ISSN: 2623-1662 (online), 2460-092X (printed)
DOI: https://doi.org/10.19109/jusifo.v10i1.21536
Vol. 10, No. 1 (2024), p.17-26

Research Article [OPEN ACCESS]

A Comparative Study of the Performance of KNN, NBC, C4.5, and Random Forest Algorithms in Classifying Beneficiaries of the Kartu Indonesia Sehat Program

Putri Nabillah*, Inggih Permana, M. Afdal, Fitriani Muttakin, Arif Marsal

ABSTRACT

This study evaluates the performance of various algorithms in determining eligible recipients for the Kartu Indonesia Sehat program. The Random Forest algorithm demonstrated the highest accuracy, precision, and recall, with values of 72.08%, 72.41%, and 99.64%, respectively. The emphasis on recall helps minimize errors in identifying eligible recipients. Additionally, the C4.5 algorithm reduced the total number of variables from 33 to 8, highlighting its computational efficiency. The findings provide valuable insights for the Social Affairs Office of Dumai City in making informed decisions regarding KIS eligibility. The results underscore the effectiveness of using algorithmic approaches to enhance the accuracy and efficiency of aid distribution processes.

Keyword: C4.5 algorithm, kartu indonesia sehat program, random forest algorithm

Received: January 10, 2024; Revised: May 19, 2024; Accepted: June 30, 2024

Corresponding Author: Putri Nabillah, Department of Information System, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia; 11950324987@students.uin-suska.ac.id
Authors: Inggih Permana, Department of Information System, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia, inggihpermana@uin-suska.ac.id; M. Afdal, Department of Information System, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia, m.afdal@uin-suska.ac.id; Fitriani Muttakin, Department of Information System, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia, fitrianimuttakin@uin-suska.ac.id; Arif Marsal, Department of Information System, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia, arif.marsal@uin-suska.ac.id

1. INTRODUCTION

The Kartu Indonesia Sehat (KIS) program, introduced by President Joko Widodo during his 2014 presidential campaign, is a major government initiative aimed at providing health insurance to Indonesian citizens, particularly the poor and disadvantaged, with costs covered by the government (Maliangga et al., 2019). This social assistance program enables low-income citizens to access free medical care at both primary and advanced healthcare facilities (Arifin et al., 2021; Lestari et al., 2020). The selection process for the KIS program remains flawed, resulting in many eligible individuals not receiving the assistance they deserve (Huriah & Nuris, 2023; Junaidi et al., 2023). For example, some participants with high electricity consumption and those with well-maintained buildings receive KIS aid, while others living in inadequate conditions who meet the criteria are not registered as beneficiaries. Consequently, many qualified individuals are not listed as KIS program recipients (Dina et al., 2023).

This study aims to implement a classification model that can evaluate whether a family qualifies for assistance. The algorithms used in this model development include KNN, NBC, C4.5, and Random Forest (Andrian et al., 2020; Azahari & Nursobah, 2021; Y. I. Kurniawan, 2018; Purwanto & Nugroho, 2023; Wuryani & Agustiani, 2021). These algorithms are commonly applied to classify social assistance beneficiaries (Badriah et al., 2021; Ramdani et al., 2022; Rosid et al., 2022). Moreover, they exhibit effective performance in data processing, computational simplicity, and satisfactory accuracy. The NBC algorithm can produce probability values (Nurdin et al., 2021). The KNN algorithm generates values based on the number of K (neighbors) tested (Ula et al., 2022). The C4.5 algorithm generates a decision tree to facilitate decision-making (Harianto & Rosiyadi, 2020), and the Random Forest algorithm, which uses multiple decision trees, is selected for its good performance and high accuracy in data classification. The classification results will be compared to determine which algorithm demonstrates the most adequate performance (Iman & Wijayanto, 2021; Nalatissifa et al., 2020; Widjiyati, 2021).

Several previous studies have utilized the KNN, NBC, C4.5, or Random Forest algorithms. For example, Adzy et al. (2023) used the NBC algorithm to classify eligibility for government health insurance aid, achieving an accuracy rate of 96.83%. Pristiawati et al. (2023) investigated the classification of beneficiaries of the rice for the poor (raskin) program using KNN, NBC, and C4.5 methods, with the C4.5 method yielding the highest accuracy and precision at 88.36% and 93.10%, respectively. Dina et al. (2023) applied the KNN, NBC, and C4.5 algorithms to classify beneficiaries of the Program Keluarga Harapan (PKH), finding that C4.5 had the highest accuracy at 80.16%. I. Kurniawan et al. (2023) utilized the Random Forest algorithm to classify beneficiaries of the raskin program, achieving the highest accuracy at 97.26%.

2. MATERIALS AND METHODS

This research consists of three main phases. The first phase involves data collection, followed by data preprocessing in the second phase, and the final phase includes modeling and evaluation. These research phases can be seen in Figure 1.

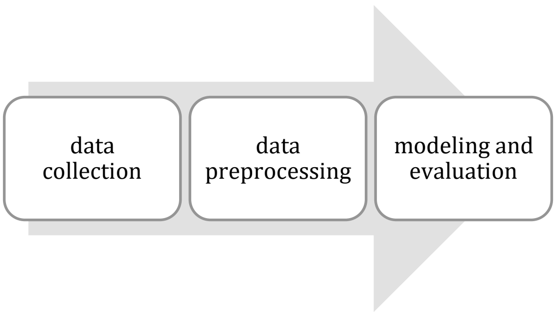

Figure 1. Research phases

2.1 Data Collection

The data for this study were sourced from the Social Affairs Office of Dumai City, Riau Province, specifically from the Integrated Social Welfare Data for the period of January to December 2022, obtained from Pangkalan Sesai, Dumai City.

2.2 Data Preprocessing

During the data preprocessing phase, data cleaning and transformation were performed. Data cleaning addressed missing values by applying the filter example operator in Rapidminer 10.1. Additionally, for data transformation, min-max normalization was employed, setting the minimum value to 0.0 and the maximum value to 1.0. This standardization was implemented using the normalize operator in Rapidminer 10.1.

2.3 Modeling and Evaluation

In this study, four classification algorithm methods were utilized: NBC, K-NN, C4.5, and Random Forest. The implementation of these algorithms was carried out using the RapidMiner application, as detailed below:

1. NBC

   The research employs the NBC process with the Laplace correction parameter to resolve the issue of obtaining a probability score of zero.

2. KNN
   The K values tested in this study are 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21. The Euclidean distance metric was used for measuring distances.
3. C4.5
   The parameters used in the C4.5 algorithm are listed in Table 1.

Table 1. Parameters of the C4.5 algorithm

Parameter	Value
Criterion	Information gain
Maximal Depth	5
Apply Pruning	Confidence=0.5

4. Random Forest
   The number of trees used is 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. The other parameters applied in the Random Forest algorithm are listed in Table 2.

Table 2. Parameters of the random forest algorithm

Parameter	Value
Criterion	Information gain
Maximal Depth	5
Apply Pruning	Confidence=0.5

The validation method used is K-Fold Cross Validation with K = 10. The data is divided into 10 parts, with each part being used as test data in turn, while the remaining parts are used as training data. Performance is assessed based on accuracy (Equation 1), precision (Equation 2), and recall (Equation 3) to provide a more stable and reliable evaluation.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad \text{(Equation 1)}$$

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad \text{(Equation 2)}$$

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad \text{(Equation 3)}$$

3. RESULTS AND DISCUSSION

The dataset acquired from the Social Affairs Office of Dumai City comprises 574 candidates for KIS assistance in the Pangkalan Sesai Subdistrict. The selection process for potential KIS recipients is based on 33 variables, detailed in Table 4 and Table 5. The dataset indicates that 148 individuals received KIS assistance, while 426 did not meet the eligibility criteria. The distribution of this data is illustrated in Table 3, and sample data are presented in Table 6.

Table 3. Data distribution

Category	Before Preprocessing	After Preprocessing
KIS Recipients	148	88
Non-KIS Recipients	426	288
Total	**574**	**376**

Table 4. KIS data variables

No	Variable	Code	Description	Details
1	Household members	AK1	Number of household members	Aligning each family member in each house
2	Building status	AK2	Building occupancy status	1. Private ownership; 2. Rented/leased; 3. Free rent; 4. Service; 5. Others
3	Land status	AK3	Land ownership status	1. Private ownership; 2. Others' assets; 3. Government land; 4. Others
4	Floor	AK4	Type of the most extensive floor	1. Marble/granite; 2. Floor tiles; 3. Parquet/vinyl/carpet; 4. Tiles/terrazzo; 5. High-quality wood/planks; 6. Temporary/bricks; 7. Bamboo; 8. Soil
5	Wall	AK5	Type of the most extensive wall	1. Wall; 2. Bamboo weaving plaster; 3. Wood; 4. Bamboo weaving; 5. Wood stem; 6. Bamboo; 7. Others
6	Wall condition	AK6	Condition of the most extensive wall	1. Good/high quality; 2. Poor/low quality
7	Roof	AK7	Type of the most extensive roof	1. Concrete/concrete tile; 2. Ceramic tile; 3. Metal tile; 4. Clay tile; 5. Asbestos; 6. Zinc; 7. Shingle; 8. Bamboo; 9. Straw/thatched/leaves
8	Roof condition	AK8	Condition of the most extensive roof	1. Good/high quality; 2. Poor/low quality
9	Number of rooms	AK9	Number of bedrooms	Arranging the number of rooms in the house
10	Drinking water	AK10	Source of drinking water	1. Branded drink; 2. Refilled water; 3. Metered tape; 4. Retail tape; 5. Bore/pump water source; 6. Protected water source; 7. Unprotected water source
11	Lighting source	AK11	Main lighting source	1. PLN electricity; 2. non-PLN electricity; 3. non-electricity
12	Power capacity	AK12	Installed power capacity	1. 450-watt electricity; 2. 900-watt electricity; 3. 1,300-watt electricity; 4. 2,200-watt electricity; 5. > 2,200-watt electricity; 6. non-electricity
13	Cooking fuel	AK13	Main cooking fuel/energy	1. Electricity; 2. Using 3kg gas; 3. Gas below 3kg; 4. City gas/biogas; 5. Fuel oil; 6. Briquettes; 7. Charcoal; 8. Firewood; 9. Not cooking
14	Toilet facility	AK14	Use of toilet facility	1. Own; 2. Shared; 3. None
15	Toilet type	AK15	Type of toilet	1. Goose neck; 2. Slope; 3. Cubluk; 4. Not using
16	Waste disposal	AK16	Final disposal place for feces	1. Tank; 2. Slope; 3. Soil hole; 4. Pond/field/river/ lake/sea; 5. Ground/paddock/garden; 6. Others
17	Gas cylinder	AK17	Owning a gas cylinder of 5.5 kg or more	1. Yes; 2. No
18	Refrigerator	AK18	Owning a refrigerator	1. Yes; 2. No
19	Air conditioner	AK19	Owning an air conditioner	1. Yes; 2. No
20	Water heater	AK20	Owning a water heater	1. Yes; 2. No
21	Telephone	AK21	Owning a home telephone (PSTN)	1. Yes; 2. No
22	Television	AK22	Owning a television	1. Yes; 2. No
23	Gold/jewelry	AK23	Owning gold/jewelry and savings (worth 10 grams of gold)	1. Yes; 2. No

Table 5. KIS data variables (continued)

No	Variable	Code	Description	Details
24	Computer/ laptop	AK24	Owning a computer/laptop	1. Yes; 2. No
25	Bicycle	AK25	Owning a bicycle	1. Yes; 2. No
26	Motorcycle	AK26	Owning a motorcycle	1. Yes; 2. No
27	Car	AK27	Owning a car	1. Yes; 2. No
28	Boat	AK28	Owning a boat	1. Yes; 2. No
29	Outboard motor	AK29	Owning an outboard motor	1. Yes; 2. No
30	Ship	AK30	Owning a ship	1. Yes; 2. No
31	Immovable assets	AK31	Household owning immovable assets	a. Land: 1. Yes; 2. No b. Dwelling elsewhere: 3. Yes; 4. No
32	Another house	AK32	Household owning another house	1. Yes; 2. No
33	Business status	AK33	Any household member owning their own/shared business	1. Yes; 2. No

Table 6. KIS sample data

AK1	AK2	AK3	...	AK7	AK8	...	Status*
5	2	2	...	6	2	...	YES
4	NULL	NULL	...	NULL	NULL	...	YES
1	NULL	NULL	...	NULL	NULL	...	NO
1	NULL	NULL	...	NULL	NULL	...	NO
1	NULL	NULL	...	NULL	NULL	...	YES
1	NULL	NULL	...	NULL	NULL	...	NO
4	2	2	...	6	2	...	NO
1	NULL	NULL	...	NULL	NULL	...	YES
2	1	1	...	6	2	...	YES
4	1	1	...	6	2	...	YES
4	2	2	...	6	2	...	YES

*Status: if YES, the household receives the KIS program; if NO, the household does not receive the KIS program

3.1 Data Preprocessing Implementation

At this stage, data containing NULL values were removed, resulting in 376 remaining rows. There are 88 records labeled YES (received KIS program) and 288 records labeled NO (did not receive KIS program). An example of dataset cleaning can be seen in Table 7, while a sample of the normalized data can be found in Table 8.

Table 7. Sample of data cleaning

No	AK1	AK2	AK3	AK4	AK5	AK6	...	AK28	AK29	AK30	AK31	AK32	AK33	Status
1	5	2	2	6	3	2	...	4	2	4	2	4	1	YES
2	4	1	1	8	3	2	...	4	2	4	1	4	2	YES
3	4	2	2	6	1	2	...	4	2	4	2	4	2	NO
4	4	2	2	6	1	2	...	4	2	4	2	4	2	NO
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
375	4	2	2	6	1	2	...	4	2	4	2	4	0	NO
376	6	1	1	6	1	2	...	4	2	4	2	4	0	NO

Table 8. Sample of data normalization

AK1	AK2	AK3	AK4	AK5	AK6	...	AK28	AK29	AK30	AK31	AK32	AK33	Status
5.0	2.0	2.0	6.0	3.0	2.0	...	4.0	2.0	4.0	2.0	4.0	1.0	YES
4.0	1.0	1.0	8.0	3.0	2.0	...	4.0	2.0	4.0	1.0	4.0	2.0	YES
4.0	2.0	2.0	6.0	1.0	2.0	...	4.0	2.0	4.0	2.0	4.0	2.0	NO
4.0	2.0	2.0	6.0	1.0	2.0	...	4.0	2.0	4.0	2.0	4.0	2.0	NO
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4.0	2.0	2.0	6.0	3.0	2.0	...	4.0	2.0	4.0	2.0	4.0	.0	NO
6.0	2.0	2.0	5.0	3.0	2.0	...	4.0	2.0	4.0	2.0	4.0	.0	YES
4.0	2.0	2.0	6.0	1.0	2.0	...	4.0	2.0	4.0	2.0	4.0	.0	NO
6.0	1.0	1.0	6.0	1.0	2.0	...	4.0	2.0	4.0	2.0	4.0	.0	NO

3.2 Modeling and Evaluation Implementation

1. Performance of KNN

   As shown in Table 9, the highest accuracy is achieved when the K value in KNN is 21, with an accuracy of approximately 71.55%. The highest precision is obtained when the K value in KNN is 7, with a precision of 72.09%. Meanwhile, the highest recall is observed when the K value in KNN is 21, with recall reaching 98.89%. Additionally, the results of the KNN confusion matrix can be seen in Table 10.

Table 9. Performance of KNN

No	K	Accuracy	Precision	Recall
1	3	60,64%	70,78%	77,14%
2	5	61,69%	70,10%	81,57%
3	7	66,72%	**72,09%**	87,83%
4	9	65,43%	70,55%	89,31%
5	11	67,82%	71,46%	92,25%
6	13	67,80%	71,30%	92,62%
7	15	69,42%	71,68 %	95,21%
8	17	70,48 %	71,87%	97,04%
9	19	70,21%	71,80%	96,69%
10	21	**71,55%**	72,04%	**98,89%**

Table 10. Confusion matrix of KNN

	true No	true Yes
Pred. No	11	50
Pres. Yes	94	221

2. Performance of NBC

   As shown in Table 11, the performance of NBC shows an accuracy of 66.49%, a precision of 71.13%, and a recall of 90.04%. Additionally, the results of the NBC confusion matrix can be seen in Table 12.

Table 11. Performance of NBC

	Accuracy	Precision	Recall
NBC	66,49%	71,13%	90,04%

Table 12. Confusion matrix of NBC

	true No	true Yes
Pred. No	6	27
Pres. Yes	99	244

3. Performance of C4.5
As shown in Table 13, the performance of the C4.5 algorithm achieved an accuracy of 70.47%, a precision of 71.84%, and a recall of 97.08%. The decision tree resulting from the C4.5 method can be seen in Figure 2. Additionally, the results of the C4.5 confusion matrix can be found in Table 14.

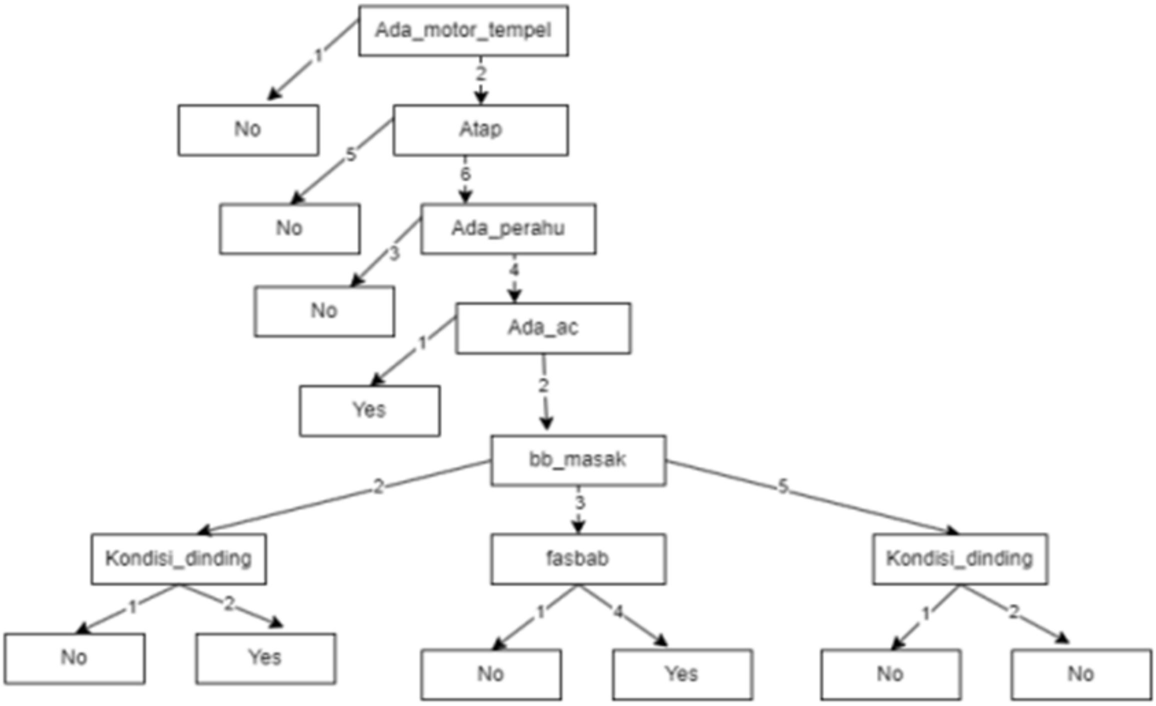

Figure 2. Decision tree of C4.5

Table 13. Performance of C4.5

	Accuracy	Precision	Recall
C4.5	70,47%	71,84%	97,08%

Table 14. Confusion matrix of C4.5

	true No	true Yes
Pred. No	2	8
Pres. Yes	103	263

4. Performance of Random Forest
As shown in Table 15, the highest accuracy is achieved when the number of trees is 50, with an accuracy of 72.08%. The highest precision is obtained when the number of trees is 10, with a precision of 72.41%. The highest recall is observed when the number of trees is 50, with a recall of 99.64%. Additionally, the confusion matrix for the random forest can be seen in Table 16.

Table 15. Performance of Random Forest

No	Number of Tree	Accuracy	Precision	Recall
1	10	71,56%	**72,41%**	97,82%
2	20	71,29%	71,99%	98,56%
3	30	71,56%	72,07%	98,93%
4	40	71,81%	72,14%	99,27%
5	50	**72,08%**	72,22%	**99,64%**
6	60	72,08%	72,22%	99,64%
7	70	72,08%	72,22%	99,64%
8	80	72,08%	72,22%	99,64%
9	90	72,08%	72,22%	99,64%
10	100	72,08%	72,22%	99,64%

Table 16. Confusion matrix of Random Forest

	true No	true Yes
Pred. No	1	1
Pres. Yes	104	270

3.3 Comparison of Algorithm Performance

As shown in Figure 3, the Random Forest algorithm achieved the highest accuracy. In this test, the Random Forest algorithm also produced the highest recall. This indicates that the model generated by the Random Forest algorithm is better at reducing errors in determining the eligible recipients of aid. The C4.5 algorithm also has the advantage of computational simplicity. Consequently, the classification achieved by the C4.5 method reduced the total number of variables from 33 to 8, providing a simpler yet effective solution.

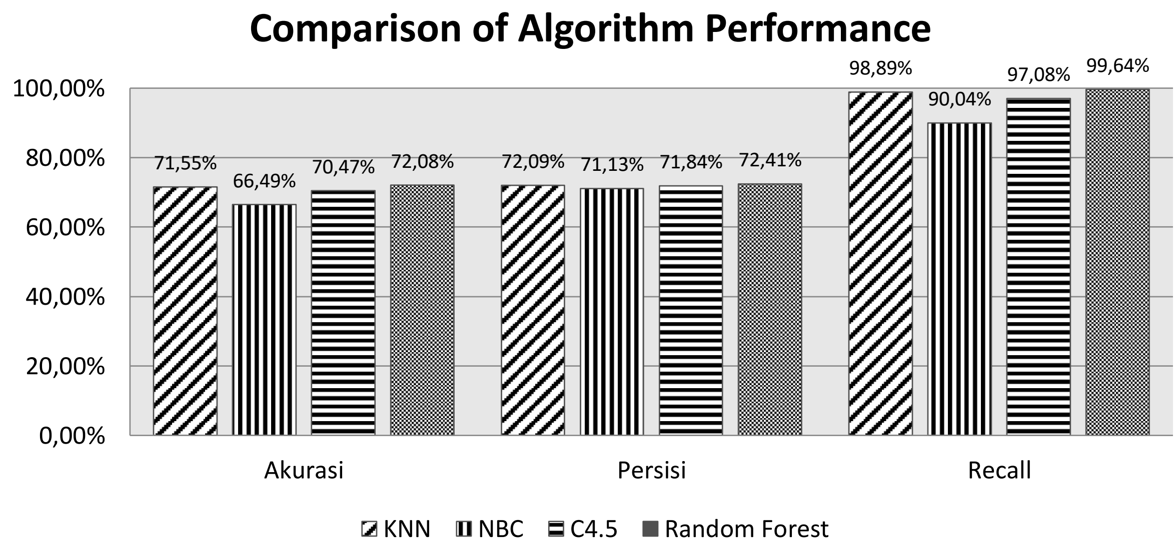

Figure 3. Performance achievements of each algorithm

4. CONCLUSION

This study reveals that the Random Forest algorithm is the best, achieving the highest recall, accuracy, and precision compared to other algorithms, with accuracy of 72.08%, precision of 72.41%, and a recall of

99.64%. In this research, recall is preferred as it helps reduce the number of eligible recipients of the KIS program being classified as ineligible. Additionally, the decision tree formed by the C4.5 method successfully reduced the total number of variables from 33 to 8. The C4.5 algorithm operates concisely in terms of computational complexity.

The results of this study can serve as a consideration for the the Social Affairs Office of Dumai City in making decisions regarding KIS recipients. Therefore, the classification outcomes of this study demonstrate how decisions can be made based on the mathematical calculations of each algorithm.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adzy, L. B., Asriyanik, A., & Pambudi, A. (2023). Algoritma naïve bayes untuk klasifikasi kelayakan penerima bantuan iuran jaminan kesehatan pemerintah daerah kabupaten sukabumi. *Jurnal Mnemonic*, *6*(1), 1–10.

Andrian, A., Steele, S., Salim, E. S., Bindan, H., Pranoto, E., & Dharma, A. (2020). Analisa metode random forest tree dan k-nearest neighbor dalam mendeteksi kanker serviks. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, *3*(2), 97–101. https://ejournal.sisfokomtek.org/index.php/jikom/article/view/73

Arifin, N. Y., Suri, G. P., Veza, O., & Setyabudhi, A. L. (2021). Dashboard penentuan penerima kartu indonesia sehat (kis) pada kelurahan binuang kampung dalam dengan metode simple additive weighting. *JURSIMA*, *9*(1), 1–9. https://ojsiibn1.indobarunasional.ac.id/index.php/jursima/article/view/239

Azahari, A., & Nursobah, N. (2021). Rekomendasi penerimaan beasiswa yayasan untuk siswa baru smk ti airlangga dengan algoritma c4.5. *Jurnal Media Informatika Budidarma*, *5*(2), 609–614. https://doi.org/10.30865/MIB.V5I2.2943

Badriah, S., Nugroho, M. F. E., Sanjaya, N., Rismawati, I., Sari, B. N., & Rozikin, C. (2021). Klasifikasi algoritma c4.5 dalam menentukan penerima bantuan covid-19 (studi kasus: desa di karawang). *Jurnal Informatika Polinema*, *7*(3), 23–28. https://doi.org/10.33795/JIP.V7I3.620

Dina, A., Permana, I., Muttakin, F., & Maita, I. (2023). Perbandingan algoritma nbc, knn, dan c4.5 untuk klasifikasi penerima bantuan program keluarga harapan. *Jurnal Media Informatika Budidarma*, *7*(3), 1079–1087. https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/6316

Harianto, H., & Rosiyadi, D. (2020). Komparasi algortima c4.5, naïve bayes dan k-nearest neighbor sebagai sistem pendukung keputusan menaikkan jumlah peserta didik. *Jurnal Informatika*, *7*(1), 55–61. https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/7250

Huriah, D. A., & Nuris, N. D. (2023). Klasifikasi penerima bantuan sosial umkm menggunakan algoritma naïve bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 360–365. https://doi.org/10.36040/JATI.V7I1.6300

Iman, Q., & Wijayanto, A. W. (2021). Klasifikasi rumah tangga penerima beras miskin (raskin)/beras sejahtera (rastra) di provinsi jawa barat tahun 2017 dengan metode random forest dan support vector machine. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, *9*(2), 178–184. https://doi.org/10.26418/JUSTIN.V9I2.44137

Junaidi, A., Agustiani, S., Agustyaningrum, C. I., & Arifin, Y. T. (2023). Klasifikasi penerima bantuan sosial menggunakan algoritma c4.5. *Jurnal Teknik Komputer AMIK BSI*, *9*(1), 77–82. https://doi.org/10.31294/JTK.V9I1.14378

Kurniawan, Y. I. (2018). Perbandingan algoritma naive bayes dan c.45 dalam klasifikasi data mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(4), 455–464. https://doi.org/10.25126/JTIIK.201854803

Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). Implementasi algoritma random forest untuk menentukan penerima bantuan raskin. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *10*(2), 421–428. https://doi.org/10.25126/JTIIK.20231026225

Lestari, W., Fatoni, F., & Hutrianto, H. (2020). Implementasi data mining untuk kartu indonesia sehat bagi masyarakat kurang mampu menggunakan metode clustering pada dinas sosial kota palembang. *Jurnal Nasional Ilmu Komputer*, *1*(4), 169–174. https://doi.org/10.47747/JURNALNIK.V1I4.163

Maliangga, M., Walewangko, E. N., & Londa, A. T. (2019). Pengaruh kebijakan pemerintah kartu indonesia pintar (kip) dan kartu indonesia sehat (kis) terhadap konsumsi rumah tangga miskin di kecamatan dumoga timur kabupaten bolaang mongondow. *Jurnal Berkala Ilmiah Efisiensi*, *19*(01). https://ejournal.unsrat.ac.id/v3/index.php/jbie/article/view/22311

Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2020). Perbandingan kinerja algoritma klasifikasi naive bayes, support vector machine (svm), dan random forest untuk prediksi ketidakhadiran di tempat kerja. *Jurnal Informatika Universitas Pamulang*, *5*(4), 578–584. https://doi.org/10.32493/INFORMATIKA.V5I4.7575

Nurdin, N., Suhendri, M., Afrilia, Y., & Rizal, R. (2021). Klasifikasi karya ilmiah (tugas akhir) mahasiswa menggunakan metode naive bayes classifier (nbc). *SISTEMASI*, *10*(2), 268.

Pristiawati, A. P., Permana, I., Zarnelly, Z., & Muttakin, F. (2023). Klasifikasi penerima bantuan beras miskin menggunakan algoritma k-nn, nbc dan c4.5. *Building of Informatics, Technology and Science (BITS)*, *5*(1), 336–344-336–344. https://doi.org/10.47065/BITS.V5I1.3617

Purwanto, A., & Nugroho, H. W. (2023). Analisa perbandingan kinerja algoritma c4.5 dan algoritma k-nearest neighbors untuk klasifikasi penerima beasiswa. *Jurnal Teknoinfo*, *17*(1), 236–243. https://doi.org/10.33365/JTI.V17I1.2370

Ramdani, A., Sofyan, C. D., Ramdani, F., Tama, M. F. A., & Rachmatsyah, M. A. (2022). Algoritma klasifikasi data mining untuk memprediksi masyarakat dalam menerima bantuan sosial. *Jurnal Ilmiah Sistem Informasi*, *1*(2), 39–47. https://doi.org/10.51903/JUISI.V1I2.363

Rosid, A., Nurdiawan, O., & Dwilestari, G. (2022). Klasifikasi penerima bantuan sosial dengan algoritma random forest untuk penanganan covid 19. *JURSIMA*, *10*(2), 114–120. https://ojsiibn1.indobarunasional.ac.id/index.php/jursima/article/view/398

Ula, M., Zulhusna, R., Fhonna, R. P., & Pratama, A. (2022). Penerapan model klasifikasi k-nearest neighbor dalam pencarian kesesuaian pekerjaan. *METIK Jurnal*, *6*(1), 18–23. https://doi.org/10.47002/METIK.V6I1.343

Widjiyati, N. (2021). Implementasi algoritme random forest pada klasifikasi dataset credit approval. *Jurnal Janitra Informatika Dan Sistem Informasi*, *1*(1), 1–7. https://doi.org/10.25008/JANITRA.V1I1.118

Wuryani, N., & Agustiani, S. (2021). Random forest classifier untuk deteksi penderita covid-19 berbasis citra ct scan. *Jurnal Teknik Komputer AMIK BSI*, *7*(2), 187–193. https://doi.org/10.31294/JTK.V7I2.10468