**FACULTY OF SCIENCE, ENGINEERING AND COMPUTING**

**School of Computer Science and Mathematics**

**MSc in Data Science**

**Author**

Darpan Sanjay Agarwal - K2279745

**Project Title**

Company Bankruptcy Predictions

**Supervisor**

Prof. Nebel, Jean-Christophe A

**Kingston University London**

**TABLE OF CONTENTS**

ABSTRACT 3

INTRODUCTION 4

OBJECTIVES 6

LITERATURE REVIEW 7

DATASET CONSIDERED 9

METHODOLOGY 9

SCHEDULE 10

RESOURCES AND BUDGET 11

ETHICS, AND DATA PROTECTION 11

REFERENCES 12

**ABSTRACT**

One of the most critical concerns in the finance field revolves around accurately predicting an organization's financial health in the future. In today's economic climate, most organizations place great importance on financial forecasting to ensure their survival and potentially avoid bankruptcy. The ability to predict bankruptcy is vital for decision-making, as it helps assess a company's solvency and its capacity to meet financial obligations.

Machine Learning techniques, particularly Deep Learning (DL) methods, have gained significant attention and have proven to be reliable indicators in financial applications, including bankruptcy prediction. In our research, we have chosen to adopt Deep Learning methods due to their effectiveness in classification tasks. Our study will compare three commonly used Deep Learning techniques those are Multilayer Perceptron model with six layers, Long-Short Term Memory (LSTM), and the Deep Belief Network

Additionally, to enhance the robustness of our predictions, we plan to leverage three ensemble classifier techniques which are K-Nearest Neighbor, Random Forest, and Support Vector Machine (SVM)

A significant challenge we face is dealing with a severely imbalanced dataset. To mitigate this issue and achieve more reliable results, we will apply appropriate balancing techniques during our analysis.

As the field of machine learning continuously evolves, we want to explore new methods, especially in Artificial Neural Network (ANN) approaches.

To address the challenges even more effectively, we are intrigued by hybrid algorithms, which have shown promise in improving accuracy. Among these is the relatively newer MOA-PSO method, which we intend to implement and compare against our established Deep Learning methods to assess their respective performances.

**INTRODUCTION**

Bankruptcy is a complex and often distressing financial situation that a company may face when it becomes unable to meet its financial obligations as detailed in *Altman & Hotchkiss (2010)*. It marks a critical point in a business's life, and its implications ripple through the economy, affecting not only the company's stakeholders but also its employees, creditors, and the wider market as described in *Bešlić, Jakšić, Bešlić Rupić & Andrić (2018)*. Understanding the factors that lead to company bankruptcy, its potential consequences, and the available solutions is vital to mitigate its impact and prevent its recurrence.

Causes of Company Bankruptcy:

Company bankruptcies do not occur overnight; they are typically the result of a combination of underlying issues that have been building up over time. Some common causes include:

1. Excessive Debt: Accumulating substantial debt without the ability to service it can lead to insurmountable financial strain, causing a company to default on its loan payments.

2. Declining Revenues: Shrinking sales, market saturation, or a failure to adapt to changing consumer demands can erode a company's profitability and jeopardize its financial stability.

3. Poor Financial Management: Inadequate financial planning, mismanagement of resources, and a lack of effective cost controls can cripple a company's ability to maintain sustainable operations.

4. Economic Downturns: External economic factors such as recessions or unfavourable market conditions can affect a company's revenue generation and viability.

5. Legal Issues: Lawsuits, regulatory non-compliance, or penalties can lead to significant financial liabilities that strain a company's financial health.

Implications of Company Bankruptcy:

The ramifications of company bankruptcy extend far beyond its financial state and can significantly impact various stakeholders:

1. Job Losses: Bankrupt companies often resort to downsizing or complete shutdown, resulting in job losses for employees, adding to the unemployment rate.

2. Creditor Losses: Unsecured creditors may face substantial losses as they may not be able to recover their investments or debts fully.

3. Market Confidence: A high-profile bankruptcy can erode market confidence, impacting investor sentiment and potentially leading to a wider economic downturn.

4. Industry Impact: Bankruptcy in a specific industry can trigger a domino effect, affecting other companies in the sector and causing a ripple effect in the market.

5. Legal Proceedings: Bankruptcy proceedings involve complex legal processes that may further drain the company's remaining resources.

Solutions to Company Bankruptcy:

While bankruptcy is a critical juncture, there are avenues available to address and manage the situation:

1. Restructuring: Companies may opt for financial restructuring, which involves renegotiating debts, altering payment terms, or selling off assets to become financially viable again.

2. Bankruptcy Filing: In some cases, companies may need to file for bankruptcy protection, such as Chapter 11 in the United States, to reorganize and protect themselves from creditors while they attempt to recover.

3. Acquisition or Merger: A stronger company may acquire the bankrupt company, providing the necessary funds and operational expertise to revive it.

4. Liquidation: In dire circumstances, liquidation may be the only option, where the company's assets are sold off to repay creditors.

Company bankruptcy is a complex and multifaceted financial crisis that can have wide-ranging consequences for various stakeholders. It emphasizes the importance of prudent financial management, adaptability, and contingency planning for businesses to safeguard against such situations described in *Zhang, Liu, Heidari, Wang, Chen & Wang (2021)*. By understanding the causes, implications, and available solutions for company bankruptcy, we can work towards building a more resilient and stable business environment.

As time passes, numerous companies have a strong interest in gathering financial information data. Since bankruptcy occurrences are infrequent but not unheard of, the data tends to be highly imbalanced.

Numerous studies have concentrated on addressing the issue of insufficient patterns within minority classes, specifically bankrupt companies in our context. This deficiency significantly impacts classifiers, leading to a decline in their reliability and performance as discussed by *Ganganwar (2012)*. The reason for this is that these methods often construct models biased towards predicting the majority class. To tackle this problem, various balancing techniques have been proposed to rectify data imbalances.

Furthermore, building upon research initiated by authors *Aljawazneh, Mora, García-Sánchez & Castillo-Valdivieso (2021)* regarding company bankruptcy prediction, we compare the effectiveness of different classification methods in predicting the financial status of Taiwanese companies. In this research, we utilize a classification approach called Metaheuristic Optimization-based Artificial Neural Network (MOAANN). This method is built upon the principles of the Particle Swarm Optimizer (PSO) which is widely discussed by *Khurma, Aljarah, Sharieh, Mirjalili & Evolopy-fs (2020).* Furthermore, author *Cheng & Jin (2014)* gives us an understanding of the technique known as Competitive Swarm Optimiser (CSO) which will help us to investigate cost sensitivity.

**OBJECTIVES**

* **OBJECTIVE**
  + The primary objective of our project is to develop an efficient predictive model capable of accurately identifying and forecasting the probability of a company or an organization going bankrupt.
  + Our aim is to offer valuable insights to stakeholders, including investors and regulatory bodies, empowering them to make well-informed conclusions and reduce financial risks associated with the possibility of bankruptcy.
* **TASK OVERVIEW**
  + Data Collection:
    - We will collect relevant financial data from various sources and meticulously cleanse and pre-process it to ensure its quality and suitability for analysis.
    - Given that our dataset exhibits an imbalance, we will employ undersampling and oversampling techniques to balance it and improve its stability.
  + Feature Selection:
    - Identifying crucial variables and features that strongly correlate with a company's health and bankruptcy risk will be a pivotal step in our analysis.
  + Model Development:
    - We will construct models using suitable machine learning or statistical techniques.
    - We have selected Deep Learning techniques of Deep Belief Network, a Multilayer Perceptron model with six layers, and Long-Short Term Memory (LSTM).
    - For the classifier ensemble, we will utilize K-Nearest Neighbor methods, Support Vector Machine (SVM), and Random Forest.
    - To further explore possibilities, we will also develop the MOA-PSO method within the framework of Artificial Neural Network methodology.
  + Evaluation:
    - To gauge the efficacy of our models, we will compare their results with previous related works employing similar techniques.
    - We will be using machine learning evaluation techniques to access the performance of our models which includes F1-score, precision, accuracy, and recall.

**LITERATURE REVIEW**

Predicting if a company or an organization will go bankrupt in the foreseeable future has been a topic of great interest in both academic research and practical application. The ability to anticipate the financial distress of a company is crucial for investors and regulators to make conclusive decisions. Over the years, numerous studies have been conducted to develop accurate and reliable bankruptcy prediction models. This literature review aims to present a comprehensive overview of the most significant research and approaches employed in predicting company bankruptcies.

Numerous researchers have explored the issue of imbalanced datasets in the literature using various machine-learning approaches. One such method developed to address binary class imbalance is the density weight method support vector machine (SVM) has been described by *Hazarika & Gupta (2021)*. Several artificial and real-world datasets were utilized to evaluate SVM's performance, and the results were compared with those of both SVM and fuzzy SVM, which were based on affinity and class probability.

Ensemble techniques, which combine multiple models to make predictions, have become popular in bankruptcy prediction. Bagging and boosting methods, such as Gradient Boosting Machines (GBM), have been utilized to improve model accuracy and reduce overfitting. Ensemble methods have demonstrated superior performance in handling imbalanced data, which is common in bankruptcy prediction where the number of non-bankrupt firms often outweighs the number of bankrupt ones.

To tackle the imbalance challenge, newer Deep Learning (DL) methods have also been considered. The authors *Aljawazneh, Mora, García-Sánchez & Castillo-Valdivieso (2021)* conducted a thorough and detailed comparison of three DL methods: Deep Belief Network (DBN), Long-Short term memory (LSTM), and a multilayer perception model (MLP). Another approach discussed by author *Gnip, P (2021)* is that of the selective oversampling method (SOA), which identifies outliers to differentiate the most common samples from the minority class and uses them for syntactic oversampling.

The authors *Hinton, Osindero, & Teh (2006)* give us a detailed idea of the concept of a Deep Belief Network (DBN). Authors *S. Hochreiter & Schmidhuber (1997)* also deeply explains the method of Long-Short term memory (LSTM) in their research paper. Multilayer perception model (MLP) techniques are also described to a great extent in the paper written by authors *A.A. Kasgari, Divsalar, Javid & Ebrahimian (2013).*

Given the high cost of running trained machine learning models due to GPU consumption, cost-sensitive techniques have gained popularity to address imbalanced datasets. Classifiers like logistic regression, decision trees, and extreme gradient boosting were explored by *Mienye & Sun (2021)*, which showed better results than standard algorithms. In another study, decision trees were used as a boosting method to improve business failure prediction, and a weight XGBoost technique was discussed by author *Zou, Y. (2022)* to overcome the class imbalance.

Recognizing the limitations of using only financial data, researchers have explored the incorporation of non-financial information to enhance bankruptcy prediction models. Social media sentiment analysis, customer reviews, and macroeconomic indicators have been used as supplementary data to provide a more holistic view of a company’s financial, health. These approaches have shown promising results in increasing and accuracy of bankruptcy prediction models.

With the advancements in computational power and data availability, researchers have turned to new machine-learning techniques. Artificial Neural Networks (ANNs) were among the early methods applied to bankruptcy prediction due to their ability to learn complex patterns from the data provided. Subsequent studies incorporated decision trees, support vector machines, and random forests, all of which demonstrated improved predictive performances compared to traditional methods. These methods have demonstrated better performance than gradient-based algorithms as shown in *Ansari*, Ahmad, Bakar & Yaakub *(2020)*. Additionally, the effects of MOA on imbalanced datasets were discussed by authors *Al-Badarneh, Habib, Aljarah & Faris (2020)*, where a PSO algorithm was used as an optimizer for predicting bankruptcy in a neural network architecture. Authors *Mahendru, Garg, Sharma & Srivastava (2021)* describe in detail the effects of neural network architecture on bankruptcy predictions.

We have examined a dataset comprising Taiwanese companies, encompassing 6819 entries collected over a decade from 1999 to 2009. Among these entries, 6599 companies (approximately 97%) are non-bankrupt, while 220 companies (roughly 3%) are labeled as bankrupt. Clearly, the dataset exhibits significant class imbalance. It includes 95 financial health indicators and a single-class label indicating the bankruptcy status of each company. Detailed information about our dataset can be found in *Liang, Lu, Tsai & Shih (2016)*.

Bankruptcy prediction remains a critical task in financial analysis and decision-making. This literature review highlights the different bankruptcy prediction models and their approaches. The incorporation of non-financial data and the utilization of ensemble techniques have further advanced the predictive capabilities of these models. However, despite this significant progress, challenges persist, including imbalanced data and the interpretability of complex models. Our research efforts are focused on addressing these challenges to create more robust and practical bankruptcy prediction models that can better serve owners, investors, creditors, and stakeholders in the financial industry.

**DATASET CONSIDERED**

We have examined a dataset comprising Taiwanese companies, encompassing 6819 entries collected over a decade from 1999 to 2009. Among these entries, 6599 companies (approximately 97%) are non-bankrupt, while 220 companies (roughly 3%) are labeled as bankrupt. Clearly, the dataset exhibits significant class imbalance. It includes 95 financial health indicators and a single-class label indicating the bankruptcy status of each company. Detailed information about our dataset can be found in *Liang, Lu, Tsai & Shih (2016)*.

**METHODOLOGY**

1. The literature review on 'Company Bankruptcy prediction' involved a thorough investigation of existing research papers related to our chosen topic. We analyzed the various methodologies and evaluation techniques used by different authors in their studies.

2. Throughout history, numerous methods have been employed to predict company bankruptcy. However, many of these methods have become outdated as new financial challenges arise, necessitating the adoption of newer approaches. For our predictive model, we have deliberately opted to utilize Deep Learning (DL) methods and explore the potential of an Artificial Neural Network (ANN) hybrid model.

3. Selecting an appropriate dataset is a critical aspect of building our model, as its success hinges on the quality and relevance of the data. We managed to identify three company datasets from three different countries which were Poland, Taiwan, and Spain. After careful consideration, we chose the Taiwan Companies Dataset due to its greater volume of relevant studies and ease of comparison with our work.

4. Having finalized our dataset and methods, we can now proceed with the actual construction and training of our models.

5. The training process is of utmost importance but also time-consuming. We face the challenge of incorporating various probabilities and utilizing different methods to ensure our model's accuracy.

6. To assess the performance of our models, we plan to employ two evaluation methods. We will be using machine learning evaluation techniques to access the performance of our models which includes F1-score, precision, accuracy, and recall. Secondly, we will compare our machine-learning results with the findings from previous research.

7. All the methods, models, and conclusions will be summarized comprehensively in our findings. We will develop a detailed report, covering each step of the project, to provide a clear understanding of our research.

**SCHEDULE (18TH JUNE 2023 – 31ST DECEMBER 2023):**

1. Project Definition and Proposal Submission:

a. 18th June 2023 – 19th July 2023

b. Duration: 31 days

2. Literature Review:

a. 20th July 2023 – 5th August 2023

b. Duration: 16 days

3. Model Building and Training:

a. 6th August 2023 – 20th September 2023

b. Duration: 45 days

4. Performance Enhancement and Evaluation:

a. 21st September 2023 – 31st October 2023

b. Duration: 40 days

5. Final Findings and Dissertation Report:

a. 1st November 2023 – 31st December 2023

b. Duration: 60 days

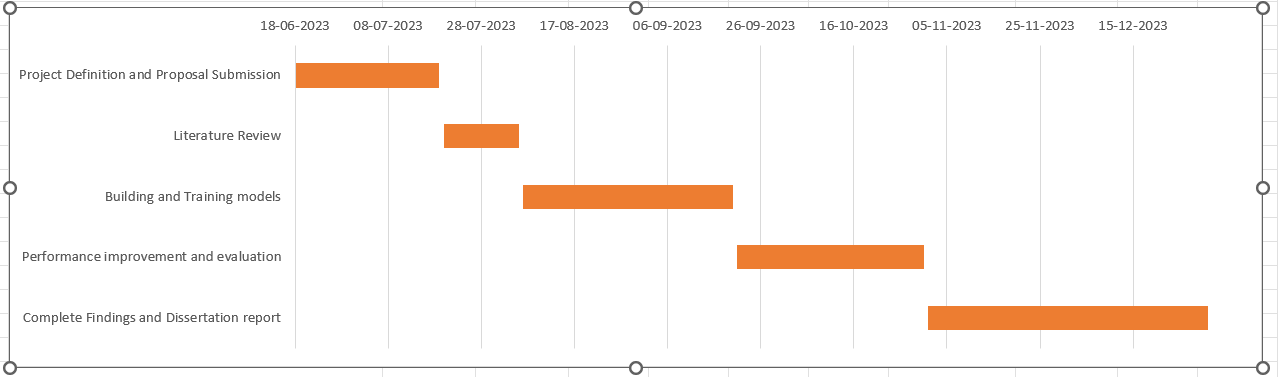


Fig. 1: Project Timeline

**RESOURCES AND BUDGET**

1. We require a comprehensive dataset containing the necessary financial indicators for our research, encompassing both bankrupt and non-bankrupt company data.

2. To facilitate the training process, we will utilize high-performance computing power, particularly a powerful GPU. This will significantly expedite the model training.

3. For the development of our models, we will use multiple different software and libraries which includes Python for development, google collab as our editor, and frameworks such as TensorFlow and Keras.

4. The Taiwan Company dataset which we will be referring to for our research is available to the public for free research and development.

5. To meet our higher computational powers requirements to develop and train our models, we will be accessing Kingston University's Library and computer facility.

**ETHICS, AND DATA PROTECTION**

• Our research relies on a publicly available Taiwan Company dataset [18], which has undergone extensive investigation by multiple authors using various machine learning methods.

• By using the Taiwan Company dataset which is available for free, we have made sure that we comply with the strict General Data Protection Regulation (GDPR) laws.

• The entirety of our models and research findings used from our academic work will be stored on a website called “BOX”.

• The Features of “BOX” will allow us to safeguard our data and findings as it will delete all the material stored in 6 months.

• We prioritize ethical considerations and are committed to utilizing the data responsibly, contributing positively to society through our research findings.

• Our intention is to use the data solely for research purposes, avoiding any misuse and instead, aiding future research endeavors.

**REFERENCES**

Altman, E.I.; Hotchkiss, E. (2010) Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt; John Wiley & Sons: Hoboken, NJ, USA; Volume 289.

Bešlić Obradović, D.; Jakšić, D.; Bešlić Rupić, I.; Andrić, M. (2018) Insolvency prediction model of the company: The case of the Republic of Serbia. Econ. Res.-Ekon. Istraž., 31, 139–157.

Zhang, Y.; Liu, R.; Heidari, A.A.; Wang, X.; Chen, Y.; Wang, M.; Chen, H. (2021) Towards augmented kernel extreme learning models for bankruptcy prediction: Algorithmic behavior and comprehensive analysis. Neurocomputing, 430, 185–212.

Ganganwar, V. (2012) An overview of classification algorithms for imbalanced datasets. Int. J. Emerg. Technol. Adv. Eng., 2, 42–47.

H. Aljawazneh, A. M. Mora, P. García-Sánchez and P. A. Castillo-Valdivieso (2021) "Comparing the Performance of Deep Learning Methods to Predict Companies’ Financial Failure," in IEEE Access, vol. 9, pp. 97010-97038, doi: 10.1109/ACCESS.2021.3093461.

Hazarika, B.B.; Gupta, D. (2021) Density-weighted support vector machines for binary class imbalance learning. Neural Comput. Appl., 33, 4243–4261.

Gnip, P.; Vokorokos, L.; Drotár, P. (2021) Selective oversampling approach for strongly imbalanced data. PeerJ Comput. Sci., 7, e604.

Mienye, I.D.; Sun, Y (2021) Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. Inform. Med. Unlocked, 25, 100690.

Zou, Y.; Gao, C.; Gao, H (2022) Business Failure Prediction Based on a Cost-Sensitive Extreme Gradient Boosting Machine. IEEE Access, 10, 42623–42639.

G. E. Hinton, S. Osindero and Y.-W. Teh (1997) "Long short-term memory", Neural Comput., vol. 9, no. 8, pp. 1735-1780.

A. A. Kasgari, M. Divsalar, M. R. Javid and S. J. Ebrahimian (2013) "Prediction of bankruptcy Iranian corporations through artificial neural network and probit-based analyses", Neural Comput. Appl., vol. 23, no. 3, pp. 927-936.

Al-Badarneh, I.; Habib, M.; Aljarah, I.; Faris, H. (2020) Neuro-evolutionary models for imbalanced classification problems. J. King Saud Univ. Comput. Inf. Sci., 34 (Pt A), 2787–2797.

Khurma, R.A.; Aljarah, I.; Sharieh, A.; Mirjalili, S. Evolopy-fs (2020) An open-source nature-inspired optimization framework in python for feature selection. In Evolutionary Machine Learning Techniques; Springer: Berlin/Heidelberg, Germany; pp. 131–173.

Cheng, R.; Jin, Y. (2014) A competitive swarm optimizer for large scale optimization. IEEE Trans. Cybern. **2014**, 45, 191–204.

Mahendru, K.; Garg, G.; Sharma, A.; Srivastava, R (2021) Evolutionary Methods for Bankruptcy Prediction: A Study on Indian Firms. In Soft Computing for Problem Solving; Springer: Berlin/Heidelberg, Germany; pp. 303–313.

Liang, D.; Lu, C.C.; Tsai, C.F.; Shih, G.A (2016) Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. Eur. J. Oper. Res., 252, 561–572.