Roadmap for AI-Powered Loan Eligibility Advisory Chatbot

Phase 1:Project Initiation & Requirement Gathering

Tasks:

1. Define the project scope, key stakeholders, and user requirements.
2. Finalize the technology stack: Python, Streamlit, AWS, and specific libraries (e.g., XGBoost, SHAP).
3. Establish project management and collaboration tools (e.g., Trello, Jira).

Deliverables:

1. Requirement Specification Document.
2. High-level System Architecture Diagram.
3. Project Plan with defined timelines and milestones.

Phase 2: Data Preparation & Exploration

Tasks:

1. Collect and evaluate datasets for loan eligibility and credit scoring.
2. Clean and preprocess data: handle missing values, normalize features, and encode categorical variables.
3. Address Imbalanced Data: Implement strategies such as oversampling (SMOTE) or using class weights during model training.
4. Split the dataset into training, validation, and test sets.

Deliverables:

1. Clean and preprocessed dataset.
2. Data Dictionary outlining all features and their descriptions.

Phase 3: Backend Machine Learning Model Development

Tasks:

1. Model Selection & Training: Train multiple models including XGBoost, Random Forest, and a Neural Network.
2. Performance Evaluation: Evaluate models using a comprehensive set of metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC.
3. Explainability: Integrate a SHAP-based module to explain model predictions.
4. Model Security: Implement data privacy and security best practices to protect sensitive user information.

Deliverables:

1. Final trained and serialized machine learning model.
2. Detailed Model Evaluation Report with comparative performance metrics.
3. Integrated Explainability Module.

Phase 4: Database & User Management

Tasks:

1. Set up a relational database using AWS RDS for PostgreSQL.
2. Design the database schema with tables for Users, UserDocuments, and LoanEligibilityResults.
3. Implement secure CRUD operations for user registration, login authentication (using password hashing), and storing loan results.

Deliverables:

1. Database Schema Design Document.

2. Functional user authentication and data storage modules.

Phase 5: Chatbot & OCR Integration

Tasks:

1. Develop the conversational flow for the chatbot.
2. Implement OCR (Tesseract) to extract data from uploaded documents (Aadhaar/PAN).
3. Integrate the chatbot with the backend ML model to pass user inputs and return predictions.

Deliverables:

1. Functional Chatbot with a defined conversational flow.
2. Working OCR integration for document data extraction.

Phase 6: Web Interface & UI Development

Tasks:

1. Build the front-end user interface using Streamlit.
2. Develop dedicated pages for user login/registration, the chatbot interface, and a results dashboard.
3. Ensure a responsive, intuitive, and simple UI design.

Deliverables:

1. Interactive Streamlit-based web application with all pages.

Phase 7: Report Generation

Tasks:

1. Create a module to automatically generate professional PDF reports.
2. Populate the report with user information, input data, the final prediction, and SHAP-based explanation visuals.
3. Develop a system to store the PDF link in the database and make it downloadable.

Deliverables:

1. Automated PDF generation module.

Phase 8: Deployment & Infrastructure

Tasks:

1. Cloud Hosting: Deploy the Streamlit application and its backend components on AWS.
2. Host the ML model on AWS Lambda for a serverless, scalable solution.
3. Store all uploaded documents and generated PDFs in AWS S3.
4. Configure the database in AWS RDS.

Deliverables:

1. A fully deployed, publicly accessible web application.
2. Cloud architecture diagram.

Phase 9: Testing, Refinement & Go-Live

Tasks:

1. Conduct comprehensive Unit, Integration, and User Acceptance Testing (UAT).
2. Gather and implement user feedback from UAT sessions.
3. Address any bugs or performance issues.
4. Finalize all documentation.

Deliverables:

1. A fully functional, bug-free, and accessible AI-Powered Loan Eligibility Chatbot System.