**MINOR PROJECT**



**Project Report On Predicting Loan Defaulters Using Machine Learning**

**Under the able guidance of: Submitted by:**

**Mrs. Kritika Harshita Kaur Chugh**

**Ishika Bansal**

**Akshay Goel**

**Yuvraj Singh Rawat**

DECLARATION

We, Ishika Bansal, Yuvraj Singh Rawat, Harshita Kaur Chugh and Akshay Goel, pursuing B.Tech, from Guru Tegh Bahadur Institute of Technology, GGSIPU, hereby declare that the work presented in this dissertation is the outcome of my own work is the correct to the best of my knowledge. The work presented does not infringe any patented work and has not been submitted to any other university or anywhere else for the award of degree or any professional diploma.

**(Student signature) (Student signature)**

**Ishika Bansal Akshay Goel**

**(Student signature) (Student signature)**

**Harshita Kaur Chugh Yuvraj Singh Rawat**

Certified that the above statement made by the student is correct to the best our knowledge and belief.

**Kritika**

**Assistant professor**

**CERTIFICATE**

This is to certify that, Yuvraj Singh Rawat, Ishika Bansal, Akshay Goel and Harshita Kaur Chugh, pursuing B.Tech, from Guru Tegh Bahadur Institute of Technology, GGSIPU, have undergone project from AUGUST TO NOVEMBER 2024 . They have successfully completed the project on “**Predicting Loan Defaulters Using Machine Learning for Enhanced Risk Management**” under the supervision of Mrs. KRITIKA(ASSISTANT PROFESSOR).

**ACKNOWLEDGMENT**

We Harshita Kaur Chugh, Ishika Bansal, Akshay Goel and Yuvraj Singh Rawat express my deep sense of gratitude and reverence to the esteemed organization, Guru Tegh Bahadur Institute of Technology, for giving us the opportunity to undertake training and providing us with the great infrastructure and facilities. I express my profound gratitude to my mentor for providing guidance and expert supervision for this project, which helped in completing the project in time, without their suggestion and guidance it would have been difficult to complete the assigned task. I am truly thankful for their support and timely help solving problems and queries related to my project.

**Introduction, Types of Loans and their applications**

Loan default prediction has become increasingly important as the demand for digital loans grows, and financial institutions strive to manage risk effectively. Traditional methods of evaluating loan applications often lack efficiency, resulting in increased financial risks and losses due to loan defaults. To address this challenge, machine learning models offer a data-driven and automated solution for predicting loan defaults. By leveraging borrower information, transaction history, and other relevant features, these models can help financial institutions make informed lending decisions and reduce default rates. This approach not only optimizes risk management but also enhances the overall efficiency of the loan approval process.

Machine learning models such as Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machines (SVM) are utilized in developing predictive systems. These models can learn patterns from historical data and accurately forecast the likelihood of a loan applicant defaulting. By implementing these advanced models, the financial sector can transform its risk assessment strategies, ensuring higher accuracy and reliability. This paper explores the use of machine learning techniques in predicting loan defaults, evaluates the models’ performance, and identifies the most effective approach for enhancing the decision-making process.

Loans come in various forms, each designed to meet specific financial needs. Below are the major types of loans and their applications:

1. **Personal Loans**  
   These are unsecured loans provided to individuals based on their creditworthiness. Personal loans can be used for various purposes, such as consolidating debt, medical expenses, home improvement, or unexpected expenses. They often have flexible repayment terms, but interest rates may vary based on the borrower's credit score.
2. **Home Loans (Mortgages)**  
   Home loans are long-term loans provided for purchasing or constructing a home. They are secured by the property itself, and the borrower repays the loan in installments over an extended period. Mortgage loans typically have lower interest rates due to the security provided by the property.
3. **Auto Loans**  
   These loans are specifically designed for purchasing vehicles. Auto loans can be secured or unsecured, with the vehicle acting as collateral in secured loans. They have relatively shorter repayment terms compared to mortgages and often come with competitive interest rates.
4. **Student Loans**  
   Student loans are aimed at financing higher education. These loans may be offered by governments or private financial institutions. They usually have lower interest rates and flexible repayment plans that accommodate the financial situation of students, including deferred payments until after graduation.
5. **Business Loans**  
   Business loans are provided to entrepreneurs and businesses to fund operations, expansion, or new projects. These loans can be secured (e.g., with business assets) or unsecured, and they vary in terms of interest rates and repayment schedules depending on the risk assessment of the business.
6. **Credit Card Loans**  
   These are revolving loans where individuals can borrow up to a certain limit using their credit card. They are convenient for short-term expenses but often come with high-interest rates if not repaid promptly.
7. **Payday Loans**  
   Payday loans are short-term loans intended to cover immediate cash needs until the borrower’s next payday. These loans typically have high-interest rates and fees, making them a high-risk option for consumers.
8. **Secured and Unsecured Loans**
   * **Secured Loans**: These loans require collateral, such as property or assets, reducing the lender’s risk. Examples include mortgages and auto loans.
   * **Unsecured Loans**: These loans do not require collateral, relying solely on the borrower’s creditworthiness. Examples include personal and credit card loans.

Each type of loan serves different purposes, and its suitability depends on the borrower’s financial needs and ability to repay. Machine learning models applied in predicting loan defaults must consider the type of loan and borrower characteristics to provide accurate predictions tailored to these diverse loan categories.

**Study Area and Dataset Used**

The study area for this project focuses on predicting loan defaults within a financial institution. The objective is to analyze and predict default risk based on loan applicants’ historical data and financial information. The dataset includes details such as:

* **Loan Information**: Loan amount, type of loan (e.g., personal, auto), loan term, and interest rate.
* **Borrower Information**: Credit score, income level, employment status, and debt-to-income ratio.
* **Historical Data**: Payment history, number of previous defaults (if any), and delinquency status.

The dataset used is sourced from a financial database that comprises anonymized records of borrowers. It includes a wide range of features, such as applicant demographics, loan details, and credit history, essential for training machine learning models to predict default risk accurately.

By using a comprehensive dataset, the study aims to build a robust predictive model capable of assessing various factors contributing to loan default, thus aiding lenders in making informed lending decisions.

**Features inside Dataset Used**

## The dataset used for predicting loan defaulters includes a diverse set of features that capture the financial and demographic profile of loan applicants, as well as information about the loans themselves. These features include:

## Loan Information:

## Loan Amount: The principal amount borrowed by the applicant.

## Interest Rate: The percentage rate charged on the loan amount.

## Loan Term: The duration of the loan in months or years.

## Loan Type: Type of loan (e.g., personal, auto, mortgage).

## Borrower Information:

## Credit Score: A numeric value representing the borrower’s creditworthiness.

## Annual Income: The annual income of the borrower.

## Employment Status: Whether the borrower is employed, self-employed, or unemployed.

## Debt-to-Income Ratio: The ratio of the borrower’s total debt compared to their income.

## Demographic Information:

## Age: The age of the applicant.

## Marital Status: Whether the applicant is single, married, or divorced.

## Dependents: The number of dependents supported by the borrower.

## Loan History:

## Delinquency Status: The number of past delinquencies or missed payments.

## Previous Defaults: The number of previous loan defaults by the applicant.

## Account Balance: The account balance history over time.

## These features collectively provide a comprehensive view of the applicant's financial behavior and risk profile, which helps machine learning models make accurate predictions regarding the likelihood of loan default.

**Methodology**



Figure 1Workflow of the proposed approach

**Source Code**

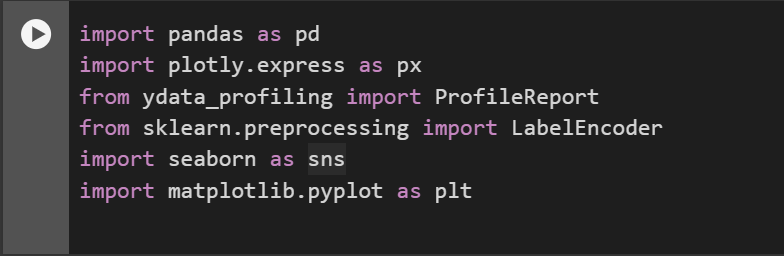


Figure 2 Importing Libraries

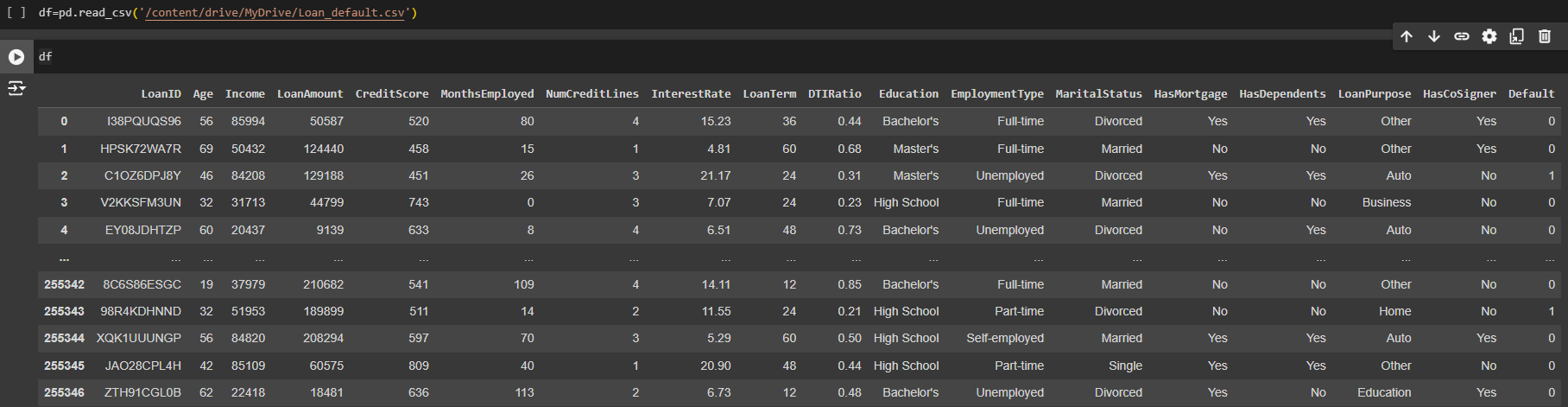


Figure 3 Reading Dataset

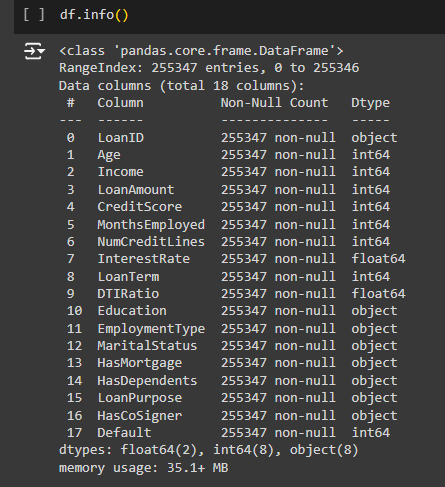


Figure 4 Info on DataFrame

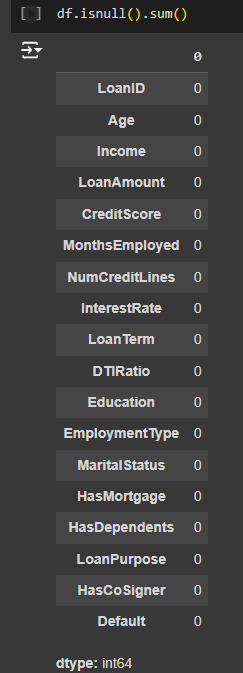


Figure 5 Checking Null Values

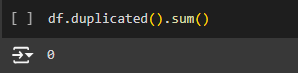


Figure 6 Checking Duplicates

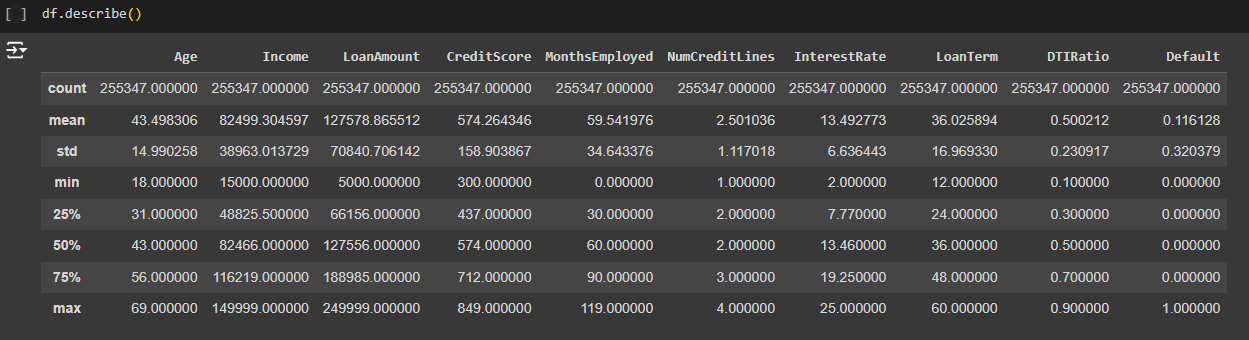


Figure 7 Describing the dataframe

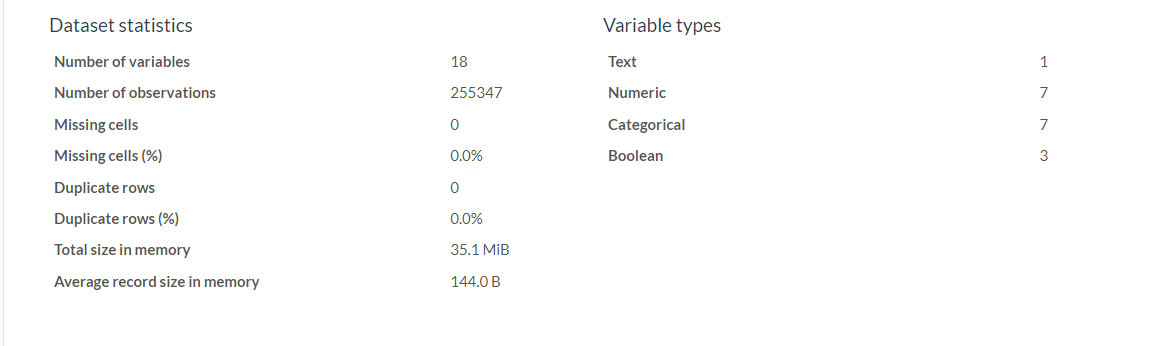


Figure 8 Overview of DataSet

**RESULT**

We focused on developing a model to classify rainfall events using radar data. Our process included:

1. **Data Processing:** We read and converted raw radar files into a usable format using PyART and PyScanF.
2. **Geospatial Analysis:** We plotted the latitude and longitude of observatories to visualize their locations.
3. **Data Extraction:** We extracted reflectivity data corresponding to these observatories and converted dBZ values to a linear scale.
4. **Rain Rate Calculation:** We calculated the rain rate using the Marshall Palmer relationship based on the extracted reflectivity data.
5. **Feature Extraction and Cleaning:** We extracted relevant features and performed data cleaning to prepare the dataset.
6. **Model Training:** We split the dataset into training and testing sets, and applied various machine learning classifiers to predict rainfall events.
7. **Classification and Evaluation:** Our classifiers distinguished between rain (1) and no rain (0) events, and we compared the precision of each model to determine the most accurate one.

This approach allowed us to develop a reliable model for classifying rainfall events, demonstrating our capabilities in data processing, machine learning, and model evaluation.

**REFERENCES**

1. [ScienceDirect Article on Loan Default Prediction](https://www.sciencedirect.com/science/article/abs/pii/S2214785321035148)(2021)
2. [IEOM Society Proceedings - Loan Default Analysis Using ML](https://ieomsociety.org/proceedings/2022orlando/328.pdf)(2022)
3. [Machine Learning Models for Loan Prediction: A Study](https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1319711&dswid=5585)(2019)
4. [Loan Risk Analysis Using Data-Driven ML Models](https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012042/pdf)(2020)
5. [Loan Default Prediction Model - A Research Study](https://www.researchgate.net/publication/368807480_Loan_Default_Prediction_Model)(2023)
6. [Advanced Techniques in Loan Default Prediction](https://www.sciencedirect.com/science/article/pii/S2666764923000218)(2023)
7. [Machine Learning Approaches for Loan Approval Prediction](https://www.researchgate.net/publication/372909313_Prediction_of_Loan_Approval_in_Banks_using_Machine_Learning_Approach)(2023)
8. [Machine Learning Techniques for Loan Default Prediction](https://www.doria.fi/handle/10024/182846)(2021)
9. [Loan Default Risk Prediction Using Machine Learning](https://sifisheriessciences.com/journal/index.php/journal/article/view/415/398)(2023)
10. [A Comparative Study of Loan Default Prediction Techniques](https://sifisheriessciences.com/journal/index.php/journal/article/view/1647)(2023)