horizontal line

Project Summary

|  |  |
| --- | --- |
| Batch details | DSE Online May 2022 |
| Team members | Gokul Mohan  Megha Gowda  Shaik Asheesh  Shreya Panicker  Ujjwal Kohli |
| Domain of Project | Big Data Analytics in Finance |
| Proposed project title | Bankruptcy Prediction using Financial Ratio’s and Corporate Governance Indicators |
| Group Number | Group 2 |
| Team Leader | Shaik Asheesh |
| Mentor Name | Animesh Tiwari |

Date: 06 Oct 2022

Table of Contents

|  |  |  |
| --- | --- | --- |
| Sl NO | Topic | Page No |
| 1 | Overview | 2 |
| 2 | Business problem goals | 3 |
| 3 | Topic survey in depth | 4 |
| 4 | Critical assessment of topic survey | 5 |
| 5 | Methodology to be followed | 5 |
| 6 | References | 7 |

Project Details

[OVERVIEW (50-100 words) 2](#_Toc116121414)

[BUSINESS PROBLEM STATEMENT (GOALS) 3](#_Toc116121415)

[TOPIC SURVEY IN BRIEF (200-250 words) 4](#_Toc116121416)

[CRITICAL ASSESSMENT OF TOPIC SURVEY (50-100 words) 5](#_Toc116121417)

[METHODOLOGY TO BE FOLLOWED (Explain each steps from 1-5) 5](#_Toc116121418)

[REFERENCES 7](#_Toc116121419)

# OVERVIEW (50-100 words)

Bankruptcy [3] means a firm or business is unable to pay its debts. Businesses file bankruptcy because they are insolvent, unable to pay their bills in both the short term and long term. There are many theories about the root causes of business bankruptcy, but most important and significant cause is poor financial practices. In this project the aim is to predict whether a business or firm goes bankrupt or not.

as we are going to predict bankrupt using data, we will need quality factors which will have significant impact on the dependent variable bankruptcy. So according to [1], In this project we will include Financial Ratios (FR) and corporate governance indicators (CGI) to predict the bankruptcy of financial institutions. The scope of this project is to asses bankruptcy with seven different categories of FRs and five different categories of CGIs

# BUSINESS PROBLEM STATEMENT (GOALS)

1. Business Problem Understanding

Bankruptcy [3] is a major issue for business and can have a negative impact both on the enterprise itself and the global economy. In short, bankruptcy prediction is a very important task for many related financial institutions. In general, the aim is to predict the likelihood that a firm may go bankrupt with quality factors such as the FR categories of solvency and profitability and the CGI categories of board structure and ownership structure. Financial institutions are in need of effective prediction models in order to make appropriate lending decisions.

1. Business Objective – optimum days to solve a ticket

The objective is to build a model which will predict the likelihood of a firm going bankrupt

1. Approach

* Data Understanding

The data set is taken from [2] which has 95 features and around 6900 rows approximately.

* Data Pre-Processing

1. We will check for the data types of all the features and then adjust the data types w.r.t features
2. Next step is to check for missing values, if there are any present we will fill them with most logical approach

* Exploratory Data Analysis

1. Perform univariate analysis for the applicable variables
2. Perform bivariate analysis for the applicable pairs
3. Perform multivariate analysis if required
4. Check for outliers and remove them with using either IQR method, Z-score method or replace outliers with mean, median whichever is most logical
5. After this step we will have the raw data set cleaned

* Feature Engineering

1. We will check if we can extract data from features and increase the column number

* Model Building

1. Split the data into training and testing data
2. We will build basic model with all the features and see the performance
3. We will then remove insignificant variables and build another model to compare with the basic model
4. We will use different ML algorithms such as Logistic Regression, Decision Trees classifier, Random Forest classifier, KNN

* Model Evaluation

1. We will use confusion matrix to find the accuracy of the model
2. We will also include ROC curve to evaluate the model

* Model Optimization

1. We can optimize the parameters of the ML algorithms using Grid Search CV and find best possible fit for the training data
2. Conclusions

By implementing the resultant models built using the above methods, we can suggest and predict the possibility of a firm or business going bankrupt.

# TOPIC SURVEY IN BRIEF (200-250 words)

1. Problem understanding

Bankruptcy [3] is a major issue for business and can have a negative impact both on the enterprise itself and the global economy. In short, bankruptcy prediction is a very important task for many related financial institutions. In general, the aim is to predict the likelihood that a firm may go bankrupt with quality factors such as the FR categories of solvency and profitability and the CGI categories of board structure and ownership structure. Financial institutions are in need of effective prediction models in order to make appropriate lending decisions.

1. Current solution to the problem

There are some techniques which have been employed to develop bankruptcy prediction models, including statistical and machine learning techniques (Balcaen and Ooghe, 2006, Kumar and Ravi, 2007, Lin et al., 2012, Verikas et al., 2010). In general, financial ratios (FRs), recognized as one of the most important factors affecting bankruptcy prediction, are used to develop prediction models (Altman, 1968, Beaver, 1966, Ohlson, 1980). There are also some models which considered CGI to predict the bankruptcy.

1. Proposed solution to the problem

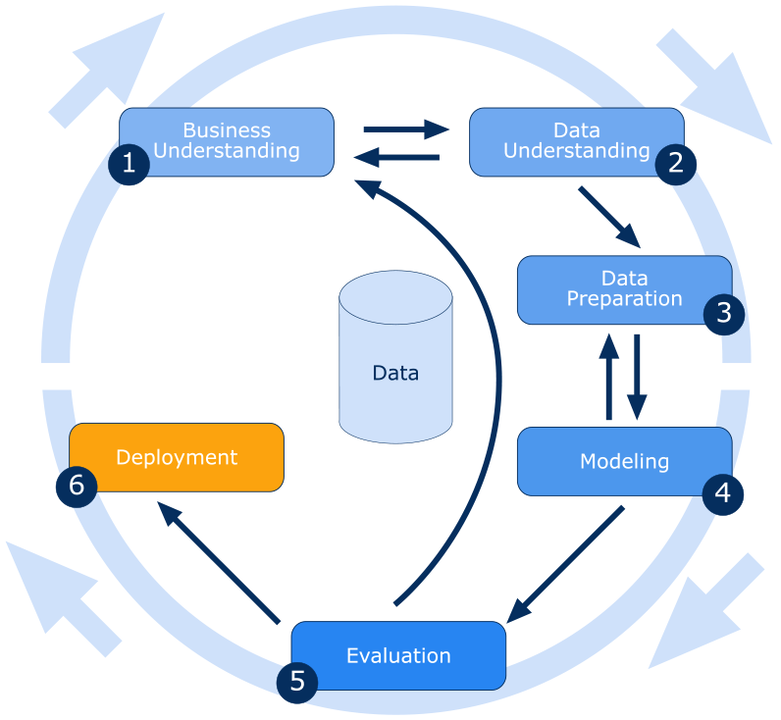
In this project we will combine FR’s and CGI’s together to predict the bankruptcy. As FR and CGI are significant features, by combining them we will get a far better prediction model and therefore increasing the accuracy of the classification model built.

1. Reference to the problem
   1. <https://www.kaggle.com/code/edacebeci/predict-bankruptcy>
   2. <https://www.sciencedirect.com/science/article/abs/pii/S1059056014000057>
   3. <https://www.jstor.org/stable/2978933#:~:text=The%20prediction%20of%20corporate%20bankruptcy%20is%20used%20as%20an%20illustrative,discriminant%20statistical%20methodology%20is%20employed.>

# CRITICAL ASSESSMENT OF TOPIC SURVEY (50-100 words)

* Since bankruptcy prediction has become very important part in a firm’s growth, it will help the firm in planning their future budgets and forecasting their financial condition. This could help the company to take pre meditated actions to prevent future losses. In a similar way it can also help the customers to monitor their spending. This model can also help in improving the allocation of resources for a firm.
* As our model consists of both the Financial Ratios and Corporate Governance Indicators, the FRs can help in understanding the financial condition of the firm while the CGIs can help in proper management of the resources in pursuit of set goals. With better knowledge of the firm’s future, risk management and internal control systems can be managed efficiently. This would certainly improve the Corporate Governance of the company i.e., better financing, firm valuation and performance of the firm.
* Our aim is to devise a model that would add value to the business by prioritizing the key indicators as well as by considering all metrics. Thereby, it adds value to the financial institution and prevents the possibility of going bankrupt.

# METHODOLOGY TO BE FOLLOWED (Explain each steps from 1-5)



*\* If deployment is out of scope to the team or not advised by the mentor, please opt to leave out the step no 6.*

Business Understanding:

Bankruptcy[3] is a major issue for business and can have a negative impact both on the enterprise itself and the global economy. In short, bankruptcy prediction is a very important task for many related financial institutions. In general, the aim is to predict the likelihood that a firm may go bankrupt with quality factors such as the FR categories of solvency and profitability and the CGI categories of board structure and ownership structure. Financial institutions are in need of effective prediction models in order to make appropriate lending decisions.

Data Understanding:

The data set is taken from [2] which has 95 features and around 6900 rows approximately. The data includes FR and CGI which are independent variables help in predicting the Dependent variable Bankruptcy

Modelling:

* + 1. Split the data into training and testing data
    2. We will build basic model with all the features and see the performance
    3. We will then remove insignificant variables and build another model to compare with the basic model
    4. We will use different ML algorithms such as Logistic Regression, Decision Trees classifier, Random Forest classifier, KNN for building the machine learning model

Evaluation:

1. We will use confusion matrix(type1 and type2 errors) to find the accuracy of the model
2. We will also include ROC curve to evaluate the model

Deployment:

1. We will make use of FLASK framework to build a web application, which takes the information required to predict Bankruptcy and then depending on the provided info it will give us the prediction.
2. We will make use of HEROKU/GCP for the deployment

# REFERENCES

1. <https://www.sciencedirect.com/science/article/abs/pii/S0377221716000412>
2. <https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction>
3. <https://blogs.worldbank.org/allaboutfinance/the-challenges-of-bankruptcy-reform>

**NOTES FOR PROJECT TEAM**

*Sample Reference for Datasets (to be filled by team and mentor)*

|  |  |
| --- | --- |
| Original owner of data | <https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction> |
| Data set information | * The shape of data set is 95 features, 6820 rows * Data set is taken from Kaggle which includes FR and CGI of business/firms |
| Any past relevant articles using the dataset | <https://www.sciencedirect.com/science/article/abs/pii/S0957417421009659> |
| Reference | <https://www.sciencedirect.com/science/article/abs/pii/S0377221716000412> |
| Link to web page | <https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction> |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*