

ABSTRACT

HANDLING OF IMBALANCE DATA WITH RANDOM OVERSAMPLING (ROS) IN CLASSIFICATION OF DIABETIC PATIENTS USING SUPPORT VECTOR MACHINE (SVM)

By

Dhifa Zhafirah

Diabetes mellitus is a health problem that often occurs in Indonesia, especially in Lampung Province. This disease is a condition in which the body does not produce enough or use the insulin hormone that carries glucose into the body's cells. The purpose of this research is to create a machine-learning model that can detect diabetes early using a Support Vector Machine (SVM). However, the dataset used in the research has data imbalance problems. Therefore, Random Oversampling (ROS) is used to overcome this problem. The results obtained from this study, ROS is able to handle imbalance data so that the accuracy value obtained reaches 96.43% (excellent classification) with the C-Classification model and Radial Basis Function (RBF) kernel, as well as sigma one and cost one parameter for the training data scheme 90% and 10% testing data. This accuracy value increases sharply compared to without ROS, which is only around 76%.

Keywords : Diabetes Melitus, Imbalance Data, Random Oversampling, Support Vector Machine

ABSTRAK

PENANGANAN *IMBALANCE* DATA DENGAN *RANDOM OVERSAMPLING* (ROS) PADA KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN *SUPPORT VECTOR MACHINE* (SVM)

Oleh

Dhifa Zhafirah

Diabetes melitus adalah salah satu masalah kesehatan yang sering terjadi di Indonesia khususnya Provinsi Lampung. Penyakit ini merupakan suatu kondisi dimana tubuh tidak cukup untuk menghasilkan atau menggunakan hormon insulin yang membawa glukosa ke dalam sel-sel tubuh. Tujuan penelitian ini adalah membuat model *machine learning* yang dapat mendeteksi dini penyakit diabetes menggunakan *Support Vector Machine* (SVM). Namun, pada dataset yang digunakan dalam penelitian memiliki masalah ketidakseimbangan data (*imbalance data*). Oleh karena itu, digunakan *Random Oversampling* (ROS) untuk mengatasi masalah tersebut. Hasil yang diperoleh dari penelitian ini, ROS mampu menangani *imbalance* data sehingga nilai akurasi yang didapatkan mencapai 96.43% (*excellent classification*) dengan model *type C-Classification* dan kernel *Radial Basis Function* (RBF), serta parameter *sigma* 1 dan *cost* 1 untuk skema data latih 90% dan data uji 10%. Nilai akurasi ini meningkat tajam dibandingkan tanpa ROS yang hanya sekitar 76%.

Kata Kunci : Diabetes Melitus, Ketidakseimbangan Data, Random Oversampling, Support Vector Machine