Jurnal Teknologi dan Komputasi
Vol.06, No.1, Maret 2024

https://journalversa.com/s/index.php/jtk

KESESUAIAN PEMILIHAN JURUSAN SISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA SMK PLUS AL MUSYARROFAH

Kamaluddin Mustofa[1], Denni Kurniawan[2]
[1,2]Universitas Budi Luhur
Email: 2111601627@student.budiluhur.ac.id[1], denni.kurniawan@budiluhur.ac.id[2]

Abstrak

Proses pemilihan jurusan merupakan tahap kritis bagi siswa karena dapat memengaruhi motivasi dan hasil belajar mereka selama bersekolah, terutama di Sekolah Menengah Kejuruan (SMK). Tantangan ini semakin signifikan dengan munculnya banyak sekolah baru di berbagai kota dan kabupaten di Indonesia, khususnya di Provinsi DKI Jakarta. Calon siswa sering kali memilih jurusan bukan berdasarkan minat pribadi, yang kemudian dapat berdampak pada penurunan nilai, terutama pada pelajaran produktif atau kompetensi tertentu. Untuk mengatasi masalah ini, diperlukan suatu sistem kesesuaian jurusan yang dapat memberikan rekomendasi berdasarkan kemampuan siswa melalui atribut-atribut tertentu. Dalam penelitian ini, dilakukan proses klasifikasi kesesuaian jurusan menggunakan metode Naive Bayes dan k-Nearest Neighbor dengan menggunakan data dari 238 siswa kelas sepuluh (X) tahun pelajaran 2023/2024, yang mencakup 9 atribut yang relevan. Proses pengujian dilakukan dengan komposisi data latih dan data uji sebanyak lima perbandingan, yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50. Hasil penelitian menunjukkan bahwa komposisi 80:20 memberikan hasil terbaik, dengan k-Nearest Neighbor mencapai tingkat recall, akurasi, dan presisi sebesar 100%. Di sisi lain, Naive Bayes Classifier menghasilkan tingkat recall sebesar 61%, dengan akurasi mencapai 73%. Hasil ini menunjukkan bahwa k-Nearest Neighbor lebih unggul dalam memprediksi kesesuaian jurusan dibandingkan dengan Naive Bayes pada kondisi tersebut.
Kata Kunci: Pemilihan Jurusan, Naive Bayes, K-Nearest Neighbor.

Abstract

The process of selecting a major is a critical stage for students because it can influence their motivation and learning outcomes while attending school, especially at Vocational High Schools (SMK). This challenge is becoming more significant with the emergence of many new schools in various cities and districts in Indonesia, especially in DKI Jakarta Province. Prospective students often choose majors not based on personal interests, which can then result in lower grades, especially in productive subjects or certain competencies. To overcome this problem, a major suitability system is needed that can provide recommendations based on student abilities through certain attributes. In this research, a department suitability classification process was carried out using the Naive Bayes and k-Nearest Neighbor methods using data from 238 tenth grade (X) students for the 2023/2024 academic year, which included 9 relevant attributes. The testing process was carried out with a composition of training data and test data in five comparisons, namely 90:10, 80:20, 70:30, 60:40, and 50:50. The research results show that the 80:20 composition provides the best results, with k-Nearest Neighbor achieving recall, accuracy and precision levels of 100%. On the other hand, the Naive Bayes Classifier produces a recall rate of 61%, with an accuracy of 73%. These results indicate that

k-Nearest Neighbor is superior in predicting major suitability compared to Naive Bayes in these conditions.
Keywords: *Major Selection, Naive Bayes, K-Nearest Neighbor.*

1. PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) di Indonesia, sejajar dengan Sekolah Menengah Atas (SMA), memiliki peran yang berbeda. Jika SMA lebih ditujukan untuk melanjutkan pendidikan ke tingkat universitas, SMK dirancang untuk mempersiapkan siswa untuk dunia kerja setelah lulus atau memberikan opsi melanjutkan pendidikan ke tingkat lebih tinggi. Dengan demikian, fokus pembelajaran di SMK lebih terkait dengan pengembangan keterampilan praktis dibandingkan dengan pemahaman pengetahuan umum seperti di SMA. Proses pemilihan jurusan di SMK dimulai sejak pendaftaran, berbeda dengan SMA yang seringkali dimulai pada semester 3 atau saat naik kelas XI. Kondisi ini dipengaruhi oleh peningkatan jumlah sekolah negeri dan swasta di setiap kota dan kabupaten, terutama di Provinsi DKI, yang menciptakan persaingan ketat dalam penerimaan di SMK. Informasi dari berbagai sumber pendidikan dan pengalaman peneliti menunjukkan bahwa ketidaksesuaian dalam pemilihan jurusan di SMK dapat berdampak pada fluktuasi nilai siswa di setiap semester. Hal ini menekankan pentingnya mendukung siswa dalam pemilihan jurusan yang sesuai dengan minat dan kemampuan mereka untuk meminimalkan ketidakcocokan dan meningkatkan kinerja akademis mereka.(Kusumadewi et al., 2020)

Pemilihan penjurusan yang sesuai akan meningkatkan minat dan memberikan kenyamanan seseorang dalam belajar. Dengan dasar kemampuan yang sama diharapkan dalam kegiatan pembelajaran dapat berjalan dengan lancar dan tidak mengalami kesulitan serta dapat meningkatkan minat serta prestasi belajar peserta didik. Sebaliknya, kurangnya minat belajar akibat kesalahan dalam memilih jurusan.(Istighfar et al., 2023)

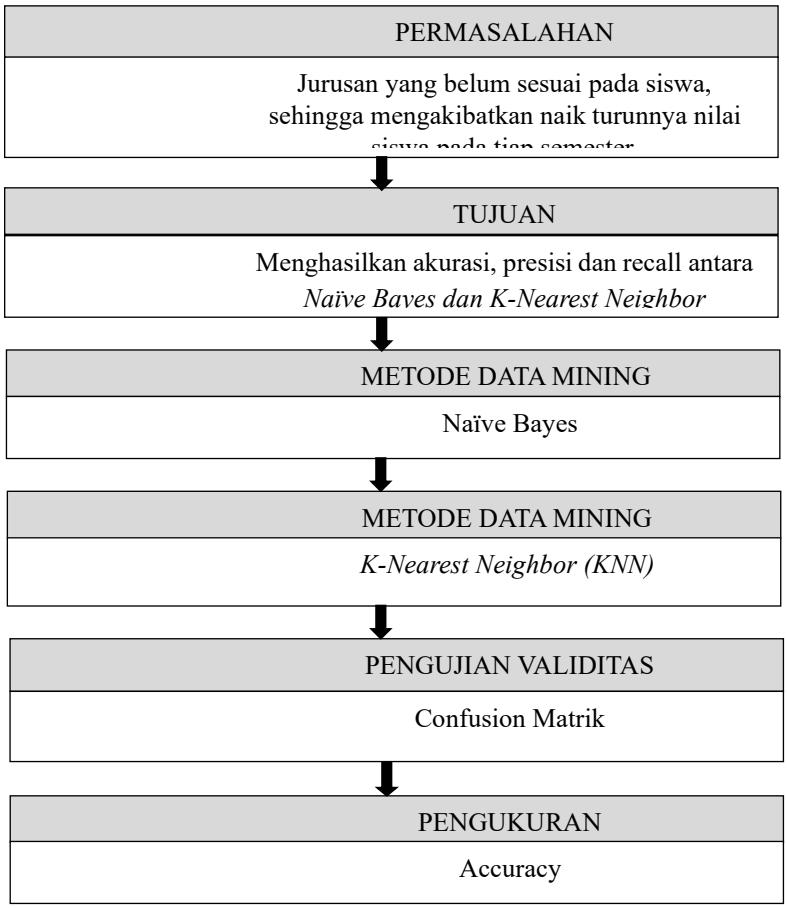
Berdasarkan informasi tersebut kemampuan dari siswa sendiri dapat membuat keputusan yang bertolak belakang dengan kemampuan siswa tersebut, beragam informasi mengenai jurusan SMK telah banyak tersedia di media cetak maupun di internet sehingga memberikan kemudahan untuk mendapatkan informasi tersebut. Akan tetapi, informasi yang tersedia hanya memberikan penjelasan secara umum, seperti: profil, biaya, lokasi, dan informasi umum lainnya. Dan informasi tersebut belum sepenuhnya membantu memberikan masukan mengenai

jurusan yang diinginkan oleh calon siswa sesuai dengan kemampuan, minat dan kesukaan calon siswanya.(Mohamad Andri Rasyid et al., 2023)

Algoritma yang akan digunakan dalam klasifikasi penjurusan di SMK Plus Al Musyarrofah adalah *k-Naerest Neighbors (K-NN)* dan *Naive Bayes.* Dikarenakan pada penelitian ini menggunakan klasifikasi yang dimana akan dibandingkan tingkat akurasi, presisi dan recall dari atribut yang dimiliki yaitu atribut yang akan digunakan untuk klasifikasi kesesuaian jurusan siswa yaitu, NISN, Nama, Nilai Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, Tes seleksi masuk dan nilai rata-rata keseluruhannya dan jurusan. Yang mana *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.(Fakhri et al., n.d.) Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Ilustrasi berikut ini memberikan gambaran perbedaan antara *accuracy* dan *precision*. (Lestari et al., 2019)Adapun untuk menentukan kesesuaiannya akan dilihat hasil dari Nilai Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, Tes seleksi masuk dan nilai rata-rata keseluruhannya turun ada kemungkinan siswa tersebut sudah merasa kalua jurusan yang ia pilih itu kurang sesuai dengan apa yang dia inginkan dan diharapkan pada awal pendaftaran.(Yudhi Putra & Ismiyana Putri, n.d.)

Kondisi saat ini, yang dihadapi pada lokasi peneliti setiap kali masuk semester 3 (tiga) atau 4 (empat) ada siswa yang nilainya naik turun dan mengakibatkan siswa tersebut cenderung merenung dan merasa kalau siswa tersebut salah memilih jurusan saat mendaftar dahulu sehingga nilainya tidak sesuai dengan keinginan dan pembelajaran atau cara belajarnyapun tidak focus pada materi yang diajarkan (Nuraeni et al., 2023).Sekolah sudah mencoba berbagai macam metode saat pembelajaran dan khusus siswa tersebut sering di aktifkan dalam kelas, tapi karena siswanya sudah merasa tidak sesuai dengan keinginannya atau salah jurusan yang berdampak pada tidak semangat mempelajari mata pelajaran dari jurusannya yang mengakibatkan Nilainya naik turun dalam setiap semester atau tiap ada tugas-tugas dari guru mata pelajarannya.(Sayhidin et al., 2023)

2. METODE PENELITIAN



Data Mining

Data mining adalah proses menggali informasi berharga dari dataset besar atau kompleks dengan menggunakan teknik analisis statistik, matematika, dan komputasi. Tujuannya adalah mengenali pola tersembunyi, koneksi, dan wawasan yang mendukung pengambilan keputusan, prediksi, dan pemahaman lebih mendalam tentang data.(Merawati, 2019) Seorang ahli data mining merupakan individu yang memiliki pengetahuan dan keterampilan dalam menerapkan teknik data mining untuk mengeksplorasi informasi dari dataset. Mereka memiliki pemahaman mendalam tentang statistik, matematika, algoritma komputasi, dan domain data yang diteliti. Ahli ini memanfaatkan berbagai alat dan teknik, seperti pengelompokan, klasifikasi, regresi, asosiasi, dan lainnya, untuk menganalisis data dengan tujuan menemukan pola yang berguna.(Widaningsih, 2019) Penerapan data mining sangat luas, ditemukan dalam berbagai sektor seperti bisnis, ilmu pengetahuan, kesehatan, keuangan, dan banyak lainnya. Di tengah pertumbuhan pesatnya data saat ini, keterampilan dalam menganalisis data secara efektif

melalui data mining semakin krusial untuk pengambilan keputusan yang lebih baik dan pemahaman yang lebih dalam terhadap tren yang ada. (Wulandari & Etikasari, 2019)

Klasifikasi

Klasifikasi merupakan elemen penting dalam ranah eksplorasi data. Teknik klasifikasi telah banyak dimanfaatkan dalam berbagai konteks penelitian. Klasifikasi merupakan pendekatan untuk mengelompokkan data yang melibatkan analisis data latih menggunakan algoritma pengklasifikasian. Terdapat beberapa algoritma klasifikasi yang umum digunakan, termasuk di antaranya Klasifikasi C4.5, Naïve Bayes, KNN, dan SVM (Muhidin & Casdi, 2019).

Klasifikasi melibatkan dua peran, yakni pembentukan model sebagai contoh yang akan diarsipkan untuk digunakan, serta penerapan model untuk mengidentifikasi, mengelompokkan, atau meramalkan objek data dalam format yang sederhana untuk disimpan. Aktivitas menemukan pola yang menggambarkan data penting disebut sebagai proses klasifikasi. Beberapa metode klasifikasi dalam eksplorasi data mencakup C4.5, Naïve Bayes, KNN, dan SVM (Indriyani et al., 2023).

Naïve Bayes

Klasifikasi Naïve Bayes adalah metode pengelompokan yang berasal dari prinsip Teorema Bayes. Teknik ini menggunakan konsep probabilitas dan statistik yang dikembangkan oleh ilmuwan Inggris, yaitu Thomas Bayes. Pendekatan ini fokus pada perkiraan probabilitas kejadian di masa depan berdasarkan pengalaman di masa lalu, oleh karena itu sering disebut sebagai Teorema Bayes. Salah satu karakteristik utama dari Klasifikasi Naïve Bayes adalah adanya asumsi yang sangat sederhana (naïf) tentang independensi antara setiap kondisi atau kejadian.(Hardoni et al., 2021) Salah satu keunggulan penggunaan metode ini adalah bahwa jumlah data pelatihan yang diperlukan relatif kecil untuk menentukan parameter-parameter yang dibutuhkan dalam proses klasifikasi. Karena metodenya mengasumsikan independensi antar-variabel, hanya diperlukan informasi tentang variasi suatu variabel dalam kelas tertentu untuk melakukan klasifikasi, bukan seluruh matriks kovariansi. (Rani, 2021)

$$P(Ci|X) = \frac{P(X|Ci)XP(Ci)}{P(X)} \qquad (2.1)$$

Keterangan :

X : Data menggunakan group yang belum diketahui

Ci : Hipotesis data X adalah suatu group spesifik

P(CiX) : Probabilitas Hipotesis Ci menurut syarat X

P(Ci) : Probabilitas Hipotesis Ci

P(XCi) : Probabilitas X menurut syarat tersebut

P(X) : Probabilitas menurut X

K-Nearest Neighbors

k-Nearest Neighbors melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik ***c*** pada ruang dimensi banyak. Data baru yang diklasifikasi selanjutnya diproyeksikan pada ruang dimensi banyak yang telah memuat titik-titik c data pembelajaran. Proses klasifikasi dilakukan dengan mencari titik ***c*** terdekat dari ***c-baru*** (*Nearest Neighbor*). Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak euclidean. Berikut beberapa formula yang digunakan dalam algoritma knn :

1) *EuclideanDistance*

   Jarak Euclidean adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi. Dapat dilihat pada Rumus dibawah ini.

   $$d = \sqrt{(x2 - x1) + (y2 - y1)^2} \qquad (2.2)$$

2) *HammingDistance*

   Jarak Hamming adalah cara mencari jarak antar 2 titik yang dihitung dengan panjang vektor biner yang dibentuk oleh dua titik tersebut dalam block kodebiner

3) *ManhattanDistance*

   Manhattan Distance atau Taxicab Geometri adalah formula untuk mencari jarak ***d*** antar 2 vektor ***p, q*** pada ruang dimensi ***n***

4) *Minkowski Distance*

   Minkowski distance adalah formula pengukuran antar 2 titik pada ruang vektor normal yang merupakan hibridisasi yang mengeneralisasi euclidean distance dan mahattan distance. Teknik pencarian tetangga terdekat disesuaikan dengan dimensi data, proyeksi, dan kemudahan implementasi oleh pengguna.

3. HASIL DAN PEMBAHASAN

Teknik pemilihan sampel yang digunakan pada penelitian ini adalah teknik sampling jenuh. Alasan sampling jenuh karena merupakan teknik pengambilan sampel dimana kita menggunakan semua anggota dari populasi menjadi sampel. karena data yang digunakan adalah data seluruhnya dan jumlahnya tidak terlalu besar yaitu sebanyak 238 sample data siswa-siswi kelas sepuluh (X) tahun pelajaran 2023/2024 pada sekolah menengah kejuruan SMK Plus Al Musyarrofah.

M1

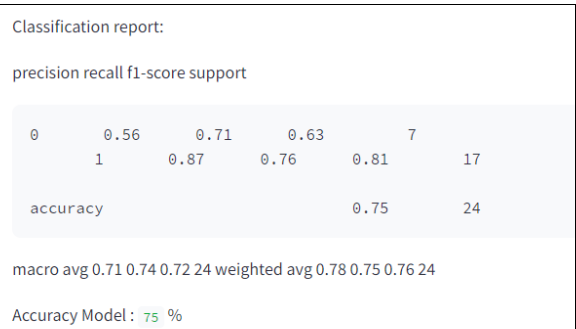
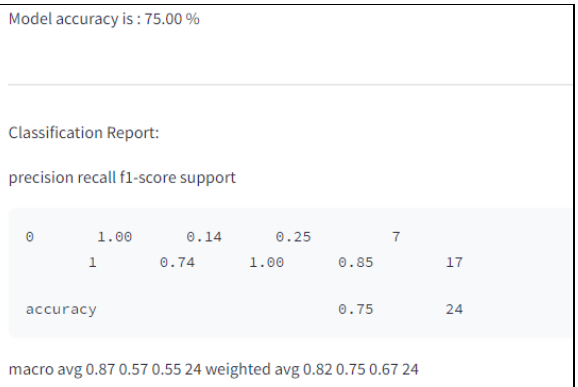
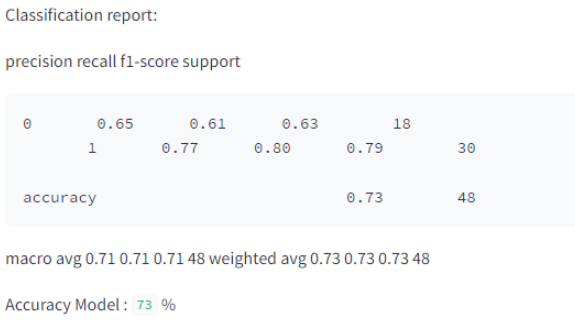
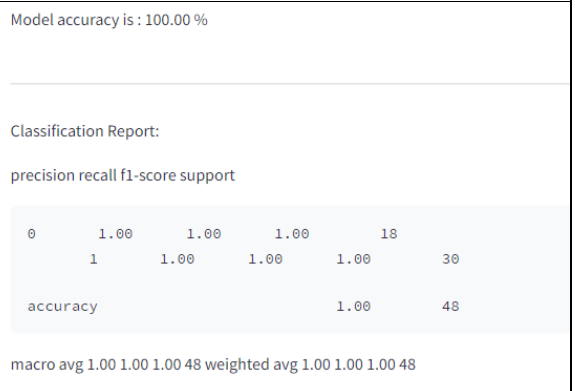
	A	B	C	D	E	F	G	H	I	J
1	NO	NISN	INISIAL NAMA	B.INDO	B.INGG	MATEMATIKA	IPA	TES SELEKSI	RATA-RATA	JURUSAN
2	1	******9742	DPL	80	76	72	84	80	78,4	TKJ
3	2	******1586	MNA	82	68	70	80	82	76,4	BDP
4	3	******8103	SPP	78	82	76	78	88	80,4	TKJ
5	4	******0975	MRS	84	80	72	78	78	78,4	TKJ
6	5	******7299	MAA	84	82	76	72	74	77,6	BDP
7	6	******9800	ALS	86	78	74	68	70	75,2	BDP
8	7	******0121	RFR	82	80	74	78	80	78,8	TKJ
9	8	******0412	ARP	82	74	76	78	80	78	TKJ
10	9	******5150	ARM	74	78	84	76	78	78	TKJ
11	10	******9166	REP	86	74	72	76	78	77,2	BDP
12	11	******4790	DAP	78	80	78	74	76	77,2	BDP
13	12	******4853	GOA	80	76	70	64	66	71,2	BDP
14	13	******4462	RMP	86	78	76	78	80	79,6	TKJ
15	14	******3648	MHH	82	74	70	78	80	76,8	BDP
16	15	******6978	AMM	80	80	80	80	82	80,4	TKJ
17	16	******1612	RFF	84	82	82	76	78	80,4	TKJ
18	17	******7147	LMK	86	78	70	76	78	77,6	BDP
19	18	******1667	ARH	70	78	70	70	72	72	BDP
20	19	******8042	BGG	68	74	76	74	76	73,6	BDP
21	20	******7738	RAA	80	74	72	82	84	78,4	TKJ
22	21	******4730	INN	80	80	82	60	62	72,8	BDP
23	22	******4823	HPP	80	76	80	68	60	72,8	BDP
24	23	******6307	YAA	78	78	80	80	82	79,6	TKJ
25	24	******1040	MFS	80	76	80	70	72	75,6	BDP
26	25	******3108	FAF	84	74	76	70	82	77,2	BDP
27	26	******5684	AMM	80	84	68	68	70	74	BDP

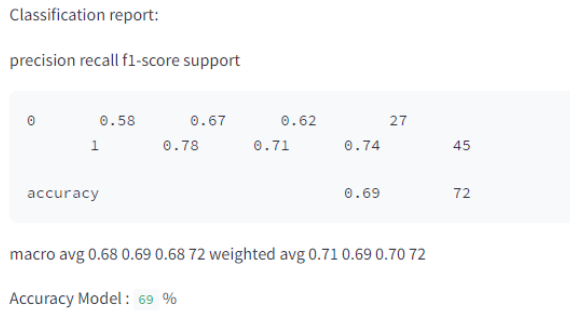
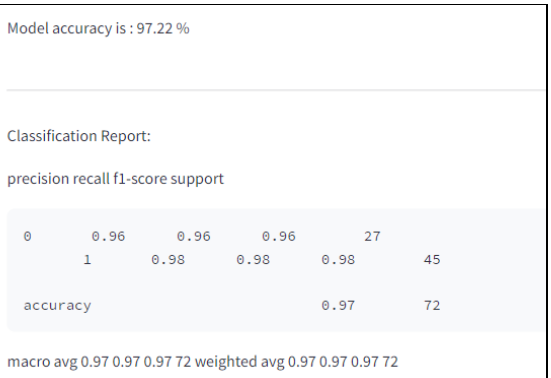
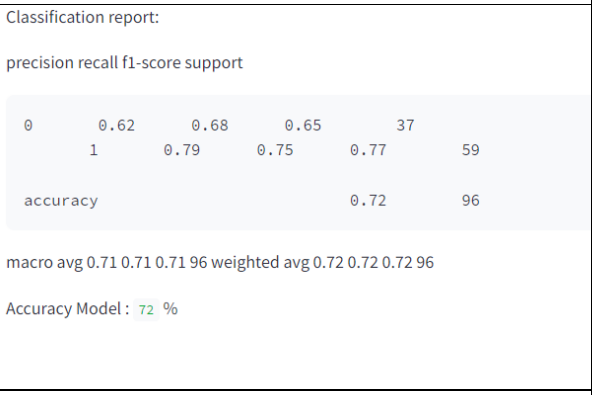
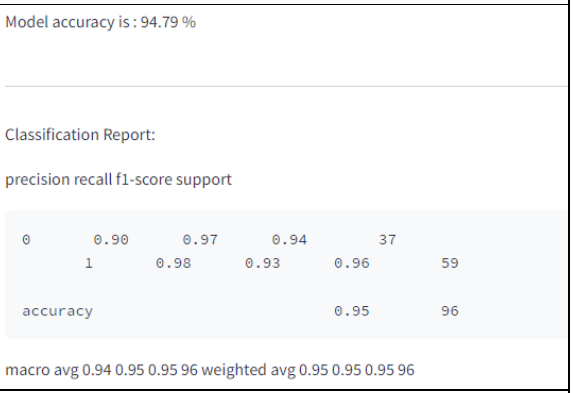
DATA TESIS BARU DATA SISWA

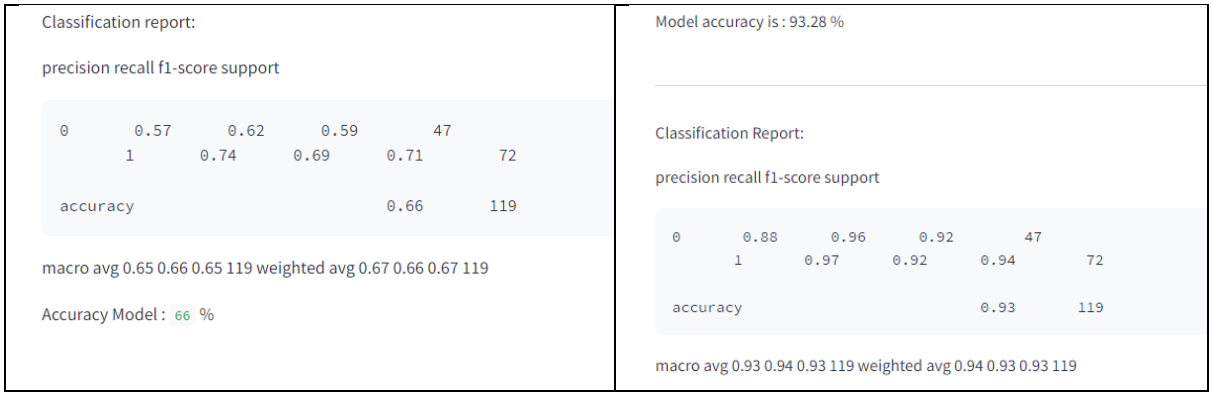
Ready

tabel 1 Dataset yang digunakan pada penelitian saat ini

Hasil dari percobaan awal dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 90:10, yang terdiri dari 304 data latih dan 34 data uji menunjukkan bahwa algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 75%. Selain itu, terdapat pencapaian nilai presisi sebesar 71%, recall sebesar 74%, dan nilai f1-score mencapai 72%, sebagaimana tergambar pada gambar berikut:	Hasil dari percobaan awal dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 90:10, yang terdiri dari 304 data latih dan 34 data uji menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) berhasil mencapai tingkat akurasi sebesar 75%. Selain itu, terdapat pencapaian nilai presisi sebesar 87%, recall sebesar 57%, dan nilai f1-score mencapai 55%, sebagaimana tergambar pada gambar berikut:

	
Hasil dari percobaan kedua dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 80:20, yang terdiri dari 270 data latih dan 68 data uji menunjukkan bahwa algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 73%. Selain itu, terdapat pencapaian nilai presisi sebesar 71%, recall sebesar 71%, dan nilai f1-score mencapai 71%, sebagaimana tergambar pada gambar berikut:	Hasil dari percobaan kedua dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 80:20, yang terdiri dari 270 data latih dan 68 data uji menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) berhasil mencapai tingkat akurasi sebesar 100%. Selain itu, terdapat pencapaian nilai presisi sebesar 100%, recall sebesar 100%, dan nilai f1-score mencapai 100%, sebagaimana tergambar pada gambar berikut:
	
Hasil dari percobaan ketiga dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 70:30, yang terdiri dari 237 data latih dan 101 data uji menunjukkan bahwa algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 69%. Selain itu, terdapat pencapaian nilai presisi sebesar 68%, recall sebesar 69%, dan nilai f1-score mencapai 68%, sebagaimana tergambar pada gambar berikut	Hasil dari percobaan ketiga dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 70:30, yang terdiri dari 237 data latih dan 101 data uji menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) berhasil mencapai tingkat akurasi sebesar 97%. Selain itu, terdapat pencapaian nilai presisi sebesar 97%, recall sebesar 97%, dan nilai f1-score mencapai 97%, sebagaimana tergambar pada gambar berikut:

	
Hasil dari percobaan keempat dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 60:40, yang terdiri dari 203 data latih dan 135 data uji menunjukkan bahwa algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 72%. Selain itu, terdapat pencapaian nilai presisi sebesar 71%, recall sebesar 71%, dan nilai f1-score mencapai 71%, sebagaimana tergambar pada gambar berikut:	Hasil dari percobaan keempat dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 60:40, yang terdiri dari 203 data latih dan 135 data uji menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) berhasil mencapai tingkat akurasi sebesar 95%. Selain itu, terdapat pencapaian nilai presisi sebesar 94%, recall sebesar 95%, dan nilai f1-score mencapai 95%, sebagaimana tergambar pada gambar berikut:
	
Hasil dari percobaan kelima dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 50:50, yang terdiri dari 169 data latih dan 169 data uji menunjukkan bahwa algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 66%. Selain itu, terdapat pencapaian nilai presisi sebesar 65%, recall sebesar 66%, dan nilai f1-score mencapai 65%, sebagaimana tergambar pada gambar berikut:	Hasil dari percobaan kelima dengan total 338 sampel data dan komposisi data latih dan data uji sebesar 50:50, yang terdiri dari 169 data latih dan 169 data uji menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) berhasil mencapai tingkat akurasi sebesar 93%. Selain itu, terdapat pencapaian nilai presisi sebesar 93%, recall sebesar 94%, dan nilai f1-score mencapai 93%, sebagaimana tergambar pada gambar berikut:



A. Analisis Hasil Pengujian

Berdasarkan hasil uji coba perbandingan komposisi antara data latih dan data uji diatas, dapat diketahui bahwa algoritma K-Nearest Neighbors (KNN) memberikan nilai akurasi tertinggi dibandingkan dengan algoritma Naïve Bayes di masing-masing algoritma dengan hasil perbandingan sebagai berikut:

Tabel 4.8 Hasil Percobaan Perbandingan Data

Percobaan Perbandingan Data	Akurasi	
	Naïve Bayes	K-NN
90:10	**75%**	75%
80:20	73%	**100%**
70:30	69%	97%
60:40	72%	95%
50:50	66%	93%

Dari kelima pengujian diatas dapat diketahui bahwa algoritma yang terbaik adalah algoritma K-Nearest Neighbors (KNN) dengan komposisi data latih dan data uji 80:20. Jika dilihat dari perbandingan nilai presisi, *recall,* dan *f1-score* diperoleh hasil sebagaimana tabel berikut:

Tabel 4.9 Hasil Pengujian Terbaik

		Percobaan Perbandingan Data				
Algoritma		90:10	**80:20**	70:30	60:40	50:50
Naïve Bayes	*Precision*	71%	**71%**	68%	71%	65%
	Recall	74%	**71%**	69%	71%	66%

	F1-Score	72%	**71%**	68%	71%	65%
K-NN	*Precision*	87%	**100%**	97%	94%	93%
	Recall	57%	**100%**	97%	95%	94%
	F1-Score	55%	**100%**	97%	95%	93%

Dari data hasil pengujian diatas didapat hasil bahwa algoritma K-Nearest Neighbors (KNN) memiliki nilai akurasi tertinggi pada komposisi pembagian data latih dan data uji 80:20, dimana dengan kelima percobaan yang telah di lakukan, nilai akurasi tertinggi yang didapatkan sebesar 100%, nilai presisi tertinggi sebesar 100%, *recall* sebesar 100%, dan *f1-score* sebesar 100%.

B. Implikasi Penelitian

Berdasarkan penelitian yang telah dilakukan dapat memberikan implikasi penelitian terhadap beberapa aspek diantaranya adalah aspek sistem, aspek manajerial, dan aspek penelitian selanjutnya.

1) Aspek Sistem

Untuk menerapkan hasil penelitian dibutuhkan dukungan sistem yang baik, sehingga pihak yang berkepentingan dapat menggunakan hasil penelitian. Oleh karena itu dibutuhkan sarana dan prasarana yang memadai baik dari sisi hardware dan software yang digunakan. Dalam penelitian ini memiliki minimal spesifikasi sebagai berikut:

Tabel 4.1 Kebutuhan Software

Software	Minimum *Requirement*
Sistem Operasi	Microsoft Windows 2010
Visual studio code	Version 1.86
Pemograman python	Streamlit

Adapun spesifikasi minimum *hardware* yang dibutuhkan untuk mendukung sistem *data mining* dapat dilihat pada tabel berikut:

Tabel 4. 2 Kebutuhan Hardware

Hardware	Minimum *Requirement*
Processor	Core i7
Memory	16 GB
Harddisk	256

2) Aspek Manajerial

Berdasarkan hasil pengukuran dan evaluasi menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) akurat dalam pengklasifikasian jurusan siswa sehingga algoritma ini dapat memberikan solusi bagi pihak sekolah dalam menentukan Jurusan siswa. kontribusi pada penelitian ini adalah penambahan 5 atribut yang meliputi: rata-rata nilai tes masuk, nilai Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA.

3) Aspek Penelitian Selanjutnya

Penelitian ini masih memiliki kekurangan dan keterbatasan, guna acuan penelitian selanjutnya dapat mempertimbangkan hal-hal sebagai berikut:

1. Mengkaitkan aspek-aspek yang belum dibahas dalam penelitian ini sehingga didapatkan penelitian yang lebih dalam.
2. Hasil dari penelitian ini perlu diterapkan dengan menggunakan metode algoritma yang lain serta dikembangkan dengan algoritma optimasi lainnya.
3. Penelitian ini juga dapat dikembangkan setiap tiga sampai empat tahun sekali agar dapat menyesuaikan dengan kondisi dan situasi yang ada.

4. KESIMPULAN

Dari identifikasi masalah yang telah dijelaskan pada latar belakang, dapat disimpulkan bahwa masalah utama adalah kecenderungan calon siswa dalam memilih jurusan tanpa mempertimbangkan minat pribadi, yang kemudian berdampak negatif pada penurunan nilai, terutama pada pelajaran produktif atau kompetensi tertentu.

a. Penelitian ini menghasilkan temuan bahwa melalui evaluasi akurasi dengan menggunakan data dari 238 siswa-siswi kelas sepuluh pada tahun pelajaran 2023/2024, model dengan perbandingan data uji 80:20 memberikan hasil terbaik. Pada perbandingan tersebut, metode k-Nearest Neighbor menunjukkan kinerja yang unggul dengan mencapai recall, akurasi, dan presisi sebesar 100%. Sebaliknya, Naïve Bayes Classifier memiliki recall sebesar 61% dan akurasi 73%. Dengan temuan ini, dapat diakui bahwa k-Nearest Neighbor lebih efektif dalam mengidentifikasi dan mengklasifikasikan data, memberikan kinerja yang lebih superior dibandingkan dengan Naïve Bayes Classifier dalam konteks penelitian ini.

b. Kesimpulan dari hasil penelitian ini menunjukkan bahwa penerapan model k-Nearest Neighbor memiliki potensi sebagai solusi untuk membantu calon siswa dalam memilih jurusan berdasarkan minat pribadi mereka. Diharapkan, hal ini dapat meningkatkan prestasi akademis mereka dan memberikan arah yang lebih tepat dalam pengembangan karir di masa depan.

DAFTAR PUSTAKA

Bantuan, P., Bpjs, I., Fajar, K., Putro, S., Utami, E., Hartanto, A. D., Yogyakarta, A., Disetujui, D. D., Kunci, K., Pbi, :, Bayes, N., & Pso, D. (2022). Klasifikasi Neive Bayes Berbasis Particle Swarm Optimization Untuk Prediksi. *Journal Computer Science*, *1*(1).

Brilliant, M., Nurhasanah, I. A., & Rahmadaniah, D. (n.d.). *Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan (Study Kasus Smks Pgri 2 Pringsewu)*.

Fakhri, J., Sunge, A. S., Zy, A. T., & Pelita Bangsa, U. (n.d.). *Perancangan Klasifikasi Algoritma Naive Bayes Pada Data Pemilihan Jurusan Siswa* (Vol. 11, Issue 2).

Hardoni, A., Rini, D. P., & Sukemi, S. (2021). Integrasi SMOTE Pada Naive Bayes Dan Logistic Regression Berbasis Particle Swarm Optimization Untuk Prediksi Cacat Perangkat Lunak. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(1), 233.

Homepage, J., A'yuniyah, Q., & Reza, M. (n.d.). *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru*.

Indriyani, S., Fatchan, M., & Firmansyah, A. (2023). Prediksi Harga Logam Mulia Dengan Pendekatan Algorithma Naïve Bayes Dan Pso. In *JINTEKS* (Vol. 5, Issue 1).

Istighfar, F., Negara, A. B. P., & Tursina, T. (2023). Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *11*(1), 77.

Kusumadewi, V. A., Cholissodin, I., & Adikara, P. P. (2020). *Klasifikasi Jurusan Siswa menggunakan K-Nearest Neighbor dan Optimasi dengan Algoritme Genetika (Studi Kasus: SMAN 1 Wringinanom Gresik)* (Vol. 4, Issue 4).

Lestari, B. A., Hasbi, M., & Susyanto, T. (2019). Pemilihan Sekolah Terbaik Dengan Menggunakan Metode K-Nearest Neighbors Dan Taxonomic Matcher. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *6*(2).

Maulana, G. (2023). Penerapan Algoritma Machine Learning Untuk Penjurusan Siswa Sekolah Menengah Kejuruan Berdasarkan Nilai Raport dan Psikotest. *Jurnal Ilmu Teknik Dan Komputer*, *07*(01), 56.

Merawati, D. (2019). Penerapan Data Mining Penentu Minat Dan Bakat Siswa Smk Dengan Metode C4.5. *JURNAL ALGOR*, *1*(1).

Mohamad Andri Rasyid, R. K., Riyanto, A., & Widyawati, R. (2023). Implementasi Algoritma Naïve Bayes Untuk Sistem Rekomendasi Pemilihan Fakultas Di Universitas Amikom Yogyakarta Implementation Of The Naïve Bayes Algorithm For The Faculty Selection Recommendation System At Universitas Amikom Yogyakarta. In *JIKOM: Jurnal Informatika dan Komputer* (Vol. 13, Issue 1).

Muhabatin, H., Prabowo, C., Ali, I., Lukman Rohmat, C., Rizki Amalia, D., sitasi, C., & Rizki, D. (2021). Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO. *Informatics for Educators and Professionals*, *5*(2), 156–165.

Muhidin, A., & Casdi, M. (2019). *SIGMA-Jurnal Teknologi Pelita Bangsa Optimasi Algoritma Naïve Bayes Berbasis Particle Swarm Optimization (Pso) Dan Stratified Untuk Meningkatkan Akurasi Prediksi Penyakit Diabetes* (Vol. 10).

Novaldy, F., & Herliana, A. (2021). Penerapan Pso Pada Naïve Bayes Untuk Prediksi Harapan Hidup Pasien Gagal Jantung. *JURNAL RESPONSIF*, *3*(1), 37–43.

Nuraeni, S., Syam, S. P. A., Wajdi, M. F., Firmansyah, B., & Malkan, M. (2023). Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari. *G-Tech: Jurnal Teknologi Terapan*, *7*(1), 89–95.

Pambudi, A., & Abidin, Z. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *JDMSI*, *4*(1), 1–14.

Rani, H. A. D. (2021). Optimasi Particle Swarm Optimization Pada Naïve Bayes Untuk Prediksi Kondisi Kelahiran Bayi. *Jurnal Dialektika Informatika (Detika)*, *2*(1), 28–33.

Rifai, A., Aulianita, R., Stmik,), Jakarta, N. M., & Jakartai, N. M. (n.d.). Komparasi Algoritma Klasifikasi C4.5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk

Penentuan Resiko Kredit. In *Journal Speed-Sentra Penelitian Engineering dan Edukasi* (Vol. 10). CDROM.

Saepudin, S., Muslih, M., Studi Sistem Informasi, P., Nusa Putra, U., Raya Cibolang Kaler No, J., & Sukabumi, K. (2019). Pemilihan Jurusan Dengan Metode K-Nearest Neighbor Untuk Calon Siswa Baru. In *Jurnal Rekayasa Teknologi Nusa Putra* (Vol. 5, Issue 2).

Sayhidin, D., Haris, G., & Juliane, C. (2023). *Implementasi Data Mining Tingkat Kepemimpinan Siswa dengan K-Nearest Neighbor, Decision Tree, dan Naïve Bayes*. *7*(1), 199–206.

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25.

Widiastuti, N. A., Azhar, M., & Mulyo, H. (2023). Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Pada Peserta Didik Baru. *Jurnal SIMETRIS*, *14*(2).

Wulandari, N., & Etikasari, P. (2019). Analisis Minat Belajar Siswa Pada Lembaga Pendidikan Pendidikan Indonesia Amerika Perumnas 3 Bekasi Dengan Metode C4.5. In *Jurnal Rekayasa Informasi* (Vol. 8, Issue 1).

Yudhi Putra, M., & Ismiyana Putri, D. (n.d.). *Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI* (Vol. 16, Issue 2).

Yusuf, D., Mubarak, Y., Pangesti, A. R., Wulansari, N., & Zulqornain, R. (2023). Perbandingan Metode Naive Bayes Classifier Dan Decision Tree C4.5 Dalam Mencari Pola Minat Pemilihan Jurusan Di Madrasah Aliyah (Studi Kasus:MA El-Bayan Majenang). In *Journal Sistem Informasi, dan Teknologi Informasi* (Vol. 2, Issue 1).