Journal of Economics AND MANAGEMENT SCIENTIES
https://jems.ink ISSN 2655-1934 (print), 2655-6685 (online)

Predicting Book Return Delays in Airlangga University Library: A Machine Learning Approach

Ayu Triandari[1], and Imam Yuadi[2]*

[1,2] *Universitas Airlangga, Indonesia*

ARTICLE HISTORY

Received: 27 September 25
Final Revision: 10 October 25
Accepted: 19 October 25
Online Publication: 31 December 25

KEYWORDS

Book Return Delays Prediction, Machine Learning, Naïve Bayes, Circulation Documents, Library Management Systems

KATA KUNCI

Prediksi Keterlambatan Pengembalian Buku, Pembelajaran Mesin, Naïve Bayes, Dokumen Sirkulasi, Sistem Manajemen Perpustakaan

CORRESPONDING AUTHOR

ayu.triandari-2024@fisip.unair.ac.id

DOI

10.37034/jems.v8i1.254

A B S T R A C T

This research aims to predicting the delay of book return in the Airlangga University library by using machine learning algorithms. With the consideration of approximately 1600 circulation documents from January to December 2023, several algorithms including Naïve Bayes, Support Vector Machine (SVM), Random Forest, Logistic Regression, Neural Networks, and Gradient Boosting are utilized. The Naïve Bayes model prove to be the most effective model by 92.7% accuracy and 97.7% precision in predicting return delay. The analysis of feature importance has demonstrated that a handful of features, especially days overdue, loan duration, and return date, are the main predictive variables for delay prediction in book returns outcomes. From this study, the Naïve Bayes can be an effective predictor of book return delays in Airlangga University library in order to improve user satisfaction, potentially notifying user in advance and offering alternatives. This study provided a promising picture about machine learning applications in library management systems for practical resource allocation and service quality improvement related to book return delays.

A B S T R A K

Penelitian ini bertujuan untuk memprediksi keterlambatan pengembalian buku di Perpustakaan Universitas Airlangga dengan menggunakan algoritma pembelajaran mesin. Dengan mempertimbangkan sekitar 1600 dokumen sirkulasi dari Januari hingga Desember 2023, beberapa algoritma termasuk Naïve Bayes, *Support Vector Machine* (SVM), *Random Forest*, Regresi Logistik, Jaringan Syaraf Tiruan, dan *Gradient Boosting* digunakan. Model Naïve Bayes terbukti menjadi model yang paling efektif dengan akurasi 92,7% dan presisi 97,7% dalam memprediksi keterlambatan pengembalian. Analisis kepentingan fitur telah menunjukkan bahwa beberapa fitur, terutama hari keterlambatan, durasi pinjaman, dan tanggal pengembalian, merupakan variabel prediktif utama untuk prediksi keterlambatan dalam hasil pengembalian buku. Dari studi ini, Naïve Bayes dapat menjadi prediktor yang efektif untuk keterlambatan pengembalian buku di Perpustakaan Universitas Airlangga untuk meningkatkan kepuasan pengguna, berpotensi memberi tahu pengguna sebelumnya dan menawarkan alternatif. Studi ini memberikan gambaran yang menjanjikan tentang aplikasi pembelajaran mesin dalam sistem manajemen perpustakaan untuk alokasi sumber daya praktis dan peningkatan kualitas layanan terkait keterlambatan pengembalian buku.

1. Introduction

Library functions are critical in ensuring that learning activities are supported and that persons get access to the knowledge resources [1]. Another important aspect of library collection management is the timely return of borrowed materials. This activity is critical in ensuring equity in access and satisfaction among users [2].

Book return delays tend to be a recurrent source of problems in almost all academic libraries and affect resource availability and general library service quality [3]. It strikes directly at the core principles of library management, calling for effective resource allocation, user satisfaction, and strategies to keep a well-working collection [4]. The return of most books very often goes late into the days that are mainly important for library operations. Returns being delayed mean that other users will not be able to access and use the book during that period of time, hence affecting the process of managing the book collection as well as resulting in a significant cost, including replacement costs [5], [6]. Further, over time, users become cynical owing to chronic late returns and therefore may not visit the library [7]. Traditionally using reminders for due dates, overdue fees, and manual tracking, libraries manage book returns [8].

However, these approaches have failed pretty much because of systemic delays [9]. The analysis of the data in relation to the mining of patterns in the use of libraries is quite essential as it can help in making well-informed decisions in regards to the users' data in the libraries [10]. Thus, the application of automation algorithm such as machine learning is necessary to predict the delay in book returns.

Meanwhile, several studies have explored machine learning in different activities in libraries. For instance, certain research looked into the use of machine learning in book borrowing patterns forecasting and estimation and thus providing sufficient evidence on the relevance of these methods in predicting circulation [11], [12]. Different research examined the possibility of employing neural networks for book recommendation system in universities and the scope of applicability of the models when the borrowing transactions are substitutive [13], [14]. Researcher utilized machine learning algorithm to predict the favourite books, showcasing the potential of these techniques for personalized library services [15]. Furthermore, machine learning has successfully predicted the availability of books which can help libraries manage inventory more effectively, ensuring that books are available when needed [16], [17]. Although exploration has been undertaken towards advanced machine learning models for several library tasks, machine learning applications to predict book return delays in university libraries remains underexplored.

The research gap highlights the need for further investigation into feature engineering adapted to predict book return delays, as well as performance comparison of different machine learning algorithms in this particular task. This study aims to contribute to this gap by identifying and developing as well as evaluating a machine learning model which will adequately determine the delay with regards to book returns in a case study at Airlangga University library.

The structure of the article "Predicting Book Return Delays in Airlangga University Library: A Machine Learning Approach" consists of several key sections. It begins with the introduction explains the importance of libraries and the issue of book return delays, as well as the need for machine learning applications. The literature review discusses previous research related to the use of machine learning in libraries and identifies research gaps in predicting book return delays. The methods section covers the dataset, exploratory data analysis, data preprocessing, algorithm selection, and model training. The results and discussion present performance metrics, feature importance analysis, and a comparison of the Naïve Bayes and SVM models. The conclusion summarizes the findings and practical implications, concluding with acknowledgments, conflict of interest statements, author contributions, and references.

2. Research Method

This section describes about the data acquisition as well as exploratory data, data preprocessing, the rationale behind the chosen machine learning algorithms, and the model training process employed in this study. Figure 1 shown the flowchart describes the typical steps involved in machine learning research about predicting book return delays in a case study at Airlangga University library.

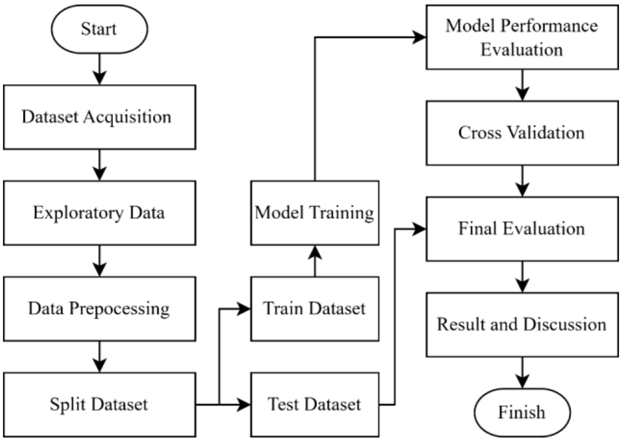

Figure 1. Flowchart of the research methodology

2.1. Dataset Acquisition

The dataset for this study is obtained from the Airlangga University library's circulation records. It encompasses a comprehensive collection of historical borrowing transactions, spanning from January to December in the year 2023 and comprises the records of approximately 1600 borrowed books circulation. Each transaction records includes a variety of information are partially shown in Table 1.

Table 1. Partial raw data on daily book circulation in the year 2023 at Airlangga University library

Inventory number	Bibliographic Data	Date Burrowed	Due Date	Date Returned
3556062111	*Mengenai hukum: suatu pengantar* Yogyakarta Liberty 2005 LAW 340	04/01/2023	08/02/2023	26/01/2023
3006082111	*Bab-bab tentang penemuan hukum* Jakarta: Citra Aditya Baktu, 1998 LAW 340	04/01/2023	08/02/2023	26/01/2023

2.2. Exploratory Data

Based on the statistics of book circulation data, the diagram of the monthly books borrowed and overdue volume of the library change in 2023 is shown in Figure 2. This figure suggest that volume borrowed book and overdue rates in Airlangga University library are influenced by academic calendars and seasonal patterns.

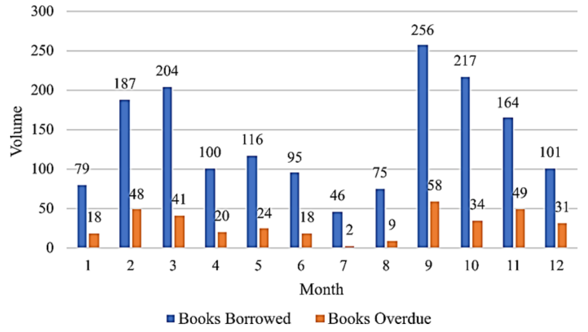

Figure 2. Monthly book circulation and overdue statistics, Airlangga University Library, 2023

2.3. Data Preprocessing

In order to prepare the dataset for training machine learning models, the following data preprocessing steps will be discussed below.

2.3.1. Handling Missing Values

The missing values within features will be addressed using appropriate techniques such as imputation with mean/median for numerical data or creating a separate "unknown" category for categorical data.

2.3.2. Feature Engineering

The features will be extracted from dataset as partially shown in Table 1 to enhance predictive ability. The features data for predicting delays in the book returns include borrow date, due date, return date, loan duration, days overdue, book category, classification number, and year published.

2.3.3. Data Target

The target variable for the purpose of predicting delays in the return of books has been labelled into two classes as on time and overdue.

2.4. Machine Learning Algorithm Selection

The next important step in the selection of machine learning algorithms for the predictions of book return delays relies on understanding the dataset and the goals to achieve. In this study, several machine learning algorithms are explored and compared the prediction results; each with its own strong point.

Gradient Boosting is the first of these: it offers a robust technique for ensemble learning, and it is pretty capable in finding complex structure in data [18]. This is because it combines lots of weak models to yield a strong predictor, thus making it classically very good at improving accuracy. While, Random Forest, an ensemble one which has other powerful ways to deal with data, being slightly better in classification tasks [19]. Whereas this included handling big data sources, it had reduced risk of overfitting, which can be a common pitfall in machine learning.

Logistic Regression is the best straightforward one among the rest for simpler cases of binary classification problems [20], [21]. Easy interpretation is a great advantage when it comes to understanding what influences book return delays. By contrast, Support Vector Machine (SVM) would be more flexible, since SVM could be used for both classification and regression tasks, while doing well in a high-dimensional space, this property made it another important means of analyzing complex data, such as book circulation data in library [22].

Also explored Neural Networks, which learned the complex relationships in the vast dataset quite well [22]. The flexibility associated with the networks enables them to model intricate patterns that cannot be detected by other algorithms. Finally, we added Naïve Bayes, a probabilistic classifier based on Bayes' theorem [23]. In fact, such classifier would be very handy when working on text classification or similar tasks, thus making significant inroads in prediction based on probabilities.

Ultimately, the choice of which algorithm to use will depend on several key factors: how the model to be, the computational resources available, and how easily we can interpret the results. By weighing these considerations, in order to select the best approach for predicting book return delays effectively.

2.5. Model Training and Validation

The dataset will be randomly split into a training set (75% of dataset) and a testing set (25% of dataset). Since the 75%/25% split is a common compromise the provide sufficient data for training while retaining enough data for robust testing [24]. The training set will be used to train the selected machine learning models. According Figure 1, thus the book circulation in year 2023 from January to September will be utilized as training dataset (75% of dataset), remains as the testing dataset. K-fold cross-validation (with k=10) will be applied to assess model generalization and prevent overfitting. Model performance will be evaluated based on metrics such as classification accuracy (CA), precision, train time and area under the curve (AUC), ensuring a comprehensive assessment of predictive capability.

3. Result and Discussion

This result section about the performance models, the features factor analysis, and the evaluation the machine learning models into testing dataset.

3.1. Performance Metrics and Model Comparison

The Table 2 shown the performance metrics of several machine learning algorithms applied to the task of predicting book return delays. The metrics considered are Accuracy (CA), Precision, Train Time, and AUC. Overall, Naïve Bayes emerges as the top-performing model with the highest accuracy (0.927) and precision (0.977). This suggests that Naïve Bayes is effective in correctly identifying both true positives and true negatives, making it a reliable model for this prediction task.

Table 2. Performance of several machine learning algorithm for prediction book return delay

Model	CA	Precision	Train time	AUC
Naïve Bayes	0.927	0.977	0.051	0.968
kNN	0.918	0.908	0.055	0.916
Logistic Regression	0.889	0.864	0.116	0.875
Random Forest	0.835	0.816	0.337	0.828
SVM	0.314	0.233	0.653	0.314
Gradient Boosting	0.957	0.924	0.671	0.928

The kNN and Logistic Regression models shown relatively strong performance, with accuracy and precision values above 0.8. However, these two models are outperformed by Naïve Bayes model. On the other hand, Random Forest and SVM shows significantly lower performance, particularly in terms of precision. This indicates that both Random Forest and SVM models may struggle to accurately identify true positives, leading to a higher number of false positives.

In terms of training time, Naïve Bayes and kNN are the most efficient models, requiring less than 0.1 seconds to train. Logistic Regression and Random Forest have longer training time, while Gradient Boosting takes the longest. But this model ends up having the best accuracy of all the others. The AUC (Area Under the Curve) metric provides insights into the model's ability to distinguish between positive and negative classes. Naïve Bayes model demonstrates strong performance with an AUC of 0.968 (see Table 2), indicating its ability to accurately rank instances based on their probability of being positive. From the data availed, the model that best fits the prediction of delays in book returns is the above Naïve Bayes models as far as accuracy, precision, and computational efficiency are concerned.

Table 3. Confusion matrix of Naïve Bayes model

		Predicted	
		On Time	Overdue
Actual	On Time	1046	8
	Overdue	6	252

Table 3 shown the confusion matrix for the Naïve Bayes model reveals a high level of accuracy in predicting book return delays. The model correctly classified 1046 books as being returned on time and 252 books as being overdue. With only 8 false positives and 6 false negatives, the model demonstrates strong performance in both sensitivity (correctly identifying overdue return books) and specificity (correctly identifying on-time return books).

3.2. Feature Importance Analysis

The feature factors analysis obtained through training proses in Naïve Bayes model. Figure 3 shows the ranking of the importance of various feature factors that affect the books return delay. From the figure 3, we can see that Days Overdue feature has the highest score, indicating that the number of days overdue is the most significant factor in predicting whether a book will be returned late or not. Other factors such as Loan Duration (days) feature and Return Date feature also have quite high scores, indicating that the duration of the loan and the actual return date also affect delays. Factors such as Borrow Date, Due Date, Book Category, Classification Number (DDC), and Year Published have lower scores, meaning that these features influence on the delay in returning books is not as great as other features.

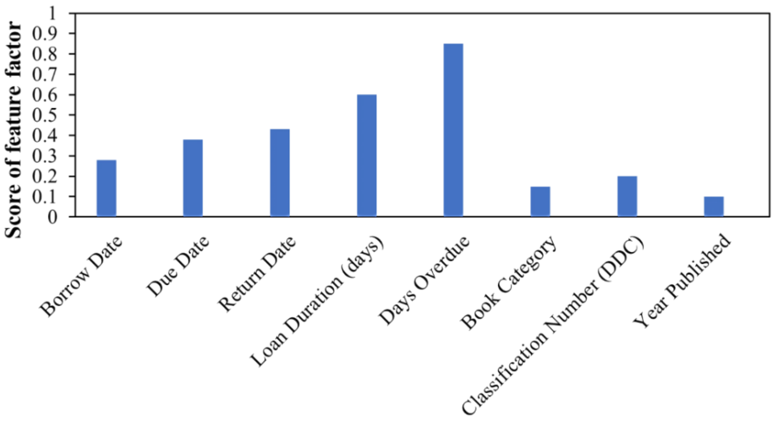

Figure 3. Ranking of feature factors for book return delay prediction

3.3. Comparation of Naïve Bayes and SVM Model Prediction on Books Return Delay

In order to evaluate the prediction performance of books return delay, thus we utilize the testing dataset (25% dataset,) to SVM and Naïve Bayes model, where the SVM has the worst in term of accuracy. The Table 4 show partial results of predicting book return delays using SVM and Naïve Bayes models. It compares the actual label of book return delay status with the predictions made by both models. The data reveals that the Naïve Bayes model tends to be more accurate in predicting overdue or on time of books return status than SVM, even Naïve Bayes prediction results almost agree with the actual label. For example, the book category Civil Law (the fourth row of data in Table 4) should have been returned on time to the Airlangga University library, but the SVM model predicted a late return, while Naïve Bayes had the correct prediction. For accumulative actual versus prediction results of SVM and Naïve Bayes model was shown in Figure 4.

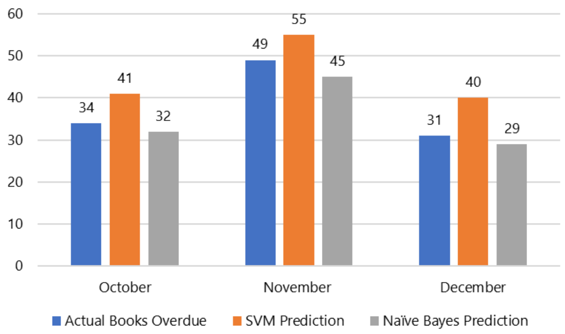

Figure 4. Comparison of actual and predicted overdue books using SVM and Naïve Bayes on testing data

Table 4. Partial testing dataset of prediction book delay result of SVM and Naïve Bayes, including the actual label of book return delay status

Borrow Date	Return Date	Days Overdue	Book Category	Label of Book Return Delay	Prediction Results of SVM	Prediction Results of Naive Bayes
19/10/2023	23/10/2023	0	Law-Philosophy	On Time	On Time	On Time
19/10/2023	23/10/2023	0	Law - Philosophy	On Time	On Time	On Time
19/10/2023	23/10/2023	0	Islamic Ethics	On Time	On Time	On Time
20/10/2023	06/11/2023	0	Civil Law	On Time	Overdue	On Time
20/10/2023	02/01/2024	21	Democratic Government	Overdue	Overdue	Overdue
20/10/2023	11/12/2023	0	Political Rights	Overdue	Overdue	Overdue
20/10/2023	23/10/2023	0	Bank & Banking-Religious Aspects	On Time	On Time	On Time
20/10/2023	23/10/2023	18	Risk Management	Overdue	Overdue	Overdue
20/10/2023	23/10/2023	0	Intellectual Property	On Time	On Time	On Time
20/10/2023	23/10/2023	0	*Kekayaan Intelektual*	On Time	On Time	On Time
20/10/2023	23/10/2023	0	Trademark - Law and Legislation	On Time	On Time	On Time
20/10/2023	23/10/2023	0	Intellectual Property	On Time	On Time	On Time
20/10/2023	23/10/2023	0	Japanese	On Time	On Time	On Time
20/10/2023	23/10/2023	0	Asia, Southeastern-History	On Time	On Time	On Time
20/10/2023	03/01/2024	0	Social Fiction	On Time	On Time	On Time
20/10/2023	27/10/2023	0	America Fiction	On Time	On Time	On Time
23/10/2023	08/12/2023	0	Alderian Psychology	On Time	Overdue	Overdue
23/10/2023	08/12/2023	0	Social Capital - Sociological Aspects	On Time	On Time	On Time

Figure 4 show a bar chart which compares the actual number of overdue books to the predicted number of overdue books, using a testing dataset, for October, November, and December 2023. The SVM model consistently overestimated the actual number of overdue books, while Naive Bayes tends to bend just a little closer to reality. This indicates that the Naive Bayes model can predict overdue books quite accurately; it may be a more trustworthy option for such case due to its usual not very conservative forecasts. The SVM again suffered from most inaccurate metrics in terms of defaulting overdue books, having most estimates crossing the threshold quite widely. For instance, in November 2023, the real overdue books increased to 49, while SVM indicated the figure at 55.

This research findings indicate that the Naïve Bayes model is the most effective in predicting book return delays based on performance metrics such as accuracy (92.7%), precision (97.7%), and AUC (96.8%). This advantage reflects the model's ability to accurately identify both on-time and late book returns while minimizing classification errors. Additionally, its very short training time (0.051 seconds) makes Naïve Bayes a computationally efficient choice compared to other models like Gradient Boosting, which, despite having the highest accuracy (95.7%), requires a longer training time (0.671 seconds). The k-Nearest Neighbors (kNN) and Logistic Regression models also demonstrated good performance with accuracy above 88%, but they fall short compared to Naïve Bayes in terms of precision and training time efficiency. On the other hand, models like Support Vector Machine (SVM) showed the worst performance, with an accuracy of only 31.4%, making them unsuitable for this task.

These findings are consistent with previous literature, which shows that Naïve Bayes often performs well in probabilistic classification tasks, especially when the dataset has clearly independent features [25]. However, these results also highlight the limitations of the SVM model, which may be due to its sensitivity to kernel parameters and the need for more complex parameter tuning [26]. Feature analysis shows that Days Overdue is the most significant factor in predicting book return delays, followed by Loan Duration and Return Date. Features such as Borrow Date, Due Date, book category, classification number (DDC), and publication year have a smaller impact on predictions. This insight provides libraries with guidance to focus on managing loan durations and return dates to reduce delays.

Moreover, minimizing book return delays has a significant impact on user satisfaction in libraries [27] and the use of technology, along with effective overdues policies and procedures, plays a crucial role in improving user satisfaction and reducing book return delays [17]. As in research conducted which states that delays in returning books have a significant impact on borrowing library collections, with an influence of 41.47% on the efficiency of circulation services [28]. This research supports the importance of predicting delays as a preventive measure to improve the efficiency of library services. Likewise, certain research highlights the importance of using technology such as automatic notifications before due dates to reduce delays [29]. This is relevant to current research findings which show that predictive models such as Naïve Bayes can be used to identify users who are at risk of being late, so that libraries can provide early warning.

4. Conclusion

In this research, it was successful in applying a machine learning method where Naive Bayes algorithm was able to carry out the prediction of the delayed return of books on Airlangga University library. The model showed an encouraging result by being above the many machine learning models with respect to accuracy and precision computation efficiency where it has 92.7% Naïve Bayes accuracy in prediction and 97.7% precision. In terms of computational efficiency measured through train time, the Naïve Bayes is 0.051 seconds. The analysis also highlighted the predictive importance of several features. These are category of book, publication year, and loan duration, further seen to greatly contribute to prediction accuracy. The understanding of these features will enable Airlangga University library or other university libraries to address the approaches regarding to collections that tend to delay in returns. Such findings have practical implications for library services, to be specific in Airlangga University. Airlangga University library can optimize library operations, improve resource allocation, and then enhance user satisfaction by using insights from this study.

References

[1] Xu, S. (2020). Association Rule Model of On-demand Lending Recommendation for University Library. *Informatica*, *44*(3), 395–399. https://doi.org/10.31449/INF.V44I3.3295

[2] Iwhiwhu, B., & Okorodudu, P. (2012). Public Library Information Resources, Facilities, and Services: User Satisfaction with the Edo State Central Library, Benin-City, Nigeria. *Library Philosophy and Practice*. Retrieved January 13, 2025, from https://digitalcommons.unl.edu/libphilprac/747

[3] Seger, R., Allen, L., & Lenny, A. (2011). A Publisher's Perspective on PDA. *Against the Grain*, *23*(3), 11. https://doi.org/10.7771/2380-176X.5889

[4] Ranganathan, S. R. (1964). Classification Research in India 1957—1963. Trend Report. *Libri*, *13*(3–4), 181–190.

[5] Stouthuysen, K., Swiggers, M., Reheul, A. M., & Roodhooft, F. (2010). Time-driven activity-based costing for a library acquisition process: A case study in a Belgian University. *Library Collections, Acquisitions, and Technical Services*, *34*(2–3), 83–91. https://doi.org/10.1016/J.LCATS.2010.05.003

[6] Siguenza-Guzman, L., Van den Abbeele, A., Vandewalle, J., Verhaaren, H., & Cattrysse, D. (2014). Using Time-Driven Activity-Based Costing to Support Library Management Decisions: A Case Study for Lending and Returning Processes. *The Library Quarterly*, *84*(1), 76–98. https://doi.org/10.1086/674032

[7] Joy, I. I., & Idowu, A.-I. (2014). Utilization and User Satisfaction of Public Library Services in South-West, Nigeria in the 21st Century: A Survey. *International Journal of Library Science*, *2014*(1), 1–6. https://doi.org/10.5923/J.LIBRARY.20140301.01

[8] Tyler, D. C., Falci, C., Melvin, J. C., Epp, M. L., & Kreps, A. M. (2013). Patron-Driven Acquisition and Circulation at an Academic Library: Interaction Effects and Circulation Performance of Print Books Acquired via Librarians' Orders, Approval Plans, and Patrons' Interlibrary Loan Requests. *Collection Management*, *38*(1), 3–32. https://doi.org/10.1080/01462679.2012.730494

[9] Sens, J. M., & Fonseca, A. J. (2013). A Skeptic's View of Patron-Driven Acquisitions: Is it Time to ask the Tough Questions? *Technical Services Quarterly*, *30*(4), 359–371. https://doi.org/10.1080/07317131.2013.818499

[10] Silwattananusarn, T., Kulkanjanapiban, P., & Athakravisunthorn, K. L. (2020). Mining and Analyzing Patron's Book-Loan Data and University Data to Understand Library Use Patterns. *International Journal of Information Science and Management*, *18*(2), 151–172. https://arxiv.org/abs/2008.03545v1

[11] Sun, J. (2021). Prediction and Estimation of Book Borrowing in the Library: Machine learning. *Informatica*, *45*(1), 163–168. https://doi.org/10.31449/INF.V45I1.3431

[12] Zhang, H. (2024). Machine Learning Algorithms for Predicting and Estimating Book Borrowing in University Libraries. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *28*(5), 1204–1209. https://doi.org/10.20965/JACIII.2024.P1204

[13] Choi, J., Han, C., Yang, H., Hong, Y., Jeon, S., & Zhu, Y. (2021). Embedding-based neural network models for book recommendation in university libraries. *CEUR Workshop Proceedings*, *2871*, 25–32.

[14] Shi, X., Hao, C., Yue, D., & Lu, H. (2023). Library book recommendation with CNN-FM deep learning approach. *Library Hi Tech*, *42*(5), 1559–1578. https://doi.org/10.1108/LHT-08-2022-0400/FULL/XML

[15] Daimari, D., Mondal, S., Brahma, B., & Nag, A. (2023). Favorite Book Prediction System Using Machine Learning Algorithms. *Journal of Applied Engineering, Technology and Science*, *4*(2), 983–991. https://doi.org/10.37385/JAETS.V4I2.1925

[16] Daimari, D., Narzary, M., Mazumdar, N., & Nag, A. (2021). A machine learning based book availability prediction model for library management system. *Library Philosophy and Practice*, *4982*.

[17] Geetha, K., Rao, G. S., Kaur, C., & Kumar, K. K. (2022). Machine learning based library management system. *6th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2022 - Proc.)*, 1031–1034. https://doi.org/10.1109/ICECA55336.2022.10009423

[18] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD *International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[19] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

[20] Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019). Logistic Regression Model Optimization and Case Analysis. *Proceedings of the IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT 2019)*, 135–139. https://doi.org/10.1109/ICCSNT47585.2019.8962457

[21] Makalic, E., & Schmidt, D. F. (2010). Review of Modern Logistic Regression Methods with Application to Small and Medium Sample Size Problems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6464 LNAI, 213–222. https://doi.org/10.1007/978-3-642-17432-2_22

[22] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science, 2*(3), 1–21. https://doi.org/10.1007/S42979-021-00592-X/FIGURES/11

[23] Zhang, H. (2004). *The Optimality of Naive Bayes*. Retrieved from www.aaai.org

[24] Liu, H., & Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing, 2*(4), 357–386. https://doi.org/10.1007/S41066-017-0049-2/FIGURES/11

[25] Arar, Ö. F., & Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing Journal, 59*, 197–209. https://doi.org/10.1016/j.asoc.2017.05.043

[26] Liu, S., Liu, Y., Shan, L., Wang, Q., Sun, Y., & He, L. (2024). Hybrid Conditional Kernel SVM for Wire Rope Defect Recognition. *IEEE Transactions on Industrial Informatics*, *20*(4), 6234–6244. https://doi.org/10.1109/TII.2023.3344135

[27] Basak, S., & Yesmin, S. (2021). Students' attitudes towards library overdue fines in an academic library: A study in a private university setting in Bangladesh. *IFLA Journal, 47*(1), 91–100. https://doi.org/10.1177/0340035220944948

[28] Pradita, E. (2019). *Pengaruh Keterlambatan Pengembalian Buku Terhadap Peminjaman Koleksi Di Dinas Perpustakaan Provinsi Sumatera Selatan* (Doctoral dissertation, Program Studi Ilmu Perpustakaan).

[29] Argha, H. (2024). Analisis Penyebab Keterlambatan Dalam Pengembalian Koleksi Di Layanan Sirkulasi Perpustakaan Universitas Wahid Hasyim. *Literasiana*, *2*(01).