

Penerapan Algoritma Knn Untuk Analisis Minat Dan Bakat Siswa

Implementation of the KNN Algorithm for Student Interest and Talent Analysis

Muhammad Al Rasyidin¹, Budi Triandi^{b,2}

^{a,b}Universitas Potensi Utama, Jl. KL. Yos Sudarso Km. 6,5 No. 3-A, Tanjung Mulia, Medan, 20241, Indonesia
M.alrasyidin05@gmail.com¹, buditriandi@gmail.com²

ABSTRAK

Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (KNN) dalam menganalisis minat dan bakat siswa di SMP Negeri 24 Medan berdasarkan data histori siswa. Metode yang digunakan meliputi pengumpulan data minat dan bakat dari hasil kuesioner serta nilai akademik, yang kemudian diolah menggunakan algoritma KNN untuk melakukan klasifikasi. Model dievaluasi berdasarkan metrik akurasi, presisi, dan recall guna menilai performanya dalam memprediksi kategori minat dan bakat siswa. Hasil penelitian menunjukkan bahwa model KNN mencapai akurasi sebesar 83,70%, dengan performa terbaik pada kelas Penyanyi (presisi 87,16%) dan terendah pada kelas Programmer (79,31%). Nilai recall tertinggi terdapat pada kelas Dokter (85,80%), sedangkan terendah pada kelas Insinyur (80,72%). Kesalahan prediksi umumnya terjadi antara kelas Programmer, Insinyur, dan Dokter karena kemiripan fitur antar kelas. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma KNN dapat digunakan secara efektif untuk mengidentifikasi pola minat dan bakat siswa, serta memberikan dasar bagi sekolah dalam merancang program pembelajaran dan bimbingan karir yang lebih terarah.

Kata Kunci: Data Mining, K-Nearest Neighbor, Students' Interests and Talents, Prediction, Learning Programs, Career Guidance.

ABSTRACT

This study aims to apply the K-Nearest Neighbor (KNN) algorithm to analyze students' interests and talents at SMP Negeri 24 Medan based on historical student data. The method involves collecting data on students' interests and talents through questionnaires and academic scores, which are then processed using the KNN algorithm for classification. The model is evaluated using accuracy, precision, and recall metrics to assess its performance in predicting categories of interests and talents. The results show that the KNN model achieved an accuracy of 83.70%, with the best performance in the Singer class (precision of 87.16%) and the lowest in the Programmer class (79.31%). The highest recall value was observed in the Doctor class (85.80%), while the lowest was in the Engineer class (80.72%). Misclassifications frequently occurred among the Programmer, Engineer, and Doctor classes due to overlapping features. Overall, the findings indicate that the KNN algorithm can be effectively utilized to identify patterns of students' interests and talents, providing a valuable foundation for schools in designing more targeted learning programs and career guidance strategies.

Keywords: Data Mining, K-Nearest Neighbor, Students' Interests and Talents, Prediction, Learning Programs, Career Guidance.

Info Artikel :

Disubmit: 11 September 2025

Direview: 10 Oktober 2025

Diterima : 3 November 2025

Copyright © 2025 – senadimu. All rights reserved.

1. PENDAHULUAN

Dalam dunia pendidikan, pengembangan potensi setiap siswa menjadi salah satu prioritas utama yang harus diperhatikan oleh para pendidik. Setiap siswa memiliki minat dan bakat yang berbeda, yang dapat berpengaruh besar terhadap keberhasilan mereka dalam proses pembelajaran dan perkembangan

karir di masa depan. Namun, untuk dapat mengenali dan mengembangkan minat serta bakat siswa dengan baik, diperlukan alat yang mampu menganalisis data secara tepat dan efisien, [1].

Di sinilah data mining (penambangan data) memainkan peran yang sangat penting. Data mining adalah proses yang digunakan untuk menemukan pola tersembunyi atau hubungan antara data yang besar dan kompleks [2]. Tujuannya adalah untuk mendeteksi hubungan yang tidak terlihat secara langsung, memungkinkan pembuatan prediksi, dan identifikasi tren yang dapat membimbing pengambilan keputusan. Dalam konteks pendidikan, data yang dimaksud meliputi berbagai informasi terkait perilaku siswa, nilai akademik, minat, serta faktor sosial dan psikologis lainnya, [3].

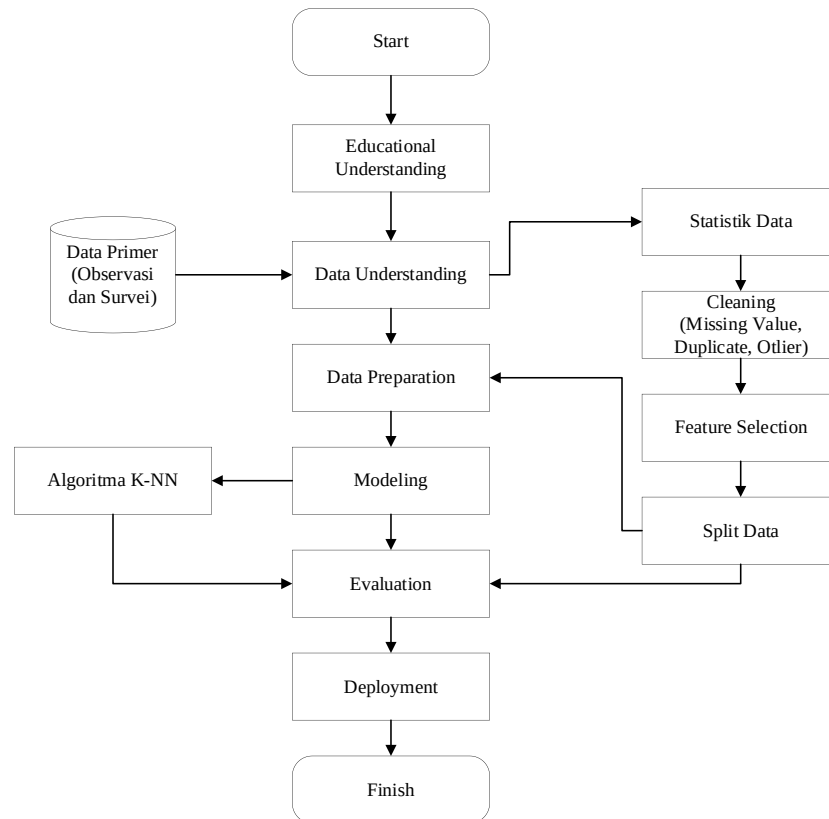
Ketersediaan data yang terbatas dan kualitas data yang kurang baik dapat mempengaruhi akurasi model machine learning dalam mengidentifikasi minat dan bakat siswa. [4]. Pemilihan metode dan algoritma yang tidak sesuai dapat menghasilkan model yang kurang efektif dalam mengelompokkan minat dan bakat siswa. [5]. Minat dan bakat siswa dapat berubah seiring waktu, dan model yang tidak dapat menyesuaikan diri dengan perubahan ini mungkin menjadi kurang akurat. [6]. Mengatasi permasalahan-permasalahan di atas memerlukan pendekatan yang holistik, termasuk peningkatan kualitas data, pemilihan metode dan algoritma yang sesuai, serta kemampuan model untuk beradaptasi dengan perubahan minat dan bakat siswa.

Salah satu teknik dalam data mining yang sering digunakan untuk analisis dan pengelompokan data adalah algoritma K-Nearest Neighbor (KNN). Algoritma KNN bekerja dengan cara mengklasifikasikan data berdasarkan kedekatannya dengan data lain yang sudah diketahui labelnya [7] Dengan menggunakan teknik ini, maka dapat mengidentifikasi pola minat dan bakat siswa berdasarkan data historis dan karakteristik mereka, serta memprediksi potensi yang mungkin mereka miliki di masa depan. Siswa sering kali memiliki minat dan bakat yang tersembunyi atau belum ter gali sepenuhnya, dan memanfaatkan data yang tersedia dapat membantu mendeteksi hal tersebut secara lebih akurat. Analisis ini tidak hanya memberikan manfaat dalam memahami potensi siswa, tetapi juga dapat meningkatkan efektivitas pengajaran, merancang program pengembangan yang lebih personal, dan memberikan panduan yang lebih baik bagi siswa dalam merencanakan masa depan mereka.

Penelitian ini bertujuan menghasilkan sebuah model prediksi atas dasar model klasifikasi untuk memberikan pengalaman atas dasar pembelajaran data untuk memberikan pola dan prediksi sehingga dapat memberikan referensi secara akurat dalam menentukan sebuah strategi pendidikan agar tepat sasaran. Menganalisis pola minat dan bakat siswa atas dasar data histori dari siswa dengan algoritma KNN. Untuk memprediksi minat dan bakat siswa di masa depan, sehingga dapat diberikan bimbingan yang lebih tepat sesuai dengan kecenderungan mereka. Melakukan analisis data untuk merancang program pembelajaran yang lebih personal. Untuk memberikan wawasan yang berguna bagi pihak sekolah dalam menentukan kebijakan pendidikan yang lebih tepat sasaran, seperti pemilihan program ekstrakurikuler atau strategi pembelajaran. Menyediakan data yang dapat digunakan untuk memberi panduan karir lebih baik bagi siswa berdasarkan minat dan bakat yang teridentifikasi melalui algoritma KNN.

2. METODE

Metode penelitian ini disusun secara sistematis untuk memastikan akurasi dan relevansi data untuk mendukung analisis yang lebih mendalam. Penelitian ini juga menggunakan data primer, adapun kerangka metodologi penelitian dapat dilihat pada Gambar 1 berikut:



Gambar. 1 Kerangka Metodologi Penelitian Analisis Minat Bakat Siswa

Berdasarkan gambar. 1 tentang kerangka metodologi penelitian dalam penerapan Data Mining dengan Algoritma K-Nearest Neighbor (KNN) untuk Analisis Minat dan Bakat Siswa meliputi:

1. *Educational Understanding*

Pada tahap ini, dilakukan pemahaman awal mengenai tujuan dan kebutuhan dari penelitian, yaitu untuk menganalisis minat dan bakat siswa.

- Input: Data primer yang diperoleh melalui observasi dan survei kepada siswa dengan jumlah data sebanyak 5000 yang tersimpan dalam minat_bakat_siswa.csv
- Tujuan: Memahami kebutuhan pendidikan, seperti membantu siswa menentukan jalur karier yang sesuai berdasarkan minat dan bakat mereka.

2. *Data Understanding*

Tahap ini bertujuan untuk mengenal lebih dalam karakteristik dataset yang digunakan.

Beberapa langkah yang dilakukan:

- Statistik Data: Menganalisis data secara deskriptif untuk memahami distribusi data, hubungan antar atribut, serta mencari pola atau outlier yang relevan.
- Cleaning Data: Menyelesaikan masalah pada data, seperti menangani *missing values*, data duplikat, atau data yang tidak valid.
- Tujuan: Memastikan data memiliki kualitas yang baik sebelum masuk ke tahap selanjutnya.

3. *Data Preparation*

Pada tahap ini, dataset dipersiapkan untuk mendukung proses modeling. Langkah-langkahnya meliputi:

- a. *Feature Selection*: Memilih atribut atau fitur yang relevan untuk analisis minat dan bakat siswa. Contohnya adalah nilai akademik, ekstrakurikuler, preferensi karier, dan lainnya.
- b. *Split Data*: Dataset dibagi menjadi data pelatihan dan data pengujian, dengan proporsi tertentu (misalnya 80:20), untuk memastikan model dapat dievaluasi secara adil.
- c. Tujuan: Membentuk dataset yang siap digunakan untuk membangun model K-NN.

4. *Modeling*

Pada tahap ini, algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk membangun model prediksi.

- a. Input: Data yang telah diproses pada tahap sebelumnya.
- b. Tujuan: Melatih model untuk dapat mengklasifikasikan siswa ke dalam kategori minat dan bakat tertentu berdasarkan atribut yang telah dipilih.

5. *Evaluation*

Model yang telah dibuat diuji dan dievaluasi menggunakan data pengujian.

- a. Metode Evaluasi: Menggunakan metrik evaluasi seperti akurasi, precision dan recall untuk menilai performa model.
- b. Tujuan: Memastikan model yang dihasilkan memiliki kinerja yang baik dan dapat digunakan dalam skenario nyata.

6. *Deployment*

Tahap akhir adalah penerapan hasil analisis ke dalam basis aturan yang dapat digunakan oleh pihak terkait.

- a. Output: Basis aturan yang berisi pola atau klasifikasi terkait minat dan bakat siswa yang dihasilkan dari model K-NN.
- b. Tujuan: Menyediakan panduan yang dapat digunakan oleh guru, konselor, atau institusi pendidikan dalam menginterpretasikan hasil klasifikasi minat dan bakat siswa, sehingga membantu dalam pengambilan keputusan pendidikan atau pengembangan siswa secara lebih terarah.

7. *Bahan Penelitian*

Bahan penelitian yang digunakan dalam studi ini terdiri dari beberapa elemen penting yang mendukung analisis data dan proses pengambilan keputusan. Berikut adalah deskripsi bahan penelitian yang digunakan meliputi, Data utama yang digunakan dalam penelitian ini diperoleh melalui pengumpulan data langsung dari sumber, yaitu siswa di SMP 45 Medan.

- a. Informasi demografi meliputi ID Siswa, Usia dan Jenis Kelamin.
- b. Nilai akademik meliputi Nilai Matematika, Nilai Bahasa Indonesia
- c. Data ekstrakurikuler Informasi tentang jenis kegiatan ekstrakurikuler yang diikuti siswa digunakan untuk memahami minat mereka di luar kegiatan akademik. Seperti olahraga, seni
- d. Minat dan bakat data ini merupakan inti dari penelitian untuk mengidentifikasi. Minat siswa, preferensi siswa terhadap bidang tertentu, seperti seni, olahraga, teknologi, atau sains. Minat ini biasanya diperoleh melalui survei atau wawancara. Bakat siswa, kemampuan alami atau keterampilan yang menonjol pada siswa, yang diidentifikasi melalui pengamatan langsung atau aktivitas yang diikuti siswa.
- e. Preferensi karier Informasi ini mencakup bidang pekerjaan atau jalur pendidikan yang diinginkan siswa di masa depan. Misalkan Siswa dengan minat di bidang teknologi dan nilai Matematika yang tinggi mungkin memilih karier di bidang teknik atau IT. Siswa dengan nilai Bahasa Indonesia tinggi dan minat pada seni mungkin memiliki preferensi karier di bidang komunikasi atau sastra.

8. *Data Statistik Pendukung*

Data statistik pendukung yang digunakan mencakup, Statistik deskriptif seperti rata-rata nilai

akademik, distribusi jenis kelamin, dan proporsi siswa dalam kegiatan ekstrakurikuler. Korelasi antar atribut untuk menentukan fitur yang paling relevan dalam analisis minat dan bakat.

ID Siswa	Uraian	Jenis Kelamin	Nilai Matematika	Nilai Bahasa Indonesia	Ekstrakurikuler	Minat	Bakat	Preferensi Karier
1	51	15	Perempuan	79	100	Klub IT	Teknologi	Programmer
2	52	15	Perempuan	82	75	Klub IT	Teknologi	Programmer
3	53	15	Laki-laki	82	70	Olimpiade Matematika	Seni	Matematika
4	54	15	Laki-laki	84	70	Olimpiade Matematika	Koridor	Matematika
5	55	15	Laki-laki	93	70	Klub IT	Choirage	Programmer
6	56	13	Laki-laki	83	85	Jurnalistik	Seni	Menulis
7	57	16	Laki-laki	75	70	Padiuan Suara	Choirage	Vokal
8	58	15	Perempuan	89	89	Olimpiade Matematika	Sains	Matematika
9	59	15	Perempuan	100	90	Seni	Menggambar	Desainer Grafis
10	60	14	Laki-laki	83	70	Seni	Sosial	Menggambar
11	61	14	Laki-laki	96	70	Jurnalistik	Sains	Menulis
12	62	15	Perempuan	70	90	Klub IT	Sains	Programming
13	63	14	Perempuan	70	100	Klub IT	Sains	Programming
14	64	17	Perempuan	83	83	Olimpiade Matematika	Seni	Matematika
15	65	16	Perempuan	87	100	Klub IT	Seni	Programming
16	66	15	Laki-laki	85	90	Seni	Teknologi	Menggambar
17	67	16	Perempuan	80	80	Klub IT	Sosial	Programming
18	68	15	Laki-laki	100	70	Jurnalistik	Sains	Menulis
19	69	14	Laki-laki	89	70	Olimpiade Matematika	Choirage	Matematika
20	70	14	Laki-laki	89	100	Jurnalistik	Choirage	Menulis
21	71	16	Laki-laki	70	82	Padiuan Suara	Sains	Vokal

Gambar. 2 Dataset Primer minat_bakat_siswa.csv

Bahan-bahan penelitian pada gambar. 2 diharapkan dapat memberikan hasil yang valid dan relevan dalam memahami pola minat dan bakat siswa di SMP Negeri 24 Medan, serta mendukung pengembangan sistem analisis berbasis data yang lebih terarah.

A. Tinjauan Pustaka

1) Gap Penelitian

Penelitian mengenai penerapan teknik data mining, termasuk algoritma K-Nearest Neighbor (KNN), untuk menganalisis minat dan bakat siswa. Penelitian ini bertujuan untuk menggali potensi siswa berdasarkan data yang dihasilkan dari tes minat, hasil ujian, dan kegiatan ekstrakurikuler. Dalam penelitian ini, data siswa yang digunakan untuk klasifikasi terdiri dari beberapa fitur, seperti hasil ujian, minat pada mata pelajaran tertentu, dan keterlibatan dalam kegiatan ekstrakurikuler.

Hasil penelitian menunjukkan bahwa penggunaan KNN memberikan hasil yang cukup akurat dalam mengklasifikasikan siswa ke dalam kategori tertentu, seperti minat dalam bidang sains, seni, atau olah raga. Berdasarkan pengelompokan ini, rekomendasi jalur pendidikan yang tepat dapat diberikan, baik itu dalam pemilihan jurusan atau karier di masa depan. Penelitian ini menggaris bawahi bahwa algoritma KNN efektif untuk membantu menentukan minat dan bakat siswa, meskipun ada tantangan dalam hal akurasi ketika jumlah data yang digunakan sangat besar dan kompleks [8].

Penelitian yang mengimplementasikan algoritma KNN untuk menganalisis minat dan bakat siswa dalam memilih bidang studi dan karier. Dalam studi ini, para peneliti mengumpulkan data dari hasil ujian, tes bakat, serta preferensi ekstrakurikuler siswa. Penelitian ini bertujuan untuk memprediksi apakah seorang siswa cenderung lebih tertarik atau berbakat di bidang teknologi, seni, atau sosial, berdasarkan pola data yang ditemukan.

Hasil eksperimen menunjukkan bahwa KNN berhasil mengidentifikasi minat siswa dengan tingkat akurasi yang tinggi. Algoritma ini mampu mengelompokkan siswa berdasarkan kemiripan dalam nilai ujian dan kecenderungan mereka terhadap bidang tertentu. Selain itu, algoritma KNN memberikan informasi yang berguna bagi pendidik untuk menyarankan jalur pendidikan yang sesuai bagi setiap siswa. Penelitian ini juga menunjukkan pentingnya memilih nilai K yang tepat dalam KNN, karena nilai K yang terlalu kecil atau terlalu besar dapat mempengaruhi kualitas klasifikasi [9].

Penelitian yang dilakukan oleh Zhang et al. fokus pada penerapan algoritma K-Nearest Neighbor (KNN) dalam pendidikan, terutama dalam menganalisis minat siswa terhadap berbagai

mata pelajaran dan kegiatan ekstrakurikuler. Dalam studi ini, KNN digunakan untuk mengklasifikasikan siswa berdasarkan preferensi mereka dalam berbagai kategori, seperti sains, teknologi, seni, dan humaniora. Data yang digunakan mencakup nilai akademik siswa, partisipasi dalam kegiatan ekstrakurikuler, serta hasil tes minat dan bakat.

Hasil penelitian ini menunjukkan bahwa KNN tidak hanya efektif untuk mengklasifikasikan siswa berdasarkan minat mereka, tetapi juga untuk memprediksi bagaimana siswa akan berperforma di bidang studi tertentu jika mereka memilih untuk mengejar jalur pendidikan tersebut. Penelitian ini mengungkapkan bahwa dengan menggunakan KNN, sistem dapat memberikan rekomendasi yang lebih personal untuk siswa mengenai jalur pendidikan yang sesuai dengan minat dan bakat mereka [10].

Liu et al. mengembangkan sistem rekomendasi untuk menganalisis minat dan bakat siswa dengan menggunakan algoritma KNN. Sistem ini dirancang untuk memberikan rekomendasi mengenai program pendidikan yang paling sesuai dengan minat dan bakat siswa berdasarkan data yang dikumpulkan dari hasil ujian, tes bakat, serta informasi demografis siswa. Dalam penelitian ini, para peneliti mengimplementasikan KNN untuk mengklasifikasikan siswa dalam beberapa kategori berdasarkan kecenderungan mereka terhadap berbagai bidang akademik dan karier.

Penelitian ini menunjukkan bahwa KNN dapat secara akurat mengelompokkan siswa ke dalam kategori minat yang berbeda, serta memberikan rekomendasi yang lebih baik dibandingkan dengan metode lainnya. Hasilnya menunjukkan bahwa penerapan KNN dalam analisis minat dan bakat siswa dapat meningkatkan kualitas keputusan pendidikan dan memberikan panduan yang lebih tepat untuk pengembangan diri siswa [11].

Prakash dan Singh melakukan penelitian yang berfokus pada penggunaan algoritma KNN untuk mengklasifikasikan siswa berdasarkan minat dan bakat mereka dalam bidang seni dan sains. Penelitian ini bertujuan untuk mengetahui apakah KNN dapat membantu memprediksi minat siswa terhadap bidang tertentu berdasarkan data yang terkumpul dari hasil ujian, prestasi ekstrakurikuler, dan hasil tes bakat.

Dalam eksperimen mereka, KNN terbukti efektif dalam memberikan rekomendasi tentang jalur pendidikan yang paling sesuai dengan bakat dan minat siswa. Berdasarkan data yang digunakan, algoritma KNN mampu memberikan hasil yang lebih tepat dan konsisten dalam mengidentifikasi bidang studi atau karier yang sesuai dengan kecenderungan siswa. Penelitian ini juga menunjukkan pentingnya penggunaan teknik pemilihan fitur yang tepat untuk meningkatkan akurasi klasifikasi [12].

Penelitian oleh Hasan dan Ahmed menginvestigasi penerapan KNN dalam menganalisis minat dan bakat siswa, dengan fokus pada penerapan di sekolah menengah. Mereka menggunakan data yang mencakup hasil tes akademik, kegiatan ekstrakurikuler, dan preferensi pribadi siswa untuk mengklasifikasikan siswa ke dalam kategori minat yang relevan, seperti bidang sains, seni, atau olahraga.

Hasil dari penelitian ini menunjukkan bahwa KNN memiliki kemampuan untuk mengklasifikasikan siswa berdasarkan kesamaan karakteristik mereka, sehingga dapat memberikan rekomendasi yang lebih terarah dalam memilih program pendidikan atau jalur karier. Penelitian ini mengungkapkan bahwa penerapan KNN pada data pendidikan memungkinkan untuk memberikan panduan yang lebih personal bagi siswa, berdasarkan karakteristik dan bakat yang dimiliki [13].

Dapat disimpulkan bahwa Penelitian-penelitian yang ada menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) dapat diterapkan dengan efektif dalam menganalisis minat dan bakat siswa. KNN memungkinkan untuk mengelompokkan siswa berdasarkan kesamaan dalam hal nilai akademik, keterlibatan dalam kegiatan ekstrakurikuler, serta hasil tes bakat. Dengan demikian, penerapan KNN dalam pendidikan memberikan manfaat yang signifikan dalam menyediakan rekomendasi yang lebih personal dan tepat mengenai jalur pendidikan dan karier yang sesuai dengan minat dan bakat siswa.

2) Data Mining

Data mining adalah proses untuk menemukan pola, informasi, dan hubungan yang tersembunyi dalam sejumlah besar data menggunakan teknik statistik, algoritma, dan model komputasi [14]. Data mining mengubah data mentah menjadi informasi yang lebih bernilai untuk mendukung pengambilan keputusan di berbagai bidang, termasuk bisnis, kesehatan, dan pendidikan. Dalam konteks pendidikan, data mining dapat digunakan untuk mengidentifikasi pola belajar siswa, memprediksi hasil akademik, serta menganalisis minat dan bakat siswa untuk merancang strategi pendidikan yang lebih efektif [15].

Menurut Chaurasia dan Pal. dalam penerapan data mining pada bidang pendidikan, terdapat berbagai metode yang digunakan untuk analisis, termasuk klasifikasi, klustering, regresi, dan prediksi. Salah satu teknik yang paling banyak digunakan untuk klasifikasi adalah algoritma K-Nearest Neighbor (KNN), yang memungkinkan identifikasi pola berdasarkan kedekatan atau kemiripan data yang ada [16].

3) Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) adalah metode pembelajaran mesin yang sering digunakan dalam klasifikasi dan regresi. KNN bekerja dengan cara menghitung jarak antara data baru dengan data yang ada dalam dataset untuk menentukan kelas atau nilai yang sesuai. Dalam klasifikasi, data yang baru akan dikategorikan ke dalam kelas mayoritas dari K tetangga terdekat berdasarkan pengukuran jarak tertentu, seperti jarak Euclidean, Manhattan, atau Minkowski [17].

Kelebihan utama dari KNN adalah kesederhanaan implementasinya dan tidak memerlukan proses pelatihan model yang rumit. Namun, kekurangannya adalah bahwa KNN sangat bergantung pada ukuran dan dimensi data, sehingga dapat mengalami masalah dalam efisiensi waktu dan komputasi ketika jumlah data sangat besar [18]. Dalam konteks analisis minat dan bakat siswa, KNN dapat digunakan untuk mengelompokkan siswa berdasarkan kesamaan karakteristik atau preferensi mereka terhadap mata pelajaran atau kegiatan tertentu, seperti yang dilakukan dalam penelitian oleh [19].

Algoritma K-NN dalam menghitung Jarak antara data uji dan data latih menggunakan *Euclidean Distance* untuk menghitung jarak antara data uji dan data latih.

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

Di mana:

- xi: Nilai fitur dari data uji.
- yi: Nilai fitur dari data latih.
- n: Jumlah fitur.

Contoh: Jika data uji memiliki $x=(13,92,95)$ dan data latih memiliki $y=(15,99,88)$ dengan menggunakan persamaan (1) maka perhitungan seperti berikut ini

$$d(x, y) = \sqrt{(13 - 15)^2 + (92 - 99)^2 + (95 - 88)^2}$$

Setelah semua jarak dihitung, urutkan data latih berdasarkan jarak dari data uji, dari yang paling kecil ke yang paling besar.

Contoh Hasil Pengurutan:

1. Data Latih A: Jarak = 2.5
2. Data Latih B: Jarak = 3.0
3. Data Latih C: Jarak = 3.5

Tentukan Kelas Data Uji Berdasarkan Label Mayoritas

1. Pilih K data terdekat dari hasil pengurutan.
2. Tentukan kelas data uji berdasarkan label mayoritas dari K tetangga terdekat.

Contoh: Jika K = 3, tetangga terdekat memiliki kelas:

- a) Atlet
- b) Atlet
- c) Dokter

Maka kelas mayoritas adalah Atlet, sehingga data uji diklasifikasikan sebagai Atlet.

4) Penerapan Data Mining dalam Pendidikan

Dalam dunia pendidikan, penerapan data mining telah berkembang pesat dalam beberapa tahun terakhir. Salah satu aplikasi yang paling sering digunakan adalah untuk memprediksi hasil akademik siswa, memberikan rekomendasi kurikulum, serta menganalisis faktor-faktor yang mempengaruhi prestasi dan minat siswa. Menurut [20], data mining dapat digunakan untuk menganalisis data siswa yang besar dan kompleks, termasuk data nilai, kehadiran, data demografis, serta aktivitas ekstrakurikuler untuk memberikan wawasan yang lebih baik tentang kebutuhan pendidikan siswa [21].

Salah satu penerapan utama dari data mining dalam pendidikan adalah untuk melakukan analisis minat dan bakat siswa. Dengan menganalisis berbagai data terkait minat akademik dan aktivitas ekstrakurikuler siswa, institusi pendidikan dapat memberikan rekomendasi yang lebih tepat mengenai jalur pendidikan yang sesuai dengan kemampuan dan potensi mereka. Sebagai contoh, KNN digunakan untuk mengklasifikasikan siswa berdasarkan pola nilai dan minat mereka, dan hasil analisis ini bisa digunakan untuk memberikan panduan yang lebih personal tentang mata pelajaran atau karier yang mungkin lebih cocok bagi siswa tersebut [22]

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan RapidMiner dengan kemampuannya untuk menyediakan antarmuka yang mudah digunakan serta berbagai fitur yang mendukung proses analisis data secara menyeluruh, mulai dari pengolahan data hingga evaluasi model.

A. Deskripsi Dataset

Dataset yang digunakan minat_bakat_siswa.csv dalam penelitian ini berisi data preferensi karier siswa yang terdiri dari beberapa atribut yang relevan untuk analisis. Dataset ini dirancang untuk mengidentifikasi pola yang dapat membantu memprediksi karier yang sesuai bagi siswa berdasarkan minat, bakat, dan faktor lainnya.

Row No.	Preferensi Karier	ID Siswa	Usia	Jenis Kelamin	Nilai Matem...	Nilai Bahasa...	Ekstrakurik...	Minat	Bakat
1	Programmer	S1	15	Perempuan	79	100	Klub IT	Teknologi	Programming
2	Programmer	S2	15	Perempuan	82	73	Klub IT	Teknologi	Programming
3	Insinyur	S3	15	Laki-laki	82	70	Olimpiade M...	Seni	Matematika
4	Insinyur	S4	15	Laki-laki	84	78	Olimpiade M...	Sosial	Matematika
5	Programmer	S5	15	Laki-laki	93	70	Klub IT	Olahraga	Programming
6	Dokter	S6	13	Laki-laki	83	85	Jurnalistik	Seni	Menulis
7	Penyanyi	S7	16	Laki-laki	75	70	Paduan Suara	Olahraga	Vokal
8	Insinyur	S8	15	Perempuan	93	89	Olimpiade M...	Sains	Matematika
9	Desainer Grafis	S9	15	Perempuan	100	98	Seni	Sains	Menggambar
10	Desainer Grafis	S10	14	Laki-laki	83	70	Seni	Sosial	Menggambar
11	Dokter	S11	14	Laki-laki	96	76	Jurnalistik	Sains	Menulis
12	Programmer	S12	15	Perempuan	78	96	Klub IT	Sains	Programming
13	Programmer	S13	14	Perempuan	70	100	Klub IT	Sains	Programming
14	Insinyur	S14	17	Perempuan	83	83	Olimpiade M...	Seni	Matematika
15	Programmer	S15	16	Perempuan	87	100	Klub IT	Seni	Programming

Gambar. 3 Collecting Dataset minat_bakat_siswa.csv

Pada gambar.3 yang ditampilkan merupakan sebuah dataset yang terdiri dari 5.000 baris data dengan 9 atribut, termasuk satu atribut target, yaitu "Preferensi Karier," yang merepresentasikan pilihan karier siswa seperti Programmer, Insinyur, Dokter, Penyanyi, dan lain-lain. Atribut lainnya meliputi "ID Siswa" sebagai identifikasi unik setiap siswa, "Usia" yang menunjukkan usia siswa dalam tahun, serta "Jenis Kelamin" yang mencatat jenis kelamin siswa, baik Perempuan maupun Laki-laki. Selain itu, terdapat atribut nilai akademik seperti "Nilai Matematika" dan "Nilai Bahasa," serta "Ekstrakurikuler," yang menggambarkan aktivitas yang diikuti siswa seperti Klub IT, Olimpiade Matematika, dan Jurnalisme. Dua atribut terakhir adalah "Minat," yang menunjukkan ketertarikan siswa terhadap bidang tertentu seperti Teknologi, Seni, atau Sains, dan "Bakat," yang mencerminkan kemampuan spesifik siswa seperti Programming, Menulis, atau Menggambar.

B. Penentuan Label

Dari beberapa atribut dikategorikan sebagai *polynomial* untuk data kategoris dan integer untuk data numerik. Tidak ada kesalahan dalam data, sehingga format data dianggap sesuai. Setelah memeriksa dan menyesuaikan jenis data, berikutnya adalah melakukan penentuan label Preferensi Karier ditentukan sebagai target yang akan dianalisis dan diprediksi.

	Nilai Mate...	Nilai Bahasa...	Ekstrakurik...	Minat	Bakat	Preferensi ...
	Integer	Integer	polynomial	polynomial	polynomial	polynomial label
1	79	100	Klub IT	Teknologi	Programming	Programmer
2	82	73	Klub IT	Teknologi	Programming	Programmer
3	82	70	Olimpiade Mate...	Seni	Matematika	Insinyur
4	84	78	Olimpiade Mate...	Sosial	Matematika	Insinyur
5	93	70	Klub IT	Olahraga	Programming	Programmer
6	83	85	Jurnalistik	Seni	Menulis	Dokter
7	75	70	Paduan Suara	Olahraga	Vokal	Penyanyi
8	93	89	Olimpiade Mate...	Sains	Matematika	Insinyur
9	100	98	Seni	Sains	Menggambar	Desainer Grafis
10	83	70	Seni	Sosial	Menggambar	Desainer Grafis
11	96	76	Jurnalistik	Sains	Menulis	Dokter

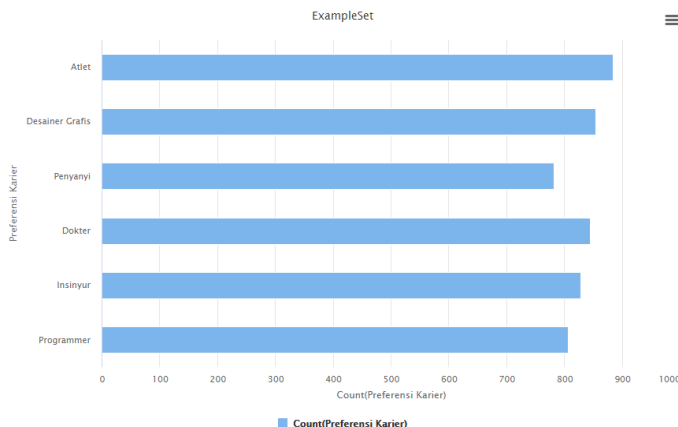
Gambar. 4 Penentuan Label pada Dataset Minat Bakat Siswa

Gambar. 4 menunjukkan proses penentuan label pada dataset minat dan bakat siswa menggunakan RapidMiner. Atribut Preferensi Karier (ditandai dengan kotak merah) ditetapkan sebagai label, yang berfungsi sebagai target klasifikasi dalam analisis. Label ini merupakan variabel kategoris yang mencerminkan pilihan karier siswa, seperti Programmer, Insinyur, Dokter, Desainer Grafis, dan lainnya. Atribut lain dalam dataset, seperti Nilai Matematika, Nilai Bahasa Inggris, Ekstrakurikuler, Minat, dan

Bakat, digunakan sebagai *features* (fitur) untuk membantu memprediksi label. Proses ini penting untuk memastikan bahwa algoritma pembelajaran mesin dapat dilatih dengan benar untuk mengklasifikasikan data berdasarkan atribut yang relevan.

C. Distribusi Data

Distribusi data dalam dataset dapat dilihat dari kolom statistik pada gambar. Berikut adalah analisis distribusi data dari Preferensi Karier yang merupakan feature target dalam model prediksi minat bakat siswa. Atribut ini merupakan data nominal dengan nilai yang paling sering muncul (modus) adalah Atlet (884 data) dan nilai yang paling sedikit adalah Penyanyi (782 data). Data ini menunjukkan distribusi yang relatif merata antara berbagai preferensi karier.



Gambar. 5 Grafik Distribusi data variabel Preferensi Karier

Grafik pada Gambar. 5 menunjukkan distribusi data variabel Preferensi Karier berdasarkan jumlah masing-masing kategori. Kategori-kategori tersebut mencakup Programmer, Insinyur, Dokter, Penyanyi, Desainer Grafis, dan Atlet. Grafik ini memperlihatkan bahwa jumlah data relatif merata di setiap kategori, dengan sedikit variasi dalam jumlah. Hal ini menunjukkan bahwa dataset memiliki representasi yang cukup seimbang untuk setiap preferensi karier, sehingga dapat mendukung analisis dan prediksi yang adil tanpa adanya bias signifikan terhadap salah satu kategori. Keseimbangan ini penting dalam memastikan akurasi dan reliabilitas model prediksi yang diterapkan.

Nominal values

Index	Nominal value	Absolute count	Fraction
1	Atlet	884	0.177
2	Desainer Grafis	854	0.171
3	Dokter	845	0.169
4	Insinyur	829	0.166
5	Programmer	806	0.161
6	Penyanyi	782	0.156

Close

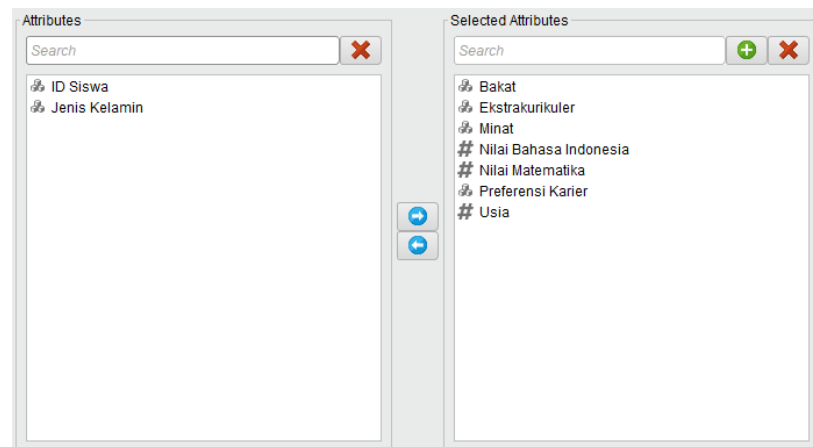
Gambar. 6 Distribusi data variabel Preferensi Karier

Hasil analisis pada distribusi data "Preferensi Karier" menunjukkan bahwa terdapat enam kategori utama dengan proporsi yang cukup merata. Kategori dengan jumlah tertinggi adalah Atlet dengan 884

data (17,7%), diikuti oleh Desainer Grafis sebanyak 854 data (17,1%), dan Dokter sebanyak 845 data (16,9%). Selanjutnya, kategori Insinyur memiliki 829 data (16,6%), disusul oleh Programmer dengan 806 data (16,1%), dan kategori dengan jumlah terendah adalah Penyanyi sebanyak 782 data (15,6%). Distribusi ini menunjukkan bahwa tidak ada kategori yang mendominasi secara signifikan, sehingga data memiliki representasi yang seimbang untuk setiap pilihan karier. Hal ini menjadi indikasi bahwa analisis lebih lanjut dapat dilakukan tanpa bias dominasi dari satu kategori tertentu.

D. *Feature Selection* (Pemilihan Fitur)

Feature Selection (Pemilihan Fitur) dalam konteks ini dilakukan untuk memilih atribut-atribut yang relevan dan signifikan untuk proses analisis data dan pembuatan model terlihat pada gambar.7.

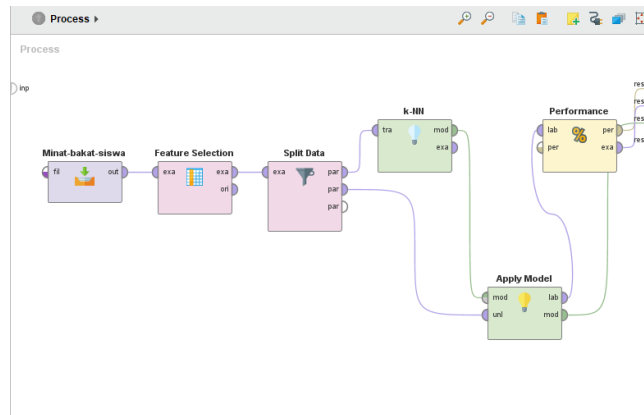


Gambar. 7 Proses *Feature Selection*

Pada proses *feature selection* yang ditunjukkan pada gambar, dilakukan pemilihan atribut untuk analisis lebih lanjut. *Atribut yang tidak relevan atau kurang signifikan, seperti ID Siswa dan Jenis Kelamin, telah dikeluarkan dari analisis karena tidak memberikan kontribusi langsung terhadap prediksi atau klasifikasi pada data "Preferensi Karier".*

E. Pembentukan Model Prediksi dengan K-NN

Dalam dunia pendidikan, memahami minat dan bakat siswa merupakan hal yang sangat penting untuk mendukung pengembangan potensi individu. Untuk mendukung proses ini, diterapkan sebuah model prediksi berbasis algoritma *k-Nearest Neighbors* (k-NN). Model ini dirancang untuk mengidentifikasi kecenderungan minat dan bakat siswa dengan menganalisis berbagai atribut yang relevan, seperti nilai akademik, usia, kegiatan ekstrakurikuler, dan preferensi karier. Dengan pendekatan berbasis data, model ini diharapkan dapat memberikan prediksi yang akurat dan mendukung pengambilan keputusan yang lebih baik dalam proses bimbingan dan konseling siswa. Berikut adalah tahapan yang dilakukan dalam membangun model ini terlihat pada gambar. 8.



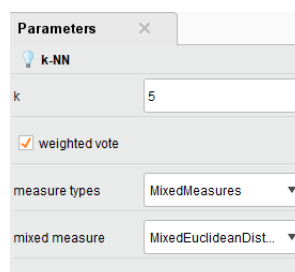
Gambar. 8 Model Prediksi Minat Bakat Dengan k-NN

Gambar. 8 menunjukkan alur proses pembangunan model prediksi minat dan bakat siswa menggunakan algoritma *k-Nearest Neighbors* (k-NN). Berikut penjelasan setiap komponen dalam diagram:

1. *Minat-bakat-siswa*: Komponen ini berfungsi sebagai input dataset yang berisi data terkait atribut siswa, seperti minat, bakat, nilai akademik, usia, dan preferensi karier.
2. *Feature Selection*: Tahapan ini digunakan untuk memilih atribut-atribut yang relevan dengan tujuan prediksi. Atribut yang tidak signifikan atau kurang relevan dikeluarkan untuk meningkatkan efisiensi dan akurasi model.
3. *Split Data*: Dataset dibagi menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (testing data). Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk mengevaluasi performa model.
4. *k-NN*: Komponen ini adalah algoritma inti yang digunakan untuk membangun model prediksi. k-NN bekerja dengan mencari k tetangga terdekat dalam ruang fitur untuk menentukan kategori atau nilai prediksi.
5. *Apply Model*: Setelah model dilatih, komponen ini menerapkan model yang telah dibangun pada data pengujian untuk menghasilkan prediksi.
6. *Performance*: Komponen ini digunakan untuk mengevaluasi performa model. Metrik evaluasi, seperti akurasi, precision, recall, dan F1-score, dihitung untuk menilai seberapa baik model dalam memprediksi minat dan bakat siswa.

Alur ini mencerminkan langkah-langkah sistematis dalam membangun, menguji, dan mengevaluasi model prediksi berbasis k-NN untuk mendukung proses pengambilan keputusan dalam konteks pendidikan.

Pengaturan parameter dalam algoritma k-NN untuk melakukan prediksi terhadap minat bakat siswa ditunjukkan pada gambar. 9.



Gambar. 9 Parameter Algoritma k-NN Dalam Model Prediksi Minat Bakat Siswa

Konfigurasi parameter pada algoritma *k-Nearest Neighbors* (k-NN) yang ditunjukkan pada gambar IV.10 digunakan untuk membangun model prediksi minat dan bakat siswa. Nilai $k = 5$ menunjukkan bahwa model akan mempertimbangkan lima tetangga terdekat untuk menentukan kategori atau kelas suatu data uji. Dengan menggunakan jumlah tetangga ini, model bertujuan untuk menangkap pola yang cukup representatif tanpa terlalu sensitif terhadap data *outlier* atau *noise*.

Aktivasi opsi *weighted vote* memberikan bobot yang lebih besar pada tetangga yang lebih dekat dengan data uji. Pendekatan ini memastikan bahwa keputusan klasifikasi lebih dipengaruhi oleh data yang memiliki jarak paling kecil, sehingga meningkatkan relevansi dalam prediksi. Selain itu, metode pengukuran jarak yang dipilih adalah *MixedEuclideanDist* dalam kategori *MixedMeasures*, yang sangat sesuai untuk dataset dengan atribut campuran (numerik dan kategorikal). Dengan metode ini, atribut numerik dihitung menggunakan jarak Euclidean, sementara atribut kategorikal diolah menggunakan metode jarak yang relevan, sehingga menghasilkan analisis yang lebih komprehensif.

F. Evaluasi Model

Evaluasi model dilakukan untuk mengevaluasi kinerja model klasifikasi *k-Nearest Neighbors* (k-NN) yang telah dibangun. Proses pengujian melibatkan pembagian data menjadi dua bagian utama, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun model, sementara data uji digunakan untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, digunakan metode pembagian data dengan proporsi tertentu, misalnya 80% untuk pelatihan dan 20% untuk pengujian, guna memastikan bahwa hasil evaluasi mencerminkan kemampuan model secara obyektif.

Metode evaluasi yang diterapkan dalam pengujian ini melibatkan penghitungan metrik-metrik kinerja seperti akurasi, *presisi* dan *recall*. Selain itu, *confusion matrix* digunakan untuk menganalisis performa model secara lebih rinci, termasuk kemampuan model dalam mengklasifikasikan masing-masing kelas. Pengujian ini bertujuan untuk menentukan apakah model k-NN mampu mengklasifikasikan data dengan benar sesuai dengan kelas target, seperti Programmer, Insinyur, Dokter, Penyanyi, Desainer Grafis, atau Atlet terlihat pada gambar. 10.

accuracy: 83.70%

	true Programmer	true Insinyur	true Dokter	true Penyanyi	true Desainer G...	true Atlet	class precision
pred. Programm...	138	7	8	5	7	9	79.31%
pred. Insinyur	5	134	5	3	6	5	84.81%
pred. Dokter	5	12	145	8	7	4	80.11%
pred. Penyanyi	3	5	2	129	3	6	87.16%
pred. Desainer ...	6	2	4	9	142	4	85.03%
pred. Atlet	4	6	5	2	6	149	86.63%
class recall	85.71%	80.72%	85.80%	82.69%	83.04%	84.18%	

Gambar. 10 *Confusion Matrix* dan Statistik Evaluasi Model k-NN

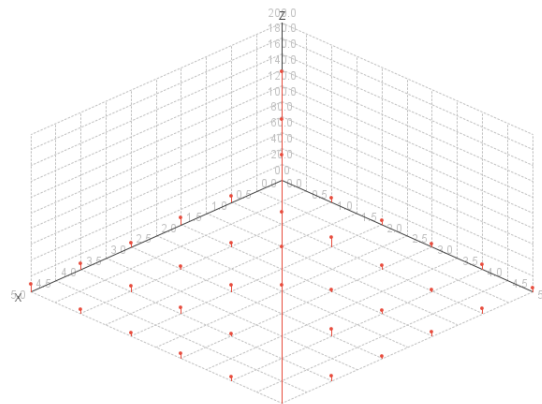
Berikut adalah hasil rekapitulasi pengujian model prediksi minat bakat siswa dapat dilihat pada tabel. 1

Tabel. 1 Rekapitulasi Pengujian Model Prediksi

No	Class	Presision (%)	Recall (%)
1	Programmer	79.31	85.71
2	Insinyur	84.81	80.72

3	Dokter	80.11	85.80
4	Penyanyi	87.16	82.69
5	Desainer Grafis	85.03	83.04
6	Atlet	86.63	84.18

Confusion Matrix (x: true class, y: pred. class, z: counters)



Gambar. 11 Grafik *Confusion Matrix* dan Statistik Evaluasi Model k-NN

Berdasarkan Grafik matriks konfusi (*confusion matrix*) pada gambar. 11 untuk hasil klasifikasi menggunakan model K-Nearest Neighbors (KNN) dengan akurasi sebesar 83.70%. Matriks ini menggambarkan performa model dalam memprediksi 6 kelas, yaitu Programmer, Insinyur, Dokter, Penyanyi, Desainer Grafis, dan Atlet. Dari hasil analisis, terlihat bahwa kelas Penyanyi memiliki performa terbaik dengan presisi sebesar 87.16%, yang menunjukkan bahwa model jarang salah memprediksi kelas lain sebagai Penyanyi. Di sisi lain, kelas Programmer memiliki presisi terendah sebesar 79.31%, yang mengindikasikan bahwa banyak sampel dari kelas lain yang salah diprediksi sebagai Programmer.

Pada metrik recall, kelas Dokter menonjol dengan nilai tertinggi sebesar 85.80%, yang menunjukkan bahwa model mampu mengenali sebagian besar sampel dari kelas ini dengan baik. Sebaliknya, kelas Insinyur memiliki recall terendah sebesar 80.72%, yang menunjukkan banyak sampel Insinyur salah diklasifikasikan ke kelas lain, seperti Programmer dan Dokter. Kesalahan prediksi ini mencerminkan adanya tumpang tindih pada fitur antar kelas tersebut.

Dari distribusi kesalahan, terlihat bahwa kelas Programmer sering diklasifikasikan sebagai Insinyur atau Dokter, sementara kelas Insinyur sering salah diklasifikasikan sebagai Dokter. Sebaliknya, kelas Penyanyi memiliki distribusi kesalahan yang paling kecil, menunjukkan bahwa karakteristik fitur kelas ini lebih unik dan mudah dikenali oleh model. Secara umum, kelas Desainer Grafis dan Atlet juga menunjukkan performa yang cukup baik dengan nilai presisi dan recall yang tinggi.

4. KESIMPULAN

Pemanfaatan k-NN untuk Klasifikasi Karier, Metode k-NN mampu mengklasifikasikan preferensi karier siswa berdasarkan atribut seperti usia, jenis kelamin, nilai akademik (Matematika dan Bahasa Indonesia), kegiatan ekstrakurikuler, minat, dan bakat. Model ini menunjukkan kemampuan dalam mengidentifikasi pola-pola relevan yang berkontribusi terhadap kecenderungan karier siswa. Dataset yang Komprehensif dan Relevan, Dataset minat_bakat_siswa.csv yang digunakan mencakup 5.000 entri

dengan 9 atribut penting, yang secara keseluruhan memberikan gambaran menyeluruh tentang karakteristik siswa. Proses pembersihan dan normalisasi data memastikan bahwa dataset dalam kondisi optimal untuk pelatihan dan pengujian model. *RapidMiner* sebagai Alat Analisis, Penggunaan *RapidMiner* terbukti efektif dalam mendukung proses analisis data, mulai dari import dataset, penentuan label (target variable), hingga pelatihan dan evaluasi model. Antarmuka visual yang intuitif mempermudah proses analisis bahkan bagi pengguna non-programmer. Evaluasi Model, Model k-NN dievaluasi menggunakan metrik akurasi, presisi, dan recall, yang memberikan gambaran menyeluruh mengenai performa prediksi. Secara umum, k-NN menunjukkan hasil klasifikasi yang cukup baik dalam memprediksi preferensi karier siswa berdasarkan atribut yang dianalisis. Penelitian ini memberikan kontribusi penting dalam bentuk sistem pendukung keputusan berbasis data yang dapat membantu siswa, guru BK, maupun institusi pendidikan dalam memberikan rekomendasi karier yang lebih terpersonalisasi dan berbasis data.

Berdasarkan hasil analisis dan klasifikasi risiko kredit berbasis data mining menggunakan algoritma XGBoost, diperoleh nilai akurasi sebesar 87%, presisi 82%, recall 93%, dan F1-Score 87%. Nilai akurasi yang cukup tinggi ini menunjukkan bahwa model mampu memprediksi risiko kredit dengan tingkat ketepatan yang baik. Tingginya nilai recall, terutama untuk kategori nasabah berisiko tinggi, mengindikasikan bahwa model efektif dalam mengidentifikasi calon debitur yang berpotensi gagal bayar, sehingga dapat membantu lembaga keuangan meminimalisasi kerugian. Namun, nilai presisi yang sedikit lebih rendah dibanding recall mengindikasikan adanya kemungkinan terjadinya false positive, yaitu nasabah yang sebenarnya layak mendapatkan kredit namun diklasifikasikan sebagai berisiko. Hal ini perlu menjadi perhatian karena dapat berdampak pada peluang bisnis yang hilang. Mengingat profil risiko nasabah dan kondisi ekonomi dapat berubah seiring waktu, evaluasi ulang model secara berkala diperlukan untuk memastikan performa tetap optimal. Model ini dapat diintegrasikan ke dalam *credit scoring system* lembaga keuangan, di mana hasil prediksi risiko kredit digunakan sebagai salah satu indikator untuk menyetujui atau menolak pengajuan kredit. Lakukan analisis terhadap nasabah yang termasuk dalam kategori FP dan FN untuk mengetahui pola kesalahan *model*. Hasil analisis ini dapat digunakan untuk perbaikan seleksi fitur dan *preprocessing data*. Gunakan *k-fold cross-validation* agar evaluasi model tidak bergantung pada satu set data uji saja, sehingga hasilnya lebih mencerminkan kemampuan model di berbagai skenario data.

UCAPAN TERIMA KASIH

Puji syukur penulis ucapkan kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini yang merupakan syarat untuk mendapatkan gelar sarjana Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama. Dan mengucapkan terimakasih yang sebesar-besarnya kepada Kaprodi Informatika Bapak Dr. Budi Triandi, M.Kom. dan Universitas Potensi Utama Medan yang membantu dalam penerbitan karya ilmiah ini.

REFERENSI

- [1] Huda, K., & Munastiwi, E. (2020). Strategi orang tua dalam mengembangkan bakat dan kreativitas anak di era pandemi Covid-19. *Jurnal Pendidikan Glasser*, 4(2), 80-87.

- [2] Rueangkit, P., Thaebanpakul, C., Sakboonyarat, B., & Prayote, A. (2024, December). Educational data mining: factors influencing medical student success and the exploration of visualization techniques. In *Frontiers in Education* (Vol. 9, p. 1390892). Frontiers Media SA.
- [3] López-Meneses, E., López-Catalán, L., Pelicano-Piris, N., & Mellado-Moreno, P. C. (2025). Artificial Intelligence in Educational Data Mining and Human-in-the-Loop Machine Learning and Machine Teaching: Analysis of Scientific Knowledge. *Applied Sciences*, 15(2), 772.
- [4] Mardiansa, M., Sari, H. L., & Prahasti, P. (2023). Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering. *Jurnal Multidisiplin Dehasen (MUDE)*, 2(4), 693-702.
- [5] Dhamawan, W. (2023). Algoritma K-Means Clustering Pada Pengelompokan Minat Bakat Siswa SMK PGRI 2 Karawang. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 10(1), 38-50.
- [6] Arofah, S. N., & Marisa, F. (2018). Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering. *JOINTECS (Journal Inf. Technol. Comput. Sci)*, 3(2), 85-90.
- [7] Aziza, L. N., Astuti, R. Y., Maulana, B. A., & Hidayati, N. (2024). Application of the K-Nearest Neighbor Algorithm for Food Security Classification in Central Java Province. *MALCOM Indones. J. Mach. Learn. Comput. Sci. J*, 4(2), 404-412.
- [8] Gupta, M., & Kharb, P. (2020). Application of Data Mining Techniques for Analyzing Student Interest and Talent. *International Journal of Educational Technology in Higher Education*, 17(1), 10-17.
- [9] Sharma, M., Jain, R., & Kumar, P. (2021). Application of KNN Algorithm for Interest and Talent Prediction in Students. *International Journal of Computer Applications*, 176(3), 38-43.
- [10] Zhang, Y., Liu, F., & Wang, Z. (2021). Improving Student Learning Outcomes with KNN Algorithm: A Case Study. *Educational Research Review*, 15(2), 112-120.
- [11] Liu, X., Zhang, L., & Yang, J. (2022). Exploring Student Preferences Using Data Mining Algorithms. *Journal of Educational Data Mining*, 14(4), 56-68.
- [12] Prakash, P., & Singh, A. (2023). Student Interest Classification Using K-Nearest Neighbor Algorithm. *Journal of Educational Research and Practice*, 21(1), 25-32.
- [13] Hasan, M., & Ahmed, S. (2023). Student Talent and Interest Classification Using KNN Algorithm. *International Journal of Educational Computing*, 9(2), 42-50.
- [14] Seth, S., Gupta, S., & Sharma, R. (2021). Educational Data Mining and its Application in Student Performance Analysis. *International Journal of Advanced Research in Education and Technology*, 8(2), 34-44.
- [15] Jain, S., Rajput, A., & Verma, S. (2020). Analysis of Student Data using Data Mining Techniques for Better Educational Decision Making. *Advances in Computer Science and Engineering*, 19(3), 45-60.
- [16] Chaurasia, S., & Pal, R. (2020). Data Mining Techniques in Education: A Review. *Journal of Data Science and Analytics*, 8(3), 210-220.
- [17] Zhang, Y., Liu, F., & Wang, Z. (2021). Improving Student Learning Outcomes with KNN Algorithm: A Case Study. *Educational Research Review*, 15(2), 112-120.
- [18] Sharma, M., Jain, R., & Kumar, P. (2022). Application of KNN Algorithm for Interest and Talent Prediction in Students. *International Journal of Computer Applications*, 176(3), 38-43.
- [19] Gupta, M., & Kharb, P. (2020). Application of Data Mining Techniques for Analyzing Student Interest and Talent. *International Journal of Educational Technology in Higher Education*, 17(1), 10-17.

- [20] Dastin, P., Shah, A., & Kumar, R. (2021). Educational Data Mining: A Survey of the Techniques and Applications. *International Journal of Educational Technology*, 15(2), 115-129.
- [21] Hasan, M., & Ahmed, S. (2023). Student Interest Classification Using K-Nearest Neighbor Algorithm. *Journal of Educational Research and Practice*, 21(1), 25-32.
- [22] Liu, X., Zhang, L., & Yang, J. (2022). Exploring Student Preferences Using Data Mining Algorithms. *Journal of Educational Data Mining*, 14(4), 56-68.