

Imputasi missing values menggunakan algoritma fuzzy c-means pada klasifikasi data Penyakit Paru Obstruktif Kronik (PPOK) = Missing values imputation based on fuzzy c-means algorithm for classification of Chronic Obstructive Pulmonary Disease (COPD)

Kiki Aristiawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494352&lokasi=lokal>

Abstrak

Penyakit Paru Obstruktif Kronik (PPOK) merupakan salah satu penyebab kematian terbanyak dengan jumlah kematian sekitar 3 juta jiwa atau setara dengan 5,2% dari jumlah kematian di seluruh dunia. Untuk itu diperlukan penelitian lebih lanjut mengenai PPOK, namun data yang dikumpulkan dalam penelitian biasanya tidak memuat semua data yang diinginkan. Hilangnya informasi dalam data dikenal dengan istilah missing values yang dapat menjadi masalah untuk semua jenis analisis data. Salah satu cara untuk menangani missing values yaitu dengan melakukan proses imputasi data pada tahap preprocessing untuk mendapatkan data lengkap yang diharapkan mampu meningkatkan akurasi dari analisis data yang dilakukan. Pada penelitian ini dilakukan imputasi menggunakan mean dan Fuzzy C-Means (FCM). FCM merupakan metode pengelompokan yang memungkinkan satu bagian data menjadi milik dua atau lebih kelompok berdasarkan nilai keanggotaannya. Data lengkap hasil imputasi diuji menggunakan metode klasifikasi Decision Tree dengan persentase data training 50%-90% untuk melihat performa dari metode mean dan FCM. Berdasarkan penelitian ini diperoleh nilai akurasi, presisi, dan recall tertinggi untuk klasifikasi data PPOK yang diimputasi menggunakan mean masing-masing sebesar 64,7%, 71%, dan 65%. Sedangkan nilai akurasi, presisi, dan recall tertinggi untuk klasifikasi data PPOK yang diimputasi menggunakan FCM masing-masing sebesar 78,9%, 85%, dan 79%. Hasil ini menunjukkan bahwa FCM membantu Decision Tree untuk mengklasifikasikan data lebih baik dengan nilai imputasi yang lebih baik untuk menggantikan missing values.

<hr>

Chronic Obstructive Pulmonary Disease (COPD) is one of the most causes of death in the world with around 3 million deaths, equivalent to 5.2% of deaths worldwide. For this reason, further research needs to be done on CPOD, but the data collected in the study often does not contain all the desired data. Loss information in data is called as a missing values which can be a problem for all types of data analysis. One way to handle missing values is by doing the imputation process at the preprocessing stage to obtain complete data which is expected to increase the accuracy of the data analysis performed. In this study, imputation was done using mean and FCM. FCM is a clustering method that allows one part of the data to belong to two or more groups based on their membership function. The complete dataset was trained with Decision Tree classifier with the percentage of data training 50%-90% to observe the performance in terms of accuracy, precision, and recall for mean and FCM method. Based on this study, the highest value of accuracy, precision, and recall for classification of COPD data imputed using mean of 64.7 %, 71 % and 65 % respectively. While the highest value of accuracy, precision, and recall for classification of COPD data imputed using FCM is 78.9 %, 85 %, and 79 % respectively. These results indicate that FCM helps Decision Tree to classify data better with better imputation values to replace missing values.