

ABSTRAK

Ayu Nandia Lestari, I Gusti (2023), *Model Rekomendasi Pemilihan Konsentrasi Pada Program Studi Dengan Kombinasi Metode Multiclassifier Voting (Studi Kasus : Program Studi Sistem Informasi)*.

Tesis ini sudah disetujui dan diperiksa oleh pembimbing I : Prof. Dr. Ir. Dewa Gede Hendra Divayana, S.Kom., M.Kom. dan Pembimbing II : Kadek Yota Ernanda Aryanto, S.Kom., M.T., Ph.D.

Kata kunci : *Machine Learning, Oversampling, SMOTE, Multiclassifier-Voting, Program Studi*.

Proses pemilihan konsentrasi seringkali rumit dan membingungkan, terutama ketika mahasiswa dihadapkan pada banyak pilihan konsentrasi yang tersedia. Selain itu, permasalahan distribusi kelas dalam *dataset* menunjukkan tidak seimbang di mana kelas minoritas memiliki representasi yang sangat sedikit dibandingkan dengan kelas mayoritas. Oleh karena itu, untuk membantu mahasiswa dalam memilih konsentrasi yang sesuai dengan kemampuannya maka, dalam penelitian ini dikembangkan sebuah model rekomendasi pemilihan konsentrasi program studi dengan kombinasi *Multiclassifier Voting* dan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*. Pada penelitian ini menggunakan *Ensemble Learning* jenis *voting* dengan *hard voting*. Dimana algoritma yang akan dikombinasikan yaitu metode klasifikasi *Support Vector Machine, Decision Tree*, dan *Random Forest*. Metode klasifikasi *Random Forest* bertujuan menangani dataset yang beragam seperti fitur pada nilai matakuliah. Lalu pada metode *Support Vector Machine* digunakan mendukung klasifikasi untuk pengolahan data berdasarkan nilai matakuliah umum dengan matakuliah khusus persemester yang menyangkut dengan pemilihan konsentrasi. Pada metode *Decision Tree* menangani masalah dataset distribusi yang tidak teratur. Sedangkan metode *Ensemble Learning* digunakan untuk meningkatkan nilai akurasi dari model dengan teknik *voting* berdasarkan nilai dari tiga metode klasifikasi. *SMOTE* digunakan untuk menghasilkan sampel sintetis pada kelas minoritas, dengan tujuan untuk meningkatkan jumlah sampel dan memperluas representasi kelas minoritas. Kemudian performa dari masing-masing algoritma klasifikasi termasuk *ensemble learning* dievaluasi menggunakan metrik evaluasi yang relevan, seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil perhitungan dengan *SMOTE* mendapatkan peningkatan jumlah dataset dari awalnya 563 menjadi 912 yang terbagi untuk empat kelas yaitu ET memiliki jumlah data sebanyak 228, GM sebanyak 228, ES sebanyak 228 dan IS sebanyak 228. Selanjutnya hasil eksperimen menunjukkan bahwa penggunaan kombinasi *SMOTE* atau data yang sudah di *oversampling* dengan metode klasifikasi dapat meningkatkan performa akurasi model. Berdasarkan pengujian diperoleh hasil akurasi yang paling baik yaitu pada metode klasifikasi *Random Forest* dengan *SMOTE* yaitu 88% dan tanpa *oversampling* 82%, kemudian klasifikasi *Decision Tree* dengan *SMOTE* yaitu 82% dan tanpa *oversampling* 75%, lalu klasifikasi *SVM* dengan *SMOTE* yaitu 85% dan tanpa *oversampling* yaitu 77%. lalu klasifikasi *Ensembl Voting* dengan *SMOTE* yaitu 87% dan tanpa *oversampling* yaitu 80%.

ABSTRACT

Ayu Nandia Lestari, I Gusti (2023), Recommendation Model For Choosing Concentrations In A Study Program Using The Multiclassifier Voting Method Combination (Case Study: Information Systems Study Program)

This thesis has been approved and reviewed by: Primary Advisor: Prof. Dr. Ir. Dewa Gede Hendra Divayana, S.Kom., M.Kom. Secondary Advisor: Kadek Yota Ernanda Aryanto, S.Kom., M.T., Ph.D.

Keyword : Machine Learning, Oversampling, SMOTE, Multiclassifier-Voting, Program Studi.

The process of selecting a concentration is often complex and confusing, particularly when students are faced with numerous available concentration choices. Moreover, class distribution issues in the dataset indicate an imbalance, where minority classes are significantly underrepresented compared to the majority classes. Therefore, to assist students in choosing a concentration that suits their abilities, this research develops a recommendation model for selecting a study program concentration using a combination of Multiclassifier Voting and the Synthetic Minority Over-sampling Technique (SMOTE) approach. In this study, an Ensemble Learning technique, specifically hard voting, is employed. The algorithms to be combined include Support Vector Machine, Decision Tree, and Random Forest classification methods. The Random Forest classification method aims to handle diverse datasets, such as course grade features. The Support Vector Machine method is used for classifying data based on common course grades with specific concentration-related courses per semester. The Decision Tree method is employed to address irregular dataset distribution issues. Ensemble Learning is used to enhance the accuracy of the model through a voting technique based on the values from these three classification methods. SMOTE is used to generate synthetic samples in the minority class to increase the sample count and broaden the representation of the minority class. The performance of each classification algorithm, including ensemble learning, is evaluated using relevant evaluation metrics such as accuracy, precision, recall, and F1-score. The results from SMOTE show an increase in the dataset size from 563 to 912 samples, divided into four classes: ET with 228 samples, GM with 228 samples, ES with 228 samples, and IS with 228 samples. Furthermore, the experimental results indicate that using a combination of SMOTE or oversampled data with classification methods can enhance the model's accuracy performance. Based on the testing, the best accuracy results were obtained as follows: Random Forest classification with SMOTE achieved 88% accuracy and 82% without oversampling, Decision Tree classification with SMOTE achieved 82% accuracy and 75% without oversampling, SVM classification with SMOTE achieved 85% accuracy and 77% without oversampling, and Ensemble Voting classification with SMOTE achieved 87% accuracy and 80% without oversampling.