

# AI-Powered Loan Eligibility Advisor Bot – Model Research

## Objective

The objective of this research is to identify a suitable machine learning model that can predict loan eligibility based on applicant information and documents. The model should provide accurate results, handle different types of data, and remain explainable so users understand the decisions.

## Data Sources

The model uses a mix of personal, financial, and document-based data:

- **Personal Information:** Name, Age, Employment Type, Location.
- **Financial Information:** Monthly Income, Existing EMIs, Liabilities, Credit Score.
- **Loan Preferences:** Loan Type, Amount Requested, Tenure.
- **Documents:**
  - **Pay Slips** → Income verification, deductions, allowances.
  - **Bank Statements** → Spending and repayment patterns.
  - **Aadhaar/PAN** → Identity verification.

## Preprocessing

Before training, the data must be cleaned and transformed:

1. Validate required fields such as income, age, and loan type.
2. Standardize numeric values like income and loan amount.
3. Encode categorical values like employment type and loan type.
4. Extract salary information from **pay slips** using OCR and verify it against declared income.
5. Engineer new features such as repayment capacity and income stability.

## Chosen Model – Gradient Boosting (XGBoost)

considered multiple algorithms like Logistic Regression, Decision Trees, and Random Forests. After analysis, **XGBoost** was chosen as the primary model because:

- It is highly accurate for structured financial data.
- It handles missing values and categorical variables effectively.
- It is widely used in the finance industry for risk prediction.
- It provides feature importance, helping explain which factors influence approval decisions.

## Model Workflow:

- Input: Preprocessed applicant data and extracted pay slip details.
- Prediction: Loan eligibility status, approval probability, suggested loan amount and tenure.
- Output: Transparent results with reasons behind the decision using explainability tools.

## Evaluation Metrics

To measure model performance, we use:

- **Accuracy** – overall correctness of predictions.
- **Precision** – ensuring approved applicants are truly eligible.
- **Recall** – capturing all genuinely eligible applicants.
- **F1 Score** – balance between precision and recall.
- **ROC-AUC** – ability to differentiate between eligible and ineligible applicants.

## Bias & Fairness Considerations

- Avoid discrimination based on gender, location, or age beyond legal lending norms.
- Ensure consistency when handling pay slips of different formats.
- Regular re-training with updated datasets to reduce historical bias.

## Deployment Plan

1. Train the model using historical loan and salary slip datasets.
2. Save the trained model as a reusable file (.pkl / .joblib).
3. Deploy through a **Flask / FastAPI backend** to serve predictions.
4. Connect to the chatbot frontend for real-time interaction.
5. Use SHAP/LIME explainability so users know why their loan was approved or rejected.
6. Continuously update the model with feedback and real outcomes.

## Conclusion

The final model for the Loan Eligibility Advisor Bot is **XGBoost**. It balances performance, reliability, and transparency while making use of structured financial information and unstructured data like pay slips. This single-model approach ensures a practical, scalable, and explainable solution.

