Prediction of University Student Performance Based on Tracer Study Dataset Using Artificial Neural Network

Zahrina Aulia Adriani[1], Irma Palupi[2]

[1,2]School of Computing, Telkom University, Bandung
[1]zahrinaadriani@students.telkomuniversity.ac.id, [2]irmapalupi1@telkomuniversity.ac.id,

Abstract
In order to increase student performance, several universities use machine learning to analyze and evaluate their data so that it enables to improve the quality of education in the university. To get a new insight from the tracer study dataset as the relevance between university performance and student capability with business and industries work, the author will develop a model to predict student performance based on the tracer study dataset using Artificial Neural Network (ANN). For obtaining attributes that correspond to labels, Phi Coefficient Correlation will be used to select the attributes with high correlation as Feature Selection. The author is also performing the oversampling method using Synthetic Minority Oversampling Technique (SMOTE) because this dataset is imbalanced and evaluates the model using K-Fold Cross-Validation. According to K-Fold Cross Validation, the result shows that K = 3 has a low standard deviation of evaluation score and it's the best candidate of K to split the dataset. The average standard deviation is 0.038 for all score evaluations (Accuracy, Precision, Recall, and F-1 Score). After applied SMOTE to treating the imbalanced dataset with the data splitting 65 training data and 35 testing data, the accuracy value increases by 10% from 0.77 to 0.87.

Keywords: Artificial Neural Network, Imbalanced Dataset, K-Fold Cross-Validation, Student Performance, Tracer Study.