

Title : A Probabilistic Approach for Missing Data Imputation

Author Names : Muhammed Nazmul Arefin, Abdul Kadar Muhammad Masum

Abstract: In the context of data analysis, missing data imputation is a vital issue due to the typically large scale and complexity of the datasets. It often results in a higher incidence of missing data. So, addressing missing data through the imputation technique is essential to ensure the integrity and completeness of the data. It will ultimately improve the accuracy and validity of the data analysis. The prime objective of this study is to propose an imputation model. This paper presents a method for imputing missing employee data through a combination of features and probability calculations. The study utilized employee datasets that were collected from the Kaggle along with primary data collected from RMG factories located in Chittagong. The suggested algorithm demonstrated a notable level of accuracy on the datasets, and the average accuracy for each identified technique was also quite satisfactory. This study contributes to the existing body of research on missing data imputation in big data analysis and offers practical implications for handling missing data in different datasets. Usage of this technique will enhance the accuracy of data analysis and decision-making in organizations.

Keywords: Missing Data Imputation, Data Analysis, Big Data, Employee Dataset, Probability-Based Imputation, Feature Combination,

<https://doi.org/10.1155/2024/4737963>

<https://onlinelibrary.wiley.com/doi/10.1155/2024/4737963>