

ABSTRAK

Tiroid merupakan salah satu kelenjar yang memiliki peran penting dalam tubuh. Namun, terdapat gangguan pada kelenjar tiroid menyebabkan adanya penyakit tiroid yaitu hipertiroid dan hipotiroid . Klasifikasi merupakan salah satu teknik dari data mining yang dapat digunakan untuk membuat prediksi. Dalam klasifikasi ini menggunakan metode *Least Square Support Vector Machine* (LSSVM) dan seleksi fitur *Information Gain*. Data yang digunakan merupakan data pasien yang bersumber dari *UCI Machine Learning* dengan 31 fitur dan 3 kelas yaitu Normal, Hipertiroid , dan Hipotiroid. Dalam dataset ini terdapat jumlah data kelas yang tidak seimbang dan jumlah dimensi fitur yang tinggi. Tujuan dari penelitian ini dilakukan adalah mengetahui kinerja seleksi fitur *Information Gain* dan pengaruh penggunaan SMOTE untuk mengatasi data yang tidak seimbang. Pengujian ini menggunakan data latih dan data uji 90:10. Hasil menunjukkan penggunaan *Synthetic Minority Over-Sampling Technique* (SMOTE) dan seleksi fitur memberikan evaluasi kinerja tertinggi dengan akurasi 97.59%, *precision* 97.59%, *recall* 97.59%, dan f1-score sebesar 97.58%.

Kata kunci : *Information Gain*, Klasifikasi, LSSVM, Penyakit Tiroid.

ABSTRACT

Thyroid is one of the glands that has an important role in the body. However, there are disorders in the thyroid gland that cause thyroid disease, namely hyperthyroid and hypothyroid. Classification is one of the data mining techniques that can be used to make predictions. In this classification using the Least Square Support Vector Machine (LSSVM) method and Information Gain feature selection. The data used is patient data sourced from UCI Machine Learning with 31 features and 3 classes namely Normal, Hyperthyroid, and Hypothyroid. In this dataset there is an unbalanced amount of class data and a high number of feature dimensions. The purpose of this research is to determine the performance of Information Gain feature selection and the effect of using SMOTE to overcome unbalanced data. This test uses 90:10 training data and test data. The results show the use of Synthetic Minority Over-Sampling Technique (SMOTE) and feature selection provides the highest performance evaluation with 97.59% accuracy, 97.59% precision, 97.59% recall, and f1-score of 97.58%.

Keywords: Classification, Information Gain, LSSVM, Thyroid Disease.