

Imputation Analysis of Time-Series Data Using a Random Forest Algorithm



Nur Najmiyah Jaafar , Muhammad Nur Ajmal Rosdi,
Khairur Rijal Jamaludin, Faizir Ramlie, and Habibah Abdul Talib

Abstract Missing data poses a significant challenge in extensive datasets, particularly those containing time-series information, leading to potential inaccuracies in data analysis and machine learning model development. To address the issue, this paper compared and evaluated four imputation methods: MissForest, MICE, Simplefill, and Softimpute which utilized Random Forest Algorithm. The research examines the impact of missing ratios and temporal variations on the performance of the imputation methods. The results indicated that MissForest consistently outperformed other methods, exhibiting the lowest RMSE values and a high coefficient of determination (R^2), indicating its accuracy and ability to explain the variation in the data. Furthermore, graphical analyses demonstrated the stability of MissForest over time, while MICE and Simplefill showed higher sensitivity to date changes. Softimpute demonstrated relative consistency but slightly lower performance compared to MissForest. Overall, this study highlights the effectiveness of MissForest as the preferred imputation method for AVL time-series data.

Keywords Imputation · Missing data · Random forest · MissForest · Time-series data

1 Introduction

In recent years, time-series data analysis has gained significant attention across various domains, including finance, health care, and environmental studies. Time-series data often contain missing values, which can arise due to various reasons such as sensor malfunction, data transmission errors, or incomplete data collection process

N. N. Jaafar (✉) · M. N. A. Rosdi

Faculty of Engineering Technology Manufacturing and Mechatronics, Universiti Malaysia Pahang
Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
e-mail: najmiyah@ump.edu.my

M. N. A. Rosdi · K. R. Jamaludin · F. Ramlie · H. A. Talib

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Jalan Sultan Yahya
Petra, 54100 Kuala Lumpur, Malaysia