Loan Status Prediction Using Machine Learning and FastAPI Deployment

# Introduction

In modern financial services, automated systems for assessing loan applications are essential. This project addresses the need for a predictive model that determines whether a loan should be approved (Y) or rejected (N) based on key applicant features. The model is intended to optimize the decision-making process in financial institutions, minimizing bias and improving turnaround time.  
  
The solution involves data preprocessing, exploratory data analysis (EDA), supervised machine learning model training and evaluation, and deployment using FastAPI. This end-to-end approach reflects practical, production-ready data science implementation.

# Dataset Overview

The dataset used comprises 381 loan applications, each containing 13 attributes. After handling missing values, the dataset was reduced to 308 rows. The features include applicant income, education level, credit history, and marital status, which are used to predict loan approval status.

# Exploratory Data Analysis (EDA) and Interpretation

Various visualizations revealed insights such as the skewness of applicant income, education-based loan amount differences, property area-based loan distributions, and credit history impact. Urban residents and graduates received higher loan amounts, and credit history had the strongest influence on approval decisions.

# Data Preprocessing

Categorical variables were label-encoded, numerical features were standardized, and five features were selected for training: Married, ApplicantIncome, Education, LoanAmount, and Credit\_History.

# Model Training and Evaluation

Three algorithms were evaluated using GridSearchCV:  
- Logistic Regression: 88.7%  
- K-Nearest Neighbors: 90.3%  
- Support Vector Machine: 90.3% (Best Model with C=0.01, kernel='linear')

# Model Export

The trained SVM model and StandardScaler object were exported using joblib for deployment in the FastAPI application.

# FastAPI Deployment

The API exposes a /predict/ endpoint that accepts JSON input for prediction. The model returns 1 for approved and 0 for rejected loans. The app is accessible through the interactive Swagger UI at http://127.0.0.1:8000/docs.

# Key Takeaways

Credit History is the most significant factor for loan approval. Education and marital status influence approval rates and loan amounts. The SVM model proved effective and was deployed using FastAPI.

# Conclusion

This project showcases a robust end-to-end machine learning pipeline and API deployment. It reflects the real-world application of data science in the FinTech domain, suitable for operational deployment or academic presentation.