POLITECNICO DI TORINO

Master of Science in
Communications and Computer Networks Engineering

Master Thesis

Application of Corporate Governance Index to Predict Bankruptcy of Companies using Machine Learning

Supervisor

PERBOLI GUIDO

Co supervisor

VELARDOCCHIA FILIPPO

Candidate

ZINNIA MONDAL

July 2022

Abstract

Bankruptcy Prediction of companies is becoming increasingly significant in recent times as it enables stakeholders to act quickly and reduce their financial losses. To create bankruptcy prediction models, many machine learning techniques have been applied using financial features. However, there have been comparatively less research on how non-financial features like Corporate Governance indices can be used to predict a company's performance. Hence, this thesis is motivated by the need of further research within bankruptcy prediction influenced by Governance indices using traditional machine learning models and neural networks as there has been very less research using only the governance indices as the contributing features. These governance indices can be for example the age of the company, the number of shareholders, the number of board members etc of a company.

In our thesis we aim to predict the bankruptcy of 160,000 Italian small and medium sized enterprises and understand the impact of Governance indices as important features related to bankruptcy prediction in companies. We try to understand which machine learning models predict the bankruptcy with maximum accuracy using only non-financial variables.

As complex machine learning models are considered as a black box, we use Explainable Artificial intelligence to understand the most important governance indices that are contributing to the prediction of bankruptcy in each company as a greater number of stakeholders have started asking justification for the reason behind the prediction made by the models. We also discover the common patterns that is followed among all the companies in terms of the features contributing to the maximum accuracy.

Acknowledgements

I would firstly like to thank my supervisors. I thank Professor Guido Perboli (DIGEP) for giving me the opportunity to work on this thesis and at ARISK Srl and, thanks to Filippo Velardocchia, my co-supervisor for guiding and mentoring me constantly throughout the thesis period of six months.

I would also like to thank my friend Harshit for helping and motivating me in every conceivable way and making each struggle seem a little easier.

Finally, I would like to thank my parents who have been my biggest supporter and believer and sacrifice so much every day just to see my dreams come true.

Contents

Chapter 1

Introduction

1.1 Significance of Corporate Governance and Bankruptcy prediction

The contribution of corporate governance in successfully running organizations have been widely documented in academic literature.

Corporate governance is the set of rules, procedures, and processes that guide and control a company. It balances the interests of a company's numerous stakeholders, which include shareholders, top management executives, consumers, suppliers, financiers, the government, and the community.

Bankruptcy or corporate failure can have negative consequences for both companies and the global economies. Business practitioners, investors, governments, and academics have long looked at techniques to predict the likelihood of a company failing in order to minimize the financial losses associated with bankruptcy (Balleisen, 2001 [1]; Zywicki, 2007 [2]).

A very strong example of Corporate Governance failure leading to bankruptcy is the case of the Lehman Brothers in 2008, the fourth-largest investment bank in the United States with 25,000 employees worldwide. It had $639 billion in assets and $613 billion in liabilities. The bank became a symbol of the excesses of the 2007-08 Financial Crisis, as it was consumed by the subprime catastrophe, which spread through financial markets and cost an estimated $10 trillion in lost economic activity (Wiggins et al., 2014 [3]).

While there has been interest in corporate bankruptcy in the accounting and finance literature, the focus has been primarily on predicting bankruptcy based on financial data (e.g. Beaver, 1966 [4]; Altman, 1968 [5]); and there has been very less research that considers the relationship of corporate governance on the bankruptcy risk. This has been attributed to the time-consuming process of gathering accurate governance data.

1.2 Significance of using Machine Learning in Bankruptcy prediction:

In general bankruptcy prediction becomes increasingly important as it allows stakeholders to take early actions and limit their financial loses. Various statistical and machine learning techniques have been used to develop bankruptcy prediction models using financial ratios (FRs), which are considered as one of the most important factors influencing bankruptcy prediction and are commonly employed to construct prediction models. However recent studies have shown that corporate governance indicators (CGIs) also play a vital role in predicting bankruptcy (Liang et al., 2013 [6]; Yeh & Woidtke, 2005 [7]).

In Corporate Governance research, Corporate Governance indices are used to see whether they can predict a company's performance. These Corporate Governance indices are nation-specific indices that use country-specific governance factors that represent local norms, institutions, and data availability, and show that these indices accurately predict firm market value in each country.

In this thesis the data collected by Arisk Srl, is used to understand the impact of Governance indices, as important features related to bankruptcy prediction in companies. The dataset consists of data from more than 160,000 Italian Small and Mid-Size Enterprises (SMEs) that were live and operational by the end of 2018, along with 3000 bankrupt company's data in the period 2001-2018 (Guido Perboli, Ehsan Arabnezhad, 2021 [8]). This thesis is motivated by the need of further research within bankruptcy prediction influenced by Governance indices using traditional machine learning models and neural networks as there has been very less research using only the governance indices as the contributing features.

1.3 Industrial Setting of bankruptcy Prediction by Arisk (Guido Perboli, et al., 2021 [9])

The skeleton of economies might be said to be small and medium-sized businesses (SMEs). SME assistance for economic growth and innovation, competitiveness, and employment are all crucial for the economy of European regions, as (F. Azevedo, et al, 2016 [9]) point out. In example, SMEs account for more than 99.8% of all firms in the EU in 2018. Additionally, SMEs account for the majority of the rise in value-added, per the yearly report (SME Envoy Network. 2019 [10], 60 %). On the one hand, SMEs benefit from greater flexibility than larger businesses and foster innovation. On the other hand, they deal with a variety of issues, including limited funding, a lack of talented and highly skilled workers, and limited access to credit (Z. Vladimirov, 2017 [11]). As the authors of (N. Lee, 2015 [12]) point out, the impacts of the 2008 financial crisis have drawn attention in particular to the accessibility of financial capital for these businesses.

Based on the previous research of (Perboli, et al 2021, [8]) on the use of a machine learning based DSS to predict mid- and long-term crisis in companies, we base our study as part of the Artificial Intelligence based DSS by Arisk to predict bankruptcy of companies using non-financial features like Governance indices. It assists decision-makers in assessing the financial status of enterprises requesting for financial assistance or funding and anticipating the likelihood of bankruptcy far in advance (e.g., local government, banks, and financial institutions). It will assist in determining the effectiveness of local, state, and federal financial policies as well as the best way to allocate available funds. (Perboli, et al 2021, [9])

This section outlines the technique used for data analysis, which is based on Perboli and Arabnezhad's work.

A training and tuning module and a prediction server make up the two separate portions of the overall DSS system. In addition to obtaining public financial data from databases, the training and tuning module may additionally collect additional data via Arisk Srl's private interface. After that, the data are aggregated, normalized, and cleaned up. They are additionally separated into sets known as core and non-core. The machine learning pipeline's feature selection phase bases its decisions on the first set of data. The latter are secondary data that have been taxonomically arranged based on SHELL (M. Cantamessa, et al., 2018 [13]). Non-core data are not directly incorporated into the predictor but are used to simulate perturbations to the machine learning features.

Utilizing financial statement data from approximately 160.000 Italian SMEs that are active and functioning as of the end of 2018, as well as data from over 3,000 bankrupt companies spanning the years 2001-2018, the machine learning module has been trained.

The DSS developed by Arisk Srl adds to the existing literature (E.I. Altman, et al., 2014 [14]; H. Son et al, 2019 [15]) on the connection between bankruptcy and non-

financial characteristics. Both financial and non-financial variables are considered, and their consequences are examined, by the prediction model. These facts, which relate to the most recent year that was available at the time of the research, are taken from the AIDA database (Bureau van Dijk, Aida, 2020[16]), (i.e., 2019).

The prediction model looks at the effects of organizational elements on financial performance, including seniority, the number of shareholders and decision-makers, the existence of an external audit, and familiarity issues. It so provides a tool, called the governance index, that assesses the adequacy of the governance model and, as a result, corporate organizations.

1.4 Research Question

Bankruptcy prediction using Machine Learning has generated substantial amount of literature in the last few years. Machine learning models such as Random Forests, Support Vector Machines (SVM), and Gradient Boosted Trees have been found to be particularly effective for predicting bankruptcy. Statistical models were compared to machine learning (ML) models, by Barboza, Kimura, and Altman. They discovered that Random Forests outperformed Alman's Z-score model by a wide margin (Barboza et al., 2017 [17]). However, little research has been conducted to investigate the impact of corporate governance measures on bankruptcy prediction, particularly in the aftermath of the Global Financial Crisis using Machine Learning.

Based on the research of numerous bankruptcies and scams based on corporate governance failure and the resultant literature as mentioned in the preceding section, the following research questions are presented:

Using Corporate governance indices, do traditional machine learning models or Deep Learning models predict the bankruptcy of companies with maximum accuracy?

Which are the most important governance indices contributing to the prediction of bankruptcy of companies? Is there a common pattern that is followed among all the companies in terms of the features contributing to the maximum accuracy?

This thesis aims to predict the bankruptcy of 160,000 Italian Small and Mid-Size Enterprises in the period of 2001-2018. To explore the research question, it has been divided into three parts:

1. Testing the machine learning models with different combination of features to find out the most important features (governance indices) that predict the mean with maximum accuracy.
2. Comparing the accuracy of prediction by traditional machine learning models with Deep learning model.
3. Finding out the pattern of the most important governance indices that contribute to the prediction of the bankruptcy for each individual company and identifying a common pattern between them.

In reference to the research questions, the rest of the thesis is structured as follows:

Section 2: Literature review of past research exploring the different aspects of Machine Learning application in bankruptcy prediction and also models in the Italian Market.

Section 3: Description of the Data and the features (governance indices), and feature analysis and Data Preprocessing

Section 4: Methodology; Detailed study of the traditional Machine Learning Models and Deep Learning Model used and comparison of the features and the models

Section 5: Testing all the models on the entire unseen dataset

Section 6: Model Analysis using shap on the global interpretation of the model.

Section 7: Description of the analysis and results obtained from the above experiments in relation to the research questions and Finding out the common pattern of most important features in all the companies that contributed to the prediction of the bankruptcy

Section 8: Conclusion, final comments, and implications of the thesis work.

Chapter 2

Literature Review

The standard models of bankruptcy prediction are presented in the first section of our literature review. The second section introduces specialized models with unique characteristics related to bankruptcy prediction. In the final section of this section, we look at locally customized models designed for the Italian market.

2.1 Relevance of Bankruptcy Prediction and Corporate Governance

Over the past 35 years, financial academic research on bankruptcy prediction has gained prominence. The global economy has seen multiple business cycles and financial crises throughout this time, including the Asian Financial Crisis of 1996, the Dot-com boom of the late 1990s, and most recently, the Global Financial Crisis of 2007–2008. These incidents caused a string of bankruptcies, which had the unintended consequences of increasing unemployment, lowering economic production, and writing down asset values.

There are a number of factors that have contributed to the importance of bankruptcy prediction in corporate finance literature, as well as among policymakers, market participants, and society at large.

Numerous members of society are significantly impacted by bankruptcies, and there are major economic and social consequences involved. People experience employment and income losses. The social stigma associated with unemployment may also have a negative impact on one's wellbeing. Asset write-offs occur as a result of shareholders' claims being subordinated to the company's assets and their likelihood of being repaid. Debt holders can make claims on corporate assets based on their level of seniority, but they typically won't be able to get their full-face value back. To make up for lost productivity, governments must offer compensation to the unemployed, retrain them if an industry is sagging, and work to boost commercial activity in other sectors.

Overview of the main bankruptcy prediction models:

There are numerous models for predicting bankruptcy nowadays. According to Nyambuu and Bernard (2015) [18], these models can be roughly categorized into five categories: i) accounting-based models; (ii) credit spread models; (iii) company value models; (iv) rating agency models; and (v) alternative models.

i) Accounting-based models:

Econometric models use a variety of financial ratios as regressors. A bankrupt and comparable non-bankrupt data group are often compared by the models. The models produce an index score, such as a Z-score or an O-score, which serves as a stand-in for the chance of default. (Univariate, Beaver (1967) [19]; Risk Index, Tamari (1966) [20]; MDA, Altman (1968) [5]; Conditional probability, Ohlson (1980) [21])

ii) Credit spread models:

Investigates the difference in interest rates between risk-free debt with a same maturity and debt that is close to default. The spread will show how much investors must be paid for taking on the debt and will, therefore, implicitly show the

likelihood of default; that is, the wider the spread, the greater the likelihood of default. Critics have argued that the credit spread is influenced by factors besides default probability (Hull and White (2000) [22]).

iii) Company value models

Assumes that the firm's capital structure incorporates and converts the possibility of default into the stock price. The Black and Scholes (1973) option model pricing methodology can be used to price the synthetic derivatives that the model creates for the firm's debt structure. The main criticisms of this model center on its reliance on financial statements, which can be somewhat manipulated, and the fact that changes in share prices can be caused by a wide range of endogenous and exogenous events. (Merton (1974) [24]; Black and Scholes (1973) [23])

iv) Rating agency models

Creates a credit rating that is converted into an alphabetical letter ranging from AAA (Best credit rating) to D. (default). The underlying approach is hidden from the public and blends objective analyst analysis with historical financial data. (Fitch Moody’s Standard and Poor’s)

v) Alternative models

Kealhofer, McQuown, and Vasicek created a proprietary model, which Moody's Analytics acquired in 2002. It integrates many default-risk modeling approaches, including the statistical and structural models. The distance between the asset value at which the business defaults and its default point, as determined by the model's Distance-to-Default (DD) measure, is the number of standard deviations. The result is the Expected Default Frequency (EDF), which is a function of DD and takes into account valuation, capital structure, and the overall market situation. (Kealhofer Et al, 2002 [25])

However, for the further literature study we will focus on the more widely available models and which are more relevant to our study.

2.2 Literature Study of Corporate Governance:

Corporate governance, which establishes the policies, procedures, and best practices for balancing competing stakeholder interests, is frequently referred to as "the system through which firms are directed and governed" (Cadbury, 1992 [36]). Due to the separation of ownership and control and the presumption that both sides are interested in maximizing their own utility, the typical problem with corporate governance is when owners and self-serving managers have conflicting interests (Jensen & Meckling, 1976 [26]). In other words, managers have incentives to stray from what is best for the company and pursue opportunistic behavior as a result of the separation of ownership, which eventually reduces value for the owner. The "principal-agent problem" refers to this antagonistic situation (Jensen & Meckling, 1976). The relationship between a company's senior management (agents) and the shareholders is highlighted in financial literature (principals). Corporate governance focuses on establishing the rules and processes for decision-making while allocating rights and obligations to various stakeholders in an organization, including the board, management, and shareholders (European Central Bank, 2004).

2.3 Standard Models

2.3.1 Early Adaptations

Before the 1990s, bankruptcy prediction models were primarily statistical models that used univariate, multivariate, and logit & probit regression techniques. Some of the most prominent studies are listed below:

The Bureau of Business Research (BBR) released a bulletin in 1930 that contained the findings of a research of ratios of failed manufacturing enterprises. 24 ratios from 29 different companies were analyzed to see what the common traits of failed businesses were. The average ratios were calculated using the ratios of the 29 companies. After that, the ratios of each firm were compared to the average ratios to see if the failed enterprises shared any common characteristics or trends. The research discovered eight ratios that were regarded good indications of a company's "increasing fragility." In 1932, FitzPatrick [27] analyzed 13 failure-to-success ratios (19 for each firm status). When compared to "standard" ratios and ratio trends, he discovered that successful companies had favorable ratios and failing companies had unfavorable ratios in most situations. Merwin published his study on small firms in 1942 [28]. When comparing successful and failing companies, he found that the losing companies showed indicators of weakness as early as four or five years before they failed. In 1945, Chudson [29] investigated the patterns of financial structure to see if there was a "normal" pattern. On a broad, economy-wide scale, he reported that there was no "typical" pattern to financial structure. However, he discovered "that there is a clustering of ratios within particular industry, size, and profitability groupings." The results are important for the development of bankruptcy prediction models, even though the study did not expressly address

bankruptcy prediction. His findings, for example, suggest that models built for general use across industries may not be as effective as industry-specific models (Bellovary, J. L., Giacomino, D. E., & Akers, M. D., 2007 [30])

Beaver used univariate analysis to predict bankruptcy in 1966, testing the predictive ability of 30 financial ratios one at a time. Using the univariate approach, he discovered substantial differences in various characteristics between bankrupt and non-bankrupt enterprises. Over a five-year period, he conducted research on a sample of 706 businesses (Beaver, 1966 [19]).

The Altman Z-score is the most well-known and widely used model in the literature (Altman, 1968 [5]). Using Multivariate Discriminant Analysis, he created the Z-score model for bankruptcy prediction based on predetermined ratios. Altman chose five ratios from the original list that were the best predictors of overall performance. Altman put the model to the test on a group of 66 manufacturing companies, half of which were insolvent. The results one year before to bankruptcy were extremely accurate, with 95% of the cases correctly identified. However, for more than 2 years prior to the bankruptcy the accuracy yielded were extremely poor. On the original data, Altman ran a hold-out sample test. While the accuracy of these tests was 96% when applied to bankrupt companies, it was only 79% when applied to non-bankrupt companies.

In 1977, Altman et al. [31] developed a new bankruptcy classification model, known as the ZETA-model which was constructed using a multivariate method, with a study of both linear and quadratic structures, like the prior Z-score model. It which took into consideration changes in the sizes of companies that went bankrupt as well as a broader model that included the retail industry as well as manufacturing.

Changes in financial reporting standards and accepted accounting practices were also factored into the new model. They gathered data from 53 bankrupt businesses and a matched sample of 58 non-bankrupt businesses, with the non-bankrupt businesses matched to the bankrupt businesses by industry and year. One year before to bankruptcy, the ZETA-score revealed an overall accuracy score of 92.8%. For predictions prior to 2-5 years of bankruptcy, the accuracy of the non-bankrupt firms remained high as for the bankrupt firms the accuracy decreased for each lagged year. The study concluded that the linear structure was better than the quadratic structure and the linear structure in the ZETA-model performs far better for 3-5 years prior to bankruptcy when compared to Altman's original Z-score model.

Ohlson developed a logit bankruptcy predictor model using a sample of 105 bankrupt enterprises in 1980, of which 17 percent were listed as bankrupt and the rest were not. He created 3 models for the consecutive number of years prior to bankruptcy. The first model predicted bankruptcy one year prior with an accuracy of 96.12%. The second model predicted two years prior with an accuracy of 95.55%. The third model which predicted three years prior with an accuracy of 92.84% (Ohlson, 1980 [21]). His model is comparable to Beaver (1966), but significantly less accurate than Altman's Z-score model.

A fresh line of inquiry into the categorization potential of several corporate governance variables has developed in bankruptcy prediction research since the 1990s. According to a number of studies (Fich & Slezak, 2008 [32]; Parker et al., 2002 [33]; Chan, 2016 [34]), governance features have a considerable impact on the likelihood of bankruptcy and can be utilized to differentiate between the two groups. In Taiwan, Chen (2008) [35] compares a model that includes extra corporate governance variables to one that only uses classic financial ratio bankruptcy prediction models and finds that the latter's accuracy is increased by 2.9

percentage points. When corporate governance variables are taken into account, other research, including Daily and Danton (1994) [37], find no significant difference in accuracy. The algorithm is still more than 90% accurate at predicting bankruptcy, though. Last but not least, Simpson and Gleason (1999) [38] discover that only particular corporate governance variables, like traits of the CEO and board, have a substantial impact on bankruptcy prediction.

2.3.2 Machine Learning models

In the 1990s, various statistical machine learning algorithms such as discriminant analysis and logistic regression outperformed prior statistical models, kicking off the evolution of bankruptcy models. Machine learning models such as Random Forests, Support Vector Machines (SVM), and Gradient Boosted Trees have been proven to be particularly useful in bankruptcy prediction.

Decision Tree: After Quinlan (1986) [39] created the iterative dichotomizer 3, DT became a crucial machine learning tool (ID3). DT recursively partitions (RPA) the collection of data for the categorization of firms after measuring the discriminant power of sample variables using entropy (Quinlan, 1986). Quinlan (1993) [40] later created the sophisticated version known as Classifier 4.5. (C4.5). The decision rules are induced by DT. The Heuristics are typically used to identify where the rules should be placed in the decision tree (Jeng et al., 1997 [41]). For instance, Liquidity will be placed above or assessed before profitability if it is determined to be more significant liquidity.

In order to predict corporate bankruptcy, statistical modeling-based methods predominated until recently; however, recently, models based on machine learning have been published (Linden et al., 2015 [42]). Recently, machine learning models have been successfully used for a range of classification and regression problems,

and they routinely outperform traditional classification methods (Krizhevsky et al., 2012 [43]). The aim of bankruptcy prediction is to evaluate a company's financial standing and future prospects. For a certain period of time, this task can be treated as a two-class classification problem (Zieba et al., 2016 [44]). Businesses either succeed or fail throughout the allotted time period. The difficulty lies in determining which of these two possible outcomes is more likely.

Thanks to financial reporting rules and public demands for transparency, there is a variety of information available on organizations' financial position (Bredart, 2014 [45]). Due to the amount of data, the area is perfect for sophisticated data-intensive processing methods (Qiu et al., 2016 [46])

The foundation for this analysis is provided by the findings of the Zi eba et al. (2016) study. Zi et al. suggest a machine learning strategy and demonstrate excellent prediction performance for the problem of predicting firm insolvency. They projected insolvency for Polish industrial businesses between 2000 and 2013 using financial parameters in their analysis. In their paper, Zieba et al. demonstrate the appalling performance of a neural network-based classifier. Given that neural networks recently showed exceptional performance in classification tasks generally and in circumstances of bankruptcy prediction, this is unexpected (Bredart, 2014).

Back propagation-trained neural networks are the most widely used method for solving this type of problem (Tsai and Wu, 2008 [47]). In a study of small and medium-sized Belgian enterprises, it was shown that using just a few easily accessible financial data as inputs to an artificial neural network might yield pretty good results (Bredart, 2014).

In an instance, Becerra et al (2005)'s [48] use of a similar approach to examine British corporate bankruptcies between 1997 and 2000 and Shah and Murtaza's (2000) [49] use of a neural network to forecast US company bankruptcy between 1992 and 1994 served as inspirations.

Recently, ensemble classifiers have also drawn some interest. Ensemble classifiers can successfully predict bankruptcy and greatly outperform competing algorithms, as shown by Alfaro et al. (2008) [50] and Zieba et al. (2016).

Behr and Weinblat (2017) examined 446,464 business statements from the balance sheets of many nations, including Italy, Germany, France, Britain, Portugal, and Spain, using three data mining techniques: logit, decision trees, and random forest. The author used accuracy, specificity, sensitivity, and precision as evaluation criteria; no particular resampling technique was used. The decision tree model and the logit model both performed worse than expected, according to the results, which also showed that the random forest model performed better than both. (Behr, et al, 2017 [51])

2.3.2.1 Ensemble methods

The multi-classifier method, sometimes referred to as the ensemble approach, combines several computer processes to enhance performance. Nanni and Lumini (2009) conducted a series of tests and discovered that the ensemble method outperformed stand-alone models in classification performance for credit score and bankruptcy prediction. Australian credit data, German credit data, and Japanese credit data were among the financial datasets employed in these investigations. (Nanni L, Lumini A, 2009 [52])

Boosting and bagging are two crucial ensemble method approaches. Boosting is a technique where a base classifier is first created from the original dataset. Based on the performance of the base classifier, the distribution of the training dataset is then changed, and the next base classifier is trained using the modified sample distribution. The weights assigned to each training set can be used to generate a set of bootstrap samples from the original data (Begley J, Et al, 1996 [53]). Kim and Upneja (2014) employed adaboost, a popular boosting technique, to predict

restaurant financial issues with success (Kim S Y, Et al, 2014 [54]). Instead of Boosting, Bagging employs Bootstrap, which generates random subsets of data by sampling from a given dataset. A technique created by several different classifiers executes a subroutine of its learners before combining them using a model averaging methodology to decrease the model's overfitting (Breiman L, 1996 [55]). Decision Tree, a more traditional machine learning model, serves as the foundation for Random Forest (RF), a common Bagging technique (DT). Kruppa et al. (2013) developed a comprehensive technique to estimate credit risk by individual default probability by applying the RF, which performed better than the LR (Kruppa J, et al, 2013 [56]).

2.3.2.2 Neural Networks

A neural network (NN) is one of the most popular machine learning approaches and is probably where other computational techniques got their start (Barboza F, et al, 2017 [17]). It is similar to human neural processing in that it contains numerous layers, with the first layer determined by the input variables and the last layer producing the output variables. The majority of the output variables are made up of the tag or label of each sample. The use of NN to predict bankruptcy has been the subject of numerous studies. Zhao et al. (2014) created an autonomous credit scoring system with outstanding accuracy (87 percent) and effectiveness using Multi-Layer Perceptron Neural Network (MLPNN) and an experiment with German credit data (Zhao Z, et al, 2015 [57]). An ensemble model, which is produced by mixing numerous single NN, may perform better than a single classifier. Tsai and Wu (2008) investigated and compared the performance of a single NN classifier and an ensemble NN classifier on the prediction of credit score and bankruptcy. The many classifiers may not be more efficient in binary classification scenarios, however, as the multiple (ensemble) NN classifier did not consistently outperform the best NN classifier (Tsai C F, et al, 2008 [47]).

2.4 Bankruptcy Prediction in the Italian Market

According to Instituto Nazionale di Statistica, Industrial Districts (ID), where 24.4 percent of enterprises and 24.5 percent of employees are employed, make up around one-fourth of the productive system in Italy, a nation recognized for its large concentration of unlisted manufacturing SMEs. Another trait is the split of Italy's regions into the North and the South.

In a study of insolvency prediction analysis of Italian small firms, we see the use of machine learning and Neural Networks for predicting bankruptcy for Italian small and medium sized enterprises, with Gradient Boosting giving the best accuracy and Neural Networks giving good performance but with high computational effort. (Agostino Di Ciaccio et al, 2019 [58])

In the paper Machine Learning based DSS for Mid and Long Term company crisis prediction we see the methods used to forecast company crisis of Italian Small and Medium Sized Enterprises(SMEs) for upto 60 months using operational and financial data. We also see the use of different Machine Learning methods in a Decision Support System developed by Arisk to analyse the Italian economic System and validate public policy pertaining to the COVID-19-related economic disruption.

It demonstrates how a two-phase dataset construction technique and a suitable feature section procedure may yield a Random Forest approach with precision over 85% and a prevision horizon up to 60 months. As a result, not only does the created machine learning predictor perform best for forecasts made over the conventional literature time horizon of 12 months, but also for predictions made over the medium-term (3 years) and long-term (5 years). The method was used to evaluate the long-term effects of the COVID-19 disease in Piedmont on the economic

environment and to mimic Italian government policies. It was eventually included in a DSS. (Perboli et al, 2020 [8])

In a further study by Arisk and some policy makers of the regional government of Piedmonte we also see the use of Machine Learning to access Public policies to support Italian SMEs where it shows how different Artificial Intelligence techniques can be used to assist decision makers for creating and deploying regional policies. (Perboli et al, 2021 [9])

In a study by the Banca d'Italia we see the use of Explainable Artificial Intelligence (XAI) for interpreting default forecast models based on Machine Learning where some XAI methods are used to clarify the random forest corporate default forecasting model used by Italian non-financial enterprises in Moscatelli et al. (2019) [59]. The techniques shed light on the relative significance of financial and credit indicators to forecast financial hardship in enterprises. Additionally, it analyzes how ML models might improve the accuracy of credit evaluation for borrowers with less established credit links, such as smaller businesses. (Giuseppe Cascarino, et al, 2022 [60])

Chapter 3

The Data

Time Period: (2001-2018)

For this study, a large dataset of 160,000 small and sized enterprises in Italy was used by ARISK which were live and operational by the end of 2018 and 3000 more bankrupt companies covering the period of 2001-2018. (Perboli et al, 2021 [9])

In addition to financial data, Arisk's Decision Support System can gather, classify, and incorporate various types of threats. It gathers data on budget and financial data, firm organization data, and cash flow and supply chain management risk matrices for families.

The training and tuning module gathers data from public databases such as public financial data (in Italy, the Italian Camera di Commercio), a set of indexes and ratios from AIDA Bureau Van Dick (Bureau Van Dijk, 2020 [16]) and whether available data from Arisk's proprietary interface to gather additional data. The data is then cleaned normalised and merged. The core data represents the machine learning module's features, whereas non-core data are not directly incorporated into the machine learning.

3.1 Data Source: The Training Data as collected by the DSS of Arisk

Limited companies and joint-stock companies are the two types of companies. There are bankrupted companies with revenues ranging from one million to forty million euros in at least one of the last five years before bankruptcy and a company lifetime of at least ten years. The last 5 years of financial data collected for each company is saved in 5 different data sets that are roughly made up of 3000 companies. A bankrupt company is removed from the dataset if it has fewer than five official financial reports.

The missing values have been replaced by zeroes on the training and test data and the standard scaling is applied to both.

If the data set was constructed in this manner, the data set would have become highly imbalanced, affecting the outcome of our machine learning model in terms of recall of the confusion matrix. So, to mitigate the negative impact of this, 6, 000 active companies are selected from a total of 160, 000 and was merged. By doing so, the data set's imbalanced nature is maintained but in a controlled manner, which means precision for recall is sacrificed because finding all companies that are most likely to declare bankruptcy is more important.

Now, for each year of information from bankrupted companies, the same sample of active companies is added, and the final data set is constructed, which has 5 parts (year 1, year 2, year 3, year 4, and year 5) and 8959 companies.

3.2 Features

In the beginning of the process when the financial information of the organisations was collected, there were more than 170 financial and operational features for each company. So, the dimensions of the data are removed by an iterative feature removal process. At every step one feature was removed if the precision score of a simple classification task didn't change more than 1 %. By repeating this, more than 150 features were removed and 15 of the most important features were left.

3.3 Dataset for Prediction of Bankruptcy using Governance Indices

In this study, we are using non-financial features for the prediction of bankruptcy, more specifically the different Governance indices of the companies so the dataset that we worked on contained only the Governance indices as features which we train on and predict the mean of the indices based on which we can predict whether a company is bankrupt or not.

As described in the previous section, the dataset consists of data of 160,000 Italian SMEs, which each have identifying columns, that is the different governance indices such as the age of the company, the stakeholders of the company and so on.

The machine learning models are trained on this dataset of 160,000 companies with a training set of 9000 company samples. Then the models are tested for accuracy on the entire set of companies and finally tested to predict the bankruptcy of the 8959 companies.

3.4 Description of the Features or the Governance Key Performance Indicators:

A detailed list of the feature set can't be given due to a non-disclosure agreement. However, the below list gives a general idea of the types of features used to train our models.

- Partita IVA: Serial number given to the companies for privacy
- Number of Shareholders of a company
- Gender of shareholders
- Number of Managers and Top positions
- Number of Board Members
- Statutory board of Auditors
- Advisory Committee
- Number of Decision-making Shareholders
- Percentage of share in the same family
- Company age
- Revenues 2 years ago minus last year
- Revenues last year

These are some of the features we work with for training our machine learning models. We experiment with different combinations of these features and calculate the accuracy of the model and finally test the models by the best combination of features that return the maximum accuracy.

3.5 Data Pre-processing

The dataset containing the governance indices of 160,000 companies did not require any further cleaning or pre-processing as most of the data was processed when the data was collected by Arisk.

The dataset containing the information of 8641 companies contained some missing values. There are features of mixed data type integer of 64-bit and float 64-bit and 2 object types.

There are 23 missing values from the column LOGATT12 which we fill with the mean value of the column.

3.6 Feature Analysis

Data analysis or feature analysis is important to have a better perspective of the data we are going to work with. In our study we use multivariate analysis.

3.6.1 Multivariate analysis

When analyzing data statistically, it is crucial to consider the linkages and organizational structure of the multivariate measures because each experimental unit is subjected to numerous measurements.

In 1928, Wishart presented his paper. The accurate distribution of the sample covariance matrix for the multivariate normal population marks the beginning of MVA.

In the 1930s, a substantial body of theoretical fundamental work on multivariate analysis was generated by R.A. Fischer, Hotelling, S.N. Roy, B.L. Xu, and others.

In the fields of biology, education, and psychology, it was frequently used at the time.

Feature selection or dimensionality reduction seeks to reduce duplicate or irrelevant features by selecting more representative features with stronger discriminatory power for a given dataset (Dash & Liu, 1997; Guyon & Elisseeff, 2003)

Selecting the right features to train the model on is a very important step of machine learning because it impacts the performance of the models used. In our study we are focused on working with the non-zero features or governance indices in the dataset but finding the most important features and non-relevant features related to the problem to predict the target variable is why we need feature analysis.

Feature Analysis and selection of the right features does not only improve the accuracy of the model but also reduces overfitting. With less redundant data, there is less chance of making decisions based on noise. It also reduces the training time of the model as there are fewer data points to train from and reduces the complexity of the algorithm.

3.6.2 Feature Importance:

The dataset has a total of 21 features including the target variable ‘mean’ to be predicted. Using the feature importance property of any model we can find the most important features in the dataset. The feature importance function is an inbuilt class that comes with regression-based machine learning models, and it returns an importance score for each feature, so if the score is high, the feature is important and vice versa.

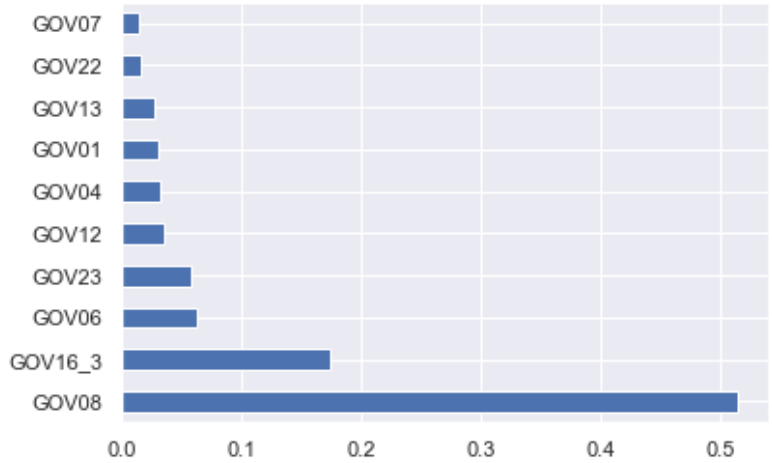


Fig 3. 1 shows the top 10 feature importance of the governance indices

For our dataset the feature importance function returns the 10 most import features as we can see in the figure 3.1 with GOV08 and GOV16_3 having the maximum importance respectively.

3.6.2 Correlation matrix with Heatmap

A corelation matrix visualizes the 2-dimensional relationship between the features or the features and the target variable. Corelation can be positive, i.e., two variables move in the same direction (both increases) or negative when the variables move in opposite directions (one increases, and one decreases). A heatmap associated with a corelation matrix helps to identify the features most related to each other and the target variable.

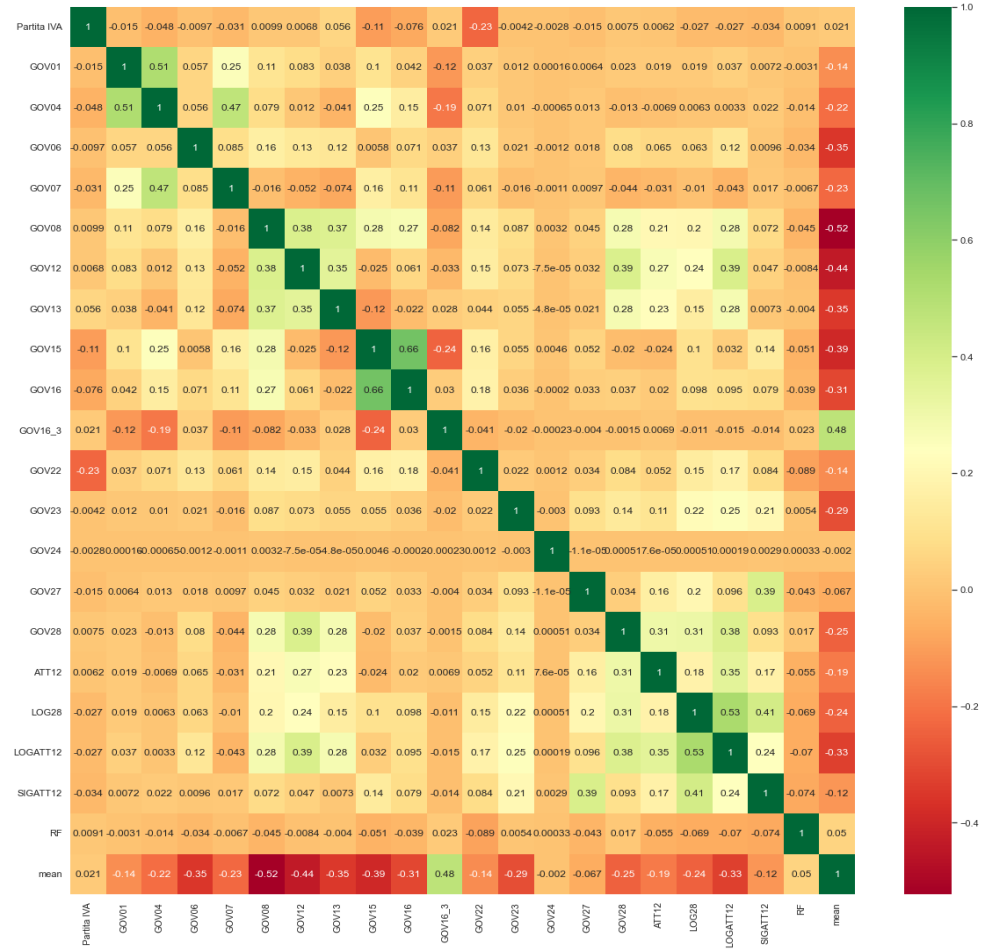


Fig 3.2. Shows the corelation matrix of all the features

From the Fig 3.2 we can again see that GOV08 has the maximum corelation with all other features and GOV16_3 has the maximum corelation with the target variable ‘mean’.

Since we will be working with the non-zero features of the dataset, we will further use the different machine learning models for feature analysis of different combination of features and find out the best combination.

3.7 Data Preprocessing

For the Machine learning and neural network models we do some pre-processing on the dataset containing 157000 companies.

The dataset doesn't contain any missing values as it had been cleaned after being extracted from the public databases by Arisk. However, we do some basic pre-processing according to the requirements of our experiments.

There are many ways to encode categorical variables which cannot be processed by machine learning models. We do one hot encoding on particular feature GOV06. It is not a categorical value; however, it has 3 unique values which represents 3 forms of GOV06.

One-hot-encoding is an encoder which is a part of the Scikit-learn library. Using this method, each category value is transformed into a new column and given a 1 or 0 (true/false) value.

'1' (denoting true) will appear in rows with the first column value, while '0' will appear in rows with other values (indicating false). The same applies to additional rows where the value matches the column value.

Although this method resolves the hierarchy/order difficulties, it has the drawback of increasing the number of columns in the data collection. If a category column contains a lot of distinct values, the number of columns may increase significantly. As we only encoded one feature in our study, it was manageable, but when numerous columns are encoded, it becomes very difficult to handle.

As a result of the encoding we have 3 more features representing GOV06 which will be used to train and test the models.

3.7.1 Handling Missing Values:

We are some missing values in the LOGATT12 column of the dataset containing 8641 companies.

Given that many machine learning algorithms do not allow missing values, managing missing data properly during the preparation of the dataset is crucial.

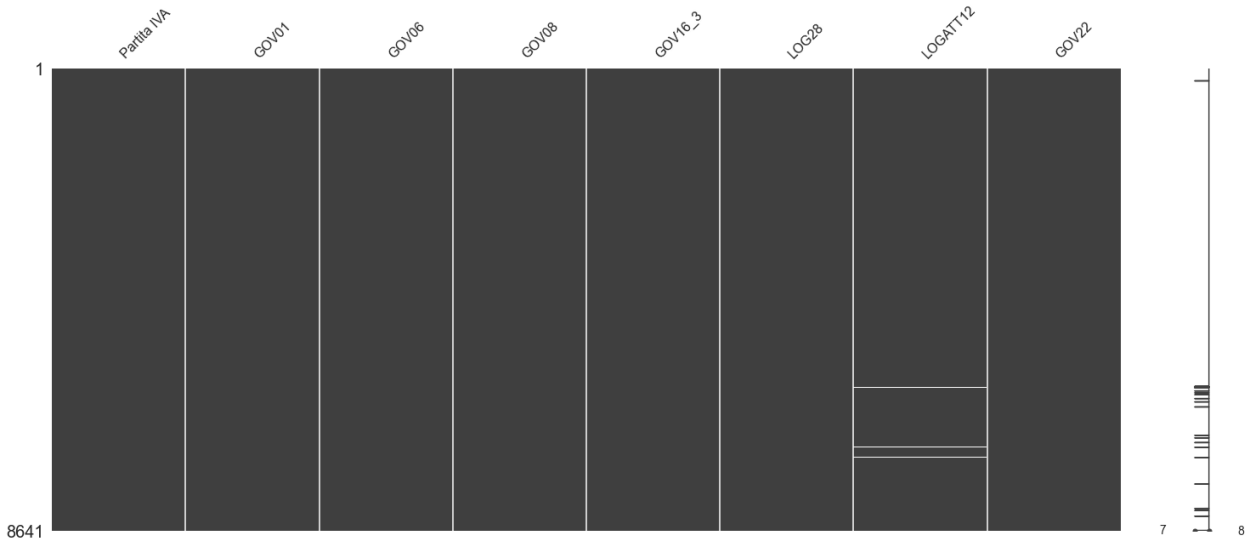


Figure 3.3 missingno matrix

Using the library missingno we can visualize the missing values from LOGATT12 column.

We cannot delete the rows containing the missing values because each row represents the governance indices of a different company.

3.7.2 Imputing the missing values with Mean

The columns in the dataset that include continuous numeric values can be replaced by the mean, median, or mode of the remaining values in the column. This method can prevent data loss as compared to the earlier one. The two estimations given above can be swapped out as a statistical solution to the missing values (mean, median). For our study we will use the mean of the LOGATT12 column to replace the missing values.

Chapter 4

Methodology

The first part of the chapter covers the introduction to Machine Learning and the study of traditional and deep learning algorithms. The second part compares the models used for this study and their detailed discussion.

We outline the theoretical foundation upon which the models are created in the section that follows. We first discuss the techniques from a classification perspective. This work does not attempt to demonstrate or derive every model's mathematical component.

4.1 Introduction to Machine Learning

By definition, machine learning is a type of artificial intelligence that uses algorithms and statistical methods to learn from data and continuously increase its accuracy to perform a certain set of tasks imitating human or natural behaviour.

Mostly machine learning models work as a learning process by analysing patterns in data which are used to make a classification or prediction decision from the information provided. The models work by using a loss function which determines how much the predicted value has deviated from the actual value and hence determine the accuracy of the models. Finally, an optimization function is used to reduce the errors in prediction. The evaluation and optimization processes are repeated until the desired accuracy is obtained.

4.1.1 The different machine learning methods

In our study we use supervised and unsupervised learning for prediction of bankruptcy. The section below gives a general idea of the different kinds of machine learning methods and the techniques used in our study.

4.1.1.1 Supervised Learning

In supervised learning, the machine learning algorithms train by using labelled data that maps an input to an output in the training part. (Hastie et al., 2009 [61])

Supervised learning is classified mainly into two categories:

- In Regression problems the output variable is real or continuous in nature.
- In Classification problems the output variable is a discrete or categorical value.

Some examples of supervised machine learning models are: Logistic regression, support vector machines, random forest, decision tree etc.

Regression models can be further divided into linear and non-linear regression models:

- In Linear regression models there is a constant and a parameter multiplied by an independent variable

$$y = a + bx \qquad \textit{Eqn 4. 1}$$

where y is the dependent variable, a is the constant and x is the independent variable.

- In Non-linear regression models, the equation does not follow the rules of linear regression using a curve to fit the equation with the data

$$y = f(X, \beta) + \in \qquad \textit{Eqn 4. 2}$$

where β is the estimated non-linear parameter, $\in$ is the error and f is the regression function.

In this thesis, non-linear supervised learning models are one of the methods used to predict the bankruptcy of the companies because we are working with a regression problem which is more efficiently solved using supervised learning.

In our study we use also use supervised learning Artificial neural network to see if it can outperform regression models as the dataset is very large with a lot of complex non-linear relationships.

4.1.1.2 Unsupervised Learning

In unsupervised machine learning, the model infers hidden patterns from data. The models are trained with unlabelled datasets and there is no human supervision. They find similarities and differences in the underlying structure of data and learn from their own experiences of interacting with the data by clustering the unlabelled datasets. (Friedman et al., 2001)

Unsupervised learning is mainly divided into two categories:

- In Clustering, objects with maximum similarities are clustered with the same category labels.
- In Association the unsupervised model learns to find relations between the different variables in a dataset and associated them together.

Some examples of unsupervised models are, k-means clustering, principal component analysis, neural networks etc.

4.1.1.3 Semi supervised Learning

Semi-supervised learning falls somewhere between supervised and unsupervised machine learning. It uses a mixture of a smaller labelled data to train the model and extracts features from a larger unlabelled dataset. The unlabelled data is clustered into groups with similarities using unsupervised machine learning algorithms and then the data are labelled using the small, labelled dataset available using supervised learning.

There are two different types of semi-supervised learning:

- Inductive machine learning also known as concept learning is a type of algorithm that learns from labelled data and generalises to new data.
- Transductive machine learning, estimates the values of the unknown function for points of interest in the given data and describes a new inference.

Some examples of semi-supervised models are, k-nearest neighbours, neural networks in natural language processing etc.

4.1.1.4 Reinforcement Learning

Reinforcement learning is a feedback-based machine learning technique that trains a model to make a sequence of decisions using trial and error. There is no previous training, and the datasets are unlabelled, hence the model learns by experience only.

There are two types of reinforcement learning:

- Positive reinforcement: Here a behaviour is added such that the expected tendency of a positive behaviour will increase.
- Negative reinforcement: Here a particular behaviour is strengthened by stopping or avoiding an unexpected behaviour.

Some examples of reinforcement learning algorithms are, Q-learning, State-Action-Reward-State-Action (SARSA), Deep Q Network (DQN) etc.

4.2 Traditional Machine Learning methods

In this section we describe in detail the concepts of traditional machine learning techniques used in our study. We have used Random Forest regressor model and Decision Tree regressor model which are based on supervised learning and compared their accuracy.

4.2.1 Decision Tree Regressor Model

Decision tree is a type of supervised predictive modelling which can be used for both regression and classification problems in organisational decision making. It uses a tree like structure and the dataset is broken down into increasingly smaller subsets while the associated decision tree is gradually developed. (Bishop, 2006)

For a regression prediction problem, the decision tree uses predictive analysis to forecast output from unknown input data and gives an output which is continuous. The model is trained such that it understands the relationship between the independent variables and the outcome.

The root node represents the entire sample, and this gets divided into smaller sets. Splitting of the root node leads to the interior nodes which represent the feature of the dataset. The leaf nodes are the terminal nodes which require no further splitting.

The decision tree works by answering true and false questions and arriving to the final leaf node or the prediction. This is done by asking questions about the data and narrowing down the possible answers. When the model is training on the dataset it learns to map data to the output through these questions predicting any relationship between the data and the target variable. After the model is trained it learns to ask the most accurately related questions on the test data and predicts the most accurate answers hence the prediction is an estimate of the training data provided to the model.

Since bankruptcy prediction is a problem based on regression and the predicted variable should be continuous in nature, answering questions in the form of true and false doesn't solve the problem, so we employ the Decision Tree Regressor Model.

The Decision Tree Regressor Model uses different kinds of errors like the MAE (Mean Absolute error), MSE (Mean squared error) and the RMSE (Root Mean Squared Error) that calculate how much the model's prediction have deviated from the actual values. The error here is calculated as:

$$Error = \gamma_t - \hat{\gamma}_t \qquad \textit{Eqn 4.3}$$

where γ_t is the actual historical data and $\hat{\gamma}_t$ is the forecasted data.

Mean Absolute Error: MAE is a statistical error that finds the absolute value of the difference between each pair of the actual historical data and the data predicted by the model and returns the average.

$$MAE = \sum_{i=1}^{n} \frac{|Error|_i}{n} \quad \textit{Eqn 4.4}$$

Mean Squared Error: MSE also returns the average of the absolute value of the square of the difference between the pairs of the actual values and the predicted values. The error is squared to prevent from cancelling each other.

$$MSE = \sum_{i=1}^{n} \frac{(Error^2)_i}{n} \quad \textit{Eqn 4.5}$$

Root Mean Square Error: RMSE is the square root of the MSE.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Error^2)_i}{n}} \quad \textit{Eqn 4.6}$$

4.2.1.1 Mathematical concept of Decision Tree Model:

There are several algorithms used to make a decision tree. For a regression model, the algorithm CART is used. It uses mean square error as a measure to choose feat ures in a regression model. The target variable to predict is continuous or discrete. There are three statistical calculation involved to generate a Decision tree regresso r model:

- Standard Deviation for building the branches of the tree
- Coefficient of Variation to decide when to stop branching

- Average value in terminal nodes

4.2.1.2 Feature analysis using Decision tree

In the previous chapter we did a generalised feature analysis using the feature importance function and the corelation matrix. In this section we will try to determine the best combination of features the Decision Tree model can be trained based on the different errors as well as the accuracy of the model in predicting the target variable.

The data obtained from Arisk is imported using pandas and stored in a pandas dataframe. Then we assign the different government indices to the independent variable and the mean column to the dependent variable.

In the next step the dataset is split into the training and test set to avoid bias while training the algorithms with a test size of 0.3, i.e., 30% of the total number of samples chosen.

First, we choose 12,000 samples used which is 3600 companies for test set. The remaining 8400 companies will be used as a training set for building the model.

Then we choose 15,000 samples used which is 4500 companies for test set. The remaining 10,500 companies will be used as a training set for building the model.

The random state is set to 42 so that we get the same training and test sets across different execution.

The DecisionTreeRegressor class is imported from the scikit-learn library and then the independent variable train set, and dependent variable train set is fitted on it. Next, we predict the results on the test set.

We have used different combination of features:

- Only non-zero features
- Only non-zero and log features
- All the features

For this combination of features, we will see the errors returned by the Decision Tree Regressor Model and their accuracy.

For **only non-zero** features and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.1, Errors for non zero features DT

Samples	MAE	MSE	RMSE
12,000	0.06190277777777778	0.007278368055555554	0.08531335215284624
15,000	0.06288425925925926	0.007417669753086419	0.08612589478830637

The following graphs shows the actual versus predicted values based on the errors on the above table 4.1:

The red line represents the actual values and the blue line represents the predicted values

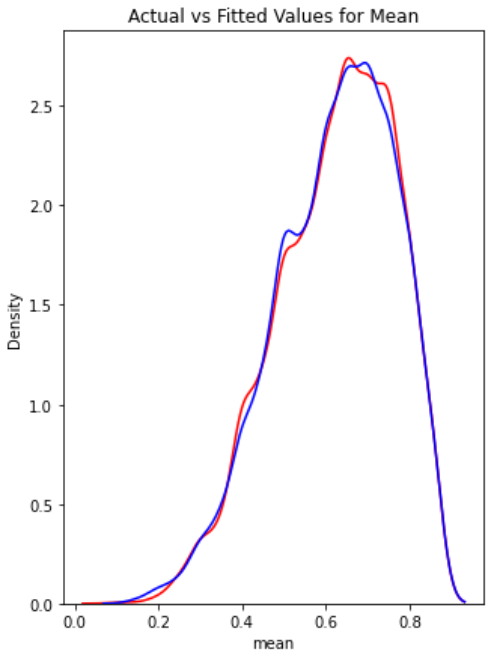


Fig 4.1 : for 12000 samples, Accuracy: 88.53%

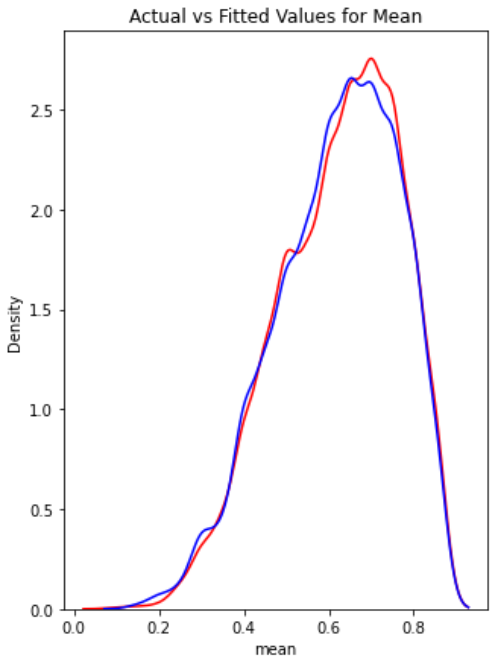


Fig 4.2 : for 15000 samples, Accuracy: 88.58%

For **only non-zero and log features** and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.2: Errors for non zero and log features DT

Samples	MAE	MSE	RMSE
12,000	0.0483111111111 1115	0.0053544444444 44444	0.073174069481233 88
15,000	0.0459444444444 44475	0.0051180555555 555545	0.071540586771115 84

The following graphs shows the actual versus predicted values based on the errors on the above table 4.2:

The red line represents the actual values and the blue line represents the predicted values

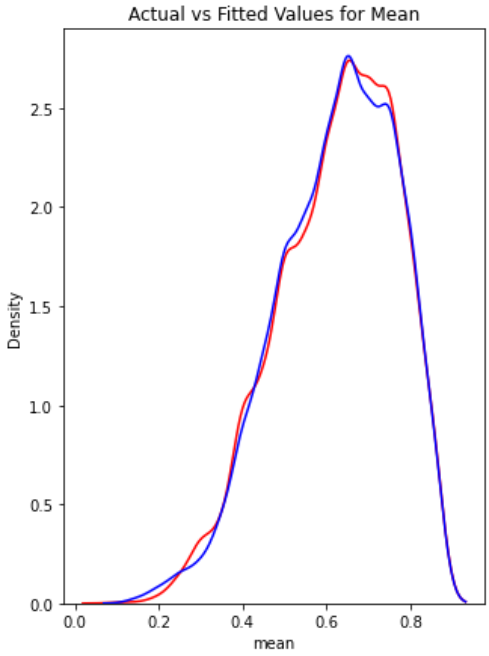


Fig 4.3, for 12000 samples, Accuracy: 90.77%

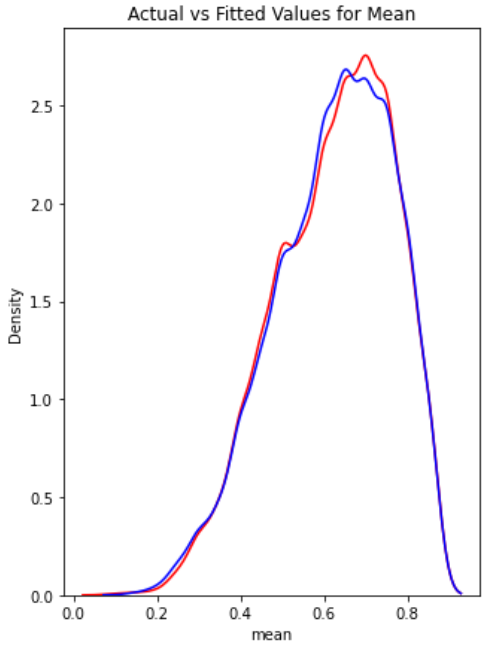


Fig 4.4, for 15000 samples, Accuracy: 91.1%

For **all features** and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.3, Errors for all res features DT

Samples	MAE	MSE	RMSE
12,000	0.020211111111111388	0.0016883333333333334	0.04108933357129723
15,000	0.022194444444444617	0.0018902777777777776	0.043477324869151936

The following graphs shows the actual versus predicted values based on the errors on the above table 4.3:

The red line represents the actual values, and the blue line represents the predicted values

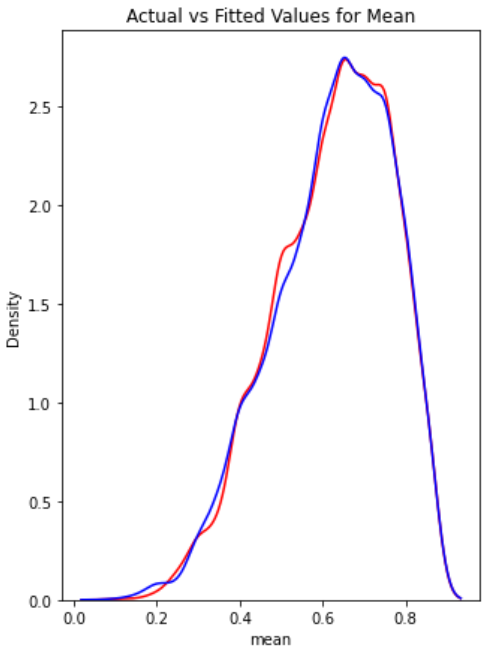


Fig 4.5, for 12000 samples, Accuracy: 95.43%

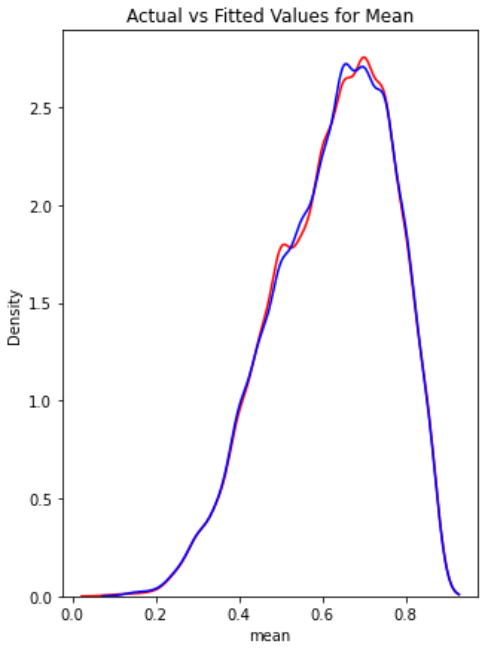


Fig 4.6, for 15000 samples, Accuracy: 95.8%

4.2.2 Random Forest Regressor model

The Random Forest Regressor model is based on ensemble learning. In ensemble learning multiple machine learning models are trained and their predictions are

combined for maximum accuracy, Ho (1995). There are two types of ensemble learning, Bagging models such as Random Forest which work only on a fraction of the entire dataset and Boosting models such as AdaBoost which work on the entire dataset.

It is called a Random Forest because a forest is created of random decision trees. So, the random forest combines the output of several decision trees to generate the final output. In our case we are using 1000 random decision trees as the number of estimators. We are fixing the random state of the random forest to 42 so that the same decision forests are selected across different executions.

Just like the Decision tree regressor model, the Random Forest uses errors like Mean square error, Mean absolute error and Root mean square error to calculate how much the predicted values have deviated from the actual values.

4.2.2.2 Feature analysis using Random Forest Regressor Model

The dataset is split into training and test set. To avoid bias while training the model, the test set is set to 0.3 that is 30% of the total samples we will train our model on which is 3600 of the 12000 samples used and 4500 of 15000 samples. The remaining data is used as the training set. The random state is set to 42 so that we get the same training and test sets across different execution.

The scikit-learn library is used to import the RandomForestRegressor class, which is then fitted with the independent variable train set and the dependent variable train set. The results of the test set are then predicted.

We have used different combination of features:

- Only non-zero features

- Only non-zero and log features
- All the features

For this combination of features, we will see the errors returned by the Random Forest Regressor Model and their accuracy.

For **only non-zero** features and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.4, Errors for non-zero features RF

Samples	MAE	MSE	RMSE
12,000	0.04801045004960315	0.0039006543061367734	0.06245521840596487
15,000	0.04689941169793168	0.0037702021605824096	0.0614019719600471

The following graphs shows the actual versus predicted values based on the errors on the above table 4.4:

The red line represents the actual values, and the blue line represents the predicted values

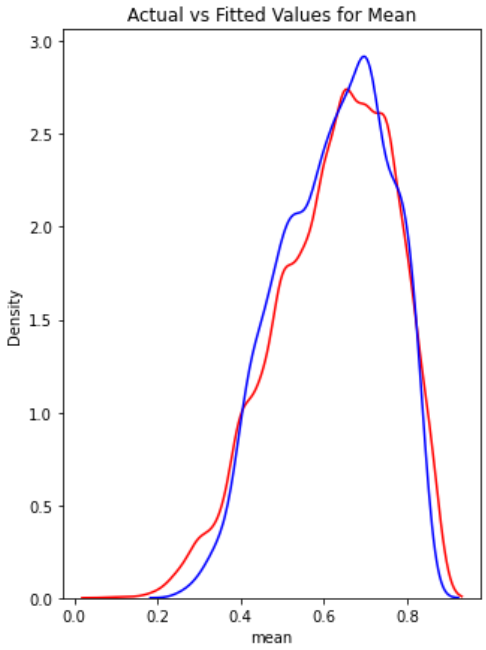


Fig 4.7, for 12000 samples, Accuracy: 91.23%

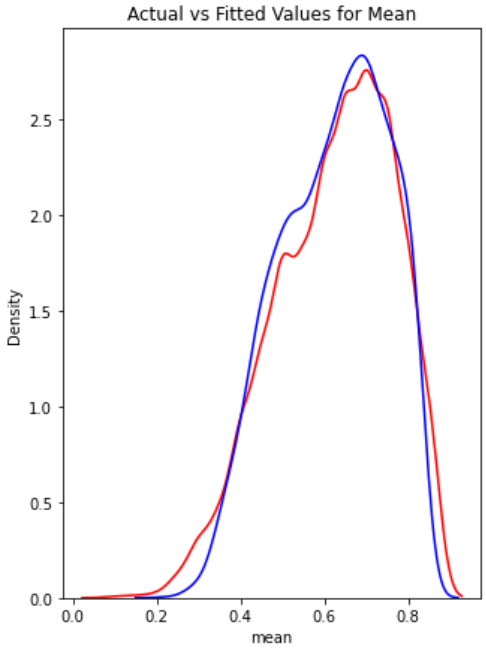


Fig 4.8, for 15000 samples, Accuracy: 91.35%

For **only non-zero and log** features and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.5, Errors for non-zero and log features RF

Samples	MAE	MSE	RMSE
12,000	0.037584236111111585	0.002806152999305563	0.052973134693970705
15,000	0.037712255555555582	0.002795424367222183	0.05287177287761607

The following graphs shows the actual versus predicted values based on the errors on the above table:

The red line represents the actual values, and the blue line represents the predicted values

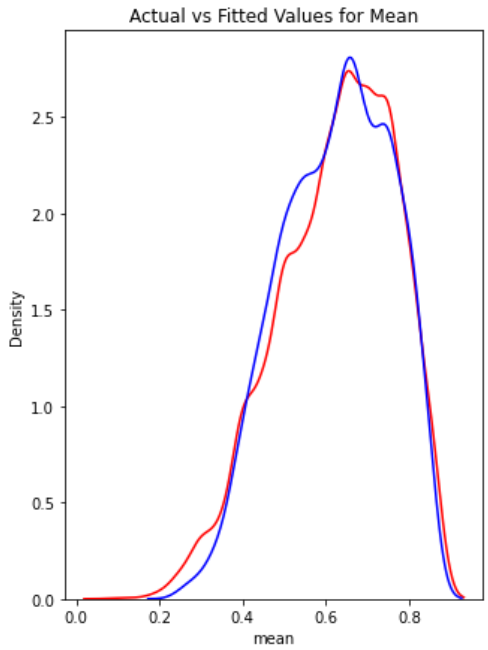


Fig 4.9, for 12000 samples, Accuracy: 92.88%

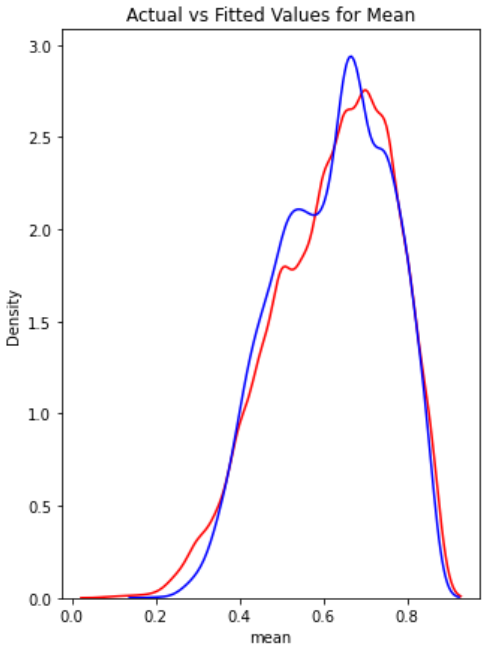


Fig 4.10, for 15000 samples, Accuracy: 92.85%

For **all** features and a sample of 12000 and 15000 companies respectively, the errors are returned as follows:

Table 4.6, Errors for all features RF

Samples	MAE	MSE	RMSE
12,000	0.0189210833333347	0.0009758796069444532	0.031239071800302474
15,000	0.0178540111111112743	0.0009218993927777822	0.03036279619497819

The following graphs shows the actual versus predicted values based on the errors on the above table 4.6:

The red line represents the actual values, and the blue line represents the predicted values

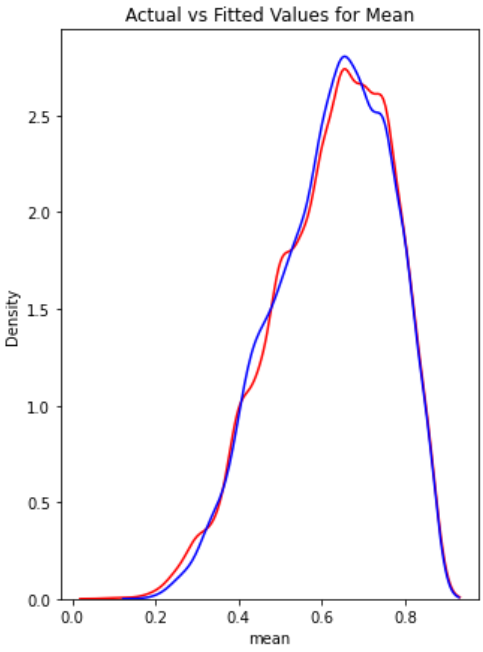


Fig 4.11, for 12000 samples, Accuracy: 96.21%

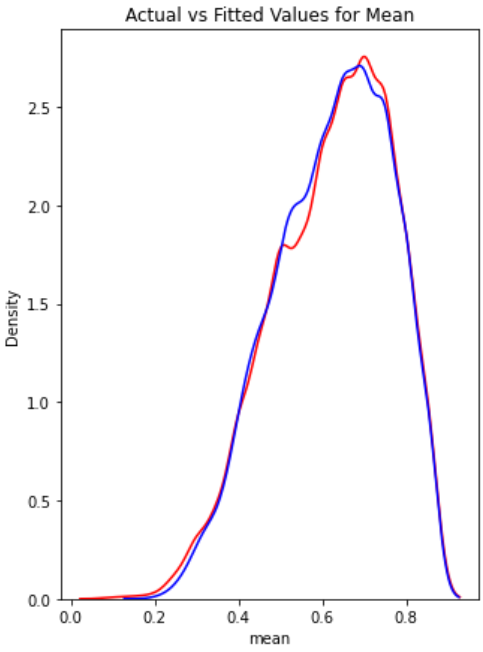


Fig 4.12, for 15000 samples, Accuracy: 96.36%

4.2.3 Comparison of accuracies for Decision Tree and Random Forest Models:

Out[75]:

	Accuracy	model
0	91.229557	Random forest Regressor with only non zero fea...
1	92.877445	Random forest Regressor non zero and log features
2	96.213082	Random forest Regressor with all features
3	95.403988	Decision Tree Regressor with only non zero fea...
4	91.101197	Decision Tree Regressor non zero and log features
5	95.431914	Decision Tree Regressor with all features

Table 4.7, Accuracies for 12,000 samples

Out[59]:

	Accuracy	model
0	91.346202	Random forest Regressor with only non zero fea...
1	92.845247	Random forest Regressor non zero and log features
2	96.362280	Random forest Regressor with all features
3	88.576844	Decision Tree Regressor with only non zero fea...
4	90.774006	Decision Tree Regressor non zero and log features
5	95.797623	Decision Tree Regressor with all features

Table 4.8, Accuracies for 15,000 samples

From the figures above we can see that after feature analysis using regression analysis, over all Random Forest Regressor Model has a higher accuracy in both the cases. The accuracy is highest when we use all the features, that is 17 features however our study emphasises on only non-zero features. So out of the other two cases, the combination of non-zero features with log features is giving a higher accuracy.

So, we will use these 10 features for the final training and testing of the dataset.

4.2.3 Artificial Neural Networks

Artificial neural networks, a class of algorithms inspired by the structure and operation of the brain, are the focus of the machine learning discipline known as deep learning. It is made up of sigmoid neuron layers that have been piled together to create a larger architecture.

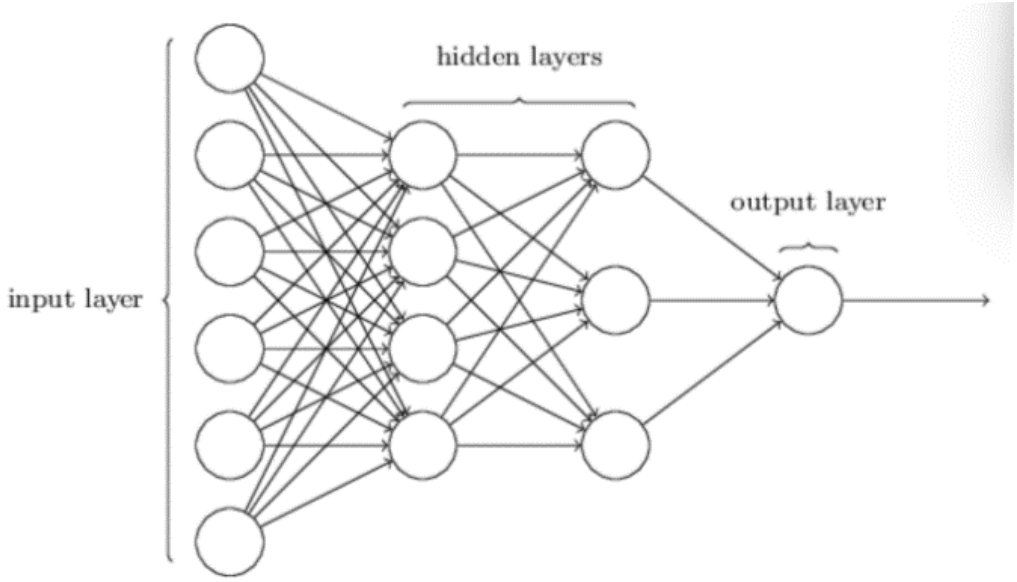
Sigmoid Neuron:

Sigmoid neurons are the basic building blocks of Artificial Neural Networks. Here the output function is much smoother than the step function so a small change in input only leads to a small change in the output.

The use of sigmoid neurons is applicable to non-linear regression problems. The output of the sigmoid neuron is given as,

$$\frac{1}{1+\exp\left(-\sum_j w_j x_j - b\right)} \qquad \text{Eqn 4.7}$$

4.2.3.1 Structure of an Artificial Neural Network



The figure 4.13, shows the structure of an Artificial Neural Network (Matel, et al, 2019 [62])

As shown in Fig 4.13 the ANN that we are using has 2 hidden layers, with number of neurons equal to 2/3rd of the input features and one output layer with one neuron as we have only one target variable to predict.

The hyperparameters used to construct the ANN is explained below:

- A series of stacked ANN layers is produced using the "Sequential" module from the Keras library.
- The "Dense" module of Keras is used to define each connected layer where the following arguments are passed:

    - The number of neurons or layer of neurons which we calculate as 2/3rd of the number of input features

- The input dimension or the input predictor features which is expected by the first layer only, from the second layer onwards, the sequential layer passes the output of the previous layers as input.
- Kernal initializer which can of types normal, glorot_uniform etc. It calculates the weight of the neurons. We have used normal algorithm for kernel initialization.
- An activation function is used to specify how the calculations are to be made inside each neuron. We use the function relu however there are other kinds of activation methods such as tanh, sigmoid etc.
- We specify the batch size which decides how many rows of data will be passed before the neural network will start adjusting it's weights based on the errors returned.
- When also specify the number of epochs or the number of times the neural network looks through the entire dataset in numbers of the specified batch size.

4.3.3 Hyperparameter tuning of Artificial Neural Network:

It is crucial to determine the ideal sizes for batch size and epoch because they have a direct impact on the performance of the model. Overfitting or underfitting might result from poor values. In this study Manual grid search is used however there are other techniques of tuning too.

There isn't a general guideline that can be used to determine the number of layers, neurons, etc. after looking at certain data. We experiment with several parameters and pick the one that results in the maximum accuracy.

The more computationally complex the network is, the longer it will take to run. Therefore, it is important to always seek out the best accuracy using the fewest layers/neurons.

For 12,000 samples,

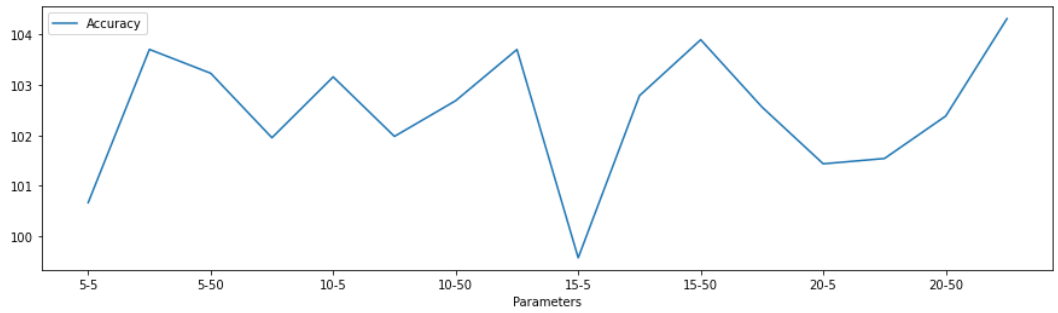


Fig 4.14, shows batch size 15 with epoch 50 giving the maximum accuracy, so we will use this to train the model for 12,000 samples

For 15,000 samples,

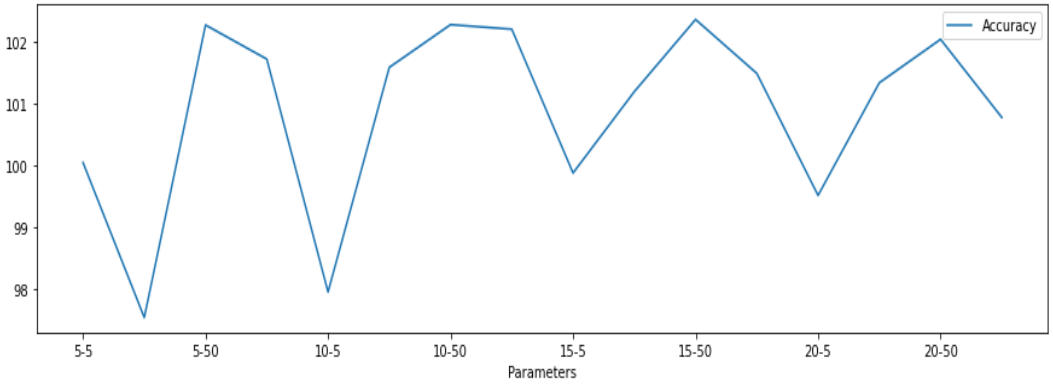


Fig 4.15, also shows batch size 15 with epoch 50 giving the maximum accuracy, so we will use this to train the model for 15,000 samples

With the best set of hyperparameters we train the neural network model.

4.3.4 How the neural networks learn:

The weighted total of the inputs is what neural networks are. Thus, adjusting these weights is the foundation of neural network learning. A process for updating the weights is required. It depends on how well the neural network is functioning. How well the predictions based on the actual features that need to be anticipated turn out is referred to as the neural network's performance. By traversing the neural network and ascertaining each neuron's value, the value at the output layer is determined. Forward propagation is the term used for this neural network crossover procedure.

We introduce loss function, which determines how poorly a neural network performs, for measuring performance. In our model we use Mean Absolute percent error to see how much the predicted value has deviated from the actual value and then calculate the accuracy of the model.

Gradient descent is used to determine the weights that need to be changed in the neural network. It is used to adjust the neural network's weights in relation to the determined loss once the loss has been calculated. The weights are updated by back propagating through the neural network based on how much each neuron contributed relatively to the initial output. Until every neuron in the network has received a loss signal that represents their proportionate contribution to the overall loss, this procedure is repeated, layer by layer.

The model is trained using 12000 and 15000 samples respectively. 30% of the samples are used for the test set. The random set is set to 42 so that it uses the sample samples for all executions.

We use the non-zero and log features which has been obtained as the best combination of features after regression analysis to train the model. So, we use 7 neurons and input dimension as 10 which are the features used to train the model.

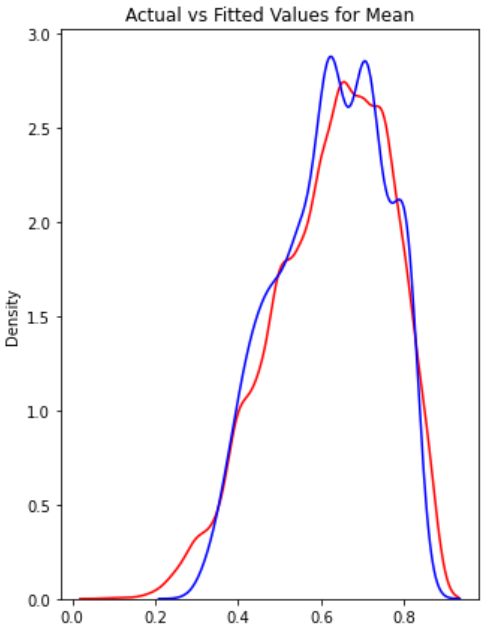


Fig 4.16, for 12000 samples, Accuracy: 91.4%

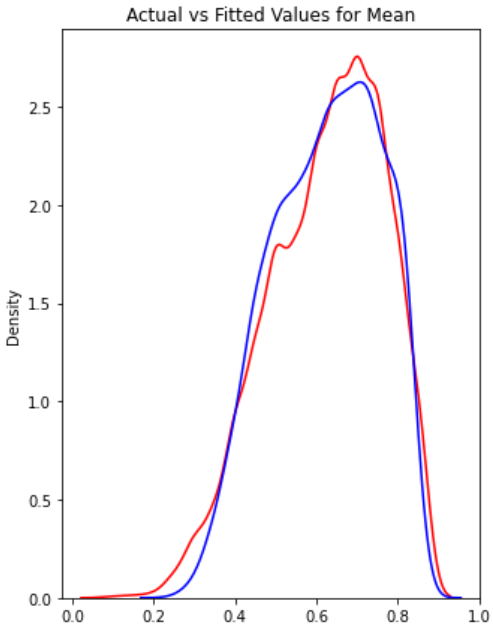


Fig 4.17, for 15000 samples, Accuracy: 91.6%

We can see the accuracy of the Artificial neural network model is greater than the Decision tree regressor model but lesser than the Random Forest regressor model.

Now, we will apply the model on the entire dataset and see how they perform.

Chapter 5

Testing the Machine Learning Models

In this chapter we are going to test the traditional machine learning models and the neural networks on the entire dataset of 157,000 companies based on the best combination of features we have obtained from our feature analysis in the previous chapter. The model that gives the best accuracy will be used on the final dataset of 8641 companies to predict the bankruptcy.

5.1 Decision Tree Regressor Model:

The model was trained using 15000 samples because it gives more accuracy. Now we test it on the entire dataset.

Fig 5.1 Density plot of accuracy for Decision tree on entire dataset

In the above fig 5.1 we see a plot of the actual versus predicted values where the actual values are in red and predicted values in blue.

The model gives an accuracy of 91.5 % with the following errors:

Mean Absolute Error: 0.04448332070473926

Mean Squared Error: 0.0049549668578996635

Root Mean Squared Error: 0.07039152546933233

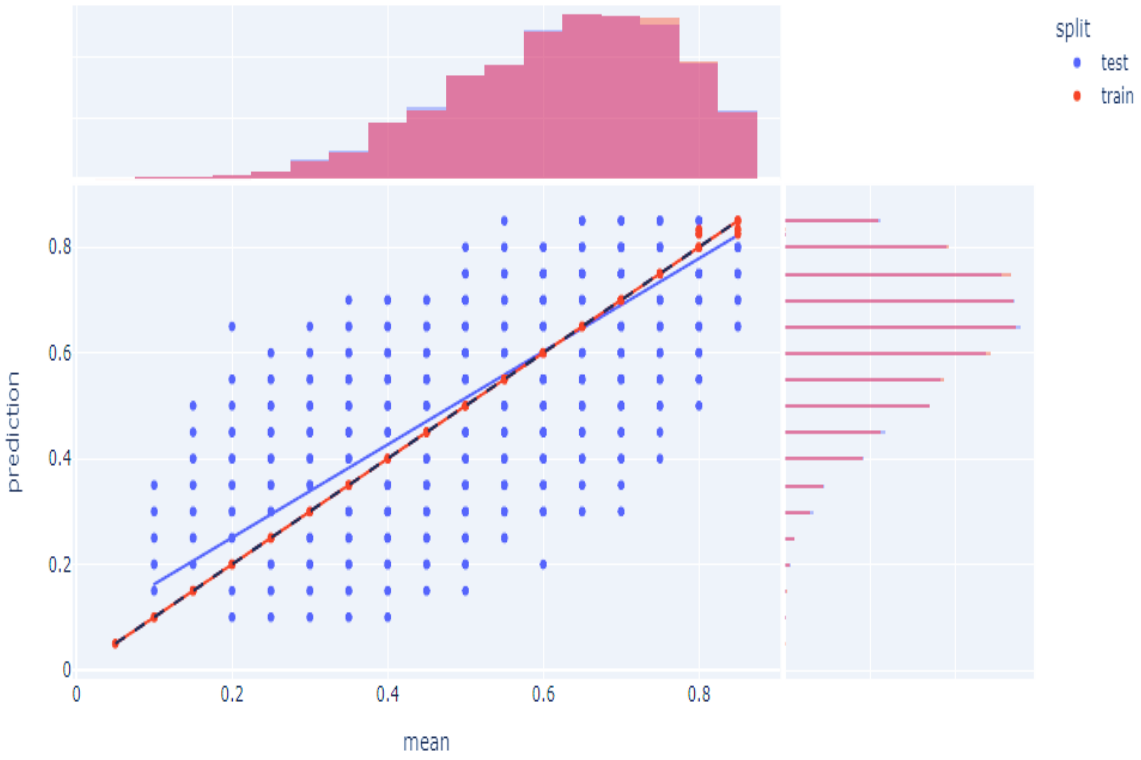


Fig: 5.2 prediction error analysis of Decision Tree Model

The above fig 5.2, shows enhanced prediction error analysis. The histogram shows the prediction bias in our model. It also shows how the model does compared to the optimal fit(black line) in the middle.

We can see the blue line almost coinciding with the black line which shows the model is performing well. There are less red dots compared to the blue dots because the red dots represent the training values which are less than blue dots representing the test set.

The dots coinciding with the black line are where the predicted values match exactly with the real values and the further the dots go from the black line the more difference there is between the actual and predicted values.

5.2 Random Forest Regressor Model

The model was trained using 15000 samples because it gives more accuracy. Now we test it on the entire dataset.

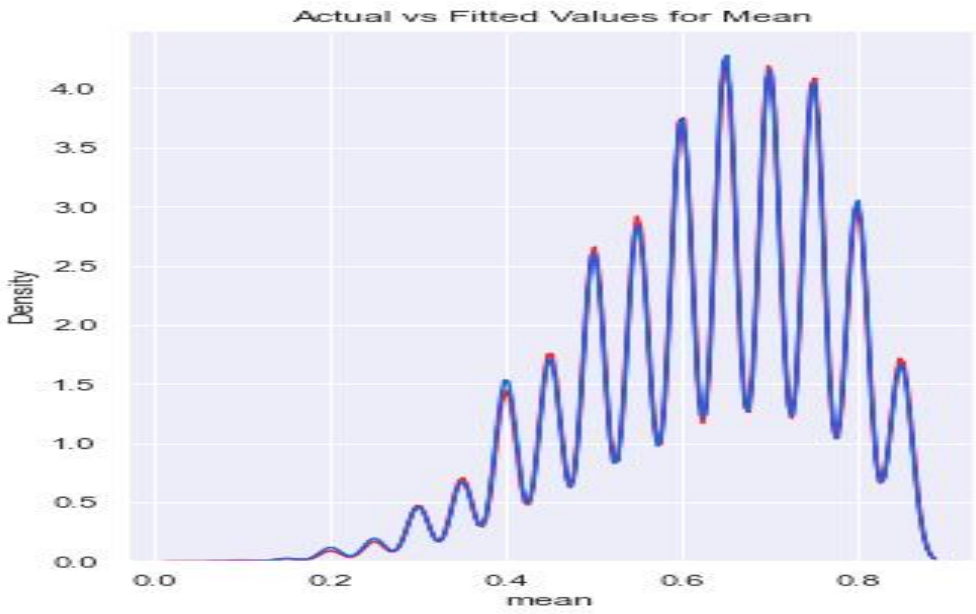


Fig 5.3 Density plot of accuracy for Random Forest on entire dataset

In the above fig 5.3 we see a plot of the actual versus predicted values where the actual values are in red and predicted values in blue.

The model gives an accuracy of 93.17 % with the following errors:

Mean Absolute Error: 0.03576763949290398

Mean Squared Error: 0.002633290638980329

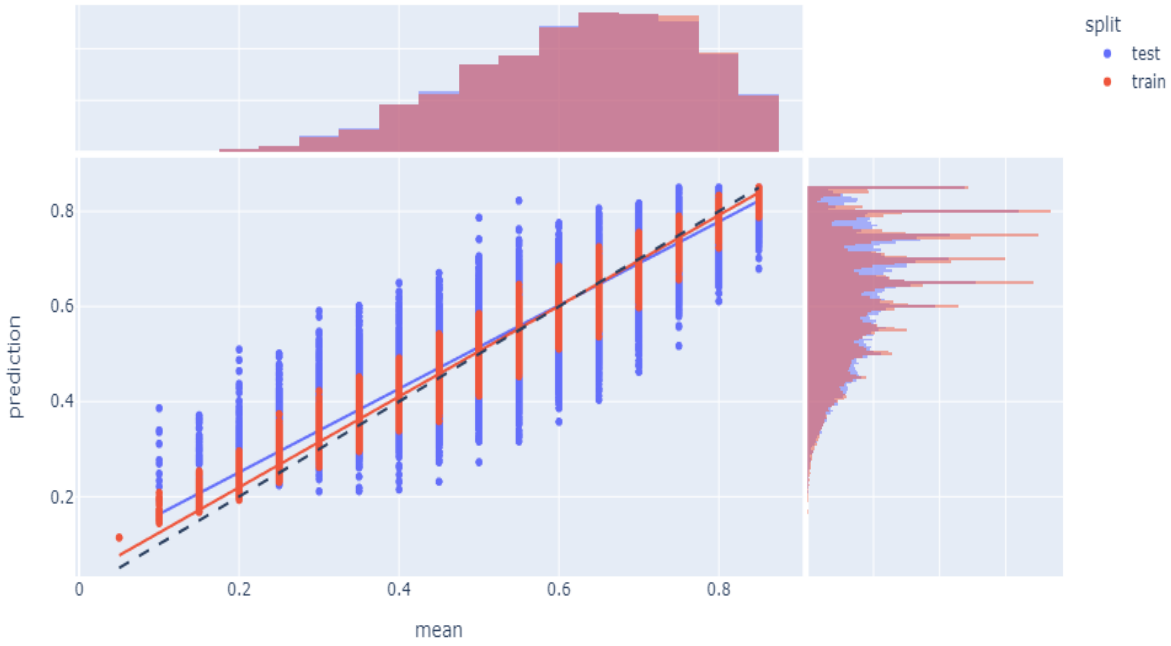


Fig 5.4 prediction error analysis of Random Forest Model

The above fig 5.4, shows enhanced prediction error analysis. The histogram shows the prediction bias in our model. It also shows how the model does compared to the optimal fit(black line) in the middle.

We can see the blue line almost coinciding with the black line which shows the model is performing well. There are less red dots compared to the blue dots because the red dots represent the training values which are less than blue dots representing

the test set. We can see more red dots coinciding the blue dots which means the accuracy of the Random Forest is better than the Decision Tree.

The dots coinciding with the black line are where the predicted values match exactly with the real values and the further the dots go from the black line the more difference there is between the actual and predicted values.

5.3 Artificial Neural Network

Initially after the model was trained using 15000 companies

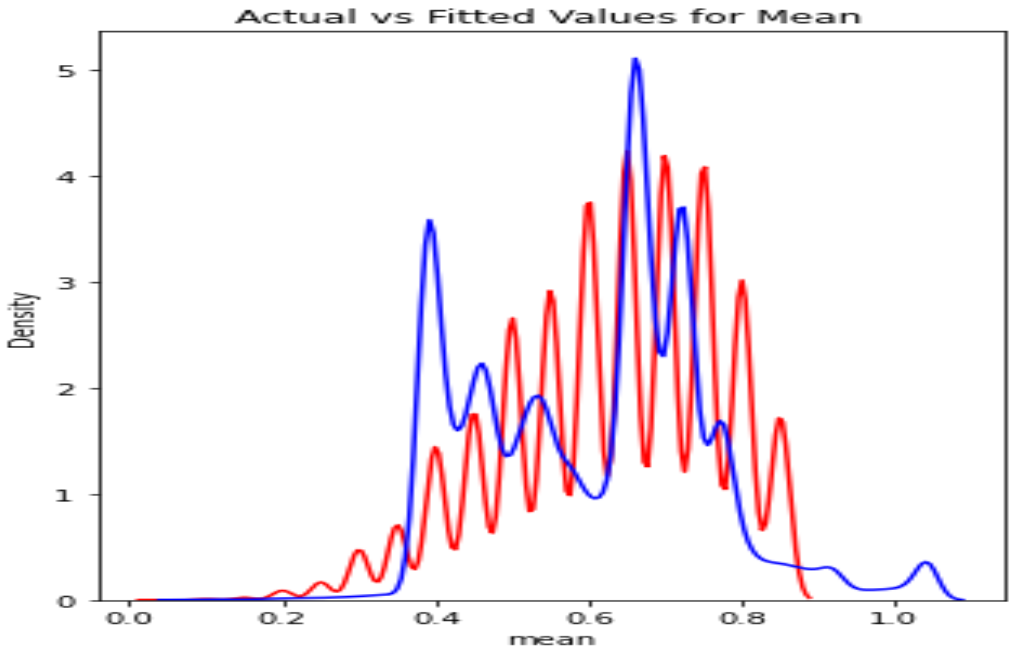


Fig 5.5 Density plot of accuracy of ANN for entire dataset

In the above fig 5.5 we see a plot of the actual versus predicted values where the actual values are in red and predicted values in blue.

The model gives an accuracy of 60.32% with the following errors:

Mean Absolute Error: 0.2833535139924623

Mean Squared Error: 0.09661425405005991

Root Mean Squared Error: 0.31082833533971754

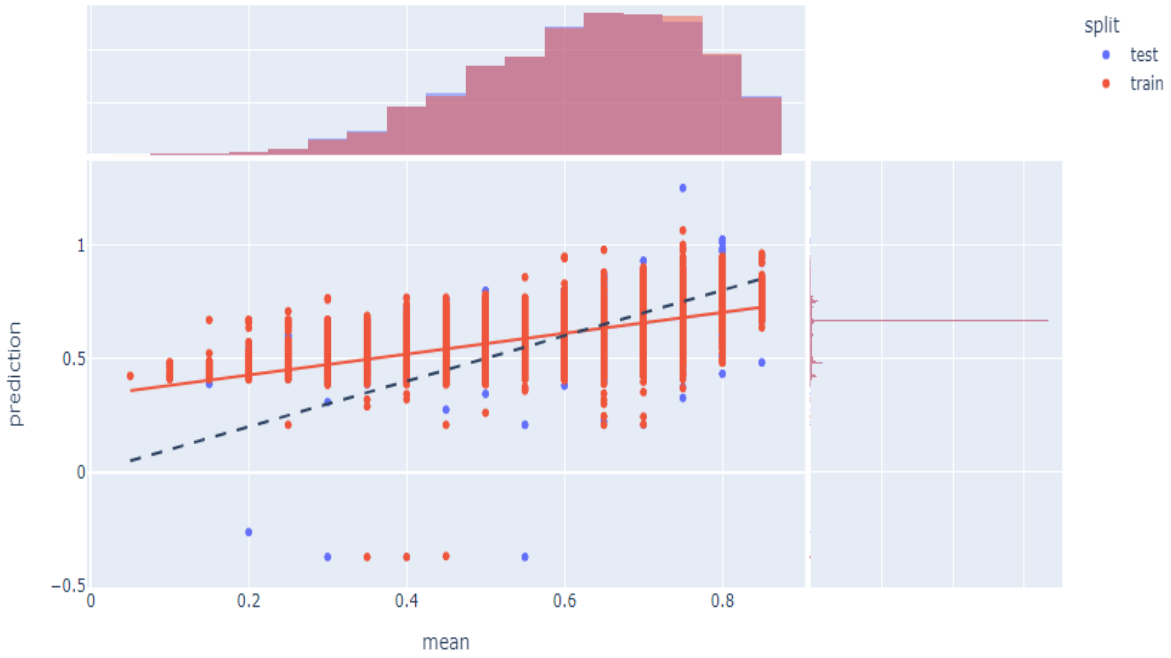


Fig 5.6, prediction error analysis of ANN

The above fig 5.6, shows enhanced prediction error analysis. The histogram shows the prediction bias in our model. It also shows how the model does compared to the optimal fit (black line) in the middle.

We can see the blue not coinciding with the black line which shows the model is not performing well. There is a lot of prediction bias in the model due to overfitting

as we can see from Fig 5.5 which shows certain sharp rises in the curves showing the model is not performing well on the new data.

5.3.1 Overfitting in Neural Networks

When a model is overfitted, it attempts to learn both the noise in the training data and too many details from it. The model's performance on unknown or test datasets suffers as a result. As a result, the training dataset's characteristics and patterns are not generalized by the network.

Deep neural networks are prone to overfitting since they learn millions or billions of parameters while building the model. A model with this many parameters may overfit the training set of data due to its significant capability.

The main answer to the overfitting problem is to simplify the model. Among other approaches, we can simply lower the number of neurons or remove layers from the network to achieve this.

We will explain the different ways we have tried to reduce over fitting and improve the performance of the Artificial Neural Network.

5.3.2 Data Augmentation

We have increased the size of the training sample from 15,000 to 50,000 because the network usually has more control over the training data when the training data is minimal in size.

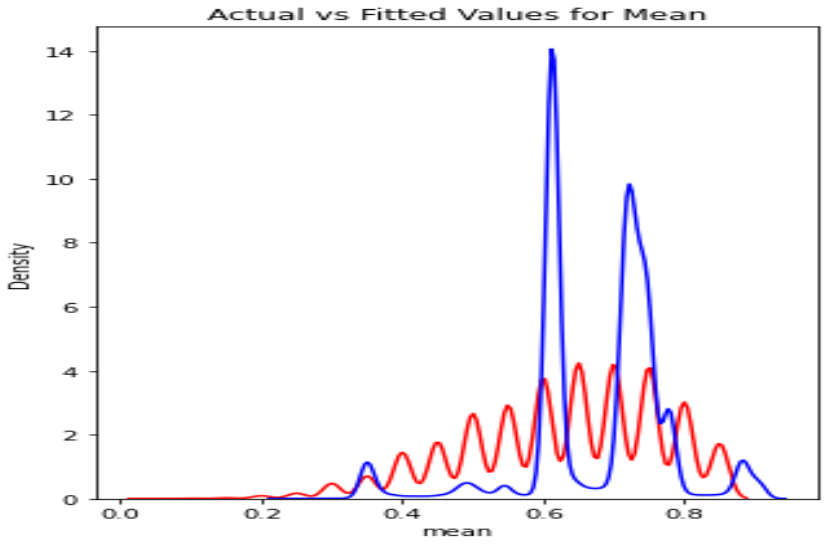


Fig 5.7 Accuracy plot of ANN after regularization

In the above fig 5.7 we see a plot of the actual versus predicted values where the actual values are in red and predicted values in blue.

The model gives an accuracy of 86.62% with the following errors:

Mean Absolute Error: 0.09574125003196753

Mean Squared Error: 0.015970276700061698

Root Mean Squared Error: 0.1263735601305182

We can already see a significant improvement in the accuracy of the ANN after it i s trained with a larger training sample.

5.3.3 Less Number of Neurons

We try training the Neural network with lesser number of neurons = 4 because if there are fewer parameters for the network to learn, it cannot memorize all the data points and must generalize. but the accuracy decreases to 78.63% so we will not use this method.

So we can see after testing the Decision Tree Regressor Model, Random Forest Regressor Model and Artificial Neural Network that the Random Forest Performs the best on the entire dataset of 157000 companies to predict the mean values.

Based on the prediction given by the Random forest model with accuracy of 93.17% we will calculate the bankruptcy of the 157000 companies as the this model has the highest accuracy.

5.4 Calculation of Bankruptcy for 157000 companies

We calculate bankruptcy by labelling the data as 1 if the mean is greater than or equal to 0.5 and 0 if less than 0.5

Row Labels	Count of Predicted Bad Governance
1	127354
0	29697
Grand Total	**157051**

Table 5.1, Predicted Bad Governance

Row Labels	Count of Actual Bad Governance
1	127782
0	29269
Grand Total	**157051**

Table 5.2, Actual Bad Governance

Here 1 signifies the bankrupt companies and 0 signifies the non-bankrupt companies. We can see that 127354 companies have a high risk of bad governance leading to the probability of bankruptcy while 29697 have no or a low risk as predicted by the Random Forest model using non-financial features. When comparing with the actual bankruptcy data, the result is quite accurate.

Row Labels	Count of Low Risk
FALSE	156435
LR	616
Grand Total	**157051**

Table 5.3, Predicted Bad Governance (LR)

Row Labels	Percent of Low Risk
FALSE	99.61%
LR	0.39%
Grand Total	**100.00%**

Table 5.4, Actual Bad Governance (LR)

If the mean values fall under 0.3, it can be categorized as low risk of bad governance leading to the probability of bankruptcy. We can see from the above tables that only 616 have a low risk of bankruptcy.

Row Labels	Count of medium risk
FALSE	127970
MR	29081
Grand Total	**157051**

Table 5.5, Predicted Bad Governance (MR)

Row Labels	Count of medium risk
FALSE	81.48%
MR	18.52%
Grand Total	**100.00%**

Table 5.6, Actual Bad Governance (MR)

If the mean values fall between 0.3 and 0.5, it can be categorized as medium risk of bad governance leading to the probability of bankruptcy. We can see from the above tables that only 29081 have a medium risk of bankruptcy.

Chapter 6

Model analysis using SHAP

In this section we will use a tool called shap to do a descriptive model analysis of the Random Forest regressor Model which is the final model used to predict bankruptcy.

6.1 What is Explainable Artificial Intelligence (XAI)?

People are asking this question increasingly frequently as more people become aware of the potential impacts of artificial intelligence. Simply defined, explainable AI is AI that is easy for people to understand. AI provides a justification for its decisions and actions. It offers the chance for explanation of the decision-making processes used by robots. This helps people trust and comprehend what is happening rather than feeling as though their information is being misused or utilized without their consent. This is important because many people are concerned about how much artificial intelligence is being used in daily life, especially in healthcare. We need to be able to trust these systems if we're going to rely on them.

6.1.1 How does Explainable AI work?

A prediction made by an AI system is considered to be explainable if the assumptions underlying it can be explained. A subset of the more generic term "interpretability" for artificial intelligence is "explainable AI." Interpretability enables us to understand what a model is learning, the other information it has to offer, and the motivations behind its decisions in the context of the real-world problem we're trying to solve. When model metrics are insufficient, interpretability

is required. Model interpretability enables us to predict how a model will perform in various test scenarios by comparing it to its training environment.

Increased transparency, dependability, equity, and accountability in AI systems are benefits of explainable AI. Explainable AI systems may be helpful when attempting to understand the reasoning behind a certain prediction or decision produced by machine learning algorithms. The graphic below shows the workflow where explainability fits in.

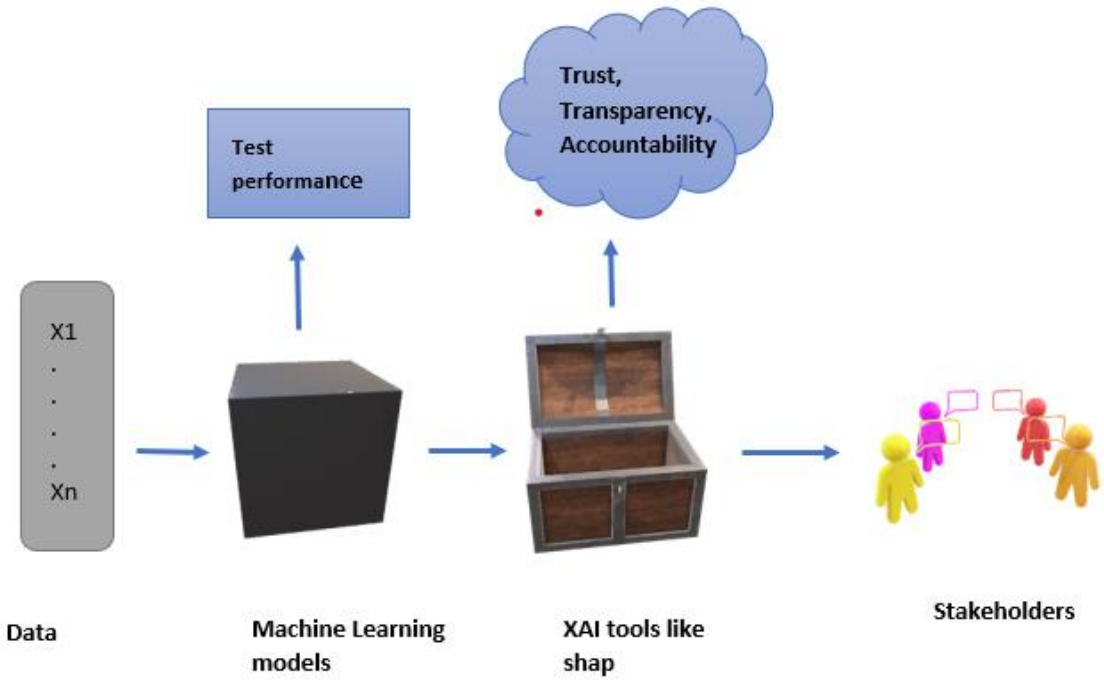


Fig: 6.1 Explainable Artificial intelligence analysing ML models [self-made image]

This kind of AI has become more significant as more stakeholders started to question the predictions made by it. They want to know how the forecasts were

made before relying on them and taking action. The image below demonstrates the need for understandable AI.

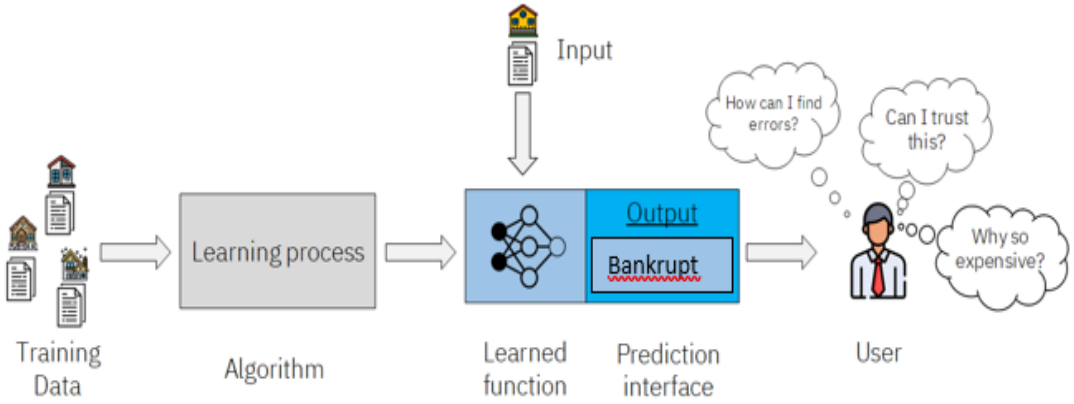


Fig 6.2, Explainable AI incorporated to the workflow, Source [63]

Explainable AI systems may be helpful in situations involving accountability, such as with autonomous vehicles, because if something goes wrong, a person is still accountable for their actions. To train explainable AI models, explainability approaches are utilized, which use textual descriptions that can be comprehended by humans to explain the reasoning behind a model's prediction. Explainability approaches are currently used in numerous artificial intelligence domains, including natural language processing (NLP), computer vision, medical imaging, health informatics, and many more.

The key contrast between explainable AI and AI is that explainable AI is a subset of artificial intelligence that justifies its decisions. Explainable artificial intelligence systems can simulate these processes by using explainability approaches because people have a significant impact on inference and conclusion-making. (P. Linardatos, 2020 [64])

A machine learning model that was trained using any of the methods, such as decision trees, random forests, and linear regression, can be helped by explainable AI by helping to provide the explanation for a decision made by the model. This functionality is offered by a number of tools, and it is one of the explainability tactics that is most frequently used in practice. For instance, you can utilize local interpretations to defend the decisions produced by a machine learning model using the explainability tools SHapley Additive exPlanations (SHAP) and LIME. Here's a brief rundown of how these tools work:

- LIME stands for Local Interpretable Model-Agnostic Explanations. The open-source explanation LIME was developed by researchers at Carnegie Mellon University. It may be used to explain the predictions of any machine learning model and has been used to explain everything from credit scoring systems to self-driving cars. LIME works by modifying the input data and seeing how the output of the model changes as a result. In the vicinity of an example x, LIME tries to fit a locally comprehensible model that is in agreement with the outcomes of the original model, f(x), given the example x. (Alex Gramegna et al, 2021 [65])
- The SHAP (SHapley Additive exPlanations) framework, which is based on game theory, is built on the Shapley values from cooperative game theory. It combines optimal credit allocation with regional explanations using the original Shapley values from game theory and their derivatives. [66]

  In our study we use Shap as the Explainable AI tool so we will dive deeper into the theory behind the framework and then see how it can explain the ML model used in our study.

6.2 Shapley values and SHAP – SHapley Additive explanation

Shapley values originated from the concept of game theory in 1953 [68] for solving the following question:

People with a variety of skills are working together for a shared goal. How should the group divide the reward fairly?

6.2.1 Calculation of Shapley values

We take into account a game with D players in which each feature $j \in \{1, \ldots, D\}$ is a player, and we wish to value their input. Each of the two possible coalitions, S, has a characteristic function, and there are 2^D possible coalitions.

$$v : 2^D \rightarrow R \qquad \textit{Eqn 6.1}$$

The equation for calculating the Shapley value of N players is given as follows:

$$\emptyset_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} (v(S \cup \{i\}) - v(S)) \qquad \textit{Eqn 6.2}$$

A set N of n players are present in a coalitional game. Additionally, we have a function v that provides the value (or payout) for any subset of those players. For example, if S is a subset of N, v(S) returns the value of that subset. Therefore, we can use the equation to get the payment for player I or the Shapley value, in a coalitional game (N, v).

According to the theory, if player N performs much better than the other, then $v(S \cup N)$ is invariably higher than v(S), and as a result, $\emptyset_N(v) \gg 0$.

Let the team N = $\{A, B, C, D\}$ players in a game and

$$i = D \qquad \textit{Eqn 6.3}$$

Then the shap formula can be rewritten as,

$$\emptyset_i(v) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \binom{|N|-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S)) \qquad \textit{Eqn 6.4}$$

If we exclude D we can make subsets from $\{A, B, C\}$

$$\begin{matrix} & A & AB & \\ \emptyset & B & BC & ABC \\ & C & CA & \end{matrix} \qquad \textit{Eqn 6.5}$$

So, different subsets can be made where $\emptyset$ is the null set.

From equation 6.4, we can see for a given subset S it's value will be compared when the value of i is included. This gives the marginal value of player i in the subset

$$\begin{matrix} & \Delta v_A, D & \Delta v_{AB}, D & \\ \Delta v_\emptyset, D & \Delta v_B, D & \Delta v_{BC}, D & \Delta v_{ABC}, D \\ & \Delta v_C, D & \Delta v_{CA}, D & \end{matrix} \qquad \textit{Eqn 6.6}$$

These subsets represent how D performs in the overall game. So, we need to calculate 8 different marginal values.

$\binom{|N|-1}{|S|}^{-1}$ calculates how many permutations of each subset size is possible when excluding i.

When calculating shapley value of D we have |N|-1 = 3.

By applying scaling factor to each marginal value,

$$\begin{matrix} & \frac{1}{3}\Delta v_A, D & \frac{1}{3}\Delta v_{AB}, D & \\ 1\Delta v_\emptyset, D & \frac{1}{3}\Delta v_B, D & \frac{1}{3}\Delta v_{BC}, D & 1\Delta v_{ABC}, D \\ & \frac{1}{3}\Delta v_C, D & \frac{1}{3}\Delta v_{CA}, D & \end{matrix} \qquad \textit{Eqn 6.7}$$

We are averaging out the impact of the other team members for each subgroup size by applying this scaling factor. This indicates that regardless of the make-up of these teams, we can capture the average marginal contribution of D when added to teams of sizes 0, 1, 2, and 3.

Final scaling is done by dividing the marginal values by 4 as that is the number of players in our game. This averages the effects of other players for each subset size.

Therefore, shap value of D,

$$\emptyset_D(v) = \frac{1}{4}\sum \begin{pmatrix} & \frac{1}{3}\Delta v_A, D & \frac{1}{3}\Delta v_{AB}, D & \\ 1\Delta v_\emptyset, D & \frac{1}{3}\Delta v_B, D & \frac{1}{3}\Delta v_{BC}, D & 1\Delta v_{ABC}, D \\ & \frac{1}{3}\Delta v_C, D & \frac{1}{3}\Delta v_{CA}, D & \end{pmatrix} \quad \textit{Eqn 6.8}$$

After we do this for each player, we will know the contribution of each player in winning or losing the game.

6.2.2 Tree Shap

In our study we are going to use an algorithm called Tree Shap for calculating the shap values.

Thanks to Tree SHAP, we can now explain the model's behavior, namely how each feature affects the model's output. In this instance, each result or forecast is seen as the accumulation of the contributions from each unique attribute.

It provides a variety of tools, especially through complex graphs, for thoroughly analyzing the model predictions.

- Global interpretability is achieved with summary graphs. These demonstrate the general behaviour of the model's features and assist us in

identifying the traits that have the biggest and most noticeable impacts on the output.

- Local data analysis can be done using force/dependence graphs. These show how the features behave precisely in a single model prediction, allowing us to understand each feature's unique influences on the outcome.
- It can be used with several machine learning models, such as: The explainers SHAP offers cover the bulk of machine learning techniques. These classes are a collection of explanations for several machine learning techniques, each of which is grouped under a separate heading. The element that describes how the model behaves is known as the explainer. [66]

6.3 Interpretation of our Random Forest Regressor Model for Bankruptcy prediction

We apply Tree shap on the Random Forest model which predicts the bankruptcy of 157000 companies and see the following interpretations:

6.3.1 Variable importance plot for global interpretability

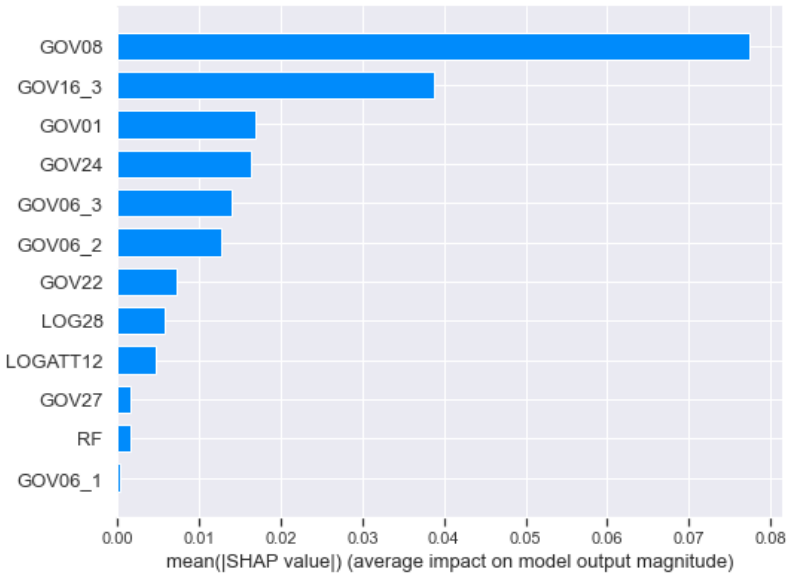


Fig: 6.3 Variable importance plot

- From the figure above we can see the most important features on the y-axis in descending order
- The shap values are represented on the x-axis
- GOV08 has the highest feature importance
- If we compare this fig with fig in our initial feature analysis on the data, where we see the indepenpent variables who have the highest corelation with the dependent variable matches with this global interpretability plot, that is GOV08 has the highest feature importance followed by GOV16_3.

6.3.2 Summary Plot:

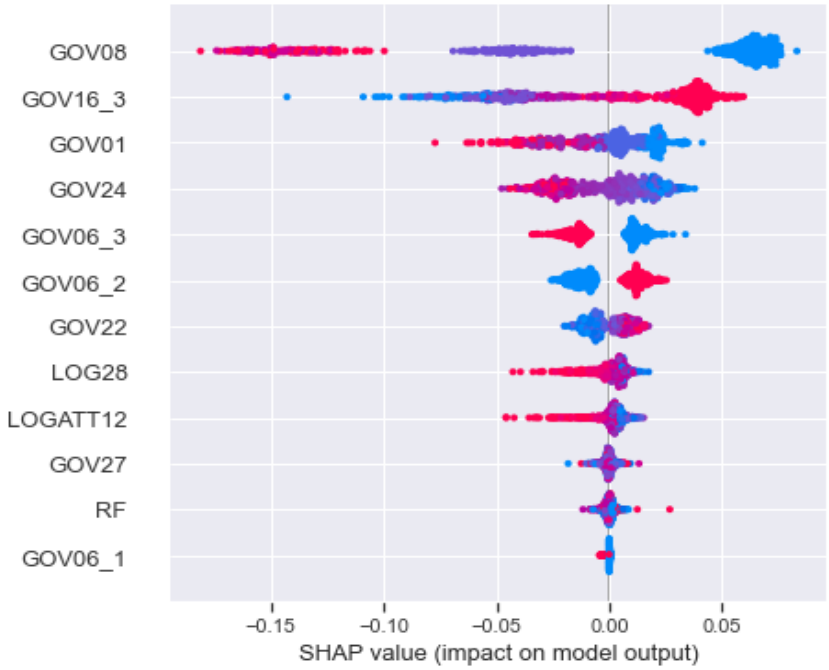


Fig 6.4, Summary plot

- From the figure above we can see the most important features on the y-axis in descending order
- The shap values are represented on the x-axis

- the feature value with **colors**. A high value is represented with red, while a low value with blue.
- Each point represents a prediction value.
- The lower the GOV08 value, the higher the relevance
- The higher the GOV16_3 value, the higher the relevance

6.3.3 Force Plot

Force plot is used to analyse a single model prediction. We see prediction for 2 companies

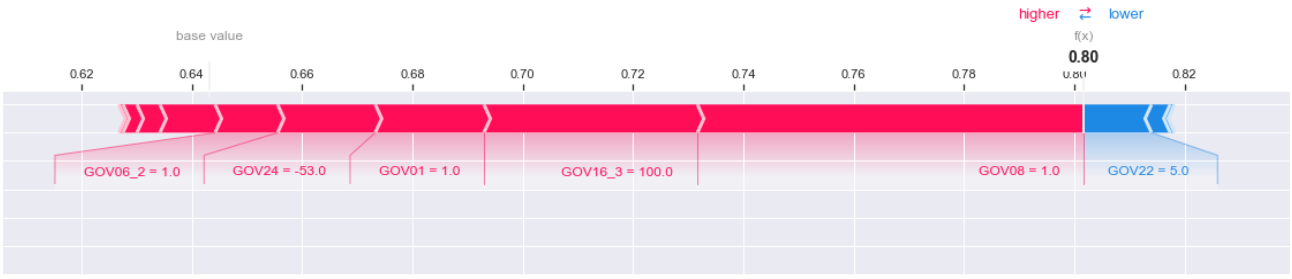


Fig 6.5, model output value = 0.8

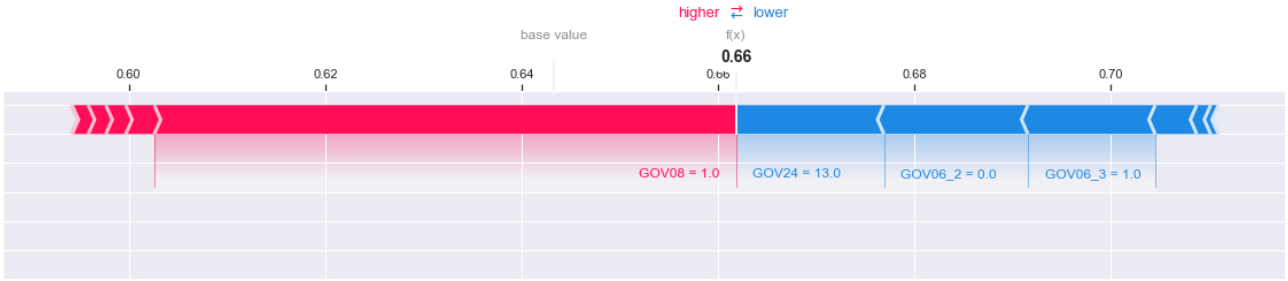


Fig 6.6, model output value = 0.66

In the above plots we can see the following observations:

- The model output values for each company
- The base value is the value that would be anticipated if no features for the current output were known.
- How each feature impacts the output

There are red and blue arrows that point to each feature.

Every arrow here denotes:

- The effect of the feature on the model is indicated by a larger arrow.
- how the feature affects the model: A rise in the model output value is indicated by a red arrow, whereas a decrease is indicated by a blue arrow.

We can also see a global representation of the above force plots:

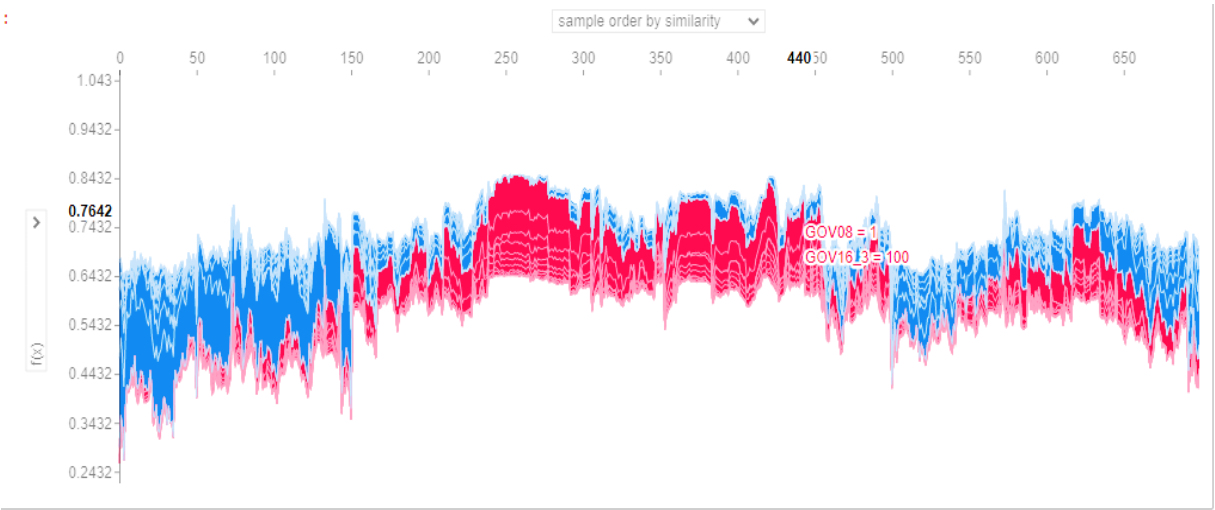


Fig: 6.7, Global representation of force plots

6.3.4 Partial Dependence plot

The partial dependence plot is an additional plot beneficial for the local interpretability.

The plots compare a chosen feature with another and illustrates whether there is an interaction between the two features.

Combining game theory and machine learning models requires matching the input features of a model with the players in a game and the function of the model with the game's rules. Since a player can choose whether to play or not in game theory, we need a way for a feature to "join" or "not join" a model. According to the most often used definition, a feature "joins" a model when its value is known, and it does not join the model when its value is uncertain. We integrate out the other features using a conditional expected value formulation to evaluate an existing model f when just a subset S of features are included in the model. [67]

$$E|f(X)|\,X_s = x_s \qquad \textit{Eqn 6.8}$$

Because we witness the characteristics in S, we are aware of their values in the equation.

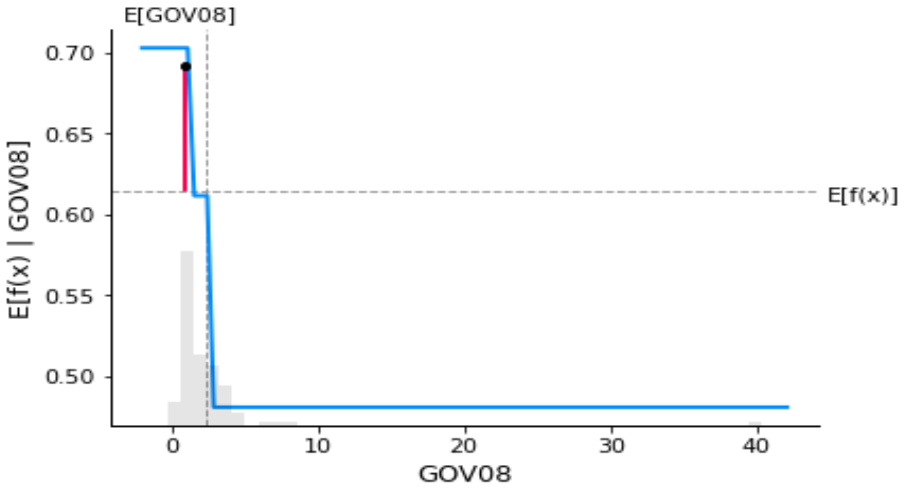


Fig: 6.8, Partial Dependence plot

Because the standard partial dependence plot and SHAP values have such a strong relationship, we can precisely trace out a mean-centered version of the partial dependence plot for a given feature by plotting the SHAP value for that feature throughout the whole dataset:

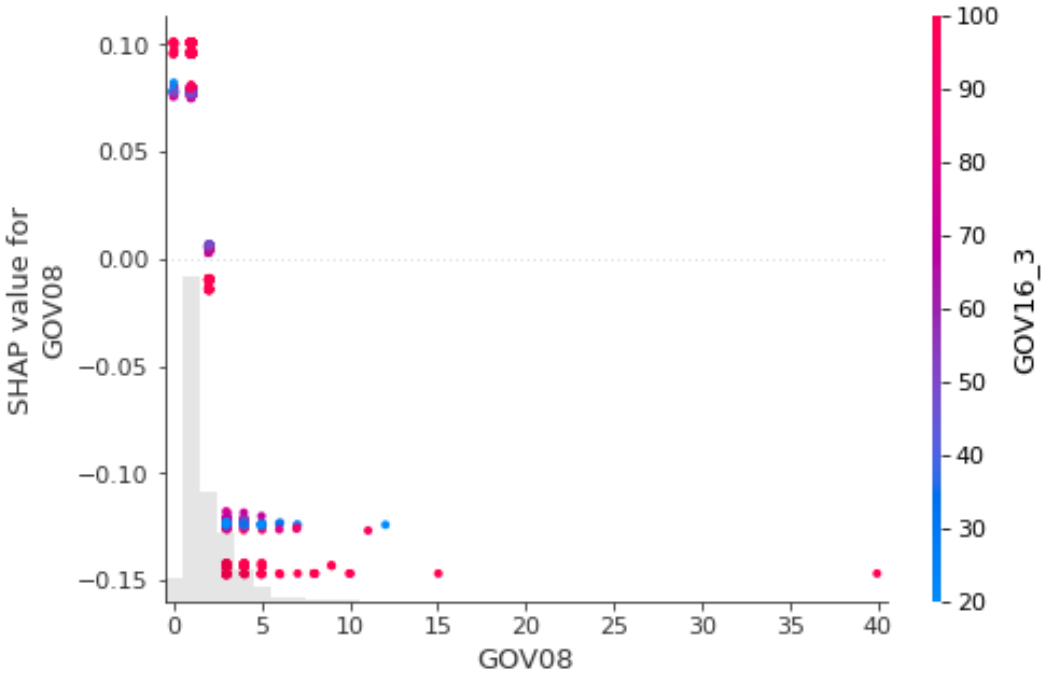


Fig 6.9, Interaction between two highly important features, GOV08 and GOV16_3

6.3.5 Heatmap

The heatmap shows the global relevance of all the features with respect to the shap values.

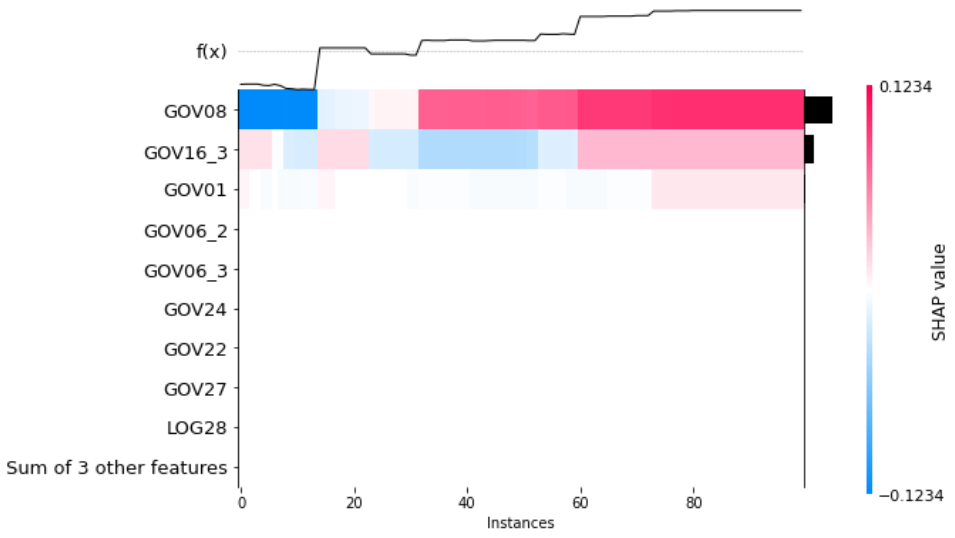


Fig 6.10 heatmap of shap values

Chapter 7

Results

In this chapter first we will see the final results related to bankruptcy prediction for a dataset containing 8641 companies and without the target variable mean. The machine learning model is trained on the dataset containing 157000 Italian companies based on which we find the model with the maximum accuracy which is then used to predict the mean variable of this final dataset.

In the next part we will see the shap analysis of the model used for the prediction to full proof the decisions made by our machine learning model and find the pattern of the most important features that have contributed to the prediction of bankruptcy in each company.

7. 1 Answering the first research question:

Using Corporate governance indices, do traditional machine learning models or Deep Learning models predict the bankruptcy of companies with maximum accuracy?

Based on my experiments with both traditional machine learning models and Deep learning model on the dataset containing the governance indices of 157000 companies we have concluded that, traditional machine learning models have a higher accuracy of prediction on unseen or new data compared to Deep Learning methods like Neural Networks due to the following reasons:

- The traditional machine Learning models like Decision tree regressor model and Random Forest Regressor model have higher accuracy than Artificial Neural Networks as it has less computational complexity.
- The Artificial Neural Network model has higher efficiency while training the model but when testing the model on unseen data, its efficiency reduces significantly due to overfitting. We have tried to regularize the model by increasing the training sample size but still its accuracy is less that both the traditional machine learning models.
- Hence even with a lesser number of training samples and less computational time and complexity, the traditional machine learning models have a higher accuracy than the Deep learning model.

Machine learning models	Decision Tree Regressor		Random Forest Regressor		Artificial Neural Network	
Size of training samples	8400	10500	8400	10500	10500	35000
Accuracy on seen data	91.1%	90.7%	92.84%	92.87%	91.4%	91.6%
Accuracy on unseen data	For 10, 500 91.5%		For 10,500 93.17%		For 35,000 86.62%	

Table 7.1, Comparison of the ML models

From the table 7.1 above we can also see that the Random Forest Regressor model has the highest accuracy on unseen data so we will use it the predict the bankruptcies on our final dataset.

7.1.1 Bankruptcy prediction on new data

Here we use the Random Forest model to predict the mean values of the companies based on which we will predict the bankruptcy of the companies.Since we have encoded the GOV06 feature into 3 different features, we will try to see its corelation with other features for this dataset of 8641 companies.

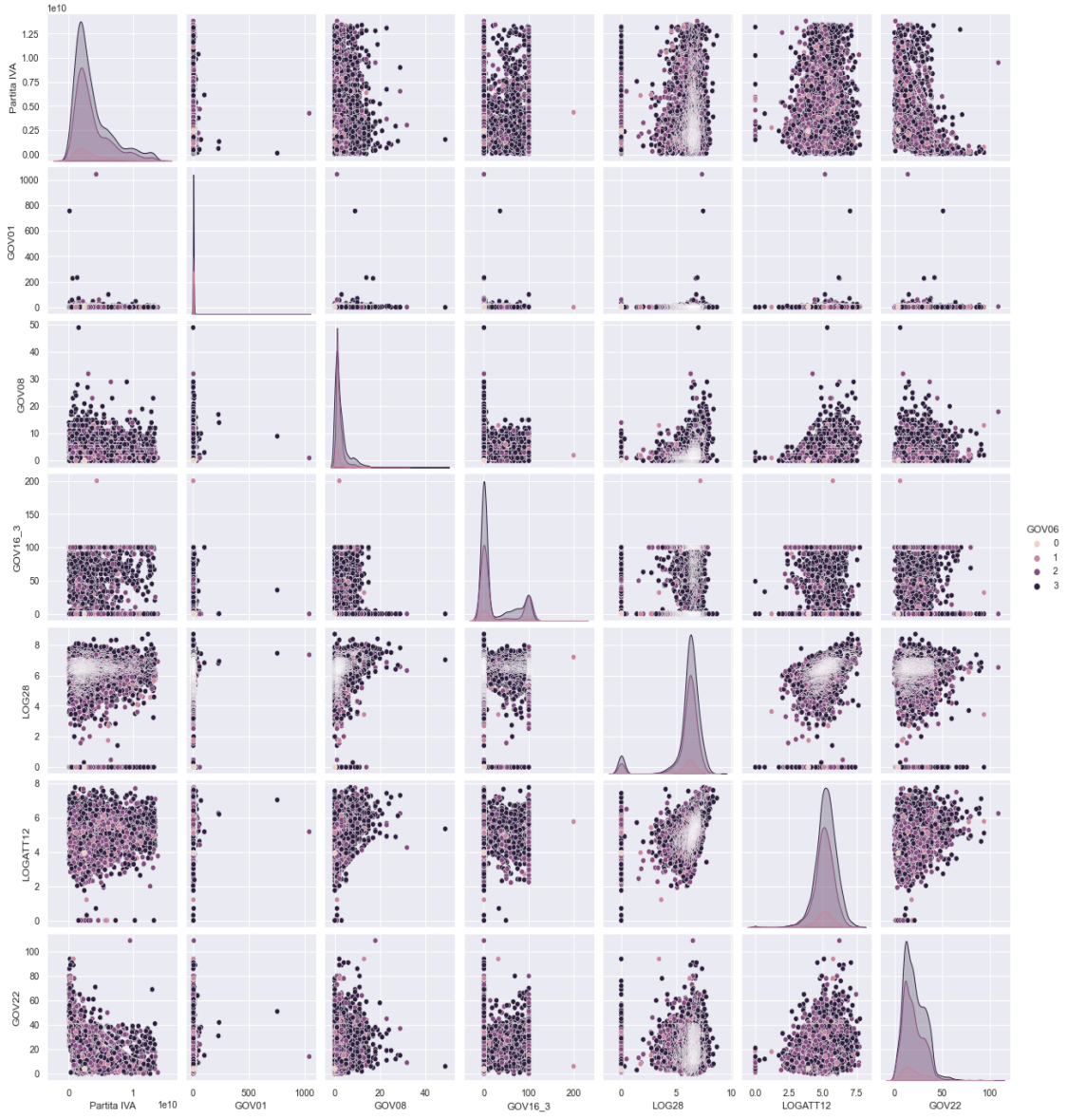


Fig 7.1, Pair plot with GOV06

- We can see from the plot that GOV06 has very less corelation with GOV01
- It has the highest corelation with GOV22

After predicting the mean values, we get the following results:

Row Labels	Count of Bad Governance
1	5699
0	2942
Grand Total	**8641**

Table 7.2, Count high risk '1'

Row Labels	Count of Bad Governance
1	65.95%
0	34.05%
Grand Total	**100.00%**

Table 7.3, Percentage high risk '1'

From the table we can see that 5699 companies represented as 1 has a high risk of bankruptcy and 2942 companies represented as zero have a lesser risk of bankruptcy.

If the predicted mean value is greater than or equal to 0.5 the company has a high risk of bankruptcy

If the predicted mean value is greater than 0.3 or less than 0.5 the company has a medium risk of bankruptcy

If the predicted mean value less than or equal to 0.3 the company has a low risk of bankruptcy.

Row Labels	Count of Low Risk
FALSE	8453
LR	188
Grand Total	**8641**

Table 7.4, Count Low risk

Row Labels	Count of Low Risk
FALSE	97.82%
LR	2.18%
Grand Total	**100.00%**

Table 7.5, Percentage Low risk

Row Labels	Count of Medium Risk
FALSE	5887
MR	2754
Grand Total	**8641**

Table 7.6, Count Medium risk

Row Labels	Count of Medium Risk
FALSE	68.13%
MR	31.87%
Grand Total	**100.00%**

Table 7.7, Percentage Medium risk

7.2 Model Analysis using Shap

Answering the second research question:

Which are the most important governance indices contributing to the prediction of bankruptcy of companies? Is there a common pattern that is followed among all the companies in terms of the features contributing to the maximum accuracy?

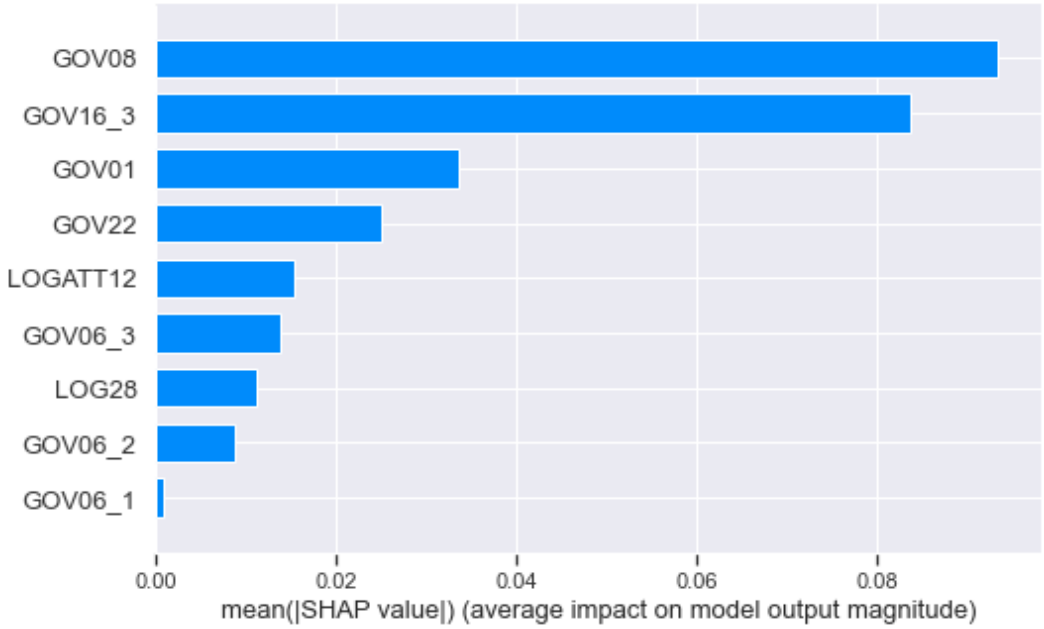


Fig 7.2, Variable importance plot for global interpretability

From the fig we can clearly see that:

- GOV08 has the highest feature importance
- If we compare this fig with fig in our initial feature analysis on the data, where we see the indepenpent variables who have the highest corelation with the dependent variable matches with this global interpretability plot, that is GOV08 has the highest feature importance followed by GOV16_3

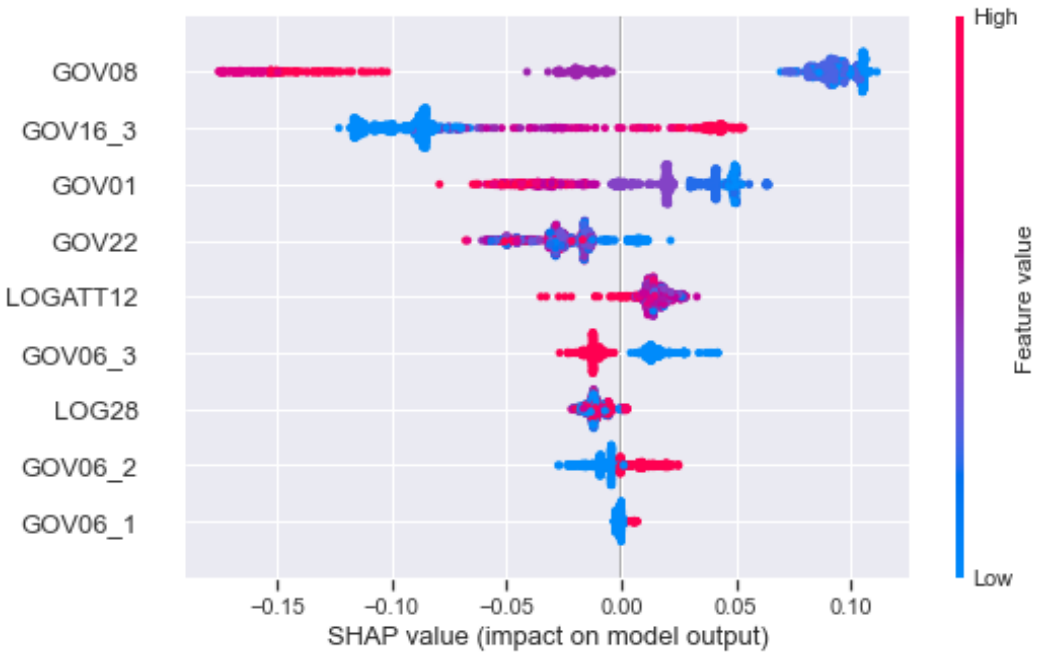


Fig 7.3, Summary plot

From the above summary plot we can see that:

- The lower the GOV08 value, the higher the relevance
- The higher the GOV16_3 value, the higher the relevance

7.2.1 Finding patterns of features in the data

After doing shap analysis on the final dataset with predicted bankruptcy we found the patterns of features that contributed most to the bankruptcy prediction in each company.

Row Labels	Count of Merge
GOV08,GOV16_3,GOV01,LOGATT12,LOG28,GOV06_3,GOV22,GOV06_2,GOV06_1	437
GOV16_3,GOV08,GOV01,LOGATT12,GOV06_3,LOG28,GOV22,GOV06_2,GOV06_1	345
GOV16_3,GOV08,LOGATT12,GOV01,LOG28,GOV06_3,GOV06_2,GOV22,GOV06_1	309
GOV08,GOV16_3,GOV01,LOG28,LOGATT12,GOV22,GOV06_2,GOV06_3,GOV06_1	306
GOV08,GOV16_3,LOGATT12,GOV01,LOG28,GOV22,GOV06_3,GOV06_2,GOV06_1	242
GOV16_3,GOV08,GOV01,LOGATT12,GOV06_3,LOG28,GOV22,GOV06_1,GOV06_2	221
GOV08,GOV16_3,LOGATT12,GOV01,LOG28,GOV06_3,GOV22,GOV06_2,GOV06_1	219
GOV08,GOV16_3,GOV01,LOG28,LOGATT12,GOV06_3,GOV22,GOV06_2,GOV06_1	213
GOV08,GOV16_3,LOGATT12,GOV01,GOV06_3,GOV22,LOG28,GOV06_2,GOV06_1	208
GOV08,GOV16_3,GOV01,LOG28,LOGATT12,GOV22,GOV06_3,GOV06_2,GOV06_1	186

Table 7.8, shows the comman patterns in shap values in descending order

Shows the most reapeated patterns in data of the most important features contributing to the prediction of bankruptcy in each company.

7.2.2 Individual feature analysis using shap to understand the reason behind the prediction of the Random Forest Model:

From the common patterns in data found by shap, we try to understand further which are the most important features that can lead to bankruptcy in each company

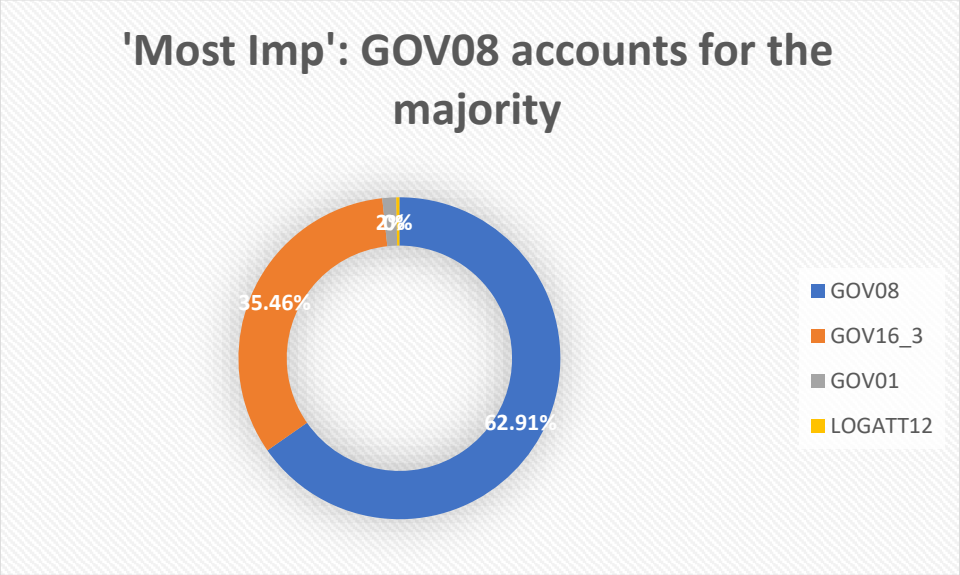


Fig 7.4, Showing the donut chart that GOV08 is the most important feature for 62.91% of the total companies

Row Labels	Count of Most Imp
GOV08	5436
GOV16_3	3064
GOV01	120
LOGATT12	21
Grand Total	**8641**

Table 7.9, Most Imp features

Row Labels	Count of Most Imp
GOV08	62.91%
GOV16_3	35.46%
GOV01	1.39%
LOGATT12	0.24%
Grand Total	**100.00%**

Table 7.10, Percentage

- We can see from the above fig and table that GOV08 is the most important features the prediction of bankruptcy in 5436 companies
- GOV16_3 is the most important feature for 3064 companies
- Followed by GOV01 and LOGATT12

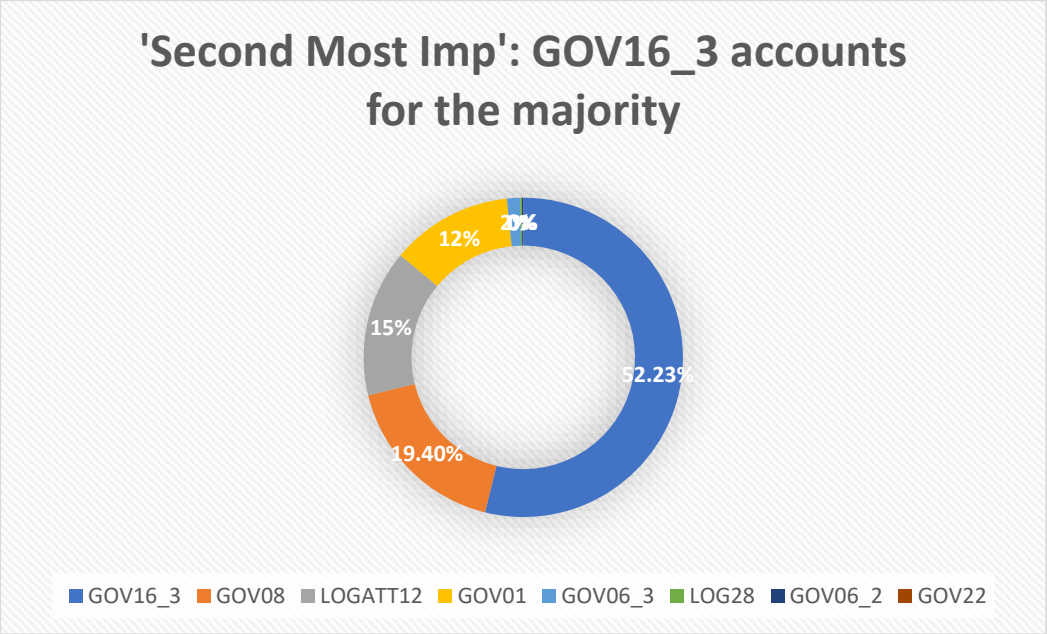


Fig 7.5, Showing 16_03 is the second most important feature in majority of the companies which is 52.23% of the total companies

Row Labels	Count of 2nd Imp
GOV16_3	4513
GOV08	1676
LOGATT12	1259
GOV01	1084
GOV06_3	90
LOG28	10
GOV06_2	7
GOV22	2
Grand Total	**8641**

Table 7.11, 2nd imp features

Row Labels	Count of 2nd Imp
GOV16_3	52.23%
GOV08	19.40%
LOGATT12	14.57%
GOV01	12.54%
GOV06_3	1.04%
LOG28	0.12%
GOV06_2	0.08%
GOV22	0.02%
Grand Total	**100.00%**

Table 7.12: percentage

- We can see from the above fig and table that GOV16_3 is the second most important features the prediction of bankruptcy in 4513 companies
- GOV08 is the second most important feature for 1676 companies

- Followed by LOGATT12, GOV01, GOV06_3 etc

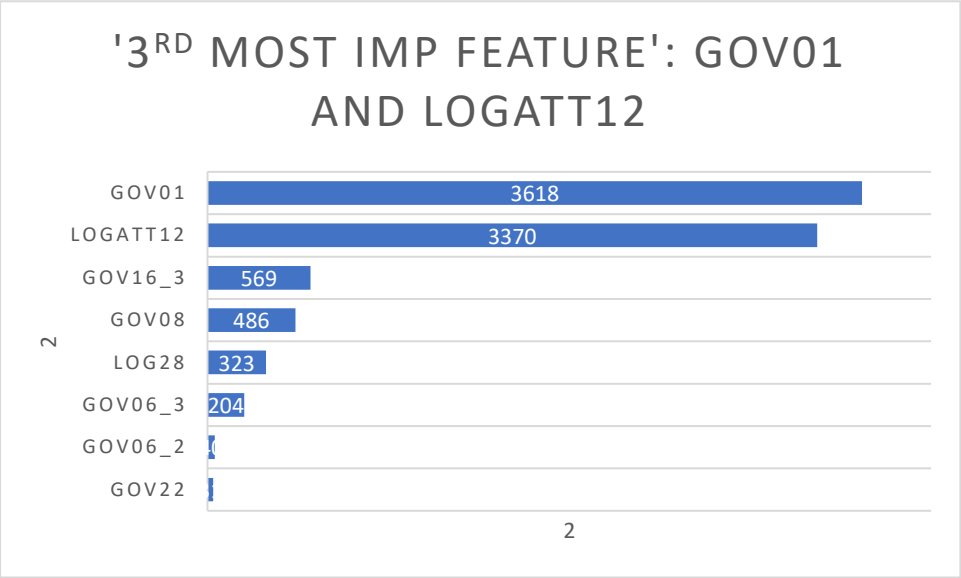


Fig 7.6, shows GOV01 and LOGATT12 are the third most important features in 41.87% and 39% of the companies respectively

Row Labels	Count of 3rd Imp
GOV01	3618
LOGATT12	3370
GOV16_3	569
GOV08	486
LOG28	323
GOV06_3	204
GOV06_2	40
GOV22	31
Grand Total	**8641**

Table 7.13, 3rd imp features

Row Labels	Count of 3rd Imp
GOV01	41.87%
LOGATT12	39.00%
GOV16_3	6.58%
GOV08	5.62%
LOG28	3.74%
GOV06_3	2.36%
GOV06_2	0.46%
GOV22	0.36%
Grand Total	**100.00%**

Table 7.14, Percentage

- We can see from the above fig and table that GOV1 is the third most important features the prediction of bankruptcy in 3618 companies
- LOGATT12 is the third most important feature for 3370 companies

- Followed by GOV16_3, GOV08, LOG28 etc

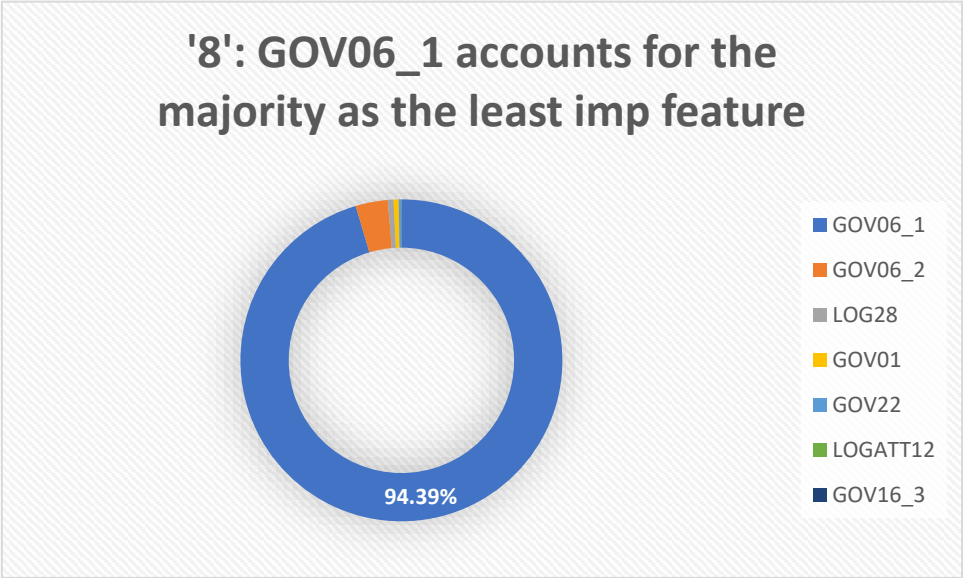


Fig 7.7, shows GOV06_1 is the least important feature in 94.39% of the companies

Row Labels	Count of least imp
GOV06_1	8156
GOV06_2	357
LOG28	52
GOV01	46
GOV22	22
LOGATT12	7
GOV16_3	1
Grand Total	**8641**

Table 7.15: Least Imp feature

Row Labels	Count of least imp
GOV06_1	94.39%
GOV06_2	4.13%
LOG28	0.60%
GOV01	0.53%
GOV22	0.25%
LOGATT12	0.08%
GOV16_3	0.01%
Grand Total	**100.00%**

Table 7.16, percentage

- We can see from the above fig and table that GOV06_1 is the least important feature the prediction of bankruptcy in 8156 companies

Chapter 8

Conclusion

In this thesis we have discussed about the significance of Bankruptcy and the significance of Corporate Governance indexes in the prediction of bankruptcy. As Bankruptcy prediction becomes increasingly significant for stakeholders to take early action and limit financial loses, more number of financial and non-financial data are being analysed especially for Small and Medium Sized Enterprises where the risks of Bankruptcy are higher than larger enterprises.

However, most of the work on bankruptcy prediction have been based on financial data or related to a time series analysis. So, this thesis is motivated by the need of bankruptcy prediction using non-financial data such as Corporate Governance indices as there has been very less work in this field due to the high complexity of the data. The Corporate governance indices represent the quality of a company's Governance structure, and they are calculated by a set of rules decided by experts. These governance indices help us to determine the non-financial health of a company and whether they can lead to bankruptcy in future.

We have used the data which consists of governance data collected by Arisk from public databases in the period of 2001-2018 for 157000 Italian SMEs and 8641 Italian SMEs that have applied for economical help from the financial institutions in the Piedmonte region. We compare traditional Machine Learning Models like Random Forest and Decision Tree with Deep Learning models like Artificial Neural Network and find out the model with maximum accuracy. The models are trained

on varying samples and tested on the entire 157000 companies concluding that traditional machine learning models are performing significantly better than Neural networks. This can be attributed to the overfitting of the neural network model and high complexity of the algorithm which does not work well of a very large dataset of unseen data. We can increase the accuracy of the neural network by regularizing the model and increasing the sample size however it further increases the time of computation and complexity of computation, thus, in general decreasing the efficiency of the model by increasing the dependency on seen data and defeating the purpose of prediction on new-unseen data.

So, we have determined that for our study, the Random Forest Regressor Model gives the maximum test accuracy of 93.17% and can almost precisely understand the nature and complex relationship of the features in the data and predict the target variable with significantly higher efficiency than the other models.

Thus, the Random Forest model is used to predict the nature of the governance structure which can lead to bankruptcy for the 8641 companies even after receiving economical help. We have determined that 65% of the companies have a bad governance structure with a very high risk of probability of bankruptcy in future, 32% of the companies have a bad governance structure with a medium risk of probability of bankruptcy in future and only 2 % have a low risk of probability of bankruptcy in future due to bad governance. These results provide the financial institutions with the opportunity of improving the governance structure of the company so that they might avoid the probability of bankruptcy in future.

So finally, using a tool called SHAP which is based on Explainable Artificial Intelligence to decipher the black box of the Random Forest Regression model which we have used to predict the bad governance structure and in turn the risk of bankruptcy in future for the companies, such that the stakeholders can trust the prediction made by the model chosen by us. Using SHAP we can clearly see that GOV08 and GOV16_3 are the two most important features that are being used to

predict the risk of bankruptcy which corresponds to our initial feature analysis of the most important features, thus proving that our Random Forest Regression Model is giving the accurate predictions.

We do further analysis of the model to find out any patterns of data among the most important features. GOV08 is the most important feature with maximum contribution for approx. 63% of the companies where as GOV16_3 is the second most important feature for approx. 52% of the companies, followed by GOV01 and LOGATT12 as the 3rd most important feature for approx. 42% and 39% of the companies respectively. We can also determine that GOV06_1 is the least important feature for approx. 95% of the companies.

We also can clearly see the 10 most repeated patterns of data and with this information each company can work with the combination of the most important feature to improve their governance structure thus reducing the chances of bankruptcy or determine the best variables of governance structure that can contribute to the success of the company.

Future applications of this work may include the addition of other risk sources such as cybersecurity, climate and seismic data which combined with financial factors can provide a dynamic evaluation of a company's operational risks.

The work can also be applied to post COVID-19 data with the addition of features relating with the pandemic such as health, temporary layoff of workers etc which can help to determine the probability of risk of bankruptcy in the future.

References and Bibliography

[1] Balleisen, E. J. (2001). *Navigating Failure: Bankruptcy and Commercial Society in Antebellum America* (New edition). The University of North Carolina Press.

[2] Zywicki, Todd J. (2007), Consumer Bankruptcy. George Mason Law & Economics Research Paper No. 07-09, *Encyclopedia of Law and Society: American and Global Perspectives, Forthcoming*, Available at SSRN: https://ssrn.com/abstract=960822

[3] Wiggins, R. Z., Piontek, T., & Metrick, A. (2014). The Lehman Brothers Bankruptcy A: Overview. *SSRN Electronic Journal*.

[4] Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, *4*, 71. https://doi.org/10.2307/2490171

[5] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

[6] Liang, Q., Xu, P. T., & Jiraporn, P. (2013). Board Characteristics and Chinese Bank Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2251496

[7] Yeh, Y. H., & Woidtke, T. (2005). Commitment or entrenchment?: Controlling shareholders and board composition. *Journal of Banking & Finance*, *29*(7), 1857–1885. https://doi.org/10.1016/j.jbankfin.2004.07.004

[8] Perboli, G., & Arabnezhad, E. (2021). A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications*, *174*, 114758. https://doi.org/10.1016/j.eswa.2021.114758

[9] G. Perboli, A. Tronzano, M. Rosano, L. Tarantino and F. Velardocchia, "Using machine learning to assess public policies: a real case study for supporting SMEs development in Italy," 2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR), 2021, pp. 1-6, doi: 10.1109/TEMSCON-EUR52034.2021.9488581.

[10] European Parliament, Directorate-General for Internal Policies of the Union, Azevedo, F., Haase, D., *Enhancing the competitiveness of SMEs : research for REGI Committee*, European Parliament, 2016, https://data.europa.eu/doi/10.2861/427679

[11] SME Envoy Network. 2019 Report of the SME Envoy Network to competitiveness council. European Commission. Available at https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=39603&no=6.

[12] Vladimirov, Z. (2017). The EU industrial policy and SME development in Central and Eastern Europe.

[13] N. Lee, H. Sameen and M. Cowling, “Access to finance for innovative

SMEs since the financial crisis", Research Policy, vol. 44(2), pp.370–380, 2015.

[13] Cantamessa, M., Gatteschi, V., Perboli, G., & Rosano, M. (2018). Startups' Roads to Failure. *Sustainability*, *10*(7), 2346. https://doi.org/10.3390/su10072346

[14] Altman, E.I., Cizel, J. and Rijken, H.A. (2014) Anatomy of Bank Distress: The Information Content of Accounting Fundamentals within and across Countries. http://ssrn.com/abstract=2504926

[15] Son, Hwijae & Hyun, C. & Phan, D. & Hwang, Hyung. (2019). Data Analytic Approach for Bankruptcy Prediction. Expert Systems with Applications. 138. 10.1016/j.eswa.2019.07.033.

[16] Bureau van Dijk, Aida - Italian companies database, 2020 https://aida.bvdinfo.com

[17] Barboza, Flavio & Kimura, Herbert & Altman, Edward. (2017). Machine Learning Models and Bankruptcy Prediction. Expert Systems with Applications. 83. 10.1016/j.eswa.2017.04.006.

[18] Nyambuu, U., & Bernard, L. (2015). A Quantitative Approach to Assessing Sovereign Default Risk in Resource-Rich Emerging Economies. *International Journal of Finance & Economics*, *20*(3), 220–241.

[19] Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research, 4*, 71-111

[20] Tamari, M. (1966). Financial Ratios as a Means of Forecasting Bankruptcy. *Management International Review, 6*(4), 15-21

[21] Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research, 18*(1), 109-131

[22] Hull, John & White, Alan. (2000). Valuing Credit Default Swaps I: No Counterparty Default Risk. Journal of Derivatives. 8. 10.3905/jod.2000.319115.

[23] Black, F. and Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 8, 637-654. http://dx.doi.org/10.1086/260062

[24] Merton, Robert C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." Journal of Finance 29, no. 2 (May 1974): 449-470. (Chapter 12 in Continuous-Time Finance.)

[25] Kealhofer, Lisa. (2002). Changing Perceptions of Risk: The Development of Agro-Ecosystems in Southeast Asia. American Anthropologist. 104. 178 - 194. 10.1525/aa.2002.104.1.178.

[26] Jensen, Michael C. & Meckling, William H., 1976. "Theory of the firm: Managerial behavior, agency costs and ownership structure," Journal of Financial Economics, Elsevier, vol. 3(4), pages 305-360, October.

[27] Fitzpatrick, F. (1932) A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firm. Certified Public Accountant, 6, 727-731.

[28] Merwin, C.L. (1942) Financing Small Corporations in Five Manufacturing Industries, 1926-1936: A Dissertation in Economics. Financing Small

Corporations in Five Manufacturing Industries, 1926-36. National Bureau of Economic Research.

[29] Walter A. Chudson, 1945. "The Current Ratio and Interrelationships of Working Capital Items," NBER Chapters, in: The Pattern of Corporate Financial Structure: A Cross-Section View of Manufacturing, Mining, Trade, and Construction, 1937, pages 67-80, National Bureau of Economic Research, Inc.

[30] Bellovary, J.L., Giacomino, D.E. and Akers, M.D. (2007) A Review of Bankruptcy Prediction Studies: 1930 to Present. Journal of Financial Education, 1, 3-41.

[31] Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. Journal of Banking and Finance, 1, 29-54. https://doi.org/10.1016/0378-4266(77)90017-6

[32] Fich, Eliezer & Slezak, Steve. (2008). Can Corporate Governance Save Distressed Firms from Bankruptcy? An Empirical Analysis. Review of Quantitative Finance and Accounting. 30. 225-251. 10.1007/s11156-007-0048-5.

[33] Parker, S., Peters, G. F., & Turetsky, H. F. (2002). Corporate governance and corporate failure: A survival analysis. *Journal of Corporate Governance, 2*(2), 4-12

[34] Chan, C.-Y., Chou, D.-W., Lin, J.-R., & Liu, F.-Y. (2016). The role of corporate governance in forecasting bankruptcy: Pre- and post-SOX enactment. *North American Journal of Economics and Finance, 35*, 166-188.

[35] Chen, H.-H. (2008). The Timescale Effects of Corporate Governance Measure on Predicting Financial Distress. *Review of Pacific Basin Financial Markets and Policies , 11*(1), 35-46

[36] Cadbury Report . (1992). *Committee on The Financial Aspects of Corporate Governance: Report with Code of Best Practice* . London : Gee Publishing

[37] Daily, C. M., & Dalton, D. R. (1994). Bankruptcy and corporate governance: The impact of board composition and structure. *Academy and Management Journal, 37*(6), 1603-1607

[38] Simpson, W. G., & Gleason, A. E. (1999). Board structure, ownership, and financial distress in banking Firms. *International Review of Economics and Finance, 8*(3), 281-292

[39] Quinlan, J.R. Induction of decision trees. *Mach Learn* 1, 81–106 (1986). https://doi.org/10.1007/BF00116251

[40] Salzberg, S.L. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach Learn* 16, 235–240 (1994). https://doi.org/10.1007/BF00993309

[41] Bingchiang Jeng, Yung-Mo Jeng, and Ting-Peng Liang. 1997. FILM: a fuzzy inductive learning method for automated knowledge acquisition. Decis. Support Syst. 21, 2 (Oct. 1997), 61–73. https://doi.org/10.1016/S0167-9236(97)00019-5

[42] Linden, Ariel & Yarnold, Paul & Nallamothu, Brahmajee. (2016). Using machine learning to model dose-response relationships. Journal of Evaluation in Clinical Practice. 22. 860-867. 10.1111/jep.12573.

[43] Krizhevsky, A., Sutskever, I., Hinton, G.E., et al. (2012) ImageNet Classification with Deep Convolutional Neural Networks. International Conference on Neural Information Processing Systems, Lake Tahoe, 3-6 December 2012, 1097-1105.

[44] Zięba, Maciej & Tomczak, Sebastian & Tomczak, Jakub. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications. 58. 10.1016/j.eswa.2016.04.001.

[45] Brédart, Xavier. (2014). Financial Distress and Corporate Governance: The Impact of Board Configuration. International Business Research. 7. 10.5539/ibr.v7n3p72.

[46] Qiu, J., Wu, Q., Ding, G. *et al.* A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.* 2016, 67 (2016). https://doi.org/10.1186/s13634-016-0355-x

[47] Tsai, Chih-Fong and Jhen-Wei Wu. "Using neural network ensembles for bankruptcy prediction and credit scoring." *Expert Syst. Appl.* 34 (2008): 2639-2649.

[48] Victor M Becerra, Roberto KH Galvão, and Magda Abou-Seada. Neural and wavelet network models for financial distress classification. Data Mining and Knowledge Discovery, 11(1):35–55, 2005

[49] Jaymeen R Shah and Mirza B Murtaza. A neural network based clustering procedure for bankruptcy prediction. American Business Review, 18(2):80, 2000.

[50] Esteban Alfaro, Noelia García, Matías Gámez, and David Elizondo. Bankruptcy forecasting: An empirical comparison of adaboost and neural networks. Decision Support Systems, 45(1):110–122, 2008.

[51] Andreas Behr & Jurij Weinblat, 2017. "Default prediction using balance-sheet data: a comparison of models," Journal of Risk Finance, Emerald Group Publishing, vol. 18(5), pages 523-540, November.

[52] Nanni, Loris & Lumini, Alessandra. (2009). An experimental comparison of ensemble classifiers for bankruptcy prediction and credit scoring. Expert Systems with Applications. 36. 3028-3033. 10.1016/j.eswa.2008.01.018.

[53] Begley J, Ming J, Watts S. Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models[J]. Review of accounting Studies, 1996, 1(4): 267-284.

[54] Kim S Y, Upneja A. Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models[J]. Economic Modelling, 2014, 36: 354-362.

[55] Breiman L. Bagging predictors[J]. Machine learning, 1996, 24(2): 123-140

[56] Kruppa J, Schwarz A, Arminger G, et al. Consumer credit risk: Individual probability estimates using machine learning[J]. Expert Systems with Applications, 2013, 40(13): 5125-5131

[57] Zhao Z, Xu S, Kang B H, et al. Investigation and improvement of multi-layer perceptron neural networks for credit scoring[J]. Expert Systems with Applications, 2015, 42(7): 3508-3516.

[58] Di Ciaccio, Agostino & Cialone, Giovanni. (2019). Insolvency Prediction Analysis of Italian Small Firms by Deep Learning. International Journal of Data Mining & Knowledge Management Process. 9. 1-12. 10.5121/ijdkp.2019.9601.

[59] Moscatelli, Mirko & Parlapiano, Fabio & Narizzano, Simone & Viggiano, Gianluca. (2020). Corporate Default Forecasting with Machine Learning. Expert Systems with Applications. 161. 113567. 10.1016/j.eswa.2020.113567.

[60] Cascarino, Giuseppe and Moscatelli, Mirko and Parlapiano, Fabio, Explainable Artificial Intelligence: Interpreting Default Forecasting Models Based on Machine Learning (March 16, 2022). Bank of Italy Occasional Paper No. 674, Available SSRN: https://ssrn.com/abstract=4090707 or http://dx.doi.org/10.2139/ssrn.4090707

[61] Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome & Franklin, James. (2004). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Math. Intell.. 27. 83-85. 10.1007/BF02985802.

[62] Matel, Erik & Vahdatikhaki, Faridaddin & Hosseinyalamdary, Siavash & Evers, Thijs & Voordijk, Hans. (2019). An artificial neural network approach for cost estimation of engineering services. International Journal of Construction Management. 1-14. 10.1080/15623599.2019.1692400.

[63] Explainable AI incorporated to the workflow, https://towardsdatascience.com/explainable-artificial-intelligence-14944563cc79

[64] Linardatos, Pantelis, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2021. "Explainable AI: A Review of Machine Learning Interpretability Methods" *Entropy* 23, no. 1: 18. https://doi.org/10.3390/e23010018

[65] Gramegna, Alex & Giudici, Paolo. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. Frontiers in Artificial Intelligence. 4. 10.3389/frai.2021.752558.

[66] https://github.com/slundberg/shap

[67] https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

[68] Von Neumann, J. and Morgenstern, O. (1953) Theory of Games and Economic Behavior. 3rd Edition, Princeton University Press, Princeton