Loan Status Prediction Documentation

Project Overview

This notebook builds a machine learning model to predict whether a loan application will be approved Y or not N. The dataset contains applicant details such as gender marital status income education credit history and property area. The workflow includes data preprocessing categorical encoding training and evaluating the classifier.

Contents and Notebook Flow

1 Importing Libraries

2 Loading and Exploring the Dataset

3 Data Cleaning Handling Missing Values

4 Encoding Categorical Variables

5 Splitting Data into Training and Testing Sets

6 Model Training SVM Model Only

7 Model Evaluation Accuracy Precision Recall

8 Predicting Loan Status for New Applicants

Explanation of Libraries Used

import numpy as np numerical computations

import pandas as pd data manipulation

from sklearn.model_selection import train_test_split splitting data

from sklearn.preprocessing import LabelEncoder encoding categorical variables

from sklearn import svm SVM classifier

from sklearn.metrics import accuracy_score accuracy evaluation

Data Preprocessing Steps

Checked and filled missing values using mean or mode.

Converted categorical columns such as Gender Married Education Self Employed Property Area and Loan Status into numeric values using replace or LabelEncoder.

Split dataset into feature matrix X and target Y where Y equals Loan Status.

Ensured correct input shape for model training and prediction.

Model Training

This project uses Support Vector Machine SVM as the primary and only classification model.

SVM Training Code

classifier = svm.SVC kernel equals linear

classifier.fit X train Y train

Model Evaluation

The trained SVM model was evaluated using the test dataset.

Y pred = classifier.predict X test

accuracy = accuracy_score Y test Y pred

Common Errors and Fixes

1 SettingWithCopyWarning or KeyError

Fix using .loc

loan_dataset.loc Colon Loan Status equals loan_dataset Loan_Status dot replace N 0 Y 1

2 ValueError in predict

Ensure input shape is correct.

x new = X test.iloc double bracket 0 double bracket

classifier.predict x new

SVM Concept Summary

SVM identifies a hyperplane that best separates two classes with maximum margin.

Key Concepts

Hyperplane decision boundary dividing classes

Margin distance between hyperplane and nearest points

Support Vectors critical points defining the margin

Project Outcome

Built an SVM based loan approval prediction model.

Completed preprocessing encoding training and evaluation.

Achieved strong accuracy with linear SVM kernel.

Future Improvements

Use cross validation for better model reliability.

Evaluate ensemble models like Random Forest or Gradient Boosting.

Tune hyperparameters with GridSearchCV.

Deploy using Flask or Streamlit for user interaction.

Author

Satyam Gajjar