Model Research: AI-Powered Loan Eligibility Advisor (Open-Source)

1. Problem Understanding

Loan eligibility is a binary classification task: predict whether an applicant's loan request should be approved (1) or reject

Challenges:
- Imbalanced data (fewer rejections than approvals)
- Regulatory need for transparency → explainable AI is mandatory
- Scalability → must handle real-time web-based inference

2. Dataset Research

All datasets considered are open-source:
- German Credit Data (UCI ML Repository)
- Loan Prediction Dataset (Kaggle)
- Credit Approval Dataset (OpenML)

Features used: Income, Employment status, Credit score/history, Loan amount, Loan term, Property type, Education, De

3. Model Candidates (All Open-Source)

• Logistic Regression (scikit-learn): Simple, interpretable but weak with non-linear patterns
• Decision Tree (scikit-learn): Easy to explain but prone to overfitting
• Random Forest (scikit-learn): Robust, good accuracy, less interpretable
• XGBoost (open-source): High accuracy, handles missing values, scalable, needs tuning

4. Model Evaluation Metrics

Since wrong predictions can cost money + trust, evaluation must go beyond accuracy:
- Accuracy: overall correctness
- Precision: % of predicted approvals that were correct
- Recall: % of actual eligible applicants correctly approved
- F1-score: balance between precision & recall
- ROC-AUC: ability to distinguish approvals vs rejections
(All computed using scikit-learn open-source metrics)

5. Explainability & Transparency

For open-source explainability:
- SHAP (shap library): Global + Local explanations
- LIME (lime library): Local feature importance
- matplotlib / seaborn: Visualizations

This ensures transparent, explainable predictions.

6. Chosen Model

- XGBoost: Selected as primary model (best accuracy)
- Logistic Regression: Baseline for comparison
- Random Forest: Backup interpretable model

7. Integration Plan (Open-Source Only)

Data Processing: Pandas, NumPy, Scikit-learn
ML Training & Inference: Scikit-learn, XGBoost

Explainability: SHAP, LIME
Visualization: Matplotlib, Seaborn
Report Generation: ReportLab / FPDF
Chatbot: Rasa (open-source NLP) OR HuggingFace Transformers
Web App: Streamlit
Deployment: Streamlit Cloud / Docker / Heroku

8. Expected Outcomes

- Loan Approval Prediction Engine (open-source ML models)
- Explainable AI Visuals (SHAP, LIME)
- Financial Chatbot (Rasa/HuggingFace)
- PDF Credit Report (ReportLab/FPDF)
- Streamlit Web App for real-time loan eligibility checking