

Implementasi Metode Imputasi Ganda Dynamically Ordered Attribute Trees (DOAT) untuk Estimasi Missing Values pada Data Kategorik = Implementation of Dynamically-Ordered Attribute Tree (DOAT) Multiple Imputation Method to Estimate Missing Values in Categorical Data

Vima Majesta Tingginehe, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20498233&lokasi=lokal>

Abstrak

ABSTRAK

Data yang tidak lengkap merupakan salah satu masalah yang sering muncul dalam penelitian medis, yang dapat mengakibatkan statistik inferensial tidak akurat. Salah satu cara yang dapat dilakukan untuk mengatasi masalah tersebut adalah dengan mengestimasi missing value yang terdapat pada data. Metode ini disebut imputasi. Beberapa metode imputasi telah dikembangkan untuk memperkirakan nilai yang hilang, dan salah satu metode imputasi yang paling sering digunakan adalah imputasi pohon keputusan. Pohon keputusan adalah salah satu metode pembelajaran mesin terawasi yang paling sederhana, yang digunakan untuk mengklasifikasikan dan meregresi data. Namun selain kepraktisan metode pohon keputusan untuk imputasi, metode ini kurang tepat jika muncul nilai-nilai yang hilang baik pada data latih maupun data pengujian secara bersamaan. Untuk mengatasi keterbatasan tersebut, metode imputasi pohon keputusan kemudian dikembangkan menjadi metode imputasi pohon atribut yang dipesan secara dinamis (DOAT). DOAT pertama kali diperkenalkan oleh Nick Street dan Jing Wang pada tahun 2009, untuk membangun sejumlah pohon keputusan yang digunakan untuk memperkirakan nilai yang hilang secara bertahap dalam data pelatihan. Selanjutnya tingkat akurasi estimasi akan digunakan untuk menentukan pohon keputusan akhir sebagai pohon keputusan terakhir yang melakukan imputasi. Setelah pohon keputusan akhir diperoleh dari data pelatihan, maka nilai-nilai yang hilang pada data pengujian akan diestimasi menggunakan sejumlah pohon keputusan yang dihasilkan dari data pelatihan, dimana pohon keputusan akhir yang dihasilkan pada data pelatihan juga merupakan pohon keputusan final. pohon keputusan yang digunakan dalam imputasi pengujian data. Pada penelitian ini dikembangkan metode imputasi DOAT dengan algoritma CART untuk membangun pohon keputusan untuk mengestimasi missing value pada data kategori penyakit. Kinerja metode imputasi DOAT akan dievaluasi berdasarkan tingkat akurasi menggunakan pohon keputusan. Hasil penelitian ini menunjukkan bahwa metode imputasi DOAT tidak mengurangi tingkat akurasi metode klasik yang sudah memiliki data lengkap.

ABSTRACT

Incomplete data is one of the problems that often arises in medical research, which can result in inaccurate inferential statistics. One way that can be done to overcome this problem is to estimate the missing value contained in the data. This method is called imputation. Several imputation methods have been developed to estimate missing values, and one of the most frequently used imputation methods is decision tree imputation. Decision trees are one of the simplest supervised machine learning methods, which are used to classify and regress data. However, in addition to the practicality of the decision tree method for imputation, this method is not appropriate if missing values appear in both training data and test data simultaneously. To

overcome these limitations, the decision tree imputation method was later developed into a dynamically ordered attribute tree (DOAT) imputation method. DOAT was first introduced by Nick Street and Jing Wang in 2009, to construct a number of decision trees that are used to estimate the incrementally missing values in training data. Furthermore, the level of accuracy of the estimate will be used to determine the final decision tree as the last decision tree that performs imputation. After the final decision tree is obtained from the training data, the missing values in the test data will be estimated using a number of decision trees generated from the training data, where the final decision tree generated in the training data is also the final decision tree. decision tree used in the imputation of data testing. In this study, the DOAT imputation method with the CART algorithm was developed to build a decision tree to estimate the missing value in disease category data. The performance of the DOAT imputation method will be evaluated based on the level of accuracy using a decision tree. The results of this study indicate that the DOAT imputation method does not reduce the level of accuracy of the classical method which already has complete data.