

Project Proposal

Question/needs:

This data was collected to analyze and study the issue of the client's failure to pay the installments as a result of a sudden change in his financial situation or mismanagement of benefiting from financing and the failure to allocate part of the income to pay the installments.

Data description:

This data for this project is available at Kaggle by link:

<https://www.kaggle.com/zaurbegiev/my-dataset>

The dataset consists of 10000 rows and 18 columns.

A brief definition of the columns is given below:

1. Loan id	Identification number given to each loan.
2. Customer Id	Identification number given to each customer.
3. Current Loan Amount	Contains the current loan number.
4. Term	Contain short term or long term.
5. Credit Score	Balance number.
6. Annual Income	Balance number during the year.
7. Years in current job	Number of years.
8. Home Ownership	Contains housing type.
9. Purpose	Contains the type of purpose.
10. Monthly Debt	Contains the debt amount.
11. Years of Credit History	Contains the number of years.
12. Months since last delinquent	Number of stuck months.
13. Number of Open Accounts	Contains the numeric.

14. Number of Credit Problems	Contains the numeric.
15. Current Credit Balance	Contains the numeric.
16. Maximum Open Credit	Contains the numeric.
17. Bankruptcies	Contains the numeric.
18. Tax Liens	Contains the numeric.

Tools:

The first step to visualize the dataset in this project is to create a correlation matrix to study the numeric data relationship. I will be planning to use pandas and numpy library for visualization the data. also I will use Logistic regression, Decision Tree and Random Forest if I need it. those are the tools I can think of in the begin. However, going through the camp I will come up with model approaches will be help me.