



The 13th Undergraduate in Applied Mathematics Conference (UAMC 2025)
Department of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang

Comparison of Loan Approval Prediction Models Using Feature Selection with Machine Learning

Sirikorn Sunthep, Tanatchaya Tueyot and Sirilak Srichaichana
Advisor: Acting Capt.Dr.Chalermrat Nontapa
Data Science Research Center, Faculty of Science, Chiang Mai University

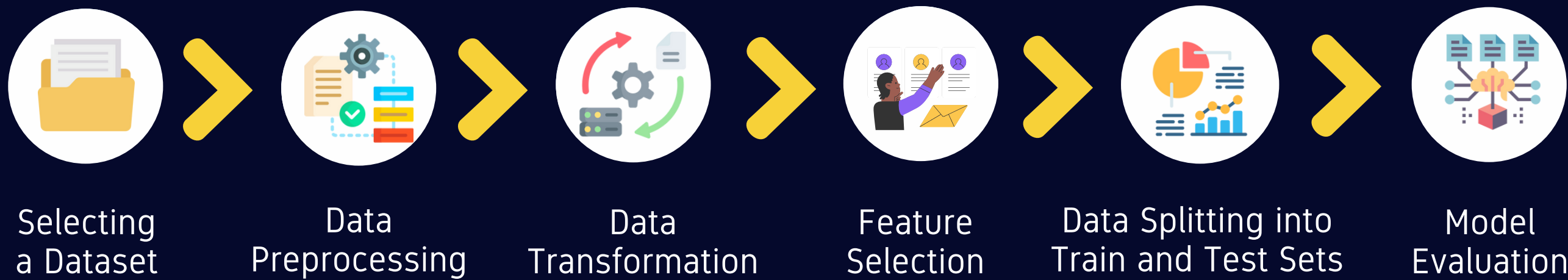
ABSTRACT

The primary objective of this study is to develop the most suitable model for predicting loan approval using machine learning techniques to enhance the efficiency of loan applicant screening and reduce the risk of granting loans that may not be repaid this study utilizes a dataset of 4,269 records and applies four machine learning models Random Forest, LightGBM, CatBoost, and XGBoost to compare their performance and identify the most accurate model for loan approval prediction additionally the Stepwise Selection method is employed to select the most influential variables affecting loan approval reducing model complexity and improving analytical accuracy the results indicate that the LightGBM model performs the best achieving an Accuracy of 99.06% a Precision of 99.36% a Recall of 98.11% and an F₁-Score of 98.73% compared to the other models these findings suggest that LightGBM is the most suitable choice for predicting loan approval the insights from this study can be utilized to enhance the accuracy of loan approval processes reduce the rate of non-performing loans and enable financial institutions to manage risks more effectively furthermore future research may incorporate Deep Learning techniques to further enhance analysis and prediction capabilities improving accuracy and minimizing errors in assessing applicants' credit risk.

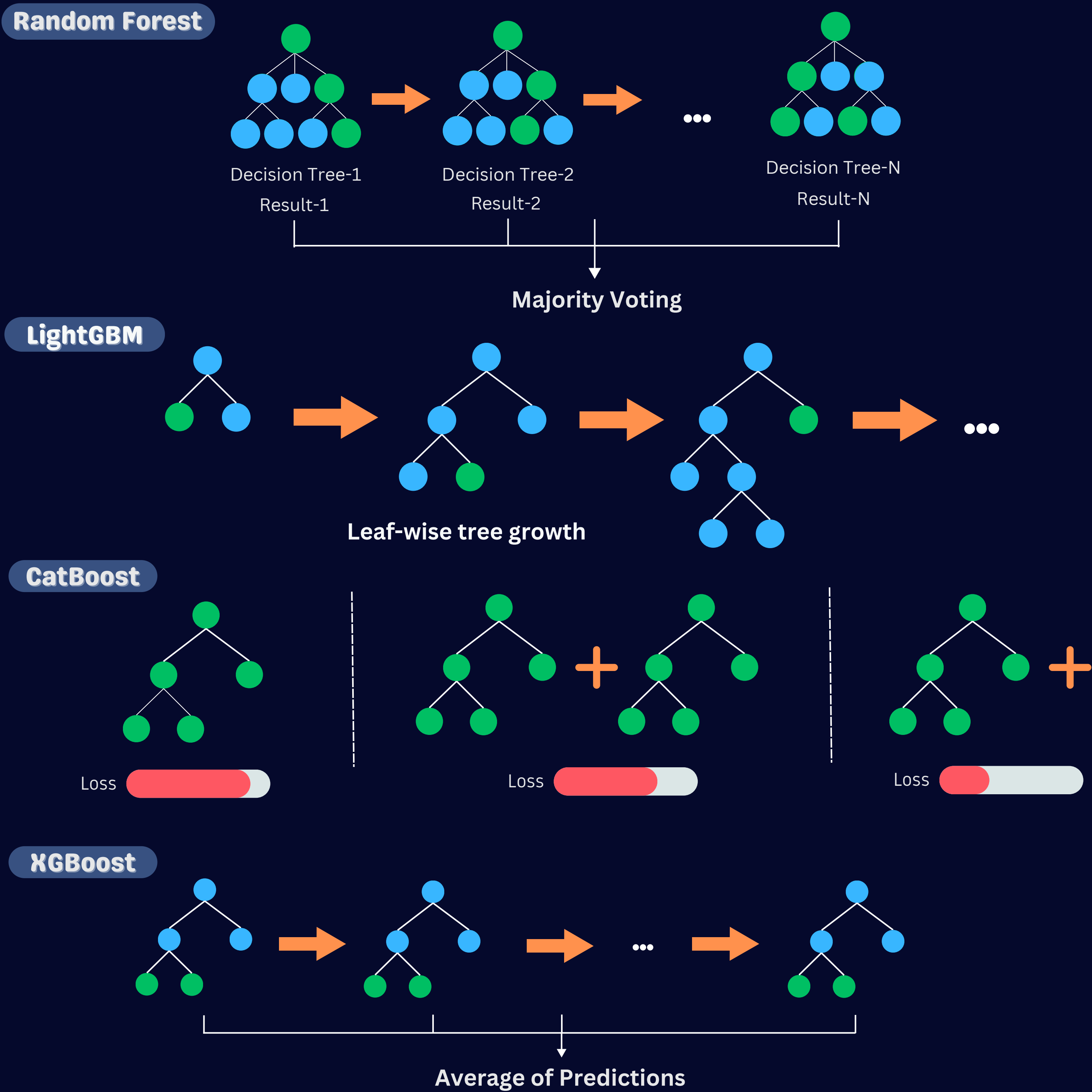
INTRODUCTION

Household debt in Thailand has been continuously increasing, making loan applications a common option for many individuals to manage debt or cover expenses exceeding their current income. However, an inaccurate loan approval process can lead to financial risks and higher default rates. This study focuses on identifying key factors influencing loan approval and predicting loan approval by testing four models: Random Forest, LightGBM, CatBoost, and XGBoost, to find the most effective model. The findings from this study will enhance the efficiency of loan applicant screening, reduce the risk of loan defaults, and support more informed decision-making for financial institutions.

METHODOLOGY



MODELS



EVALUATION CRITERIA

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

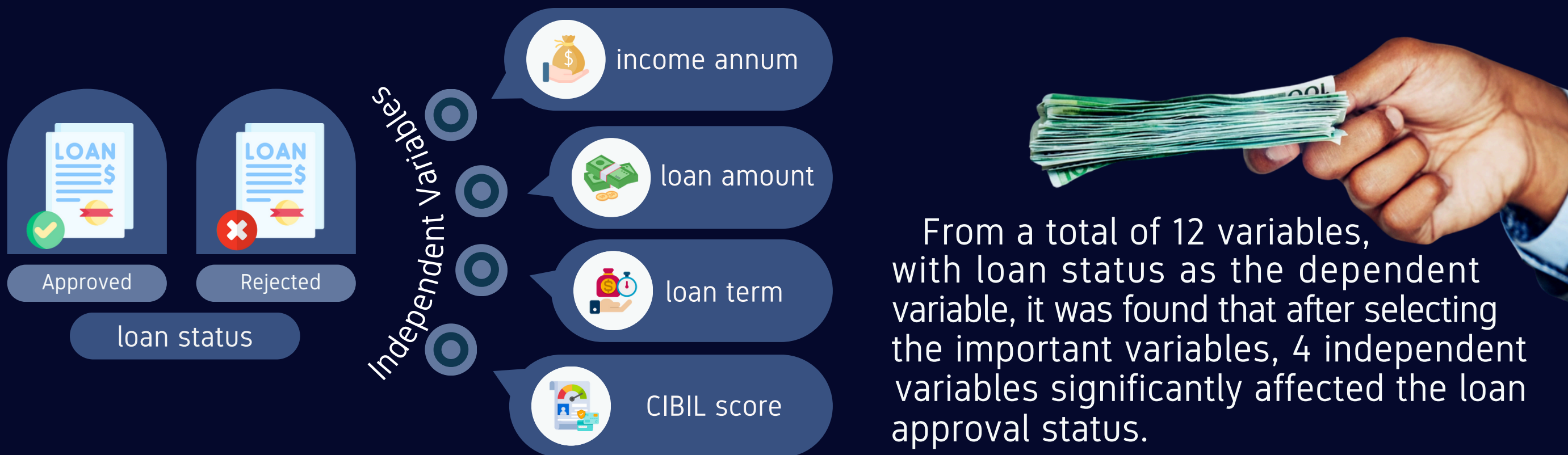
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F_1\text{-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

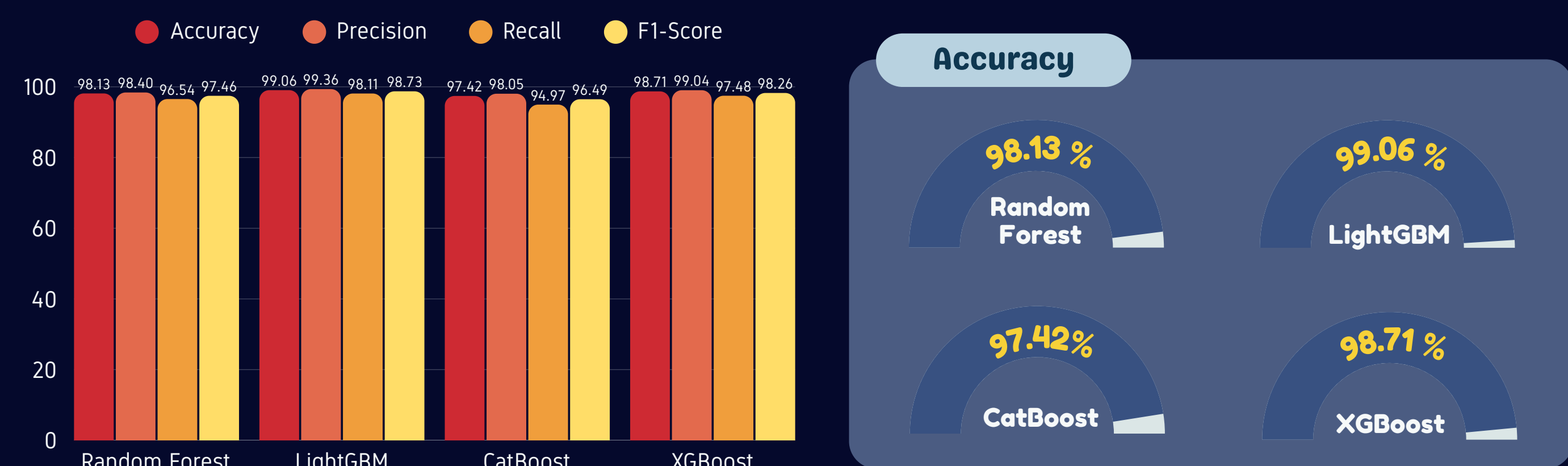
OBJECTIVES

- To study the relationships among various factors influencing loan approval.
- To compare and build four prediction models for loan approval based on various factors, and ultimately select the most suitable model for accurate prediction.

FACTORS



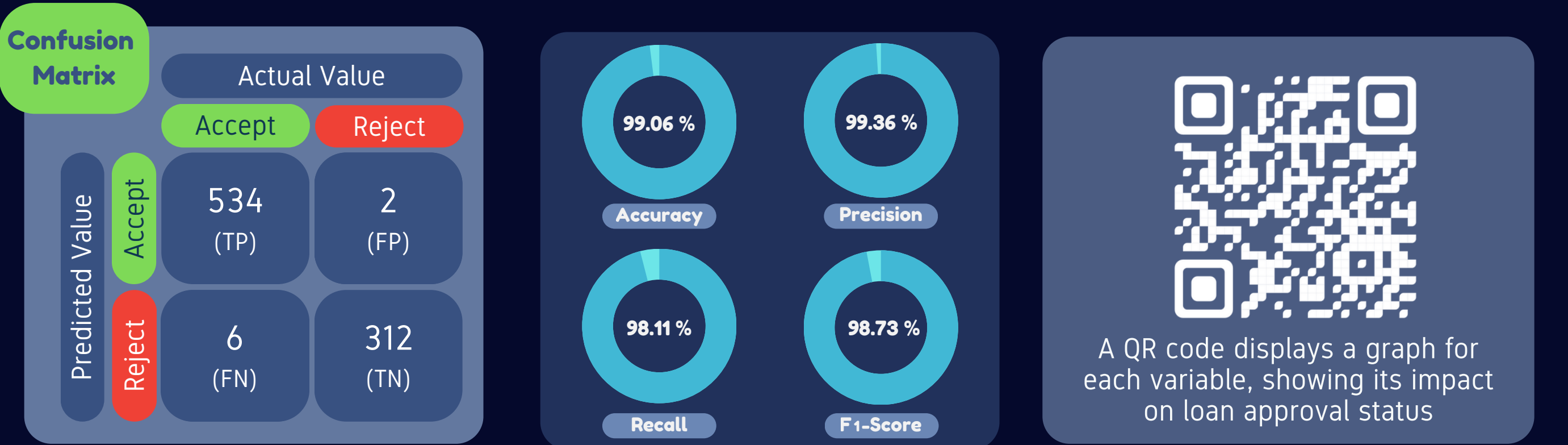
RESULT



The test results indicated that the LightGBM model performed the best, achieving the highest values for Accuracy, Precision, Recall and F₁-Score, with an Accuracy of 99.06%, a Precision of 99.36%, a Recall of 98.11%, and a F₁-Score of 98.73%. This highlights LightGBM's superior performance across key evaluation metrics, confirming it as the most suitable model for providing accurate and dependable loan approval predictions in this study.

CONCLUSION & DISCUSSION

The results of the confusion matrix showed that LightGBM correctly predicted loan approval in 99.06% of cases, accurately classifying 4,229 out of 4,269 people, while misclassifying 0.94% of cases, or 40 people.



From the study, it was found that the LightGBM model was the most suitable for predicting the occurrence of loan approval status. This model achieved performance metrics with an Accuracy of 99.06%, Precision of 99.36%, Recall of 98.11%, and F₁-Score of 98.73%. These findings align with the research of (Xingzhe Dong, 2022; Teuku Risky Noviandy, et al., 2024), which also demonstrated that the LightGBM model outperformed other models, showing similarly high predictive performance metrics, consistent with this study.

References

[1] Noviandy, T. R., Idroes, G. M., & Hardi, I. (2024). Enhancing loan approval decision-making: An interpretable machine learning approach using LightGBM for digital economy development. Malaysian Journal of Computing, 9(1), 1734-1745.
[2] Dong, X. (2022). Loan default prediction based on machine learning (LightGBM model). BCP Business & Management, 25, 457-468.