Print ISSN: 2461-1409
Journal Sensi Online ISSN: 2655-5298

Research Article Open Access (CC–BY-SA)

Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Mahasiswa Berpotensi Dropout

Nurwati[*1], Nur Azizah[2],Yudi Santoso[3]
[1,3]Program Studi Sistem Informasi Fakultas Teknologi Informasi, Universitas Budi Luhur Jakarta
[2]Program Studi Sistem Informasi Fakultas Sain dan Teknologi, Universitas Raharja Tangerang
Email: [*1] nurwati@budiluhur.ac.id ,[2]nur.azizah@raharja.info, [3] yudi.santoso@budiluhur.ac.id

Abstrak

Peramalan atau prediksi mahasiswa berpotensi dropout digunakan untuk memonitor jumlah mahasiswa aktif agar perkuliahan lancar dan lulus tepat waktu. Prediksi menggunakan algoritma K-NN digunakan karena salah satu kelebihannya yaitu tangguh terhadap training data yang noise dan efektif apabila data latih nya besar. Setelah didapat hasil dari proses K-NN lalu dilakukan pengujian menggunakan confusionmatrix menghasilkan nilai akurasi 0,83. Nilai presisi 1 dan nilai recall 0,78.

Kata kunci–dropout, K-NN, prediksi

Abstract

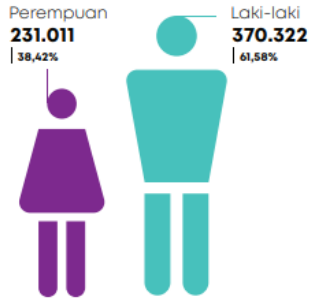
Forecasting or prediction of potential dropout students is used to monitor the number of active students so that lectures runs moothly and graduate on time. Prediction using the K-NN algorithm is used because one of its advantages is that it is tough against noisy training data and is effective when the training data is large. After obtaining the results from the K-NN process, testing was carried out using the confusion matrixt oproduce an accuracy valueof 0,83. The precision value is 1 and there call valueis 0,78.

Keywords*–dropout, K-NN, prediction*

1. PENDAHULUAN

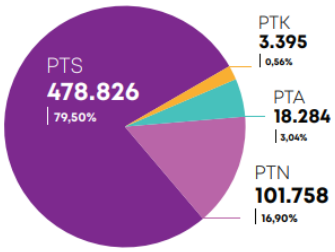
Melanjutkan ke perguruan tinggi merupakah impian dari sebagaian besar siswa setelah lulus dari Sekolah Menengah Atas. Namun banyak alasan yang membuat mereka/siswa tidak melanjutkan. Bagi mereka yang melanjutkan ke perguruan tinggi baik yang diterima di Perguruan Tinggi Negeri, Perguruan Tinggi Kedinasan, Perguruan Tinggi Agama dan Perguruan Tinggi Swasta bukan hal yang mudah untuk dapat menuntaskan kuliahnya, masih perlu ketekunan, usaha dan doa agar selesai kuliah tepat waktu.

Tidak lulus dalam kuliah atau sering dikenal dengan drop out/DO salah satu momok yang ditakuti mahasiswa.Mahasiswa yang dinyatakan drop out jika tidak memenuhi persyaratan akademik untuk kelulusan, kegiatan ini dilakukan setiap semester, agar kelancaran studi mahasiswa bisa dideteksi sejak dini[1]. Sehingga fakultas sejak semester 5 (lima) sudah menyeleksi mahasiswa yang sudah mencapai batas minimal SKS yang terkumpul. Beberapa ketentuan mahasiswa dinyatakan DO diantaranya[2] mahasiswa yang tidak mencapai batas minimal SKS dianggap mengundurkan diri atau dikenakan penghentian studi, mahasiswa yang tidak mendaftar ulang sebagai mahasiswa paling lama dua tahun akademik berturut-turut tanpa pemberitahuan tertulis dianggap DO, mahasiswa yang melakukan tindak pidana otomatis dianggap mengundurkan diri atau dikenakan penghentian studi.



Gambar 1 Jumlah angka putus kuliah berdasarkan jenis kelamin diambil dari statistik pendidikan tinggi 2020[3]

Gambar 1 menjelaskan sebanyak 38,42% jenis kelamin perempuan putus kuliah pada tahun 2020. Laki-laki sebanyak 79,50% yang putus kuliah di tahun 2020.



Gambar 2 Jumlah angka putus kuliah berdasarkan kelompok Pembinadiambil dari statistik pendidikan tinggi 2020[3]

Dari gambar 2 menjelaskan bahwa dari pengelompokan angka putus kuliah berdasarkan kelompok Pembina sebanyak 0,56% dari PTK (Perguruan Tinggi Kedinasan), dari PTA (Perguruan Tinggi Agama) sebanyak 3.04%, dari PTN(Perguruan Tinggi Negeri) sebanyak 16.90% dan PTS (Perguruan Tinggi Swasta) menyumbang 79,50%.

Berdasarkan gambar 1 dan 2 tersebut maka bagaimana memprediksi mahasiswa yang berpotensi tidak melanjutkan kuliah / dropout sehingga mampu memonitoring mahasiswa aktif dan memotivasi mahasiswa yang berpotensi dropout sehingga selesai kuliah dengan tepat waktu.

Prediksi mahasiswa drop out pada penelitian ini menggunakan K-NearestNeighbor (K-NN). Metode k-NearestNeighbor(k-NN) merupakanmetode dari algoritma supervised yang mana hasil proses dari queryinstance yang terklasifikasi berdasarkan kebanyakan dari kelas label. Algoritma k-NN diproses dengan jarak terpendek dari queryinstance ke data training sampai mendapatkan nilai k-NN-nya. Untuk menghitung jarak dengan tetangga (neighbor) digunakan rumus EuclidianDistance[1]. Algoritma KNN bersifat sederhana, bekerja dengan berdasarkan pada jarak terpendek dari sampel uji (testing sample) ke sampel latih (trainingsample) untuk menentukan KNN nya. Setelah mengumpulkan KNN, kemudian diambil mayoritas dari KNN untuk dijadikan prediksi dari sample uji[4].

Dikutip dari (Rini Nuraini Sukmana, dkk 2020) [4]KNN memiliki beberapa kelebihan yaitu tangguh terhadap training data yang noise dan efektif apabila data latih nya besar. Pada fase training, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi data trainingsample. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk testing data atau yang klasifikasinya tidak diketahui. Jarak dari vektor baru yang ini terhadap seluruh vektor trainingsample dihitung dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

Beberapa ulasan mengenai penelitian prediksi menggunakan algoritma KNN yang digunakan untuk prediksi produk kemasan skincara yang paling diminati[5], pemantuan dan evaluasi kelulusan mahasiswa [6], prediksi mahasiswa berpotensial berhenti kuliah [7], prediksi kelulusan siswa SMK [8].

Tabel 1 Penelitian terkait prediksi menggunakan metode K-NearestNeighbor

Paper	Tujuan penelitian	Hasil penelitian
1	Memprediksi produk kemasan skincare yang paling banyak diminati sehingga berguna untukmempermudah pihak perusahaan dalam perencanaan penyediaan stok produk[5].	Hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma K-NearestNeighbor, terdiri dari atribut Kategori Produk, Kuantitas dan Bulan. Didapatkan hasil prediksi penjualan tertinggi pada produk kemasan skincare dengan 7 kategori produk yaitu Lipgloss Tube pada bulan Juli & Maret), CreamBottle (Februari), Essential Oil Bottle (Oktober), SprayBottle (September), Powder Box (Januari), PumpBottle (Agustus & Oktober) dan Tube (Agustus & Desember). Hasil pengujian perhitungan akurasi menggunakan Rapidminer untuk mengetahui penjualan beberapa bulan mendatang diperoleh hasil nilai akurasi 80%.
2	Pemantauan dan evaluasi terhadap kelulusan mahasiswa dengan menggunakan klasifikasi data mining[6].	Memprediksi kelulusan mahasiswa dengan menggunakan algoritma klasifikasi data mining K-NearestNeighbor dengan mengklaster data k=1, k=2, k=3, k=4, dan k=5. Hasil yang diperoleh dengan cluster data k=5 accuracy adalah 85,15% dan nilai AUC adalah 0.888 adalah akurasi paling tinggi.

3	Prediksi mahasiswa yang berpotensi berhenti kuliah secara sepihak dengan melihat beberapa kriteria dan menggali informasi terhadap data mahasiswa yang berpotensi untuk berhenti kuliah dengan menerapkan algoritma K-NN[7].	Algoritma K-NN merekam data lama dan melihat kemiripan terhadap data baru dalam upaya pengenalan pola mahasiswa berhenti kuliah, hasil yang didapatkan dari data kuliah baru menunjukan kemiripan data dengan data lama mahasiswa yang berhenti kuliah dengan kemiripan nilai terdekat dari kasus lainnya yaitu 17,3815 dengan 19,98875 sehingga hasil yang didapatkan keputusan mahasiswa data baru tersebut memutuskan kemungkinan berhenti kuliah.
4	Metode K- NearestNeighbor untuk melakukan prediksi kelulusan siswa pada SMK Swasta Anak Bangsa[8].	Hasil penelitian yang diperoleh adalah nilai K=5 dengan tingkat akurasi sebesar 93,55 % yang ditetapkan sebagai K-Optimal. Nilai K=5 diterapkan pada algoritma K-NearestNeighbor untuk prediksi kelulusan siswa berdasarkan kehadiran, sikap dan nilai pengetahuan.

2. METODE PENELITIAN

Dalam penelitian ini diawali dengan langkah pertama yaitu identifikasi masalah lalu pengumpulan data,pencarian literatur dengan tinjauan pustaka, dilanjutkan dengan langkah persiapan dan pemilihan data,selanjutnya lakukan proses K-NN dan terakhir hasil dan pembahasan lalu membuat kesimpulan dan saran.

Gambar 3. Langkah Penelitian

Langkah penelitian pada gambar 3 penjelasannya yaitu dimulai dengan langkah identifikasi masalah. Identifikasi masalah dibuat untuk menentukan ruang lingkup masalah, merumuskan masalah dan mencari solusi dari masalah yang terjadi.Selanjutnya langkah pengumpulan data. Data yang digunakan adalah data Indeks Prestasi mahasiswa mulai semester 1(satu) hingga semester 5 (lima) dimulai dari semester Ganjil tahun 2019/2020 sampai dengan semester Ganjil tahun 2021/2022. Langkah penelitian selanjutnya melakukan tinjauan pustaka dengan mengumpulkan artikel/jurnal/penelitian berkaitan dengan topik penelitian. Dilanjutkan dengan persiapan dan pemilihan data untuk memproses prediksi mahasiswa yang dropout menggunakan K-NN. Setelah itu dilakukan proses K-NN dengan nilai K=5. Langkah berikutnya menuliskan kembali hasil dari proses K-NN dan pembahasannya kemudian melakukan kesimpulan dan saran.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Data yang digunakan adalah data Indeks Prestasi mahasiswa mulai semester 1(satu) hingga semester 5 (lima) dimulai dari semester Ganjil tahun 2019/2020 sampai dengan semester Ganjil tahun 2021/2022.

Untuk menghitung jarak dengan tetangga (neighbor) digunakan rumus EuclidianDistance[1].

$$di = \sqrt{\sum_{i=1}^{p}(X_{2i} - X_{1i})^2}$$

Gambar 4 Rumus EuclideanDistance[1]

Dimana; x1=sampel data
x2=data uji atau data testing
i=variabel data
d= jarak
p=dimensi data

3.2. Pembahasan

Penelitian ini mencari nilai euclidean dari keterkaitan level kelulusan mahasiswa dengan menggunakan data mahasiswa yaitu nilai indeks prestasi dari semester pertama sampai semesterlima kemudian dilanjutkan pengujian dari data testing menggunakan tehnikconfusionmatrixsehingga hasil pengolahan data dapat terlihat dengan memanfaatkan nilaiindeks prestasi di semester ke berapa dan nilai k yang dianggap terbaik yang nantinya akan menghasilkantingkat pengujian terbaik.

3.2.1. Persiapan dan pemilihan data

Data penelitian ini diambil dari data akademik di salah satu Universitas swasta Tangerang, Bantendari Fakultas Ekonomi dan Bisnis . Data ini terdiri dari data trainingdan data testing (tabel 2).

Tabel 2 Tabel data mahasiswa angkatan 2019

No	NIM	Sem1	Sem2	Sem3	Sem4	Sem5
1	1981528151	0	0	0	0	0

2	1981526345	2,94	3,28	3,41	3,41	3,37
3	1981526300	3,62	3,76	3,78	3,79	3,71
4	1981525754	3,34	3,6	3,7	3,75	3,73
...						
28	1981528471	3,82	3,88	3,91	3,92	3,90
29	1981523870	3,41	3,62	3,70	3,74	3,75
30	1981528304	3,89	3,93	3,95	3,95	3,94
31	1981526006	0	1,64	2,00	0	2,79
32	1981526819	0	0	0	0	0
33	1981526794	3,29	3,49	3,42	3,39	3,37
34	1981528402	3,72	3,83	3,87	3,86	3,86
35	1981522921	3,90	3,88	3,86	3,79	3,83

Keterangan field yang digunakan NIM : Nomor Induk Mahasiswa, Sem1 : Nilai indeks prestasi semester 1, Sem2 : Nilai indeks prestasi semester 2, Sem3 : Nilai indeks prestasi semester 3, Sem4: Nilai indeks prestasi semester 4, Sem5 : Nilai indeks prestasi semester 5.

3.2.2. Proses K-NN

Langkah awal pada proses K-NN adalah menentukan jumlah jarak yang akan menjadi parameter,penelitian ini menggunakan 5 (lima). Jumlah tetangga terdekat (K=5). Tabel berikut merupakan data mahasiswa yang berhenti kuliah karena nilai indeks prestasi 0(nol)atau tidak melanjutkan kuliah (tabel 3).

Tabel 3 data mahasiswa lengkap

No	NIM	Sem1	Sem2	Sem3	Sem4	Sem5	Status
1	1981528151	0	0	0	0	0	Berhenti
2	1981526345	2,94	3,28	3,41	3,41	3,37	Lanjut
3	1981526300	3,62	3,76	3,78	3,79	3,71	Lanjut
4	1981525754	3,34	3,6	3,7	3,75	3,73	Lanjut
...							
28	1981528471	3,82	3,88	3,91	3,92	3,90	Lanjut
29	1981523870	3,41	3,62	3,70	3,74	3,75	Lanjut
30	1981528304	3,89	3,93	3,95	3,95	3,94	Lanjut
31	1981526006	0	1,64	2,00	0	2,79	Lanjut

32	1981526819	1,22	0	0	0	0	Berhenti
33	1981526794	3,29	3,49	3,42	3,39	3,37	Lanjut
34	1981528402	3,72	3,83	3,87	3,86	3,86	Lanjut
35	1981522921	3,90	3,88	3,86	3,79	3,83	Lanjut

Tabel 4 tabel testing data mahasiswa

No	NIM	Sem1	Sem2	Sem3	Sem4	Sem5	Status
1	1981526125	3,45	3,62	3,65	3,66	3,66	?
2	1981525855	0	0	0	0	0	?
3	1981526530	2,67	2,79	2,58	2,79	2,69	?
4	1981523752	2,77	3,14	3,23	3,26	3,2	?
5	1981528211	2,94	3,19	0	0	0	?
6	1981528142	2,65	2,72	2,35	0	0	?
7	1981525707	2,93	3,30	3,34	3,14	3,34	?
8	1981528232	3,04	3,32	3,42	3,50	3,47	?
9	1981525875	3,68	3,76	3,75	3,75	3,77	?

Langkah selanjutnya mencari *euclid (queryinstance)* untuk masing-masing training data dengan rumus*EuclidianDistance*, tabel 5.

Tabel 5 Tabel testing data uji coba 1

16	1981526334	3.03	3,37	3,48	3,58	3,54	?

$$d1 = \sqrt{(0-3.03)^2 + (0-3,37)^2 + (0-3,48)^2 + (0-3,58)^2 + (0-3,54)^2} = \quad 7,61552$$

$$d2 = \sqrt{(2,94-3,03)^2 + (3,28-3,37)^2 + (3,41-3,48)^2 + (3,41-3,58)^2 + (3,37-3,54)^2} = \quad 0,28089$$

Tabel 6SquareInstancetoQueryDistance (SItQD)

No	NIM	Sem1	Sem2	Sem3	Sem4	Sem5	SItQD
1	1981528151	0	0	0	0	0	7,61552
2	1981526345	2,94	3,28	3,41	3,41	3,37	0,28089
3	1981526300	3,62	3,76	3,78	3,79	3,71	0,81437

4	1981525754	3,34	3,6	3,7	3,75	3,73	0,5125
...							
28	1981528471	3,82	3,88	3,91	3,92	3,90	1,1334
29	1981523870	3,41	3,62	3,70	3,74	3,75	0,57009
30	1981528304	3,89	3,93	3,95	3,95	3,94	1,2534
31	1981526006	0	1,64	2,00	0	2,79	5,26717
32	1981526819	1,22	0	0	0	0	7,21744
33	1981526794	3,29	3,49	3,42	3,39	3,37	0,38807
34	1981528402	3,72	3,83	3,87	3,86	3,86	1,01025
35	1981522921	3,90	3,88	3,86	3,79	3,83	1,13561

Selanjutnya membuat tabel pengurutan jarak euclid terkecil pada tabel 7,

Tabel 7 Tabel pengurutan jarak euclid terkecil

No	NIM	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	SItQD	UJET	ANN	Y=Class
1	1981528151	0	0	0	0	0	7,61552	34	Tidak	Tidak
2	1981526345	2,94	3,28	3,41	3,41	3,37	0,28089	4	Ya	Ya
3	1981526300	3,62	3,76	3,78	3,79	3,71	0,81437	16	Ya	Ya
4	1981525754	3,34	3,6	3,7	3,75	3,73	0,5125	10	Ya	Ya
...	...	....	...	...	...	...	....	...	...	...
28	1981528471	3,82	3,88	3,91	3,92	3,90	1,1334	20	Ya	Ya
29	1981523870	3,41	3,62	3,70	3,74	3,75	0,57009	14	Ya	Ya
30	1981528304	3,89	3,93	3,95	3,95	3,94	1,2534	22	Tidak	Ya
31	1981526006	0	1,64	2,00	0	2,79	5,26717	28	Tidak	Ya
32	1981526819	1,22	0	0	0	0	7,21744	33	Tidak	Tidak
33	1981526794	3,29	3,49	3,42	3,39	3,37	0,38807	8	Ya	Ya
34	1981528402	3,72	3,83	3,87	3,86	3,86	1,01025	18	Ya	Ya
35	1981522921	3,90	3,88	3,86	3,79	3,83	1,13561	21	Tidak	Ya

Keterangan: SItQD : SquareInstancetoQueryDistance, UJET : Urutan Jarak Euclid terkecil, ANN : Apakah termasuk Nearest Neighbor

Dengan memanfaatkan K-NN dapat diprediksikan dengan perhitungan nilai queryinstance. Pada urutanjarak yang terdekat dari 1 (satu) sampai 5 (lima) dengan nilai K yang digunakan adalah k=5, maka diketahui ada mahasiswa yang dropout dan ada mahasiswa yang lanjut kuliah. Sehingga testing datadapat dilihat pada tabel 8,

Tabel 8 Tabel hasil klasifikasi

No	NIM	Sem1	Sem2	Sem3	Sem4	Sem5	Klasifikasi
1	1981526125	3,45	3,62	3,65	3,66	3,66	Ya
2	1981525855	0	0	0	0	0	Tidak
3	1981526530	2,67	2,79	2,58	2,79	2,69	Ya
4	1981523752	2,77	3,14	3,23	3,26	3,2	Ya
5	1981528211	2,94	3,19	0	0	0	Tidak
6	1981528142	2,65	2,72	2,35	0	0	Ya
7	1981525707	2,93	3,30	3,34	3,14	3,34	Ya
8	1981528232	3,04	3,32	3,42	3,50	3,47	Ya
9	1981525875	3,68	3,76	3,75	3,75	3,77	Ya

3.2.3. Pengujian data

Pengujian data untuk mengukur akurasi menggunakan confusion matrix dengan rumus dikutip dari Raharjo Putra Kurniadi, dkk (2021)[9].

$$AKURASI = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$PRESISI = \frac{TP}{TP + FP}$$

$$RECALL = \frac{TP}{TP + FN}$$

Gambar 4 rumus confusionmatrix

Hasil yang didapat sebagai berikut:

Tabel 9 Hasil confusion matrix

Keterangan	Nilai
Akurasi	0,83
Presisi	1
Recall	0,78

Dari hasil confusion matrix nilai akurasi 0,83. Nilai presisi dengan nilai 1 dan nilai recall sebesar 0,78 menggunakan K-NN dengan nilai k=5.

4. KESIMPULAN

Berdasarkan pembahasan di atas, kesimpulan penelitian ini adalah:

1. Proses prediksi mahasiswa dropout dengan penerapan algoritma K-NN yang akan diterapkan di Fakultas Ekonomi dan Bisnis.
2. Dengan memanfaatkan data mahasiswa lengkap sebanyak 35 record dan data testing sebanyak 9 record dengan nilai k=5 di prediksi sebanyak 6 mahasiswa tidak melanjutkan kuliah/dropout dan sebanyak 29 mahasiswa melanjutkan kuliah.
3. Pengujian menggunakan confusionmatrix menghasilkan nilai akurasi 0,83. Nilai presisi 1 dan nilai recall 0,78.

5. SARAN

Adapun saran yang dapat diberikan untuk pengembangan dan perbaikan penelitian ini adalah menambahkan atribut nilai kehadiran dan nilai prilaku pada proses prediksi mahasiswa yang berpotensi tidak melanjutkan kuliah/dropout.

DAFTAR PUSTAKA

[1] Jasmir, D. Zaenal, P. A. J, and E. Rasywir, "Prediksi Mahasiswa Drop Out Dengan Menggunakan Algoritma Klasifikasi Data Mining," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 82–87, 2018, [Online]. Available: http://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1864.

[2] U. A. Medan, S. O. Prosedur, and M. D. Out, "SISTEM PENJAMINAN MUTU INTERNAL (SPMI) Page 70." pp. 70–72, 2015.

[3] Kemendikbud, "Statistik Pendidikan Tinggi (Higer Education Statistic) 2020," *PDDikti Kemendikbud*. pp. 81–85, 2020, [Online]. Available: https://pddikti.kemdikbud.go.id/publikasi.

[4] R. N. Sukmana, Abdurrahman, and Y. Wicaksono, "Implementasi K-Nearest Neighbor Untuk Menentukan Prediksi Penjualan (Studi Kasus : PT Maksiplus Utama Indonesia)," *J. Teknol. Inf. dan Komun. Vol. 8 No. 2, Desember 2020*, vol. 8, no. 2, pp. 31–38, 2020.

[5] M. Arifanto and E. Santoso, "Implementasi Metode K-Nearest Neighbor untuk prediksi penjualan kemasan skincare pada PT. Universal Jaya Perkasa," *Technologic*, vol. 10, no. 8, pp. 1–9, 2015.

[6] A. J. Nathan and A. Scobell, "Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa," *Foreign Aff.*, vol. 91, no. 5, pp. 1–9, 2012.

[7] Y. Yunita, "Implementasi K-Nearest Neighbor Dalam Prediksi Mahasiswa Berhenti Kuliah," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 866, 2021, doi: 10.30865/mib.v5i3.3049.

[8] S. R. Rani, S. R. Andani, and D. Suhendro, "Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa pada SMK Anak Bangsa," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 670, 2019, doi: 10.30645/senaris.v1i0.73.

[9] R. P. Kurniadi, V. P. Widartha, and U. Telkom, "Perbandingan Akurasi Algoritma K-Nearest Neighbor Dan Logistic," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9757–9764, 2021.