Project 2: Loan Default Risk Prediction System

Tools

Python | Statistics | ML | Deployment (Streamlit/FastAPI)

Business Problem

A fintech company wants to build a system that predicts whether a loan applicant is likely to default or not based on financial, demographic, and credit history features. The goal is to minimize loan default risk using a predictive model and deploy it for business users to use during customer onboarding.

Key Business Objectives

- • Identify the most significant factors affecting loan defaults
- • Build a classification model to predict loan status (Default/No Default)
- • Optimize model using statistical testing, feature selection, and hyperparameter tuning
- • Deploy the model with an interactive front end (Streamlit or FastAPI) for entering customer data and viewing predictions

Steps to Follow

- EDA & Statistical Summary
    - - Outlier treatment, normality tests, correlation, multicollinearity checks
- Feature Engineering
    - - Binning, encoding, handling imbalance
- Model Building
    - - Try Logistic Regression, Random Forest, and XGBoost
    - - Use GridSearchCV
- Evaluation
    - - AUC-ROC, Precision, Recall, Confusion Matrix
- Deployment
    - - Build UI using Streamlit or FastAPI
    - - Host the app on Render or Heroku
    - - Include a model interpretation section using SHAP/feature importances

Kaggle Dataset

https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset

Bonus Challenges (for Hard Level)

- • Add a "what-if" scenario engine in the deployed app
- • Perform hypothesis testing (e.g., Do defaulters have statistically lower income?)
- • Build a PDF report generator for predictions

Skills Practiced

- • End-to-end ML pipeline
- • Statistical reasoning
- • Model deployment
- • Dashboard for predictions

Summary Table

Project	Tools Involved	Domain	Skills Covered / Outcome
Sales & Customer Insights	Python, MySQL, Power BI	E-Commerce	Data Cleaning, SQL joins, Power BI dashboards Executive-level BI dashboard
Loan Default Prediction	Python, Statistics, ML, Deployment	FinTech	EDA, ML modeling, Streamlit, Statistics Deployed predictive system with insights