Loan Eligibility Prediction in machine learning

Tarun Kumar Arcot(1NT18CS174), Rakshith R (1NT19CS413)

Abstract

A loan eligibility prediction using decision tree and Naive Bayes algorithms

Introduction

Machine learning (ML) is **a type of artificial intelligence (AI)** that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. The proposal is to predict whether the specified person will be paying the loan or not. The machine learning methods are Decision tree and Naive Bayes algorithms to predict the loan eligibility criterion,

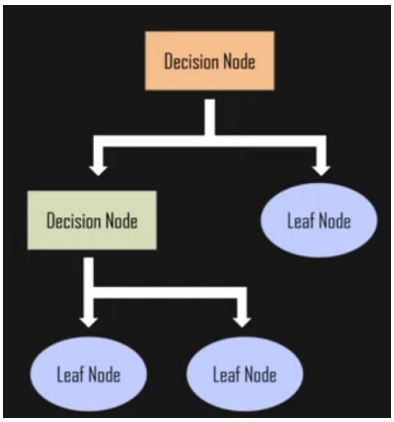
Data Set

There are two types of data sets used in the project 1)training data set and 2)testing data set. The data set includes:-
1) Loan_ID 2) Gender 3) Married 4) Dependents 5) Education 6) Self_Employed 7) ApplicantIncome 8) CoapplicantIncome 9) LoanAmount 10) Loan_Amount_Term 11) Credit history 12) Property_area 13) Loan_Status

Machine Learning Methods

1) Decision tree:-

A **Decision Tree** has many analogies in real life and turns out, it has influenced a wide area of **Machine Learning**, covering both **Classification** and **Regression**. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. But in this we will use decision tree for classification.



2) Naive Bayes:-

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems. It is mainly used in *text classification* that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**. Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

Assessment:-

1) k-fold cross-validation
2)Holdout methodology

Discuss the assessment methodology you will use to validate the models.

Presentation and Visualization

Will be using graphs to show the presentation

Bibliography

[1] https://en.wikipedia.org/wiki/Decision_tree

[2] https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[3] https://www.coursera.org/lecture/big-data-machine-learning/using-a-validation-set-Pb8Cl

[4] https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset

[5] https://medium.com/devcareers/loan-prediction-using-selected-machine-learning-algorithms-1bdc00717631