

Implementasi Metode *Naïve Bayes* Dalam Pengelompokan Buku Di Pondok Belajar Terang Bahari

Diah Ayu Kisworo Putri^{1*}, T.H. Dwiati Wismarini²
 Teknik Informatika, Universitas Stikubank, Semarang, Indonesia
 *e-mail *Corresponding Author*: pdiah068@gmail.com

Abstract

The system of grouping books according to the desired genre and order of arrangement needs to be done well in order to ensure easy access to the desired books, ensure order and control over the existence of the book collection owned, and facilitate the operationalization of book management. The Bayesian method is a decision-making method that is more widely used, because it is easy to understand, only requires simple code, and is faster in calculating. In this study, an inventory process or book data collection was carried out using the Naïve Bayes Classifier method which will be implemented at the Terang Bahari reading corner. The dataset consists of 8 attributes, namely shelf number, book code, field, book serial number, copy number, complete number, book title, author. Based on 179 book data grouped into 72 test data, with the Confusion Matrix test parameters an accuracy value of 62.5% was obtained. Thus, the Naïve Bayes Classifier method applied is quite good and the books are organized according to their respective fields.

Keywords: *Grouping books; Naïve Bayes method; accuracy.*

Abstrak

Sistem pengelompokan buku sesuai dengan genre maupun urutan susunan yang dikehendaki perlu dilakukan dengan baik agar dapat menjamin kemudahan akses terhadap buku yang dikehendaki, menjamin keteraturan dan kontrol terhadap keberadaan koleksi buku yang dimiliki, serta memudahkan operasionalisasi pengelolaan buku. Metode *Bayes* merupakan metode mengambil keputusan yang lebih banyak dipakai, karena mudah dipahami, hanya memerlukan kode yang sederhana, serta lebih cepat dalam menghitung. Dalam penelitian ini, dilakukan sebuah proses inventaris atau pendataan buku dengan menggunakan metode *Naïve Bayes Classifier* yang akan diterapkan di Pojok baca Terang Bahari. Dataset terdiri dari 8 atribut yaitu nomor rak, kode buku, bidang, nomor urut buku, nomor kopian, nomor komplit, judul buku, pengarang. Berdasarkan 179 data buku yang dikelompokkan ke dalam 72 data uji, dengan parameter uji Confusion Matrix diperoleh nilai akurasi sebesar 62,5%. Dengan demikian metode *Naïve Bayes Classifier* yang diterapkan sudah cukup baik dan buku tertata sesuai dengan bidangnya masing-masing.

Kata kunci: *Pengelompokan buku; Metode naïve bayes; Akurasi.*

1. Pendahuluan

Salah satu cara yang dapat dilakukan untuk membangun kualitas manusia melalui peningkatan pengetahuan adalah dengan mengembangkan budaya literasi melalui kebiasaan membaca buku-buku berkualitas. Menurut Sitepu [1] buku adalah kumpulan kertas berisi informasi, tercetak, disusun secara sistematis, dijilid serta bagian luarnya diberi pelindung terbuat dari kertas tebal atau karton sebagai sampul.

Kemudahan akses terkait keberadaan buku berkualitas yang relatif mudah saat ini baik secara on-line maupun off-line menciptakan satu kondisi yang kondusif bagi mereka yang sadar bahwa setiap dari kita harus mengupayakan peningkatan kualitas diri secara mandiri yang salah satunya bisa diperoleh melalui membangun kebiasaan baik membaca buku. Keberadadan perpustakaan umum, maupun keberadaan pojok baca yang hadir secara fisik maupun sumber lain yang dapat diakses secara daring adalah bentuk kemudahan yang bisa dicapai untuk memenuhi kebutuhan akan buku-buku yang berkualitas.

Divisi Pojok Baca merupakan salah satu ujung tombak bagi Pondok Belajar untuk mengupayakan peningkatan pengetahuan dan pemahaman warga melalui penyediaan buku-buku bacaan yang dapat menjadi sarana perluasan ilmu dan pengetahuan bagi kelompok sasaran yang

terdiri dari berbagai klasifikasi dan latar belakang. Sejauh ini, penyediaan buku di Pojok Baca Terang Bahari banyak diperoleh dari dukungan para donatur dari berbagai kalangan dan latar belakang yangakhirnya juga berpengaruh pada jenis dan kondisi buku di Pojok Baca. Oleh karena itu pengunjung dari Pondok Belajar juga berasal dari berbagai usia dari mulai anak-anak hingga remaja dari berbagai kalangan. Ada lebih dari 2000 buku yang saat ini ada di Pondok Belajar Terang Bahari dibawah pengelolaan Divisi Pojok Baca namunbelum mendapatkan perlakuan yang semestinya jadi masih dalam kondisi acak dan belum terklasifikasikan dengan baik. Mengacu pada kejelasan dan kemudahan menemukan buku maka klasifikasi utama buku akan didasarkan pada jenis buku diantaranya Novel, ilmu pengetahuan, buku anak, dan lain sebagainya.

Disisi lain, pengaturan buku yang tentu berjumlah sangat banyak dalam suatu perpustakaan, pojok baca maupun fasilitas membaca lain memerlukan satu sistem yang kemudian berdampak pada menjamin keberlangsungan kualitas pelayanan yangbaik. Pentingnya keberadaan sistem pengaturan ini salah satunya perlu dilakukan terhadap sistem pengelompokan buku sesuai dengan genre maupun urutan susunan yang dikehendaki dengan beberapa pertimbangan diantaranya menjamin kemudahanakses terhadap buku yang dikehendaki, menjamin keteraturan dan kontrol terhadap keberadaan koleksi buku yang dimiliki serta memudahkan operasionalisasi pengelolaan buku. Oleh karenanya penerapan sistem pengelolaan terlebih yang berbasis komputerisasi akan sangat berguna dalam membantu pengelolaan lembaga yang menyediakan jasa penyewaan buku dalam jumlah besar yang salah satu diantaranya adalah Pondok Belajar Terang Bahari yang didalam nya terdapat satu divisi yang disebut sebagai Pojok Baca.

Metode *Bayes* adalah salah satu metode komputasi yang dapat digunakan dalam tugas-tugas klasifikasi di mana kita ingin mengelompokkan data ke dalam beberapa kelas yang sudah ditentukan berdasarkan fitur-fitur yang diamati dalam data. Metode ini telah meluas digunakan dalam klasifikasi data pada berbagai bidang bisnis, seperti pada bidang medis [2]-[4], bidang bisnis keuangan [5]-[7], bidang pertanian [8]-[10], dan berbagai bidang bisnis lainnya [11][12].

Pada penelitian ini akan diuji penggunaan metode *Naïve Bayes* dalam mengklasifikasi jenis buku. Studi kasus di lakukan pada Pojok Baca Terang Bahari.

2. Tinjauan Pustaka

Penelitian yang dilakukan oleh Vidya Rizqiyani, dkk menggunakan algoritma *Naïve Bayes* yang membahas sistem klasifikasi dan pencarian buku yang dirancang untuk menentukan kategori dan pencarian buku untuk mengurangi beban kerja pengelola buku. Hasil penelitian menggunakan metode *Naïve Bayes* tersebut dihasilkan 70 dokumen relevandan 5 dokumen tidak relevan dengan nilai recall 88,20%, presisi 94,56%, F-measure90,46%, dan akurasi sebesar 97,78%. Dari sinij dapat disimpulkan bahwa algoritma *Naïve Bayes* dapat diugunakan untuk otomatisasi klasifikasi judul buku perpustakaan [13].

Penelitian yang dilakukan oleh Agus Setiawan, dkk dengan menggunakan metode *naïve bayes* ini bertujuan untuk mengembangkan aplikasi klasifikasi dan pencarian buku yang elbih mudah dan akurat untuk menyelesaikan masalah yang berkaitan dengan klasifikasi dan pencarian buku. Ada lima kategori buku, yaitu: Pemrograman komputer, jaringan komputer, database, multimedia dan sistem operasi dengan mengimplemntasikan metode NBC dan lima kategori tersebut maka dibuatlah aplikasi klasifikasi dan pencarian di Perpustakaan Daerah Kalimantan Timur [14].

Penelitian yang dilakukan oleh Atri Nurani, dkk menggunakan metode *naïve bayes classifier* diperoleh skor akurasi proses klasifikasi buku refrensi sebesar 63%. Dalammelakukan penelitian mengklasifikasikan daftar isi buku referensi, sistem dipengaruhi oleh beberapa factor, seperti model data dan jumlah data latih [15].

Penelitian yang dilakukan oleh Nisriina Nur Hasanah, dkk menggunakan algoritma *K-means clustering*, karena buku-buku yang tersedia berasal dari sumbangan, maka pengawasan terhadap ketersediaan buku-buku tertentu masih belum lengkap. Misalnya, buku yang disumbangkan oleh donator buku berkaitan dengan ilmu Teknik, sedangkan bukuyang diminta adalah buku tentang ilmu tumbuha. Salah satu caranya adalah denganmelakukan data mining menggunakan metode *K-Means clustering* yang membantu pengelola perpustakaan dalam melakukan pengelompokan buku berdasarkan banyaknya peminjaman dan banyaknya ketersediaan eksemplar buku. Data yang digunakan dalam penelitian ini adalah data Riwayat peminjaman buku perpustakaan politeknik LPP Yogyakarta kurang lebih selama 2 tahun. Hasil akhir dari penelitian adalah informasi tentangbuku-buku yang paling laris, cukup laris, dan kurang laris [16].

Penelitian yang dilakukan oleh Annisa Putri Ayudhitama dengan menggunakan metode *Pengenalan Wajah Dengan Face-API.js Berbasis CNN dan GeolokasiBimo Akbar F.*

bayes. Pengujian metode ini menggunakan 50 judul data latih dengan 10 kategori yang adadan 50 data uji. Model pengujian menggunakan ketelitian dan hasil percobaan mencapai 90%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode Niave Bayes dapat digunakan untuk mengambil keputusan yang tepat untuk memecahkan masalah dari banyaknya solusi yang ada. [17].

3. Metodologi

3.1 Metode Penelitian

Metode penelitian menggunakan pendekatan kualitatif dengan pendekatan kualitatif deskriptif yang memberikan data deskriptif dalam bentuk kata-kata lisan dan tulisan tentang orang-orang dan objek yang diamati oleh peneliti. Penggunaan pendekatan penelitian ini disesuaikan dengan tujuan utama penelitian yaitu mendeskripsikan perkembangan koleksi perpustakaan Pojok Baca Terang Bahari.

Subjek penelitian adalah orang yang memberikan informasi tentang data yang dicari. Subjek penelitian merupakan sumber utama penelitian yang memiliki informasi tentang variable yang diteliti. Mengenai pendapat penelitian ini, yang menjadi subjek peneliti adalah ketua Pondok Belajar dan penanggung jawab divisi Pojok Bacayang bekerja sebagai pustakawan. Adapun yang menjadi objek penelitian ialah bagaimana cara mengembangkan manajemen pengelolaan koleksi buku sesuai genre atau jenis/ kriteria buku dan apa saja masalah yang dihadapi dalam proses pengelompokan koleksi buku sesuai genre atau jenis di perpustakaan Pojok Baca Terang Bahari.

Dalam penggalan data pada penelitian ini ada dua jenis, yaitu data Primer dan data Sekunder. Data Primer adalah informasi yang dikumpulkan langsung dari sumber primer, misalnya wawancara. Informasi ini diperoleh, apa yang terjadi Ketika informasi diberikan melalui saran komunikasi pengumpulan informasi utama di perpustakaan Pojok Baca Terang Bahari. Tentang pengembangan koleksi di Pondok Belajar Terang Bahari. Informasi tentang keterbatasan pengembangan koleksi. Data sekunder adalah informasi yang diperoleh secara tidak langsung dalam penelitian. Informasi ini biasanya berupa Data ini biasanya diperoleh dari studi pustaka yang berupa referensi dan dokumen dari hasil pengamatan penelitian yang berfungsi untuk melengkapi data primer. Sekilas tentang ruang belajar perpustakaan Pojok Baca Terang Bahari. Sarana dan prasarana perpustakaan Pojok Baca Terang Bahari. Teknik pengumpulan data penelitian ini terdiri dari observasi, wawancara, dan dokumentasi.

Teknik pengolahan data menggambarkan langkah-langkah atau proses yang dilakukan dalam pengolahan data untuk keperluan penelitian. Metode pengolahan data menggambarkan prosedur pengolahan data sesuai dengan pendekatan yang digunakan. Karena metode kualitatif digunakan dalam penelitian ini, maka metode pengolahan data dilakukan dengan memaparkan data dalam kalimat yang teratur, runtut, logis, tidak tumpang tindih dan mengesankan untuk memudahkan pemahaman dan interpretasi data. Diantaranya melalui tahap review data (editing), klasifikasi (classification), review (verifikasi), dan penarikan kesimpulan (conclusion). Pengolahan data yang diterima melalui teknik berikut: edit (verifikasi informasi), klasifikasikan (klasifikasi), verifikasi (verifikasi), dan simpulkan (kesimpulan). Teknik analisis data penelitian ini adalah reduksi data, penyajian dan penarikan kesimpulan atau verifikasi. Metode penelitian ini adalah pembersihan data. Pada langkah ini, peneliti membersihkan nilai-nilai kosong yaitu integrasi data dimana peneliti menggabungkan sebagian data dari materi yang berbeda menjadi materi baru dan menggunakan fungsi reduksi data untuk menghilangkan informasi yang tidak perlu.

3.2 Algoritme Naïve Bayes

Formula dasar algoritma *Naïve Bayes* didasarkan pada Teorema Bayes. Formula ini umumnya digunakan untuk menghitung probabilitas kelas (target) berdasarkan fitur atau atribut yang diamati dalam data. Dalam kasus klasifikasi, akan dihitung probabilitas bahwa suatu contoh data termasuk dalam suatu kelas tertentu berdasarkan nilai-nilai fitur yang diamati. Formula eorema Bayes dapat ditulis berikut [18]:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \dots\dots\dots (1)$$

Dimana:

$P(A|B)$ = Probabilitas kelas A (target), diberikan fitur atau atribut B

$P(B|A)$ = Probabilitas fitur atau atribut B, diberikan kelas A

$P(A)$ = Probabilitas Prior A (tanpa mempertimbangkan fitur)

$P(B)$ = Faktor normalisasi (Probabilitas dari fitur yang diamati)

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Tahap pertama yang akan dilakukan dalam melakukan penelitian analisis sentiment ialah pengumpulan data yang berasal dari Pondok Belajar Terang Bahari. Dalam pengambilan data peneliti difasilitasi oleh pengurus Pondok Belajar Terang Bahari untuk melakukan pengumpulan sebanyak 179 data dan 10 bidang buku secara gratis. Gambar 1 merupakan sampel data awal yang dikumpulkan, yang disimpan dengan format csv.

#	A	B	C	D	E	F	G	H
1	No. Rak	KODE	BIDANG	Nomor Urut buku	Nomor oplan	Nomor komplit	judul	pengarang
2	1 N	NOVEL		1	1 N001-001-1	THE MASK OF TROY	DAVID GIBINS	
3	1 N	NOVEL		2	1 N002-002-1	THE PERFECT WORLD	CLARISSA GOENAWAN	
4	1 N	NOVEL		3	1 N003-003-1	HANSEL AND GRETEL	LOUISE MURPHY	
5	1 N	NOVEL		4	1 N004-004-1	SENANDUNG BIJAK	ACHMAD SURIAWINATA	
6	1 N	NOVEL		5	1 N005-005-1	PEREMPJUAN KEDUA	SITI ROFIKOH	
7	1 N	NOVEL		6	1 N006-006-1	HIGH SCHOOL MUSICAL	PETER BARSOCCHINI	
8	1 N	NOVEL		7	1 N007-007-1	BUNGA	KORRIE LAYUN RAMPAN	
9	1 N	NOVEL		8	1 N008-008-1	HANSEL AND GRETEL	LOUISE MURPHY	
10	1 N	NOVEL		9	1 N009-009-1	KOALA KUMAL	RADITYA DIKA	
11	1 N	NOVEL		10	1 N010-010-1	FAIRISH	ESTI KINASIH	
12	1 N	NOVEL		11	1 N011-011-1	DRAMATURGY DOFIMA	FARIS RAHMAN	
13	1 N	NOVEL		12	1 N012-012-1	THE PHANTOM OF THE OPERA	GASTON LEROUX	
14	1 N	NOVEL		13	1 N013-013-1	THE BEAUTIFUL GEEK	TRIA BARMAWI	
15	1 N	NOVEL		14	1 N014-014-1	SEMUA INDAH PADA WAKTUNYA	EKO HARTONO	
16	1 N	NOVEL		15	1 N015-015-1	REFRAIN	WINA EFENDI	
17	1 N	NOVEL		16	1 N016-016-1	REMEMBER WHEN	WINA EFENDI	
18	1 N	NOVEL		17	1 N017-017-1	KACU	EN HA	
19	1 N	NOVEL		18	1 N018-018-1	CHLOE QUEEN OF DENIAL	NAOMI NASH	
20	1 N	NOVEL		19	1 N019-019-1	NEIL		
21	1 N	NOVEL		20	1 N020-020-1	GAIMAN M IS FOR MAGIC	TEDDY KRISTIANSEN	
22	2 G	KOMIK		1	1 G001-001-1	LUPUS DRAKULI KUPER	HILMAN	
23	2 G	KOMIK		2	1 G002-002-1	LUPUS DIH, SEREM	HILMAN	
24	2 G	KOMIK		3	1 G003-003-1	DIARY JEANY	JEANY SHANTI DEVI	
25	3 E	ENSIKLOPEDIA		1	1 E001-001-1	ANAK HEBAT TUMBUHAN	PT BHUANA ILMU POPULER	

Gambar 1. Tampilan Data Yang Telah dikumpulkan

4.2 Preprocessing Data

Pada tahapan *preprocessing* dapat melalui berbagai proses karena ada beberapa judul buku yang tidak menggunakan Bahasa baku dan juga terdapat angka, tanda baca, dan lain sebagainya.

1. Case Folding

Case Folding merupakan tahapan dalam preprocessing yang digunakan untuk menyamaratakan semua kata dalam bentuk lowercase atau huruf kecil untuk mempermudah dalam proses selanjutnya. Dalam tahapan ini menggunakan method `lower()` dan `strip()` untuk menghapus whitespace (karakter kosong). Pada gambar 4.3 berikut merupakan kode program tahap case folding dan tabel 4.1 merupakan hasil dari preprocessing tahap case folding.

```

Case Folding

def casefolding(text):
    text = text.lower()
    text = text.strip(' ')
    text = re.sub(r'[?|$.|!|_|:|"](-+,]', '', text)
    return text

tweet_df['case_folding'] = tweet_df['judul'].apply(casefolding)
tweet_df.head(20)
    
```

Gambar 2. Kode Program Case Folding

Hasil dari kode program pada Gambar 2 seperti disajikan pada Tabel 1.

Tabel 1 Hasil Tahapan *Case Folding*

index	BIDANG	judul	case_folding
0	NOVEL	THE MASK OF TROY	the mask of troy
1	NOVEL	THE PERFECT WORLD	the perfect world
2	NOVEL	HANSEL AND GRETEL	hansel and gretel
3	NOVEL	SENANDUNG BIJAK	senandung bijak
4	NOVEL	PEREMPUAN KEDUA	perempuan kedua
5	NOVEL	HIGH SCHOOL MUSICAL	high school musical
6	NOVEL	BUNGA	bunga
7	NOVEL	HANSEL AND GRETEL	hansel and gretel
8	NOVEL	KADAL KUMAL	kadal kumal
9	NOVEL	FAIRISH	fairish
10	NOVEL	DRAMATURGY COFINA	dramaturgy cofina
11	NOVEL	THE PHANTOM OF THE OPERA	the phantom of the opera
12	NOVEL	THE BEAUTIFUL GEEK	the beautiful geek
13	NOVEL	SEMUA INDAH PADA WAKTUNYA	semua indah pada waktunya
14	NOVEL	REFRAIN	refrain
15	NOVEL	REMEMBER WHEN	remember when
16	NOVEL	KACU	kacu
17	NOVEL	CHLOE QUEEN OF DENIAL	chloe queen of denial
18	NOVEL	NEIL	neil
19	NOVEL	GAMAM M IS FOR MAGIC	gamam m is for magic

Show | 25 | per page

2. Tokenization

Tokenization merupakan tahapan dalam preprocessing yang digunakan untuk memisahkan kata atau yang biasa disebut token. Tahap tokenization dilakukan dengan bantuan library dari python yaitu nltk.tokenize dengan module yaitu RegexpTokenizer. Gambar 3 merupakan kode program tahap tokenization dan tabel 2 merupakan hasil dari preprocessing tahap tokenization.

```

Tokenization
0 d
from nltk.tokenize import RegexpTokenizer
tokenization = RegexpTokenizer(r'\w+|${0-9}+|\S+')

tweet_df['tokenization'] = tweet_df['case_folding'].apply(tokenization.tokenize)
tweet_df.head()

```

Gambar 3. Kode Program Tokenization

Tabel 2. Hasil Tahapan Tokenization

BIDANG	judul	case_folding	tokenization
0	NOVEL	THE MASK OF TROY	the mask of troy [the, mask, of, troy]
1	NOVEL	THE PERFECT WORLD	the perfect world [the, perfect, world]
2	NOVEL	HANSEL AND GRETEL	hansel and gretel [hansel, and, gretel]
3	NOVEL	SENANDUNG BIJAK	senandung bijak [senandung, bijak]
4	NOVEL	PEREMPUAN KEDUA	perempuan kedua [perempuan, kedua]

3. Stemming

Stemming yaitu tahapan dengan tujuan untuk mengubah bentuk kata menjadi ke bentuk dasar atau menghilangkan kata imbuhan yang terdapat didepan atau di belakang dan diganti menjadi kata dasar. Dalam melakukan tahapan stemming dibutuhkan sebuah library dari python yaitu sastrawi. Hasil dari proses tahapan stemming akan menghapus kata-kata yang memiliki imbuhan yang terdapat diawal atau diakhir. Tabel 3 merupakan dari hasil tahapan stemming.

Tabel 3. Contoh Perubahan Kata Tahap Stemming

Kata Asal	Perubahan Kata
waktunya	Waktu
tumbuhan	Tumbuh
pemecah	Pecah
kesenian	Seni

Kata Asal	Perubahan Kata
pengalaman	Alam
seruku	Seru
Kehidupan	Hidup
kecerdikan	cerdik
penyelamat	Selamat
sebelum	Belum
berbeda	Beda
belajar	ajar
mengantre	Antre
Mendunia	Dunia
kesepian	sepi
nyanyian	Nyanyi
kuulang	Ulang
menunggu	tunggu

Untuk menginstall library digunakan perintah pip install sastrawi. Gambar 4 merupakan kode program tahap stemming dan tabel 4 merupakan hasil dari preprocessing tahap stemming judul buku.

```

Stemming

from sklearn.pipeline import Pipeline
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(text):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in text:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = ' '.join(do)
    print(d_clean)
    return d_clean

tweet_df['clean_data'] = tweet_df['stemming'].apply(stemming)
tweet_df.head()
    
```

Gambar 4. Kode Program Stemming

Tabel 4. Hasil Tahapan Stemming

BIDANG	Judul	case_folding	tokenization	stemming	clean_data
0 NOVEL	THE MASK OF TROY	the mask of troy	[the, mask, of, troy]	[the, mask, of, troy]	the mask of troy
1 NOVEL	THE PERFECT WORLD	the perfect world	[the, perfect, world]	[the, perfect, world]	the perfect world
2 NOVEL	HANSEL AND GRETEL	hansel and gretel	[hansel, and, gretel]	[hansel, and, gretel]	hansel and gretel
3 NOVEL	SENANDUNG BIAK	senandung bijak	[senandung, bijak]	[senandung, bijak]	senandung bijak
4 NOVEL	PEREMPUAN KEDUA	perempuan kedua	[perempuan, kedua]	[perempuan, dua]	perempuan dua

4.3 Klasifikasi Naïve Bayes Classifier

Pada tahap klasifikasian menggunakan metode *Naïve Bayes Classifier* untuk memperoleh bidang yang berkelas novel, komik, ensiklopedia, cerita fiksi, kamus, kerajinan, majalah, komputer, bidang keilmuan lainnya, Pendidikan. *Naïve Bayes Classifier* digunakan untuk melakukan suatu prediksi terhadap suatu kasus berdasarkan hasil dari klasifikasi yang diperoleh. Untuk menentukan nilai kelas setiap bidang buku dapat dihitung dengan formula 1.

$$\text{Probabilitas prior novel} = \frac{\text{Jumlah kelas novel}}{\text{Jumlah keseluruhan data}} = \frac{20}{179} = 0,111$$

Probabilitas prior komik	$= \frac{\text{Jumlah kelas komik}}{\text{Jumlah keseluruhan data}} = \frac{3}{179} = 0,016$
Probabilitas prior ensiklopedia	$= \frac{\text{Jumlah kelas cerita ensiklopedia}}{\text{Jumlah keseluruhan data}} = \frac{13}{179} = 0,072$
Probabilitas prior cerita fiksi	$= \frac{\text{Jumlah kelas cerita fiksi}}{\text{Jumlah keseluruhan data}} = \frac{91}{179} = 0,508$
Probabilitas prior kamus	$= \frac{\text{Jumlah kelas kamus}}{\text{Jumlah keseluruhan data}} = \frac{9}{179} = 0,050$
Probabilitas prior kerajinan	$= \frac{\text{Jumlah kelas kerajinan}}{\text{Jumlah keseluruhan data}} = \frac{10}{179} = 0,055$
Probabilitas prior majalah	$= \frac{\text{Jumlah kelas majalah}}{\text{Jumlah keseluruhan data}} = \frac{10}{179} = 0,055$
Probabilitas prior komputer	$= \frac{\text{Jumlah kelas komputer}}{\text{Jumlah keseluruhan data}} = \frac{5}{179} = 0,027$
Probabilitas prior keilmuan lainnya	$= \frac{\text{Jumlah kelas keilmuan lainnya}}{\text{Jumlah keseluruhan data}} = \frac{9}{179} = 0,050$
Probabilitas prior Pendidikan	$= \frac{\text{Jumlah kelas pendidikan}}{\text{Jumlah keseluruhan data}} = \frac{9}{179} = 0,050$

Selanjutnya dilakukan pembagian data latih dan data uji dengan menggunakan library dari python yaitu *scikit-learn*. Pembagian data yang diperoleh dengan perbandingan 6:4 yang artinya 60% untuk data latih dan 40% untuk data uji. Dalam melakukan pembagian data digunakan model *sklearn.model_selection* dan library yaitu *train_test_split*. Gambar 5 merupakan kode program untuk pembagian data.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(cos_sin, tweet_df['BIDANG'], test_size=0.4, random_state=42)
```

Gambar 5. Kode Program Pembagian Data

Dari proses ini pembagian data didapatkan hasil untuk jumlah data uji dan data latih, seperti pada Gambar 6.

```
print("Jumlah Data Uji:", X_test.shape)
print("Jumlah Data Latih:", X_train.shape)
print("Jumlah Data Uji:", y_test.shape)
print("Jumlah Data Latih:", y_train.shape)

Jumlah Data Uji: (72, 179)
Jumlah Data Latih: (107, 179)
Jumlah Data Uji: (72,)
Jumlah Data Latih: (107,)
```

Gambar 6. Hasil Pembagian Data

Berikut perhitungan untuk hasil pembagian data latih dan data uji.

$$\text{Data Uji} = \frac{40}{100} \times 179 = 71,6 = 72$$

$$\text{Data Latih} = \frac{60}{100} \times 179 = 107,4 = 107$$

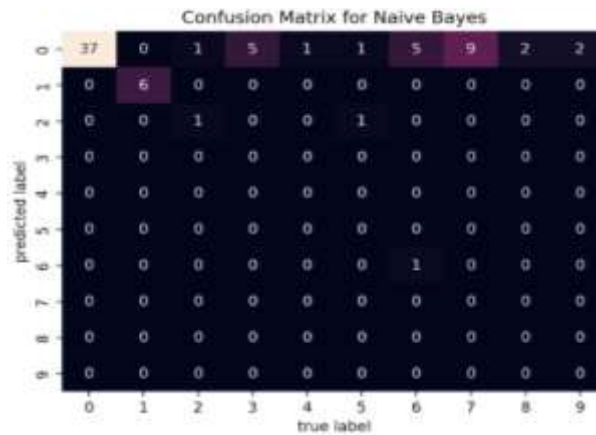
Klasifikasi *Naïve Bayes Classifier* menggunakan library yang berasal dari *scikit-learn* dengan import library *from sklearn.naive_bayes import MultinomialNB*. Berikut gambar 7 merupakan kode program proses klasifikasi dengan metode *Naïve Bayes Classifier*.

```
clf = MultinomialNB().fit(X_train, y_train)
predicted = clf.predict(X_test)
```

Gambar 7. Kode Program Klasifikasi

4.4 Pengujian Model

Setelah proses klasifikasi, tahap selanjutnya melakukan pengujian model yang bertujuan untuk mengetahui performa yang dihasilkan dari metode tersebut. Hasil dari klasifikasi tersebut akan divisualisasikan dalam bentuk confusion matrix. Confusion matrix adalah tabel yang digunakan untuk mengukur model klasifikasi menentukan perhitungan dari prediksi benar atau salah. Dalam pengujian model klasifikasi data yang diambil berasal dari data uji yang berjumlah 72 data kelas dari 179 data kelas. Modul yang digunakan dalam proses ini ialah *sklearn.metrics* dengan library *confusion_matrix*. Gambar 8 merupakan tampilan dari confusion matrix.



Gambar 8. Hasil Confusion Matrix

Setelah mengetahui hasil dari confusion matrix maka tahapan selanjutnya perhitungan untuk menentukan berapa nilai akurasi yang diperoleh dengan menggunakan metode *Naive Bayes Classifier*. Untuk menentukan nilai akurasi dibutuhkan modul *sklearn.metrics* dengan library *accuracy score*. Dengan menggunakan library tersebut mendapatkan akurasi sebesar 0,625 atau 62,5%. Gambar 9 merupakan kode program untuk menghitung nilai akurasi.

```
clf = MultinomialNB().fit(X_train, y_train)
predicted = clf.predict(X_test)

print("Hasil Akurasi Dengan Naive Bayes :")
print("Accuracy:" , accuracy_score(y_test,predicted))
```

Hasil Akurasi Dengan Naive Bayes :
Accuracy: 0.625

Gambar 9. Hasil Nilai Accuracy

4.5 Evaluasi Model

Berdasarkan hasil dari confusion matrix di atas maka didapatkan nilai Precision, Recall, dan F1-Score untuk kelas bdang buku. Hasil dari Evaluasi model perhitungan tersebut dapat dilihat pada Tabel 5.

Tabel 5. Nilai Evaluasi Model

Klasifikasi	Precision	Recall	F1_Score
Novel	0	0	0
Komik	0	0	0
Ensiklopedia	100%	100%	100%
Cerita fiksi	59%	100%	100%
Kamus	50%	50%	50%
Kerajinan	0	0	0
Majalah	100%	17%	29%
Komputer	0	0	0
Keilmuan lainnya	0	0	0
Pendidikan	0	0	0

Dari hasil performa klasifikasi diatas sehingga dihasilkan nilai rata-rata untuk setiap performa klasifikasi. Maka dari itu diperoleh rata-rata nilai precision, recall, dan F1-score beserta hasil akurasi pada tabel 6.

Tabel 6. Hasil Rata-Rata Evaluasi Model

Accuracy	Precision	Recall	F1-Score
62,5 %	30,8 %	26,6 %	25,2 %

5. Simpulan

Dalam penelitian ini, dilakukan proses inventaris atau mendata buku menggunakan Teknik Naïve Bayes Classifier di Pondok Belajar Terang Bahari. Dataset terdiri dari 8 atribut yaitu no rak, kode buku, bidang, nomor urut buku, nomor kopian, nomor komplit, judul buku, pengarang. Berdasarkan 179 data buku yang dikelompokkan ke dalam 72 data uji, mendapatkan nilai akurasi sebesar 62,5 % dengan parameter uji Confusion Matrix. Dengan demikian metode Naïve Bayes Classifier yang diterapkan sudah cukup baik dan buku tertata sesuai dengan bidangnya masing-masing.

Daftar Referensi

- [1] B. P. Sitepu, "Pengembangan Taman Bacaan Masyarakat Sebagai Sumber Belajar," *Jurnal Ilmiah Visi P2TK PAUD NI*, Vol. 7, No. 1, Pp.42-56, Juni 2012.
- [2] A. Byna, & M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes". *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, vol. 9, no. 3, pp. 407-411, 2020.
- [3] H. Purnomo, A.B. Kusdina, & W. Apriandari, "Model Aplikasi Diagnosis Penyakit Kulit pada Anjing Menggunakan Metode Naïve Bayes". *Progresif: Jurnal Ilmiah Komputer*, vol. 16, no. 2, pp, 155-164, 2020..
- [4] D.F. Risa, F. Pradana, F.A. Bachtiar, "Implementasi Metode Naive Bayes untuk Mendeteksi Stres Siswa Berdasarkan Tweet pada Sistem Monitoring Stres". *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 6, pp. 1301-1306, 2021
- [5] D. Fitrihanah, S. Dwiasnati, & K.A. Baihaqi, "Penerapan Metode Machine Learning untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi Naïve Bayes". *Faktor Exacta*, vol 14, no. 2, pp. 92-99, 2021.
- [6] M. Rianto, R. Rusdiah, & H. Ichwan, "Penerapan Data Mining Dengan Metode Naïve Bayes Dan Learning Vector Quantization Credit Rating Dalam Memprediksi Kelayakan Pemberian Kredit Oleh PT. BPR Lebak Sejahtera". *Respati*, vol. 17, no. 1, pp. 69-76, 2022.
- [7] T. Ramadhan, D. Wahiddin, & E. Awal, "Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes". *Scientific Student Journal for Information, Technology and Science*, vol. 4, no. 1, pp. 82-87, 2021.
- [8] M. Zunaidi, U. F. S. S. Pane, & A.H. Nasyuha, "Analisis Teorema Bayes Dalam Mendiagnosa Penyakit Tanaman Pisang". *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, pp. 1302-1308, 2021.
- [9] I. B. K. D. S. Negara, I.P.K. Negara, & N.Y. Arso, "PREDIKSI HASIL PANEN PADI DI KABUPATEN JEMBRANA MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER". *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 3, pp. 260-265, 2023
- [10] F.P. Dara, & H. Haerudin, "Penerapan Metode Naive Bayes dalam Pengambilan Keputusan Pemilihan Bibit Cokelat Terbaik Pada Petani Kutacane". *OKTAL: Jurnal Ilmu Komputer dan Sains*, vol. 1, no. 9, pp. 1460-1466, 2022.
- [11] A. A. A. Arifin, W. Handoko, & Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan". *J-Com (Journal of Computer)*, vol. 2, no. 1, pp. 21-26, 2022.
- [12] T. Taufiq, & Y. Yudihartanti, "Penerapan Theorema Bayes Pada Penilaian Kelayakan Angkutan Kota". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 111-122, 2021.
- [13] V. Rizqiyani, A. Mulwinda, And R. D. M. Putri, "Klasifikasi Judul Buku Dengan Algoritma Naïve Bayes Dan Pencarian Buku Pada Perpustakaan Jurusan Teknik Elektro," *Jurnal Teknik Elektro*, Vol. 9, No. 2, Pp.60-65, Juli - Desember 2017.
- [14] A. Setiawan, I. F. Astuti, And A. H. Kridalaksana, "Klasifikasi Dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier (Nbc) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur)," *Jurnal Informatika Mulawarman*, Vol. 10, No. 1, Pp.1-10,

- Februari2015.
- [15] A. Nurani, B. Susanto, And U. Proboyekti, "Implementasi Naïve Bayes Classifier Pada Porgram Bantu Penentuan Buku Referensi Matakuliah", *Jurnal Informatika*, Vol. 3, No. 2, November 2007.
- [16] N. N. Hasanah & A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik Lpp Yogyakarta)," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, Vol. 4, No. 2, Pp. 300–311, Juli 2022, Doi: 10.47233/Jteksis.V4i2.499.
- [17] A. P. Ayudhitama, "Sistem Pengelompokan Kategori Buku Berbasis Metode Bayes Pada Sistem Informasi Perpustakaan Jurusan Teknik Elektro UM," *Jurnal Informatika Polinema*, Vol. 5, No. 3, Pp.108-112, Mei 2019.
- [18] E. Prasetyo, *Data Mining*, Yogyakarta: Penerbit ANDI, 2012