Smart Loan Approval

Forecasting

**Submitted by:**

|  |  |
| --- | --- |
| Madhav A Das | AM.EN.U4AIE22131 |
| Sanvia saj ps | AM.EN.U4AIE22163 |
| P Abhiram Reddy | AM.EN.U4AIE22162 |
| Suryanarayan | AM.EN.U4AIE22130 |

ABSTRACT

Informal Description:

In Loan Approval prediction, the aim is to fit a model with the capability to find whether the loan applications will be approved or not using a supervised technique. We will be able to achieve this by making use of the available datasets, feeding them to the model and making it learn the patterns inside it. We will also evaluate using multiple metrics.

Formal Description:

* Task: Our task is to predict whether the loan will be approved (positive) or denied (negative).
* Experience: We will train our model on the available datasets that have many features like income, loan amount, and credit score. etc., related to loan application data along with the loan approval outcome (positive or negative).
* Performance: We measure the performance of our model using multiple metrics like accuracy, precision, and f1-score, which helps us know how well our model will perform on new unseen data.

Assumptions:-

* **Representative Data**: The assumption is that the dataset used for training our model represents the real-world data. And we are also assuming that nobody messed around with the data randomly.
* **Independence of Observations**: Assuming that the loan applications are independent of each other, which means that the approval or denial of an application doesn’t show any impact on other applications.
* **Static Approval Criteria:** Assuming that the criteria followed by the lenders to approve or deny a loan remain constant and don’t change over time.

INTRODUCTION

Motivation**:-**

The main motivation behind choosing this project is our curiosity to learn and understand the process of creating a machine-learning model used for classification.

These days loans have become common for individuals whether it be either for education or business. When we apply for a loan, it takes 3 or more days to get approval for it, and most of us are uncertain whether it will be approved or not. So our motive is to help customers know faster about the status of their loan application or help the bank officials to make decisions faster using this kind of model.

Benefits of solution:-

* Helps to make appropriate decisions based on the output, which means reducing the chance of approving high-risk applications and improving the chance of approving low-risk applicants.
* Applicants can receive faster feedback about their loan approval it frees them from uncertainty or unnecessary anxiety.
* Even if the number of applicants gradually increases there will be no compromise in performance since it’s a machine learning model.
* Use of a machine learning model helps in making the output consistent, reducing the chance of human bias or any other inequality

Solution use:-

This loan approval prediction model can be integrated with the available loan processing system so that when a loan application is sent it automatically takes the data into the model and be able to predict the approval status. Using this, applicants can receive faster responses about their loan status. It will also help in saving time by reducing the need for manual work.

Dataset finalization

**Dataset 1**

Source link: [Kaggle dataset - 1](https://www.kaggle.com/datasets/ninzaami/loan-predication)

Dataset description: The dataset contains information relevant to loan approval

Features and their description:

**Loan\_ID**: It’s a unique id value for each loan application

**Gender**: The gender of the applicant

**Married**: Marital status of the applicant

**Dependents**: Number of individuals financially dependent on the applicant

**Education**: It’s a feature informing whether the applicant is a graduate or not

**Self\_Employed**: Feature that informs whether the applicant is self-employed or not

**ApplicantIncome**: Income of the applicant

**CoApplicantIncome**: Income of an individual who applies for a loan jointly with the primary applicant

**LoanAmount**: Amount of the loan requested by the applicant

**Loan\_Amount\_Term**: The length of time over which a loan is scheduled to be repaid

**Credit\_History**: Refers to a record of an individual's or entity's borrowing and repayment behaviour.

**Proprty\_Area**: Type of area where a property is situated(urban or rural)

**Loan\_Status**: This is our TARGET VARIABLE informing whether the loan is approved or not.

**Dataset 2:**

Source link: [Kaggle dataset - 2](https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset)

Dataset description: The dataset contains information relevant to loan approval

Features and their description:

**loan\_id**: It’s a unique id values for each loan application

**no\_of\_dependents**: Number of individuals financially dependent on the applicant

**education**: It’s a feature informing whether the applicant is a graduate or not

**self\_employed**: Feature that informs whether the applicant is self-employed or not

**income\_annum**: Income of the applicant

**loan\_amount**: Amount of the loan requested by the applicant

**loan\_term**: The length of time over which a loan is scheduled to be repaid

**cibil\_score**: Refers to a credit score generated by the Credit Information Bureau Limited

**residential\_assets\_value:** Estimated value of residential properties owned by the applicant

**commercial\_assets\_value:** Estimated value of commercial properties owned by applicant

**luxury\_assets\_value:** Estimated value of luxurious assets properties owned by applicant

**bank\_asset\_value:** Refers to the total value of assets held by a bank

**loan\_status:** This is our TARGET VARIABLE informing whether the loan is approved or not.

**Dataset 3:**

Source link: [Kaggle dataset - 3](https://www.kaggle.com/datasets/chilledwanker/loan-approval-prediction)

Dataset description: The dataset contains information relevant to loan approval

Features and their description:

**person\_age:** Age of the applicant

**person\_income:** Income of the applicant

**person\_home\_ownership:** Refers to the homeownership status of the applicant

**person\_emp\_length:** Refers to the length of time an individual has been employed

**loan\_intent:** Refers to the purpose or intention behind the loan application.

**loan\_grade:** Refers to the grade which indicates the risk level of the loan

**loan\_amnt:** Amount of the loan requested by the applicant

**loan\_int\_rate:** Refers to the interest rate associated with the loan

**loan\_status:** This is our TARGET VARIABLE informing whether the loan is approved or not.

**loan\_percent\_income:** represents the ratio of the loan amount applied by the applicant to the applicant’s income (in %)

**cb\_person\_default\_on\_file:** Indicates whether the person has a record of defaulting on a loan according to their credit bureau file

**cb\_person\_cred\_hist\_length**: Refers to the length of the credit history for the individual, as recorded in their credit bureau file