World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:18, No:04, 2024

Open Science Index, Educational and Pedagogical Sciences Vol:18, No:04, 2024 publications.waset.org/abstracts/168151.pdf

Imputing Missing Data in Electronic Health Records: A Comparison of Linear and Non-Linear Imputation Models

Authors : Alireza Vafaei Sadr, Vida Abedi, Jiang Li, Ramin Zand

Abstract : Missing data is a common challenge in medical research and can lead to biased or incomplete results. When the data bias leaks into models, it further exacerbates health disparities; biased algorithms can lead to misclassification and reduced resource allocation and monitoring as part of prevention strategies for certain minorities and vulnerable segments of patient populations, which in turn further reduce data footprint from the same population – thus, a vicious cycle. This study compares the performance of six imputation techniques grouped into Linear and Non-Linear models on two different realworld electronic health records (EHRs) datasets, representing 17864 patient records. The mean absolute percentage error (MAPE) and root mean squared error (RMSE) are used as performance metrics, and the results show that the Linear models outperformed the Non-Linear models in terms of both metrics. These results suggest that sometimes Linear models might be an optimal choice for imputation in laboratory variables in terms of imputation efficiency and uncertainty of predicted values.

Keywords : EHR, machine learning, imputation, laboratory variables, algorithmic bias
Conference Title : ICML 2024 : International Conference on M-Learning
Conference Location : Venice, Italy
Conference Dates : April 04-05, 2024