

Prediction of Book Borrowing Patterns Using LSTM-XGBoost for Collection Planning at UINSU Library

Andri Ismail Sitepu*, Zulham Sitorus, Muhammad Iqbal

Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ¹*adrimstp@gmail.com, ²zulhamsitorus@dosen.pancabudi.ac.id, ³muhammadiqbal@dosen.pancabudi.ac.id

Email Penulis Korespondensi: adrimstp@gmail.com*

Submitted: 28/08/2029; Accepted: 27/09/2026; Published: 30/09/2026

Abstract– University libraries face challenges in evidence-based collection planning due to the underutilization of borrowing circulation data. This study proposes a hybrid prediction model that combines Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) to predict book borrowing patterns at the Library of Universitas Islam Negeri Sumatera Utara (UINSU). The proposed approach employs XGBoost to learn relationships among borrowing-related features and generate prediction outputs, which are subsequently used as additional input features for the LSTM model to capture temporal borrowing patterns. The dataset consists of 3,961 physical book borrowing transactions collected between 2024 and 2025. After preprocessing and feature engineering, the data were chronologically sorted and divided into 80% training data and 20% testing data using a time-based split to preserve the temporal sequence. Model performance was evaluated on the testing data using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Experimental results demonstrate that the proposed hybrid XGBoost–LSTM model achieved an RMSE of 0.3879 and an MAE of 0.1868, indicating good predictive performance with relatively low error rates. The comparison between actual and predicted borrowing values further shows that the proposed model is capable of capturing borrowing trends effectively. The developed model can serve as the foundation for a Decision Support System (DSS) to support data-driven collection planning and acquisition strategies in academic libraries.

Keywords: book borrowing prediction; LSTM; XGBoost; hybrid model; library collection planning; time series

1. INTRODUCTION

Libraries function as scientific and cultural institutions that provide various collections, reading areas, study spaces, as well as virtual knowledge centers and information production hubs. Libraries play an important role in education and literacy development, while also being regarded as social, cultural, and community assets where various academic, social, and cultural activities take place [1] [2]. University libraries are strategic units serving as driving forces for academic activities, research, and knowledge development. One indicator of a library's success is the high number of book loans and frequency of user visits. Library managers need to understand the extent to which users are aware of the library's strategic role in supporting the educational process [3] [4], [5], [6].

According to Rose-wiles academic libraries face large and ever-increasing volumes of circulation data [7]. Circulation data is frequently used to analyze collection characteristics or identify book usage status. Collection evaluation through circulation statistical analysis is essential for understanding the current status of collections and predicting usage patterns to improve quality.

The main challenge in modern collection management is no longer solely related to the acquisition process, but rather how libraries can project collection needs more precisely based on real and measurable user behavior [1]. Traditional prediction methods such as linear regression are not always capable of capturing the complexity of borrowing patterns that are non-linear and influenced by various factors such as changes in academic needs, course dynamics, literature trends, and fluctuating user preferences [8]

The use of a hybrid model combining Long Short Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) represents a highly promising approach. LSTM, as a variant of Recurrent Neural Networks, is effective at capturing temporal dynamics including seasonal trends and historical patterns in time series data [9], [10], [11] [12]. Meanwhile, XGBoost offers strong predictive modeling capabilities through a decision tree-based ensemble approach, which is robust in handling complex numerical data and more resistant to overfitting [13] [14].

This study aims to formulate a prediction model for book borrowing patterns at the UINSU Library as a basis for decision-making in collection planning. Currently, the UINSU Library possesses large amounts of borrowing data from years of daily circulation activities that have not yet been optimally utilized as a source for predictive analysis. This research addresses this gap by developing a hybrid LSTM–XGBoost model to predict book borrowing patterns using historical circulation data from the UIN Sumatera Utara (UINSU) Library. The proposed model is expected to provide accurate forecasting results that can assist librarians in optimizing collection planning, improving resource allocation, and supporting data-driven decision-making for sustainable library management.

2. RESEARCH METHODOLOGY

2.1 Research Approach and Type

This study uses an experimental approach through the process of designing, training, and testing prediction models based on machine learning and deep learning. The research type is quantitative, aimed at analyzing and predicting book borrowing patterns at the UINSU Library based on numerical data derived from the transformation of book borrowing transactions.

2.2 Dataset and Research Variables

The object of this study is the historical data of physical book borrowing at the UINSU Library. The data used is secondary data obtained from the library information system with the main attributes: member ID, member name, book title, loan date, and return date. The total dataset consists of 3,961 borrowing transactions for the period 2024–2025.

Table 1. Sample of Book Borrowing Transaction Data.

No	Member ID	Member Name	Book Title	Loan Date	Return Date
1	0301223089	Indah Permata Sari Sitorus	Metodologi penelitian kuantitatif...	2024-01-02	2024-01-09
2	0301213074	Rizky Nurfadillah	Pembelajaran Berbasis Inkuiri	2024-01-02	2024-01-09
3	0301212195	Fatimah Zahra Chaniago	Strategi Pembelajaran Inovatif Kontemporer	2024-01-02	2024-01-09
4	0206202080	M. Bahri Al Soddik Saputra	Si Putih	2024-01-02	2024-01-09
5	0441233004	Saidatun Nisa	Terjemah Tafsir Al Maraghi 5	2024-01-03	2024-01-17
...	...	...	...	...	...
3961	0401223051	Akhmed Syarifatul Raflus	12 Tokoh Pengubah Dunia	2025-10-02	2025-10-09

Research variables are grouped into non-temporal variables (member identity, book characteristics, borrowing frequency) and temporal variables (daily, weekly, monthly borrowing patterns). The input variables consist of encoded member identity (Anggota_enc), encoded book title (Judul_enc), borrowing duration, borrowing frequency per member, borrowing day, borrowing month, and the prediction output generated by XGBoost. The target variable is the borrowing status (0 or 1), which represents the observed borrowing outcome for each transaction.

2.3. Research Stages

This study follows four main stages:

- Data Collection:** collecting book borrowing datasets from the UINSU library system, including transaction data, member data, and collection metadata.
- Data Pre-processing:** Converting loan and return dates into datetime format, calculating borrowing duration, encoding categorical variables using Label Encoding, sorting transactions chronologically, and exporting the processed dataset.
- Feature Engineering:** Generating borrowing frequency for each member, borrowing frequency for each book, and extracting temporal features including borrowing day and borrowing month.
- Modeling:** Splitting the dataset chronologically into 80% training data and 20% testing data, training the XGBoost model, generating XGBoost prediction outputs, and using these outputs as additional input features for the LSTM model.
- Model Evaluation:** Evaluating prediction performance on the testing dataset using RMSE and MAE.

2.4. Hybrid Stacking Modeling Method

The hybrid stacking approach is implemented with the following workflow:

2.4.1. Non-Temporal Feature Extraction with XGBoost

The XGBoost model processes encoded member identity, encoded book title, and member borrowing frequency to learn non-temporal borrowing patterns. XGBoost is chosen for its computational efficiency and consistent performance in handling large datasets using a gradient boosted decision trees approach.

Target value predictions are generated from the sum of the outputs of K decision trees using the equation:

$$\hat{y}^{(t+h)} = \sum f_k(x_i), f_k \in F (k = 1 \text{ to } K)$$

XGBoost optimizes an objective function consisting of a loss function (evaluating prediction error) and a regularization term (controlling model complexity to reduce the risk of overfitting).

2.4.2. Temporal Pattern Modeling by LSTM

LSTM is a type of RNN designed to address the problem of exploding and vanishing gradients [15]. LSTM has a memory cell and gate mechanisms that allow the model to retain important long-term information while discarding irrelevant information. The LSTM mechanism consists of:

- Forget Gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Input Gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$, $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
- Cell State Update: $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$
- Output Gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$, $h_t = o_t \cdot \tanh(C_t)$

2.4.3. Final Prediction Integration (Stacking)

The prediction generated by XGBoost is not used as the final prediction. Instead, it is incorporated as an additional feature (`xgb_output`) together with temporal attributes (borrowing duration, borrowing day, and borrowing month). These combined features are then used as input for the LSTM model, which produces the final borrowing prediction. This stacking strategy enables XGBoost to capture non-temporal relationships while allowing LSTM to learn temporal dependencies from historical borrowing behavior.

3.5 Evaluation Metrics

Model evaluation Ajder (2025) is carried out using two main metrics commonly used in time series research [12]. These metrics were calculated using the testing dataset obtained from an 80:20 chronological train and test split in order to assess the generalization capability of the proposed model on previously unseen data.

Table 2. Prediction Model Evaluation Metrics

Metric	Description	Hybrid Model Value
RMSE	Root Mean Square Error measures the average prediction error while assigning greater penalties to larger errors. This metric is useful for evaluating model stability and sensitivity to significant deviations.	0.3879
MAE	Mean Absolute Error measures average absolute difference between the actual and the predicted values. MAE is easy to interpret and provides a direct indication of prediction accuracy.	0.1868

The proposed hybrid XGBoost–LSTM model achieved an RMSE value of 0.3879 on the testing dataset. RMSE evaluates the magnitude of prediction errors while assigning a greater penalty to larger deviations between the predicted and the actual values. The obtained result indicates that the model was able to maintain relatively stable predictive performance, even when encountering variations in borrowing behavior across different periods.

In addition, the model achieved an MAE value of 0.1868, indicating that the average absolute difference between the predicted and actual borrowing status values was approximately 0.19 units. Since MAE directly reflects the average prediction error, the relatively low value suggests that the proposed model was capable of generating predictions that closely matched the observed borrowing outcomes for most transactions.

The evaluation was performed using an 80:20 chronological train–test split, where the earliest 80% of borrowing records were used for training and the remaining 20% were reserved for testing. This evaluation strategy provides a more realistic assessment of the model's generalization capability because the testing data were not used during the training process.

Overall, the obtained RMSE and MAE values demonstrate that the hybrid XGBoost–LSTM model can effectively capture borrowing patterns from historical circulation data. These findings indicate that the proposed approach has the potential to support evidence-based collection planning and decision-making processes in academic libraries by providing reliable predictions of future borrowing behavior.

3. RESULT AND DISCUSSION

3.1. Dataset Description

The dataset used in this study comes from the historical book borrowing data of the UINSU Library. The raw data contains key information including borrowing dates, return dates, member identities, book titles, and borrowing status. Based on initial exploration, the dataset has a consistent and complete structure (3,961 entries, indexed 0 to 3960) with a total of 7 columns.

The loan date and return date attributes were converted to datetime format to enable calculation of the borrowing duration. Duration is calculated as the difference in days between the return date and the loan date. Categorical attributes such as member IDs and book titles were transformed into numerical form using label encoding so they could be processed by machine learning algorithms.

Prior to model development, the borrowing transactions were sorted chronologically according to the loan date. This chronological ordering preserves the temporal dependency of the borrowing records and provides a suitable foundation for subsequent time-based data partitioning and sequence learning using the hybrid XGBoost–LSTM model.

The dataset characteristics observed in this section demonstrate that the borrowing records from the UINSU Library provide a sufficiently rich and structured foundation for predictive modeling. The completeness of 3,961 borrowing transactions minimizes the need for extensive data cleaning while preserving the integrity of the original circulation records. The conversion of date attributes into numerical duration values and the label encoding of categorical fields are standard preprocessing steps that align with best practices in machine learning pipelines for library circulation data [1]. Furthermore, the chronological ordering of transactions ensures that temporal dependencies are maintained throughout the modeling process, supporting realistic prediction scenarios and reducing the risk of information leakage during model evaluation.

3.2. Feature Engineering Results

The feature engineering stage was carried out to enrich the information contained in the borrowing data. Several new features were created:

- a. Member borrowing frequency (freq_member): describes the intensity of borrowing made by each library member.
- b. Book borrowing frequency (freq_book): indicates the popularity level of a book title based on the number of borrowings.
- c. Temporal features (day and month): extracted from the borrowing date to capture seasonal and periodic patterns in book borrowing activity.

These engineered features were utilized during the hybrid modeling process. Encoded member identity, encoded book title, and member borrowing frequency were first processed by the XGBoost model to generate predictive outputs. The resulting XGBoost predictions were then combined with temporal features, including borrowing duration, borrowing day, and borrowing month, as input variables for the LSTM model.

The feature engineering process significantly enriched the information contained in the original borrowing records. Member borrowing frequency represents individual borrowing behavior, while temporal attributes such as borrowing day and borrowing month enable the model to identify periodic borrowing trends. Instead of directly relying solely on the engineered variables, the proposed hybrid architecture first utilizes XGBoost to learn non-temporal relationships and subsequently incorporates its prediction output as an additional feature for the LSTM model. This sequential integration allows the hybrid model to capture both static borrowing characteristics and temporal borrowing dynamics more effectively than using either approach independently.

3.3. Hybrid XGBoost-LSTM Model Implementation Results

The prediction model was developed using a hybrid XGBoost–LSTM architecture. Prior to model training, all borrowing transactions were chronologically sorted according to the borrowing date to preserve the temporal characteristics of the dataset. The dataset, consisting of 3,961 borrowing transactions, was then divided into two subsets using a chronological hold-out strategy. Approximately 80% of the data (3,168 records) were allocated as the training set, while the remaining 20% (793 records) were reserved as the testing set. This chronological partition was selected to simulate a real forecasting scenario in which historical borrowing records are used to predict future borrowing behavior, thereby preventing information leakage that may occur when random sampling is applied to time-series data.

The XGBoost model was first trained using non-temporal features, namely encoded member ID, encoded book title, and member borrowing frequency. The prediction generated by XGBoost (xgb_output) was subsequently integrated as an additional feature for the LSTM model together with borrowing duration, borrowing day, and borrowing month. This stacking strategy enables the LSTM network to learn temporal borrowing patterns while simultaneously utilizing the non-linear relationships captured by XGBoost.

The LSTM training process was conducted for 20 epochs using the training dataset. The training loss consistently decreased from 0.2010 in the first epoch to 0.0173 in the twentieth epoch, indicating that the model progressively learned the underlying borrowing patterns throughout the optimization process.

Epoch	1/20	99/99	1s	1ms/step	loss: 0.2010
Epoch	2/20	99/99	0s	1ms/step	loss: 0.0336
Epoch	3/20	99/99	0s	1ms/step	loss: 0.0319
Epoch	4/20	99/99	0s	1ms/step	loss: 0.0314
Epoch	5/20	99/99	0s	1ms/step	loss: 0.0310
Epoch	6/20	99/99	0s	1ms/step	loss: 0.0306
Epoch	7/20	99/99	0s	1ms/step	loss: 0.0302
Epoch	8/20	99/99	0s	1ms/step	loss: 0.0297
Epoch	9/20	99/99	0s	1ms/step	loss: 0.0292
Epoch	10/20	99/99	0s	1ms/step	loss: 0.0286
Epoch	11/20	99/99	0s	1ms/step	loss: 0.0277
Epoch	12/20	99/99	0s	1ms/step	loss: 0.0276
Epoch	13/20	99/99	0s	1ms/step	loss: 0.0260
Epoch	14/20	99/99	0s	1ms/step	loss: 0.0253
Epoch	15/20	99/99	0s	1ms/step	loss: 0.0243
Epoch	16/20	99/99	0s	1ms/step	loss: 0.0227
Epoch	17/20	99/99	0s	1ms/step	loss: 0.0209
Epoch	18/20	99/99	0s	1ms/step	loss: 0.0199
Epoch	19/20	99/99	0s	1ms/step	loss: 0.0179
Epoch	20/20	99/99	0s	1ms/step	loss: 0.0173

Figure 1. LSTM Training Loss Graph over 20 Epochs

The training results demonstrate a substantial improvement during the early stage of learning, where the loss rapidly decreased from 0.2010 in the first epoch to 0.0336 in the second epoch. This significant reduction suggests that the proposed hybrid model quickly captured the dominant relationships within the borrowing data. After this initial improvement, the loss continued to decrease gradually across subsequent epochs, reaching 0.0173 at the final epoch.

The smooth and monotonic decline of the training loss indicates that the optimization process converged successfully without exhibiting unstable oscillations or divergence. Unlike the previous implementation, no increase in the loss value was observed during the final epochs, suggesting that the selected training configuration remained stable throughout the learning process. Although the training loss demonstrates successful model learning, predictive performance was further evaluated using the unseen testing dataset through Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to assess the model's generalization capability.

The implementation results confirm the effectiveness of the proposed hybrid XGBoost–LSTM architecture in integrating both non-temporal and temporal borrowing information. XGBoost effectively captures nonlinear relationships among encoded member identity, encoded book title, and borrowing frequency, while its prediction output is utilized as an additional feature for the LSTM model. The consistently decreasing training loss across all 20 epochs demonstrates stable optimization and successful convergence, indicating that the proposed hybrid architecture provides a reliable computational framework for predicting book borrowing patterns.

3.4. Prediction Model Performance Analysis

Model performance in this study was evaluated using the testing dataset obtained from the chronological 80:20 train and test split. Two evaluation metrics were employed, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Based on the testing results, the proposed hybrid XGBoost–LSTM model achieved an RMSE of 0.3879 and an MAE of 0.1868.

The obtained RMSE indicates that the proposed model maintained a relatively low prediction error while assigning greater penalties to larger deviations between predicted and actual borrowing status. Meanwhile, the MAE value demonstrates that the average absolute prediction error remained relatively small, indicating consistent predictive performance across the testing dataset.

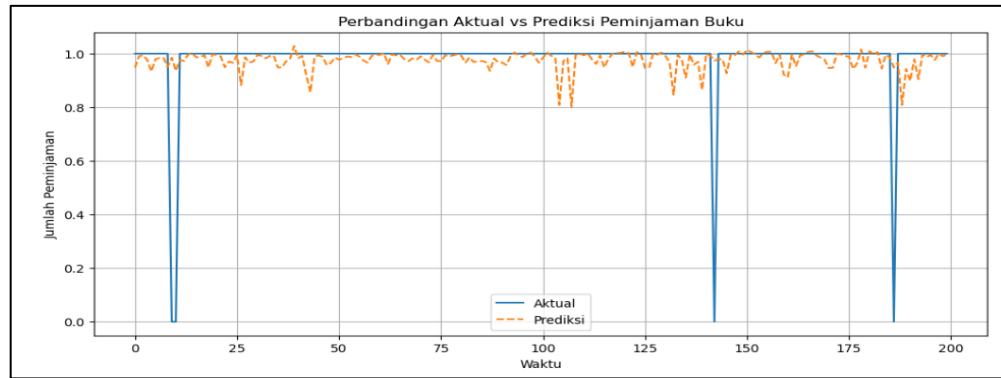


Figure 2. Comparison Graph of Actual Data and Hybrid XGBoost-LSTM Model Predictions

Figure 2 illustrates the comparison between the actual borrowing status and the predicted values on the testing dataset. The prediction curve generally follows the overall pattern of the actual data, although several deviations are still observed at certain points. These discrepancies are expected because borrowing behavior may also be influenced by external factors that were not included in the current feature set, such as academic schedules, curriculum changes, promotional activities, and seasonal library usage patterns. Overall, the evaluation result demonstrate that the proposed hybrid XGBoost - LSTM model achieved satisfactory predictive performance on previously unseen borrowing transactions. Although the prediction errors are higher than those observed during model training, this difference is expected because the testing dataset contains data that were not used during the learning process. The combination of an RMSE of 0.3879 and an MAE of 0.1868 indicates that the model is capable of generalizing reasonably well while maintaining stable predictive performance. These findings suggest that the proposed hybrid model has strong potential to support evidence-based collection planning and decision-making within the UINSU Library.. The deviations observed at specific time points are attributable to external demand shocks that are absent from the feature set, such as the influence of library promotional activities or changes in faculty reading requirements, which are inherently difficult to capture without additional contextual data [12].

3.5. Hybrid Model Contribution Analysis

The integration of XGBoost and LSTM provides several advantages over using a single prediction model. XGBoost effectively captures complex non-linear relationships among static borrowing features, such as encoded member identity, encoded book title, and borrowing frequency, while LSTM is designed to model temporal dependencies from sequential borrowing records. By combining the strengths of both models through a stacking mechanism, the proposed hybrid architecture is able to exploit both static and temporal information simultaneously.

In this study, the dataset was divided chronologically into training and testing subsets, ensuring that future borrowing records were never used during model training. This experimental design provides a more realistic evaluation of forecasting performance because predictions are generated from previously unseen data. The obtained RMSE of 0.3879 and MAE of 0.1868 indicate that the hybrid model maintains relatively low prediction errors despite the more challenging evaluation setting.

This hybrid approach aligns with the research of Ajder et al. (2025), which states that a combination of tree-based models and deep learning can improve prediction stability on data that is dynamic and fluctuating [12]. Research by Salam et al. (2025) also confirms that the hybrid XGBoost-LSTM model is consistently more accurate than single models or conventional methods such as linear regression and random forest [8].

The results of this study are consistent with the findings of Liu (2025), which showed that the hybrid LSTM-XGBoost model has proven capable of capturing non-linear interactions as well as temporal structures, especially when combined with hyperparameter optimization [9]. This confirms that integrating both models provides more stable, accurate, and adaptive performance compared to a single model.

The contribution analysis reinforces the theoretical advantage of ensemble and hybrid architectures over single-model approaches in the context of dynamic time series forecasting. The complementary nature of XGBoost and LSTM means that each component addresses the other's weaknesses: XGBoost compensates for LSTM's limited ability to leverage categorical metadata, while LSTM fills the gap left by XGBoost in modeling temporal order dependencies. This division of labor within the stacking framework is consistent with the broader machine learning principle that model diversity, when properly orchestrated, yields superior generalization performance [13]. The alignment of this study's findings with provides external validity and suggests that the hybrid LSTM-XGBoost paradigm is emerging as a robust standard for sequential prediction tasks characterized by mixed feature types. The additional value brought by hyperparameter tuning [16].

Overall, the proposed hybrid architecture demonstrates the complementary strengths of XGBoost and LSTM. XGBoost contributes efficient learning of complex relationships from structured borrowing attributes, whereas LSTM captures sequential borrowing behavior over time. The combination of these complementary

capabilities enables the hybrid model to produce more stable and reliable predictions than either model could achieve independently, making it suitable for forecasting book borrowing patterns in academic libraries.

3.6. Practical Implications

The results of this study have several practical implications for academic library management. First, prediction results can support collection development by identifying books with potentially high borrowing demand, allowing librarians to prioritize procurement and optimize collection availability. Second, forecasting borrowing patterns can assist library managers in planning circulation services and resource allocation during periods of increased borrowing activity, such as the beginning of academic semesters.

Furthermore, the proposed hybrid XGBoost–LSTM model can be integrated into the existing library information system as a Decision Support System (DSS) module. The predicted borrowing values can provide librarians with quantitative evidence to support collection planning, circulation management, and long-term policy formulation. By utilizing historical borrowing data through predictive analytics, library decision-making can gradually shift from experience-based practices toward evidence-based management.

Although the proposed model demonstrates satisfactory predictive performance, the current study relies primarily on historical borrowing transactions and basic borrowing characteristics. Future research may further improve prediction accuracy by incorporating additional explanatory variables, such as academic calendars, course enrollment, faculty reading lists, seasonal events, or promotional activities, which may influence borrowing behavior but were not included in the present dataset.

The practical implications outlined in this section underscore the transformative potential of data-driven forecasting models in the operational management of academic libraries. The ability to anticipate surges in borrowing demand during specific academic periods enables library administrators to proactively reallocate resources, procure high-demand titles in advance, and schedule staffing accordingly all of which directly contribute to improved service quality and cost efficiency. The proposed integration of the hybrid model as a DSS module represents a scalable and sustainable pathway for embedding predictive analytics into existing library information systems without requiring a complete overhaul of institutional infrastructure [1]. Furthermore, the shift toward evidence-based collection management aligns with international trends in academic library governance, where data transparency and measurable outcomes are increasingly demanded by institutional stakeholders. The findings of this study therefore not only advance the technical literature on hybrid predictive modeling but also provide actionable insights that library practitioners can adopt to modernize collection planning workflows.

3.7. Research Limitations

Despite the promising results obtained in this study, several limitations should be acknowledged. First, the dataset was collected from a single academic library, namely the UINSU Library, which may limit the generalizability of the proposed model to other university libraries with different borrowing characteristics and user behavior. Future studies should validate the proposed approach using datasets from multiple institutions to evaluate its broader applicability. Second, the prediction model relies primarily on historical borrowing transactions and basic borrowing attributes. External factors that may influence borrowing behavior, such as academic calendars, examination periods, library promotional activities, curriculum changes, or special campus events, were not incorporated into the current feature set. Including these contextual variables may further improve prediction accuracy and better explain fluctuations in borrowing demand. Third, although this study adopted a chronological 80% training and 20% testing split to evaluate the model on previously unseen data, the evaluation was limited to a single hold-out testing scenario. More comprehensive validation techniques, such as walk-forward validation or time-series cross-validation, could provide a more robust assessment of the model's generalization capability across different time periods.

Finally, the current implementation employed a relatively simple LSTM architecture consisting of a single LSTM layer with 64 hidden units. Future studies may investigate deeper architectures, hyperparameter optimization, or alternative hybrid approaches to further enhance forecasting performance. Overall, these limitations provide opportunities for future research to improve both the robustness and practical applicability of hybrid machine learning models for academic library demand forecasting.

4. CONCLUSION

This study proposed a hybrid XGBoost–LSTM model to predict book borrowing patterns using historical circulation data from the UINSU Library. The proposed approach combines the strengths of XGBoost in modeling complex non-linear relationships among structured borrowing attributes with the capability of LSTM to capture temporal dependencies in sequential borrowing data. Prior to model development, the dataset was preprocessed, relevant features were engineered, and the data were chronologically divided into 80% training data and 20% testing data to simulate a realistic forecasting scenario. The experimental results demonstrate that the proposed hybrid model achieved an RMSE of 0.3879 and an MAE of 0.1868 on the testing dataset. These results indicate that the model is capable of producing reasonably accurate predictions while maintaining relatively low prediction

errors. Furthermore, the comparison between actual and predicted borrowing values shows that the model successfully captures the overall borrowing trend despite slight deviations during periods influenced by external factors. From a practical perspective, the proposed hybrid model can support evidence-based collection planning by helping librarians identify books with potentially high borrowing demand, optimize procurement decisions, and improve circulation management. The model also has the potential to be integrated into the library information system as a Decision Support System (DSS), enabling data-driven decision-making for sustainable academic library management. Future research is encouraged to enhance the proposed model by incorporating additional contextual variables, including academic calendars, examination schedules, library promotional activities, and other external factors that influence borrowing behavior. Moreover, more comprehensive validation techniques, such as walk-forward validation or time-series cross-validation, should be employed to further evaluate the model's generalization performance. Future studies may also investigate hyperparameter optimization and alternative hybrid architectures to improve forecasting accuracy and increase the practical applicability of predictive analytics in academic libraries.

REFERENCES

- [1] J. Saptari, "Implementasi Perpustakaan Cerdas," *Media Inf.*, vol. 32, no. 1, pp. 38–48, 2023, doi: <https://doi.org/10.22146/mi.v32i1.6986>.
- [2] E. R. Pebriansyah and T. A. Fitri, "PREDICTION OF LIBRARY BOOK BORROWING PATTERNS USING THE," vol. 7, no. 4, 2025.
- [3] M. R. R. Putra and F. Muttakin, "Prediksi Jumlah Pengunjung Perpustakaan Daerah Kabupaten Batang dengan Menggunakan Metode Fuzzy Time Series Chen-Hsu," *J. Math. UNP*, vol. 8, no. 1, pp. 110–119, 2023, [Online]. Available: <http://ejournal.unp.ac.id/students/index.php/mat/%0A1>.
- [4] A. Kumar and Y. T. A., "Proposing a Conceptual Framework for Social Listening in Libraries : A Potential Game Changer to Engage Gen Z and Centennial Users," *DE GRUYTER*, vol. 9, no. 20240012, pp. 1–14, 2025, doi: <https://doi.org/10.1515/opis-2024-0012>.
- [5] Y. Prasetya and R. Rahmi, "Pilot Study on User Experience Analysis of Universitas Indonesia Library Website," *Tik Ilmeu J. Ilmu Perpust. dan Inf.*, vol. 7, no. 1, p. 105, 2023, doi: 10.29240/tik.v7i1.6453.
- [6] S. Endarti, "Perpustakaan sebagai Tempat Rekreasi Informasi," *J. Abdi pustaka*, vol. 2, no. 1, pp. 23–28, 2022, doi: <https://doi.org/10.24821/jap.v2i1.6990>.
- [7] S. Khademizadeh, Z. Nematollahi, and F. Danesh, "Library and Information Science Research Analysis of book circulation data and a book recommendation system in academic libraries using data mining techniques," *Libr. Inf. Sci. Res.*, vol. 44, no. 4, p. 101191, 2022, doi: 10.1016/j.lisr.2022.101191.
- [8] Z. A. Salam, A. Saiku, and J. L. Buliali, "Prediksi curah hujan dengan model hybrid xgboost- lstm berdasarkan data pengamatan permukaan," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 4, pp. 3731–3739, 2025, [Online]. Available: <https://jurnal.stkipggritlungagung.ac.id/index.php/jupi>
- [9] X. Liu, "Machine Learning with Applications Enhanced prediction of karst spring discharge using a hybrid LSTM-XGBoost model optimized with grid search," *Mach. Learn. with Appl.*, vol. 22, no. September, p. 100740, 2025, doi: 10.1016/j.mlwa.2025.100740.
- [10] K. Ito, H. Iima, and Y. Kitamura, "LSTM forecasting foreign exchange rates using limit order book," *Financ. Res. Lett.*, vol. 47, no. PA, p. 102517, 2022, doi: 10.1016/j.frl.2021.102517.
- [11] S. Kasus, P. Susu, and S. Harian, "Perbandingan Kinerja LSTM dan Prophet untuk," vol. 9, no. 3, pp. 428–435, 2023.
- [12] A. Ajder, H. A. A. Hamza, and R. Ayaz, "Wavelet-Enhanced Hybrid LSTM-XGBoost Model for Predicting Time Series Containing Unpredictable Events," *IEEE Access*, vol. 13, no. January, pp. 58671–58679, 2025, doi: 10.1109/ACCESS.2025.3556540.
- [13] M. Surono *et al.*, "Hybrid XGBoost-SVM Model untuk Sistem Pendukung Keputusan dalam Prediksi Penyakit Diabetes," vol. 4, no. 3, pp. 443–454, 2025, doi: 10.55123/insologi.v4i3.5410.
- [14] A. F. Alkayes and T. Sugihartono, "Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025 Abstrak Comparison of XGBoost and LSTM Algorithms in Tesla Stock Price Prediction Using 2025 Data Abstract," vol. 5, no. 6, pp. 1563–1573, 2025.
- [15] Z. Sitorus, M. Iqbal, D. Nasution, and R. F. Wijaya, "Penerapan Deep Learning dan Analisis Sentimen terhadap Gap Kompetensi Lulusan Lembaga Pendidikan dan Pelatihan Vokasi terhadap Dunia Kerja dengan Metode Long Short-Term Memory (LSTM)," vol. 6, no. 2, pp. 0–11, 2025, doi: 10.47065/bit.v5i2.2029.
- [16] X. Deng *et al.*, "Bagging – XGBoost algorithm based extreme weather identification and short-term load forecasting model," *Energy Reports*, vol. 8, pp. 8661–8674, 2022, doi: 10.1016/j.egyr.2022.06.072.