



Implementation of K-Nearest Neighbor Classification Algorithm as Decision Support Method in Staple Food Aid Distribution to Society

Implementasi Algoritma Klasifikasi K-Nearest Neighbor Sebagai Metode Pendukung Keputusan Dalam Distribusi Bantuan Sembako Pada Masyarakat

Alfiah^{1*}, Budi Prasetyo²

¹Information System, Faculty of Mathematics and Natural Science, State University of Semarang

²Information System, Faculty of Mathematics and Natural Science, State University of Semarang

E-Mail: ¹Alfiahts00@gmail.com, ²bprasetiyo@maill.unnes.ac.id

Makalah: Diterima 15 Maret 2024; Diperbaiki 15 Maret 2024; Disetujui 15 Maret 2024
Corresponding Author: Alfiah

Abstrak

Sembako adalah sembilan jenis kebutuhan pokok yang umumnya diperlukan oleh orang Indonesia untuk kebutuhan sehari-hari. Dengan pendistribusian sembako tersebut, dimaksudkan untuk memenuhi kebutuhan makanan pokok secara lebih terjangkau dan mudah, terutama bagi masyarakat yang mengalami kesulitan ekonomi. Namun, masih terjadi ketidaktepatan dalam pembagiannya, di mana orang yang seharusnya menerima bantuan tidak mendapatkannya, sementara yang tidak memenuhi kriteria justru menerima bantuan. Oleh karena itu, penting untuk melakukan pengelompokan penerima bantuan sembako berdasarkan data yang telah diproses. Metode data mining cocok digunakan untuk mengidentifikasi pola dalam data, karena dapat mengekstraksi informasi penting yang sebelumnya tidak diketahui. Hasil eksperimen menunjukkan bahwa algoritma K-NN memiliki nilai akurasi yang tinggi pada K=1 dan K=2 sebesar 81.08%, namun cenderung menurun dengan peningkatan K. Evaluasi model menunjukkan kinerja baik dalam mengenali kasus "Ya", tetapi perlu diperbaiki dalam mengenali kasus "Tidak". Dengan demikian, penggunaan algoritma K-NN dalam klasifikasi distribusi bantuan sembako dapat memberikan hasil yang memadai dengan pemilihan parameter yang tepat.

Kata Kunci: Sembako, Klasifikasi, K-Nearest Neighbor, Distribusi, Data Mining

Abstract

"Sembako" is a term for nine essential food or drink items generally needed by the Indonesian population for daily sustenance. The distribution of "sembako" is intended to make basic food more affordable and accessible, especially for economically disadvantaged communities. However, there are still issues with its distribution, where those who should receive assistance may not, while those who do not meet the criteria receive it. Therefore, it is important to categorize "sembako" recipients based on processed data. Data mining methods are suitable for identifying patterns in data, as they can extract important information that was previously unknown. Experimental results show that the K-NN algorithm has the highest accuracy at K=1 and K=2, reaching 81.08%, but tends to decrease with higher K values. Model evaluation indicates good performance in recognizing "Yes" cases but needs improvement in recognizing "No" cases. Thus, the use of the K-NN algorithm in classifying "sembako" distribution can provide adequate results with the right parameter selection. Keyword: Sembako, Classification, K-Nearest Neighbor, Distribution, Data Mining

1. PENDAHULUAN

Sembako adalah sembilan jenis kebutuhan pokok yang umumnya dibutuhkan oleh orang Indonesia untuk kebutuhan sehari-hari. Sembako biasanya terdiri dari gula, beras, minyak goreng, tepung terigu, garam, telur, daging sapi, daging ayam, dan susu. Sembako sendiri merupakan salah satu bentuk dari bantuan sosial yang merupakan program pemberian bantuan kepada masyarakat yang mengalami risiko sosial dengan menggunakan anggaran pemerintah. Bantuan sosial ini akan didistribusikan oleh kelurahan atau desa yang diberikan oleh pemerintah melalui Kementerian Sosial. Program bantuan sosial ini diharapkan dapat meningkatkan kesejahteraan

masyarakat dan meringankan pengeluaran yang diperlukan oleh keluarga yang miskin dan rentan untuk memenuhi kebutuhan pangannya. Oleh karena itu, sudah sepatutnya program ini diberikan kepada kelompok masyarakat yang benar-benar layak untuk menerimanya [1][2].

Namun, masih terdapat permasalahan dimana pendistribusian bantuan program sembako tidak tepat sasaran, dimana orang yang seharusnya menerima bantuan tidak mendapatkannya, sementara orang yang tidak memenuhi kriteria malah menerima bantuan. Beberapa penyebab diantaranya yaitu ketidakjelasan kriteria penerima manfaat dan ketidakmampuan dalam melakukan penyeleksian. Ketidakjelasan kriteria penerima manfaat dapat terjadi ketika aturan atau persyaratan untuk menerima bantuan tidak ditetapkan dengan jelas. Sehingga proses seleksi yang dilakukan secara tidak terstruktur atau tidak memiliki mekanisme yang jelas dapat memudahkan terjadinya manipulasi data, yang pada akhirnya dapat menyebabkan beberapa penerima yang sebenarnya tidak memenuhi syarat mendapatkan bantuan [3].

Oleh karena itu, untuk menghindari ketidakmerataan dalam distribusi bantuan sembako, penting untuk melakukan klasifikasi warga. Tujuannya adalah memastikan bahwa bantuan sosial hanya diberikan kepada yang membutuhkan dan tidak disalahgunakan oleh pihak yang tidak bertanggung jawab. Proses klasifikasi ini dapat dilakukan dengan menggunakan data mining, yang merupakan metode untuk mengidentifikasi dan mengambil pola-pola berguna dari data [4]. Pola-pola yang ditemukan dapat digunakan untuk memprediksi dalam pengambilan keputusan di masa mendatang. Salah satu metode yang berguna untuk menyelesaikan permasalahan dalam data mining adalah algoritma K-Nearest Neighbor (K-NN). Algoritma ini digunakan untuk mengklasifikasikan objek dalam data berdasarkan kedekatannya dengan objek baru atau tetangga terdekatnya [5].

Pada tahun 2021, Saifur Rohman Cholil dan rekan-rekannya melakukan penelitian tentang klasifikasi seleksi penerima beasiswa menggunakan algoritma K-Nearest Neighbor (KNN). Hasil perhitungan dari penelitian tersebut menunjukkan tingkat akurasi sebesar 90,5%, yang dievaluasi dengan metode confusion matrix [6]. Penelitian yang dilakukan oleh Sumarlin pada tahun 2015 tentang penggunaan algoritma KNN untuk membantu keputusan klasifikasi penerima beasiswa PPA dan BBM menunjukkan bahwa hasil presentasi rata-rata mencapai di atas 80%, dengan nilai K optimal adalah K=6 [5]. Penelitian lainnya yang dilakukan oleh Hazmira Yozza, dkk pada tahun 2023, mengenai klasifikasi program sembako di daerah Payobasung, Sumatera Barat, didapatkan hasil akurasi sebesar 97.14% dengan nilai K=3 [7]. Hal tersebut menandakan bahwasanya penggunaan algoritma K-Nearest Neighbor dapat dipergunakan dengan baik untuk proses klasifikasi.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode klasifikasi K-Nearest Neighbor (K-NN) sebagai pendekatan dalam memprediksi distribusi bantuan sembako pada masyarakat. Metode ini dipilih karena kemampuannya yang sederhana namun efektif dalam mengklasifikasikan data berdasarkan kedekatannya dengan data latih. Pendekatan ini diimplementasikan menggunakan perangkat lunak RapidMiner, yang menyediakan alat analisis data dan pemodelan prediktif yang dapat memfasilitasi proses eksplorasi data, pemrosesan data, serta pengujian dan evaluasi model. Pada penelitian ini, data yang digunakan adalah *data dummy* yang dihasilkan oleh ChatGPT untuk mensimulasikan distribusi bantuan sembako pada masyarakat dengan berbagai skenario dan kondisi yang mungkin terjadi. Perlu digaris bawahi bahwa *data dummy* yang digunakan dalam penelitian ini tidak merepresentasikan data aktual dari distribusi bantuan sembako di lapangan. Adapun beberapa tahapan dalam proses penelitian ini yaitu (1) Perencanaan; (2) Pengumpulan data; (3) Pre-processing (4) proses klasifikasi; (5) dan hasil klasifikasi [13].



Gambar 1. Tahapan Penelitian

2.1 Data Mining

Data mining adalah proses eksplorasi data yang menggunakan kecerdasan buatan dan statistika untuk menemukan pola atau informasi penting dari kumpulan data [8]. Namun, untuk mengidentifikasi pola tersebut memerlukan penggunaan metode atau algoritma tertentu. Hasil dari proses data mining ini dapat berupa opsi atau alternatif yang mendukung pengambilan keputusan [9].

2.2 Klasifikasi

Dalam ilmu data, klasifikasi adalah suatu proses yang dimaksudkan untuk mengelompokkan data ke dalam klasifikasi atau kelompok tertentu berdasarkan karakteristik atau fitur yang mereka miliki. Dengan melakukan klasifikasi, data dapat diorganisir secara terstruktur, memudahkan analisis, dan mendukung dalam pengambilan keputusan [10].

2.3 Algoritma K-Nearest Neighbor

Algoritma K-NN merupakan salah satu algoritma klasifikasi dalam *machine learning* yang menggunakan data pelatihan untuk mengklasifikasikan objek baru dengan setiap data pelatihan untuk menentukan kelas yang paling sesuai. Dalam pengklasifikasian tersebut diperlukan sebuah dataset yang terdiri dari *data training* dan *data testing* [11][12].

2.5 Evaluasi Model

Tahapan ini merupakan tahap evaluasi hasil akurasi dari metode K-Nearest Neighbor dengan menganalisis data training dan data testing, kemudian memvalidasi hasilnya menggunakan teknik *K-Fold Cross Validation* [14][15]. Untuk membuktikan kinerja algoritma digunakan persamaan sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (1)$$

Dan untuk hasil akurasi akan dihitung menggunakan persamaan sebagai berikut:

$$Akurasi = \frac{Jumlah\ klasifikasi\ benar}{Jumlah\ data\ uji} \times 100\% \quad (2)$$

3. HASIL DAN ANALISIS

3.1 Pengumpulan Data

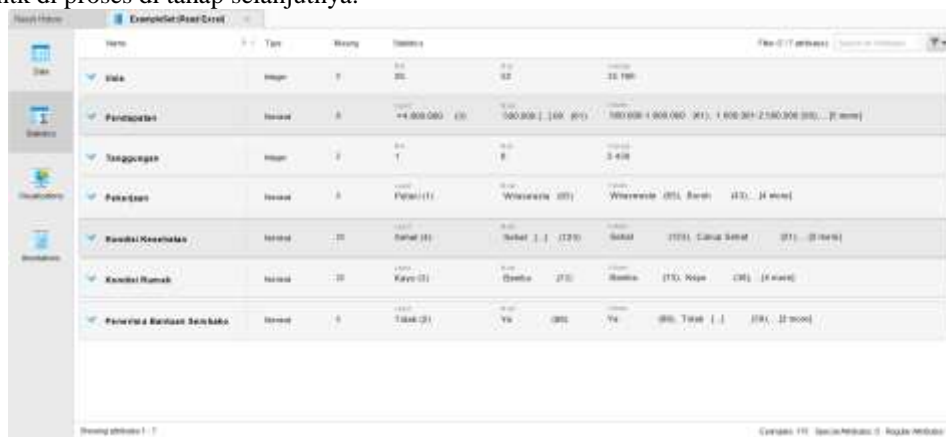
Untuk mendapatkan data yang akan diolah dan dianalisis, tahap pertama dalam proses penelitian adalah tahap pengumpulan data.. Data yang digunakan pada penelitian ini menggunakan data dummy yang dihasilkan oleh ChatGPT sebanyak 175 data. Data tersebut berisi atribut usia, pendapatan, tanggungan, pekerjaan, kondisi kesehatan, dan kondisi rumah. Adapun data yang digunakan dapat dilihat pada tabel berikut ini:

Tabel 1. Data Calon Penerima Bantuan Sembako

No	Usia	Pendapatan	Tanggungan	Pekerjaan	Kondisi Kesehatan	Kondisi Rumah
1	34	<500.000	2	Buruh	Sehat	Bambu
2	37	2.500.001-4.000.000	6	Wiraswasta	Sehat	Bambu
3	33	<500.000	1	Petani	Sehat	Bambu
4	38	1.000.001-2.500.000	4	Buruh	Sehat	Kayu
5	42	1.000.001-2.500.000	2	Wiraswasta	Sehat	Bata
6	31	500.000-1.000.000	2	Pedagang	Cukup Sehat	
7	39	500.000-1.000.000	3	Wiraswasta	Sehat	Bata
8	34	1.000.001-2.500.000	3	PNS	Sehat	Bambu
9	29	<500.000	1	Buruh		Kayu
10	46	500.000-1.000.000	2	Wiraswasta	Sehat	Bata
11	32	500.000-1.000.000	1	Petani	Cukup Sehat	Bambu
12	37	1.000.001-2.500.000	2	Buruh	Sehat	Kayu
13	43	1.000.001-2.500.000	3	Wiraswasta	Sehat	Bata
14	26	<500.000	2	Petani	Sehat	Bambu
15	29	500.000-1.000.000	3	Pedagang	Cukup Sehat	Bata
...
173	32	500.000-1.000.000	1	Petani	Sehat	Bambu
174	37	1.000.001-2.500.000	3	Buruh		Kayu
175	43	1.000.001-2.500.000	3	Wiraswasta	Sehat	Bata

3.2 Preprocessing

Setelah dilakukan pengumpulan data, dilakukan pre-processing data yang mana data perlu dilakukan pembersihan sebelum data tersebut diolah dan diproses. Data yang dihasilkan akan menjadi lebih ringkas dan relevan untk di proses di tahap selanjutnya.



Gambar 2. Hasil Statistik Data



Gambar 3. Data Cleaning untuk missing value

Setelah dilakukan tahap pembersihan (cleaning), jumlah data dalam dataset berkurang dari 175 menjadi 121 data. Dari total tersebut, terdapat 84 data untuk pelatihan (training) dan 37 data untuk pengujian (testing).

Tahap ini adalah tahap transformasi yang dimaksudkan untuk mengubah data menjadi format yang lebih sederhana agar sesuai dengan proses data mining. Hasil dari proses transformasi data tersebut adalah sebagai berikut:

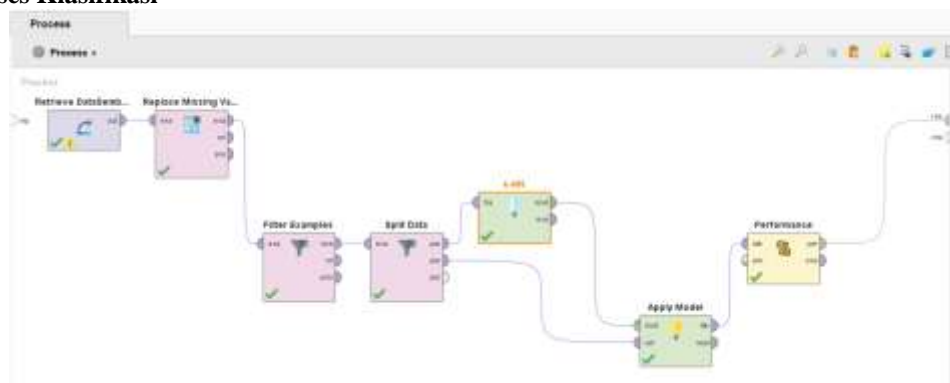
Tabel 2. Data Training

No	Usia	Pendapatan	Tanggungan	Pekerjaan	Kondisi Kesehatan	Kondisi Rumah
1	34	<500.000	2	Buruh	Sehat	Bambu
2	33	<500.000	1	Petani	Sehat	Bambu
3	38	1.000.001-2.500.000	4	Buruh	Sehat	Kayu
4	42	1.000.001-2.500.000	2	Wiraswasta	Sehat	Bata
5	46	500.000-1.000.000	2	Wiraswasta	Sehat	Bata
6	43	1.000.001-2.500.000	3	Wiraswasta	Sehat	Bata
7	26	<500.000	2	Petani	Sehat	Bambu
8	29	500.000-1.000.000	3	Pedagang	Cukup Sehat	Bata
9	41	500.000-1.000.000	1	Wiraswasta	Sehat	Bambu
10	36	2.500.001-4.000.000	5	PNS	Sehat	Kayu
..
84	39	1.000.001-2.500.000	4	Wiraswasta	Sehat	Bata

Tabel 3. Data Testing

No	Usia	Pendapatan	Tanggungan	Pekerjaan	Kondisi Kesehatan	Kondisi Rumah	Hasil
1	37	2.500.001-4.000.000	6	Wiraswasta	Sehat	Bambu	Ya
2	39	500.000-1.000.000	3	Wiraswasta	Sehat	Bata	Ya
3	34	1.000.001-2.500.000	3	PNS	Sehat	Bambu	Ya
4	32	500.000-1.000.000	1	Petani	Cukup Sehat	Bambu	Ya
5	37	1.000.001-2.500.000	2	Buruh	Sehat	Kayu	Tidak
6	52	2.500.001-4.000.000	2	Wiraswasta	Sehat	Bambu	Ya
7	33	500.000-1.000.000	1	Wiraswasta	Sehat	Bata	Tidak
8	35	500.000-1.000.000	3	Wiraswasta	Sehat	Bambu	Ya
9	45	500.000-1.000.000	2	Wiraswasta	Sehat	Bambu	Tidak
10	28	500.000-1.000.000	3	Buruh	Cukup Sehat	Kayu	Ya
..
37	33	<500.000	1	Petani	Sehat	Bambu	Tidak

3.3 Proses Klasifikasi



Gambar 4. Desain Model Pengujian Menggunakan Rapid Miner

Pada proses klasifikasi ini dilakukan pengimplementasian algoritma K-Nearest Neighbor menggunakan Tools RapidMiner. Implementasi algoritma K-NN akan dilakukan pada data testing yang berjumlah 37 data yang sudah memiliki label. Pada data tersebut dilakukan prediksi dan kemudian akan dilakukan klasifikasi akurasi

berdasarkan nilai K. Nilai K merupakan jarak terdekat untuk menentukan hasil prediksi. Dalam klasifikasi ini, percobaan dilakukan dengan menggunakan berbagai variasi nilai, diantaranya K=1 hingga K=5, dengan nilai K=1 dan K=2 memiliki nilai akurasi tertinggi.

Tabel 4. Nilai K dan Akurasi

Nilai K	Akurasi
K = 1	81.08%
K = 2	81.08%
K = 3	75.68%
K = 4	72.97%
K = 5	70.27%

Dengan nilai K=1, model mencapai tingkat akurasi sebesar 81.08%. Dalam hal ini, model berhasil mengidentifikasi 15 data sebagai True Positive, 1 data sebagai False Positive, 3 data sebagai True Negative, dan 10 data sebagai False Negative.

Tabel 5. Confusion Matrix K-NN K=1

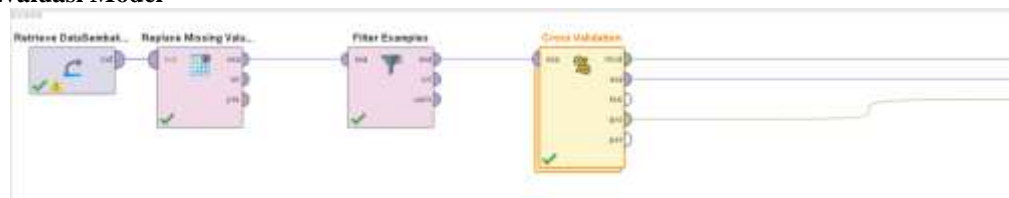
	Positive	Negative
Positive	15	1
Negative	10	3

Dengan nilai K=2, model mencapai tingkat akurasi sebesar 81.08%. Dalam hal ini, model berhasil mengidentifikasi 16 data sebagai True Positive, 2 data sebagai False Positive, 3 data sebagai True Negative, dan 9 data sebagai False Negative.

Tabel 6. Confusion Matrix K-NN K=2

	Positive	Negative
Positive	16	2
Negative	9	3

3.4 Evaluasi Model



Gambar 5. Desain model untuk Cross Validation

accuracy: 87.63% +/- 8.94% (micro average: 87.60%)

	true Ya	true Ya	true Tidak	true Tidak	class-precision
pred Ya	58	1	7	0	87.60%
pred Ya	2	18	0	2	81.82%
pred Tidak	3	0	30	0	90.91%
pred Tidak	0	0	0	0	0.00%
class recall	92.08%	94.74%	81.08%	0.00%	

Gambar 6. Hasil Cross Validation untuk keseluruhan data

Secara keseluruhan, Model ini memiliki akurasi sebesar **87.63%** dengan **deviasi** sekitar **8.94%** (micro average: 87.60%). Akurasi ini mengukur seberapa baik model dalam memprediksi secara keseluruhan. Secara keseluruhan, model ini memiliki kinerja yang baik dalam mengenali kasus “Ya”, tetapi perlu diperbaiki dalam mengenali kasus “Tidak”. Evaluasi ini membantu kita memahami seberapa baik model dalam mengklasifikasikan data dan mengidentifikasi area yang perlu diperbaiki.

4. KESIMPULAN

Klasifikasi dengan menggunakan algoritma K-Nearest Neighbor (K-NN) digunakan untuk memprediksi distribusi bantuan sembako pada masyarakat. Dengan menggunakan metode K-NN, penelitian ini berhasil mencapai tingkat akurasi tertinggi sebesar 81.08% dengan nilai $K=1$ dan $K=2$. Namun, nilai akurasi menurun dengan peningkatan nilai K , yang menandakan pentingnya pemilihan parameter yang tepat dalam algoritma K-NN. Hasil evaluasi menunjukkan bahwa model memiliki kinerja baik dalam mengenali kasus "Ya", tetapi perlu diperbaiki dalam mengenali kasus "Tidak". Dengan demikian, penggunaan algoritma K-NN dalam klasifikasi distribusi bantuan sembako memiliki potensi untuk meningkatkan akurasi dalam pemilihan penerima bantuan yang layak, namun perlu perhatian dalam penentuan parameter agar dapat mencapai kinerja optimal.

REFERENSI

- [1] Rahayu, N. A., Ginting, B. S., & Simanjuntak, M. (2021). SISTEM PENDUKUNG KEPUTUSAN SELEKSI PENERIMAAN BANTUAN PROGRAM SEMBAKO MENGGUNAKAN METODE SMART (SIMPLE MULTI ATTRIBUTE RATING TECHNIQUE) (STUDI KASUS: DINAS SOSIAL KOTA BINJAI). *Jurnal Sistem Informasi Kaputama*, 5(1).
- [2] Pahrudin, P., & Harianto, K. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2276>
- [3] Adhe, P. N. (2022, June 16). *Bansos Tidak Tepat Sasaran Adalah Maladministrasi*. Ombudsman Republik Indonesia. <https://ombudsman.go.id/perwakilan/news/r/pwkinternal--bansos-tidak-tepat-sasaran-adalah-maladministrasi>
- [4] Fitri, M., Chezy Vermila, dan W., & Pengajar Program Studi Agribisnis, S. (n.d.). *ANALISIS KEPUASAN KONSUMEN TERHADAP PRODUK SEMBAKO PADA MINIMARKET JUAN DI KECAMATAN BENAI KABUPATEN KUANTAN SINGINGI*.
- [5] Sumarlin. (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM Sumarlin STIKOM Uyelindo Kupang. *Jurnal Sistem Informasi Bisnis (1)*
- [6] S. R., Cholil, T., Handayani, R., Prathivi, & T. Ardianita (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- [7] YOZZA, H., AZIZAH, N. M., YULIANTI, L., & RAHMI, I. (2023). The Classification of "Program Sembako" recipients in Payobasung West Sumatra based on the K-nearest neighbors classifier. *Jurnal Natural*, 23(2), 83–91. <https://doi.org/10.24815/jn.v23i2.29738>
- [8] Mardi, Y. (n.d.). Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika* (2) 213–219.
- [9] F. Handayani (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. doi: 10.34010/jati.v12i1
- [10] I. Budiman and R. Ramadina (2015). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. 1–5
- [11] A'yuniyah, Q., & Reza, M. (2022). IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering*, 3(1), 39–45.
- [12] Fasnuri, H.A.D., Yuana, H., & Chulkamdi, M.T. (2022). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Klasifikasi Penyakit Diabetes Melitus Studi Kasus : Warga Desa Jatitengah. *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, 16(2). 133 – 142.
- [13] Priadana, S. M., & Sunarsi, D. (2021). *Metode Penelitian Kuantitatif*. Pascal Books.
- [14] Hafid, H. (2023). Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia. In *Journal of Mathematics* (Vol. 6, Issue 2). <http://www.ojs.unm.ac.id/jmathcos>
- [15] Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81–86. <https://doi.org/10.33096/ilkom.v12i2.507.81-86>