



Research Article

Library Transaction Data as a Predictor of User Retention in Academic Libraries: A Logistic Regression Approach

Comfort Oluwapelumi Egunyemichn^{1*}  and Ifeoluwa Adenike Ayoola² 

¹University of Ilesha, Ilesha, Osun State, Nigeria

²Bowen University, Iwo, Osun State, Nigeria

Article Information

Abstract

Article History

Received: 2 May 2026

Revised: 6 June 2026

Accepted: 30 June 2026

Published online: 10 July 2026

Keywords

Academic Libraries

Library Transaction Data

User Retention

Predictive Analytics

Logistic Regression

Library Users

Evidence-Based Library Management

Correspondence*

comfort_awolusi@unilesa.edu.ng

ORCID

Comfort Oluwapelumi Egunyemichn 

<https://orcid.org/0009-0009-7407-0718>

Ifeoluwa Adenike Ayoola 

<https://orcid.org/0009-0006-1157-9750>

Academic libraries generate substantial transaction data through circulation systems, online public access catalogues, electronic resource platforms, library access records, and reference services. Despite the availability of these data, many academic libraries continue to use them mainly for descriptive reporting rather than predictive decision-making. This study examined the extent to which library transaction data predict user retention in an academic library using binary logistic regression. The study adopted a correlational research design and used data from 500 registered users selected from four faculties and sixteen departments of the University of Ilesha. The predictor variables were borrowing frequency, OPAC search frequency, electronic resource access, library visit frequency, reference service usage, traditional library circulation use, and user category. User retention was coded as a binary outcome, with retained users coded 1 and inactive users coded 0. Descriptive statistics were used to examine transaction patterns, while binary logistic regression was used to determine significant predictors and the overall contribution of the predictors to retention. The results showed that electronic resource access and library visit frequency were statistically significant positive predictors of user retention, while borrowing frequency approached significance. The overall model was statistically significant, $\chi^2(7) = 151.26$, $p < .001$, with a McFadden pseudo R^2 of 0.289. The classification results indicated an overall accuracy of 76.0%, with a sensitivity of 72.2% and specificity of 78.1%. The study concludes that routine library transaction data can provide useful evidence for identifying patterns associated with continued library use and supporting evidence-based user engagement strategies. The study recommends the systematic use of transaction data for user monitoring, service planning, and targeted interventions.

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

Academic libraries remain important components of the teaching, learning, and research functions of higher education institutions. They provide access to scholarly information resources, research support, information literacy instruction, digital resources, institutional repositories, and physical learning spaces. The development of information and communication technologies has also changed the way academic libraries acquire, organise, provide, and assess information resources. Contemporary academic libraries operate through combinations of print collections, electronic resources, integrated library systems, online catalogues,

institutional repositories, discovery platforms, and other digital services (Association of College and Research Libraries [ACRL], 2023; Cox *et al.*, 2019).

The continued effectiveness of an academic library depends substantially on whether registered users continue to make meaningful use of the resources and services provided. The library user remains central to library and information services because the purpose of the library is closely connected with satisfying users' information needs (Aina, 2004; Nwalo, 2003). Consequently, assessment of academic library performance increasingly requires attention to patterns of actual use rather than relying exclusively on the

size of collections, number of registered users, or quantity of services provided. This is particularly important in an environment where students and researchers can obtain information through search engines, open-access repositories, electronic databases, institutional learning platforms, and emerging artificial intelligence tools.

User retention in academic libraries refers in this study to the continued active use of library resources and services by registered users over a defined period. It differs from registration because a user may register with a library without subsequently making sustained use of its resources. It also differs from occasional use because retention implies continuity of engagement. Continued use may occur through borrowing library materials, accessing electronic resources, visiting the library, searching the catalogue, or consulting library staff. Measuring continued use therefore provides a behavioural indication of the extent to which registered users remain engaged with the library.

The importance of sustained library use has been recognised in studies examining the relationship between library engagement and student outcomes. Haddow (2013), for example, examined academic library use and student retention using institutional data and found that library use was associated with student retention. Murray *et al.*, (2016) also found that academic library service use was associated with undergraduate student retention. Soria *et al.*, (2014) reported relationships between library use, academic achievement, and retention among first-year undergraduate students. Stemmer and Mahan (2016) similarly examined relationships between library use and student outcomes. More recently, O'Kelly *et al.*, (2023) examined the relationship between library instruction and student retention and reported a positive relationship between library instruction and subsequent student reenrolment.

These studies demonstrate the value of library use data for assessing the contribution of academic libraries. However, much of the existing work has focused on establishing relationships between library engagement and broader educational outcomes. There is a need for additional research that examines whether routine library transaction data can be used specifically to estimate the likelihood that a registered user will remain active. Such an approach changes the analytical question from simply asking how many users visited, borrowed materials, or accessed databases to asking whether observable patterns of use can help identify users who are more likely to remain engaged.

Academic libraries routinely generate transaction data through integrated library systems and associated platforms. Borrowing records, catalogue searches, electronic resource access, library visits, and reference consultations can provide behavioural evidence about how users interact with library services. These records have an important advantage over purely self-reported data because they capture recorded behaviour rather than relying entirely on users' recollection or stated intentions. Such data can therefore support

evidence-based library management when they are collected, organised, analysed, and interpreted appropriately (Booth, 2002; Eldredge, 2006).

The growing availability of operational data has also increased interest in predictive analytics. Predictive analytics involves the use of historical data and statistical or computational models to estimate the probability of future events or behaviours (Shmueli & Koppius, 2011). In higher education, analytics has been applied to student progression, academic performance, and retention. Academic libraries can similarly use their transaction data to identify patterns of engagement and support service planning. Oakleaf *et al.*, (2017) argued that academic library data can contribute to institutional learning analytics and provide evidence of the library's contribution to student success.

Binary logistic regression is suitable where the outcome variable has two categories. In the present study, user retention is classified into retained and inactive users. Logistic regression estimates the probability of an outcome from a set of predictor variables and produces coefficients that can be transformed into odds ratios. These odds ratios make the results relatively straightforward to interpret because they indicate how changes in a predictor are associated with the odds of the outcome while other predictors are held constant (Hosmer *et al.*, 2013). The method is therefore appropriate for examining whether borrowing, catalogue searching, electronic resource access, physical library visits, reference service use, circulation activity, and user category are associated with user retention.

The study focuses on the University of Ilesa Library because the library's transaction records provide a practical basis for examining the relationship between recorded user behaviour and continued library use. The study uses 500 registered users drawn from four selected faculties and sixteen departments. The analysis is intended to provide empirical evidence on the usefulness of library transaction data for predicting user retention and to demonstrate how routine operational records can support evidence-based management decisions.

A. Statement of the Problem

Academic libraries collect substantial amounts of transaction data through integrated library systems, circulation records, online catalogues, electronic resource platforms, library access records, and reference services. However, these data are often used primarily for administrative reporting, annual statistics, collection assessment, and service evaluation. Their potential for identifying patterns associated with continued or declining library use is less frequently examined.

A registered library user is not necessarily an active library user. Some users register and continue to borrow materials and use other library services, while others register but make little or no subsequent use of library resources. Some may visit the library for reading, internet access, or other facilities

without borrowing materials. Without systematic analysis of transaction behaviour, library managers may find it difficult to distinguish between these patterns and identify users whose engagement is declining.

The problem is therefore the limited use of routinely generated library transaction data to examine and predict user retention. Descriptive statistics can show the number of books borrowed, database accesses, library visits, or catalogue searches, but they do not by themselves estimate the probability of continued user engagement. Logistic regression provides a suitable approach because it can examine several behavioural variables simultaneously and estimate their relationship with a binary retention outcome.

This study consequently applies binary logistic regression to library transaction data to determine the factors associated with user retention and evaluate the predictive performance of the resulting model.

B. Objectives of the Study

The main objective of the study is to examine the extent to which library transaction data predict user retention in an academic library using binary logistic regression.

The specific objectives are to:

1. Identify the library transaction variables associated with user retention;
2. Determine the significant predictors of user retention among registered library users; and
3. Evaluate the predictive performance of the binary logistic regression model.

C. Research Questions

The study answers the following research questions:

1. What library transaction variables are associated with user retention?
2. Which variables are the significant predictors of user retention?
3. How accurately does the logistic regression model classify retained and inactive users?

D. Research Hypothesis

H₀: Library transaction variables do not have a statistically significant relationship with user retention among registered academic library users.

II. LITERATURE REVIEW

A. Concept of User Retention in Academic Libraries

User retention concerns the continued engagement of registered users with library resources and services over time. In an academic library, registration provides evidence that an individual has established membership, but it does not necessarily demonstrate sustained use. Retention is therefore

better assessed through repeated interaction with library resources and services.

The concept of retention is related to the broader literature on loyalty and continued use. Noh and Kim (2021) examined factors contributing to university library user loyalty and demonstrated the relevance of continued user engagement to university library services. Although loyalty and retention are not identical constructs, both emphasise continued relationships between users and library services. In academic libraries, sustained engagement can be observed through repeated borrowing, electronic resource access, catalogue use, library visits, and interactions with library staff.

The centrality of users to library services has long been established in African library and information science literature. Aina (2004) and Nwalo (2003) emphasised the importance of users and their information needs in determining the purpose and effectiveness of library services. Contemporary developments have expanded the channels through which users interact with libraries. Users can now engage with academic libraries physically and remotely through electronic resources, catalogues, repositories, and online services.

Research has also established relationships between library use and student outcomes. Haddow (2013) found an association between academic library use and student retention. Murray *et al.*, (2016) found that library service use was associated with undergraduate student retention. Soria *et al.*, (2014) reported relationships among first-year students' library use, academic achievement, and retention. These findings support the relevance of library engagement as an important area of academic library assessment.

However, continued library use should not be treated as identical to institutional student retention. A student may remain enrolled without using the library, while a registered library user may stop using library services but remain enrolled at the university. The present study therefore focuses specifically on library user retention, operationalised through continued active use of library materials and services. User retention is also relevant to resource planning. Academic libraries commit institutional resources to print collections, electronic databases, information systems, infrastructure, and professional services. Evidence concerning continued use can assist managers in assessing demand and planning services. ACRL (2023) emphasises assessment and evidence-based decision-making in higher education libraries. Transaction data can contribute to this process by providing measurable evidence of actual service use.

B. Library Transaction Data and Evidence-based Library Management

Library transaction data consist of records generated through users' interactions with library systems and services. Depending on the systems available in a library, such records may include borrowing transactions, catalogue searches,

electronic resource access, library visits, reference consultations, reservations, and other service interactions.

Evidence-based librarianship encourages the use of reliable evidence alongside professional knowledge in library decision-making (Booth, 2002; Eldredge, 2006). Traditional library assessment has often relied on descriptive indicators such as circulation totals, gate counts, reference transactions, and electronic resource statistics. These measures remain useful because they show patterns of service demand. However, descriptive statistics alone do not estimate which users are likely to continue using library services.

Transaction data provide an opportunity to examine user behaviour over time. For example, a user who records repeated electronic resource access and frequent library visits may exhibit a different engagement pattern from a registered user with no recorded transactions. Similarly, changes in borrowing or catalogue searching may indicate changes in information-seeking behaviour.

Research involving library data has demonstrated that transaction records can contribute to broader institutional assessment. Soria *et al.*, (2014) used library use indicators to examine relationships with academic achievement and retention. Haddow (2013) examined library use and student retention using institutional records. Murray *et al.*, (2016) similarly examined library services as predictors of student retention. These studies provide support for the use of recorded library behaviour as an analytical source.

The increasing use of integrated library systems has further expanded the volume and variety of data available to academic libraries. Systems such as Koha and other integrated library platforms can record circulation and catalogue transactions, while separate electronic resource platforms can provide access records. The availability of these data creates opportunities for more systematic analysis.

The challenge is that data availability does not automatically produce useful evidence. Transaction variables must be clearly defined, records must be cleaned, and the timing of predictors and outcomes must be specified. Privacy and ethical concerns must also be addressed when individual user records are analysed. Robertshaw and Asher (2019) drew attention to ethical issues associated with library learning analytics, making privacy, confidentiality, and responsible data use important considerations in any predictive library study.

C. Predictive Analytics in Academic Libraries

Predictive analytics uses historical observations to estimate the probability of future events or behaviours (Shmueli & Koppius, 2011). Its application in higher education has increased as institutions have gained access to large administrative and learning datasets. Siemens and Long (2011) described learning analytics as an important approach

to understanding learners and improving educational decision-making.

Academic libraries generate data that can contribute to such analytical activities. Oakleaf *et al.*, (2017) examined the relationship between academic libraries and institutional learning analytics and argued that library data can contribute to institutional understanding of student success. Library transaction records can therefore complement other institutional information when appropriate data-sharing arrangements and privacy safeguards exist.

Predictive analysis is different from descriptive reporting. A circulation report may show that a particular number of books were borrowed during a semester. A predictive model can examine whether borrowing frequency, electronic resource access, library visits, and other variables are associated with the probability of a user remaining active. The latter provides information that may be more useful for proactive service management.

Previous studies have established associations between library engagement and student outcomes. Haddow (2013), Murray *et al.*, (2016), Soria *et al.*, (2014), and Stemmer and Mahan (2016) provide evidence of the relationship between library use and educational outcomes. O'Kelly *et al.*, (2023) further examined library instruction and student retention. However, the research emphasis has often been on demonstrating library value or explaining relationships with student outcomes rather than developing models focused specifically on continued library use.

This creates an opportunity for academic libraries to examine their own operational data. A predictive model can potentially assist in identifying patterns of engagement that distinguish active users from users with limited engagement. Such information can inform user education, service promotion, electronic resource support, and outreach strategies.

Predictive analytics should nevertheless be applied carefully. A model does not establish causation simply because a predictor is statistically significant. A significant relationship indicates that a variable contributes information about the probability of the outcome within the model. It does not necessarily mean that changing the variable will cause retention to increase. This distinction is important when translating statistical findings into library management decisions.

D. Application of Logistic Regression in Library and Information Science

Logistic regression is appropriate when the dependent variable is categorical and particularly when the outcome has two categories. The method estimates the probability of an event occurring as a function of one or more predictor variables (Hosmer *et al.*, 2013). In this study, the event is user retention.

The basic model expresses the log odds of retention as a linear combination of the predictor variables:

$\text{logit}(p) = \ln[p/(1 - p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$
 where p represents the probability that a user is retained, β_0 is the intercept, and β_1 to β_k represent the coefficients associated with the predictor variables.

One advantage of logistic regression is interpretability. Exponentiating a coefficient produces an odds ratio. An odds ratio greater than 1 indicates that higher values of a predictor are associated with higher odds of the outcome, while an odds ratio below 1 indicates lower odds, assuming other variables remain constant (Hosmer *et al.*, 2013).

Murray *et al.*, (2016) provide a particularly relevant example because their study used logistic regression to examine academic library services and student retention. Haddow (2013) also used quantitative modelling to examine the relationship between library use and student retention. These studies demonstrate the usefulness of regression-based approaches in analysing library-related outcomes. The present study differs in its focus on library user retention as the dependent variable. Rather than treating university enrolment or academic persistence as the outcome, it examines whether the user's own library transaction behaviour can distinguish retained users from inactive users.

E. Research Gap

Existing literature demonstrates that library engagement is associated with important educational outcomes and that library transaction records can provide evidence of user behaviour. Studies by Haddow (2013), Soria *et al.*, (2014), Murray *et al.*, (2016), Stemmer and Mahan (2016), and O'Kelly *et al.*, (2023) provide substantial evidence of the relationship between library use and student outcomes. However, there remains a need for research that focuses specifically on predicting continued library use using routine

transaction indicators. Much of the literature examines student retention, academic achievement, or library value rather than the retention of library users as a distinct behavioural outcome. There is also a need for additional empirical evidence from African academic library settings.

The present study addresses this gap by using transaction variables generated through library operations to predict user retention. The study combines physical and digital indicators, including borrowing, OPAC searches, electronic resource access, library visits, reference services, circulation activity, and user category. This approach provides a behavioural perspective on retention and demonstrates the application of logistic regression to academic library transaction data.

III. RESEARCH METHODOLOGY

A. Research Design

The study adopted a correlational research design. The design was appropriate because the study examined relationships among naturally occurring library transaction variables and user retention without manipulating the variables. Binary logistic regression was used because the dependent variable had two categories: retained and inactive users (Creswell & Creswell, 2018; Hosmer *et al.*, 2013).

B. Population of the Study

The population comprised registered users of the University of Ilesa Library. The study included registered users who had library membership records, regardless of whether they made frequent or limited use of library resources and services.

For the study, four faculties were selected: Faculty of Social and Management Sciences, Faculty of Arts, Faculty of Education, and Faculty of Basic Medical Sciences. Four departments were selected from each faculty, producing sixteen departments.

TABLE I FACULTIES AND DEPARTMENTS SELECTED FOR THE STUDY

Faculty	Departments Selected
Social and Management Sciences	Economics, Accounting, Political Science, Public Administration
Arts	English, History and International Studies, Religious Studies, Linguistics
Education	Educational Management, Guidance and Counselling, Science Education, Adult Education
Basic Medical Sciences	Anatomy, Physiology, Medical Biochemistry, Nursing Science

C. Sample Size and Sampling Technique

A sample of 500 registered library users was used for the study. The users were selected from the sixteen departments across the four selected faculties. A multistage sampling procedure was applied. First, the four faculties were selected. Second, four departments were selected from each faculty. Third, registered library users were selected from the departments using proportionate allocation so that departments with larger numbers of registered users contributed more users to the sample. The use of 500 users

provided sufficient observations for the binary logistic regression model containing the specified predictors.

D. Sources and Method of Data Collection

Data were obtained from library transaction records and relevant user records. The transaction indicators included borrowing records, OPAC searches, electronic resource access, library visits, reference service use, and traditional circulation transactions. User category was obtained from the relevant user registration records.

The use of transaction records provided behavioural measures of library engagement. Where records were maintained through different service points, the relevant records were organised using anonymised user identifiers before analysis.

E. Description of Variables

The dependent variable was user retention. A retained user was coded 1, while an inactive user was coded 0. Retained

users were those who demonstrated active use of library materials during the defined study period. Registered users without sustained active material use were classified as inactive.

The independent variables comprised borrowing frequency, OPAC search frequency, electronic resource access, library visit frequency, reference service usage, traditional library circulation use, and user category.

TABLE II DESCRIPTION OF STUDY VARIABLES

Variable	Type	Description / Measurement
Retained User (outcome)	Binary	1 = active library user; 0 = inactive user
Borrowing Frequency	Count	Number of items borrowed during the study period
OPAC Search Frequency	Count	Number of online catalogue searches
Electronic Resource Access	Count	Number of recorded electronic resource logins/downloads
Library Visit Frequency	Count	Number of recorded physical library visits
Reference Service Usage	Count	Number of reference or enquiry consultations
Traditional Library Circulation Use	Count	Number of print-material checkout/return transactions
User Category	Categorical	Undergraduate or Staff

Traditional circulation use was retained as a separate predictor because it represents desk-mediated print circulation activity, whereas borrowing frequency was treated as the broader frequency of items borrowed. The distinction should be maintained in the underlying records to prevent duplication of the same transaction measure.

F. Model Specification

Binary logistic regression was used to estimate the probability of user retention. The model was specified as:

$$\text{logit}(p) = \ln[p/(1 - p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7$$

where:

p = probability of user retention;

β_0 = constant;

X_1 = borrowing frequency;

X_2 = OPAC search frequency;

X_3 = electronic resource access;

X_4 = library visit frequency;

X_5 = reference service usage;

X_6 = traditional library circulation use; and

X_7 = user category, with undergraduate users serving as the reference category.

The coefficients were exponentiated to obtain odds ratios. Statistical significance was assessed at the 0.05 level.

G. Method of Data Analysis

The data were analysed using SPSS. Descriptive statistics, including frequencies, percentages, and means, were used to summarise user characteristics and transaction patterns. Binary logistic regression was then used to examine the

relationship between the predictor variables and user retention.

Model significance was assessed using the likelihood ratio or Omnibus test. McFadden's pseudo R^2 was used as an indicator of model fit. To assess out-of-sample predictive performance, the 500 cases were divided into a training subset (80%, $n = 400$) used to estimate the model and a holdout test subset (20%, $n = 100$) used to evaluate classification accuracy, sensitivity, specificity, and the Area Under the ROC Curve (AUC).

H. Ethical Considerations

The study treated user records confidentially. Personally, identifying information was excluded from the analytical dataset and replaced with anonymised identifiers. The data were used for research purposes and handled in accordance with institutional requirements concerning confidentiality and responsible use of user records. Ethical safeguards are particularly important in library analytics because transaction records can reveal individual patterns of information use (Robertshaw & Asher, 2019).

IV. DATA PRESENTATION, ANALYSIS, AND DISCUSSION OF FINDINGS

A. Demographic and Descriptive Characteristics

Table III shows that 131 users (26.2%) were from Basic Medical Sciences, 130 (26.0%) from Social and Management Sciences, 128 (25.6%) from Education, and 111 (22.2%) from Arts. The distribution provided representation across all four selected faculties.

TABLE III DISTRIBUTION OF RESPONDENTS BY FACULTY

Faculty	Frequency	Percentage (%)
Basic Medical Sciences	131	26.2
Social and Management Sciences	130	26.0
Education	128	25.6
Arts	111	22.2
Total	500	100.0

TABLE IV DISTRIBUTION OF RESPONDENTS BY USER CATEGORY

User Category	Frequency	Percentage (%)
Undergraduate	447	89.4
Staff	53	10.6
Total	500	100.0

Table IV shows that undergraduates constituted 447 respondents (89.4%), while staff accounted for 53 (10.6%). The sample was therefore predominantly undergraduate.

TABLE V DISTRIBUTION OF RESPONDENTS BY ENGAGEMENT GROUP

Engagement Group	Frequency	Percentage (%)
Active users	180	36.0
Registered but not actively using materials	274	54.8
Facility-only users	46	9.2
Total	500	100.0

Table V indicates that 180 users (36.0%) were active users, 274 (54.8%) were registered but not actively using library materials, and 46 (9.2%) were facility-only users. For the binary retention analysis, active users were coded as retained,

TABLE VII LOGISTIC REGRESSION COEFFICIENTS AND ODDS RATIOS

Predictor	B	S.E.	p-value	Odds Ratio	95% CI
Constant	-3.289	0.314	<.001	0.037	—
Borrowing Frequency	0.095	0.054	.080	1.100	0.989–1.224
OPAC Search Frequency	0.040	0.050	.419	1.041	0.944–1.148
Electronic Resource Access	0.118	0.048	.014*	1.125	1.024–1.235
Library Visit Frequency	0.103	0.037	.005*	1.109	1.031–1.192
Reference Service Usage	0.113	0.092	.220	1.119	0.935–1.341
Traditional Library Circulation Use	0.089	0.068	.189	1.093	0.957–1.248
User Category: Staff	0.513	0.437	.240	1.671	0.710–3.932

*Significant at $p < .05$. Reference category for user category = Undergraduate.

Model fit: McFadden pseudo $R^2 = 0.289$; likelihood ratio $\chi^2(7) = 151.26$, $p < .001$.

The likelihood ratio test was significant, indicating that the full model provided a significantly better fit than the intercept-only model. The McFadden pseudo R^2 of 0.289 indicates a meaningful contribution of the predictors to explaining variation in retention status.

producing 180 retained users (36.0%) and 320 inactive users (64.0%).

B. Descriptive Statistics of Predictor Variables

TABLE VI MEAN TRANSACTION INDICATORS BY RETENTION STATUS

Variable	Inactive (n=320)	Retained (n=180)
Borrowing Frequency	2.97	6.00
OPAC Search Frequency	3.86	6.64
Electronic Resource Access	3.96	7.82
Library Visit Frequency	5.91	10.89
Reference Service Usage	1.14	2.71
Traditional Library Circulation Use	2.21	4.94

Table VI shows that retained users recorded higher mean values on all six transaction indicators. The largest differences occurred in library visit frequency and electronic resource access, indicating stronger physical and digital engagement among retained users.

C. Logistic Regression Results

The regression results show that electronic resource access and library visit frequency were significant positive predictors of user retention.

Electronic resource access had an odds ratio of 1.125, while library visit frequency had an odds ratio of 1.109. Borrowing frequency was positive but did not reach the 0.05 significance level. The remaining predictors were also positive but statistically non-significant.

D. Model Predictive Performance

To assess predictive performance independently of the model-estimation data, the holdout test subset of 100 users (64 inactive, 36 retained) described in section 3.7 was used for classification.

TABLE VIII CLASSIFICATION TABLE
(HOLDOUT TEST SUBSET, N=100)

Actual	Predicted Inactive	Predicted Retained	Total
Inactive	50	14	64
Retained	10	26	36
Total	60	40	100

The classification results show that the model correctly classified 76.0% of the cases in the holdout test subset. Sensitivity was 72.2%, indicating that the model correctly identified most retained users, while specificity was 78.1%, indicating stronger identification of inactive users. The reported AUC of 0.822 indicates good discrimination between retained and inactive users.

E. Answers to Research Questions and Test of Hypothesis

1. *Research Question One:* What library transaction variables are associated with user retention?

Electronic resource access and library visit frequency were statistically significant positive predictors of retention. Borrowing frequency also had a positive relationship with retention but did not reach statistical significance at the 0.05 level. OPAC search frequency, reference service usage, traditional library circulation use, and user category were not statistically significant in the regression model.

2. *Research Question Two:* Which variables are the significant predictors of user retention?

Electronic resource access and library visit frequency were the statistically significant predictors. Their odds ratios indicate that increases in these variables were associated with higher odds of retention, holding other variables constant.

Although staff users had an odds ratio of 1.671, the relationship was not statistically significant. It should therefore not be interpreted as a statistically established predictor.

3. *Research Question Three:* How accurately does the logistic regression model classify retained and inactive users?

On the holdout test subset, the model correctly classified 76.0% of the cases. It recorded a sensitivity of 72.2% and specificity of 78.1%, while the reported AUC of 0.822 indicates good discriminatory performance.

4. Test of Hypothesis

H_0 : Library transaction variables do not have a statistically significant relationship with user retention.

The likelihood ratio test produced $\chi^2(7) = 151.26$, $p < .001$. Since the p-value is below 0.05, the null hypothesis is rejected. The result indicates that the set of library transaction

variables, considered jointly, has a statistically significant relationship with user retention.

F. Discussion of Findings

The study found that retained users recorded higher mean values across all transaction indicators than inactive users. This indicates that continued library engagement is reflected in multiple forms of interaction, including borrowing, catalogue searching, electronic resource use, physical visits, reference consultations, and circulation activity. The result supports the general position in previous research that active library engagement is associated with meaningful educational and institutional outcomes (Haddow, 2013; Murray *et al.*, 2016; Soria *et al.*, 2014).

Electronic resource access was a significant predictor of retention. This finding suggests that users who make greater use of electronic resources are more likely to remain active library users. The result is consistent with the changing structure of academic library services, where electronic resources have become an important channel for accessing scholarly information. It also supports the broader argument that academic library engagement should be assessed across both physical and digital service environments (Cox *et al.*, 2019).

Library visit frequency was also a significant positive predictor. Users who visited the library more frequently had higher odds of being retained. This finding indicates that physical library engagement remains relevant despite the increasing availability of electronic information resources. The library continues to provide physical spaces and services that support study, information seeking, access to materials, and interaction with library staff.

Borrowing frequency had a positive coefficient and an odds ratio greater than 1, but its p-value of .080 did not meet the 0.05 threshold. This means that the observed positive relationship was insufficient to establish borrowing frequency as a statistically significant independent predictor in the model. Its higher mean among retained users nevertheless indicates that borrowing remains associated with stronger engagement at the descriptive level.

OPAC search frequency was also higher among retained users but was not statistically significant after the other variables were included. One possible interpretation is that catalogue searching overlaps with other forms of library engagement. A user may search the catalogue because the user intends to borrow materials or access other services. Consequently, the independent contribution of OPAC searches may become less pronounced when other behavioural variables are controlled.

Reference service usage and traditional circulation use showed higher mean values among retained users but were not statistically significant predictors in the regression model. This difference between descriptive and inferential results is

important. A variable can show a noticeable difference between two groups but fail to make a statistically significant independent contribution when several related predictors are considered simultaneously.

The overall model was statistically significant, with a likelihood ratio $\chi^2(7)$ of 151.26 and $p < .001$. This indicates that the predictors jointly provide useful information for distinguishing retained users from inactive users. The McFadden pseudo R^2 of 0.289 further indicates that the model accounts for a meaningful portion of the variation in retention status, although transaction variables do not capture every factor that may influence continued library use.

The predictive classification results, drawn from the holdout test subset, also support the usefulness of the model. An overall accuracy of 76.0% indicates that the model correctly classified more than three-quarters of the evaluated cases. Sensitivity of 72.2% means that the model identified a substantial proportion of retained users, while specificity of 78.1% indicates a somewhat stronger ability to identify inactive users. The reported AUC of 0.822 further indicates good discrimination.

The findings are consistent with previous research showing that library use is associated with student and institutional outcomes. Haddow (2013) and Murray *et al.*, (2016) demonstrated the importance of library engagement in relation to student retention, while Soria *et al.*, (2014) found relationships between library use and academic success. The present study extends this line of research by treating library user retention itself as the outcome and using multiple transaction indicators as predictors. The findings also support the relevance of evidence-based library management. Rather than relying solely on general assumptions about library use, managers can examine transaction patterns to identify differences between retained and inactive users. Such analysis can inform service planning, user engagement programmes, electronic resource promotion, and targeted communication.

The results should, however, be interpreted as associations rather than evidence of causation. For example, the significant relationship between library visits and retention does not establish that increasing the number of visits will necessarily cause users to remain active. Other factors, including academic workload, information needs, awareness of library services, course requirements, satisfaction, and access to alternative information sources, may influence user behaviour. Future research can incorporate these factors alongside transaction records.

V. SUMMARY, CONCLUSION AND RECOMMENDATIONS

A. Summary

The study examined library transaction data as predictors of user retention in an academic library using binary logistic

regression. Data from 500 registered users were analysed across four selected faculties and sixteen departments. The study examined borrowing frequency, OPAC search frequency, electronic resource access, library visit frequency, reference service usage, traditional library circulation use, and user category.

Descriptive results showed that retained users recorded higher mean values across all six transaction indicators. The regression analysis showed that electronic resource access and library visit frequency were statistically significant positive predictors of retention. Borrowing frequency was positive but did not reach statistical significance, while OPAC searches, reference service use, traditional circulation use, and staff status were not statistically significant independent predictors.

The overall logistic regression model was statistically significant, and the predictive classification results, based on the holdout test subset, showed 76.0% accuracy, with sensitivity of 72.2% and specificity of 78.1%. These findings indicate that transaction data contain useful information for distinguishing retained users from inactive users.

B. Conclusion

The study concludes that library transaction data provide useful behavioural evidence for predicting user retention in academic libraries. Electronic resource access and library visit frequency emerged as significant predictors, indicating that both digital and physical engagement contribute important information when distinguishing retained users from inactive users.

The significant overall model demonstrates that library transaction variables, considered jointly, are associated with user retention. The results also show the value of moving beyond simple descriptive statistics toward analytical approaches that can identify patterns in user behaviour.

The findings provide a basis for academic library managers to use routinely generated transaction data in service assessment and user engagement planning. However, predictive results should be interpreted as evidence of association rather than direct evidence that increasing any particular transaction will cause retention to increase.

C. Recommendations

Based on the findings, the study recommends that

1. Academic library management should strengthen electronic resource services. Since electronic resource access was a significant predictor of retention, libraries should ensure that users have reliable access to subscribed databases, electronic journals, digital collections, and related services.
2. The library should encourage meaningful physical library engagement. Library visit frequency was a significant predictor of retention. User education,

reading-space improvement, orientation programmes, and relevant physical services can be used to encourage continued library engagement.

3. Library transaction data should be incorporated into routine assessment. Management should regularly examine borrowing, catalogue use, electronic resource access, visits, and reference transactions rather than relying exclusively on annual descriptive statistics.
4. Predictive models should support targeted user engagement. Transaction patterns can help identify users with declining engagement and guide appropriate communication or user-support activities.
5. Libraries should maintain clear distinctions among transaction variables. Borrowing frequency and traditional circulation use should be operationally defined so that the same transaction is not counted twice under different predictor names.
6. Future studies should include additional explanatory variables. Variables such as user satisfaction, awareness of library services, perceived usefulness, academic workload, information literacy, and access to alternative information sources may improve understanding of user retention.
7. User privacy should remain central to library analytics. Transaction data should be anonymised where possible, securely stored, accessed only for legitimate purposes, and handled according to institutional ethical and data-protection requirements.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The authors confirm that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

Aina, L. O. (2004). *Library and information science text for Africa*. Third World Information Services.

- Association of College and Research Libraries. (2023). *Standards for libraries in higher education*. American Library Association. <https://www.ala.org/acrl/standards/standardslibraries>.
- Booth, A. (2002). From EBM to EBL: Two steps forward or one step back? *Medical Reference Services Quarterly*, 21(3), 51–64. https://doi.org/10.1300/J115v21n03_06.
- Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of AI on academic libraries. *Library Hi Tech*, 37(3), 418–435. <https://doi.org/10.1108/LHT-08-2018-0105>.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- Eldredge, J. D. (2006). Evidence-based librarianship: The EBL process. *Library Hi Tech*, 24(3), 341–354. <https://doi.org/10.1108/07378830610692118>.
- Haddow, G. (2013). Academic library use and student retention: A quantitative analysis. *Library & Information Science Research*, 35(2), 127–136. <https://doi.org/10.1016/j.lisr.2012.12.002>.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>.
- Murray, A., Ireland, A., & Hackathorn, J. (2016). The value of academic libraries: Library services as a predictor of student retention. *College & Research Libraries*, 77(5), 631–642. <https://doi.org/10.5860/crl.77.5.631>.
- Noh, Y., & Kim, H. (2021). Developing a model to identify the factors contributing to user loyalty of university libraries. *The Journal of Academic Librarianship*, 47(5), 102386. <https://doi.org/10.1016/j.acalib.2021.102386>.
- Nwalo, K. I. N. (2003). *Fundamentals of library practice: A manual on library routines*. Stirling-Horden.
- Oakleaf, M., Whyte, A., Lynema, E., & Brown, M. (2017). Academic libraries and institutional learning analytics: One path to integration. *The Journal of Academic Librarianship*, 43(5), 454–461. <https://doi.org/10.1016/j.acalib.2017.08.008>.
- O'Kelly, M. K., Jeffries, J., Hobscheid, M., & Passarelli, R. (2023). Correlation between library instruction and student retention: Methods and implications. *College & Research Libraries*, 84(1), 85–99. <https://doi.org/10.5860/crl.84.1.85>.
- Robertshaw, M. B., & Asher, A. (2019). Unethical numbers? A meta-analysis of library learning analytics studies. *Library Trends*, 68(1), 76–101. <https://doi.org/10.1353/lib.2019.0031>.
- Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553–572. <https://doi.org/10.2307/23042796>.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*, 40(1), 84–91. <https://doi.org/10.1016/j.acalib.2013.12.002>.
- Stemmer, J. K., & Mahan, D. M. (2016). Investigating the relationship of library usage to student outcomes. *College & Research Libraries*, 77(3), 359–375. <https://doi.org/10.5860/crl.77.3.359>.