

ABSTRACT

Metacognition affects how students behave in class and influences the studying performance level. For detecting the metacognition factors of a student by Educational Data Mining (EDM), it is better to be implemented in the same class, the same subject or in the same learning environment level. But the limitation of this implementation is formed a small dataset. Small dataset has main weakness called class imbalance. The serious effect of class imbalance is low performance prediction in learning algorithm, because most of them only focus on majority class.

This research proposed a new algorithm whose combinations of some prior researches in class imbalanced handling and was tested in metacognition dataset whose respondents from the students of Information Technology UGM. The working principle of this algorithm was balancing the instances number of both majority and minority classes by SMOTE (Synthetic Minority Oversampling Technique), and also adding the dataset information by synthetic attributes through MTD (Mega Trend Diffusion). This algorithm was then tested by ensemble algorithm approach called Stacking between Support Vector Machine (SVM) and Naive Bayes.

The result of this work shows that this proposed algorithm has more optimal performance in class imbalance handling in metacognition dataset than some prior researches. Moreover in 95% of the degree of belief, there is enough proofs to show that testing by Stacking has better performance than without Stacking in some performance metrics. From this research is also proved that all metacognition factors affecting academic performance based on PCA (Principal Component Analysis) feature selection. Academic performance is lower when the metacognitive awareness is lower too.

Key words: stacking, class imbalance, MTD, SMOTE, metacognition.

INTISARI

Faktor metakognisi mempengaruhi bagaimana siswa berperilaku di kelas dan mempengaruhi tingkat performa di dalam akademik. Untuk mendeteksi pengaruh metakognisi dengan menggunakan Educational Data Mining (EDM), sebaiknya diimplementasi pada level kelas, mata pelajaran atau level lingkungan pendidikan yang sama. Namun imbas dari implementasi tersebut adalah terbentuknya dataset kecil. Kekurangan dari dataset kecil adalah ketidakseimbangan kelas. Efek serius dari ketidakseimbangan kelas adalah rendahnya performa prediksi oleh algoritme klasifikasi karena kebanyakan dari mereka hanya fokus pada kelas mayoritas.

Penelitian ini mengajukan sebuah algoritme baru yang bekerja dengan mengkombinasikan berbagai penelitian sebelumnya dan diuji pada dataset metakognisi dengan responden mahasiswa program studi Teknologi Informasi UGM. Prinsip kerja algoritme ini berupa penyeimbangan jumlah kelas mayoritas dan minoritas dengan memanfaatkan SMOTE (Synthetic Minority Oversampling Technique) dan penambahan informasi dataset dengan atribut sintesis dengan MTD (Mega Trend Diffusion). Algoritme tersebut diuji dengan pendekatan algoritme *ensemble* yakni Stacking antara Support Vector Machine (SVM) dan Naive Bayes.

Hasil dari penelitian ini menunjukkan bahwa algoritme yang diajukan dapat lebih optimal menangani ketidakseimbangan kelas di dataset metakognisi dibandingkan berbagai penelitian sebelumnya. Selain itu dengan tingkat kepercayaan 95%, terdapat cukup bukti untuk menunjukkan bahwa pengujian dengan Stacking lebih unggul dibandingkan dengan pengujian tanpa Stacking di berbagai metrik performa. Dari penelitian ini pula dapat dibuktikan bahwa seluruh atribut faktor metakognisi mempengaruhi performa akademik berdasarkan seleksi fitur Principal Component Analysis (PCA). Semakin rendah performa akademik seseorang, maka semakin rendah kesadaran metakognisinya.

Kata kunci –stacking, ketidakseimbangan kelas, MTD, SMOTE, metakognisi.