Classification Model from Scratch: Loan Approval Prediction

Please go through the document carefully to understand the requirements and our expectations.

Scope of Work

The goal of this task is to build a **machine learning pipeline** from scratch to predict **loan approval** based on applicant data. The pipeline should include all necessary steps from data preprocessing to model evaluation.

Steps to Include

1. **Data Preprocessing**:
    - Handle missing values
    - Apply encoding to categorical variables
    - Perform feature scaling (standardization or normalization)
2. **Exploratory Data Analysis (EDA)**:
    - Analyze the dataset using descriptive statistics, visualizations, and correlation analysis
3. **Feature Selection/Engineering**:
    - Select important features and/or create new ones to improve model performance
4. **Model Training**:
    - Train at least two models to achieve the same.
5. **Model Evaluation**:
    - Evaluate models using common metrics such as:
        - **Accuracy**
        - **Precision**
        - **Recall**
        - **F1 Score**

Deliverables Expectations

- A **Python script** or **Jupyter notebook** (.py or .ipynb)
- **README.md** file that includes:
    - Summary of your approach
    - Key challenges/decisions
    - Model comparison
- Output files (e.g., test predictions in .csv or .json format)
- A link to the **Kaggle Loan Prediction Dataset** used in your solution

- A Google Drive folder link will be provided to save the following items
    - Code
    - Sample data
    - Any required assets or models

Optional

- Create a **Streamlit app** for easy interaction with the model (to predict loan approval from user input)

Technologies / Tools

You may use any of the following (or similar):

- Python
- Scikit-learn (for preprocessing, model training, and evaluation)
- Pandas (for data manipulation)
- Matplotlib / Seaborn (for EDA and visualizations)
- Streamlit (if creating an interactive app)
- Kaggle Dataset: Loan Prediction Dataset

README Expectations

Ensure your README includes:

- Project title: **Classification Model from Scratch: Loan Approval Prediction**
- Instructions to:
    - Install dependencies
    - Run the code
- Summary of your approach (overview of the pipeline)
- Key challenges and decisions made (e.g., handling missing data, feature engineering techniques, etc.)
- Model comparison results (include a table or description of the model performance)
- A link to the **Kaggle Loan Prediction Dataset** is used

Strictly Avoid

- Submitting AI-generated repositories or README files with no real implementation or understanding.
- Copy-pasting code from ChatGPT, GitHub Copilot, or other tools without customization or comprehension.

- Submitting cloned tutorials with zero modifications.
- Private repositories or links with no access granted.
- Incomplete or undocumented codebases.