

Klasifikasi Mahasiswa Berprestasi Menggunakan *Fuzzy C-Means* Dan *Naive Bayes*

Rezki Nurul Jariah S.Intam¹⁾, Wulandari²⁾, Andi Akram Nur Risal^{3)*}, Dewi Fatmarani Suriyanto⁴⁾

^{1), 2), 3), 4)} Teknik Komputer, Universitas Negeri Makassar

Jl. Mallengkeri Raya, Parang Tambung, Kec. Tamalate, Kota Makassar, Sulawesi Selatan
Email : rezkinuruljariah@gmail.com¹⁾, wlndry.nurdin@gmail.com²⁾, akramandi@unm.ac.id³⁾,
dewifatmaranis@unm.ac.id⁴⁾

ABSTRACT

Success in the world of education is often associated with successful academic achievements. Therefore, processing information is very important to determine the selection of students who excel. However, study programs and student services often face difficulties in recognizing students who have achievements. In this research, outstanding students from the Faculty of Engineering, Makassar State University were determined using the Naive Bayes classification method combined with the Fuzzy C-Means (FCM) method to identify data patterns before classification. The criteria measured are GPA, achievements achieved, organizations attended, and the number of Semester Credit Units (SKS) that have been programmed. By using the Confusion Matrix, the evaluation results show an accuracy level of 98%, recall of 97%, precision of 100%, and F1-Score of 99%.

Keywords : Classification, Student, Performance, Fuzzy C-Means, Naive Bayes

ABSTRAK

Keberhasilan dalam dunia pendidikan seringkali dikaitkan dengan prestasi akademik yang berhasil dicapai. Oleh karena itu, pengolahan informasi sangat penting dilakukan untuk menentukan pemilihan mahasiswa yang berprestasi. Namun, seringkali program studi dan layanan mahasiswa menghadapi kesulitan dalam mengenali mahasiswa yang memiliki prestasi. Pada penelitian ini dilakukan penentuan mahasiswa berprestasi Fakultas Teknik Universitas Negeri Makassar menggunakan metode klasifikasi Naive Bayes yang dikombinasikan dengan metode Fuzzy C-Means (FCM) untuk melakukan identifikasi pola data sebelum diklasifikasi. Kriteria yang diukur adalah nilai IPK, Prestasi yang pernah dicapai, organisasi yang pernah diikuti, dan jumlah Satuan Kredit Semester (SKS) yang telah diprogram. Dengan menggunakan Confusion Matriks, hasil evaluasi menunjukkan tingkat akurasi sebesar 98%, recall sebesar 97%, precision sebesar 100%, dan F1-Score sebesar 99%.

Kata Kunci : Klasifikasi, Mahasiswa, Prestasi, Fuzzy C-Means, Naive Bayes

1. Pendahuluan

Menurut Kamus Besar Bahasa Indonesia (KBBI), Prestasi merujuk kepada pencapaian yang telah diperoleh sebagai hasil dari upaya, pekerjaan, atau tindakan yang telah dilakukan. Suatu individu dapat dianggap berprestasi, jika mereka memiliki kemampuan untuk bersaing dan menduduki posisi teratas di antara individu-individu yang lain (Hakim, 2019). Bagi mahasiswa, prestasi merupakan hal yang sangat penting karena dapat digunakan untuk menunjukkan potensi yang mereka miliki (Suwardi et al., 2021).

Keberhasilan dalam dunia pendidikan seringkali dikaitkan dengan prestasi akademik yang berhasil dicapai (Kusumastuti, 2020). Oleh karena itu, faktor-faktor yang mempengaruhi, memprediksi, memediasi, atau menjadi penyebab variasi prestasi akademik sangatlah penting untuk menentukan mahasiswa berprestasi (Kusumastuti, 2020). Indikator prestasi yang dicapai ditunjukkan melalui Indeks Prestasi (IP) dan Indeks Prestasi Kumulatif (IPK) serta tingkat ketepatan dalam menyelesaikan studi (Suwardi et al., 2021).

Pada umumnya, penentuan mahasiswa berprestasi dilakukan untuk memberi penghargaan ataupun untuk menentukan mahasiswa yang berhak memperoleh beasiswa berprestasi. Oleh karena itu, pengolahan informasi sangat penting dilakukan untuk menentukan pemilihan mahasiswa yang berprestasi. Hal ini tentunya melibatkan pemetaan mahasiswa berprestasi berdasarkan kriteria penilaian seperti Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), serta partisipasi dalam lomba akademik atau non-akademik (Hakim, 2019). Namun, seringkali program studi dan layanan mahasiswa menghadapi kesulitan dalam mengenali mahasiswa yang memiliki prestasi (Sovia et al., 2020).

Pada penelitian sebelumnya, telah dilakukan penelitian menggunakan klasifikasi *naive bayes*, dengan berbagai studi kasus dan data yang berbeda-beda, penelitian tersebut antara lain yaitu, penelitian pertama menggunakan metode *naive bayes* untuk mengklasifikasikan pencapaian nilai mahasiswa. Pada penelitian ini menggunakan 5 atribut antara lain yaitu Nomor Induk Mahasiswa (NIM), mata kuliah, Bobot nilai, kode mata kuliah, dan data nilai mahasiswa, dimana data-data tersebut diperoleh secara langsung dari operator jurusan. Hasil dari penelitian ini memiliki tingkat akurasi sebesar 99% (Batuallo et al., 2023). Kemudian penelitian berikutnya guna memprediksi mahasiswa berprestasi dengan menggunakan memperhitungkan akumulasi nilai dari semester satu sampai dengan nilai yang sedang ditempuh. Hasil dari penelitian ini menyatakan bahwa pengklasifikasian mahasiswa berprestasi memberikan hasil yang valid dan mendapatkan akurasi yang tinggi (Puspita & Aminah, 2022). Selanjutnya, penelitian untuk menentukan konsentrasi mahasiswa menggunakan variabel usia, jenis kelamin, Indeks Prestasi Kumulatif (IPK). Penelitian ini mengklasifikasikan mahasiswa ke dalam 3 konsentrasi yaitu Pemrograman, Multimedia, dan Jaringan Komputer. Hasil dari penelitian ini memiliki akurasi sebesar 84.27% (Wibowo et al., 2022).

Penelitian berikutnya masih menggunakan metode yang sama yaitu *naive bayes* guna mengklasifikasikan kelulusan mahasiswa tepat waktu. Kriteria yang digunakan dalam penelitian ini yaitu NIM, IPK, dan Satuan Kredit Semester (SKS). Hasil dari penelitian ini mendapatkan hasil akurasi sebesar 75% (Yustira et al., 2021). Kemudian, terdapat pula penelitian guna menyeleksi penerima beasiswa PPA, dimana dalam penelitian ini menggunakan 3 kriteria yaitu IPK, keaktifan organisasi, dan pekerjaan orang tua. Hasil dari penelitian ini berhasil mengambil keputusan untuk menentukan mahasiswa yang layak menerima beasiswa dengan menggunakan kriteria yang telah ditetapkan sebelumnya (Dita & Chairunisyah, 2021). Namun pada kelima penelitian diatas hanya menggunakan 1 metode saja dan data yang digunakan tergolong sedikit.

Oleh karena itu, dalam penelitian ini, penentuan mahasiswa berprestasi di Fakultas Teknik Universitas Negeri Makassar dilakukan dengan menggabungkan metode klasifikasi Naive Bayes dan metode Fuzzy C-Means (FCM) untuk mengidentifikasi pola data sebelum dilakukan klasifikasi. Kriteria yang diukur adalah nilai IPK, Prestasi yang pernah dicapai, organisasi yang pernah diikuti, dan jumlah SKS yang dilulusi. Adapun untuk penentuan tingkat akurasinya, dilakukan perbandingan dari hasil pengelompokan menggunakan metode *Fuzzy C-Means* dengan hasil pengklasifikasian menggunakan metode *Naive Bayes*, dan dievaluasi menggunakan *confusion matrix*.

2. Tinjauan Pustaka

2.1. Prestasi Mahasiswa

Prestasi mahasiswa merujuk pada pencapaian atau pengakuan yang diperoleh oleh mahasiswa dalam berbagai bidang, baik dalam hal akademik maupun non-akademik (Mona & Yunita, 2021). Beberapa kriteria umum yang biasanya digunakan untuk menilai prestasi mahasiswa meliputi :

- a. IPK (Indeks Prestasi Kumulatif), Merupakan ukuran yang sering digunakan untuk mengukur prestasi akademik mahasiswa. Semakin tinggi IPK, semakin baik prestasi akademik mahasiswa. IPK yang tinggi dapat menunjukkan tingkat pemahaman dan pencapaian akademik yang kuat (Sari et al., 2021).
- b. Jumlah SKS yang dilulusi, mencerminkan komitmen dan usaha dalam menyelesaikan mata kuliah. Mahasiswa yang menyelesaikan lebih banyak SKS dengan nilai yang baik menunjukkan kedisiplinan dan ketekunan dalam mencapai prestasi akademik (Sari et al., 2021).
- c. Partisipasi dalam organisasi, mahasiswa dapat berperan dalam pengembangan keterampilan interpersonal, kepemimpinan, dan kemampuan berkolaborasi. Aktivitas aktif dalam organisasi-organisasi ini mencerminkan prestasi di luar aspek akademik dan kemampuan untuk memberikan kontribusi dalam kehidupan kampus (Hendra, 2018).

d. Prestasi yang pernah dicapai, seperti prestasi di bidang olahraga, seni, kegiatan sosial, atau prestasi akademik tambahan, juga memiliki dampak penting. Ini tidak hanya menambah nilai, tetapi juga mewakili dedikasi dan prestasi di luar lingkungan kelas (Tl et al., 2017).

2.2. Clustering

Clustering adalah proses mengelompokkan sekelompok objek ke dalam kelas atau pola objek yang memiliki kesamaan. Clustering adalah teknik yang sangat signifikan dalam analisis data (Kurniawan & Defit, 2020). Clustering memisahkan data ke dalam kelompok-kelompok berdasarkan kemiripan atribut-atribut di antara kumpulan data tersebut, dengan atribut yang serupa tersebut direpresentasikan sebagai titik-titik dalam ruang multidimensi (Alhapizi et al., 2020). Dalam clustering terdapat beberapa metode pengelompokan salah satunya yaitu metode Fuzzy C-Means.

Pengelompokan menggunakan algoritma FCM merupakan metode pengelompokan berbasis perangkat lunak yang menghitung probabilitas (tingkat keanggotaan) dari suatu titik data tertentu dalam kelompok, yang terdiri dari poin-poin dengan tingkat kesamaan yang cukup besar. Dalam algoritma FCM, perhitungan jarak digunakan untuk mengukur sejauh mana kemiripan antara titik data untuk menentukan probabilitas bahwa titik data tersebut masuk dalam suatu cluster tertentu. Algoritma FCM konvensional ini berlandaskan pada penggunaan jarak Euclidean (Krasnov et al., 2023). Adapun langkah-langkah dari FCM antara lain (Fadiah et al., 2023) :

1. Menginputkan dataset. Langkah pertama adalah melakukan penginputan data yang telah terkumpul. Dataset tersebut terdiri dari matriks yang berukuran $n \times m$ (dengan n sebagai jumlah sampel data, dan m sebagai jumlah atribut setiap data). X_{ij} data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, \dots, m$).
2. Menentukan beberapa parameter:
 - a) Jumlah Cluster = 2
 - b) Pangkat = 3
 - c) Maksimum Iterasi = 1000
 - d) Kesalahan terkecil yang diharapkan = $1e-5$
 - e) Fungsi Objektif awal = $P_0 = 0$
 - f) Iterasi awal = $t = 1$
3. Membangkitkan tingkat keanggotaan secara acak dengan rumus μ_{ik} , di mana $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; Tingkat keanggotaan ini kemudian menjadi bagian dari matriks partisi awal U .
4. Menghitung pusat cluster dengan Menggunakan rumus yang telah ditentukan, yaitu

$$V_{kj} = \frac{\sum_i^n = 1 (\mu_{ik})^w \times x_{ij}}{\sum_i^n = 1 (\mu_{ik})^w} \dots\dots(1)$$

5. Menghitung fungsi objektif (Pt) menggunakan rumus:

$$P_t = \sum_{t=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \dots\dots(2)$$

6. Menghitung perubahan matriks partisi

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w}{\sum_{k=1}^c \sum_{k=1}^m \left[(X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \dots\dots(3)$$

7. Memeriksa kondisi berhenti dengan memeriksa perbedaan antara P_t dan P_{t-1} . Jika perbedaan tersebut kurang dari nilai kesalahan terkecil atau jumlah iterasi telah mencapai batas maksimum ($t > \text{MaxIter}$), maka iterasi dihentikan. Jika kondisi tersebut tidak terpenuhi, langkah berikutnya adalah menambahkan nilai t sebanyak 1 dan mengulangi Langkah ke-4 dari algoritma FCM.

2.3. Klasifikasi

Klasifikasi merupakan teknik dalam pembelajaran terawasi (*supervised learning*) yang digunakan untuk memprediksi kategori asal suatu objek (Hakim Tanjung, 2022). Klasifikasi melibatkan penggambaran dan penentuan perbedaan antara kelas-kelas data dengan tujuan agar model dapat memprediksi kelas dari suatu objek yang labelnya belum diketahui. Proses klasifikasi meliputi dua tahap utama, yaitu pengembangan model klasifikasi dan penerapannya dalam memprediksi kelas dari data baru.

Salah satu teknik klasifikasi yang banyak digunakan adalah metode *Naïve Bayes*, yang terkenal, dikarenakan metode ini menunjukkan tingkat akurasi yang tinggi dan mampu mengklasifikasi sejumlah besar data (Wikarsa et al., 2022). Metode *Naïve Bayes* didasarkan pada penerapan teorema Bayes, yang merujuk pada perhitungan probabilitas yang ditemukan oleh Thomas Bayes (Putri & Wijayanto, 2022). Teknik ini melibatkan perhitungan probabilitas untuk kategori tertentu dan seluruh data yang tersedia, dan perhitungan probabilitas ini dapat diwujudkan dalam persamaan (Fanani, 2020):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots\dots(4)$$

Keterangan:

- B = Data yang memiliki kelas yang belum diketahui
- A = Asumsi bahwa data B termasuk dalam suatu kelas spesifik.
- $P(A|B)$ = Probabilitas dari A berdasarkan kondisi B.
- $P(B|A)$ = Probabilitas dari B berdasarkan kondisi A
- $P(A)$ = Probabilitas dari A
- $P(B)$ = Probabilitas dari B

Tahapan Algoritma *Naïve Bayes* (Fanani, 2020):

1. Membaca data latih
2. Menghitung jumlah kelas.
3. Menghitung jumlah kasus yang sama dengan kelas yang sama (Probabilitas).
4. Mengalihkan semua nilai hasil sesuai dengan data baru yang dicari kelas-nya. Bandingkan hasil class pilih dengan nilai terbesar.

2.4. *Confusion Matrix*

Confusion Matrix merupakan alat ukur yang digunakan untuk mengukur sejauh mana kemampuan mengklasifikasi dalam mengidentifikasi fitur dari berbagai kelas (Hardoni et al., 2021). *Confusion matrix* berbentuk matriks dua dimensi yang memvisualisasikan perbandingan antara hasil prediksi dengan realitas, seperti yang diperlihatkan pada tabel berikut (Alkhussayid & Ferdiansyah, 2022).

Tabel 1. *Confusion Matrix*

Classification	Group	
	+	-
+	<i>True Positives</i>	<i>False Negatives</i>
-	<i>False Positives</i>	<i>True Negatives</i>

Perhitungan akurasi, presisi, recall pada *confusion matrix* (Alkhussayid & Ferdiansyah, 2022).

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots(5)$$

$$Presisi = \frac{TP}{TP+FP} \dots\dots(6)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots(7)$$

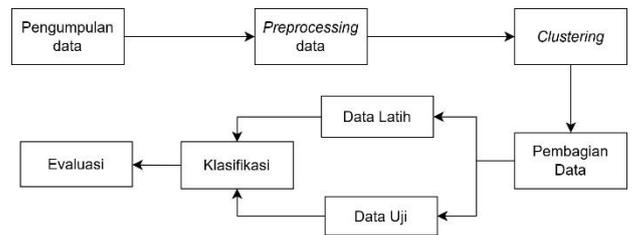
Keterangan:

- a. Akurasi Merupakan rasio prediksi yang benar dari keseluruhan prediksi..
- b. Presisi adalah proporsi kasus positif yang terklasifikasi dengan benar dari semua kasus yang diklasifikasi sebagai positif.
- c. TP (*True Positive*) adalah jumlah kasus positif yang diklasifikasikan dengan benar oleh pengklasifikasi.
- d. TN (*True Negative*) adalah jumlah kasus negatif yang diklasifikasikan dengan benar oleh pengklasifikasi.

- e. FP (*False positive*) adalah jumlah kasus negatif yang salah diklasifikasikan sebagai positif oleh pengklasifikasi.
- f. FN (*False Negative*) adalah jumlah kasus positif yang salah diklasifikasikan sebagai negatif oleh pengklasifikasi.

3. Metodologi

Dalam Penelitian ini, digunakan metode yang melibatkan serangkaian proses secara berurut, antara lain yaitu tahap pengumpulan data, tahap *preprocessing*, tahap *clustering*, tahap pembagian data, tahap klasifikasi, dan terakhir yaitu tahap evaluasi. Tahap-tahap tersebut dapat dilihat pada Gambar 1.



Gambar 1. Tahap-Tahap Penelitian

3.1. Pengumpulan data

Pada tahap ini, dilakukan proses pengumpulan data yang akan digunakan untuk klasifikasi. Data yang dikumpulkan mencakup informasi mengenai IPK, jumlah SKS, partisipasi dalam organisasi, dan prestasi yang pernah dicapai oleh mahasiswa semester 5. Pengumpulan data dilakukan melalui distribusi kuesioner kepada mahasiswa Fakultas Teknik Universitas Negeri Makassar. Dari distribusi kuesioner tersebut, berhasil dikumpulkan data dari 300 responden.

3.2. *Preprocessing*

Pada tahap ini dilakukan proses *preprocessing* data pada kolom IPK, Penghargaan, Organisasi, dan SKS, dimana tahap ini dilakukan untuk membersihkan data, menghilangkan kesalahan maupun duplikasi yang terdapat pada data yang telah terkumpul, dan juga menyamakan format pada data agar data mudah untuk diimplementasikan ke dalam algoritma. dimana format data yang digunakan yaitu, menggunakan titik pada nilai IPK, dan menggunakan angka numerik di setiap data.

3.3. *Clustering*

Tahap *clustering* dilakukan untuk memberikan label pada data dengan menggunakan metode *clustering Fuzzy C-Means*. Pada awal tahap *clustering*, jumlah *cluster* diinisialisasi dengan 2. Setelah proses *clustering* selesai, langkah selanjutnya adalah mengidentifikasi *cluster* yang mewakili mahasiswa berprestasi dan *cluster* yang mewakili mahasiswa yang kurang berprestasi.

3.4. Pembagian Data

Pada tahap ini dilakukan pembagian data yang telah terkumpul ke dalam 2 jenis data yaitu data latih dan data uji, dimana data latih sebesar 70% dan data uji sebesar 30% dari data yang ada. Data latih digunakan untuk melakukan pelatihan atau pengajaran pada algoritma yang akan digunakan untuk mengklasifikasi, sedangkan data uji digunakan untuk menguji seberapa akurat algoritma yang telah dilatih.

3.5. Klasifikasi

Pada tahap ini, algoritma Naive Bayes diimplementasikan untuk mengelompokkan data. Penelitian ini menggunakan dua kelas, yaitu kelas mahasiswa berprestasi dan kelas mahasiswa kurang berprestasi. Algoritma Naive Bayes digunakan untuk mengklasifikasikan data berdasarkan atribut-atribut yang telah dikumpulkan sebelumnya, seperti IPK, jumlah SKS, partisipasi dalam organisasi, dan prestasi. Dengan menggunakan metode ini, penelitian dapat menentukan model klasifikasi setiap mahasiswa ke dalam salah satu dari dua kelas tersebut berdasarkan karakteristik mereka menggunakan data latih. Setelah itu dilakukan pengujian sejauh mana model *Naive Bayes* mampu

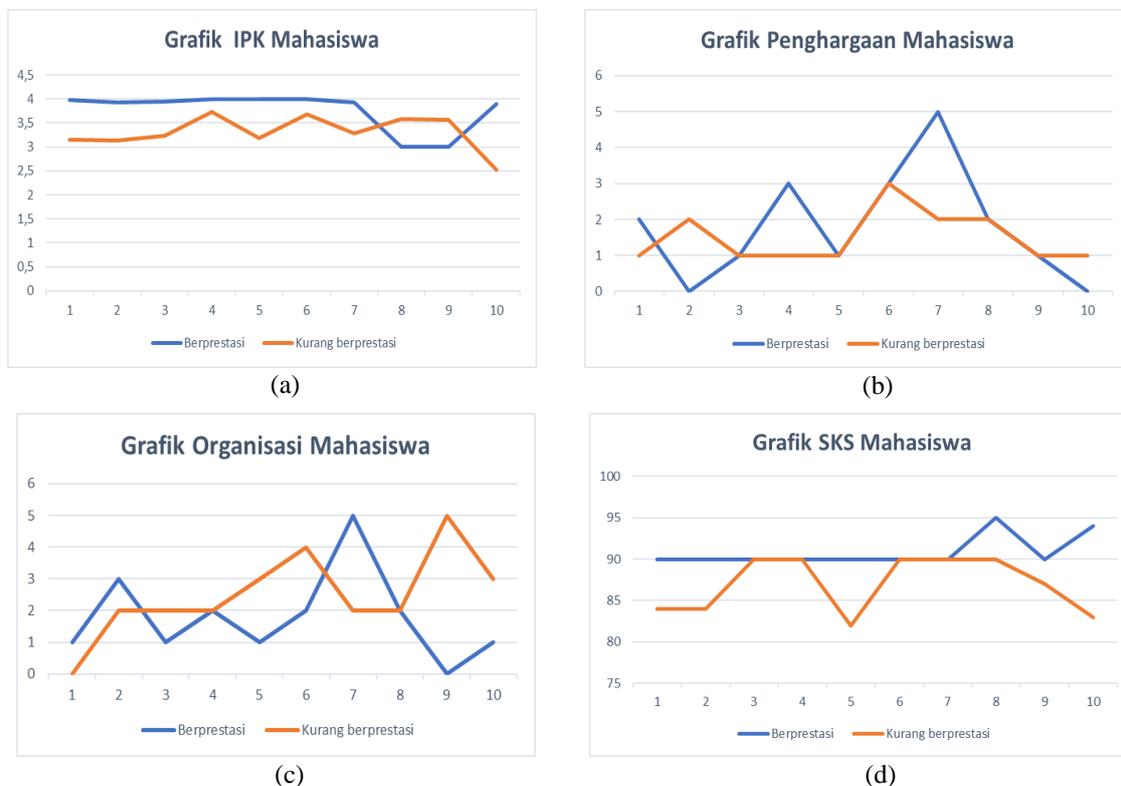
mengklasifikasikan mahasiswa dengan benar menggunakan data uji.

3.6. Evaluasi

Tahap evaluasi adalah langkah penting dalam pengembangan model klasifikasi untuk menilai sejauh mana kinerja model yang dibentuk. Dalam penelitian ini digunakan sebuah metode yang dikenal sebagai *confusion matrix*. *Confusion matrix* memberikan informasi dalam mengevaluasi sejauh mana model klasifikasi mampu memprediksi kelas data secara tepat (Xu et al., 2020). Dengan menggunakan matriks ini, kita dapat mengidentifikasi berapa banyak prediksi yang benar (*true positives* dan *true negatives*) serta sejauh mana model kita keliru dalam memprediksi kelas data (*false positives* dan *false negatives*). *Confusion matrix* juga dimanfaatkan untuk menghitung akurasi, presisi, *recall* dan *F1-score* dari model yang dibentuk.

4. Pembahasan

Dalam penelitian ini, sebelum melakukan tahapan klasifikasi, dilakukan tahap pelabelan data dengan menggunakan algoritma *Fuzzy C-means*. Hasil dari pelabelan dari 300 data mahasiswa, didapatkan data yang memiliki label “Kurang Berprestasi” sebanyak 72 data dan data yang memiliki label “Berprestasi” sebanyak 228 data.



Gambar 2. Grafik Data Mahasiswa Berprestasi dan Tidak Berprestasi

Berdasarkan Gambar 2(a), dalam grafik IPK, dapat diamati bahwa mahasiswa yang berprestasi memiliki rentang nilai antara 3 hingga 4, sementara mahasiswa yang kurang berprestasi memiliki rentang nilai mulai dari 2,5 dan tidak mencapai nilai 4. Pada Gambar 2(b) grafik penghargaan mahasiswa menunjukkan perbedaan yang signifikan antara mahasiswa yang berprestasi dan yang kurang berprestasi. Mahasiswa berprestasi memiliki pola naik turun yang sangat berubah dengan nilai berkisar dari 0 hingga 5, sementara mahasiswa yang kurang berprestasi memiliki rentang nilai antara 1 hingga 3. Kemudian pada Gambar 3(c) grafik organisasi mahasiswa, terlihat bahwa baik mahasiswa berprestasi

maupun yang kurang berprestasi memiliki rentang nilai yang sama, yaitu antara 0 hingga 5. Sedangkan Gambar 4(d) grafik SKS mahasiswa menunjukkan bahwa mahasiswa berprestasi memiliki rentang nilai antara 90 hingga 95, sementara mahasiswa kurang berprestasi memiliki rentang nilai antara 83 hingga 90.

Setelah melakukan tahapan pelabelan menggunakan *Fuzzy C-Means*, kemudian dilakukan tahapan implementasi klasifikasi dengan metode Naive Bayes. Pada penelitian ini, data yang digunakan akan dibagi menjadi data latih sebanyak 70% dan data latih sebanyak 30%.

Tabel 2. Data Latih

Data Ke-	IPK	Penghargaan	Organisasi	SKS	Label
282	3.18	3	2	68	Kurang Berprestasi
44	3.93	5	4	90	Berprestasi
272	3.16	1	2	64	Kurang Berprestasi
189	3.71	1	0	90	Berprestasi
....
251	2.70	0	0	64	Kurang Berprestasi
192	3.53	1	2	90	Berprestasi
47	3.93	0	1	90	Berprestasi

Pada tabel 2 menampilkan contoh data latih yang digunakan, dengan total 210 data latih. Dari jumlah tersebut, terdapat 158 data yang memiliki label

berprestasi, sementara 52 data lainnya memiliki label kurang berprestasi.

Tabel 3. Data Uji

Data Ke-	IPK	Prestasi	Organisasi	SKS	Label	Prediksi
208	3.60	2	2	90	Berprestasi	Berprestasi
188	3.56	3	2	90	Berprestasi	Berprestasi
161	2.86	2	3	82	Berprestasi	Kurang Berprestasi
221	3.50	1	2	90	Berprestasi	Berprestasi
....
136	2.66	3	2	83	Berprestasi	Kurang Berprestasi
255	3.13	1	3	61	Kurang Berprestasi	Kurang Berprestasi
239	2.64	1	2	68	Kurang Berprestasi	Kurang Berprestasi

Tabel 3 menampilkan contoh data uji yang digunakan untuk mengukur kinerja algoritma yang sudah di-latih sebelumnya. Total data uji yang digunakan adalah 90, dengan 70 data teridentifikasi sebagai mahasiswa berprestasi dan 20 data lainnya teridentifikasi sebagai mahasiswa kurang berprestasi.

Berikut ini adalah hasil implementasi dari klasifikasi menggunakan Naive Bayes untuk menentukan mahasiswa berprestasi.

Tabel 4. Confusion Matriks Hasil Pengujian

Data Aktual	Data Prediksi	
	Kurang Berprestasi	Berprestasi
Kurang Berprestasi	20	0

Berprestasi	2	68
-------------	---	----

Berdasarkan tabel , dapat dilihat bahwa hasil dari pengujian menggunakan algoritma naive bayes berhasil mengklasifikasikan mahasiswa berprestasi dan kurang berprestasi dimana pada data uji mahasiswa kurang berprestasi terdapat 20 data yang semuanya berhasil diprediksi. Adapun untuk mahasiswa berprestasi terdapat 70 data, namun dari hasil klasifikasi hanya 69 data yang berhasil diprediksi sebagai mahasiswa berprestasi dan 2 data salah diprediksi sebagai mahasiswa kurang berprestasi. Dapat dilihat pada tabel 2 pada data ke 161 dan 136, data tersebut yang seharusnya menjadi mahasiswa berprestasi, tetapi yang terprediksi menunjukkan mahasiswa kurang berprestasi, hal tersebut

terjadi karena dalam data menunjukkan bahwa nilai IPK-nya tidak mencapai standar yang biasanya dianggap sebagai ciri dari mahasiswa berprestasi. Dari hasil pengujian tersebut, didapatkan akurasi mencapai 98%, *recall* mencapai 97%, *precision* mencapai 100%, dan *F1-Score* mencapai 99%.

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan didapatkan sejumlah atribut yang dapat digunakan untuk

memprediksi mahasiswa berprestasi. Atribut tersebut meliputi jumlah IPK, jumlah penghargaan, jumlah organisasi, dan jumlah SKS. Dengan memanfaatkan Fuzzy C-Means dalam tahap preprocessing untuk memberikan label pada data, dan Naïve Bayes sebagai algoritma untuk memprediksi mahasiswa berprestasi, hasil evaluasi menunjukkan tingkat akurasi sebesar 98%, *recall* sebesar 97%, *precision* sebesar 100%, dan *F1-Score* sebesar 99%.

Daftar Pustaka

- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), Article 1. <https://doi.org/10.51519/journalsea.v1i1.10>
- Alkhussayid, M. D. & Ferdiansyah. (2022). Implementasi Algoritma K-Nearest Neighbors Pada Penentuan Jurusan Siswa. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(1), Article 1. <https://doi.org/2685-998X>
- Batuallo, C. R., Mewengkang, A., & Kaparang, D. R. (2023). Klasifikasi Pencapaian Nilai Mahasiswa Berdasarkan Jenis Mata Kuliah Menggunakan Algoritma Naive Bayesian Classifisier. *Eduetik : Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, 3(4), 499–511. <https://doi.org/10.53682/edutik.v3i4.7606>
- Dita, C. A. P., & Chairunisyah, P. (2021). Penerapan Naive Bayesian Classifier Dalam Penyeleksian Beasiswa PPA (2). 2(2), Article 2.
- Fadiah, N., Jariah, R. N., & Surianto, D. F. (2023). Analisis Penentuan Mahasiswa Berprestasi Fakultas Teknik UNM Menggunakan Metode Fuzzy C-Means. *Jurnal Sistem dan Teknologi Informasi (JustiIN)*, 11(3), 480. <https://doi.org/10.26418/justin.v11i3.66167>
- Fanani, M. R. (2020). Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa. *Jurnal DISPROTEK*, 11(1), Article 1.
- Hakim, L. (2019). Perancangan Aplikasi Penilaian Mahasiswa Berprestasi Universitas XYZ Menggunakan Algoritma K-Means Clustering. *Jurnal SITECH: Sistem Informasi dan Teknologi*, 2(1), Article 1. <https://doi.org/10.24176/sitech.v2i1.3337>
- Hakim Tanjung, D. Y. (2022). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Data Pengisian ATM. *INFOSYS (INFORMATION SYSTEM) JOURNAL*, 7(1), Article 1. <https://doi.org/10.22303/infosys.7.1.2022.12-24>
- 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), Article 1. <https://doi.org/10.30865/mib.v5i1.2616>
- Hendra, F. (2018). Peran Organisasi Mahasiswa Dalam Meningkatkan Mutu Pembelajaran Keterampilan Berbahasa Arab. *Arabiyat : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 5(1), Article 1. <https://doi.org/10.15408/a.v5i1.7480>
- Krasnov, D., Davis, D., Malott, K., Chen, Y., Shi, X., & Wong, A. (2023). Fuzzy C-Means Clustering: A Review of Applications in Breast Cancer Detection. *Entropy*, 25(7), Article 7. <https://doi.org/10.3390/e25071021>
- Kurniawan, H., & Defit, S. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal (2). 1(2), Article 2.
- Kusumastuti, D. (2020). Kecemasan dan Prestasi Akademik pada Mahasiswa. *Analitika*, 12(1), Article 1. <https://doi.org/10.31289/analitika.v12i1.3110>
- Mona, S., & Yunita, P. (2021). Faktor-Faktor yang Berhubungan Dengan Prestasi Belajar Mahasiswa.
- Puspita, D., & Aminah, S. (2022). IMPLEMENTASI NAIVE BAYES UNTUK SISTEM PREDIKSI MAHASISWA BERPRESTASI. *Jurnal Ilmiah Teknosains*, 8(2).
- Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, 11(1), Article 1. <https://doi.org/10.34010/komputika.v11i1.4350>
- Sari, P. N., Kurniawati, Y., & Amalita, N. (2021). Faktor yang Mempengaruhi Kualitas IPK Mahasiswa di Jurusan Matematika FMIPA UNP Menggunakan Regresi Logistik (1). 4(1), Article 1.
- Sovia, R., Mandala, E. P. W., & Mardhiah, S. (2020). Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 6(2), Article 2. <https://doi.org/10.26418/jp.v6i2.37759>
- Suwardi, D. M., Ahman, E., Machmud, A., & Iswanti, I. (2021). Hubungan Kecerdasan Emosional dan

- Spiritual Terhadap Prestasi Akademik Mahasiswa. *JPEKA: Jurnal Pendidikan Ekonomi, Manajemen dan Keuangan*, 5(1), Article 1.
<https://doi.org/10.26740/jpeka.v5n1.p61-70>
- Tl, D. I., Widowati, A. I., & Surjawati, S. (2017). Faktor-Faktor Yang Mempengaruhi Prestasi Akademik: Studi Kasus Pada Mahasiswa Program Studi Akuntansi Universitas Semarang. *Jurnal Dinamika Sosial Budaya*, 18(1), Article 1.
<https://doi.org/10.26623/jdsb.v18i1.557>
- Wibowo, M. F. S., Puspitasari, N. F., & Satya, B. (2022). Penerapan Data Mining Dan Algoritma Naïve Bayes Untuk Pemilihan Konsentrasi Mahasiswa Menggunakan Metode Klasifikasi. *Journal of Information System Management (JOISM)*, 3(2), Article 2.
<https://doi.org/10.24076/joism.2022v3i2.680>
- Wikarsa, L., Angdresey, A., & Kapantow, J. D. (2022). Implementasi Metode Naïve Bayes Dan Lexicon-Based Approach Untuk Mengklasifikasi Sentimen Netizen Pada Tweet Berbahasa Indonesia. *Jurnal Ilmiah Realtech*, 18(1), Article 1.
<https://doi.org/10.52159/realtech.v18i1.5>
- Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>
- Yustira, N., Witarsyah, D., & Sutoyo, E. (2021). *Implementasi Algoritma Naïve Bayes Classification Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu (Studi Kasus: Program Studi Sistem Informasi Universitas Telkom)*.