

ABSRTACT

THE COMPARISON IMPUTATION METHOD : MEAN METHOD AND K NEAREST NEIGHBOR (KNN) METHOD TO OVERCOME THE MISSING DATA OF SURVEY DATA

By

DESY NUR FITRIANA MURJITO

One of the problems in survey is often experienced is there are some units that do not respond to some of the questions so that it makes data are incomplete or data are missing. Imputation method is one of the ways to overcome the missing data. Mean imputation and K Nearest Neighbor Imputation are two method that can be used to this research. The purpose of the research is to compare the imputation method to estimate the missing data with Mean Imputation and K Nearest Neighbor (KNN) Imputation and search which method is better among two methods. Based on simulation study with 1000 replication KNN Imputation method has a smaller the average value of Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) than Mean Imputation. Therefore, it is concluded that KNN Imputation method is better than Mean Imputation Method.

Keywords : Missing Data, Imputation, Mean Imputation, K Nearest Neighbor Imputation, Mean Square Error, Mean Absolute Percentage Error

ABSRTAK

PERBANDINGAN METODE IMPUTASI: METODE *MEAN* DAN METODE *K NEAREST NEIGHBOR* (KNN) UNTUK MENGATASI DATA HILANG PADA DATA SURVEI

Oleh

DESY NUR FITRIANA MURJITO

Salah satu masalah yang sering dialami di dalam survei adalah ditemukannya unit-unit yang tidak merespon sejumlah pertanyaan yang diajukan, sehingga menyebabkan data tidak lengkap atau data hilang. Salah satu cara mengatasi data hilang adalah metode imputasi. Dalam penelitian ini digunakan metode Imputasi *Mean* dan Imputasi *K Nearest Neighbor*. Tujuan penelitian ini adalah membandingkan metode imputasi untuk mengestimasi data hilang dengan metode Imputasi *Mean Imputation* dan Imputasi *K Nearest Neighbor* (KNN) dan mencari metode mana yang lebih baik di antara kedua metode tersebut. Berdasarkan hasil studi simulasi dengan 1000 kali ulangan diperoleh bahwa metode Imputasi KNN menghasilkan rata-rata nilai Mean Square Error (MSE) dan Mean Absolute Percentage Error (MAPE) yang lebih kecil dibandingkan metode Imputasi *Mean*. Dengan demikian, maka dapat disimpulkan bahwa metode imputasi KNN lebih baik daripada metode imputasi *Mean*.

Kata Kunci : Data Hilang, Imputasi, *Mean Imputation*, *K Nearest Neighbor Imputation*, *Mean Square Error*, *Mean Absolute Percentage Error*