

SDBI
Post Graduate Diploma
Data Science & Business Analytics

PROJECT
CREDIT RISK MODELLING

Team Members:

- 1. Pratik Doshi - 15**
- 2. Jinay Gada - 17**
- 3. Swara Vedak - 10**
- 4. Vineet Jadhav - 13**
- 5. Kareena Zaveri - 19**

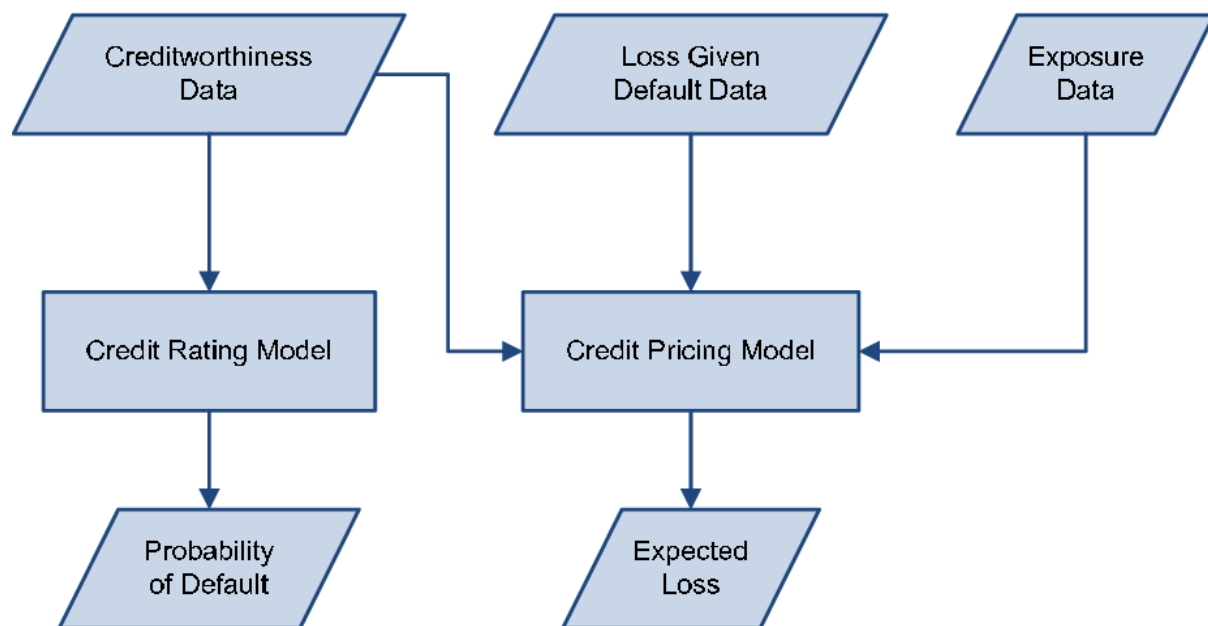
Credit Risk Modelling Table Of Content

- 1. INTRODUCTION**
- 2. PROBLEM STATEMENT**
- 3. OBJECTIVES**
- 4. SURVEY OF TECHNOLOGY**
- 5. COMPONENTS OF CREDIT RISK**
- 6. MACHINE LEARNING MODELS**
- 7. LIMITATIONS**
- 8. FRAMEWORK FOR CREDIT RISK ANALYSIS**
- 9. CREDIT RISK MANAGEMENT IN BANKS**
- 10. CREDIT RISK PRACTICES**
- 11. PUBLIC PRACTICES**
- 12. HOW TO ACHIEVE CREDIT CONTROL**
- 13. RISKS FOR CUSTOMERS**
- 14. ELEMENTS FOR CREDIT RISK MODELLING**
- 15. HOW TO RESTRICT BAD CREDIT**
- 16. OPPORTUNITIES IN CREDIT RISK MODELLING**
- 17. DATASET**
- 18. REPORT**
- 19. CONCLUSION**

ABSTRACT

Credit risk modeling is a crucial process used by financial institutions to assess the likelihood of borrowers defaulting on their loans or credit products. This involves the use of statistical methods and financial models to analyze a wide range of data related to the borrower, including their credit history, income, and other financial information. By quantifying the risk of default, credit risk modeling helps financial institutions to make informed lending decisions and to minimize potential losses. This abstract provides an overview of credit risk modeling and highlights its importance in the financial industry.

Keywords- credit risk modeling, default risk, statistical methods, financial models, borrower creditworthiness, expected losses, financial institutions, lending decisions, data analysis.



1. Introduction

Credit risk modeling is a vital process for financial institutions to assess the risk of default on loans and credit products. This modeling involves analyzing a borrower's credit history, financial data, and other factors to determine the likelihood of repayment and potential losses in case of default. Credit risk modeling helps financial institutions make informed lending decisions by providing insight into a borrower's creditworthiness and their ability to repay a loan.

There are several types of credit risk modeling, including quantitative and qualitative models. Quantitative models use statistical methods and financial models to analyze data and calculate the probability of default and expected losses. Qualitative models use expert judgment and subjective assessments to evaluate a borrower's creditworthiness and the risk of default.

The use of credit risk modeling offers several benefits to financial institutions. It provides a systematic and objective approach to assessing risk, improves credit risk management, and helps reduce losses from default. By understanding the risk of default and the potential losses associated with it, financial institutions can make informed decisions about lending, pricing, and risk management.

Overall, credit risk modeling is an essential tool for financial institutions in managing their credit risk and making informed lending decisions. It provides a framework for assessing the risk of default and helps institutions to make better credit risk management decisions. As such, it plays a crucial role in the efficient functioning of financial markets and the broader economy.

Keywords- credit risk modeling, default risk, quantitative models, qualitative models, statistical methods, financial models, credit risk management, lending decisions, risk management, financial institutions, creditworthiness, potential losses, objective approach.

2. Problem Statement

Credit risk modeling is a vital tool used by financial institutions to identify and assess the risk of default on loans and other credit products. The problem statement for credit risk modeling is to develop a predictive model that can accurately assess the creditworthiness of borrowers, estimate the likelihood of default, and quantify potential losses in the event of default. This model must be developed and implemented in a way that is consistent with regulatory requirements and industry best practices, while also aligning with the institution's risk management objectives and business goals. The model should be regularly monitored and refined to ensure that it continues to accurately assess credit risk and provide actionable insights to support sound credit decisions.

3. Objectives

The objectives of credit risk modeling are to enable financial institutions to make informed credit decisions and manage credit risk effectively. The primary objective is to assess the likelihood of default by a borrower and quantify the potential losses associated with that default. To achieve this objective, the credit risk model should consider a range of factors that influence credit risk, such as the borrower's credit history, income, and other financial information.

Another objective of credit risk modeling is to develop a framework for managing credit risk across the institution's portfolio of loans and credit products. This involves developing risk management policies and procedures that are consistent with regulatory requirements and industry best practices, as well as implementing appropriate risk monitoring and reporting systems. The credit risk model should also enable the institution to identify and manage concentrations of credit risk across its portfolio, such as exposure to particular industries or geographic regions.

In addition to managing credit risk, credit risk modeling can also support pricing decisions and inform business strategy. By assessing the credit risk of different borrowers and credit products, financial institutions can set appropriate pricing to balance risk and return. Credit risk modeling can also help institutions to identify opportunities for growth and expansion while managing the associated credit risk. Overall, the objectives of credit risk modeling are to enable financial institutions to make informed credit decisions, manage credit risk effectively, and support the institution's business goals and strategy.

4. Survey of Technology

A survey of technology for credit risk modeling projects would involve an assessment of the various tools and platforms available for developing and implementing credit risk models. This survey could include both proprietary and open-source software, as well as cloud-based solutions and data analytics platforms.

Some of the key technologies that are commonly used for credit risk modeling include statistical modeling languages such as R and Python, as well as machine learning libraries and frameworks like scikit-learn, TensorFlow, and Keras. These tools enable data scientists and analysts to develop and train predictive models using various algorithms, including decision trees, random forests, logistic regression, and neural networks.

Other technologies that may be used in credit risk modeling projects include data visualization and reporting tools, such as Tableau and Power BI, as well as database management systems like MySQL and Oracle. Cloud-based platforms like Amazon Web Services and Microsoft Azure can also be leveraged to support the storage, processing, and analysis of large volumes of data.

The choice of technology for a credit risk modeling project will depend on a range of factors, including the complexity of the models, the size and structure of the data, and the specific requirements of the institution. It is important to carefully evaluate the available technologies and select those that are best suited to the needs of the project.

5. COMPONENTS OF CREDIT RISK:

The components for credit risk - Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD) typically include:

1. **Probability of Default (PD):** The PD component of credit risk refers to the likelihood of a borrower defaulting on their loan or credit obligation within a specific period. The PD is typically calculated using statistical models that consider factors such as the borrower's credit history, financial performance, and industry trends.
2. **Loss Given Default (LGD):** LGD refers to the expected loss that a lender will incur in the event of borrower default. This component of credit risk takes into account the collateral, if any, that is held against the loan or credit product, and the amount that can be recovered through liquidation or sale of the collateral.

3. Exposure at Default (EAD): EAD refers to the total exposure that a lender has to a borrower at the time of default. This component of credit risk considers the outstanding balance of the loan or credit product, as well as any additional exposure that may be associated with the borrower's creditworthiness.

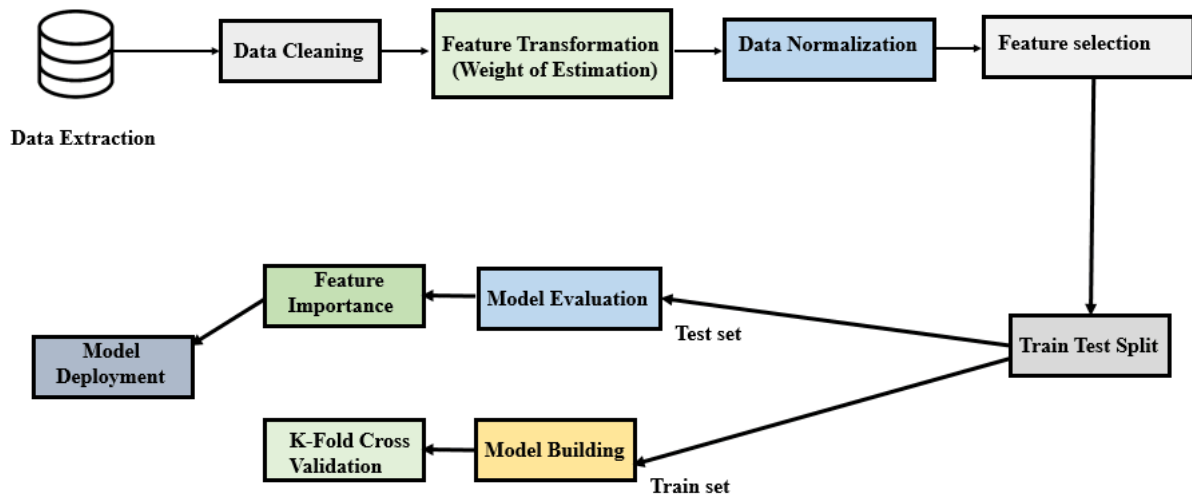


6. Machine Learning Models:

There are several machine learning models that can be used for credit risk modeling, including:

1. Logistic Regression: A popular model used for binary classification problems such as predicting whether a borrower will default or not. It is easy to interpret and can handle both categorical and numerical variables.
2. Decision tree: is a popular machine learning model used in credit risk modeling. A decision tree is a tree-like model where each node represents a feature or attribute, and the edges represent the decision rules or conditions. The leaves of the tree represent the final decision or classification.

These are just a few examples of machine learning models that can be used for credit risk modeling. The choice of model depends on the specific requirements of the problem and the nature of the data being analyzed.



1. Logistic Regression:

Logistic regression is a commonly used machine learning model in credit risk modeling. It is a statistical method used for binary classification, where the goal is to predict the probability of a binary outcome, such as whether a borrower will default or not.

In credit risk modeling, logistic regression can be used to predict the probability of default based on various borrower characteristics such as credit history, income, and debt-to-income ratio. Logistic regression can handle both categorical and numerical variables and is useful for identifying the factors that contribute to the likelihood of default.

One advantage of logistic regression is that it provides interpretable results, allowing lenders to understand the factors that are driving the prediction. Logistic regression models can also be updated with new data, making them flexible and adaptable to changing market conditions.

However, logistic regression assumes that the relationship between the predictor variables and the outcome variable is linear, which may not be the case in some credit risk modeling scenarios. Additionally, logistic regression may not perform well when there are interactions between the predictor variables. In such cases, more complex machine learning models like random forests or gradient boosting may be more appropriate.

2. Decision Tree:

The decision tree is a popular machine learning model used in credit risk modeling. A decision tree is a tree-like model where each node represents a feature or attribute, and the edges represent the decision rules or conditions. The leaves of the tree represent the final decision or classification.

In credit risk modeling, decision trees can be used to predict the probability of default by analyzing various factors that influence creditworthiness, such as credit history, income, and debt-to-income ratio. Decision trees can handle both categorical and numerical variables and are especially useful for identifying non-linear relationships between variables.

Decision trees can also be used to explain the decision-making process, making it easier for lenders to understand why a particular decision was made. Moreover, decision trees are computationally efficient and can handle large datasets, which is useful in credit risk modeling where large amounts of data are often analyzed.

However, decision trees can be prone to overfitting, which occurs when the model is too complex and fits the training data too closely, leading to poor performance on new data. This can be mitigated by using techniques such as pruning and ensemble methods like random forests, which use multiple decision trees to make a more accurate prediction.

7. LIMITATIONS:

Credit risk modeling has several limitations that can affect its accuracy and usefulness:

1. **Limited Data Availability:** Credit risk modeling relies on historical data to predict the likelihood of default, but the quality and quantity of available data can be limited. This can result in biased or inaccurate models, especially when there are significant changes in the economic or regulatory environment.
2. **Data Quality Issues:** The quality of credit risk data can also be a challenge, as there may be errors or inconsistencies in the data that can affect the accuracy of the model.
3. **Lack of Transparency:** Credit risk models can be complex and difficult to interpret, making it challenging for lenders to understand how the model is making its predictions.
4. **Unforeseen Events:** Credit risk modeling is based on historical data, but unforeseen events such as natural disasters, political events, or pandemics can disrupt the economy and financial markets, leading to unexpected default rates.

5. Regulatory and Legal Issues: Credit risk modeling is subject to regulatory and legal requirements, which can limit the use of certain data or techniques and require lenders to explain their decisions.

6. Model Drift: Credit risk models can become outdated or inaccurate over time, especially if the underlying economic or market conditions change.

8. FRAMEWORK FOR CREDIT RISK ANALYSIS

A framework for credit risk management in banks typically involves the following components: Risk identification and assessment: This involves identifying and assessing credit risks associated with various lending activities, such as loans, credit lines, and other financial products. The bank must identify the potential risks associated with each borrower, such as their credit history, income, and other relevant factors.

1. Risk measurement and monitoring: The bank must quantify the risks associated with each lending activity and develop metrics to monitor and measure the level of risk exposure. This includes measuring credit risk by evaluating the probability of default and potential loss given default.
2. Risk mitigation: Once the risks have been identified and measured, the bank must implement appropriate risk mitigation

strategies, such as setting appropriate credit limits, collateral requirements, and loan pricing.

3. Portfolio management: The bank must manage its portfolio of loans and other credit products to ensure that it is diversified and appropriately balanced. This involves regularly assessing the performance of the portfolio and making necessary adjustments to maintain an optimal balance of risk and return.
4. Regulatory compliance: Banks must comply with regulatory requirements related to credit risk management, including maintaining appropriate capital reserves and reporting requirements.

Overall, a comprehensive credit risk management framework helps banks to mitigate risk, maintain a healthy portfolio, and comply with regulatory requirements. It is a crucial component of a bank's overall risk management strategy and ensures the institution's long-term financial health and stability.

9. Credit Risk Management In Banks

Credit risk is a crucial factor in the banking industry. Banks offer various credit products to customers, such as loans and credit cards, which expose them to the risk of default. Credit risk management in banks involves identifying, assessing, and controlling the risk of losses resulting from customer defaults on credit obligations.

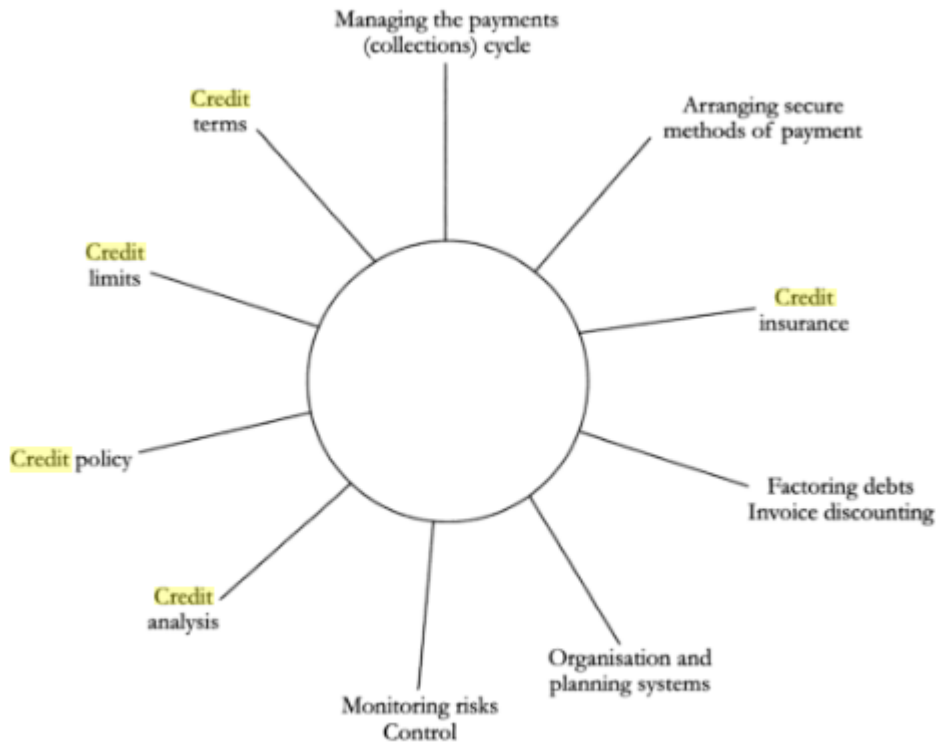
The credit risk management process begins with the evaluation of a customer's creditworthiness. Banks use credit scoring models to assess the likelihood of a borrower defaulting on a loan or credit product. These models take into account factors such as credit history, income, debt-to-income ratio, and other relevant financial information.

Once a borrower's creditworthiness is assessed, banks must decide on the appropriate credit limit and interest rate for the loan or credit product. This decision is based on the risk of default, the bank's own risk appetite, and market conditions.

After a loan or credit product is issued, banks must monitor the borrower's creditworthiness and payment history. This helps them identify early warning signs of default and take appropriate action to mitigate the risk of losses.

In addition to individual credit risk management, banks also engage in portfolio risk management. This involves assessing the credit risk of the entire portfolio of loans and credit products issued by the bank. Portfolio risk management helps banks balance the risk and return of their credit products and ensure that they have adequate capital to cover potential losses.

Overall, credit risk is a critical aspect of banking, and effective credit risk management is essential for banks to maintain financial stability and profitability. By properly assessing and managing credit risk, banks can offer credit products to customers while minimizing the risk of losses from defaults.



10. Credit Risk Practices

A. Private Practices

Credit risk management is essential for all businesses, including those in private practice. In private practice, credit risk management involves evaluating the creditworthiness of clients and determining the risk of non-payment for services rendered.

The following are some elements of credit risk management in private practice:

1. **Client Evaluation:** It involves the assessment of a client's financial stability, credit history, and reputation. It is crucial to ensure that clients have a strong financial standing and a history of timely payments.

2. **Credit Policies:** Developing credit policies that outline the terms and conditions of payment for services rendered is crucial. The policies should cover areas such as credit limits, payment terms, and consequences for non-payment.

3. **Credit Monitoring:** Regularly monitoring the creditworthiness of clients is vital to identify any potential risks. This can be done through credit checks and regular communication with clients.

4. **Credit Control:** It involves ensuring that clients adhere to the credit policies set out by the business. Credit control measures can include sending reminders for payments, applying late payment fees, or even ceasing services for clients with a history of non-payment.

5. **Credit Reporting:** Maintaining accurate records of credit transactions and reporting any late payments or defaults to credit bureaus is crucial. This helps to identify potential risks and prevent future losses.

Overall, credit risk management is a critical aspect of private practice operations. It helps to minimize the risk of non-payment and ensures that the business operates in a financially stable and sustainable manner.

B. Public Practices

Credit risk management is an important aspect of public practice as well. Public institutions such as government agencies, regulatory bodies, and central banks also deal with credit risks in their operations. The management of credit risk in public practice involves the identification, measurement, monitoring, and control of credit risk exposure in government-owned financial institutions and entities.

One of the main objectives of credit risk management in public practice is to ensure the stability of financial systems and prevent financial crises. For example, central banks play a crucial role in ensuring the stability of the banking system by supervising and regulating banks' credit risk management practices. They also provide liquidity support to banks during times of financial stress to prevent systemic risks.

In addition, public institutions may also be involved in providing credit to specific sectors such as agriculture, small and medium-sized enterprises, and infrastructure projects. In such cases, credit risk management is crucial to ensure that credit is allocated efficiently and effectively and that the risk of default is minimized. Overall, credit risk management in public practice is essential to ensure financial stability and promote economic growth. Public institutions must develop robust credit risk management frameworks and adhere to best practices to effectively manage credit risk exposure.

11. How to achieve Credit Control

Achieving credit control involves implementing strategies and procedures to manage credit risk and ensure the timely collection of outstanding debts. Some steps to achieve credit control include:

1. **Credit policy:** Establishing a clear and effective credit policy is crucial in managing credit risk. This policy should outline the criteria for extending credit, credit limits, payment terms, and procedures for assessing and monitoring creditworthiness.
2. **Credit assessment:** Conducting a thorough credit assessment is important to evaluate the creditworthiness of customers before

extending credit. This assessment should include a review of financial statements, credit reports, and other relevant information.

3. Credit monitoring: Regularly monitoring the creditworthiness of customers can help identify potential risks and take necessary action. This includes tracking payment patterns, credit utilization, and changes in financial status.

4. Credit terms: Setting clear and reasonable credit terms is important to ensure timely payments. This includes specifying due dates, late payment penalties, and interest rates.

5. Collection procedures: Establishing effective collection procedures can help ensure the timely payment of outstanding debts. This includes setting up a process for reminders, follow-ups, and escalation of non-payment.

6. Credit reporting: Reporting credit information to credit bureaus can help deter delinquent behavior and improve credit control. It also provides a valuable source of information for future credit assessments.

Overall, effective credit control requires a systematic approach to managing credit risk and ensuring timely collection of outstanding debts. It involves implementing clear policies and procedures, regular monitoring and assessment, and proactive collection strategies.

12. Credit Risk and Interest Rate

Credit risk and interest rates are interconnected as they both impact the profitability and financial health of a bank. When a bank lends money, it faces the risk that the borrower may not repay the loan. This is known as credit risk. To compensate for this risk, banks charge an interest rate on loans. The interest rate reflects the cost of borrowing money and compensates the bank for the risk it takes on by lending money.

The level of credit risk associated with a loan can impact the interest rate that a bank charges. For example, a borrower with a poor credit history may be considered high risk, and therefore the bank may charge a higher interest rate to compensate for the increased credit risk. Conversely, a borrower with a good credit history may be considered low risk, and the bank may charge a lower interest rate. Changes in interest rates can also impact credit risk. When interest rates increase, it becomes more expensive for borrowers to repay their loans. This can increase the risk of default, leading to higher credit risk for banks. On the other hand, when interest rates decrease, it becomes easier for borrowers to repay their loans, reducing the risk of default and lowering credit risk for banks.

In summary, credit risk and interest rates are closely linked as they both impact the profitability and risk profile of banks. Banks use credit risk models to assess the likelihood of default and set interest rates accordingly to manage their credit risk.

	Bad debts	Performing loans	Net income required	Interest income required	Interest rate
	(1)		(2)	(1) + (2)	
%	\$m	\$m	\$m	\$m	%
0	0	100	10	10	(10 ÷ 100) 10.0
1	1	99	10	11	(11 ÷ 99) 11.1
2	2	98	10	12	(12 ÷ 98) 12.2
5	5	95	10	15	(15 ÷ 95) 15.8
10	10	90	10	20	(20 ÷ 90) 22.2

13. Risks For Customers

Credit risk poses several risks to customers.

Firstly, customers may experience difficulties in repaying their loans, leading to default. This can result in penalties, additional fees, and damage to their credit score, making it more challenging to access credit in the future.

Secondly, customers may be subject to higher interest rates if they are deemed to be a higher credit risk. This can result in higher monthly payments and additional costs over the life of the loan.

Finally, customers may face challenges in obtaining credit altogether if they have a poor credit history or are considered a high credit risk. This can limit their ability to make necessary purchases or investments, potentially impacting their financial well-being.

It is essential for customers to understand their credit risk and take steps to manage it appropriately, such as by making timely payments, maintaining a good credit history, and avoiding taking on too much debt.

14. ELEMENTS FOR CREDIT RISK MODELLING

Credit risk management involves various elements that are essential for mitigating credit risk. Some of the critical elements of credit risk management are:

1. **Risk Identification:** This involves identifying and understanding the various risks associated with the lending portfolio. It is essential to understand the type and extent of risk present in the portfolio to manage it effectively.

2. **Risk Measurement:** This involves quantifying the risks identified. It is necessary to have a standard methodology for measuring and comparing different types of risk to prioritize and allocate resources to manage them.

3. **Risk Monitoring:** This involves tracking the identified risks and measuring them on an ongoing basis to ensure that the risks are within acceptable levels. It helps to detect any adverse trends or changes in the risk profile.

4. **Risk Mitigation:** This involves implementing appropriate risk mitigation strategies to reduce the identified risks to acceptable levels. The mitigation strategies may include reducing exposure, transferring risk, or enhancing risk management processes.

5. Risk Reporting: This involves reporting and communicating the risk profile, risk exposure, and risk management strategies to stakeholders, including regulators, shareholders, and customers.

6. Risk Culture: This involves establishing a risk culture that fosters risk awareness and encourages proactive risk management across the organization. It includes setting up policies, procedures, and processes that promote sound risk management practices.

Effective credit risk management involves a comprehensive approach that incorporates all these elements to minimize the impact of credit risk on the organization.

15. How to restrict Bad Credit

To restrict bad credit, financial institutions can take several steps including:

1. Implementing strict credit policies: Financial institutions can establish strict credit policies, including credit score thresholds and loan-to-value (LTV) ratios, to ensure that only creditworthy borrowers are approved for loans.
2. Conducting thorough credit assessments: Financial institutions can conduct comprehensive credit assessments of borrowers, including analyzing credit histories, financial statements, and other relevant data to evaluate their creditworthiness.
3. Setting appropriate interest rates: Financial institutions can set appropriate interest rates based on the credit risk of the borrower. Higher-risk borrowers can be charged higher interest rates to compensate for the increased risk.
4. Implementing risk management strategies: Financial institutions can use risk management strategies such as diversification of the loan portfolio, hedging, and securitization to reduce the overall risk of default.

5. Monitoring credit performance: Financial institutions can closely monitor the credit performance of borrowers, including tracking payment histories, and taking corrective actions if needed.

6. Using credit insurance: Financial institutions can use credit insurance to protect themselves against losses from default.

By implementing these measures, financial institutions can minimize their exposure to bad credit and reduce the risk of significant losses from default.

16. Challenges in Credit Risk

There are several challenges in credit risk management, some of which include:

1. **Data Quality:** The accuracy and completeness of data used for credit risk modeling are crucial. Poor data quality can lead to inaccurate risk assessments and poor lending decisions.
2. **Economic Conditions:** Economic conditions can have a significant impact on credit risk. Recessions, inflation, and other macroeconomic factors can affect borrowers' ability to repay loans and increase the risk of default.
3. **Regulatory Changes:** Changes in regulations, such as Basel III, can affect the way financial institutions manage credit risk. Compliance with these regulations can be costly and time-consuming.
4. **Technology:** The increasing use of technology in credit risk management has led to new challenges, such as data security and privacy concerns. There is also a need for financial institutions to continually update their technology to keep up with changing business needs and regulatory requirements.
5. **Human Factors:** The role of human judgment in credit risk management cannot be ignored. The challenge is to strike a balance between the use of technology and human judgment to make accurate and timely lending decisions.

6. **Portfolio Diversification:** Portfolio diversification is an essential strategy to manage credit risk. However, the challenge is to ensure that the portfolio is diversified across different industries and regions to avoid concentration risk.

Overall, credit risk management is a complex and evolving field that requires financial institutions to continuously monitor and adapt to changing economic conditions, regulatory requirements, and technological advancements.

17. Opportunities in Credit Risk Modelling

There are several opportunities in credit risk modeling, including

1. **Improved Risk Assessment:** Credit risk modeling provides financial institutions with a systematic and objective approach to assessing risk, allowing them to make informed lending decisions. By using advanced data analytics and machine learning techniques, institutions can improve their risk assessment capabilities and identify high-risk borrowers more accurately.

2. **Increased Efficiency:** Credit risk modeling can help financial institutions to streamline their lending processes, reducing costs and improving efficiency. By automating credit decision-making and risk assessment, institutions can reduce the time and resources required to process loan applications, allowing them to focus on other critical tasks.

3. **Better Customer Service:** Credit risk modeling can also help financial institutions to improve customer service by providing more personalized and tailored lending solutions. By analyzing borrower

data and credit history, institutions can offer customized loan products and services that meet their customers' specific needs and preferences.

4. Enhanced Risk Management: Credit risk modeling can also help financial institutions to improve their risk management capabilities by identifying potential risks and vulnerabilities in their lending portfolios. By monitoring and analyzing credit risk data on an ongoing basis, institutions can take proactive measures to mitigate risks and reduce losses from default.

Overall, credit risk modeling provides financial institutions with numerous opportunities to improve their lending processes, reduce risk, and enhance customer service. By embracing these opportunities, institutions can stay competitive in an ever-changing financial landscape and meet the evolving needs of their customers.

18. About Our Dataset

The dataset contains 11 features and 1 target variable. The target variable is called 'loan_status' and determines whether the said observation pertains to a defaulted loan. The 11 variables are as follows:

1. Borrower's age (person_age)
2. Borrower's income (person_income)
3. Borrower's housing status (person_home_ownership)
4. Borrower's period of employment in years (person_emp_length)
5. Loan Purpose (loan_intent)
6. Loan Amount (loan_amnt)
7. Interest rate (loan_int_rate)
8. Income as % of loan (loan_percent_income)
9. Historic Default (cb_person_default_on_file)
10. Credit History Length (cb_person_cred_hist_length)

Model and Performance Comparison:

Logistic Regression and Decision Tree Classifier models are used to predict default. The performance of both these models is assessed using Accuracy, Precision, Recall, F1 score, and AUC. Both the models are implemented and evaluated in Python.

Decision Trees

Accuracy 0.9062819725803152

Precision: 0.9253424657534246

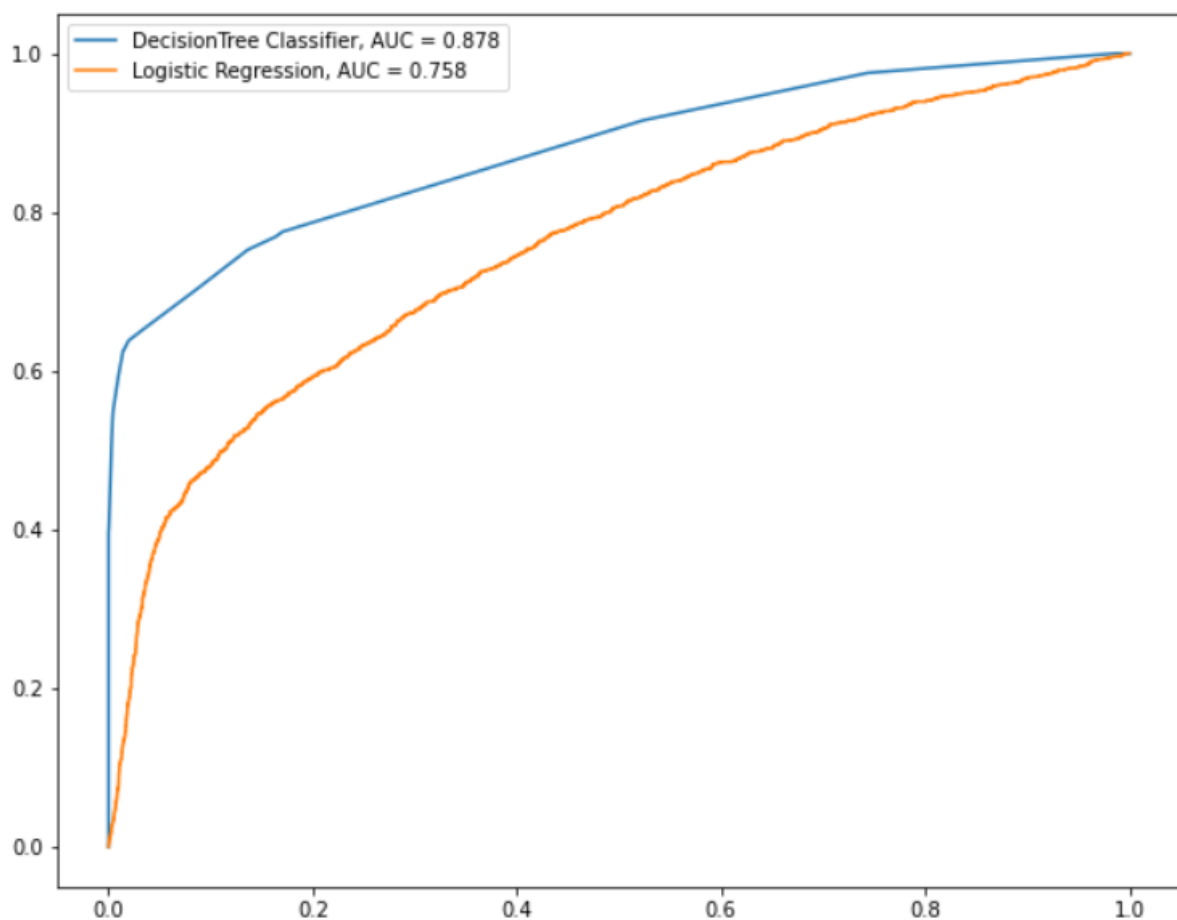
Recall: 0.6260426320667285

F1 score: 0.7468214483139856

Logistic Regression

Accuracy 0.7992633517495396
Precision: 0.7168141592920354
Recall: 0.15013901760889714
F1 score: 0.2482758620689655

The results of both these models are published here. The Decision Tree model is outperforming the Logistic Regression model on all parameters. Furthermore, the ROC curves of both these models are plotted and the Area Under the Curve (AUC) is computed.



The Decision Tree curve is higher than the Logistic Regression curve. This implies that Decision Tree provides a superior tradeoff between True Positive Rate and False Positive Rate. The area under the Decision Tree curve is 0.878 and that for Logistic Regression is 0.758.

Performance Evaluation metrics:

Accuracy, precision, recall, F1 score, AUC (Area Under the Curve), and ROC (Receiver Operating Characteristic) curve are performance metrics used to evaluate the performance of binary classification models.

Accuracy:

Accuracy is a performance metric used to evaluate the performance of a classification model. It represents the percentage of correctly classified instances over the total number of instances. In other words, it measures how often the model correctly predicts the correct class of an instance.

For example, if a model correctly predicts the class of 90 out of 100 instances, the accuracy of the model is 90%. It is calculated as:

Accuracy = (number of correctly classified instances) / (total number of instances)

Number of Correctly Classified Instances = (True Positives + True Negatives) / (True Positives + False Positives + True Negatives + False Negatives)

While accuracy is a useful metric, it can be misleading in certain situations, especially when the dataset is imbalanced or when the cost of misclassification is not the same for all classes. In such cases, other performance metrics such as precision, recall, F1 score, AUC, and ROC curve may be more appropriate.

Precision:

Precision is a performance metric that measures the proportion of true positives (correctly identified instances) among all instances that are identified as positive by a classification model. In other words, precision quantifies the model's ability to correctly identify positive instances, without falsely labeling negative instances as positive.

Precision is particularly useful when the cost of a false positive is high, meaning that misclassifying a negative instance as positive can have severe consequences. For example, in a medical diagnosis context, a false positive may lead to unnecessary and potentially harmful treatments for a patient who does not actually have a disease. In such cases, a high-precision model is preferred to minimize false positives.

Precision is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (FP):

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

A perfect precision score of 1.0 indicates that all instances labeled as positive are positive, while a score of 0.0 means that none of the instances labeled as positive are positive.

Recall:

Recall, also known as sensitivity, is a performance metric used in machine learning and classification problems that measure the proportion of actual positive instances that are correctly identified by the model. In other words, it is the ratio of true positives to the total number of actual positives.

Recall is particularly useful in scenarios where correctly identifying all positive instances is more important than minimizing false positives. For example, in a medical diagnosis task, correctly identifying all cases of a disease (true positives) is more critical than wrongly diagnosing a healthy person as having the disease (false positives).

Recall ranges from 0 to 1, where a value of 1 indicates perfect recall, meaning all actual positive instances were correctly identified by the model. Conversely, a recall of 0 indicates that the model failed to identify any actual positive instances correctly.

F1-Score:

The F1 score is a measure of a model's accuracy that combines both precision and recall into a single value. It is calculated as the harmonic mean of precision and recall, with values ranging from 0 to 1.

The F1 score is particularly useful in cases where both precision and recall are important, such as in binary classification problems where the classes are imbalanced.

It provides a balance between precision and recall, giving an overall evaluation of the model's performance. A higher F1 score indicates better performance, with a perfect score of 1 indicating that the model has perfect precision and recall.

The F1 score is calculated as:

$$\text{F1 score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

AUC (Area Under the Curve):

AUC (Area Under the Curve) is a performance metric that is commonly used in binary classification problems. It is used to evaluate the quality of a model's predictions and is calculated by computing the area under the Receiver Operating Characteristic (ROC) curve.

The ROC curve is a plot of the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The TPR represents the proportion of actual positive cases that are correctly classified as positive, while the FPR represents the proportion of actual negative cases that are incorrectly classified as positive.

The AUC ranges from 0 to 1, with a higher value indicating better performance. An AUC of 0.5 indicates random guessing, while an AUC of 1 indicates perfect classification. AUC is a useful metric when dealing with imbalanced datasets or when the cost of misclassification is not the same for all classes, as it provides an overall measure of model performance that is insensitive to class imbalance.

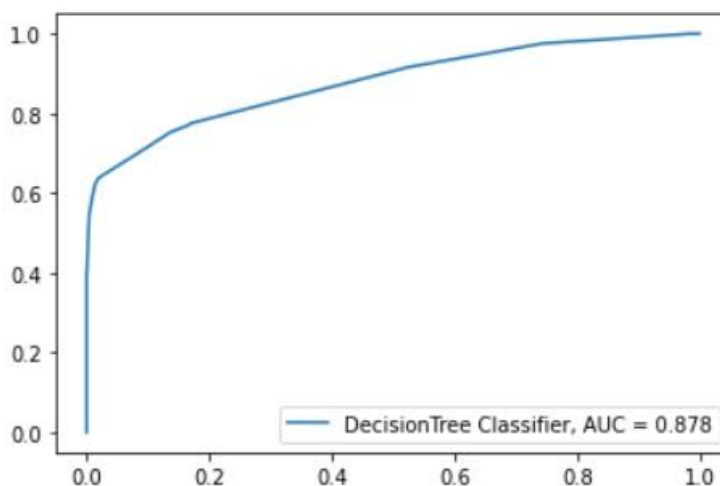
In summary, AUC is a widely used metric for evaluating the performance of binary classification models, providing a summary of the model's ability to distinguish between positive and negative classes at various classification thresholds.

19. REPORT:

1. Decision Tree Report

```
Decision Trees
Accuracy 0.9062819725803152
Precision: 0.9253424657534246
Recall: 0.6260426320667285
F1 score: 0.7468214483139856
```

```
<matplotlib.legend.Legend at 0x2c34399fcd0>
```



Accuracy: 0.9062

Precision: 0.9253

Recall: 0.6260

F1 Score: 0.7468

2. Logistic Regression Report

Logistic Regression

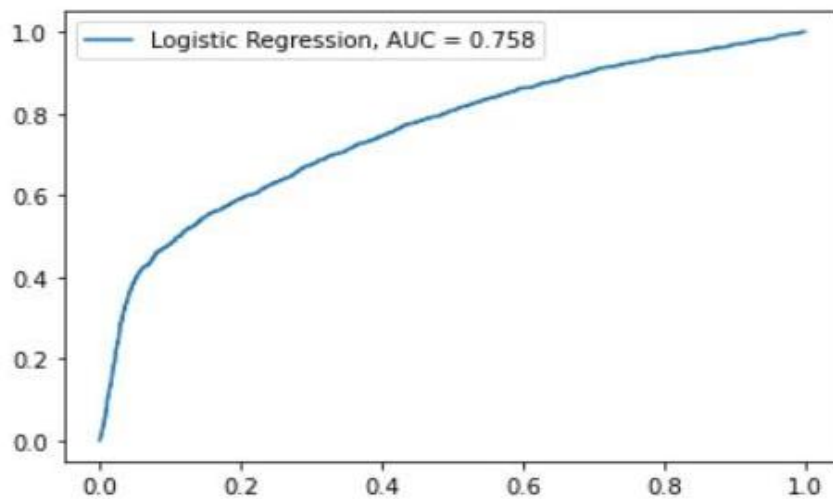
Accuracy 0.7992633517495396

Precision: 0.7168141592920354

Recall: 0.15013901760889714

F1 score: 0.2482758620689655

<matplotlib.legend.Legend at 0x2c346d4e160>



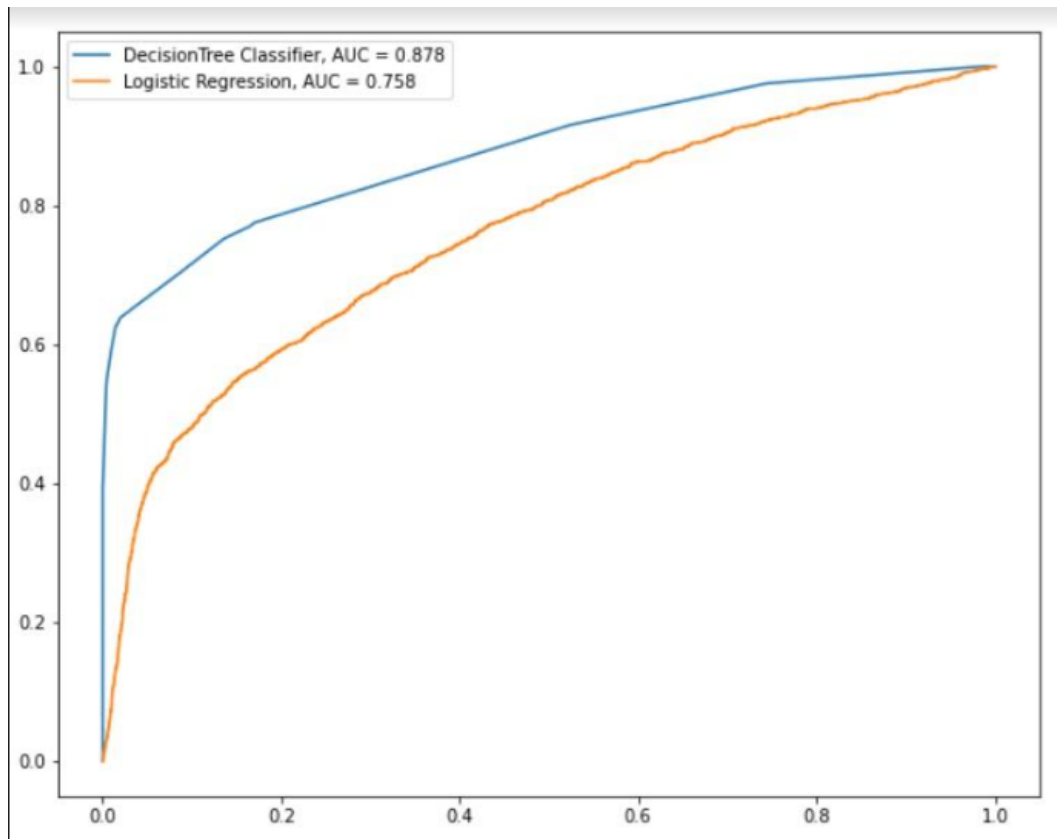
Accuracy: 0.7992

Precision: 0.7168

Recall: 0.1501

F1 Score: 0.2482

3. Comparing Logistic Regression & Decision Tree



Decision Tree, (AUC): 0.878

Logistic Regression, (AUC): 0.758

Model and Performance Comparison

Logistic Regression and Decision Tree Classifier models are used to predict default. The performance of both these models is assessed using Accuracy, Precision, Recall, F1 score, and AUC. Both models are implemented and evaluated in Python.

20. CONCLUSION:

In conclusion, credit risk modeling plays a vital role in the lending and credit decision-making process of financial institutions.

Using various quantitative and qualitative models, financial institutions can analyze borrower data and assess the risk of default, which helps them make informed decisions about lending, pricing, and risk management. Credit risk modeling offers several benefits, including reducing losses from default, improving credit risk management, and providing a systematic approach to assessing risk.

However, credit risk modeling also faces several challenges, such as data quality issues, model validation, and regulatory compliance. Despite these challenges, advancements in technology and data analytics offer significant opportunities to enhance credit risk modeling and improve the efficiency and effectiveness of credit risk management.

Overall, credit risk modeling is an essential tool for financial institutions in managing their credit risk and making informed lending decisions, contributing to the stability and growth of financial markets and the broader economy.