

Implementasi ensemble peta swadaya untuk imputasi yang hilang nilai-nilai = Implementation of a self-help map ensemble for lost imputations values

Kathan Gerry Vivaldi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493902&lokasi=lokal>

Abstrak

Nilai yang hilang adalah nilai yang tidak disimpan dalam variabel tertentu dalam pengamatan. Nilai yang hilang dapat ditemukan dalam data di hampir semua bidang penelitian dan dapat mempersulit analisis data. Self-Organizing Maps (SOM) adalah metode clustering berbasis jaringan saraf yang dapat digunakan sebagai metode imputasi, di mana SOM menyalahkan nilai-nilai yang hilang dengan menggeneralisasi pengamatan mengandung nilai yang hilang. Ensemble Self-Organizing Maps (E-SOM) adalah pengembangan metode imputasi SOM, di mana metode E-SOM diterapkan kerangka ensemble dengan menggunakan beberapa SOM untuk meningkatkan kemampuan generalisasi. Dalam tesis ini metode E-SOM dan SOM diimplementasikan untuk imputasi nilai yang hilang dalam data Penyakit Jantung Afrika Selatan dengan menggunakan 15 ansambel dan berbagai variasi dalam jumlah neuron. Pada data imputasi kedua metode ini kemudian dibentuk oleh model klasifikasi Hutan Acak dan dilakukan evaluasi kinerja model yang dibentuk menggunakan nilai akurasi dalam data pengujian. Hasil evaluasi menunjukkan bahwa untuk model yang terbentuk dari data imputasi E-SOM menghasilkan nilai akurasi yang lebih baik untuk variasi 20, 30, 50, 60, dan 80 neuron dalam klasifikasi data pengujian. Sedangkan untuk variasi 40 neuron, model terbentuk dari data imputasi SOM menghasilkan nilai akurasi yang lebih baik, dan untuk variasi 70 neuron, kedua metode menghasilkan nilai akurasi yang sama. Selain itu, menerapkan berbagai kombinasi variasi dalam jumlah neuron dan jumlah ansambel dalam metode ini imputasi E-SOM. Model Random Forest dihasilkan dari data dari imputasi E-SOM dengan kombinasi 60 neuron dan 5 ansambel menghasilkan nilai akurasi paling optimal.

<hr>

Missing values are values that are not stored in certain variables in the observation. Missing values can be found in data in almost all fields of research and can complicate data analysis. Self-Organizing Maps (SOM) is a neural network based clustering method that can be used as an imputation method, where SOM blames missing values by generalizing observations contains missing values. Ensemble Self-Organizing Maps (E-SOM) is the development of the SOM imputation method, in which the E-SOM method is applied to an ensemble framework by using multiple SOMs to improve generalization capabilities. In this thesis the E-SOM and SOM methods are implemented for the imputation of missing values in South African Heart Disease data with using 15 ensembles and various variations in the number of neurons. In the imputation data the two methods are then formed by the Random Forest classification model and an evaluation of the performance of the model is formed using the accuracy values in the test data. The evaluation results show that the model formed from E-SOM imputation data produces better accuracy values for variations of 20, 30, 50, 60, and 80 neurons in the classification of test data. As for the variation of 40 neurons, the model formed from SOM imputation data produces a better accuracy value, and for the variation of 70 neurons, both methods produce the same accuracy value. Other than that, applying

various combinations of variations in the number of neurons and the number of ensembles in this method imputation E-SOM. The Random Forest model is generated from data from the E-SOM imputation with a combination of 60 neurons and 5 ensembles producing the most optimal accuracy value