

Project Proposal: Loan Default Prediction Using KNN and K means

Introduction:

The aim of this project is to develop a machine learning model to predict loan defaults based on a dataset from Kaggle. Here is the link:

<https://www.kaggle.com/datasets/hemanthsai7/loandefault?select=train.csv>

The dataset contains information on loan applicants, including demographic information, credit history, and other relevant factors. We will use two machine learning algorithms, KNN, and K-means, to analyze the dataset. We will use sklearn package in Python to implement these two algorithms. The tool we use is Jupyter Notebook.

Data Cleaning and Preprocessing:

The first step of the project is to clean and preprocess the dataset. This includes removing any missing values, normalizing the data, and transforming variables as needed. We will also select 5-7 features from the 35 columns to use in our analysis. If two features are strongly linearly dependent, one feature will be dropped.

Machine Learning Algorithms:

Next, we will apply two machine learning algorithms to the preprocessed dataset to predict the loan status. The first algorithm is KNN, which will use the selected features to classify each loan applicant as either default or non-default. The second algorithm is K-means, which also classifies but without target labels.

Evaluation and Comparison:

Once the algorithms have been applied, we will evaluate their performance calculating the rate of precision, recall, and f score. We will draw the ROC curve and see the area under the curve. We will then compare the performance of the KNN and K-means algorithms to determine their relative strengths and weaknesses.

Conclusion:

The results of this project will provide valuable insights into the effectiveness of KNN and K-means algorithms for loan default prediction. This information can be used to improve lending practices and reduce the risk of loan defaults. In addition, the project will demonstrate the value of machine learning in data analytics and its potential to help solve real-world problems.