

ABSTRACT

Name : Fina Sifaul Nufus
NIM : 14230025
Study of Program : Computer Science
Faculty : Information Technology
Levels : Master Degree (S2)
Concentration : *Software Engineering & Data Science*
Title : *Cross-project Defect Prediction* pada Dataset AEEEM
Menggunakan Hybrid SMOTE–Tomek dan Ensemble
Learning

This study aims to improve the performance of *Cross-project Defect Prediction* (CPDP) on the AEEEM dataset by applying a hybrid *preprocessing* approach that combines feature normalization, dimensionality reduction using PCA, class balancing with SMOTE–Tomek, and decision *threshold* tuning. Experiments were conducted on five AEEEM projects, namely EQ, JDT, LC, ML, and PDE, under two main scenarios: *single-source* CPDP and *multi-source* CPDP. Random Forest and Support Vector Machine (SVM) were used as classification models, while performance was evaluated using *F1-score* and AUC metrics. The results show that the *multi-source* scenario generally provides more stable performance than the *single-source* scenario. Furthermore, the proposed hybrid *preprocessing* approach significantly improves *F1-score* compared to the *baseline*, especially for datasets with highly imbalanced class distributions. The ablation study confirms that the performance improvement is achieved through the combined effect of class balancing and *threshold* tuning rather than a single *preprocessing* component.

Keywords: AEEEM Dataset, Cross-project Defect Prediction, Hybrid Preprocessing, Machine Learning, SMOTE–Tomek