CREDIT RISK ASSESSMENT AND LOAN PREDCTION

Group Members

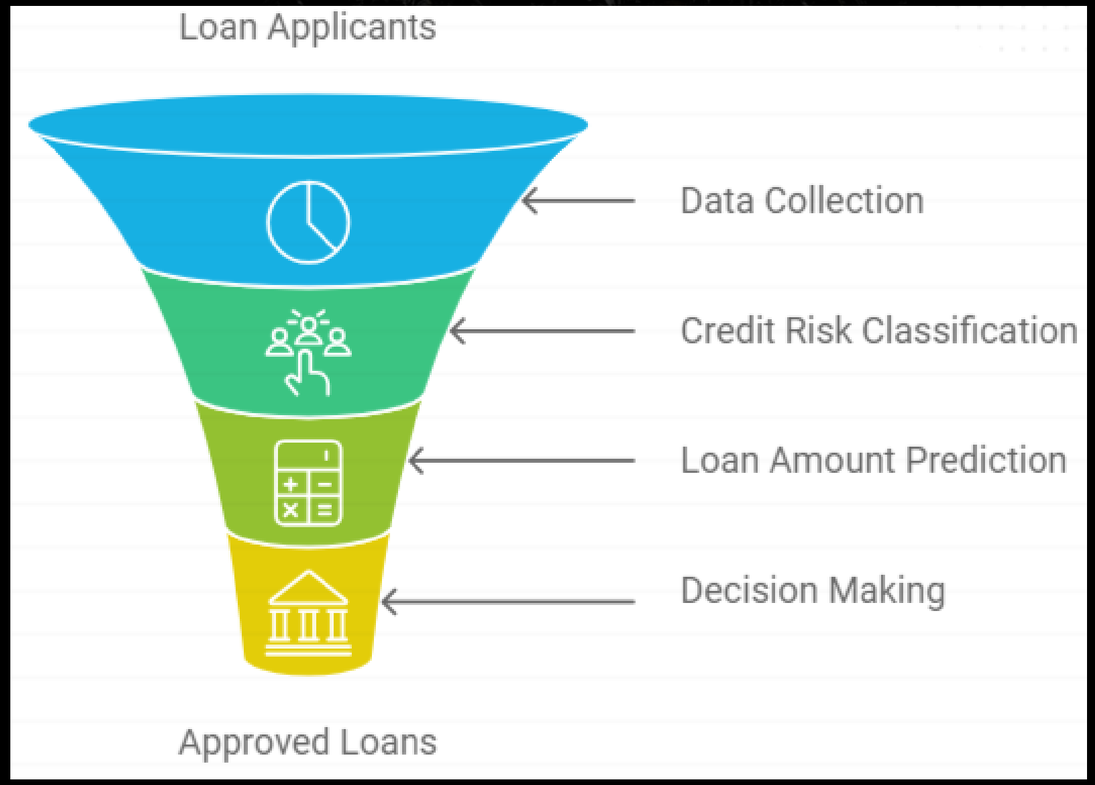
- Grace Gitau
- Kelvin Letimalo
- Grace Wacheke
- David Chege
- Faith Wanjala

BUSINESS PROBLEM

Inaccurate credit risk assessment causes financial losses and burdens.

Goal: Protect financial health of banks.

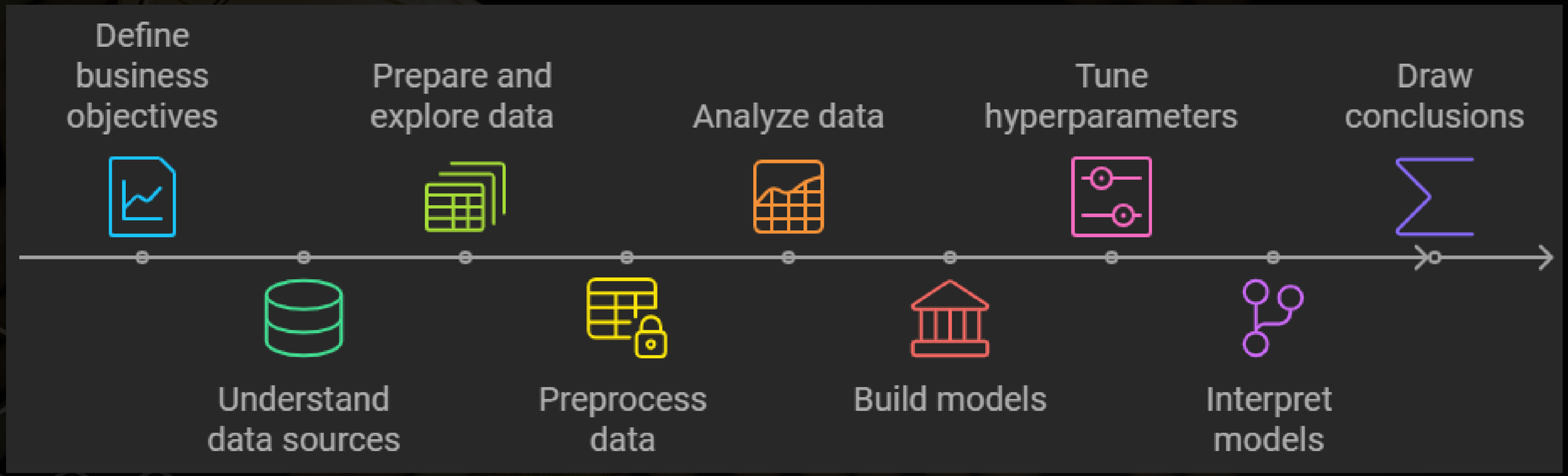
- **Credit risk classification by grouping appliicants based on credit worthiness**
- **Create a predictive model to estimate loan amount based on financial health indicators e.g income level.**



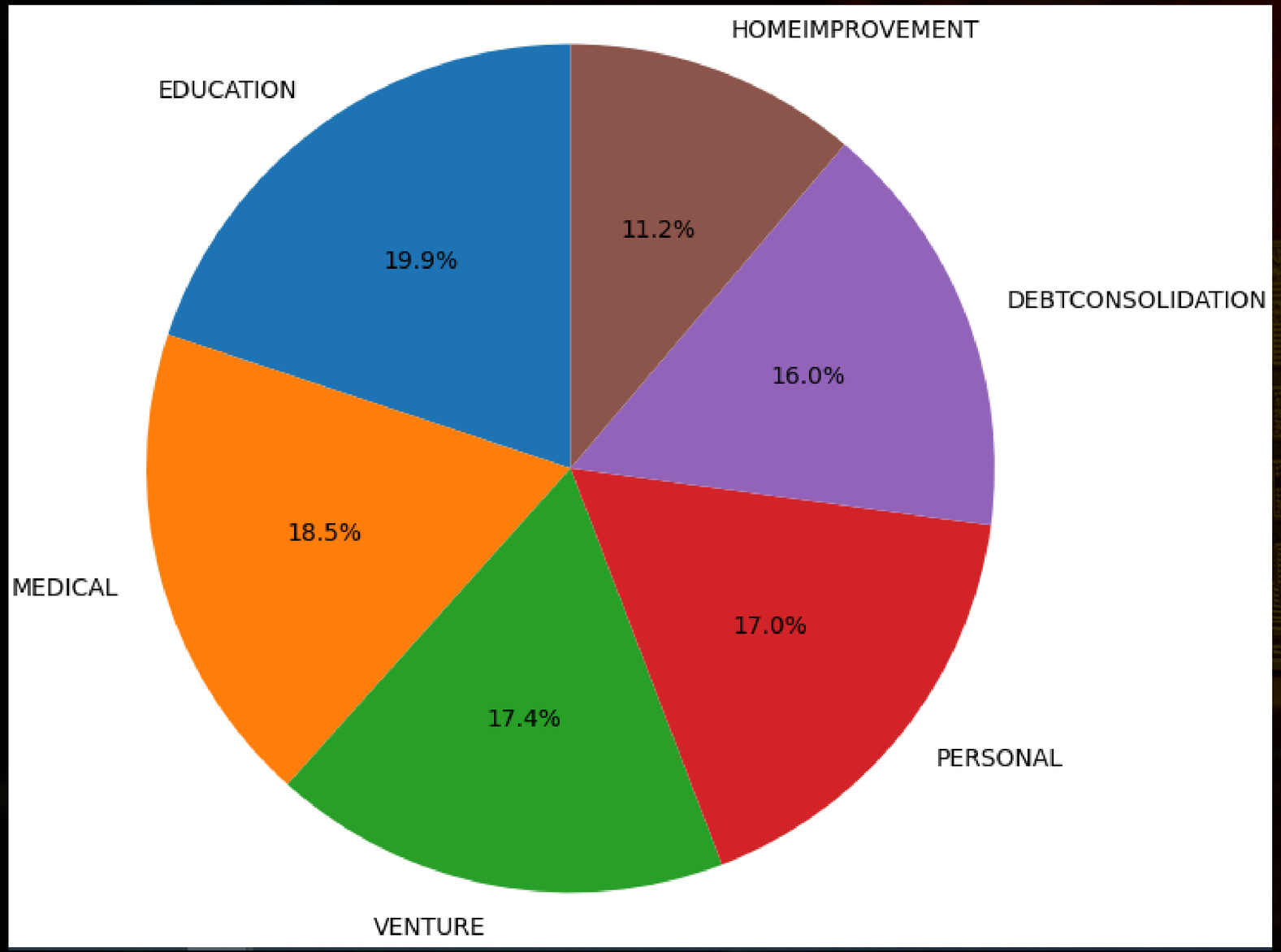
DATA UNDERSTANDING

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

Process

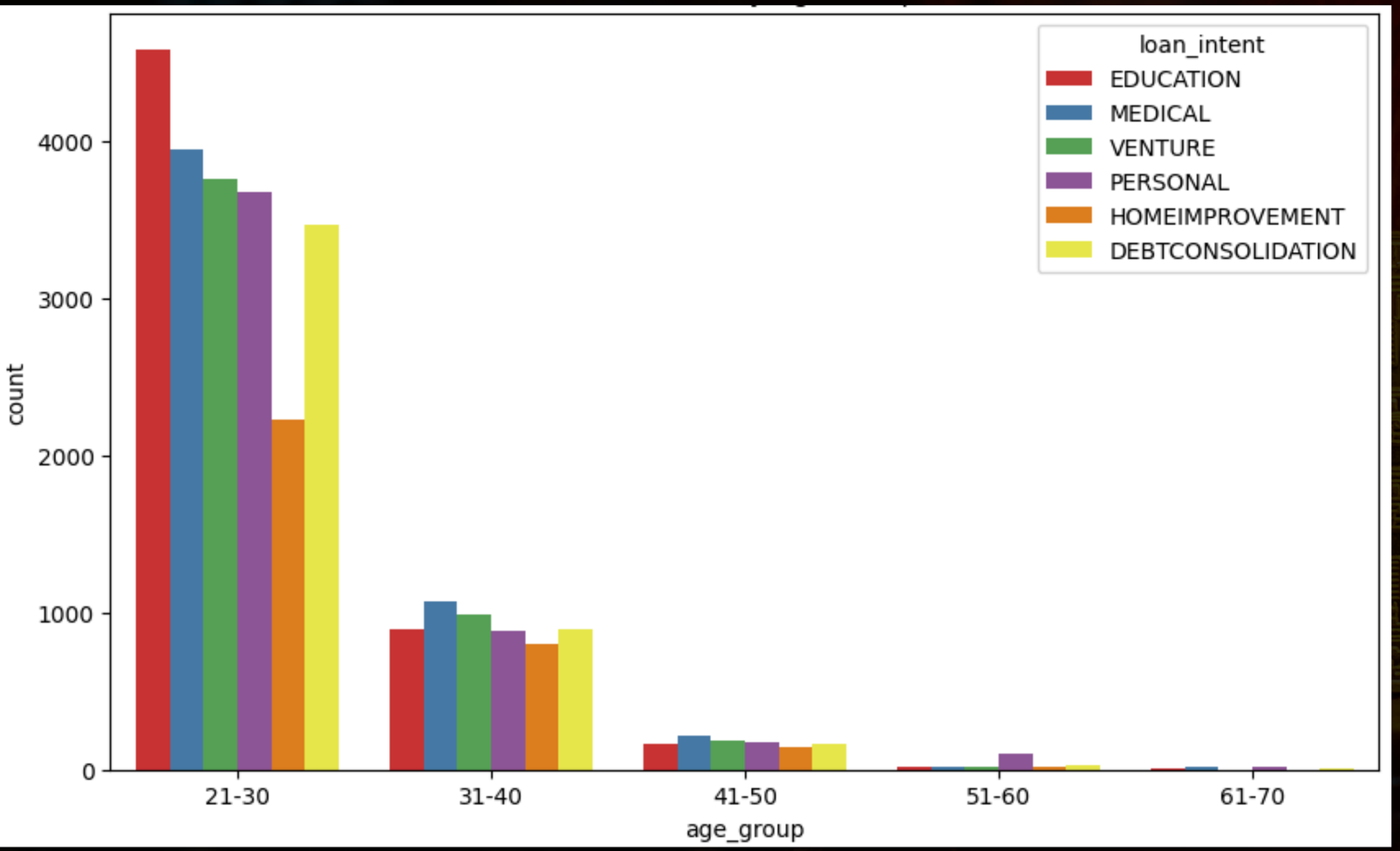

Define business objectives
Prepare and explore data
Analyze data
Tune hyperparameters
Draw conclusions
Understand data sources
Preprocess data
Build models
Interpret models

Univariate Analysis



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

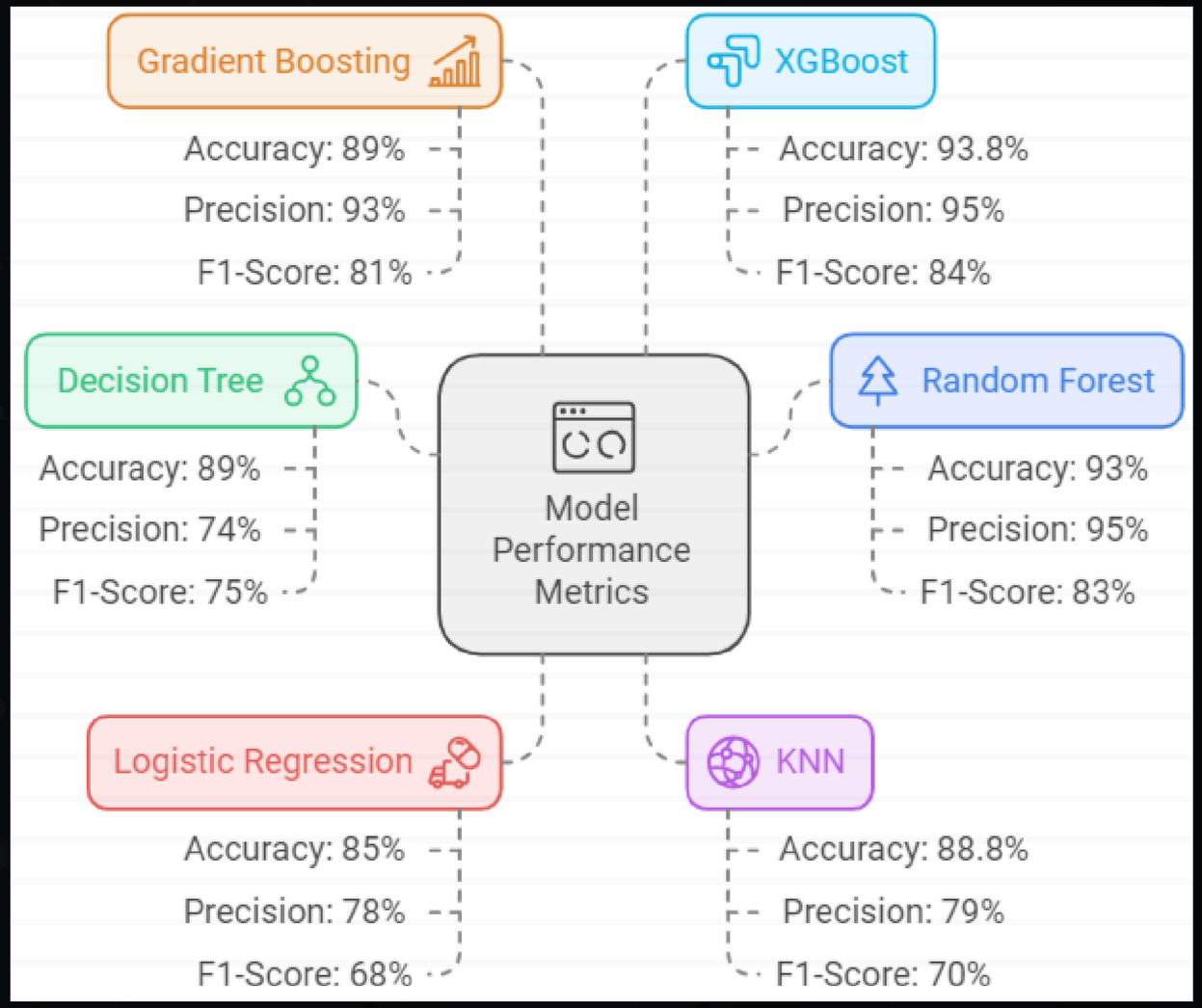
Bivariate Analysis



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

Models

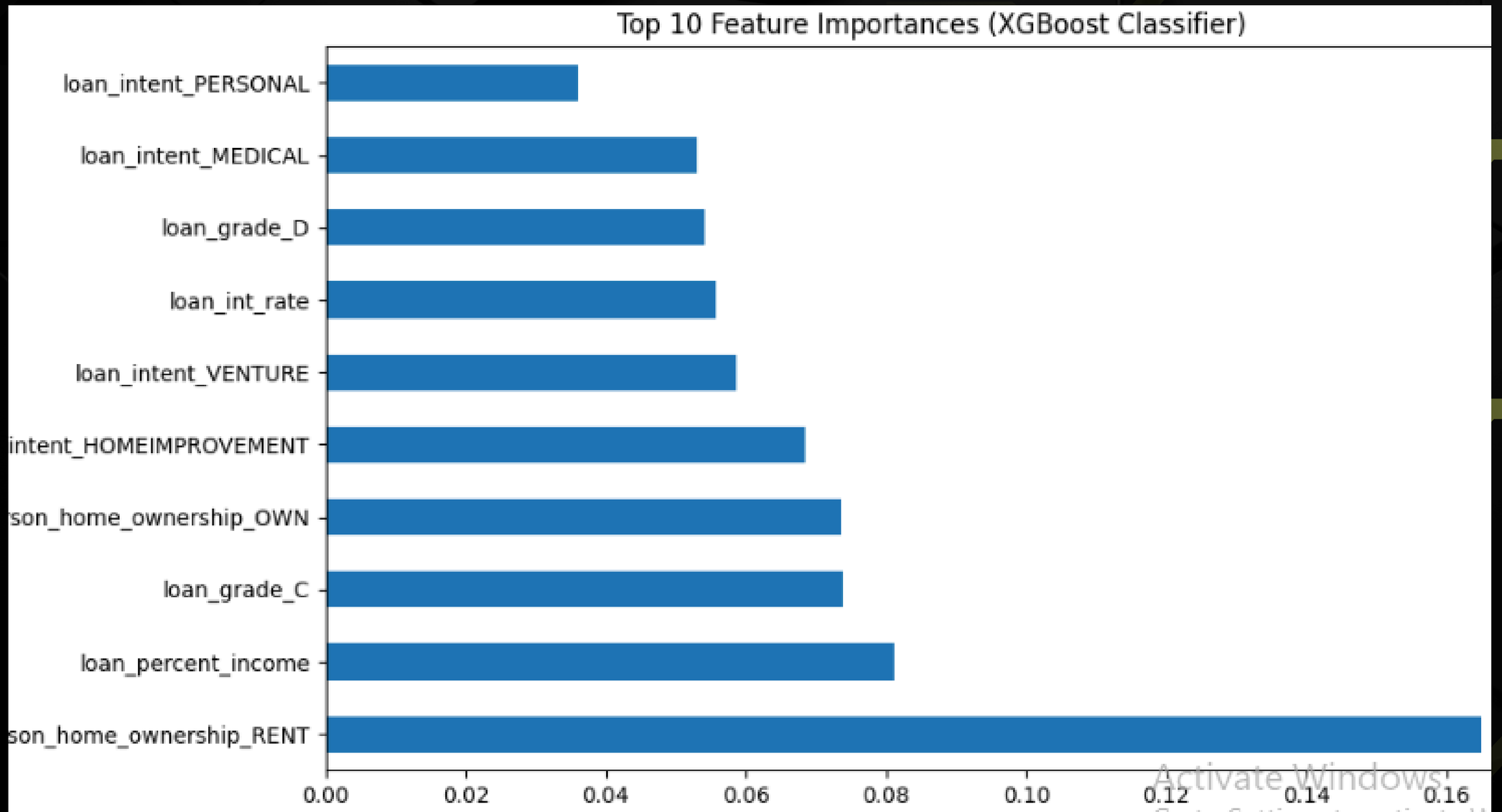
Credit risk classification



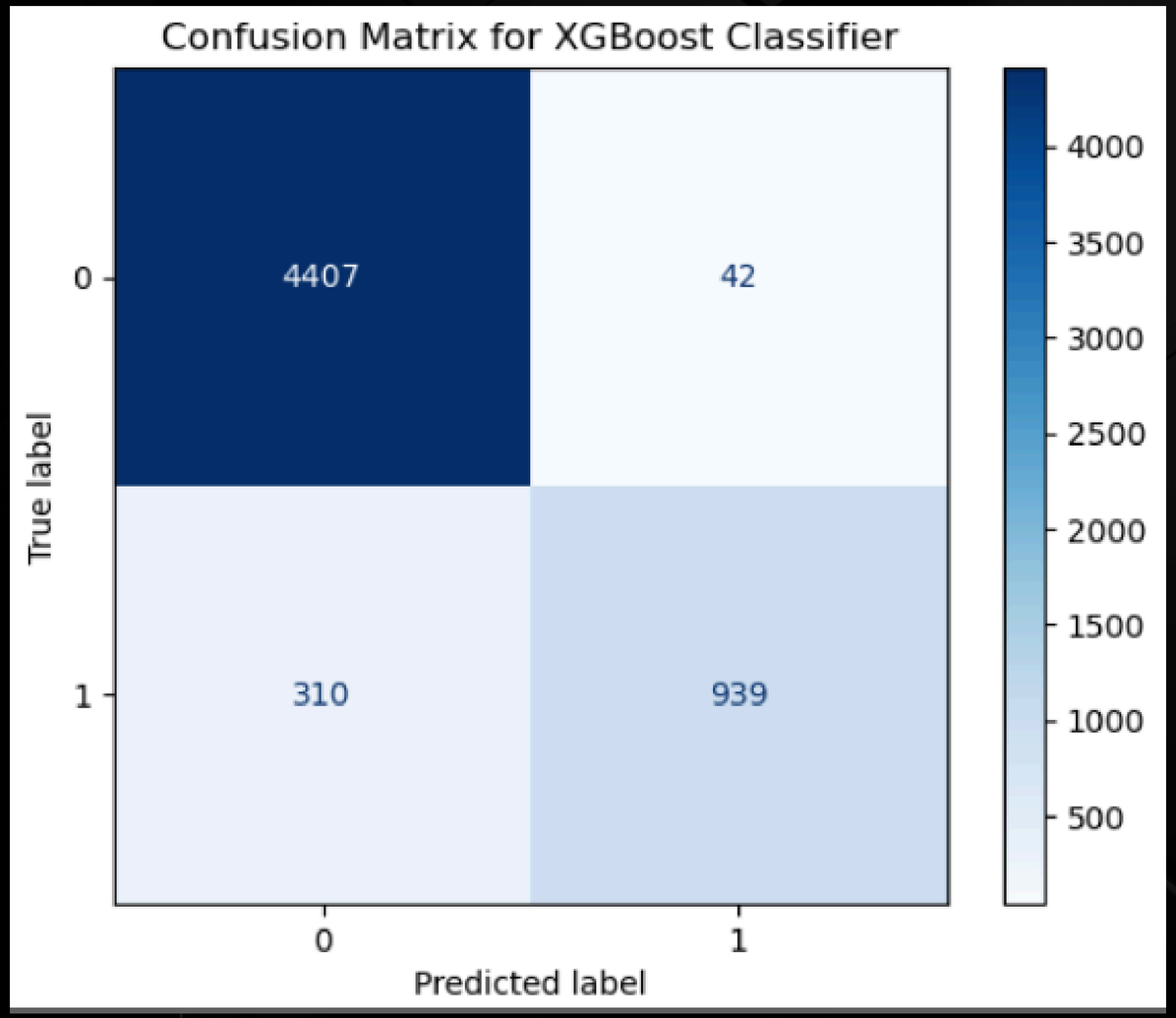
Metric scores for the various models.

next slide →

Top ten important features in XGBoost classifier model



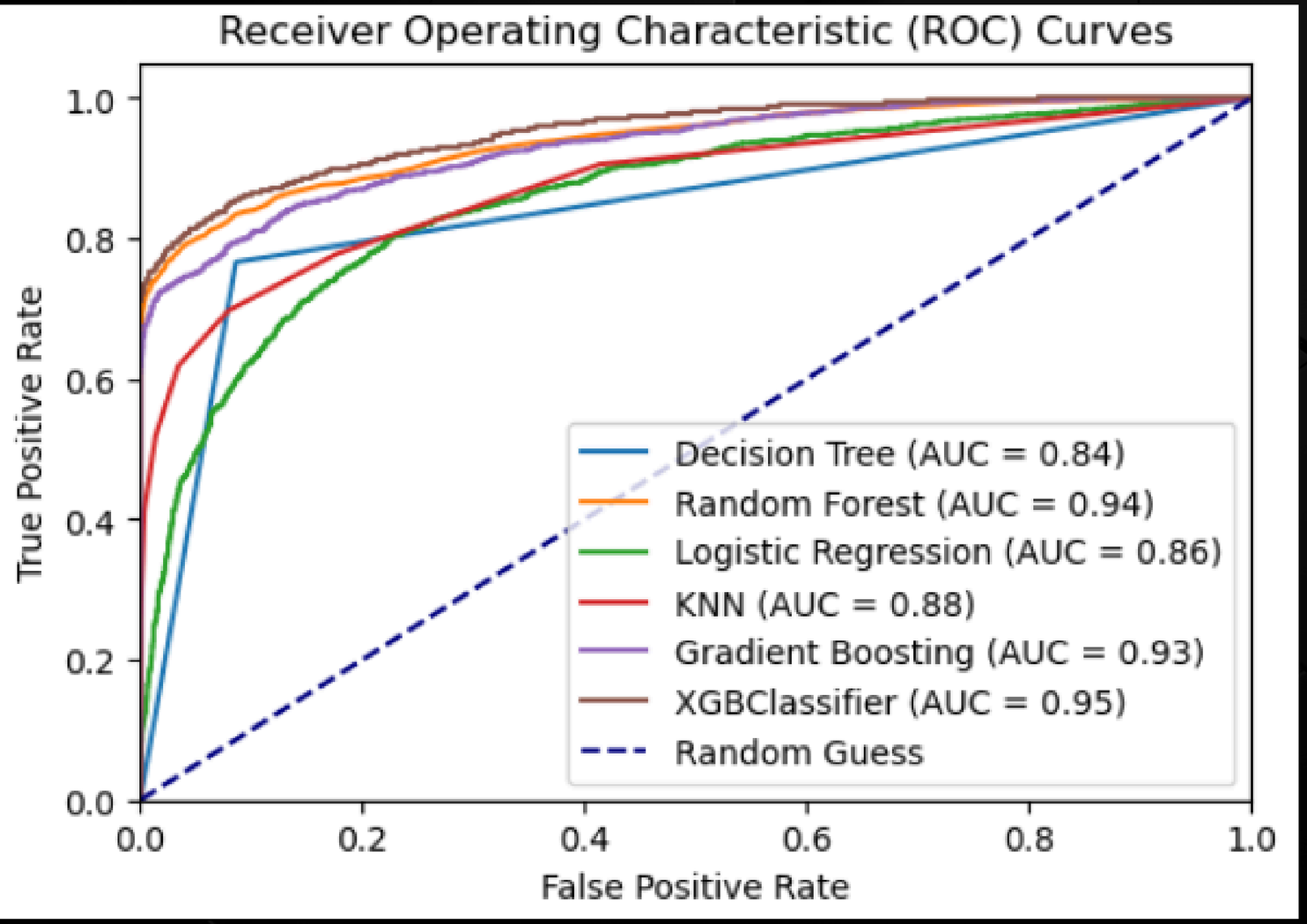
XGBoost classifier confussion matrix



The confussion matrix indicates that 4,407 paid loans and 939 unpaid loans were correctly classified and only 310 loans were misclassified.

next slide →

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

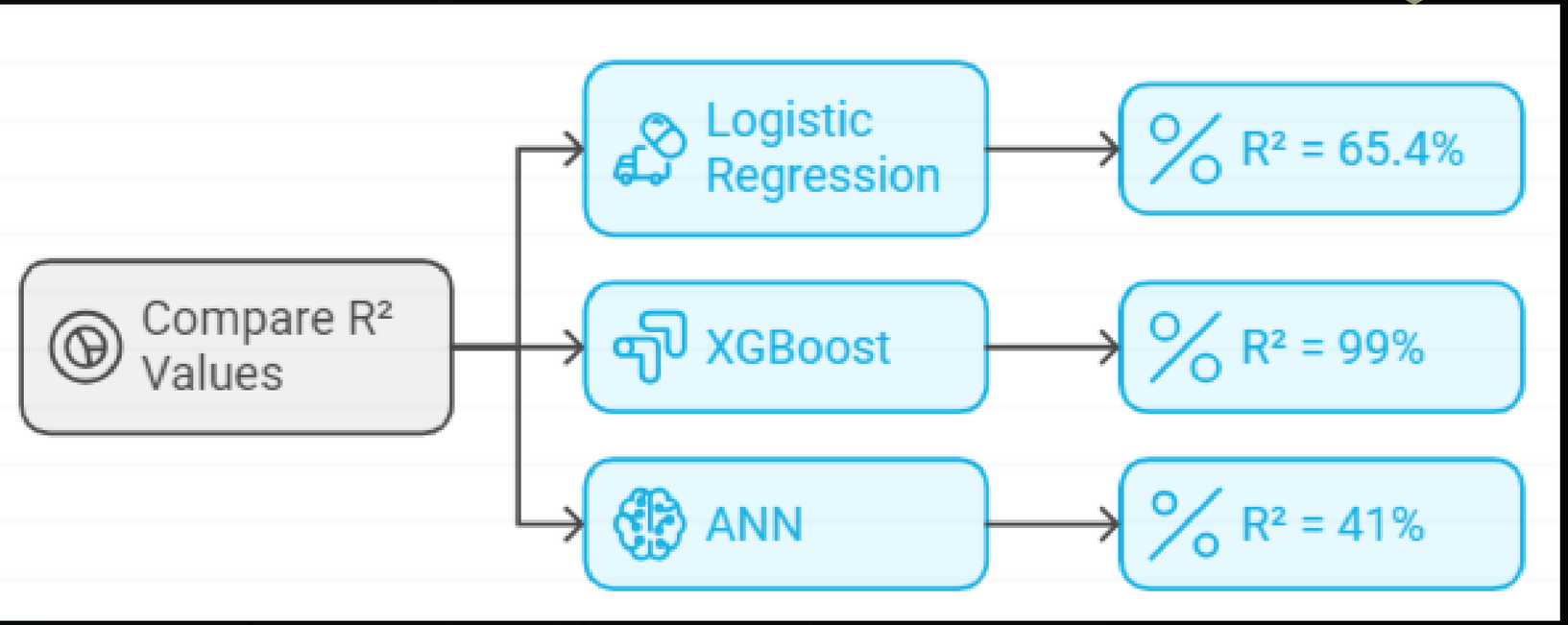
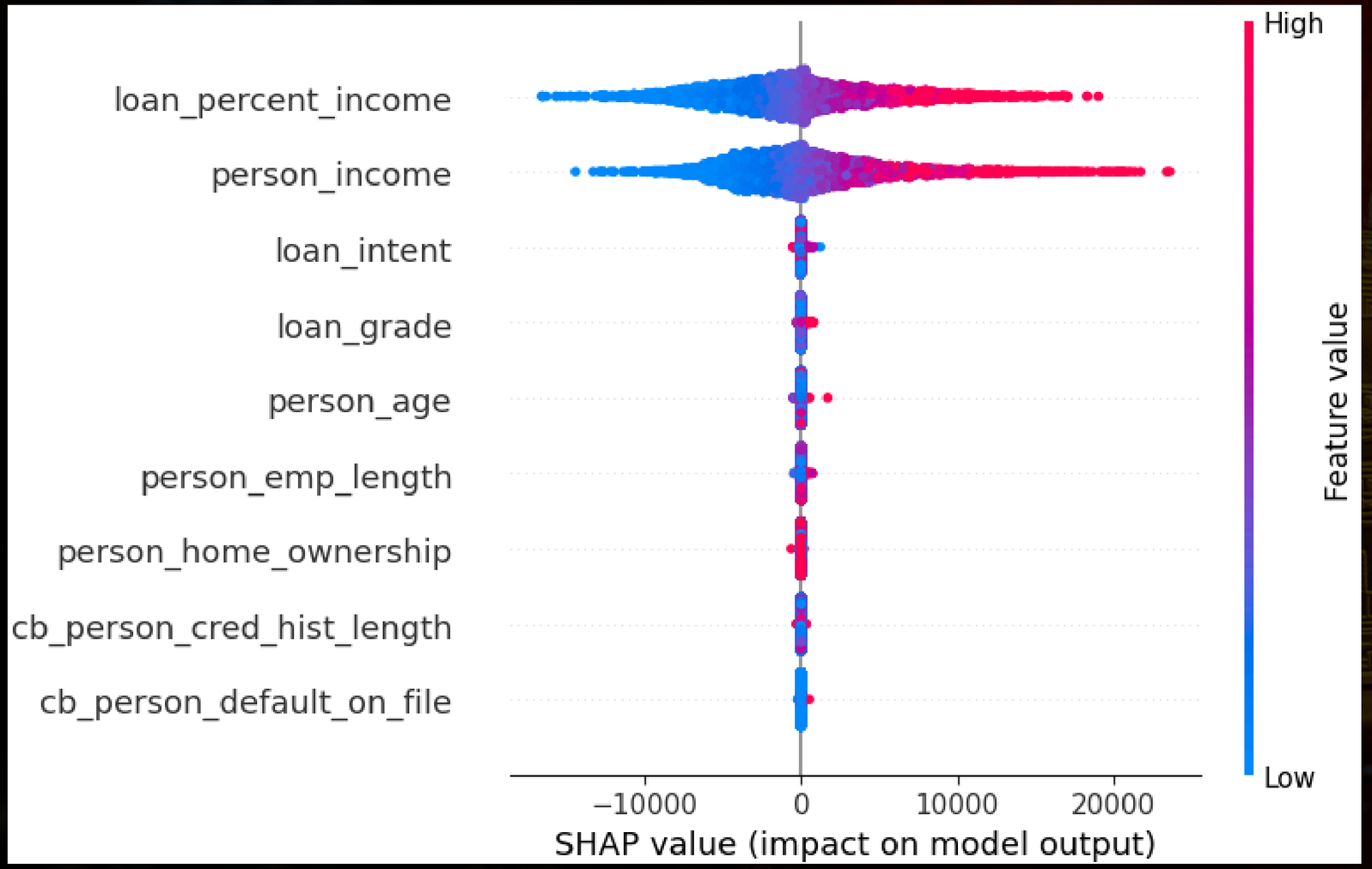


Chart of R squared values in percentage form for the various models. As evident from the chart XGBoost has the highest R squared suggesting it explains 99% of the variance in the target variable.

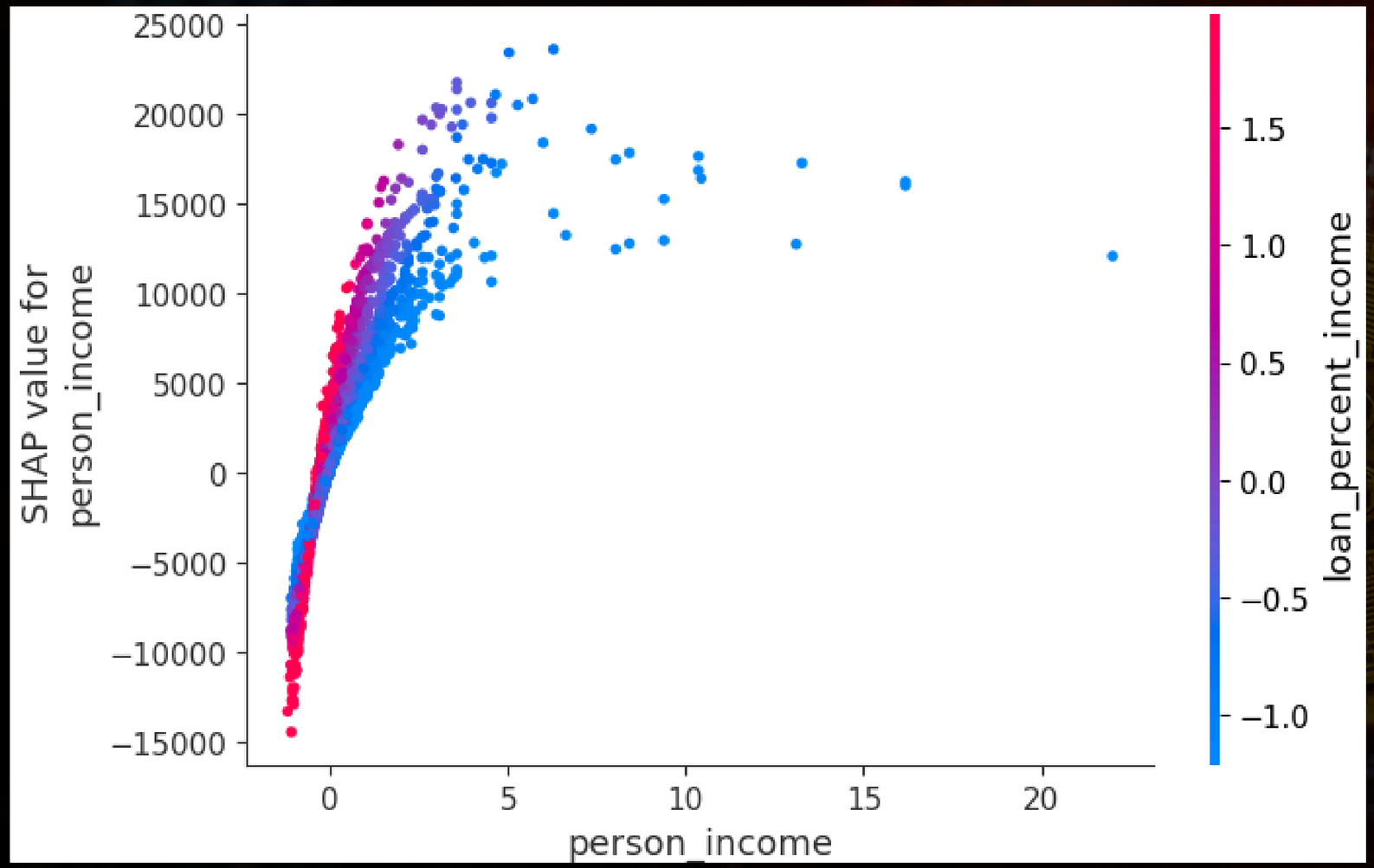
Model Interpretability



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.

next slide →

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 precentage of income.

next slide →

Recommendations

- Develop tailored loan pruducts 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.

Thank you.

grace.gitau2@student.moringaschool.com

0745021147