

# Pemanfaatan Data Mining Untuk Memprediksi Kelulusan Mata Kuliah dan Referensi Strategi Pembelajaran

**Andry Firdiansyah<sup>1</sup>, Ibnu Al Ikrom<sup>2</sup>, Moh. Khamdanni<sup>3</sup>, Wahyu Cahyo Utomo<sup>4</sup>**

<sup>1,2,3,4</sup>Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: <sup>1</sup>[1a9n1d1@gmail.com](mailto:1a9n1d1@gmail.com), <sup>2</sup>[ibnu.alikrom19@gmail.com](mailto:ibnu.alikrom19@gmail.com), <sup>3</sup>[khamdannimoh@gmail.com](mailto:khamdannimoh@gmail.com),

<sup>4</sup>[wahyu.utomo@unpkediri.ac.id](mailto:wahyu.utomo@unpkediri.ac.id)

**Abstrak** – Memanfaatkan teknologi dalam pendidikan bisa membantu kita agar lebih efektif dari segi tenaga maupun waktu. Pilihan strategi pembelajaran yang tepat, termasuk referensi strategi pembelajaran, merupakan langkah penting dalam perancangan kurikulum di tingkat akademik. Efektivitasnya berpengaruh pada bagaimana metode pengajaran dan pendekatan pembelajaran yang relevan memfasilitasi pemahaman. Penelitian menggunakan dataset dengan atribut nilai dalam mata kuliah yang pernah di tempuh mahasiswa yaitu Pns, Alpro 1, Alpro 2, Diskrit, PBO, dan Aljabar yang memengaruhi keputusan dan melibatkan proses pelatihan model Naive Bayes Dan bisa berpotensi memengaruhi keputusan. Oleh karena itu Metode Naive Bayes akan diimplementasikan untuk memodelkan hubungan probabilistik antara atribut dan kelas keputusan. Hasil penelitian ini diharapkan dapat memberikan wawasan praktis dengan berbasis aplikasi Android. Algoritma memiliki nilai evaluasi 90% accuracy, 1 recall, 89% precision, dan skor F1-0,67, dengan nilai tersebut maka Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat membuat prediksi kelulusan dalam mata kuliah data mining dengan kinerja yang baik.

**Kata Kunci** — Data Mining, Kelulusan, Naive Bayes

## 1. PENDAHULUAN

Pendidikan tinggi merupakan fondasi kritis dalam pembentukan sumber daya manusia yang berkualitas dan berdaya saing. Dalam upaya meningkatkan kualitas pendidikan, pengelolaan data mahasiswa menjadi krusial, terutama dalam memahami faktor-faktor yang memengaruhi kelulusan mereka. Seiring dengan perkembangan teknologi informasi, pendekatan Data Mining menjadi solusi efektif untuk menggali pola dan hubungan dalam data besar, termasuk data akademis. Suatu sistem dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa. Namun, beberapa perguruan tinggi tidak memiliki sistem untuk memprediksi keterlambatan kelulusan mahasiswa, yang membuat mereka tidak dapat mencegah hal ini terjadi. Memanfaatkan teknologi di era seperti ini adalah suatu hal yang dapat dipergunakan dimana dosen juga dapat lebih mengefisienkan tenaga maupun waktu dalam mengambil Keputusan.

Metode Naive Bayes memberikan kerangka kerja untuk menghitung probabilitas suatu kejadian berdasarkan informasi yang ada. Keunikan metode ini terletak pada gagasan naif bahwa setiap atribut atau fitur dalam data tidak bergantung satu sama lain (independent), Sementara asumsi bahwa atribut ini independen pada data sebenarnya jarang terjadi, hal ini dibuktikan oleh sejumlah penelitian empiris.. [1] Beberapa peneliti telah membuat sebuah sistem untuk memprediksi tingkat kelulusan mahasiswa. Dalam penelitian [2] yang berjudul “Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa” yang dilakukan oleh Pratama, Wihandika, dan Ratnawati pada tahun 2018, menghasilkan akurasi sebesar 80,55% dengan menggunakan data training sebesar 170 data. Sedangkan pada penelitian [3] berjudul “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika” yang dilakukan oleh Putri dan Waspada pada tahun 2018, menghasilkan akurasi sebesar 62.44%. Jenis kelamin, asal daerah, IPK, dan TOEFL adalah beberapa variabel yang digunakan. Terdapat beberapa penelitian menggunakan metode naive bayes [4] [5] [6]. Hasil penelitian sebelumnya kurang akurat dan belum termasuk dalam kategori yang baik.

Oleh karena itu, solusi yang efektif diperlukan untuk membantu dosen membuat proses pengumpulan nilai menjadi lebih mudah dan efektif, sehingga dosen dapat memaksimalkan dalam mengefisienkan waktu untuk pengambilan keputusan menggunakan Metode *Naive Bayes*.

## 2. METODE PENELITIAN

### 2.1. Pengumpulan Data

Dengan menggunakan pendekatan analisis dokumen daftar nilai yang sudah ditempuh oleh mahasiswa, penelitian ini dapat memperoleh pemahaman mendalam terhadap konteks dan kompleksitas topik yang sedang diselidiki. Dengan demikian, penelitian ini dapat memperkuat hasilnya dengan memberikan dasar kuat untuk interpretasi dan generalisasi. penelitian dapat memastikan bahwa variabilitas dan kedalaman informasi yang

dibutuhkan untuk mendukung keputusan. Keseluruhan, analisis dokumen menjadi landasan yang kritis dalam memberikan konteks dan mendukung hasil penelitian ini.

Tabel 1. Contoh Data Nilai Mata Kuliah

Nama	PNS	Alpro1	Alpro2	Diskrit	PBO	Aljabar Linear	Data Mining
Alvi Nuril Hidayah	A	B+	A	B+	A	A	A
Wahyu Tia F.	A-	B+	A	B+	A	A	A
Mochamad Yuda T.	A	A	A	A	A	A	A
Moh. Khamdanni	A	B	B	A	B+	A-	A
Donny Firdani	A	A	A	A	A-	B+	B

Dari Tabel 1 menunjukkan bahwa delapan variabel digunakan dalam penelitian, yaitu: PNS, Alpro 1 dan 2, diskrit, PBO, Aljabar, dan Data Mining, dengan ketentuan 100-96 (A), 95-91 (A-), 90-86 (B+), 85-81 (B), 80-76 (B-), 75-71 (C+), 70-66 (C), 65-61 (C-), 60-56 (D+), 55-51 (D), 40-36 (D-), 35-31 (E+), 30-26 (E), 25-0 (E-)

### 2.2. Algoritma Naïve Bayes

*Naïve Bayes Classifier* dikenal sebagai pengklasifikasi *bayesian* sederhana dan telah menjadi model probabilistik yang penting dan telah berhasil dalam praktiknya. Walaupun memiliki asumsi independensi yang kuat, *Naïve Bayes Classifier* telah terbukti efektif dalam klasifikasi dalam bentuk teks, diagnosa medis dan manajemen kinerja computer. *Teorema Bayes* ditunjukkan pada persamaan 1 berikut [1].

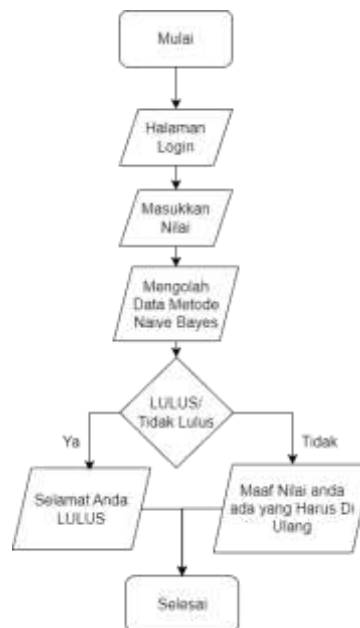
$$P(C|X) = \frac{P(X|C).P(C)}{P(X)}$$

Keterangan :

- P = Probabilitas
- C = Kelas atau Kondisi
- X = Variabel atau Pengamatan

### 2.3. Rancangan Sistem

Panduan yang membantu tim pengembang untuk membuat sistem yang berfungsi sesuai yang diinginkan. Perancangan sistem bisa diibaratkan sebagai langkah-langkah merancang rencana atau gambaran detail tentang bagaimana suatu sistem akan dibuat. Ini seperti membuat peta atau sketsa yang menjelaskan bagaimana semua bagian sistem akan berinteraksi. Saat merancang sistem, kita memikirkan bagaimana pengguna akan berinteraksi dengan sistem itu, bagaimana data akan disimpan, dan bagaimana berbagai bagian sistem akan bekerja satu sama lain. Perancangan sistem membantu kita membuat keputusan tentang teknologi apa yang akan kita gunakan dan bagaimana semuanya akan diatur sehingga sistem berjalan dengan baik.



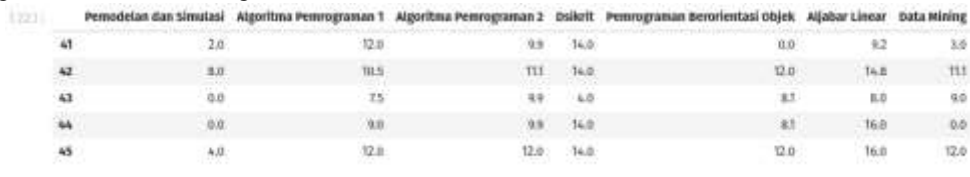
Gambar 1. Rancangan Sistem

Gambar 2.1 menunjukkan alur kerja dari aplikasi yang mendeteksi kelulusan yang menggunakan algoritma Naive Bayes berbasis android. Beginilah dengan melihat halaman dan memasukkan semua nilai dari data daftar nilai. Setelah semua data dimasukkan, hasilnya akan menunjukkan apakah Mahasiswa tersebut Lulus atau tidak.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Informasi yang digunakan dalam penelitian ini diperoleh melalui ekstraksi data dari website resmi universitas, yaitu <https://siakad2.unpkediri.ac.id/>. Gambar 2 menampilkan tangkapan layar dari dataset yang berisi informasi mengenai nilai mahasiswa.



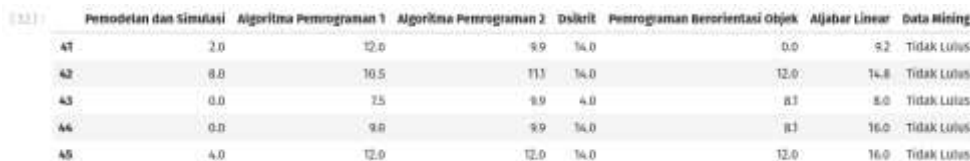
	Pemodelan dan Simulasi	Algoritma Pemrograman 1	Algoritma Pemrograman 2	Diskrit	Pemrograman Berorientasi Objek	Aljabar Linear	Data Mining
41	2.0	12.0	9.9	14.0	0.0	8.2	3.0
42	8.0	10.5	11.1	14.0	12.0	14.8	11.1
43	0.0	7.5	9.9	4.0	8.1	8.0	9.0
44	0.0	9.0	9.9	14.0	8.1	16.0	0.0
45	4.0	12.0	12.0	14.0	12.0	16.0	12.0

Gambar 2. Dataset

Pada gambar 2, terlihat contoh 5 data terakhir dalam dataset ini. Dalam kerangka penelitian ini, digunakan enam variabel untuk mengidentifikasi kelulusan dalam mata kuliah Data Mining. Variabel yang dipertimbangkan meliputi Pemodelan dan Simulasi, Algoritma dan Pemrograman 1, Algoritma dan Pemrograman 2, Diskrit, Pemrograman Berorientasi Objek, dan Aljabar Linear. Sementara itu, variabel Data Mining dijadikan variabel dependen. Dalam konteks penilaian, nilai yang sama dengan atau melebihi 5 mengindikasikan bahwa mahasiswa telah berhasil lulus mata kuliah Data Mining. Sebaliknya, nilai di bawah 5 menunjukkan bahwa mahasiswa belum memenuhi syarat untuk lulus dalam mata kuliah Data Mining.

#### 3.2. Transformasi Data

Proses transformasi data pada nilai mata kuliah Data Mining dilakukan dengan mengonversi setiap nilai mahasiswa. Jika nilai tersebut sama dengan atau lebih dari 5, hal ini mencerminkan bahwa mahasiswa telah berhasil lulus dalam mata kuliah Data Mining. Sebaliknya, apabila nilai berada di bawah 5, hal tersebut menunjukkan bahwa mahasiswa belum mencapai standar kelulusan untuk mata kuliah Data Mining.



	Pemodelan dan Simulasi	Algoritma Pemrograman 1	Algoritma Pemrograman 2	Diskrit	Pemrograman Berorientasi Objek	Aljabar Linear	Data Mining
41	2.0	12.0	9.9	14.0	0.0	8.2	Tidak Lulus
42	8.0	10.5	11.1	14.0	12.0	14.8	Tidak Lulus
43	0.0	7.5	9.9	4.0	8.1	8.0	Tidak Lulus
44	0.0	9.0	9.9	14.0	8.1	16.0	Tidak Lulus
45	4.0	12.0	12.0	14.0	12.0	16.0	Tidak Lulus

Gambar 3. Tranformasi Data

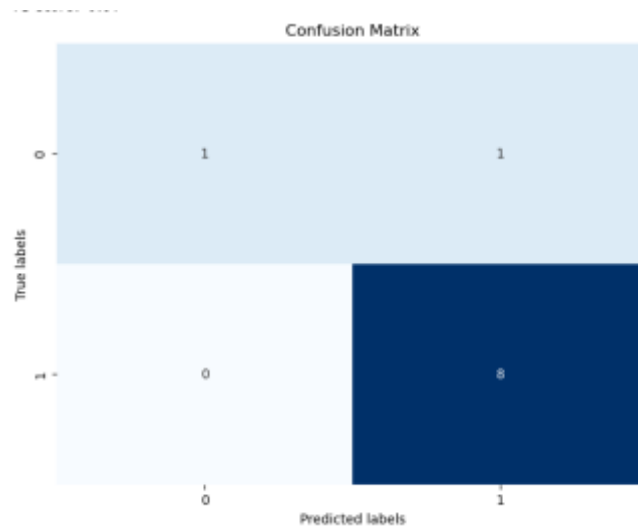
Pada gambar 3.2, dapat terlihat dengan jelas bagaimana setiap nilai mahasiswa pada mata kuliah Data Mining telah diubah. Nilai yang mencapai atau melebihi 5 menunjukkan bahwa mahasiswa berhasil lulus, sedangkan nilai di bawah 5 mengindikasikan bahwa mahasiswa belum memenuhi syarat kelulusan untuk mata kuliah Data Mining."

#### 3.3. Evaluasi Sistem

Evaluasi sistem menjadi suatu langkah kritis dalam mengevaluasi kinerja suatu sistem. Dalam proses evaluasi ini, beberapa metrik utama seperti akurasi (accuracy), presisi (precision), recall, dan F1 Score digunakan sebagai alat pengukur. Akurasi memberikan gambaran keseluruhan tentang sejauh mana sistem dapat memberikan prediksi yang benar. Presisi memberikan informasi tentang sejauh mana hasil positif dari sistem benar, sementara recall mengukur kemampuan sistem untuk mendeteksi seluruh hasil yang benar. F1 Score, sebagai gabungan dari presisi dan recall, memberikan perspektif yang seimbang mengenai kinerja sistem secara keseluruhan. Penggunaan metrik-metrik ini tidak hanya memberikan gambaran tentang keberhasilan prediksi, tetapi juga membantu dalam mengevaluasi sejauh mana sistem mampu mengatasi kesalahan dan memberikan solusi yang akurat.

### 3.3.1. Confusion Matrix

Dalam konteks judul proyek "Pemanfaatan Data Mining Untuk Memprediksi Kelulusan Mata Kuliah dan Referensi Strategi Pembelajaran", confusion matrix berperan penting sebagai alat evaluasi untuk mengukur seberapa efektif model dapat memprediksi kelulusan mahasiswa berdasarkan nilai mata kuliah tertentu. Confusion matrix memberikan gambaran yang jelas tentang keberhasilan model dalam mengidentifikasi mahasiswa yang lulus atau tidak lulus. Dengan menganalisis True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), kita dapat memahami sejauh mana model dapat memberikan prediksi yang akurat dan memberikan dasar untuk mengoptimalkan kinerja model guna meningkatkan akurasi prediksi kelulusan mahasiswa.



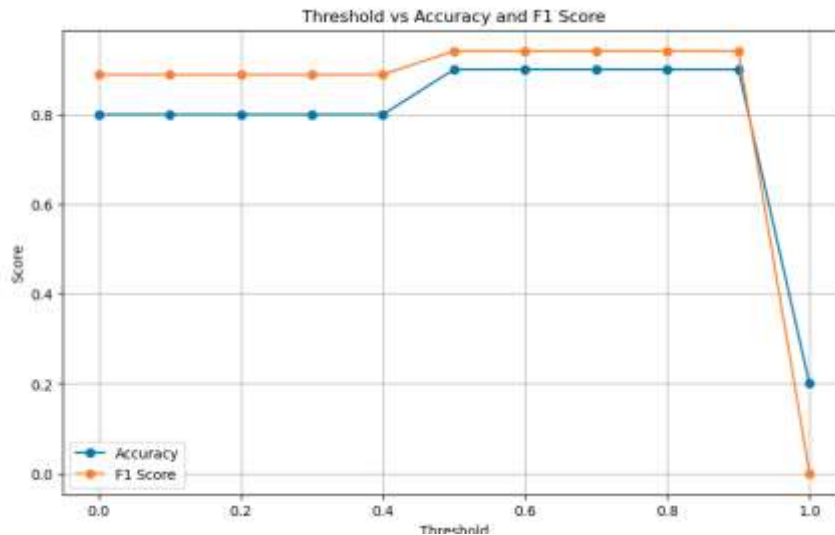
Gambar 4, Confusion Matrix

Mengacu pada Gambar 3.3, kita dapat melakukan perhitungan akurasi, recall, presisi, dan F1 Score-nya sebagai berikut :

- Accuracy =  $\frac{TP + TN}{\Sigma \text{ data}} = \frac{1+8}{1+1+0+8} = \frac{9}{10} = 0.90 \times 100\% = 90\%$
- Precision =  $\frac{TP}{TP+FP} = \frac{1}{1+1} = \frac{1}{2} = 0.5 \times 100\% = 50\%$
- Recall =  $\frac{TP}{TP+FN} = \frac{1}{1+0} = \frac{1}{1} = 1$
- F1 Score =  $\frac{TP}{TP + \frac{1}{2}(FP+FN)} = \frac{1}{1 + \frac{1}{2}(1+0)} = \frac{1}{1.5} = 0.67$

#### 3.2.1. Threshold vs Accuracy and F1 Score

Pada Gambar 3.4, memvisualisasikan bagaimana perubahan nilai ambang batas (threshold) pada model klasifikasi Naive Bayes dapat mempengaruhi metrik evaluasi seperti akurasi dan F1 Score. Nilai ambang batas yang lebih tinggi cenderung meningkatkan akurasi tetapi dapat mengorbankan recall, sementara nilai ambang batas yang lebih rendah dapat meningkatkan recall tetapi dengan biaya penurunan precision. Grafik membantu pemahaman tentang keseimbangan optimal antara precision dan recall, serta memungkinkan pemilihan threshold yang sesuai dengan kebutuhan spesifik tugas klasifikasi, terutama dalam kasus ketidakseimbangan kelas.



Gambar 5. Threshold vs Accuracy and F1 Score

### 3.2. Implementasi Algoritma Naive Bayes kedalam Django API

Algoritma Naive Bayes dapat diimplementasikan kedalam Application Programming Interface (API). Pada penelitian ini akan digunakan Django sebagai framework untuk membangun API Servernya. Berikut langkah implementasi algoritma Naive Bayes kedalam Django Framework API:

- Mengimpor semua library yang diperlukan untuk melaksanakan langkah-langkah implementasi algoritma.
- Memuat model algoritma Naive Bayes yang telah disiapkan sebelumnya.
- Melakukan request dan menampilkan respons dari API Django dalam format Javascript Object Notation (JSON). Percobaan dilakukan dengan menggunakan aplikasi Postman.



Gambar 3.1 Implementasi Algoritma Naive Bayes kedalam Django API

Keterangan :

Gambar 3.5 menunjukkan percobaan request terhadap Django API Server dan ditampilkan dalam bentuk URL.

Key	Value
<input checked="" type="checkbox"/> periode_simulasi	0
<input checked="" type="checkbox"/> algoritma_programan_1	0
<input checked="" type="checkbox"/> algoritma_programan_2	0
<input checked="" type="checkbox"/> deskrit	0
<input checked="" type="checkbox"/> perograman_benar_tata_objek	0
<input checked="" type="checkbox"/> aljabar_linear	7
Key	Value

Gambar 6. tampilan dalam bentuk URL

Keterangan :

Gambar 3.6 menampilkan key yang dikirim kepada Django API Server untuk mendapatkan hasil prediksi. Key disini adalah variabel independen yang digunakan sebagai predictor untuk mendapatkan hasil prediksi.

- Memeriksa hasil respons untuk memastikan apakah model algoritma telah dijalankan atau belum.
- Jika API Django berfungsi dengan baik, respons yang ditampilkan akan sesuai dengan model yang sudah dilatih, sebagaimana tampak dalam contoh gambar 8.



Gambar 7. Tampilan Sesuai Dengan Model Yang Sudah Dilatih

Keterangan :

Pada gambar 3.7 ditampilkan berupa response hasil prediksi dari Django API Server dalam bentuk JSON. Hasil prediksi ditampung dalam JSON Object dan dimuat dalam key bernama “prediction” dan “Tidak lulus” merupakan value dari keynya sekaligus hasil prediksi yang dihasilkan dari metode algoritma Naive Bayes di dalam Django API Server. Hasil menunjukkan bahwa dengan nilai Pemodelan dan Simulasi: 9, Algoritma dan pemrograman 1: 8, Algoritma dan Pemrograman 2: 8, Diskrit: 5, Pemrograman Berorientasi Objek: 6, serta Aljabar Linear: 7 diprediksi “Lulus” dalam mata kuliah Data Mining.

### 3.3. Tampilan Uji Coba Menggunakan Aplikasi Mobile

Tampilan antarmuka aplikasi mobile berbasis android ditampilkan pada Gambar X. Pada tampilan awal aplikasi disajikan 6 form yang harus diisi oleh user. Isian ini merupakan variabel-variabel dari keenam variabel predictor yang diperlukan untuk mendapatkan hasil prediksi.



Gambar 8. Tampilan Uji Coba Menggunakan Aplikasi Mobile

Pada Gambar 8 ditunjukkan hasil prediksi “Lulus” yang berarti aplikasi android sudah berjalan sesuai dengan apa yang diprediksikan didalam Django API Server. Hasil yang diprediksi hasilnya sama seperti pada aplikasi Postman. Hal ini menandakan bahwa aplikasi android sudah berhasil terhubung dengan Django API Server dengan baik.

#### 4. SIMPULAN

Penelitian ini fokus pada prediksi tingkat kelulusan mahasiswa pada mata kuliah Data Mining dengan menggunakan variabel prediktor, antara lain Pemodelan dan Simulasi, Algoritma Pemrograman 1, Algoritma Pemrograman 2, Diskrit, Pemrograman Berorientasi Objek (PBO), dan Aljabar Linear. Variabel-variabel ini merepresentasikan nilai mahasiswa, di mana nilai 5 atau lebih dianggap sebagai indikator kelulusan, sementara nilai di bawah 5 menandakan ketidakkelulusan. Eksperimen menggunakan algoritma Naive Bayes menghasilkan tingkat akurasi sebesar 90%, precision sebesar 50%, recall sebesar 1, dan F1 score sebesar 0.67. Temuan ini mengindikasikan bahwa pendekatan Naive Bayes efektif dalam memprediksi kelulusan mahasiswa pada mata kuliah Data Mining. Walaupun demikian, penelitian selanjutnya dapat mempertimbangkan integrasi metode lain guna meningkatkan akurasi prediksi dan mendalami pemahaman faktor-faktor yang memengaruhi kelulusan.

#### 5. SARAN

Peneliti menyadari bahwa penelitian ini masih banyak kekurangan dalam penulisan, sistem aplikasi, dan teori yang digunakan. Dari penelitian ini mungkin masih perlu atau adanya penelitian tambahan. Guna meningkatkan kualitas dan efektivitas proses pembelajaran, implementasi sistem pendukung keputusan dalam konteks akademis dan pengembangan strategi pembelajaran menjadi langkah yang strategis dengan memanfaatkan perkembangan jaman. Perlu terus mendorong inovasi dan pengembangan aplikasi yang lebih pintar dan terintegrasi. Seiring perkembangan teknologi, kita dapat menciptakan lingkungan belajar yang mendukung, dinamis, dan memberikan pengalaman belajar yang lebih baik bagi generasi mendatang.

#### 6. REFERENCES

- [1] D. A. Putra and K. Mia, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA," *Seminar Nasional Teknoka*, vol. V, p. 35, 2020.
- [2] A. Pratama, R. C. Wihandika and D. Eka, "Implementasi algoritme suport vector mechine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 4, pp. 1704-1708, 2018.
- [3] W. Musu, A. Ibrahim and H. Heriadi, "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5," *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi*, vol. 10, no. 1, pp. 186-195, 2021.
- [4] N. Purwati and A. D. Januanti, "Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naive Bayes," *Ilmu Komputer Unila Publishing Network all rights reserved*, vol. 2, no. 1, pp. 126-137, 2021.
- [5] N. Ramsari and A. R. Firmansyah, "SISTEM PENDUKUNG KEPUTUSAN UNTUK MENINGKATKAN MUTU KINERJA PROGRAM STUDI MELALUI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS FRAMEWORK LARAVEL (STUDI KASUS : FAKULTAS ILMU KOMPUTER DAN INFORMATIKA UNIVERSITAS NURTANIO BANDUNG)," *Program Studi Teknik Informatika*, pp. 6-18.
- [6] R. W. and B. P. Adhi, "Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa pad Program Studi Pendidikan Teknik Informatika dan Komputer dengan Menggunakan Algoritma Naive Bayes," *Jurnal Pintar*, vol. IV, no. 1, 2020.
- [7] R. P. S. Putri and I. Waspada, "Penerapan Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, vol. 4, no. 1, pp. 1-7, 2018.
- [8] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Technique*, Amsterdam: Elsevier, 2011.
- [9] D. Dahri, F. Agus and D. M. Khairina, "Metode Naive Bayes Untuk Penentuan Penerimaan Beasiswa Bidikmisi Universitas Mulawarman," *Jurnal Informatika Mulawarman*, vol. 11, no. 2, p. 29, 2016.
- [10] F. Elfaladonna and A. Rahmadani, "Analisa Metode Classification-Decission Tree dan Algoritma C4.5 untuk Memprediksi Penyakit Diabetes dengan Menggunakan Aplikasi Rapid Miner," *SINTECH (Science and Information Technology)*, vol. 2, no. 1, pp. 10-17, 2019.