Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 4 (2022) 570 - 576 ISSN Media Electronic: 2580-0760

K Nearest Neighbor Imputation Performance on Missing Value Data Graduate User Satisfaction

Abdul Fadlil[1], Herman[2], Dikky Praseptian M[3]
[1]Department of Electrical Engineering, Industrial Technology, Universitas Ahmad Dahlan
[2,3]Master Program of Informatics, Industrial Technology, Universitas Ahmad Dahlan
[1] fadlil@mti.uad.ac.id, [2] hermankaha@mti.uad.ac.id, [3] dikky2107048008@webmail.uad.ac.id*

Abstract

A missing value is a common problem of most data processing in scientific research, which results in a lack of accuracy of research results. Several methods have been applied as a missing value solution, such as deleting all data that have a missing value, or replacing missing values with statistical estimates using one calculated value such as, mean, median, min, max, and most frequent methods. Maximum likelihood and expectancy maximization, and machine learning methods such as K Nearest Neighbor (KNN). This research uses KNN Imputation to predict the missing value. The data used is data from a questionnaire survey of graduate user satisfaction levels with seven assessment criteria, namely ethics, expertise in the field of science (main competence), foreign language skills, foreign language skills, use of information technology, communication skills, cooperation, and self-development. The results of testing imputation predictions using KNNI on user satisfaction level data for STMIK PPKIA Tarakanita Rahmawati graduates from 2018 to 2021. Where using the five k closest neighbors, namely 1, 5, 10, 15, and 20, the error value of the k nearest neighbors is 5 in RMSE is 0, 316 while the error value using MAPE is 3,33 %, both values are smaller than the value of k other nearest neighbors. K nearest neighbor 5 is the best imputation prediction result, both calculated by RMSE and MAPE, even in MAPE the error value is below 10%, which means it is very good.

Keywords: Graduate user, Imputation, KNN, Missing Value, Satisfaction.

1. Introduction

Implementation of a more specific tracer study on the assessment of graduate users is very much needed by universities because it is a feedback medium from graduate users in an effort to improve education systems and management. Tracer study, in this case, the assessment of graduate users towards universities, are carried out in almost the same way in every university, namely by distributing questionnaires to the agencies/companies/institutions where the alumni work. The agency/company/institution is asked to assess each alumnus. Assessment is usually carried out by superiors in the field of alumni work so that the assessment can be carried out more objectively. The questionnaire media used varied at each university, such as providing paper questionnaires, google forms, or applications both desktop/website/mobile owned by the assessed universities. STMIK PPKIA Tarakanita Rahmawati is one of several private universities in North Kalimantan that conducts tracer study, in this case, the assessment of graduate users on graduates from universities. PPKIA, the name that is usually attached to STMIK PPKIA Tarakanita Rahmawati, evaluates graduates still using a questionnaire in the form of a sheet of paper folded into an envelope to maintain confidentiality which is then handed over to graduate users to be returned after being filled out. The results of the graduate user assessment questionnaire are usually recapitulated as college evaluation material if there is a bad assessment of graduates. However, in the process of recapitulating the graduate user questionnaire, there is an important problem, namely missing value,

Missing data or a missing value is a condition where there are incomplete or empty values on one or more criteria. A missing value is a common problem in most scientific research in fields such as Biology, Medicine, or Climate Science. They can arise from various sources such as sample handling error, low signal-to-noise ratio, measurement error, non-response or deleted deviant values. [1], Rubin (2022) defines missing data based on three loss mechanisms: data are missing completely at random (MCAR) when the probability of a case having an error value for the variable does not depend on the

Accepted: 23-06-2022 | Received in revised: 31-07-2022 | Published: 22-08-2022

known value or missing data; data are missing at random (MAR) when the probability of a case having a missing value for a variable may depend on the known value but not on the missing data value itself; data is missing not at random (MNAR) when the probability of an instance having a missing value for a variable can depend on the value of that variable[2]. Missing value creates an element of ambiguity when analyzing data and which can affect the nature of statistical estimators and result in loss of power and misleading conclusions [3]. So that it is so important to handle missing values in data processing in other processes to obtain information.

Missing value has occurred several times in the graduate user assessment questionnaire for college graduates.A missing value that often occurs is the loss of some values on certain assessment attributes, empty questionnaires are almost never found or not assessed at all. When confirmed regarding a questionnaire whose attributes do not have complete scores, most graduate users answered with answers that they could not judge and forgot. In the case of forgetting, it may still be possible to reload, but in the case of not being able to judge, this needs a way to solve the problem. Cases cannot judge, for example, on the example of the attribute of foreign language proficiency assessment, because not all companies/institutions and agencies use foreign languages at graduate users are very difficult to assess this. The form of missing values can vary, such as the most frequently encountered data is empty/NaN, 0, and -. There is a lot of literature with various methods that have been used or applied to deal with missing observations or missing values [4]. These methods are divided into four main categories as follows [5]. The first is the deletion of all datasets that have missing values. This approach is the simplest for dealing with missing values by removing incomplete data from the data set and analyzing only the available data. Deletion is done listwise or pairwise [6]. Despite the simplicity of this method, the removal of too much data can significantly hinder the analysis and reduce the statistical significance of its conclusions, which then adversely affects the prediction process [2]. Second, the single imputation method, by replacing the missing values with statistical estimates using one calculated value such as, mean, median, min, max, and mode [2]. The three model-based imputation methods, such as the maximum likelihood method [7] and expectancy maximization [8], where more than one plausible value is used to predict one missing value observation. Fourth, machine learning methods such as K Nearest Neighbor [9], In this study, the authors used K Nearest Neighbor Imputation (KNNI) to predict the missing value of the questionnaire data for assessing graduate user satisfaction for college graduates. The purpose of predicting the missing value in this study is as part of the pre-processing process, that later the graduate user satisfaction level data can be grouped or classified to get more optimal results than without the imputation process. This study uses two evaluation methods, namely Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to test the accuracy of the predicted missing value,

2. Research Methods

2.1. Object of research

The object of the research used questionnaire data on the level of satisfaction of graduate users or tracer study from users of STMIK PPKIA graduates, Tarakanita Rahmawati. The graduate user satisfaction data has seven assessment criteria in the form of questions as shown in Table 1 with 5 assessment models as shown in Table 2. The assessment criteria are based on 7 aspects that are indicators of graduate user satisfaction assessment in the BAN-PT accreditation guidelines [10] , criteria This is also used by STMIK PPKIA Tarakanita Rahmawati in making a graduate user satisfaction questionnaire. The dataset used consists of 100 graduate user data filled in by agencies/institutions/companies in 2018 to 2021.

Table 1. Assessment Criteria

No	Criteria	Information
1	A1	Ethics
2	A2	Expertise in the field of science (main competence)
3	A3	Foreign language skills
4	A4	Use of information technology
5	A5	Communication skills
6	A6	Cooperation
7	A7	Self-development

Table 2. Assessment Model

No	Rating Model	Score
1	Very good	5
2	Well	4
3	Enough	3
4	Not enough	2
5	Very less	1

2.2. Research Stages

The research stage begins with preparing a research data set, namely data *tracer study alumni STMIK PPKIA* Tarakanita Rahmawati. The data is divided into two parts, the first data contains complete data that will be used as *training* data, the second data contains complete data which we will eliminate some of the data on certain attributes to become data *testing*, this stage is called pre-processing data. Imputed KNN implementation is carried out in two ways, namely manually calculating with *Microsoft Office Excel* as control data and the main calculation using the scikit-learn KNN *imputer library,* The evaluation was carried out to test the accuracy by comparing the actual data and the predicted data from the KNN *imputer library* from scikit-learn. E -valuation uses two methods, namely *Root Mean Squared Error*

DOI: https://doi.org/10.29207/resti.v6i4.4173

(RMSE) and *Mean Absolute Percentage Error* (MAPE), where the smallest *error value will be* the best. The stages of the research can be seen in Figure 1.

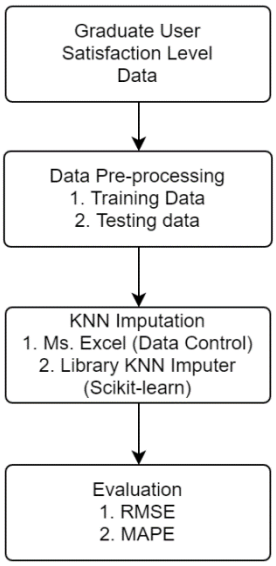

Figure 1. Research Stages

2.3, K Nearest Neigbor Imputation (KNNI)

The imputation method with K Nearest Neigbor (KNN) is one of the most popular methods for solving missing value problems [11] . KNN is popular because of its simplicity and proven effectiveness in many missing value imputation problems [12] . The superiority of the imputation method with KNN can be used to predict 2 types of data, discrete data (mode value) and continuous data (*mean* value), Imputing with KNN does not require the formation of a forecasting model for each data criterion that has data *missing values,* The weakness of imputation using KNN is that when looking for the most appropriate observations with observations that have *missing values,* imputation with KNN will search all *training* data or *datasets,* [9] . This weakness will affect when a large number of *datasets* or *training data* are used, so it will take a long time to observe, Even so, imputation with KNN is still a good method for imputing data on missing *values,* [13] . The sequence of steps in the process of finding the value of *missing value* with imputation KNN [14] .

First, determine K, the number of closest observations used. Second, Calculate the distance between observations that have *missing value* in jth with other observations that do not have a *missing value* on the variable in accordance with the calculation of the *Euclidean distance* in the formula 1 [15] .

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (1)$$

Where, d(x_i,x_j) is the distance from i to the center of cluster j, Xi is the *training data,* Xj is the *testing data,* n is the number of attributes. k is the attribute, x_{ik} is the ith data on the kth attribute, and x_{jk} is the jth centroid of the kth attribute. Third, look for the shortest K observations based on the smallest distance value. The value of j in the shortest k observations will be used in the imputation process for observations that have a *missing value*. Fourth, calculate the *weight* of all the k shortest observations. The closest observation will get the highest score. Fifth, calculate the *average value* in the shortest k observations that do not have a *missing value* using the formula 2 [16] .

$$X_j = \frac{1}{k}\sum_{k=1}^{k} V_{kj} \quad (2)$$

Where, X_j is the weighted average, V_{kj} = the value of the complete data on the *missing variable value,* and k is the closest observation used. Sixth, carry out the imputation process for *missing* values on observations that have *missing values* using the average value *obtained* at stage 5,

2.4 Root Mean Squared Error (RMSE)

RMSE is one of several methods to measure the accuracy of prediction results or to evaluate prediction techniques. RMSE states the *average value* of the sum of the cubes of the predictions [17] . The smallest value of the RMSE calculation shows that the distribution of the values obtained from the predictions is close to the distribution of the observed values. Written by Makridakis [18] , one of several measures of *error* in prediction is the mean of the square root or RMSE, and the RMSE calculation can be seen in formula 3 [19] .

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}} \quad (3)$$

Where, RMSE is the *Root Mean Square Error,* is the Number of Samples, is the Actual Value of the Index, and = Predicted Value of the Index.

2.5 Mean Absolute Percentage Error (MAPE)

MAPE is the result of the percentage *error* of the forecasting or prediction model. The lower the MAPE percentage value, the lower the forecast or prediction *error,* on the contrary, the higher the MAPE percentage value, the higher the forecast or prediction *error,* The imputation results are very good if the MAPE *error percentage value* is below 10%, while the imputation results are good if the MAPE *error percentage value* is between 10% to 20%. The MAPE *error* percentage value can be calculated using the formula 4 [20] .

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{X_i - F_i}{X_i}\right|}{n} 100\% \quad (4)$$

Where, X_i is the actual data for the i -th period, Fi is the predicted result for the i-period, and n is the number of time periods.

DOI: https://doi.org/10.29207/resti.v6i4.4173

3. Results and Discussions

The dataset in this study consisted of 100 *sample* questionnaire data on the level of user satisfaction of STMIK PPKIA Tarakanita Rahmawati graduates from 2018 to 2021. The data is considered by the authors to be sufficient to represent the distribution of assessments from respondents. The limitation of 100 data is also carried out as a form of time efficiency in the KNNI calculation process because as has been written KNNI will calculate from all *datasets, of* course, the more data, the longer the calculation process. Susanti [14] concluded that her research on imputation is that the more missing values compared to the dataset, the smaller the accuracy obtained. *Training data* consists of 90 data and *testing data* consists of 10 data, which means the author only uses 10 data. % data will be used as *testing data.*

3.1. Training Data

training data used is 90 initial data from user satisfaction data for graduates at STMIK PPKIA Tarakanita Rahmawati from 2018 to 2021. The following snippets of data from the training data that will be used can be seen in Table 3.

Table 3. Training Data

No	Alternative	A1	A2	A3	A4	A5	A6	A7
1	R1	5	5	4	5	5	5	5
2	R2	4	4	4	5	4	4	4
3	R3	4	4	4	4	4	4	4
4	R4	4	4	3	4	5	4	4
5	R5	5	4	4	4	4	4	4
6	R6	4	4	3	4	4	4	3
7	R7	4	4	4	5	4	4	4
8	R8	5	5	4	5	5	5	5
9	R9	5	5	3	5	5	5	5
10	R10	4	4	3	4	5	5	4
..	..	..	..	..	..	..	..	..
81	R81	4	4	3	3	4	4	3
82	R82	4	4	3	4	3	3	3
83	R83	5	5	4	5	5	5	5
84	R84	4	4	3	4	4	4	4
85	R85	4	4	4	4	4	4	4
86	R86	4	5	4	4	5	5	4
87	R87	5	5	3	4	5	5	4
88	R88	5	5	4	5	5	5	5
89	R89	5	5	4	5	5	5	5
90	R90	4	4	3	4	4	4	4

3. 2. Testing Data

testing data used are the last 10 data from graduate user satisfaction data. The data is not data that has a missing value but will randomly remove some values on several attributes or assessment criteria. This is done so that the data still has its original value so that later it can be measured the level of prediction accuracy using the imputation method that the researchers used in this study. The data before it has a missing value can be seen in Table 4 and the data that has been deleted has some values in the sense that it has a missing value, written with "NaN" can be seen in Table 5.

Table 4. Data Testing (without *missing value*)

No	Alternative	A1	A2	A3	A4	A5	A6	A7
1	R91	5	5	5	5	5	5	5
2	R92	4	4	3	4	4	4	3
3	R93	5	5	4	4	5	5	5
4	R94	4	4	3	4	4	4	4
5	R95	4	4	3	2	3	3	4
6	R96	5	4	3	3	5	5	5
7	R97	4	4	3	4	3	3	3
8	R98	5	5	3	5	5	5	4
9	R99	5	5	3	5	5	5	5
10	R100	5	4	2	3	4	4	4

Table 5. Testing Data (with *missing values*)

No	Alternative	A1	A2	A3	A4	A5	A6	A7
1	R91	NaN	5	5	5	5	5	5
2	R92	4	NaN	3	4	4	4	3
3	R93	5	5	NaN	4	5	5	5
4	R94	4	4	3	NaN	4	4	4
5	R95	4	4	3	2	NaN	3	4
6	R96	5	4	3	3	5	NaN	5
7	R97	4	4	3	4	3	3	NaN
8	R98	5	5	3	5	5	NaN	4
9	R99	5	5	3	5	NaN	5	5
10	R100	5	4	2	NaN	4	4	4

3. 3. Calculation of KNNI

The application of K Nearest Neighbor Imputation at this stage calculates manually using *Microsoft Office Excel*. This calculation is used as control data from the main calculation using the KNN *imputer library* from scikit-learn. The first step is to determine the K value of the nearest neighbor to be used. This study uses four K values, namely K=1, K=5, K=10, and K=20. So that later it will be able to calculate which K has the best accuracy value. The second step is to calculate the distance value between the testing data and the training data using the *encludian distance calculation* as in equation 1. The following is an example of calculating the distance between alternatives R1 and R91.

$$R1R91= \frac{6}{7}\sqrt{(5-5)^2+(4-5)^2+(5-5)^2+(5-5)^2+(5-5)^2+(5-5)^2}$$

$$R1R91 = \frac{6}{7}\sqrt{1}$$

$$R1R91 = 0.857142857$$

The following 20 results of the calculation of the distance of the R91 testing data with all *training data*

DOI: https://doi.org/10.29207/resti.v6i4.4173

that have been sorted from the smallest distance can be seen in Table 6.

Table 6. Distance Calculation Results

No	Alternative 1	Alternative 2	Distance
1	R91	R7	0
2	R91	R23	0
3	R91	R28	0
4	R91	R32	0
5	R91	R39	0
6	R91	R57	0
7	R91	R84	0
8	R91	R1	0.857142857
9	R91	R5	0.857142857
10	R91	R12	0.857142857
11	R91	R15	0.857142857
12	R91	R22	0.857142857
13	R91	R24	0.857142857
14	R91	R52	0.857142857
15	R91	R54	0.857142857
16	R91	R59	0.857142857
17	R91	R78	0.857142857
18	R91	R82	0.857142857
19	R91	R10	1.212183053
20	R91	R11	1.212183053

The next step is to calculate the weight of each K value that we have previously determined using equation 2 which will then become the imputed value. The following is an example of calculating the weight of R91 against *a dataset* with k=5.

$$R91K5=\frac{1}{k}(R7A1 + R23A1 + R28A1 + R32A1 + R39A1)$$

$$R91K5=\frac{1}{5}(4 + 4 + 5 + 5 + 4)$$

$$R91K5=4{,}4$$

imputed value of each K value in the R91 *testing data, which* can be seen in Table 7.

Table 7. Value of R91. Imputation Results

No	K	Imputation Results
1	1	4
2	5	4.4
3	10	4.5
4	15	4.6
5	20	4.55

Steps for calculating distances and calculating weights are carried out on all *testing data* on all *training data* so as to produce all imputed results from all testing data, which can be seen in Table 8.

Table 8. Value of Imputed Results

No	Alternative	current	K1	K5	K10	K15	K20
1	R91	5	5	4.6	4.6	4.6	4.6
2	R92	4	4	4.6	4.6	4.53	4.55
3	R93	4	4	3.8	4	4	3.75
4	R94	4	5	4.4	4.5	4.33	4.25
5	R95	3	4	4.4	4.6	4.73	4.65
6	R96	5	4	4.2	4.4	4.4	4.45
7	R97	3	5	4.6	4.7	4.53	4.5
8	R98	5	5	4.8	4.7	4.67	4.5
9	R99	5	5	5	4.9	4.73	4.75
10	R100	3	4	3.2	4	4	4.2

3. 4. KNNI Calculation with the KNN Imputer Library

Calculations are carried out using the KNN *imputer library* from scikit-learn as used by Yazan Jian [21] and also Laboni Akter [22] , the first step by importing numpy which stands for *Numerical Python* function *Python library* which is used to create single and multidimensional array class objects. Then import the KNN *imputer library* from scikit-learn. The command to import numpy is “import numpy as np” where np is just a variable. The command to import KNN Imputer from scikit learn is "from sklearn.impute import KNNImputer".

Then enter *the dataset* containing the *training* data and *testing data* used, the command snippet to enter the *dataset* is as follows "X = [[5,5,4,5,5,5,5], [4,4,4,5,4,4,4], …., [5,5,3,5,5,5,5], [5,4,2,3,4,4,4]]" where X is a variable to accommodate *the dataset array*. Next, determine the desired value of k nearest neighbors, according to this study using five k values, namely 1, 5, 10, 15 and 20. The command determines the value of k as follows "imputer = KNNImputer(n_neighbors=20)". Where *imputer* is a variable and 20 is the selected value of k.

Finally, just call the imputed results from the KNN *imputer library*, the command to display the results is as follows “imputer.fit_transform(X)”, X is the variable that holds the dataset. The result of imputation on data R91 with k nearest neighbors 10 on criteria A1 is 5.

The recap of the imputation results on the R91 to R100 testing data using the k nearest neighbors 1, 5, 10, 15, and 20 shows almost the same results as manual calculations, only I show slightly different results in the hundredth or smaller value like on R9 2 k 20. Where the manual calculation shows the results of 4.55 while the KNN *imputer is* 4.1. The following is a complete recap of the imputation results using the KNN imputer in Table 9.

Table 9. Value of Imputation Results (KNN Imputer)

No	Alternative	current	K1	K5	K10	K15	K20
1	R91	5	5	5	5	5	5
2	R92	4	4	4	4	4	4.1
3	R93	4	3	4	3.9	4.1	4.1
4	R94	4	4	4	4	3.9	4
5	R95	3	4	4	4.1	4.1	4.1
6	R96	5	4	5	4.7	4.6	4.7
7	R97	3	3	3	3.7	3.7	3.7
8	R98	5	5	5	4.7	4.8	4.8
9	R99	5	5	5	5	4.9	4.9
10	R100	3	3	3	4	4	4

3. 5. Evaluation of KNNI Results

Evaluation uses two methods, namely *Root Mean Squared Error* (RMSE) and *Mean Absolute Percentage Error* (MAPE), where the smallest *error value will be the best.* The data used are actual data from graduate user satisfaction data and predictive data from the imputed KNN imputer *library* from scikit-learn. The

DOI: https://doi.org/10.29207/resti.v6i4.4173

RMSE calculation uses the formula in equation 3 and MAPE in equation 4. Where the best accuracy value is the smallest *error value*, the graduate user satisfaction data accuracy test shows that the k closest neighbors 5 are the best results. The error value of the k nearest neighbor 5 on the RMSE is 0.316 while the error value using MAPE is 3.33%, both values are smaller than the other k values of the other closest neighbors. This means that the k nearest neighbors 5 are the best imputation prediction results, both calculated by RMSE and MAPE, even in MAPE the *error value* is below 10% which means it is very good. The results of the *error values* with RMSE and MAPE can be seen in Table 10.

Table 10. Evaluation of RMSE & MAPE

No	Alternative	RMSE	MAPE
1	K1	0.547722558	7.83%
2	K5	0.316227766	3.33%
3	K10	0.537587202	10.78%
4	K15	0.541294744	11.23%
5	K20	0.534789678	11.03%

3.6 Comparison of KNNI and Statistical Methods

The author will compare the results of *missing value predictions* with KNNI and simple statistical methods that are widely used as imputation methods, namely *mean*, median, *most frequent* /mode. Calculations with statistical methods also use the *library* from scikit-learn. The following are the results of missing value predictions with KNNI when compared with the statistical method, which can be seen in graphical form in Figure 2.

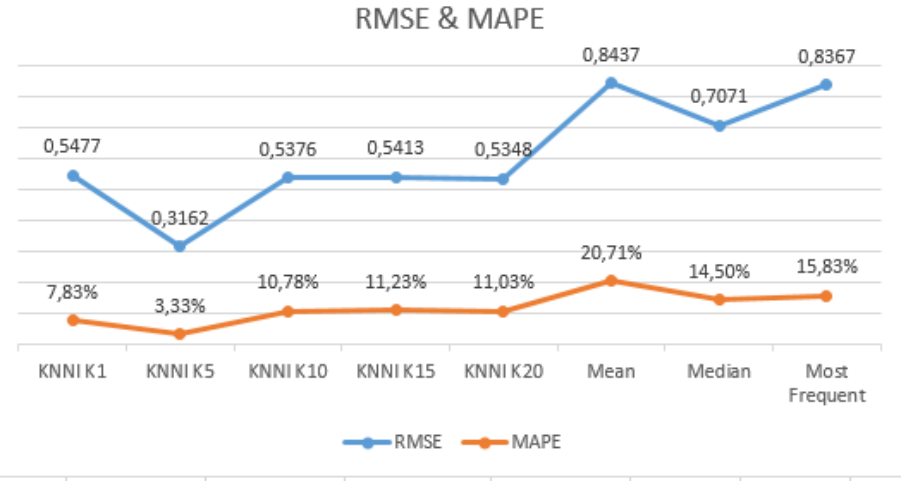

Figure 2. Comparison of KNNI Accuracy and Statistical Methods

Seen from Figure 2, the accuracy of using either RMSE or MAPE for all KNNI scores (K1, K5, K10, K15, and K20) shows much better results than all statistical method models (mean, median, and most frequent). In the statistical method the best accuracy value in the median model is the RMSE value of 0.7071 and the MAPE 14.5%, meaning that the value is still not below 10%.

4. Conclusion

Based on the results of the discussion of the results of the imputation prediction test using K Nearest Neighbor Imputation (KNNI) on user satisfaction data from STMIK PPKIA Tarakanita Rahmawati graduates from 2018 to 2021. Where using the five k closest neighbors, namely 1, 5, 10, 15, and 20 shows the results the error value of the k nearest neighbor 5 on the RMSE is 0.316 while the error value using MAPE is 3.33%, both values are smaller than the k values of the other closest neighbors. This means that the k nearest neighbors 5 are the best imputation prediction results, both calculated by RMSE and MAPE, even in MAPE the error value is below 10% which means it is very good. In further research, the KNNI will be compared with other imputation prediction methods.

Reference

[1] S. P. Mandel J, "A Comparison of Six Methods for Missing Data Imputation," *J. Biom. Biostat.*, vol. 06, no. 01, 2015, doi: 10.4172/2155-6180.1000224.
[2] G. Vink, "Roderick J. Little and Donald B. Rubin: Statistical Analysis with Missing Data," *Psychometrika*, 2022, doi: 10.1007/s11336-022-09856-8.
[3] R. S. Somasundaram and R. Nedunchezhian, "Evaluation of Three Simple Imputation Methods for Enhancing Preprocessing of Data with Missing Values," *Int. J. Comput. Appl.*, vol. 21, no. 10, 2011, doi: 10.5120/2619-3544.
[4] S. Awawdeh, H. Faris, and H. Hiary, "EvoImputer: An evolutionary approach for Missing Data Imputation and feature selection in the context of supervised learning," *Knowledge-Based Syst.*, vol. 236, 2022, doi: 10.1016/j.knosys.2021.107734.
[5] R. B. Kline, *TXTBK Principles and practices of structural equation modelling Ed. 4* ***. 2015.
[6] A. Basu, "Book Review: Missing Data: A Gentle Introduction, by Patrick E. McKnight, Katherine M. McKnight, Souraya Sidani, and Aurelio José Figueredo, New York: Guilford, 2007," *Am. J. Eval.*, vol. 28, no. 3, 2007, doi: 10.1177/1098214007306655.
[7] W. DeSarbo and V. R. Rao, "A Constrained Unfolding Methodology for Product Positioning," *Mark. Sci.*, vol. 5, no. 1, 1986, doi: 10.1287/mksc.5.1.1.

DOI: https://doi.org/10.29207/resti.v6i4.4173

[8] N. M. Laird, "Missing data in longitudinal studies," *Stat. Med.*, vol. 7, no. 1–2, 1988, doi: 10.1002/sim.4780070131.
[9] G. E. A. P. A. Batista and M. C. Monard, "A study of k-nearest neighbour as an imputation method," *Front. Artif. Intell. Appl.*, vol. 87, 2002.
[10] BAN-PT, "Akreditasi Perguruan Tinggi Kriteria dan Prosedur 3.0," *Badan Akreditasi Nas. Perguru. Tinggi*, p. 18, 2019.
[11] S. Y. Siregar, S. St, T. Toharudin, B. Tantular, S. Si, and M. Si, "Performa Metode K Nearest Neighbor Imputation (Knni) Untuk Menangani Multivariate Missing Data," pp. 1–7, 2013.
[12] S. G. Liao *et al.*, "Missing value imputation in high-dimensional phenomic data: Imputable or not, and how?," *BMC Bioinformatics*, vol. 15, no. 1, 2014, doi: 10.1186/s12859-014-0346-6.
[13] P. J. García-Laencina, J. L. Sancho-Gómez, A. R. Figueiras-Vidal, and M. Verleysen, "K nearest neighbours with mutual information for simultaneous classification and missing data imputation," *Neurocomputing*, vol. 72, no. 7–9, 2009, doi: 10.1016/j.neucom.2008.11.026.
[14] S. Susanti, S. Martha, and E. Sulistianingsih, "K NEAREST NEIGHBOR DALAM IMPUTASI MISSING DATA," *Bul. Ilm. Math. Stat. dan Ter.*, 2018.
[15] G. M. Susanto, S. Kosasi, D. David, G. Gat, and S. M. Kuway, "Sistem Referensi Pemilihan Smartphone Android Dengan Metode Fuzzy C-Means dan TOPSIS," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, 2020, doi: 10.29207/resti.v4i6.2584.
[16] I. Hidayatin, S. Adinugroho, and C. Dewi, "Pengelompokan Wilayah berdasarkan Penyandang Masalah Kesejahteraan Sosial (PMKS) dengan Optimasi Algoritme K-Means menggunakan Self Organizing Map (SOM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, 2019.
[17] E. Sartika, "ANALISIS METODE K NEAREST NEIGHBOR IMPUTATION (KNNI) UNTUK MENGATASI DATA HILANG PADA ESTIMASI DATA SURVEY," *TEDC*, 2018.
[18] S. Makridakis, S. Wheelwright C, and V. E. McGee, "Metode dan Aplikasi Peramalan," *Bin. Aksara*, 1999.
[19] Moch Farryz Rizkilloh and Sri Widiyanesti, "Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, 2022, doi: 10.29207/resti.v6i1.3630.
[20] I. M. Yudha Arya Dala, I. K. Gede Darma Putra, and P. Wira Buana, "Forecasting Cases of Dengue Hemorrhagic Fever Using the Backpropagation, Gaussians and Support-Vector Machine Methods," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, 2021, doi: 10.29207/resti.v5i2.2936.
[21] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "Using Machine Learning to Predict Diabetes Complications," 2021. doi: 10.1109/BioSMART54244.2021.9677649.
[22] L. Akter and N. Akhter, "Ovarian Cancer Prediction from Ovarian Cysts Based on TVUS Using Machine Learning Algorithms," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 95, 2022. doi: 10.1007/978-981-16-6636-0_5

DOI: https://doi.org/10.29207/resti.v6i4.4173