ISBA 4755 Fall 2024 FINAL SEMESTER PROJECT (Group)

Project: Loan Approval Prediction using Machine Learning and Model Deployment

Business Context: Loan Processing

Loan processing is a critical function in financial institutions, involving the evaluation of an applicant's eligibility for a loan based on various factors such as income, credit history, and financial stability. Accurate and efficient loan approval processes are essential for minimizing risk, improving customer satisfaction, and ensuring regulatory compliance.

How This Application Helps

For Consumers: Predictive loan approval models provide faster decisions, reducing the waiting time for applicants. This improves transparency and allows customers to plan their finances with confidence.

For Businesses: Automating loan approval decisions helps financial institutions streamline operations, reduce costs, and lower the risk of defaults by identifying high-risk applicants more effectively.

This application bridges the gap between customer needs and business objectives, fostering trust and efficiency in loan processing systems.

Project Objective:
The objectives of this Machine Learning model building and local deployment project are:

(1) to develop a binary classification model to predict whether a loan application will be approved based on the applicant's details.
(2) to make the final trained model persistent by storing it as a serialized object (.pkl or .sav object)
(3) to demonstrate real-time prediction capability by allowing it to accept loan application user inputs from a browser and displaying the predicted approval result.

You will go through the entire machine learning workflow, including data exploration and pre-processing, model building, evaluation, and reporting.

Learning Outcomes:
By completing this assignment, you will:
1. Understand the process of preparing data for machine learning models.
2. Apply classification algorithms to solve a real-world problem.
3. Evaluate model performance using appropriate metrics.
4. Deploy the trained model locally integrated with a browser-based UI to demonstrate real-life application of the model

5. Gain hands-on experience in presenting insights and recommendations (bonus)

Dataset Description:
The dataset contains loan application records with the following features:
Number of columns: 13
Number of records: 4269
Column Names:
1. loan_id
2. no_of-dependents
3. education
4. self_employed
5. income_annum
6. loan_amount
7. loan_term
8. cibil_score (credit score)
9. residential_asset_value
10. commercial_asset_value
11. luxury_asset_value
12. bank_asset_value
13. loan_status (Target Variable)

Data File and location:

loan_approval_dataset.csv (384.34 kB)
Download from Brightspace project assignment folder

Guideline Steps:

A. Model Development Program (Python – Jupyter Notebook)

Step 1: Data Preprocessing
1. Load the dataset and inspect for missing values.
2. Handle missing values using appropriate methods (e.g., imputation).
3. Convert categorical variables into numerical ones using encoding techniques.
4. Normalize or standardize numerical features if necessary*.

Step 3: Model Development
1. Split the dataset into training (80%) and testing (20%) sets.
2. Train at least two different binary classification models (e.g., Decision Tree, Random Forest, logistic regression etc.).
3. Compare the performance of the models using the testing set.

Step 4: Model Evaluation
1. Use evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

2. Select the best-performing model based on these metrics. Justify the basis of your final selection.

Step 5: Insights and Interpretability
Identify the most important features contributing to the model's predictions.
1. Research and learn how to generate a ranked order of important features based on "Feature Importance" function available in Scikit Learn.
2. Select TOP EIGHT predictive features based on this ranking
3. Explain the significance of these features in the context of loan approval.

Step 6: Rebuilt the ML Model using the previously determined "best" algorithm (Step 4) and TOP EIGHT features (Step 5.2)

Step 7: Save the "trained model" from Step 6 in .pkl or .sav format on to local drive

B. Model Deployment and User Input Program (Python – <u>not Jupyter Notebook</u>)

Step 1: Use StreamLit functions to render the TOP EIGHT input features and receive user inputs of hypothetical new loan applications, one loan at a time.

Step 2: Load the saved serialized trained model (Step 7) and use the model too generate prediction for the inputted loan data (eight fields)

Step 4: Display the results of loan approval in an user-friendly, business-like meaningful way

C. Write a report summarizing:

   - The preprocessing steps and rationale.
   - Key findings from the data exploration / analysis.
   - Evaluation results and comparison of models.
   - Business insights and recommendations.
   - What new knowledge you learned on your own by doing the project in addition to what was covered in the lecture

 D: Submission

- Submit the following:
  1. Code: A well-documented Jupyter Notebook. The notebook must be fully run with all outputs of the cells. All warning messages need to be addressed and should not be left in the code.
  2. Report: A structured PDF or Word document summarizing the project as described in Section C.

Grading Rubric:

Criteria	Weight
Exploratory Data Analysis	5%
Data Pre-processing	5%
Model Building	25%
Model Evaluation & Selection	5%
Model Deployment Success	25%
User Interface Design Quality	25%
Report Quality and Insights	10%

Deadline:

All submissions are due by 10-Dec-2024. Late submissions will incur a penalty of 10%.

Resources:
1. Scikit-learn Documentation (https://scikit-learn.org/)
2. Matplotlib and Seaborn for Visualization (https://seaborn.pydata.org/)
3. Pandas and Numpy for Data Analysis (https://pandas.pydata.org/)
4. Streamlit documentation (https://docs.streamlit.io/)
5. Streamlit tutorials (https://docs.streamlit.io/develop/tutorials)

For questions or assistance, reach out to me during office hours or via email.

Good luck, and enjoy the project, and don't forget to add it to your resume!