Department of Electrical and Computer Engineering
North South University

Directed Research

Predicting Company Bankruptcy and Financial Status Analysis Using Machine Learning and Deep Learning

Maqsudul Mahmud Fahim **2011312642**

Zubair Mahmood Sowrab **2011657042**

Tanjim Imtial **2014215642**

Md. Zakaria Khan **2013412042**

Faculty Advisor:

Rifat Ahmed Hassan

Lecturer

ECE Department

Spring, 2024

LETTER OF TRANSMITTAL

June, 2024

To

Dr. Rajesh Palit
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: **Submission of Directed Research Report on "Predicting Company Bankruptcy and Financial Status Analysis Using Machine Learning and Deep Learning"**

Dear Sir,

With due respect, we would like to submit our **Directed Research Report** on **"**Predicting Company Bankruptcy and Financial Status Analysis Using Machine Learning and Deep Learning**"** as a part of our BSc program. The report deals with forecasting Bankruptcy using Machine Learning. The project was a very useful study, as it helped us to apply our skills and implement a system to assist in real world problems. We have ensured to meet the requirements demanded by the report to the maximum extent.

We will be highly obliged if you kinDLy receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

.....

Maqsudul Mahmud Fahim
ECE Department
North South University, Bangladesh

.....

Zubair Mahmood Sowrab
ECE Department
North South University, Bangladesh

.....

Tanjim Imtial
ECE Department
North South University, Bangladesh

.....

Md. Zakaria Khan
ECE Department
North South University, Bangladesh

APPROVAL

Maqsudul Mahmud Fahim (2011312642), Zubair Mahmood Sowrab (2011657042), Tanjim Imtial (2014215642) and Md. Zakaria Khan (2013412042) from Electrical and Computer Engineering Department of North South University, have worked on the Directed Research Project titled "Predicting Company Bankruptcy and Financial Status Analysis Using Machine Learning and Deep Learning" under the supervision of Rifat Ahmed Hassan Sir for partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor's Signature

…….

Rifat Ahmed Hassan

Lecturer

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

Chairman's Signature

…….

Dr. Rajesh Palit

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

DECLARATION

This is to declare that the project done is our original work and has not been submitted anywhere else for receiving a degree, a diploma or any other award of such sort. The information related to the project will not be disclosed and will remain confidential unless allowed by the course instructor. The previous works done in this field have been referenced and cited. The plagiarism policy, as stated by the project supervisor, has been maintained.

Students' names & Signatures

1. Maqsudul Mahmud Fahim

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

2. Zubair Mahmood Sowrab

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

3. Tanjim Imtial

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

4. Md. Zakaria Khan

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Rifat Ahmed Hassan, Lecturer, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. The authors would also like to thank their friends and loved ones for their countless sacrifices and continual support.

ABSTRACT

Predicting Company Bankruptcy and Financial Status Analysis Using Machine Learning and Deep Learning

Bankruptcy is one of the biggest financial issues that hampers the economy. It is a process in which a business cannot repay debts and want relief from their creditors. The following paper uses machine learning techniques to predict the chances that a company goes bankrupt based on various factors. For predicting the bankruptcy, we have used various machine learning techniques: SVM, XGBoost, Random Forest, Logistic Regression and Decision Tree. On performing our experiments with the models, we found Random Forest to be the best technique for predicting bankruptcy. The model produced a Training accuracy of 99.1%. We also applied different deep learning methods such as DNN, GRU, TCN to compare ML models performance. We found that each of these 3 DL models produce a training accuracy of approximately 98%. Bankruptcy has been on the rise over the past decade and it is very important to predict and take precautions for a business. The paper presents the optimal method to forecast bankruptcy and financial analysis. The project not only improves on traditional econometric models to do the same, but also presents the best approach to improve on existing works which use machine learning to predict bankruptcy and financial conditions. The research will help businesses predict a business' or an individual's financial situation in a much better way than ever before and help them thrive in the economic sphere.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background and Motivation

Bankruptcy is a process in which a company can longer legally afford to repay their debts to their creditors. The process starts with the debtor filing a petition. Their assets are valued and measured. These assets are used to repay the loans.

The Bankruptcy Act, 1997 is what defines bankruptcy in Bangladesh. An institution or an individual can be declared bankrupt, if they commit an act of bankruptcy according to the ninth section of this act.

There are three main types of bankruptcies:

Chapter 7 bankruptcy: This is the most common type of bankruptcy. Most individuals or businesses file a chapter 7 bankruptcy. Such a type of bankruptcy allows people to remove their unsecured debts which include bills and credit cards.

Chapter 11 bankruptcy: This kind of bankruptcy is the second most common bankruptcy. Individuals or businesses file for this type of bankruptcy with the goal of restructuring themselves and remaining a legit business.

Chapter 13 bankruptcy: Chapter 13 bankruptcy is very uncommon and is rarely filed by a business or an individual. This is a type of bankruptcy similar to Chapter 7. This bankruptcy can be filed by all those types of businesses in Chapter 7, but have too much capital or monetary assets. It is mainly filed to take precautions and build repayment plans.

According to dun&bradstreet's 2023 Global Bankruptcy Report, the number of bankruptcies grew by over 60% after 2022. The number was 50% in 2021 and 40% in 2020. Therefore, it is clear that bankruptcy is on the rise and there is a need to predict, forecast and tackle this problem. Such a factor affects the economy greatly in a country.

The economy of Bangladesh, especially after the recent events, has taken a massive hit. Businesses are struggling to find their feet again and there is an overall uncertainty regarding the

monetary issues in the country. The following research aims to provide a new way to predict the company's future monetary condition and plan out a repayment or restructuring plan.

1.2 Purpose and Goal of the Project

Most of the techniques used for predicting bankruptcy are outdated. There needs to be a better model to predict and help businesses plan their finances. The project aims to do just that. The research aims to keep in mind the rapid development of Bangladesh in the technological and economic sphere, and produce a result that will overall benefit the country's overall monetary condition.

The use of technology in economy and financing is still greatly lacking in a country like Bangladesh, and we hope to bridge the gap even just a little bit through helping businesses and individuals in predicting bankruptcy. The project aims to:

- Contribute to the growth and development of businesses positively.
- Produce a good economic outcome for the country through preventing bankruptcies.
- Provide the best possible method to predict bankruptcies.
- Find the best machine learning models to predict bankruptcy and provide a basis for future works.
- Use the research to replace primitive and outdated techniques to predict bankruptcy.
- Allow businesses and individuals to take precautions and plan out their finances.

The project is different from existing works because it uses different machine learning and deep learning models which have not been used before. The project uses new techniques other than primitive bankruptcy prediction techniques. The project also looks to improve on existing machine learning and deep learning techniques by tuning the models. The research comes at a time when businesses are struggling in Bangladesh and we hope to bridge the gap between the economy and technology through our research.

1.3 Organization of the Report

The paper is divided into 5 chapters. Chapter 1 is the introduction which describes the goals of the project and introduces the research. The next chapter, chapter 2, presents the literature review and talks about existing works in this field. Chapter 3 describes the methodology for this paper.

It shows the designs, components and implementations for our project. Chapter 4 presents the analysis, results and discussion of our project which will discuss the performance of all the models used in this project. Finally, Chapter 5 concludes the paper with summary, limitations of the project and future improvements that can be done to develop the research.

Chapter 2 Research Literature Review

2.1 Existing Research and Limitations

While there have been works done on bankruptcy and financial status prediction in the past, the existing works do not use the number of models used in this research. One such work is More et. al. (2023) [1], which used the SVM model to predict bankruptcy with an accuracy of 0.96. Similarly Narvekar et. al. (2021) [2] found XGBoost model to be the best model for forecasting bankruptcy. Wang (2017) [3] found neural network with dropout to have the most accurate results. On the other hand, Shetty et. al. (2022) [4] used the XGBoost and SVM models to predict Bankruptcy. Clement et. al. (2022) [5] used an ensemble method to predict the most accurate results. The Random Forest model performed the best for Joshi et. al. (2023) [6]. Soui et. al. (2019) [7] used stacked auto-encoders to find the bankruptcy forecasts.

Although the above researches and works have predicted bankruptcy quite accurately, there are areas to even increase the accuracy. Our research has used fine-tuning and optimization wherever it was possible. The accuracy has been increased to the maximum extent possible. We have also used more machine learning and deep-learning algorithms than the number of models used in existing works. We have also used more evaluation metrics to compare the accuracy among different models. Therefore, our research aims to fill the gaps left by previous works and present the best technique possible for predicting company bankruptcy and financial situation.

Chapter 3 Methodology

3.1 System Design

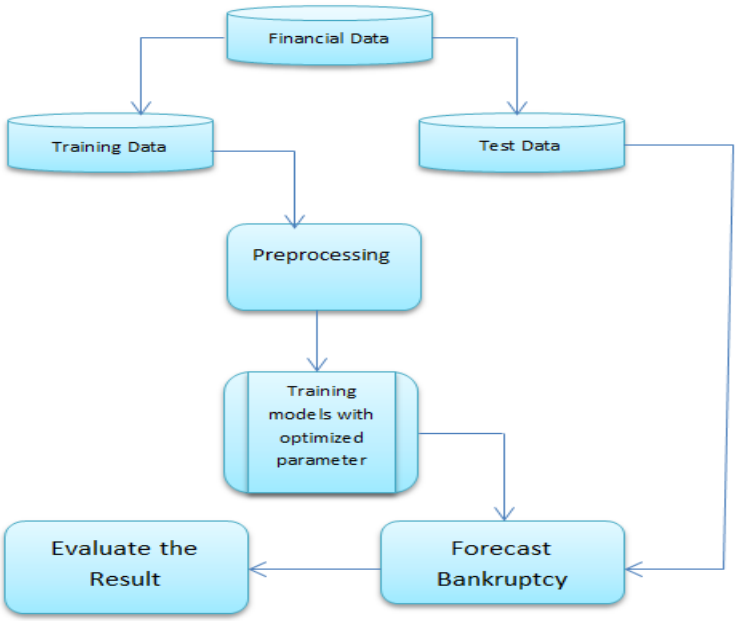

Figure 1. Workflow of the Project

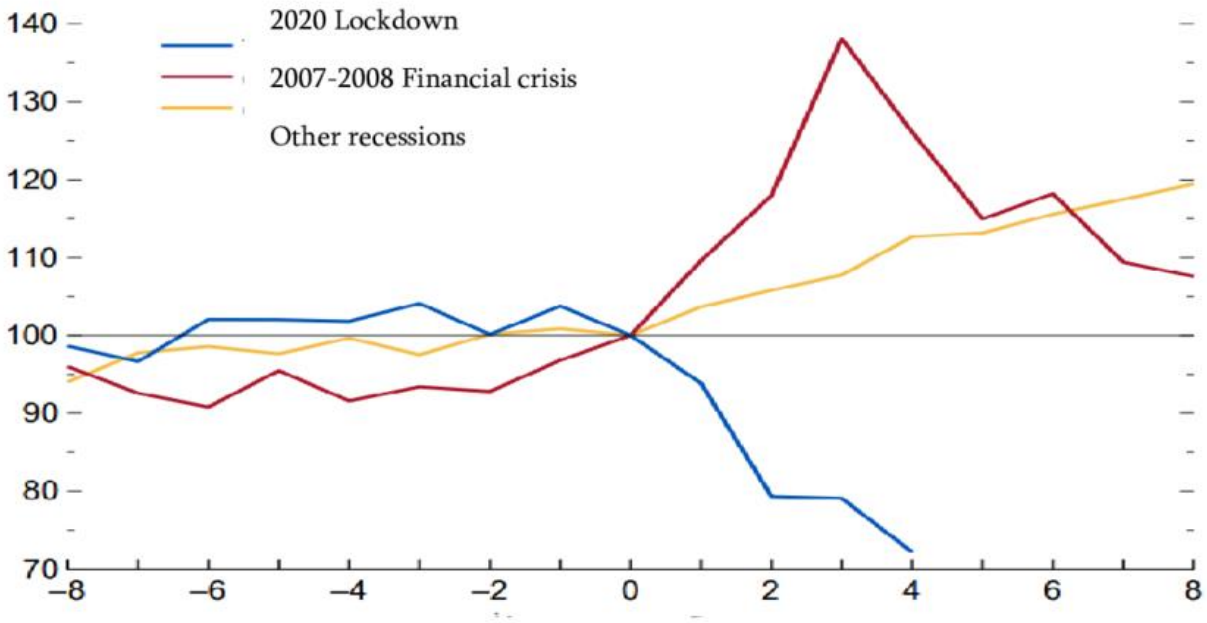

Figure 2. Global Bankruptcy Chart

3.2 Hardware and/or Software Components

Empirical Models:

Data Preprocessing: Before applying ML and DL methods, we firstly pre-processed our dataset using minmax scaler technique which is used to balance the value of each feature so that our model was not biased by a particular class or feature. We also applied Standardscaler method to scale our data and SMOTE (synthetic minority oversampling technique) to balance our dataset before applying different deep learning methods.

Support Vector Machine: This model is a supervised machine learning model. The model uses a hyperplane to separate the data points optimally. It maximizes the margin between points of particular classes. SVM was used for this project because it has better stability and explanatory power [8].

Random Forest: Random forest model is a collection of decision trees. By forming a collection it prevents overfitting. The trees are randoMLy selected using random sets of data. We chose random forest because it already achieved an accuracy of over 90% in Li et. al. (2021) [9], and we tuned it even more and get even better results.

Logistic Regression: This model is also a supervised model. The logistic regression algorithm is used in problems related to binary classification. Logitsic regression is a robust and versatile model that does not require the normality in data distribution to predict financial conditions [10]. Therefore, we chose the model as part of our project.

Ensemble Method: Ensemble method is a method where the different machine learning methods are combined to form a new model. By doing so, the accuracy of the new model formed increases. In this project we used the ensemble of the tuned and balanced random forest models.

Dense Neural Network: This is a deep learning model where the layers are interconnected. Each layer is connected to the layer before. These layers contain nodes. The nodes join the inputs

from different layers and use a weighted coefficient to change the value. The probabilistic nature of neural networks makes it ideal for forecasting bankruptcy and financial conditions [11].

Gated Recurrent Unit: GRU is another deep learning method that we used in our project. Gated recurrent unit is a model that is an extension of recurrent neural networks. GRU is useful in working with sequential data. GRU works on the basis of gating. There are two gates working here: the reset and the update gate. Due to this mechanism, GRU is ideal for predicting bankruptcy as it can be used for modeling dependencies in the long run and also understand how much of exiting information is required to predict bankruptcy [12].

Temporal Convolutional Network: This deep learning neural network model is used to work on trajectory related tasks. These models have a better and longer memory than normal convolutional neural networks. We used TCN in an attempt to convert data to a topological network.

TABLE I. A SAMPLE SOFTWARE/HARDWARE TOOLS TABLE

Tool	Functions	Other similar Tools (if any)	Why selected this tool
Tensor Flow	**Creating new models, model training and importing new tools.**	**PyTorch**	**Is open source, efficient, and scalable.**
Pandas	**Data management, cleaning and sorting.**	**Numpy, SQL**	**Interoperable, easy to use and fast.**
Numpy	**Mathematical and statistical computations, building graphs.**	**MATLAB, pandas**	**Robust, fast and inteoperable.**
Matplotlib	**Visualize data, plot graphs and charts.**	**Seaborn**	**Easy to implement and integrate in code, labeled accurate visualizations.**
Jupyter	**Used as IDE to write and edit code.**	**VSCode, Google Colab**	**Easy to use, proper interface, easy to collaborate.**
Google Sheets	**Used to store the dataset.**	**Microsoft Excel**	**Convenient, easy to integrate, good interface.**
Imblearn	**HanDLing imbalance among classes. (SMOTE)**	**ADASYN, NearMiss**	**Removes imbalance in dataset, provides easier to use dataset.**

Google Colab	**Used as IDE to write and execute code.**	**Jupyter, VSCode**	**Cloud-based, interoperable, fast and easier sharing.**
Scikit-Learn	**Importing libraries, tools and models.**	**PyTorch, Tensor Flow**	**Easy to use, open source, accessible, contains all necessary libraries.**

3.3 Hardware and/or Software Implementation

To start off, we first imported all the necessary modules and libraries using scikit-learn, tenor flow, numpy, pandas and imblearn. We then pre-processed the data. We used standard scaling techniques to do so.

The project used Support Vector Machine, Random Forest, Logistic Regression and Ensemble Method to forecast bankruptcy. We used different fine tuning and optimization techniques as mentioned above.

The outputs obtained were then visualized. The matplotlib module was used for visualizing the output. We built bar graphs and scatter plots to visualize the predictions and compare them to the training data. We used log loss, R2 score, confusion matrix, training accuracy and validation accuracy as the metrics to compare accuracy among different models.

To execute our project on forecasting inflation, we use dataset which contained the following features: ROA(C) before interest and depreciation before interest, ROA(A) before interest and % after tax, ROA(B) before interest and depreciation after tax, Operating Gross Margin, Realized Sales Gross Margin, Operating Profit Rate, Pre-tax net Interest Rate, After-tax net Interest Rate, Non-industry income and expenditure/revenue, Net Income to Total Assets, Total assets to GNP price, No-credit Interval, Gross Profit to Sales, Net Income to Stockholder's Equity, Liability to Equity, Degree of Financial Leverage (DFL), Interest Coverage Ratio (Interest expense to EBIT), Net Income Flag and Equity to Liability. We used the ‘Bankrupt’ factor as the target variable.

We scaled the data using the Standard Scaler technique. Then we split the data into training and testing data. We also fit the data as required. The models used in the research were Support

Vector Machine (SVM) with SMOTE, ensemble method and tuned random forest. We also used logistic regression with class weight adjustment and increased iteration. To select the features, we used correlation method.

For hyperparameter optimization, we used different tuning methods. We used the SMOTE technique to increase the accuracy of SVM and decrease the imbalance in class. We tuned the random forest method by using grid search, defining a parameter grid and using the best estimator or model. To use the ensemble method, we combined the tuned and balanced random forest models. The logistic regression model also gave good results when paired with class weight adjustment and increased iterations.

We then performed the experiments for deep learning. We used the dense neural network, gated recurrent unit (GRU), and Temporal Convolutional Network (TCN) models to predict bankruptcy.

Chapter 4 Investigation/Experiment, Result, Analysis and Discussion

All of the models used in this project, with proper tuning and hyperparameterization, performed exceptionally well. The models' accuracy were as follows:

TABLE II. TRAINING ACCURACY OF MODELS

Models	Training Accuracy
SVM	0.916
Random Forest	0.991
Logistic Regression with Class-weight adjustment and increased iterations	0.936
Ensemble Method	0.987
Dense Neural Network	0.980
GRU	0.985
TCN	0.988

We have applied nine machine learning models in total. But finally, we considered the four best models among them based on their performance. If we observe the training summary model, we can see that SVM model after applying SMOTE, which is a popular technique in ML used to deal with imbalance dataset, ended up achieving an accuracy of about 91%. We also applied random forest which is a powerful ML method. Initially when we applied Random Forest, we saw a case of overfitting there. So, we started hyperparameter tuning on the model and we obtained an accuracy of 99%, Random Forest model gave the best performance among all models we applied here. We also applied Logistic Regression here but initially we saw that our model was performing abnormally and this was happening because the models during that time became biased to the majority class of the dataset. So, we started to adjust the weights of classes in our dataset and then increase the number of training iterations and the logistic regression finally was able to give a balance performance with an accuracy of 93%. Then we also applied ensemble method by combining both balanced Random Forest Classifier and regular Random Forest Classifier and we obtained an accuracy of 98%.

The following figure shows the performance of each of the models used in the research project. The figure displays the training accuracy and validation accuracy of the models used.

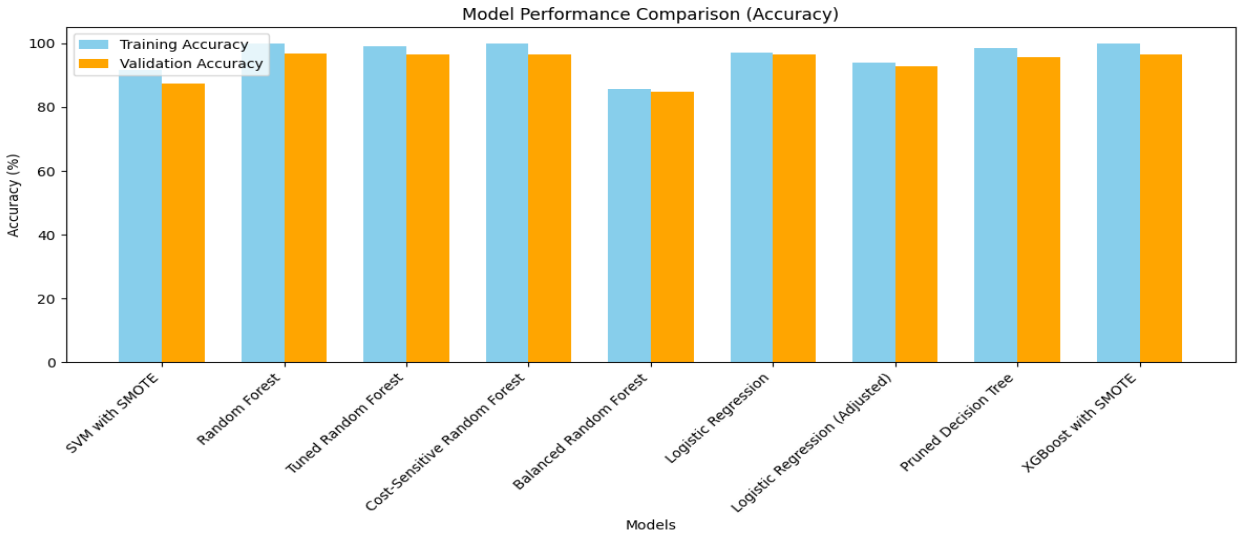

Figure 3. Training vs. Validation Accuracy

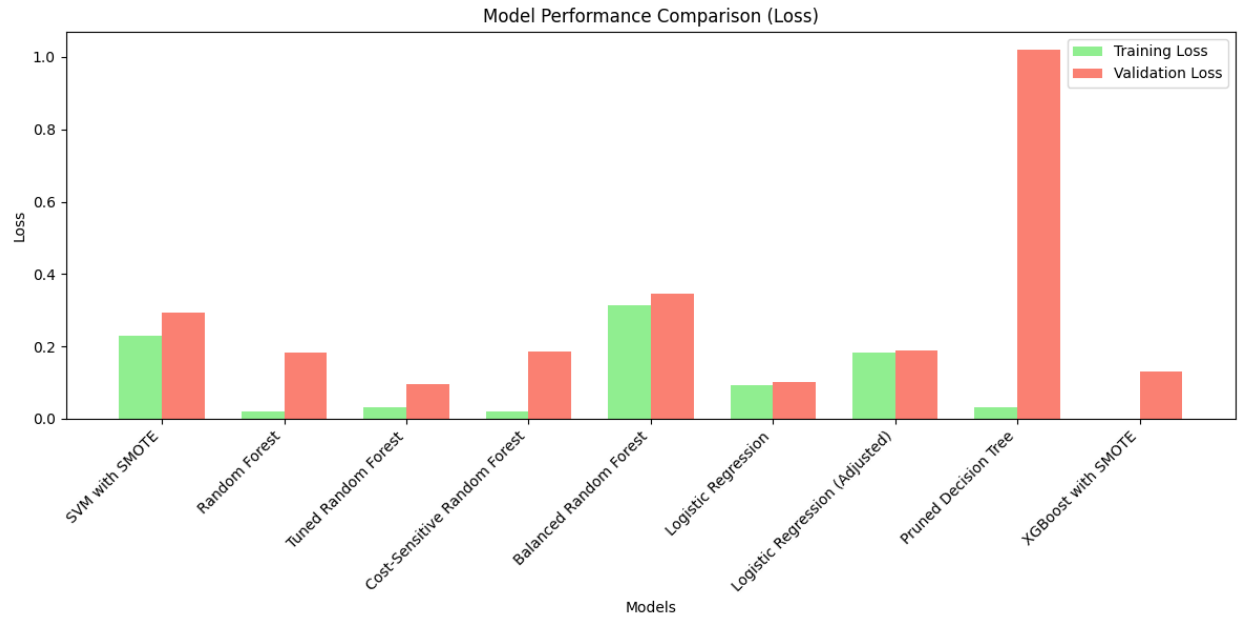

Figure 4. Training vs. Validation Loss

The above figure displays the training loss and validation loss of the models used. If we observe the bar chart above, we can see that Pruned Decision Tree method performed unexpecteDLy in terms of validation loss. On the other hand, models like SVM with smote, Tuned Random Forest, Logistic regression etc. provided a significantly better performance with small amount of validation and training loss.

The confusion matrices of the machine learning models have also been presented below:

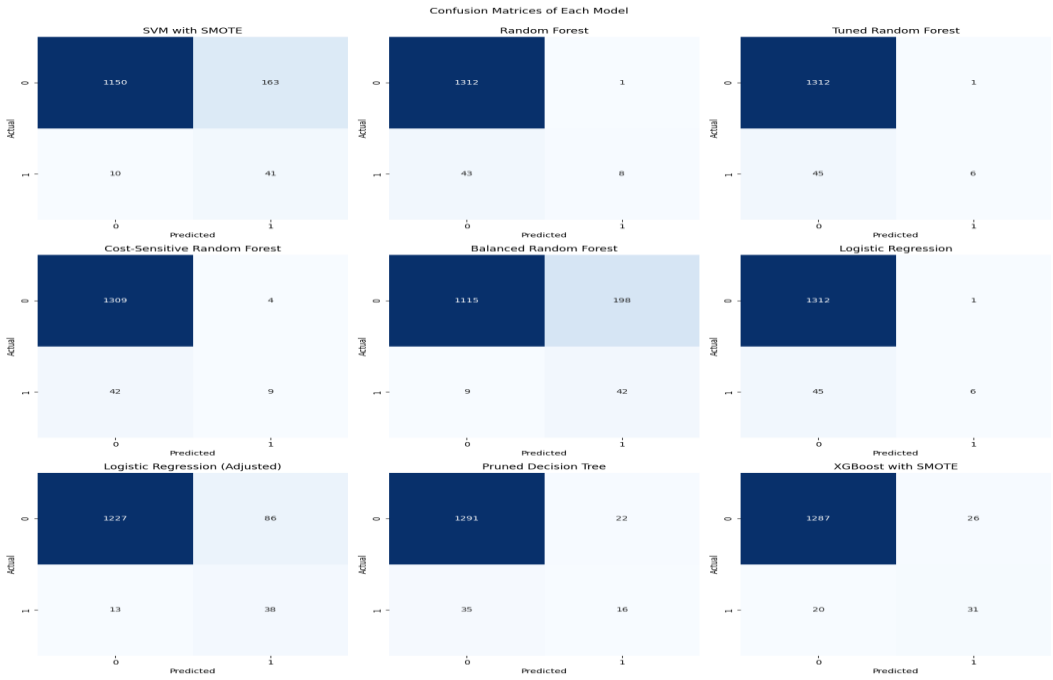

Figure 5. Confusion Matrices

The below model compares the performance of the three deep learning models:

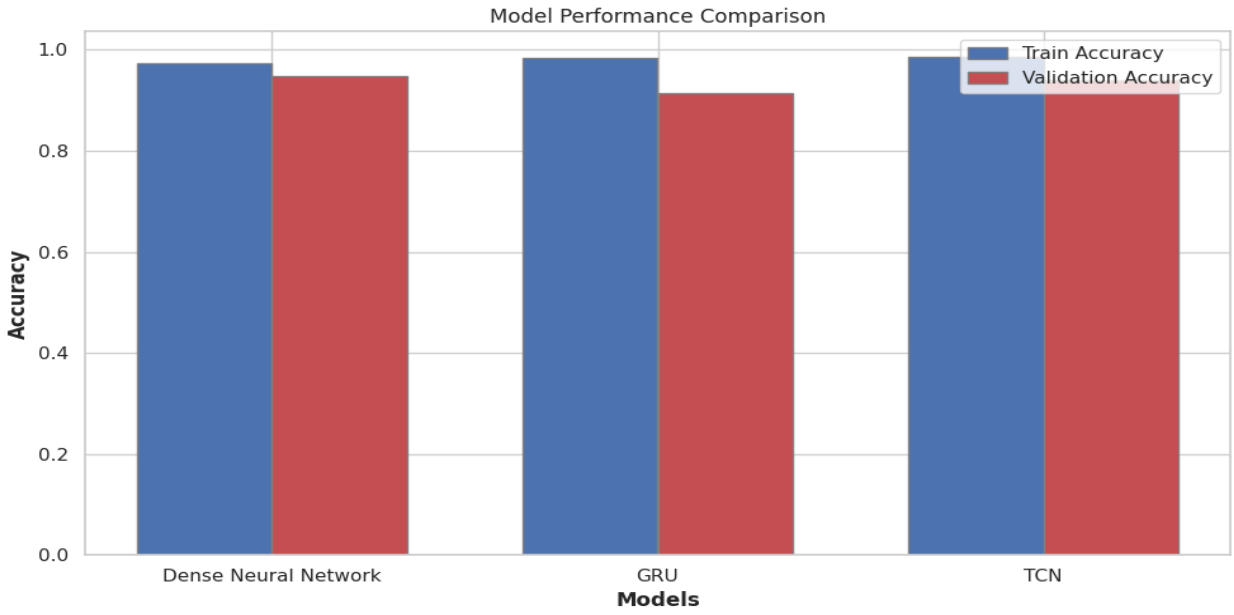

Figure 6. Deep Learning models' training vs. validation accuracy

The accuracy during training for deep learning models is presented below. This graph is graphically demonstrating the performance of each model during training and validation process per epochs.

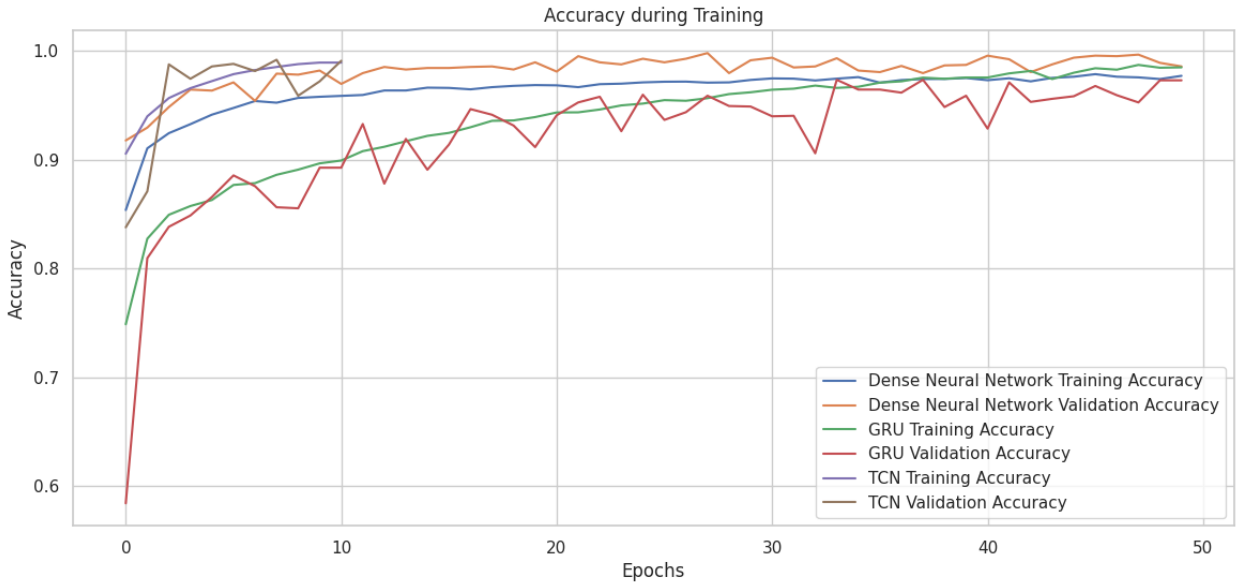

Figure 7. Deep Learning models' accuracy during training

Chapter 5 Conclusions

8.1 Summary

Bankruptcy is one of the biggest problems faced by businesses. There is a need to accurately predict and take measures to avoid bankruptcy. Businesses need to maintain a healthy financial condition to stay afloat and keep working. The above paper presents such a way to tackle this exact problem.

Technology and finance go hand-in-hand. The above paper presented four machine learning methods to predict and prevent bankruptcy. The paper presents four machine learning methods to forecast bankruptcy: Support Vector Machine, Random Forest, Logistic Regression and Ensemble Method. The deep learning models used were dense neural network, gated recurrent unit (GRU), and Temporal Convolutional Network (TCN).

The paper presented an accurate prediction of bankruptcy by tuning the aforementioned models. The models performed exceptionally well as compared to previous works. The research aimed to provide the best possible technique to predict bankruptcy and, to improve and fill in the gaps in existing works.

The paper compared the accuracy of different models using various comparison metrics. The models have all performed well. But in particular, the random forest model performed the best with an almost perfect accuracy close to 0.98 (training accuracy).

8.2 Limitations

The project has some limitations. All the models have performed particularly well, but in different circumstances with even more factors involved, other models which have not been mentioned in the paper could be used to obtain even better results.

The project has used financial data from a few companies. The addition of more companies with their financial details could have given even better results. The research paper has provided a model to predict bankruptcy but it is still not usable by the general public as there is no interface or any other way to access and use this method.

The project has also used the above machine learning and deep learning models with tuning and hyperparameterization techniques to predict bankruptcy and financial analysis. The research project could have used more deep learning and machine learning models to predict the same.

8.3 Future Improvement

To improve the project, we will be using even more data to get even more accurate results. We hope to work with companies and other financial institutions to provide this project as a way to prevent bankruptcies and poor financial conditions.

The project can be improved further by using more machine and deep learning models and tuning them to the maximum extent. The project can also use a user interface where companies can enter their financial data and check their financial situations and stability.

With this project, we hope to contribute positively to the economy of this country and bring about a wave of economic changes combined with technology. To do so, we hope to work and add even more features to our project.

The project can be further improved by adding and comparing primitive econometric models to the machine learning and deep learning models used in this research project. The paper can also be improved through adding inputs and summaries from companies and how much these companies and businesses would be willing to incorporate machine learning into their daily financial tasks.

References

[1] Manisha More, Rajasmita Panda, Bapurao Bandgar, and Mayuri More. Bankruptcy prediction using machine learning: A new technological approach to prevent corporate bankruptcy through well deployed streaMLit based application. In 2023 International Conference for Advancement in Technology (ICONAT), pages 1–5. IEEE, 2023.

[2] Aditya Narvekar and Debashis Guha. Bankruptcy prediction using machine learning and an application to the case of the covid-19 recession. Data Science in Finance and Economics, 1(2):180–195, 2021.

[3] Nanxi Wang et al. Bankruptcy prediction using machine learning. Journal of Mathematical Finance, 7(04):908, 2017.

[4] Shekar Shetty, Mohamed Musa, and Xavier Br´edart. Bankruptcy prediction using machine learning techniques. Journal of Risk and Financial Management, 15(1):35, 2022.

[5] Claudiu Clement, Mihaela David, and Danut,-Vasile Jemna. Bankruptcy prediction using machine learning–a meta-analysis. Editorial board, page 62, 2022.

[6] Shreya Joshi, Rachana Ramesh, and Shagufta Tahsildar. A bankruptcy prediction model using random forest. In 2018 second international conference on intelligent computing and control systems (ICICCS), pages 1–6. IEEE, 2018.

[7] Makram Soui, Salima Smiti, Mohamed Wiem Mkaouer, and Ridha Ejbali. Bankruptcy prediction using stacked auto-encoders. Applied Artificial Intelligence, 34(1):80–100, 2020.

[8] Jae H Min and Young-Chan Lee. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert systems with applications, 28(4):603–614, 2005.

[9] Yachao Li and Yufa Wang. Machine learning methods of bankruptcy prediction using accounting ratios. Open Journal of Business and Management, 6(1):1–20, 2017.

[10] Milica Vuk˘cevi´c, Milan Laki´cevi´c, Boban Melovi´c, Tamara Backovi´c, and Branislav Dudi´c. Modern models for predicting bankruptcy to detect early signals of business failure: Evidence from montenegro. Plos one, 19(5):e0303793, 2024.

[11] ZR Yang, Marjorie B Platt, and Harlan D Platt. Probabilistic neural networks in bankruptcy prediction. Journal of business research, 44(2):67–74, 1999.

[12] Michał Thor and Łukasz Postek. Gated recurrent unit network: A promising approach to corporate default prediction. Journal of Forecasting, 43(5):1131–1152, 2024