

PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI UNTUK PREDIKSI PENYAKIT TIROID

Maulana Luthfi

ABSTRAK

Tiroid merupakan salah satu bagian tubuh yang sangat penting bagi manusia. Gangguan tiroid seringkali sulit dideteksi karena memiliki gejala yang sangat mirip dengan penyakit lain. Akibatnya pasien seringkali tidak menyadari memiliki masalah pada kelenjar tiroid.

Penelitian ini bertujuan untuk melakukan prediksi penyakit tiroid menggunakan data penyakit tiroid dari Garavan Institute dan J. Ross Quinlan. Data tersebut akan diolah menggunakan proses data mining menggunakan metode klasifikasi.

Metode klasifikasi yang digunakan pada penelitian ini adalah algoritma K-Nearest Neighbor, Naïve Bayes, Decision Tree, dan Random Forest. Penelitian ini juga menggunakan teknik *undersampling* untuk mengatasi jumlah kelas yang tidak seimbang. Keluaran dari sistem adalah *performance measure* yang disajikan dalam bentuk persentase berdasarkan nilai – nilai pada *confusion matrix* dari setiap algoritma terhadap data penyakit tiroid.

Berdasarkan hasil yang diperoleh, kinerja algoritma Random Forest memperoleh hasil paling baik dibandingkan dengan algoritma lainnya. Random Forest mendapat persentase akurasi sebesar 95,16%, sedangkan KNN, Naïve Bayes, dan Decision Tree hanya mendapat persentase akurasi sebesar 87,10%, 93,55%, dan 90,32%. Namun jika dilihat dari waktu proses pelatihan dan pengujian model, Decision Tree memiliki waktu proses paling cepat dibandingkan dengan algoritma lainnya. Decision Tree membutuhkan waktu 0,0091 detik, sedangkan KNN, Naïve Bayes, dan Random Forest membutuhkan waktu 0,0273 detik, 0,0103 detik, dan 0,0162 detik.

Oleh karena itu jika ingin mendapatkan nilai yang maksimal, maka dapat menggunakan algoritma Random Forest. Namun jika ingin menggunakan waktu proses yang lebih cepat, maka dapat menggunakan algoritma Decision Tree untuk melakukan prediksi penderita penyakit tiroid.

Kata kunci : *Tiroid, Data Mining, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Random Forest, Undersampling, Confusion Matrix, Performance Measure*

COMPARISON OF CLASSIFICATION ALGORITHM PERFORMANCE FOR THYROID DISEASE PREDICTION

Maulana Luthfi

ABSTRACT

Thyroid is a very important part of the body for humans. Thyroid disorders are often difficult to detect because their symptoms are very similar to those of other diseases. As a result, patients often do not realize they have problems with the thyroid gland.

This study aims to predict thyroid disease using thyroid disease data from the Garavan Institute and J. Ross Quinlan. The data will be processed using the data mining process using the classification method.

The classification methods used in this study are the K-Nearest Neighbor, Naïve Bayes, Decision Tree, and Random Forest algorithms. This study also uses undersampling techniques to overcome the unequal number of classes. The output of the system is a performance measure that is presented in the form of a proportion based on the values in the confusion matrix of each algorithm against thyroid disease data.

Based on the results obtained, the performance of the Random Forest algorithm obtained the best results compared to other algorithms. Random Forest got an accuracy proportion of 95.16%, while KNN, Naïve Bayes, and Decision Tree only got an accuracy proportion of 87.10%, 93.55% and 90.32%. However, when viewed from the training process time and the testing model, Decision Tree has the fastest processing time compared to other algorithms. Decision Tree takes 0.0091 seconds, while KNN, Naïve Bayes, and Random Forest takes 0.0273 seconds, 0.0103 seconds and 0.0162 seconds.

Therefore if you want to get the maximum value, you can use the Random Forest algorithm. However, if you want to use a faster processing time, you can use the Decision Tree algorithm to predict thyroid disease sufferers.

Keywords : *Thyroid, Data Mining, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Random Forest, Undersampling, Confusion Matrix, Performance Measure*