Missing Data Imputation Technique

Author: Bonumaddi Naga Pravallika

Department of Computer Science & Engineering

National Institute of Technology, Rourkela, Odisha – 769008

Contact: 120cs0121@nitrkl.ac.in

Under the Supervision of

Prof. Pabitra Mohan Khilar

Department of Computer Science & Engineering

National Institute of Technology, Rourkela, Odisha – 769008

Contact: pmkhilar@nitrkl.ac.in

September 1st, 2023

Abstract

This research aims to improve missing data imputation in computer science engineering datasets, focusing on healthcare data. The proposed approach integrates Class Center-Based Missing Value Imputation (CCMVI), sampling-based methods, and the Firefly Algorithm (FA) to address limitations and enhance imputation accuracy. The research objectives involve developing C3-FA, evaluating predictive, distributional, and classification accuracy, and analyzing the impact of missing data characteristics. The proposal fills knowledge gaps in imputation techniques, particularly concerning attribute correlation and missing data mechanisms. The methodology covers data preprocessing, imputation, and evaluation. The conclusion highlights the potential to advance imputation techniques and contribute to computer science engineering.

Keywords: missing data imputation, mean substitution, Class center-based missing value imputation Normalized and Outlier Handling (CCMVI_NOH), Sampling Based Decision Tree, Enhanced Firefly Optimization.

Reference Papers:

Adaptive Multiple imputations of missing values using the class center: Springer Link, 2022, Link1

A class center-based approach for missing value imputation: Science Direct, Elsevier, 2020, Link2

Name: Bonumaddi Naga Pravallika **Prof. Pabitra Mohan Khilar**

Roll no: 120CS0121 **Supervisor**