# Application of the Naïve Bayes Algorithm in Classifying the Reading Interests of Regional Library Visitors

*Murlena[1*], Wandi Syahindra[2]*

*[1]Universitas Pat Petulai, Bengkulu, Indonesia*
*[2]Institut Agama Islam Negeri Curup, Bengkulu, Indonesia*

## Article Information

## A B S T R A C T

Reading interest is a key indicator in assessing the success of library services. However, manually understanding visitors' preferences poses a challenge for library managers. This study aims to classify the reading interests of regional library visitors by employing the Naïve Bayes algorithm, a widely-used classification method in data mining. The research data includes visit records and book borrowing data from a regional library. Through a quantitative approach, this study analyzes reading interest patterns and evaluates the performance of the Naïve Bayes algorithm in classifying these interests. The analysis results show that the algorithm achieves an accuracy of 65%, with a precision of 62%, recall of 63%, and F1-score of 63%. These findings are expected to assist libraries in formulating better-targeted collection management and service policies, contributing to the overall improvement of reading interest in the community. This study contributes to the field by providing a practical, data-driven solution for libraries to enhance service quality through a better understanding of visitor preferences. Furthermore, it demonstrates the applicability of the Naïve Bayes algorithm in a non-commercial context, encouraging future research on data-driven approaches in library management to support literacy and educational development.

## 1.   Introduction

The role of libraries in supporting learning and fostering a reading culture in society is undeniable. As information hubs and knowledge sources, libraries not only serve as places to access books but also as agents of change in shaping reading habits within the community [1]. However, with the rapid development of digital technology and the increasing use of social media and electronic devices, the reading interest among the younger generation has been declining [2]. This condition has led to a decrease in library visits and low book borrowing rates, which indirectly impact literacy and students' academic performance [3].

This challenge indicates that libraries need to adopt more innovative approaches to understand visitors' reading interest patterns and develop programs that can attract public interest, particularly among students, to return to libraries. This is where the utilization of modern technology becomes crucial in managing and analyzing visitor data. The Naïve Bayes algorithm, known for its simplicity and effectiveness in data classification, presents a solution that can help libraries map visitors' reading preferences based on book borrowing data [4]. The use of this technology can lead to more efficient, data-driven strategies to increase the public's reading interest [5].

This research focuses on applying the Naïve Bayes algorithm to classify the reading interests of regional library visitors based on available data, such as the types of books borrowed, visitors' backgrounds, and their preferences. Through this data analysis, libraries can better understand visitor preferences and design more

relevant programs and book collections. This technology not only benefits library managers in handling visitor data but also enhances the visitor experience by providing collections that align with their reading interests [6].

The purpose of this study is to apply the Naïve Bayes algorithm to classify the reading interests of regional library visitors based on book borrowing data. Through this analysis, libraries are expected to more easily identify the most popular types of books among visitors, as well as the demographic characteristics of visitors with high reading interest. This research also aims to contribute to the development of data-driven library management strategies that can improve visitor engagement and strengthen the role of libraries in promoting literacy. The practical benefits of this study include optimizing library collection management, enhancing visitor services, and planning more effective programs to encourage reading interest.

This study focuses on visitors to regional library X who borrowed books between April and August 2023. The data used includes the types of books borrowed, visitors' demographic backgrounds (such as gender and occupation), and their preferences for specific types of books. The Naïve Bayes algorithm is used to classify visitors based on this data, aiming to provide a clearer understanding of reading interests. The study also includes an analysis of the relationship between demographic factors and book preferences, and how these factors influence reading interest.

This research offers novelty by combining the Naïve Bayes algorithm approach to analyze library visitors' reading interests, a method that is still rarely used in this context. Most studies utilizing this algorithm typically focus on commercial sectors, such as e-commerce or health forecasting [7], whereas its application in libraries and education is still minimal. Thus, this research provides a significant contribution to technology-based library management and opens up new opportunities for further research on the use of technology in enhancing library services.

A literature review reveals studies that strengthen this research, including An Implementation of Naive Bayes Classifier. This study elaborates on the application of Naïve Bayes in probabilistic classification scenarios across various cases, emphasizing that the simplicity of Naïve Bayes enables efficient use in large datasets and rapid classification tasks, making it an ideal choice for data management within library or educational environments [8]. Additionally, in the study titled Comparison of SVM and Naïve Bayes for Sentiment Classification Using BERT Data, Naïve Bayes is compared to other models such as Decision Trees, highlighting Naïve Bayes's effectiveness in fast classification contexts, especially when paired with tailored text-processing techniques. This relevance is evident in library applications where Naïve Bayes can efficiently classify user preferences based on their reading interests [9].

In another study, Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis, Naïve Bayes is shown to be suitable for classifying user preferences on digital library platforms. Through optimization methods such as feature selection and performance evaluation at the categorical level, Naïve Bayes can be applied effectively in reading interest classification with high accuracy [10].

Reading interest refers to an individual's ability to engage with themselves in understanding the meaning of reading material, offering insightful awareness of the importance of reading. A high interest in reading is crucial for individuals to recognize the value of this activity. Society's enthusiasm for reading is reflected in their reading habits, which provide significant benefits by helping individuals develop a deeper understanding of the importance of reading [11].

Reading interest can also be defined as a strong and deep attention, accompanied by a sense of enjoyment, toward the activity of reading, which motivates an individual to do it voluntarily. In the context of students, there are generally two factors that influence their performance: internal and external factors. Internal factors include the lack of motivation or intention to read, while external factors involve the social and environmental influences surrounding the student [12]. The influence of technological development on

students' habits is evident in the increased time they spend using trending technologies such as smartphones, the internet, and portable games, which can impact their study habits. If students become caught in a dilemma between technological advancement and emotional control, their academic performance may decline. Additionally, the reality on the ground shows that the availability of books in regional libraries is increasingly limited, further reducing society's reading interest. Even when available, the books are often outdated collections [13].

In the process of problem-solving and knowledge discovery, there are several general classifications such as Estimation, Association, Classification, Clustering, and Prediction. Several classification techniques used for problem-solving include: the C4.5 Algorithm, K-Nearest Neighbor Algorithm, ID3, Naïve Bayes Classification, and CART (Classification And Regression Tree) [14].

Data mining is a method used to process large-scale data. As such, data mining plays a crucial role in various fields, including industry, finance, weather forecasting, as well as science and technology. Data mining is defined as a series of activities aimed at discovering interesting patterns from large amounts of data that can be stored in databases, data warehouses, or information storage systems [15]. It is also described as a process of analyzing large data sets to discover hidden patterns, trends, and valuable information. Through statistical techniques and algorithms, data mining supports better decision-making across fields such as business, healthcare, and science [16]. Data mining, also referred to as KDD (Knowledge Discovery in Databases), involves a series of structured stages, with data mining being one of the key steps in the KDD process [17].

The Naïve Bayes algorithm is a classification method based on the application of Bayes' theorem, assuming that each feature is independent of the others. This approach is used to predict the category or class of data based on probabilities, calculating the likelihood of each class based on the features present. The strength of this algorithm lies in its simplicity and speed in handling large datasets, although the assumption of feature independence is often not met in practice [18].

Bayes' theorem provides a way to calculate the probability of a hypothesis based on new information. It is expressed as follows:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Where:

$P(H|E)$: Posterior probability of hypothesis $H$ given evidence $E$.

$P(E|H)$: Likelihood of evidence $E$ given hypothesis $H$.

$P(H)$: Prior probability of hypothesis $H$.

$P(E)$: Total probability of evidence $E$.

The steps for classification using the Naïve Bayes algorithm include collecting data by preparing a dataset with features (attributes) and target classes to predict. The prior probability for each class in the dataset is calculated using the following formula:

$$P(C_K) = \frac{\text{Number of data in class } C_k}{\text{Total number of data}}$$

In addition, the theory of data mining serves as an important foundation in this research. Data mining is the process of extracting valuable patterns from large data sets, and in the context of libraries, it is used to identify reading interest patterns based on book borrowing data. This process involves several stagesuch as data cleaning, integration, selection, and transformation before applying the Naïve Bayes algorithm to classify the results. This theory is relevant because it enables libraries to better utilize their existing data, rather than just storing it without further analysis [19].

Thus, this research not only builds on previous studies but also introduces a new approach by applying the Naïve Bayes algorithm to address the challenges faced by libraries in increasing visitors' reading interest.

## 2. Method

This study adopts a quantitative approach, utilizing techniques to classify the reading interests of regional library visitors. The Naïve Bayes algorithm was chosen as the classification method due to its efficiency in analyzing large datasets. A quantitative approach was employed to measure the relationships between variables related to reading interest, such as the types of books borrowed, gender, and the occupation of library visitors.

Data used in this study are private data obtained from book borrowing records at regional library X during the period from April to August 2023. The data consists of 311 visitors, including variables such as the types of books borrowed, gender, occupation, and reading interest. Each piece of data was categorized and processed for further analysis using the Naïve Bayes algorithm.

The data used in this study consists of library visitor data, which includes demographic information such as gender, age, and occupation. This data is considered representative in determining general reading interest patterns, as demographic characteristics can influence preferences for certain categories of books. By analyzing visitor data without requiring specific borrowing records, this study can identify reading interests through the preference patterns of different demographic groups.

Using the Naïve Bayes algorithm, this study classifies the types of books most favored by visitors based on this demographic information. The algorithm works by calculating the probability of each class (book category) relative to the existing visitor data, resulting in a classification that reveals the book categories most relevant to or preferred by various demographic groups among library visitors.

The data collection procedure involved recording book borrowing histories at the regional library. The data collection process included the following steps:

1. Book Borrowing Record Collection: Data was retrieved from the library information system, which tracks the types of books borrowed, borrowing dates, gender, and occupations of visitors.
2. Data Validation: Each data entry was verified to ensure no duplication or missing information. Incomplete or irrelevant data was removed from the dataset.
3. Data Preprocessing: The collected data was cleaned, normalized, and transformed into a suitable format for further analysis.

The research flow diagram illustrates the stages from data collection to result evaluation. This study used private data from 311 library visitors who borrowed books at regional library X. The data was analyzed using the Naïve Bayes algorithm to predict the public's reading interest in the library. The steps involved in analyzing the data to address this issue are shown in Figure 1.
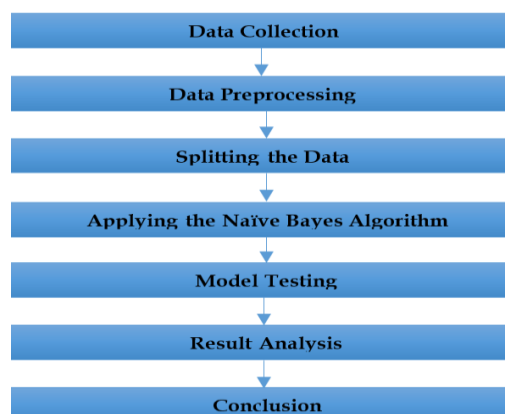


**Figure 1.** Research Flow Diagram

The model evaluation was conducted using several metrics to assess its performance:

1. Accuracy: Measures the percentage of correct predictions from the total test data.
2. Precision: Measures the model's accuracy in predicting positive reading interest.
3. Recall: Measures how many relevant data points were correctly predicted by the model.
4. F1-Score: Combines precision and recall to provide an overall performance summary of the model.

Testing Method: After applying and testing the model, the results from the Naïve Bayes algorithm were compared with results from similar studies using other methods, such as predictions in healthcare and e-commerce. These results were then interpreted to explore how this approach could be applied more broadly beyond the library context.

## 3. Results and Discussion

The data used to predict reading interest comes from the visitors of regional library X, who borrowed books from April to August 2023. The data from these visitors is displayed in Table 1.

**Table 1.** Regional Library X Visitor Data

| No | Gender | Book Title | Book Type | Borrowing Date | Work | Address |
|---|---|---|---|---|---|---|
| 1 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 2 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 3 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 4 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 5 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 6 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 7 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 8 | P | Bahasa Indonesia | Sastra | 11/04/2023 | Pelajar | SMAN 3 RL |
| 9 | P | Psikologi Konsling | Ensiklopedia | 12/04/2023 | Mahasiswa | IAIN CURUP |
| 10 | P | Psikologi Konsling | Ensiklopedia | 12/04/2023 | Mahasiswa | IAIN CURUP |
| 11 | l | The Magic of Talking | Novel | 12/04/2023 | Umum | Rejang Lebong |
| 12 | L | Bumi Manusia | Novel | 12/04/2023 | Umum | Dusun Teladan |
| 13 | L | Skripsi Kelar Dalam 30 Hari | Ensiklopedia | 12/04/2023 | Mahasiswa | Adirejo |
| 14 | P | Cepat Mahir Menjadi Youtuber | Ensiklopedia | 12/04/2023 | Pelajar | Rejang Lebong |
| 15 | L | Langkah Mudah Membuat Iklan Fb | Ensiklopedia | 12/04/2023 | Pelajar | Rejang Lebong |
| 16 | P | Sejarah Nasional Indonesia | Sosial | 12/04/2023 | Pelajar | Rejang Lebong |
| 17 | P | Merayakan Kehilangan | Novel | 12/04/2023 | Pelajar | Rejang Lebong |
| … | …......... | …....................... | …................. | …................. | …................. | …...................... |
| … | …......... | …....................... | …................. | …................. | …................. | …...................... |

| No | Gender | Book Title | Book Type | Borrowing Date | Work | Address |
|---|---|---|---|---|---|---|
| 310 | L | Diet Gampang Tanpa Sebuah Larangan | Referensi | 24/08/2023 | Pelajar | Air Meles Bawah |
| 311 | L | Resep Sehat Cantik Ala Selebritis | Referensi | 24/08/2023 | Umum | BTN Air Bang |

Based on the data above, the attributes to be used for the classification analysis were determined. The attributes are as follows:

a. Book Genre: The titles of the books borrowed by the visitors were grouped by genre. In regional library X, there are 11 book genres: Religion, Short Stories, Encyclopedias, Mini Encyclopedias, Dictionaries, Health, Motivation, Novels, References, Literature, and Social Sciences.

b. Gender: Classified into two categories: Male and Female.

c. Occupation: Divided into four categories:
   1. Student: Includes elementary, junior high, high school, and vocational school students.
   2. College Student: University or college students.
   3. Teacher: Includes lecturers, teachers, and school principals.
   4. General Public: Includes individuals such as NGO members, military personnel, and housewives.

d. Interest: This is the target attribute or class, which consists of two categories: "yes" and "no."

After preprocessing and transforming the data, the dataset for library visitors is shown in Table 2.

**Table 2.** Regional Library Visitor Dataset

| No | Book Type | Gender | Occupation | Interest |
|---|---|---|---|---|
| 1 | Ensiklopedia | Female | University Student | Yes |
| 2 | Ensiklopedia | Female | University Student | Yes |
| 3 | Ensiklopedia | Male | University Student | Yes |
| 4 | Ensiklopedia | Male | Student | No |
| 5 | Ensiklopedia | Male | General | No |
| 6 | Motivasi | Female | University Student | No |
| 7 | Novel | Female | University Student | Yes |
| 8 | Novel | Female | Student | No |
| 9 | Novel | Female | General | No |
| 10 | Novel | Male | University Student | No |
| 11 | Novel | Male | General | Yes |
| 12 | Referensi | Female | General | Yes |
| 13 | Referensi | Male | General | No |
| 14 | Referensi | Male | University Student | Yes |
| 15 | Sastra | Female | Student | Yes |
| 16 | Sosial | Female | Student | No |
| 17 | Sosial | Male | University Student | No |
| 18 | Agama | Female | University Student | Yes |
| 19 | Agama | Female | Student | Yes |
| 20 | Agama | Male | Student | Yes |
| … | ….. | …........... | …......... | …........ |
| 113 | Sosial | Male | Student | No |

The following steps describe the process of analyzing the data using Python programming:

1.  Import the necessary libraries such as NumPy, Pandas, and Scikit-learn.

```
[5]  import numpy as np
     import pandas as pd

     from sklearn.preprocessing import LabelEncoder
     from sklearn.model_selection import train_test_split
     from sklearn.preprocessing import StandardScaler
     from sklearn.naive_bayes import GaussianNB
     from sklearn.metrics import confusion_matrix
     from sklearn.metrics import classification_report
     from sklearn.metrics import accuracy_score
```

**Figure 2.** Python Code

2.  Reading the dataset file, which is stored in an Excel file.

```
dataset = pd.read_excel('datasetuji.xlsx')
dataset.head()
```

|   | Jenis Buku | Jenis Kelamin | Pekerjaan | Minat |
|---|------------|---------------|-----------|-------|
| 0 | Ensiklopedia | Perempuan | Mahasiswa | Ya |
| 1 | Ensiklopedia | Perempuan | Pelajar | Ya |
| 2 | Ensiklopedia | Laki-laki | Mahasiswa | Ya |
| 3 | Ensiklopedia | Laki-laki | Pelajar | Tidak |
| 4 | Ensiklopedia | Laki-laki | Umum | Tidak |

**Figure 3.** Regional Library Visitor Dataset

3.  The dataset is visualized in the form of a graph:

```
from matplotlib import pyplot as plt
dataset['Jenis Buku'].plot(kind='hist', bins=20, title='Jenis Buku')
plt.gca().spines[['top', 'right',]].set_visible(False)
```
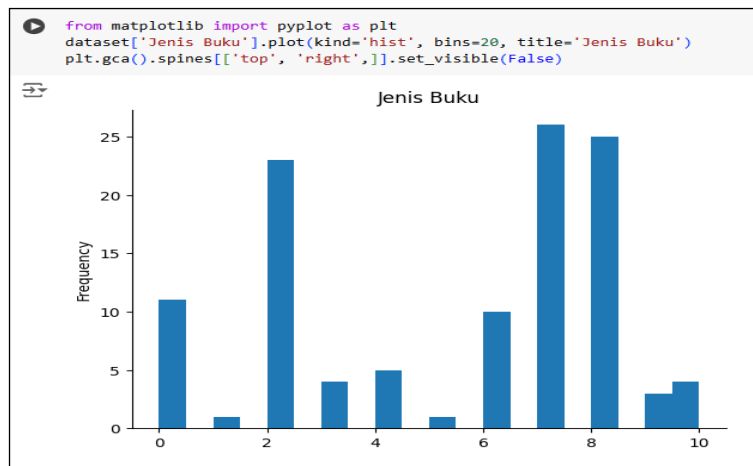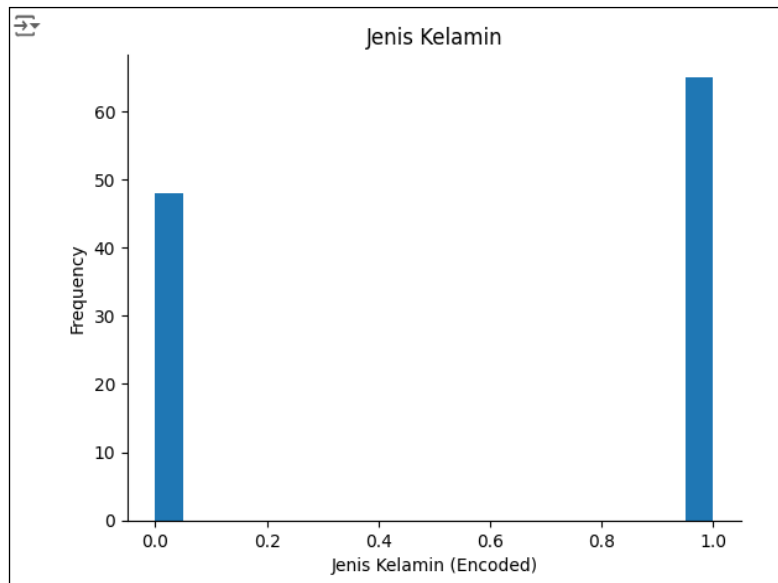


**Figure 4.** Dataset of Book Genres

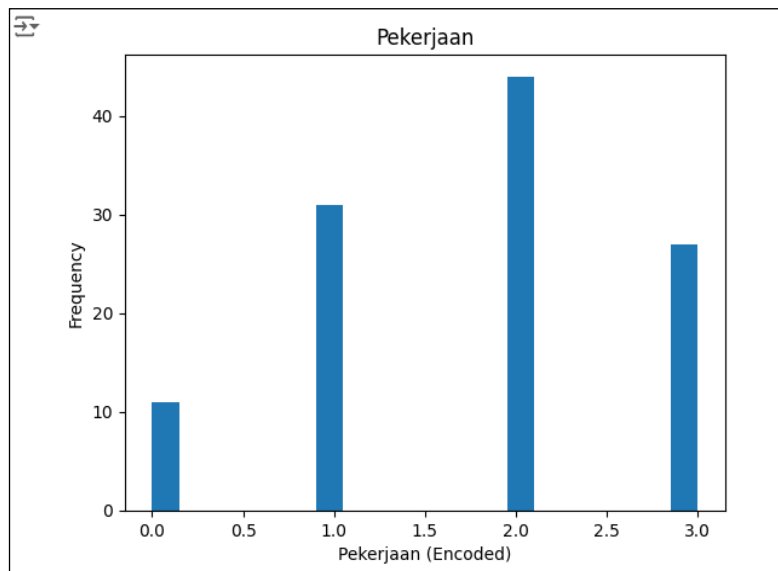**Figure 5.** Dataset of Genders



**Figure 6.** Dataset of Occupations

4. To get detailed information about the dataset, the following script is used to display dataset information.

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113 entries, 0 to 112
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Jenis Buku    113 non-null    object
 1   Jenis Kelamin 113 non-null    object
 2   Pekerjaan     113 non-null    object
 3   Minat         113 non-null    object
dtypes: object(4)
memory usage: 3.7+ KB
```
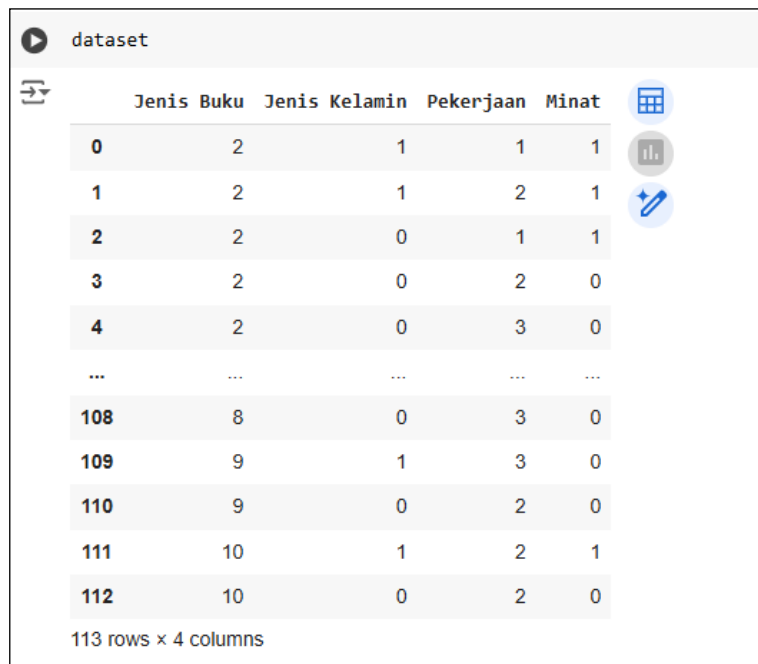
**Figure 7.** Python Dataset Info Code

5. Check if there is any missing data in the dataset.



**Figure 8.** Dataset Missing Values Check

6. The dataset is displayed in numerical form.



dataset

| | Jenis Buku | Jenis Kelamin | Pekerjaan | Minat |
|---|---|---|---|---|
| 0 | 2 | 1 | 1 | 1 |
| 1 | 2 | 1 | 2 | 1 |
| 2 | 2 | 0 | 1 | 1 |
| 3 | 2 | 0 | 2 | 0 |
| 4 | 2 | 0 | 3 | 0 |
| ... | ... | ... | ... | ... |
| 108 | 8 | 0 | 3 | 0 |
| 109 | 9 | 1 | 3 | 0 |
| 110 | 9 | 0 | 2 | 0 |
| 111 | 10 | 1 | 2 | 1 |
| 112 | 10 | 0 | 2 | 0 |

113 rows × 4 columns

**Figure 9.** Numerical Representation of Dataset

7. The data is split into training and testing sets, as shown in the following script.

```
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.3, random_state=123)

print("x_train = ", len(x_train))
print("x_test = ", len(x_test))
print("y_train = ", len(y_train))
print("y_test = ", len(y_test))

x_train =  79
x_test =  34
y_train =  79
y_test =  34
```

**Figure 10.** Data Splitting Script

8.  The Naïve Bayes algorithm is then implemented.



```
[21] classifier = GaussianNB()
     classifier.fit(x_train,y_train)
```
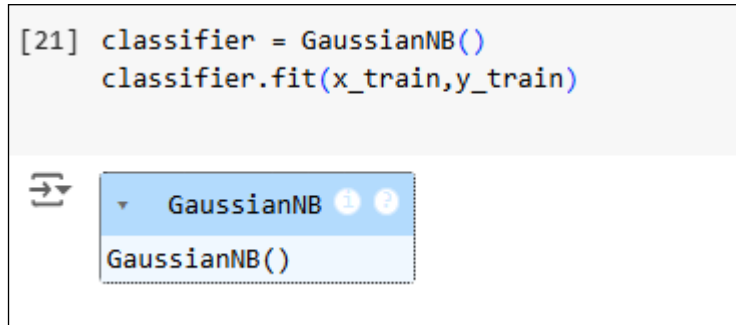
GaussianNB
GaussianNB()

**Figure 11.** Naïve Bayes Algorithm Implementation

9.  A confusion matrix is generated to assess the model's performance.



```
[24] cm = confusion_matrix(y_test,y_pred)
     print(cm)

     [[ 7  5]
      [ 7 15]]
```
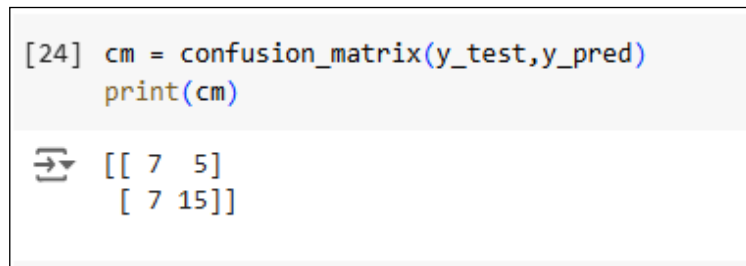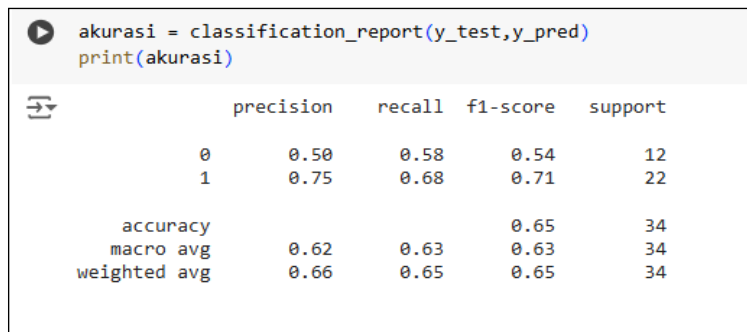
**Figure 12.** Confusion Matrix

10. The classification report shows that the model achieved an accuracy of 65%, with a precision of 62%, recall of 63%, and an F1-score of 63%.



```
akurasi = classification_report(y_test,y_pred)
print(akurasi)

              precision    recall  f1-score   support

           0       0.50      0.58      0.54        12
           1       0.75      0.68      0.71        22

    accuracy                           0.65        34
   macro avg       0.62      0.63      0.63        34
weighted avg       0.66      0.65      0.65        34
```
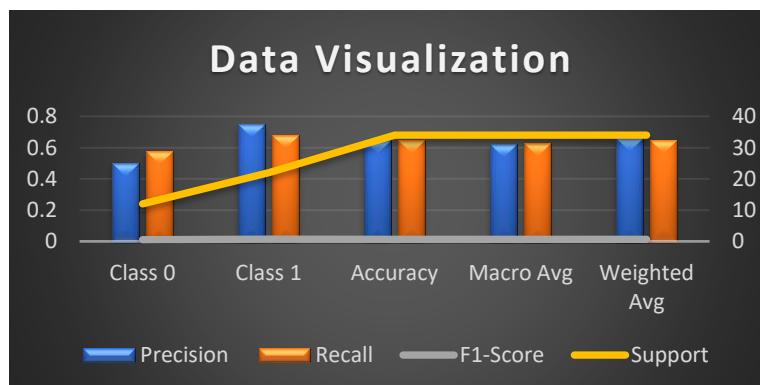
**Figure 13.** Model Accuracy Results



**Figure 14.** Visualization of Model Accuracy Results

The classification model demonstrates moderate performance with an overall accuracy of 65%. In evaluating each class, the model performs better in identifying class 1, with a precision of 0.75, recall of 0.68, and an F1-Score of 0.71. In contrast, the model's performance for class 0 is relatively lower, with a precision of 0.50, recall of 0.58, and an F1-Score of 0.54, indicating more misclassifications for this class. The macro average (precision 0.62, recall 0.63, F1-Score 0.63) provides an overall view of model performance by giving equal weight to each class, while the weighted average (precision 0.66, recall 0.65, F1-Score 0.65) better reflects the data distribution, as class 1 has more instances than class 0. In conclusion, although the model performs reasonably well on class 1, further optimization is required to improve accuracy and consistency in classifying instances of class 0.

## 4. Conclusion

This study successfully implemented the Naïve Bayes algorithm to classify the reading interest of regional library visitors. The results show that the algorithm is capable of classifying reading interest with an accuracy of 65%, precision of 62%, recall of 63%, and an F1-score of 63%. While the results are reasonably good, there is room for improvement, especially in enhancing precision and recall to achieve more accurate and relevant predictions in line with the library's needs.

The findings of this study suggest that utilizing the Naïve Bayes algorithm to analyze reading interest can significantly contribute to mapping library visitors' preferences. By adopting a data-driven approach, library managers can design programs and provide book collections that are more aligned with visitors' preferences, ultimately aiming to improve public reading interest.

Therefore, this study provides valuable contributions to the development of reading interest analysis methods using the Naïve Bayes algorithm, which can be adapted by other libraries as a strategy to increase visitor engagement and strengthen the culture of literacy in society. The study can be further developed by exploring hybrid algorithms or comparing Naïve Bayes with other classification methods, such as Decision Trees or Support Vector Machines, to enhance predictive accuracy in a library context.

## References

[1] W. Wardiana, A. Fadli, and D. Wahyudiati, "Hubungan Pemanfaatan Perpustakaaan Sekolah dan Minat Baca terhadap Hasil Belajar Siswa Kelas XII IPS di MA AL-Ijtihad Danger Lombok Timur," *JS*, vol. 10, no. 2, Art. no. 2, Dec. 2021, doi: 10.20414/schemata.v10i2.4090.

[2] Z. Azman, "Dakwah Bagi Generasi Milenial Melalui Media Sosial," *Jurnal Khabar: Komunikasi dan Penyiaran Islam*, vol. 3, no. 2, Art. no. 2, 2021, doi: 10.37092/khabar.v3i2.350.

[3] A. Karimah, N. Alfatikarahma, and A. Fauziah, "Studi Literatur: Peran Penting Literasi Membaca dalam Upaya Meningkatkan Karakter Positif Siswa Sekolah Dasar," *Indo-MathEdu Intellectuals Journal*, vol. 5, no. 1, pp. 623–634, Jan. 2024, doi: 10.54373/imeij.v5i1.670.

[4] Y. Wijaya, "Implementasi data mining dengan metode support vector machine pada aplikasi fundraising UKM LDK Syahid," bachelorThesis, Fakultas Sains Teknologi UIN Syarif Hidayatullah Jakarta, 2024. Accessed: Sep. 25, 2024. [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/80019

[5] A. Habibillah, T. Terttiaavini, and A. Heryati, "Pengembangan Perpustakaan Digital Untuk Meningkatkan Minat Membaca Siswa SD Negeri 8 Rantau Bayur Palembang," *Klik - Jurnal Ilmu Komputer*, vol. 3, no. 1, Art. no. 1, Mar. 2022, doi: 10.56869/klik.v3i1.340.

[6] - Elmelia Syari, "Pengelolaan Perpustakaan Dalam Meningkatkan Minat Baca Siswa Di Sekolah Menengah Atas Negeri 1 Kampar," skripsi, UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU, 2024. Accessed: Sep. 25, 2024. [Online]. Available: http://repository.uin-suska.ac.id/82948/

[7] A. Ahmad Fauzi *et al.*, *Pemanfaatan Teknologi Informasi di Berbagai Sektor Pada Masa Society 5.0*. PT. Sonpedia Publishing Indonesia, 2023.

[8] F.-J. Yang, "An Implementation of Naive Bayes Classifier," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2018, pp. 301–306. doi: 10.1109/CSCI46756.2018.00065.

[9] S. Rana, R. Kanji, and S. Jain, "Comparison of SVM and Naïve Bayes for Sentiment Classification using BERT data," in *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, Nov. 2022, pp. 1–5. doi: 10.1109/IMPACT55510.2022.10029067.

[10] K. Chong and N. Shah, "Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis," *International Journal of Advanced*

*Computer Science and Applications (IJACSA)*, vol. 13, no. 12, Art. no. 12, 33/30 2022, doi: 10.14569/IJACSA.2022.0131213.

[11] R. Salam and H. Usman, "Pengaruh Minat Baca dan Ketersediaan Sumber Belajar Terhadap Prestasi Belajar Siswa Kelas IV SDN Centre Mangalli Kecamatan Pallangga Kabupaten Gowa," 2021.

[12] R. K. Sari *et al.*, *Metodologi Penelitian Pendidikan*. Sada Kurnia Pustaka, 2023.

[13] H. Maulana, T. Repelita, I. Purnamasari, and S. Azzahwa, "Pengaruh Minat Baca dan Ketersedian Sumber Belajar Terhadap Prestasi Belajar".

[14] M. Murlena and D. Apriana, "Penerapan Data Mining Untuk Memprediksi Ketersediaan Stok Produk HNI HPAI Menggunakan Algoritma C4.5," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 2, no. 1, Art. no. 1, Jun. 2022, doi: 10.29240/arcitech.v2i1.5271.

[15] M. Anggraeni, "Pengembangan Dan Penerapan Repositori Institusi Di Perguruan Tinggi Dilihat Dari Dua Fungsi Manajemen: Studi Kasus di UI dan IPB," p. 14, 2014.

[16] M. Murlena and W. Syahindra, "DATA MINING PENGOLAHAN PENEMPATAN LIBRARY BOOKS MENGGUNAKAN METODE ASSOCIATION RULE DENGAN ALGORITMA APRIORI," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 5, no. 2, Art. no. 2, Sep. 2020, doi: 10.24252/instek.v5i2.16203.

[17] M. Imron, "Penerapan Data Mining Algoritma Naives Bayes dan Part Untuk Mengetahui Minat Baca Mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto," vol. 10, no. 2, 2017.

[18] D. Lianda and N. S. Atmaja, "Prediksi Data Buku Favorit Menggunakan Metode Naïve Bayes (Studi Kasus: Universitas Dehasen Bengkulu)," *pseudocode*, vol. 8, no. 1, pp. 27–37, Mar. 2021, doi: 10.33369/pseudocode.8.1.27-37.

[19] E. Rahma, *Akses dan Layanan Perpustakaan: Teori dan Aplikasi*. Kencana, 2018.