**A PROJECT REPORT**

*Submitted by*

# GANESH KUMAR M (711619104010) SIVASAMY E (711619104046)

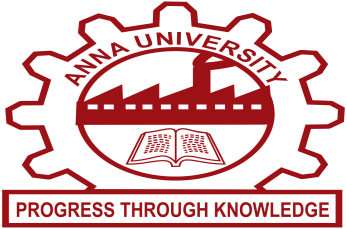
# SUBBU KUTTY B (711619104049)

*In partial fulfilment for the award of the degree* *of*

**BACHELOR OF ENGINEERING**

*in*

# COMPUTER SCIENCE AND ENGINEERING



**KATHIR COLLEGE OF ENGINEERIING**

**“WISDOM TREE”, NEELAMBUR, COIMBATORE – 641 062**

**ANNA UNIVERSITY: CHENNAI – 600 025 JUNE 2022**

**ANNA UNIVERSITY: CHENNAI 600 025**

# BONAFIDE CERTIFICATE

Certified that this project report “**SENTIMENT ANALYSIS IN TEXT MINING”** is the bonafide work of

## “GANESH KUMAR M (711619104010)

### SIVASAMY E (711619104046) and

**SUBBUKUTTY B (711619104049)”** who carried out the project under my supervision.

## SIGNATURE SIGNATURE Dr. S. JAGADEESH KUMAR, M.E.,Ph.D., Ms .KAVITHA , ME.,(Ph.D.) HEAD OF THE DEPARTMENT, PROJECT GUIDE,

### Professor and Head, Assistant Professor,

Department of CSE, Department of CSE,

Kathir College of Engineering Kathir College of Engineering

Coimbatore – 641 062 Coimbatore – 641 062

External viva voce held on \_\_\_\_\_\_\_\_\_\_

**INTERNAL EXAMINER EXTERNAL EXAMINER**

# ACKNOWLEDGEMENT

We express our immense gratitude to **Thiru E.S. KATHIR, Chairman, Kathir Institutions**, Coimbatore for giving us an opportunity to study in their prestigious institution and to take up the project in Partial fulfillment of the Regulation for the B.E Program.

We would like to express our deepest gratitude to **Thirumathi LAVANYA KATHIR, Secretary, Kathir Institutions**, Coimbatore for the soul support in our studies.

We would bound to express our gratitude to **Dr. R.UDAIYAKUMAR, M.E.,Ph.D., Principal, Kathir College of Engineering**, Coimbatore for their permission and constant encouragement throughout our course.

It is great pleasure to express our sincere and wholehearted gratitude to Professor **Dr. S. JAGADEESHKUMAR, M.E.,Ph.D., Head of the Computer Science and Engineering Department, and Dr.AYYAVOO MITHILA, M.E.,(Ph.D.,),Project coordinator and associate professor of Computer Science and Engineering Department**, for their constant suggestion and encouragement in the project work.

We also express our heartfelt thanks to **Ms.KAVITHA, M.E.,(Ph.D.,), Assistant Professor and Project Guide, Department of Computer Science and Engineering** for being supportive throughout the tenture of our project.

We also thank all our Faculty Members and Non Teaching Staff Members of Department of Computer Science and Engineering and our Lovable Parents and Friends who contribute many suitable ways for achieving final results.

# ABSTRACT

Financial distress prediction is a crucial area of research in the field of finance, and the use of machine learning techniques has gained considerable attention in recent years. In this study, the primary focus is to predict financial distress of Indian automobile sector using machine learning techniques. The study considers data from the companies that are performing well and the companies that has faced bankruptcy or insolvency. The study focuses on three types of machine learning models: random forest, decision tree and XGBoost. By using machine learning algorithms to predict financial distress, investors can make more informed investment decisions and policymakers can design better policies to mitigate the impacts of financial distress on the sector. Overall, this study demonstrates the potential of machine learning techniques in predicting financial distress in the automobile sector, and provides valuable insights for future research in this area.

**LIST OF FIGURES**

## FIGURE NO TITLE PAGE NO

1.1 Machine Learning 4

1.2 Python 5

3.1 Block diagram of Proposed system 15

4.1 System Architecture 17

4.2 Use Case Diagram 18

4.3 Data Flow Diagram 19

5.1 Matplot Graph for the Dataset 23

5.2 Seaborn Graph for Keywords of Joy 24

5.3 Wordcloud with Keywords of Joy 26

5.4 Neattext applied Dataset 27

5.5 Eli 5 Representation 27

6.1 Module 29

7.1 Performance and Evaluation 30

**LIST OF ABBREVIATIONS**

**CMS -** Content Management System

**HTML -** Hyper Text Markup Language

**XML -** Extensible Markup Language

**JSON -** JavaScript Object Notation

**FTP -** File Transfer Protocol

**IMAP -** Internet Mobile Access Protocol

**NLP -** Natural Language Processing

**MAT -** Mental Arithmetic Task

### MIST - Montreal Imaging Stress Task

**HPA -** Hypothalamic Pituitary Adrenal

**OCR -** Optical Character Recognition

**CSV -** Comma Separated Values

**GPU -** Graphical Processing Unit

**TPU -** Text Processing Utility

**COLAB -** Google Colaboratory

**NUMPY -** Numerical Python

**API -** Application Programming Interface

## TABLE OF CONTENTS

## CHAPTER NO TITLE PAGE NO

**ABSTRACT iii**

## LIST OF FIGURES iv LIST OF ABBREVIATIONS v 1 INTRODUCTION 1

1.1 Motivation of the Project

1.2 Objective of the Project

1.3 Machine Learning

1.4 Python

1.5 Natural Language Processing

**2 LITERATURE SURVEY**  **7**

## 3 SYSTEM ANALYSIS 12

3.1 Problem Definition

3.2 Existing System

3.2.1 Methodology

3.2.2 Disadvantages of Existing System

3.3 Proposed System

3.3.1 Advantages of Proposed System

## 4 SYSTEM DESIGN 17

4.1 System Architecture

4.2 Use Case Diagram

4.3 Data Flow Diagram

## 5 SYSTEM DESCRIPTION 20

5.1 Source Platform

5.2 Pandas in Python

5.3 Libraries used in python

1. **MODULE 28**

6.1 Module Description

1. **PERFORMANCE AND EVALUATIONS**   **30**

1. **CONCLUSION AND FUTURE WORK**   **31**

**CHAPTER 1**

# INTRODUCTION

## Motivation of the Project:

## In the current scenario, financial management has much greater significance. The systematic financial management begins with proper collection of funds and ends with effective use of those funds by generating ROI with enhancing market value of the share. The proper financial management leads to longer survival of the entity. The longer survivals depend on year-on-year performance of the entity. If the entity’s financial structure is not taken proper care, the unit grows weak leading toward the end of corporate failure. Bankruptcy is a legal status by which a person or an entity declares that it is unable to pay off its debts to creditors. Financial distress is a stage where the entity faces before declaring bankruptcy. Financial distress is the condition where the entity or the individual cannot be able to generate sufficient revenues or income, making it unable to meet the financial obligations. In India, Bankruptcy is governed by the Insolvency and Bankruptcy Code, 2016(IBC), which replaced the earlier laws on bankruptcy and insolvency in India. The IBC is a legal framework in India to insolve resolution of corporate entities, partnerships and individuals.



## 1.2 Objective of the Project

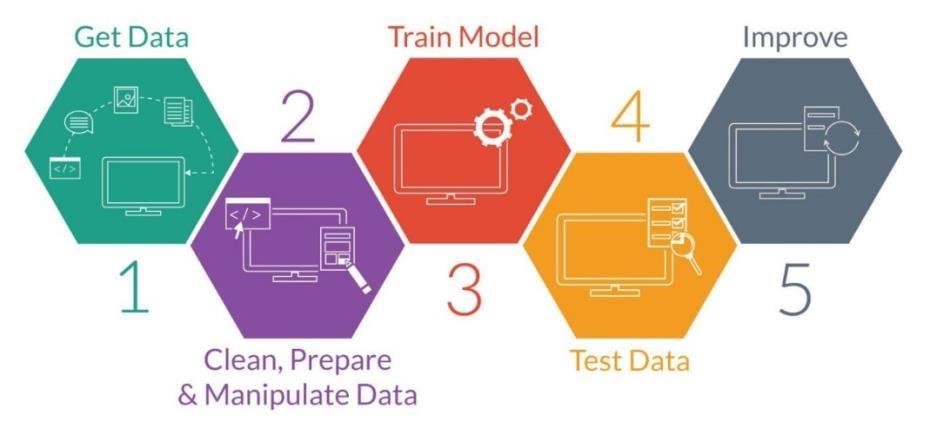
The Main Objective of this Project was to detect the bankruptcy of the company before it gone through one, It also helps the company to get more cautious about their moves in the market and the financial loss they are going to face if there is any.

By Using Machine Learning Algorithms, and Python, we can detect the input (liquidity ratio, solvency ratio, return of equity given by the user. Bankruptcy algorithm is used to find the stability of the company in the market. Also known as financial distress management , bankruptcy analysis  [a](https://www.telusinternational.com/articles/sentiment-analysis-101)llows to determine whether a result of the company is positive, negative or neutral by performing certain calculations by the given input by the user in order to find the stability of the company in the market. The main purpose of financial distress prediction is to analyse the current state of the company in the market used to maintain the stability of the company in the market.

There are several algorithms to find the bankruptcy of the company but there is not any accuracy prediction technique in order to improve our result or to find the accuracy of the result the result was just simply made. The actions are taken accordingly, so that leads to actions without any choices, that was the main reason for building the model to find the accuracy of the model

## 1.3 Machine Learning

Machine learning (ML) is the process of using mathematical models of data to help a computer learn without direct instruction. It’s considered a subset of artificial intelligence (AI). Machine learning uses algorithms to identify patterns within data and those patterns are then used to create a data model that can make predictions.



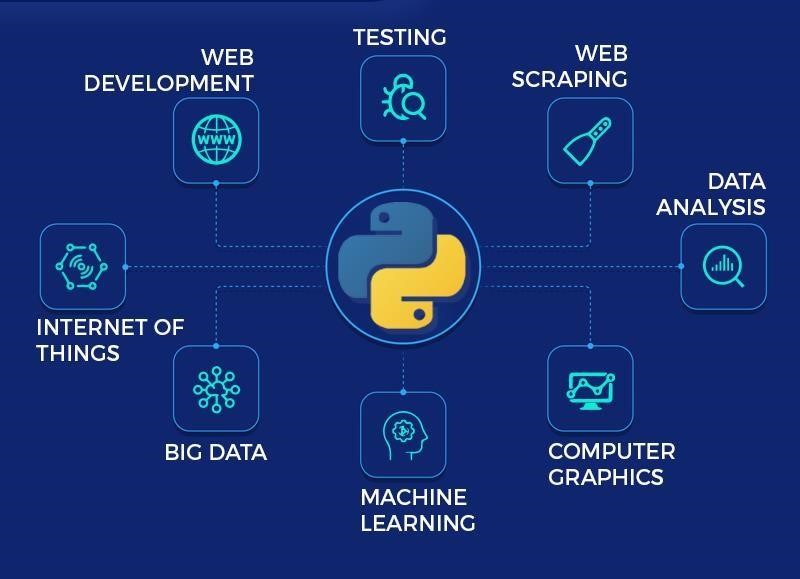
## 1.1 MACHINE LEARNING

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

## 1.4 Python

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry. Python Programming Language is very well suited for Beginners, also for experienced programmers with other programming languages like C++ and Java.



## 1.2 PYTHON

Python is used in many application domains such as Web and Internet Development, Education, Software Development etc.

For Web development choices such as Frameworks such as Django and Pyramid.

Micro-frameworks such as Flask and Bottle.

Advanced content management systems such as Plone and django CMS.

Python's standard library supports many Internet protocols:

HTML and XML

JSON

E-mail processing.

Support for FTP, IMAP, and other Internet protocols.

Easy-to-use socket interface.

## 

**CHAPTER 2**

**LITERATURE SURVEY**

**Corporate Bankruptcy Prediction: An Approach Towards Better Corporate World - Talha Et all (2020):**

10,000 North American companies were taken for the study and five years financial data before bankruptcy was taken. Different ML techniques such as Logistic Regression, Decision Tree, Random Forest, ANN, CNN, boosting etc. has been used. Random forest has given higher accuracy.

The area of corporate bankruptcy prediction attains high economic importance, as it affects many stakeholders. The prediction of corporate bankruptcy has been extensively studied in economics, accounting and decision sciences over the past two decades. The corporate bankruptcy prediction has been a matter of talk among academic literature and professional researchers throughout the world. Different traditional approaches were suggested based on hypothesis testing and statistical modelling. Therefore, the primary purpose of the research is to come up with a model that can estimate the probability of corporate bankruptcy by evaluating its occurrence of failure using different machine learning models. As the dataset was not well prepared and contains missing values, various data mining and data pre-processing techniques were utilized for data preparation. Within this research, the task of resolving the issues induced by the imbalance between the two classes is approached by applying different data balancing techniques. We address the problem of imbalanced data with the random under sampling and Synthetic Minority Over Sampling Technique (SMOTE). We used five machine learning models (support vector machine, J48 decision tree, Logistic model tree, random forest and decision forest) to predict corporate bankruptcy earlier to the occurrence. We use data from 2009 to 2013 on Poland manufacturing corporates and selected the 64 financial indicators to be broken down. The main finding of the study is a significant improvement in predictive accuracy using machine learning techniques. We also include other economic indicators ratios, along with Altman’s Z-score variables related to profitability, liquidity, leverage and solvency (short/long term) to propose an efficient model. Machine learning models give better results while balancing the data through SMOTE as compared to random under sampling. The machine learning technique related to decision forest led to 99% accuracy, whereas support vector machine (SVM), J48 decision tree, Logistic Model Tree (LMT) and Random Forest (RF) led to 92%, 92.3%, 93.8% and 98.7% accuracy, respectively, with all predictive financial indicators. We find that the decision forest outperforms the other techniques and previous techniques discussed in the literature. The proposed method is also deployed on the web to assist regulators, investors, creditors and scholars to predict corporate bankruptcy

**Bankruptcy Prediction Using Survival Analysis Technique: Yuri Zelenkov (2020):**

Around 2450 Russian companies has been taken for analysis Models used are Random Survival Forest, Multi-task Neural Network, Aalen’s additive Regression, Weibull Accelerated Failure Model etc.

Currently, there is an extensive set of bankruptcy prediction models, but almost all of them are classification based, i.e., they allow to estimate the posterior probability that a particular firm will fail, given its financial characteristics. The expected time to failure is not considered explicitly. On the other hand, there is a survival analysis that deals with the time of the occurrence of the event of interest (while this event may not occur during observation). However, despite its popularity in the medical and technical sciences, survival analysis is relatively rarely used in predicting financial failure. Even when it is applied, most authors use the simplest form of a model. The goal of our work is to evaluate the applicability of survival analysis to bankruptcy prediction. We compare a few state-of-art statistical and machine learning models using a real dataset. Our findings confirm that survival analysis allows (1) to extract from given data valuable information regarding the dynamics of risks and (2) to estimate the impact of features.

To summarize all of the above survival analysis is a handy tool that can be applied to bankruptcy prediction problem. It is especially true for machine learning models, such as the RSF. Survival analysis models not only allow to reasonably accurately identify potential bankruptcy but assess the de-pendence of risks on times using censoring data. What is especially valuable, quite mature free tools are currently available, which could potentially reduce the cost of introducing SA models into industrial operation.

**CHAPTER 3**

# SYSTEM ANALYSIS

## 3.1.1 Problem statement

To analyse the number of algorithms and find the best one to work with and also to obtain the efficient way to avoid bankruptcy.

**3.1.2 Problem statement definition**

There are several algorithms to find the bankruptcy of the company and also have several methods to calculate the bankruptcy but there is not any actual method to calculate the accuracy of the algorithm. This model will run several algorithms on the use case scenario and find the efficient algorithm for the right situation to obtain the better result so that the result calculated.to improve the stability of the company.

**Analysis**

Using the machine learning model to predict the right algorithm for the right company at the right situation.

**Idea**

Analyse the companies listed in NSE and BSE to predict whether the company gone to bankruptcy or not.

**Plan**

Gather a dataset which contains lists of companies in the BSE and NSE and use various algorithms of machine learning to predict the accuracy of the model.

**Realization**

Every algorithm which presented are not suitable for every situation and the types of industries available. So, predicting the suitable algorithm for the suitable situation yields more accuracy

## 3.2 Existing System

### 3.2.1 Methodology

***1. Altman Z-score method***

In 1968, Professor Altman developed a bankruptcy prediction model. The Z-score is a formula which combines 5 financial ratios to predict the probability of a company going bankrupt within two years. The Z-score formula gives each of these ratios a important factor, which reflects the relative importance of each ratio in predicting bankruptcy. The formula is as

**Z-Score = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E**

where,

1. = Working Capital / Total Assets
2. = Retained Earnings / Total Assets
3. = EBIT / Total Assets
4. = Market Value of Equity / Book Value of Total

Liabilities

1. = Sales / Total Assets

***2. Springate Method***

Developed by Gordon Springate in 1978, this model selects four out of nineteen common financial ratios to determine the likelihood of entity failing. At the calculation, if the companies with Springate score lower than 0.862 are classified as “failed companies” and if the companies scores more than 0.862 then the companies are classified as not distress (healthy). The formula to calculate Springate score is

**S = 1,03X1 + 3,07X2 + 0,66X3 + 0,4X4** where,

X1 = Working Capital / Total Assets

X2 = Net Profit Before Interest and Taxes / Total

Assets

X3 = Net Profit Before Taxes / Current Liability

X4 = Sales / Total Assets

***3. Zmijewski Method***

The Zmijewski score is another model for predicting bankruptcy of enterprises based on metrics like performance, leverage and financial liquidity. The ratios which are included are

summarized in this model with the following formula

**X = -4,3 – 4,5X1 + 5,7X2 – 0,004X3** where,

X1 = Earnings after Tax / Total Assets

X2 = Total Debt / Total Assets

X3 = Current Assets / Current Liabilities

**1. Sample Data**

The objective of the research is to predict whether the entity will go bankrupt or not. TO proceed, collect the financial ratios and non-financial ratios of Automobile and automobile related companies from its balance sheet, income statement, cash flow statement market share and other relevant variables for the past 3 years.

The Financial ratios are

1. ***Current Ratio=Current Assets/Current Liabilities***
2. ***Debt-to-Equity Ratio = Total debt / Shareholder’s equity***
3. ***Return-on-Assets = Net Income / Total Assets***
4. ***Return-on-Equity = Net Income / Shareholder’s Equity***

**2. Data Preprocessing**

Preprocess the collected data by cleaning, formatting and transforming to suitable format for the machines to understand. This also includes removing missing values, punctuation, and handling outliers. Putting together all the collected data randomize it. This helps to make sure the model is evenly distributed and the ordering does not affect the learning process.

Visualize the data to understand how it is structured and understand the relationship between various variables present in the dataset. Splitting the dataset into two set – a training set and a testing set. The training set is the set for the model to learn and the testing set is to check the accuracy of the model after training.

**3. Choosing a model**

A machine learning model determines the output you get after running a machine learning algorithm on the collected data. It is important to choose a model which is relevant to the task at hand. Since this is prediction-based research, use models such as Linear Regression, Decision Tree etc.

**4.Train the Model**

Predicting the financial distress comes under supervised algorithm. The main goal of a supervised algorithm is to take some data with known values and to create a model with those values. Since this is a financial related, the companies or entities which already faced bankruptcy or insolvency are considered as the training data (data with known values). The algorithm does the learning and the model contains the learned values. The

learned values include

**1. Distress value more than 2.9**

No danger of bankruptcy and the company is financially safe.

**2. Distress value from 2.77 to 2.90**

The company is on alert to work for the betterment in terms of solvency of the company.

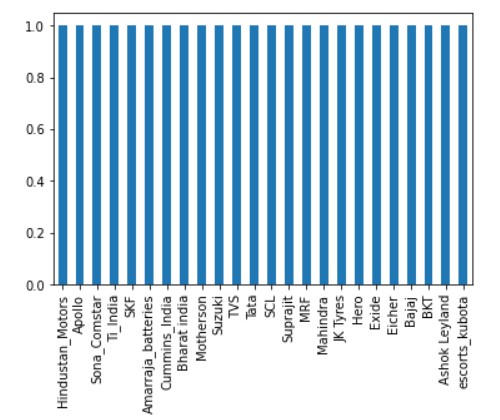
**3. Distress values from 1.8 to 2.77**  The values range depicts the signs of grey areas in the company. The company may go bankrupt within two years if no action is taken.

**4. Distress value below 1.8**

This value depicts the definite failure and closure of the company.

**5. Evaluate the model**

Evaluation is always good in any field. In the case of machine learning, it is considered as best practice. Evaluate the trained model by testing with previously unseen data using appropriate evaluation metrics such as accuracy, precision, recall and F1score.



## 

### 

### 3.3.1 Advantages of the System

* The Main advantage of the System, it is an prediction model for the financial distress of the company by the inputs given by the user .
* Various Libraries and packages in python were implemented to represent the model.
* Provides the accurate result and also calculates the accuracy of the algorithm in order to give the accurate result for the better calculations.
* Simple Classification algorithms for calculate the bankruptcy a group of three algorithms were used to calculate the accuracy of the algorithm performed .
* Simply input is needed for the model to calculate the financial distress status it does not care about the company so that any small scale and large scale industry can calculate their financial status.

**MODULE 1: FRAMING THE DATASET**

**Introduction:**

In machine learning, framing refers to the process of defining the problem one seeks to solve. It involves outlining the goals of a given project and shaping the rest of the machine learning process. The stage provides some sort of checklist before proceeding with the Subsequent stages. Framing in datasets refers to the primary function of the data link layer and it provides a way to transmit data between the connected devices. Framing uses frames to send or receive data. The data link layer receives the packets from the network layer and converts them into frames.

Datasets play a key role in machine learning. They contain data that can be used to teach a computer how to learn and forecast future events using the patterns discovered in the dataset. Many datasets make predictions about the future based on the past. There are many different kinds of datasets that can be used as machine learning resources. Homogeneous datasets make up the majority of them. Datasets that are structured, less noisy, and adequately cleansed will produce results with good accuracy.

**Objective of the datasets:**

We can identify the input (liquidity ratio, solvency ratio, and return of equity given by the user) by using the algorithms and the altman-z score method. The algorithms used in the financial distress management it calculates the input values to calculate the financial distress. The stability of the business in the market is determined by the bankruptcy algorithm. Bankruptcy analysis, often referred to as financial distress management, enables users to execute certain calculations on user-provided information to evaluate if a firm's result is beneficial, detrimental, or null in order to determine the stability of the company in the market. Analysis of the company's existing position in the market, which is used to maintain the company's stability in the market, is the primary goal of financial distress prediction.

There are a number of algorithms to determine whether a company would file for bankruptcy, but there is no accuracy prediction technique to enhance our results or determine the accuracy of the results—the results were just randomly generated. The fundamental goal of constructing the model was to determine its accuracy because when activities are conducted in accordance, there are no choices available to verify the decision taken by the algorithm.

**Datasets:**

1. **Liquidity ratio**

The liquidity ratio was an important variable to calculate the financial status of the company the share holders calculate the company’s stability by analyzing the financial stability of the company in order to buy the shares of that particular company.

**Liquidity Ratio = Current Assets ÷ Current Liabilities**

Other than this single formula the liquidity ratio can be calculated by using other three formulae.

**Current ratio**

The current ratio was called current because it deals with the current assets and the current recent transaction of that particular company. By calculating the current ratio the investor can calculate the short term dept of the company. Current ratios inputs are mentioned in the balance sheet of the company in the annual report by observing that the investor can calculate the financial stability of the company.

The current ratio of the company not like the liquidity ratio of the company results of the current ratio of the company mostly calculated on the recent current event so the result of the current ratio of the company for investing will be more accurate than the liquidity ratio.

* It compares all the current elements mentioned in the balance sheet of the company’s annual report.
* These are updated every year the other financial ratios face a light change in the year end but the current ratio is completely independent of changing in the year end it m ay vary on the returns or the profits or the products bought for the rotation of the next financial year in order to maintain the production scale.
* The current ratio helps the share holders to calculate the short term dept of the company it helps the repaying rate of the company. So that the share holders may buy the shares accordingly.
* The main disadvantage of the current ratio is to the same category industries can be calculated easily, Whereas, the different categories find the difficulty of calculating the current ratio of the company in combined.

C**urrent ratio = Current assets / Current liabilities.**

1. **Debt-to-equity (D/E)**

Debt-to-equity (D/E) ratio is used to evaluate [a company’s financial leverage](https://www.investopedia.com/ask/answers/040915/what-considered-good-net-debttoequity-ratio.asp) and is calculated by dividing a company’s total liabilities by its [shareholder equity](https://www.investopedia.com/terms/s/shareholdersequity.asp). D/E ratio is an important value in corporate finance. It is a measure of the company’s debt and the financial rotation of the company beyond the company’s resources. Debt-to-equity ratio is a certain type of [gearing ratio](https://www.investopedia.com/terms/g/gearingratio.asp).

* Debt-to-equity (D/E) ratio compares a company’s total liabilities with the shareholder equity and can be used to calculate on debt.
* D/E ratios may vary form one industry to another industry so the D/E ratio should be calculated for the companies under the same area. It will be difficult to calculate the D/E ratio for the companies in the different sectors.
* According to the D/E ratio under similar companies if the ratio is high the company can take risk accordingly whereas in the vice-versa case the company should play a safe play in order to avoid the critical situation.
* Investors will often modify the D/E ratio to consider only long-term debt because it leads to higher risk than short-term obligations.

**3.solvency ratio**

The solvency ratio is an important metric used to evaluate a company's ability to meet its long-term debt obligations. This ratio is often used by prospective lenders and investors to assess the creditworthiness and financial health of a company.

Some common solvency ratios include the debt-to-assets ratio, interest coverage ratio, equity ratio, and debt-to-equity (D/E) ratio. Each of these ratios provides insight into different aspects of a company's financial health.

While liquidity ratios focus on a company's short-term ability to meet its obligations, solvency ratios take a longer-term view. It's important to note that these ratios are best used when compared over time or against other companies in the same industry.Additionally, the return on assets ratio is another key financial metric that measures how efficiently a company is using its assets to generate profit. This ratio is calculated by dividing net income by total assets. A higher return on assets ratio indicates that a company is generating more profit per dollar of assets.

**Return on assets ratio = Net income / Total assets**

1. **return on equity**

Return on equity (ROE) is a financial ratio that measures a company's profitability by dividing its net income by shareholders' equity. Shareholders' equity represents the amount of assets that a company's shareholders would receive if all of the company's liabilities were paid off.

ROE is an important measure of a company's financial performance as it indicates how effectively a company is using its equity financing to generate profits. A higher ROE is generally considered better, as it shows that the company's management is efficient at generating profits from the funds invested by shareholders.

To calculate ROE, you divide a company's net income by its shareholders' equity. ROEs can vary widely depending on the industry or sector in which the company operates, so it's important to compare a company's ROE to other companies in the same industry to get a better sense of its financial health.

1. **All are financial ratio:**

The Relationship between Liquidity and Financial Distress

Liquidity is often measured using current ratio (current assets divided by current liabilities). Current ratio measures the ability of a company to fulfill its short-term liabilities with its current assets. This means that the higher the ratio value, the better is the company’s ability to meet its current liabilities (which are soon due).

The Relationship between Activity Ratio and Financial Distress

The activity ratio measures the effectiveness of company in utilizing its assets, or the level of efficiency of resource utilization. Activity ratio can be used as a proxy for total assets turnover. It measures the turnover of the assets over the sales. Assets used for operating activities will increase production. The higher the total assets turn over value, the higher the ability of company to increase sales, so the lower the company’s potential to experience financial distress. Total assets turn over can predict financial distress companies

How is the liquidity ratio capable of forecasting financial difficulties? The liquidity ratio measures a company's ability to meet its current requirements. In other words, it is the ability of a company to repay (current assets) all the short-term loan obligations (current liabilities). If the value of the ratio is higher, it indicates that the company has more short-term assets than short-term liabilities

**Liquidity ratios:**

**Leverage ratios:**

[Leverage ratios](https://corporatefinanceinstitute.com/resources/knowledge/finance/leverage-ratios/) measure the amount of capital that comes from debt. In other words, leverage financial ratios are used to evaluate a company’s debt levels. Common leverage ratios include the following

The [debt ratio](https://corporatefinanceinstitute.com/resources/knowledge/finance/debt-to-asset-ratio/) measures the relative amount of a company’s assets that are provided from debt

**Debt ratio = Total liabilities / Total assets**

The [debt to equity ratio](https://corporatefinanceinstitute.com/resources/knowledge/finance/debt-equity-ratio-formula/) calculates the weight of total debt and financial liabilities against shareholders’ equity:

**Debt to equity ratio = Total liabilities / Shareholder’s equity**

**Efficiency ratios:**

Efficiency ratios, also known as activity financial ratios, are used to measure how well a company is utilizing its assets and resources. Common efficiency ratios include:

The [asset turnover ratio](https://corporatefinanceinstitute.com/resources/knowledge/finance/asset-turnover/) measures a company’s ability to generate sales from assets:

**Asset turnover ratio = Net sales / Average total assets**

The [inventory turnover ratio](https://corporatefinanceinstitute.com/resources/knowledge/finance/inventory-turnover/) measures how many times a company’s inventory is sold and replaced over a given period:

**Inventory turnover ratio = Cost of goods sold / Average inventory**

**Profitability ratios:**

[Profitability ratios](https://corporatefinanceinstitute.com/resources/knowledge/finance/profitability-ratios/) measure a company’s ability to generate income relative to revenue, balance sheet assets, operating costs, and equity. Common profitability financial ratios include the following:

The [gross margin ratio](https://corporatefinanceinstitute.com/resources/knowledge/finance/gross-margin-ratio/) compares the gross profit of a company to its net sales to show how much profit a company makes after paying its cost of goods sold: **Gross margin ratio = Gross profit / Net sales**

**Module 2 ALGORITHMS**

**Introduction**

Machine Learning is a subfield of a artificial intelligence that involves the development of algorithms and statistical mode that enable computers to improve their performance in tasks through experience. Machine learning starts with data numbers, photos or text. It is a growing technology which enables computers to learn automatically from past data. Machine Learning uses various algorithms for building mathematical models and making predication using historical data or information.

Algorithms:

1. Altman Z-score method

Professor Altman created a model for predicting bankruptcy in 1968. The Z-score is a method that combines five financial ratios to estimate the likelihood that a business would fail within two years. Each of these ratios is assigned an important element by the Z-score algorithm, which shows the proportionate weight each ratio has in the prediction of bankruptcy. The equation reads as

Z-Score = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E

where,

A = Working Capital / Total Assets

B = Retained Earnings / Total Assets

C = EBIT / Total Assets

D = Market Value of Equity / Book Value of Total

Liabilities

E = Sales / Total Assets

Zeta

Zeta, which is the Z score, is the outcome of the model that determines whether the company being analyzed is under financial distress or not and if it's likely to go bankrupt.

Zeta values lie between O-4, with the upper value being flexible.

X1= Working Capital / Total asset Ratio

Working capital is a financial metric representing operating liquidity available to a business and is measured as the difference between current assets and current liabilities.

Total assets are taken from the company's annual financial report filings. The higher the ratio, the better.

X2= Retained Earning / Total asset Ratio

Retained earnings (RE) is the amount of net income left over for the business after it has paid out dividends to its shareholder. The decision to provide retained earnings is with management and is vital to the company's value.

X3=EBIT / Total asset Ratio

EBITi.e., earnings before interests and taxes, is calculated by subtracting interest and tax expenses from revenue generated by a company during the fiscal year.

It can also be referred to as operating profit or profit before interests and taxes. Calculating this ratio can infer how well a company uses its assets to generate profits. Hence, determining the profitability of the company.

X4 =Market value of equity / Total asset ratio

It is not the same as book value, but it is also known as market capitalization. It is calculated by multiplying the total number of outstanding shares by one share's market value. It can also be referred to as the total value given by the investment community to a business.

X5 = Total sales / Total Liabilities Ratio

Total sales refer to the total income generated by a company by sales of all goods and services. At the same time, total liabilities comprise current liabilities and non-current liabilities.

Liabilities that are meant to be paid within a year are current liabilities, and those which are payable after a year are non-current liabilities.

Range of Z score and observations :

• According to the model, a company with a score greater than 2.99 means it's in a safe zone and under no financial distress.

• A company with a score of 1.8 < Z > 3, i.e., greater than 1.8 but less than 3, puts the company into a gray zone, meaning the company is under financial distress and has a high chance of going bankrupt shortly.

• A company with a score of 0 < Z > 1.8, i.e., greater than zero but less than 8, puts into a distress zone, making it most vulnerable to bankruptcy because of financial distress

**2. Springate Method**

This model, created by Gordon Springate in 1978, chooses four of the nineteen most typical financial ratios to assess the chance of an entity failing. Companies are categorised as "failed companies" at the time of computation if their Springate scores are less than 0.862 and as "healthy" if their scores are greater than 0.862. The following formula determines the Springate score:

S = 1,03X1 + 3,07X2 + 0,66X3 + 0,4X4 where,

X1 = Working Capital / Total Assets

X2 = Net Profit Before Interest and Taxes / Total

Assets

X3 = Net Profit Before Taxes / Current Liability

X4 = Sales / Total Assets

**3. Zmijewski Method**

Another model for predicting company failure based on factors like performance, leverage, and financial liquidity is the Zmijewski score. The following formula serves as a summary of the contained ratios in this model:

X = -4,3 – 4,5X1 + 5,7X2 – 0,004X3 where,

X1 = Earnings after Tax / Total Assets

X2 = Total Debt / Total Assets

X3 = Current Assets / Current Liabilities

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. in classification problems , Random Forest builds decision tree from various samples and uses their majority vote for classification. In Regression problems, it builds decision tree from various samples and uses their average for regression. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. it is applied in various industries such as banking and e-commerce to predict behaviour and outcomes.

Decision trees Algorithm:

Decision trees algorithm is a popular supervised learning algorithm that is used for classification and Regression Problem. It generate the outcomes as optimized result based upon the tree structure with the condition or rules. Decision trees are versatile machine learning algorithms capable of performing both regression and classification tasks and even work in case of tasks which have multiple outputs.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior training data.

It is considered to be the most understandable Machine Learning algorithm, and it can be easily interpreted. Unlike most Machine Learning algorithms, it works effectively with non-linear data.

**Advantage of Decision tree :**

• Easy to understand

• Requires minimum data cleaning

• No constraint on the data type

**Disadvantage of Decision tree :**

• Possibility of overfitting

DecisionTreeRegressio (): It is the decision tree regressor function used to build a decision tree model in Machine Learning using python. The DecisionTreeRegressor () function looks like this:

DecisionTreeRegressor (criterion = ‘mse’, random\_state =None , max\_depth=None, min\_samples\_leaf=1,)

• criterion: This function is used to measure the quality of a split in the decision tree regression. By default, it is ‘mse’ (the mean squared error), and it also supports ‘mae’ (the mean absolute error).

• max\_depth: This is used to add maximum depth to the decision tree after the tree is expanded.

• min\_samples\_leaf: This function is used to add the minimum number of samples required to be present at a leaf node.

**XGBoost Algorithm :**

XGBoost or Extreme Gradient Boosting is a machine learning algorithm that is used for the implementation of gradient boosting decision trees. It is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models. XGBoost is an implementation of gradient-boosting decision trees. It has been used by data scientists and researchers worldwide to optimize their machine-learning models. XGBoost is known for its accuracy and has been shown to outperform other machine learning algorithms in many predictive modeling tasks

**Module 3 Literature Survey**

**1. Corporate Bankruptcy Prediction: An Approach Towards Better Corporate World - Talha Et all (2020):** An Approach Towards Better Corporate World" by Talha et al. (2020) explores the importance of corporate bankruptcy prediction models for companies and the wider economy.

The authors argue that predicting bankruptcy is crucial for companies to take appropriate measures to avoid it or minimize its impact. They point out that bankruptcy not only affects the company itself, but also its stakeholders such as employees, shareholders, creditors, and the economy as a whole.

The article highlights various bankruptcy prediction models and techniques, including financial ratios, statistical methods, and machine learning algorithms. The authors discuss the advantages and limitations of each method, and emphasize the importance of selecting the most appropriate model for a particular company or industry.

The article also provides a comprehensive review of the literature on corporate bankruptcy prediction, including studies that have applied various prediction models to different industries and countries. The authors summarize the main findings and conclusions of these studies and suggest future research directions.

Overall, the article stresses the importance of bankruptcy prediction for companies and the economy, and provides a valuable resource for researchers and practitioners interested in this field. The authors suggest that the development of more accurate and effective bankruptcy prediction models can contribute to a better corporate world by helping companies to avoid bankruptcy and promoting financial stability and sustainability.

**2. Bankruptcy Prediction Using Survival Analysis Technique: Yuri Zelenkov (2020):** It presents a study on predicting bankruptcy using a survival analysis technique called Cox proportional hazards regression.

The author argues that traditional bankruptcy prediction models, such as discriminant analysis and logistic regression, have limitations in predicting bankruptcy due to the dynamic nature of financial data. Survival analysis, on the other hand, takes into account the time dimension and allows for the analysis of censored data, which can be particularly useful in predicting bankruptcy. The article provides a comprehensive review of the literature on bankruptcy prediction and survival analysis, and explains the Cox proportional hazards regression model and its assumptions in detail. The author then applies the model to a dataset of bankrupt and non-bankrupt firms from the Korean stock market and compares the results with traditional bankruptcy prediction models.The results of the study show that the Cox proportional hazards regression model outperforms traditional bankruptcy prediction models in terms of accuracy and stability. The model also provides insights into the factors that influence the time to bankruptcy, such as profitability, liquidity, and leverage.

Overall, the article highlights the importance of using advanced statistical techniques such as survival analysis in bankruptcy prediction and provides a valuable contribution to the literature on the topic. The author suggests that the Cox proportional hazards regression model can be used by financial analysts and investors to identify companies that are at risk of bankruptcy and make informed investment decisions.

**3. Financial distress analysis in Indian Automobile industry -Judging Financial Health of 12 companies from auto Ancillary Sector of India using Altman’s Z Score Model**.

This literature review study looks into the present state of the Indian automobiles industry's finances and determines if Altman's Z score model can accurately predict the sector's corporate difficulties. Present analysis reveals that automobile industry under our study was just on the range of intermediate zone. indicates that overall financial performance of automobile sector in India is at present viable as Z score indicates but may lead to corporate bankruptcy in near future unless regulatory measures are undertaken immediately. In various capital markets around the world, numerous investigations have made an effort to enhance and replicate Altman's initial findings. This analysis of the literature examines the current financial situation of the Indian automobile industry and assesses whether Altman's Z score model can reliably forecast the sector's corporate issues. Current research indicates that the Automobile sector was only operating in the range of the middle zone. Numerous studies have attempted to improve and replicate Altman's initial findings in other capital markets around the world. Therefore, we are using a machine learning method that uses the Z-Score Altman model to predict financial distress.

1. **Detection of Financial Distress in the Indian Automobile industry**

explores the financial distress and bankruptcy prediction in the Indian automobile industry using financial ratios and machine learning techniques.

The authors argue that financial distress is a common phenomenon in the automobile industry, and it is essential for stakeholders such as investors, creditors, and regulators to detect and respond to financial distress in a timely manner to avoid financial losses and negative impacts on the industry and the economy.

The paper provides a comprehensive review of the literature on financial distress and bankruptcy prediction in the automobile industry, and describes the financial ratios commonly used for bankruptcy prediction. The authors then apply machine learning algorithms, including decision trees, support vector machines, and artificial neural networks, to a dataset of Indian automobile companies to compare their performance in predicting financial distress.

The results of the study show that machine learning techniques outperform traditional financial ratios in predicting financial distress in the Indian automobile industry. The authors suggest that stakeholders can use the machine learning models to monitor the financial health of companies and take appropriate measures to prevent financial distress.

Overall, the paper highlights the importance of detecting and responding to financial distress in the Indian automobile industry and provides a valuable resource for researchers and practitioners interested in bankruptcy prediction and machine learning techniques. The authors suggest that the findings can be applied to other industries and countries as well.