JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 7, No. 4. September 2025 DOI: https://doi.org/10.34288/jri.v7i4.409
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2025

PREDICTION OF LIBRARY BOOK BORROWING PATTERNS USING THE RANDOM FOREST ALGORITHM

Ega Ranaldi Pebriansyah[-1], Susanti[-2], Rahmiati[-3], Triyani Arita Fitri[-4]

Teknik Informatika
Universitas Sains Dan Teknologi Indonesia
egaranaldi93@gmail.com[-1], susanti@usti.ac.id[-2], rahmiati@usti.ac.id[-3], triyani@usti.ac.id[-4]

Abstract

Libraries play a crucial role in supporting the improvement of public literacy by providing reading materials tailored to users' needs and interests. One of the challenges faced by the Bukit Batu District Public Library is that the collection acquisition analysis process is not yet based on comprehensive borrowing patterns, potentially resulting in inaccurate results. This study aims to predict book borrowing patterns and classify collections into popular and unpopular categories using the Random Forest algorithm. Historical book borrowing data from 2019 to 2024 was used as the primary source in the model training and testing process. Testing was conducted with three data sharing ratios, namely 70:30, 80:20, and 90:10, which resulted in prediction accuracy of 89.19%, 88.69%, and 86.74%, respectively. Based on the analysis results, mathematics books were identified as the most popular collection with 146 borrowings, while social studies books were categorized as unpopular with 122 borrowings. These findings are expected to serve as a reference for libraries in formulating more effective, efficient, and data-based collection management strategies, thereby increasing the relevance and attractiveness of collections for users and supporting the optimization of library services.

Keywords: Library, Borrowing Pattern, Random Forest, Prediction, Popular Collection

Abstrak

Perpustakaan memegang peran penting dalam mendukung peningkatan literasi masyarakat melalui penyediaan bahan bacaan yang sesuai dengan kebutuhan dan minat pengguna. Salah satu tantangan yang dihadapi Perpustakaan Umum Kecamatan Bukit Batu adalah proses pengadaan koleksi yang belum berbasis pada analisis pola peminjaman secara menyeluruh, sehingga berpotensi kurang tepat sasaran. Penelitian ini bertujuan memprediksi pola peminjaman buku dan mengklasifikasikan koleksi ke dalam kategori populer dan tidak populer menggunakan algoritma Random Forest. Data historis peminjaman buku dari tahun 2019 hingga 2024 digunakan sebagai sumber utama dalam proses pelatihan dan pengujian model. Pengujian dilakukan dengan tiga rasio pembagian data, yaitu 70:30, 80:20, dan 90:10, yang menghasilkan akurasi prediksi berturut-turut sebesar 89,19%, 88,69%, dan 86,74%. Berdasarkan hasil analisis, buku matematika MTK teridentifikasi sebagai koleksi paling populer dengan 146 kali peminjaman, sedangkan buku IPS masuk kategori tidak populer dengan 122 kali peminjaman. Temuan ini diharapkan dapat menjadi acuan bagi pihak perpustakaan dalam merumuskan strategi pengelolaan koleksi yang lebih efektif, efisien, dan berbasis data, sehingga dapat meningkatkan relevansi dan daya tarik koleksi bagi pengguna serta mendukung optimalisasi layanan perpustakaan.

Kata kunci: Perpustakaan, Pola Peminjaman, Random Forest, Prediksi, Koleksi Populer

INTRODUCTION

Libraries serve as centers of information and learning that play a vital role in enhancing the literacy and knowledge of the community (Mahmuda et al., 2023). They provide a wide range of information collections, especially books, that the public can use to support learning, research, and knowledge development. As literacy centers, libraries play a strategic role in fostering a reading culture and improving the quality of human resources, as exemplified by the Public Library of Bukit Batu District (Rosman & Nining, 2022). The Public Library of Bukit Batu District is a Technical Implementation Unit (UPT) under the

Department of Library and Archives of Bengkalis Regency, Riau Province. Located on Jalan Hang Leqiu, Sejangat Village, the library acts as an information and literacy service center for the community in Bukit Batu District. In carrying out its function, the library is responsible for providing reading materials that align with the needs and interests of visitors. However, in practice, book procurement is still conducted without analyzing borrowing patterns, leading to a mismatch between the available collections and user preferences. This affects the effectiveness of services and the utilization of collections. Therefore, this study aims to analyze borrowing patterns to identify books that are considered popular or unpopular. The results of this research are expected to form the basis for planning more targeted, efficient, and appropriate book acquisitions for future library users. To address this issue, an analysis using the Random Forest algorithm is needed to predict library book borrowing patterns. The Random Forest algorithm is considered effective for prediction and classification tasks, especially in the context of library borrowing patterns. As a supervised learning method, Random Forest constructs multiple decision trees and combines their results to produce accurate and stable predictions. It can process complex and heterogeneous data, both numeric and categorical, which are common in library information systems. Moreover, Random Forest provides feature importance analysis that allows researchers to identify the most influential factors in borrowing behavior. Thus, the use of Random Forest in this study not only aims to build an accurate prediction model but also to extract valuable knowledge from historical borrowing data for strategic decision-making in library management. Previous studies have also applied Random Forest for prediction tasks. Previous studies have applied the Random Forest algorithm in various domains with notable results. For instance, (Daimari et al, 2021) used it to predict book availability in libraries and achieved 80% accuracy. (Fatunnisa and Marcos, 2024) applied it to predict on-time graduation of vocational school students, reaching 100% accuracy. (Putri, 2024) used it for stroke risk prediction with 99% accuracy, while (Wibowo and Rohman, 2022) applied it to predict heart failure with 82% accuracy. Although these studies demonstrate the effectiveness of the Random Forest algorithm, there remains a research gap specifically in the library science domain, especially regarding user borrowing behavior. Most prior works either focus on resource availability or are conducted outside the library context altogether. Even in library-related studies, the focus has rarely been on using borrowing pattern analysis to inform strategic book collection planning. This study addresses that gap by applying the Random Forest algorithm to historical borrowing data from a public library, aiming to identify popular and unpopular collections. The findings are intended to directly support data-driven decision-making in library collection development.

RESEARCH METHODS

This study applies a machine learning approach using the Random Forest algorithm to predict book borrowing patterns at the Bukit Batu District Public Library. The research process is carried out systematically through several key stages, starting from data collection to model evaluation. The overall research flow is illustrated in the following diagram:

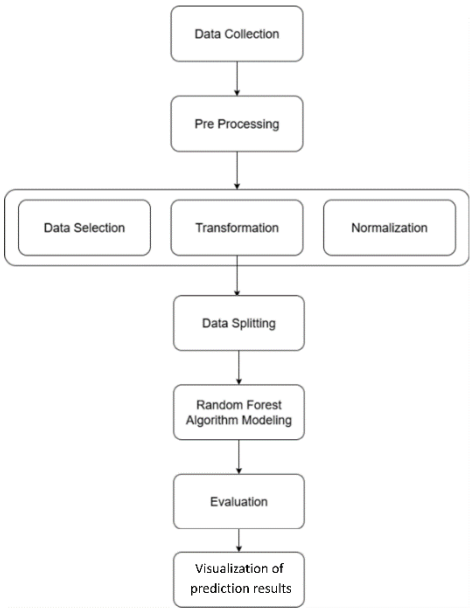


Figure 1. Research Stages

1. **Data Collection**
   The first stage in this research involves collecting data from the historical records of book borrowings at the Bukit Batu District Public Library, Bengkalis Regency, Riau Province, covering the period from 2019 to 2024. The dataset contains a total of 1,348 entries, comprising information related to the books borrowed and user profiles. The

data collection technique employed includes direct observation of the library's digital borrowing system and interviews (Novi Rudiyanti et al., 2025) with the head of the library and administrative staff. This ensures that the data used are both relevant and representative of actual user behavior over time.

2. Data Preprocessing

Preprocessing is a crucial step in preparing raw data for effective modeling (Dhewayani et al., 2022). It ensures data consistency, accuracy, and suitability for machine learning algorithms (Normawati & Prayogi, 2021). This study applies the following preprocessing steps:

a. Data Selection: From the original dataset, only relevant variables were selected for analysis. These include: book_category (e.g., fiction, education, religion) book_title borrower_status (public, student, college student) age of the borrower (Suryati, 2022)
b. Data Transformation: As the dataset includes categorical variables, such as book category and borrower status, these were converted into numerical representations using Label Encoding. This allows the machine learning model to process them appropriately (Permana et al., 2024).
c. Normalization: The age variable was normalized using Min-Max Scaling to ensure that the numerical values fall within a similar range(Prasojo & Haryatmi, 2021). Additionally, missing or null values were handled through appropriate data cleaning techniques to avoid bias or errors during modeling (Normawati & Prayogi, 2021).

3. Data Splitting

After preprocessing and EDA, the dataset is split into training and testing subsets. The study explores three different data splitting ratios: 70:30 80:20 90:10 (Setiawan et al., 2024).

These scenarios are used to evaluate the consistency and reliability of the model across various proportions of training data (Natzir, 2023). The goal is to ensure that the model is trained adequately while being tested on unseen data to assess its generalization capability.

4. Random Forest Modeling

The core of this study is the application of the Random Forest algorithm, a robust ensemble method under supervised learning (Sinambela et al., 2023). This algorithm constructs multiple decision trees from randomly sampled subsets of the training data, and the final prediction is made through majority voting (Nurul Chairunnisa, 2023). Advantages of using Random Forest include:

a. High accuracy and stability in prediction.
b. Ability to handle both numerical and categorical data.
c. Built-in feature importance, which helps identify the most influential variables (e.g., book category) in the prediction process.
d. Resistance to overfitting, especially in high-dimensional data.

The target variable is the label indicating whether a book is categorized as "popular" or "not popular", which is determined by borrowing frequency and user engagement.

5. Model Evaluation

To assess the performance of the trained model, the study uses a Confusion Matrix, which compares actual vs. predicted values (Fadli et al., 2024). From this matrix, the following metrics are derived:

a. Accuracy
   Proportion of correct predictions to total predictions (Arya Darmawan et al., 2023).
b. Precision
   The ability of the model to correctly identify positive instances (e.g., predicting a book is popular when it truly is) .
c. Recall (Sensitivity)
   The ability of the model to capture all actual positive instances (Rahmayanti et al., 2022).
d. F1-Score
   The harmonic mean of precision and recall, which balances both metrics (Ariyono Setiawan et al., 2023).

These evaluation metrics help determine whether the model performs reliably and can be used as a decision-support tool in planning book procurement and collection development strategies.

6. Visualization of prediction results

At this stage, researchers visualize the prediction results using the random forest algorithm. This involves displaying the most and least popular books in the form of a bar graph based on the data obtained (Wibowo & Rohman, 2022).

RESULTS AND DISCUSSION

1. Data Collection

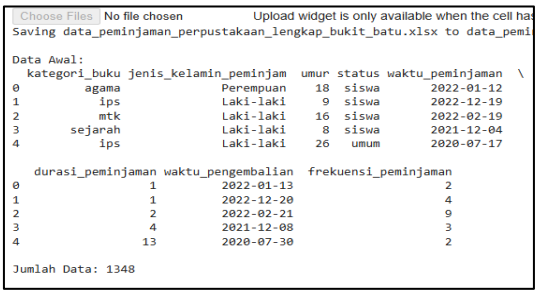
The data collection stage is the initial step in predicting library book borrowing patterns using the random forest algorithm. It requires a dataset to be processed and used in the prediction. For this data collection, the author directly obtained a private dataset from the Bukit Batu District Public Library in Riau Province, comprising 1,348 records with 8 Variable.

2. Importing Libraries

The first step in the machine learning-based data analysis process. It begins with importing the libraries needed to: manage and process data, perform visualization, share data and train models, perform evaluation, and visualize the model's decision results. This step is crucial as a foundation before starting to read the data, clean it, label it, and train the predictive model.

3. Enter Dataset

This step is part of the data collection process for the analysis. In Google Colab, because it cannot directly access local files like in Jupyter Notebook, it is necessary to manually retrieve data from the user's computer. This step is crucial because it marks the beginning of the entire analysis process, namely reading and understanding the data structure that will be further analyzed in the preprocessing, EDA, modeling, and evaluation stages. The following is the result of calling the dataset in Google Collab.



```
Choose Files  No file chosen        Upload widget is only available when the cell has
Saving data_peminjaman_perpustakaan_lengkap_bukit_batu.xlsx to data_pemi

Data Awal:
  kategori_buku jenis_kelamin_peminjam  umur status waktu_peminjaman  \
0         agama               Perempuan    18  siswa       2022-01-12
1           ips               Laki-laki     9  siswa       2022-12-19
2           mtk               Laki-laki    16  siswa       2022-02-19
3       sejarah               Laki-laki     8  siswa       2021-12-04
4           ips               Laki-laki    26   umum       2020-07-17

   durasi_peminjaman waktu_pengembalian  frekuensi_peminjaman
0                  1          2022-01-13                     2
1                  1          2022-12-20                     4
2                  2          2022-02-21                     9
3                  4          2021-12-08                     3
4                 13          2020-07-30                     2

Jumlah Data: 1348
```

Figure 2. Research Stages

4. Preprocessing

Preprocessing is the initial stage in data analysis or machine learning, aiming to prepare raw data so that it can be processed effectively by models or algorithms. This step is crucial to ensure the quality of the data that will be used for model training and evaluation.

a. Data Selection

The first step in preprocessing is data selection, where only relevant columns or variables are chosen for further analysis and modeling. This step aims to focus solely on features that significantly contribute to the modeling process, thereby reducing noise and excluding irrelevant attributes. In this study, three columns were selected from the original dataset :

- kategori_buku: indicating the category or genre of the borrowed book,
- status: the status of the borrower (e.g., student, public, general),
- umur: the age of the borrower.

These features were considered sufficient and relevant for predicting borrowing trends.

Table 1. Sample of Dataset After Feature Selection

Variable	Data Sample		
kategori_buku	agama	ips	mtk
jenis_kelamin_peminjam	Perempuan	Laki-laki	Laki-laki
umur	18	9	16
status	siswa	siswa	siswa
waktu_peminjaman	2022-01-12	2022-12-19	2022-02-19
durasi_peminjaman	1	1	2
waktu_	2022-01-	2022-	2022-

pengembalian	13	12-20	02-21
frekuensi_ peminjaman	2	4	9

b. Data Normalization
After selecting relevant columns, the next step is data normalization. This process ensures consistent formatting for textual values, particularly in the kategori_buku and status columns. The normalization procedures include:
- Converting all characters to lowercase.
- Removing unnecessary whitespace from the beginning and end of each string.

This step is important to prevent recognition errors during modeling caused by differences in text format, such as "Fiksi" vs "fiksi" vs " FIKSI ".

Table 2. Sample Data Before and After Normalization

kategori_ buku (before)	***status (before)***	***kategori _buku (after)***	***status (after)***
Fiksi	*Siswa*	*fiksi*	*siswa*
AGAMA	*Umum*	*agama*	*umum*
sejarah	*Masyarakat*	*sejarah*	*masyarakat*
Komik	*siswa*	*komik*	*siswa*

c. Data Transformation
At this stage, popularity labels are assigned to each row of the dataset to indicate whether a book is classified as "popular" or "not popular", based on predefined rules. The assigned labels are stored in a new column called label_popularitas, which becomes the target variable for classification. Next, categorical variables are transformed into numerical values using Label Encoding from the Scikit-Learn library. This step is necessary because most machine learning algorithms, including Random Forest, require numeric input.

Table 3. Sample Data After Transformation and Labeling

kategori_encoded	**status_encoded**	**label_encoded**
0	2	1
3	1	1
7	2	1
2	0	1
5	1	0

In order to make the determination of the frequency of borrowing clearer, the researcher explains it in the following table.

Table 4. Borrowing Frequency Parameters

frekuensi_ peminjaman	**frekuensi_ kategori**	**populer**
5	Rendah (0)	Tidak Populer (0)
8	Rendah (0)	Tidak Populer (0)
10	Tinggi (1)	Tidak Populer (0)
13	Tinggi (1)	Populer (1)
20	Tinggi (1)	Populer (1)

5. Data Splitting
Data Splitting This step is a crucial part of the machine learning process, where the dataset is divided into two main subsets: training data and testing data.
The division process involves:
a. X refers to the features used as input for the model, which includes book category, user status, and age.
b. y refers to the target label to be predicted, i.e., the popularity status of the book (“popular” or “not popular”), which has been encoded into numeric values using LabelEncoder.

The data is split using the train_test_split function from the Scikit-learn library:
a. X_train and y_train are used to train the model.
b. X_test and y_test are used to evaluate the model’s performance.

The test_size=0.2 parameter means that 20% of the data is used for testing and the remaining 80% for training. The random_state=42 ensures reproducibility, allowing the same data split each time the code is run. In this study, the researcher experimented with three different data split ratios: 70:30 (70% training, 30% testing) 80:20 90:10.
The purpose is to compare the model's performance across different data proportions and determine which configuration yields the best results.

6. Modeling Using the Random Forest Algorithm

This phase involves training the prediction model using the Random Forest algorithm, which is an ensemble learning method that combines multiple decision trees. This step is the core of the prediction process, where the model is trained to recognize patterns between the input variables and the popularity label.

The results can be seen in the image below :

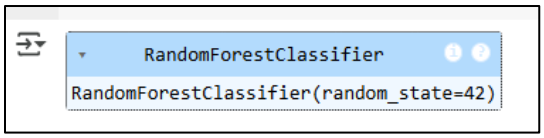


Figure 3. Modeling Using the Random Forest Algorithm

7. Model Evaluation

As for the results obtained, the researcher conducted an evaluation of the split data with a ratio of 70:30, 80:20 and 90:10, the following are the results of the evaluation of the 70:30 ratio which can be seen in the following image.

a. Evaluation Result for 70:30 Split

Akurasi: 89.19%

Classification Report:

	precision	recall	f1-score	support
0	0.98	0.86	0.92	640
1	0.76	0.96	0.85	304
accuracy			0.89	944
macro avg	0.87	0.91	0.88	944
weighted avg	0.91	0.89	0.89	944

Figure 4. Model evaluation results for 70:30 split

The test results using 70:30 data were 89.19%. The visualization of the results of the random forest algorithm evaluation using a 70:30 data ratio can be seen in the confusion matrix below.

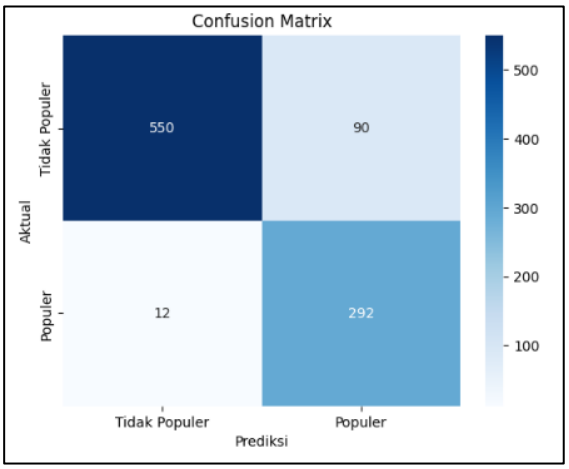


Figure 5. Confusion Matrix for 70:30 split

b. Evaluation Result for 80:20 Split

The following are the results of the evaluation of the 80:20 ratio which can be seen in the following image.

Akurasi: 88.69%

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.87	0.91	732
1	0.77	0.93	0.84	347
accuracy			0.89	1079
macro avg	0.87	0.90	0.88	1079
weighted avg	0.90	0.89	0.89	1079

Figure 6. Model evaluation results for 80:20 split

The test results using 80:20 data were 88.69%. The visualization of the results of the random forest algorithm evaluation using the 80:20 data ratio can be seen in the confusion matrix below.

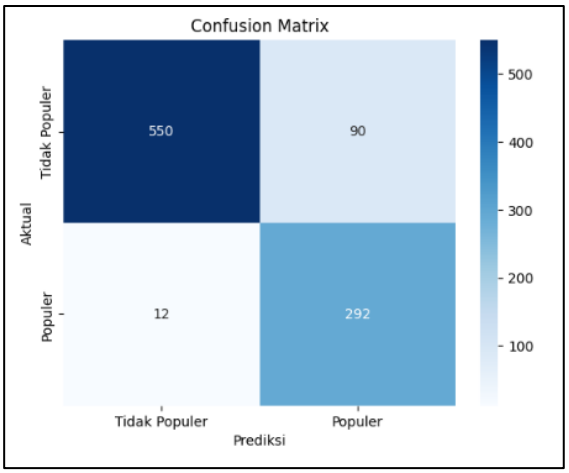


Figure 7. Confusion Matrix for 80:20 split

c. Evaluation Result for 90:10 Split
The following are the results of the evaluation of the 90:10 ratio which can be seen in the following image.

```
Akurasi: 86.74%

Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.87      0.90       823
           1       0.76      0.86      0.81       391

    accuracy                           0.87      1214
   macro avg       0.84      0.87      0.85      1214
weighted avg       0.87      0.87      0.87      1214
```

Figure 8. Model evaluation results for 90:10 split

The test results using 90:10 data were 86.74%. The visualization of the results of the random forest algorithm evaluation using a 90:10 data ratio can be seen in the confusion matrix below.

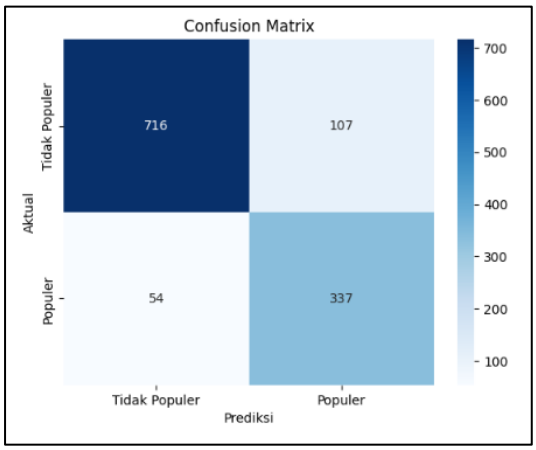


Figure 9. Confusion Matrix for 90:10 split

The following is a summary of the comparison of the results of the prediction test using the random forest algorithm which can be seen in the following table:

Table 5. Prediction Test Results

Splitting Data	***Akurasi***	***Presisi***	***Recall***	***F1-Score***
70:30	89,19%	87%	91%	89%
80:20	88,69%	87%	90%	88%
90:10	86,74%	84%	87%	85%

Based on the evaluation results in the table, the results show that data splitting significantly affects model performance. With a 70:30 data split, the model achieved the highest accuracy of 89.19% with a precision of 87%, a recall of 91%, and an F1-score of 89%. This demonstrates a good balance between the model's ability to correctly recognize positive data and avoid prediction errors. When the data proportion was changed to 80:20, model performance decreased slightly with an accuracy of 88.69% and an F1-score of 88%. Although the decrease was not significant, this result indicates that reducing the test data had a small impact on overall performance. With a 90:10 data split, the performance decrease became more pronounced with an accuracy of 86.74% and an F1-score of 85%. This decrease is likely due to the reduced amount of training data, so the model did not obtain enough information to optimally learn patterns. Overall, the 70:30 data split proved to provide the most optimal and balanced results for the tested models.

8. Visualization of prediction results
The next step is to visualize the prediction results. This visualization displays which books are popular and which are not. Here are the results:

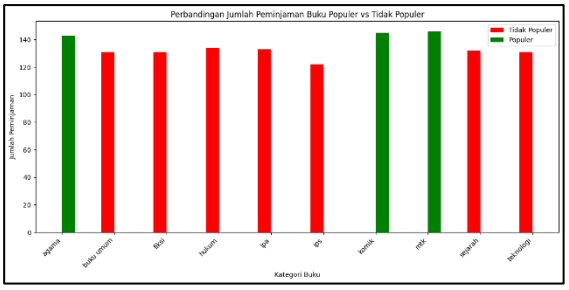


Figure 10. Popular and Unpopular Book Prediction Results Graph

Based on the graph above, we can conclude that the most popular book in the library is the Mathematics book with a total of 146 borrowings, while the unpopular book is the Social Studies book with a total of 122 borrowings and the remaining borrowings are filled by other books.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study successfully demonstrates that the Random Forest algorithm can be effectively applied to predict book borrowing patterns at the Public Library of Bukit Batu District, Riau Province. The model was developed using historical borrowing data from 2019 to 2024, aiming to classify books into two categories: popular and unpopular. The results indicate that the model is capable of accurately predicting the book categories based on previous borrowing patterns. Based on the performance evaluation using three data splitting ratios 70:30, 80:20, and 90:10 differences were observed in terms of accuracy and metric balance. With a 70:30 ratio, the model achieved the highest accuracy of 89.19%, along with a precision of 87%, recall of 91%, and an F1-Score of 89%, reflecting an optimal balance between prediction precision and completeness. Using the 80:20 ratio, accuracy slightly decreased to 88.69% with an F1-Score of 88%, although the performance remained relatively stable. In contrast, the 90:10 ratio resulted in a further drop in accuracy to 86.74% with an F1-Score of 85%, likely due to the reduced amount of training data, which limited the model's ability to recognize patterns effectively. Overall, the 70:30 ratio yielded the best results among the three scenarios tested. The findings also reveal that the MTK book was the most popular in the library, with the highest number of borrowings at 146 times. On the other hand, the IPS book was the least popular, with only 122 borrowings, while the remaining borrowings were distributed among other books. These results provide a clear insight into borrowing preferences at the library and can serve as a reference for future book collection management..

Suggestion

For further research, it is recommended that the same dataset be used for comparative analysis by applying various other classification algorithms, such as Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), or Neural Network..

REFERENCES

Ariyono Setiawan, Efendi, A. M. K. T. P., & Wibowo, U. L. N. (2023). Check On-Time Performance of Domestic Airlines Using Random Forest Machine Learning Analysis. *Technium Social Sciences Journal*, *47*, 379–397.

Arya Darmawan, M. B., Dewanta, F., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot. *TELKA - Telekomunikasi Elektronika Komputasi Dan Kontrol*, *9*(1), 52–61. https://doi.org/10.15575/telka.v9n1.52-61

Daimari, D., Narzary, M., Mazumdar, N., & Nag, A. (2021). Prediksi Ketersediaan Buku Berbasis Machine Learning untuak Sistem Manajemen Perpustakaan. *Library Philosophy and Practice*, *2021*.

Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, *12*(1), 64–77. https://doi.org/10.34010/jati.v12i1.6674

Fadli, A., Limbong, T., Priskila, R., & Handrianus Pranatawijaya, V. (2024). Penggunaan Algoritma Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(3), 3773–3779. https://doi.org/10.36040/jati.v8i3.9791

Fatunnisa, A., & Marcos, H. (2024). Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest. *Jurnal Manajemen Informatika (JAMIKA)*, *14*(1), 101–111. https://doi.org/10.34010/jamika.v14i1.12114

Mahmuda, F., Armys Roma Sitorus, M., Widyastuti, H., & Ely Kurniawan, D. (2023). Clustering Profil Pengunjung Perpustakaan (Studi Kasus Perpustakaan BP Batam). *Journal of Applied Informatics and Computing (JAIC)*, *1*(1), 2548–6861. http://jurnal.polibatam.ac.id/index.php/JAIC

Natzir, S. M. (2023). Prediksi Banjir menggunakan Naive Bayes Di Sleman. *Jurnal Teknologi Informasi*, *14*(2), 59–64. https://doi.org/10.52972/hoaq.vol14no1.p59-64

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *5*(2), 697–711.

Novi Rudiyanti, Mela Aprillia, Fanesha Rahma

Fitri, & Pupung Purnamasari. (2025). TEKNIK PENGUMPULAN DATA: OBSERVASI, WAWANCARA DAN KUESIONER. *JISOSEPOL: Jurnal Ilmu Sosial Ekonomi Dan Politik*, *3*(1), 132–138. https://doi.org/10.61787/zk322946

Nurul Chairunnisa. (2023). Prediksi Kemampuan Pembayaran Klien Home Credit Menggunakan Model Random Forest, Decision Tree, Dan Logistic Regression. *Jurnal Elektronika Dan Teknik Informatika Terapan (JENTIK)*, *1*(3), 140–147. https://doi.org/10.59061/jentik.v1i3.383

Permana, A. A., Taufiq, R., Destriana, R., & Nur'aini, A. (2024). Implementasi Algortima Naïve Bayes Untuk Prediksi Kelulusan Mahasiswa. *Jurnal Teknik*, *13*(1), 65–70. https://jurnal.umt.ac.id/index.php/jt/article/view/10996

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 79–89. https://doi.org/10.25077/teknosi.v7i2.2021.79-89

Putri, M. (2024). Prediksi Penyakit Stroke Menggunakan Machine Learning Dengan Algoritma Random Forest. *Jurnal Infomedia: Teknik Informatika*, *9*(2), 16–21.

Rahmayanti, A., Rusdiana, L., & Suratno, S. (2022). Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa. *Walisongo Journal of Information Technology*, *4*(1), 11–22. https://doi.org/10.21580/wjit.2022.4.1.9654

Rosman, Nining, H. (2022). Peningkatan Kemampuan Otomasi Perpustakaan bagi Pustakawan Madrasah di Provinsi Riau. *BIDIK: Jurnal Pengabdian Kepada Masyarakat*, *2*(2), 51–57. https://doi.org/10.31849/bidik.v2i2.9822

Setiawan, A., Febrio Waleska, R., Adji Purnama, M., Rahmaddeni, & Efrizoni, L. (2024). Komparasialgoritmak-Nearest Neighbor(K-Nn), Support Vector Machine(Svm), Dan Decision Treedalam Klasifikasi Penyakit Stroke. *Jurnal Informatika & Rekayasa Elektronika)*, *7*(1), 107–114. http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900

Sinambela, D. P., Naparin, H., Zulfadhilah, M., & Hidayah, N. (2023). Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin. *Jurnal Informasi Dan Teknologi*, *5*(3), 58–64. https://doi.org/10.60083/jidt.v5i3.393

Suryati, P. (2022). Analisis Pola Peminjaman Buku Dengan Menggunakan Algoritma Apriori. *JIKO (Jurnal Informatika Dan Komputer)*, *5*(1), 17. https://doi.org/10.26798/jiko.v5i1.509

Wibowo, A., & Rohman, A. (2022). *Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ*. *8*(200).