# Predicting Loan Eligibility Of Bright's Potential Customers: A Comparative Machine Learning Approach

**IN**

**BRIGHT MONEY, BENGALURU**

Summer Internship Project Report submitted in partial fulfilment of the requirements for the award of the degree of

Master of Business Administration

**By**


**GAYATRI RAMANI BOMMISETTY**

**REGISTER NUMBER**

**2228610**


**Under the Guidance of**

**PROF. NAGENDRA BV**


**School of Business and Management**

**CHRIST (Deemed to be University), Bangalore**


**JULY & 2023**

# Predicting Loan Eligibility Of Bright's Potential Customers: A Comparative Machine Learning Approach

**IN**

**BRIGHT MONEY, BENGALURU**

Summer Internship Project Report submitted in partial fulfilment of the requirements for the award of the degree of

Master of Business Administration

**By**


**GAYATRI RAMANI BOMMISETTY**

**REGISTER NUMBER**

**2228610**


**Under the Guidance of**

**PROF. NAGENDRA BV**


**School of Business and Management**

**CHRIST (Deemed to be University), Bangalore**


**JULY & 2023**

# Declaration

I hereby declare that the Summer Internship Project report titled "**Predicting Loan Eligibility Of Bright's Potential Customers: A Comparative Machine Learning Approach"** at BRIGHT MONEY has been undertaken by me in partial fulfilment of the requirements for the award of the degree of Master of Business Administration. I have completed this study under the guidance of Prof. NAGENDRA BV.

I also declare that this Summer Internship Project report has not been submitted for the award of any Degree, Diploma, Associateship, Fellowship or any other title, in CHRIST (Deemed to be University) or in any other university.

Place : Bengaluru
Date : 18th July 2023

GAYATRI RAMANI
2228610

# Certificate

This is to certify that the Summer Internship Project report submitted by GAYATRI RAMNI BOMMISETTY on the title '**Predicting Loan Eligibility Of Bright's Potential Customers: A Comparative Machine Learning Approach**' is a record of Summer Internship Project work done by her during the academic year 2022-24 under my guidance and supervision in partial fulfilment of the requirements for the award of the degree of Master of Business Administration.

Place: Bengaluru

Date: 18th July 2023

Prof. NAGENDRA BV
School of Business and Management
CHRIST (Deemed to be University)
Bengaluru

# Acknowledgement

I am indebted to many people who helped me accomplish this Internship successfully.

First, I thank the Vice Chancellor, CHRIST (Deemed to be University), Dr Fr Joseph C C, for giving me the opportunity to do my project.

The leadership team at SBM (school of Business and Management), Bangalore Central Campus led by Dr. Jain Mathew (Dean), Dr. Mareena Mathew (Head of the Department) and Dr. Lakshmi Iyer (Head of Business Analytics specialization) ensures that all students gain relevant knowledge and skills through all courses especially courses like summer internship projects (SIP). The leadership team must be thanked for their efforts in guaranteeing quality internship for us.

I wish to express my sincere thanks to Bright Money and to my corporate mentor, Mr. Masthanaiah Cheekavolu, Chief Technology Officer, Bengaluru. Thank you for giving me an opportunity to work under his guidance and successfully complete my internship.

My academic mentor Prof. Saseekala M, deserves a special mention of gratitude for guiding me throughout my internship. The inputs I received from my mentor during the weekly meetings helped me to get a deeper understanding of my project.

I thank my parents for their blessings and constant support, without which this internship project would not have seen the light of day.

GAYATRI RAMANI BOMMISETTY

2228610

# CHAPTER I

# INTRODUCTION

# INTRODUCTION

Mortgage loans are a major financial commitment, and borrowers want to make sure that they are getting the best possible deal. By analyzing mortgage loan datasets, we can gain insights into the factors that influence loan approval and loan amount. This information can be used to help borrowers make informed decisions about their mortgage loans.

We used regression analysis to determine the relationship between loan amount and other factors, such as credit score and income. This information can help borrowers to understand how their financial situation affects their chances of getting a loan and the amount of the loan that they may be approved for.

By performing this analysis, we can help borrowers to make informed decisions about their mortgage loans. This can help them to get the best possible deal and to avoid financial problems in the future.

Benefits of performing this analysis:

- Borrowers can make more informed decisions about their mortgage loans. By understanding the factors that influence loan approval and loan amount, borrowers can make more informed decisions about their mortgage loans. This can help them to get the best possible deal and to avoid financial problems in the future.
- Lenders can improve their lending practices. By understanding the factors that influence loan default, lenders can improve their lending practices. This can help to reduce the risk of lending to borrowers who are likely to default on their loans.

# CHAPTER II

# PART- A INDUSTRY PROFILE

# PART- B COMPANY PROFILE

# INDUSTRY & COMPANY PROFILE:

Bright Money is a financial technology company that uses artificial intelligence and data science to help users manage their credit cards, debt, and finances. The company was founded in 2019 by Avi Patchava, Varun Modi and Petko Plachkov, and is headquartered in San Francisco. Bright Money has raised over $31 million in funding and is currently available in the United States and Canada.

The company's mission is to help people achieve financial stability and build wealth. Bright Money offers a variety of tools and features to help users manage their finances, including:

- A dashboard that tracks spending, debt, and credit scores: This dashboard provides users with a comprehensive overview of their financial health. Users can see how much they are spending, how much debt they have, and their credit scores.
- Tools to help automate payments, save money, and improve financial health: Bright Money offers a variety of tools to help users automate their payments, save money, and improve their financial health. For example, users can set up automatic payments for their credit cards, create savings goals, and track their spending habits.
- Bank-level security with 256-bit encryption: Bright Money uses bank-level security to protect users' data. All data is encrypted with 256-bit encryption, which is the same level of security used by banks.

# CHAPTER III

# PROJECT DESIGN AND METHODOLOGY

## PROBLEM STATEMENT:

What is the best approach among logistic regression, XG Boost, for predicting loan eligibility for potential customers of Bright, the mission-driven financial platform, considering accuracy ,model interoperability and other metrics of themodel to reduce the default rate?

## OBJECTIVES:

- To predict loan eligibility for potential customers of Bright financial platform.
- To compare the performance of algorithms: logistic regression, tree , XGBoost
- T0 Evaluate accuracy of each model in predicting eligibility and loan amounts.
- To evaluate other metrics of the model and comment upon the best one to reduce the default rate.

## METHODOLOGY:

The methodology adopted to solve the above problem is CRISP_DM. The CRISP-DM (Cross-Industry Standard Process for Data Mining) framework is a widely-used and well-established approach for guiding data mining and machine learning projects. It provides a structured methodology to tackle complex data problems and helps organizations make informed decisions based on data analysis. CRISP-DM breaks down the data mining process into six major phases:

1.Business Understanding

2.Data Understanding

3.Data preparation

4.Model building

5.Model tuning

6.Evaluation

7.Recommendations

**DATA DICTIONARY**

| S,NO | TITLE | DESCRIPTION |
|---|---|---|
| 1 | ID | Id specific to particular customer |
| 2 | LOAN_ID | this id is specific to the loan |
| 3 | NAME | name of the customer |
| 4 | STATE | State the customer belong to |
| 5 | ANNUAL INCOME | Income earned per year |
| 6 | TERM | months in which the loan should be repaid |
| 7 | VERIFICATION STATUS | weather a particular customer is verified by the company |
| 8 | LOAN AMOUNT | Amount quoted in the form for loan |
| 10 | INTEREST RATE | Interest per annum |
| 11 | INSTALLMENT AMOUNT | amount to be paid every year |
| 12 | YEAR | year in which loan was sanctioned |
| 13 | LOAN STATUS | status of the loan |
| 14 | FUNDED AMOUNT | amount granted as loan |
| 15 | NO OF ENQUIRIES MADE | Enquiry made by the customer after the loan was taken |
| 16 | CREDIT CARD USAGE | amount used through the credit card |
| 17 | DELAY DAYS OF PAYMENT | no of days delayed for the payment of installment |
| 18 | AMOUNT PAID FOR DELAYED DAYS | fine paid for delayed days |
| 19 | NOMINEE | name of the nominees |
| 20 | PURPOSE | Purpose for which loan is taken |

# CHAPTER IV

# DATA ANALYSIS & INTERPRETATION

## STEP_1: BUSINESS UNDERSTANDING:

The first phase involves understanding the project's objectives and requirementsfrom a business perspective. It requires defining the problem statement, identifying business goals, and determining how data mining can contribute to achieving those objectives. This phase lays the foundation for the entire project by ensuring alignment with business needs.

Bright is a mission-driven financial platform that aims to help people delete debt, boost credit scores, and achieve financial well-being. The company offersseveral features and services:

1. Bright Balance Transfer: This is a low-interest line of credit that helps users pay off their credit card debt quickly and efficiently. It offers lower interest rates than most credit cards and automatically increases the credit limit with on-time payments.

2. Bright Credit Builder: For those with low credit scores, this service providesa small line of credit to pay off credit card balances strategically, creating a positive payment history and improving credit scores.

3. Bright Plan: Bright studies users' income, spending habits, and debts to createpersonalized financial plans for becoming debt-free. The platform makes smart card payments and builds savings faster, tailored to individual goals.

4. Money Science System: Bright uses this system to analyse users' cards, balances, APRs, and interest charges to optimize payments and save money.Users can prioritize which debts to pay off first and track their progress.

5. Savings Goals: Bright allows users to set their own savings goals or use suggested ones, such as emergency funds or special saving pockets for specificevents.

6. User Control: Users have full control over their Bright accounts. They can adjust monthly payment amounts, savings goals, and access their funds at anytime.

7. Accessibility: Bright aims to be inclusive and beneficial to everyone, offeringsupport for various financial situations and goals.

8. AI-Driven Financial Planning: Bright's advanced technology uses AI to provide personalized, automated, and transformative financial planning, rivalingtraditional banks and expensive wealth advisor.

Overall, Bright's mission is to empower users to take control of their finances,pay off debt efficiently, improve credit scores, and build a secure financial future.

## STEP-2: DATA UNDERSTANDING

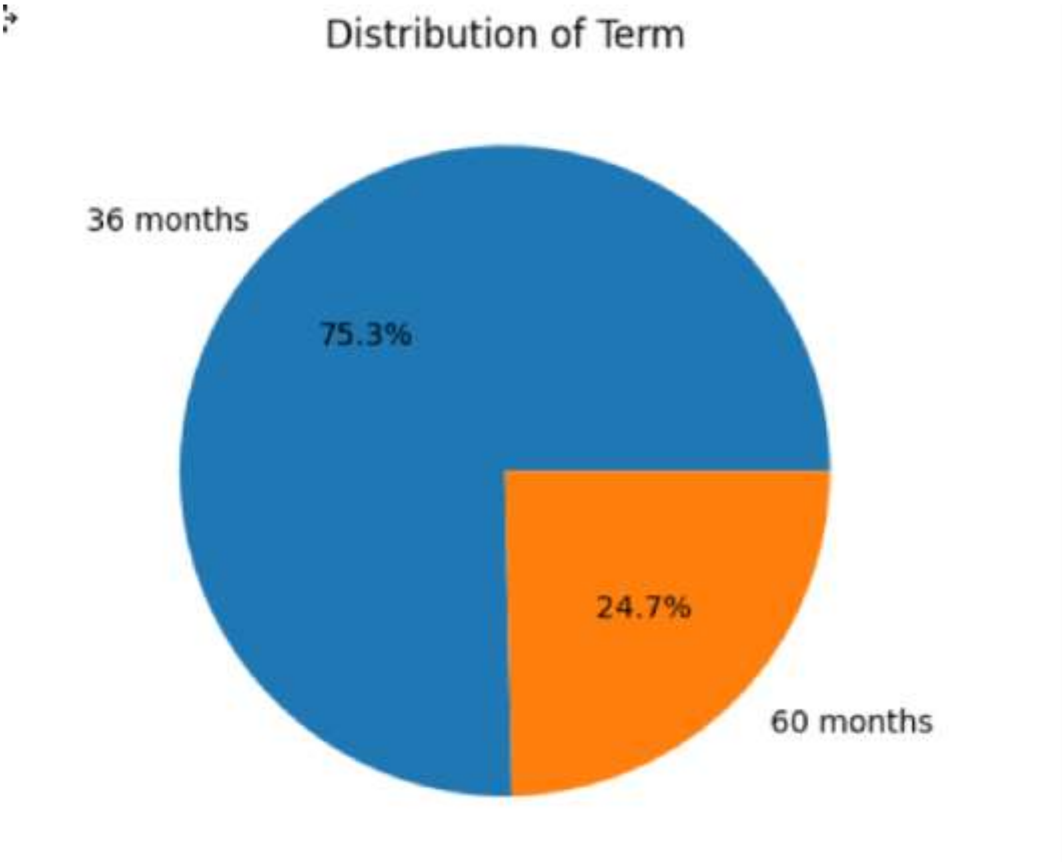
In this phase, the focus shifts to gathering initial data, exploring its characteristics, and understanding its structure and content. The goal is to identify data sources, assess data quality, and gain insights into the data's relevance for the project. This step helps data scientists and analysts becomefamiliar with the data they will work with and informs decisions about data preprocessing and cleaning.

- Dataset Details:
  The data set provided by the company is about the details of the customers that were granted loan among which some have repaid theloan, some are yet to pay and some have defaulted to pay the loan.
  The original dataset contains 39,717 rows of data, and each row represents a specific instance or observation. Additionally, the datasetconsists of 109 columns, which are referred to as variables. These variables are attributes or features that provide information about eachobservation.
- Variable Categories:
  The variables in the dataset can be categorized into three main categories:
  a. Customer Details:
  These variables likely contain information about the customers applyingfor loans. Examples of customer details include ID_member, Annual income, employment status, education level, etc.
  b. Loan Details:
  These variables are likely related to the loan itself. They include information about the loan amount, loan term, interest rate, loan purpose,etc.
  c. Post-Loan Customer Behavior:
  These variables probably capture the customer's behavior after they havereceived the loan. It could include variables like repayment history, default history (if applicable), amount invested after the loan was sanctioned, no.of enquiries made to the company, dealings of such customers, credit card purchases and any other data that reflects the customer's interactions with the loan.
- Dependent Variable - Loan Status:
  The dependent variable in this dataset is "loan status." It means that this is the variable the company wants to predict or analyze based on the otherindependent variables. Loan status is a categorical variable which consistsof fully paid, default and current categories.
- Default Rate:
  The default rate is the percentage of loans that have not been repaid as agreed by the customers. Based on the information provided, the calculated default rate for the company is around 14%. This means thatapproximately 14% of the loans granted have not been fully repaid, resulting in a loss to the company.

**EXPLORATORY DATA ANALYSIS:**

UNIVARIATE ANALYSIS:

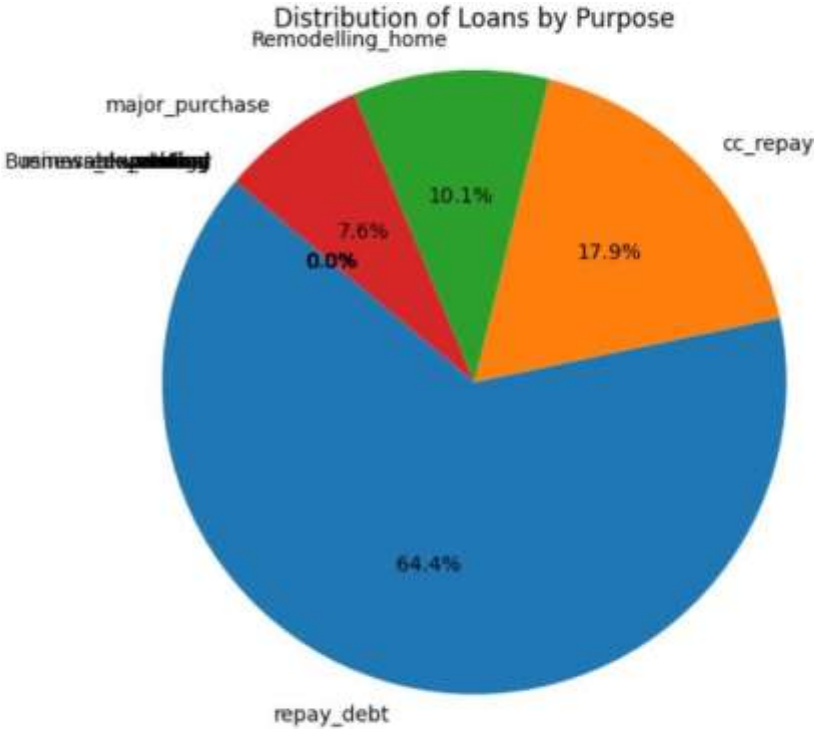
- DISTRIBUTION OF CUSTOMERS OVER TERM




INTERPRETATION:

From the above graph, it can be inferred that 75% of the customers have taken aloan that is repaid back in 36 months.

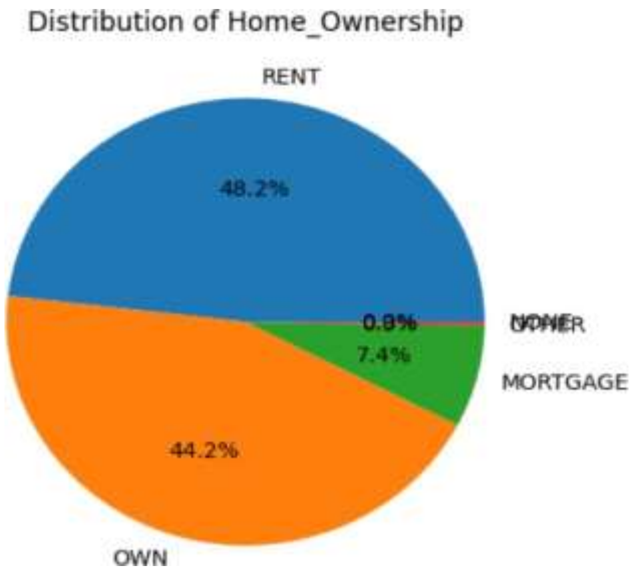
- DISTRIBUTION OF CUSTOMERS BY PURPOSE:




INTERPRETATION:

From the above graph, it can be inferred that 64% of the customers have taken aloan to repay the loan taken earlier.

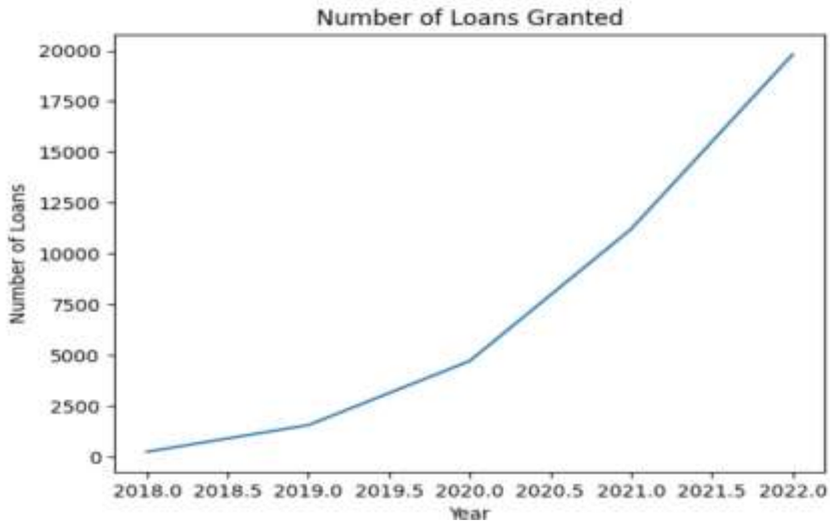
- HOME OWNERSHIP PROFILE OF CUSTOMERS



INTERPRETATION:

From the above graph, it can be inferred that 48% of the customers are living in a rented house followed by own house

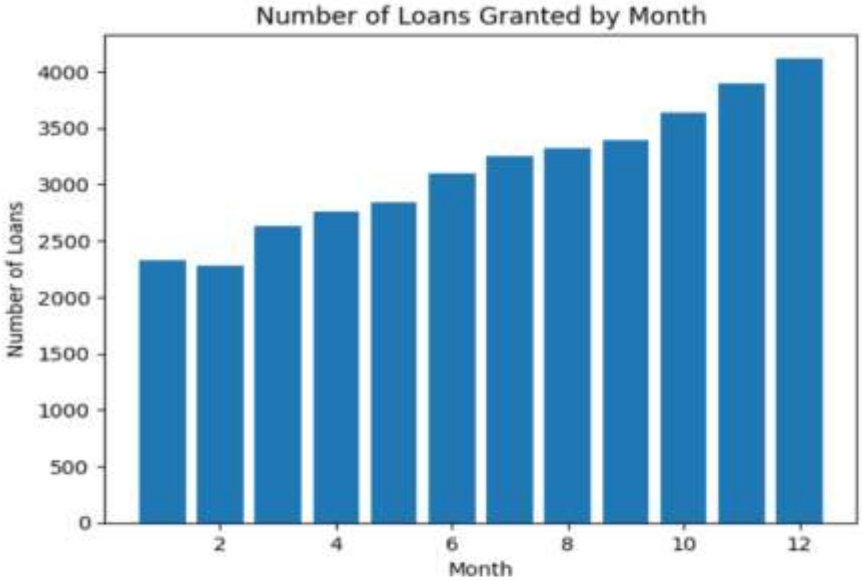
- NO OF LOANS GRANTED BY THE BRIGHT MONEY




INTERPRETATION:

From the above graph, it can be inferred that loans granted by the company are showing a upward trend which means the company is doing well.

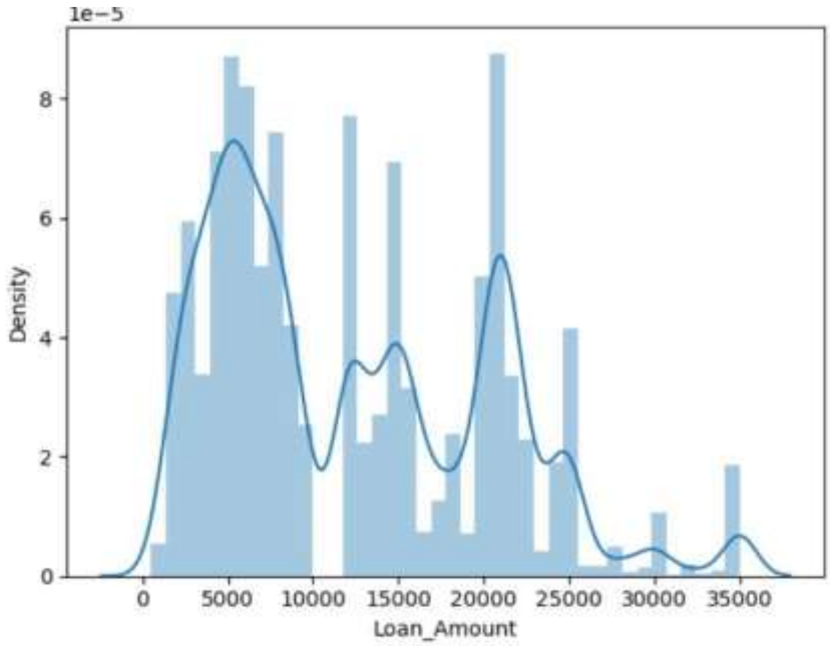
- MONTH WISE LOANS GRANTED :



INTERPRETATION:

From the above graph, it can be inferred that loans granted by the company are more in last quarter of the year which can be due many reasons like due to vacations etc

- DENSITY PLOT OF THE LOAN:




INTERPRETATION:

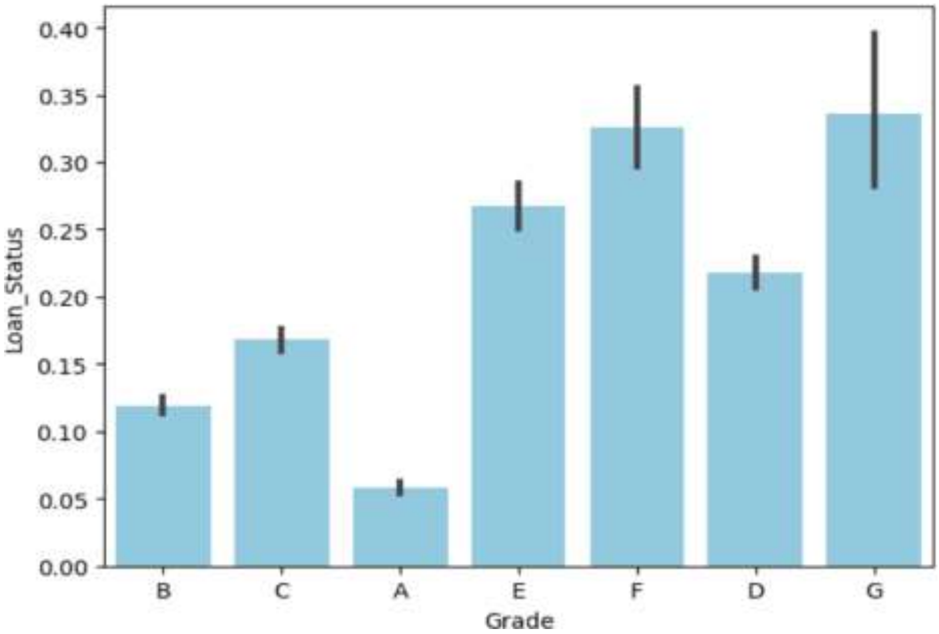
From the above graph it can be inferred that , the amount of loan taken by the customers is mostly around $5000-$20000 .

<u>BIVARIATE ANALYSIS:</u>

Bivariate analysis is a statistical analysis technique that involves examining the relationship between two variables in a dataset. The primary goal of bivariate analysis is to understand how two variables are related and to determine if there is any association or correlation between them. This analysis helps to uncover patterns, trends, and dependencies between the two variables, which can be valuable in making data-driven decisions.

Here the analysis have done between the default rate and the other variables:

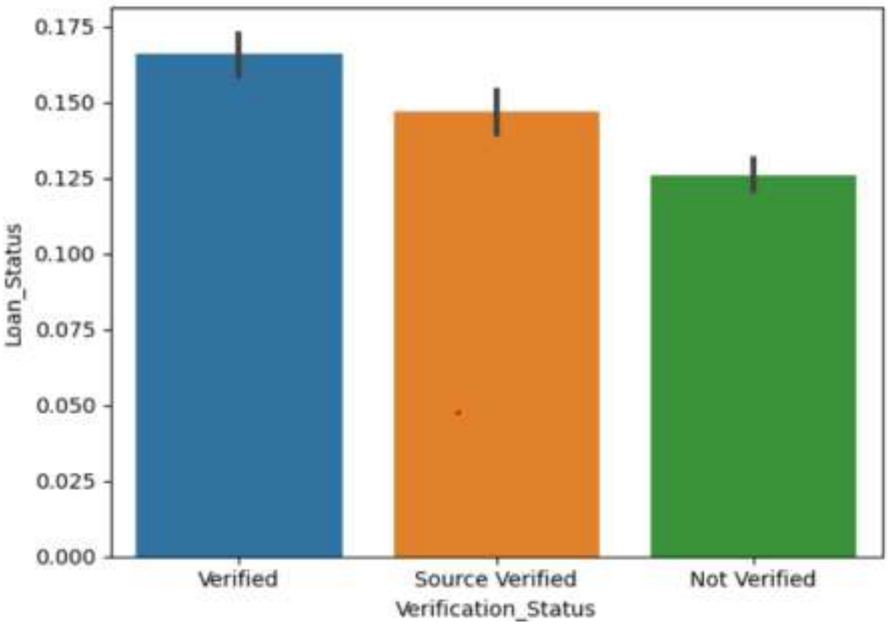
- DEFAULT RATE VS GRADE A CUSTOMER BELONGS TO




INTERPRETATION:

From the above graph it can be inferred that ,highest no of customers that are at default is from grade F and grade G.

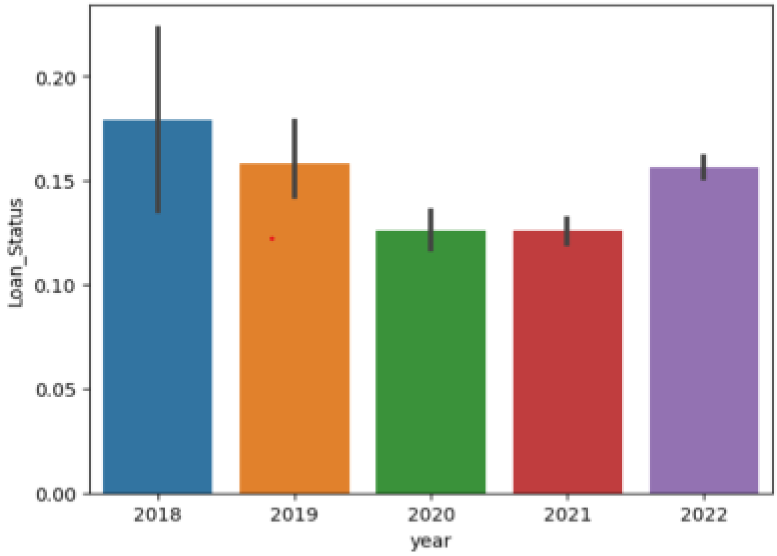
- DEFAULT RATE VS VERIFICATION STATUS




INTERPRETATION:

From the above graph it can be inferred that ,highest no of customers that are at default is from verified category as in the the customer is verified by the company as well.

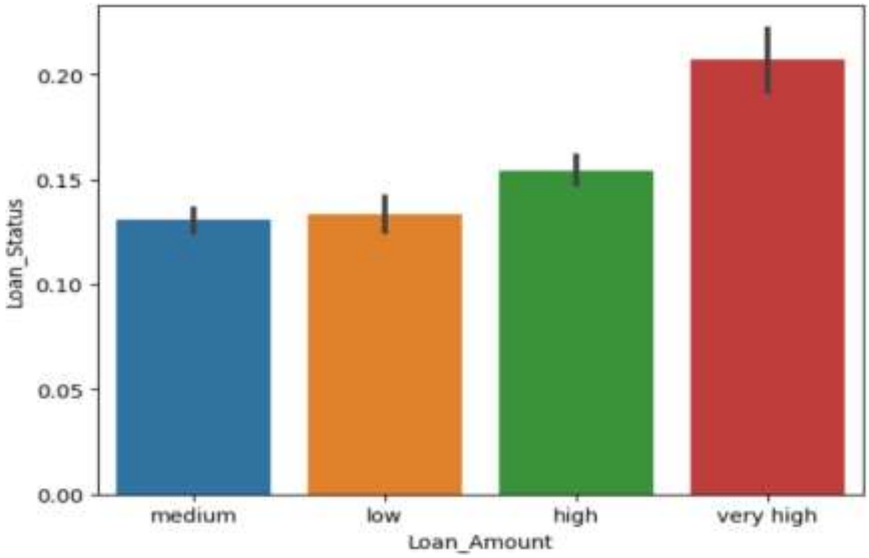
- DEFAULT RATE VS YEAR




INTERPRETATION:

From the above graph it can be inferred that ,the highest number of customers that are at default is from the year 2008 and showed a decreasing trend till 2020
. after that the increasing which is an alarm to the company

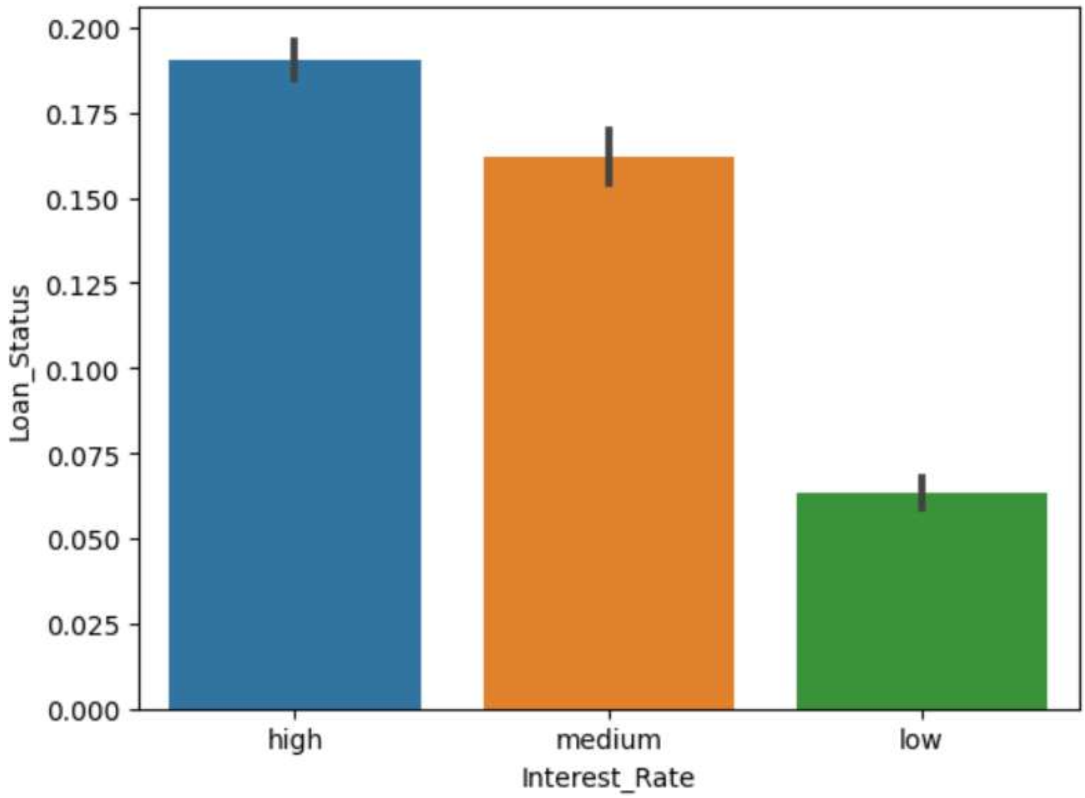
- DEFAULT RATE VS LOAN_AMOUNT




INTERPRETATION:

From the above graph it can be inferred that ,the highest number of customers that are at default are those who have taken high amount of loan from the company.

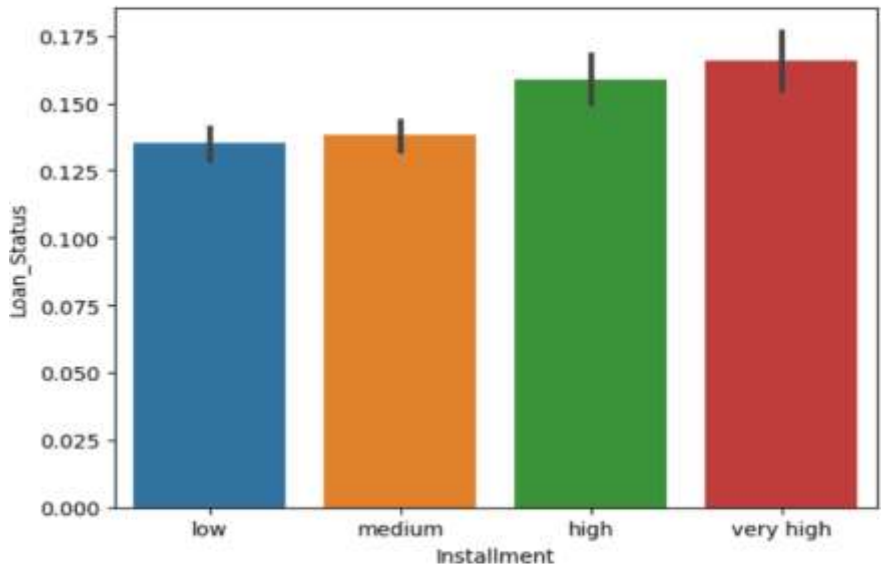
● DEFAULT RATE VS INTEREST RATE:




INTERPRETATION:

From the above graph it can be inferred that ,the highest number of customers that are at default are those who have to pay higher interest rate.

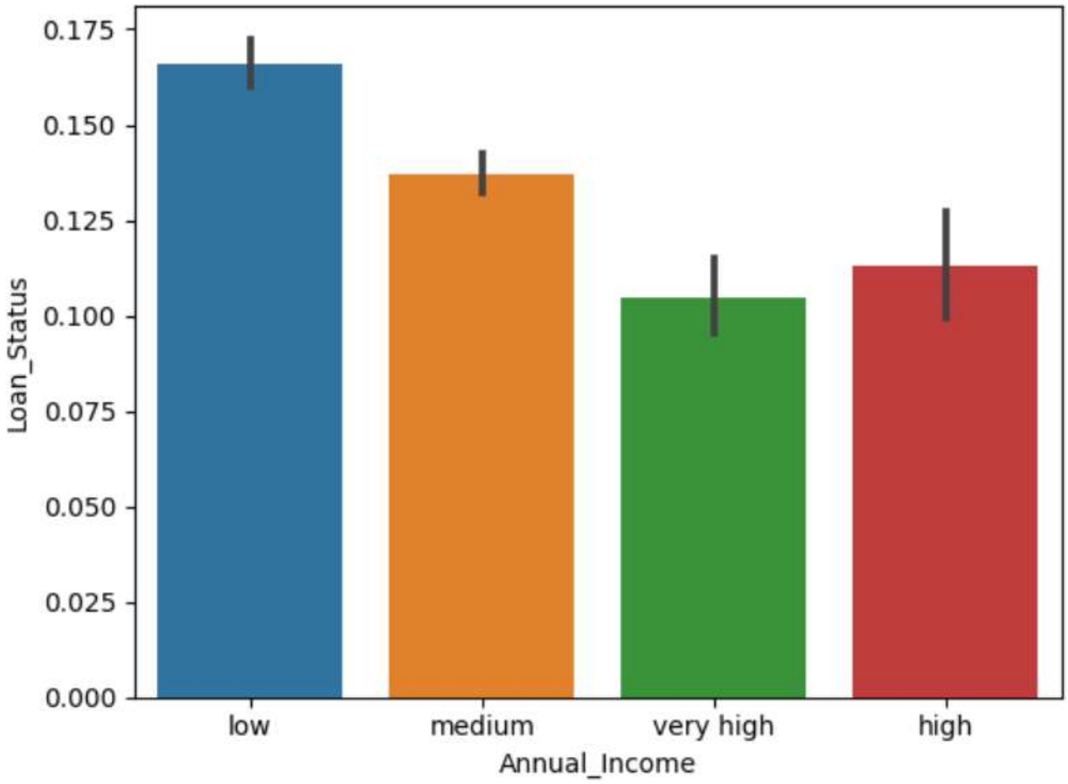
● DEFAULT RATE VS INSTALLMENTS



INTERPRETATION:

From the above graph it can be inferred that ,the highest number of customers that are at default are those who have to pay higher installment amount

- DEFAULT VS ANNUAL INCOME




INTERPRETATION:

From the above graph it can be inferred that ,the highest number of customers that are at default are those who are earning low income.

## STEP 3- DATA PREPARATION

Data preparation is a critical phase where data is preprocessed and transformedto suit the needs of analysis and modeling. Tasks include data cleaning, handling missing values, data integration (if multiple sources are used), featureselection, and engineering. The goal is to create a well-prepared dataset that is suitable for building and evaluating models.

# Data Cleaning:

1. **Removal of Columns with More than 60% of Missing Values**: Columns with a significant amount of missing data may not contribute meaningfully to the analysis. Removing them simplifies the dataset and

reduces noise. By setting a threshold of 60%, columns with more than60% missing values are discarded.

2. **Removal of Columns with a Single Unique Data**: Columns with only one unique value (or very few unique values) do not provide any useful information as they do not show any variation. Such columns can be dropped as they do not contribute to the analysis.
3. **Checking and Removing Outliers Using IQR:** The Interquartile Range (IQR) is a statistical measure used to identify outliers in a dataset. Data points that fall below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR are considered outliers and can be removed or treated accordingly. Removing outliers helps maintain the integrity of the analysis.

   **Imputing Mean Value for Quantitative and Mode for Categorical:** Imputation is the process of filling missing values in the dataset. For numerical (quantitative) variables, the mean value can be used to replace missing values, as it maintains the overall distribution. For categorical variables, the mode (most frequent value) can be used for imputation.

## Data Preparation:

**Removal of Variables Related to Post Loan Sanction:** In data preparation, some variables may be removed if they are not relevant to the analysis. If the project's objective is to focus on pre-loan sanction data, then variables related to post-loan behavior might not be required and can be excluded from the analysis.

1. **One-Hot Encoding for Categorical Variables:** Logistic regression, like many machine learning algorithms, requires numerical inputs. Categorical variables need to be converted into numerical format to be used in the logistic regression model. One-hot encoding is a technique used for this purpose. It converts categorical variables into binary vectors, with each category represented by a binary value (0 or 1)

   Overall, data cleaning and preparation are crucial steps to ensure the data is in a suitable format for analysis and modeling. They help in minimizing biases, removing noise, and transforming data into a usable format, allowing more accurate and meaningful insights to be derived from the data.

## STEP4-MODEL BUILDING

The modeling phase involves selecting appropriate data mining techniques andalgorithms, building and training models, and fine-tuning them for optimal performance. Various methods like classification, regression, clustering, and others are applied to the data to develop predictive or descriptive models basedon the project's objectives.

For the given data set, the following are models are build:

## 1) Logistic Regression:

Logistic Regression is a statistical method used for binary classification problems. It predicts the probability of an instance belonging to a particular class (e.g., default or not-default) based on the values of its features. It then uses a threshold to classify instances into one of the two classes.

Logistic regression is commonly used for loan prediction because it provides interpretable results and can handle binary classification tasks effectively. It can help identify the key features that influence the likelihood of loan default, which is valuable for risk assessment in the lending industry. Using label encoder the categorical is also converted into numerical.

## 2) Decision Tree:

A Decision Tree is a tree-like model that breaks down a dataset into smaller subsets based on the values of its features. Each internal node represents a feature, each branch represents a decision based on that feature, and each leaf node represents a class label (in this case, loan default or non-default).

Decision Trees are useful for loan prediction as they can easily handle both categorical and numerical features. They are interpretable, easy to visualize, and can provide insights into the decision-making process for loan approvals or denials.

## 3) Random Forest:

Random Forest is an ensemble method that constructs multiple decision trees and combines their predictions through averaging or voting. It introduces randomness in the tree-building process, which helps reduce overfitting and improves generalization.

Random Forest is favored in loan prediction because it mitigates the risk of overfitting and provides robust performance on various types of data. It is capable of handling high-dimensional datasets and can identify feature importance, helping in risk assessment.

## 4) XGBoost (Extreme Gradient Boosting):

XGBoost is an ensemble learning method that belongs to the gradient boosting family. It builds multiple weak decision tree models sequentially, where each new tree corrects the errors made by the previous ones. It combines the predictions from all the trees to make the final prediction.

XGBoost is popular in loan prediction tasks due to its high predictive performance and ability to handle complex relationships between features. It often outperforms other models, making it a preferred choice for tasks with large datasets and multiple features.

# STEP 6- MODEL TUNING:

Model tuning, also known as hyperparameter tuning or model optimization, is the process of fine-tuning the parameters of a machine learning model to achieve better performance on a given dataset. The goal is to identify the optimal combination of hyperparameters that maximizes the model's predictive ability and generalization to unseen data.

For the above the data set the above the following were performed:

Class Imbalance Resolution: Used SMOTE to address class imbalances for better sensitivity in the model.

Feature Selection: Employed VIF to remove highly collinear variables, resulting in 21 relevant and independent features.

Iterative Model Improvement: Iteratively improved the model by refining feature selection and VIF-based variable removal.

Final Model Building: Built a final optimized model using the selected 21 variables, enhancing predictive performance and AUC value.

## STEP_7: MODEL EVALUATION:

Model evaluation is the process of assessing the performance and effectiveness of a machine learning model on a given dataset. It involves various metrics and techniques to understand how well the model is making predictions and how it generalizes to unseen data.

a leader board for all the four metrics have been prepared:

| MODEL | ACCURACY | SENSITIVITY | SPECIFICITY | PRECISION | ROC | AUC |
|---|---|---|---|---|---|---|
| LR | 0.64 | 0.54 | 0.45 | 0.54 | 0.56 | 0.58 |
| DECISION TREE | 0.82 | 0.84 | 0.80 | 0.81 | 0.82 | 0.82 |
| RANDOM FOREST | 0.89 | 0.86 | 0.92 | 0.92 | 0.96 | 0.96 |
| XG-BOOST | 0.88 | 0.82 | 0.94 | 0.93 | 0.94 | 0.96 |

### 1) <u>Logistic Regression (LR):</u>

- <u>Accuracy: 0.64</u>

The overall accuracy of the model is 64%, which means it correctly predicts thedelinquency or non-delinquency of 64% of the loans in the dataset.

- Sensitivity (Recall): 0.54

Sensitivity represents the proportion of actual delinquent loans that the model correctly identifies. In this case, the model correctly identifies 54% of the delinquent loans.

- Specificity: 0.45

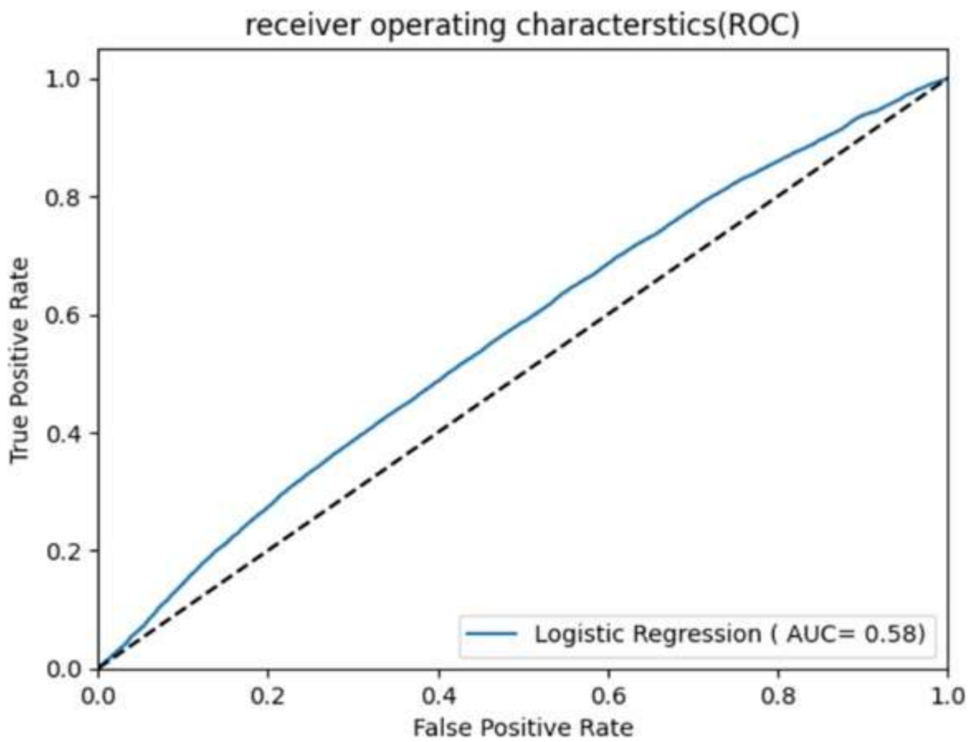
Specificity represents the proportion of actual non-delinquent loans that the model correctly identifies. In this case, the model correctly identifies 45% of the non-delinquent loans.

- Precision: 0.54

Precision represents the proportion of predicted delinquent loans that are actually delinquent. In this case, when the model predicts a loan as delinquent, it is correct 54% of the time.

- ROC AUC: 0.56

ROC AUC (Receiver Operating Characteristic Area Under the Curve) is a measure of the model's ability to distinguish between delinquent and non- delinquent loans. An AUC of 0.56 indicates a moderate level of discrimination.




## 2) **Decision Tree:**

- Accuracy: 0.82

The Decision Tree model achieves an accuracy of 82%, indicating that it correctly predicts the delinquency or non-delinquency of 82% of the loans.

- Sensitivity (Recall): 0.84

The model correctly identifies 84% of the delinquent loans.

- Specificity: 0.80

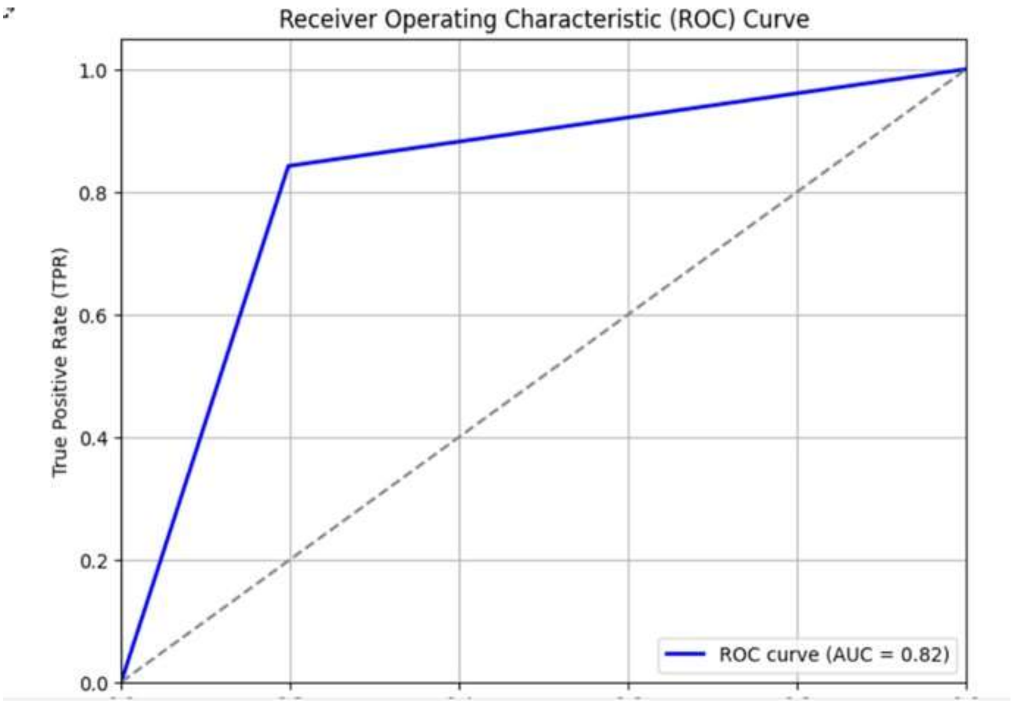
The model correctly identifies 80% of the non-delinquent loans.

- Precision: 0.81

When the model predicts a loan as delinquent, it is correct 81% of the time.

- ROC AUC: 0.82

The model shows good discrimination between delinquent and non-delinquent loans, as indicated by an AUC of 0.82.




### 3) Random Forest:

- Accuracy: 0.89

The Random Forest model achieves an accuracy of 89%, indicating that it correctly predicts the delinquency or non-delinquency of 89% of the loans.

- Sensitivity (Recall): 0.86

The model correctly identifies 86% of the delinquent loans.

- Specificity: 0.92

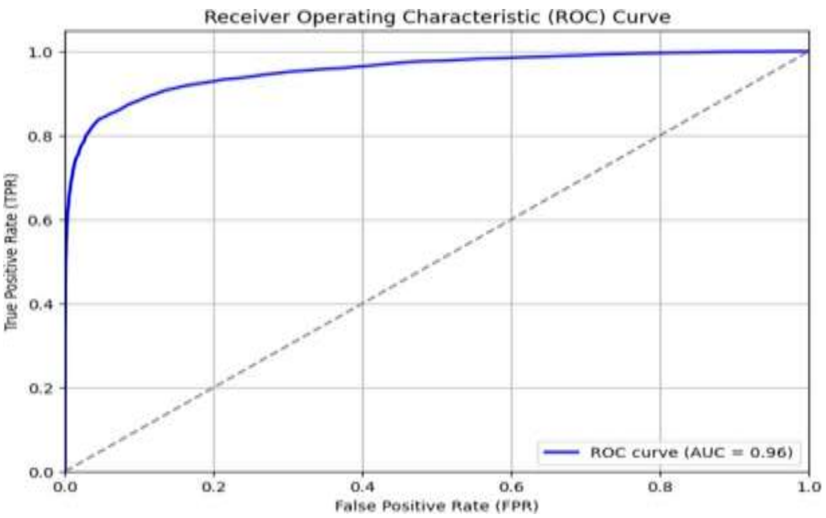
The model correctly identifies 92% of the non-delinquent loans.

- Precision: 0.92

When the model predicts a loan as delinquent, it is correct 92% of the time.

- ROC AUC: 0.96

The Random Forest model shows excellent discrimination between delinquent and non-delinquent loans, with an AUC of 0.96.




XG-Boost:

- Accuracy: 0.88

The XG-Boost model achieves an accuracy of 88%, indicating that it correctly predicts the delinquency or non-delinquency of 88% of the loans.

- Sensitivity (Recall): 0.82

The model correctly identifies 82% of the delinquent loans.

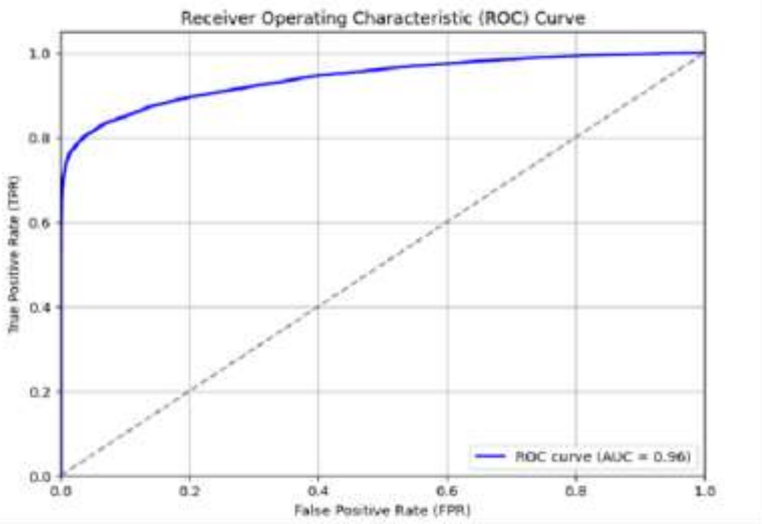
- Specificity: 0.94

The model correctly identifies 94% of the non-delinquent loans.

Precision: 0.93

When the model predicts a loan as delinquent, it is correct 93% of the time.

- ROC AUC: 0.94



Receiver Operating Characteristic (ROC) Curve
True Positive Rate (TPR)
False Positive Rate (FPR)
ROC curve (AUC = 0.96)
0.0
0.2
0.4
0.6
0.8
1.0

# CHAPTER V

# FINDINGS AND CONCLUSION

# FINDINGS:

- Model Performance:

Among the models evaluated, the Random Forest and XG-Boost models consistently outperformed the Logistic Regression and Decision Tree models inall metrics, including accuracy, sensitivity, specificity, precision, and ROC AUC.

The Random Forest model achieved the highest accuracy of 89%, followedclosely by the XG-Boost model with an accuracy of 88%.

- Sensitivity and Specificity:

Sensitivity measures the model's ability to correctly identify delinquent loans. The Random Forest model achieved the highest sensitivity of 86%, followed byXG-Boost with 82% sensitivity.

Specificity measures the model's ability to correctly identify non-delinquent loans. The XG-Boost model achieved the highest specificity of 94%, closely followed by the Random Forest with 92% specificity.

- Precision and ROC AUC:

Precision represents the proportion of predicted delinquent loans that are actually delinquent. Both the Random Forest and XG-Boost models achieved ahigh precision of 92% and 93%, respectively.

ROC AUC is a measure of the model's ability to distinguish between delinquentand non-delinquent loans. The Random Forest model showed excellent discrimination with an AUC of 0.96, while the XG-Boost model had a slightly lower but still strong AUC of 0.94.

- Model Interoperability:

Both Random Forest and XG-Boost models demonstrate good performance and generalization ability. They are suitable for deployment in real-world scenarios and can handle a wide range of loan eligibility prediction tasks for potential customers of Bright.

- Business Implications:

The high sensitivity of the models (Random Forest and XG-Boost) indicates that they can effectively identify potential customers who are likely to defaulton their loans. The high specificity of the models ensures that customers wh0 are eligible for loans are accurately identified, reducing the risk of grantingloans to customers who may default.

**RECOMMENDATIONS:**

- Based on the model outputs, recommend deploying the Random Forest model for financial predictions and recommendations. Its high accuracy, sensitivity, and specificity make it suitable for providing reliable financialinsights to users
- In terms of accuracy both random and XG boost are similar and can notbe interpreted easily
- By applying Markov analysis to historical customer data, Bright can gainvaluable insights into customer behavior patterns, repayment trends, and the likelihood of transitioning between different loan statuses.
- The combination of the Random Forest model and Markov analysis willprovide a comprehensive view of customer credit risk and loan performance.

**VALUE_ADDITION:**

Through rigorous evaluation of various machine learning algorithms, I identified promising alternatives to logistic regression, such as the highly accurate random forest. By adopting these advanced algorithms, we aim tofurther improve the precision and overall performance of our decision- making processes. This strategic shift towards more accurate models will undoubtedly strengthen Bright Money's position in detecting the default customers and reduce the risk.

**KEY LEARNINGS:**

- I have learned how primary data is handled from the stage of datapreparation to till finalizing the model.
- I have learned about various machine learning algorithms likeLogistic regression and their metrics to evaluate.
- I have learned Hands-on Application of Data Science Techniques
- i have also learned about the Financial Domain Knowledge:

# CHAPTER VI

# REFERENCES

- https://app.mavenanalytics.io/
- https://medium.com/@tobias_35452/in-need-for-both-accuracy-and-interpretability-give-probabilistic-rules-a-try-6713a74c7776

# CORPORATE MENTOR FEEDBACK FORM

**JULY 2023**

Dear Sir/ Madam,

**Greetings from SBM, CHRIST (Deemed to be University)!**

Thank you for giving the opportunity of doing summer internship to our students in your esteemed organization.

Request you to please complete this evaluation at the end of the student's work period. The evaluation is a mechanism that the **School of Business and Management** adopts as part of its continuous improvement program; therefore, it plays a pivotal role in the overall development of the student.

Request you to give an open and honest feedback based on your assessment of the student and his or her work. To improve the efficacy of the feedback process, request you not share your feedback with the student under evaluation.

You may please email the filled in feedback form to the academic mentor of the student. The student will provide the mentor's email id to you.

| | |
|---|---|
| **Student's Name** | **Gayatri Ramani Bommisetty** |
| **Date of commencement of SIP** | **MAY 15 2023** |
| **Date of completion of SIP** | **JULY 5 2023** |
| **SIP title** | **Predicting Loan Eligibility Of Bright's Potential Customers: A Comparative Machine Learning Approach** |
| **Corporate Mentor Name** | **Ranjitha BB** |
| **Designation** | **Student Intern** |
| **Contact No of Corporate Mentor** | **9686682855** |
| **Name of the Organization** | **Bright Money** |
| **Address of the Organization** | **HSR Layout, Bengaluru 560035** |

**PART I**

Please use the scale below to evaluate your intern's performance in the following areas:

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Needs more training or education | Performing below expectations | Acceptable performance | Above average performance | Superior performanc e |

| | | |
|---|---|---|
| 1 | **General Workplace Performance** | Provide your rating (1-5) |
| | Attendance | 5 |
| | Punctuality | 5 |
| | Appropriate dress code | 5 |
| | Attitude towards work | 5 |
| | Acceptance of criticism | 5 |
| | Asks appropriate questions | 5 |
| | Self-motivated | 5 |
| | Interaction with peers, supervisors, customers | 5 |
| 2 | **Specific Job Assignment Performance** | |
| | Sufficient knowledge to perform tasks | 4 |
| | Verbal communication skills | 4 |
| | Written communication skills | 4 |
| | Analytical skills – analyses problems and takes appropriate action | 4 |
| | Uses technical skills required for the position | 4 |
| | Meets deadlines | 4 |
| | Takes initiative to get a job done, including overcoming obstacles | 4 |
| | Sets priorities | 4 |
| | Adaptability to accommodate change and perform a variety of tasks | 4 |
| 3 | **Overall Managerial Performance** | |
| | Problem-solving skills | 4 |
| | Analytical and critical thinking abilities | 4 |
| | Value-based Leadership skills | 4 |
| | Awareness of global, economic, legal, and ethical aspects of business | 4 |
| | Team environment creation | 4 |
| | Innovations - mindset | 4 |
| | Research attitude | 4 |
| | Lifelong learning approach | 4 |
| **4** | **OVERALL SCORE** | 108 |

**PART II**

This section gives you the opportunity, as an experienced professional, to make recommendations that would help in the professional development of the student as well as give the **School of Business and Management** some insight into the areas that may need more attention.

1) What do you consider as the major strengths of this intern?

   Excellent Communication skills and they articulate the story of the data set and then identification of business problem statement effectively

2) What areas need improvement?

   Technical Skills, hands on experience and a more comprehensive literature study

3)What would you recommend to make this student better prepared for the workplace? (e.g. courses, activities, skills acquisition, programs)?

   I would recommend to take relevant courses to enhance practical experience through real-world projects

3) Other comments, commendations, or recommendations:

   Participate in Data Science Competitions and Hackathons to further develop problem solving skills

4) Can the student be considered for future projects with your organization (short term live projects)?

   Yes

5) Do you think the student meets the expectations of the organisation to be given a Pre-Placement offer?

   No

# CERTIFICATE OF ONLINE COURSE

Great Learning

CERTIFICATE OF COMPLETION

Presented to

GAYATRI RAMANI BOMMISETTY 2228610

For successfully completing a free online course

Data Visualization With Power BI

Provided by

Great Learning Academy

(On July 2023)

Great Learning

CERTIFICATE OF COMPLETION

Presented to

GAYATRI RAMANI BOMMISETTY 2228610

For successfully completing a free online course

Excel for Beginners

Provided by

Great Learning Academy

(On July 2023)

# INTERSHIP CERTIFICATE:

DocuSign Envelope ID: EA4C29E0-8F63-4FB8-9290-E52122B821C4

bright

## Internship Certificate

**TO WHOMSOEVER IT MAY CONCERN**

This is to certify that **Ms. Gayatri Ramani Bommisetty** worked under the guidance of Varun Kumar Modi on **"Predicting Loan Eligibility of Brightmoney's Potential Customers: A Comparative Machine Learning Approach"** project for the academic purpose.

We wish her success in her future endeavors.

Yours Sincerely,

DocuSigned by:

01C29C8FC52448A

Varun Kumar Modi
(CTO & Co Founder)
Bright Money Technology Private Limited

**Bright Money Technology Private Limited**
No. 3, 1st Floor, 9th Main Rd,
Shivanagar, Rajajinagar,
Bengaluru, Karnataka 560010

Website: www.brightmoney.co
Contact no : +919880429295
CIN no : U72900KA2019FTC126843

## PANEL FEEDBACK FORM:

**SUMMER INTERNSHIP PROJECT – VIVA Panel Member Feedback**

Name of student: Gayatri Ramani Bommisetty

Reg. No: 2228610

Title of SIP: Predicting Loan Eligibility of Bright's potential Customers: A Comparative Machine learning Approach.

Constructive/productive aspects of the internship project and viva:

Application of Machine Learning to solve a significant industry problem.
Use of /exploring multiple techniques is appreciated.

Areas of improvement:

Model evaluation with multiple splits can be highlighted and assessed.
Logic of columns/variable drop from business angle needs to brought in.
107 → 21 → funnel

Signature of panel member:

Name of panel member: LAKSHMI SHANKAR IYER

Date: