

**UNIVERSITI TEKNOLOGI MARA**

**TECHNICAL REPORT**

**CORPORATE DEFAULT RISK PREDICTION USING DECISION  
TREE, RANDOM FOREST AND GRADIENT BOOSTED  
MODELS**

**NUR SABRINA SHAMSUL –2020820002  
FAYYADH IZZUDDIN AHMAD SHAH –2020878552  
NURUL ZAHIDAH BINTI RIDZUAN –2020898312  
(P27M23)**

**Report submitted in partial fulfillment of the requirement  
for the degree of  
Bachelor of Science (Hons.) (Mathematics)  
College of Computing, Informatics and Mathematics**

**17TH JULY 2023**

## **ACKNOWLEDGEMENTS**

Firstly, we are grateful to Allah S.W.T for giving us the strength to continue our journey in completing this project successfully within the given time. Much time and energy had been sacrificed to achieve our goals in this research.

We would like to express our sincere gratitude to our supervisor, Madam Norliza Muhamad Yusof, who patiently guided us from the beginning until the end of our project. Her ideas, enthusiasm, patience, insights and unceasing ideas are truly a big help in finishing this project. Her immense knowledge and profound experience has enabled us to complete this project successfully. We could not imagine having a better supervisor than her.

We would also like to thank our lecturers, Dr Nurul Liyana Abdul Aziz, Dr Noraimi Azlin Mohd Nordin, and Madam Nur Lina Abdullah for guiding us since last semester until now for our report writing. Without their continuous guidance, this report will not be as good as it is now. Not only them, without our friends' and families' encouragement since the beginning of this project, we would not have the strength to finish it. Thank you for not giving up on us with all of the support.

Last but not least, a big appreciation to ourselves who endured a lot of challenges to complete this study from the start until the end. We are so strong to keep up all night, skipping meals, brainstorming together just to get a better result for our study. Therefore, we hope this project will be appreciated by people as it is by us.

Thanks for all of your encouragement!

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS.....</b>	<b>1</b>
<b>TABLE OF CONTENTS.....</b>	<b>2</b>
<b>LIST OF TABLES.....</b>	<b>4</b>
<b>LIST OF FIGURES.....</b>	<b>5</b>
<b>ABSTRACT.....</b>	<b>5</b>
<b>CHAPTER 1.....</b>	<b>1</b>
<b>INTRODUCTION.....</b>	<b>1</b>
1.1 Problem Statement.....	1
1.2 Objectives.....	2
1.3 Significant and Benefit of Study.....	2
1.4 Scope and Limitation of Study.....	3
<b>CHAPTER 2.....</b>	<b>4</b>
<b>LITERATURE REVIEW.....</b>	<b>4</b>
2.1 Default Risk for Corporates.....	4
2.2 KMV-Merton Model and its Application.....	5
2.3 Decision Tree (DT) Models.....	7
2.4 Comparison of DT, RF and GB Models.....	12
<b>CHAPTER 3.....</b>	<b>14</b>
<b>METHODOLOGY AND IMPLEMENTATION.....</b>	<b>14</b>
3.1 Setting Up Data.....	14
3.1.1 Data Collection.....	14
3.1.2 Setting Up Independent and Dependent Variable.....	16
3.1.3 Calculate The Financial Ratios.....	16
3.1.4 Estimating Probability of Default.....	17
3.1.5 Data Storage.....	17
3.2 Data Implementation.....	17
3.2.1 Predicting Default and Non-Default of Corporations.....	19
3.2.2 Determining the Models Accuracy.....	20
<b>CHAPTER 4.....</b>	<b>22</b>
<b>RESULT AND DISCUSSION.....</b>	<b>22</b>
4.1 The Prediction of Default and Non-Default of Corporations.....	22
4.1.1 Decision Tree Model.....	22
4.1.2 Random Forest Model.....	26
4.1.3 Gradient Boosted Model.....	31
4.2 The Best Models to Predict The Corporations Default.....	34

<b>CHAPTER 5.....</b>	<b>37</b>
<b>CONCLUSION AND RECOMMENDATION.....</b>	<b>37</b>
<b>REFERENCES.....</b>	<b>39</b>

## **ABSTRACT**

Predicting a corporate default is essential to avoid failure in business operations. Default is said as an unpredicted event that may cause a failure in paying debts on time. This study aims to predict the default and non-default of corporations using Decision Tree (DT), Random Forest (RF) and Gradient Boosted (GB) models. Not only that, the best models used in predicting the corporate default are determined by using cross validation on the models performance evaluation (classification). A sample of financial data of 24 Malaysian companies consisting of companies listed as Practice Note 17 (PN17) and main market companies in the Bursa Malaysia board is used. The data is utilized to estimate the prediction of the companies by using the KMV-Merton model. It is also used to calculate variables such as Net Profit Margin, Cash Flows, Cash Coverage Ratio, Current Ratio, GDP, Currency Rates and Volatility. Then, analysis is done on the dependent and independent variable by using Rapidminer Studio Educational software. As a result, the accuracy of RF is 84.17% which is the highest followed by the accuracy of GB and DT is 79.17% and 75.83% respectively. Therefore, RF is the best model to use for predicting default and non-default companies. This study encourages increasing the use of machine learning models in default forecasting. Thus, It is recommended to increase more data based on the most relevant independent variables that are used in the study. Then, by using a combination of different machine learning models and software can yield a better performance and improve the overall effectiveness of the models.