KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

Department of Computer Science and Engineering (AI)

“Loan Default Prediction”

Submitted By :

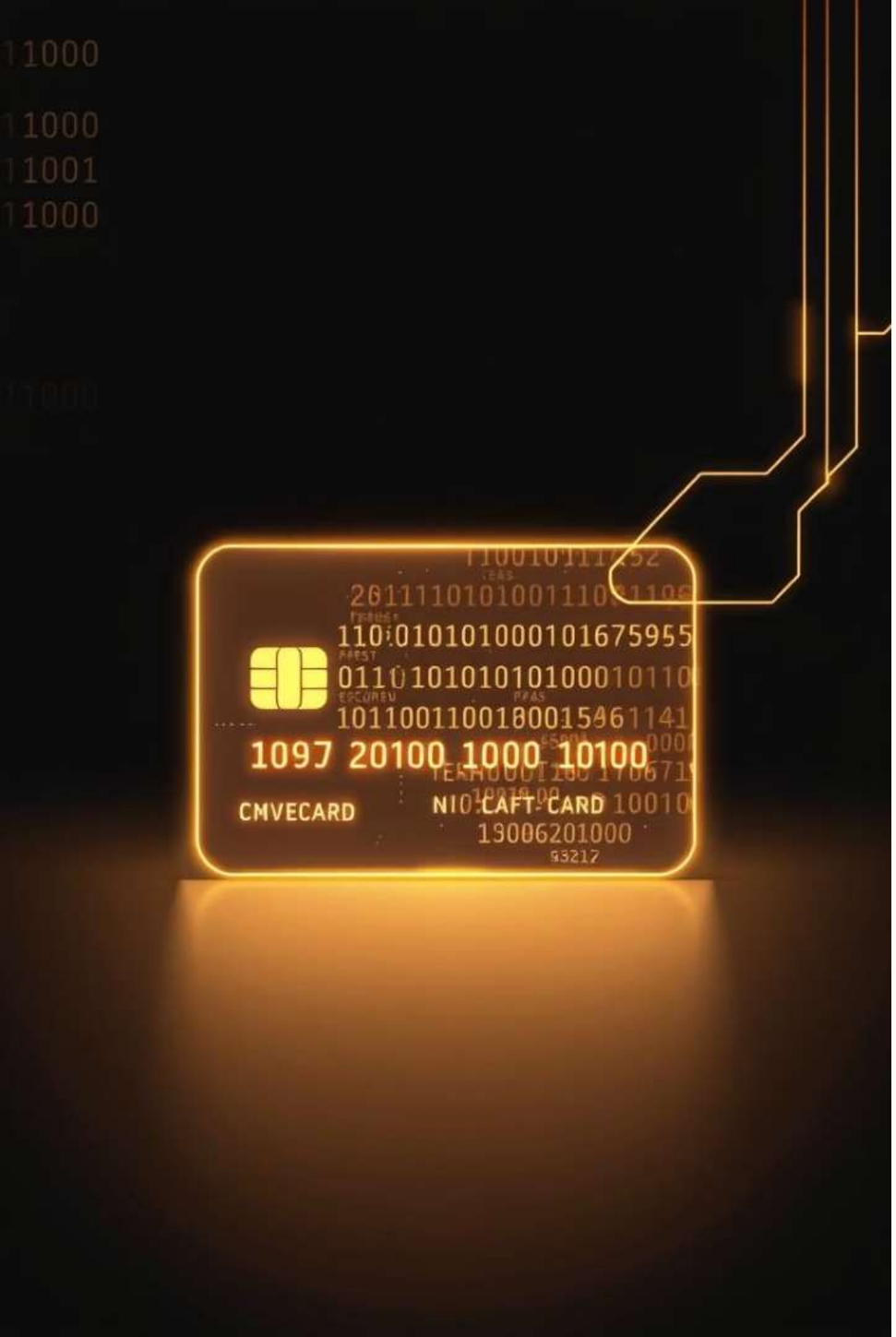
Atharv Tayal
Aviral Dixit
Ayush Goyal
Dhairya Goel

Branch : CSEAI

Section : B

Session : 2024-2025

Submitted To : Mr. Shivansh Prasad



Loan Default Prediction: A Classification Approach

Predicting loan defaults is crucial for lenders. This project uses machine learning to enhance the loan approval process. We aim to identify risk early.

The Problem: Loan Default Risk

$140B

Annual Losses

Loan defaults cost lenders billions every year.

30%

Risk Reduction

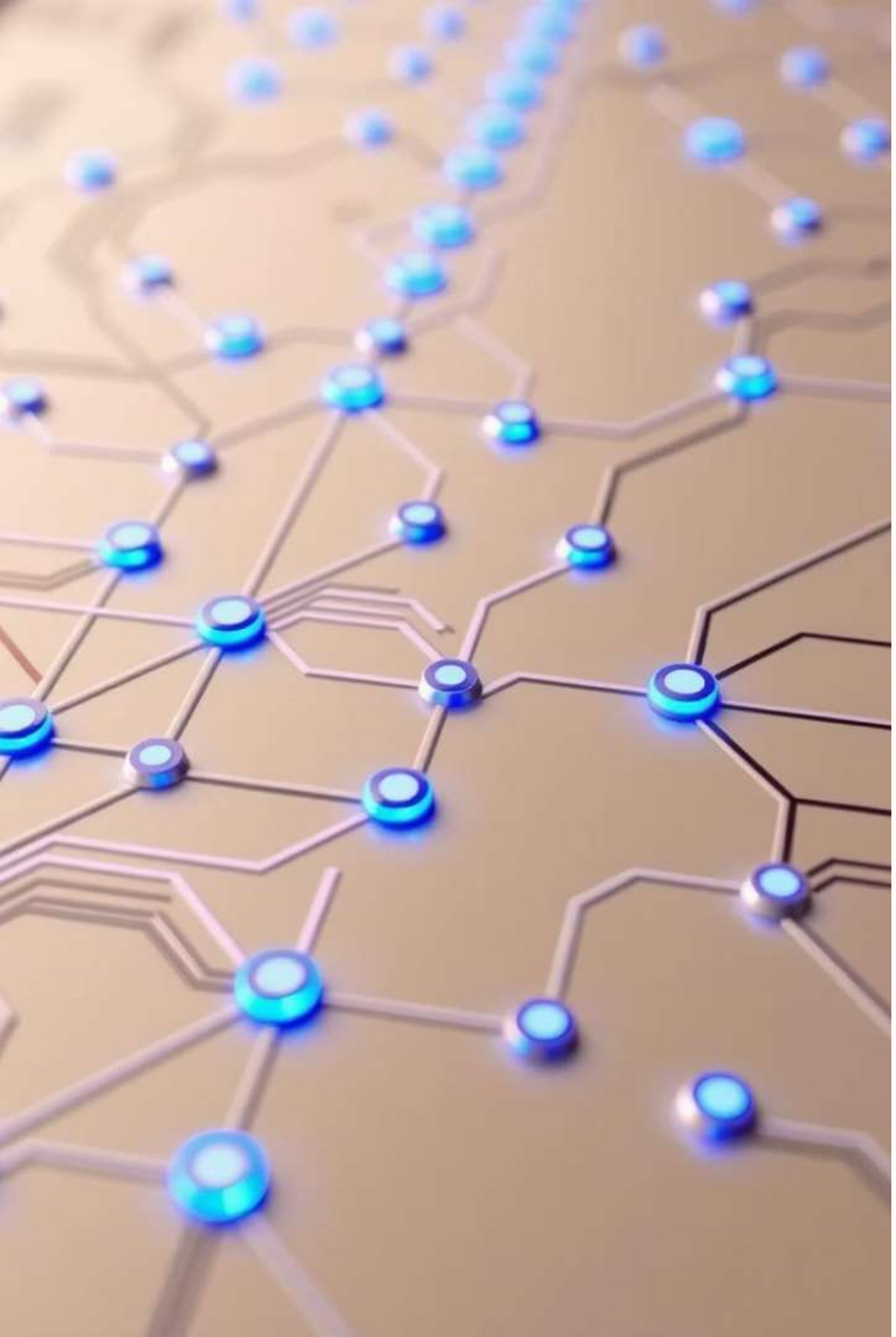
Accurate predictions can significantly reduce these losses.

15%

Fair Access

Improved models lead to fairer credit access for borrowers.

Predicting defaults helps manage risk and refine pricing strategies. This ultimately benefits both lenders and applicants.



Methodology: Classification Algorithms

1

Logistic Regression

A statistical model for binary outcomes.

2

Decision Trees

Flowchart-like structures for decision making.

3

Random Forest

An ensemble method for improved accuracy.

We utilized Logistic Regression, Decision Trees, and Random Forest. Data was split 80/20 for training/testing. Performance was measured by accuracy, precision, recall, and F1-score with 5-fold cross-validation.

Data Preprocessing and Feature Engineering

Missing Values

Imputed using mean or median strategies.

Categorical Encoding

One-hot encoding for nominal variables.

Feature Scaling

StandardScaler applied for normalization.

Feature Selection

Leveraged domain knowledge for key inputs.

We prepared the dataset by handling missing values and transforming variables. This crucial step ensures data quality and model effectiveness.

Results: Model Performance

Model	Accuracy	F1-Score
Logistic Regression	80%	75%
Decision Trees	72%	68%
Random Forest	83%	80%

Random Forest demonstrated the best performance among all models. It achieved the highest accuracy and F1-score, proving its robustness.

Feature Importance: Random Forest

Credit History

Contributed 35% to prediction accuracy.

Income

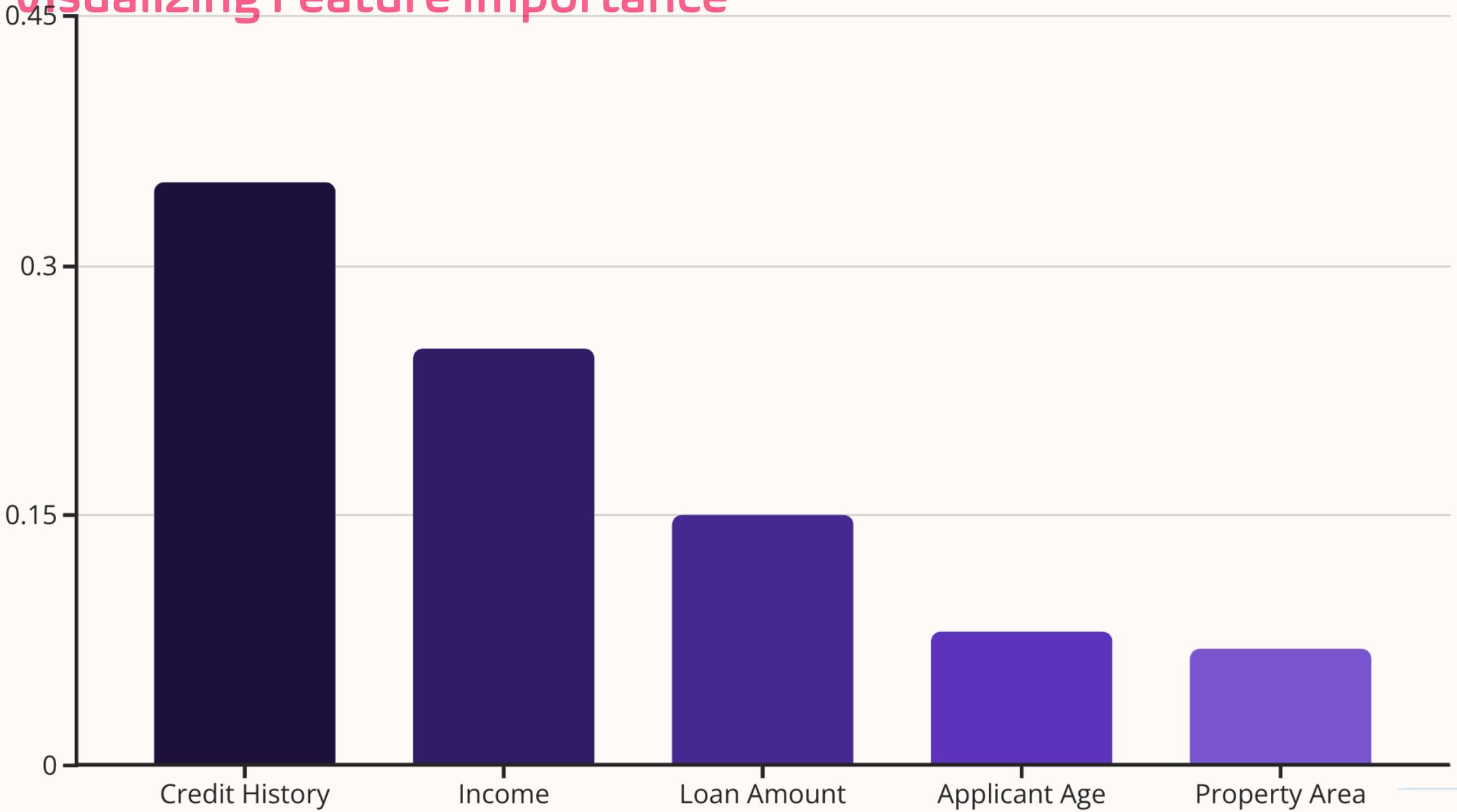
Accounted for 25% of feature importance.

Loan Amount

Weighted at 15% in the model.

Credit History, Income, and Loan Amount are the most influential features. These insights are crucial for understanding loan risk factors.

Visualizing Feature Importance



Conclusion and Future Work

Key Takeaways

Random Forest excels in predictive performance. Feature importance guides understanding.

Next Steps

Hyperparameter tuning can optimize models further. More data will improve accuracy.

These models enable more informed lending decisions. They contribute to a more robust financial system.

Thank You