

# Web of Science



Look Up Full Text



Save to EndNote online

Add to Marked List

1 of 1

## Performance Analysis of Machine Learning Algorithms for Missing Value Imputation

By: Abidin, NZ (Abidin, Nadzurah Zainal)<sup>[1]</sup>; Ismail, AR (Ismail, Amelia Ritahani)<sup>[1]</sup>; Emran, NA (Emran, Nurul A.)<sup>[2]</sup>

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

Volume: 9 Issue: 6 Pages: 442-447

Published: JUN 2018

Document Type: Article

### Abstract

Data mining requires a pre-processing task in which the data are prepared, cleaned, integrated, transformed, reduced and discretized for ensuring the quality. Missing values is a universal problem in many research domains that is commonly encountered in the data cleaning process. Missing values usually occur when a value of stored data absent for a variable of an observation. Missing values problem imposes undesirable effect on analysis results, especially when it leads to biased parameter estimates. Data imputation is a common way to deal with missing values where the missing value's substitutes are discovered through statistical or machine learning techniques. Nevertheless, examining the strengths (and limitations) of these techniques is important to aid understanding its characteristics. In this paper, the performance of three machine learning classifiers (K-Nearest Neighbors (KNN), Decision Tree, and Bayesian Networks) are compared in terms of data imputation accuracy. The results shows that among the three classifiers, Bayesian has the most promising performance.

### Keywords

Author Keywords: Data Mining; Imputation; Machine Learning; K-Nearest Neighbors; Decision Tree; Bayesian Networks

### Author Information

Reprint Address: Ismail, AR (reprint author)

- + Int Islamic Univ Malaysia, Kulliyah Informat & Commun Technol, Dept Comp Sci, POB 10, Kuala Lumpur 50728, Malaysia.

### Addresses:

- + [ 1 ] Int Islamic Univ Malaysia, Kulliyah Informat & Commun Technol, Dept Comp Sci, POB 10, Kuala Lumpur 50728, Malaysia
- + [ 2 ] Univ Teknikal Malaysia Melaka UTeM, Fac Informat & Commun Technol, Durian Tunggal 76100, Melaka, Malaysia

### Funding

Funding Agency	Grant Number
IIUM Research Initiative Grants Scheme (RIGS)	RIGS16-346-0510

[View funding text](#)

### Publisher

SCIENCE & INFORMATION SAI ORGANIZATION LTD, 19 BOLLING RD, BRADFORD, WEST YORKSHIRE,

### Citation Network

In Web of Science Core Collection

0

Times Cited

[Create Citation Alert](#)

19

Cited References

[View Related Records](#)

### Use in Web of Science

Web of Science Usage Count

1

Last 180 Days

1

Since 2013

[Learn more](#)

This record is from:

Web of Science Core Collection - Emerging Sources Citation Index

[Suggest a correction](#)

If you would like to improve the quality of the data in this record, please [suggest a correction](#).

00000, ENGLAND

**Categories / Classification**

Research Areas: Computer Science

Web of Science Categories: Computer Science, Theory &amp; Methods

[See more data fields](#)

◀ 1 of 1 ▶

**Cited References: 19**Showing 19 of 19 [View All in Cited References page](#)*(from Web of Science Core Collection)*

1. **Proper imputation techniques for missing values in data sets** Times Cited: 1  
 By: Aljuaid, Tahani; Sasi, Sreela.  
 INT C DAT SCI ENG IC Published: 2016
2. **[Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation](#)** Times Cited: 5  
 By: Allasonniere, Stephanie; Kuhn, Estelle  
 COMPUTATIONAL STATISTICS & DATA ANALYSIS Volume: 91 Pages: 4-19 Published: NOV 2015
3. Title: [not available] Times Cited: 8  
 By: Barga, Roger; Fontama, Valentine; Tok, Wee Hyong; et al.  
 Predictive analytics with Microsoft Azure machine learning Published: 2015  
[\[Show additional data\]](#)
4. **[Nearest neighbor imputation algorithms: a critical evaluation](#)** Times Cited: 6  
 By: Beretta, Lorenzo; Santaniello, Alessandro  
 BMC MEDICAL INFORMATICS AND DECISION MAKING Volume: 16 Supplement: 3 Article Number: 74 Published: JUL 25 2016
5. Title: [not available] Times Cited: 1  
 By: Emran; Nurul, A.  
 Data completeness measures, Pattern Analysis, Intelligent Security and the Internet of Things Pages: 117-130 Published: 2015
6. **[Decision Tree: Review of Techniques for Missing Values at Training, Testing and Compatibility](#)** Times Cited: 1  
 By: Gavankar, Sachin; Sawarkar, Sudhirkumar  
 2015 THIRD INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, MODELLING AND SIMULATION (AIMS 2015) Book Series: International Conference on Artificial Intelligence Modelling & Simulation Pages: 122-126 Published: 2015
7. **Data Mining-K-Nearest Neighbors** Times Cited: 1  
 Group Author(s): Gerardnico  
 CC Attribution-Noncommercial-Share Alike 4.0 International Published: 2017
8. **Basic bayesian methods** Times Cited: 1  
 By: Glickman, Mark E; Dyk, Van; David, A.  
 Topics in Biostatistics Pages: 319-338 Published: 2017
9. **The prevention and handling of the missing data** [\(View record in KCI-Korean Journal Database\)](#) Times Cited: 23