



PROJECT REPORT

**OPTIMIZING RANDOM FOREST FOR MULTICLASS
CREDIT SCORING CLASSIFICATION WITH SMOTE AND
FEATURE SELECTION**

RAFFI DZAKY MAHENDRA
22.K3.0003

Faculty of Computer Science
Soegijapranata Catholic University
2026

ACKNOWLEDGMENT

Praise and gratitude to Allah SWT for all His grace, gifts, and guidance, enabling the author to successfully complete this thesis. This thesis was written as one of the requirements for obtaining a Bachelor's degree in Faculty of Computer Science Soegijapranata University, Semarang.

The process of writing this thesis was neither quick nor easy. Therefore, the author fully acknowledges that the success of this thesis would not have been possible without the assistance, support, and guidance of various parties.

The author would like to express sincere gratitude to Dr. Yulianto Tejo Putranto S.T., M.T., as the thesis supervisor, for his valuable guidance, constructive feedback, and continuous support throughout the thesis writing process. His insights, suggestions, and encouragement played an important role in helping the author refine the research methodology, improve the experimental process, and strengthen the analysis of the research results.

The author also extends appreciation to the beloved family, especially the parents, for their endless prayers, encouragement, and moral support. Their patience and motivation became the main source of strength that enabled the author to remain committed and complete this thesis.

In addition, the author would like to thank all friends and parties who cannot be mentioned individually for their support and encouragement throughout the academic journey and the completion of this thesis. Their presence, discussions, and support have been invaluable in overcoming various challenges during the research process.

Finally, the author acknowledges that this thesis still has limitations. Therefore, constructive suggestions are highly appreciated for future improvements. It is hoped that this thesis can provide benefits and contributions to the development of knowledge, particularly in the fields of artificial intelligence and machine learning.

ABSTRACT

Credit scoring is the process of evaluation and interpretation stages carried out by financial services institutions to determine a person's creditworthiness based on their financial history data. The main challenge in conducting credit scoring classification lies in the large number of interrelated features and the imbalance in class distribution. Most previous studies have only focused on improving algorithm performance, applying feature selection methods, and applying balancing data methods separately so that the effectiveness of the combination of these approaches, especially in the context of multiclass classification has not received adequate attention. This study attempts to fill this gap by optimizing the Random Forest algorithm through the application of the SMOTE method and two feature selection methods, namely Recursive Feature Elimination (RFE) and Mutual Information. A total of six modeling scenarios were used to evaluate the effect of using the feature selection and balancing data methods with the Random Forest algorithm to identify the best scenario in performing multiclass credit scoring classification. Model evaluation was carried out using various evaluation metrics such as accuracy, macro precision, macro recall, and macro F1-Score. The results of this study showed that the use of RFE provided the best performance with an accuracy of 83.21%, macro precision of 82.24%, macro recall of 82.86%, and macro F1-Score of 82.53%.

Keyword: Credit Scoring, Random Forest, SMOTE, Recursive Feature Elimination, Mutual Information