



ABSTRACT

Coronary artery disease (CAD) is one of the most dangerous diseases in the world, including in Indonesia. CAD occurs due to narrowing or blockage of the coronary arteries which is usually caused by Atherosclerosis. Atherosclerosis, also called hardening or blockage of the arteries, is a buildup of cholesterol or fat found in the inner walls of the arteries which limits blood flow from the heart and causes the arteries to function abnormally.

Various studies have been conducted with the aim of predicting the characteristics of this disease. One example of this study uses the Z-Alizadeh Sani dataset which consists of 55 attributes with two classification results, CAD and Normal. With too many attributes contained in this dataset, it becomes a problem in the development of CAD research. With many attributes contained in the dataset does not guarantee that all of these attributes have a high effect on the dataset as a whole. Feature selection is one way to reduce the number of attributes by removing attributes that have a low effect on the dataset. In this study, the Boruta method is used as a feature selection to minimize the attributes in the dataset.

This research shows that by reducing the attributes in the dataset through the feature selection process, the selected data from attributes with high importance to the dataset can produce a pretty good accuracy value of 92% using the SVM classification method using the Boruta feature selection approach. Other than measuring performance through accuracy, the Boruta method also produces more stable results in sensitivity and specificity values for CAD analysis compared to the Extra Tree, Linear SVC and Chi-Square feature selection methods with values more than 85%.

Key Word – Coronary Artery Disease, Feature Selection, Boruta, Classification, Extra Tree, Chi-Square, SVM, Linear SVC



INTISARI

Penyakit arteri koroner (CAD) adalah salah satu penyakit paling berbahaya di seluruh dunia, termasuk di Indonesia. CAD terjadi karena penyempitan atau penyumbatan arteri koroner yang biasanya disebabkan oleh aterosklerosis. Aterosklerosis, juga disebut pengerasan atau penyumbatan arteri, adalah penumpukan kolesterol atau lemak yang ditemukan di dinding bagian dalam arteri yang membatasi aliran darah dari jantung dan menyebabkan arteri berfungsi secara tidak normal.

Berbagai penelitian telah dilakukan dengan tujuan untuk memprediksi sifat dan karakteristik penyakit ini. Salah satu contoh penelitian ini menggunakan dataset Z-Alizadeh Sani yang terdiri dari 55 atribut dengan dua hasil klasifikasi, CAD dan Normal. Dengan terlalu banyak atribut yang terkandung dalam dataset ini menjadi satu masalah dalam pengembangan penelitian CAD. Dengan banyak atribut yang terkandung dalam dataset tidak menjamin bahwa semua atribut ini memiliki efek yang tinggi terhadap dataset secara keseluruhan. Pemilihan fitur adalah salah satu cara untuk mengurangi jumlah atribut yang ada dengan meninggalkan atribut yang memiliki efek tinggi pada dataset. Dalam penelitian ini, metode Boruta digunakan sebagai pemilihan fitur untuk meminimalkan atribut pada dataset.

Hasil pengujian yang dilakukan menunjukkan jika dengan mengurangi atribut dalam dataset melalui proses pemilihan fitur, data-data yang terpilih dari atribut dengan kepentingan yang tinggi dengan dataset mampu menghasilkan nilai akurasi yang cukup baik yakni sebesar 92% menggunakan metode klasifikasi SVM dengan menggunakan pendekatan seleksi fitur Boruta. Selain dengan mengukur kinerja melalui nilai akurasi, metode Boruta juga menghasilkan nilai sensitivitas dan spesivitas yang lebih stabil terhadap pengujian analisis CAD dibandingkan dengan metode seleksi fitur *Extra Tree*, *Linear SVC* dan *Chi-Square* dengan nilai melebihi 85%.

Kata Kunci – Coronary Artery Disease, Seleksi fitur, Boruta, Klasifikasi, *Extra Tree*, *Chi-Square*, SVM, *Linear SVC*