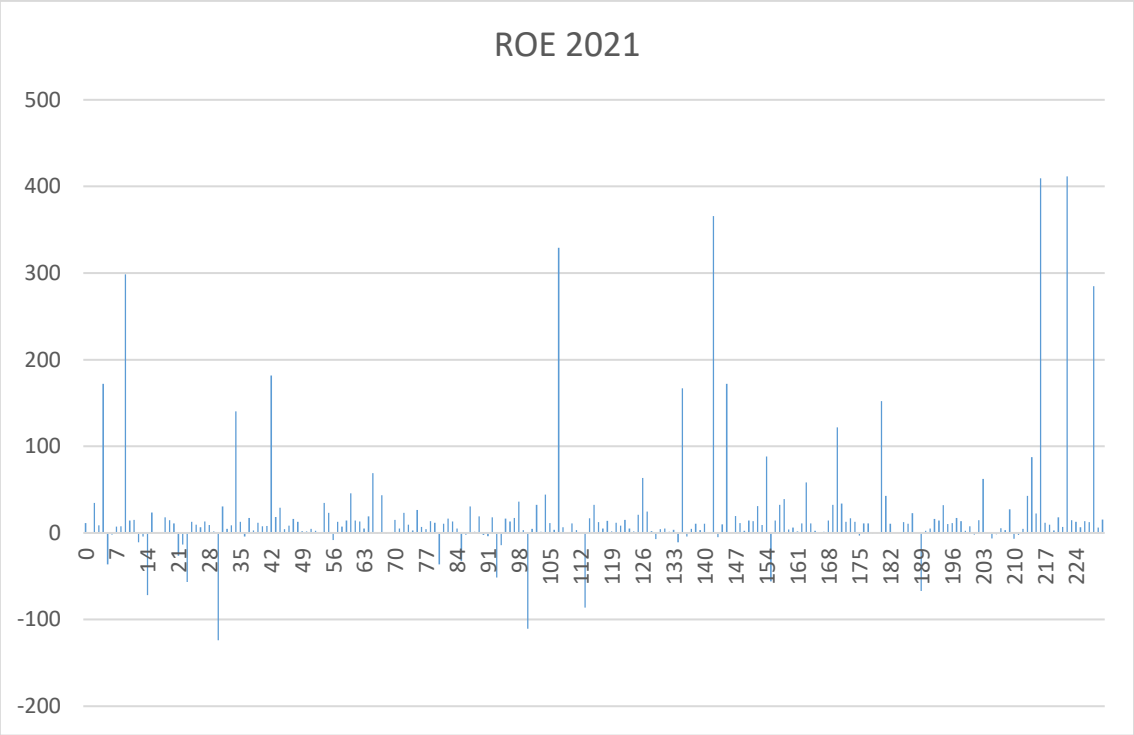


Figure 18. Return on equity for the year 2021.

The figure 19 below shows the return on equity for the year 2022. The highest return on equity for the year 2022 are the year with the labels 108 (Hain celestial group inc), 113 (Sanderson farm inc) and 223 (CVS health corporation). The lowest ones are of the year are 55(RGC resources) and 92 (Aviat networks inc).

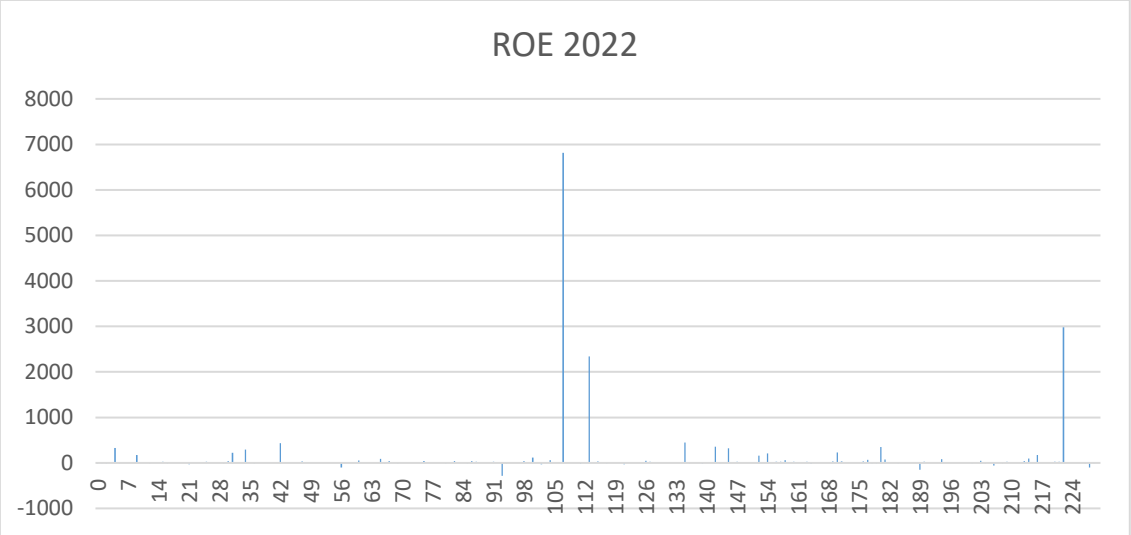


Figure 19. Return on equity for the year 2022.

The figure 20 below shows the total forecast for the return on equity. The x-axis shows the values for the forecast whereas, the x-axis shows the company values. The values of the ROE are depicted on the X-axis, while the Y-axis represents the companies'. There are in total 228 companies under consideration. The companies that fall within 100 - 105 and 215 -220 shows the highest return on equity. These companies include General Motors, Grante Falls Energy Inc., Greenbrier Companies Inc, Grief Inc, and Griffon Corp. Similarly, companies that fall within the 215 – 220 range are United Natural Foods Inc, United Parcel Services Inc., Until Corp, Universal Technical Institute, and Uranium Energy Corp as shown in Figure 20.

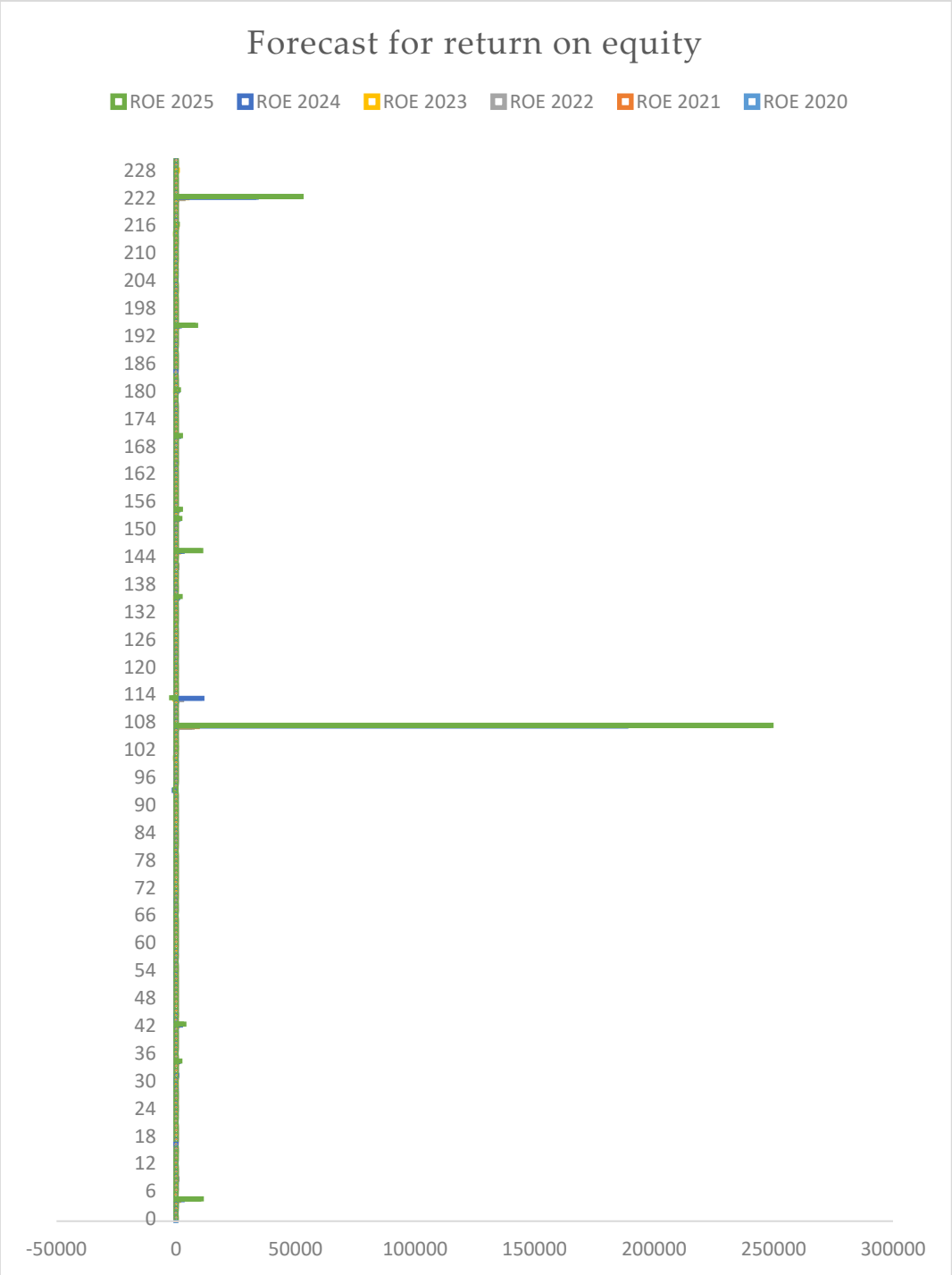


Figure 20. Forecast value for the return on equity.

6.3 Forecast for revenue.

Figure 11 below shows the forecast for revenue. In the graph, the companies are in X-axis and the Y-axis shows companies with the highest values. There are in total 216 companies. The forecasts for the highest revenue companies were from 9-12. These companies are Alliance Resources Partners, Amazon Inc. and America's Car Mart inc. The average company forecast was on average below 1000000000. The average value of the company is below 100000000.

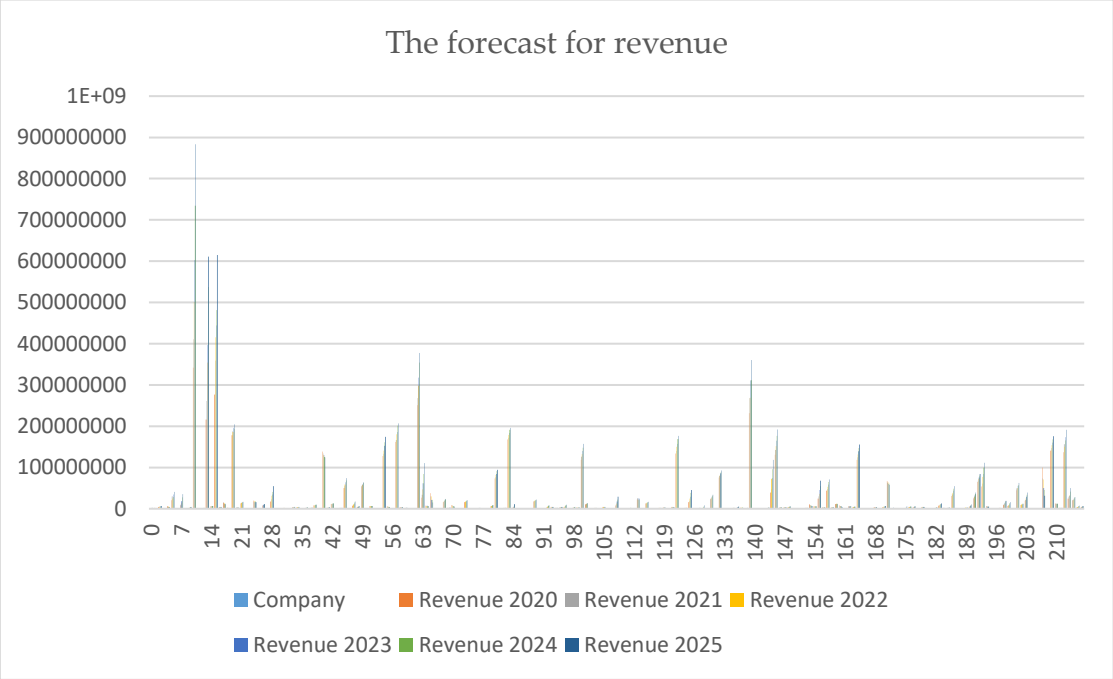


Figure 21. Forecast value for the revenue**.**

6.4 Forecast for Earnings per Share (diluted and basic)

The forecast results for earnings per share are shown in Figure 15 below. The x-axis shows the number of companies. The total number of companies is 232. Likewise, the y-axis shows the earnings per share values. The highest returns per share are from companies between 85 and 92. The companies included are Extreme Networks, Exxon Monil Corp, Facebook, Inc, Framers Bro, Ferrellgas Partners, Ford Motors, Foresight Energy LP.

Likewise, the medium highest numbers are between 51-57 and 10-15. The companies between 51-57 are, Charter Communications, Chaveron Corporation, China HGS, Chiptole Mexican Grill Inc and Ciena Corp. Similarly, the companies which are in group 10-15 are, Amech Systems, Amercia'a Car Mart Inc, American Outdoor Brands Corporation, AmerisourceBergen Corp, Analog Devices Inc., and Apple Inc.

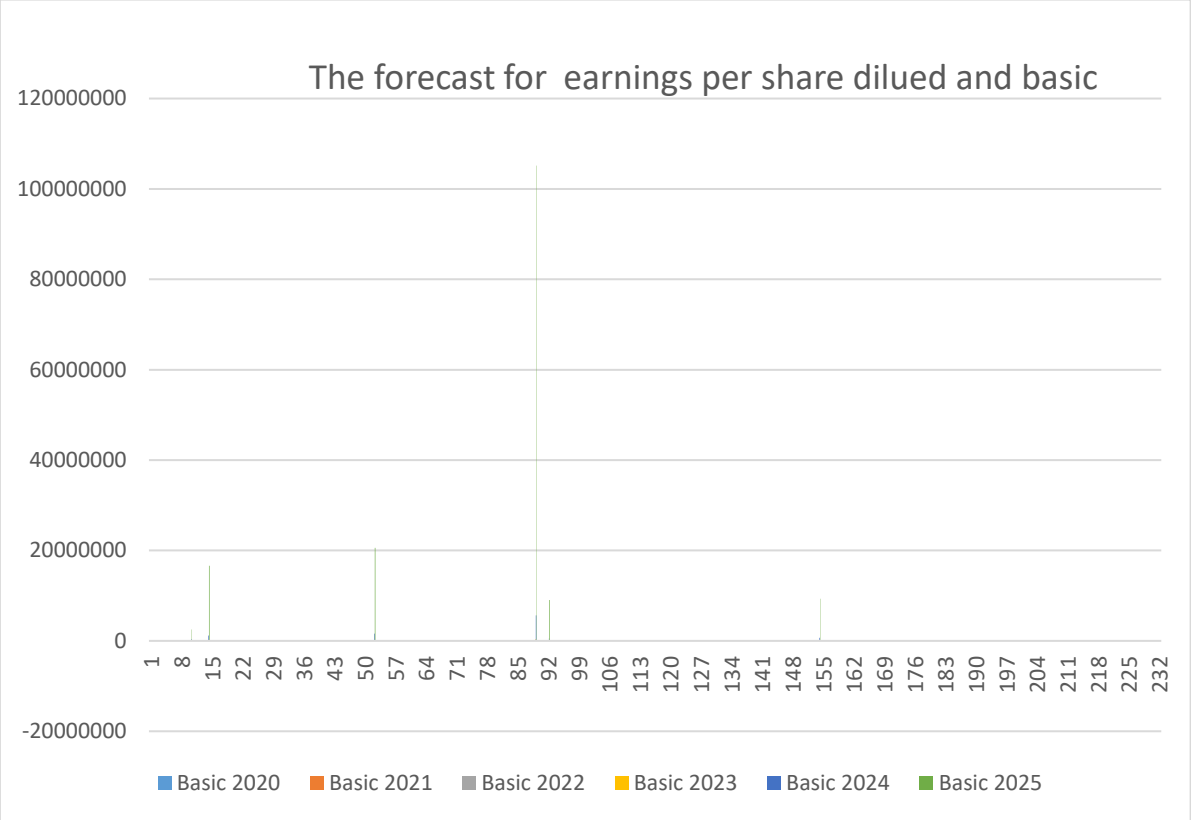


Figure 22. Forecast value for the earnings per share.

The earnings per share for the year 2020 is shown in the figure 23 below. The x-axis shows the company names. The Y-axis shows the EPS values. The highest earnings per share for the year 2020 is with the company label 154 (Enerpac tool group corp).

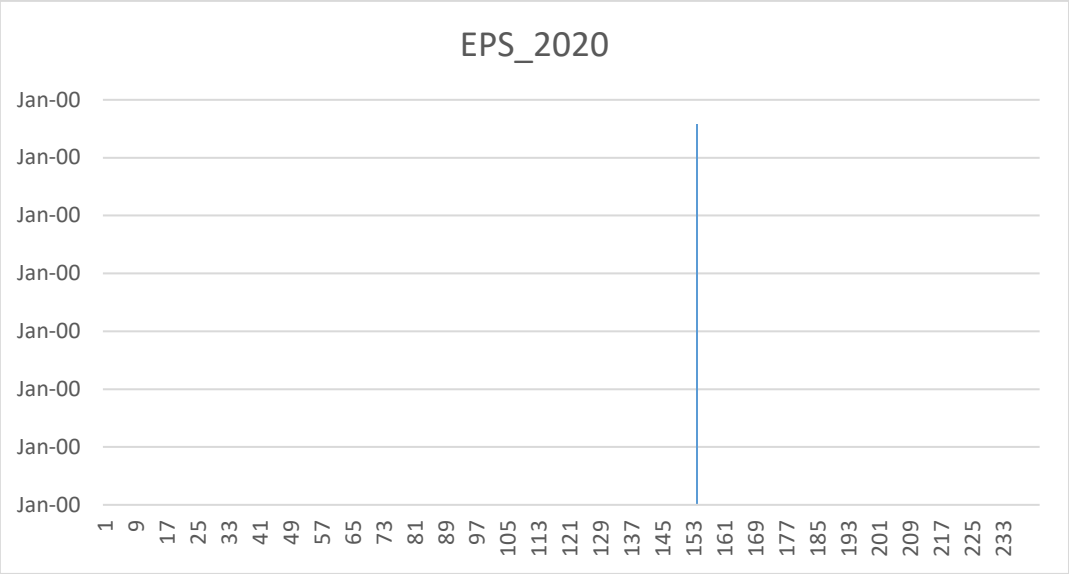


Figure 23. Earnings per share for the year 2020

The figure 24 below shows the earnings per share for the year 2021. The x-axis shows the company names whereas the y-axis shows the earnings per share. The company with the highest eps for the year 2021 is the year 154 (Enerpac tool group corp).

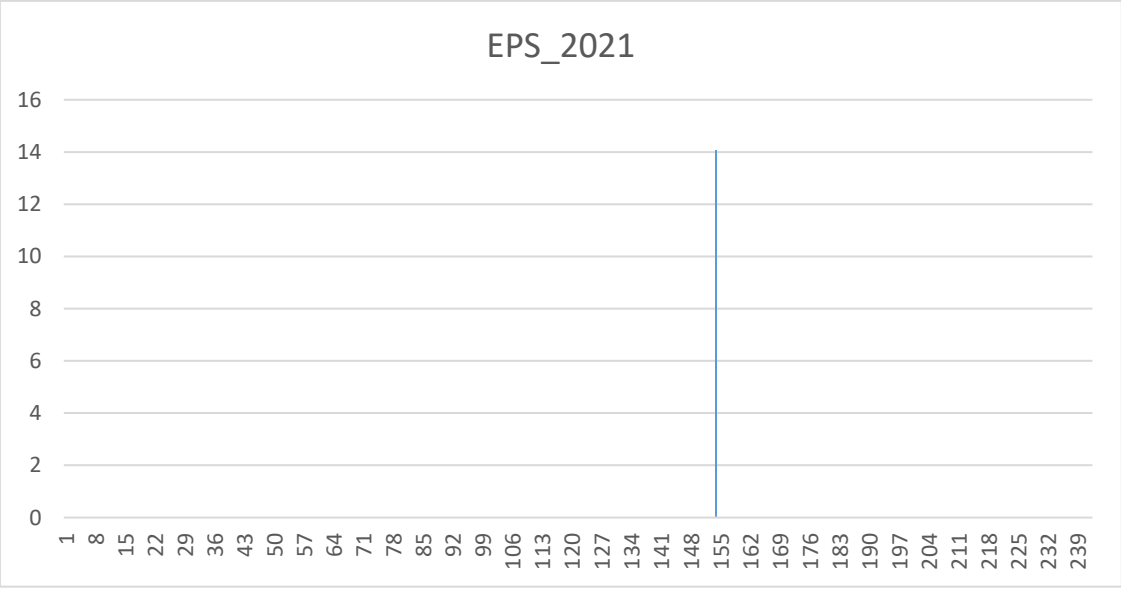


Figure 24. Earnings per share for the year 2021

The figure 25 below shows the earnings per share for the year 2022. The x-axis shows the company labels, and the y-axis shows the earnings per share for the given values. The highest values are with the company label 15 (Colrox co).

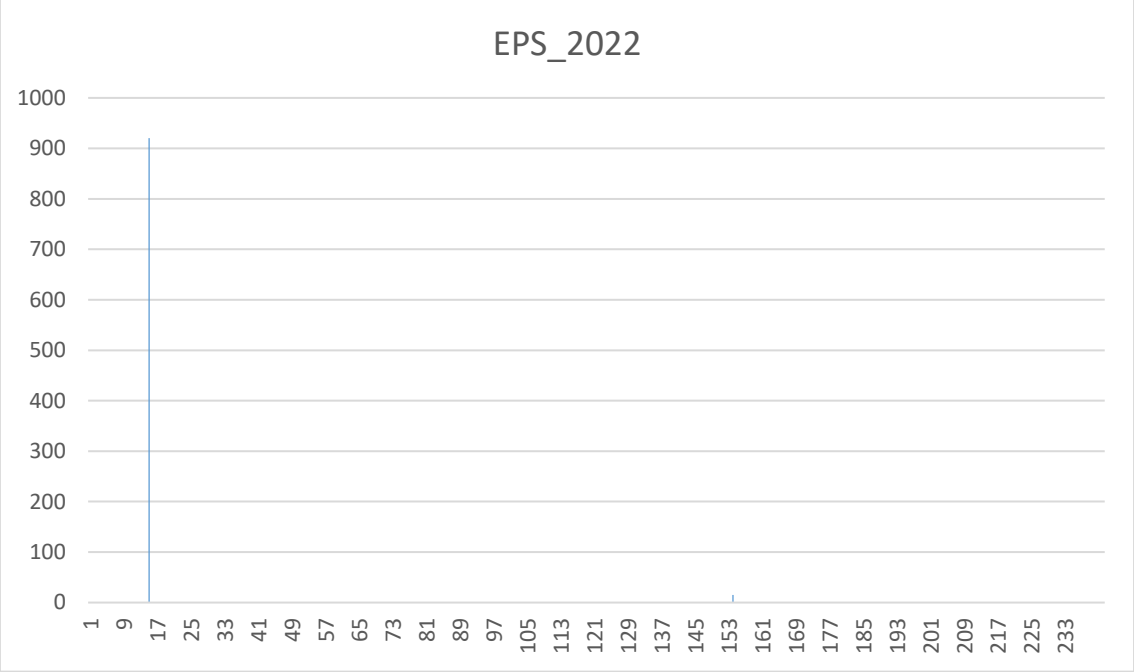


Figure 25. Earnings per share for the year 2022

The figure 26 below shows the earnings per share for the year 2023. The x-axis shows the earnings per share for the year 2023 whereas, the y-axis shows the earnings per share values. The highest values of earnings per share for the company is the 154 (Enerpac tool group corp).

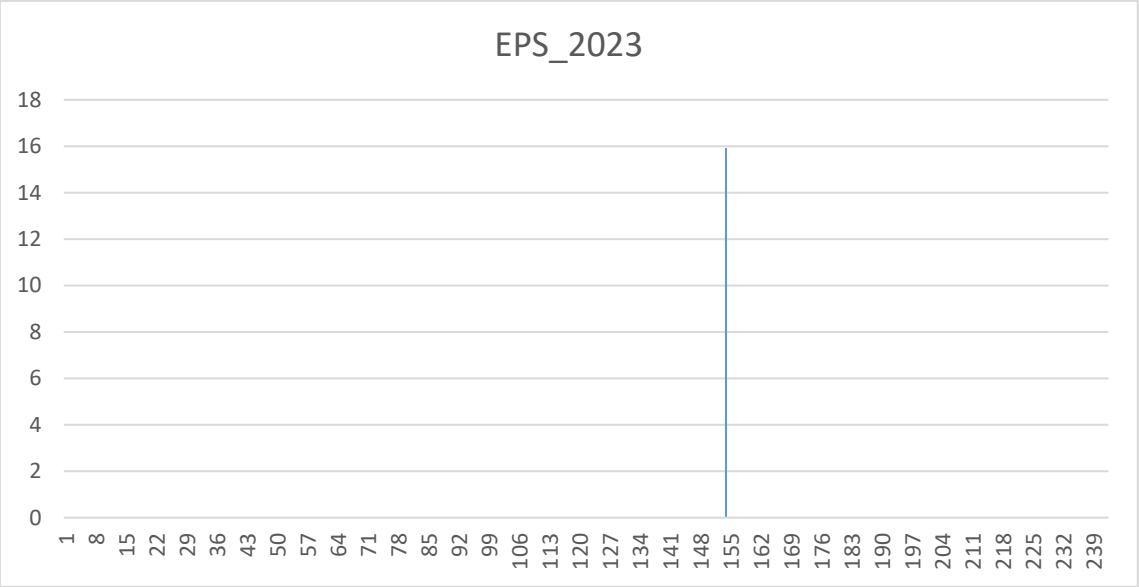


Figure 26. Earnings per share for the year 2023

The figure 27 below shows the earnings per share for the year 2024. The x-axis shows the company names, and the y-axis shows the values for the companies. The highest value for the company for the year 2022 is the 154 (Enerpac tool group corp).

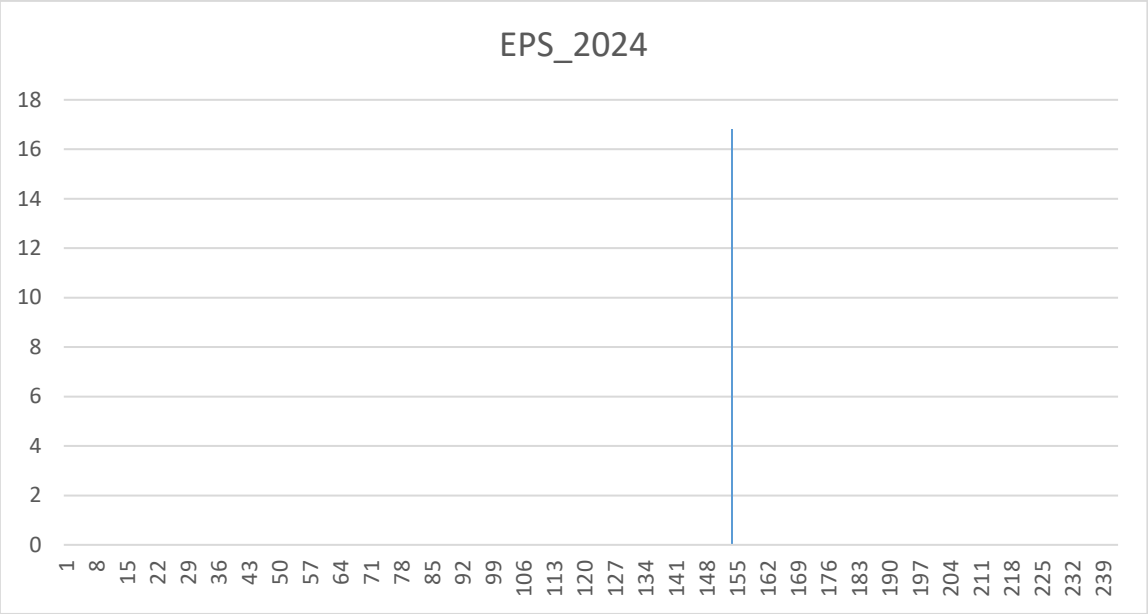


Figure 27. Earnings per share for the year 2024

The figure 28 below shows the earnings per share for the year 2025. The x-axis shows the company names and the y-axis the values for the earnings per share. The highest value of the company is the year 154 (Enerpac tool group corp).

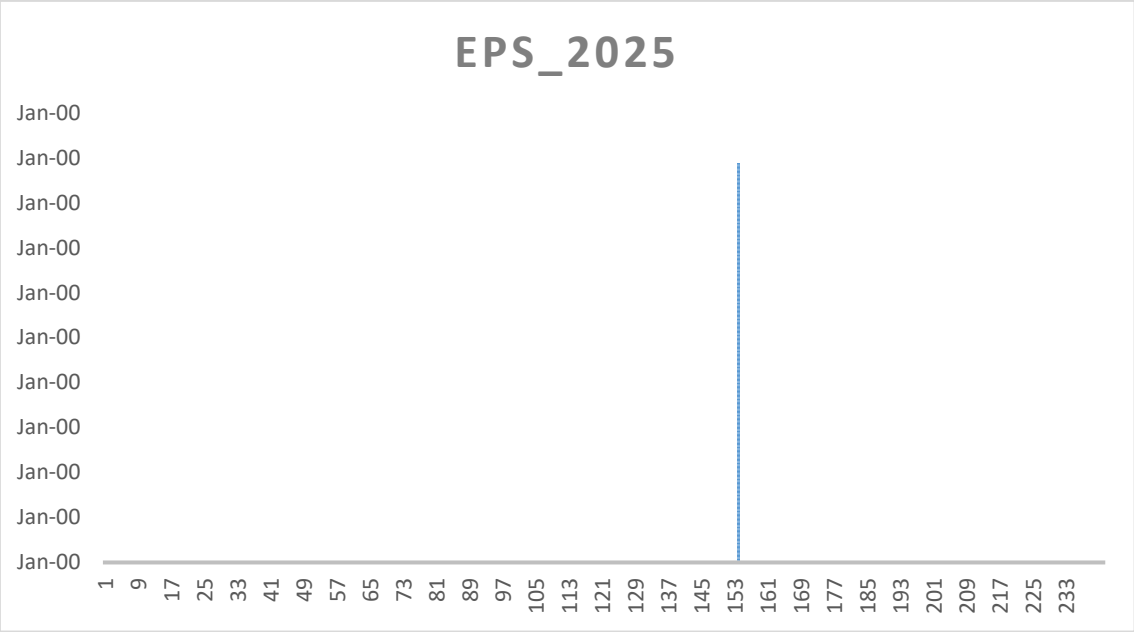


Figure 28. Earnings per share for the year 2025

6.5 Classification based on return on equity and earnings per share.

The classification of the companies was based on the 5% percentage rule. The classification based on return on equity was 50.5 percent. These are companies which have good performance. Similarly, 41.8% of companies are in the worst performance category. Neutral performance comprises 7.7% of the total companies. The figure 16 below shows the pie-chart for the performance measure based on ROE.

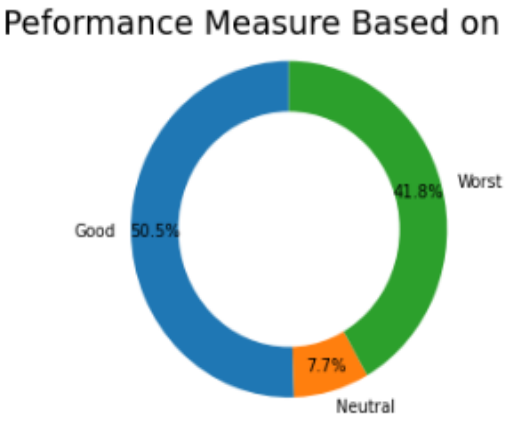
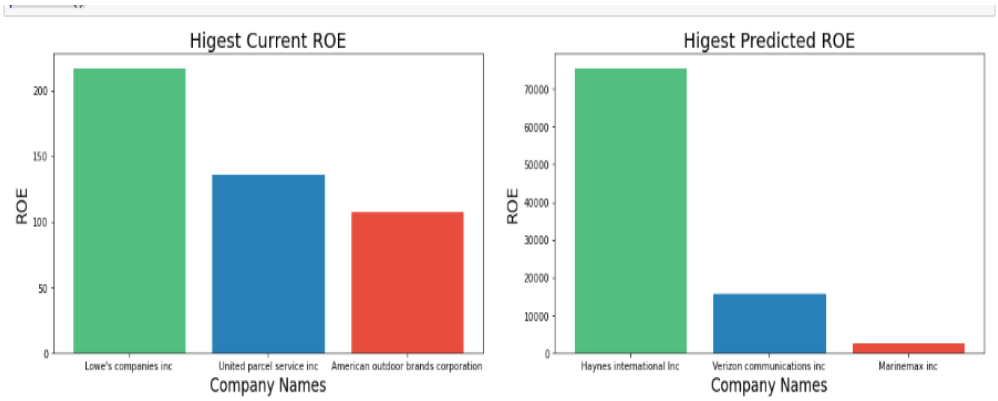


Figure 29. Performance measure based on return on equity.

The highest current ROE is Lowe's Companies Inc., United Parcels Service Inc. and American Outdoor Brands Corporation. The highest predicted companies are: - Haynes international Inc., Verizon Communications Inc. and Marinemax Inc. The Figure below shows the return on equity.



c

Figure 30. Highest current return on equity and predicted return on equity

The classification based on earnings per share shows the performance measure based on EPS. Good company performance is 51%. The worst company performance is 44.3%. The neutral figure is 4.6%. This can be seen in the figure below.

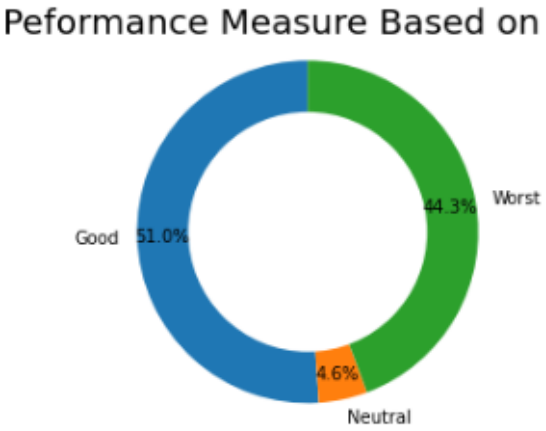


Figure 31. Performance measure based on earnings per share.

The highest performance-based classification for per earnings per share companies are Cardinal Eethanol LLC, Granite Falls Energy Inc and AutoZone. The highest predicted Eps are Farmer's Bro, AmerisourceBergen Crop and Meritor Inc.

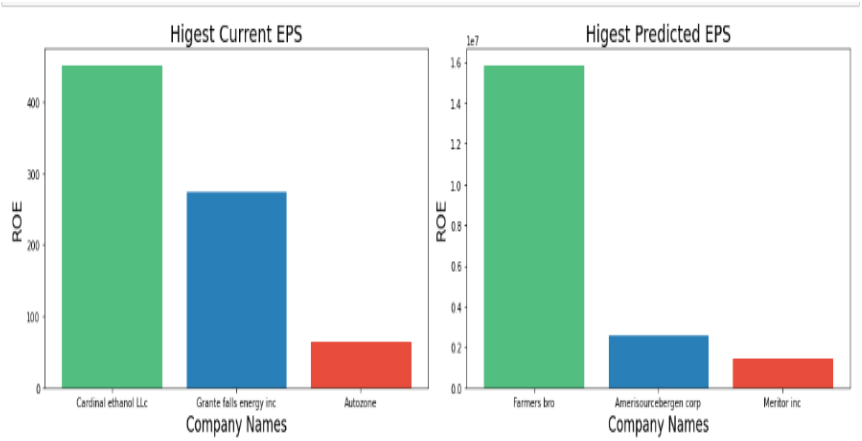


Figure 32. Highest current EPS and highest predicted EPS

7 IMPORTANCE OF RESULTS

The results show the company's performances, which were based on return on assets, return on equity, revenue, and earnings per share. It shows the future profitable or non-profitable situation of the company. Likewise, it is also useful for investors. The investors or share market analysts can make decisions about the investment opportunities based on the forecasting results. Any individual without knowledge of investment can use the code for finding a forecast or better investment opportunities.

The results can also be compared with profitability performance, market value performance, growth performance and employee performance. The values for ROE, EPS, and revenue can be compared with Table 1. Thus, comparing the values can produce the firm's measurements in respect to profitability, market value, growth, and employee performances. The purpose of the classification of the forecasting was to produce immediate information about the company's performance: - the company performance can be analysed immediately. The financial statement forecasting includes necessary information about income, equity etc. These values can be forecast to obtain the necessary forecasting results for the future. Machine learning is a new and emerging subject: - its application to the financial sectors can provide profitable outcomes. This thesis proves that its application can be profitable.

8 MANAGERIAL IMPLICATIONS

This study shows the results of 250 companies. The managers of the listed companies can check the companies regarding return on assets (ROA), return on equity (ROE), revenue, and earnings per share (EPS). The companies which have the highest ROA are Brooks international, Deere and Co International, Inc, Kennametal Corp, Netapp Corp and Nordson Corp. The companies which are in ROE are General motors, Grante Falls Energy Inc, Green Brier Companies Inc, Grief Inc, Griffon Inc, United Natural Foods Inc, United Parcel Services Inc, Until Corp, Universal Technical Institute, Uranium Energy Corp. The highest revenue companies are Amazon, America Art. Similarly, the highest earnings per share (EPS) are Charter Com, Chaveron Corporation, China HGS and Chiptole Mexicam Grills.

The company managers can then check the future 5 years forecasts of return on assets, return on equity, revenue and earnings per share. Likewise, they can also check the lowest values of the companies. The code provided in Appendix 1 and Appendix 2 can also be helpful for other companies which are not listed in this study. Forecasts can still be applied to new companies in future years. Table 1 in this study can also be used to compare the results. The conditions are suitable for profitability performance, market value performance, growth performance and employee performance. The highest ROA means high profitability performance and market value performance and vice-versa. Likewise, the highest ROE means high profitability performance. Similarly, high revenue means high employee satisfaction, and high earnings per share means high market value performance.

9 CONCLUSIONS

The aim of the thesis was to measure firm performances. In our study, the particular interest was on profitability performance, market value performance and employee satisfaction. The notable point is that large capital companies had good performances in all three measures. Similarly, another goal was to utilize a machine learning algorithm which was successfully implemented using the time-series forecasting method and long short-term memory model. Similarly, machine learning techniques were used for forecasting. This can be further used for investment purposes. The investors can easily check earnings per share (EPS), return on assets (ROA), return on equity (ROE), and revenue. New companies can be added using the same process mentioned in this study.

In the theoretical part of the thesis, forecasting methods and balance sheets were measured. Likewise, industrial classification systems were also measured. In the empirical research framework, the supervised learning methods and financial indicators were explained.

The process was implemented in six steps. The first step was for the conversion of group of data into individual company data. Then each model for the company was given a company label. The model was then used for testing and validation. 80 % of the data was used for training and 20% for validation. Juypter software was implemented for the application. Then the models were trained and results collected in separate Excel files.

The results show that most companies which were based on 5% ROE showed good performance category. 51% good performance. 41% neutral performance and 8% worst performance. Likewise, the forecast results based on return on equity, return on assets and revenue shows on average good results.

The advantages of this code are that, the methods can be useful for investors and analysts for investment purposes. Similarly, Excel files can be added and made bigger by

adding other companies' firm performance measurements. These measurements can help for profitable investment opportunities.

10 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

There are few limitations involved in the empirical part of the thesis. The time for completion of the research study was limited. Thus, the data availability of the companies was restricted. The reason for that is because of the time for completion of the thesis: - the time period for the study was short. Therefore, the data were limited to those companies whose information could easily be searched. The accuracy and prediction of the results could have been more extensive if more data variables for the companies had been available.

REFERENCES

Alpaydin, E. (2020). Introduction to machine learning (4thed). Cambridge, Massachu setts: MIT Press. Retrieved from https://books.google.fi/books?hl=en&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&dq=Ethem+Alpaydin,+E.+(2004).+Introduction+to+Machine+Learning.+Cabridge,+MA:+MIT+Press.&ots=F3RSaX4wCh&sig=4MpmMw51xrVsLVvQ3ukF70YKKCY&redir_esc=y#v=onepage&q=Ethem%20Alpaydin%2C%20E.%20(2004).%20Introduction%20to%20Machine%20Learning.%20Ca bridge%2C%20MA%3A%20MIT%20Press.&f=false

Amel, Z. A., Calliess, P.J., Kaise, D., & Roberts, S. (2020). Machine learning-based financial statement analysis, *Social Science Research Network*, 1-69. Retrieved from https://dx.doi.org/10.2139/ssm.3520684

Armstrong, J. S. (2001). Principles of forecasting: A handbook for researchers and practitioners. Kluwer Academic Publishers, 1-18. Retrieved from https://faculty.wharton.upenn.edu/wp-content/uploads/2012/04/PoF-Selecting-Forecasting-Methods.pdf

Breiman, L., Friedman, H. J., Olshen, A. R., & Stone, J. C. (1993). Classification and regression trees. 1-354.
Retrieved from https://books.google.fi/books?hl=en&lr=&id=MGlQDwAAQBAJ&oi=fnd&pg=PP1&ots=gBkhjLISfL&sig=bSSyuD9Dz_AIGJUfEKY6rixWR6A&redir_esc=y#v=onepage&q&f=false

Brownlee, J. (2018). A gentle introduction to exponential smoothing for time series forecasting in python. Retrieved from: https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/

Brownlee, J. (2020). Autoregression models for time series forecasting with python. Retrieved from https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/

Brownlee, J. (2020). How to develop LSTM models for time series forecasting. Retrieved from https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

Carl, W. (2008). *Survey of accounting*. Cincinnati: South-Western College Publication, *128–132*. *ISBN 978-0-324-65826-2*.

Chang, X., Gao, M., Wang, Y., & Hou, X. (2013). Seasonal autoregressive integrated moving average model for precipitation time series. Journal of mathematics and statistics, 8(4), 500-505. Retrieved from https://thescipub.com/pdf/jmssp.2012.500.505.pdf

Chen, Y.Y., Lin, H.Y., Kung, C.C., Chung, M.H., & Yen, H.I. (2019). Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes. *US National Library of Medicine National Institutes of Health, 19*(9), doi: 10.3390/s19092047

Fernando, J., & Kindness, D. (2020). Earnings per share-EPS definition. Retrieved from https://www.investopedia.com/terms/e/eps.asp

Fernando, J., & Mansa, J. (2020). Return on equity, what is return on equity (ROE)? Retrieved from: https://www.investopedia.com/terms/r/returnonequity.asp#what-is-return-on-equity-roe

Fernando, J., & Margaret, J. (2020). Balance sheet. Investopedia, Retrieved from: https://www.investopedia.com/terms/b/balancesheet.asp

Frank, K. R. (1989). Investment analysis and portfolio management (3rded). The Dryden Press.

Fuchs, K. (2017). Machine learning: Classification models. Retrieved from https://medium.com/fuzz/machine-learning-classification-models-3040f71e2529

Hanias, M., Curtis, P., & Thalassinos, E. (2012). Time series prediction with neural networks for the Athens Stock Exchange indicator. *European Research Studies Journal, 0*(2), 23-32. Retrieved from https://ideas.repec.org/a/ers/journl/vxvy2012i2p23-32.html

Hasanaj, P., & Kuqi, B. (2019). Analysis of financial statements. *Humanities and Social Science Research, 2*(2), 17. Retrieved from https://www.researchgate.net/publication/333880039_Analysis_of_Financial_Statements

Hayes, A. (2020). Moving average. Retrieved from https://www.investopedia.com/terms/m/movingaverage.asp

Hyndman, R. J., & Athanasopolpus, G. (2018). Forecasting: principles and practice (2nd ed). Otexts: Melbourne, Australia. Retrieved from https://otexts.com/fpp2/

Kia, N. A., Fathia, M., & Gholamian, R. M. (2012). Using MLP and RBF neural networks to Improve the Prediction Exchange Rate Time series with ARIMA. *International Journal of Information and Electronics Engineering, 2*(4), 543-546. Retrieved from https://www.researchgate.net/publication/290466117_Using_MLP_and_RBF_Neural_Networks_to_Improve_thePrediction_of_Exchange_Rate_Time_Series_with_ARIMA

Kotsiantis, S.B. (2007). Supervised machine learning: A review of classification techniques, University of Peloponnese, Greece, 31, 249-268. Retrieved from https://datajobs.com/data-science-repo/Supervised-Learning-[SB-Kotsiantis].pdf

Miyakawa, D., & Miyauchi, Y. (2017). Forecasting firm performance with machine learning: Evidence from Japanese firm-level data. *The Research Institute of Economy, Trade and Industry, 17*, 1-26. Retrieved from https://www.rieti.go.jp/jp/publications/dp/17e068.pdf

Murphy, B. C., & Margaret, J. (2020). Financial statements, Investopedia, Retrieved from: https://www.investopedia.com/terms/f/financial-statements.asp#what-are-financial-statements

Nyce, C. (2007). Predictive analytics White Paper. *American institute for Chartered Property Casually Underwrites/Insurance Institute of America,* 1-15. Retrieved from https://www.the-digital-insurer.com/wp-content/uploads/2013/12/78-Predictive-Modeling-White-Paper.pdf

Patrick, R. D., Ralph, N., Barry, J. E., & Susan, W. B. (2002). Interpretation and application of generally accepted accounting principles. John Wiley & Sons, Inc. New York.

Qiu, Y. X. (2007). On predictive models with company annual reports. “PhD (Doctor of Philosophy) thesis, University of lowa, Retrieved from http://doi.org/10.17077/etd.dw68q7gz

Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modelling. Retrieved from https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43905.pdf

Santos, J.B., & Brito, L.A.L. (2012). Toward a subjective measurement model for firm performance. *Brazilian Administration Review, 9* (6), 1-24. Retrieved from https://www.researchgate.net/publication/262467405

Timothy, J. G., & Joseph, D. A. (2003). Financial management principles & practice (3rd ed). Prentice Hall

Tuovila, A., & Kindness, D. (2020). Financial analysis form. Retrieved from: https://www.investopedia.com/terms/f/financialanalysis.asp#:~:text=Financial%20analysis%20is%20the%20process,to%20warrant%20a%20monetary%20investment.

Wallace, M. (2018). How financial forecasting benefits your business, Retrieved from: https://iammarketing.com/how-financial-forecasting-benefits-your-business/#:~:text=Financial%20forecasting%20helps%20you%20identify,you%20devise%20necessary%20avoidance%20strategies.

Wilson, A. (2019). A brief introduction to supervised learning. Retrieved from https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590

Xin, Y. Q., Srinivasan, P., & Street, N. (2006). Exploring the forecasting potential of company annual reports. *Annual Meeting of the American Society for Information Science and Technology (ASIST)*, 3-8. Retrieved from: http://eprints.rclis.org/8842/

APPENDICES

Appendix 1. Code for forecasting


```
{
 "cells": [
  {
   "cell_type": "markdown",
   "metadata": {},
   "source": [
      ]
  },
  {
   "cell_type": "markdown",
   "metadata": {},
   "source": [
    "### Run in the sequence, in class cell it will ask you for Comapny name and year. and it will give you\n",
    "### ROA, ROE, REVENUE, Both EPS of that given company for the year you will give"
   ]
  },
  {
   "cell_type": "markdown",
   "metadata": {},
   "source": [
    "# Imports"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 19,
   "metadata": {},
   "outputs": [],
   "source": [
    "import os\n",
    "import pandas as pd\n",
    "import re\n",
    "\n",
    "from pandas import ExcelWriter\n",
    "from pandas import ExcelFile\n",
    "\n",
    "\n",
    "from numpy import split\n",
    "from numpy import array\n",
    "\n",
```

```
 "from math import sqrt\n",
 "from sklearn.metrics import mean_squared_error\n",
 "from matplotlib import pyplot\n",
 "from keras.models import Sequential\n",
 "from keras.models import load_model\n",
 "from keras.layers import Dense\n",
 "from keras.layers import Flatten\n",
 "from keras.layers import LSTM\n",
 "import matplotlib.pyplot as plt\n",
 "from matplotlib.pyplot import figure"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Required Functions"
 ]
},
{
 "cell_type": "code",
 "execution_count": 20,
 "metadata": {},
 "outputs": [],
 "source": [
  "def split_dataset(data):\n",
  "    _80_percent = int((80/100)*len(data))\n",
  "    train, test = data[0:_80_percent], data[_80_percent:]\n",
  "    train = array (split (train, len(train)/2))\n",
  "    test = array (split (test, len(test)/2))\n",
  "    return train, test\n",
  "\n",
  "def to_supervised(train, n_input, n_out=1):\n",
  "    data = train.reshape((train.shape[0]*train.shape[1], train.shape[2]))\n",
  "    X, y = list(), list()\n",
  "    in_start = 0\n",
  "    for _ in range(len(data)):\n",
  "        in_end = in_start + n_input\n",
  "        out_end = in_end + n_out\n",
  "        if out_end <= len(data):\n",
  "            x_input = data[in_start:in_end, 0]\n",
  "            x_input = x_input.reshape((len(x_input), 1))\n",
  "            X.append(x_input)\n",
  "            y.append(data[in_end:out_end, 0])\n",
  "        in_start += 1\n",
```

```
   "    return array(X), array(y)\n",
   "\n",
   "\n",
   "\n",
   "def forecast (model, history, n_input):\n",
   "    data = array(history)\n",
   "    data = data.reshape((data.shape[0]*data.shape[1], data.shape[2]))\n",
   "    input_x = data[-n_input:, 0]\n",
   "    input_x = input_x.reshape((1, len(input_x), 1))\n",
   "    yhat = model.predict(input_x, verbose=0)\n",
   "    yhat = yhat[0]\n",
   "    return yhat\n",
   "\n",
   "def split_dataset_EPS(data):\n",
   "    train, test = data[0:3], data[3:]\n",
   "    train = array(split(train, len(train)/1))\n",
   "    test = array(split(test, len(test)/1))\n",
   "    return train, test"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "### Reading all avalibale company names so we can check given company name is valid or not "
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 21,
  "metadata": {},
  "outputs": [],
  "source": [
   "companies_names_ROA_ = os.listdir('Data_sets/company_transform_data_ROA/')\n",
   "companies_names_ROE = os.listdir('Data_sets/company_transform_data_ROE/')\n",
   "companies_names_EPS = os.listdir('Data_sets/company_transform_data_Basic_EPS/')"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
```

```
   "### Removing the extension from end "
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 22,
  "metadata": {},
  "outputs": [],
  "source": [
   "companies_names_ROA_REVENUE = [fn[:-4] for fn in companies_names_ROA_]\n",
   "companies_names_for_ROE =     [fn[:-4] for fn in companies_names_ROE]\n",
   "companies_names_for_EPS =     [fn[:-4] for fn in companies_names_EPS]"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# Revenue"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 23,
  "metadata": {},
  "outputs": [],
  "source": [
   "def find_REVENUE(file_name, year):\n",
   "          Year_list  =  [2010,  2011,  2012,  2013,  2014,  2015,  2016,  2017,  2018,
2019,2020,2021,2022,2023,2024,2025]\n",
   "    actual_values = []\n",
   "    predicted_values = []\n",
   "    n_input , n_out  = 2 , 2\n",
   "    \n",
   "    file_path = \"Data_sets/company_transform_data_Revenue/\"\n",
   "    file_path += file_name+'.csv' \n",
   "                    dataset     =     pd.read_csv(file_path,     infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
   "    \n",
   "    train, test = split_dataset(dataset.values)\n",
   "    \n",
   "    \n",
   "    model = load_model('Models/Revenue_model/'+file_name+'.hS')\n",
   "    \n",
   "    history = [x for x in train]\n",
```

```
   "    \n",
   "    predictions = []\n",
   "    for i in range(len(test)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "        predictions.append(yhat_sequence)\n",
   "        history.append(test[i, :]) \n",
   "\n",
   "    yhat_sequence = forecast(model, history, n_input)\n",
   "    predictions.append(yhat_sequence)\n",
   "    year_var = Year_list[10]\n",
   "    for i in range(int(len(Year_list[10:])/2)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "                        history.append(array((array([yhat_sequence[0],   dataset.Company[0],
year_var]),\n",
   "                          array([yhat_sequence[1], dataset.Company[0], year_var+1])))\n",
   "        year_var += 2\n",
   "    \n",
   "    history = array(history)\n",
   "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
   "        \n",
   "    index = Year_list.index(year)\n",
   "    print(\"Revenue     : {:>12}\".format(float(history[index][0])))\n",
   "    "
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# ROA"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 24,
  "metadata": {},
  "outputs": [],
  "source": [
   "def find_ROA(file_name, year):\n",
   "          Year_list  =  [2010,  2011,  2012,  2013,  2014,  2015,  2016,  2017,  2018,
2019,2020,2021,2022,2023,2024,2025]\n",
   "    actual_values = []\n",
   "    predicted_values = []\n",
   "    n_input , n_out  = 2 , 2\n",
   "    verbose, epochs, batch_size = 0, 699, 1\n",
```

```
   "    \n",
   "    file_path = \"Data_sets/company_transform_data_ROA/\"\n",
   "    file_path += file_name+'.csv' \n",
   "                    dataset    =    pd.read_csv(file_path,    infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
   "    \n",
   "    train, test = split_dataset(dataset.values)\n",
   "    \n",
   "    \n",
   "    model = load_model('Models/ROA_model/'+file_name+'.hS')\n",
   "    \n",
   "    history = [x for x in train]\n",
   "    \n",
   "    predictions = []\n",
   "    for i in range(len(test)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "        predictions.append(yhat_sequence)\n",
   "        history.append(test[i, :]) \n",
   "\n",
   "    yhat_sequence = forecast(model, history, n_input)\n",
   "    predictions.append(yhat_sequence)\n",
   "    year_var = Year_list[10]\n",
   "    for i in range(int(len(Year_list[10:])/2)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "                    history.append(array((array([yhat_sequence[0],   dataset.Company[0],
year_var]),\n",
   "                        array([yhat_sequence[1], dataset.Company[0], year_var+1])))\n",
   "        year_var += 2\n",
   "    \n",
   "    history = array(history)\n",
   "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
   "        \n",
   "    index = Year_list.index(year)\n",
   "    print(\"ROA        : {:>12}\".format(float(history[index][0])))\n",
   "    "
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# ROE"
  ]
 },
 {
```

```
  "cell_type": "code",
  "execution_count": 25,
  "metadata": {},
  "outputs": [],
  "source": [
   "def find_ROE(file_name, year):\n",
   "        Year_list = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018,
2019,2020,2021,2022,2023,2024,2025]\n",
   "    actual_values = []\n",
   "    predicted_values = []\n",
   "    n_input , n_out  = 2 , 2\n",
   "    \n",
   "    file_path = \"Data_sets/company_transform_data_ROE/\"\n",
   "    file_path += file_name+'.csv' \n",
   "                    dataset = pd.read_csv(file_path, infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
   "    \n",
   "    train, test = split_dataset(dataset.values)\n",
   "    \n",
   "    \n",
   "    model = load_model('Models/ROE_model/'+file_name+'.hS')\n",
   "    \n",
   "    history = [x for x in train]\n",
   "    \n",
   "    predictions = []\n",
   "    for i in range(len(test)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "        predictions.append(yhat_sequence)\n",
   "        history.append(test[i, :])\n",
   "\n",
   "    yhat_sequence = forecast(model, history, n_input)\n",
   "    predictions.append(yhat_sequence)\n",
   "    year_var = Year_list[10]\n",
   "    for i in range(int(len(Year_list[10:])/2)):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "                    history.append(array((array([yhat_sequence[0],   dataset.Company[0],
year_var]),\n",
   "                        array([yhat_sequence[1], dataset.Company[0], year_var+1])))\n",
   "        year_var += 2\n",
   "    \n",
   "    history = array(history)\n",
   "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
   "        \n",
   "    index = Year_list.index(year)\n",
   "    print(\"ROE        : {:>12}\".format(float(history[index][0])))"
```

```
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Basic EPS"
 ]
},
{
 "cell_type": "code",
 "execution_count": 26,
 "metadata": {},
 "outputs": [],
 "source": [
  "def find_Basic_EPS(file_name, year):\n",
  "    Year_list = [2015, 2016, 2017, 2018, 2019,2020,2021,2022,2023,2024,2025]\n",
  "    actual_values = []\n",
  "    predicted_values = []\n",
  "    n_input , n_out  = 1 , 1\n",
  "    verbose, epochs, batch_size = 0, 700, 2\n",
  "    file_path = 'Data_sets/company_transform_data_Basic_EPS/'\n",
  "    file_path += file_name+'.csv' \n",
  "                    dataset     =     pd.read_csv(file_path,     infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
  "    \n",
  "    train, test = split_dataset_EPS(dataset.values)\n",
  "    \n",
  "    model = load_model('Models/B_EPS_model/'+file_name+'.hS')\n",
  "    \n",
  "    history = [x for x in train]\n",
  "    \n",
  "    predictions = []\n",
  "    for i in range(len(test)):\n",
  "        yhat_sequence = forecast(model, history, n_input)\n",
  "        predictions.append(yhat_sequence)\n",
  "        history.append(test[i, :]) \n",
 "\n",
  "    yhat_sequence = forecast(model, history, n_input)\n",
  "    predictions.append(yhat_sequence)\n",
  "    \n",
  "    \n",
  "    year_var = Year_list[5]\n",
  "    for i in range(len(Year_list[5:])):\n",
  "        yhat_sequence = forecast(model, history, n_input)\n",
```

```
   "                              history.append(array([array([yhat_sequence[0],dataset.Com-
pany[0],year_var])]))\n",
   "       year_var += 1\n",
   "       \n",
   "    index = Year_list.index(year)\n",
   "    print(\"Basic EPS   : {:>12}\".format(float(history[index][0][0])))    "
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# Diluted_EPS"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 27,
  "metadata": {},
  "outputs": [],
  "source": [
   "def find_Diluted_EPS(file_name, year):\n",
   "    Year_list = [2015, 2016, 2017, 2018, 2019,2020,2021,2022,2023,2024,2025]\n",
   "    actual_values = []\n",
   "    predicted_values = []\n",
   "    n_input , n_out  = 1 , 1\n",
   "    verbose, epochs, batch_size = 0, 700, 2\n",
   "    file_path = 'Data_sets/company_transform_data_Diluted_EPS/'\n",
   "    file_path += file_name+'.csv' \n",
   "                    dataset     =     pd.read_csv(file_path,     infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
   "    \n",
   "    train, test = split_dataset_EPS(dataset.values)\n",
   "    \n",
   "    model = load_model('Models/D_EPS_model/'+file_name+'.hS')\n",
   "    \n",
   "    history = [x for x in train]\n",
   "    \n",
   "    predictions = []\n",
   "    for i in range(len(test)):\n",
   "       yhat_sequence = forecast(model, history, n_input)\n",
   "       predictions.append(yhat_sequence)\n",
   "       history.append(test[i, :]) \n",
   "\n",
   "    yhat_sequence = forecast(model, history, n_input)\n",
```

```
   "    predictions.append(yhat_sequence)\n",
   "    \n",
   "    \n",
   "    year_var = Year_list[5]\n",
   "    for i in range(len(Year_list[5:])):\n",
   "        yhat_sequence = forecast(model, history, n_input)\n",
   "                                    history.append(array([array([yhat_sequence[0],dataset.Com-
pany[0],year_var])]))\n",
   "        year_var += 1\n",
   "        \n",
   "    index = Year_list.index(year)\n",
   "    print(\"Diluted EPS : {:>12}\".format(float(history[index][0][0])))"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# Main Code run this after running all of the above cells"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "### This is will take 2 inputs  \n",
   "   ###  1 : Name of company\n",
   "   ###  2 : Year (Year must be between 2020 to 2025) because it is predicting next 6
years(2020 to 2025)\n",
   "### Out put of this will be ROA, Revenue, ROE, Diluted EPS, Basic EPS of given com-
pany for given year"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 28,
  "metadata": {},
  "outputs": [
   {
    "name": "stdout",
    "output_type": "stream",
    "text": [
     "Enter the Company : Alico inc\n",
     "Enter the Year : 2025\n",
     "Year : 2025\n",
```

```
    "ROA        : 1.6915656328201294\n",
    "Revenue    : 127349.703125\n",
    "ROE        : 7.735899448394775\n",
    "Diluted EPS : 0.4321373701095581\n",
    "Basic EPS   : 0.44920438528060913\n"
   ]
  }
 ],
 "source": [
  "file_name = input(\"Enter the Company : \")\n",
  "year = int(input(\"Enter the Year : \"))\n",
  "print(\"Year : {}\".format(year))\n",
  "if file_name in companies_names_ROA_REVENUE:\n",
  "    find_ROA(file_name,year)\n",
  "    find_REVENUE(file_name,year)\n",
  "if file_name in companies_names_for_ROE:\n",
  "    find_ROE(file_name,year)\n",
  "if file_name in companies_names_for_EPS:\n",
  "    find_Diluted_EPS(file_name,year)\n",
  "    find_Basic_EPS(file_name,year)"
 ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": []
 }
],
"metadata": {
 "kernelspec": {
  "display_name": "neural_networks",
  "language": "python",
  "name": "neural_netwroks"
 },
 "language_info": {
  "codemirror_mode": {
   "name": "ipython",
   "version": 3
  },
  "file_extension": ".py",
  "mimetype": "text/x-python",
  "name": "python",
  "nbconvert_exporter": "python",
```

```
   "pygments_lexer": "ipython3",
   "version": "3.7.6"
  }
 },
 "nbformat": 4,
 "nbformat_minor": 4
}
```

Appendix 2: Code for classification


```
{
 "cells": [
  {
   "cell_type": "markdown",
   "metadata": {},
   "source": [
    "# imports"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 1,
   "metadata": {},
   "outputs": [
    {
     "name": "stderr",
     "output_type": "stream",
     "text": [
      "Using TensorFlow backend.\n"
     ]
    }
   ],
   "source": [
    "import os\n",
    "import pandas as pd\n",
    "import re\n",
    "\n",
    "from pandas import ExcelWriter\n",
    "from pandas import ExcelFile\n",
    "\n",
    "\n",
    "from numpy import split\n",
    "from numpy import array\n",
    "\n",
    "from math import sqrt\n",
    "from sklearn.metrics import mean_squared_error\n",
    "from matplotlib import pyplot\n",
    "from keras.models import Sequential\n",
    "from keras.models import load_model\n",
    "from keras.layers import Dense\n",
    "from keras.layers import Flatten\n",
    "from keras.layers import LSTM\n",
    "import matplotlib.pyplot as plt\n",
```

```
 "from matplotlib.pyplot import figure"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Required Functions"
 ]
},
{
 "cell_type": "code",
 "execution_count": 2,
 "metadata": {},
 "outputs": [],
 "source": [
  "def split_dataset(data):\n",
  "    _80_percent = int((80/100)*len(data))\n",
  "    train, test = data[0:_80_percent], data[_80_percent:]\n",
  "    train = array(split(train, len(train)/2))\n",
  "    test = array(split(test, len(test)/2))\n",
  "    return train, test\n",
  "\n",
  "def to_supervised(train, n_input, n_out=1):\n",
  "    data = train.reshape((train.shape[0]*train.shape[1], train.shape[2]))\n",
  "    X, y = list(), list()\n",
  "    in_start = 0\n",
  "    for _ in range(len(data)):\n",
  "        in_end = in_start + n_input\n",
  "        out_end = in_end + n_out\n",
  "        if out_end <= len(data):\n",
  "            x_input = data[in_start:in_end, 0]\n",
  "            x_input = x_input.reshape((len(x_input), 1))\n",
  "            X.append(x_input)\n",
  "            y.append(data[in_end:out_end, 0])\n",
  "        in_start += 1\n",
  "    return array(X), array(y)\n",
  "\n",
  "\n",
  "def forecast(model, history, n_input):\n",
  "    data = array(history)\n",
  "    data = data.reshape((data.shape[0]*data.shape[1], data.shape[2]))\n",
  "    input_x = data[-n_input:, 0]\n",
  "    input_x = input_x.reshape((1, len(input_x), 1))\n",
  "    yhat = model.predict(input_x, verbose=0)\n",
```

```
"    yhat = yhat[0]\n",
"    return yhat\n",
"\n",
"def split_dataset_EPS(data):\n",
"    train, test = data[0:3], data[3:]\n",
"    train = array(split(train, len(train)/1))\n",
"    test = array(split(test, len(test)/1))\n",
"    return train, test\n",
"\n",
"def find_ROE(file_name, year):\n",
"        Year_list = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019,2020,2021,2022,2023,2024,2025]\n",
"    actual_values = []\n",
"    predicted_values = []\n",
"    n_input , n_out  = 2 , 2\n",
"    \n",
"    file_path = \"Data_sets/company_transform_data_ROE/\"\n",
"    file_path += file_name+'.csv' \n",
"                dataset = pd.read_csv(file_path, infer_datetime_format=True, parse_dates=['datetime'], index_col=['datetime'])\n",
"    \n",
"    train, test = split_dataset(dataset.values)\n",
"    \n",
"    \n",
"    model = load_model('Models/ROE_model/'+file_name+'.hS')\n",
"    \n",
"    history = [x for x in train]\n",
"    \n",
"    predictions = []\n",
"    for i in range(len(test)):\n",
"      yhat_sequence = forecast(model, history, n_input)\n",
"      predictions.append(yhat_sequence)\n",
"      history.append(test[i, :])\n",
"\n",
"    yhat_sequence = forecast(model, history, n_input)\n",
"    predictions.append(yhat_sequence)\n",
"    year_var = Year_list[10]\n",
"    for i in range(int(len(Year_list[10:])/2)):\n",
"      yhat_sequence = forecast(model, history, n_input)\n",
"                history.append(array((array([yhat_sequence[0], dataset.Company[0], year_var]),\n",
"                    array([yhat_sequence[1], dataset.Company[0], year_var+1]))))\n",
"      year_var += 2\n",
"    \n",
"    history = array(history)\n",
```

```
  "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
  "    return history\n",
  "      \n",
  "#    index = Year_list.index(year)\n",
  "#    print(\"ROE       : {:>12}\".format(float(history[index][0])))\n",
  "\n",
  "def find_REVENUE(file_name, year):\n",
  "        Year_list = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018,
2019,2020,2021,2022,2023,2024,2025]\n",
  "    actual_values = []\n",
  "    predicted_values = []\n",
  "    n_input , n_out  = 2 , 2\n",
  "    \n",
  "    file_path = \"Data_sets/company_transform_data_Revenue/\"\n",
  "    file_path += file_name+'.csv' \n",
  "                dataset = pd.read_csv(file_path, infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
  "    \n",
  "    train, test = split_dataset(dataset.values)\n",
  "    \n",
  "    \n",
  "    model = load_model('Models/Revenue_model/'+file_name+'.hS')\n",
  "    \n",
  "    history = [x for x in train]\n",
  "    \n",
  "    predictions = []\n",
  "    for i in range(len(test)):\n",
  "      yhat_sequence = forecast(model, history, n_input)\n",
  "      predictions.append(yhat_sequence)\n",
  "      history.append(test[i, :]) \n",
  "\n",
  "    yhat_sequence = forecast(model, history, n_input)\n",
  "    predictions.append(yhat_sequence)\n",
  "    year_var = Year_list[10]\n",
  "    for i in range(int(len(Year_list[10:])/2)):\n",
  "      yhat_sequence = forecast(model, history, n_input)\n",
  "                history.append(array((array([yhat_sequence[0],  dataset.Company[0],
year_var]),\n",
  "                    array([yhat_sequence[1], dataset.Company[0], year_var+1])))\n",
  "      year_var += 2\n",
  "    \n",
  "    history = array(history)\n",
  "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
  "    \n",
  "    return history\n",
```

```
  "\n",
  "def find_Basic_EPS(file_name, year):\n",
  "    Year_list = [2015, 2016, 2017, 2018, 2019,2020,2021,2022,2023,2024,2025]\n",
  "    actual_values = []\n",
  "    predicted_values = []\n",
  "    n_input , n_out  = 1 , 1\n",
  "    verbose, epochs, batch_size = 0, 700, 2\n",
  "    file_path = 'Data_sets/company_transform_data_Basic_EPS/'\n",
  "    file_path += file_name+'.csv' \n",
  "                    dataset     =     pd.read_csv(file_path,     infer_datetime_format=True,
parse_dates=['datetime'], index_col=['datetime'])\n",
  "    \n",
  "    train, test = split_dataset_EPS(dataset.values)\n",
  "    \n",
  "    model = load_model('Models/B_EPS_model/'+file_name+'.hS')\n",
  "    \n",
  "    history = [x for x in train]\n",
  "    \n",
  "    predictions = []\n",
  "    for i in range(len(test)):\n",
  "        yhat_sequence = forecast(model, history, n_input)\n",
  "        predictions.append(yhat_sequence)\n",
  "        history.append(test[i, :]) \n",
  "\n",
  "    yhat_sequence = forecast(model, history, n_input)\n",
  "    predictions.append(yhat_sequence)\n",
  "    \n",
  "    \n",
  "    year_var = Year_list[5]\n",
  "    for i in range(len(Year_list[5:])):\n",
  "        yhat_sequence = forecast(model, history, n_input)\n",
  "                                    history.append(array([array([yhat_sequence[0],dataset.Com-
pany[0],year_var])]))\n",
  "        year_var += 1\n",
  "        \n",
  "    history = array(history)\n",
  "    history = history.reshape((history.shape[0]*history.shape[1], history.shape[2]))\n",
  "    \n",
  "    return history\n",
  "    \n",
  "#    index = Year_list.index(year)\n",
  "#    print(\"Basic EPS   : {:>12}\".format(float(history[index][0][0])))    \n",
  "    \n",
  "\n",
  "def find_average(numbers):\n",
```

```
   "    total_len = len(numbers)\n",
   "    sum_of_all = 0\n",
   "    for num in numbers:\n",
   "        sum_of_all += num\n",
   "        \n",
   "    return sum_of_all/total_len    \n",
   "\n",
   "def find_class(curent_value, new_value):\n",
   "    percentage = ((new_value-curent_value)/abs(curent_value))*100\n",
   "    if percentage >= 5:\n",
   "        return 'Good'\n",
   "    elif percentage <= -5:\n",
   "        return 'Worst'\n",
   "    else:\n",
   "        return 'Neutral'"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# lists of all companies"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": 3,
  "metadata": {},
  "outputs": [],
  "source": [
   "companies_names_ROA_ = os.listdir('Data_sets/company_transform_data_ROA/')\n",
   "companies_names_ROE = os.listdir('Data_sets/company_transform_data_ROE/')\n",
   "companies_names_EPS = os.listdir('Data_sets/company_transform_data_Basic_EPS/')"
  ]
 },
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "# Removeing the extension from names of files"
  ]
 },
 {
```

```
 "cell_type": "code",
 "execution_count": 4,
 "metadata": {},
 "outputs": [],
 "source": [
  "companies_names_ROA_REVENUE = [fn[:-4] for fn in companies_names_ROA_]\n",
  "companies_names_for_ROE =     [fn[:-4] for fn in companies_names_ROE]\n",
  "companies_names_for_EPS =     [fn[:-4] for fn in companies_names_EPS]"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Finding same companies names in all three categories"
 ]
},
{
 "cell_type": "code",
 "execution_count": 5,
 "metadata": {},
 "outputs": [],
 "source": [
  "same_companies = set(companies_names_ROA_REVENUE).intersection(set(compa-
nies_names_for_ROE))\n",
  "same_companies              =              set(same_companies).intersection(set(compa-
nies_names_for_EPS))"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Calulate performance Based on ROE"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "performance_of_companies = [find_ROE(x,2020) for x in same_companies]"
 ]
},
```

```
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "performances_classes = []\n",
   "Expected_values = []\n",
   "current_values = []\n",
   "\n",
   "for Pc in performance_of_companies:\n",
   "    current_value = Pc[9][0]\n",
   "    new_number    = find_average(Pc[10:,0])\n",
   "    performances_class = find_class(current_value, new_number)\n",
   "    \n",
   "    performances_classes.append(performances_class)\n",
   "    Expected_values.append(new_number)\n",
   "    current_values.append(current_value)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "per_based_ROE_dic = {\"Company\": list(same_companies), \"Current ROE\": cur-
rent_values, \"Expected ROE\":Expected_values, \"Performance\":performances_clas-
ses}"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "ROE_Based_Peformance = pd.DataFrame(data=per_based_ROE_dic)\n",
   "ROE_Based_Peformance.head(10)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
```

```
  "outputs": [],
  "source": [
   "ROE_Based_Peformance.to_csv('Performance Based On ROE.csv', index=False)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {
   "scrolled": true
  },
  "outputs": [],
  "source": [
   "performance_count = ROE_Based_Peformance.groupby('Performance').count()\n",
   "performance_count"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "plt.pie(performance_count['Company'], labels= performance_count.index, au-
topct='%1.1f%%', startangle=90, pctdistance=0.85)\n",
   "\n",
   "#draw circle\n",
   "centre_circle = plt.Circle((0,0),0.70,fc='white')\n",
   "fig = plt.gcf()\n",
   "fig.gca().add_artist(centre_circle)\n",
   "# Equal aspect ratio ensures that pie is drawn as a circle\n",
   "plt.axis('equal')  \n",
   "plt.title('Peformance Measure Based on ROE', fontsize= 20)\n",
   "\n",
   "plt.show()"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "higest_values_and_index_c = sorted([(x,i) for (i,x) in enumerate(current_values)], re-
verse=True)[:3]\n",
```

```
   "higest_values_and_index_e = sorted([(x,i) for (i,x) in enumerate(Expected_values)], reverse=True)[:3]"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "roe_values_c = []\n",
   "roe_values_index_c = []\n",
   "for i in higest_values_and_index_c:\n",
   "    roe_values_c.append(i[0])\n",
   "    roe_values_index_c.append(i[1])\n",
   "    \n",
   "roe_values_e = []\n",
   "roe_values_index_e = []\n",
   "for i in higest_values_and_index_e:\n",
   "    roe_values_e.append(i[0])\n",
   "    roe_values_index_e.append(i[1])"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "comname_c = [list(same_companies)[index] for index in roe_values_index_c]\n",
   "comname_e = [list(same_companies)[index] for index in roe_values_index_e]"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "fig = plt.figure(figsize=(17,5))\n",
   "\n",
   "ax1 = fig.add_subplot(1,2,1)\n",
   "barlist1 = ax1.bar(comname_c,roe_values_c)\n",
   "barlist1[0].set_color('#52BE80')\n",
   "barlist1[1].set_color('#2980B9')\n",
```

```
 "barlist1[2].set_color('#E74C3C')\n",
 "ax1.set_xlabel('Company Names', fontsize=17)\n",
 "ax1.set_ylabel('ROE', fontsize=17)\n",
 "#ax.set_ylim(-30,60)\n",
 "ax1.set_title(\"Higest Current ROE\", fontsize=20)\n",
 "\n",
 "ax2 = fig.add_subplot(1,2,2)\n",
 "barlist2 = ax2.bar(comname_e,roe_values_e)\n",
 "barlist2[0].set_color('#52BE80')\n",
 "barlist2[1].set_color('#2980B9')\n",
 "barlist2[2].set_color('#E74C3C')\n",
 "ax2.set_xlabel('Company Names', fontsize=17)\n",
 "ax2.set_ylabel('ROE', fontsize=17)\n",
 "#ax.set_ylim(-30,60)\n",
 "ax2.set_title(\"Higest Predicted ROE\", fontsize=20)\n",
 "    \n",
 "plt.tight_layout()\n",
 "plt.show()"
]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "## Calulate performance Based on EPS"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "performance_of_companies = [find_Basic_EPS(x,2020) for x in same_companies]"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "performances_classes = []\n",
  "Expected_values = []\n",
  "current_values = []\n",
```

```
   "\n",
   "for Pc in performance_of_companies:\n",
   "    current_value = Pc[4][0]\n",
   "    new_number    = find_average(Pc[4:,0])\n",
   "    performances_class = find_class(current_value, new_number)\n",
   "    \n",
   "    performances_classes.append(performances_class)\n",
   "    Expected_values.append(new_number)\n",
   "    current_values.append(current_value)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "per_based_ROE_dic = {\"Company\": list(same_companies), \"Current EPS\": current_values, \"Expected EPS\":Expected_values, \"Performance\":performances_classes}"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "ROE_Based_Peformance = pd.DataFrame(data=per_based_ROE_dic)\n",
   "ROE_Based_Peformance.head(10)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "ROE_Based_Peformance.to_csv('Performance Based On EPS.csv', index=False)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
```

```
  "outputs": [],
  "source": [
   "performance_count = ROE_Based_Peformance.groupby('Performance').count()\n",
   "performance_count"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "plt.pie(performance_count['Company'], labels= performance_count.index, au-
topct='%1.1f%%', startangle=90, pctdistance=0.85)\n",
   "\n",
   "#draw circle\n",
   "centre_circle = plt.Circle((0,0),0.70,fc='white')\n",
   "fig = plt.gcf()\n",
   "fig.gca().add_artist(centre_circle)\n",
   "# Equal aspect ratio ensures that pie is drawn as a circle\n",
   "plt.axis('equal')  \n",
   "plt.title('Peformance Measure Based on EPS', fontsize= 20)\n",
   "\n",
   "plt.show()"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "higest_values_and_index_c = sorted([(x,i) for (i,x) in enumerate(current_values)], re-
verse=True)[:3]\n",
   "higest_values_and_index_e = sorted([(x,i) for (i,x) in enumerate(Expected_values)],
reverse=True)[:3]"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "roe_values_c = []\n",
```

```
 "roe_values_index_c = []\n",
 "for i in higest_values_and_index_c:\n",
 "    roe_values_c.append(i[0])\n",
 "    roe_values_index_c.append(i[1])\n",
 "    \n",
 "roe_values_e = []\n",
 "roe_values_index_e = []\n",
 "for i in higest_values_and_index_e:\n",
 "    roe_values_e.append(i[0])\n",
 "    roe_values_index_e.append(i[1])"
]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "comname_c = [list(same_companies)[index] for index in roe_values_index_c]\n",
  "comname_e = [list(same_companies)[index] for index in roe_values_index_e]"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "fig = plt.figure(figsize=(17,5))\n",
  "\n",
  "ax1 = fig.add_subplot(1,2,1)\n",
  "barlist1 = ax1.bar(comname_c,roe_values_c)\n",
  "barlist1[0].set_color('#52BE80')\n",
  "barlist1[1].set_color('#2980B9')\n",
  "barlist1[2].set_color('#E74C3C')\n",
  "ax1.set_xlabel('Company Names', fontsize=17)\n",
  "ax1.set_ylabel('ROE', fontsize=17)\n",
  "#ax.set_ylim(-30,60)\n",
  "ax1.set_title(\"Higest Current EPS\", fontsize=20)\n",
  "\n",
  "ax2 = fig.add_subplot(1,2,2)\n",
  "barlist2 = ax2.bar(comname_e,roe_values_e)\n",
  "barlist2[0].set_color('#52BE80')\n",
  "barlist2[1].set_color('#2980B9')\n",
  "barlist2[2].set_color('#E74C3C')\n",
```

```
  "ax2.set_xlabel('Company Names', fontsize=17)\n",
  "ax2.set_ylabel('ROE', fontsize=17)\n",
  "#ax.set_ylim(-30,60)\n",
  "ax2.set_title(\"Higest Predicted EPS\", fontsize=20)\n",
  "    \n",
  "plt.tight_layout()\n",
  "plt.show()"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Calulate performance Based on Revenue"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "performance_of_companies = [find_REVENUE(x,2020) for x in same_companies]"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "performances_classes = []\n",
  "Expected_values = []\n",
  "current_values = []\n",
  "\n",
  "for Pc in performance_of_companies:\n",
  "    current_value = Pc[9][0]\n",
  "    new_number    = find_average(Pc[10:,0])\n",
  "    performances_class = find_class(current_value, new_number)\n",
  "    \n",
  "    performances_classes.append(performances_class)\n",
  "    Expected_values.append(new_number)\n",
  "    current_values.append(current_value)"
 ]
},
```

```
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "per_based_ROE_dic = {\"Company\": list(same_companies), \"Current REVENUE\": current_values, \"Expected REVENUE\":Expected_values, \"Performance\":performances_classes}"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "ROE_Based_Peformance = pd.DataFrame(data=per_based_ROE_dic)\n",
   "ROE_Based_Peformance.head(10)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "ROE_Based_Peformance.to_csv('Performance Based On REVENUE.csv', index=False)"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "performance_count = ROE_Based_Peformance.groupby('Performance').count()\n",
   "performance_count"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
```

```
  "outputs": [],
  "source": [
   "plt.pie(performance_count['Company'], labels= performance_count.index, au-
topct='%1.1f%%', startangle=90, pctdistance=0.85)\n",
   "\n",
   "#draw circle\n",
   "centre_circle = plt.Circle((0,0),0.70,fc='white')\n",
   "fig = plt.gcf()\n",
   "fig.gca().add_artist(centre_circle)\n",
   "# Equal aspect ratio ensures that pie is drawn as a circle\n",
   "plt.axis('equal')  \n",
   "plt.title('Peformance Measure Based on Revenue', fontsize= 20)\n",
   "\n",
   "plt.show()"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "higest_values_and_index_c = sorted([(x,i) for (i,x) in enumerate(current_values)], re-
verse=True)[:3]\n",
   "higest_values_and_index_e = sorted([(x,i) for (i,x) in enumerate(Expected_values)],
reverse=True)[:3]"
  ]
 },
 {
  "cell_type": "code",
  "execution_count": null,
  "metadata": {},
  "outputs": [],
  "source": [
   "roe_values_c = []\n",
   "roe_values_index_c = []\n",
   "for i in higest_values_and_index_c:\n",
   "    roe_values_c.append(i[0])\n",
   "    roe_values_index_c.append(i[1])\n",
   "    \n",
   "roe_values_e = []\n",
   "roe_values_index_e = []\n",
   "for i in higest_values_and_index_e:\n",
   "    roe_values_e.append(i[0])\n",
   "    roe_values_index_e.append(i[1])"
```

```
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "comname_c = [list(same_companies)[index] for index in roe_values_index_c]\n",
  "comname_e = [list(same_companies)[index] for index in roe_values_index_e]"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  "fig = plt.figure(figsize=(17,5))\n",
  "\n",
  "ax1 = fig.add_subplot(1,2,1)\n",
  "barlist1 = ax1.bar(comname_c,roe_values_c)\n",
  "barlist1[0].set_color('#52BE80')\n",
  "barlist1[1].set_color('#2980B9')\n",
  "barlist1[2].set_color('#E74C3C')\n",
  "ax1.set_xlabel('Company Names', fontsize=17)\n",
  "ax1.set_ylabel('ROE', fontsize=17)\n",
  "#ax.set_ylim(-30,60)\n",
  "ax1.set_title(\"Higest Current Revenue\", fontsize=20)\n",
  "\n",
  "ax2 = fig.add_subplot(1,2,2)\n",
  "barlist2 = ax2.bar(comname_e,roe_values_e)\n",
  "barlist2[0].set_color('#52BE80')\n",
  "barlist2[1].set_color('#2980B9')\n",
  "barlist2[2].set_color('#E74C3C')\n",
  "ax2.set_xlabel('Company Names', fontsize=17)\n",
  "ax2.set_ylabel('ROE', fontsize=17)\n",
  "#ax.set_ylim(-30,60)\n",
  "ax2.set_title(\"Higest Predicted Revenue\", fontsize=20)\n",
  "    \n",
  "plt.tight_layout()\n",
  "plt.show()"
 ]
},
{
```

```
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": []
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": []
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": [
  """\n",
  "    [[-1.23000000e-01  2.01000000e+02  2.01000000e+03]\n",
  "     [ 2.30200000e+00  2.01000000e+02  2.01100000e+03]\n",
  "     [ 9.50100000e+00  2.01000000e+02  2.01200000e+03]\n",
  "     [ 6.00100000e+00  2.01000000e+02  2.01300000e+03]\n",
  "     [-3.38970000e+01  2.01000000e+02  2.01400000e+03]\n",
  "     [-6.41200000e+01  2.01000000e+02  2.01500000e+03]\n",
  "     [-3.55080000e+01  2.01000000e+02  2.01600000e+03]\n",
  "     [ 2.06600000e+00  2.01000000e+02  2.01700000e+03]\n",
  "     [-4.13790000e+01  2.01000000e+02  2.01800000e+03]\n",
  "     [-2.23080000e+01  2.01000000e+02  2.01900000e+03]\n",
  "     [ 9.15831470e+00  2.01000000e+02  2.02000000e+03]\n",
  "     [ 8.56932163e+00  2.01000000e+02  2.02100000e+03]\n",
  "     [-2.90708084e+01  2.01000000e+02  2.02200000e+03]\n",
  "     [-7.27347031e+01  2.01000000e+02  2.02300000e+03]\n",
  "     [ 8.83634806e-01  2.01000000e+02  2.02400000e+03]\n",
  "     [-2.37545753e+00  2.01000000e+02  2.02500000e+03]]\n",
  """
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "metadata": {},
 "outputs": [],
 "source": []
```

```
  }
 ],
 "metadata": {
  "kernelspec": {
   "display_name": "Python 3",
   "language": "python",
   "name": "python3"
  },
  "language_info": {
   "codemirror_mode": {
    "name": "ipython",
    "version": 3
   },
   "file_extension": ".py",
   "mimetype": "text/x-python",
   "name": "python",
   "nbconvert_exporter": "python",
   "pygments_lexer": "ipython3",
   "version": "3.7.6"
  }
 },
 "nbformat": 4,
 "nbformat_minor": 4
}
```

Appendix 3: The list of companies

S. No.	Sector	Agriculture
1.	Agriculture	Alico_inc
2.	Agriculture	Cal_Maine_foods_inc
3.	Agriculture	S_and _W_Seed_Co
4.	Automotive dealers	America's car_mart inc.
5.	Automotive dealers	Autozone
6.	Automotive dealers	Marinemax_inc
7.	Chemicals and ailed	Advaxis_inc
8.	Chemicals and ailed	Alexion_Pharmaceuticals_Inc.
9.	Chemicals and ailed	Applied_DNA_Sciencs_inc
10.	Chemicals and ailed	Avid_Bioservices_inc
11.	Chemicals and ailed	Biogen_Inc
12.	Chemicals and ailed	Bio-techne_ corporation
13.	Chemicals and ailed	Cabot_corp
14.	Chemicals and ailed	Cabot_microelectronics_corp
15.	Chemicals and ailed	Cardinal_ethanol_LLc
16.	Chemicals and ailed	Clorox_co

17.	Chemicals and ailed	Estee_lauder_companies_Inc.
18.	Chemicals and ailed	Grante_falls_energy_inc
19.	Chemicals and ailed	H.B. Fuller_company
20.	Chemicals and ailed	Innovation_pharmaceuticals_inc.
21.	Chemicals and ailed	Landec_corp
22.	Chemicals and ailed	Lannett_company_inc.
23.	Chemicals and ailed	Mei_Pharma
24.	Chemicals and ailed	Meridian_bio_sciences inc,
25.	Chemicals and ailed	Allance_Resources_partners
26.	Chemicals and ailed	Arch_Coal_Inc.
27.	Coal mining	Amrnstrong_in
28.	Coal mining	Cnx_resources_corporation
29.	Coal mining	Foresight_energy_LP
30.	Coal mining	Charter_communications
31.	Coal mining	Idt_corp
32.	Communications services	Comcast_corporation
33.	Communications services	Sirius_xm_holdings_inc.

34.	Communications services	T-mobile_US_inc.
35.	Communications services	Walt_disney_company
36.	Communications services	Viacom
37.	Communications services	D.R.Horton_Inc
38.	Communications services	Jacobs_Engineering_group_inc
39.	Construction	Quanex_building_products_corporations
40.	Construction	Chiptole_mexican_grill_inc
41.	Eating and drinking places	Cracker_barrel_old_country_store_inc.
42.	Eating and drinking places	Darden_resturants_inc
43.	Eating and drinking places	Jack_in_the_box_inc
44.	Eating and drinking places	Starbucks
45.	Eating and drinking places	Adtalem_global_educational_ inc.
46.	Educational services	Franklin_covey_co

47.	Educational services	K12_inc
48.	Educational services	Universal_Technical_institute
49.	Educational services	Atmos_Energy_Corp
50.	Electricity and sanitary services	MS energy corp
51.	Electricity and sanitary services	DTE_energy_corp
52.	Electricity and sanitary services	National_Fuel_gas_company
53.	Electricity and sanitary services	New_jersey_resources_corp
54.	Electricity and sanitary services	Pure_cycle_corporation
55.	Electricity and sanitary services	RGC_resources_inc
56.	Electricity and sanitary services	Sharps_complaince_corp
57.	Electricity and sanitary services	Spire_inc
58.	Electricity and sanitary services	UGI_corp
59.	Electricity and sanitary services	Unitil_corp
60.	Electricity and sanitary services	Acuity_brands_Inc

61.	Electronics	Advanced_micro_devices
62.	Electronics	Analog_devices_INC
63.	Electronics	Avaya_holdings_corp.
64.	Electronics	Aviat_networks_inc
65.	Electronics	Brady_corp
66.	Electronics	Ciena_corp
67.	Electronics	Cisco_Systems_Inc
68.	Electronics	Clearfield_Inc
69.	Electronics	Comtech_telecommunications_corp
70.	Electronics	Cree_inc
71.	Electronics	Emcore_corp
72.	Electronics	Extreme_networks_inc.
73.	Electronics	Fuelcell_energy_inc.
74.	Electronics	Intel_corp
75.	Electronics	Jabil_inc
76.	Electronics	Kimball_international_inc
77.	Electronics	Kla_corporation
78.	Electronics	Kulicke_and_soffa_industries_inc

79.	Electronics	Lantroix_inc
80.	Electronics	Lightpath_technologies_inc
81.	Electronics	Lsi_industries_inc
82.	Electronics	Maxim_integrated_products_inc
83.	Electronics	Maxlinear_inc
84.	Electronics	Methode_electronics_inc
85.	Electronics	Micron_technology_inc
86.	Electronics	Napco_security_technologies_inc
87.	Electronics	Photronics_inc
88.	Electronics	Plexus_corp
89.	Electronics and computer equiptments except computer equiptment	Acuity_brands_Inc
90.	Electronics and computer equiptments except computer equiptment	Analog_devices_INC
91.	Electronics and computer equiptments except computer equiptment	Avaya_holdings_corp.

92.	Electronics and computer equiptments except computer equiptment	Aviat_networks_inc
93.	Electronics and computer equiptments except computer equiptment	Dakrotonics_inc
94.	Electronics and computer equiptments except computer equiptment	Dobly_labrotories_inc
95.	Electronics and computer equiptments except computer equiptment	Emcore_corp
96.	Electronics and computer equiptments except computer equiptment	Extreme_networks_inc.
97.	Electronics and computer equiptments except computer equiptment	Fuelcell_energy_inc.
98.	Electronics and computer equiptments except computer equiptment	IEC_electronics_corp
99.	Electronics and computer equiptments except computer equiptment	Aecom

100.	Engineering, accounting and	Arrowhead_pharmaceuticals_Inc
101.	Engineering, accounting and	Paychex_inc
102.	Engineering, accounting and	Tetra_tech_inc
103.	Engineering, accounting and	Brown_forman_corp
104.	Food and kindered products	Campbell_soup_co
105.	Food and kindered products	Central_garden_and_pet_company
106.	Food and kindered products	Farmers_bro
107.	Food and kindered products	General_Mills_Inc
108.	Food and kindered products	Hain_celestial_group_inc
109.	Food and kindered products	Hormel_foods_corp
110.	Food and kindered products	J_and_J_Snack_Food_corp
111.	Food and kindered products	Lancaster_colony_corp
112.	Food and kindered products	National_beverage_corp

113.	Food and kindered products	Sanderson_farms_inc
114.	Food and kindered products	Sanfilippo_john
115.	Food and kindered products	Tyson_foods
116.	Food and kindered products	Pope_resources_Ltd
117.	Forestry	Bassett_furniture_industries_inc
118.	Furniture	Ethan_allen_interriors_inc
119.	Furniture	Felexsteel_industries_inc
120.	Furniture	La_z_boy_inc
121.	Furniture	Miller_herman_inc
122.	Furniture	Casey_general_store
123.	General food stores	Ingles_market
124.	General food stores	Village_super_market
125.	General food stores	Costco_wholesale_corp
126.	General merchandise	Pricesmart_Inc
127.	General merchandise	Tuesday_morning_corp.
128.	General merchandise	Mercury_systems_Inc

129.	Industrial, computer equiptment	Applied_materials_Inc
130.	Industrial, computer equiptment	Lam_research_corp
131.	Industrial, computer equiptment	Briggs_and_straton_Inc
132.	Industrial, computer equiptment	Amech_systems
133.	Industrial, computer equiptment	Apple_Inc
134.	Industrial, computer equiptment	Brooks_automation_Inc
135.	Industrial, computer equiptment	Deere_and_co
136.	Industrial, computer equiptment	Digi_international_inc
137.	Industrial, computer equiptment	Kennametal_Inc
138.	Industrial, computer equiptment	Netapp_inc
139.	Industrial, computer equiptment	Nordson_Corp
140.	Industrial, computer equiptment	Strandex_international_corp
141.	Industrial, computer equiptment	Super_micro_computers_Inc

142.	Industrial, computer equiptment	Toro_company
143.	Industrial, computer equiptment	Western_digital_corp
144.	Industrial, computer equiptment	Albany
145.	Textile industries	Boot_Barn_Holdings_Inc.
146.	Textile industries	Crown_crafts_Inc
147.	Textile industries	Culp_Inc.
148.	Textile industries	Dixie
149.	Textile industries	Interface_Inc
150.	Textile industries	Tapestry_Inc
151.	Textile industries	American_outdoor_brands_corporation
152.	Leather and leather related	Carpenter_technology_corp
153.	Leather and leather related	Commercial_Metal
154.	Leather and leather related	Enerpac_tool_group_corp
155.	Leather and leather related	General_finance_corporation

156.	Leather and leather related	Grief_Inc
157.	Leather and leather related	Griffon_corp
158.	Leather and leather related	Haynes_international_Inc
159.	Leather and leather related	Insteel_industries_Inc
160.	Leather and leather related	Mueller_water_products_inc
161.	Leather and leather related	Schnitzer_steel_industris_Inc
162.	Leather and leather related	Shiloh_industries_inc
163.	Leather and leather related	Worthington_industries_Inc
164.	Leather and leather related	Uranium_energy_corp
165.	Leather and leather related	Netflix_Inc.
166.	Metal mining	Evolution_petroleum_corporation
167.	Metal mining	Helmerich_and_payne_Inc
168.	Oil and gas extraction companies	Matrix_service
169.	Oil and gas extraction companies	Panhanle_oil_and_gas_Inc
170.	Oil and gas extraction companies	Valero_energy_corp
171.	Oil and gas extraction companies	Phillips_66

172.	Oil and gas extraction companies	Marathon_petroleum_corporation
173.	Oil and gas	Westrock_company
174.	Oil and gas	Kansas_city
175.	Oil and gas	Norfolk_southern_corp
176.	Oil and gas	China_HGS
177.	Paper industries	J_W_Mays_Inc
178.	Rail transportation	Terreno_realty_corp
179.	Rail transportation	1-800_Flowers_com_inc
180.	Real estate	Amazon_com_inc
181.	Real estate	Barnes_&_Noble_inc
182.	Real estate	Ebay_inc
183.	Retail	Sally_beauty_holdings_inc
184.	Retail	Star_group
185.	Retail	Suburban_propane_partners
186.	Retail	Walgreens_boot_ alliance_inc
187.	Retail	Nike_inc.
188.	Retail	Eaton_Vance_corp

189.	Retail	Intercontinental_exchange_inc
190.	Retail	Value_line_Inc
191.	Rubber	Aar_corp
192.	Security and commodity brokers, dealers, exchanges and services	Boeing_company
193.	Security and commodity brokers, dealers, exchanges and services	Navistar_corporation
194.	Security and commodity brokers, dealers, exchanges and services	Ford_motors
195.	Transporation equiptments	General_motors_company
196.	Transporation equiptments	Greenbrier_companies_inc
197.	Transporation equiptments	Heico_corp
198.	Transporation equiptments	Lear_corp
199.	Transporation equiptments	Meritor_inc
200.	Transporation equiptments	Moog_inc
201.	Transporation equiptments	Oshkosh_corporation

202.	Transporation equiptments	Strattec_security_corp
203.	Transporation equiptments	Thor_industries_inc
204.	Transporation equiptments	Transdigm_group_incorporation
205.	Transporation equiptments	Winnebago_industries_inc
206.	Transporation equiptments	Amerisourcebergen_corp
207.	Transporation equiptments	Applied_industrial_technologies_inc
208.	Transporation equiptments	Avnet_inc
209.	Transporation equiptments	Beacon_roofing_supply inc
210.	Wholesales	Calavo_growers_inc
211.	Wholesales	Cardinal_health_Inc
212.	Wholesales	Fastenal_company
213.	Wholesales	Ferrellgas_partners
214.	Wholesales	MSC_industrial_direct_co
215.	Wholesales	Patterson_companies
216.	Wholesales	Scansource_Inc

217.	Wholesales	Synnex_corporation
218.	Wholesales	Sysco_corp
219.	Wholesales	United_Natural_foods inc
220.	Wholesales	AT_ and_T
221.	Wholesales	Verizon_communications_inc
222.	Wholesales	Caterpillar_inc
223.	Telecommunications	cvs_health_corporation
224.	Telecommunications	Facebook_inc
225.	Pharmaceuticals	Fedex_corp
226.	Heavy equiptments	General_electric_company
227.	Retail health care	Home_depot_inc
228.	Social media	Hp_inc
229.	Post delivery	International_business_machines_corp
230.	Conglomerate	johnson_and_johnson
231.	Retaliing	Pfizer_inc
232.	Computer hardware	Kroger_co
233.	Pharmaceuticals	Lowe's_companies_inc
234.	Retail	Mckesson_corporation

235.	Retail	Microsoft_corporation
236.	Health care	Pepsico_Inc
237.	Software development	Procter_and_gamble_co
238.	Food processing	Target_corp
239.	Pharmaceuticals	United_parcel_service_inc.