Loan Application Status Prediction

Introduction Loan Prediction Problem

Welcome to this article on Loan Prediction Problem. Below is a brief introduction to this topic to get you acquainted with what you will be learning.

The Objective of the Article

This article is designed for people who want to solve binary classification problems using [Python](https://www.analyticsvidhya.com/blog/2022/05/working-with-dynamodb-in-python-using-boto3/). By the end of this article, you will have the necessary skills and techniques required to solve such problems. This article provides you with sufficient theory and practice knowledge to hone your skills.

**Expectations from the Article**

. Introduction to the problem

1. Problem Definition  
   2.  Data Analysis  
   3.  EDA Concluding Remarks  
   4.  Pre-processing Pipeline  
   5.  Building Machine Learning Models  
   6.  Concluding Remarks

INTRODUCTION - A loan is the core business part of banks. The main portion the bank’s profit is directly come

from the profit earned from the loans. Though bank approves loan after a regress process of verification and testimonial but still there's no surety whether the chosen hopeful is the right hopeful or not. This process takes fresh time while doing it manually. We can prophesy whether that particular hopeful is safe or not and the whole process of testimonial is automated by machine literacy style. Loan Prognostic is really helpful for retainer of banks as well as for the hopeful also

Problem Definition-

Bank employees check the details of applicant manually and give the loan to eligible applicant. Checking the details of all applicants takes lot of time. The artificial neural network model for predict the credit risk of a bank. The Feed- forward back propagation neural network is used to forecast the credit default. The method in which two or more classifiers are combined together to produce a ensemble model for the better prediction. They used the bagging and boosting techniques and then used random forest technique. The process of classifiers is to improve the performance of the data and it gives better efficiency. In this work, the authors describe various ensemble techniques for binary classification and also for multi class classification. The new technique that is described by the authors for ensemble is COB which gives effective performance of classification but it also compromised with noise and outlier data of classification. Finally they concluded that the ensemble based algorithm improves the results for training data set Drawback of Existing System Checking details of all applicants consumes lot of time and efforts. There is chances of human error may occur due checking all details manually.

2.DATA ANALYSIS-

**Collect data on loan applications including information about the: borrower, loan amount & term, credit score, income, employment status, and other factors.**

**Exploratory Data Analysis (EDA)-**

* **Perform descriptive statistics to understand the distribution and correlation of the variables**
* **Use visualization tools to explore the relationship between the variables and the target variable (Loan\_Status: 1|0)**

**Pre-processing Pipeline-**

* Handling missing values, outliers, errors
* Dropping unnecessary variables (Loan\_ID)
* Encoding categorical variables (2 ways: 2-label encoding and one-hot encoding
* Split the data into training (80%) and validation|test sets (20%)
* Train the model using the training data
* elect an appropriate machine learning algorithm to build the loan approval prediction model
* Train various machine learning models, using the training set
* Evaluate their performance on the testing set

.  Building Machine Learning Models

* Evaluate the performance of the model using the validation data
* Evaluate various metrics (accuracy, precision, recall, F1 score, ROC-AUC)
* Choose the best-performing model
* Tune its hyperparameters (k-fold cross validation, validation set approach)
* Evaluate the final model on the testing set
* Report its performance metrics
* Deploy the model in a production environment
* Use it to predict loan approval decisions
* This will allow the bank to use the model in a web application or integrate it into its existing systems

**Variables:**

**Numerical variables**

1. ApplicantIncome: Loan applicant income
2. CoapplicantIncome: Income of applicant’s spouse, if applicable
3. LoanAmount: Amount of loan (in thousands)
4. Loan\_Amount\_Term: Length of loan (in months)
5. Dependents: Number of dependents of the applicant
6. Credit History: Credit history of individual’s repayment of their debts

**Categorical variables**

1. Loan\_ID: Unique identificator used for anonymity
2. Loan\_Status: Was loan approved or not? - “Y” / ”N”
3. Gender: “Male” / “Female”
4. Married: Is applicant married? - “Yes” / “No”
5. Education: “Graduate” / “Not Graduate”
6. Self\_Employed: Is applicant self-employed? - “Yes” / “No”
7. Property\_Area: Area of property - “Urban” / “Semiurban” / “Rural”