

Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Diagnosis Penyakit Alzheimer

*Implementation of K-Nearest Neighbor Algorithm for Classification of Alzheimer's
Disease Diagnosis*

Roudlotul Jannah Alfirdausy¹, Saiful Bahri²

¹Program Studi Matematika, UIN Sunan Ampel

²Program Studi Biologi, UIN Sunan Ampel

E-mail: ¹09030220048@student.uinsby.ac.id, ²saifulsi@uinsby.ac.id

Abstrak

Menurut World Alzheimer Report 2021, Alzheimer's Disease International (ADI) memperkirakan bahwa 75% penderita demensia tidak terdiagnosis secara global, dengan angka tersebut diyakini meningkat hingga 90% di beberapa negara berpenghasilan rendah dan menengah. Saat ini ada dua jenis demensia yaitu vaskular dan non-vaskular. Demensia vaskular, juga dikenal sebagai Alzheimer, adalah Perubahan yang terjadi pada tubuh yang disebabkan oleh penyumbatan pembuluh darah di otak dapat menyebabkan melemahnya ingatan manusia. K-Nearest Neighbor (KNN) adalah suatu teknik yang digunakan untuk mengelompokkan objek berdasarkan kesamaannya dengan objek terdekat. Penelitian ini bertujuan untuk menangani masalah efektivitas dan akurasi dalam mendeteksi penyakit Alzheimer dengan mengimplementasikan algoritma KNN untuk diagnosis penyakit Alzheimer. Berdasarkan analisis hasil klasifikasi dengan pembagian data testing dan training 70:30 menggunakan metode KNN dengan K=3, K=5 dan K=7 menghasilkan nilai akurasi sebesar 99% , sensitivitas sebesar 100%, serta spesifitas sebesar 96%.

Kata kunci: K-Nearest Neighbor, Penyakit Alzheimer

Abstract

As per the World Alzheimer Report 2021 by Alzheimer's Disease International (ADI), around 75% of people suffering from dementia are globally undiagnosed, and this number is expected to rise to 90% in certain low- and middle-income countries. Vascular and non-vascular are the two types of dementia currently known. Vascular dementia, which is also known as Alzheimer's, is a bodily change caused by the blockage of blood vessels in the brain that can result in a decline in human memory. K-Nearest Neighbor (KNN) is a method used to cluster objects based on their similarity to the nearest object. The objective of this research is to improve the effectiveness and accuracy of Alzheimer's disease diagnosis by applying the KNN algorithm. Based on the analysis of classification results by dividing the testing and training data 70:30 using the KNN method with K=3, K=5 dan K=7, the accuracy achieved is 99%, with sensitivity at 100% and specificity at 96%.

Keywords: K-Nearest Neighbor, Alzheimer Disease

1. PENDAHULUAN

Menurut Alzheimer's Disease International (ADI) berdasarkan World Alzheimer Report 2021, memperkirakan bahwa 75% penderita Demensia tidak terdiagnosis secara global, dengan angka tersebut diyakini meningkat hingga 90% di beberapa negara berpenghasilan rendah dan menengah. Diperkirakan penderita Demensia meningkat menjadi 139 juta jiwa pada tahun 2050, berdasarkan angka terbaru Organisasi Kesehatan Dunia (WHO) [1]

Menurut perkiraan WHO, pada tahun 2011 terdapat sekitar satu juta jiwa penderita Alzheimer di Indonesia. Keadaan tersebut terus berkembang seiring berjalannya waktu dan

harapan hidup orang Indonesia meningkat. Banyak orang tidak melaporkan kondisinya karena tidak tahu bahwa dirinya mengidap penyakit Alzheimer atau Demensia. Saat ini demensia terdiri dari demensia vaskular atau yang dikenal dengan alzheimer serta demensia non-vaskular. Alzheimer, adalah Perubahan yang terjadi pada tubuh yang disebabkan oleh penyumbatan pembuluh darah di otak, diperkirakan 75% demensia vaskuler (Alzheimer) disebabkan oleh stroke sumbatan. [2]

Metode klasifikasi merupakan pengelompokan objek berdasarkan ciri-ciri setiap objek. [3]. Klasifikasi adalah suatu proses di mana objek dari data diberikan kategori atau kelas tertentu yang telah ditentukan sebelumnya. Penggunaan klasifikasi secara luas dalam berbagai untuk mengelompokkan profil, mendeteksi penipuan, mendiagnosis kondisi medis, memprediksi penjualan, dan banyak lagi [4].

K-Nearest Neighbor (KNN) adalah metode atau suatu teknik yang digunakan untuk mengelompokkan objek berdasarkan kesamaannya dengan jarak objek terdekat pada data training. Metode ini memudahkan untuk menentukan klasifikasi atau prediksi suatu variabel dengan mencari objek terdekat.[5] KNN mengklasifikasikan dengan memperhatikan kesamaan antara objek dan kelasnya. Oleh karena itu, penggunaan metode KNN untuk klasifikasi memberikan akurasi yang lebih karena mempertimbangkan jarak antara objek dan kelasnya.[6]

Pada penelitian sebelumnya mengenai Klasifikasi Kanker Payudara Menggunakan Algoritma KNN [7], Penelitian ini digunakan untuk mengambil keputusan berdasarkan hasil yang didapat untuk menentukan kebijakan yang akan diambil dalam penanganan pasien kanker payudara. Hasilnya menunjukkan persentase yang cukup baik yaitu menghasilkan accuracy sebesar 72,62%.

Selain itu, penelitian sebelumnya mengenai Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes dan Correlation Based Feature Selection[8], Setelah dilakukan pemodelan dengan algoritma Naïve diperoleh nilai akurasi sebesar 93,83%.

Dalam penelitian ini, implementasi algoritma KNN untuk diagnosis penyakit Alzheimer dapat memberikan manfaat yang besar dalam bidang kesehatan, terutama dalam membantu dokter dan tenaga medis dalam mempercepat proses diagnosis penyakit Alzheimer. Dengan adanya model KNN, maka dapat dilakukan prediksi keberadaan penyakit Alzheimer.

2. METODE PENELITIAN

2.1 Klasifikasi

Data mining adalah studi tentang pola dalam sejumlah besar data dengan tujuan mengklasifikasikan, menilai, memprediksi, mencocokkan, dan mengelompokkannya. Pentingnya data mining terletak pada perannya dalam pencarian informasi basis data atau KDD (Knowledge Discovery in Databases), yang bertujuan untuk mengidentifikasi informasi berharga dari kumpulan data yang besar. Tugas data mining meliputi klasifikasi, evaluasi, prediksi, pengelompokan, dan pemetaan. [9]

Klasifikasi merupakan metode dalam data mining yang memisahkan data ke dalam kelompok atau kategori yang telah ditetapkan sebelumnya. Metode klasifikasi menggunakan pembelajaran terpantau, di mana data pelatihan yang sudah diberi label digunakan untuk menghasilkan aturan-aturan yang dapat mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan sebelumnya. [9]

Proses klasifikasi data biasanya melibatkan dua tahapan, yaitu tahap pembelajaran (training) di mana algoritma klasifikasi dibuat untuk menganalisis data pelatihan dan mewakili mereka dalam bentuk aturan klasifikasi, dan tahap klasifikasi di mana data uji digunakan untuk memprediksi akurasi dari aturan klasifikasi yang telah dibuat.[10]

2.2 K-Nearest Neighbor

Metode *K-Nearest Neighbor* atau K-NN merupakan metode pengklasifikasian yang bersifat global, dalam arti semua data di setiap kelas harus diperhitungkan dalam proses learning untuk menemukan suatu model (*clasifier*) menggunakan berbagai macam parameter[11]. Metode

K-NN bekerja dengan mencari jumlah k item data atau sampel (dari semua sampel latih) yang paling dekat dengan sampel masukan, kemudian memilih sampel berikutnya dengan jumlah sampel terbanyak diantara k sampel. Secara sederhana, K-NN mengurutkan sampel berdasarkan pemindaian. Penentuan k pola terdekat dilakukan berdasarkan jarak, *similarity* atau *disimilarity* tergantung jenis atributnya.[12]

Berikut adalah langkah-langkah untuk menghitung metode K-Nearest Neighbor (KNN)[13]:

1. Menentukan parameter K .
2. Menghitung jarak antara data training dan data testing: Gunakan rumus jarak Euclidean pada rumus (1) untuk menghitung jarak antara titik data pelatihan dan titik data pengujian. Jarak ini mengukur seberapa dekat atau jauhnya dua titik dalam ruang fitur.

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \quad (1)$$

Keterangan:

D = jarak kedua titik

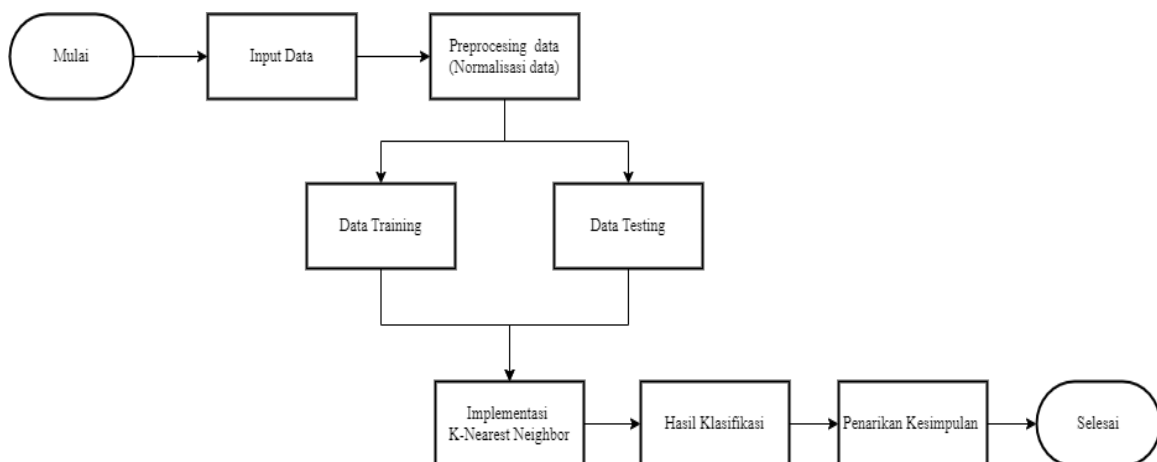
a = koordinat titik awal

b = koordinat titik akhir

3. Mengurutkan jarak yang terbentuk dari yang terkecil hingga terbesar.
4. Menentukan jarak terdekat dengan mengambil K jarak terdekat dari data pengujian yang telah diurutkan.
5. Memasangkan kelas yang bersesuaian: Identifikasi kelas dari data pelatihan yang sesuai dengan tetangga terdekat yang telah dipilih.
6. Mencari jumlah kelas dari tetangga terdekat: Hitung jumlah masing-masing kelas dari tetangga terdekat.
7. Tetapkan kelas yang paling umum sebagai kelas data yang akan dievaluasi

2.3 Prosedur Penelitian

Adapun tahapan dalam klasifikasi menggunakan algoritma KNN dengan langkah berdasarkan diagram alur berikut.



Gambar 1 Flowchart Klasifikasi Menggunakan K-Nearest Neighbor

Pada tahap preprocessing, dilakukan proses normalisasi data, digunakan metode StandardScaler. Metode ini berperan penting dalam menyesuaikan skala nilai dari seluruh atribut dalam data sehingga dapat disesuaikan dengan skala tertentu. [14]

Dalam penelitian ini, hasil klasifikasi dilakukan dengan menggunakan matriks kebingungan (*confusion matriks*) yang digunakan untuk menghitung akurasi, recall (sensitivity), dan Specificity. Matriks konfusi yang terdiri dari empat bagian sebagai true positive (TP), false positive (FP), true negative (TN) dan false negative (FN) digunakan untuk menghitung akurasi, recall (sensitivity), dan Specificity [15]. Perhitungan hasil klasifikasi sebagai berikut.

- a. Menghitung nilai akurasi:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2)$$

- b. Menghitung nilai sensitivitas:

$$\text{Sensitivitas} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

- c. Menghitung nilai Specifisitas:

$$\text{Specifisitas} = \frac{TN}{TN+FP} \times 100\% \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset Penelitian

Penelitian ini dilakukan dengan mengambil data dari website <https://www.kaggle.com/datasets/brsdincer/alzheimer-features>. Data terdiri 331 data penderita penyakit Alzheimer dengan 2 kelas yaitu kelas Alzheimer [1] sebanyak 146 data dan Normal [0] sebanyak 109 data dengan menggunakan 6 fitur diantaranya :

- a. Umur : Usia pasien pengidap penyakit Alzheimer
- b. MMSE : Mini Mental State Examination
- c. CDR : Clinical Dementia Rating
- d. rTIV : Estimated total intracranial volume
- e. nWBV : Normalize Whole Brain Volume
- f. ASF : Atlas Scaling Factor

Tabel 1 Data Penyakit Alzheimer

Umur	MMSE	CDR	eTIV	nWBV	ASF	Group
78	30	0.5	1315	0.707	1.335	1
61	30	1	1957	0.734	0.897	1
62	30	0.5	1928	0.731	0.91	1
68	30	0.5	1556	0.713	1.128	1
66	30	0.5	1446	0.78	1.214	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
71	16	1	1492	0.725	1.176	1
73	16	1	1478	0.696	1.188	1

78	15	2	1401	0.7	1.253	1
75	30	1	1169	0.742	1.501	1
77	30	1	1159	0.733	1.515	1

3.2 Statistik Deskriptif

Tahap pertama sebelum memulai proses klasifikasi data, analisis statistik deskriptif merupakan tahap awal yang penting untuk dilakukan guna mendapatkan gambaran yang jelas mengenai data yang akan digunakan. Analisis ini meliputi berbagai informasi tentang data. Dengan melakukan analisis statistik deskriptif, dapat memahami karakteristik data secara keseluruhan sehingga dapat digunakan untuk melakukan pengklasifikasian.

Tabel 2 Analisis Deskriptif Fitur Alzheimer

Variabel	n	Min	Mean	Max	Std. Deviasi
Umur	331	60	76,71	98	7.615
MMSE	331	4	27,21	30	3.814
CDR	331	0	0.2946	2	0.386
eTIV	331	1106	1491	2004	179.940
nWBV	331	0,6440	0,7302	0,8970	0.0373
ASF	331	0,876	1,194	1,587	0.141

3.3 Normalisasi Data

Sebelum menerapkan metode K-Nearest Neighbor untuk analisis klasifikasi, penting untuk melakukan normalisasi data. Normalisasi data menggunakan Standard Scaler atau dikenal dengan. Prosedur ini mengubah setiap nilai fitur menjadi z-score, di mana z-score adalah berapa standar deviasi jauhnya suatu nilai dari rata-rata fitur. Dengan menggunakan Standard Scaler, setiap fitur akan memiliki rata-rata sekitar 0 dan standar deviasi sekitar 1. Hasil dari normalisasi data dapat ditemukan pada Tabel 3.

Tabel 3 Normalisasi Data Penyakit Alzheimer

Umur	MMSE	CDR	eTIV	nWBV	ASF	Group
0.473684	1.0	0.25	0.232739	0.326425	0.645570	1
0.026316	1.0	0.50	0.947661	0.466321	0.029536	1
0.052632	1.0	0.25	0.915367	0.450777	0.047820	1
0.210526	1.0	0.25	0.501114	0.357513	0.354430	1
0.157895	1.0	0.25	0.378619	0.704663	0.475387	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.526316	0.615385	0.50	0.697105	0.212435	0.192686	1
0.578947	0.615385	0.50	0.693764	0.279793	0.195499	1
0.342105	0.576923	0.50	0.187082	0.435233	0.704641	1
0.578947	0.538462	0.25	0.398664	0.196891	0.454290	1
0.394737	0.538462	0.50	0.415367	0.067358	0.437412	1

3.4 Pengujian dan Analisis

Pada pengujian ini, dilakukan pengujian dengan data uji dengan skenario pengujian pada pembagian data menggunakan split rasio pembagian data 70:30, 25:25 dan 80:20 dengan berbagai parameter nilai k = 1, 2, 3, 4 dan 5. Setiap data uji diuji dengan tujuan untuk menguji apakah sistem dapat mendeteksi atau mendiagnosis penyakit sesuai dengan informasi yang terdapat pada data training.

Tabel 4: Akurasi Hasil Pengujian

Split Ratio	Parameter (K)	Akurasi
70:30	2	98%
	3	99%
	5	99%
	7	99%
75:25	2	97.6%
	3	98.8%
	5	98.8%
	7	98.8%
80:20	2	97%
	3	98.2%
	5	98.2%
	7	98.2%

Akurasi tertinggi yaitu 99% didapatkan pada parameter K=3, K=5 dan K=7 dengan split rasio 70:30. Dengan Confussion Matrix berikut

Tabel 5: Confussion Matrix Hasil Pengujian

Aktual	Prediksi	
	Positif	Negatif
Positif	54 (TP)	1 (FN)
Negatif	0 (FP)	46 (TN)

Berdasarkan tabel 5, dapat dihitung hasil evaluasi model K-NN menggunakan rumus (2), (3), dan (4) sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \\
 &= \frac{54 + 46}{54 + 1 + 0 + 46} \times 100\% \\
 &= \mathbf{99\%}
 \end{aligned}$$

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \times 100\%$$

$$\begin{aligned} &= \frac{54}{54 + 1} \times 100\% \\ &= \mathbf{98.2\%} \end{aligned}$$

$$\begin{aligned} \text{Spesifisitas} &= \frac{TN}{TN + FP} \times 100\% \\ &= \frac{46}{46 + 0} \times 100\% \\ &= \mathbf{100\%} \end{aligned}$$

Berdasarkan hasil penelitian sebelumnya menggunakan algoritma *Naïve Bayes*[8] menghasilkan tingkat akurasi sebesar 93,83% dalam deteksi Penyakit Alzheimer. Namun, dengan data yang sama juga mengungkapkan bahwa pendekatan klasifikasi menggunakan metode K-Nearest Neighbor (KNN) dengan variasi nilai K yaitu K=3, K=5, dan K=7, menghasilkan tingkat akurasi yang lebih tinggi, mencapai 99%. Oleh karena itu, pendekatan KNN dengan variasi K ini dapat dianggap sebagai metode yang lebih unggul dalam mendeteksi penyakit Alzheimer berdasarkan data yang dianalisis, dibandingkan dengan pendekatan *Naïve Bayes*.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian implementasi algoritma KNN untuk diagnosis penyakit Alzheimer didapatkan kesimpulan sebagai berikut:

1. Penerapan metode K-Nearest Neighbor untuk diagnosis penyakit Alzheimer didapatkan hasil klasifikasi yang baik. Dimana Implementasi menggunakan metode K-Nearest Neighbor untuk Diagnosis Penyakit Alzheimer dengan pembagian data testing dan training 70:20 terdapat 230 data training dan 101 data testing.
2. Berdasarkan hasil evaluasi confusion matrix diketahui bahwa untuk mendiagnosa penyakit alzheimer menggunakan metode K-Nearest Neighbor dengan K=3, K=5 dan K=7 menghasilkan nilai akurasi sebesar 99% dimana dari 101 data uji terdapat hanya 1 data yang tidak sesuai, sensitivitas sebesar 98%, serta spesifitas sebesar 100%.
3. Pendekatan KNN dengan variasi nilai K dapat dianggap sebagai metode yang lebih unggul dalam mendeteksi penyakit Alzheimer berdasarkan data yang dianalisis.

DAFTAR PUSTAKA

- [1] R.-N. P. Gauthier S, Webster C, Servaes S, Morais JA, *World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support*. 2022.
- [2] Y. E. Muliatie, N. Jannah, and S. Suprpti, "Pencegahan Demensia/Alzheimer Di Desa Prigen, Kecamatan Prigen, Kabupaten Pasuruan," *Pros. Konf. Nas. Pengabd. Kpd. Masy. dan Corp. Soc. Responsib.*, vol. 4, pp. 379–387, 2021, doi: 10.37695/pkmsr.v4i0.1308.
- [3] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, p. 134, 2018.
- [4] D. Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Informatika, 2019.
- [5] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
- [6] Y. I. Claudy, R. S. Perdana, and M. A. Fauzi, "Klasifikasi Dokumen Twitter Untuk

- Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (knn),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2761–2765, 2018, [Online]. Available: <https://www.researchgate.net/publication/322959490>
- [7] N. Meilani and O. Nurdiawan, “Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor,” *J. Wahana Inform.*, vol. 2, no. 1, pp. 177–187, 2023.
- [8] S. K. Wildah, S. Agustiani, M. R. R. S, W. Gata, and H. M. Nawawi, “Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes Dan Correlation Based Feature Selection,” *J. Inform.*, vol. 7, no. 2, pp. 166–173, 2020, doi: 10.31294/ji.v7i2.8226.
- [9] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [10] F. Harahap, “Penerapan data Mining dalam Pemilihan Mobil menggunakan Algoritma C4.5,” *Ijccs*, no. x, pp. 11–20, 2018, [Online]. Available: <https://voi.stmik-tasikmalaya.ac.id/index.php/voi/article/viewFile/99/42>
- [11] W. Yustanti, “Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah,” *J. Mat. Stat. dan komputasi*, vol. 9, no. 1, pp. 57–68, 2012.
- [12] A. Abdullah and K. Sandi, “SISTEM PREDIKSI RASA BUAH JERUK MENGGUNAKAN METODE k-NEAREST NEIGHBOR,” *Method. J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 2, pp. 7–13, 2021, doi: 10.46880/mtk.v7i2.457.
- [13] A. R. D.N, K. Auliasari, and Y. A. Pranoto, “Implementasi Metode K-Nearest Neighbor (Knn) Untuk Seleksi Calon Karyawan Baru (Studi Kasus : Bfi Finance Surabaya),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 14–20, 2020.
- [14] M. R. Andryan, M. Fajri, and N. Sulistyowati, “Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (Svm) Untuk Diagnosis Penyakit Kanker Payudara,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, p. 1, 2022, doi: 10.26798/jiko.v6i1.500.
- [15] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.