Jurnal PTI (Jurnal Pendidikan Teknologi Informasi)
Fakultas Keguruan dan Ilmu Pendidikan Universitas Putra Indonesia
https://jpti-upiyptk.org/ojs
2025 Vol. 12 No. 1 Hal: 7-13 p-ISSN: 2355-9977, e-ISSN: 2685-3914

Perbandingan Metode KNN dan Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Pada Mata Kuliah Probstat

Yuyun Nabilawati Rumbia[1]✉, Raihanfitri Adi Kalipaksi[2], Alvito Gabbriel Saputra[3], Muhammad Dzacky[4], Alif Rifa'i[5], Anindita Septiarini[6], Novianti Puspitasari[7]

[1,2,3,4,5] Informatika, Fakultas Teknik, Universitas Mulawarman
yuyunabila10@gmail.com

Abstract

This study aims to compare the performance of two classification algorithms in data mining, namely K-Nearest Neighbor (KNN) and Naive Bayes, on raw numerical data that has not undergone normalization. The main focus of this research is to classify the graduation status of students in the Probability and Statistics course for the 2022 cohort in the Informatics Study Program at Mulawarman University. The dataset used is original data obtained directly from the lecturer of the course. It consists of 136 entries, each with three numerical attributes and one classification label, which is either "Pass" or "Fail." The labels were determined based on the official graduation standards applied at Mulawarman University. In the model training and testing process, the data was split using a 70:30 ratio, with 70% used as training data and the remaining 30% as testing data. This process was carried out using the Scikit-learn library in Python. Unlike most previous approaches that typically involve data normalization or standardization before classification, this study specifically examines the performance of the two algorithms on raw, unprocessed data. The performance evaluation was conducted using four main metrics: accuracy, precision, recall, and F1-score. The test results show that the KNN algorithm outperformed Naive Bayes in all evaluation metrics. KNN achieved an accuracy of 94.87%, while Naive Bayes only reached 87.80%. In addition, the precision, recall, and F1-score values of KNN were consistently higher. Therefore, it can be concluded that KNN is more effective and reliable in classifying raw numerical data without normalization, particularly in the context of predicting student graduation in a specific course.

Keywords: k-nearest neighbor (knn), naive bayes, data mining, graduation classification, probability and statistics

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi dalam data mining, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes, terhadap data numerik mentah yang tidak melalui proses normalisasi. Fokus utama dari penelitian ini adalah untuk mengklasifikasikan status kelulusan mahasiswa dalam mata kuliah Probabilitas dan Statistika untuk angkatan tahun 2022 pada Program Studi Informatika, Universitas Mulawarman. Dataset yang digunakan merupakan data asli yang diperoleh langsung dari dosen pengampu mata kuliah tersebut. Dataset ini terdiri atas 136 entri yang masing-masing memiliki tiga atribut numerik dan satu label klasifikasi, yaitu "Lulus" atau "Tidak Lulus". Penentuan label klasifikasi dilakukan berdasarkan standar kelulusan resmi yang berlaku di Universitas Mulawarman. Dalam proses pelatihan dan pengujian model, data dibagi menggunakan rasio 70:30, yakni 70% data digunakan sebagai data latih (training data) dan 30% sisanya sebagai data uji (testing data). Proses ini dilakukan dengan bantuan library Scikit-learn di Python. Berbeda dengan sebagian besar pendekatan sebelumnya yang cenderung melakukan normalisasi atau standarisasi data terlebih dahulu sebelum klasifikasi, penelitian ini secara khusus menguji performa kedua algoritma terhadap data mentah yang tidak diproses terlebih dahulu. Evaluasi performa dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa algoritma KNN memberikan hasil yang lebih baik dibandingkan dengan Naive Bayes pada seluruh metrik evaluasi. KNN berhasil mencapai akurasi sebesar 94,87%, sedangkan Naive Bayes hanya mencapai 87,80%. Selain itu, nilai presisi, recall, dan F1-score dari KNN juga secara konsisten lebih tinggi. Dengan demikian, dapat disimpulkan bahwa KNN lebih efektif dan andal dalam mengklasifikasikan data numerik mentah yang tidak dinormalisasi.

Kata kunci: k-nearest neighbor (knn), naive bayes, data mining, klasifikasi kelulusan, probabilitas dan statistika

1. Pendahuluan

Perkembangan dalam ilmu data telah memberikan dampak yang besar dalam berbagai bidang dan pendidikan merupakan salah satunya. Salah satu objek dalam dunia pendidikan adalah Mahasiswa yang dalam menyelesaikan masa studinya diharuskan untuk lulus dalam setiap mata kuliah yang diambil. Ketidakmampuan untuk menyelesaikan mata kuliah akan mempengaruhi waktu tempuh dari masa studi sehingga evaluasi kelulusan mata kuliah diperlukan untuk menilai kinerja dan kemampuan mahasiswa [1].

Kelulusan mahasiswa dalam suatu mata kuliah dapat dipastikan dengan nilai yang didapatkan, karena itu mahasiswa harus mendapat nilai yang memenuhi standar

Diterima: 14-04-2025| Revisi: 21-04-2025 | Diterbitkan: 30-04-2025 | doi: 10.35134/jpti.v12i1.228

kelulusan yang telah ditetapkan oleh masing-masing perguruan tinggi. Komponen nilai mata kuliah mahasiswa yang dikumpulkan terdiri dari lima yaitu nilai tugas, nilai afektif, nilai ujian tengah semester, ujian akhir semester dan nilai proyek. Kemampuan dan pemahaman mahasiswa juga dapat diukur dengan nilai yang di dapat, karena itu diperlukan prediksi kelulusan mata kuliah mahasiswa menggunakan nilai-nilai yang telah didapatkan untuk melihat mahasiswa yang lulus dan tidak.

Probabilitas adalah bagian matematika yang membahas perihal ukuran kebolehjadian terjadinya suatu peristiwa yang ada dalam kehidupan [2]. Statistika adalah kumpulan data dalam bentuk tabel(daftar) atau diagram yang menggambarkan atau berkaitan dengan suatu masalah tertentu [3].

Data mining merupakan metode pencarian pola yang didapatkan dari data untuk dikelola menjadi sebuah informasi [4]. Data mining diperlukan untuk menggali informasi yang berada di dalam data. Dalam mendapatkan informasi data mining menggunakan konsep matematika, statistic, kecerdasan buatan dan pembelajaran mesin [5].

K Nearest Neighbor (KNN) merupakan metode klasifikasi yang menentukan kategori suatu data berdasarkan mayoritas kategori dari sejumlah tetangga terdekatnya [6]. metode ini mengklasifikasikan data uji dengan membandingkan jarak kedekatannya terhadap data latih. Semakin dekat suatu data uji dengan data latih tertentu, maka semakin besar kemungkinan data uji tersebut dikategorikan ke dalam kelas yang sama dengan data latih tersebut [7].

Naive Bayes adalah metode klasifikasi yang memprediksi label atau kelas dan berdasar pada teori Bayesian yang memperkirakan bahwa nilai atribut tidak bergantung dengan nilai yang lainnya. Proses klasifikasi data terdiri dari dua Langkah yang pertama ialah pelatihan dengan dataset training. Langkah Kedua klasifikasi dengan dataset yang tidak diketahui label atau kelasnya [8]

Pada penelitian klasifikasi gender berdasarkan mata, metode KNN menunjukkan keunggulan yang lebih baik dibandingkan Naive Bayes. Hal ini ditunjukkan melalui nilai akurasi yang lebih tinggi, khususnya ketika menggunakan kombinasi fitur HOG dan HSV dengan penerapan teknik cropping pada mata. kombinasi tersebut mampu meningkatkan representasi fitur visual dari mata, sehingga metode KNN dapat mengelompokkan data dengan tingkat akurasi yang lebih baik. Oleh karena itu, dalam konteks klasifikasi gender berdasarkan mata, metode KNN lebih unggul dibandingkan Naïve Bayes [9].

Pada penelitian Perbandingan Analisis Prediksi Kepuasan Pengguna Layanan GoFood ditemukan bahwa metode KNN memiliki hasil yang lebih baik dengan tingkat Accuracy 98,80% dan Recall 100% sedangkan Naïve Bayes mendapatkan nilai Accuracy 94,10% dan Recall 94,43%. Dalam penelitian ini proses pengolahan data menggunakan tools RapidMiner [10]

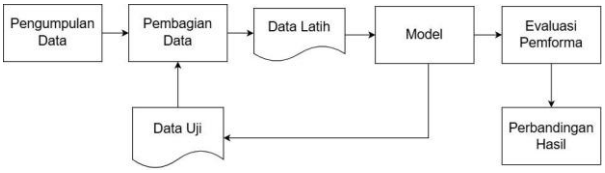
Hasil Perbandingan Metode Naive bayes dengan Metode KNN menunjukkan Bahwa Metode Naive bayes memiliki performa yang lebih baik saat mengklasifikasi text dengan hasil pengujian yang menghasilkan performa dengan rata rata 84% sedangkan Metode KNN menghasilkan performa dengan rata-rata 73% [11].

Pada penelitian klasifikasi keluarga miskin yang menggunakan tools RapidMiner dalam pengolahan data metode Naïve Bayes lebih efektif dengan nilai akurasi 99,26%, precision 100%, recall 85,41% dan AUC 0,978 dibanding metode KNN yang mendapat nilai akurasi 98,52%, precision 100%, recall 90,82%, AUC 0,993. [12]

Penelitian ini dilakukan karena belum adanya studi yang secara khusus membandingkan performa algoritma klasifikasi K-Nearest Neighbor (KNN) dan Naive Bayes terhadap nilai mata kuliah Probabilitas dan Statistika (Probas) mahasiswa Program Studi Informatika Universitas Mulawarman angkatan 2022. Data yang digunakan dalam penelitian ini merupakan data asli yang diperoleh langsung dari dosen pengampu, sehingga hasil analisis yang dihasilkan memiliki nilai validitas yang tinggi. Selain itu, implementasi kedua algoritma dilakukan menggunakan aplikasi rapid miner, yang memberikan fleksibilitas dan akurasi dalam proses klasifikasi. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem prediksi akademik berbasis machine learning di lingkungan perguruan tinggi.

Tujuan dari penelitian yang dilakukan ialah untuk mengetahui perbandingan tingkat akurasi metode KNN dan Naive Bayes dalam klasifikasi kelulusan mata kuliah mahasiswa dan penerapan metode KNN serta Naive Bayes dalam klasifikasi kelulusan mata kuliah.

2. Metodologi Penelitian



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Dataset yang kami peroleh berasal dari wawancara dan diskusi dengan Pak Prof. Dr. Fahrul Agus, S.Si., M.T, selaku dosen pengampu mata kuliah Probabilitas dan Statistika angkatan 2022, dengan jumlah data sebanyak 136 entri yang memiliki atribut nilai kuis, UTS, dan UAS, serta label “Lulus” atau “Tidak Lulus” yang menggunakan patokan nilai akademik Universitas Mulawarman yang menyatakan jika nilai akhir. suatu matakuliah adalah >60 maka dinyatakan “Tidak Lulus”

atau mendapatkan nilai huruf "D" dan mewajibkan mahasiswa mengulang mata kuliah tersebut . Jenis data yang digunakan merupakan data numerik, yang sesuai untuk keperluan analisis dan pengolahan lebih lanjut.

2.2. Pra-pemprosesan Data

Pada tahap pra-pemrosesan, kami tidak melakukan penanganan terhadap data kosong (missing value) karena data yang diperoleh telah lengkap. Selain itu, normalisasi data juga tidak dilakukan, karena tujuan dari penelitian ini adalah untuk membandingkan kinerja dua metode klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes, dengan menggunakan data asli. Selanjutnya, data dibagi (split) menggunakan Python dengan bantuan library Scikit-Learn, Scikit-learn merupakan *library* python yang sering digunakan untuk memudahkan pemrosesan sebuah data kelebihan dari scikit-learn performa yang *optimized* karena library ini menggunakan perhitungan yang efisien dengan menggunakan Numpy (Numerical Python) dan Scipy (Sientific Python) [13]. khususnya fungsi train_test_split. Melalui fungsi ini, data diacak dan dibagi dengan rasio 70:30, di mana 70% digunakan untuk pelatihan (training) dan 30% untuk pengujian (testing).

2.3. K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) adalah metode klasifikasi yang menentukan kelas suatu data berdasarkan jarak terdekat ke data lain dalam dataset. Termasuk algoritma non-parametrik dan berbasis contoh, KNN dianggap sebagai salah satu metode paling sederhana dalam data mining. Algoritma ini menghitung jarak antara data baru dan seluruh data latih, kemudian memilih 5 tetangga terdekat (K=5) untuk menentukan kelas berdasarkan mayoritas suara. Pemilihan nilai K sangat mempengaruhi hasil klasifikasi, karena nilai yang terlalu kecil bisa sensitif terhadap noise, sedangkan nilai terlalu besar bisa menyebabkan bias tinggi [14]. Algoritma KNN ini umum digunakan karena mampu melakukan prediksi berdasarkan data uji kemudian diproses menggunakan data latih yang menghasilkan nilai akurasi yang tinggi dalam klasifikasi data [15].

$$d_i = \sqrt{\sum_{i=1}^{p}(x_1 - x_2)^2} \quad (1)$$

2.4. Naive Bayes

Naïve Bayes adalah teknik klasifikasi probabilitas sederhana, menerapkan teorema Bayes, yang mampu menangani data kuantitatif dan diskrit untuk menghitung perkiraan probabilitas yang diperlukan untuk jenis analisis [16]. Metode Naïve Bayes didasarkan pada asumsi bahwa atribut-atribut dalam data adalah independen satu sama lain. Naïve Bayes mengelompokkan data dengan menghitung probabilitas kemunculan kelas target berdasarkan atribut-atribut yang terkait, serta memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa lampau [17].

$$P(C_i|X) = \frac{P(X|C_i)\,.P(C_i)}{P(X)} \quad (2)$$

2.5. Rapid Miner

Rapid Miner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Rapid Miner merupakan perangkat lunak independen yang digunakan untuk menganalisa data dan mesin penambangan data, yang dapat diintegrasikan dengan berbagai bahasa pemrograman secara mudah. RapidMiner menyediakan tampilan (UI) yang ramah pengguna, sehingga memudahkan pengguna saat menggunakannya. Tampilan yang terdapat pada RapidMiner disebut Perspective [18].

2.6. Model Evaluasi

Pada evaluasi, model yang dipakai adalah confusion matrix. Confusion Matrix merupakan metode yang digunakan untuk mengevaluasi tingkat akurasi suatu model klasifikasi dalam membedakan data berdasarkan kelasnya. Melalui pengukuran akurasi ini, performa model klasifikasi dapat diketahui secara menyeluruh. Nilai akurasi diperoleh dari data pelatihan dan dinyatakan dalam bentuk persentase, yang menunjukkan proporsi prediksi yang sesuai dengan kelas sebenarnya dari data tersebut [19]. Sedangkan, hasil uji dari model klasifikasi yang akan dianalisa menggunakan table atau sering disebut (confusion matrix) dapat memperoleh nilai accuracy, recall, precision, serta f1-score. Berikut Confusion matrix yang dievaluasi dengan tabel yang menerangkan bahwa hasil klasifikasi dari jumlah data testing yang benar dan dari jumlah data testing yang salah [20].

Tabel 1. Confusion Matrix

Kelas Prediksi	Kelas Sebenarnya	
	+	-
+	*True Positives*	*False Positives*
-	*False Negatives*	*True Negatives*

2.6.1. Akurasi

Akurasi adalah ukuran seberapa sering model membuat prediksi yang benar secara keseluruhan. Ini dihitung dengan membagi jumlah prediksi yang benar dengan total seluruh prediksi (baik yang benar maupun salah).

$$\text{Akurasi} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (3)$$

2.6.2. Presisi

Presisi mengukur seberapa banyak dari prediksi positif yang benar-benar positif. Dengan kata lain, ini menunjukkan ketepatan model dalam memprediksi kelas positif. Presisi juga berarti keadaan yang berdekatan antara hasil test yang di dapat dari suatu kondisi yang telah ditetapkan, semakin mendekati nilai

hasil pengulangan maka semakin tinggi tingkat presisinya [21].

$$\text{Presisi} = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \quad (4)$$

2.6.3. Recall

Recall, atau yang dikenal juga sebagai tingkat perolehan, merupakan kemampuan suatu sistem dalam menampilkan seluruh hasil yang relevan berdasarkan kata kunci yang dimasukkan [22].

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (5)$$

2.6.4. F1-Score

F1-score merupakan rata-rata harmonis antara precision dan recall yang telah diberi bobot. Nilai F1-score umumnya lebih rendah dibandingkan dengan nilai akurasi karena mempertimbangkan secara seimbang kedua metrik, yaitu precision dan recall, dalam proses perhitungannya [23].

$$\text{F1-Score} = 2\ \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (6)$$

3. Hasil dan Pembahasan

3.1. Persiapan Data

Data ini diperoleh dari Prof. Dr. Fahrul Agus, S.Si., M.T, selaku dosen pengampu mata kuliah Probabilitas dan Statistika Angkatan 2022 Informatika Universitas Mulawarman. Data berjumlah 136 entri dengan 3 atribut dan 1 label, berikut adalah tampilan dari data

Tabel 2. Data awal

No	Nim	Kuis	UTS	UAS	Keterangan
1	2209106001	90	72	30	Tidak Lulus
2	2209106002	90	63	67	Lulus
3	2209106003	90	67	77	Lulus
4	2209106004	90	62	46	Tidak Lulus
…	…	…	…	…	…
…	…	…	…	…	…
133	2209106135	90	89	55	Lulus
134	2209106136	90	92	45	Lulus
135	2209106138	90	79	70	Lulus
136	2209106139	90	55	58	Lulus

3.2. Pre-processing Data

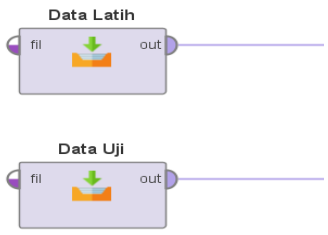
Pada tahap preprocessing data, dilakukan pembagian dan pengacakan data agar distribusi data menjadi lebih merata antara data latih dan data uji. Pembagian ini dilakukan dengan perbandingan 70:30 menggunakan library Python, yaitu sklearn, dengan memanfaatkan fungsi train_test_split. Berikut ini adalah algoritma deskriptifnya

Algoritma 1. Split Data

Input: *dataset.csv*
Output: *data_latih.csv, data_uji.csv*
Initialization *df, X, y, X_train, X_test, y_train, y_test, train_set, test_set*
Get df
df ← read_csv("probas abc 20222.csv")
df ← sort(df berdasarkan NIM)
reset_index(df)
Get fitur dan label
X ← df[['Kuis', 'UTS', 'UAS']]
y ← df['keterangan']
Split data
[X_train, X_test, y_train, y_test] ← train_test_split(X, y, test_size = 0.3, shuffle = True, random_state = 10)
train_set ← df dengan indeks X_train
test_set ← df dengan indeks X_test
write_csv(train_set, "data_latih.csv")
write_csv(test_set, "data_uji.csv")
print "File 'data_latih.csv' dan 'data_uji.csv' sudah disimpan."

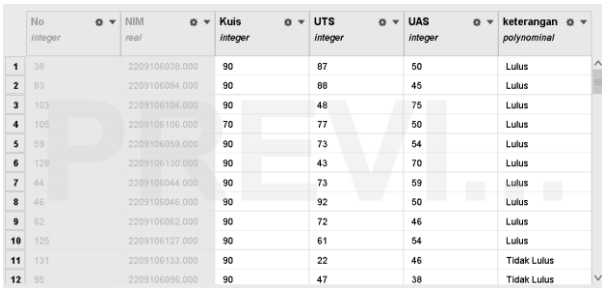
3.3. Hasil Data Mining

Pada tahap klasifikasi, data latih dan data uji dimuat ke dalam RapidMiner melalui operator Read CSV. Operator ini digunakan untuk membaca dataset dalam format CSV yang telah dipersiapkan sebelumnya.



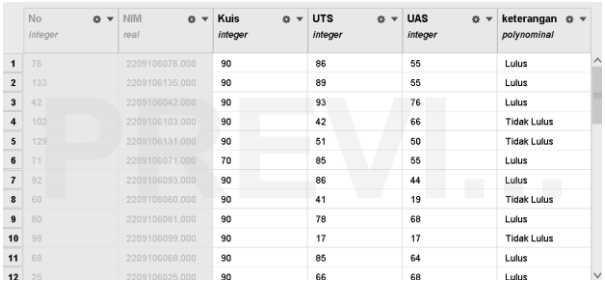
Gambar 2. Operator Read CSV

Dalam proses ini, atribut "No" dan "NIM" tidak disertakan dalam pemodelan karena tidak memiliki kontribusi prediktif terhadap target klasifikasi. Kedua atribut tersebut hanya berfungsi sebagai identitas dan tidak relevan untuk proses pembelajaran mesin.



	No integer	NIM real	Kuis integer	UTS integer	UAS integer	keterangan polynominal
1	38	2209106038.000	90	87	50	Lulus
2	83	2209106084.000	90	88	45	Lulus
3	103	2209106104.000	90	48	75	Lulus
4	105	2209106106.000	70	77	50	Lulus
5	59	2209106059.000	90	73	54	Lulus
6	126	2209106130.000	90	43	70	Lulus
7	44	2209106044.000	90	73	59	Lulus
8	46	2209106046.000	90	92	50	Lulus
9	62	2209106062.000	90	72	46	Lulus
10	125	2209106127.000	90	61	54	Lulus
11	131	2209106133.000	90	22	46	Tidak Lulus
12	95	2209106096.000	90	47	38	Tidak Lulus

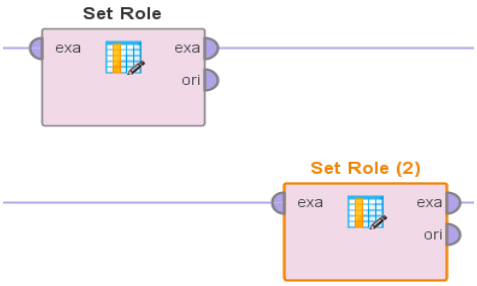
Gambar 2. Preview data latih



	No integer	NIM real	Kuis integer	UTS integer	UAS integer	keterangan polynominal
1	76	2209106076.000	90	86	55	Lulus
2	133	2209106135.000	90	89	55	Lulus
3	42	2209106042.000	90	93	76	Lulus
4	102	2209106103.000	90	42	66	Tidak Lulus
5	129	2209106131.000	90	51	50	Tidak Lulus
6	71	2209106071.000	70	85	55	Lulus
7	92	2209106093.000	90	86	44	Lulus
8	60	2209106060.000	90	41	19	Tidak Lulus
9	80	2209106081.000	90	78	68	Lulus
10	98	2209106099.000	90	17	17	Tidak Lulus
11	68	2209106068.000	90	85	64	Lulus
12	25	2209106025.000	90	66	68	Lulus

Gambar 3. Preview data uji

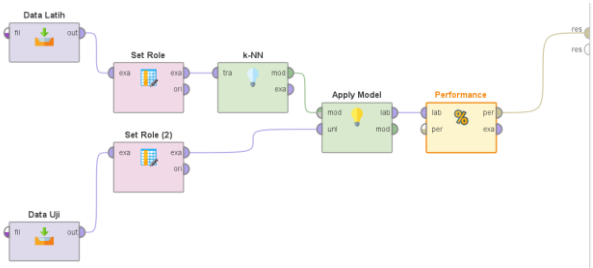
Selanjutnya, penetapan atribut target (label) dilakukan menggunakan operator Set Role, dengan menetapkan atribut "keterangan" sebagai label klasifikasi. Hal ini bertujuan agar RapidMiner mengenali kolom tersebut sebagai target dalam proses pelatihan model.



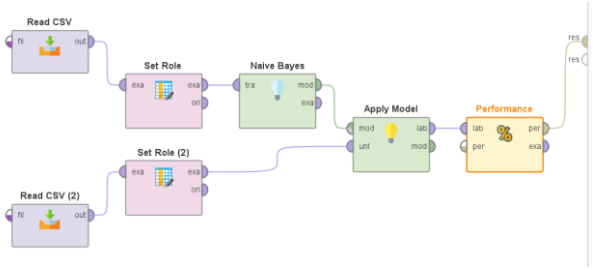
Gambar 4. Operator Set Role

Dua algoritma klasifikasi digunakan dalam penelitian ini, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes. Kedua model dilatih menggunakan dataset latih, kemudian diuji terhadap dataset uji menggunakan operator Apply Model.

Evaluasi kinerja model dilakukan dengan operator Performance (Classification) untuk memperoleh metrik performa seperti akurasi, precision, recall, dan confusion matrix.



Gambar 5. Proses Klasifikasi KNN



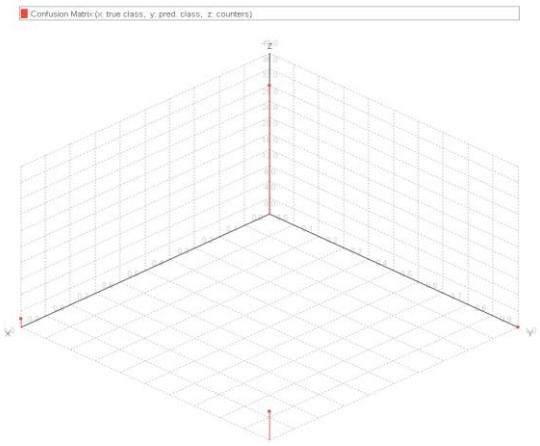
Gambar 5. Proses Klasifikasi Naive Bayes

Proses kemudian dieksekusi, dan sistem menampilkan hasil klasifikasi berupa matrik evaluasi kinerja model seperti akurasi, precision, dan recall.

accuracy: 95.12%

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	32	2	94.12%
pred. Tidak Lulus	0	7	100.00%
class recall	100.00%	77.78%	

Gambar 6. Confusion Matrix KNN



Gambar 7. Plot View KNN

PerformanceVector

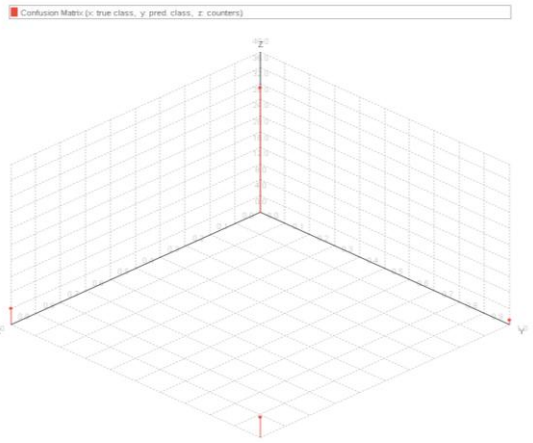
```
PerformanceVector:
accuracy: 95.12%
ConfusionMatrix:
True:   Lulus   Tidak Lulus
Lulus:  32      2
Tidak Lulus:    0       7
```

Gambar 8. Performance Vector KNN

accuracy: 87.80%

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	31	4	88.57%
pred. Tidak Lulus	1	5	83.33%
class recall	96.88%	55.56%	

Gambar 9. Confusion Matrix Naive Bayes



Gambar 10. Plot View Naive Bayes

PerformanceVector

```
PerformanceVector:
accuracy: 87.80%
ConfusionMatrix:
True:   Lulus   Tidak Lulus
Lulus:  31      4
Tidak Lulus:    1       5
```

Gambar 11. Performance Vector Naive Bayes

Berdasarkan hasil evaluasi ini, performa masing-masing model dapat dibandingkan untuk menentukan algoritma klasifikasi yang lebih optimal terhadap klasifikasi kelulusan mahasiswa di mata kuliah Probabilitas dan Statistika.

3.4. Evaluasi

Pada tahap Evaluasi kami membandingkan hasil akurasi, performance, dan f1-score dari kedua hasil sebelumnya

Tabel 3. Perbandingan Antara KNN dan Naive Bayes

No	Algoritma	Performance Vector
1.	K-Nearest Neighbors	94.12%
2.	Naive Bayes	88.57%

Sekarang kita akan menghitung akurasi, presisi, recall, dan f1-score dari kedua hasil analisis model tadi

3.4.1. Evaluasi KNN

Tabel 4. Confusion Matrix KNN

Kelas Prediksi	Kelas Sebenarnya	
	+	-
+	*32*	*2*
-	*0*	*5*

Hasil klasifikasi dari algoritma yang digunakan dievaluasi menggunakan metrik kinerja berupa akurasi, presisi, recall, dan F1-score. Berdasarkan confusion matrix yang diperoleh, diketahui bahwa:

True Positive (TP) = 32
False Negative (FN) = 2
False Positive (FP) = 0
True Negative (TN) = 5

Dengan nilai-nilai tersebut, diperoleh hasil sebagai berikut

Tabel 5. Model evaluasi KNN

Model Evaluasi	Persentase
Akurasi	94.87%
Presisi	100%
Recall	94.12%
F1-Score	96.98%

Hasil ini menunjukkan bahwa KNN memiliki kinerja yang sangat baik, terutama dalam meminimalkan kesalahan klasifikasi positif palsu (False Positive) dan mendeteksi mayoritas kasus positif secara akurat.

3.4.2. Evaluasi Naive Bayes

Tabel 6. Confusion Matrix Naive Bayes

Kelas Prediksi	Kelas Sebenarnya	
	+	-
+	*31*	*4*
-	*1*	*5*

Dari confusion matrix di atas, didapatkan:

True Positive (TP) = 31
False Negative (FN) = 4
False Positive (FP) = 1
True Negative (TN) = 5

Maka, perhitungan model evaluasi kinerja model adalah sebagai berikut:

Tabel 7. Model evaluasi Naive Bayes

Model Evaluasi	Persentase
Akurasi	87.80%
Presisi	96.88%
Recall	88.57%
F1-Score	92.53%

Berdasarkan hasil tersebut, model Naive Bayes menunjukkan kinerja yang cukup baik dengan akurasi 87,80%. Meskipun presisinya tinggi (96,88%), terdapat sedikit penurunan pada nilai recall (88,57%) yang menunjukkan bahwa beberapa kasus positif gagal dikenali (False Negative = 4). Nilai F1-Score sebesar 92,53% menandakan model tetap memiliki keseimbangan yang baik antara presisi dan recall, namun tidak sebaik model KNN yang sebelumnya diuji

4. Kesimpulan

Berdasarkan hasil evaluasi kinerja model terhadap klasifikasi kelulusan mahasiswa pada mata kuliah Probabilitas dan Statistika angkatan 2022 Program Studi Informatika Universitas Mulawarman, diperoleh perbandingan performa antara metode K-Nearest Neighbor (KNN) dan Naive Bayes. Hasil pengujian menunjukkan bahwa metode KNN memiliki performa yang lebih unggul dibandingkan Naive Bayes dalam seluruh matrik evaluasi.

Tabel 8. Perbandingan Model evaluasi KNN dan Naive Bayes

Model Evaluasi	KNN	Naive Bayes
Performance Vector	94.12%	88.57%
Akurasi	94.87%	87.80%
Presisi	100%	96.88%
Recall	94.12%	88.57%
F1-Score	96.98%	92.53%

Dengan demikian, dapat disimpulkan bahwa KNN lebih unggul dalam mengidentifikasi pola kelulusan mahasiswa dan menghasilkan prediksi yang lebih tepat, khususnya ketika fitur numerik digunakan secara langsung tanpa normalisasi.

Perbedaan performa ini mengindikasikan bahwa KNN lebih sesuai untuk kasus klasifikasi berbasis nilai numerik yang memiliki distribusi tidak terlalu kompleks dan tidak memerlukan asumsi distribusi data tertentu, seperti yang dibutuhkan oleh Naive Bayes. Menariknya, meskipun data yang digunakan dalam penelitian ini merupakan data asli tanpa melalui proses normalisasi, algoritma KNN tetap mampu memberikan performa yang sangat baik dengan akurasi tinggi. Hal ini menunjukkan bahwa KNN cukup tangguh dan efektif dalam menangani data numerik mentah, serta direkomendasikan sebagai metode klasifikasi yang lebih akurat dalam konteks serupa, khususnya untuk

pengambilan keputusan berbasis penilaian akademik mahasiswa.

Daftar Rujukan

[1] Ahmad, N., Hafizh, S., & Sulthanah, R. (2024). Prediksi kelulusan mata kuliah mahasiswa Teknologi Informasi menggunakan algoritma K-Nearest Neighbor. Jurnal Manajemen Informatika (JAMIKA), 14(2), 135–149. https://doi.org/10.34010/jamika.v14i2.12454

[2] Maulana, C. (2023). Pengembangan modul matakuliah statistika dan probabilitas berbasis kontekstual. Jurnal Pengembangan Rekayasa dan Teknologi, 7(1), 1–10. https://doi.org/10.26623/jprt.v19i1.7745

[3] Rusmini, R., & Mazali, M. R. (2024). Peranan visual thinking berbasis computational thinking dalam penyelesaian masalah statistik dan probabilitas. Jurnal Cendekia: Jurnal Pendidikan Matematika, 8(2), 1342–1350. https://doi.org/10.31004/cendekia.v8i2.3304

[4] Shah, K., Shah, N., Sawant, V., & Parolia, N. (Eds.). (2023). Practical data mining techniques and applications (1st ed.). Auerbach Publications. https://doi.org/10.1201/9781003390220

[5] Margareta Sari, A., Rosita, D., & Subastian, E. (2025). Pengembangan media pembelajaran menggunakan Lumio by Smart pada mata pelajaran TIK siswa kelas VII SMP Negeri 37 Samarinda. *SABER: Jurnal Teknik Informatika, Sains dan Ilmu Komunikasi*, 3(2), 1–8. https://doi.org/10.59841/saber.v3i2.690 Amik Veteran Journal+3

[6] Dijaya, R., & Setiawan, H. (2023). *Buku ajar pengolahan citra digital*. Umsida Press. https://doi.org/10.21070/2023/978-623-464-075-5

[7] Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing* (3rd ed.). Pearson Education. https://doi.org/10.5555/1076432

[8] Martantoh, E., & Yanih, N. (2022). Implementasi metode Naïve Bayes untuk klasifikasi karakteristik kepribadian siswa di sekolah MTS Darussa'adah menggunakan PHP MySQL. *Jurnal Teknologi dan Sistem Informasi*, 3(2), 166–175. https://doi.org/10.33365/jtsi.v3i2.2896

[9] Kurniawan, C., & Irsyad, H. (2022). Perbandingan metode K-Nearest Neighbor dan Naïve Bayes untuk klasifikasi gender berdasarkan mata. *Jurnal Algoritme*, 2(2), 82–91. https://doi.org/10.35957/algoritme.v2i2.2358

[10] Yuliarina, A. N., & Hendry, H. (2022). Comparison of prediction analysis of GoFood service users using the KNN & Naive Bayes algorithm with RapidMiner software. *Jurnal Teknik Informatika (JUTIF)*, 3(4), 847–856. https://doi.org/10.20884/1.jutif.2022.3.4.294

[11] Berton, F. T., Ratnawati, D. E., & Rahman, M. A. (2024). Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk analisis sentimen terhadap ulasan aplikasi Threads. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(1), 1–10. https://doi.org/10.1234/jptiik.v8i1.14109

[12] Pamungkas, W. A. B., & Asnawi, M. F. (2022). Studi komparasi algoritma Naive Bayes dan K-Nearest Neighbor untuk klasifikasi keluarga miskin di Desa Ngadiwarno Kecamatan Sukorejo Kabupaten Kendal. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, *3*(2), 1–8. https://doi.org/10.55123/storage.v3i2.3600

[13] Fahmi, M. N. (2023). Implementasi Machine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning). Sains Data: Jurnal Studi Matematika dan Teknologi, 1(2), 87–96. https://doi.org/10.52620/sainsdata.v1i2.31

[14] Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi algoritma klasifikasi K-Nearest Neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *Indonesian Journal on Computer and Information Technology (IJCIT)*, 6(2), 118–127. https://doi.org/10.31294/ijcit.v6i2.10438

[15] Sari, D. P., & Ardika, I. K. (2022). Implementasi algoritma K-Nearest Neighbor untuk prediksi gizi buruk pada balita. *Jurnal Teknologi dan Sistem Informasi*, 3(2), 176–185. https://doi.org/10.33365/jtsi.v3i2.2897

[16] Fauzi, A. (2025). Aplikasi sistem pakar dengan metode Naive Bayes untuk mendeteksi penyakit diabetes. *Jurnal Manajemen Informatika (JAMIKA)*, 15(1). https://doi.org/10.34010/jamika.v15i1.12391

[17] Yusnita, A., Lailiyah, S., & Saumahudi, K. (2021). Penerapan algoritma Naïve Bayes untuk penerimaan peserta didik baru. *Jurnal Informatika Wicida*, *10*(1), 11–16. https://doi.org/10.46984/inf-wcd.1194

[18] Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan aplikasi RapidMiner untuk prediksi nilai tukar rupiah terhadap US Dollar dengan metode linear regression. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(1), 8–17. https://doi.org/10.25077/TEKNOSI.v7i1.2021.8-17

[19] Helmud, E., Fitriyani, F., & Romadiana, P. (2023). Classification comparison performance of supervised machine learning random forest and decision tree algorithms using confusion matrix. *Jurnal Sisfokom (Sistem Informasi dan Komputer), 13*(1), 39–45. https://doi.org/10.32736/sisfokom.v13i1.1985

[20] Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. African Journal of Biomedical Research, 27(4S). https://doi.org/10.4314/ajbr.v27i4s.4345

[21] Malik, Y. (2024). Akurasi dan presisi analisis kadar nikel (Ni) pada sampel nikel laterit menggunakan X-Ray Fluorescence Spectometry (XRF). *Sains: Jurnal Kimia dan Pendidikan Kimia*, 12(2), 87–94. https://doi.org/10.47191/sains.v12i2.47

[22] Nengsih, W. (2020). *Analisa recall dan precision menggunakan VSM pada kasus text mining*. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 5(1). https://doi.org/10.30743/infotekjar.v5i1.2663

[23] Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Zami, F. A. (2022). Analisis perbandingan algoritma Naive Bayes Classifier dan Support Vector Machine untuk klasifikasi hoax pada berita online Indonesia. *Jurnal Masyarakat Informatika (JMASIF)*, 13(2), 85–94. https://doi.org/10.14710/jmasif.13.2.2022.85-94