

## Klasifikasi Pola Peminjam Buku Berdasarkan Profesi Menggunakan Algoritma Naïve Bayes

Febri Rosita Dewi<sup>1\*</sup>  
Ade Eviyanti<sup>2</sup>  
Arif Senja Fitrani<sup>3</sup>  
Ika Ratna IndraAstutik<sup>4</sup>

<sup>1,2,3,4</sup>Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Jl. Mojopahit No.666 B,  
Sidowayah, Celep, Kec. Sidoarjo, Kabupaten Sidoarjo, Jawa Timur, 61215, Indonesia.

<sup>1</sup>febrirositadewi@gmail.com, <sup>2</sup>adeeviyanti@umsida.ac.id, <sup>3</sup>asfjim@umsida.ac.id,

<sup>4</sup>ikaratna@umsida.ac.id

### \*Penulis Korespondensi:

Febri Rosita Dewi  
febrirositadewi@gmail.com

### Abstrak

Sebagai pusat literasi dan pembelajaran, perpustakaan menghadapi tantangan dalam memahami pola peminjaman buku untuk memenuhi kebutuhan pengguna yang beragam. Masalah utama yang dihadapi adalah kurangnya analisis berbasis data dalam mengoptimalkan layanan dan koleksi perpustakaan. Penelitian ini bertujuan untuk mengklasifikasikan pola peminjaman buku berdasarkan profesi dengan menerapkan algoritma Naive Bayes, memanfaatkan data dari Dinas Perpustakaan Sidoarjo tahun 2023. Data terdiri dari 4476 transaksi dengan atribut seperti profesi, kategori buku, dan tingkat minat baca. Penelitian ini dilaksanakan melalui beberapa fase, yaitu pengumpulan data, preprocessing, pemrosesan menggunakan algoritma Gaussian dan Multinomial Naive Bayes, serta evaluasi model. Dengan pengujian pada berbagai rasio data (90:10, 80:20, 75:25, dan 50:50), hasil menunjukkan bahwa Gaussian Naive Bayes memberikan akurasi tertinggi sebesar 97% pada skenario random dataset. Temuan utama menunjukkan bahwa profesi pelajar, mahasiswa, dan ibu rumah tangga mendominasi kategori minat baca tinggi, sedangkan dokter dan peneliti memiliki minat baca yang lebih rendah. Kontribusi unik penelitian ini terletak pada penerapan analisis berbasis data untuk mendukung pengelolaan perpustakaan. Hasil penelitian memberikan wawasan strategis untuk pengembangan layanan berbasis data yang lebih responsif, optimalisasi koleksi sesuai kebutuhan profesi, serta peningkatan efektivitas program literasi. Penelitian ini diharapkan menjadi langkah awal dalam memanfaatkan teknologi data mining untuk mengatasi tantangan modern dalam pengelolaan perpustakaan.

**Kata Kunci:** Klasifikasi; Minat Baca; Naïve Bayes; Perpustakaan; Profesi

### Abstract

*As centers of literacy and learning, libraries face challenges in understanding book lending patterns to meet the needs of diverse users. The main problem faced is the lack of data-based analysis in optimizing library services and collections. This research aims to classify book borrowing patterns based on profession using the Naive Bayes algorithm, utilizing data from the Sidoarjo Library Service in 2023. The data consists of 4476 transactions with attributes such as profession, book category, and level of reading interest. This research was conducted in several phases, namely data collection preprocessing, processing using Gaussian and Multinomial Naive Bayes algorithms, and model evaluation. By testing on various data ratios (90:10, 80:20, 75:25, and 50:50), the results show that Gaussian Naive Bayes provides the highest accuracy of 97% in the random dataset scenario. The main findings show that students, university students and housewives dominate the high reading interest category, while doctors and researchers have lower reading interest. The unique value of this research is in its application of data-based analysis to support library management. The research results provide strategic insight for developing more responsive data-based services, optimizing collections according to professional needs, and increasing the effectiveness of literacy programs. This research is anticipated to serve as the initial phase in utilizing data mining technology to overcome modern challenges in library management.*

**Keywords:** Classification; Library; Naïve Bayes; Profession; Reading Interest

---

## 1. Pendahuluan

Sebagai institusi layanan publik serta media literasi, Perpustakaan memegang peran signifikan dalam memperkaya wawasan publik. Sebagai media literasi, perpustakaan berfungsi sebagai

sumber pembelajaran sepanjang hayat yang mendukung perbaikan mutu pendidikan masyarakat[1]. Perkembangan teknologi, yang dipengaruhi oleh ilmu pengetahuan, membuka banyak peluang inovasi. Agar perpustakaan tetap relevan, penting untuk memahami karakteristik dan kebutuhan peminjam buku sehingga layanan dapat disesuaikan. Mengetahui pengguna dan pola peminjaman buku membantu perpustakaan mengoptimalkan koleksi dan layanannya.

Dinas Perpustakaan Sidoarjo di Jawa Timur menyediakan layanan perpustakaan dengan koleksi buku yang lengkap dan beragam, melayani masyarakat dari berbagai profesi seperti pelajar, mahasiswa, pegawai, wirausaha, dan lainnya. Data profesi peminjam buku dicatat dan disimpan sebagai data transaksi yang digunakan untuk mengklasifikasikan peminjam berdasarkan profesi, membantu perpustakaan mengelola koleksi, memenuhi kebutuhan informasi, dan mengembangkan koleksi buku sesuai kebutuhan peminjam[2].

Untuk mendukung upaya ini, diperlukan metode yang efektif dalam mengklasifikasikan peminjam buku berdasarkan profesi mereka. Salah satu teknik yang dapat dimanfaatkan adalah Teorema Bayes, khususnya algoritma Naive Bayes. Teorema Bayes digunakan untuk menghitung sejumlah probabilitas dengan menggabungkan nilai-nilai dataset yang tersedia.[3] Penelitian ini menggunakan dua varian dari model Naive Bayes, yaitu Gaussian Naive Bayes dan Multinomial Naive Bayes, yang dipilih berdasarkan karakteristik data yang akan diproses. Gaussian Naive Bayes digunakan untuk atribut numerik yang diasumsikan mengikuti distribusi normal, sementara Multinomial Naive Bayes lebih cocok untuk atribut kategori, seperti profesi, yang sering ditemukan dalam data teks atau data dengan kategori terbatas. Kedua model ini memiliki keunggulan masing-masing, di mana Gaussian Naive Bayes dapat menangani distribusi data yang kontinu dengan lebih baik, sedangkan Multinomial Naive Bayes lebih efisien dalam menangani data yang terdistribusi dalam kategori diskrit[4].

Metode Naive Bayes memiliki keunggulan karena cukup dengan sedikit data pelatihan untuk menghitung parameter yang wajib dalam proses klasifikasi. Dalam banyak situasi dunia nyata yang penuh tantangan, Naive Bayes biasanya berperforma lebih baik dari yang diharapkan .[5]

Beberapa riset sebelumnya memberikan konteks yang relevan untuk penelitian ini. Penelitian pertama menunjukkan bahwa penggunaan algoritma Apriori dapat mengidentifikasi kombinasi buku yang paling sering dipinjam, dengan temuan bahwa 32,56% peminjam buku fiksi juga meminjam buku tentang seni dekoratif [6]. Penelitian kedua menerapkan metode Naive Bayes untuk memprediksi minat baca berdasarkan usia, dengan akurasi 80%, menunjukkan efektivitas metode ini dalam aplikasi serupa[7]. Selanjutnya penelitian ketiga menggambarkan penggunaan algoritma C4.5 untuk klasifikasi minat baca, menghasilkan akurasi sebesar 88,03% [8] Sementara itu penelitian keempat menyatakan bahwa metode Naive Bayes efisien dalam mengklasifikasikan buku berdasarkan kategori, dengan akurasi 100% pada sistem yang diuji[9]

Namun, berdasarkan penjabaran penelitian terdahulu, dapat ditarik kesimpulan bahwa fokus penelitian sebelumnya lebih mengarah pada klasifikasi jenis buku, minat baca, atau pola peminjaman buku. Belum ada penelitian yang secara spesifik meneliti klasifikasi peminjam berdasarkan profesi. Penelitian ini berfokus untuk menjawab kekosongan tersebut dengan pemanfaatan data yang *up-to-date* dan melakukan pengambilan data secara manual di Dinas Perpustakaan Sidoarjo. Fokus pada klasifikasi peminjam berdasarkan profesi merupakan parameter baru yang belum pernah dijadikan fokus penelitian sebelumnya. Dengan demikian, gap penelitian ini terletak pada kurangnya kajian yang mengklasifikasikan peminjam buku berdasarkan profesinya, yang dapat memberikan wawasan lebih dalam mengenai preferensi peminjaman buku berdasarkan latar belakang profesi.

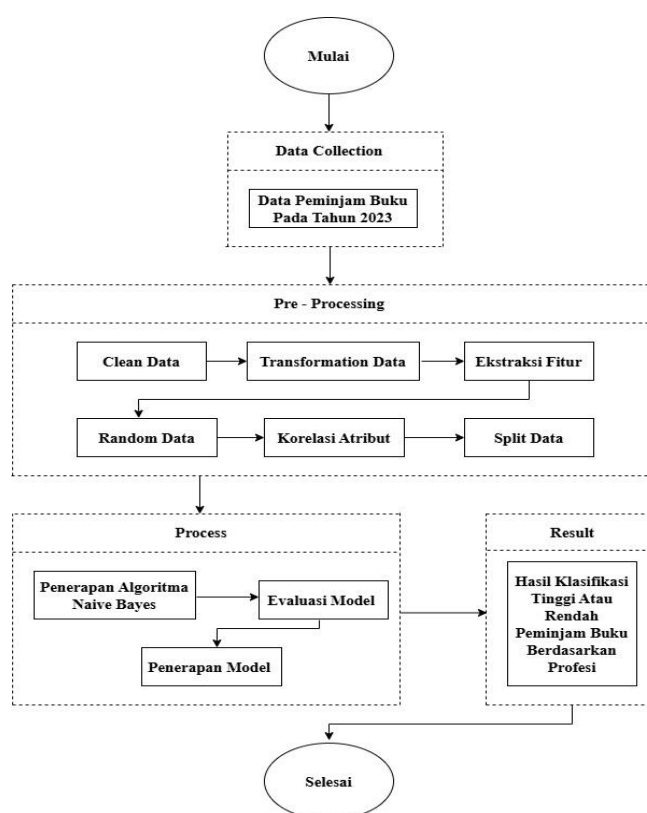
Proses klasifikasi ini memberikan wawasan tentang kebiasaan membaca peminjam dari berbagai profesi, seperti minat guru pada buku pendidikan atau literatur anak, serta minat profesional teknologi pada buku teknis dan sains. Hasil dari klasifikasi ini dapat digunakan untuk

mengoptimalkan koleksi, meningkatkan layanan, dan merancang program sesuai kebutuhan peminjam. Implementasi algoritma *Naive Bayes* di perpustakaan diharapkan dapat meningkatkan efektivitas dan efisiensi layanan perpustakaan [10]. Dengan memahami pola dan preferensi peminjam berdasarkan profesi, perpustakaan dapat mengembangkan strategi yang lebih tepat dalam penyediaan koleksi dan layanan.[11].

Secara keseluruhan, penerapan algoritma Naive Bayes dalam klasifikasi peminjam buku berdasarkan profesi di perpustakaan daerah diharapkan dapat mendukung tujuan utama perpustakaan sebagai pusat literasi dan pendidikan masyarakat.

## 2. Metode Penelitian

Penelitian ini berjenis eksperimen maka pendekatan yang dipakai dalam studi ini adalah metode naïve bayes sebagai model pengukuran untuk mengklasifikasi peminjam buku berdasarkan profesi peminjam buku. [12]



**Gambar 1.** Alur Metode Penelitian

Pada gambar 1 terdapat 4 tahapan pada penelitian ini yaitu data collection, pre-processing, process, dan result. Berikut penjelasan setiap tahapan tersebut. Langkah awal dimulai dengan proses pengumpulan data. Data didapatkan setelah melakukan wawancara langsung dengan pihak Dinas Perpustakaan dan Kearsipan Kabupaten Sidoarjo sehingga diketahui data transaksi peminjam buku. Data yang digunakan berasal dari Dinas Perpustakaan dan Kearsipan Kabupaten Sidoarjo, yang memiliki jumlah 4476 pada tahun 2023. Tabel 1 menunjukkan variabel – variable dataset yang terdiri dari 19 variabel dan 1 atribut target[13].

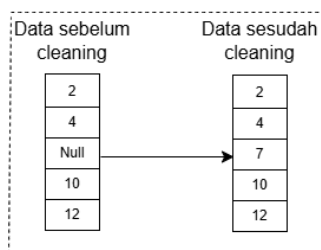
*Preprocessing* adalah langkah penting dalam analisis data yang bertujuan untuk menyiapkan data dipersiapkan untuk keperluan pemodelan. Pada penelitian ini, tahap preprocessing mencakup beberapa proses, termasuk *clean data*, *transformasi data*, ekstraksi fitur, random data, korelasi atribut, dan split data. Berikut adalah penjelasan detail tentang masing-masing tahapan. [14]

**Tabel 1.** Variabel Dataset

No. Variabel	Kelompok Variabel	Keterangan
X1	Id Peminjam	Id Traksaksi Peminjam [X1]
X2	Gender	Jenis Kelamin Peminjam [X2] 0= Laki – Laki, 1= Perempuan
X3	Lokasi	Lokasi Peminjaman Buku [X3] 0= Pusat
X4 – X13	Kategori Buku	Karya Umum [X4], Filsafat [X5], Agama [X6], Ilmu Sosial [X7], Bahasa [X8], Ilmu Murni [X9], Ilmu Terapan [X10], Kesenian [X11], Kesusastraan [X12], Sejarah [X13] 0 = Tidak, 1 = Ya
X14	Profesi	Profesi Peminjam Buku [X14] 0=Belum Bekerja, 1=Dokter, 2=Dosen, 3=Guru, 4=Ibu Rumah tangga, 5=Karyawan, 6=Lainya, 7=Mahasiswa, 8=Pegawai Negri, 9=Pegawai Swasta, 10=Pelajar, 11=Peneliti, 12=Pensiunan, 13=Perangkat Desa, 14=Seniman, 15=Swasta, 16=Tni/Polri, 18=Umum, 19=Wirawasta, 20=Wirausaha
X15	Usia	Usia Peminjam Buku [X15]
X16 - X17	Tanggal Pinjam dan Tanggal Kembali	Tanggal Pinjam Buku [16], Tanggal Kembali Buku [17]
Y	Minat Baca	Tingkat Minat Baca, Tinggi Atau Rendah 0=Rendah, 1=Tinggi
X19	Selisih Hari	Selisih Hari Pengembalian Buku [X19]

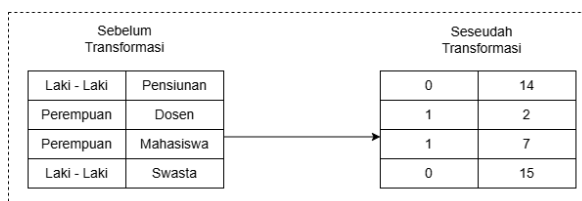
## a. Clean Data

Clean data atau pembersihan data merupakan usaha untuk menghapus kesalahan, inkonsistensi, dan informasi yang tidak relevan di mana data hilang dan duplikat diperiksa. Jika ada data hilang, mereka dapat diimputasi dengan median atau dihapus. [15]

**Gambar 2.** Clean Data

## b. Transformation Data

Data *transformation* di fase ini, data akan disesuaikan menjadi format yang relevan untuk data mining. Transformasi data merubah data category bakal dikelola menjadi format numerik. [16]

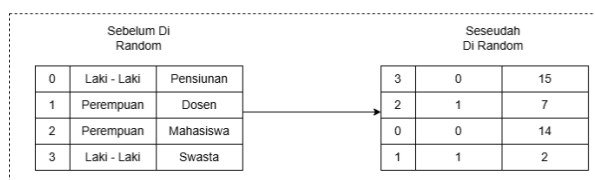
**Gambar3.** Transformation Data

c. Ekstraksi Fitur

Proses mengidentifikasi dan memilih fitur-fitur yang relevan untuk model klasifikasi.

d. Random Data

Random dataset adalah dataset yang elemen-elemen datanya disusun secara acak sehingga tidak memiliki pola tertentu. Dapat meningkatkan kinerja model, peningkatan validasi dan generalization.



**Gambar 4.** Random Dataset

e. Korelasi Atribut

Dalam data mining, correlation attribute merujuk pada tingkat hubungan atau keterkaitan di antara dua atau lebih variabel dalam suatu dataset. Korelasi antara atribut dapat membantu menemukan pola atau keterkaitan penting di antara data, yang kemudian bisa digunakan untuk pengambilan keputusan, pemodelan prediktif, atau pengelompokan.

f. Split Data

Data yang telah di proses selanjutnya dibagi 80 % sebagai data *training* dan 20% sebagai data *testing*. [17] Data *training* merupakan data yang dipakai untuk menjalankan prosedur pelatihan model. Data *testing* merupakan data yang digunakan untuk menilai sejauh mana model efektif. [18]

Process data dalam konteks pemodelan menggunakan algoritma naïve bayes yang mencakup serangkaian tahapan yang sistematis untuk mengolah data, menerapkan algoritma, dan mengevaluasi hasil. Berikut adalah rincian dari tahapan tersebut :

a. Penerapan Algoritma Naïve bayes

Naïve Bayes termasuk dalam teknik yang bisa diterapkan untuk mengelompokkan data. Klasifikasi Bayesian adalah teknik statistik yang digunakan untuk mengestimasi peluang keanggotaan suatu kelas [12]. Bentuk umum teorema bayes dapat dilihat pada persamaan seperti yang akan dijelaskan berikut :

$$P(H|x) = \frac{P(x|H)}{P(x)} P(H) \quad (1)$$

Keterangan Persamaan (1):

X : Data yang kategorinya belum ditetapkan

H : Hipotesis yang berkaitan dengan data

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (Probabilitas posterior)

P(H) : Probabilitas hipotesis H

(P(X|H) : Probabilitas X berdasarkan hipotesis H

P(X) : Probabilitas X

b. Evaluasi Model

Penilaian model dilakukan untuk mengukur performa model melibatkan penggunaan matriks seperti *confusion matrix* dan berbagai *matrix* evaluasi seperti akurasi, presisi, recall dan F1-score. Pada penelitian ini dengan menerapkan evaluasi 2 model yaitu model *Gaussian Naïve bayes* dan *Multinomial Naïve bayes* [19]. Berikut adalah perhitungan *confusion matrix*.

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+TN} \quad (4)$$

TP (True Positive) : Jumlah data yang memiliki nilai aktual positif dan diprediksi sebagai positif.  
 FP (False Positive) : Jumlah data yang memiliki nilai aktual negatif tetapi diprediksi sebagai positif.  
 FN (False Negative) : jumlah data yang memiliki nilai aktual positif namun diprediksi sebagai negatif.  
 TN (True Negative) : Jumlah data yang memiliki nilai aktual negatif dan diprediksi sebagai negatif.

### c. Penerapan Model

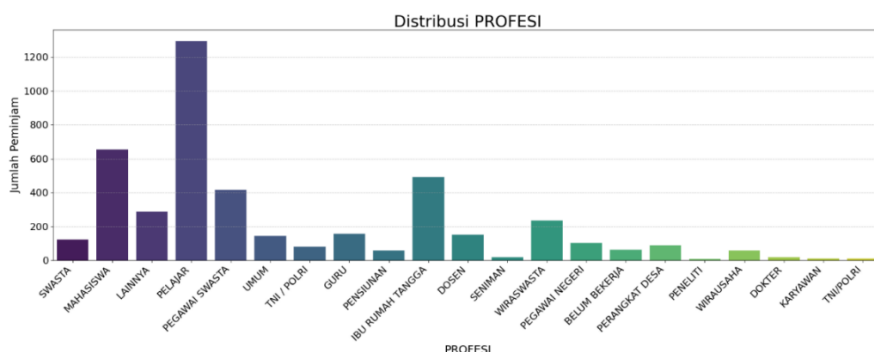
Setelah model dievaluasi dan dioptimalkan, model yang telah dilatih diterapkan pada data baru untuk mengklasifikasikan profesi peminjam. Hasil klasifikasi dapat digunakan untuk memberikan rekomendasi buku yang sesuai dengan profesi peminjam.

Result adalah hasil klasifikasi peminjam buku berdasarkan waktu pengembalian yang dikelompokkan menjadi dua kategori :

- Tinggi : peminjam yang mengembalikan buku dalam waktu kurang dari 10 hari. Ini menunjukkan bahwa peminjam memiliki disiplin yang baik dan minat baca yang tinggi.
- Rendah : peminjam yang mengembalikan buku dalam waktu lebih dari 10 hari. Ini dapat mengindikasikan kurangnya minat atau ketidakdisiplinan dalam pengembalian buku.

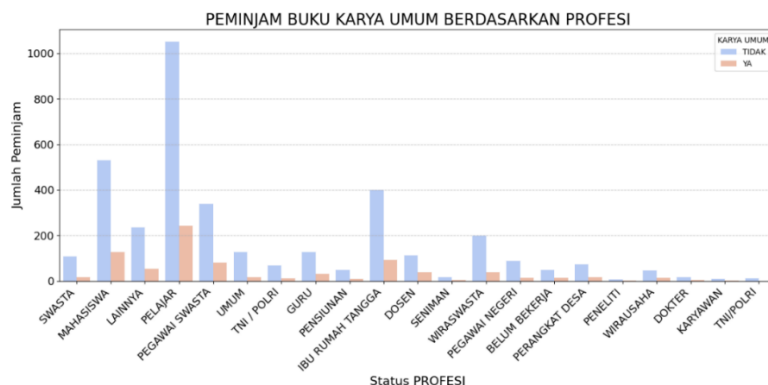
## 3. Hasil dan Pembahasan

Penelitian ini mengungkapkan beragam profil peminjam buku di perpustakaan berdasarkan data yang disajikan dalam bentuk grafik. Berikut pembahasan hasil yang ditemukan:



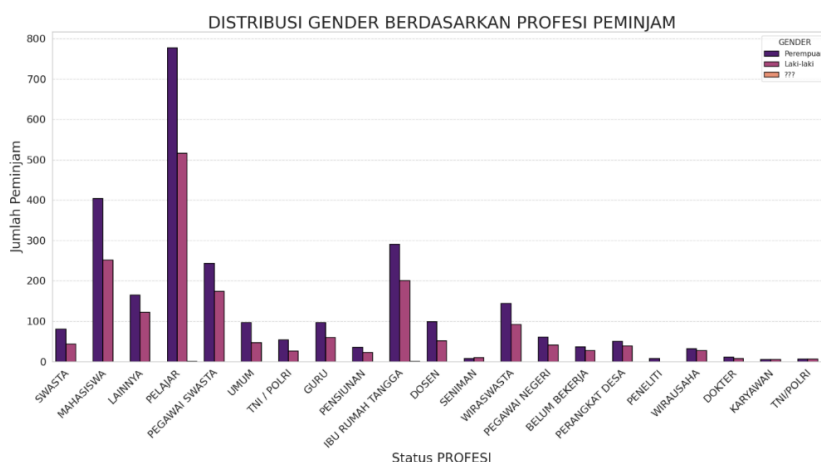
**Gambar 5.** *Distribusi Profesi*

Pertama, Analisis distribusi peminjam berdasarkan profesi, seperti yang ditunjukkan pada Gambar 2, mengungkapkan bahwa kelompok Pelajar merupakan peminjam terbanyak dengan jumlah sekitar 1.294 orang. Diikuti oleh Mahasiswa sebanyak 655 orang, Ibu Rumah Tangga sebanyak 493 orang, dan Pegawai Swasta sebanyak 417 orang. Temuan ini menunjukkan bahwa perpustakaan ini berhasil menarik minat dari berbagai kelompok profesional, terutama yang berasal dari latar belakang akademik dan swasta.



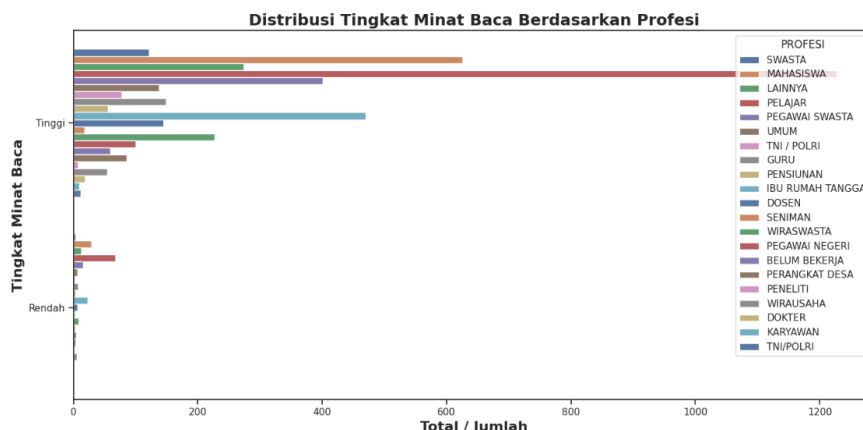
**Gambar 6.** Peminjam Buku Karya Umum Berdasarkan Profesi

Kedua, Grafik pada gambar 3 adalah menggambarkan peminjam buku karya umum berdasarkan profesi peminjam. Warna biru pada grafik menunjukkan "Karya Umum Tidak", artinya profesi tersebut tidak meminjam buku. Sementara warna oranye menunjukkan "Karya Umum Ya", yang berarti profesi tersebut meminjam buku. Profesi dengan tingkat peminjaman buku karya umum tertinggi adalah "Pelajar", ditunjukkan oleh tingginya balok oranye. Sementara profesi seperti "TNI/POLRI", "Seniman", dan "Pensiunan" memiliki tingkat peminjaman yang relatif rendah, ditunjukkan oleh rendahnya balok biru.



**Gambar 7.** Distribusi Gender Berdasarkan Profesi Peminjam

Ketiga, Grafik pada gambar 4 grafik ini menyajikan data mengenai distribusi gender berdasarkan profesi peminjam. Melalui visualisasi ini, dapat dilihat jumlah peminjam dari berbagai jenis profesi, yang diklasifikasikan berdasarkan jenis kelamin (wanita dan pria). Tampak bahwa ada variasi yang cukup signifikan dalam proporsi gender di antara profesi-profesi tersebut. Beberapa profesi didominasi oleh perempuan, seperti Pelajar dan Mahasiswa/i, sedangkan profesi lain seperti TNI/POLRI didominasi oleh laki-laki.

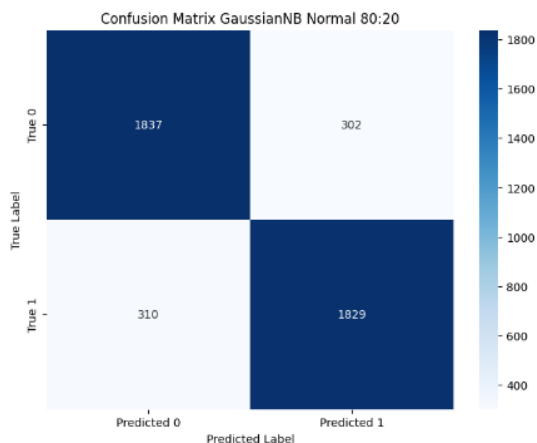


**Gambar 8.** Distribusi Tingkat Minat Baca Berdasarkan Profesi Peminjam

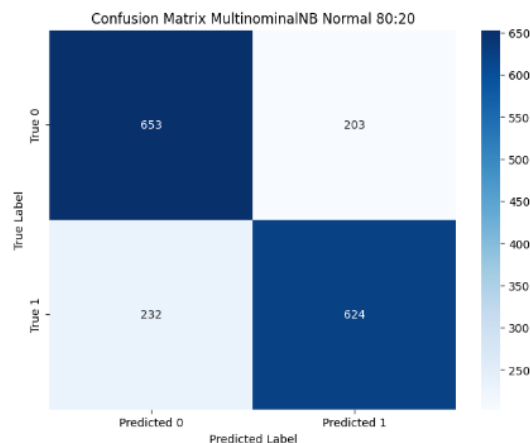
Keempat, grafik pada gambar 5 Grafik ini menyajikan data mengenai distribusi tingkat minat baca berdasarkan profesi. Melalui visualisasi ini, dapat dilihat total jumlah peminjam dari berbagai jenis profesi. Tampak bahwa profesi Pelajar memiliki tingkat minat baca tertinggi, diikuti oleh Mahasiswa, Ibu rumah tangga, Pegawai swasta dan Lainnya. Sementara itu, profesi Dokter, Peneliti, dan lainnya menunjukkan tingkat minat baca yang relatif rendah dibandingkan profesi lainnya.

### Hasil Pengujian 2 Model, Skenario Normal, Rasio 80:20

Setelah tahap metode penelitian selesai, langkah berikutnya adalah memproses data ke tahap pemodelan klasifikasi. Sebelum itu, data akan dipisahkan menjadi dua bagian, yaitu data latih dan data uji, dengan proporsi 80:20 [20]



**Gambar 11.** Confusion Matrix Gaussian Nb



**Gambar 12.** Confusion Matrix Multinomial NB

Pada gambar 11, menunjukkan hasil confusion matrix pada model Gaussian Naive Bayes. Terdapat 1837 data yang Dikategorikan dengan tepat ke kelas rendah, 1829 data yang Dikategorikan dengan tepat ke kelas tinggi., 302 data yang salah dikategorikan ke kelas tinggi, dan 310 data yang salah dikategorikan ke kelas rendah.

Pada gambar 12, menunjukkan hasil confusion matrix pada model Multinomial Naive Bayes. Terdapat 653 data yang dikategorikan dengan tepat ke kelas rendah, 624 data yang dikategorikan dengan tepat ke kelas tinggi, 203 data yang salah dikategorikan ke kelas tinggi, dan 232 data yang salah dikategorikan ke kelas rendah.

**Tabel 2.** *Matrix Evaluasi Model*

Performa	Gaussian NB	Multinomial NB
Accuracy	0.85%	0.75%
Precision Negatif (0)	0.85	0.74
Precision Positif(1)	0.85	0.75
Recall Negariff(0)	0.87	0.76
Recall Positif(1)	0.85	0.73
F-1 Score Negatif (0)	0.86	0.75
F-1 Score Positif (1)	0.85	0.74
Support Negatif (0)	856	856
Support Positif (1)	856	856

Pada matrix evaluasi di Tabel 2, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 85%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 75%.

### Perbandingan Model, Rasio Data dan Skenario Pengujian

Setelah tahap pada metode penelitian selesai, langkah selanjutnya adalah melakukan perbandingan dengan model yang berbeda, serta menguji dengan rasio data yang bervariasi dan skenario (normal, random, korelasi). Penulis menggunakan *tools Google Colab* sebagai alat untuk melakukan pengujian ini. Dalam penelitian ini, penulis membandingkan kinerja dua model Naive Bayes, yaitu *Gaussian Naive Bayes* dan *Multinomial Naive Bayes*. Pengujian dilakukan dengan menggunakan rasio pembagian data 90:10, 80:20, 75:25, dan 50:50. Berikut adalah hasil pengujian yang menunjukkan kinerja masing-masing model pada berbagai rasio dan skenario yang telah ditentukan.

### Pengujian Skenario Normal

Pengujian skenario normal (*normal scenario testing*) dalam konteks model pembelajaran mesin seperti Multinomial Naive Bayes dan Gaussian Naïve Bayes biasanya merujuk pada pengujian model dengan data yang dianggap "normal" atau representatif dari kondisi sebenarnya.

#### a) Pengujian 1 rasio 90:10

**Tabel 3.** *Hasil Pengujian Skenario Normal Rasio 90:10*

Performa	Model Gaussian NB		Model Multinomial NB	
	0	1	0	1
Precision	0.85	0.85	0.73	0.74
Recall	0.85	0.85	0.75	0.72
F1 - score	0.85	0.85	0.74	0.73
Supprot	428	428	856	428
Accuracy	0.85		0.73	

Pada hasil pengujian di Tabel 2, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 85%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 73%.

#### b) Pengujian 2 rasio 75:25

**Tabel 4.** *Hasil Pengujian Skenario Normal Rasio 75:25*

Performa	Model Gaussian NB		Model Multinomial NB	
	0	1	0	1
Precision	0.86	0.86	0.74	0.75
Recall	0.87	0.86	0.75	0.73

F1 - score	0.86	0.86		0.75	0.74	
Supprot	1070	1069	2139	1070	1069	2139
Accuracy			0.86			0.74

Pada hasil pengujian di Tabel 4, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 86%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 74%.

c) Pengujian 3 Rasio 50:50

**Tabel 5.** Hasil Pengujian Skenario Normal Rasio 50:50

Performa	Model Gaussian NB			Model Multinomial NB		
	0	1		0	1	
Precision	0.86	0.86		0.75	0.75	
Recall	0.86	0.86		0.75	0.75	
F1 - score	0.86	0.86		0.75	0.75	
Supprot	2139	2139	4278	2139	2139	4278
Accuracy			0.86			0.75

Pada hasil pengujian di Tabel 5, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 86%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 73%.

### Pengujian Skenario Random Dataset

Pengujian skenario random dataset (*random scenario testing*) dalam konteks model pembelajaran mesin seperti *Multinomial Naive Bayes* dan *Gaussian Naive Bayes* merujuk pada pengujian model dengan menggunakan data yang dipilih secara acak dari keseluruhan dataset.

a) Pengujian 1 Rasio 90:10

**Tabel 6.** Hasil Pengujian Skenario Random Rasio 90:10

Performa	Model Gaussian NB			Model Multinomial NB		
	0	1		0	1	
Precision	0.77	0.97		0.00	0.95	
Recall	0.48	0.99		0.00	1.00	
F1 - score	0.59	0.98		0.00	0.97	
Supprot	21	427	448	24	424	448
Accuracy			0.97			0.95

Pada hasil pengujian di Tabel 6, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 97%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 95%.

b) Pengujian 2 rasio 80:20

**Tabel 7.** Hasil Pengujian Skenario Random Rasio 80:20

Performa	Model Gaussian NB			Model Multinomial NB		
	0	1		0	1	
Precision	0.76	0.97		0.00	0.96	
Recall	0.40	0.99		0.00	1.00	
F1 - score	0.52	0.98		0.00	0.98	
Supprot	40	856	896	40	856	896
Accuracy			0.97			0.76

Pada hasil pengujian di Tabel 7, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 97%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 76%.

## c) Pengujian 3 rasio 75:25

**Tabel 8.** Hasil Pengujian Skenario Random Rasio 75:25

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.89	0.97	0.00	0.96	
Recall	0.44	1.00	0.00	1.00	
F1 - score	0.59	0.98	0.00	0.98	
Supprot	55	1065	1120	48	1072 1120
Accuracy			0.97		0.96

Pada hasil pengujian di Tabel 8, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 97%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 96%.

## d) Pengujian 4 rasio 50:50

**Tabel 9.** Hasil Pengujian Skenario Random Rasio 50:50

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.80	0.98	0.00	0.75	
Recall	0.50	0.99	0.00	1.00	
F1 - score	0.61	0.98	0.00	0.98	
Supprot	105	2134	2239	103	2136 2239
Accuracy			0.97		0.95

Pada hasil pengujian di Tabel 9, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 97%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 95%.

**Pengujian Skenario Korelasi Atribut**

Pengujian skenario korelasi atribut (*attribute correlation testing*) dalam konteks model pembelajaran mesin seperti *Multinomial Naive Bayes* dan *Gaussian Naive Bayes* merujuk pada pengujian model dengan fokus pada hubungan antara atribut dalam dataset. Dengan memahami korelasi antar atribut, peneliti dapat membuat keputusan yang lebih baik mengenai pemilihan fitur, yang pada akhirnya dapat meningkatkan akurasi dan keandalan model.

## a) Pengujian 1 rasio 90:10

**Tabel 10.** Hasil Pengujian Skenario Korelasi Atribut Rasio 90:10

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.69	0.97	0.06	0.95	
Recall	0.48	0.99	0.09	0.92	
F1 - score	0.56	0.98	0.07	0.94	
Supprot	23	425	448	23	425 448
Accuracy			0.96		0.88

Pada hasil pengujian di Tabel 10, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 96%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 88%.

## b) Pengujian 2 rasio 80:20

**Tabel 11.** Hasil Pengujian Skenario Korelasi Atribut Rasio 80:20

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.58	0.98	0.09	0.95	
Recall	0.63	0.98	0.13	0.93	
F1 - score	0.60	0.98	0.11	0.94	
Supprot	46	850	448	46	850 896
Accuracy	0.96		0.89		

Pada hasil pengujian di Tabel 11, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 96%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 89%.

## c) Pengujian 3 rasio 75:25

**Tabel 12.** Skenario Korelasi Atribut Rasio 75:25

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.60	0.98	0.13	0.96	
Recall	0.62	0.98	0.21	0.93	
F1 - score	0.61	0.98	0.16	0.94	
Supprot	58	1062	1120	58	1062 1120
Accuracy	0.96		0.89		

Pada hasil pengujian di Tabel 12, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 96%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 89%.

## d) Pengujian 4 rasio 50:50

**Tabel 13.** Hasil Pengujian Skenario Korelasi Atribut Rasio 50:50

Performa	Model Gaussian NB		Model Multinomial NB		
	0	1	0	1	
Precision	0.74	0.98	0.21	0.96	
Recall	0.63	0.99	0.25	0.96	
F1 - score	0.68	0.99	0.23	0.96	
Supprot	101	2138	2239	101	2128 2239
Accuracy	0.97		0.92		

Pada hasil pengujian di Tabel 13, terlihat bahwa nilai akurasi tertinggi diperoleh dari model Gaussian Naive Bayes, yaitu sekitar 97%, sedangkan model Multinomial Naive Bayes mencapai akurasi sekitar 92%.

Penelitian ini melengkapi kekurangan dalam penelitian terdahulu dengan fokus pada klasifikasi peminjam berdasarkan profesi. Menggunakan data transaksi perpustakaan dari Dinas Perpustakaan Sidoarjo tahun 2023 dan algoritma Gaussian Naive Bayes, penelitian ini berhasil mencapai akurasi tertinggi 97% pada skenario random dataset. Hasil penelitian menunjukkan bahwa profesi pelajar, mahasiswa, dan ibu rumah tangga mendominasi kategori minat baca tinggi, sementara profesi dokter dan peneliti cenderung memiliki tingkat minat baca yang lebih rendah.

Penelitian ini juga mengisi kekurangan studi sebelumnya, seperti yang dilakukan oleh [7], yang menggunakan algoritma Naive Bayes dengan akurasi 80% untuk memprediksi minat baca namun tidak mempertimbangkan faktor profesi. Begitu pula, penelitian oleh [9], menunjukkan efektivitas Naive Bayes dalam klasifikasi buku dengan akurasi mencapai 100%, tetapi tidak mengidentifikasi karakteristik peminjam. Dengan mempertimbangkan profesi sebagai salah satu faktor, penelitian

ini memberikan wawasan lebih mendalam mengenai preferensi peminjaman berdasarkan latar belakang pengguna.

Dengan memanfaatkan variabel profesi secara spesifik, penelitian ini mengisi celah yang ada dalam penelitian terdahulu. Hasilnya memberikan wawasan berharga bagi pengelolaan koleksi perpustakaan yang lebih sesuai dengan kebutuhan peminjam, serta mendukung pengembangan layanan berbasis data dan program literasi yang lebih tepat sasaran untuk setiap profesi.

#### 4. Penutup

Penelitian ini berhasil mengidentifikasi pola peminjaman buku berdasarkan profesi dengan menggunakan algoritma Naive Bayes, hasil penelitian mengungkapkan bahwa kelompok profesi seperti Pelajar, Mahasiswa, dan Ibu Rumah Tangga mendominasi kategori minat baca tinggi, sedangkan profesi seperti Dokter dan Peneliti cenderung memiliki tingkat minat baca yang lebih rendah. *Model Gaussian Naive Bayes* memberikan performa terbaik dengan akurasi mencapai 97% dalam skenario random dataset, menunjukkan efektivitas algoritma dalam memproses dataset yang tidak terstruktur.

Hasil penelitian ini membantu perpustakaan meningkatkan layanan, menyesuaikan koleksi buku sesuai kebutuhan profesi, serta mengembangkan strategi pemasaran dan program literasi yang tepat sasaran. Temuan ini juga mendukung pengambilan keputusan strategis dalam pengelolaan perpustakaan.

Untuk pengembangan penelitian di masa depan, disarankan agar ruang lingkup penelitian diperluas dengan menggunakan dataset yang lebih besar dan bervariasi agar hasil klasifikasi dapat lebih representatif. Penambahan variabel baru, seperti tingkat pendidikan atau jumlah buku yang dipinjam, dapat memperkaya analisis dan menghasilkan klasifikasi yang lebih akurat. Selain itu, penerapan algoritma lain seperti *Random Forest* atau *Decision Tree* dapat memberikan perbandingan performa untuk mendapatkan hasil optimal.

#### 5. Referensi

- [1] W. Nasihuddin, "Peran Perpustakaan Sebagai Media Literasi Digital Masyarakat," *Jurnal Politik dan Sosial Kemasyarakatan* 11, 2019.
- [2] N. S. N. Arimbi, "Pelayanan dan Fasilitas Perpustakaan Umum Daerah Sidoarjo (Perpusda)," *Jatim News*. Accessed: Jul. 03, 2024. [Online]. Available: <https://jatim.news/2024/05/04/pelayanan-dan-fasilitas-perpustakaan-umum-daerah-sidoarjo-perpusda/>
- [3] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *SMATIKA JURNAL*, vol. 13, no. 01, pp. 117–130, Jul. 2023, doi: 10.32664/smatika.v13i01.750.
- [4] E. Martantoh and N. Yanhi, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL Implementation of Naive Bayes Method for Classification of Student's Personality Characteristics at MTS Darussa'adah School Using PHP Mysql," 2022.
- [5] R. A. Anggraini, G. Widagdo, A. S. Budi, and M. Qomaruddin, "Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 7, no. 1, p. 47, Jan. 2019, doi: 10.26418/justin.v7i1.30211.
- [6] T. Y. Yanto and F. Maedjaja, "Implementasi Data Mining Untuk Menganalisis Pola Peminjaman Buku Perpustakaan Dengan Menggunakan Metode Apriori," *Jurnal InFact Sains dan Komputer*, vol. 8, no. 02, 2023.

- [7] L. Utari and Y. Ulfah, "Penerapan Metode Naïve Bayes untuk Prediksi Minat Baca Berdasarkan Usia," vol. 11, pp. 67–74, 2021, doi: 10.36350/jbs.v11i1.
- [8] E. Irfiani, Y. Kusnadi, S. Sunarti, and F. Handayanna, "Implementasi Data Mining dalam Mengklasifikasi Minat Baca Pada Perpustakaan Daerah Menggunakan Algoritma C4.5," *JOINS (Journal of Information System)*, vol. 8, no. 2, pp. 106–114, Nov. 2023, doi: 10.33633/joins.v8i2.8004.
- [9] R. N. Ainum, N. Hidayat, and A. A. Soebroto, "Klasifikasi Buku Perpustakaan menggunakan Metode Naive Bayes," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] R. N. Ainum, N. Hidayat, and A. A. Soebroto, "Klasifikasi Buku Perpustakaan menggunakan Metode Naive Bayes," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] H. Latipa Sari and I. Yati Beti, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Dalam Pengelompokkan Buku Yang Dipinjam Menggunakan Algoritma K-Means," *Media Online*, vol. 3, no. 6, pp. 925–933, 2023, doi: 10.30865/klik.v3i6.826.
- [12] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [13] I. K. A. Wiraguna, E. Setyati, and E. Pramana, "Prediksi Anak Stunting Berdasarkan Kondisi Orang Tua Dengan Metode Support Vector Machine Dengan Study Kasus Di Kabupaten Tabanan-Bali," *SMATIKA JURNAL*, vol. 12, no. 01, pp. 47–54, Jun. 2022, doi: 10.32664/smatika.v12i01.662.
- [14] A. Setiawan, R. Febrio Waleska, M. Adji Purnama, and L. Efrizoni, "Sidomulyo barat, Sidomulyo Bar," 2024. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- [15] M. Riyyan, H. Firdaus, J. H. Ronggo Waluyo, T. Timur, and J. Barat, "PERBANDINGAN ALGORITME NAÏVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," 2022. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- [16] A. Felicia Watratan, A. B. Puspita, D. Moeis, S. Informasi, and S. Profesional Makassar, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," 2020. [Online]. Available: <http://journal.isas.or.id/index.php/JACOST>
- [17] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovenisional dalam Prediksi Komplikasi Hipertensi," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 6, no. 1, pp. 37–49, Apr. 2024, doi: 10.35746/jtim.v6i1.488.
- [18] F. Rizki, M. P. Kharisma Putra, M. A. Assuja, and F. Ariany, "Implementasi Deep Learning Lenet Dengan Augmentasi Data Pada Identifikasi Anggrek," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 3, pp. 357–366, Sep. 2023, doi: 10.33365/jatika.v4i3.3652.
- [19] R. Darmawan *et al.*, "KLASIFIKASI DIAGNOSA PENYAKIT TIROID MENGGUNAKAN METODE RANDOM FOREST," 2024. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- [20] A. P. J. Dwitama, "DETEKSI UJARAN KEBENCIAN PADA TWITTER BAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING: REVIU LITERATUR," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no. 1, Aug. 2021, doi: 10.20885/snati.v1i1.5.