

Konstruksi metode imputasi decision tree-expectation maximization untuk mengestimasi missing values = Construction of decision tree-expectation maximization imputation method for estimating missing values

Iasha Zahra Nabila, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494343&lokasi=lokal>

Abstrak

Abstrak. Imputasi missing values berperan penting dalam pre-processing data untuk menghasilkan data yang lengkap dan berkualitas. Dalam penelitian ini, dilakukan kombinasi dari Decision Tree dan algoritma Expectation Maximization (EM) sebagai metode imputasi dalam mengestimasi missing values. Namun, terdapat kekurangan pada metode ini, yaitu algoritma Expectation Maximization (EM) cenderung memberikan hasil imputasi yang lebih akurat jika terdapat banyak kemiripan antar atribut. Hal ini dapat diatasi dengan mempartisi segmen horisontal menggunakan algoritma Decision Tree. EM diterapkan pada berbagai segmen horisontal dari data set agar diperoleh banyak kemiripan antar atribut. Metode yang diusulkan ini kemudian dievaluasi dengan membandingkan kinerjanya dengan imputasi menggunakan mean dalam mengestimasi missing values data numerik dan dipilih Decision Tree sebagai classifier. Data yang digunakan untuk simulasi dalam penelitian ini yaitu data set PPOK-OSA. Penelitian ini menunjukkan bahwa metode imputasi missing values yang diusulkan menghasilkan akurasi yang lebih tinggi jika dibandingkan menggunakan nilai mean dalam mengestimasi missing values. Hasil akurasi tertinggi dari metode ini yaitu 83.3%, diperoleh pada saat persentase data training 10% dan rata-rata akurasi pada berbagai persentase data training sebesar 70.3%, sedangkan hasil akurasi tertinggi menggunakan nilai mean yaitu 58.3%, diperoleh pada saat persentase data training 20% dengan rata-rata akurasi pada berbagai persentase data training sebesar 50.5%.

<hr>

Abstract. Missing values imputation plays a vital role in data pre-processing for ensuring good quality of data. In this study, we present a combination of Decision Tree and Expectation Maximization (EM) algorithm as imputation method to estimate missing values. However, there are shortcomings of method, where the Expectation Maximization algorithm tends to give more accurate imputation results if there are many similarities among attributes. Therefore, this can be overcome by partitioning the horizontal segments using the Decision Tree algorithm. We applied EM on various horizontal segments of a data set where there are many similarities among attributes. Besides, we evaluate our proposed method by comparing its performance with mean values to impute missing values, and we choose the Decision Tree as a classifier. Data used in this research is COPD-OSA data set. In this study shows that our proposed method leads to higher accuracy than mean imputation. The highest accuracy results obtained from our proposed method is 83.3% when the percentage of training data is 10% and average accuracy in various percentage of training data is 70.3%, while the highest accuracy results using the mean value is 58.3% when the training data percentage is 20% and the average accuracy in

various percentage of training data is 50.5%.