Terbit online pada laman web jurnal: **https://ejurnalunsam.id/index.php/jicom/**

JICOM
Jurnal Informatika dan Teknologi Komputer
Vol. 04 No. 01 (2023) 01 - 06 E-ISSN :2774-7115 P-ISSN: 2775-2089

Penerapan KNN Penentuan Pelanggan Baru PDAM dan Clustering K-Means Berdasarkan Wilayah

Mutammimul Ula [1], Muthmainnah[2], Ridha Maulana[3], Veri Ilhadi[4]
[1,2,3,4] Program Studi Sistem Informasi, fakultas, Universitas Malikussaleh
[1]mutammimul@unimal.ac.id*, [2]ridha.maulana170180084@unimal.ac.id, [2]muthmainnah@unimal.ac.id, [2]veri@unimal.ac.id

Abstract

The Regional Drinking Water Company (PDAM) Tirta Gunung Pase is a company that distributes and provides drinking water services for the entire region of Lhokseumawe and North Aceh. This research aims to find new customers by determining new pairs and grouping each area in determining clusters in each region. This makes it easier for leaders to view collections by customer ID and the number of household group usage. The research methodology is to collect data at PDAM Tirta Mon Pase, then look at the classification with the KNN model and group it with the K-Means method. A final result is a group of regular customers and non-permanent customers with the variable customer id, house class, and several occupants of the house, and the respective values are seen from UA with a value of 5, RB value of 4, RC value 3, RD value two and SB value 1 with test data using 15 samples. The results of testing the KNN model by entering the customer id 25768, occupants of houses in groups 5 and 4 are included in the class 2 classification. Meanwhile, the customer id is 41162 with the results of group 4 classification and class 1 occupants 3 class 1. Kmeans clustering results for households, businesses and industrial companies. K-Result, while the results of the K-means method contained three clusters which showed that cluster 1 was 46.07%, cluster 2 was 39.79%, and cluster 3 was 14.13%..

Keywords: Clustering, Classification, K-Means, KNN.

Abstrak

Perusahaan Daerah Air Minum (PDAM) Tirta Gunung Pase merupakan perusahaan yang mendistribusikan dan menyediakan layanan air minum untuk seluruh wilayah Lhokseumawe dan Aceh Utara. Tujuan penelitian untuk mengetahui pelanggan baru dalam menentukan pasang baru dan pengelompokkan tiap-tiap daerah dalam menentukan cluster pada setiap wilayah. Hal ini memudahkan pimpinan untuk melihat cluster berdasarkan ID pelanggan dan jumlah penggunaan kelompok rumah tangga. Metodologi penelitian adalah mengumpulkan data di PDAM Tirta Mon Pase kemudian melihat klasifikasi dengan model KNN dan melakukan pengelompokan dengan metode K-Means. Hasil akhir adalah kelompok pelanggan tetap dan pelanggan tidak tetap dengan variabel id pelanggan, kelas rumah, dan beberapa penghuni rumah, dan nilai masing-masing dilihat dari UA dengan nilai 5, nilai RB 4, RC nilai 3, nilai RD dua dan nilai SB 1 dengan data uji menggunakan 15 sampel. Hasil pengujian model KNN dengan memasukkan id pelanggan 25768, penghuni rumah golongan 5 dan 4 masuk dalam klasifikasi kelas 2. Sedangkan id pelanggan adalah 41162 dengan hasil klasifikasi kelompok 4 dan penghuni 3 kelas 1. Hasil kmeans clustering untuk rumah tangga, Usaha dan industri perusahaan. K-Result sedangkan hasil metode K-means terdapat tiga cluster yang menunjukkan bahwa cluster 1 sebesar 46,07%, cluster 2 sebesar 39,79%, dan cluster 3 sebesar 14,13%.

Kata kunci: Clustering, Klasifikasi, K-Means, KNN.

1. Pendahuluan

PDAM Tirta Gunung Pase merupakan perusahaan yang bergerak di bidang air minum (BUMD), permasalahan dalam penentuan status meteran yang ditetapkan oleh perusahaan. Skala penyediaan air bersih dalam kota semakin meningkat setiap tahunnya akibat pertumbuhan penduduk, kemajuan teknologi, dan peningkatan ekonomi masyarakat. sedangkan biaya air baku yang diolah dengan PDAM terus meningkat. [1]

Penentuan kriteria pelanggan baru berbeda dalam pengelompokan dan klasifikasi penentuan nilai, sehingga informasi pelanggan harus diklasifikasikan sesuai dengan tingkat nilai yang terkandung dalam range penilaian pelanggan [2][3]. Misalnya, nilai pelanggan dapat dikelompokkan ke dalam klasifikasi tingkat tinggi hingga tingkat kekurangan dan termasuk kelompok yang telah ditetapkan[4], [5].

Data mining mencari pola dan informasi untuk mengambil data dengan menggunakan teknik atau

Diterima Redaksi : 15-02-2023 | Selesai Revisi : 10-03-2023 | Diterbitkan Online : 01-04-2023

metode tertentu dalam membuat klasifikasi dan pengelompokan. [6][7]. Pengelompokkan dalam data mining digunakan menentukan area destinasi wisata di Asia Timur yang sering dikunjungi atau diminati pengunjung, pengelompokkan tersebut dibagi menjadi 2 yang terdiri dari kriteria 6, 7, 8, dan 10 atau dengan kata lain destinasi resort, area piknik , pantai, dan institusi. Keagamaan [8].

Teknik yang digunakan dalam data mining berbeda-beda, sehingga pemilihan metode yang tepat akan tergantung pada tujuan dan proses Knowledge Discovery dalam database (KDD) secara keseluruhan. Model tersebut diklasifikasikan dalam basis data Knowledge Discovery (KDD) dan dapat melihat korelasi antar setiap atribut dalam dataset. Rasio gain adalah bobot untuk setiap kualitas untuk menghasilkan dataset. Hasil nilai model KNN dengan nilai akurasi k dengan K = 3 dan K = 5 dengan discount 80%. Sehingga memudahkan perusahaan untuk mengklasifikasikan [9], [10].

Klasifikasi juga termasuk salah satu metode data mining dan dapat menentukan jenis pelanggan atau pelanggan baru. Data set dapat digunakan dalam kombinasi kategori dalam kumpulan data untuk menentukan pelanggan dan variasi atau kelas kumpulan data yang termasuk dalam suatu klasifikasi. Dalam klasifikasi juga mempunyai variabel target kategoris dalam menentukan jenis barang yang banyak dilihat pelanggan [11][12] . Sehingga perlunya monitoring dalam melihat klasifikasi dalam sebuah model dan dapat memprediksi [13].

Klasifikasi adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting. Klasifikasi *K-Nearest Neighbor* (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN.[12]

Metode KNN dapat digunakan untuk memilih input untuk model nonlinier dan juga untuk menyajikan data yang akurat untuk deret waktu dalam peramalan[14]. Ketika JST digunakan dalam klasifikasi, objek diklasifikasikan ke dalam kelas yang paling diinginkan dari k tetangga. Saat digunakan untuk regresi, nilai diberikan ke properti objek sebagai rata-rata nilai yang diketahui dari nilai k tetangga terdekatnya.[15][16].

Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru dengan jarak Eucledian [17].

Beirkut Langkah-langkah untuk melihat nilai algoritma K-Nearest Neighbor adalah sebagai berikut :

$$\sqrt{(a1-b1)^2+(a2-b2)^2+.+(an-bn)^2} \quad (1)$$

Dimana a = a1,a2,..., an, dan b = b1, b2, ..., bn mewakili n nilai atribut dari dua record.

Cluster yaitu pengelompokkan terhadap kelas, objek atau lainnya yang memiliki kemiripan satu sama lain dan yang tidak memiliki ketidakmiripan akan dimasukkan kedalam kelompok lain [16]. Model pengelompokan k-means banyak digunakan dalam pengelompokan, misalnya, klaster pelanggan—model klaster dari setiap situs milik satu set dalam satu set. Model clustering adalah salah satu metode pengelompokan data non-hierarkis yang dapat mempartisi data menjadi dua atau lebih kelompok. Metode clustering mirip dengan klasifikasi dalam pengumpulan data. Hanya saja kumpulan data yang diambil harus memiliki kesamaan tertentu dalam klasifikasi dan pengelompokan [18]. K-Means yaitu salah satu dari metode pengelompokan data non hierarki yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih [19], [20].

$$\frac{\text{Jumlah data}}{\text{jumlah class}+1} \quad (2)$$

Jumlah data : Jumlah data pada klasifikasi data
Jumlah class : Jumlah kelompok terbagi sangat tinggi, tinggi, normal, rendah, sangat rendah.

$$Vij = \frac{1}{Ni}\sum_{k=0}^{Ni} Xkj \quad (3)$$

Keterangan:
Vij : *centroid* rata-rata cluster ke-i untuk ke-j dan
Ni : jumlah anggota cluster ke-i
i, k : indeks dari cluster : j : indeks dari variabel

Selanjutnya menghitung jarak terdekat dengan centroid Jarak centroid yang digunakan adalah Euclidean Distance sebagai berikut :

$$De = \sqrt{(xi-si)^2+(yi-ti)^2} \quad (4)$$

Keterangan:
De : *Euclidean Distance* dan I : banyaknya objek
(x, y) :koordinat objek dan (s, t) :koordinat *centroid*

Confusion matriks adalah salah satu yang digunakan dalam klasifikasi.[21][22] Confusion bertujuan untuk menghitung ,accuracy, recall, precision dan F1 score dimana precision digunakan untuk mengevaluasi kemampuan sistem untuk meningkatkan peringkat yang paling relevan, Recall bertujuan untuk menghitung seberapa mampu sistem dalam menemukan yang relevan terhadap data, accuracy digunakan untuk menghitung data yang diduga benar dengan jumlah seluruh data, sedangkan error rate digunakan untuk menghitung data yang diduga salah dengan membandingkan seluruh jumlah data [23]

2. Metodelogi Penelitian

2.1. Pengumpulan Dataset

Dataset yang digunakan data pelanggan Titra mount pase Aceh utra berjumlah 190 data. Dimana terdiri dari

Diterima Redaksi : 15-02-2023 | Selesai Revisi : 10-03-2023 | Diterbitkan Online : 01-04-2023

id pelanggan, golongan, dan jumlah penghuni rumah. Seperti pada tabel berikut

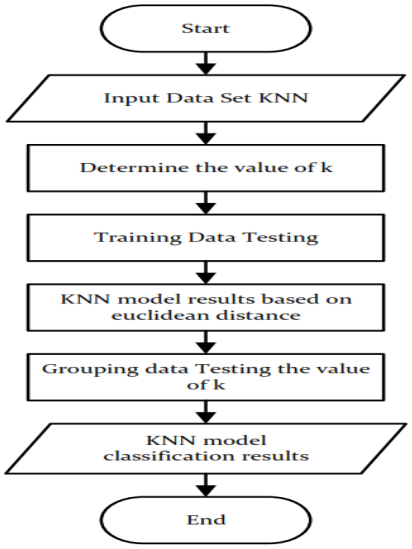
Tabel 1. Dataset PDAM

No	Id Pelanggan	Golongan	Jumlah penghuni rumah
1.	24036	RB- RT B	2
2.	11319	UA- USH B	5
3.	37746	RD-RT D	5
4.	32638	2B-RT B	2
5.	41143	RD-RT B	4
6.	38818	RB-RT B	3
7.	38819	UA-USH A	3
8.	28945	RB- RT B	2
9.	41666	RC-RT C	4
....	.....	....	...
190	21866	RC-TT C	3

Dari informasi dataset didapatkan data masih tidak terurut dari golongan maunpun pelanggannya. Dataset diatas merupakan gambaran untuk melakukan klasifikasi dengan menggunakan metode K-NN dan clustering dengan metode K-Means dengan dataset keseluruhan sebanyak 190 data.[24]

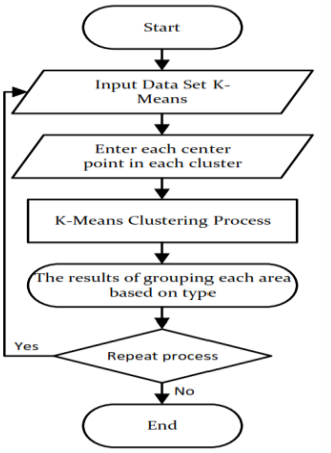
2.2. Metode Penelitian

Berikut ini skema sistem model KNN dalam kerja sistem atau alur penelitian untu melakukan klasifikasi pada metode K-NN seperti pada gambar berikut.



Gambar. 1. Skema sistem KNN

Dapat dilihat pada skema diatas, data setelah dikelompokkan menggunakan algoritma K-NN[25] selanjutnya akan diklasifikasi menggunakan Algoritma KNN. Dimulai dari penginputan data uji yang seterusnya akan dilakukan proses menentukan jumlah tetangga terdekat K, berikutnya proses menghitung jarak data uji ke data latih, proses pengurutan hasi jarak,dan menentukan kelompok data uji berdasarkan mayoritas nilai K. [26] maka melakukan pengujian klasifikasi pada data testing.[27]. Berikut merupakan skema sistem penerapan Clustering pada metode K-Means seperti pada gambar berikut.



Gambar. 2. Skema sistem K-Means

Proses skema sistem pada metode k-means dimana, Mulai: merupakan proses awal perancangan sitem. Dataset: merupakan kumpulan data yang akan digunakan dalam clustering. Selanjutnya kumpulan dataset akan digunakn untuk penginputan titik pusat clustering k-means Proses algoritma k-means clustering. Tampilan hasil clustering merupakan tampilan hasil dari proses algoritma k-means setelah proses selesai maka hasil akan di tampilkan.

3. Hasil dan Pembahasan

Perapan model knn dalam klasifikasi data pelanggan klasifikasi menggunakan algoritma knn pada penelitian ini bertujuan untuk mengklasifikasian dengan cara memprediksi kelas dari cluster yang tidak diketahui. Berikut tabel dataset pelanggan pdam tirta moun pase aceh utara.

Tabel 1. Dataset

No.	Id Pelanggan	Golongan	Jumlah Penghuni Rumah
1.	24036	4	2
2.	11319	5	5
3.	37746	2	5
4.	32638	4	2
5.	41143	2	4
6..	38818	4	3
7.	38819	5	3
8.	28945	4	2
9.	41666	3	4
10.	41673	4	5
11.	41673	1	9
12.	41728	4	4
13.	21384	5	5
14.	27606	2	3
15.	21866	3	3

Tabel diatas merupakan sampel untuk menentukaan klasifikasi golongan pembeli pdam titra moun pase dengan metode knn dan clustering dengan metode k-means.

Tabel 2. Data Uji KNN

No	id pelanggan	Golongan	jumlah penghuni rumah
1	25768	5	4
2	41162	4	4
3	22540	5	5
4	36427	3	3
5	37746	2	5

Diterima Redaksi : 15-02-2023 | Selesai Revisi : 10-03-2023 | Diterbitkan Online : 01-04-2023

Setelah menentukan cluster pada data,kemudian data akan diklasifikasikan menggunakan metode KNN. Tahap pada perhitungan menggunakan metode KNN ini adalah sebagai berikut:dengan nilai K secara random. Lalu menghitung jarak antar dua titik dengan menggunakan algoritma knn, dan Euclidean Distance sebagai tabel berikut.

Tabel 3. Hasil Data uji ke-1

No	Id pelanggan	Euclidian Distance
1	24036	2,236067977
2	11319	1
3	37746	3,16227766
4	32638	2,236067977
5	41143	3
6	38818	1,414213562
..	…	…
14	27606	3,16227766
15	21866	2,236067977

Dari hasil perhitungan data sampel dengan data uji, kemudian tiap data Memilih kedekatan terdekat. Pada penelitian ini menggunakan K=3 maka berdasarkan hasil klasifikasi dari tetangga terdekat.

Maka dapat dianalisis bahwa data uji pertama nendapatkan hasil prediksi yang masuk kedalam cluster 1, dimana data ujui kedua, tiga dan empat masuk kedalam cluster 2, dan data uji kelima mendapatkan hasil prediksi yang masuk ke dalam cluster 1. Dari hasil diatas dapat diambil kesimpulan bahwa data uji masuk kedalam klasifikasi yang ditunjukkan pada tabel berikut.

Tabel 4. Hasil Klasifikasi Metode KNN

No	Id pelangan	Golongan	Jumlah penghuni rumah	Prediksi
1.	25768	5	4	2
2.	41162	4	4	1
3.	22540	5	5	2
4.	36427	3	3	2
5.	37746	2	5	1

Perhitungan Metode K-Means

Algoritma K-means digunakan untuk menentukan Clustering pada dokumen yang ada. Pada penelitian ini akan menentukan pengelompokan berdasarkan golongan dan jumlah penghuni rumah data pelanggan PDAM tirta mount pase di aceh utara untuk menentukan golongan pelanggan PDAM. Untuk menentukan cluster menggunakan algoritma K-Means menggunakan tahap berikut ini.

Tabel 5. Golongan Pelanggan PDAM

keterangan	
UA = usaha rumah A	5
RB = rumah tangga B	4
RC = rumah tangga C	3
RD = Rumah Tangga D	2
SB = Sosial Khusus	1

Langkah selanjutnya ialah menginisalisasi data sampek pada kecamatan muara batu. Seperti pada tabel berikut.

Tabel 6. Inisiasi Data Sampel

id pelanggan	Golongan	jumlah penghuni rumah
24036	4	2
11319	5	5
37746	2	5
32638	4	2
41143	2	4
38818	4	3
38819	5	3
28945	4	2
41666	3	4
41673	4	5
41673	1	9
41728	4	4
21384	5	5
27606	2	3
21866	3	3

Langkah selanjutnya ialah menginisiasi pusat cluster dan menentukan jumlah centroid awal cluster yang dilakukan secara acak (random) seperti pada tabel berikut.

Tabel 7. Centroid Awal

Id Pelanggan	Cluster Pusat	Golongan	Jumlah Penghuni Rumah
37746	cluster 1	2	5
23370	cluster 2	3	4
24036	cluster 3	4	2

Langkah selanjutnya ialah menghitung nilai jarak terdekat (cost) dengan menggunakan persamaan Euclidian distance untuk melihat hasil cluster iterasi ke-1 pada data sampel seperti pada tabel berikut.

Tabel 8.Hasil Cluster Iterasi Ke-1

Rumah	C1	C2	C3	Minimum	cluster
3	3,605551	2,236068	1,414213562	1,414214	3
2	3,605551	2,236068	0	0	3
4	1,414214	0	2,236067977	0	2
5	2	1,414214	3	1,414214	2
9	4,123106	5,385165	7,615773106	4,123106	1
4	2,236068	1	2	1	2
5	3	2,236068	3,16227766	2,236068	2
3	2	1,414214	2,236067977	1,414214	2
3	2,236068	1	1,414213562	1	2

Langkah selanjutnya ialah melakukan proses iterasi kedua, ketiga, keempat dan kelima untuk menentukan cluster sampai hasil cluster yang akurat dan tepat.

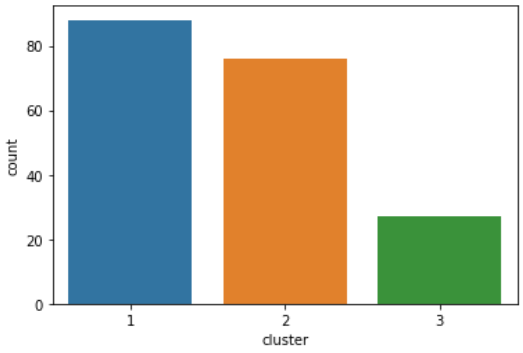
Diterima Redaksi : 15-02-2023 | Selesai Revisi : 10-03-2023 | Diterbitkan Online : 01-04-2023

Berikut merupakan implementasi hasil klasifikasi dan cluster menggunakan python.

	precision	recall	f1-score	support
1	1.00	0.95	0.97	19
2	0.80	1.00	0.89	4
3	1.00	1.00	1.00	16
accuracy			0.97	39
macro avg	0.93	0.98	0.95	39
weighted avg	0.98	0.97	0.98	39

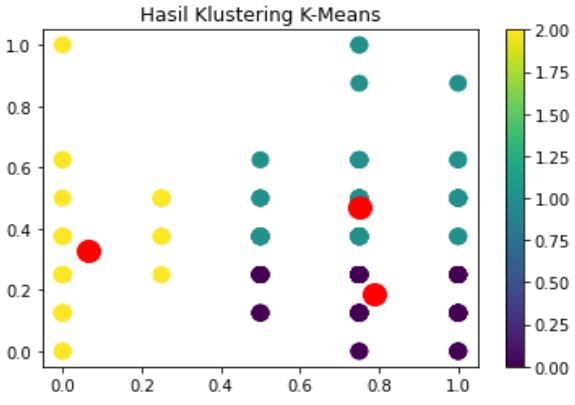
Gambar. 3. Hasil Klasifikasi

pada gambar diatas merupakan hasil klasikasi menggunakan metode K-NN untuk menentukan pelanggan yang memakai membeli pdam mount pase di aceh utara. Untuk melihat sebaran pemkainnya sebagai berikut.



Gambar. 3. Hasil kluster

Gambar 3 dilihat bahwa hasil cluster yang paling besar berada pada hasil cluster 1 yang berwarna biru, dan hasil cluster 2 yang berwarna coklat yang berati sedang dan hasil cluster 3 yang berwarna biru. Dimana dapat simpulkan bahwa pelanggan tetap dan pelanggan non tetap dapat dilihat bahwa pelanggan tetatp akan selalu membeli air minum di pdam mount pase dibandingkan dengan pelanggan non tetap. Dimana dapat dilihat pada grafik berikut.



Gambar 4. Hasil kluster

PDAM titra mount pase pelanggan yang selalu membutuhkan kebutuhan air. Hal ini dapat dilihat hari warna biru yang terus meninggkat pembelinya. Dikarenakan air merupakan kebutuhan yang sangat penting bagi masyarakat.

4. Kesimpulan

Penerapan model K-NN dalam klasifikasi pelanggan pdam tirta mount pase menghasilkan nilai akurasi sebesar 97%, recall 93%, presisi 98% dan f1-score 97%. Sehingga model KNN dapat mengklasifikasikan pelanggan dengan tingkat akurasi tinggi. Penerapan Model K-Means dalam menentukan pengelompokkan pelanggan pdam tirta mount pase dibagi menjadi 3 cluster, yaitu cluster 1 sebesar 46,073%, cluster 2 sebesar 39,79%, dan cluster 3 sebesar 14,13% sehingga efektifitas dari metode K-Means dalam pengelompokan data dalam menentukan kelompok pelanggan PDAM dapat dijadikan acuan dalam menentukan pelanggan PDAM berdasarkan kelompok sehingga pelayanan pelanggan menjadi lebih baik.

Daftar Rujukan

[1] Abdullah, D., Novita, C. Y., & Erliana, C. I. (2014). Perancangan Sistem Informasi Pengarsipan pada PDAM Tirta Mon Pase Aceh Utara. Sisfotenika, 4(2), 95-104..

[2] T. Imandasari and A. P. Windarto, "Sistem Pendukung Keputusan dalam Merekomendasikan Unit Terbaik di PDAM Tirta Lihou Menggunakan Metode Promethee," *J. Teknol. dan Sist. Komput.*, vol. 5, no. 4, pp. 159–165, Oct. 2017, doi: 10.14710/JTSISKOM.5.4.2017.159-165.

[3] Jadhav, S. D., & Channe, H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. International Journal of Science and Research (IJSR), 5(1), 1842-1845..

[4] A. P. Mutammimul Ula, Fajar Tri Tri Anjani, Ananda Faridhatul Ulva, Ilham Sahputra, "APPLICATION OF MACHINE LEARNING WITH THE BINARY DECISION TREE MODEL IN DETERMINING THE CLASSIFICATION OF DENTAL DISEASE," *J. Informatics Telecommun. Eng.*, vol. 6, no. 1, 2022.

[5] M. Ula, A. Faridhatul Ulva, M. Abdullah Ali, Y. Rilasmi Said, and S. Informasi, "APPLICATION OF MACHINE LEARNING IN DETERMINING THE CLASSIFICATION OF CHILDREN'S NUTRITION WITH DECISION TREE," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1457–1465, Sep. 2022, doi: 10.20884/1.JUTIF.2022.3.5.599.

[6] S. I. Nurhafida and F. Sembiring, "ANALISIS TEXT CLUSTERING MASYARAKAT DI TWITER MENGENAI MCDONALD'SXBTS MENGGUNAKAN ORANGE DATA MINING," *Pros. Semin. Nas. Sist. Inf. dan Manaj. Inform. Univ. Nusa Putra*, vol. 1, no. 01, pp. 28–35, Sep. 2021.

[7] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, Feb. 2022, doi: 10.34012/JURNALSISTEMINFORMASIDANILMUKOMPUTER.V5I2.2393.

[8] M. Pita, U. Sitompul, S. Sitompul, and Z. Situmorang, "Optimization of Determination Against K-Means Cluster Algorithm using Elbow Creation," *J. Tek. Inform. C.I.T Medicom*, vol. 14, no. 1, pp. 1–9, Mar. 2022, doi: 10.35335/CIT.VOL14.2022.176.PP1-9.

[9] K. Annisa, B. S. Ginting, and M. A. Syari, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat," *J. Ilmu Komput. dan Inform. Vol.*, vol. 6, no. 1, pp. 113–129, 2022.

[10] Siagian and Samsul Bahri, "Implementasi Gain Ratio Pada Metode KNN Dalam Memprediksi Penjualan Sparepart Electronic Di Service Center Panasonic Lhokseumawe," 2022.

[11] M. A. Rahman, N. Hidayat, and A. Afif Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM

Tirta Kencana Kabupaten Jombang),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Vol. 2, No. 12, Desember 2018, hlm. 6346-6353 e-ISSN*, vol. 2, no. 12, pp. 925–928, 2018.
[12] B. N. Azmi, A. Hermawan, and ..., “Analisis Pengaruh PCA Pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression,” *... (Jurnal Sist. dan ...*, vol. 7, no. 2, pp. 94–103, 2022.
[13] et al. Ula Mutammimul, “A New Model of The Student Attendance Monitoring System Using RFID Technology,” *J. Phys. Conf. Ser.*, vol. 1807, no. 1, 2021.
[14] T. Y. Prahudaya and A. Harjoko, “METODE KLASIFIKASI MUTU JAMBU BIJI MENGGUNAKAN KNN BERDASARKAN FITUR WARNA DAN TEKSTUR,” *J. Teknosains*, vol. 6, no. 2, pp. 113–123, Aug. 2017, doi: 10.22146/TEKNOSAINS.26972.
[15] S. S. A. Umri, “ANALISIS DAN KOMPARASI ALGORITMA KLASIFIKASI DALAM INDEKS PENCEMARAN UDARA DI DKI JAKARTA,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, Aug. 2021, doi: 10.33387/JIKO.V4I2.2871.
[16] S. Sarbaini, W. Saputri, Nazaruddin, and F. Muttakin, “Cluster Analysis Menggunakan Algoritma Fuzzy K-Means Untuk Tingkat Pengangguran Di Provinsi Riau,” *J. Teknol. dan Manaj. Ind. Terap.*, vol. 1, no. 2, pp. 78–84, Jun. 2022, doi: 10.55826/TMIT.V1III.30.
[17] M. Ula, A. Faridhatul Ulva, I. Saputra, I. Maulana, and A. Affiliations, “IMPLEMENTATION OF MACHINE LEARNING USING THE K-NEAREST NEIGHBOR CLASSIFICATION MODEL IN DIAGNOSING MALNUTRITION IN CHILDREN,” *MULTICA Sci. Technol. J.*, vol. 2, no. 1, pp. 94–99, Apr. 2022, doi: 10.47002/MST.V2I1.326.
[18] “Clustering of Clean Water Needs in Indonesia for the 2012-2017 Period Using the K-Means Algorithm | Saputri | SISFOTENIKA.”
[19] A. Pangestu and T. Ridwan, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS PENGELOMPOKAN PELANGGAN BERDASARKAN KUBIKASI AIR TERJUAL MENGGUNAKAN WEKA,” *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 12, no. 3, pp. 67–71, Aug. 2022, doi: 10.24853/JUSTIT.12.3.67-71.
[20] M. Ula, Khatib Sulaiman, A. Zulfikri, A. Faridhatul Ulva, and R. Achmad Rizal, “Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru,” *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 2023–336, Feb. 2023, doi: 10.33022/IJCS.V12I1.3162.
[21] A. Sumiah and N. Mirantika, “PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN NAIVE BAYES UNTUK REKOMENDASI PENENTUAN MAHASISWA PENERIMA BEASISWA PADA UNIVERSITAS KUNINGAN,” *Buffer Inform.*, vol. 6, no. 1, pp. 1–14, Jun. 2020, doi: 10.25134/BUFFER.V5I2.
[22] S. Busono, “Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa,” *J. Ilm. Teknol. Inf. Asia*, vol. 14, no. 1, pp. 31–40, Mar. 2020, doi: 10.32815/JITIKA.V14I1.400.
[23] H. Al Rasyid Harpizon *et al.*, “Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, 2022.
[24] K. Umam, S. Bin Lukman, M. Zulfahmi, and A. A. Nababan, “Perbandingan Rapid Centroid Estimation (RCE) — K Nearest Neighbor (K-NN) Dengan K Means — K Nearest Neighbor (K-NN),” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 1, pp. 79–89, Sep. 2017, doi: 10.30743/INFOTEKJAR.V2I1.166.
[25] S. T. Siska, “ANALISA DAN PENERAPAN DATA MINING UNTUK MENENTUKAN KUBIKASI AIR TERJUAL BERDASARKAN PENGELOMPOKAN PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING,” *J. Teknol. Inf. dan Pendidik.*, vol. 9, no. 1, pp. 86–93, Apr. 2016, doi: 10.24036/TIP.V9I1.55.
[26] N. Fadlia, R. Kosasih, and A. Kemacetan, “KLASIFIKASI JENIS KENDARAAN MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN),” *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 207–215, Feb. 2020, doi: 10.35760/TR.2019.V24I3.2397.
[27] R. Rachmat, Y. H. Chrisnanto, and F. R. Umbara, “Sistem Prediksi Mutu Air Di Perusahaan Daerah Air Minum Tirta Raharja Menggunakan K – Nearest Neighbors (K – NN),” *Pros. SISFOTEK*, vol. 4, no. 1, pp. 189–193, Aug. 2020.

Diterima Redaksi : 15-02-2023 | Selesai Revisi : 10-03-2023 | Diterbitkan Online : 01-04-2023