

Project Report: Loan Approval Prediction using ML

1. Introduction

This project involves the development of two machine learning models for loan approval prediction based on customer demographic and financial information. The goal is to predict whether a loan will be approved based on the given features using both a Self Organizing Map (SOM) and a Classification-based Model.

2. Dataset Overview

The dataset consists of 690 entries and 16 columns, primarily made up of numerical data, including information such as customer demographics, loan-related features, and a target class (Class), which likely represents whether a loan was approved or not. Here's a quick breakdown:

- **CustomerID:** Unique identifier for each customer.
- **Gender, Marital Status, Employment Status, Job Security, Credit History, Personal Loan:** Categorical features represented as integers.
- **Age, Income(LPA), Current Loan Amount, Number of Dependents, Number of Existing Loans, Years at Current Residence, Current Bank Balance, Collateral Value:** Numerical features.
- **Class:** The target variable indicating loan approval status.

3. Notebook 1: Loan Approval System using Self Organizing Map (SOM)

The first notebook, titled "Loan Approval System using Self Organizing Map (SOM)", begins with:

- **Introduction:** The aim is to build a loan approval system using a Self Organizing Map (SOM), a type of unsupervised learning.
- **Importing Libraries:** Standard libraries like numpy, pandas, and matplotlib are used.
- **Data Import:** The dataset is loaded from the CSV file, and features (X) and target (y) are extracted.

4. Notebook 2: Loan Approval using Classification

The second notebook, "Loan Approval using Classification.ipynb", focuses on classification-based modeling for loan approval and includes the following steps:

- **Library Imports:** Similar to the first notebook, it imports pandas, numpy, and matplotlib.
- **Dataset Import:** The dataset is loaded, and features (X) and the target variable (y) are defined.
- **Modeling:** The notebook likely builds a machine learning model for classification, though further details on model training and evaluation haven't been inspected yet.

5. Project Report Summary

Introduction: This project involves the development of two machine learning models for loan approval prediction based on customer demographic and financial information. The goal is to predict whether a loan will be approved based on the given features using both a Self Organizing Map (SOM) and a Classification-based Model.

Dataset Overview: The dataset contains 690 records with 16 features, including both customer demographics and financial indicators. The features are:

- **CustomerID:** Unique identifier.
- **Gender, Marital Status, Employment Status, Job Security, Credit History, Personal Loan:** Categorical features encoded as integers.
- **Age, Income(LPA), Current Loan Amount, Number of Dependents, Number of Existing Loans, Years at Current Residence, Current Bank Balance, Collateral Value:** Continuous numerical features.
- **Class:** The target variable indicating loan approval status (0 for disapproval, 1 for approval).

6. Methodology

3.1 Self Organizing Map (SOM) Model

Objective: Use unsupervised learning to identify patterns and group customers based on their loan approval likelihood.

Libraries Used: numpy, pandas, matplotlib.

Process:

- Data is loaded and split into features (X) and target (y).
- A Self Organizing Map is applied to the data to cluster similar data points, identifying customers with similar financial profiles.

3.2 Classification-based Model

Objective: Use supervised learning to build a model that can classify whether a loan will be approved or not.

Libraries Used: pandas, numpy, matplotlib.

Process:

- Data is loaded and split into features (X) and target (y).
- Feature preprocessing, model training, and evaluation using a classification algorithm (specific details of the algorithm are in the notebook).
- Prediction of loan approval status based on input features.

7. Results

SOM Model: Clusters the customers based on their financial patterns but is unsupervised, providing insights into potential loan approval groupings.

Classification Model: Likely trained using a decision tree, logistic regression, or a similar supervised model to classify loan approvals. Performance metrics (accuracy, precision, recall) will be detailed in the notebook.

8. Conclusion

Two approaches were explored for loan approval prediction:

- **Unsupervised Learning (SOM)** for identifying patterns in customer profiles.
- **Supervised Classification** for predicting loan approval with a focus on high accuracy and interpretability.

Further optimization of these models could involve feature engineering, model tuning, and incorporating additional evaluation metrics to improve predictions.