

Imputasi Missing Values menggunakan algoritma Hybrid Fuzzy CMeans dan Majority Vote untuk Klasifikasi Data Penyakit Paru Obstruktif Kronik (PPOK) = Missing Values Imputation using Hybrid Fuzzy C-Means and Majority Vote Algorithm for Chronic Obstructive Pulmonary Disease (COPD) Data Classification

Shafira Nur Amalia, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920534162&lokasi=lokal>

Abstrak

Dalam suatu penelitian, dibutuhkan data yang dikumpulkan dan diolah untuk memecahkan permasalahan dan membuktikan hipotesis dalam penelitian. Namun, seringkali data yang diperoleh tidak menyimpan nilai untuk suatu variabel pada observasi yang diharapkan. Data yang tidak tersimpan menyebabkan data penelitian kosong dan berdampak pada penelitian. Jika peristiwa ini terjadi, maka penelitian terindikasi memiliki missing data atau missing values. Salah satu cara untuk mengatasi missing values yaitu dengan imputasi. Imputasi bekerja dengan mengisi nilai pada missing values dengan suatu nilai estimasi yang telah dianalisis dan diputuskan untuk membuat suatu dataset lengkap. Dalam proses imputasi, seringkali ditemukan bahwa data yang digunakan untuk imputasi terkadang memiliki karakteristik yang tidak jelas atau tidak konsisten, maka salah satu solusinya adalah dengan menggunakan metode Fuzzy C-Means (FCM). Estimasi nilai-nilai missing values menggunakan model FCM menghasilkan model prediksi dengan variasi parameter yang beragam sehingga dibutuhkan pendekatan lain untuk menghasilkan model terbaik dengan parameter yang optimal. Hal inilah yang mendasari diperlukannya suatu pendekatan hybrid, yaitu dengan menggabungkan beberapa model machine learning untuk memperoleh hasil estimasi missing values terbaik. Pada penelitian ini, dilakukan implementasi Hybrid Fuzzy C-Means dan Majority Vote (Hybrid FCMMV) pada data Penyakit Paru Obstruktif Kronik (PPOK) tahun 2012-2017 yang diperoleh dari Rumah Sakit Cipto Mangunkusumo (RSCM) untuk memberikan performa imputasi yang lebih baik berdasarkan akurasi, presisi, recall, dan F-Score melalui klasifikasi metode ensemble Random Forest.

.....In a research study, collected and processed data are needed to solve problems and prove hypotheses. However, the obtained data often do not store the value for a variable in the expected observation. Data that are not stored contribute to the emptying of research data which has an impact on the research itself. If the phenomenon occurs, it indicates that the research has missing data or missing values. One way to overcome missing values is using imputation techniques. The technique works by filling in the missing values with an estimated value that has been analyzed and decided to create a complete dataset. In the process, it is often found that the data being used for imputation have unclear or inconsistent characteristics, which can be solved by implementing Fuzzy C-Means (FCM) method. The estimation of missing values using the FCM model produces predictive models with a variety of parameters, hence another approach to produce the best model with optimal parameters is needed. This underlies the need for a hybrid approach, which is acquired through combining or integrating different machine learning models to earn the best estimation result of missing values. In this study, the implementation of Hybrid Fuzzy C-Means and Majority Vote (Hybrid FCMMV) was conducted on Chronic Obstructive Pulmonary Disease (COPD) data in 2012-2017 from Cipto Mangunkusumo Hospital (RSCM)) to provide better imputation performance based on accuracy, precision, recall, and F-Score through the classification of the Random Forest ensemble method.