



Implementasi Metode *K-Nearest Neighbors* (KNN) Untuk Menentukan Jurusan Siswa di SMK Sumber Bunga

Komarul Imam¹, Zaehol Fatah²

¹ Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy

² Sistem Informasi, Universitas Ibrahimy

[1komarulaza@gmail.com](mailto:komarulaza@gmail.com), [2zaeholfatah@gmail.com](mailto:zaeholfatah@gmail.com)

Abstrak

Penentuan jurusan siswa di SMK merupakan proses penting yang dapat memengaruhi kesuksesan belajar siswa dan karier masa depannya. Proses ini sering kali membutuhkan pertimbangan berbagai faktor akademik, seperti nilai mata pelajaran utama. Penelitian ini menggunakan algoritma *K-Nearest Neighbors* (KNN) untuk membantu mengklasifikasikan siswa ke jurusan yang sesuai berdasarkan data nilai mata pelajaran, seperti Bahasa Indonesia, IPA, IPS, dan Matematika. Dengan data siswa SMK Sumber Bunga, model KNN dikembangkan dan dievaluasi untuk menentukan efektivitasnya dalam mengklasifikasi jurusan "Teknologi Komputer dan Jaringan" serta "Multimedia". Hasil evaluasi menunjukkan akurasi model mencapai 97,14%, dengan presisi dan recall yang tinggi pada kedua jurusan. Tingkat keyakinan (*confidence*) dari prediksi model juga memberikan gambaran yang jelas tentang keakuratan setiap prediksi. Hasil ini menunjukkan bahwa metode KNN dapat diimplementasikan sebagai alat bantu yang efektif untuk penentuan jurusan siswa, sehingga dapat mendukung keputusan yang lebih objektif dan sesuai dengan kemampuan akademik siswa.

Kata Kunci: *K-Nearest Neighbors*, Data Mining, evaluasi, pelajaran, SMK

PENDAHULUAN

Pada era modern ini, pemilihan jurusan yang tepat di Sekolah Menengah Kejuruan (SMK) menjadi faktor penting dalam mempersiapkan siswa untuk menghadapi dunia kerja. SMK merupakan satuan pendidikan formal yang menyelenggarakan pendidikan kejuruan pada jenjang pendidikan menengah dimana kurikulum SMK lebih banyak praktek dari pada teori (Asmoni 2018). SMK dirancang untuk memberikan keterampilan praktis yang relevan dengan industri dan lapangan kerja, sehingga pemilihan jurusan yang sesuai dengan minat dan kemampuan siswa sangat berpengaruh pada kesuksesan siswa tersebut di masa depan. Pemilihan jurusan yang tepat bagi siswa di jenjang pendidikan menengah atas merupakan langkah penting dalam menunjang perkembangan akademik dan kariernya di masa depan. Proses penjurusan yang efektif memerlukan pertimbangan dari berbagai faktor, seperti prestasi akademik. Namun, proses penentuan jurusan secara manual sering kali dipengaruhi oleh subjektivitas dan terbatasnya data yang dapat dianalisis. Oleh karena itu, diperlukan metode yang lebih objektif dan akurat dalam mendukung proses pengambilan keputusan ini (Ramadhani 2018).

Algoritma *K-Nearest Neighbor* atau yang biasa disebut KNN merupakan metode pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Algoritma ini berusaha mengklasifikasikan data baru yang belum diketahui class-nya dengan memilih data sejumlah k yang letaknya terdekat dari data baru tersebut. Class terbanyak dari data terdekat sejumlah k tersebut dipilih sebagai class yang diprediksikan untuk data yang baru. k umumnya ditentukan dalam jumlah ganjil untuk menghindari munculnya jumlah jarak yang sama dalam proses pengklasifikasian (Raymana Aprilian, Roni Habibi 2020). Metode *K-Nearest Neighbors* (KNN) telah banyak digunakan dalam klasifikasi data dan analisis prediktif, termasuk di bidang pendidikan. KNN bekerja dengan mengelompokkan data berdasarkan kemiripan atau kedekatan atribut-atribut yang relevan, seperti nilai akademik atau hasil tes minat bakat siswa, sehingga menghasilkan klasifikasi yang lebih obyektif (Kurniawan, A., & Utomo 2019). Algoritma ini sederhana namun cukup efektif dalam menangani data dengan pola yang jelas, sehingga dapat diadaptasi dengan mudah pada aplikasi penentuan jurusan siswa (Rahman, T., Iskandar, A., & Setiawan 2020).

Sejumlah penelitian sebelumnya telah membuktikan bahwa penggunaan metode KNN dalam klasifikasi pendidikan mampu memberikan hasil yang akurat. Hal ini disebabkan oleh kemampuan KNN untuk melakukan klasifikasi berdasarkan data historis dan profil siswa secara individual (Putra, R., & Sari 2021). Di samping itu, metode ini juga fleksibel untuk diterapkan pada berbagai jenis data, termasuk data kuantitatif dan kualitatif, yang sering kali digunakan dalam penentuan jurusan siswa (Nugraha 2022).

Studi ini bertujuan untuk menerapkan metode KNN dalam menentukan jurusan siswa berdasarkan data akademik dan psikologis. Implementasi KNN diharapkan dapat meningkatkan akurasi dan mengurangi kesalahan dalam proses penentuan jurusan, sehingga lebih sesuai dengan minat dan kemampuan siswa (Permana, S., & Dewi 2023). Dengan

demikian, penelitian ini diharapkan mampu berkontribusi pada optimalisasi proses penjurusan di sekolah menengah atas.

METODE

Metode Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif dengan menggunakan metode data mining untuk mengembangkan model klasifikasi yang bertujuan membantu dalam penentuan jurusan yang sesuai bagi siswa di SMK. Algoritma yang digunakan adalah *K-Nearest Neighbor* (KNN), yang bekerja dengan mengelompokkan data berdasarkan kesamaan karakteristik dengan data yang telah ada. Data historis siswa SMK, yang mencakup nilai akademik seperti Bahasa Indonesia, Matematika, IPA, IPS, digunakan sebagai data pelatihan untuk membangun model KNN. Tujuan utama dari penelitian ini adalah untuk mengevaluasi efektivitas KNN dalam memprediksi jurusan siswa dengan membandingkan kesamaan antara siswa yang baru dengan data siswa dalam set pelatihan. Model KNN ini dilatih menggunakan data historis yang sudah memiliki jurusan yang ditetapkan, kemudian digunakan untuk memprediksi jurusan siswa baru dengan mencari *K* tetangga terdekat berdasarkan perhitungan jarak antar fitur. Setelah model dilatih, dilakukan pengujian untuk menilai akurasi prediksi dengan membandingkan hasil prediksi dengan jurusan yang sebenarnya, untuk mengevaluasi sejauh mana KNN efektif dalam menyediakan rekomendasi jurusan yang objektif dan berbasis data.

Data dalam penelitian ini mencakup beberapa fitur yang menggambarkan karakteristik siswa, yang diambil dari nilai mata pelajaran inti seperti Bahasa Indonesia, IPA, IPS, dan Matematika, yang masing-masing mewakili kemampuan akademik siswa di bidang tertentu—literasi dan verbal untuk Bahasa Indonesia, potensi sains untuk IPA, kemampuan sosial untuk IPS, serta logika dan analisis untuk Matematika. Pengumpulan data dilakukan secara sistematis dari arsip sekolah, mencakup nilai rapor dan hasil tes asesmen lain yang relevan untuk penentuan jurusan. Setelah data terkumpul, langkah standarisasi dan format ulang dilakukan untuk memastikan semua data dalam format seragam yang mudah diolah, seperti skala yang konsisten untuk setiap fitur atau konversi nilai ke dalam skala persentase. Proses ini bertujuan agar data siap dianalisis oleh algoritma KNN, sehingga hasil prediksi jurusan bagi siswa dapat dilakukan dengan akurat.

Metode Pengambilan Data

Data pada tabel memiliki kolom seperti Nama Siswa, nilai-nilai mata pelajaran utama (Bahasa Indonesia, IPA, IPS, Matematika), Nilai Total, dan Jurusan yang dipilih siswa (misalnya, Teknologi Komputer dan Jaringan atau Multimedia). Dalam implementasi metode *K-Nearest Neighbors* (KNN), data ini digunakan dengan menjadikan nilai-nilai mata pelajaran sebagai fitur input, sementara kolom *Jurusan* menjadi label atau target. Proses klasifikasi KNN bekerja dengan mencari sejumlah tetangga terdekat dari data baru (siswa baru tanpa jurusan), kemudian memprediksi jurusan berdasarkan jurusan yang dimiliki tetangga-tetangga terdekat tersebut. Dengan demikian, pengambilan data dari jurnal ini dilakukan melalui langkah menyiapkan data training yang mencakup nilai siswa-siswa terdahulu, memilih fitur utama (nilai-nilai mata pelajaran), dan memproses klasifikasi menggunakan algoritma KNN untuk membantu menentukan jurusan siswa baru berdasarkan pola yang ditemukan dari data sebelumnya.

Tabel 1. Dataset Training

Nama Siswa	Nilai Bahasa Indonesia	Nilai IPA	Nilai IPS	Nilai Matematika	Nilai Total	Jurusan
A. Adi Sanjaya	52	97	76	77	302	Teknologi Komputer dan Jaringan
Ababil Maulana	95	59	61	84	299	Multimedia
ABD. FAQIH ALFARISI	86	70	77	70	303	Teknologi Komputer dan Jaringan
Abdul Baqi khoirul Umam	85	67	58	54	264	Multimedia
Abdul Fatah Ramdani	73	92	76	52	293	Multimedia
ABDUL WAHID HASIM ABELIA	76	82	56	62	276	Multimedia
WIDA SUGIARTI	77	65	61	57	260	Multimedia
ABIL FARHAT FATIHI ULUM	87	51	92	97	327	Teknologi Komputer dan Jaringan
ACH. FAUZI	55	93	89	85	322	Teknologi Komputer dan

						Jaringan
ACH. NUFIL RAMADHAN	65	51	59	80	255	Multimedia

Tabel 2. Dataset Testing

Nama Siswa	Nilai Bahasa Indonesia	Nilai IPA	Nilai IPS	Nilai Matematika	Nilai Total	Jurusan
A. Adi Sanjaya	52	97	76	77	302	
Ababil Maulana	95	59	61	84	299	
ABD. FAQIH ALFARISI	86	70	77	70	303	
Abdul Baqi khoirul Umam	85	67	58	54	264	
Abdul Fatah Ramdani	73	92	76	52	293	
ABDUL WAHID HASIM	76	82	56	62	276	
ABELIA WIDA SUGIARTI	77	65	61	57	260	
ABIL FARHAT FATIH ULUM	87	51	92	97	327	
ACH. FAUZI	55	93	89	85	322	

K-Nearest Neighbor (KNN)

KNN ialah mencari k tetangga terdekat yang nantinya digunakan untuk menentukan nilai peubah output. (Umu Sa'adah, Masithoh Yessi Rochayani, Dwi Wahyu Lestari 2021). K-Nearest Neighbor (KNN) adalah metode klasifikasi yang bekerja dengan mengidentifikasi K tetangga terdekat dari data yang ingin diprediksi, berdasarkan kesamaan karakteristik atau fitur yang dimilikinya. KNN adalah metode yang sederhana dan mudah diimplementasikan, yang mengklasifikasikan data berdasarkan jarak antara data yang ingin diprediksi dengan data yang sudah ada dalam dataset. metode ini masuk dalam teknik *lazy learning* dan banyak digunakan untuk dapat memprediksi atau klasifikasi suatu data (Yogianto, Homaidi, and Fatah 2024). Dalam konteks penelitian ini, KNN digunakan untuk memprediksi jurusan yang sesuai untuk siswa di SMK Sumber Bunga berdasarkan karakteristik akademik peserta didik. Karakteristik ini termasuk nilai mata pelajaran inti seperti Bahasa Indonesia, IPA, IPS, dan Matematika. Data historis yang berisi nilai-nilai akademik dan jurusan yang dipilih oleh siswa sebelumnya akan digunakan untuk melatih model KNN. Dengan demikian, model ini akan dapat mempelajari pola-pola hubungan antara nilai akademik dan jurusan yang sesuai, sehingga dapat memberikan prediksi yang lebih akurat untuk siswa yang akan datang.

Klasifikasi

Klasifikasi merupakan bentuk analisis data yang dapat digunakan untuk mengekstraksi model yang menjelaskan pentingnya kelas data. Pada praktiknya, klasifikasi dapat memprediksi label kategori dalam bentuk diskrit, tidak terurut. Analisis ini dapat memberikan pemahaman yang lebih baik tentang data pada umumnya (Deny Jollyla, Prihandoko, Alyauma Hajjah, Elin Haerani 2023).

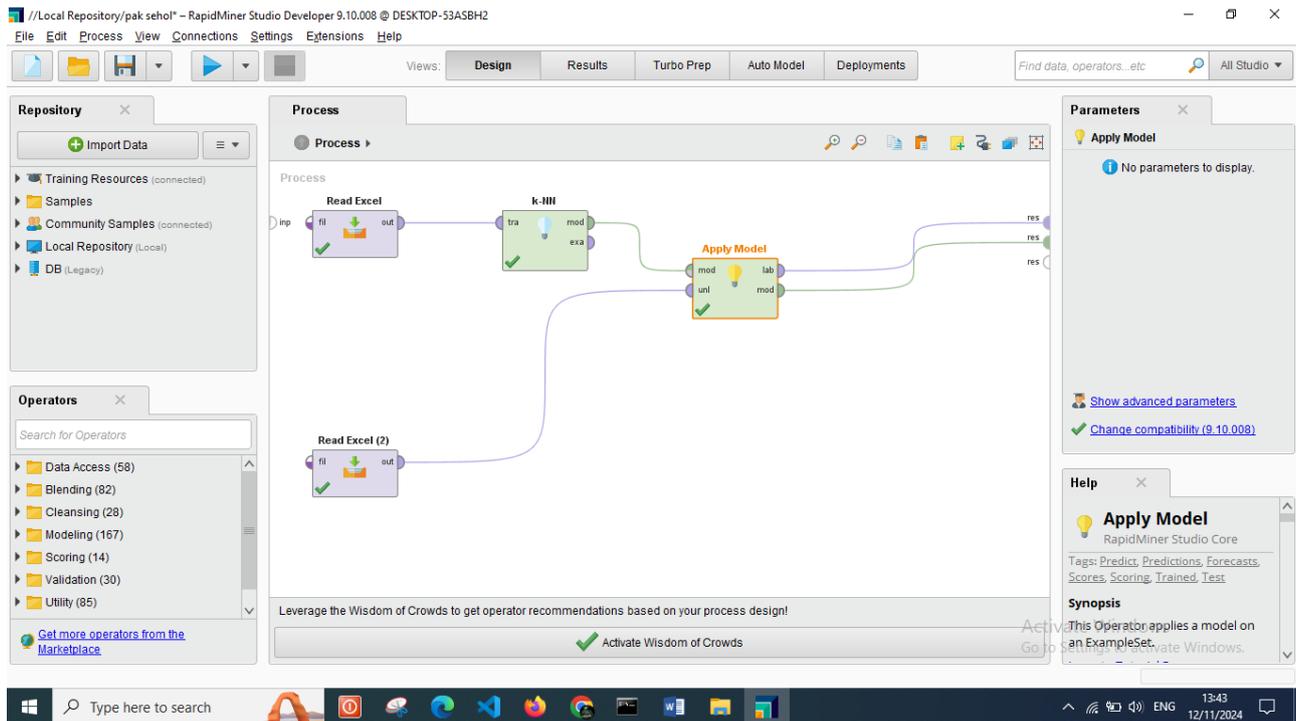
Data Mining

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalan pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basisdata (Jundanuddin et al. 2024)

HASIL DAN PEMBAHASAN

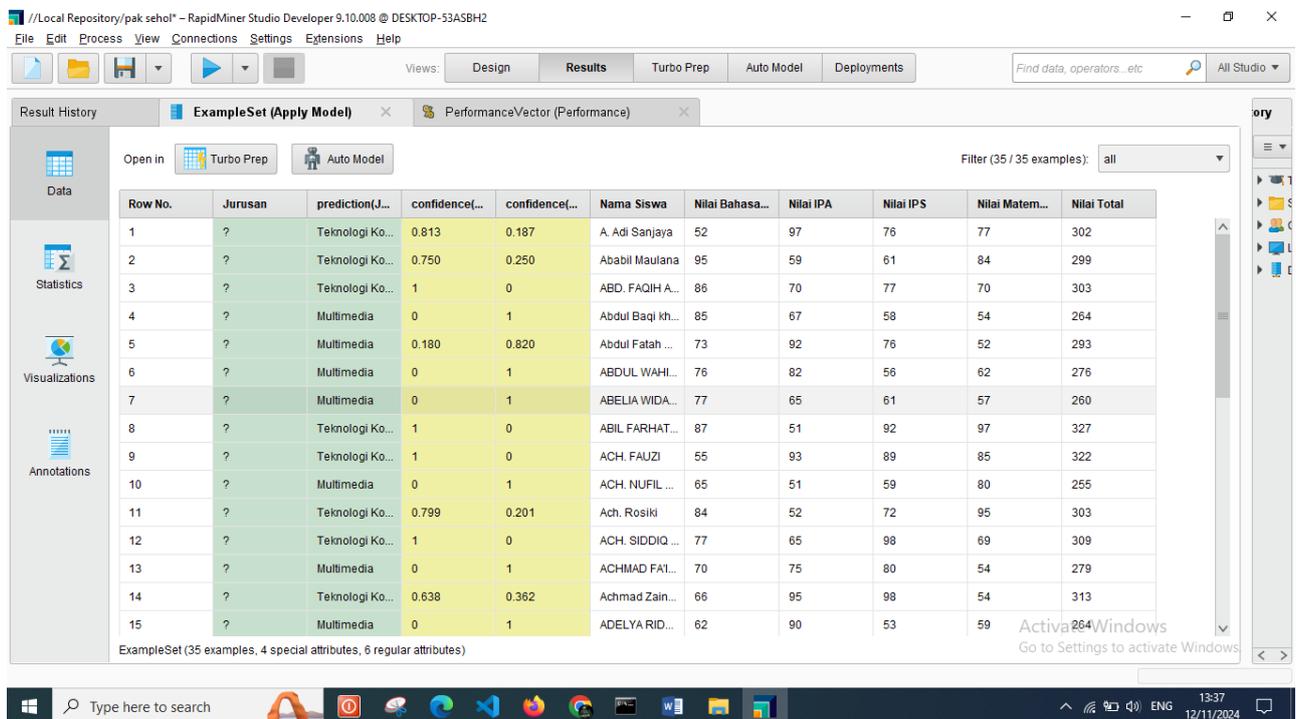
Model klasifikasi yang diterapkan adalah *K-Nearest Neighbors* (KNN), menggunakan algoritma Machine Learning RapidMiner versi 9.10. Dataset yang digunakan diambil dari kaggle

Pada Gambar 1 menjelaskan tentang penginputan proses data melalui Rapid Miner penghitungan tools rapid miner Read Excel untuk menginput data yang akan di proses, *K-Nearest Neighbors* untuk menghitung data yang akan di proses, Apply Model untuk menguji data hasil operator *K-Nearest Neighbors* model ini mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model preprocessing



Gambar 1 Proses Perhitungan Tools Rapidminer dengan Metode KNN

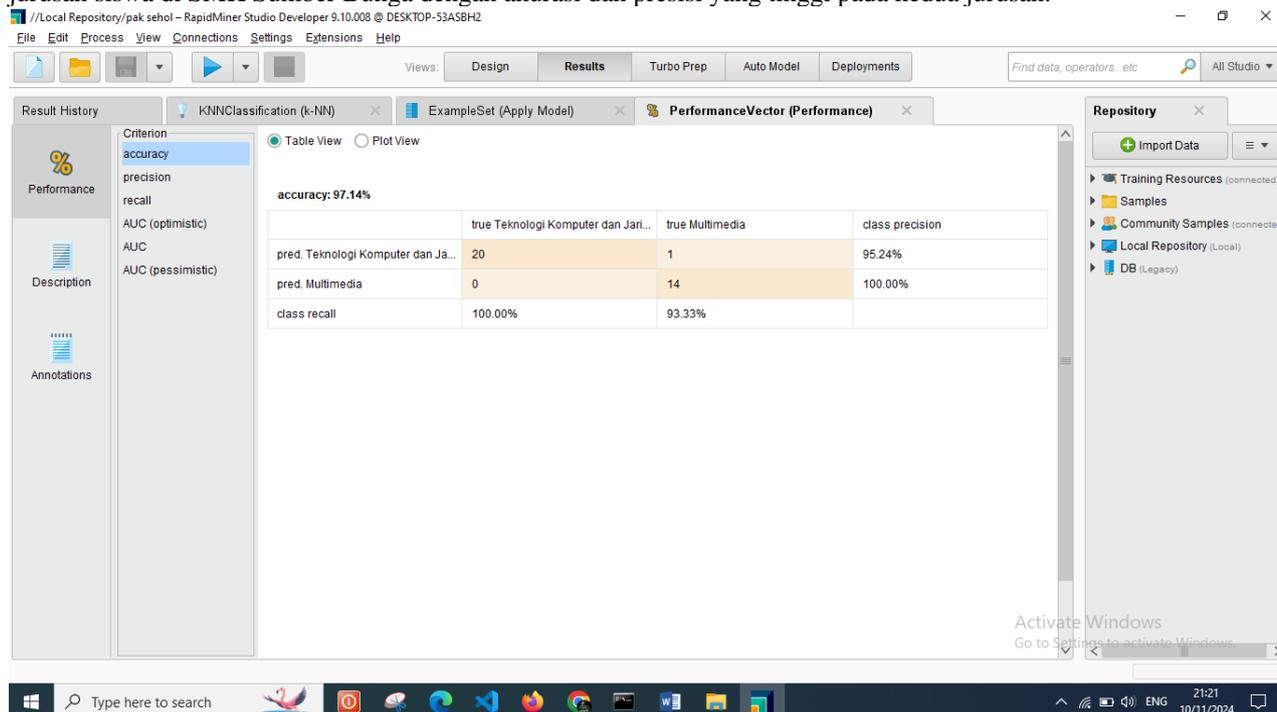
Gambar 2 menampilkan hasil prediksi model KNN dalam menentukan jurusan siswa dengan informasi tambahan berupa tingkat keyakinan (*confidence*). Kolom "Prediction (Jurusan)" menunjukkan hasil prediksi, sementara dua kolom *confidence* menunjukkan tingkat keyakinan untuk masing-masing jurusan, seperti "Teknologi Komputer dan Jaringan" dan "Multimedia".



Gambar 2 Klasifikasi

Hasil akurasi model KNN menunjukkan performa klasifikasi yang sangat baik dalam menentukan jurusan siswa dengan akurasi sebesar 97,14%, yang berarti 97,14% dari prediksi model sesuai dengan data asli. Presisi untuk jurusan "Teknologi Komputer dan Jaringan" adalah 95,24%, yang menunjukkan bahwa 95,24% dari siswa yang diprediksi pada jurusan ini benar-benar berasal dari jurusan tersebut, sedangkan presisi untuk jurusan "Multimedia" mencapai 100%, yang berarti semua prediksi model untuk jurusan ini benar. Recall untuk "Teknologi Komputer dan Jaringan" adalah 100%, yang menunjukkan bahwa model berhasil mengidentifikasi semua siswa dalam jurusan ini tanpa ada yang

terlewatkan, dan recall untuk "Multimedia" adalah 93,33%, yang berarti 93,33% siswa dalam jurusan ini terprediksi dengan benar. Secara keseluruhan, hasil ini mengindikasikan bahwa model KNN sangat andal dalam menentukan jurusan siswa di SMK Sumber Bunga dengan akurasi dan presisi yang tinggi pada kedua jurusan.



Gambar 3 Hasil Akurasi

KESIMPULAN

Metode K-Nearest Neighbor (KNN) efektif dalam membantu SMK Sumber Bunga memilih jurusan yang sesuai untuk siswa berdasarkan data seperti nilai akademik. Dengan memilih nilai K yang optimal, metode ini mampu menghasilkan tingkat akurasi tinggi dan memiliki potensi besar dalam memprediksi jurusan yang cocok bagi setiap siswa yaitu 97,14%. Selain itu, penelitian menunjukkan bahwa kualitas data sangat mempengaruhi kinerja model KNN, sehingga proses pengumpulan dan pra-pemrosesan data yang cermat sangat penting demi hasil yang akurat. Namun, kelemahan KNN, seperti ketergantungannya pada nilai K dan sensitivitasnya terhadap jumlah data, perlu diperhatikan karena dapat mengurangi akurasi prediksi jika data yang ada tidak memadai atau kurang representatif. Secara keseluruhan, KNN adalah alat bantu yang andal bagi sekolah dalam menentukan jurusan siswa secara lebih objektif, tetapi disarankan agar penelitian selanjutnya mempertimbangkan pengembangan model dan peningkatan kualitas data untuk mengoptimalkan hasil, sekaligus memperkuat sistem penentuan jurusan guna memaksimalkan potensi siswa di bidang yang tepat.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

- Asmoni. 2018. *Kebijakan Peningkatan Mutu Sekolah Menengah Kejuruan*.
- Deny Jollyla, Prihandoko, Alyauma Hajjah, Elin Haerani, Muhammad Siddik. 2023. *Algoritma Klasifikasi Untuk Pemula Solusi Python Dan RapidMiner*.
- Jundanuddin, Muhammad, Zaehol Fatah, Akhlis Munazilin, Sistem Informasi, Fakultas Sains, Universitas Ibrahimy Sukorejo, Ilmu Komputer, Fakultas Sains, and Universitas Ibrahimy Sukorejo. 2024. "Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode K-Nearest Neighbor (K-NN) Studi Kasus Pada Mi Alfagiri Silo-Jember Implementation Of Data Mining On Teacher Performance Appraisal Using K-Nearest Neighbor (K-NN) Method Case Study At Mi Alfagiri Silo-Jember." 13(105):869–80.
- Kurniawan, A., & Utomo, W. 2019. *Implementasi K-Nearest Neighbor Pada Data Klasifikasi*.
- Nugraha, F. 2022. *Pemanfaatan KNN Untuk Klasifikasi Di Dunia Pendidikan*.
- Permana, S., & Dewi, L. 2023. *Teknologi Data Mining Dalam Pengambilan Keputusan Pendidikan*.
- Putra, R., & Sari, M. 2021. "Algoritma Klasifikasi Dalam Analisis Data Pendidikan." *Media Akademika*.
- Rahman, T., Iskandar, A., & Setiawan, H. 2020. *Pengantar Data Mining Dan Aplikasinya Di Bidang Pendidikan*.
- Ramadhani, D. 2018. *Sistem Pendukung Keputusan Dalam Pendidikan*.
- Raymana Aprilian, Roni Habibi, M. Yusril Helmi Setyawan. 2020. *Algoritma KNN Dalam Memprediksi Cuaca Untuk Menentukan Tanaman Yang Cocok Sesuai Musim*. edited by R. M. Awangga. Kreatif.

- Umu Sa'adah, Masithoh Yessi Rochayani, Dwi Wahyu Lestari, Dwi Ayu Lusia. 2021. *Kupas Tuntas Algoritma Data Mining Dan Implementasinya Menggunakan R O*.
- Yogianto, Ahmad, Ahmad Homaidi, and Zaehol Fatah. 2024. "Implementasi Metode K-Nearest Neighbors (KNN) Untuk Klasifikasi Penyakit Jantung." *G-Tech: Jurnal Teknologi Terapan* 8(3):1720–28. doi: 10.33379/gtech.v8i3.4495.