

## ABSTRACT

### ***IMPLEMENTATION OF FEATURE ENGINEERING AND HYPERPARAMETER TUNING TO IMPROVE THE ACCURACY OF RANDOM FOREST MODELS ON CREDIT RISK CLASSIFICATION***

By

Nadea Putri Nur Fauzi 20110031

*Credit risk analysis and modeling using machine learning is a special challenge because the data is complex, has a large size, consists of various types of features, contains outliers and imbalanced values. For this reason, research is needed using data mining techniques, especially through creating classification models using the Random Forest algorithm. The methodology used is SEMMA (Sample, Explore, Modify, Model, and Assess) which is a popular methodology in data mining. Through the application of the SEMMA methodology, this research specifically focuses on the use of feature engineering and hyperparameter tuning to improve model accuracy. Hyperparameter tuning is used to explore various parameter combinations in Random Forest, while feature engineering is used to improve data quality. The engineering features used include oversampling and standardization. Parameter selection by hyperparameter tuning the Random Forest model using the Random Search and Grid Search methods. The dataset used is credit risk data with a size of 32581 rows with 12 columns. The results of this study show a significant increase in model accuracy after implementing feature engineering and hyperparameter tuning. The best model obtained is a model that applies feature engineering and hyperparameter tuning with Grid Search CV to get an accuracy of 97.94%, an increase of 5.38% from the baseline model.*

**Keywords:** *Classification, Random Forest, Hyperparameter Tuning, Feature Engineering, Imbalanced*