

Guide for Analyst

This is a technical guide for analysts for the vehicle loan approval system.

- The first step towards this project is data gathering. The data source is Rpubs.com where data is fetched.
- The data is stored in .CSV file on which data preprocessing steps is performed.
- The P-value hypothesis test is done on the data to support or ignore the null value hypothesis.
- Data cleaning is also performed on the dataset to remove the duplicate values.
- A missing value is also filled using the average value.
- Label encoding is also performed on the data which is used to transforming non-numerical value into its equivalent numerical data.
- For testing the model several models are used and based on the accuracy the model is selected.
 - Naïve bayes: 78% accuracy and biased
 - Random forest: 76% accuracy and biased
 - Decision tree: 66% accuracy and non-biased
 - KNN: 82% accuracy and non-biased.
- So the K nearest neighbor model is the most suitable model for the project.