AI-Powered Loan Eligibility Advisory System

■ Roadmap

Phase 1 – Data Collection & Preprocessing: Gather datasets, clean data, encode categorical variables, normalize features.

Phase 2 – Exploratory Data Analysis (EDA): Identify correlations, visualize trends, perform feature engineering.

Phase 3 – Model Development: Build baseline Logistic Regression, test advanced models (Random Forest, XGBoost, Gradient Boosting).

Phase 4 – Model Optimization & Validation: Hyperparameter tuning, cross-validation, handle class imbalance with SMOTE/weighted loss.

Phase 5 – System Integration: Flask web app, PostgreSQL database integration, PDF report generation.

Phase 6 – Deployment: Containerize with Docker, deploy on AWS/Heroku, add logging & monitoring.

■ Model Research

1. Baseline Approach (Logistic Regression): Simple, interpretable, effective for linear relationships.

2. Ensemble Models (Random Forest, XGBoost, Gradient Boosting): Capture non-linear relationships, higher accuracy, provide feature importance.

3. Neural Networks (Optional for Research): Useful for complex patterns, needs more data, may not always outperform tree models.

4. Model Evaluation Strategy: Stratified train-test split, metrics: Accuracy, Precision, Recall, F1, ROC-AUC.

5. Explainability & Fairness: Use SHAP/LIME, check for bias, ensure fairness and transparency.