

CREDIT RISK ASSESSMENT AND LOAN PREDICTION

Group 3 Members

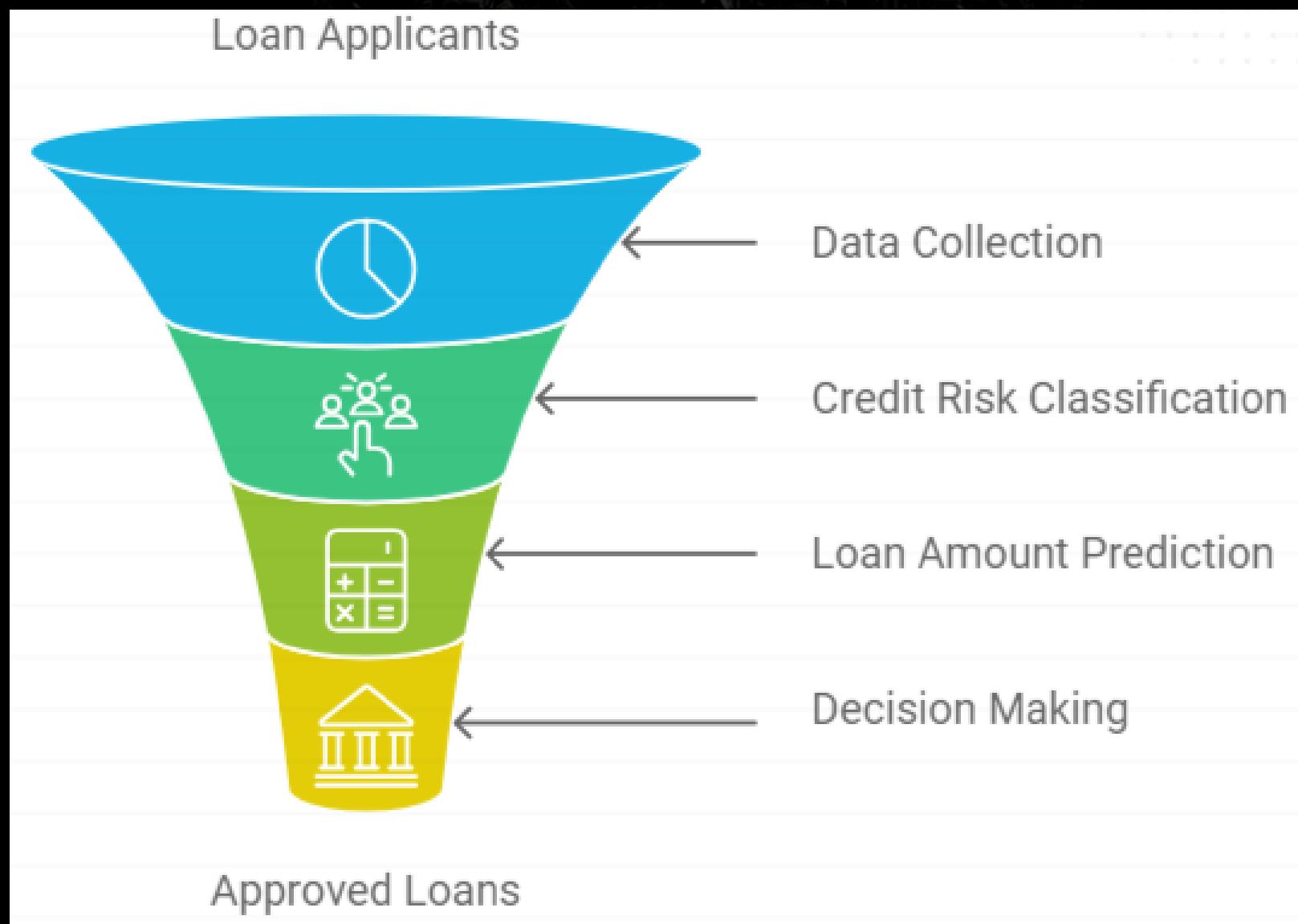
- Grace Gitau
- Kelvin Letimalo
- Grace Wacheke
- David Chege
- Faith Wanjala

BUSINESS PROBLEM

Inaccurate credit risk assessment causes financial losses and burdens.

Goal: Protect financial health of banks.

- Credit risk classification by grouping applicants based on credit worthiness
- Create a predictive model to estimate loan amount based on financial health indicators e.g income level.

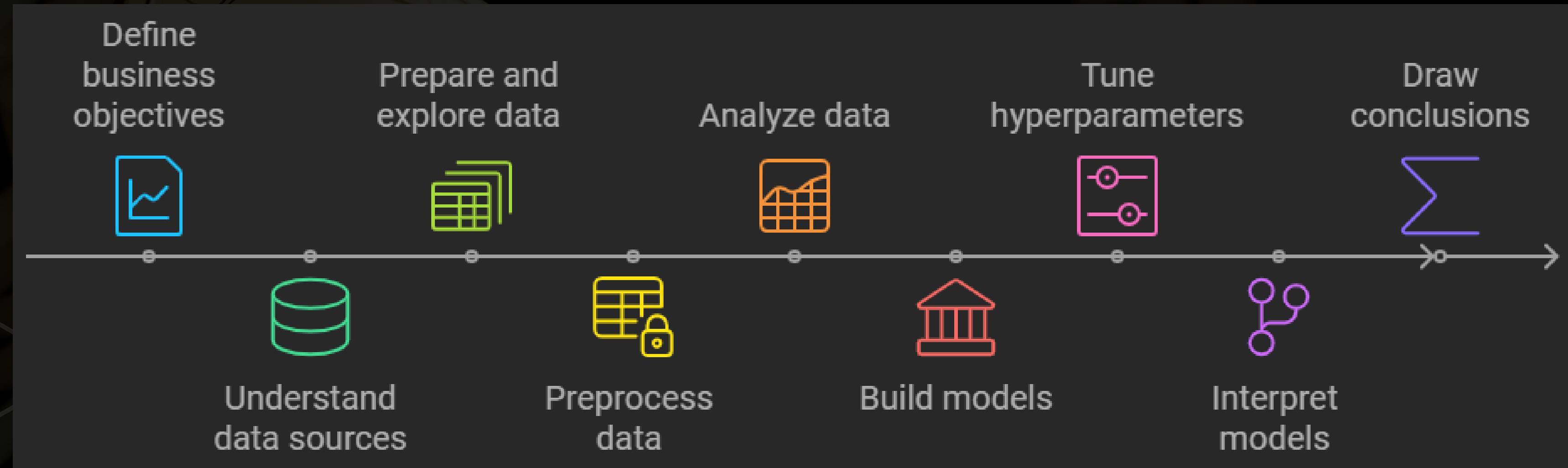


DATA UNDERSTANDING

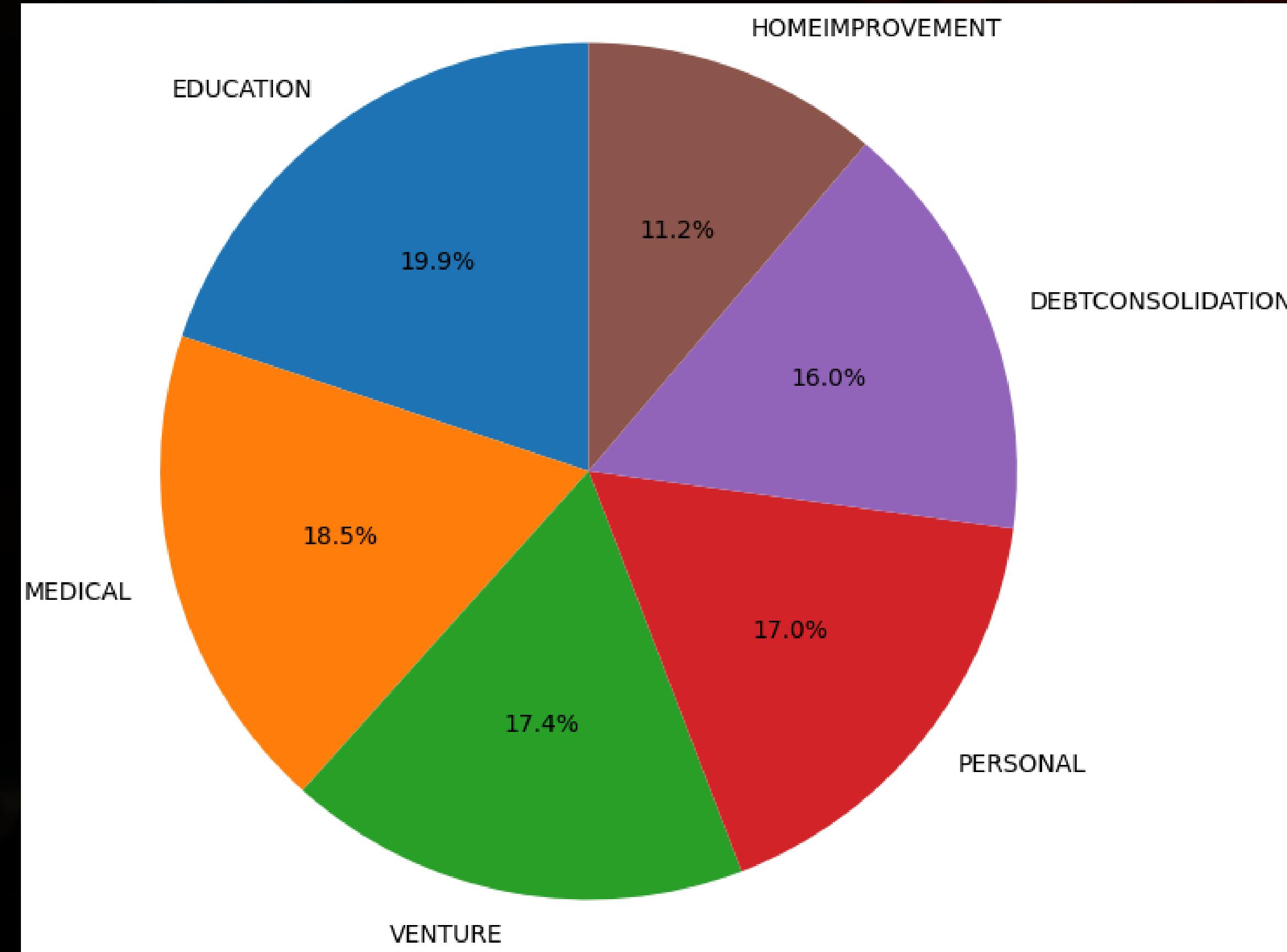


- The dataset contains 32,581 observations.
- Observations are of 12 variables.
- Loan status is our target column.

Process

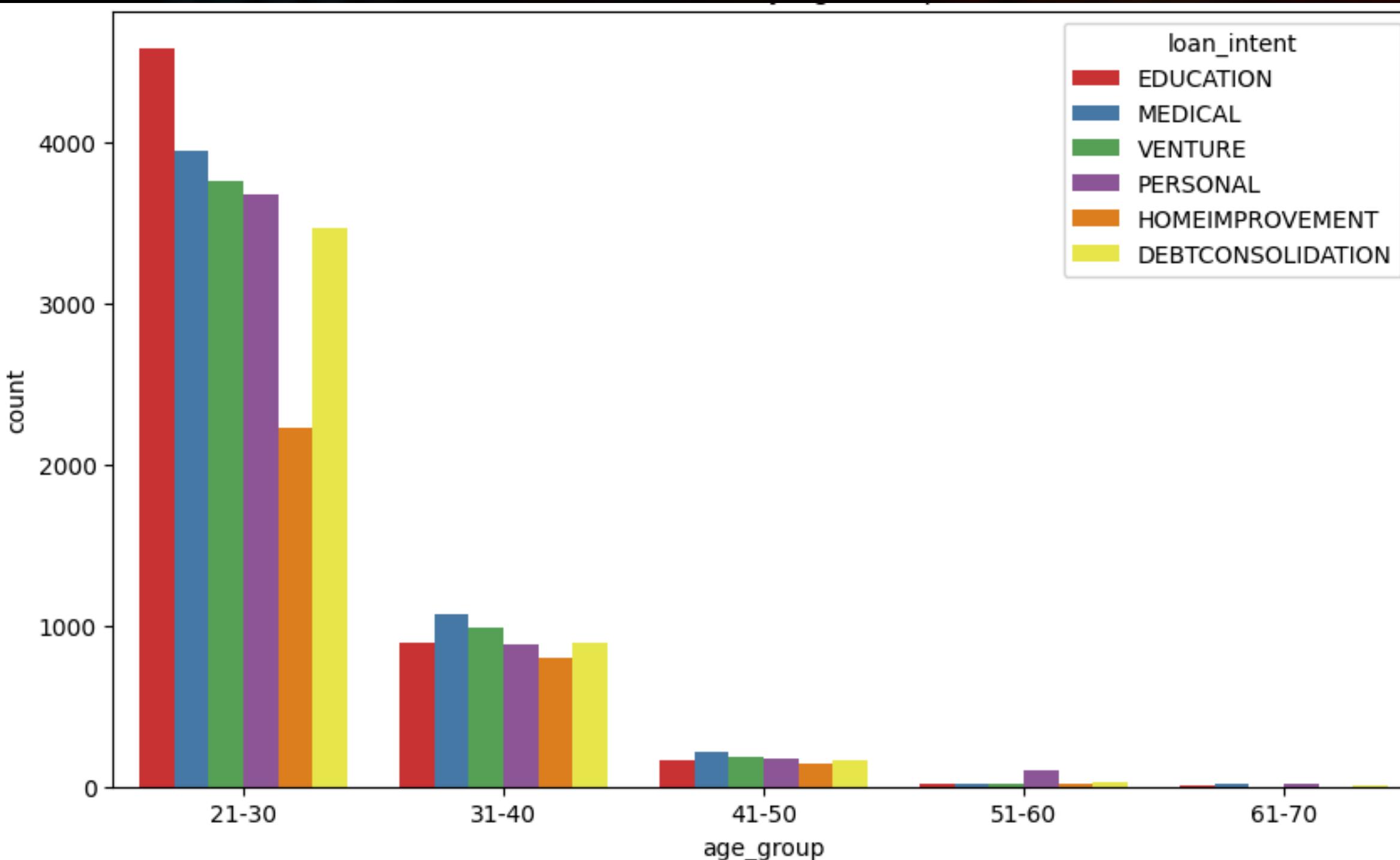


Univariate Analysis



Pie chart showing intent of use for loan having education with the highest percentage.

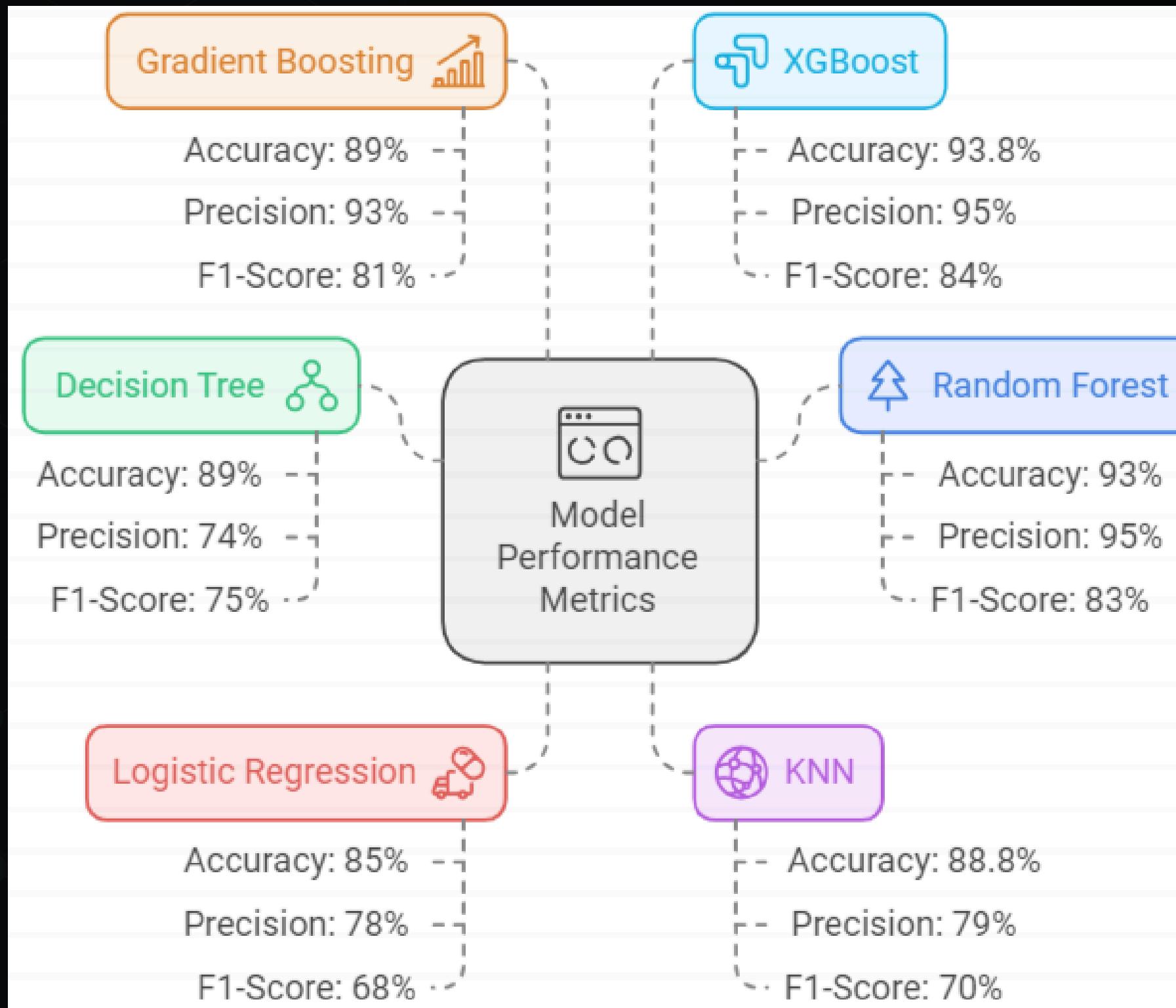
Bivariate Analysis



This graph shows patterns of different age groups and reasons for their loan acquisitions. Color code for loan reason indicated in the legend.

Models

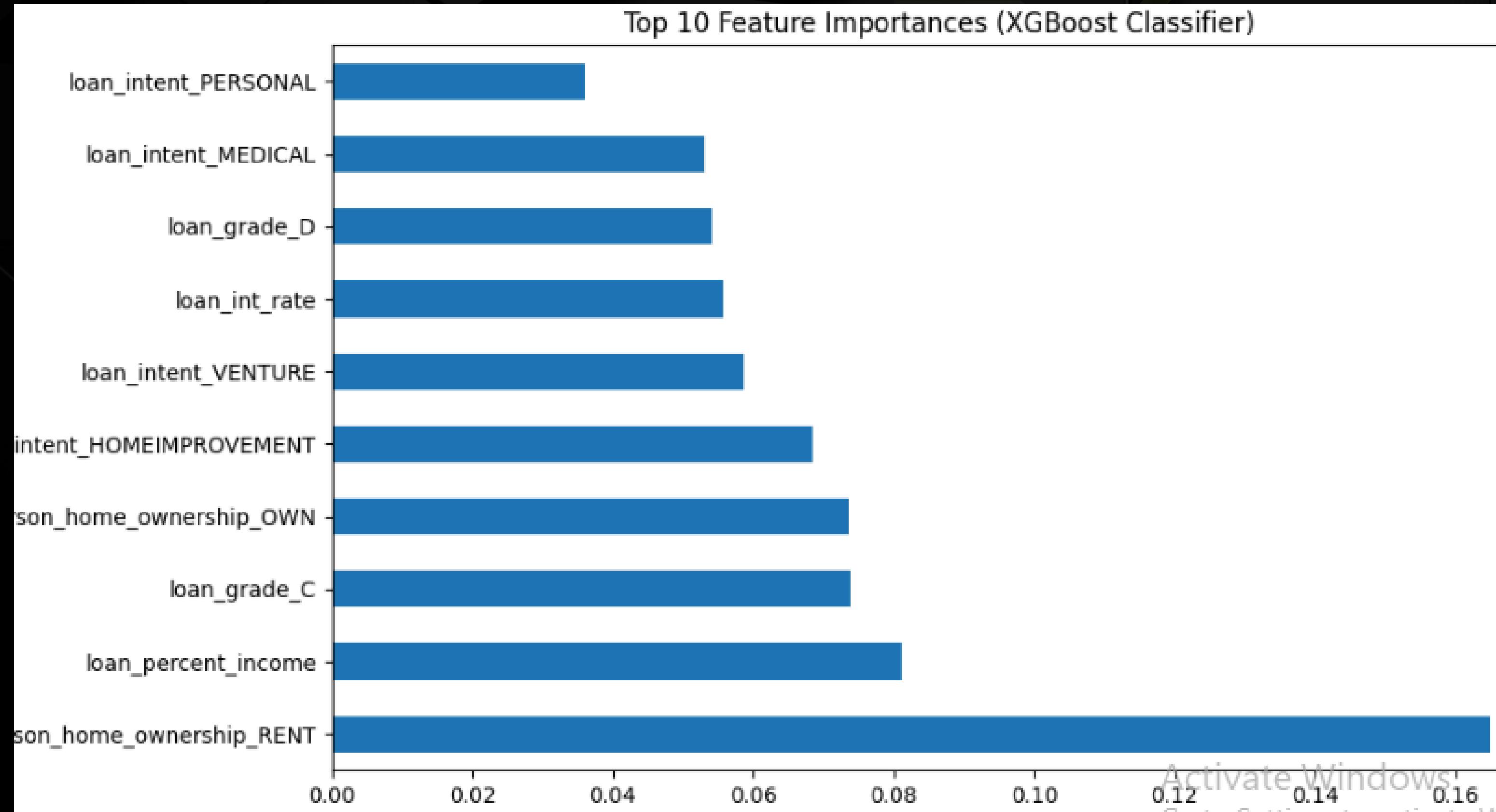
Credit risk classification



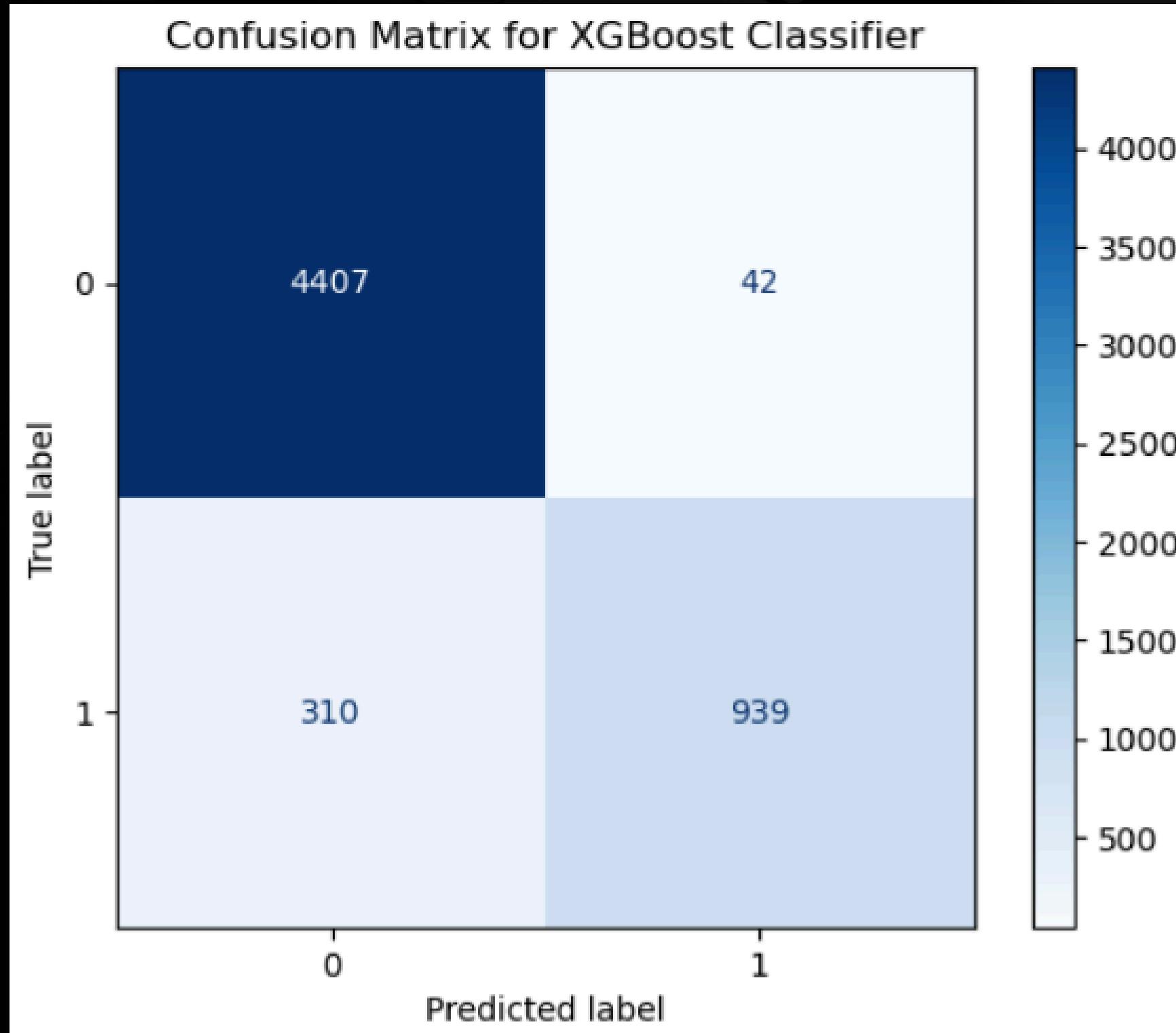
Metric scores for the various models.

next slide →

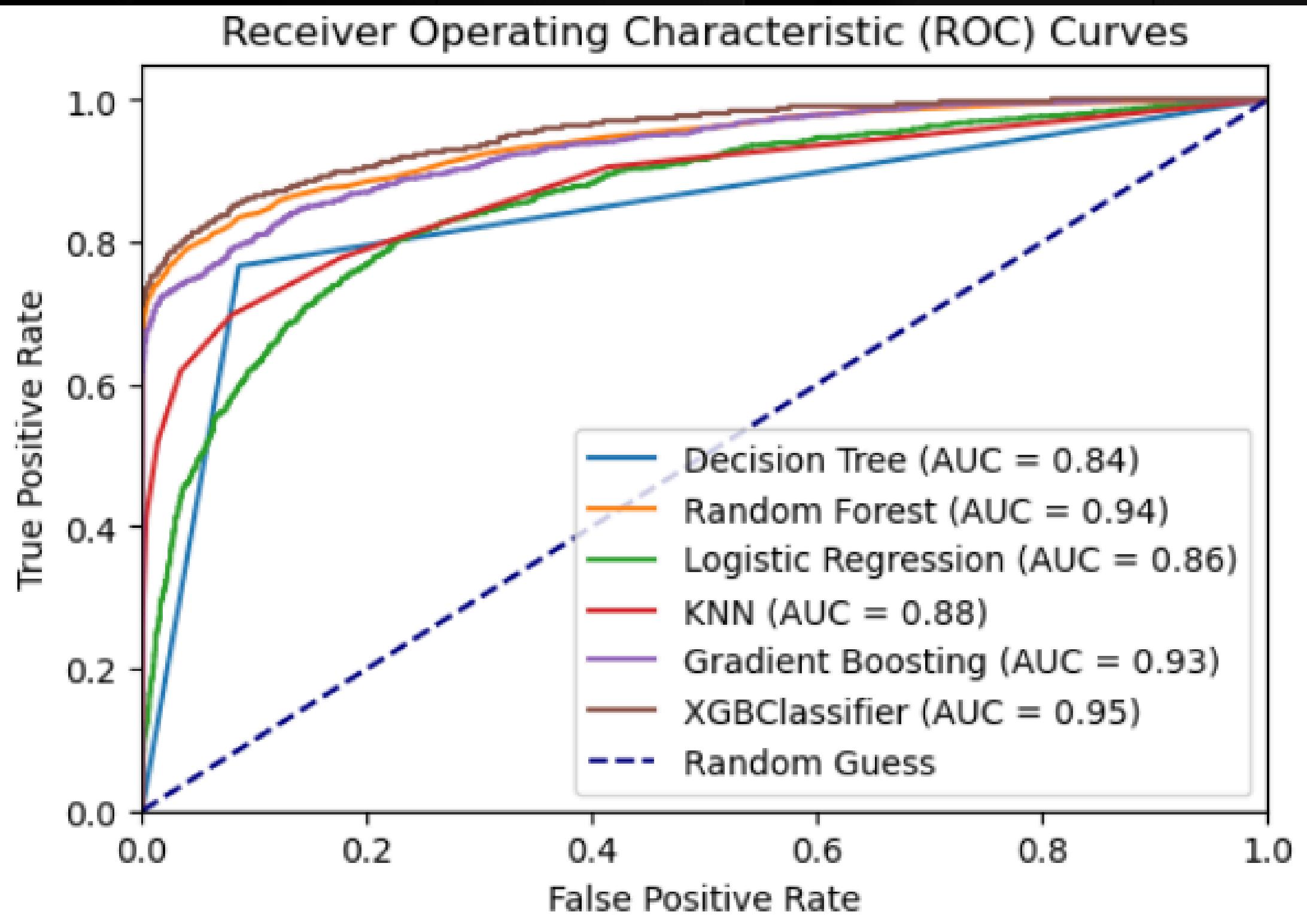
Top ten important features in XGBoost classifier model



XGBoost classifier confussion matrix



Receiver Operating Curves for all models.



XGBClassifier and Random Forest lead in performance (AUCs 0.95 and 0.94), while Decision Tree has the lowest (AUC 0.84).

Loan amount prediction models

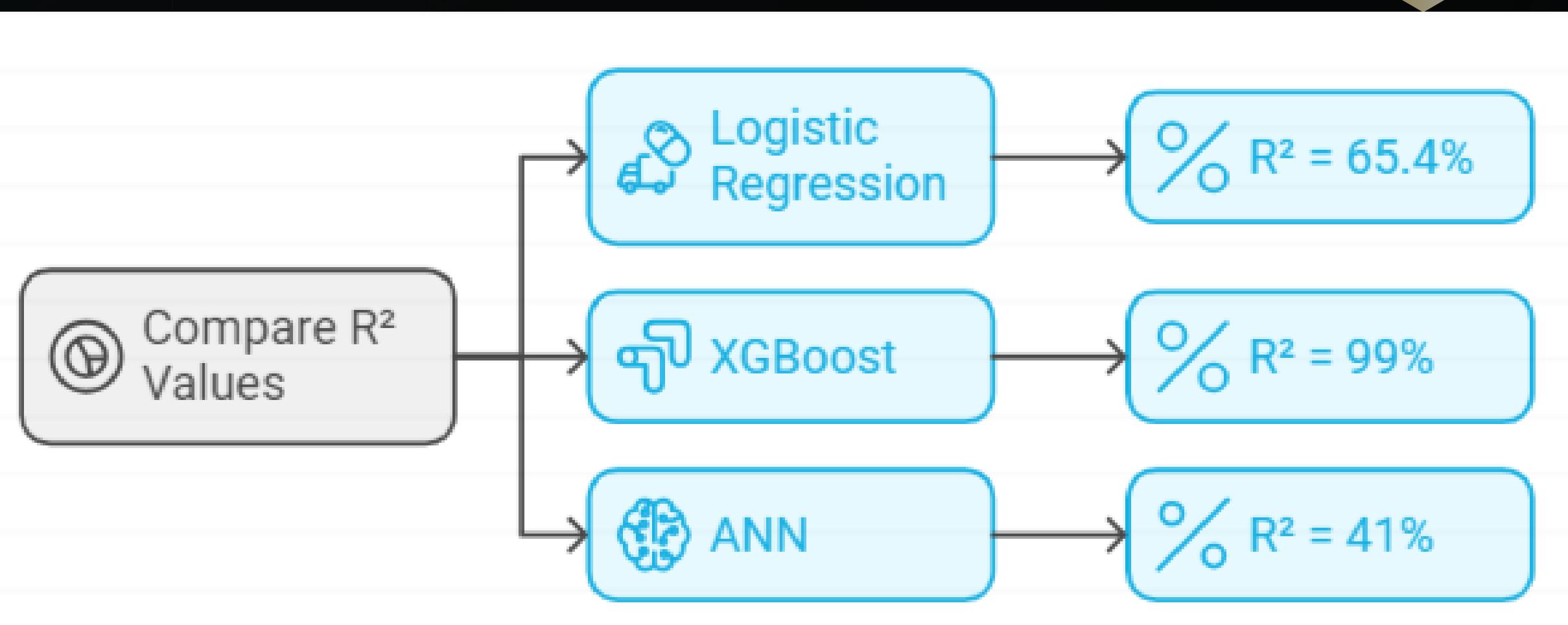
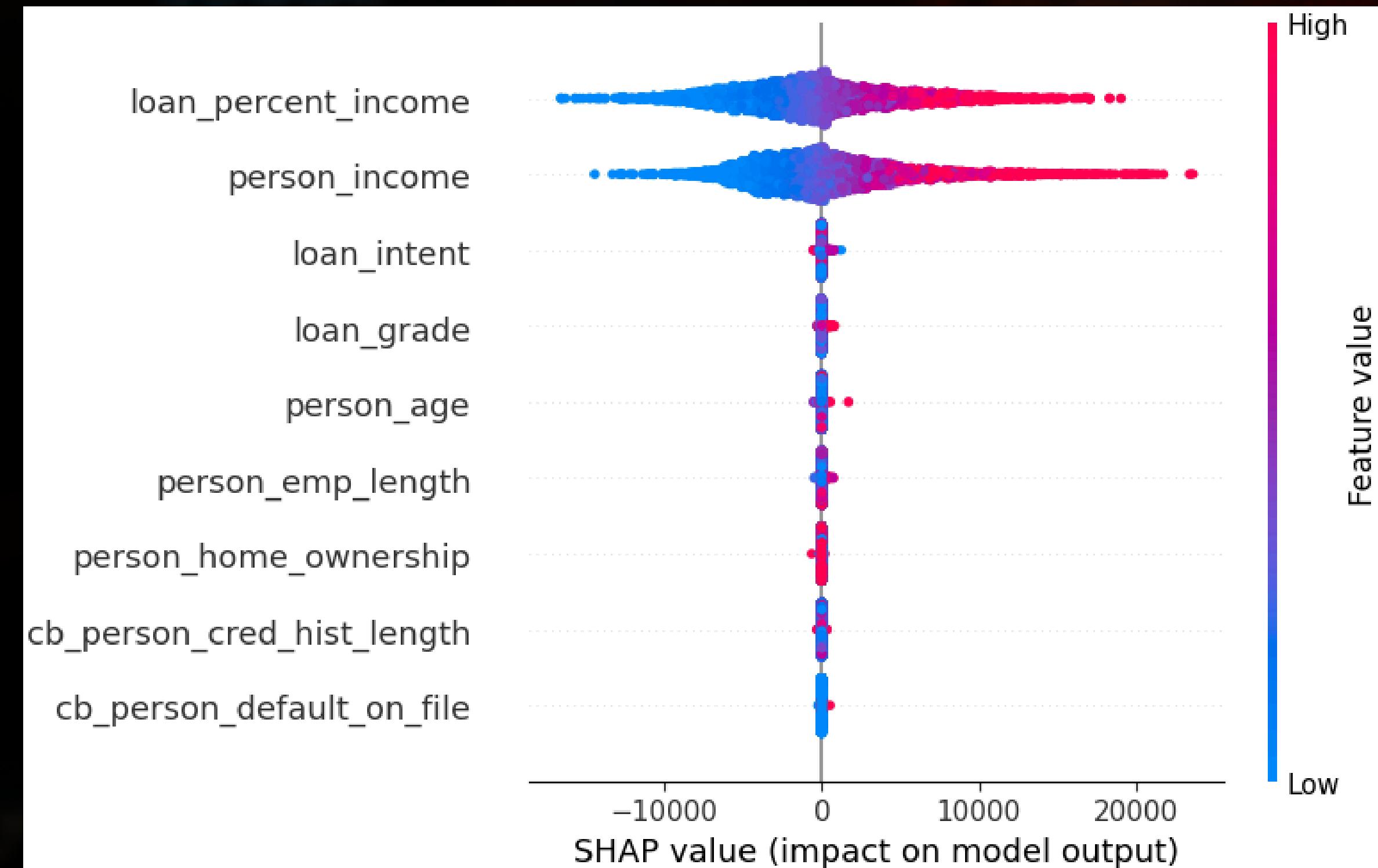


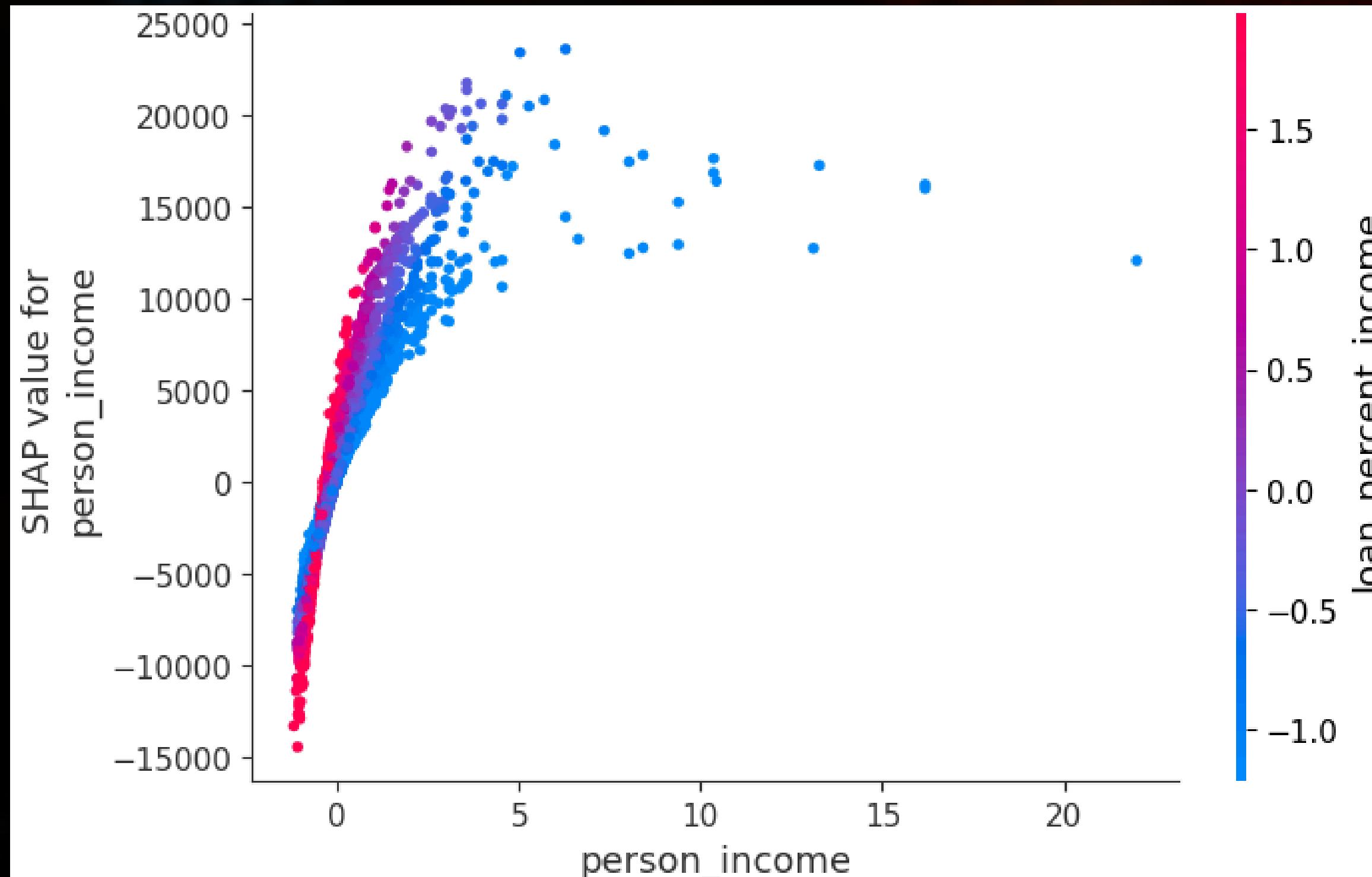
Chart of R squared values in percentage form for the various models. As evident from the chart XGBoost has the highest R squared suggesting it explains 99% of the variance in the target variable.

Model Interpretability



From plot, loan percent income and person income are the most critical features determining loan amount.
Higher loan percent income (red) - lower loan amount.
Higher person income - higher loan amount.

Model Interpretability



SHAP plot focuses on relationship between person income and its impact on predicted loan amount with color scale representing loan amount as a percentage of income.

Recommendations

- Develop tailored loan products for specific segments.
- Use machine learning models like XGBoost to assess credit risk.
- Implement financial literacy programs for borrowers.
- Regularly update model with new data.
- Explore ensemble learning for better predictions.

Deployment

The project was deployed using Streamlit, offering an interactive interface for real-time credit risk and loan amount predictions. This deployment enhances accessibility, providing valuable tools for lenders to make informed decisions.

Streamlit App



Thank you.