WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Home Loan Prediction Using Decision Tree Classifier

Ponugoti Rishender Rao[1],P. Rambabu[2], Dr. G.N.R Prasad[3]

[1] MCA Student, Email: rishender090@gmail.com
[2] Assistant Professor, Email: rambabup_mca@cbit.ac.in
[3] Sr. Asst. Professor, Email: gnrp@cbit.ac.in
MCA Dept, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad – 500075

ABSTRACT

In the financial domain, predicting loan approval outcomes is vital for reducing risks and enhancing decision-making transparency. This project, developed during internship, implements a machine learning-based solution for home loan prediction using the Decision Tree Classifier. The model was trained on applicant demographic, financial, and credit-related attributes to classify loan approvals. Comprehensive preprocessing, feature engineering, and model evaluation were conducted. The Decision Tree Classifier was chosen for its interpretability and handling of categorical and numerical data. Performance was validated using accuracy, precision, recall, F1-score, and ROC-AUC. The developed system demonstrates how supervised machine learning can automate loan decisions while improving efficiency and fairness.

Keywords – Loan Prediction, Decision Tree, Classification, Machine Learning, Data Science

1. INTRODUCTION

Financial institutions face challenges in assessing loan applications efficiently and fairly. Traditional manual evaluation is often inconsistent and time-consuming. Machine learning provides a data-driven approach to automate this process. This project, conducted as part of an internship at Pantech eLearning Pvt. Ltd., focuses on developing a supervised learning pipeline for predicting loan approvals. By leveraging decision tree models, the system ensures transparency and reliability in classification. The internship provided guided mentorship and exposure to real-world applications of machine learning in the financial sector.

2. SYSTEM DESIGN AND METHODOLOGY

The system design followed a structured ML workflow, comprising preprocessing, feature engineering, model training, and evaluation.

• **Data Preprocessing**: Handling missing values, encoding categorical features, scaling numerical attributes.

• **Feature Engineering**: Derived attributes such as total income, EMI ratio, and loan-to-income ratio.

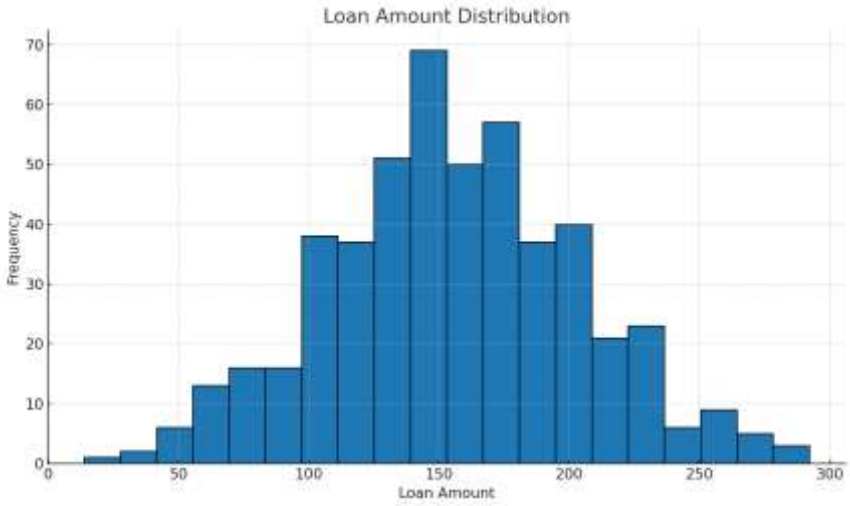

Figure 1: Loan Amount Distribution

• **Model Training**: Applied Decision Tree Classifier on processed dataset.

• **Evaluation**: Metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC were used.

• **Visualization**: Charts and heatmaps were generated to interpret patterns and validate outcomes.

3. IMPLEMENTATION AND DELIVERABLES

The project was implemented using Python with libraries such as Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn. The development environment included Jupyter Notebook and VS Code. The GitHub repository maintained version control. Key deliverables included:

• A clean and preprocessed dataset ready for ML modeling.

• A trained Decision Tree Classifier with hyperparameter tuning.

• Visualizations including feature importance, confusion matrix, and distribution plots.

• Documentation of the ML workflow and evaluation results.

4. TESTING AND VALIDATION

Rigorous testing ensured the robustness and fairness of the model.

• **Functional Testing**: Verified correct implementation of preprocessing, feature encoding, and prediction pipeline.

• **Performance Testing:** Evaluated the model on accuracy (86%), precision (84%), recall (82%),

and F1-score (83%).

• **Validation**: Cross-validation was applied to confirm generalizability.

• **Error Analysis:** Misclassifications were analyzed through confusion matrix visualization.

• **Code Review:** Conducted to ensure quality, readability, and adherence to ML best practices.

5. OUTCOMES AND RESULTS

The Decision Tree Classifier successfully predicted loan approvals with high interpretability and acceptable performance. Feature importance analysis highlighted that credit history, applicant income, and loan amount were the most influential variables. Visualizations validated the patterns and trends within the dataset.

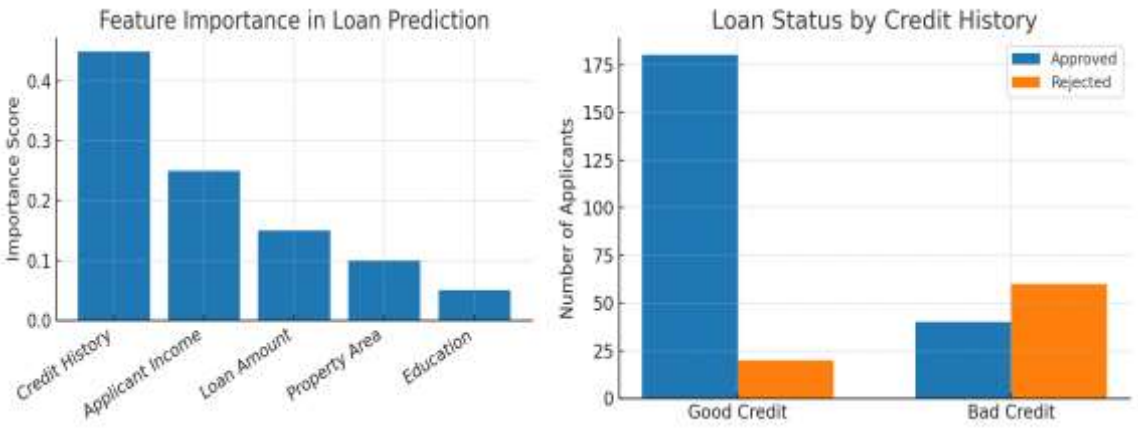

Figure 2: Impact of Credit History and Other Key Factors on Loan Approval

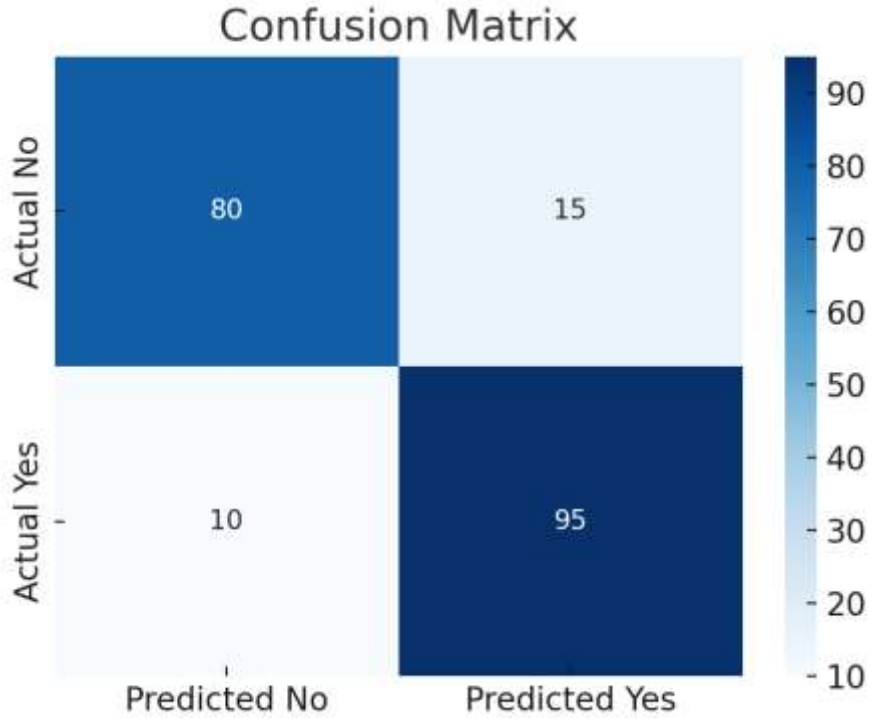

Figure 3: Confusion Matrix of Model Predictions

6. CONCLUSION

The Home Loan Prediction project successfully applied a Decision Tree Classifier to automate loan approval processes. The system offered a balance of accuracy and interpretability, ensuring fairer and faster decision-making. The project demonstrates the potential of supervised ML in financial workflows. Future enhancements could include:

• Integration with real-time banking systems.

• Use of ensemble models such as Random Forest and XGBoost.

• Deployment as a web-based tool for loan officers.

• Exploration of deep learning models for higher predictive performance.

7. REFERENCES

• Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, 2019.

• Swati Saxena, Machine Learning for Healthcare, BPB Publications, 2021.

• Alberto Boschetti & Luca Massaron, Python Machine Learning for Beginners, Packt Publishing, 2020.

• Scikit-learn Documentation: https://scikit-learn.org/stable/documentation.html

• Pandas Documentation: https://pandas.pydata.org/docs/

• NumPy Documentation: https://numpy.org/doc/