International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 5, No. 4, (2021), pp. 421-430

The K-NN method was used to assess student satisfaction with the services provided by employees of research and service institutions

Muhammad Rais Wathani[1], Dian Agustini[2], Muthia Farida[3], Mokhamad Ramdhani Raharjo[4]

[1,2,3,4]Universitas Islam Kalimantan Muhammad Arsyad Al Banjari, Banjarmasin, South Kalimantan

[1]raiswathani@uniska-bjm.ac.id,[2]dian.ftiuniskabjm@gmail.com, [3]muthiafarida59@gmail.com

Abstract

Population growth caused by the year of birth led to the classification of population groups into several generations. Classification is important because in each generation there is based on population growth has different characteristics and traits in each generation. This research was conducted to try to group generations based on provinces in Indonesia based on the number of residents owned. When researchers analyzed the data obtained from population census data conducted by the central statistics agency (BPS). The method used in generation classification grouping uses the K-Means algorithm method based on 3 clusters. Based on the results of calculations carried out for 3 clusters obtained cluster 1 has 25 provinces, cluster 2 has 3 provinces and cluster 3 has 6 provinces. Based on the 2020 census that has been conducted, the current population is generation Z, generation and Pre Boomer generation is last in line so that from the available data can provide information about mapping in 34 provinces to be able to improve communication patterns between generations and fulfill public facilities that can be used every generation.

Keywords: *Generation clustering, K-Means, Clustering*

1. Introduction

Service is an action taken to meet the needs of others (consumers, customers, guests, clients, patients, passengers, etc.) whose level of satisfaction can only be felt by those who serve and those who are served. The services provided can lead to a person's positive or negative attitude towards service providers [1]. Therefore, the service must be carried out optimally so that satisfaction arises between the person being served and the person serving. Customer satisfaction or the satisfaction of someone being served is a person's feeling of pleasure or disappointment that arises after comparing his perceptions or impressions of the performance or results of a product and his expectations [1]. If the service provided is in accordance with customer expectations, then they will feel satisfied. This research was conducted in one of the universities in the province of South Kalimantan. The performance of research and community service institutions (abbreviated as LPPM) in an institution can be assessed from the services provided by the institution to lecturers and students. From this, it can be analyzed how the services provided by the institution to lecturers, especially to students. Services that are less than optimal can affect the achievement of the institution's goals. Due to the need for a reciprocal relationship between students and the institution in order to fulfill the assigned tasks. An LPPM has several tasks, one of which is realizing planning, implementing, and coordinating research activities, application and updating of science and technology within the institution, both in the form of research and community service for the prosperity of the community and increasing the competitiveness of the nation.

ISSN : 2580-7250

Based on these problems, research was conducted to assess student satisfaction with the services of LPPM employees as a form of evaluation so that existing employees can improve services and can realize the vision and mission that has been set. Many branches of computer science can solve this case. One of them is decision support system [2]–[6], datamining [7]–[9], artificial neural network [10]–[13] and others. Each branch of science has its own advantages. Based on these problems, researchers used data mining with the K-Nearest Neighbors (KNN) method. Data mining is a series of processes that use one or more computer learning techniques to analyze and expand knowledge automatically or a series of processes to explore more values from a data set in the form of knowledge that has not been known manually [14]. The application of data mining in predicting the level of satisfaction has been carried out in previous studies. One of them is a study conducted by [15], on the assessment of the satisfaction level of visitors to the animal park as a result of this study. The results of processing the C4.5 method using the RapidMiner software where the attributes of infrastructure (C2) and service officers (C5) are the attributes that most influence the level of satisfaction of animal park visitors. Based on these reasons, it is hoped that research using the K-NN algorithm can produce an analysis and implementation of a system on student satisfaction in LPPM employee services so that it can be input to improve service quality and achieve the existing vision and mission.

2. Research and Methodology

2.1. Classification

Data classification is a process that finds the same properties in a set of objects in a database and classifies them into different classes according to a defined classification model [16-20]. The purpose of classification is to find a model from the training set that distinguishes attributes into the appropriate category or class, the model is then used to classify attributes whose class has not been previously known [20-22].

2.2. K-NN (K-Nearest Neighbor)

The K-Nearest Neighbor (KNN) algorithm is a method for classifying objects based on learning data that has the closest distance to the object. Learning data is projected into a multidimensional space, where each dimension displays the characteristics of the data [23-25]. The application of the K-NN method through several steps.

a) Determine the parameter k.
b) Calculate the distance between the data to be evaluated with all training.
c) Sort the distance formed (in ascending order).
d) Determine the closest distance to the order k.
e) Match the appropriate class.
f) Find the number of classes from the nearest neighbor and set the class as the data class to be evaluated [26].

3. Results and Discussion

3.1. Solution to problem

a) Data

After analyzing the existing problems with the data mining stages to produce a classification of student satisfaction at LPPM employee services, this analysis ends with carrying out the actual data mining process, the results achieved are to determine the classification of student satisfaction levels in LPPM employee services. This study uses the help of the RapidMiner application to facilitate the data mining process that produces information on the classification of student satisfaction.

Table 1. Criteria/Attributes

No	Attribute
1	Ability

No	Attribute
2	Appearance
3	Responsibility
4	Attitude and Behavior

From the attributes that have been determined, the results of the assessment given by the respondents have been converted into the form of training data as follows:

Table 2. Conversion Data

Ability	Appearance	Responsibility	Attitude and Behavior	Label
81,5	90	90	88	1
85	96	91	89	1
83,5	78	88	88	1
79	87	87	77	2
65	88	78	91	1
78	85	88	88	1
77	67	83	78	2
66,5	78	85	77	2
91	66	88	85	1
71	65	84	86	1
68	78	67	76	2
78	77	88	78	2
79	80	78	87	1
88	89	66	79	2
83	88	78	88	1
84	78	87	86	1
74	86	77	85	1
67	75	67	78	2
78	77	75	77	2
79	67	73	89	1
81	88	77	90	1
78	79	88	88	1
68	77	76	89	1
77	69	76	78	1
76	77	73	88	1
68	66	76	87	2

The dataset in Table 3 is obtained from the results of the conversion of the questionnaire into the form of testing data. The data was obtained by distributing questionnaires to the respondents and the respondents assessed in a range between 1-100 against each of the existing criteria which were converted into training data.

b) K-Nearet Neighbor

Based on the data mining stages for the K-Nearet Neighbor algorithm, the steps for the K-Nearest Neighbor are:

1) Determination of the value of k. The determination of the value of k used does not have a standard rule.
2) Calculate the distance between the training data and test data (Test) in the transformation stage using the Euclidean distance calculation as follows:

$$d_1 = \sqrt{(81{,}5-88)^2+(90-89)^2+(90-87)^2+(88-88)^2} = 7{,}228$$
$$d_2 = \sqrt{(85-88)^2+(96-89)^2+(91-87)^2+(89-88)^2} = 8{,}660$$
$$d_3 = \sqrt{(83{,}5-88)^2+(78-89)^2+(88-87)^2+(88-88)^2} = 11{,}926$$
$$d_4 = \sqrt{(79-88)^2+(87-89)^2+(87-87)^2+(77-88)^2} = 14{,}352$$
$$d_5 = \sqrt{(65-88)^2+(88-89)^2+(78-87)^2+(91-88)^2} = 24{,}899$$

$$d_6 = \sqrt{(78-88)^2+(85-89)^2+(88-87)^2+(88-88)^2} = 10{,}816$$
$$d_7 = \sqrt{(77-88)^2+(67-89)^2+(83-87)^2+(78-88)^2} = 26{,}851$$
$$d_8 = \sqrt{(66{,}5-88)^2+(78-89)^2+(85-87)^2+(77-88)^2} = 26{,}612$$
$$d_9 = \sqrt{(91-88)^2+(66-89)^2+(88-87)^2+(85-88)^2} = 23{,}409$$
$$d_{10} = \sqrt{(71-88)^2+(65-89)^2+(84-87)^2+(86-88)^2} = 29{,}631$$
$$d_{11} = \sqrt{(68-88)^2+(78-89)^2+(67-87)^2+(76-88)^2} = 32{,}634$$
$$d_{12} = \sqrt{(78-88)^2+(77-89)^2+(88-87)^2+(78-88)^2} = 18{,}574$$
$$d_{13} = \sqrt{(79-88)^2+(80-89)^2+(78-87)^2+(87-88)^2} = 15{,}620$$
$$d_{14} = \sqrt{(88-88)^2+(89-89)^2+(66-87)^2+(79-88)^2} = 22{,}847$$
$$d_{15} = \sqrt{(83-88)^2+(88-89)^2+(78-87)^2+(88-88)^2} = 10{,}344$$
$$d_{16} = \sqrt{(84-88)^2+(78-89)^2+(87-87)^2+(86-88)^2} = 11{,}874$$
$$d_{17} = \sqrt{(74-88)^2+(86-89)^2+(77-87)^2+(85-88)^2} = 17{,}720$$
$$d_{18} = \sqrt{(67-88)^2+(75-89)^2+(67-87)^2+(78-88)^2} = 33{,}719$$
$$d_{19} = \sqrt{(78-88)^2+(77-89)^2+(75-87)^2+(77-88)^2} = 22{,}561$$
$$d_{20} = \sqrt{(79-88)^2+(67-89)^2+(73-87)^2+(89-88)^2} = 27{,}604$$
$$d_{22} = \sqrt{(81-88)^2+(88-89)^2+(77-87)^2+(90-88)^2} = 12{,}409$$
$$d_{22} = \sqrt{(78-88)^2+(79-89)^2+(88-87)^2+(88-88)^2} = 14{,}177$$
$$d_{23} = \sqrt{(68-88)^2+(77-89)^2+(76-87)^2+(89-88)^2} = 25{,}806$$
$$d_{24} = \sqrt{(83-88)^2+(88-89)^2+(78-87)^2+(88-88)^2} = 27{,}2396$$
$$d_{25} = \sqrt{(76-88)^2+(77-89)^2+(73-87)^2+(88-88)^2} = 22$$
$$d_{26} = \sqrt{(68-88)^2+(77-89)^2+(76-87)^2+(87-88)^2} = 32{,}419$$

From the distance calculation results, complete results are obtained which are summarized in the following table:

Table 3. Results of distance calculations using euclidean distance

Ability	Appearance	Responsibility	Attitude and Behavior	Label	Distance
81,5	90	90	88	1	7,228416
85	96	91	89	1	8,660254
83,5	78	88	88	1	11,92686
79	87	87	77	2	14,3527
65	88	78	91	1	24,8998
78	85	88	88	1	10,81665
77	67	83	78	2	26,85144
66,5	78	85	77	2	26,61297
91	66	88	85	1	23,4094
71	65	84	86	1	29,63106
68	78	67	76	2	32,63434
78	77	88	78	2	18,57418
79	80	78	87	1	15,6205
88	89	66	79	2	22,84732
83	88	78	88	1	10,34408
84	78	87	86	1	11,87434
74	86	77	85	1	17,72005
67	75	67	78	2	33,71943
78	77	75	77	2	22,56103
79	67	73	89	1	27,60435
81	88	77	90	1	12,40967
78	79	88	88	1	14,17745
68	77	76	89	1	25,80698
77	69	76	78	1	27,23968

Ability	Appearance	Responsibility	Attitude and Behavior	Label	Distance
76	77	73	88	1	22
68	66	76	87	2	32,41913

3) Sequencing of the calculated data. The distance that has been obtained is then sorted again from the closest to the farthest (ascending).
4) d) Determine the test data group based on the majority label from the k closest neighbors, because the value of k = 10, the 10 smallest distances are taken.

Table 4. Calculation Results with K=1, K=5, and K=10

Ability	Appearance	Responsibility	Attitude and Behavior	Label	Distance	K=1	k=5	K=10
81,5	90	90	88	1	7,228416	1	1	1
85	96	91	89	1	8,660254		1	1
83,5	78	88	88	1	11,92686			1
79	87	87	77	2	14,3527			1
65	88	78	91	1	24,8998			
78	85	88	88	1	10,81665		1	1
77	67	83	78	2	26,85144			
66,5	78	85	77	2	26,61297			
91	66	88	85	1	23,4094			
71	65	84	86	1	29,63106			
68	78	67	76	2	32,63434			
78	77	88	78	2	18,57418			
79	80	78	87	1	15,6205			1
88	89	66	79	2	22,84732			
83	88	78	88	1	10,34408		1	1
84	78	87	86	1	11,87434		1	1
74	86	77	85	1	17,72005			
67	75	67	78	2	33,71943			
78	77	75	77	2	22,56103			
79	67	73	89	1	27,60435			
81	88	77	90	1	12,40967			1
78	79	88	88	1	14,17745			1
68	77	76	89	1	25,80698			
77	69	76	78	1	27,23968			
76	77	73	88	1	22			
68	66	76	87	2	32,41913			

5) By using the most majority nearest neighbor category, it can be predicted that the level of satisfaction in the next data can be predicted.

c) RapidMiner Implementation

Rapidminer is one of the data mining software for dataset processing to find data patterns according to the purpose of processing the data.

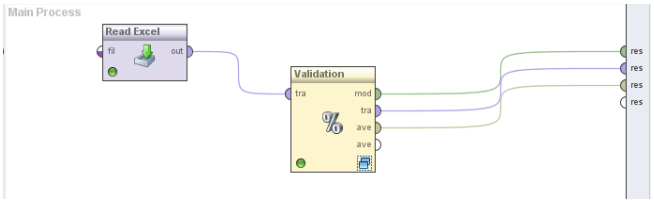

Figure 1. Input excel data into the RapidMiner process

Figure 1 shows how the process of entering data in excel format into the RapidMiner process and connecting it to the validation process.

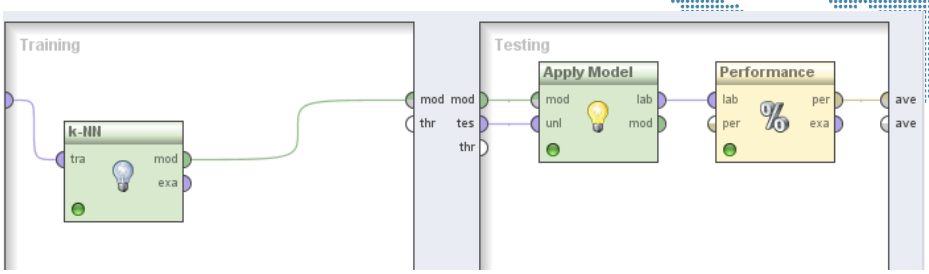

Figure 2. K-Nearet Neighbor classification sub-process model

Figure 2 is a process in the validation that is connecting the KNN algorithm with the validation test process which will then produce the values of the validation, namely accuracy data, recall and precision. The results of the accuracy of performance at validation = 5 of 81.33% for the process carried out in the model above can be seen in Figure 3.

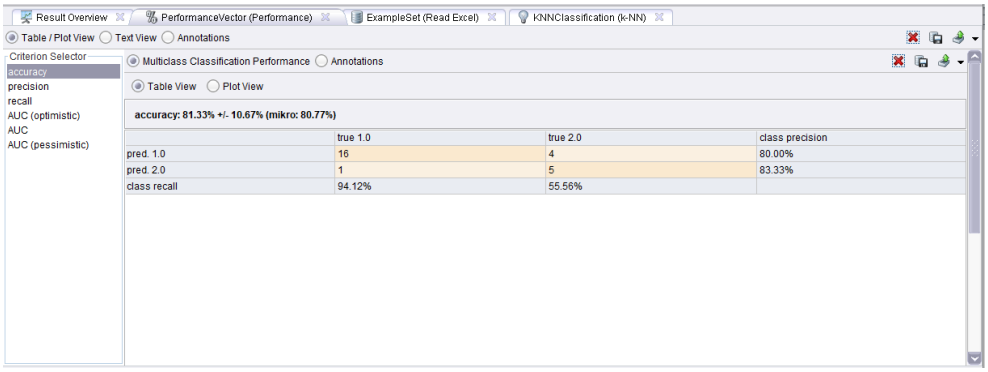

Figure 3. Accuracy value on validation = 5

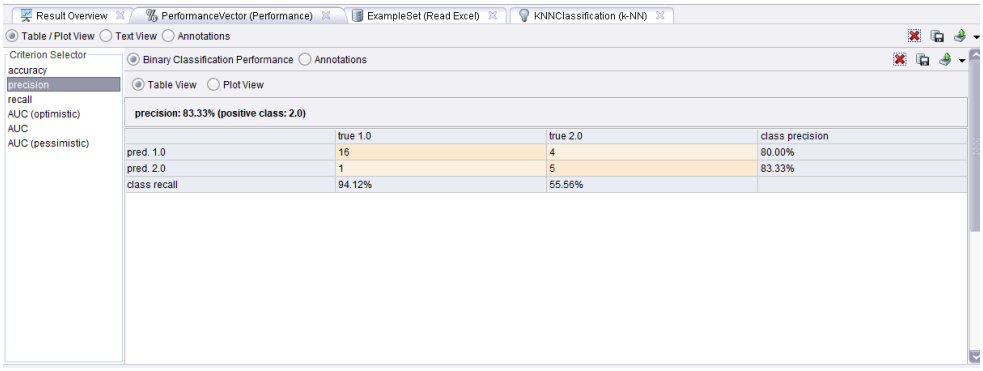

Figure 4. Validation chest precision value = 5

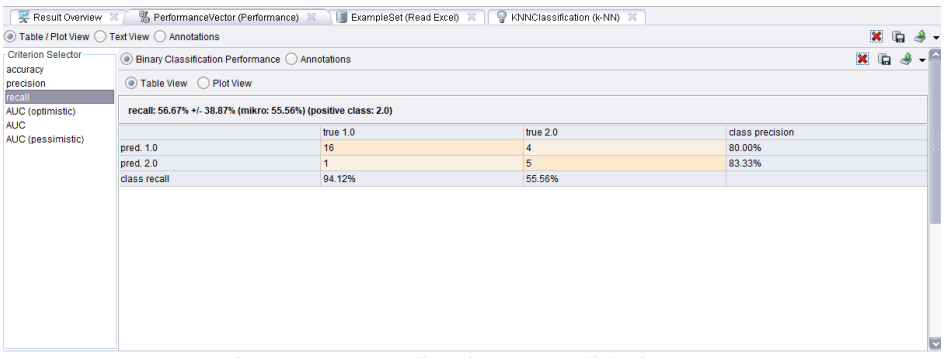

Figure 5. Recall value on validation = 5

And at validation = 10 the following results are obtained:

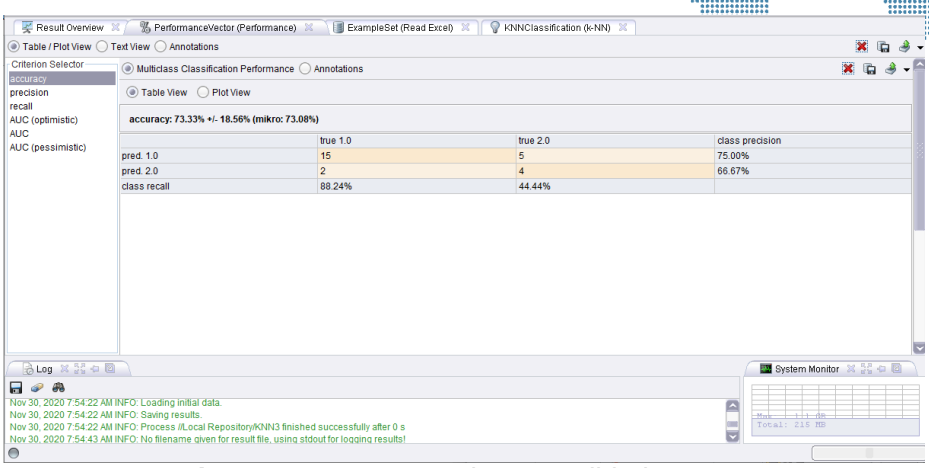

Figure 6. Accuracy value on validation = 10

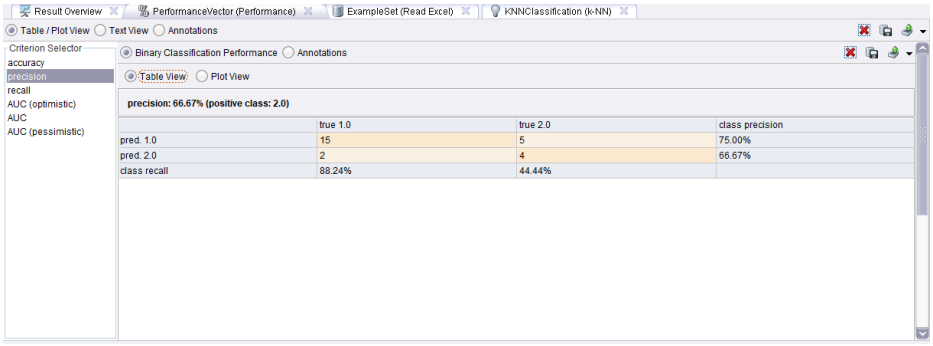

Figure 7. Validation chest precision value = 10

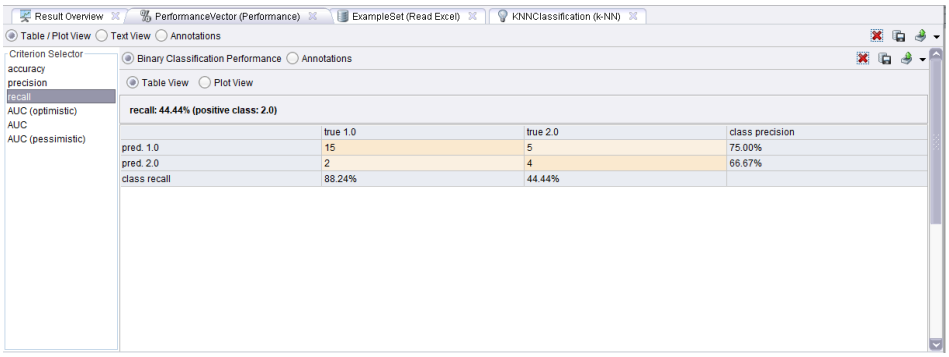

Figure 8. Recall value on validation = 10

3.2. KNN Algorithm Results

Table 5. K-NN calculation results

Ability	Appearance	Responsibility	Attitude and Behavior	Label	Distance	K=10
81,5	90	90	88	1	7,228416	1
85	96	91	89	1	8,660254	1
83,5	78	88	88	1	11,92686	1
79	87	87	77	2	14,3527	1
65	88	78	91	1	24,8998	
78	85	88	88	1	10,81665	1
77	67	83	78	2	26,85144	
66,5	78	85	77	2	26,61297	

Ability	Appearance	Responsibility	Attitude and Behavior	Label	Distance	K=10
91	66	88	85	1	23,4094	
71	65	84	86	1	29,63106	
68	78	67	76	2	32,63434	
78	77	88	78	2	18,57418	
79	80	78	87	1	15,6205	1
88	89	66	79	2	22,84732	
83	88	78	88	1	10,34408	1
84	78	87	86	1	11,87434	1
74	86	77	85	1	17,72005	
67	75	67	78	2	33,71943	
78	77	75	77	2	22,56103	
79	67	73	89	1	27,60435	
81	88	77	90	1	12,40967	1
78	79	88	88	1	14,17745	1
68	77	76	89	1	25,80698	
77	69	76	78	1	27,23968	
76	77	73	88	1	22	
68	66	76	87	2	32,41913	

From the calculation results, it is found that determining the test data group based on the majority label from the k nearest neighbors. Because the value of k = 10 then the 10 smallest distance is taken. By using the Nearest Neighbor category, which is the majority, it can be predicted that the results of the classification of satisfaction in the next data can be predicted. It was obtained from the results of the research that label 1 (Satisfied) appeared the most so that it could be determined that the classification results on student satisfaction with the services of LPMM employees were "PUAS" or students were satisfied with the services provided by LPPM employees. And here is a table of cross validation test results using the RapidMiner application.

Table 6. Cross Validation test results using RapidMiner

Validation	5	10
Accuracy	81,33%	73,33%
Precission	83,33%	66,67%
Recall	56,67%	44,44%

From the 26 existing conversion data, it can be seen that the results of testing data using cross validation are validation = 5 which has a more accurate value on the existing data.

4. Conclusion

The conclusion that can be drawn after calculating the classification of student satisfaction with LPPM employees is that the classification of student satisfaction with LPPM employees using the K-NN (Nearest Neighbor) Algorithm can be applied. From the data that has been obtained, it can be determined that the predicted results of student satisfaction for the next data are "PUAS" or students are satisfied with the services provided by LPPM employees. From the calculations that have been done, KNN is more suitable with large amounts of data or has a large amount of data so that it can obtain more accurate results.

References

[1] A. V. K, "Analisis Kepuasan Pelanggan Terhadap Kualitas Pelayanan Jasa Pos Express Di Pt. Pos Indonesia Cabang Surakarta," *Universitas Negeri Sebelas Maret*, p. 96, 2011.

[2] T. Mufizar, T. Nuraen, and A. Salama, "Sistem Pendukung Keputusan Dalam Penentuan Pertukaran Pelajar Di Sma Negeri 2 Tasikmalaya Dengan Metode

Analytical Hierarchy Process (Ahp)," *CogITo Smart Journal*, vol. 3, no. 1, p. 68, 2018, doi: 10.31154/cogito.v3i1.46.68-82.

[3] Nofriansyah, "Sistem Pendukung Keputusan Pemilihan Hotel Di Kota Palembang Dengan Metode Simple Additive Weighting (SAW)," *Jurnal Sistem Informasi (JSI)*, vol. 5, no. 1, pp. 546–565, 2013.

[4] C. F. Putri, "Pemilihan Supplier Bahan Baku Pengemas Dengan Metode AHP," *Widya Teknika*, vol. 20, no. 1, pp. 25–31, 2012.

[5] L. F. Rhozi, "Sistem Pendukung Keputusan Pemilihan Smartphone Android Menggunakan Metode Simple Additive Weighting (Saw)," 2016.

[6] M. R. Ritonga, S. Solikhun, M. R. Lubis, and A. P. Windarto, "Sistem Pakar Diagnosa Gejala Awal Penyakit Akibat Virus Pada Anak Berbasis Mobile Dengan Forward Chaining," *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, vol. 2, no. 2, pp. 140–145, 2018, doi: 10.30743/infotekjar.v2i2.298.

[7] A. Nurzahputra, M. A. Muslim, and M. Khusniati, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," vol. 16, no. 1, pp. 17–24, 2017.

[8] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," no. April, pp. 10–20, 2013.

[9] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center," vol. 3, no. 1, pp. 87–93, 2018.

[10] R. S. Suhartanto, C. Dewi, and L. Muflikhah, "Implementasi Jaringan Syaraf Tiruan Backpropagation untuk Mendiagnosis Penyakit Kulit pada Anak," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 1, no. 7, pp. 555–562, 2017.

[11] A. Wanto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation," vol. 2, pp. 37–44, 2019.

[12] A. P. Windarto *et al.*, "Model Arsitektur Neural Network Dengan Backpropogation Pada Prediksi Total Laba," vol. 05, no. 02, pp. 147–158, 2018.

[13] A. P. Windarto, M. R. Lubis, and L. R. Komprehensif, "Implementasi Jst Pada Prediksi Total Laba Rugi Komprehensif Implementation Of Neural Network In Predicting Total Comprehensive Income Of Conventional Commercial Banks Using," vol. 5, no. 4, pp. 411–418, 2018, doi: 10.25126/jtiik.201854767.

[14] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 9, no. 1, pp. 11–22, 2018, doi: 10.24176/simet.v9i1.1627.

[15] D. N. Batubara and A. P. Windarto, "Analisa Klasifikasi Data Mining Pada Tingkat Kepuasan Pengunjung Taman Hewan Pematang Siantar Dengan Algoritma," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 588–592, 2019, doi: 10.30865/komik.v3i1.1664.

[16] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 13, no. 1, p. 33, 2018, doi: 10.30872/jim.v13i1.629.

[17] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *Jurnal ULTIMATICS*, vol. 7, no. 2, pp. 98–104, 2016, doi: 10.31937/ti.v7i2.356.

[18] D. Kurniawan and A. Saputra, "Penerapan K-Nearest Neighbour dalam Penerimaan Peserta Didik dengan Sistem Zonasi," *Jurnal Sistem Informasi Bisnis*, vol. 9, no. 2, p. 212, 2019, doi: 10.21456/vol9iss2pp212-219.

[19] Y. R. Amalia, "Penerapan data mining untuk prediksi penjualan produk elektronik

terlaris menggunakan metode k-nearest neighbor," 2018.
[20] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Informatika*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
[21] R. P. Fitrianti, A. Kurniawati, and D. Agusten, "Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia," *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2019*, pp. 27–32, 2019.
[22] M. Fansyuri, "Analisa Algoritma Klasifikasi K-Nearest Neighbor Dalam (Study Kasus Pt. Trigatra Komunikatama) Analysis of K-Nearest Neighbor Classification Algorithm in Determining Accuracy Value on Customer Pleasure (Case Study on Pt . Trigatra Komunikatama)," *Humanika: Jurnal Ilmu Sosial, Pendidikan, dan Humaniora*, vol. 3, no. 1, pp. 29–33, 2020.
[23] T. C. Pratama, "Penerapan Metode K-Nearest Neighbour Dalam Menentukan Kelayakan Calon Nasabah Yang Layak Untuk Kredit Mobil (Studi Kasus : Pt . Astra International , Tbk-Toyota)," *JURIKOM (Jurnal Riset Komputer)*, vol. 5, no. 4, pp. 402–408, 2018.
[24] A. Rohman, "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa," *Neo Teknika*, vol. 1, no. 1, 2015, doi: 10.37760/neoteknika.v1i1.350.
[25] A. Bode, "K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika," *ILKOM Jurnal Ilmiah*, vol. 9, no. 2, pp. 188–195, 2017, doi: 10.33096/ilkom.v9i2.139.188-195.
[26] M. Sadli, F. Fajriana, W. Fuadi, E. Ermatita, and I. Pahendra, "Penerapan Model K-Nearest Neighbors Dalam Klasifikasi Kebutuhan Daya Listrik Untuk Masing-Masing Daerah Di Kota Lhokseumawe," *Jurnal ECOTIPE*, vol. 5, no. 2, pp. 11–18, 2018, doi: 10.33019/ecotipe.v5i2.646.