

"Klasifikasi Data Stroke Menggunakan Random Forest dengan Recursive Feature Elimination" = "Classification of Stroke Data Using Random Forest with Recursive Feature Elimination"

Fiftitah Repfian Aszhari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509676&lokasi=lokal>

Abstrak

Stroke merupakan salah satu penyakit dengan risiko kematian dan kecacatan yang tinggi. Secara umum, stroke diklasifikasikan menjadi dua jenis, yaitu stroke iskemik dan stroke hemoragik. Klasifikasi jenis stroke secara cepat dan tepat diperlukan untuk menentukan jenis pengobatan dan tindakan yang tepat guna mencegah terjadinya dampak yang lebih fatal pada pasien stroke. Pada penelitian ini, klasifikasi stroke dilakukan menggunakan pendekatan *machine learning*. Adapun data penelitian yang digunakan adalah data stroke yang terdiri atas pemeriksaan laboratorium. Pada data penelitian tersebut, terdapat berbagai komponen pemeriksaan laboratorium yang dicatat serta memungkinkan adanya suatu pemeriksaan yang kurang relevan atau informatif dalam mengklasifikasi stroke. Apabila data tersebut tidak ditangani, akan mempengaruhi kinerja serta waktu komputasi model dalam mengklasifikasi stroke. Oleh karena itu, pada penelitian ini, *Random Forest* (RF) dengan seleksi fitur *Recursive Feature Elimination* (RFE) digunakan dalam mengklasifikasi data stroke. Dengan menerapkan metode tersebut, diperoleh kinerja model yang lebih baik saat melakukan klasifikasi menggunakan sejumlah fitur yang diperoleh dari hasil seleksi fitur, dibandingkan menggunakan keseluruhan fitur dalam data stroke. Selain itu, pada penerapan metode tersebut, diperoleh kinerja model yang baik dalam mengklasifikasi data kelas stroke iskemik, akan tetapi tidak cukup baik dalam mengklasifikasi data kelas stroke hemoragik. Hal ini dikarenakan proporsi jumlah data pada kelas stroke iskemik lebih banyak dibandingkan stroke hemoragik. Dalam hal ini dibutuhkan suatu metode penanganan agar kinerja model tetap optimal dalam mengklasifikasi data kelas stroke iskemik dan stroke hemoragik. Pada penelitian ini, *Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk menyeimbangkan kedua kelas data stroke guna memperoleh kinerja model yang optimal dalam mengklasifikasi kedua kelas data stroke. Berdasarkan penerapan metode RF dengan RFE serta SMOTE dalam mengklasifikasi data stroke, diperoleh kinerja model yang lebih baik dibandingkan melakukan klasifikasi pada data stroke yang tidak diseimbangkan dengan SMOTE.

Stroke is one of the diseases with the high risk of death and disability. Stroke generally can be classified into two types, namely ischemic stroke and hemorrhagic stroke. A quick and accurate stroke classification is needed to find the right treatment to prevent a dangerous effect on the stroke patients. In this study, the stroke classification was applied using a machine learning approach. The data used in this study is stroke data that consists of laboratory examinations. The data consists of various laboratory examination components, therefore, it might be possible that some of the components are less relevant and has less informative related in classifying stroke. If the data is not well handled, it might affect the performance and computation time of the model in classifying stroke. Therefore, in this study, Random Forest (RF) with Recursive Feature Elimination (RFE) method is used to classify the stroke data. The result showed that by applying the method in classifying several amounts of features obtained from the feature selection results has better performance rather than classifying the method using all features in stroke data. Moreover, based on applying this

method, the result showed that the model has better performance in classifying ischemic stroke class data but not good enough in classifying hemorrhagic stroke class data. This result might occur because the proportion of numbers the ischemic stroke more than hemorrhagic stroke class data. Therefore, the handling method is needed to obtain optimal model performance in classifying ischemic stroke and hemorrhagic stroke class data. In this study, Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the two classes of stroke data so optimal performance of the classification model can be obtained. Based on the application of the RF with RFE methods and SMOTE in the classification of stroke data, better model performance is obtained compared to classifying the stroke data that is not balanced with SMOTE.