Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

ANALISIS PERBANDINGAN K-*NEAREST NEIGHBORS* DAN *NAIVE BAYES* UNTUK REKOMENDASI PILIHAN PROGRAM STUDI BAGI MAHASISWA

Nurwati[1*], Yudi Santoso[2]
[1,2]Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
Email: [1*]nurwati@budiluhur.ac.id, [2]yudi.santoso@budiluhur.ac.id
(* : coressponding author)

Abstrak- Kesulitan memilih program studi bagi lulusan Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK) merupakan tantangan besar bagi alumni. Tantangan ketidakpastian karir di masa depan yang sering kali menjadikan proses pemilihan program studi semakin rumit. Tantangan lainnya, banyak alumni siswa SMA belum sepenuhnya memahami minat bakat yang dimiliki. Berbeda dengan lulusan SMK yang telah mendapatkan pendidikan dan pengalaman kerja magang selama di sekolah. Tekanan orang tua, faktor biaya, keterbatasan informasi yang dimiliki alumni mengenai program studi yang diimpikan, serta memilih antara minat dan peluang karir merupakan faktor yang dipertimbangkan bersama orang tua. Dengan demikian, permasalahan penelitian ini adalah tantangan apa saja yang dihadapi oleh lulusan SMA juga SMK dalam memperoleh informasi yang memadai mengenai program studi di perguruan tinggi impian serta bagaimana lulusan mengatasi kebingungan dalam menentukan program studi yang cocok dengan bakat, minat serta prospek kerjanya. Untuk mengatasi hal ini tujuan penelitian memberikan rekomendasi untuk calon mahasiswa dalam menentukan program studi yang cocok berdasarkan latar belakang akademik dan kemampuan calon mahasiwa yang dimiliki. Rekomendasi program studi ini menggunakan metode *text mining* dengan membandingkan hasil nilai akurasi antara algoritma *K-Nearest Neighbors* (KNN) dengan algoritma Naive Bayes. Perbandingan algoritma ini mendukung pengambilan keputusan. Data set yang digunakan data mahasiswa tiga angkatan total 347 data. Selanjutnya, data dibagi menjadi data latih dan data uji. Akurasi metode KNN tercatat sebesar 81,16% dengan nilai K=2 dan proporsi data uji sebesar 40%. Akurasi Naive Bayes mencapai 82,61% program studi Teknik Informatika. Hasil akurasi tidak menunjukkan perbedaan yang signifikan, namun metode Naive Bayes menghasilkan akurasi yang lebih tinggi dibandingkan KNN.

Kata Kunci: Akurasi, *K-Nearest Neighbors (KNN)*, *Naive bayes*, Program Studi, Rekomendasi

Abstract- *The difficulty of choosing a study programme for graduates of Senior High School (SMA) and Vocational High School (SMK) is a big challenge for alumni. The challenge of career uncertainty in the future often makes the process of choosing a study programme even more complicated. Another challenge is that many high school alumni do not fully understand their interests and talents. In contrast to SMK graduates who have received education and internship work experience while at school. Parental pressure, cost factors, limited information that alumni have about their dream study programme, and choosing between interests and career opportunities are factors that are considered with parents. Thus, the problem of this research is what challenges are faced by high school graduates as well as vocational school graduates in obtaining adequate information about study programmes in dream universities and how graduates overcome confusion in determining study programmes that match their talents, interests and job prospects. To overcome this, the research aims to provide recommendations for prospective students in determining suitable study programmes based on their academic background and abilities. This study programme recommendation uses text mining methods by comparing the results of the accuracy value between the K-Nearest Neighbors (KNN) algorithm and the Naive Bayes algorithm. This algorithm comparison supports decision making. The data set used is three batches of student data totalling 347 data. Furthermore, the data is divided into training data and test data. The accuracy of the KNN method was recorded at 81.16% with a value of K = 2 and a proportion of test data of 40%. Naive Bayes accuracy reached 82.61% of Informatics Engineering study programme. The accuracy results do not show a significant difference, but the Naive Bayes method produces higher accuracy than KNN.*

***Keywords**: Accuracy, K-Nearest Neighbors (KNN), Naive Bayes, Recommendation, Study Program*

1. PENDAHULUAN

Salah satu peran teknologi pada pendidikan yang sangat penting adalah menghasilkan lulusan perguruan tinggi yang cepat mendapatkan pekerjaan setelah lulus kuliah. Ini merupakan salah satu cara promosi yang efektif untuk menarik perhatian dan mempengaruhi lulusan SMA atau SMK agar memilih perguruan tinggi tersebut. Setelah lulus sekolah, banyak calon mahasiswa yang masih merasa bingung dan ragu dalam menentukan jurusan kuliah yang sesuai dengan bakat dan minat mereka. Diantara keraguan tersebut banyak lulusan SMA atau SMK adalah belum mengenal bakat dan minat yang dimiliki, masalah biaya kuliah, tekanan orang tua dan lingkungan, ketidakpastian peluang pekerjaan dimasa depan atas program studi yang dipilih, serta minimnya informasi yang dimiliki lulusan SMA atau SMK atas pilihan program studi dan lulusannya. Berdasarkan Undang-Undang Republik Indonesia Tahun 2003 Nomor 20 pasal 1 ayat 8 tentang Sistem Pendidikan Nasional [1], jenjang pendidikan mengacu pada tahapan pendidikan yang ditentukan sesuai dengan tingkat perkembangan peserta didik, tujuan yang ingin dicapai, dan kemampuan yang akan dikembangkan [1].

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Melalui peraturan Undang-Undang tersebut diharapkan calon mahasiswa tidak mengalami kebingungan atas pilihan program studi. Kebingungan dalam memilih program studi di Universitas mengakibatkan tidak sesuai dengan latar belakang akademis mahasiswa dan kemampuan yang dimiliki. Akibatnya, banyak mahasiswa yang mengakhiri masa perkuliahan lebih awal atau pindah program studi di tengah-tengah perkuliahan [2].

Dengan demikian, masalah yang diidentifikasi adalah tantangan apa saja yang dihadapi oleh lulusan SMA dan SMK dalam memperoleh informasi yang memadai mengenai program studi kuliah yang ada, serta bagaimana cara mereka mengatasi kebingungan dalam memilih program studi yang sesuai dengan bakat, minat, dan prospek pekerjaan.

Untuk menyediakan informasi terkait program studi dan rekomendasi mata pelajaran yang dipersiapkan agar mendapat nilai bagus, maka dibuat penelitian rekomendasi pilihan program studi. Rekomendasi pilihan progrm studi membandingkan metode klasifikasi, yaitu KNN dan *Naive Bayes*. Penggunaan metode ini karena diperlukan klasifikasi program studi yang mampu mengelompokan nilai-nilai mata pelajaran, nilai eskul dan nilai Indeks Prestasi Semester (IPS) sehingga mampu menampilkan program studi secara otomatis dan akurat.

Kajian teoritis terkait rekomendasi pilihan program studi menggunakan algoritma *Naive Bayes* [3] dilakukan pada penerimaan mahasiswa STMIK Royal untuk prediksi peminatan program studi menggunakan *Naive Bayes*. Visualisasi menunjukkan data yang tidak seimbang/*Imbalance data*, dimana 92% (390 mahasiswa) memilih Program Studi Sistem Informasi dan 7,6% (32 mahasiswa) memilih program studi Sistem Komputer. Untuk mengatasi hal ini menggunakan metode *Random Undersampling* dan menghasilkan tingkat akurasi *Naive Bayes* 65%. Studi serupa menggunakan *Naive Bayes* untuk memprediksi penjurusan minat siswa SMA Yadika 5 [4], dengan tahap pengumpulan data menghasilkan 306 data. Data tersebut kemudian dibagi menjadi 245 data untuk pelatihan (*training*) dan 61 data untuk pengujian (*testing*). Pada pengumpulan data dihasilkan 9 atribut dengan model *Naive Bayes*. Berdasarkan pengujian, dihasilkan prediksi dengan tingkat akurasi 65%. Kajian ini menghasilkan aplikasi yang dibuat untuk digunakan sekolah dalam menentukan strategi pembelajaran dan menentukan pilihan minat siswa.

Mempelajari algoritma KNN sebagai kajian teoritis ke-2 digunakan juga pada implementasi metode KNN di SMAN 02 Manokwari dalam menentukan jurusan siswa [5]. Menghasilkan 79% nilai akurasi dengan jumlah 60 data *training*, nilai K=9. Rumus *confusion matrix* digunakan dengan membandingkan kelas asli dan kelas prediksi untuk pengujian algoritmanya. Proses perhitungan algoritma KNN ini menggunakan *Microsoft excel*. Jumlah data sampel sebanyak 74 data. Klasifikasi dipengaruhi oleh nilai K yang dipilih, serta pembagian antara data *training* dan data *testing* [5].

Klasifikasi jurusan siswa kelas XI menggunakan algoritma *Naive Bayes* dan KNN yang bertujuan membandingkan kedua algoritma tersebut dalam hal akurasi tertinggi dalam mengklasifikasikan jurusan IPA dan IPS, sehingga dapat membantu pihak sekolah dalam proses penentuan jurusan bagi siswa kelas XI [6]. Penelitian menggunakan data nilai semester 2 sebanyak 277 *record* dan 4 atribut mata pelajaran, yaitu PPKN, Sejarah, Prakarya, dan PAI. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki akurasi sebesar 81,82% dengan sampel data sebanyak 55 dari 277 data. Sementara itu, algoritma KNN memperoleh akurasi sebesar 92,73% dengan sampel data yang sama.

Berdasarkan kajian teoritis yang terkait tentang perbandingan dua algoritma tersebut, dengan ini kembangkan dalam penggunaan *dataset* mahasiswa yang berjumlah 347 data. Semula data awal sejumlah 2.592 data kemudian diolah dengan rumus Slovin sehingga hasil akhir didapat sejumlah itu. Rumus Slovin adalah salah satu metode pengambilan sampel yang paling populer dalam penelitian kuantitatif. Rumus ini sering digunakan untuk menentukan jumlah sampel yang dapat mewakili populasi sehingga hasil penelitian dapat digeneralisasi, dan perhitungannya tidak memerlukan tabel jumlah sampel [7]. Rumus Slovin yang digunakan ada di bawah ini (1).

$$n=\frac{N}{1+N(e)^2} \qquad (1)$$

dengan n sebagai ukuran jumlah responden, N sebagai ukuran populasi, e sebagai persentasi kelonggaran ketelitian kesalahan pengambilan sampel yang masih bisa di tolerir; e=0,1.

Berdasarkan literatur studi yang telah disebutkan, tujuan penelitian ini adalah memberikan rekomendasi kepada calon mahasiswa dalam memilih program studi yang sesuai dengan latar belakang akademik dan kemampuan yang dimiliki, hasil dari pendidikan yang telah ditempuh di SMA atau SMK. Untuk mengetahui algoritma yang dapat menghasilkan nilai akurasi yang baik, maka digunakan perbandingan 2 (dua) algoritma yaitu KNN dan *Naive Bayes*. Penelitian ini diharapkan mampu memberikan manfaat bagi para lulusan SMA atau SMK sebagai rekomendasi pilihan program studi pada saat melanjutkan kuliah hingga lulus kuliah.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

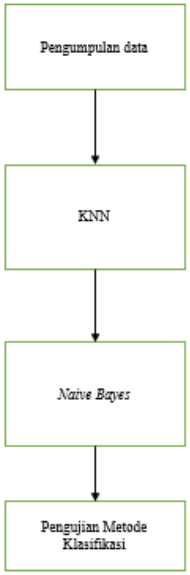
2. METODE PENELITIAN

Penelitian ini menggunakan *text mining* sebagai metode penelitian. Data yang digunakan berasal dari data calon mahasiswa dan mahasiswa. Atribut yang digunakan penelitian adalah nomor pendaftaran menjadi mahasiswa, data nilai rapor yaitu nilai matematika, nilai bahasa Indonesia, nilai pendidikan agama, nilai ekstrakurikuler sewaktu SMA atau SMK, nilai Indeks Prestasi Semester (IPS) 1 dan IPS 2 dengan target mahasiswa angkatan 2020, 2021, dan 2022. Nilai IPS diperoleh per enam bulan sekali atau tiap semester dan dosen dibantu sistem melakukan perhitungan prestasi mahasiswa [8]. Adapun tujuan perhitungan indeks prestasi akademik untuk menilai prestasi akademik mahasiswa yang dilakukan setelah melakukan ujian akhir semester dan nilai indeks yang baik memiliki nilai indeks 3.00. Nilai tersebut sudah ditetapkan sebagai standar indeks [8].

Untuk menjaga agar nilai indeks prestasi akademik mahasiswa tetap baik, diperlukan pemilihan program studi yang sesuai dengan minat, bakat, pengetahuan, dan kemampuan finansial. Dalam penelitian ini, program studi yang direkomendasikan adalah Sistem Informasi, Teknik Informatika, dan Sistem Komputer. Dengan total data yang diperoleh dari data mahasiswa sebanyak 2.592 data. Kemudian menentukan sampel penelitian menggunakan rumus Slovin didapat sejumlah 347 data yang akan diolah dalam penelitian ini. Rumus Slovin [9] digunakan dalam menentukan ukuran sampel yang dibutuhkan.

Penelitian ini mengaplikasikan algoritma KNN dan *Naive Bayes* dengan tujuan untuk membandingkan algoritma mana yang lebih efektif dalam menganalisis *dataset* siswa, sehingga dapat memberikan rekomendasi pilihan program studi.Tindakan yang menghasilkan saran bagi individu lain berdasarkan perhitungan tertentu disebut dengan rekomendasi [10]. Algoritma klasifikasi dalam *supervised learning* yang berbasis jarak dikenal dengan nama Algoritma KNN. Algoritma ini bekerja dengan membandingkan jarak antara data uji dan data latih [11], [12]. Sementara itu, algoritma *Naive Bayes* menggunakan perhitungan probabilitas, di mana *Naive Bayes* menghitung kemungkinan dari suatu kelas terhadap setiap kelompok atribut yang ada, dan menentukan kelas yang paling optimal. Algoritma *Naive Bayes* sering digunakan dalam statistika [13], [14].

Pengolahan data penelitian rekomendasi pemilihan program studi menggunakan *RapidMiner*. Gambar 1 menunjukan tahapan penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan data

Pada tahapan pengumpulan data, peneliti mengumpulkan data nilai rapor mahasiswa angkatan 2020, 2021, 2022 dan nilai IPS 1 serta nilai IPS 2. Data yang terkumpul diperoleh sebanyak 2.592 data dengan atribut yang digunakan yaitu no pendaftaran mahasiswa, nilai matematika, nilai bahasa Indonesia, nilai pendidikan agama, nilai ekstrakurikuler sewaktu SMA atau SMK, IPS 1 dan IPS 2.

2.2 Klasifikasi KNN

Setelah data terkumpul semua dilakukan pengecekan dan pemilihan ternyata masih ada data dengan nomor pendaftaran mahasiswa yang kosong tidak ada dokumen yang diminta di bagian penerimaan mahasiswa baru sehingga dilakukan pengurangan data mahasiswa yang akan digunakan. Terdapat juga data yang berisi bukan dokumen yang diminta untuk pendaftaran mahasiswa baru. Ada pula hasil dokumen yang di *scan* atau foto *blur*

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

atau tidak jelas hasilnya sehingga kesulitan membaca dokumen. Untuk mengolah data yang berjumlah besar dilakukan perhitungan menggunakan rumus Slovin sehingga didapat data berjumlah 347 data dari jumlah awal sebanyak 2.592 data. Dari hasil pengecekan data ternyata peminatan program studi Sistem Komputer tidak mencukupi jumlah maksimal hasil rumus Slovin, sehingga diputuskan tidak menggunakan program studi Sistem Komputer.

Setelah didapatkan data yang siap diolah dari masing-masing mata pelajaran dan nilai IPS lalu nilai diubah dari nilai tipe *number* menjadi tipe *string* untuk selanjutya dilakukan klasifikasi perhitungan algoritma KNN. Metode ini mengklasifikasikan objek baru berdasarkan K tetangga terdekat yang dapat digunakan untuk menentukan kategori objek tersebut [6]. Berikut adalah rumus perhitungan jarak *Euclidean* untuk mengukur jarak antara dua data (2) [11].

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

dengan d sebagai jarak, i sebagai *variabel* data, x1 sebagai sampel data, x2 sebagai data uji atau data testing dan p sebagai dimensi data.

2.3 Klasifikasi *Naive Bayes*

Klasifikasi *Naive Bayes* dilakukan setelah proses klasifikasi KNN, dengan jumlah data training dan data uji yang sama seperti pada KNN. Dalam klasifikasi *Naive Bayes*, nilai atribut kelas dianggap tidak dipengaruhi oleh nilai atribut lainnya. Hal ini didasarkan pada asumsi bahwa keberadaan satu kata dalam kalimat tidak mempengaruhi keberadaan kata lainnya [6]. Persamaan *Naive Bayes* (3), di bawah ini.

$$P(H|X)=\frac{p(X|H)_{p\langle H\rangle}}{p(x)} \qquad (3)$$

dengan X sebagai data *class* yang belum diketahui, H sebagai hipotesis data X merupakan suatu *class* spesifik, P(H|X) sebagai probabilitas hipotesis berdasar kondisi X (*posteriori probability*), P(H) sebagai probabilitas hipotesis H *(prior probability)*, P(X|H) sebagai probabilitas X berdasarkan kondisi pada hipotesis dan P(X) sebagai probabilitas X.

2.4 Pengujian Metode Klasifikasi

Pengujian menggunakan aplikasi *RapidMiner* [12] yaitu bertindak sebagai perancang proses visual untuk menganalisis ilmu data dan pembelajaran mesin pada tim dimulai dari analis sampai pakar. *RapidMiner* juga menampilkan tabel *confusion matrix* sebagai hasil pengujian. Tabel *confusion matrix* berisi klasifikasi jumlah data uji yang besar dan jumlah data uji yang salah [13], [15], [16] disajikan pada Tabel 1 di bawah ini.

Tabel 1. *Confusion Matrix*

Confusion Matrix		Prediksi	
		Positif	Negatif
Aktual	Positif	*True Possitive* (TP)	*False Negative* (FN)
	Negatif	*False Possitive* (FP)	*True Negative* (TF)

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Kegiatan penelitian yang sudah dilakukan adalah mendapatkan data penelitian dari DTI (Direktorat Teknologi Informasi) Universitas Budi Luhur. Rincian data hasil *scan* rapor calon mahasiswa Universitas Budi Luhur, disajikan pada Tabel 2.

Tabel 2. Hasil Rincian Scan Rapot Mahasiswa

No	Tahun	Jumlah mahasiswa melampirkan rapot	Jumlah *scan* rapor yang dapat dibaca
1	2020	1.458 orang	954 orang
2	2021	1.141 orang	1.020 orang
3	2022	1.177 orang	618 orang

Dari jumlah tersebut dihitung kembali menggunakan rumus Slovin untuk menentukan banyaknya *sample* yang digunakan dalam penelitian ini. Didapat hasilnya 347 data dengan nilai E=5%.

Selanjutnya dipilih kembali mahasiswa yang memilih program studi Sistem Informasi, Teknologi Informasi dan Sistem Komputer pada Fakultas Teknologi Informasi. Namun karena jumlah mahasiswa yang memilih prodi Sistem Komputer sangat sedikit hanya ada 7 atau 9 mahasiswa, maka diputuskan pemilihan program studi hanya ada Sistem Informasi dan Teknik Informatika.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Kemudian melakukan seleksi atribut untuk memilih atribut apa saja yang diperlukan dalam proses menganalisa hasil nilai akademik mahasiswa. Atribut yang diambil dari data mahasiswa baru diambil dari data rapor yaitu no daftar/no pendaftaran, nama mahasiswa, asal SMA, jenis kelamin, nilai matematika, nilai bahasa Indonesia, nilai Agama, program studi yang diambil, nilai ekstrakurikuler, nilai IPS 1 rentang nilai berkisar antara 0,01 hingga 4,00, dengan nilai IPS 2 memiliki rentang antara 0,01 hingga 4,00. Atribut yang digunakan disajikan pada Tabel 3 dibawah ini.

Tabel 3. Atribut data penelitian

Kode	Atribut	Nilai Atribut
A	No Daftar	Angka
B	Nama mahasiswa	Alfabet
C	Asal SMA atau SMK	Asal Sekolah Menengah Atas/ Kejuruan
D	Jenis kelamin	Laki-laki; Perempuan
E	Nilai matematika	0 s/d 100
F	Nilai bahasa Indonesia	0 s/d 100
G	Nilai Agama	0 s/d 100
H	Program Studi	Sistem Informasi (SI); Teknik Informatika (TI)
I	Nilai ekstrakurikuler	0 s/d 100
J	Nilai IPS1	0,01 s/d 4,00
K	Nilai IPS2	0,01 s/d 4,00

Berikut data mahasiswa yang terpilih dan terisi semua atributnya yang kemudian dimasukkan di *Excel* dan dimasukan ke database *RapidMiner*. Tabel 4 ini menampilkan data hasil rumus Slovin yang menghasilkan data sampel.

Tabel 4. Data Sampel

No	A	B	C	D	E	F	G	H	I	J	K
1	2201008360	Dimas W.	SMK TRM	Laki-laki	77	82	81	SI	79	2.96	3.51
2	2201004575	Bonifasius C.	SMK BInf	Laki-laki	77	81	90	SI	80	3.57	3.25
3	1201000204	M. Dimas G.	SMA AMj	Laki-laki	85	85	86	SI	85	3.42	2.00
4	2201034416	Wan Evan R.	SMK Khut	Laki-laki	90	90	86	SI	80	3.69	3.45
5	1201000205	Angga Sya.	SMA 85 Jkt	Laki-laki	85	83	95	SI	85	3.68	3.75
6	2201003809	M. Ichsan	SMA/MA	Laki-laki	83	76	84	SI	90	3.51	2.98
7	2202037780	Faiz Ahmad	SMA IIS	Laki-laki	70	80	90	SI	85	3.19	3.36
…	…	…	…	…	…	…	…	…	…	…	…
346	2202039562	Marcella A.	SMK	Perempuan	79	85	79	TI	90	3.80	2.98
347	2202040859	Akbar Y.	SMK	Laki-laki	81	77	76	TI	77	3.73	3.39

Pada Tabel 5 menampilkan nilai bobot program studi penelitian.

Tabel 5. Nilai Bobot Program Studi

No	Kategori	Bobot
1	Sistem Informasi	1
2	Teknik Informastika	0

3.2 Klasifikasi Algoritma KNN

Perbandingan yang digunakan penelitian ini adalah 60% data atau sejumlah 208 data untuk data *training* dan data uji sebesar 40% atau sejumlah 139 data. Kemudian melakukan perhitungan data menggunakan algoritma KNN dengan K=2, K=3, K=5, K=7 sehingga didapat perbedaan jarak menggunakan K=2. Perbandingan akurasi nilai K menggunakan aplikasi *RapidMiner* ditampilkan Tabel 6. Tabel *Euclidean distance* pada Tabel 7 di bawah ini.

Tabel 6. Perbandingan akurasi nilai K

Nilai K	Nilai *accuracy*
2	81.69%
3	79.71%
5	66.67%
7	59.42%

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Tabel 7. *Euclidean distance*

No	No Daftar	Jarak *Euclid*
1	2201008360	0.4968
2	2201004575	0.7439
3	1201000204	0.6437
4	2201034416	0.8679
5	1201000205	0.5789
6	2201002553	0.6930
7	2201004195	0.6844
8	2201004229	0.6405
9	…	…..
110	1201000021	0.5537
111	2211002684	0.4169
112	1211000023	0.6382
113	1211000061	0.8208
114	…	…
134	2211003294	0.5958
135	2211002965	0.4889
136	2211005554	0.8037
137	2211005687	0.3367
138	2201003809	0.5319
139	2202037780	0.8130

Kemudian dilakukan pengurutan data mulai dari jarak terkecil sampai jarak terbesar dengan K=2 pada Tabel 8 di bawah ini.

Tabel 8. *Euclidean distance*

No	Nomor Daftar	Jarak *Euclide*	Jurusan
1	1203000317	0.2842	TI
2	2211005687	0.3367	TI
3	2201013519	0.3605	TI
4	2211002924	0.3729	SI
5	2203068545	0.3789	TI
6	2211002684	0.4169	SI
7	1203000297	0.4236	SI
8	2201004542	0.4656	TI
9	2211002965	0.4889	TI

3.3 Klasifikasi *Naive Bayes*

Untuk klasifikasi *Naive Bayes*, digunakan data *training* dan data uji yang sama seperti pada perhitungan dengan algoritma KNN. Perhitungan *Naive Bayes* menggunakan *RapidMine*r dengan melakukan pengukuran performa data *training* dengan data uji. Data *training* yang digunakan 60% data (208 data) dan data uji sebesar 40% (139 data). Data *training* diberi nama Data *training* (*read excel*) dan data uji (*read excel* (2)) diberi nama data latih mahasiswa dengan *extension* xls. Kemudian dipilih operator *split ratio* untuk *prosentase* data. Kemudian pilih operator *Naive Bayes* dan operator *apply model*. Selanjutnya, ditambahkan operator *performance*, sehingga diperoleh hasil akurasi *Naive Bayes* dengan persentase yang berbeda antara data *training* dan data uji. Tabel 9 di bawah ini menampilkan hasil akurasi *Naive Bayes*.

Tabel 9. Hasil Akurasi *Naïve Bayes*

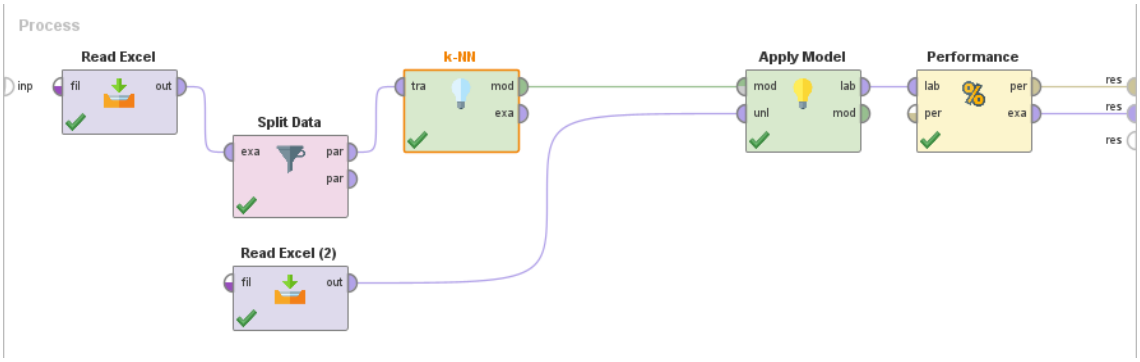
Data T*raining*	Data Uji	Hasil Akurasi
60% (208 data)	40% (139 data)	82.61%
70% (243 data)	30% (104 data)	61.54%
80% (278 data)	20% (69 data)	62.23%

3.4 Pengujian

a. Model KNN

Model KNN menggunakan prosentase 60:40, nilai K=2. Pada Gambar 2 di bawah ini model KNN dengan *RapidMiner*.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index



Gambar 2. Model KNN

Nilai *accuracy* KNN = 81,16% ditampilkan pada Tabel 10 di bawah ini. Akurasi KNN yang mencapai 81,16% menunjukkan bahwa model KNN berhasil mengklasifikasikan 81,16% data dengan tepat dari keseluruhan data yang diuji. Hal ini menandakan bahwa prediksi yang dihasilkan oleh model tersebut cukup akurat, meskipun masih terdapat beberapa kesalahan dalam prediksinya.

Tabel 10. *Accuracy* KNN

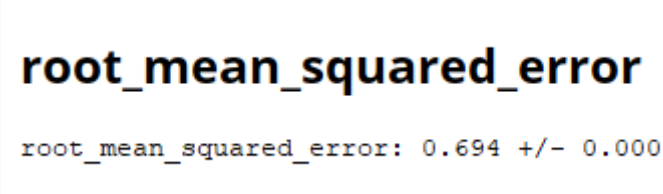
	True no	*True yes*	*Class precision*
Pred. *no*	19	6	76.00%
Pred. *yes*	7	37	84.09%
Class recall	73.08%	86.05%	

Kemudian dicari hasil prediksi KNN menggunakan *RapidMiner* yang dibandingkan dengan nilai bobot awal pada Tabel 11 di bawah ini. Hasil penilaian *Root Mean Square Error* (RMSE) algoritma KNN menggunakan operator *performance* pada *RapidMiner* sebesar 0.694. Hasil RMSE terlihat pada Gambar 3.

Tabel 11. Hasil Prediksi *K-Nearest Neighbors* (K-NN) dengan *RapidMiner*

No Daftar	Prodi awal	Prediksi prodi *Rapidminer*
2201008360	No	Yes
2201004575	No	Yes
1201000204	Yes	Yes
2201034416	Yes	No
1201000205	Yes	Yes
2201002553	Yes	Yes
2201004195	No	Yes
2201004249	Yes	Yes
2201004542	Yes	Yes
1201000021	Yes	Yes
2211002684	No	Yes
….	….	….
….	….	…
2211027327	Yes	Yes
1212000223	No	Yes
2203083627	Yes	No
2203083676	No	No
2203083809	No	No
2203083817	Yes	No
2203083866	No	No
2203083890	Yes	No

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

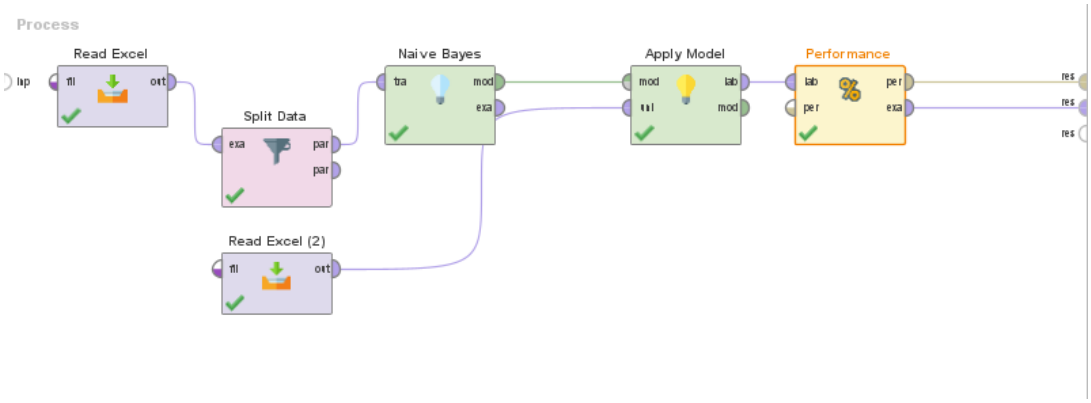


Gambar 3. RMSE

Gambar 3 RMSE sebesar 0,694 berarti rata-rata perbedaan antara nilai prediksi dan nilai sebenarnya pada model tersebut sekitar 0,694. Hal ini menunjukkan bahwa meskipun model relatif akurat, masih ada kesalahan dalam prediksi yang perlu ditingkatkan.

b. Model *Naive Bayes*

Pada Gambar 4 menampilkan model *Naive Bayes* di bawah ini dengan menggunakan prosentase 60:40.



Gambar 4. Model *Naive Bayes*

Nilai *accuracy Naive Bayes* = 82,61% ditunjukkan pada Tabel 12 di bawah ini. Akurasi *Naive Bayes* yang mencapai 82,61% menunjukkan bahwa model *Naive Bayes* berhasil mengklasifikasikan 82,61% data dengan tepat dari keseluruhan data yang diuji. Hal ini menunjukkan bahwa model ini memiliki tingkat akurasi yang sedikit lebih tinggi dibandingkan dengan KNN.

Tabel 12. *Accuracy Naive Bayes*

	True no	*True yes*	*Class prediction*
Pred. *no*	19	5	79,17%
Pred. *yes*	7	38	84,44%
Class recall	73,08%	88,37%	

Kemudian dicari hasil prediksi *Naive Bayes* menggunakan *RapidMiner* yang dibandingkan dengan nilai bobot awal, ditunjukkan pada Tabel 13 di bawah ini.

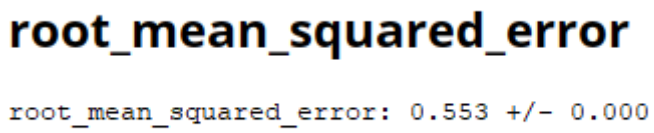
Tabel 13. Hasil Prediksi *Naive Bayes* dengan *RapidMiner*

No Daftar	Prodi awal	Prediksi prodi *RapidMiner*
2201008360	No	Yes
2201004575	No	Yes
1201000204	Yes	No
2201034416	Yes	No
1201000205	Yes	No
2201002553	Yes	Yes
2201004195	No	Yes
2201004249	Yes	Yes
2201004542	Yes	Yes
1201000021	Yes	Yes
2211002684	No	Yes
….	…..	……
….	…..	…..

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 117-126
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

No Daftar	Prodi Awal	Prediksi Prodi *RapidMiner*
2211027327	Yes	Yes
1212000223	No	Yes
2203083627	Yes	No
2203083676	No	No
2203083809	No	No
2203083817	Yes	Yes
2203083866	No	Yes
2203083890	Yes	Yes

Hasil penilaian RMSE algoritma *Naive Bayes* menggunakan operator *performance* pada *RapidMiner* sebesar 0.553 Gambar 5 di bawah ini.



Gambar 5. *Performance Vector Root Mean Square Error* (RMSE)

Gambar 5 RMSE sebesar 0,553 berarti rata-rata perbedaan antara nilai prediksi dan nilai sebenarnya pada model tersebut sekitar 0,553. Hal ini menunjukkan bahwa meskipun model relatif akurat, masih ada kesalahan dalam prediksi yang perlu ditingkatkan.

c. Perbandingan nilai *accuracy* dan nilai *precision* KNN dan *Naive Bayes*

Tabel 14 perbandingan nilai *accuracy* dan nilai *precision* algoritma KNN dan *Naive Bayes* di bawah ini.

Tabel 14. Perbandingan *accuracy* dan *precisios* K-*Nearest Neighbors* (KNN) dan *naive bayes*

Algoritma	*accuracy*	*precisious*
KNN	81,16%	76,00%
Naive Bayes	82,61%	79,17%

Hasil dari penelitian ini menghasilkan nilai RMSE yang tidak berbeda jauh antara algoritma KNN dengan algoritma *Naive Bayes*. Nilai RMSE algoritma KNN senilai 0,694 dan nilai RMSE algoritma *Naive Bayes* senilai 0,496.

Berdasarkan penelitian sebelumnya dengan ide membandingkan algoritma KNN dan *algoritma Naive Bayes* untuk untuk klasifikasi jurusan siswa kelas XI dengan jumlah 277 *record* dan menggunakan 4 (empat) atribut yaitu nilai mata pelajaran PPKN, Sejarah, Prakarya dan PAI. Data *training* sebanyak 222 data dan data *testing* sebanyak 55 data [6]. Dan pada penelitian yang dilakukan ini digunakan sebanyak 347 data. Untuk atribut yang digunakan 11 atribut yaitu no pendaftaran, nama mahasiswa, asal SMA, jenis kelamin, nilai matematika, nilai bahasa Indonesia, nilai Agama, program studi yang diambil, nilai ekstrakurikuler, nilai IPS 1, dan nilai IPS 2. Data yang telah diperoleh akan diuji menggunakan metode KNN, dengan pembagian menjadi data training dan data uji untuk menghitung nilai akurasi. Pembagian data dilakukan dengan variasi, yaitu data training 60% dan data uji 40%, data training 70% dan data uji 30%, serta data training 80% dan data uji 20%. Nilai K yang digunakan adalah 2, 3, 5, dan 7. Begitu juga pada metode Naive Bayes, dilakukan pembagian data training dan data uji untuk mencari nilai akurasi, dengan pembagian persentase data yang serupa: data training 60% dan data uji 40%, data training 70% dan data uji 30%, serta data training 80% dan data uji 20%.

4. KESIMPULAN

Merujuk hasil pengujian dari penggunaan algoritma KNN dan *Naive Bayes* untuk rekomendasi pemilihan program studi yang diusulkan, maka dapat disimpulkan bahwa berdasarkan 347 data dari jumlah awal 2.592 data diperoleh nilai *accuracy* dari algoritma KNN sebesar 81.16% dengan nilai K=2 dan jarak *euclid* 0.2842. Nilai *accuracy* dari algoritma *Naive Bayes* sebesar 82.61%. Untuk nilai *precisious* algoritma KNN sebesar 76.007% dan nilai *precisious* algoritma *Naive Bayes* sebesar 79.17%. Hasil pengujian ini cenderung merekomendasikan program studi Teknik Informatika sebagai program studi pilihan mahasiswa dan lulusan dari SMK cenderung lebih

cepat beradaptasi dan menguasai mata kuliah yang disajikan pada semester 1 dan semester 2. Hal ini diperlihatkan pada hasil IPS semester 1 dan IPS semester 2 lebih tinggi dibandingkan lulusan SMA.

Berdasarkan hasil penelitian analisa perbandingan algoritma KNN dengan algoritma *Naive Bayes* untuk rekomendasi pilihan program studi bagi mahasiswa, maka dapat disimpulkan berjalan baik karena menghasilkan nilai akurasi yang maksimal. Penelitian ini memberikan wawasan lebih dalam tentang karakteristik dan keunggulan masing-masing algoritma baik algoritma KNN maupun algoritma *Naive Bayes*.

Untuk saran yang diberikan bagi penelitian lanjutan dengan menggunakan algoritma lain selain KNN dan *Naive Bayes* serta menambahkan nilai mata pelajaran bahasa Inggris, Fisika, Kimia dan Biologi karena mata pelajaran tersebut yang diujikan untuk UTBK-SBMPTN atau UTBK SNBT jurusan Sainstek. Harapannya dapat menghasilkan kinerja algoritma lebih baik lagi.

DAFTAR PUSTAKA

[1] R. Indonesia, *Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional*. Jakarta: Sekretaris Negara Republik Indonesia, 2003. [Online]. Available: https://shorturl.at/svx8v

[2] H. N. F. Fikrillah and D. Kurniadi, "Rekomendasi Pemilihan Program Studi Menggunakan Algoritma Naïve Bayes," *J. Algoritm.*, vol. 20, no. 1, pp. 42–49, 2023, doi: 10.33364/algoritma/v.20-1.1236.

[3] W. Handoko and M. Iqbal, "Prediksi Peminatan Program Studi Pada Penerimaan Mahasiswa Baru Stmik Royal Menggunakan Naïve Bayes," *J. Sci. Soc. Res.*, vol. 4, no. 2, p. 231, 2021, doi: 10.54314/jssr.v4i2.661.

[4] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020, [Online]. Available: https://shorturl.at/wZtwb

[5] S. Nuraeni, S. P. A. Syam, M. F. Wajdi, B. Firmansyah, and M. Malkan, "Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 89–95, 2023, doi: 10.33379/gtech.v7i1.1905.

[6] M. Y. Putra and D. I. Putri, "Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI," *J. Tekno Kompak*, vol. 16, no. 2, p. 176, 2022, doi: 10.33365/jtk.v16i2.2002.

[7] A. Mardiastuti, "Mengenal Rumus Slovin, Kapan Digunakan dan Contoh Soal," Web Page. Accessed: Jan. 24, 2024. [Online]. Available: https://www.detik.com/jabar/berita/d-6253944/mengenal-rumus-slovin-kapan-digunakan-dan-contoh-soal

[8] U. Susilo and M. Arifin, "Analisis Hubungan Indeks Prestasi Semester Dan Indeks Prestasi Kumulatif Dengan Prestasi Mahasiswa Fakultas Ekonomi Universitas Kadiri," *J. Ris. Bisnis dan Ekon.*, vol. 1, no. 1, pp. 12–22, 2020, [Online]. Available: http://ojs.unik-kediri.ac.id/index.php/jimek

[9] A. Santoso, "Rumus Slovin:Panacea Masalah Ukuran Sampel?," *SuksmaJurnal Psikol. Univ. Sanat Dharma*, vol. 4, p. 6, 2023, [Online]. Available: https://e-journal.usd.ac.id/index.php/suksma/article/view/6434/3637

[10] E. Erwin, V. C. Mawardi, and J. Hendryli, "Penggunaan Metode Collaborative Filtering Based Untuk Rekomendasi Kendaraan Bermotor," *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 1, pp. 3–7, 2022, doi: 10.24912/jiksi.v10i1.17796.

[11] J. Kalyzta, M. A. Willdan, S. Halfiani, and I. Indra, "Penerapan Analisis Sentimen Ujaran Kebencian Terhadap Vaksinasi Covid-19 Pada Tweet Berbahasa Indonesia Menggunakan Algoritme K-Nearest Neighbor," *IDEALIS Indones. J. Inf. Syst.*, vol. 5, no. 2, pp. 87–97, 2022, doi: 10.36080/idealis.v5i2.2959.

[12] K. Kartarina, N. K. Sriwinarti, and N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 107–113, 2021, doi: 10.35746/jtim.v3i2.159.

[13] M. A. Fahtu Rahman, Z. R. Mair, and D. Sartika, "Klasifikasi Ulasan Pelanggan Shopee Mall Terhadap E-Commerce Penjualan Baju Batik Metode Naïve Bayes," *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 164–177, 2024, doi: 10.36080/idealis.v7i2.3178.

[14] E. Salim and A. Solichin, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Pelayanan Dinas Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes," *IDEALIS Indones. J. Inf. Syst.*, vol. 5, no. 2, pp. 79–86, 2022, doi: 10.36080/idealis.v5i2.2961.

[15] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369/348

[16] R. Rusliyawati, K. Muludi, A. Wantoro, and D. A. Saputra, "Implementasi Metode International Prostate Symptom Score (IPSS) Untuk E-Screening Penentuan Gejala Benign Prostate Hyperplasia (BPH)," *J. Sains dan Inform.*, vol. 7, no. 1, pp. 28–37, 2021, doi: 10.34128/jsi.v7i1.298.