

## ABSTRACT

Krismon Noldiantus Liwe Tude (01679220008)

THE IMPLEMENTATION OF EDUCATIONAL DATA MINING IN PREDICTING STUDENTS' ACADEMIC ACHIEVEMENT IN MATHEMATICS AT A PRIVATE ELEMENTARY SCHOOL IN TANGERANG

(xii + 68 pages; 15 figures; 13 tables, 7 attachments)

The application of Educational Data Mining (EDM) to predict academic achievement of elementary students in Mathematics is discussed in this thesis. Ten Machine Learning classifiers including eight base learners (Linear SVM, Logistic Regression, Medium KNN, Wide NN, Fine Decision Tree, Bilayered NN, Fine KNN, and Medium NN) along with two ensemble learners (Ensemble Subspace Discriminant and Ensemble Boosted Trees) are explored within the MATLAB environment utilizing a dataset consisting of various 33 academic and demographic features of 280 students. To address the imbalanced distribution in the classes, resampling methods such as Random Under-Sampling Boost (RUSBoost) and Synthetic Minority Oversampling Technique (SMOTE), and both combination as hybrid-resampling techniques are utilized.

The experiment results show that the hybrid-sampling SMOTE-RUSBoosted Tree algorithm performs the best accuracy of 75% on testing data, suggesting that a combination of oversampling and undersampling techniques effectively models imbalanced datasets. The finding underscores the potential of EDM in elementary education setting to support data-driven interventions and to improve students' achievement in Mathematics.

Keywords: Educational Data Mining, Mathematics Achievement, Classification Algorithms, Ensemble Learning, Imbalanced Class, Resampling Methods

References 87 (2009-2023)

## ABSTRAK

Krismon Noldiantus Liwe Tude (01679220008)

PENERAPAN EDUCATIONAL DATA MINING DALAM MEMPREDIKSI  
CAPAIAN AKADEMIK SISWA PADA PELAJARAN MATEMATIKA DI  
SEBUAH SEKOLAH DASAR SWASTA DI TANGERANG

(xii + 68 halaman; 15 gambar; 13 tabel, 7 lampiran)

Tesis ini membahas tentang penerapan Educational Data Mining (EDM) untuk memprediksi capaian akademik dari siswa sekolah dasar pada pelajaran Matematika. Sepuluh model *Machine Learning* termasuk delapan *base learner* (*Linear SVM*, *Logistic Regression*, *Medium KNN*, *Wide NN*, *Fine Decision Tree*, *Bilayered NN*, *Fine KNN*, dan *Medium NN*) beserta dua model *Ensemble* (*Ensemble Subspace Discriminant and Ensemble Boosted Trees*) diterapkan pada platform MATLAB menggunakan dataset yang terdiri dari 33 fitur akademik dan demografis dari 280 siswa. Untuk mengatasi distribusi kelas yang tidak seimbang, metode resampling seperti *Random Under-Sampling Boost (RUSBoost)* dan *Synthetic Minority Oversampling Technique (SMOTE)*, dan gabungan keduanya (*hybrid sampling*) juga digunakan pada tesis ini.

Hasil eksperimen menunjukkan bahwa algoritma *hybrid-sampling*, *SMOTE-RUSBoosted Tree*, mencapai akurasi terbaik sebesar 75% pada data uji. Penelitian ini menunjukkan bahwa kombinasi teknik *oversampling* dan *undersampling* secara efektif dapat memodelkan data yang tidak seimbang. Hasil temuan ini juga menggarisbawahi potensi EDM pada lingkungan pendidikan dasar untuk mendukung intervensi berbasis data dalam meningkatkan capaian akademik pada pelajaran Matematika.

Kata Kunci: *Educational Data Mining, Mathematics Achievement, Classification Algorithms, Ensemble Learning, Imbalanced Class, Resampling Methods*

Referensi: 87 (2009-2023)