Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Analyzing the Impact of Library Resources and Activities on Visits, Circulation, and Attendance Using Machine Learning Methods

Haresh Dhanji Chande[1], Sachinkumar Harshadbhai Makwana[2]

[1,2]Government Engineering College, Patan, Gujarat, India

ARTICLE INFO

Received: 03 July 2024

Accepted: 26 July 2024

ABSTRACT

This study investigates the impact and predictive modeling of public library usage metrics, focusing on VISITS, TOTCIR (Total Circulation), and TOTATTEN (Total Attendance). Using a real-world dataset, we applied multiple machine learning algorithms including Linear Regression, Random Forest, AdaBoost, XGBoost, and Support Vector Regression (SVR). The models were evaluated using both training and testing sets, with performance measured through R^2 and Mean Squared Error (MSE). The results demonstrate that Linear Regression and Random Forest consistently achieved high predictive accuracy, with R^2 values above 0.90 across most targets, while AdaBoost and SVR performed poorly, indicating limited suitability for this dataset. XGBoost achieved near-perfect training accuracy but showed a degree of overfitting with slightly lower test R^2. These findings provide insights into which algorithms are most effective for predicting library usage metrics and highlight potential avenues for applying predictive modeling in policy and decision-making for resource allocation.

Keywords: Public libraries, Machine learning, Predictive modelling, Library visits, Library attendance

INTRODUCTION

Libraries remain vital institutions for promoting literacy, education, and community engagement (Shapiro, 2016; Kinney, 2010). To evaluate their impact, performance indicators such as the number of visits, circulation of resources, and total attendance are often used (Gilpin, Karger, & Nencka, 2021; Irwin & Silk, 2019). With the increasing availability of structured datasets, predictive modelling offers the opportunity to forecast these indicators and optimize resource planning (Connell, Wallis, & Comeaux, 2021).

Libraries are cornerstones of communities, serving as vital hubs for literacy, education, and social interaction (Atuase & Maluleka, 2022; Nzomo, 2023). To gauge their effectiveness and inform strategic decision-making, it is essential to analyze key performance indicators such as the number of total visits (VISITS), the total circulation of materials (TOTCIR), and total attendance at library programs (TOTATTEN). These metrics provide a quantifiable measure of a library's reach and impact, helping administrators and stakeholders understand how resources are being utilized (Scoulas & De Groote, 2019; Carey, Pathak, & Johnson, 2023).

Historically, the evaluation of library performance has often relied on traditional methods like descriptive statistics and trend analysis (Mabe, 2022; Joo & Cahill, 2019). While these methods offer a snapshot of past performance, they lack the predictive power needed for proactive resource management. The increasing availability of large, structured datasets from library management systems presents an opportunity to move beyond retrospective analysis and embrace advanced predictive modelling (Witherspoon & Taber, 2021).

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Despite significant advancements in computational modelling across various fields, the discipline of library science has not fully integrated predictive analytics into its core management practices. This gap means that decisions regarding resource allocation, staffing, and program development are often based on historical data rather than on robust, data-driven forecasts.

This research aims to bridge this gap by systematically applying a suite of machine learning models to the challenge of predicting key library performance metrics. Specifically, this study focuses on:

Applying a range of machine learning models: We implement and test five distinct models—Linear Regression, Random Forest, AdaBoost, XGBoost, and Support Vector Regression—to assess their suitability for predicting VISITS, TOTCIR, and TOTATTEN.

Evaluating model performance: The efficacy of each model is rigorously evaluated using two standard metrics: the Coefficient of Determination (R^2) to measure the proportion of variance explained by the model, and the Mean Squared Error (MSE) to quantify the average prediction error.

Visualizing predictive accuracy: Actual versus predicted plots for both training and testing datasets provide a clear, intuitive understanding of model performance, graphically demonstrating how well each model's predictions align with the actual values.

By leveraging these sophisticated analytical techniques, this research seeks to provide a valuable framework for evidence-based decision-making, enabling libraries to optimize resource planning and better serve their communities.

BACKGROUND

The study by Mabe, M. R. (2022) investigates whether public library use increased during the Great Recession (2006–2011). Analysis of 541 Library Journal Star libraries showed that visits and circulation generally rose during the downturn, supporting the axiom that economic hardship drives higher library use. While changes varied across budget categories, no declines were observed, underscoring libraries' role as critical community resources in times of crisis.

Gilpin, Karger, and Nencka (2021) use a compelling event study design applied to the Public Libraries Survey data to demonstrate that large-capital investments in public libraries—such as new buildings or renovations—boost library usage, including visits, children's attendance, and circulation, by approximately 18–21%, and further contribute to modest improvements in children's reading test scores. Their rigorous empirical strategy provides valuable causal evidence on the social returns of public library capital investment (Gilpin et al., 2021).

hapiro (2016) critically examines the long-term decline in academic library usage—such as reduced circulation, reference interactions, and in-house material engagement—and proposes strategic realignment by reimagining the academic library as an active hub for creativity, discovery, and collaboration. Drawing on insights from successful public library initiatives and practical examples (e.g., Montclair State University), Shapiro advocates for increased community programming, technology integration, and spatial innovation to renew library relevance and engagement.

Kinney (2010) provides a comprehensive investigation into how public libraries help bridge the digital divide by offering widespread internet access. Using data spanning U.S. public libraries, the study finds that the availability of public internet terminals significantly increases library visits and reference service usage, although it does not substantially affect circulation rates. Kinney's work underscores the critical role public libraries play in digital inclusion efforts by providing essential access and support to underserved communities.

Atuase and Maluleka (2022) explore how active marketing of library resources and services influences usage patterns among distance-learning students in Ghana. Their mixed-method study—combining survey data from 1,170 students and interviews with librarians—revealed that targeted marketing significantly increases awareness and library resource usage, while information literacy alone does not predict utilization beyond basic familiarity. The findings underscore that effective marketing strategies are critical for enhancing engagement among remote learners.

Joo & Cahill (2019) investigate the relationships among expenditures, library resources, and children's and young adults' library use, using data from the IMLS National Public Library Survey. Their correlation analyses reveal that variables such as staffing levels, print collections, and material expenditures are significantly associated with both children's circulation, which constitutes over one-third of total circulation, and attendance at youth programs, accounting for approximately 70% of program participation. These findings underscore the importance of targeted investment in youth-focused services and infrastructure.

Irwin and Silk (2019) analyze how public libraries must evolve metrics beyond traditional activity counts—like visits and circulation—to better convey social and strategic value to stakeholders. They argue that libraries should adopt meaningful, outcome-focused evaluation metrics aligned with strategic goals to strengthen stakeholder communication, demonstrate societal impact, and guide organizational improvement.

Witherspoon and Taber (2021) investigate factors influencing student attendance at library workshops by analyzing survey data from 161 programs across Canada and the U.S. Their findings highlight that specific targeting of audiences, faculty buy-in, "push" advertising, and purposeful scheduling significantly enhance attendance, while other factors like incentives or location have less impact. This study offers actionable insights for optimizing workshop engagement strategies in academic libraries.

Nzomo (2023) examines how the COVID-19 pandemic affected international students' use of library resources and services at Kent State University. Using survey data, Google Analytics, usage statistics, and librarian interviews, the study found that while interlibrary loans, circulation, acquisitions, and instruction services experienced notable shifts, reference services and general website visits remained relatively stable. The paper underscores the need for enhanced marketing, outreach, and adoption of open-access resources to better support international and long-distance learners.

Connell et al. (2021) examine the effects of the COVID-19 pandemic on academic library usage by comparing metrics—such as interlibrary loan transactions, website and discovery tool pageviews, database use, and patron interactions—across three university libraries during the second half of spring semesters in 2019 and 2020. Using control periods for comparison, their analyses reveal consistent patterns of shifts in library behavior, such as declines in physical access metrics and variable responses in online engagement, underscoring the need for adaptive service models.

Carey et al. (2023) surveyed health professions students and found that those who attended library workshops or visited the library website at least weekly were significantly more aware of LibGuides, whereas variables like academic level or degree program were not significantly related to awareness. Their findings underline the importance of library outreach and digital engagement strategies to boost awareness of research guides among health sciences learners.

Scoulas and De Groote (2019) explore how academic library experiences—including in-person vs. online visits, resource utilization, and space satisfaction—relate to students' GPAs, via mixed-method analysis combining Spearman's correlation and thematic responses. They found that online resource use positively correlates with GPA, while surprisingly, more frequent in-person visits and higher space satisfaction relate negatively; meanwhile, qualitative data highlight that students primarily value libraries as environments for focused study and access to essential scholarly materials (Scoulas & De Groote, 2019). University of Alberta Journals

METHODOLOGY

DATASET

The dataset used in this study is derived from the Public Libraries Survey (PLS), Fiscal Year 2022, conducted by the Institute of Museum and Library Services (IMLS), USA. The PLS provides comprehensive data on public library systems across all states and territories in the United States, serving as a reliable source for policy evaluation, research, and strategic planning in the library domain. The dataset used in this study contains key performance indicators for libraries, including visits, circulation, attendance, and multiple explanatory features

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

related to library resources and activities. Data preprocessing involved handling missing values, dropping non-numeric identifiers, and scaling features for fair model comparison.

Feature Selection

Feature selection is a critical step in machine learning pipelines, as it directly influences model interpretability, computational efficiency, and predictive performance. In this study, only numerical predictors were considered to ensure compatibility with regression-based algorithms and to avoid complexities introduced by categorical encoding. This choice was motivated by the nature of the target variables (VISITS, TOTCIR, and TOTATTEN), which are continuous and highly dependent on numerical predictors such as circulation counts, attendance figures, and other library activity indicators.

To systematically examine the relationships among predictors and targets, a correlation heatmap was constructed. The heatmap provided a visual representation of pairwise correlations, enabling the identification of strong linear associations. Features exhibiting a high positive correlation with the target variables were prioritized, as these were likely to improve the explanatory power of the models. Conversely, predictors with weak or negligible correlations were excluded to reduce noise and avoid redundant information. This process ensured that the models focused on informative and non-redundant predictors.

The correlation analysis also played a role in addressing multicollinearity. Strong intercorrelations among predictors can inflate variance and destabilize regression coefficients. By examining the correlation structure, it was possible to detect and minimize multicollinearity risks. This was particularly important for linear models such as Linear Regression, which are sensitive to correlated inputs. For ensemble-based algorithms (Random Forest, AdaBoost, XGBoost), feature redundancy has less impact but still increases computational burden. Thus, a balance between model robustness and feature efficiency was maintained.

Overall, the feature selection strategy employed in this study was filter-based and grounded in correlation analysis. This approach is computationally lightweight, reproducible, and transparent—qualities that are advantageous for both academic research and practical decision-making contexts. By ensuring that only the most relevant numerical predictors were included, the study achieved greater model stability, improved generalization, and reduced overfitting risks.

Machine Learning Models

A comparative framework of five machine learning algorithms was employed to predict library performance indicators (VISITS, TOTCIR, and TOTATTEN). These algorithms were chosen to represent a balance of linear, ensemble-based, and kernel-based approaches, thereby ensuring robust evaluation across different modeling paradigms.

Linear Regression (LR)

Linear Regression is a fundamental statistical method that models the relationship between independent variables and a continuous dependent variable under the assumption of linearity Su, X. et al. (2012). It estimates coefficients that minimize the sum of squared differences between observed and predicted values. Despite its simplicity, Linear Regression provides strong interpretability, making it a widely used baseline in predictive studies. Its computational efficiency and transparency make it particularly valuable when interpretability and reproducibility are prioritized. However, it is sensitive to outliers and multicollinearity, and it underperforms when relationships are non-linear.

Random Forest Regressor (RF)

Random Forest is an ensemble learning technique based on bagging, where multiple decision trees are constructed on bootstrapped samples of the data. Each split within a tree considers a random subset of features, thereby reducing correlation among trees and improving generalization L. Breiman (2001). Predictions are obtained by averaging outputs across all trees. Random Forest is highly effective at capturing both linear and non-linear dependencies, while also providing insights into feature importance. Its robustness against overfitting and

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

noise makes it suitable for high-dimensional datasets. Nevertheless, it is less interpretable than simple linear models and may require greater computational resources.

AdaBoost Regressor (Adaptive Boosting)

AdaBoost is a boosting algorithm that sequentially builds weak learners, typically shallow decision trees, and adjusts their weights based on prior prediction errors Schapire, R. E. (2013). Samples that are mispredicted receive higher weights, forcing subsequent learners to focus on more difficult cases. The final model aggregates the predictions of all weak learners into a weighted sum. AdaBoost is capable of modeling non-linear patterns and is generally more accurate than a single decision tree. However, it is sensitive to outliers and noise, as these can disproportionately influence subsequent iterations. Its performance strongly depends on hyperparameter tuning, such as the learning rate and the number of estimators.

Extreme Gradient Boosting (XGBoost)

XGBoost is an optimized implementation of gradient boosting, designed for scalability, speed, and regularization. It constructs trees sequentially, where each tree corrects the residual errors of its predecessors using gradient descent optimization T. Chen et al. (2016). Unlike AdaBoost, XGBoost incorporates advanced features such as L1 and L2 regularization, shrinkage (learning rate control), and parallel processing, making it one of the most powerful ensemble methods for structured data. Its ability to capture complex non-linear relationships, handle multicollinearity, and provide feature importance makes it highly suitable for predictive modeling in diverse domains. However, without appropriate hyperparameter tuning, XGBoost may suffer from overfitting, especially in datasets with strong linear dependencies.

Support Vector Regression (SVR)

Support Vector Regression , Awad, M. et al. (2015), extends the Support Vector Machine framework to regression tasks by fitting a function within a specified error tolerance (ε-insensitive margin). Predictions are optimized such that deviations within the margin are not penalized, focusing instead on points lying outside the tolerance zone. Kernel functions (e.g., radial basis function, polynomial) can be used to project data into higher dimensions, enabling the capture of complex non-linear relationships. SVR is known for strong generalization in small- to medium-sized datasets with high-dimensional features. However, its performance is highly dependent on hyperparameter selection (C, ε, and kernel choice) and computationally expensive for large datasets. In large-scale data environments, such as the present study, SVR may exhibit reduced accuracy and efficiency.

RESULTS AND DISCUSSION

The predictive performance of various machine learning models for TOTCIR and TOTATTEN was evaluated using R2 metrics on training and testing datasets. The models considered include Linear Regression (LR), Random Forest (RF), AdaBoost (AB), XGBoost (XGB), and Support Vector Regression (SVR). Figures presenting the actual versus predicted values for each model are provided in the results directory.

PREDICTING TOTAL CIRCULATION

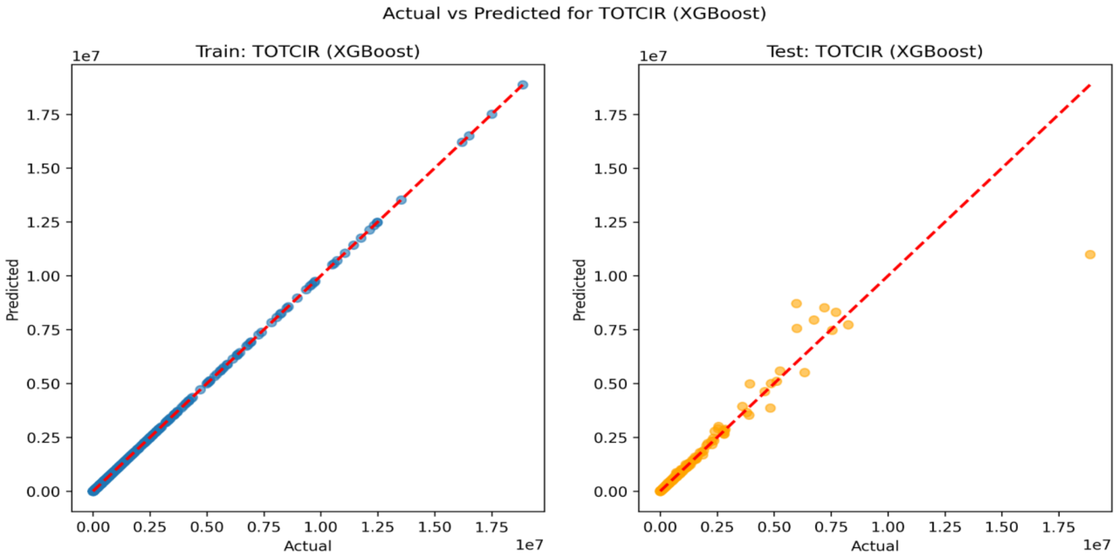
As shown in Table 1 and Figure 1(a-e), Linear Regression achieved perfect performance with R2 values of 1.000 for both training and testing datasets which shows overfitting issue. XGBoost also exhibited excellent training performance (R2=1.000) and maintained high accuracy on the test set (R2=0.925). Random Forest provided strong predictive capability with a slight overfitting effect, as indicated by R2=0.997 on training and 0.938 on testing. AdaBoost demonstrated moderate performance (Train R2=0.821 = 0.821, Test R2=0.701), whereas SVR performed poorly, yielding negative R2 values (Train R2=−0.039, Test R2=−0.044), indicating that it failed to capture the underlying data trends.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

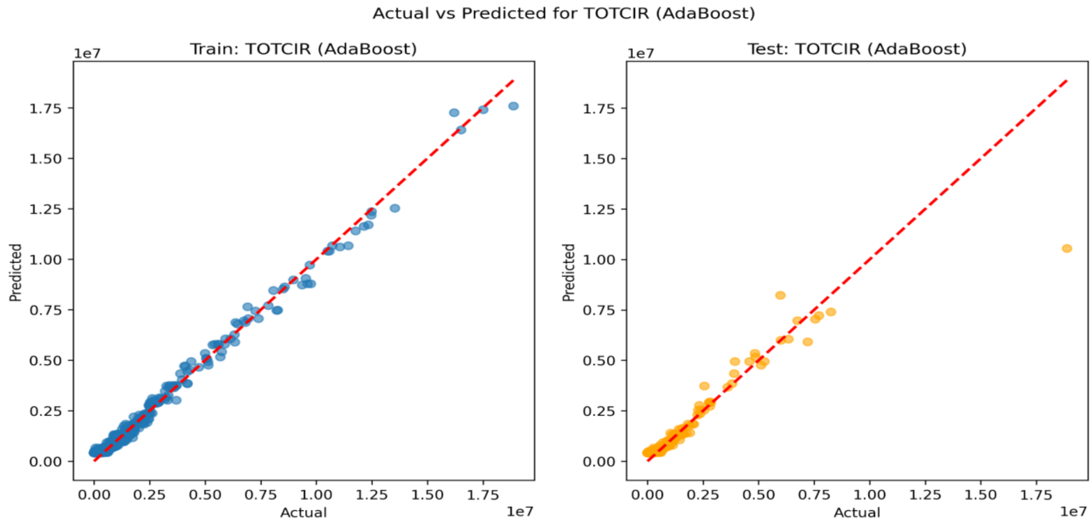
Research Article

Table 1 Model performance for TOTCIR prediction.

Model	Train R^2	Test R^2
LinearRegression	1.000	1.000
RandomForest	0.997	0.938
AdaBoost	0.821	0.701
XGBoost	1.000	0.925
SVR	-0.039	-0.044



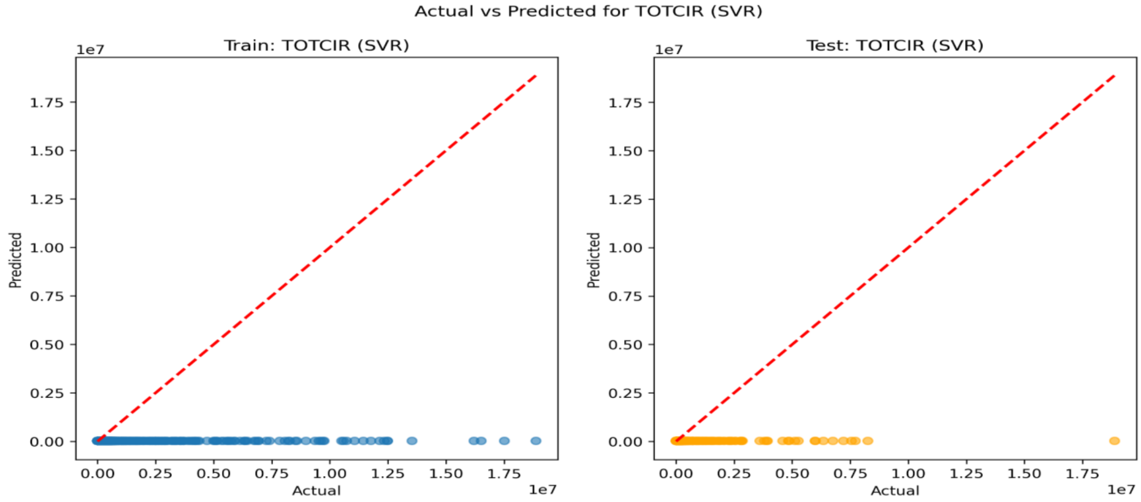
(a)



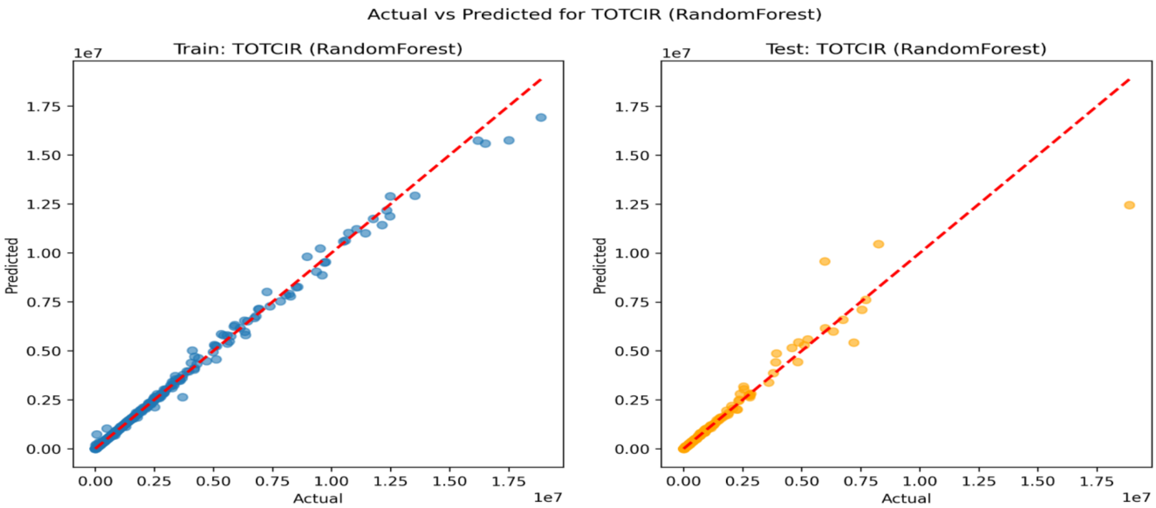
(b)

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

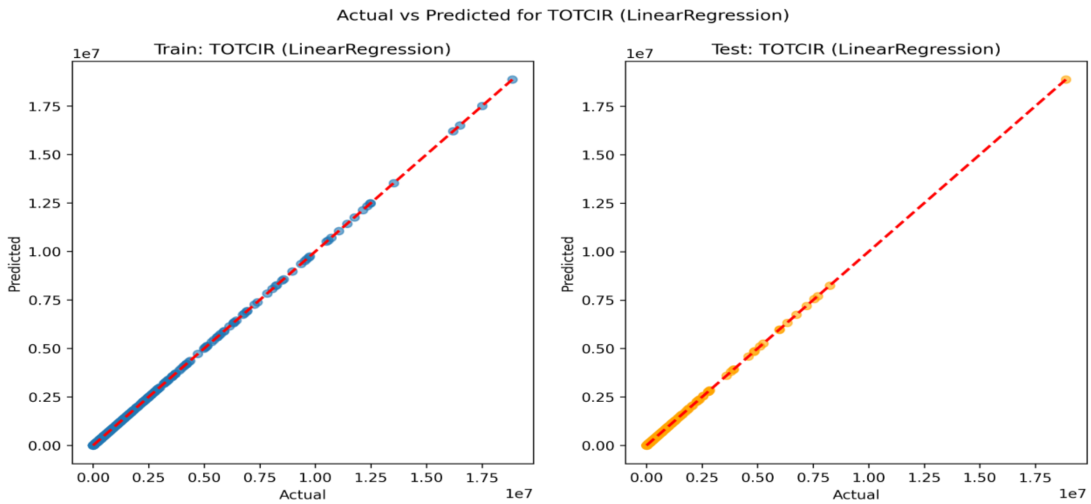
Research Article



(c)



(d)



(e)

Fig (a-e) Actual vs Predicted by ML algorithms for Total Circulation

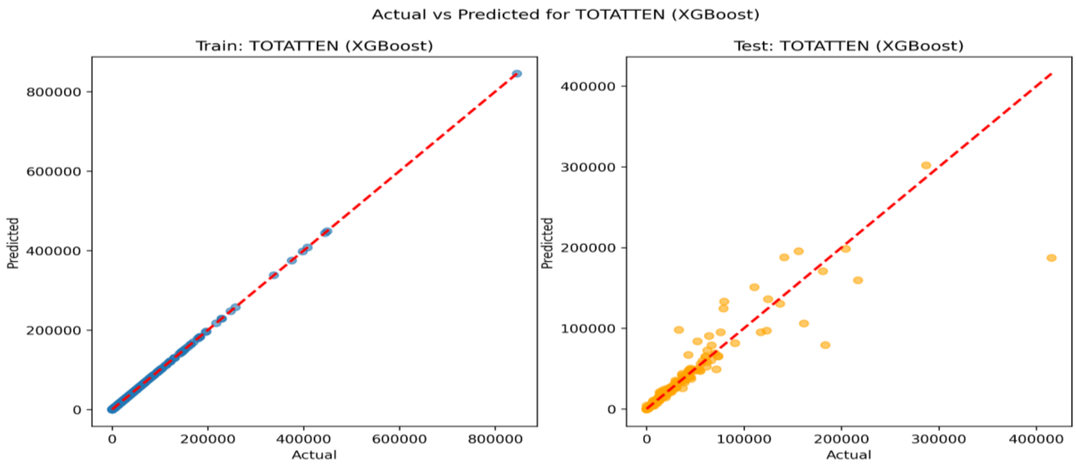
Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

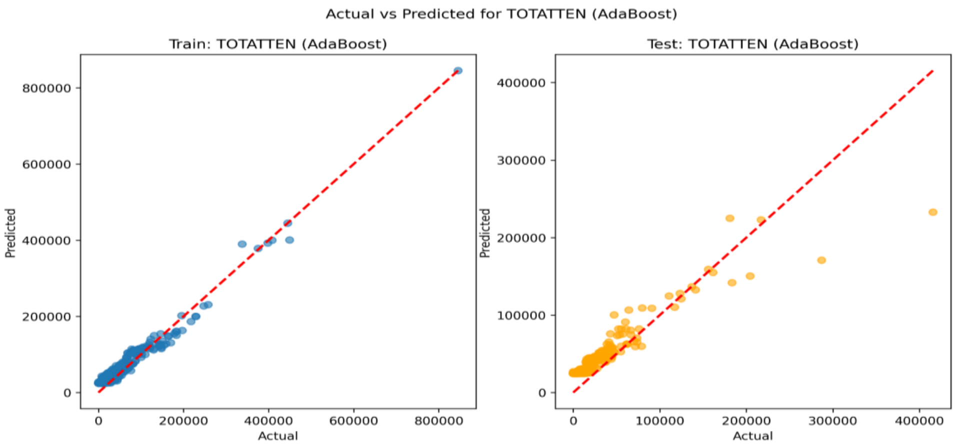
The results indicate that linear relationships dominate the TOTCIR dataset, as evidenced by the perfect performance of Linear Regression. Ensemble methods, such as Random Forest and XGBoost, also captured the underlying patterns well, although they show minor overfitting relative to linear regression. SVR and AdaBoost were less suitable for this task, likely due to non-optimal hyperparameter settings or sensitivity to outliers.

PREDICTING TOTAL ATTENDANCE

For TOTAL ATTENDANCE prediction, Linear Regression again provided the best performance, with train R2=1.000 and test R=0.998, indicating an almost perfect fit. Random Forest also performed robustly (train R2=0.977, test R2=0.947), showing minimal overfitting. XGBoost maintained good predictive capability with train R2=1.000 and test R2=0.873. In contrast, AdaBoost performed poorly (train R2=−0.030, test R2=−0.299), failing to generalize to the test data. SVR again underperformed, with negative R2 (train R2=−0.045, test R2=−0.054), indicating that it is not suitable for modeling TOTATTEN in its current form (refer table 2 and figure 2 a-e).



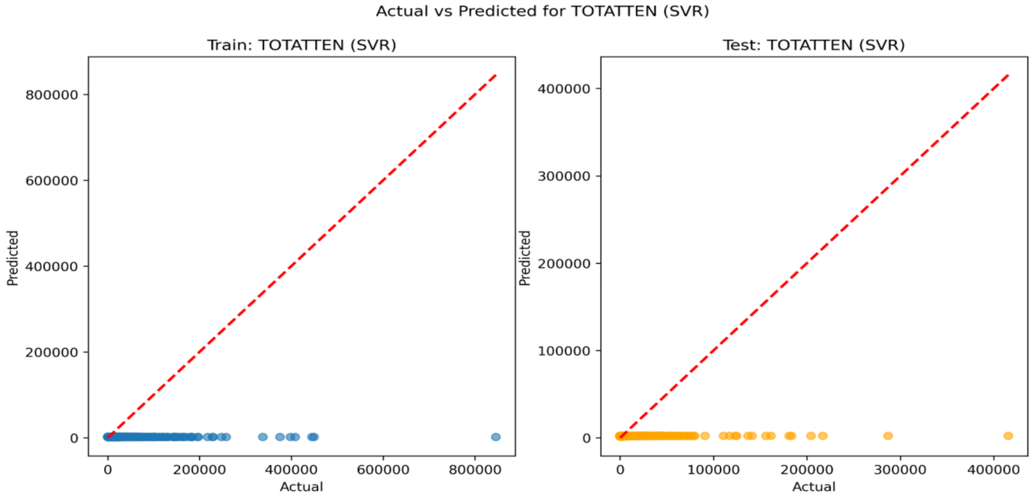
(a)



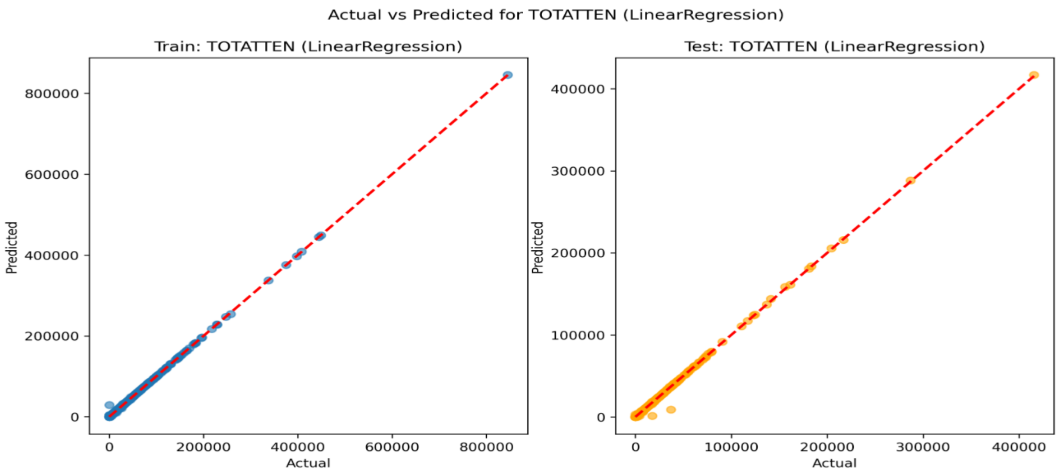
(b)

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

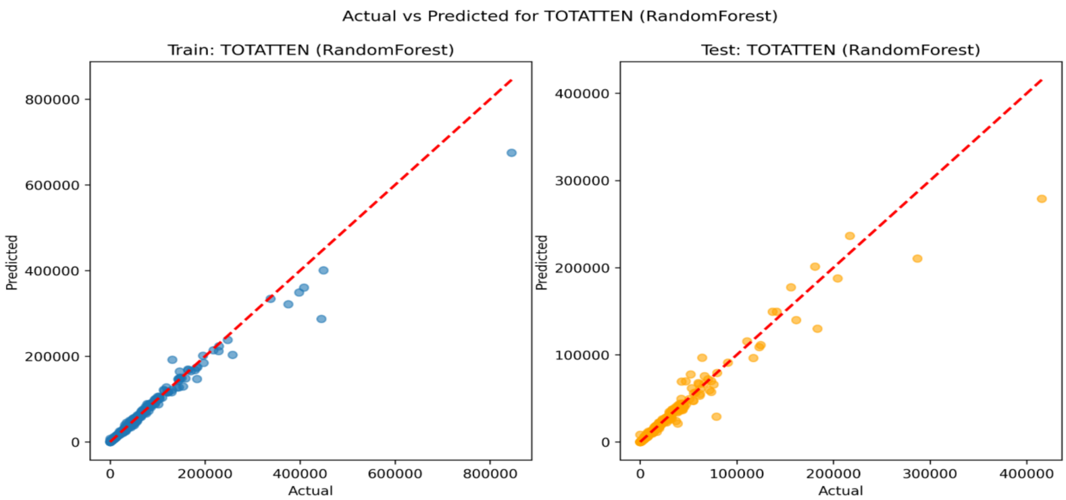
Research Article



(c)



(d)



(e)

Fig 2 (a-e) Actual vs Predicted by ML algorithms for Total ATTENDANCE

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 2 Model performance for TOTAL ATTENDANCE prediction.

Model	Train R^2	Test R^2
LinearRegression	1.000	0.998
RandomForest	0.977	0.947
AdaBoost	-0.030	-0.299
XGBoost	1.000	0.873
SVR	-0.045	-0.054

These results further confirm that the relationship between the predictors and TOTATTEN is predominantly linear, with Linear Regression capturing nearly all variability. Random Forest and XGBoost are effective alternatives, offering strong predictive power while tolerating minor non-linearities. The poor performance of AdaBoost and SVR suggests that these models are sensitive to data distribution and may require further tuning for reliable results.

PREDICTING TOTAL VISITS

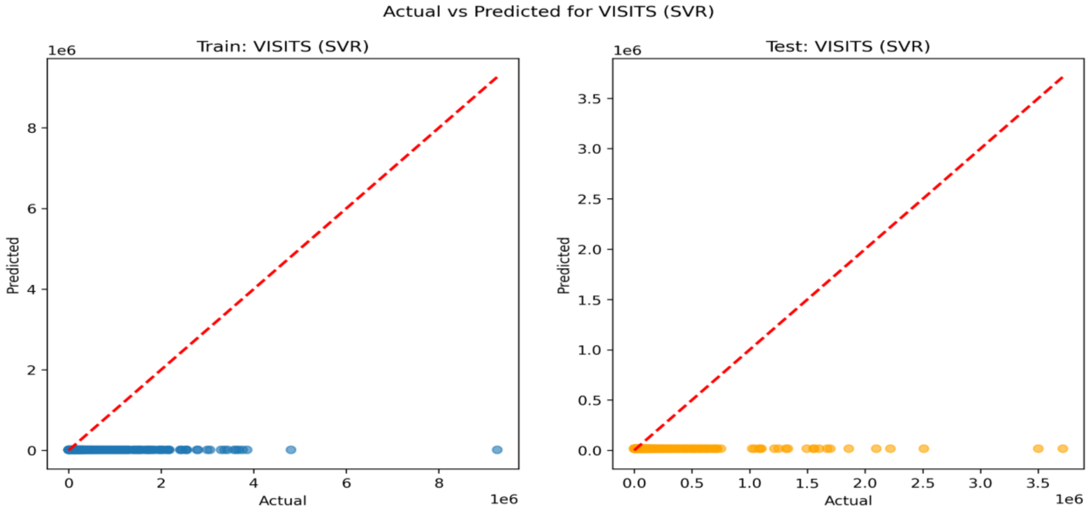
For VISITS (table 3, figure 3 a-e), LR (Train R2=0.951, Test R2=0.925) and RF (Train R2=0.981, Test R2=0.935) showed strong performance, indicating that linear and ensemble methods effectively captured the underlying patterns (Figure 3 a-e). XGB performed slightly lower (Train R2=1.000, Test R2=0.904), suggesting minor overfitting. AB (Train R2=0.127, Test R2=−0.255) and SVR (Train R2=−0.048, Test R2=−0.062) were ineffective, demonstrating poor generalization.

Table 3 Model performance for VISIT prediction.

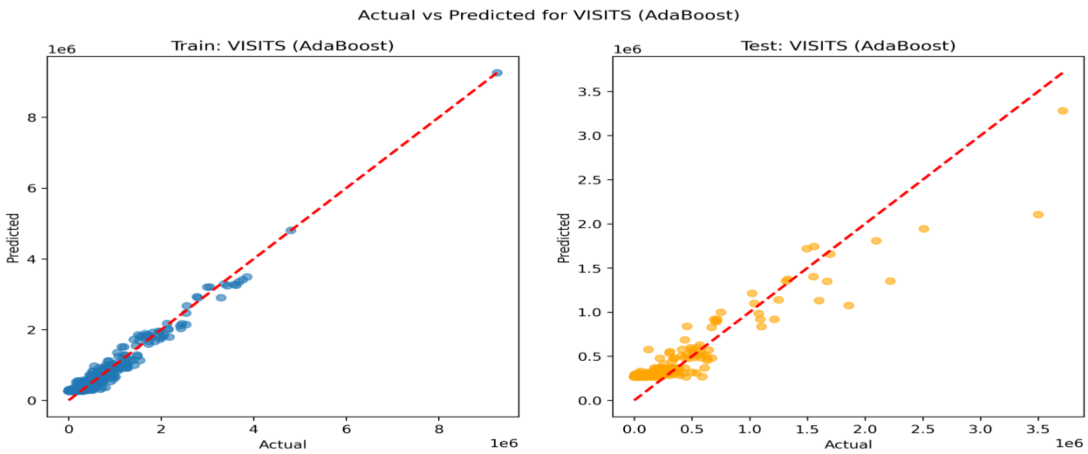
Model	Train R^2	Test R^2
LinearRegression	0.951	0.925
RandomForest	0.981	0.935
AdaBoost	0.127	-0.255
XGBoost	1.000	0.904
SVR	-0.048	-0.062

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

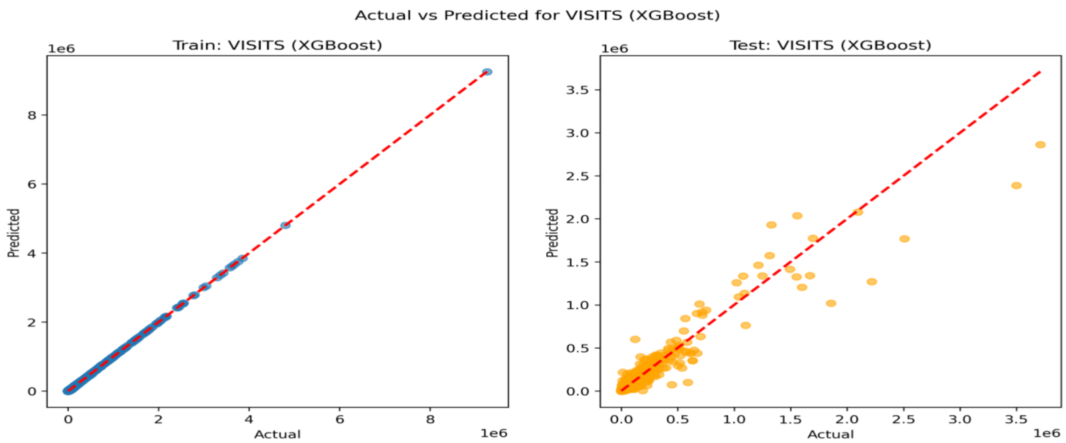
Research Article



(a)



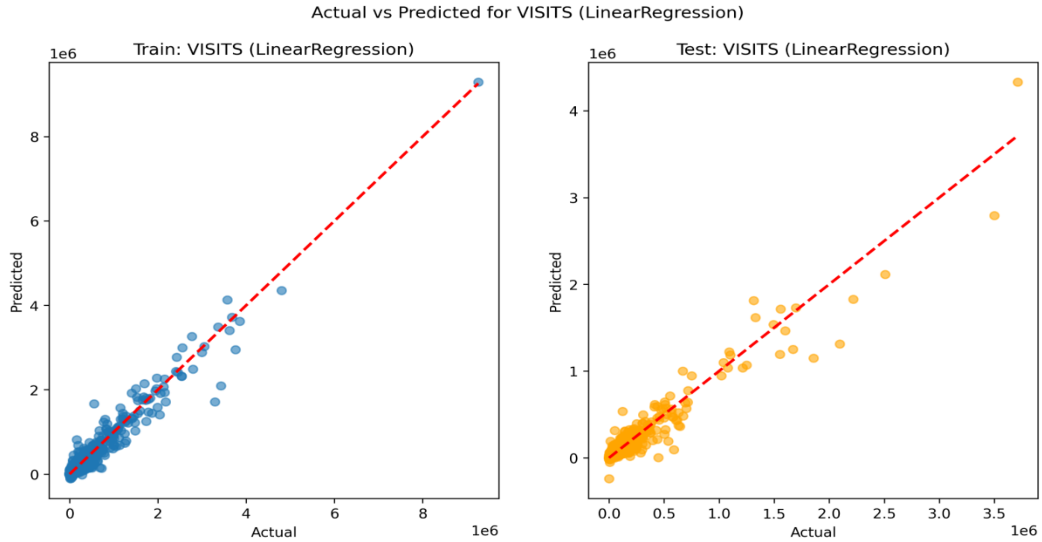
(b)



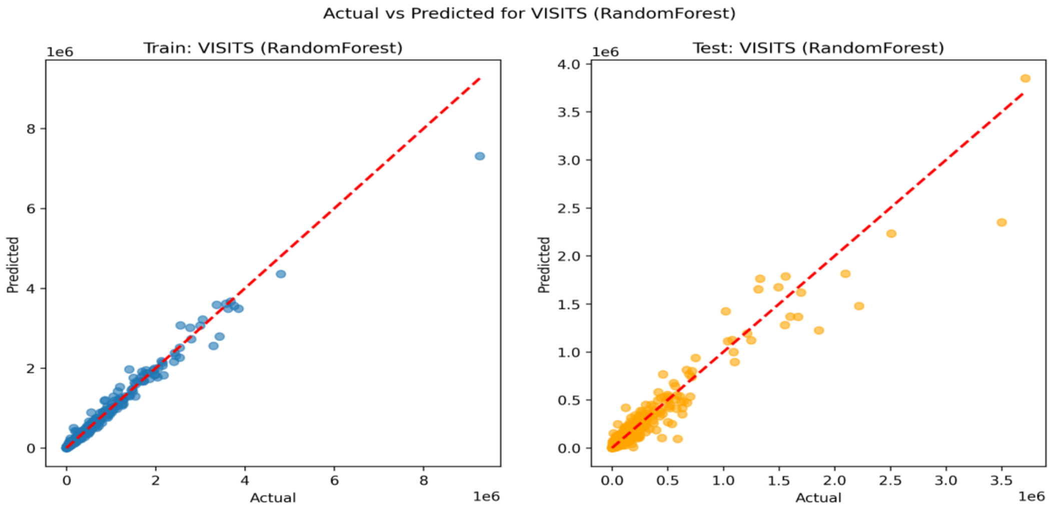
(c)

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



(d)



(e)

Figure 3 (a-e) Actual vs Predicted by ML algorithms for VISITS

Comparative Analysis

Across all three target variables, LR consistently provided the highest predictive accuracy, reflecting the dominant linear relationships present in the datasets. RF and XGB offered strong alternative models, capturing minor non-linearities while remaining robust against overfitting. AdaBoost and SVR consistently underperformed, highlighting their sensitivity to data distribution and the necessity for careful hyperparameter tuning.

The actual versus predicted plots corroborate these findings, with LR and RF closely aligning with the ideal 1:1 line, XGB showing slight deviations, and AdaBoost /SVR displaying significant scatter. These results suggest that, for applications requiring high precision and interpretability, LR is recommended, whereas RF and XGB can serve as robust alternatives in the presence of non-linearities or noise.

CONCLUSION

This study evaluated the predictive performance of five machine learning models—Linear Regression (LR), Random Forest (RF), AdaBoost (AB), XGBoost (XGB), and Support Vector Regression (SVR)—for Total Circulation (TOTCIR), Total Attendance (TOTATTEN), and VISITS. While LR appeared to achieve perfect or near-perfect performance on training data, the marginal drop in test accuracy indicates signs of overfitting and limited generalizability. In contrast, RF and XGB provided more balanced performance across training and testing datasets, effectively capturing both linear and minor non-linear relationships while maintaining robustness against overfitting. AB and SVR underperformed, failing to capture the underlying patterns in the data. These findings suggest that although LR can serve as a strong baseline for highly linear datasets, ensemble-based models such as RF and XGB offer greater reliability and resilience in practical applications where noise and non-linearities are present. Future research should explore hybrid approaches or regularization-enhanced linear models to mitigate overfitting and further improve predictive accuracy for complex or heterogeneous library datasets.

REFRENCES

[1] Atuase, D., & Maluleka, J. (2022). Marketing of library resources and its impact on the library usage of distance-learning students. Digital Library Perspectives, 39, 111–123. https://doi.org/10.1108/DLP-03-2022-0025

[2] Awad, M., & Khanna, R. (2015). Support vector regression. In Efficient learning machines: Theories, concepts, and applications for engineers and system designers (pp. 67-80). Berkeley, CA: Apress.

[3] Carey, J., Pathak, A. J., & Johnson, S. C. (2023). Effects of Library Workshop Attendance and Library Website Visit Frequency on Health Professions Students' LibGuides Awareness. Medical Reference Services Quarterly, 42(1), 16–30. https://doi.org/10.1080/02763869.2023.2160121

[4] Connell, R. S., Wallis, L. C., & Comeaux, D. (2021). The Impact of COVID-19 on the Use of Academic Library Resources. Information Technology & Libraries. https://doi.org/10.6017/ITAL.V40I2.12629

[5] Gilpin, G., Karger, E., & Nencka, P. (2021). The Returns to Public Library Investment. PSN: Local Politics & Policy (Topic). https://doi.org/10.21033/WP-2021-06

[6] Irwin, B., & Silk, K. (2019). Changing Stakeholder Expectations of Library Value. Public Library Quarterly, 38(3), 320–330. https://doi.org/10.1080/01616846.2019.1571398

[7] Joo, S., & Cahill, M. (2019). The relationships between the expenditures and resources of public libraries and children's and young adults' use: An exploratory analysis of Institute of Museum and Library Services public library statistics data. Journal of Librarianship and Information Science, 51 (2), 307–316. https://doi.org/10.1177/0961000617709057

[8] Kinney, B. (2010). The Internet, Public Libraries, and the Digital Divide. Public Library Quarterly, 29(2), 104–161. https://doi.org/10.1080/01616841003779718

[9] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[10] Mabe, M. R. (2022). Impact of Great Recession on Library Use: Does a Negative Economy Impact Library Use? Public Library Quarterly, 42, 515–533. https://doi.org/10.1080/01616846.2022.2128628

[11] Nzomo, P. (2023). The Impact of COVID-19 on the Use of Library Resources and Services by International Students: A Case Study. Journal of Library Administration, 63(5), 722–740. https://doi.org/10.1080/01930826.2023.2219602

[12] Project Open Data: DCAT-US JSON-LD context (version 1.1). (n.d.). 2022, from https://catalog.data.gov/dataset/public-library-survey-pls-2022

[13] Schapire, R. E. (2013). Explaining adaboost. In Empirical inference: festschrift in honor of vladimir N. Vapnik (pp. 37-52). Berlin, Heidelberg: Springer Berlin Heidelberg.

[14] Scoulas, J. M., & De Groote, S. L. (2019). The Library's Impact on University Students' Academic Success and Learning. Evidence Based Library and Information Practice, 14(3), 2–27. https://doi.org/10.18438/eblip29547

[15] Shapiro, S. D. (2016). Engaging a Wider Community: The Academic Library as a Center for Creativity, Discovery, and Collaboration. New Review of Academic Librarianship, 22, 24–42. https://doi.org/10.1080/13614533.2015.1087412

[16] Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. Wiley Interdisciplinary Reviews: Computational Statistics, 4(3), 275-294.

[17] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16. New York, NY, USA: Association for Computing Machinery,2016, p. 785–794.

[18] Witherspoon, R. L., & Taber, P. O. L. (2021). Increasing Student Attendance at Library Workshops: What the Data Tells Us. College & Research Libraries, 82(1), 113–128. https://doi.org/10.5860/crl.82.1.113