Loan Approval Prediction using Machine Learning

1. Aim

To build, optimize, and evaluate multiple classification models to predict whether a loan application will be Approved or Rejected using supervised machine learning techniques.

2. Sample Dataset (Attributes Description)

- Gender – Male/Female
- Married – Yes/No
- Education – Graduate/Not Graduate
- Self_Employed – Yes/No
- ApplicantIncome – Applicant income (numeric)
- CoapplicantIncome – Co-applicant income (numeric)
- LoanAmount – Loan amount (numeric)
- Loan_Amount_Term – Loan repayment term (numeric)
- Credit_History – 1 (Good), 0 (Bad)
- Property_Area – Urban/Semiurban/Rural
- Loan_Status – Approved/Rejected (Target Variable)

3. Tasks to Perform

1. Part A – Data Preprocessing: Handle missing values, encode categorical variables, create TotalIncome feature, normalize numerical features, encode target variable.
2. Part B – Exploratory Data Analysis: Generate correlation heatmap and analyze impact of Credit_History.
3. Part C – Model Building: Train Logistic Regression, Random Forest, and SVM using 70-30 split and 5-fold cross-validation.
4. Part D – Hyperparameter Tuning: Perform GridSearchCV for Random Forest (n_estimators, max_depth) and SVM (C, kernel).
5. Part E – Model Evaluation: Compute Accuracy, Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC.
6. Part F – Prediction Task: Predict loan approval for a new applicant.

1.
2. ## 2. Medical Insurance Claim Fraud Detection using Machine Learning

3. ### Aim

   To develop and evaluate multiple machine learning models to detect fraudulent insurance claims, handle class imbalance, perform feature selection, and interpret model predictions.

Tasks to perform

1. Part A – Data Preprocessing:
2. Part B - Class Imbalance Handling- Check class distribution.
3. Part C – Model Building - Logistic Regression - Random Forest - XGBoost (if available) - Support Vector Machine using Train-Test Split (80-20) 5-Fold Cross Validation
4. Part D – Hyperparameter Tuning: Perform GridSearchCV for Random Forest (n_estimators, max_depth) and SVM (C, kernel).
5. Part E – Model Evaluation: Compute Accuracy, Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC, PR Curve (important for imbalanced data).
6. Part F – Plot Feature importance, Identify top 3 fraud indicators.
7. Prediction Task: Predict loan approval for a new applicant
8. Explain why Claim_Amount and Previous_Claims may influence fraud

Predict fraud possibility for

Age = 52
Gender = Male
Policy_Tenure = 6
Claim_Amount = 280000
Hospital_Type = Private
Previous_Claims = 4
Chronic_Disease = Yes
Claim_Type = Surgery