

Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari

Septiya Nuraeni^{1✉}, Sri Putri Aulia Syam², Muhammad Farid Wajdi³, Bachtiar Firmansyah⁴, Muhammad Malkan⁵

¹ Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar, Republik Indonesia

^{2,3,4} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Papua, Republik Indonesia

⁵ Program Studi Pendidikan Informatika, Fakultas Saintek, Universitas Ivet Semarang, Republik Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 23-11-2022

Direvisi : 29-11-2022

Diterima : 01-12-2022

Kata Kunci:

Pemilihan Jurusan SMA, klasifikasi, K-NN, Confussion matriks

Keywords :

School selection, classification, k-nn, confuszion matrix.

Corresponding Author :

Septiya Nuraeni

Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar

Jalan Masjid No.22, Kota Blitar, Jawa Timur 66117

Email: humas@unublitar.ac.id

ABSTRAK

Di era revolusi 4.0 saat ini, Lembaga Pendidikan sangat membutuhkan sistem informasi yang dapat menunjang kemajuan teknologi yang ada di sekolah – sekolah, agar memudahkan dalam proses belajar mengajar yang mana pada saat ini tidak sedikit sekolah menerapkan sistem online termasuk dalam proses penerimaan murid baru. Oleh karena itu SMA N 02 Manokwari membutuhkan metode penjurusan yang tepat untuk membantu bidang kurikulum dalam menentukan jurusan murid tersebut sesuai dengan minat, nilai (point), dan kemampuan yang dimilikinya untuk membantu SMA N 2 Manokwari dalam penjurusan siswa baru, dapat menggunakan data mining dengan teknik klasifikasi, algoritma yang digunakan dalam penelitian ini yaitu K-NN. Setelah melakukan pengujian dengan menggunakan confusion matriks pada Algoritma K-NN dapat dihasilkan nilai akurasi yaitu sebesar 79% dengan nilai k=9.

ABSTRACT

Today, in the revolutionary 4.0 era, institutions desperately need an inflating system that could support technological advances in school-schools, in order to make it easier for school teaching processes, which at present are not a few schools implementing online systems, including the process of receiving new students. Therefore, SMAN 2 Manokwari will need a system of tailor-making to assist the curriculum in determining student majors according to its interests, values, and abilities to SMAN 2 manokwari in new student matters, can use the classified data mining, the method used in the research, k-nn algorithm. After conducting a trial using confusion matrix on k-nn algorithm can result in an accuracy value of 79% on k=9

PENDAHULUAN

Pendidikan merupakan pengajaran atau sebuah wadah untuk mengasah dan mengembangkan bakat atau kemampuan seseorang, Pendidikan juga dapat dikatakan sebagai sebuah pembelajaran untuk mengembangkan kreativitas dan pengetahuan seseorang (Ibeng, 2022). Sehingga Pendidikan adalah sebuah usaha untuk mendukung anak dalam menggapai cita-cita dan tujuan hidupnya dengan didasari nilai – nilai serta norma yang sesuai dengan adat dan aturan

yang berlaku (Ds, AIEd SuiaEyah, r, ild., 2011). Di era revolusi 4.0 saat ini, Lembaga Pendidikan sangat membutuhkan sistem informasi yang dapat menunjang kemajuan teknologi yang ada di sekolah – sekolah, agar memudahkan dalam proses belajar mengajar yang mana pada saat ini tidak sedikit sekolah menerapkan sistem online termasuk dalam proses penerimaan siswa baru.

Dalam hal ini, sekolah membutuhkan sistem untuk membantu siswa dalam mengembangkan bakat serta proses pembelajaran akademik sesuai dengan keinginannya, salah satunya dengan sistem penjurusan siswa baru. Sampai saat ini banyak murid SMA yang sulit untuk menentukan jurusan yang sesuai dengan keahliannya. Hal yang sering dilakukan oleh sebagian siswa untuk menghilangkan rasa bingungnya, yaitu dengan konsultasi langsung baik bersama guru walinya atau pun orang tua masing – masing. Cara tersebut sangat tidak efisien ketika banyak pendapat dari beberapa orang yang pada akhirnya akan membuat siswa semakin bingung. Penjurusan ini adalah Sebagian dari proses pemilihan pengajaran yang sesuai dengan minat dan bakat siswa (Rahayuningtyas & Winarso, 2019).

Oleh karena itu SMA N 02 Manokwari membutuhkan sistem (metode) penjurusan yang bertujuan untuk membantu bidang kurikulum dalam menentukan jurusan siswa baru sesuai minat, nilai, dan kemampuan yang dimiliki. Sejak tahun 2016 SMA N 02 Manokwari merupakan salah satu sekolah yang telah menerapkan kurikulum 2013, yang mana penjurusan tidak lagi dilakukan di kelas XI atau XII, tetapi sejak siswa duduk di bangku kelas X (Sihotang et al., 2018). Di SMA N 02 Manokwari memiliki 2 jurusan yaitu IPA dan IPS. Penjurusan ini sangat penting dilakukan karena sangat mempengaruhi proses belajar siswa, ketika siswa salah masuk dalam jurusan maka akan mempengaruhi prestasi siswa tersebut.

Berdasarkan permasalahan yang ada, maka penulis memiliki kreativitas untuk mengklasifikasikan data nilai ujian murid baru agar dapat menentukan jurusan yang tepat sehingga murid tidak salah jurusan. Algoritma yang dapat digunakan yaitu Naïve Bayes, Logistic Regression, Pohon Keputusan C4.5, dan KNN. Algoritma KNN merupakan metode supervised, yang mana hasil dari data uji yang diklasifikasikan sesuai dengan mayoritas kategori dalam KNN yang memiliki tujuan yaitu mengklasifikasi data baru sesuai dengan atribut dan training samples (Sulistiyono, 2014). KNN memiliki nilai akurasi yang terus meningkat, sehingga algoritma ini sangat baik jika digunakan untuk menentukan jurusan siswa di SMA N 02 Manokwari.

K-NN merupakan metode yang termasuk dalam pengaplikasian data yang sederhana dan sangat mudah dalam pengimplementasiannya, efektif pada data yang lebih besar, serta dapat mengklasifikasikan data dengan cepat. KNN juga kuat terhadap data training yang noise, dan dapat menghasilkan data yang lebih akurat. Algoritma ini dapat mengklasifikasikan objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Yani et al., 2020). Oleh karena itu metode KNN sesuai dengan data yang akan digunakan pada penelitian ini. KNN banyak diterapkan diberbagai bidang penelitian. Dalam (Rahmat Dian Nugraha et al., 2020) KNN digunakan untuk menyeleksi karyawan baru berdasarkan klasifikasi data yang termasuk dalam kategori lolos atau tidak dengan akurasi sebesar 91%. Sedangkan (Milzam et al., 2018) menggunakan KNN untuk mendiagnosa penyakit sapi berbasis android dengan akurasi 97,56%. KNN juga seringkali di bandingkan dengan metode lainnya untuk suatu tujuan yang spesifik.

METODE PENELITIAN

Data yang akan di proses dalam penelitian ini yaitu data nilai ujian masuk (tes) yang didapatkan dari SMA N 02 Manokwari, yang berjumlah 74 kemudian data tersebut akan digunakan untuk variabel dalam perhitungan. Dalam menentukan jurusan siswa baru terdapat beberapa atribut yaitu berdasarkan nilai Matematika, Bahasa Indonesia, IPA, IPS dan nilai Bahasa Inggris, dalam setiap atribut atau variabel terdapat pernyataan atau indikator yang menjadi tolak ukur siswa baru masuk dalam jurusan IPA atau IPS. Dalam penelitian ini teknik pengumpulan datanya dilakukan yaitu dengan mewawancarai secara langsung pihak sekolah terkait permasalahan yang ada, dan atribut yang sekolah gunakan untuk menentukan jurusan siswa baru.

Berdasarkan analisis kebutuhan data penjurusan siswa baru, 74 Data akan dijadikan sampel dengan data training 60 dan data testing 14 dengan perbandingan 80% : 20%. Dalam penelitian yang menggunakan algoritma K-NN, maka di perlukannya beberapa tahap yaitu memasukkan nilai K, kemudian memasukkan data training dan data testing, lalu melakukan normalisasi untuk seluruh data baik data training maupun data testing, metode yang digunakan yaitu dengan menggunakan *min-max*.

$$\text{Normalisasi} = \frac{\text{data}_x - \text{data}_{\min}}{\text{data}_{\max} - \text{data}_{\min}} \dots\dots\dots(1)$$

keterangan:

- a. data_x adalah data yang akan dihitung normalisasinya
- b. data_{\min} adalah data terkecil

Setelah melakukan normalisasi, kemudian menghitung jarak Euclidean, dapat di simbolkan sebagai $d_{\text{euclidean}}(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$ dengan penjelasan data i yaitu banyaknya set data dan x yaitu jumlah set dari data testing, kemudian y yaitu set data training, lalu melakukan pemilihan atau mengurutkan jarak Euclidean dari yang terkecil hingga yang terbesar sejumlah nilai k kemudian terakhir yaitu menentukan hasil klasifikasi berdasarkan tahap pengurutan jarak euclidean terbanyak(Tanjungpinang, 2021).

Untuk menentukan jurusan IPA atau IPS bobot dari setiap kriteria akan dijumlahkan dan dihitung sesuai dengan rumus K-NN.

1. Algoritma K-Nearest Neighbor

K-Nearest Neighbor adalah sebuah algoritma klasifikasi yang hasil datanya didapatkan berdasarkan jarak tetangga terdekat, tujuannya yaitu mengklasifikasikan objek baru dari atribut yang berasal dari data latih, dalam proses klasifikasi umumnya nilai K dapat menggunakan jumlah ganjil dengan pertimbangan perhitungan menggunakan persamaan 3 agar tidak adanya jarak yang sama, jarak tetangga terdekat dapat dihitung dengan *Euclidean Distance*, dengan rumus berikut(Tanjungpinang, 2021) :

$$d(x,y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \dots\dots\dots (2)$$

Keterangan :

x_i merupakan data baru, dan y_i merupakan data latih, i yaitu variabel data atau atribut, dan k yaitu dimensi data.

Beberapa tahapan dalam perhitungan *K-Nearest Neighbor* yaitu :

- a. Nilai K dapat ditentukan menggunakan rumus $k = \sqrt{n} \dots\dots\dots (3)$
- b. Menghitung jarak data testing dengan masing-masing data lainnya (data training)
- c. Urutkan jarak yang terbentuk
- d. Menentukan beberapa tetangga yang terdekat sampai urutan ke-K
- e. Menentukan kelas dari tetangga terdekat
- f. Cari dan tetapkan jumlah kelas terbanyak dari tetangga terdekat sebagai data yang akan di uji

Untuk mengetahui tingkat kesalahan yang terjadi, dapat dilihat melalui perhitungan nilai akurasi menggunakan rumus akurasi berikut (Mutiara, 2015):

$$\text{Akurasi} = \frac{\text{Jumlah pengujian benar}}{\text{jumlah data uji}} \times 100\% \dots\dots\dots (4)$$

HASIL DAN PEMBAHASAN

Untuk memastikan bidang (jurusan) yang sesuai dengan keahlian siswa yaitu “IPA” atau “IPS” di SMA N 2 Manokwari, maka dilakukan pembobotan terlebih dahulu di setiap kelasnya, kemudian dijumlahkan dan dihitung sesuai dengan rumus K-NN. Pembobotan ini dapat dilihat pada Tabel 1 berikut:

Tabel 1. Nilai (bobot) setiap kelas

No	Kategori	Bobot
1	IPA	1
2	IPS	0

Sebelum sampai dalam perhitungan K-NN langkah berikutnya adalah melakukan normalisasi pada data training (latih) dan data testing (uji) menggunakan rumus *Min – Max* seperti pada persamaan 1 diatas, dibawah ini merupakan hasil dari normalisasi data training (latih) dan data testing (uji) antara lain:

Tabel 2. Data Training Hasil Normalisasi

No	Nama	Nilai MTK	Nilai B.IND	Nilai IPA	Nilai IPS	Nilai B.ING	Bobot
1	Murid 1	0.71	0.16	0.43	0.24	0.15	1
2	Murid 2	0.90	0.6	0.76	0.63	0.84	1
3	Murid 3	0.56	0.5	0.16	0.24	0.53	0
4	Murid 4	0.79	0.33	0.5	0.63	0.61	1
5	Murid 5	0.79	0.33	0.63	0.69	0.5	1
-	-	-	-	-	-	-	-
59	Murid 59	0.64	0.53	0.93	0.66	0.80	1
60	Murid 60	0.45	0	0.1	0.09	0.34	0

Tabel 3. Data Testing Hasil Normalisasi

No	Nama	Nilai MTK	Nilai B.IND	Nilai IPA	Nilai IPS	Nilai B.ING	Bobot
1	Siswa 01	0.72	0.37	0.7	0.66	0.89	1
2	Siswa 02	0.44	0.54	0.66	0.33	0.31	1
3	Siswa 03	0.48	0.81	0.26	0.66	0.57	0
4	Siswa 04	0.03	0.24	0	0.1	0	0
5	Siswa 05	0.43	0.60	0.34	0.55	0.27	1
6	Siswa 06	0.22	0	0.16	0.2	0.05	0
7	Siswa 07	0	0.32	0.13	0	0.57	0
8	Siswa 08	0.55	0.62	0.36	0.63	0.26	0
9	Siswa 09	0.25	0.43	0.33	0.33	0.05	1
10	Siswa 010	1	0.81	0.73	0.9	1	1
11	Siswa 011	0.44	1	1	1	1	1
12	Siswa 012	0.44	0.45	0.16	0.43	0.05	1
13	Siswa 013	0.07	0.51	0.06	0.36	0.31	0
14	Siswa 014	0.74	0.81	0.26	0.83	0.89	0

Klasifikasi

Setelah melakukan pengolahan data, langkah berikutnya yaitu melakukan perhitungan data menggunakan Algoritma K-NN dengan berbagai tahap yaitu : Menentukan nilai K, dalam menentukan nilai K ini dapat ditentukan menggunakan persamaan 3 (tiga) yaitu $K = \sqrt{80} = 8.94 = 9$, sehingga pada penelitian ini akan menggunakan nilai K = 9. Selanjutnya menghitung jarak antara data testing dan data training dengan rumus *Euclidean Distance*, kemudian hasil perhitungannya dapat dilihat pada tabel berikut :

Tabel 4. Hasil Perhitungan Jarak Euclid Data Testing Satu Terhadap Data Training

No	Nama	Jarak <i>Euclidean</i>
1	Murid 1	0.55044751
2	Murid 2	0.77418944
3	Murid 3	0.56945217
4	Murid 4	0.61107042
5	Murid 5	0.57560464
-	-	-
-	-	-
59	Murid 59	0.68049774
60	Murid 60	0.82040052

Setelah melakukan perhitungan jarak, maka langkah berikutnya yaitu melakukan pengurutan data mulai dari jarak terkecil sampai jarak terbesar, selanjutnya memberikan label kelas berdasarkan dengan data training, seperti pada tabel berikut :

Tabel 5. Hasil Data yang telah di urutkan dan pelabelan Data Testing Satu

No	Nama	Jarak <i>Euclidean</i>	Jurusan
1	Murid 1	0.35410361	IPA
2	Murid 2	0.35839871	IPS
3	Murid 3	0.36328394	IPA
4	Murid 4	0.38936067	IPA
5	Murid 5	0.40003962	IPA
-	-	-	-
-	-	-	-
59	Murid 59	1.01326392	IPA
60	Murid 60	1.08627303	IPS

Kemudian tahap terakhir yaitu melakukan perhitungan Data Testing Satu yang telah diurutkan berdasarkan nilai K dan jarak terdekat. Hasil prediksi penjurusan siswa baru di SMA N 2 Manokwari dapat dilihat melalui hasil dari perbandingan kelas. Apabila jumlah data jurusan IPA lebih dominan dibandingkan jurusan IPS, maka hasilnya adalah IPA, begitupun sebaliknya.

Tabel 6. Hasil perhitungan Data Testing Satu dengan K=9

No	Nama	Jarak <i>Euclidean</i>	Jurusan
1	Murid 23	0.35410361	IPA
2	Murid 12	0.35839871	IPS
3	Murid 47	0.36328394	IPA
4	Murid 27	0.38936067	IPA
5	Murid 57	0.40003962	IPA
6	Murid 9	0.41457608	IPS
7	Murid 43	0.41766581	IPA
8	Murid 31	0.43271583	IPS
9	Murid 15	0.45429911	IPS

Dari Tabel 6 diatas dapat dilihat bahwa data dengan kelas IPA berjumlah 5 dan IPS berjumlah 4, sehingga dapat disimpulkan hasil dari prediksi data testing satu, yaitu "IPA". Hasil dari prediksi data testing keseluruhan beserta perbandingan kelas data testing asli dapat dilihat pada Tabel berikut :

Tabel 7. Hasil Prediksi Data Testing

No	Nama	Data Asli	Data Prediksi	Keterangan
1	Siswa 01	IPA	IPA	Sesuai
2	Siswa 02	IPA	IPA	Sesuai
3	Siswa 03	IPS	IPA	Tidak Sesuai
4	Siswa 04	IPS	IPS	Sesuai
5	Siswa 05	IPA	IPS	Tidak Sesuai
6	Siswa 06	IPS	IPS	Sesuai
7	Siswa 07	IPS	IPS	Sesuai
8	Siswa 08	IPS	IPS	Sesuai
9	Siswa 09	IPA	IPA	Sesuai
10	Siswa 010	IPA	IPA	Sesuai
11	Siswa 011	IPA	IPA	Sesuai
12	Siswa 012	IPA	IPA	Sesuai
13	Siswa 013	IPS	IPS	Sesuai
14	Siswa 014	IPS	IPA	Tidak sesuai

Berdasarkan Tabel 7 diatas hasil prediksi data testing sebanyak 11 kelas yang sesuai dengan data asli. sehingga dapat dihitung nilai akurasi menggunakan perhitungan persamaan 4 sehingga didapatkan hasil akurasi $\frac{11}{14} \times 100\% = 79\%$. Dari hasil perhitungan akurasi tersebut algoritma KNN dalam kasus ini memiliki nilai akurasi yang tinggi.

KESIMPULAN DAN SARAN

Kesimpulan

Penerapan algoritma KNN untuk menentukan jurusan siswa baru di SMA N 2 Manokwari mampu menghasilkan nilai akurasi yang tinggi yaitu 79% dengan jumlah data training 60, sesuai dengan pengujian pada nilai K yang di gunakan yaitu K=9. Untuk menguji algoritma ini dapat menggunakan rumus *confusion matrix* dengan cara membandingkan kelas asli dengan kelas prediksi, proses perhitungan algoritma K-NN ini menggunakan Microsoft excel. Data yang digunakan dalam perhitungan ini yaitu 74 data sampel dari SMA N 2 Manokwari. Data training dan data testing yang digunakan dengan nilai K yang ditentukan sangat berpengaruh dalam pengklasifikasiannya

Saran

Berdasarkan hasil akurasi yang masih tergolong cukup baik dengan angka 79% maka diperlukannya penelitian selanjutnya unuk dapat meningkatkan nilai akurasi lebih tinggi, dan dapat menggunakan aplikasi tambahan untuk membantu perhitungan dan menguji penggunaan algoritma K-NN. Untuk penelitian selanjutnya agar dapat menambahkan dataset dengan jumlah yang lebih besar agar dapat menambah performa algoritma yang digunakan.

REFERENSI

- Ds, AlEd SuaEyah, r, ild., P.-D. (2011). *Landasan Pendidikan* (D. & Z. J. J (ed.)). I. Bmi 1,ls Ra)a Ko@tftBmi Jaw R1 10No- 5A Barjamasir 70rt9.
- Ibeng, P. (2022). $\sqrt{\text{Pendidikan: Pengertian, Tujuan, Fungsi Menurut Para Ahli. PENDIDIKAN.CO.ID. https://pendidikan.co.id/pengertian-pendidikan/}}$
- Milzam, A., Hidayat, N., & Mahfud, moch. cholil. (2018). Sistem Pakar Diagnosis Penyakit Pada Sapi Menggunakan Metode Dempster-Shafer Berbasis Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3767–3770.
- Mutiara, I. dan A. (2015). Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4. *Klik - Kumpulan Jurnal Ilmu Komputer*, 2(2), 159–173.

- Rahayuningtyas, C., & Winarso, D. S. (2019). Implementasi Algoritma k-Nearest Neighbor untuk Penjurusan Siswa SMA. *CAHAYATECH*, 6(2), 50. <https://doi.org/10.47047/ct.v6i2.10>
- Rahmat Dian Nugraha, A., Auliasari, K., & Agus Pranoto, Y. (2020). IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus : BFI Finance Surabaya). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 4(2), 14–20. <https://doi.org/10.36040/jati.v4i2.2656>
- Salam, A., Zeniarja, J., Septiyan, R., Khasanah, U., Studi, P., Informatika, T., Komputer, F. I., Nuswantoro, U. D., & Analysis, S. (2018). *ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA)*. 480–486.
- Sihotang, J. H., Suparman, A. R., & Niken, C. (2018). Studi Perbandingan Model Pembelajaran Dan Jenis Kelamin Terhadap Hasil Belajar Kognitif Peserta Didik Kelas X IPA SMA Negeri 2 Manokwari. *Jurnal Nalar Pendidikan*, 6(2), 120–127. <https://ojs.unm.ac.id/nalar/article/view/7086>
- Sulistiyo, A. (2014). □ Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang. *Fasilkom Udimus*, 1(1), 1–5.
- Tanjungpinang, D. I. W. (2021). *Online ISSN 2746-8461*. 250–256.
- Yani, S., Jumeilah, F. S., & Kadafi, M. (2020). Algoritma K-Nearest Neighbor Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus : Kelurahan Karya Jaya). *Journal of Information Technology Ampara*, 1(2), 75–87. <https://doi.org/10.51519/journalita.volume1.issuet2.year2020.page75-87>