Factor Affecting Credit Risk

SD22061

ANG MEI YING

BSD2712 RESEARCH METHODOLOGY

RESEARCH PROPOSAL

ABSTRACT

The major activity of the banking industry is lending money to people who need money. To minimize the rate of credit risk, this research was conducted and aims to identify the factors that affect credit risk within Malaysian commercial banks. There are some factors that were determined based on literature review. However, this research paper will focus on examining the borrower's characteristics that affect credit risk. Gender, income level, education level, marital status and other borrower's characteristics are the variables that will be used to conduct the research and eventually evaluate the probability of default based on their characteristics. Quantitative research methods had been used in this research by using the data obtained from Kaggle. By analyzing borrower characteristics from 1000 loan applicants, two machine learning models were built. The logistic regression and random forest models were developed to analysis and forecast the default status of a given borrower. The expected results have shown that the random forest is the most effective machine learning approach in this study for identifying significant borrower's characteristics that affect borrower ability to repay their loan. This discovery has an important impact for the use of machine learning in banking systems, helping bankers improve their decision making in loan approval and reduce potential losses cause by loan default of borrower.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

1.1 Research Background

Commercial banks are places where majority of people do their banking activities (Kagan, 2023). Commercial bank can be defined as one of the financial institutions in which the main activities carried out by them are to accepts deposits, offers checking account services, make various loans, and provide basic financial products such as certificates of deposit and savings accounts. Loans granted and interest income generated from these loans are the sources of revenue for commercial banks (Kagan, 2023). However, the commercial banks has the potential to expose to credit risk by providing loans such as mortgages, auto loans, business loans and personal loans (Giesecke, 2005). Credit risk is the most critical risk looked by banks and the achievement of their business is depends on the accuracy estimating and efficiency managing the risk (Giesecke, 2005). Credit risk refers to the risk that a lender may not be paid back by a borrower and the probability of a financial loss resulting from a borrower's failure to repay a loan (Serwadda, 2018).

Credit Risk occur when a borrower fails to repay the loan as scheduled within the agreed timespan (Bank for International Settlements, 2006). This could arise due to various reason. The study from Sukriti and Ajay K.Shah (2019) claims that the factors that influence a borrower for timely loan repayment relating to clients' characteristic, business characteristic and lender's characteristic(Shah et al., n.d.).Other than that, economic conditions like GDP growth, unemployment and real interest rate have affect their ability to repay the loans (Berge & Boye, 2007).Combination of individual and economic factors and others financial challenges make it difficult for borrowers to repay the loan, ultimately contributing to credit risk (Haryono et al., 2016).For example, economic downturns can result in lower income levels and increased unemployment rates. Furthermore, individual personality such as a poor credit history also can has the higher probability of loan default due to lack of creditworthiness.

Besides that, there are various stakeholders involved in credit risk. At the forefront are lenders, including banks, credit unions, and other financial institutions, who provide loans to borrowers. These lenders are directly exposed to credit risk as they face the probability of financial losses if borrowers default on their loans (Bank for International Settlements, 2006). Next are borrowers who borrow the loan from lenders. Their financial status, creditworthiness, and ability to repay loans significantly impact the credit risk (Goel & Rastogi, 2023). Furthermore, investors and shareholders of financial institutions are indirectly affected by credit risk. When borrowers don't repay the loan, those institutions will lose money. This makes the institution less profitable, and they might have to set aside more money to cover the losses or reduce the dividends to shareholders (Rahman & Shahimi, 2010).

The type of research in this case study is quantitative research. There are some past studies that had been conducted on credit risk using quantitative approach which involves the collection and analysis of data related to credit risk. For example, research from David Durand, quantitative data on various factors that affect credit risk such as age of borrower, personal income, credit history, marital status and other information related to borrower are gathered to analyse the credit risk. To overcome the credit risk, various methodologies have been proposed by researchers. One of the researchers which is Turjo et al. have proposed Credit Risk Assessment Model (CRAM) which uses heatmap to select features and different machine learning models in order to predict credit risk. However, the problem with all of these existing methods is that they use only one classification model for predicting credit risk. In most cases, the single classifier-based models do not reduce the variance and identify the biases if there are any biases present in the data (Turjo et al., 2021). Therefore, the present study aims to develop a more alternative model to predict the credit risk more accurately.

Credit risk can have far-reaching effects on the community. A high level of credit risk can lead to financial instability (Ismail & Ahmed, 2023). If many borrowers default on their loans at the same time, it can strain the resources of financial institutions and even lead to bank failures or bankruptcy (Ekinci & Poyraz, 2019). Credit risk has a significant impact on the profitability of banks as a large proportion of banks' revenue and interest income is derived from loans (Twum et al., 2021). Besides that, when most of the borrowers fails to repay its debts, it will affect the economic stability for being move toward crisis and eventually cause instability of financial(Majani, 2022). Furthermore, higher credit risk can cause banks to become more caution and tighten their lending standards. This could restrict individuals and businesses from accessing funds needed for essential investments such as buying homes or starting new businesses. Uncertainty about stability of financial institutions and the ability of borrowers to repay their loans can lead to investor caution and reluctance to make long-term investments (Boye & Kwabena, 2014). This may cause economic volatility and affect inflation rates and GDP growth. Therefore, it is important to manage credit risk effectively to keep sustainable economic development within communities.

There are many studies on credit risk among banks worldwide, however, only a few have been conducted in Malaysia involving the banking sector by focusing on the commercial bank. In addition, only a few studies have been carried out to assess the impact of credit risk factors on credit performance, and no precise findings or clear conclusion are drawn about the elements of credit risk management that can reduce credit risk. Therefore, it is crucial to determine what factors that affect credit risk and develop an alternative model to predict and manage credit risk within Malaysian commercial banks.

1.2 Problem Statement

Credit risk is by far the most significant risk faced by banks. In recent years, banks and other financial institutions have faced credit risk management problem in reducing credit risk. Although there are some researchers have built predictive models using machine learning and traditional statistical models to estimates the probability that a customer will be able to pay back a loan, however there are remains a gap in understanding the comprehensive factors influencing credit risk and it is not updated to recent. Therefore, the problem addressed in this study is to identify the factor that affecting credit risk and develop a robust predictive model based on identified factor to forecast credit risk accurately and reliably, with a focus on enhancing predictive performance. By addressing this problem, the study aims to provide lenders and others financial institutions with valuable insights for making informed decisions in the realm of lending and risk management.

1.3 Research Questions

- What is the factor that affecting credit risk?
- What is the model that can be developed using machine learning to evaluate the credit risk based on the borrower characteristics?
- What is the performance of the developed model in predicting credit risk?

1.4 Research Objectives

- To identify the factors that affecting the credit risk
- To develop a model for credit risk based on the borrower characteristics using machine learning
- To investigate the performance of the developed model in predicting credit risk within Malaysian commercial banks

1.5 Scope of Research

The study aims to investigate the factors affecting credit risk within the commercial banking sector of Malaysia. By conducting a comprehensive literature review, relevant factors affecting credit risk will be identified such as economic indicators, borrower characteristics, and financial market condition specific to Malaysia. However, the scope of this research will narrow as to examine how borrower characteristic affect credit risk. The geographic focus of this study will be limited to Malaysia and the target population are Malaysian who involved in borrowing and lending money, like loan applicant, borrowers, banker, and lender. In addition, the research scope will primarily center on the application of machine learning techniques to construct predictive models. Model evaluation will be conducted to assess the predictive performance of the developed model.

1.6 Significance of Research

Financial sector is crucial to the economic development and growth across all countries. For most businesses, banking sector is regarded as the major source of finance. However, there are instances where borrowers fail to repay their bank loans which may lead to occurrence of credit risk. Therefore, it is important to manage the credit risk effectively for ensuring financial stability and economic growth.

This study is going to identify various factors that affecting credit risk. It is essential to understand how the identified factors influence borrower's ability to repay the loan. From this research, it will give a new view of knowledge to users to identify the factor affecting credit risk. This information help users to manage their finances effectively including making timely payments, reducing debt, and maintaining a healthy credit utilization ratio to improve their creditworthiness over time. Other than that, by investigating the factors affecting credit risk, the study aims to identify the relationship between the factors and credit risk and explore how different factors interact with and give impact on credit risk of banks. Thus, banks can minimize their losses associated with credit risk and maintaining financial stability.

Besides that, the development of predictive model for credit risk based on borrower's characteristics offers valuable insight for risk management in Malaysian commercial banks. Credit risk assessment using predictive modelling can help businesses or banks make informed decisions about credit approval, reduce the risk of default, increase profitability, and stay ahead of changes in the credit market. By evaluating the performance of the developed model in predicting and managing the credit risk, banks can consider it is the alternative model that can be used to determine the credit risk for each person, ultimately help them make a better lending decision.

In addition, borrowers also can benefit from the development of credit risk model. The models aim to provide a more accurate and fair assessment of a borrower's creditworthiness based on various factors such as credit history, income, and other relevant information. Therefore, an alternative developed model can help borrowers with strong credit profiles more likely to be approved for loan and obtain favourable terms, such as higher loan amounts and lower interest rates. Furthermore, credit risk models enable investors to understand the level of risk associated with their investment. It helps investors better manage their investments, ensuring the minimization of losses and maximization of return.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter contains a summary of the observations and conclusions made by previous research on factors that affect commercial banks' credit risk, as well as variables and methods they have applied. This study therefore sought to identify some of the literature that defines and elaborates theories concerning each dependent and independent variable, including bank credit risk measurements.

2.2 The concept of credit risk

Credit risk is the possibility that a borrower or counterparty does not fulfil their contractual obligations. Credit risk management aims to maximise the bank's risk adjusted return by maintaining credit risk exposures in accordance with accepted parameters (Pandey & Joshi, 2023). According to Kusnandar (2022), credit risk refers to the possibility of partial or complete loss of the outstanding debt as a result of a credit event. It also can be expressed as the fact that the contracting party will not be able to fulfil its obligations in accordance with the conditions of agreement. Exist of banks not only for accepting deposits but also to offer credit facilities, therefore inevitably exposed to credit risk. Credit risk is the most important risk that banks confronted. Accurately measuring and effectively managing of credit risk are the key feature to bring success to its operations (Naili & Lahrichi, 2022).

Credit risk is the implied value fluctuation of derivatives and loan instruments which results from changes in borrowers' and counterparties' credit quality (Torban, 2020). Bhatt et al. (2023) posits that credit risk is defined as the losses experienced due to a credit customer's inability or refusal to pay what he owes at the right time and in full. Credit risk is the risk faced by a bank when a borrower fails to meet its debt obligations on or at maturity. This risk can be interchangeably referred to as "counterparty risk" and can put banks in a difficult position if not properly managed.

Table 1 shows the summary of previous research paper about how to identifiy the factors affecting credit risk and credit risk management strategies.

Author	Title	Objective	Method	Finding
Acu Kusnandar	Credit Risk in Islamic Banking	To identify the systematic factor and unsystematic factor that affect credit risk in Islamic Banking	Literature study	There are two factors that may lead to the occurrence of credit risk, which may include unsystematic factors, systemic factors or factors that cannot be eliminated or managed.

Tesfaye Kefale Torban	Assessment of Credit Risk Management in Micro Finance Institutions: A Case of Adama Town MFIs, Ethiopia	-To assess credit risk management practices of Adama town MFIs. - To identify the challenges faced by MFIs in Adama town credit risk management.	Descriptive statistics	Credit risk management of Adama town MFIs is strongly influenced by client assessment, credit risk management and collection policies.

Table 1: Summary of the concept of credit risk

2.3 Main sources of credit risk

The main source of credit risk includes borrowers' poor credit histories, institutional limitations, fluctuations in interest rates, unfavourable borrower characteristics, inadequate management, extensive bank licensing, low capital and liquidity levels, poor loan underwriting, direct lending practices, lax credit assessment and improper lending practices (Bandyopadhyay & Saxena, 2023). It is critical because the default of even a few significant customers can result in substantial losses, potentially leading to insolvency (Rehman et al., 2019).

Abbas et al. (2019) emphasize that the bank's profitability relies on its ability to anticipate, avoid and monitor risks to mitigate potential losses. However, this can lead to an increase in substandard credits in the bank's portfolio and a decrease in profitability.

Credit risk surfaces every time a bank gives out loans, commits its funds, invests its money, or exposes itself to financial risks. This can be attributed to poor risk management practices, adverse economic conditions or many other factors (Bussmann et al., 2021). Nevertheless, there is a common belief that transactional risk, operational risk, oversight risk and portfolio risk grow with the level of credit risk. The importance of credit risk management is also highlighted by previous research and analysis as shown in Table 2.

Author	**Title**	**Objective**	**Method**	**Finding**
Arindam Bandyopadhyay & Mayuri Saxena	Interaction between credit risk, liquidity risk, and bank solvency performance: a panel study of Indian banks	To determine the relationship between credit risk and bank liquidity	Statistics method	Based on empirical result, it is proved that credit risk has a significant impact on banks' liquidity risk. Borrower defaults can result in reduced cash flow and depreciation in loan assets, thereby increasing the risk of liquidity.
Zia Ur Rehman, Noor Muhammad,	Impact of risk management strategies on the	To identify the risk management	Statistics method	The result identified 4 areas of impact on credit risk

Bilal Sarwar, Asif Raz	credit risk faced by commercial banks of Balochistan	strategies that commercial banks have in place		management (CRM): corporate governance, diversification, hedging and bank's capital adequacy ratio.

Table 2: Summary of main sources of credit risk

2.4 Credit history of borrower affect credit risk

An individual's credit history is important in determining credit risk since it show the creditworthiness of an individual (Bhilare et al., 2024). Credit history is the record of an individual's financial information which includes repayment of debt. It generally contains details like history of payment, outstanding debts, credit history duration, types of credit used, and recent credit inquiries.

A strong connection between credit history and creditworthiness has been established through elaborate research. People with a positive credit history, which entails timely payments and good management of credit, have higher credit scores and better chances of accessing loans on favourable terms (Muñoz-Cancino et al., 2023). In contrast, consumers who have no established financial credit or any credit background will have problems borrowing money from banks, and lenders will charge them high interest rates and tighter terms.

Goel & Rastogi (2023) underlined the importance of past repayment patterns in assessing credit risk. The research found out that there is a strong negative relationship between credit risk and past payment history. According to the study, borrowers with a consistent history of on-time payments are less likely to default. In addition, credit scoring models are widely used in credit risk assessment. The study by Bhilare et al (2024) shown that the credit scores obtained from historical credit data significantly contribute to the accuracy of credit risk prediction models. Higher credit scores, indicative of positive credit history, are associated with lower default probabilities and reduced credit risk. Table 3 has shown how the previous researcher evaluates borrower's creditworthiness using credit scoring model.

Author	**Title**	**Objective**	**Method**	**Finding**
Meghana Bhilare, Vishal Wadajkar & Kiran Kale	Empowering Financial Decisions: Precision Credit Scoring through Advanced Machine Learning	To predict credit scoring using machine learning	Machine learning	The research using machine learning includes Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Naive Bayes (NB). Random Forest with an accuracy of 92.5% has the higher performance compared to other machine learning models.
Akanksha Goel &	Understanding the impact of	To identify certain	review method	An objective credit scoring model need to

Shailesh Rastogi	borrowers' behavioural and psychological traits on credit default: review and conceptual model	behavioural and psychological characteristics of borrowers that are likely to impact credit risk		be developed to assess and verify the reliability of debtors based on soft information such as their behavioural and characteristics.

Table 3: Summary of credit scoring modeling from previous research

2.5 Macroeconomic factors affect credit risk

Macroeconomic factors such as economic expansion, interest rate, inflation, and unemployment have strongly impact repayment of borrower and lead to increasing credit risk which it cannot be directly controlled by banks. Research by Sharma et al. (2024) highlights the impact of economic growth on credit risk. According to the study, loan defaults rate often occur at a lower rate during periods of economic expansion as borrowers' income are rising. Thus, it enhances credit quality. Conversely, economic downturns are associated with increased credit risk due to higher unemployment rates.

The unemployment rate is a critical macroeconomic indicator affecting credit risk. Unemployment is the inability to find employment for individuals of working age. Research by Kacerová & Černohorský (2023) have found that there is a positive correlation between unemployment rates and credit risk. Borrowers who are jobless cannot afford to repay their debts since they lack a source of income. Furthermore, loss of employment during economic downturns makes the income of borrowers unstable thus increasing the risk of default on loans and credit. Additionally, the fact that credit risk is very sensitive to changes in interest rates is well demonstrated in research conducted by Wahyudi (2020), with higher rate usually linked with increased likelihood of default for variable-rate loans and mortgages.

Table 4 has shown that previous research mainly focused on analysing credit risk caused by unemployment and macroeconomic factors.

Author	**Title**	**Objective**	**Method**	**Finding**
Marketa Kacerová & Jan Černohorský	Unemployment as a Determinant of Credit Risk	To identify the relationship between credit risk and unemployment rate	Vector autoregression (VAR) model	Based on the analysis, the increase in unemployment has been found to have an increasing impact on nonperforming loans to households and nonfinancial corporations.

Hemlata Sharma, Aparna Andhalkar, Oluwaseun Ajao and Bayode Ogunleye	Analysing the Influence of Macroeconomic Factors on Credit Risk in the UK Banking Sector.	To assess the impact of macroeconomic factors on the UK banking credit risk	Machine learning	Based on result, the predictive model has achieved a 95% accuracy and has shown that unemployment and inflation rates are important indicators of credit risk in UK banking sector.

Table 4: Summary of credit risk caused by macroeconomic factors

2.6 Borrower characteristics affect credit risk

Assessing credit risk involves evaluating various borrower characteristics to determine the likelihood of default. Borrower characteristics significantly influence credit risk with income, employment stability and demographic factors playing key roles in determining default probabilities and loan performance.

A study from Sharma et al. (2024) has shown that the pivotal factors that influencing credit risk are income level and employment stability. Therefore, borrower with high income level is better able to repay their debt and meet other financial obligations. In addition, income stability is the consistency of earnings over time which income received in a fixed amount from the same sources on a regular basis without any fluctuation. A study by Kacerová & Černohorský (2023) has shown that individuals with high income level and stable income are less likely to loan default. In addition, the employment stability is the employee’s duration of stay with their current job without any interruption. Borrower who with stable employment history are more able to have a stable source of income and repay their debt on time, thereby lowering credit risk. Research by Fout et al. (2020) emphasize the importance of borrowers' ability to generate a consistent income stream to service debt obligations. High income and stable employment history lower default probability and show the borrower's capability to meet financial obligations.

Moreover, credit risk is influenced by demographic characteristics such as age, marital status and education levels. According to Tien & Nguyen (2023), younger borrowers are more likely to default due to lack of financial experience. In addition, financial stability of borrowers can be affected by their marital status and education level. Married borrowers in stable relationships have lower default rates compared to single, divorced or widowed individuals (Motedayen et al., 2022). Apart from that, a higher education level is associated with a greater financial literacy, higher earning potential and better career prospects (Tien & Nguyen, 2023). Therefore, borrowers with advanced degrees or specialized training will have lower default probabilities because of their enhanced financial knowledge and career opportunities to getting high payment job.

In order to reduce credit risk, it is crucial for banks to decide whether to approve or reject a loan application based on the borrower’s characteristics. Research about the loan prediction is conducted by Abhinav Jadhav to predict the loan approval based on borrower characteristics including ages, incomes, loans amounts, credit histories and other factors. Predictive model is developed in this research using machine learning such as logistic regression, decision trees and random forests. Figure 1 shows a small decision tree in random forest classifier. According to the result of the research paper, logistic regression model had the highest loan prediction accuracy of 80% while random forest had an accuracy of 78%.

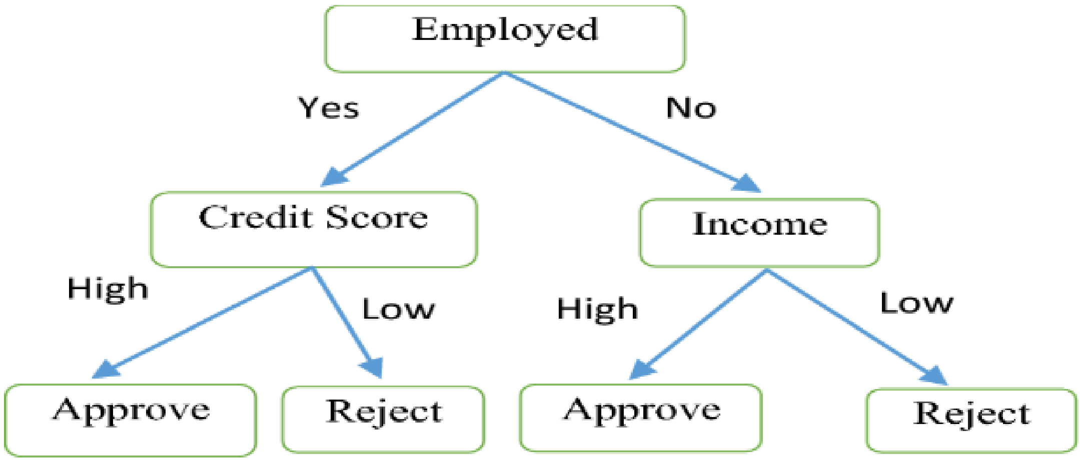

Figure 1: small decision tree in random forest classifier

Sources: LOAN APPROVAL PREDICTION USING MACHINE LEARNING (2023)

Table 5 shows the summary from previous research on how they analyse the credit risk level and probability of default based on borrower's characteristics.

Author	**Title**	**Objective**	**Method**	**Finding**
Hamilton Fout, Grace Li, Mark Palim, & Ying Pan	Credit risk of low income mortgages	To determine the rate of default by the income level of borrower	Statistical model	Defaults are more common among low and moderate income borrowers compared to those with high incomes. Based on result, lower income borrower have higher Debt-to-Income (DTI) compared to other income level, which associated with higher credit risk.
Nasrin Motedayen, Rafik Nazarian, Marjan Damankeshideh & Roya Seifipour	Identification and Evaluation of the Factors Affecting	-To identify the factors that affecting credit risk	Multinomial Logistic Regression	The results of the estimated model show that the indicators such as gender, age, loan

	Credit Risk Management Using the Multinomial Logistic Regression Model.	-To develop a predictive model using multinomial logistic regression approach.		value, repayment interval, previous loan, occupation, loan repayment term, loan extension, frequency installment amount, amount per installment, type of collateral, average balance and education level have a significant impact on the credit risk of real customers.
Abhinav Jadhav, Pranit Wankhande, Pradnya Balure, Abhishek Pawar & Dr.P.P.Halkarnikar	Loan Approval Prediction Using Machine Learning	To predict loan approval	Machine Learning	Logistic regression model has the highest accuracy compared to other machine learning model.

Table 5: Summary of credit risk affected by borrower characteristics

2.7 Loan Default Prediction using Machine Learning

Based on previous research papers, most researcher are using machine learning method to predict the loan default. Logistic Regression, Random Forests, Gaussian Naïve Bayes, Support Vector Machines are the most common machine learning approach that used in modeling.

For tasks of binary classification, involving loan default prediction, it is common practice to use Logistic Regression as a supervisory learning technique. According to Zhao & Zou (2021), the effectiveness of logistic regression models to evaluate default risks is demonstrated by analysing past data on loans and borrower characteristics. While logistic regression can provide interpretational results and simple model representation, its prediction performance may be limited with regard to nonlinear relationships or complex data patterns. According to the research paper, interest rates, loan amounts and borrower income are used to estimate the likelihood of default. The result of research paper has shown that interest rates is the predictor variable that best suited logistic regression due to its linear correlation with default.

Moreover, decision tree algorithms such as random forests have been widely used for predicting loan default because of their ability to deal with nonlinear relationships and interactions between prediction models. Random forests aggregate multiple decision trees to mitigate overfitting and improve generalization performance, making them well-suited for loan default prediction tasks. Research conducted by H. Li & Wu (2024) developed a predictive model using various machine learning techniques to predict loan defaults based on borrower characteristics. Figure 2 show the comparison of prediction performance of various machine learning models. The result from previous research has shown that random forest is the best model to predict the loan default compared to other machine learning.

Models	Accuracy (%)	Recall (%)	AUC	F1 score
KNN	83.19	74.38	0.85	0.52
SVM	55.83	51.22	0.56	0.22
Logistic	51.35	57.04	0.55	0.22
DT	84.77	87.41	0.85	0.59
RF	88.05	79.99	0.93	0.62
XGBoost	82.45	81.50	0.87	0.53
Stacking	86.76	82.80	0.93	0.61
GauNB	31.84	80.63	0.55	0.23
MLP	69.18	44.44	0.63	0.26

Figure 2: Performance of model using different machine learning method

Sources: Li & Wu (2024)

In addition, Support Vector Machine (SVM) is popular for loan default prediction. It can be used for classification as well as for regression. Research conducted by Aditya Sai Srinivas et al. (2022) to predict loan default using different machine learning techniques. One of the techniques used is Support Vector Machine (SVM). In this model, each data item is plotted as a unique point in an p-dimension and the process of classification is performed by finding the hyper-plane that differentiates the two classes. Other than SVM, other machine learning model also had been developed. The Figure 3 show the performance results of the research. From the result, it also shown that random forest has the highest accuracy compared to other models.

Algorithms	Accuracy	RMSE	Precision	Recall	AUC for ROC
LR	0.9211	0.1189	0.8922	0.8801	0.7913
RF	**0.9299**	**0.1051**	**0.9043**	**0.8929**	**0.8145**
SVM	0.8365	0.2811	0.5168	0.7189	0.5000
ANN	0.7297	0.6284	0.5973	0.6055	0.5031
Naïve B	0.9185	0.1244	0.8783	0.8757	0.8023
CART	0.8614	0.2051	0.7894	0.7944	0.7281

Figure 3: Result of model performance

Sources: Aditya Sai Srinivas et al. (2022)

Table 6 shows the summary of finding from previous researcher to predict loan default using machine learning.

Author	**Title**	**Objective**	**Method**	**Finding**
Selena Zhao & Jiying Zou	Predicting Loan Defaults Using Logistic Regression	To analyse the correlations between loan defaults and characteristics	Machine learning	Based on result, interest rate has the higher accuracy in predict the loan default

		of loan and borrower		
Huan Li & Weixing Wu	Loan default predictability with explainable machine learning	To compare machine learning models to predict the loan default risk	Machine learning	Random forest model outperformed other models with an accuracy of 88.05%.
Aditya Sai Srinivas T, Ramasubbareddy S, Govinda K	Lan Default Prediction Using Machine Learning Techniques	To compare machine learning models to predict the loan default risk	Machine learning	Random forest model outperformed other models with an accuracy of 92.99%.

Table 6: Summary of loan default prediction using machine learning

2.8 Conclusion

In conclusion, the literature review gives a comprehensive understanding of credit risk and its determining factors. Firstly, concept of credit risk refers to the possibility that borrowers or counterparties fail to repay their loan or satisfy contractual obligation and cause losses to the bank. Secondly, the main sources of credit risk, ranging from institutional capacity limitations to economic fluctuations, underscore the complexity and diversity of risk factors influencing lending operations. These resources necessitate sturdy threat control strategies to mitigate destructive influences on financial stability.

Moreover, there is a strong relationship between credit risk and borrower's credit history in which borrowers with better credit histories will having a lower probability of default. Furthermore, the dynamics of credit risk are significantly influenced by macroeconomic factors. These include fluctuations in economic growth, unemployment rates and interest rates, all of which have an impact on the probability of default. In addition, borrower characteristics such as income level, employment stability, and demographic factors considerably influence credit risk profiles and give lenders with information on how to evaluate risk by identify the suitable borrower based on their characteristics.

Furthermore, accuracy of loan default prediction can also be improved by using advanced machine learning techniques such as logistic regression and decision tree-based algorithms. Logistic regression is useful for modelling binary outcome by establishing a linear relationship between a predictor variable and a response variable. Besides that, decision tree-based algorithms can handle non-linear relationships and are able to split data into different branches, allowing for more detailed and nuanced risk assessments. Machine learning makes it possible to make better decisions by utilising large datasets and complex algorithms, ultimately helps the financial sector implement better risk mitigation techniques and reduce the probability of default.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter outlines an overview of research methodology. Starting from research design, follow by business understanding, analytic approach, data requirements, data collection, data understanding, data preparation, modeling, evaluation, deployment and lastly feedback. This chapter is important in choosing a method to investigate the research problem in order to obtain accurate and reliable results for chapter 4.

3.2 Research Design

Research design is the first step in developing and coordinating research activities. It is a "blueprint" which help researchers to answer their research question based on empirical data. This research involves identifying research objective, obtaining research goal, and developing research questions to analyse the factor affect credit risk. The first phase of the data science methodology is business understanding which focuses on gaining a clear understanding of credit risk problem in order to determine the factors that affect credit risk. Once understanding the problem of credit risk, an analytic approach needs to be defined to solve the problem. This might be a predictive model, descriptive approach, statistical analysis, or any other machine learning techniques. It is crucial to choose a proper analytic approach based on business understanding to solve the problem of credit risk effectively.

Moreover, the following stage is data requirements. During the process of data requirements, necessary data content, formats, and sources need to be identified for initial data collection based on analytic approach. In the data collection phase, relevant dataset is collected through related website, surveys, databases, APIs, or experiments in various format such as CSV files or Excel. Once data collection stage is complete, it is necessary to explore the collected data and understand its content. This involve understanding data structure, data types and data attributes.

In the data preparation stage, data is cleaned by handling for missing values, null values, or any bias data. Additionally, data also need to transform into correct format. It is crucial to ensure the data quality in data preparation stage for further use in modelling. The following phase is modeling which focuses on developing a predictive or descriptive model using machine learning. This model needs to be trained based on cleaned data until the model are able to understand the credit risk problem and achieve research goals.

Once the modeling is complete, model is evaluate based on its performance. It is crucial to check the effectiveness and quality of model and ensure it meets the business requirements. Besides, the following stage is deployment where the evaluated model needs to be deployed in real world environment and generate insight. Finally, the last stage for this methodology is feedback where the model is open to feedback from users or public. It is crucial to refine the model, accessing its impact and makes improvement if needed.

Through this data science methodology, it facilitates a deep investigation of the factors affecting credit risk and model development to solve the credit risk issue. Figure 4 show the flow chart of data science methodology.

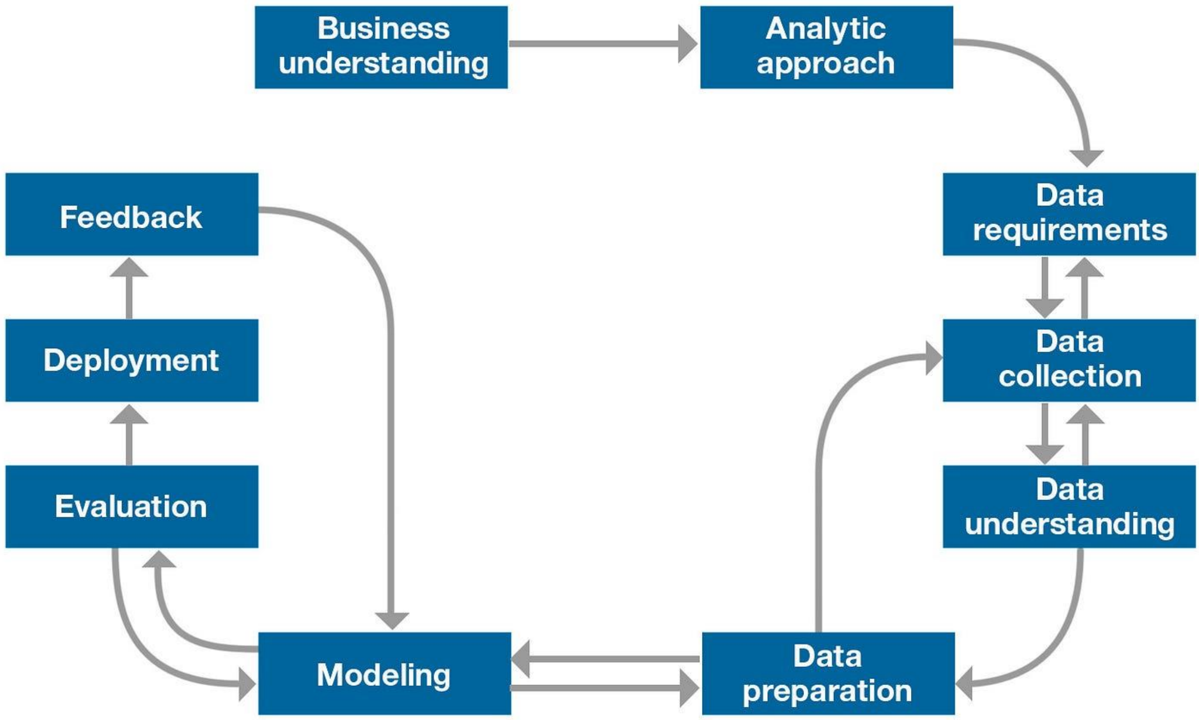

Figure 4: Data Science Methodology

Sources: Emeka Boris Ama (2019)

3.3 Business Understanding

In this research, our research scope is aims to investigate the factors affecting credit risk within the commercial banking sector of Malaysia. The question, goal and objective of project are listed to better understand the research.

Question: How to effectively identify the factor that affecting credit risk in Malaysia's commercial banks and how to develop a predictive model using machine learning to make prediction and manage the credit risk based on the borrower characteristics.

Goal: The goal of this study is to maximize borrower's loan repayment rates and reduce the credit risk within Malaysian commercial bank, eventually maintaining the financial stability of Malaysia. The goal might be achieved by creating a model that is used to calculate the probability of default based on the identified factor that affects credit risk.

Objectives: To identify the factor that affects credit risk and to develop an accurate and reliable model that can predict the credit risk using machine learning approach.

3.4 Analytic Approach

In this research, a descriptive approach is used to analyse the existing credit risk data and uncover the relationships and patterns, eventually driving meaningful insight based on the information given. Besides that, a predictive approach is also used to forecasts on the trends of credit risk. Based on data sources and related variables, a predictive model is built by using machine learning to predict the probability of default for the targeted borrower. This model is trained using large datasets, allowing for precise credit risk predictions based on borrower characteristics such as age, income level and marital status. With the use of predictive approaches, bankers or lenders can make more informed decisions on extending credit to borrowers based on predicted possibility of default on loan. It facilitates bankers to determine whether a targeted borrower is qualified to do a loan or not based on their characteristics and evaluate their ability to repay the loan. Overall, this method can reduce the probability of default, lower potential losses of bank and build a more effective credit risk management strategy to reduce credit risk.

3.5 Data Requirements

This research is targeted to be focus on examine the credit risk based on borrower's characteristics. Therefore, loan application data is needed in this research. Before banker make approval for their loan, they need to consider either they meet the requirement to apply for loan and predict the possibility of default based on borrower's characteristics such as age, gender, marital status, education status, income level, number of dependents, credit history and loan amount. By identify the main borrower characteristics that cause default, it helps to banker to improve their lending decisions and optimize their loan terms.

3.6 Data Collection

The dataset for developing predictive model using machine learning includes secondary data collecting technique. The data collected from 1000 Malaysian who wanted to apply for loans from several selected Malaysian commercial banks. Each applicant needs to provide their application data such as age, gender, marital status, education status, income level, number of dependents, credit history and loan amount. All the borrower's characteristics will be selected as the independent variable to be used to examine the probability of their default on loans. The dataset is collected from Kaggle. Figure 5 shows the existing dataset from data source Kaggle and the sample data will be considered in this research to predict the credit risk based on borrower's characteristics.

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y

Figure 5: Dataset from Kaggle

Sources: https://www.kaggle.com/datasets/krishnaraj30/finance-loan-approval-prediction-data

3.7 Data Understanding

In order to conduct the research, it is crucial to understand all the variables inside the dataset. Based on the existing dataset, there are total of 8 qualitative data and 6 quantitative data. Table 7 shows the description for each attribute in the existing dataset.

Attribute	Description
Loan_ID	Unique identifier for each applied loan
Gender	Gender of the loan applicant
Married	Marital Status of the loan applicant (Yes/No)
Dependents	Dependent counts of the loan applicant
Education	Education level of the loan applicant
Self-Employed	Self-employed (Yes/No)
ApplicantIncome	Income of the loan applicant
CoapplicantIncome	Coapplicant Income
LoanAmount	Loan amount applied in thousands
Loan_Amount_Term	Term of loan in months
Credit history	Credit history meets guidelines(0-No/1-Yes)
Property_Area	Urban/Semi-Urban/Rural

Table 7: Description of attributes

3.8 Data Preparation

Before developing a credit risk model, preparation of data is essential. The development of an accurate and robust credit risk model requires a high level of quality data. The goal of data preparation process is to clean and transform the raw data, making them more reliable to be used for analysis and modeling. During the cleaning process, the data is checked for missing values, null values, duplicate values, incorrect values, inconsistencies, outlier, and bias data. The missing values can be handled in two ways, which is deletion and imputation. When go with deletion approach, the entire row needs to delete if any missing cell exists within the row or delete the entire column when the percentage of null values is above 50%. If imputation method is used, the missing value can be replaced by mean, median or mode. However, it is better to use median compared to mean when meets normal distribution because mean is highly susceptible to outliers. Besides that, transformation is also used to transform the data into usable format. After the data is cleaned, it will be further used for training the model to predict the default possibility of borrower and assess credit risk.

3.9 Modelling

In this research, logistic regression and random forest classifier are used in developing a credit risk model and to predict the probability of loan default. These two machine learning approaches are chosen due to they are the most common method that used in previous research. This study aimed to learn about the performance and applicability of these machine learning models to the provided datasets.

3.9.1 Logistic Regression

Logistic regression is a supervised machine learning algorithm that is used for binary classification tasks. It predicts the categorical dependent variables based on the independent factors. Logical regression estimates the mathematical probability that a given applicant will default on their loan using sigmoid function and eventually the model delivers a binary outcome limited to two possible outcomes where a borrower default (1) or does not default (0). The equation of logistic regression is defined in Eq. (1) where $P(y = 1|\, x)$ is the probability of y being 1, x represents the input variable of the loan applicant and $\beta$is the coefficient of the logistic regression model. Logistic regression yields an S-shaped curve with values ranging between 0 and 1, seen in Figure 6.

$$P(y = 1|\, x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \qquad (1)$$

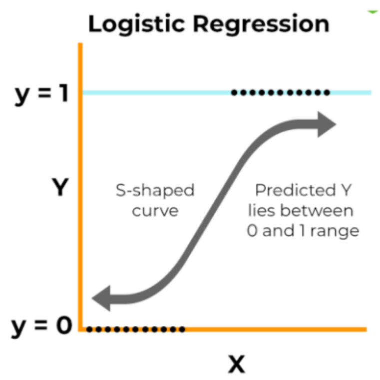

Figure 6: Logistic regression classifier

Sources: Vijay Kanade (2022)

3.9.2 Random Forest Classifier

Random Forest is a machine learning algorithm used for classification and regression tasks. It combines the output of multiple decision trees to reach a singular, more accurate result. This algorithm is able to handle complex datasets and mitigate overfitting. The data is split into training and testing sets. The training data is used to build the model while the testing data is used to evaluate its performance. Each decision tree is built using random subset of the features and a random subset of the training data. As the number of trees in the forest increases, the chance of overfitting decreases. The final prediction is made based on the majority vote across all trees. The design of the random forest classifier utilized in this study is shown in Figure 7.

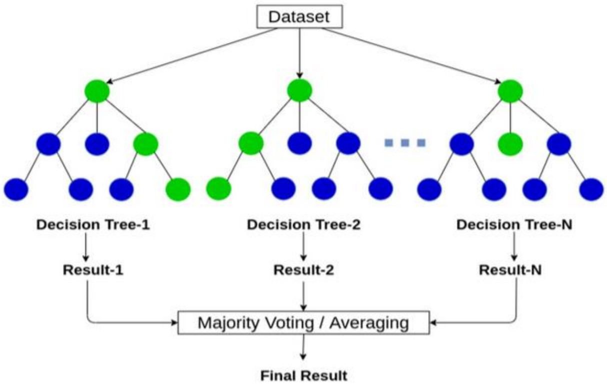

Figure 7: Random Forest Classifier architecture

Sources: Anas Brital (2021)

3.10 Evaluation

In this stage, various metrics are used to evaluate the model's performance. Classification metrics such as accuracy, precision, recall, F1 score, area under curve (AUC) and receiver operating characteristic (ROC) curve are commonly used to assess a model's capability to accurately categorise borrower that fails to repay the loan. An introduction to confusion matrix is shown in Figure 8 before proceeding to the introduction of the metrics. True positive (TP) indicates the positive class outcomes that is correctly classified by the model. False positive (FP) shows outcome where the model incorrectly classifies the negative data as positive. False negative (FN) is an outcome where the model incorrectly classifies the positive data as negative. True negative (TN) refers to an outcome where the model correctly classified the negative data.

Confusion matrix	**Positive (actual)**	**Negative (actual)**
Positive (predicted)	True positive (TP)	False positive (FP)
Negative (predicted)	False negative (FN)	True negative (TN)

Figure 8: Introduce to Confusion Matrix

Sources: Zhu et al. (2023)

Accuracy evaluation is used to measure how often the model correctly predicts the outcome. It is calculated by dividing the number of correct predictions by the total number of predictions, as shown in Eq. (2).

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \qquad (2)$$

Precision is the proportion of true positive predictions among all positive predictions as shown in Eq. (3) whereas recall refers to the proportion of true positive predictions that were identified by the model among all actual positive instances as shown in Eq. (4).

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

The F1 score is the harmonic mean of precision and recall. It combines precision and recall of model into a single metric, allowing for more precise evaluation of the model’s performance. The F1 score is expressed in Eq. (5).

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (5)$$

Figure 9 shows the AUC-ROC curve. Based on the diagram, ROC curve is plotted with True Positive Rate (TPR) against the False Positive Rate (FPR). True Positive Rate (TPR) and False Positive Rate (FPR) are defined in Eq. (6) and Eq. (7). AUC stands for the area under the ROC curve and it has a range value from 0 to 1. The closer the AUC to 1.0, the better the model is at prediction. The used of ROC-AUC curve is to measure the model’s ability to distinguish between people who have repay their loans on time or default on their loans.

$$TPR = \frac{TP}{TP + FN} \qquad (6)$$

$$FPR = \frac{FP}{FP + TN} \qquad (7)$$

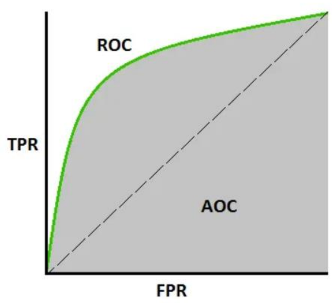

Figure 9: AUC-ROC Curve

3.11 Deployment

Once the predictive model achieves satisfactory results, it can be used in real-world settings to predict the probability of borrowers defaulting on loans. This involves implementing the predictive model in the banking industry's operational system to analyse historical data, identify patterns related to creditworthiness, help banks to enhance decision-making and eventually manage the credit risk exposure effectively.

The suggested technique used a combination of statistical and machine learning to evaluate factors associated with credit risk level. The main idea is to exploit the diversity and complementarity between different techniques to reduce variability and bias of model. Machine learning algorithms are trained on massive datasets in order to develop an effective prediction model and analyse integrated data for pattern identification. As a result, a user-friendly model is developed to provide valuable insights for credit risk management.

However, through validation in large-scale research with various populations, there are some improvements that need to be made to this model. Firstly, the model might incorporate more data sources and more precise assessments of credit risk to help lenders price credit offers and reduce charge-off. Besides that, it is also important to deploy the model in a real-time scoring environment. Adaptability of model in response to evolving economic environment and real-time monitoring of credit risk trends, banks or lenders are able to respond quickly to borrower's circumstances and their repayment capabilities and make timely credit decisions. Additionally, the model needs to be refined to be able to evaluate customers who do not have recent credit information or have sufficient credit files.

Finally, the suggested strategy can help to improve the accuracy and reliability of predictive model in order to effectively predict the possibility of default based on identified factor. The proposal intends to produce more accessible and POD-focused monitoring tool using machine learning and eventually lower credit risk within Malaysian commercial bank.

3.12 Feedback

In this stage, the information and suggestions are collected from stakeholders involved in the lending process, including borrowers, bankers, credit analysts and regulators. Raising borrower awareness of the development of machine learning towards loan default could assist the borrower to improve their loan repayment behavior. Borrowers with positive characteristics appreciate the model as it helps them easily apply for higher loan amounts, speed up the loan approval process and reduce the waiting time for access to funds.

Bankers discover the potential of machine learning to predict borrower's default probability based on their characteristics. Compared to traditional methods, bankers appreciate the ability of model to identify low risk and high-risk default borrowers, helping them to improve their decision making in lending process. Hence, reducing potential losses and increasing their profits.

Credit analysts express satisfaction with the ease of remote data collection which allows them to gather and review loan applicants' financial data including their income and credit history to assess the creditworthiness of loan applicant and their ability to repay the loan and recommend approval or denied a loan. Additionally, regulators appreciate that the model facilitates better monitoring and management of credit risk, ensuring financial stability.

3.13 Project Planning (Flow Chart and Gantt Chart)

Figure 10 shows the flow of project planning. It involves determining the problem statement and identifying the research objective, followed by literature review and data collection. Once the data was collected, it is used to train the model and evaluation is made based on performance of model. Finally, this project presents an expected result of credit risk analysis based on identified factors.

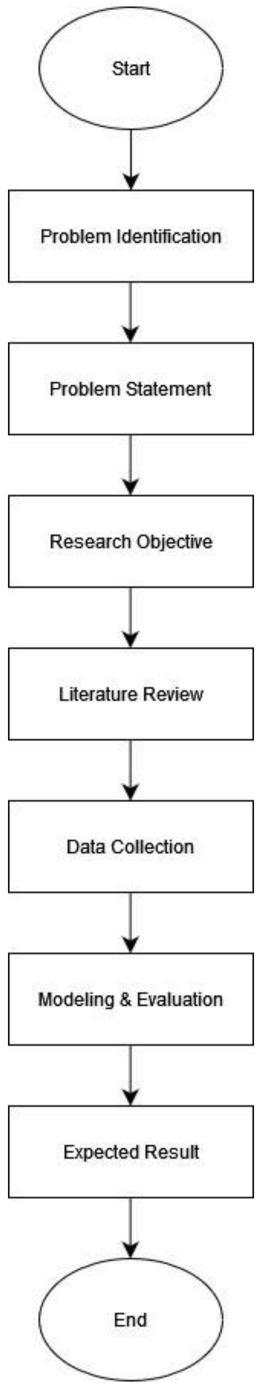

Figure 10: Project Planning Flow Chart

Figure 11 shows the project distribution by week in Gantt chart. It provides a timeline to show how the project was conducted from week 3 until week 12. The bottom of the chart shows the time frame in week while the left side of chart lists the project activities.

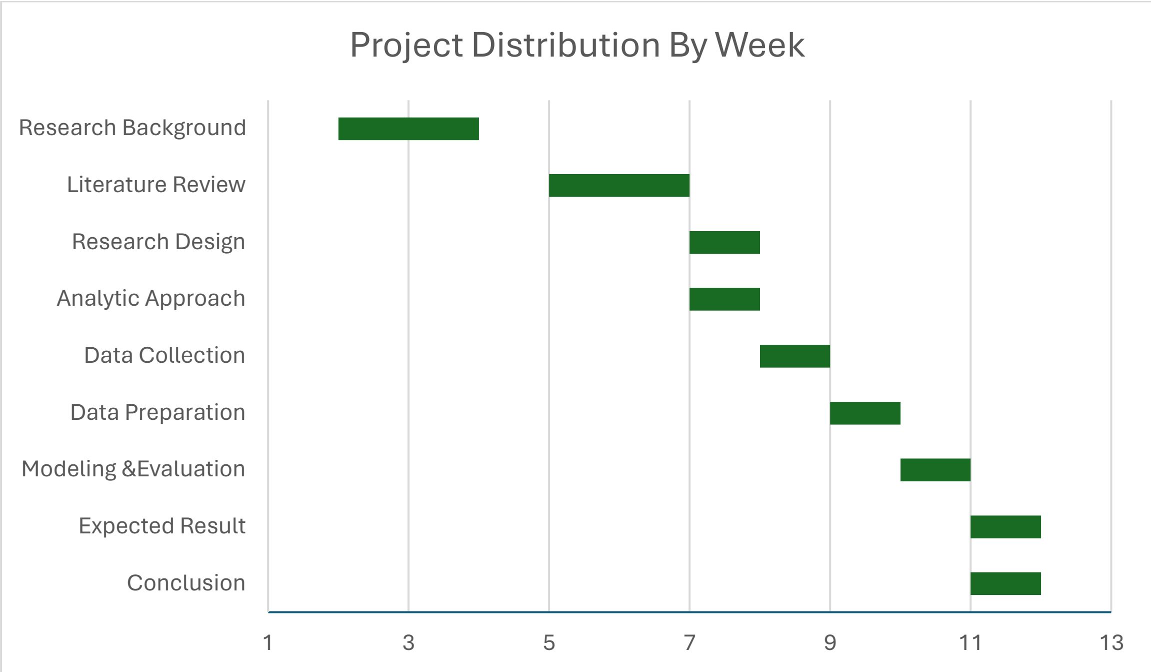

Figure 11: Project Distribution in Gantt Chart

CHAPTER 4

EXPECTED OUTCOME AND CONCLUSION

4.1 Introduction

This chapter aims to present the findings and analyses resulting from the acquired data and answering the objectives of the study. It contains two subchapters where section 4.2 will take a look at the expected result and emphasize their importance and potential impact while section 4.3 will summarize the study plan.

4.2 Expected Outcomes

Based on review from previous research papers, the factors affecting credit risk were identified. In this research, the expected factors that affects credit risk includes historical data, borrower's characteristics, and macroeconomic factors. Historical data such as borrower's credit history, financial statement, spending habit, spending amount on credit card and loan repayment history are important variables that need to be analysed in order to determine borrower's creditworthiness. Besides that, borrower's age, gender, marital status, number of dependent, employment status, income level also needs high attention as these characteristics showing their ability to repay the loan. Borrower with unstable financial status will has low ability to repay the loan. Hence, the probability of loan default is higher which may lead to the occurrence of credit risk. Moreover, credit risk also affected by macroeconomic factor such as GDP growth and downstream of economic. Since unemployment rate is higher during downstream of economic, therefore it will give impact on borrower's financial stability and ability to repay the loan.

Based on the identified factors, this research has decided to narrow down the research scope and focus more on the impact of borrower's characteristics on credit risk. Predictive models are developed using different machine learning including logistic regression and random forest. The expected outcome for the modeling is that both of the machine learning model can be used as alternative model to predict the probability of loan default.

Lastly, the performance of developed model also has been evaluated. According to previous research paper, it shows that the expected result of model using random forest will be more accurate compared to logistic regression. Therefore, random forest is expected to be the most effective machine learning approach in this research paper in order to determine the significant borrower's characteristics that are strongly associated with loan repayment or default and predict the probability of default.

4.3 Conclusion

Within Malaysian commercial banks, it is important to analyse the credit risk that cause by borrower's loan default to maintain financial stability. This study examined the factors that affect credit risk. According to the literature review, credit risk is caused by a variety of factors including historical data such as borrower's financial statement, credit history and repayment history, borrower's characteristics such as age, education level and income level and macroeconomic factors such as GDP growth. However, this study has narrowed down the scope to focus more on factor borrower's characteristics. This study suggests an innovative approach for identifying the default rate based on loan applicant's characteristics using machine learning. A predictive model is developed to determine borrower's creditworthiness and their ability of loan repayment and eventually allows for the early recognition of the default probability of a given borrower. The applicant who is in the default class has a low probability of money returning to the bank, therefore bankers can consider rejecting their loan application. In contrast, the applicant who is in the non-default class has a high probability of loan repayment, hence bankers can consider approving their loan application. Lastly, the predictive model can facilitate bankers to improve their decision making in order to reduce the default rate and overall decrease the credit risk.

REFERENCES

Abbas, F., Iqbal, S., & Aziz, B. (2019). The impact of bank capital, bank liquidity and credit risk on profitability in postcrisis period: A comparative study of US and Asia. *Cogent Economics and Finance*, *7*(1). https://doi.org/10.1080/23322039.2019.1605683

Aditya Sai Srinivas, T., Ramasubbareddy, S., & Govinda, K. (2022). Loan Default Prediction Using Machine Learning Techniques. *Lecture Notes in Networks and Systems*, *385*. https://doi.org/10.1007/978-981-16-8987-1_56

Anas Brital. (2021, September 20). *Random Forest Algorithm Explained* . https://anasbrital98.github.io/blog/2021/Random-Forest/

Bandyopadhyay, A., & Saxena, M. (2023). Interaction between credit risk, liquidity risk, and bank solvency performance: a panel study of Indian banks. *Indian Economic Review*, *58*(2). https://doi.org/10.1007/s41775-023-00202-y

Bank for International Settlements. (2006). Principles for the management of credit risk. *IFAS Extension*, *September*.

Berge, T. O., & Boye, K. G. (2007). An analysis of banks' problem loans. *Norges Bank. Economic Bulletin*, *78*(2).

Bhatt, T. K., Ahmed, N., Iqbal, M. B., & Ullah, M. (2023). Examining the Determinants of Credit Risk Management and Their Relationship with the Performance of Commercial Banks in Nepal. *Journal of Risk and Financial Management*, *16*(4). https://doi.org/10.3390/jrfm16040235

Bhilare, M., Wadajkar, V., & Kale, K. (2024). Empowering Financial Decisions: Precision Credit Scoring through Advanced Machine Learning. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(13s).

Boye, A., & Kwabena, M. (2014). Credit Risk Management in Financial Institutions: A Case Study of Ghana Commercial Bank Limited. *Research Journal of Finance and Accounting*, *5*(23).

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics*, *57*(1). https://doi.org/10.1007/s10614-020-10042-0

Ekinci, R., & Poyraz, G. (2019). The Effect of Credit Risk on Financial Performance of Deposit Banks in Turkey. *Procedia Computer Science*, *158*. https://doi.org/10.1016/j.procs.2019.09.139

Emeka Boris Ama. (2019, March 20). *Data Science Methodology*. Medium. https://medium.com/100daysofdscode/data-methodology-1fa727efbaf2

Fout, H., Li, G., Palim, M., & Pan, Y. (2020). Credit risk of low income mortgages. *Regional Science and Urban Economics*, *80*. https://doi.org/10.1016/j.regsciurbeco.2018.07.013

Giesecke, K. (2005). Credit Risk Modeling and Valuation: An Introduction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.479323

Goel, A., & Rastogi, S. (2023). Understanding the impact of borrowers' behavioural and psychological traits on credit default: review and conceptual model. In *Review of Behavioral Finance* (Vol. 15, Issue 2). https://doi.org/10.1108/RBF-03-2021-0051

Haryono, Y., Ariffin, N. Mohd., & Hamat, M. (2016). Factors Affecting Credit Risk in Indonesian Islamic Banks. *Journal of Islamic Finance*, *5*(1). https://doi.org/10.12816/0027649

Ismail, S., & Ahmed, E. (2023). The impact of liquidity risk, credit risk, and operational risk on financial stability in conventional banks in Jordan. *Uncertain Supply Chain Management*, *11*(2). https://doi.org/10.5267/j.uscm.2023.3.006

Kacerová, M., & Černohorský, J. (2023). Unemployment as a Determinant of Credit Risk. *Proceedings of the International Scientific Conference Hradec Economic Days 2023*, *13*. https://doi.org/10.36689/uhk/hed/2023-01-026

Kagan, J. (2023). *How Do Commercial Banks Work, and Why Do They Matter?* Investopedia.

Kusnandar, A. (2022). Credit Risk in Islamic Banking. *International Journal of Finance & Banking Studies (2147-4486)*, *11*(3). https://doi.org/10.20525/ijfbs.v11i3.1547

Li, H., & Wu, W. (2024). Loan default predictability with explainable machine learning. *Finance Research Letters*, *60*. https://doi.org/10.1016/j.frl.2023.104867

LOAN APPROVAL PREDICTION USING MACHINE LEARNING. (2023). *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets39658

Majani, S. I. (2022). The Relationship between Credit Risk Management and Financial Performance of Commercial Banks Listed at the Nairobi Securities Exchange, Kenya. *International Journal of Managerial Studies and Research*, *10*(5). https://doi.org/10.20431/2349-0349.1005008

Motedayen, N., Nazarian, R., Damankeshideh, M., & Seifipour, R. (2022). Identification and Evaluation of the Factors Affecting Credit Risk Management Using the Multinomial Logistic Regression Model. *Iranian Economic Review 2022*, *26(3)*.

Muñoz-Cancino, R., Bravo, C., Ríos, S. A., & Graña, M. (2023). On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance. *Expert Systems with Applications*, *218*. https://doi.org/10.1016/j.eswa.2023.119599

Naili, M., & Lahrichi, Y. (2022). Banks' credit risk, systematic determinants and specific factors: recent evidence from emerging markets. *Heliyon*, *8*(2). https://doi.org/10.1016/j.heliyon.2022.e08960

Pandey, C., & Joshi, B. K. (2023). Impact of Credit Risk Management on Profitability of Nepalese Commercial Banks. *The Harvest*, *2*(1). https://doi.org/10.3126/harvest.v2i1.54406

Rahman, A. A., & Shahimi, S. (2010). Credit risk and financing structure of Malaysian islamic banks. *Journal of Economic Cooperation and Development*, *31*(3).

Rehman, Z. U., Muhammad, N., Sarwar, B., & Raz, M. A. (2019). Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan. *Financial Innovation*, *5*(1). https://doi.org/10.1186/s40854-019-0159-8

Serwadda, I. (2018). Impact of credit risk management systems on the financial performance of commercial banks in Uganda. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *66*(6). https://doi.org/10.11118/actaun201866061627

Shah, A., Giri, S., & Shah, A. K. (n.d.). *Factors Influencing Loan Repayment in Microfinance Institutions in Bhaktapur District, Nepal*. *6*(4).

Sharma, H., Andhalkar, A., Ajao, O., & Ogunleye, B. (2024). Analysing the Influence of Macroeconomic Factors on Credit Risk in the UK Banking Sector. *Analytics*, *3*(1). https://doi.org/10.3390/analytics3010005

Tien, D. P., & Nguyen, K. Q. T. (2023). CREDIT RISK MANAGEMENT AND EFFECTIVENESS OF CREDIT ACTIVITIES: EVIDENCE FROM VIETNAM. *Journal of Eastern European and Central Asian Research*, *10*(5). https://doi.org/10.15549/jeecar.v10i5.989

Torban, T. K. (2020). Assessment of Credit Risk Management in Micro Finance Institutions: A Case of Adama Town MFIs, Ethiopia. *Srusti Management Review*, *13*(1).

Turjo, A. A., Rahman, Y., Karim, S. M. M., Biswas, T. H., Dewan, I., & Hossain, M. I. (2021). CRAM: A Credit Risk Assessment Model by Analyzing Different Machine Learning Algorithms. *ICOIACT 2021 - 4th International Conference on Information and Communications Technology: The Role of AI in Health and Social Revolution in Turbulence Era*. https://doi.org/10.1109/ICOIACT53268.2021.9563995

Twum, A. K., ZhongMing, T., Agyemang, A. O., Ayamba, E. C., & Chibsah, R. (2021). The impact of internal and external factors of credit risk on businesses: An empirical study of Chinese commercial banks. In *Journal of Corporate Accounting and Finance* (Vol. 32, Issue 1). https://doi.org/10.1002/jcaf.22482

Vijay Kanade. (2022, April 8). *What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices*. Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/

Wahyudi, S. M. (2020). The Influence of Credit Interest Rate and Credit Risk on Market Performance. *Asian Journal of Economics, Business and Accounting*. https://doi.org/10.9734/ajeba/2020/v17i430269

Zhao, S., & Zou, J. (2021). Predicting Loan Defaults Using Logistic Regression. *Journal of Student Research*, *10*(1). https://doi.org/10.47611/jsrhs.v10i1.1326

Zhu, X., Chu, Q., Song, X., Hu, P., & Peng, L. (2023). Explainable prediction of loan default based on machine learning models. *Data Science and Management*, *6*(3). https://doi.org/10.1016/j.dsm.2023.04.003

RUBRICS FOR RESEARCH PROPOSAL

CLO	Description	PLO mapping	Percentage	Marks
CLO2	Organise the research proposal by including the research elements accordingly.	PLO2: Able to develop analytical and critical thinking by utilizing Data Analytics knowledge in solving various problems. Keyword: Critical Thinking & Problem-Solving Skills C5: Synthesis	35%	70
CLO3	Explain effectively in written and oral form through research proposal presentation.	PLO5: Able to communicate ideas in appropriate forms, various mediums, and to a range of audiences in different situations effectively. Keyword: Communication A4: Organizing Values	5%	10
CLO4	Demonstrate of the ethical principles of research, ethical challenges and approval processes in research proposal and presentation.	PLO9: Able to show professional and ethical responsibility. Keyword: Ethics A3: Valuing	10%	20
TOTAL			50%	100

CLO2								
Item	**Weightage**	**Very Weak (0)**	**Weak (1)**	**Moderate (2)**	**Average (3)**	**Good (4)**	**Very Good (5)**	**Students Score**
Introduction Title and Research Background	**1**	The title and research background are poorly defined and inadequately contextualised. Little evidence exists that they are scientifically grounded. No originality. Not related to Data Science and Analytics knowledge.	Although the title and research background are defined, they are inadequately contextualised, with limited evidence that they are scientifically grounded. Little or no originality and not related to Data Science	The title and research background are defined. Some information and knowledge are scientifically grounded but lack originality and is not related to Data Science and Analytics knowledge.	The title and research background are adequately defined. Most of the information and knowledge are scientifically grounded. There is a high degree of originality and related to Data Science and Analytics knowledge.	The title and research background are original. They are defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	The title and research background are innovative and original. They are clearly defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	

			and Analytics knowledge.					
Problem Statement	**1**	The problem statement of the proposal is poorly defined and inadequately contextualised. Little evidence exists that they are scientifically grounded. No originality. Not related to Data Science and Analytics knowledge.	The problem statement of the proposal is defined, but they are inadequately contextualised, with limited evidence that they are scientifically grounded. Little or no originality and not related to Data Science and Analytics knowledge.	The problem statement of the proposal is defined. Some statements are scientifically grounded but lack originality and are not related to Data Science and Analytics knowledge.	The problem statement of the proposal is adequately defined. Most of the statements are scientifically grounded. There is a high degree of originality and related to Data Science and Analytics knowledge.	The problem statement of the proposal is original. They are defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	The problem statement of the proposal is innovative and original. They are clearly defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	
Research Questions and Research Objectives	**1**	The research question and objectives of the proposal are poorly defined and inadequately contextualised. Little evidence exists that they are scientifically grounded. No originality. Not related to Data Science and Analytics knowledge.	The research question and objectives of the proposal are defined, but they are inadequately contextualised, with limited evidence that they are scientifically grounded. Little or no originality and not related to Data Science and Analytics knowledge.	The research question and objectives of the proposal are defined. Some objectives are scientifically grounded but lack originality and are not related to Data Science and Analytics knowledge.	The research question and objectives of the proposal are adequately defined. Most of the objectives are scientifically grounded. There is a high degree of originality and related to Data Science and Analytics knowledge.	The research question and objectives of the proposal are original. They are defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	The research question and objectives of the proposal are innovative and original. They are clearly defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	

Research Scopes and Significance of Study	**1**	The research scopes and significance of study of the proposal are poorly defined and inadequately contextualised. Little evidence exists that they are scientifically grounded. No originality. Not related to Data Science and Analytics knowledge.	The research scopes and significance of study of the proposal are defined, but they are inadequately contextualised, with limited evidence that they are scientifically grounded. Little or no originality and not related to Data Science and Analytics knowledge.	The research scopes and significance of study are defined. Some information and knowledge are scientifically grounded but lack originality and is not related to Data Science and Analytics knowledge.	The research scopes and significance of study of the proposal are adequately defined. Most of the information and knowledge are scientifically grounded. There is a high degree of originality and related to Data Science and Analytics knowledge.	The research scopes and significance of study of the proposal are original. They are defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	The research scopes and significance of study of the proposal are innovative and original. They are clearly defined, contextualised, scientifically grounded and related to Data Science and Analytics knowledge.	
Literature Review	**2**	Inadequate knowledge, interpretation, and application of literature. Use of irrelevant literature.	A basic overview of the literature, with limited interpretation and application.	Adequate knowledge of the relevant literature. Minor shortcomings in the interpretation and application of the literature.	Adequate knowledge, interpretation, and application of the relevant literature.	Authoritative knowledge, coverage, interpretation, and application of the relevant literature.	Excellent knowledge, coverage, interpretation, and application of the relevant literature.	
Methodology: Appropriateness of research design and Data Science Methodology	**3**	The work displays such a low level of research design and data science methodology, unacceptable and do not have any Data Science and	Satisfactory understanding of the research design and data science methodology, reasonably, but without the Data Science and Analytics	Adequate knowledge of the research design, data science methodology and the Data Science and Analytics skills.	Demonstrates average research design, data science methodology and the Data Science and Analytics skills.	Demonstrates good research design, data science methodology and the Data Science and Analytics skills.	Demonstrates advanced research design, data science methodology and the Data Science and Analytics skills.	

		Analytics skills.	skills.					
Methodology: Appropriateness of project planning (Flow Chart & Gantt Chart)	**2**	Poorly demonstrate an appropriate and relevant understanding on project planning by presenting relevant flow chart and Gantt chart.	Demonstrate a satisfactory understanding on project planning by presenting relevant flow chart and Gantt chart.	Demonstrate an adequate understanding on project planning by presenting relevant flow chart and Gantt chart.	Demonstrate an appropriate and relevant understanding on project planning by presenting relevant flow chart and Gantt chart.	Demonstrate a good understanding on project planning by presenting relevant flow chart and Gantt chart.	Demonstrate a very good understanding on project planning by presenting good flow chart and Gantt chart.	
Expected Outcome	**1**	No logical or valid expected outcome reached.	The expected outcome is not in all respects logical and valid.	The expected outcome is partially logical and/or valid.	The expected outcome is logical and valid.	The expected outcome is logical and valid and show an awareness of the published literature.	The expected outcome is logical and valid and show a strong awareness of the authoritative published literature.	
Conclusions	**1**	No logical or valid conclusions are reached. The final summary in no way communicates the purpose and findings of the study, and the use of terminology is confusing.	The conclusions are not in all respects logical and valid. The final summary does not communicate the purpose and findings of the study.	The conclusions are partially logical and/or valid. They communicate clearly. The final summary only partially communicates the purpose and findings of the study.	The conclusions are logical and valid. The conclusions are communicated clearly and, where applicable, they are linked/related to existing beliefs. The final summary communicates the purpose and findings of the	The conclusions are logical and valid and show an awareness of the published literature. The conclusions are clearly communicated and evaluated and, where applicable, they are linked/related to existing	The conclusions are logical and valid and show a strong awareness of the authoritative published literature. The conclusions are clearly communicated and evaluated and, where applicable, they refute the	

					study.	beliefs. The final summary is relevant and communicates the purpose and findings of the study.	existing beliefs in the field. The final summary is relevant and communicates the purpose and findings of the study in clearly understandable terms.	
References	**1**	References are of poor quality or have been omitted. The format and layout are inconsistent and not in line with standard referencing techniques. Neither important nor the most recent sources appear in the bibliography.	Referencing is not in line with the standard conventions in terms of format and layout. Numerous important as well as relevant sources have been omitted.	The referencing has been done, but the layout of the bibliography is neither consistent nor in line with internationally acceptable conventions. Some important and very relevant sources have been omitted.	The referencing has been done in a proper manner and the layout of the bibliography is largely in line with internationally acceptable conventions. There are, however, a number of inconsistencies. The bibliography includes most of the important sources.	The referencing has been done in a proper manner and the layout of the bibliography is largely in line with internationally acceptable conventions. The bibliography includes the most important sources.	The referencing has been done in a proper and extensive manner. The format and layout of the bibliography are correct and in line with internationally acceptable conventions. The bibliography contains the most important and most recent sources.	
								/70

CLO3								
Item	**Weightage**	**Very Weak (0)**	**Weak (1)**	**Moderate (2)**	**Average (3)**	**Good (4)**	**Very Good (5)**	**Students Score**
Writing Style	**1**	There are serious, conspicuous, and unacceptable linguistic and typographical errors.	The organised writing style of the proposal requires serious attention, considering the numerous linguistic and typographical mistakes. These errors should definitely be rectified.	The organised writing style of the proposal requires attention. There are omissions and linguistic and/or typographical errors. Editing/revision would improve the work and errors should definitely be rectified.	The organised writing style of the proposal is of acceptable quality. There are less important omissions and linguistic and/or typographical errors. Editing would improve the text.	The organised writing style of the proposal is good quality. There are few linguistic and typographical errors, and few linguistic and/or typographical rectifications are required.	The organised writing style of the proposal is of very high quality. There are no or extremely few linguistic and typographical errors, and almost no rectifications are required.	
Writing Layout	**1**	Organise the writing style and layout are plagued by serious problems and should be reviewed.	The organized writing layout of the proposal requires serious attention, this errors should definitely be rectified.	The organized writing layout of the proposal requires attention. Editing/revision would improve the work and errors should be rectified.	The organized writing layout of the proposal is of acceptable quality. Editing would improve the text.	The organized writing layout of the proposal are of good quality.	The organized writing layout of the proposal is very high quality.	
								/10

CLO4								
Item	**Weightage**	**Very Weak (0)**	**Weak (1)**	**Moderate (2)**	**Average (3)**	**Good (4)**	**Very Good (5)**	**Students Score**
Work Responsibility	**1**	Does not perform assigned tasks within the scope of work and do not meet expectations.	Does not perform assigned tasks within the scope of work even with close supervision.	Perform assigned tasks within the scope of work with close supervision.	Perform assigned tasks within the scope of work and meets expectation.	Perform assigned tasks within the scope of work and exceeds expectation.	Perform assigned tasks beyond the scope of work and beyond expectation.	
Work Ethics	**1**	Practise very inappropriate working culture such as bad behaviour, no punctuality as well as not being efficient, productive, and ethical at work in all situations.	Practise inappropriate working culture such as bad behaviour, no punctuality as well as not being efficient, productive, and ethical at work in all situations.	Practise less appropriate working culture such as inconsistent behaviour, less punctuality as well as being less efficient, productive, and ethical at work in many situations.	Practise good working culture such as good moral, timeliness as well as being efficient, productive, and ethical at work in general.	Practise good working culture such as good moral, timeliness as well as being efficient, productive, and ethical at work in most situations.	Always practise excellent working culture such as good moral, timeliness as well as being efficient, productive, and ethical at work in all situations.	
Plagiarism	**1**	Score more than 40% in Turnitin.	Score within 36%-40% (inclusive) in Turnitin.	Score within 31%-35% (inclusive) in Turnitin.	Score 26%-30% (inclusive) in Turnitin.	Score within 21%-25% (inclusive) in Turnitin.	Score less than 20% in Turnitin.	
AI Plagiarism	**1**	Score more than 40% in Turnitin.	Score within 36%-40% (inclusive) in Turnitin.	Score within 31%-35% (inclusive) in Turnitin.	Score 26%-30% (inclusive) in Turnitin.	Score within 21%-25% (inclusive) in Turnitin.	Score less than 20% in Turnitin.	
								/20
							GRAND TOTAL	**/100**

RUBRICS FOR RESEARCH PROPOSAL PRESENTATION

CLO	Description	PLO mapping	Percentage	Marks
CLO3	Explain effectively in written and oral form through project proposal presentation	PLO5: Able to communicate ideas in appropriate forms, various mediums, and to a range of audiences in different situations effectively. Keyword: Communication A4: Organizing Values	20%	40

Item	Weightage	Very Weak (0)	Weak (1)	Moderate (2)	Average (3)	Good (4)	Very Good (5)	Students Score
Introduction: Topic and purpose of communication are introduced in the presentation slide with appropriate attention getter	**1**	Topic is not introduced. Purpose is not stated. There is no attention getter.	The topic and purpose of communication are stated but not clear in the presentation slide. There is no attention getter.	The topic and purpose of communication are stated but not clear in the presentation slide. Attention getter is fairly acceptable.	The topic and purpose of communication are made clear in the presentation slide. Attention getter is acceptable.	The topic and purpose of communication are made clear in the presentation slide. Attention getter is good.	The topic and purpose of communication are stated clearly in the presentation slide. Attention getter is excellent and leaves an impact on the audience.	
Body: Provide general and background information which correspond with the purpose. Show understanding of the topic	**1**	General and background information are not provided. Do not show understanding of the topic.	General and background information are inadequate and do not correspond with the purpose. Show poor understanding of the topic.	General and background information are slightly inadequate and fairly correspond with the purpose. Show fair understanding of the topic.	General and background information are adequate, but moderately correspond with the topic. Show average understanding of the topic.	General and background information are adequate and correspond with the topic. Show good understanding of the topic.	General and background information are adequate and correspond with the topic. Show excellent understanding of the topic.	

Body: Provide adequate, relevant and significant information of the topic (Include Literature, Methodology & Significance of the Research); Provide links between ideas	**1**	Information is inadequate, irrelevant, and insignificant.	The information is inadequate. The relevance and significance of information is poorly made. Links and connections between ideas is poorly made.	Information is present, but supporting details are inadequate. The relevance and significance of information is fairly made. Links and connections between ideas is fairly made.	Information is adequate. The relevance and significance of information is moderately made. Links and connections between ideas is moderately made.	Information is adequate. The relevance and significance of information is good. Links and connections between ideas is good.	Information is adequate. The relevance and significance of information is excellent. Links and connections between ideas is excellent.	
Closing: Summary of the topic; Memorable and impactful	**1**	There is no closing, and no summary made.	Closing highlights a few ideas and no final statement.	Closing contains some key ideas and concludes with a fair final statement.	Closing contains some key ideas and concludes with an acceptable final statement.	Closing highlights most key ideas and concludes with a good final statement.	Closing highlights all key ideas and concludes with a strong and impactful final statement.	
Appropriate language and words expression according to the topic and level of communication	**1**	Language expression used is not appropriate.	Uses inappropriate language expression to the topic and level of communication. Grammatical structure is poor.	Uses fairly appropriate language expression to the topic and level of communication. Grammatical structure is fair.	Uses moderately appropriate language expression to the topic and level of communication. Grammatical structure is moderately correct.	Uses appropriate and good language expression to the topic and level of communication. Grammatical structure is good.	Uses appropriate and excellent language expression to the topic and level of communication. Grammatical structure is excellent.	
Appropriate organisation of information in oral communication (Coherence and Cohesion)	1	There is no organisation in the communication. No coherence and cohesion of ideas.	Organisation of the communication, coherence and cohesion are poor.	Organisation of the communication, coherence and cohesion are fair.	Organisation of the communication, coherence and cohesion are moderate.	Organisation of the communication, coherence and cohesion are good.	Organisation of the communication, coherence and cohesion are excellent.	

Understand and respond to questions	**1**	Not able to understand and respond to questions.	Able to understand questions but not able to answer all questions.	Able to understand and answer some questions but not able to accurately answer the questions.	Able to understand and answer some questions satisfactorily.	Able to respond to all questions well.	Able to fully understand and respond to all questions very well.	
Adapt delivery to audience level	**1**	Not able to deliver the ideas.	Not able to deliver appropriately to the audience level.	Able to deliver ideas with limited appropriateness to the target audience and require further improvements.	Able to deliver ideas appropriately to the target audience satisfactorily.	Able to deliver ideas appropriately to the target audience well.	Able to fully deliver ideas appropriately very well.	
							TOTAL	**/40**