

PENERAPAN *BORDERLINE-SMOTE* DAN *GRID SEARCH* PADA *BAGGING-SVM* UNTUK KLASIFIKASI PENYAKIT DIABETES

ABSTRAK

Penyakit diabetes telah menjadi penyakit yang cukup berbahaya yang dapat menyebabkan kelumpuhan hingga mengancam jiwa penderitanya. Indonesia merupakan negara peringkat ke-7 diantara 10 negara yang memiliki jumlah penderita terbanyak dengan 10,7 juta penderita dari 172,2 juta total populasi orang dewasa pada tahun 2019. Salah satu langkah pencegahan yang bisa dilakukan untuk menghindari bahaya dari penyakit diabetes adalah melakukan prediksi dini yang dapat memprediksi penyakit diabetes dengan memanfaatkan *data mining* untuk membangun model klasifikasi. Salah satu algoritma klasifikasi yang telah banyak digunakan dalam penelitian untuk mendeteksi penyakit diabetes adalah *support vector machine*. Walaupun memberikan hasil generalisasi yang baik tetapi algoritma SVM memiliki kelemahan ketika diberikan data dengan kelas yang tidak seimbang dan sulitnya menentukan parameter yang optimal. Untuk mengatasi kekurangan klasifikasi pada data yang tidak seimbang dapat menggunakan metode *oversampling borderline-SMOTE* yang akan meningkatkan jumlah data pada kelas minor agar distribusi kelas menjadi seimbang. Dalam masalah optimasi parameter, dapat menggunakan metode *grid search*. Kemudian juga menerapkan algoritma *bagging* agar mendapatkan hasil klasifikasi yang lebih baik dengan menghindari *overfitting* dan mengurangi variansi. Hasil penelitian menunjukkan bahwa model yang dibentuk dengan algoritma SVM, *bagging*, *borderline-SMOTE* dan *grid search* mendapat akurasi sebesar 92,1%, nilai *precision* sebesar 95,51% untuk kelas sehat dan 86,12% untuk kelas diabetes, nilai *recall* sebesar 92,32% untuk kelas sehat dan 91,66% untuk kelas diabetes, dan nilai *f1-score* sebesar 93,39% untuk kelas sehat dan 88,81% untuk kelas diabetes.

Kata kunci: klasifikasi, diabetes, svm, *bagging*, *borderline-smote*, *grid search*

APPLICATION OF BORDERLINE-SMOTE AND GRID SEARCH ON BAGGING-SVM FOR CLASSIFICATION OF DIABETES

ABSTRACT

Diabetes has become a fairly dangerous disease that can cause paralysis to threaten the life of the sufferer. Indonesia is ranked 7th among the 10 countries with the highest number of sufferers with 10.7 million sufferers out of 172.2 million total adult population in 2019. One of the preventive steps that can be taken to avoid the dangers of diabetes is to predict which can predict diabetes by utilizing data mining to build a classification model. One of the classification algorithms that has been widely used in research to detect diabetes is a support vector machine. Although it gives good generalization results, the SVM algorithm has weaknesses when given data with unbalanced classes and it is difficult to determine optimal parameters. To overcome the lack of classification on unbalanced data, we can use the borderline-SMOTE oversampling method which will increase the amount of data in the minor class so that the class distribution becomes balanced. In parameter optimization problems, you can use the grid search method. Then also apply the bagging algorithm to get better classification results by avoiding overfitting and reducing variance. The results showed that the model formed with the SVM, bagging, borderline-SMOTE and grid search algorithms got an accuracy of 92,1%, a precision value of 95,51% for a healthy class and 86,12% for a diabetes class, a recall value of 92,32% for a healthy class and 91,66 % for diabetes class, and f1-score value of 93,39% for healthy class and 88,81% for diabetes class.

Keywords: *classification, diabetes, svm, bagging, borderline-smote, grid search*