

ABSTRACT

EVALUATION OF SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) PERFORMANCE TO OVERCOME UNBALANCED DATA CLASSIFICATION IN K-NEAREST NEIGHBOR (KNN) AND SUPPORT VECTOR MACHINE (SVM) METHODS

OLEH

WIDYA AMALIA PUTRI RISWANDHA

K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) methods are one of the most popular techniques in classification. In general, a common problem with classification is the presence of data imbalances. Data imbalance is a condition in which the distribution in the number of classes is uneven or can be interpreted as having a significant difference in the number of classes. It affects the performance of the classification results. So from that, it's important to address data imbalances, one of which is using the Synthetic Minority Oversampling Technique. (SMOTE). The study aims to study and evaluate the performance of SMOTE to address the imbalanced classification of data in KNN and SVM. Based on the results of the analysis obtained that SMOTE is effective in improving the performance of the classification of diabetic patients which is proven by the presence of increased accuracy value in the method KNN with parameter $k = 9$ obtaining the accurate value of 81.58% and in the SVM method with the RBF kernel obtaining the accurate value of 85,53%.

Keywords: Classification, Imbalance Data, KNN, SVM, SMOTE

ABSTRAK

EVALUASI PERFORMA *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE) UNTUK MENGATASI KLASIFIKASI DATA TIDAK SEIMBANG PADA METODE *K-NEAREST NEIGHBOR* (KNN) DAN *SUPPORT VECTOR MACHINE* (SVM)

OLEH

WIDYA AMALIA PUTRI RISWANDHA

Metode *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) merupakan salah satu teknik yang populer pada klasifikasi. Pada umumnya, masalah yang sering terjadi pada klasifikasi yaitu adanya ketidakseimbangan data. Ketidakseimbangan data merupakan suatu keadaan dimana distribusi pada jumlah kelas tidak merata atau dapat diartikan bahwa terdapat perbedaan yang signifikan terhadap jumlah kelas. Hal ini mempengaruhi terhadap performa hasil klasifikasi. Maka dari itu, penting untuk mengatasi masalah ketidakseimbangan data, salah satunya dengan menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Penelitian ini bertujuan untuk mengkaji dan mengevaluasi performa SMOTE untuk mengatasi klasifikasi data tidak seimbang pada KNN dan SVM. Berdasarkan hasil analisis diperoleh bahwa SMOTE efektif untuk meningkatkan performa klasifikasi penderita diabetes yang dibuktikan dengan adanya peningkatan nilai akurasi pada metode KNN dengan parameter $k = 9$ didapatkan nilai akurasi sebesar 81.58% dan pada metode SVM dengan Kernel RBF didapatkan nilai akurasi sebesar 85.53%.

Kata Kunci: Klasifikasi, Ketidakseimbangan Data, KNN, SVM, SMOTE