

Loan Eligibility Prediction

Machine Learning

Introduction

Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation.

However, they still don't have assurance if the applicant is able to repay the loan with no difficulties. In this tutorial, we'll build a predictive model to predict if an applicant is able to repay the lending company or not. We will prepare the data using Jupyter Notebook and use various models to predict the target variable



Language

We will be using Python for this course along with the below-listed libraries

1. PYTHON
2. PANDAS
3. SEABORN
4. SKLEARN





DATA

For this problem, we have three CSV files : train ,test ,and sample submission

1. Train File : will be used for training the model, i.e. our model will learn from this file. It contains all the independent variables and the target variable
2. Test file : contains all the independent variables, but not the target variable for the test data
3. Sample submission file : Contains the format in which we have to submit our predictions

Model building

1. Logistic Regression : Logistic regression is a classification algorithm. It is used to predict a binary outcome (1/0, Yes/No, True/False) given a set of independent variables.
2. Decision Tree: Decision tree is a type of supervised learning algorithm that is mostly used in classification problems. Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes
3. Random Forest: Random forest is a tree-based bootstrapping algorithm wherein a certain no.of weak decision trees are combined to make a powerful prediction model

Applications

finance company deals in all loans. They have presence across all urban, semi urban and rural area. Customer first for a loan after that company validates the customer eligibility for loan

The company wants to automate the loan eligibility process based on customer detail provided while filling online application form. The details are gender, marital status, education, number of dependents, income, loan credit history and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers.



CONCLUSION

1. In this project we've gone through a good portion of the data science pipe such as processing and modeling and we've used essential classification models such as Logistic regression , Decision tree and Random forests
2. It includes mathematical techniques and processes that are applied to historical data to study correlations , identify trends and predict possibel outcomes by quantifying the uncertainty and the characteristics of the variation

Thank
you!

