Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.1, 2026, pp.179-189

DOI: https://doi.org/10.51983/ijiss-2026.16.1.19

Digital Footprint Analytics for User Behavior Prediction in Libraries

R. Radha[1*], A. Rehash Rushmi Pavitra[2], Dr.M. Hemasundari[3], Haeedir Mohameed[4], Dr.S.N.V.J. Devi Kosuru[5] and Dr. Aida Ventkat Rao Dora[6]

[1*]Department of Data Science and Business Systems, SRM Institute of Science and Technology, Kattankulathur, India
[2]Assistant Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, Chennai, India
[3]Assistant Professor, Department of Management Studies, SRM Valliammai Engineering College, Kattankulathur Tamil Nadu, India
[4]Department of Computers Techniques Engineering, College of Technical Engineering, Islamic University in Najaf, Najaf, Iraq; Department of Computers Techniques Engineering, College of Technical Engineering, Islamic University in Najaf of Al Diwaniyah, Al Diwaniyah, Iraq
[5]Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
[6]Assistant Professor, Department of Management, Kalinga University, Naya Raipur, Chhattisgarh, India
E-mail: [1]radhar@srmist.edu.in, [2]rushmiips@gmail.com, [3]hemasundarim.mba@srmvalliammai.ac.in, [4]tech.iu.comp.haideralabdeli@gmail.com, [5]jyotsnakosuru@gmail.com, [6]ku.aidaventkatraodora@kalingauniversity.ac.in
ORCID: [1]https://orcid.org/0000-0001-9822-0949, [2]https://orcid.org/0000-0002-1220-1497, [3]https://orcid.org/0009-0007-9714-6231, [4]https://orcid.org/0009-0002-0098-5228, [5]https://orcid.org/0000-0003-1521-5701, [6]https://orcid.org/0009-0008-1572-6994
(Received 26 October 2025; Revised 24 November 2025, Accepted 08 December 2025; Available online 05 January 2026)

***Abstract* - As libraries seek to have more people satisfied as well as to manage resources better, the application of data analytics in the digital age sounds almost ubiquitous. This paper will look at the use of digital footprints to forecast behavior in libraries. This paper will build models on the behavioral pattern of users, such as logins, use of resources, search, and borrowing history, so that it can serve users better and also to maximize resource allocation. This model employs machine learning techniques, namely, supervised models, such as logistic regression and decision trees to extract patterns and relationships in user data. Incorporating digital footprint analysis in libraries enhances personalized service and helps librarians make sound administrative judgments. Libraries are used to collect data, which will be anonymized to ensure that the identities of users are not revealed and to conform to ethics. The suggested structure would enable digital libraries to more efficiently predict user actions and proactively propose the appropriate content. This case demonstrates the significance of behavioral analysis in building responsive, self-learning, and user-optimized intelligent library systems. The findings contribute to the rationale of the increasing literature on digital libraries and also highlight the development of library analytics in providing library services to academic and public libraries.**

***Keywords:* Digital Footprint, Analytics, User Behavior, Prediction, Libraries, Machine Learning, Personalization**

I. INTRODUCTION

The increasing digitalization of libraries results in greater user interaction with digital interfaces, including online catalogs, e-resources, mobile applications, and self-checkout systems. Such interactions propel data generation on a monumental scale, commonly known as digital footprints. A digital footprint includes every fragment of a user's presence in the library's information system, including search queries, clickstreams, borrowing patterns, and logins (Rowley, 2014; Haddow, 2013; Kovacevic et al., 2010). Now, libraries possess the capabilities of modern digital systems, which include the flexibility to capture, store, and analyze data on user behavior, thereby providing rich insights (Tenopir et al., 2015; Chakma, 2025). This provides an opportunity to incorporate data-driven decision-making into what were once traditional libraries. The introduction of Library Management Systems (LMS), integrated digital resource systems, and RFID tracking systems have improved the accuracy and detail with which user data is captured far beyond what was previously available (Joo & Choi, 2015; Tunc et al., 2024; Dwivedi et al., 2013). These digital footprints act as a reservoir that not only maps out resource use but also records behavioral patterns relevant to efficient and effective design in user service efforts, engagement initiatives, collection development, and marketing. From the viewpoint of a librarian, as libraries strive to meet today's users' demands, the reliance on digital footprints helps in being anticipatory and, oftentimes, pre-emptive rather than reactive to user needs (Khudhair, 2024; Kaganer et al., 2023).

Predicting user behaviour is of strategic importance for modern libraries (Ansari et al., 2021). Digital systems can help libraries anticipate user behaviors and actions, tailor resource recommendations, and boost resource service delivery with a high degree of certainty (Zhou & Zhan, 2020; Muralidharan, 2023). Indicatively, when a system can determine the patterns of a student consuming digital journals to gain access during exam dates, it will be able to provide the required content or at least remind the student. Effective behavioral prediction can help libraries optimize the allocation of resources, enabling strategic decisions such as targeted subscriptions and acquisitions according to the needs of users (Yu et al., 2017; Shamsitdinova et al., 2024). Predictive analytics streamlines operational workflows by simplifying user-facing services and enhancing service delivery decisions on the administrative side (Oakleaf, 2018). Moreover, these insights may be used to increase retention and recognize vulnerable users (e.g., those with a usage decrease) and improve demographic or academic-based outreach (Li et al., 2021; Kavitha, 2020). Libraries face increasing competition for attention within the digital realm and, as such, need to evolve into smart platforms that tailor services to user needs (Li, S et al., 2019). With data-driven societies on the rise, libraries must stay relevant, competitive, and aligned with educational trends. This is made possible through user behavior-driven predictive analytics (Borgman, 2014; Sharma & Nayak, 2018). When implemented carefully, behavior prediction enables customization while safeguarding user privacy.

This paper focuses on the application of digital footprint analytics for predicting user behavior in the library context. The study develops a framework for applying machine learning to interaction logs, borrowing history, searches, and usage frequency (Ferguson et al., 2016; Salo, 2024). The primary aim is to construct predictive models for personalized recommendations, intelligent service automation, and decision support systems in libraries. This research further examines the technical and ethical issues associated with behavior prediction, including accuracy, privacy, the ethics of transparency, and data quality. By reviewing case studies of recent technologies used in libraries and integrating contemporary library technologies with data science, the research constructs a strategy for applying predictive analytics to libraries. Addressing the research questions will demonstrate that digital footprints, when analyzed carefully through ethical lenses, can be successfully transformed into knowledge that drives action. This work enables the construction of flexible and anticipatory digital libraries that cater to students, researchers, and the general public by adapting to their needs in real time.

The remainder of the paper is organized as follows. A literature review detailing digital footprint analytics, user behavior prediction in libraries, and key challenges in the domain is provided in Section II. The methodology, data collection, analysis, and predictive model applied in the study are presented in Section III. The analysis of user behavior patterns and their relevance to library services is discussed in Section IV. In section V, the considerations of the findings and their relevance within the literature are provided, along with insights on future work to be undertaken. Concluding the paper in section VI, practical recommendations and other research areas are proposed alongside the contributions of this work.

II. Literature Review

Digital footprint analytics refers to the capture, analysis, and interpretation of data that users leave behind through their online activities. Within a library scenario, such data includes information such as catalog search actions, e-resource utilization, borrowing activities, login sessions, and page views (Seadle, 2016; Aliyari, 2024; Huvila et al., 2017). All of these traces or footprints, when taken together, can be of great assistance in decision-making, tailoring systems, and improving services on user behaviour (Nicholas et al., 2017; Puri & Lakhwani, 2013). Digital footprints are classified into two types: passive footprints that are automatically accrued as a result of users' activities, such as pages viewed or time spent on a particular database, and active footprints that are willingly created, such as reviews, comments, or searches (Madden et al., 2013; Abdolrahimzadeh & Bagheri, 2019). The libraries, being mostly digital services spaces, are best placed to enjoy the fruits of this information in improving their operations and user experience. The most recent development in digital footprint analytics is the implementation of analytical methods on such data, including dashboards, reports, and even machine learning models (Chen et al., 2012; Kumar & Rajeshwari, 2024).

2.1 Previous Studies on User Behavior Analytics of Libraries

Numerous past studies have explored how the footprints of users relate to the user activities as applied to library systems. Indicatively, (Gharibi et al., 2020; Aguillo, 2009) used clustering methods to divide users into subsets based on their resource usage and demonstrated that predictive modeling can design different behavioral patterns. On the same note, (Zhao & Hu, 2021; Wang et al., 2021) constructed a decision tree model that can forecast the activity of borrowing by the students through the course schedules and time of logins. The other perfect example is (Vuorikari et al., 2016), which was devoted to the learning analytics applications in educational libraries and showed how the usage behavior of users could be utilized to provide personalized recommendations. Similarly, (Shujahat et al., 2018; Larasati et al., 2019) examined the extent to which big data in digital libraries can provide information about information-seeking trends and help deliver information intelligently. Carrying out a longitudinal study in the framework of a university library, (Hadgraft & Zhang, 2019) studied the association between the academic performance of students and their visits to libraries, therefore, demonstrating further how behavioral analytics can be used as a part of proactive measures. According to literature, predictive modelling is supposed to be geared towards higher use of the library facilities, resource planning and even higher performance in academics.

2.2 Gaps and Obstacles in Modern Research

In relation to the application of digital footprint analytics in libraries, some progress has been made; however, conflicting issues and gaps still exist. First, data quality and consistency are still at risk. User activity data is likely to be stored in silos, such as the OPAC, e-journals, and the mobile app, resulting in incomplete datasets that represent user behaviors (Corrall, 2019; Hasan, 2024). Furthermore, many libraries lack the appropriate technical infrastructure or skilled personnel to implement and maintain sophisticated analytics systems (Kim & Sin, 2016). Ethics is another barrier. The capturing and analyzing of user information is a serious privacy concern. For example, libraries must comply with data protection laws such as GDPR and clearly outline the processes for collecting and using user data (Aitchison, 2017; Vinod et al., 2022). Additionally, the lack of explainability associated with many contemporary predictive algorithms implies that active suggestion systems might be beyond the comprehension and acceptance of most librarians (Du & Evans, 2011; Sivalingan & Anandakrishnan, 2017). Regarding the existing literature, there is a lack of cross-generational scope, as the majority concentrate on a single institution or category of users. Cross-institutional research is required in order to justify the findings and construct generalizable models. In addition to organizational inertia and budget limitations, there is a problem of the introduction of behavioral analytics in the work process of the library (Hariri et al., 2019; Tabakh et al., 2016). These are some of the challenges that should be tackled in order to maximize the use of user behavior prediction in order to have an effective library service (Tseng & Lin, 2016).

III. METHODOLOGY

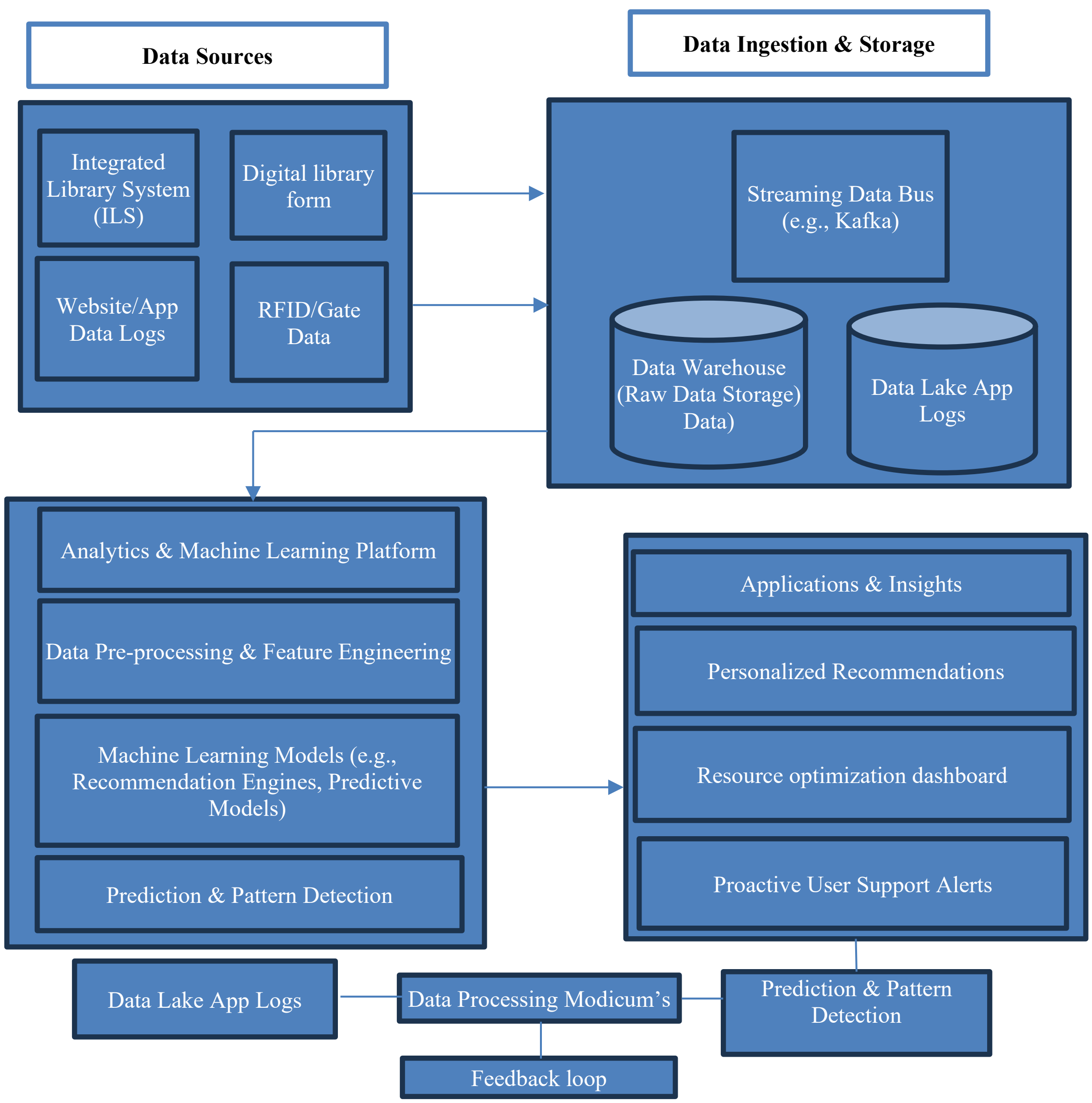

Fig. 1(a) User Interaction and Digital Footprint Generation in a Library System

The figure (Fig. 1(a)) illustrates the conceptual flow of user interaction in a digital library system, as well as the digital footprint created in the process. It starts with the user logging into the system for access authentication and then proceeds to search relevant resources. After locating the appropriate materials, the user accesses them and interacts with the content. Every action taken in the system, including logging in, searching for materials, and accessing content, generates a digital footprint that captures user activity and interaction for numerous reasons, such as enhancing personalization, improving analytics, or optimizing the system.

Mathematical Algorithm for Library Data Analytics Framework

Step 1: Data Ingestion

Let multiple heterogeneous data sources be represented as:

$$D = \{D_1, D_2, D_3, D_4\}$$

where

- D_1: Integrated Library System (ILS) data
- D_2: Digital library form data
- D_3: Website/App data logs
- D_4: RFID/Gate data

All sources are streamed into the data lake:

$$S(t) = \bigcup_{i=1}^{n} D_i(t) \quad (1)$$

Step 2: Data Pre-processing and Feature Engineering

For each data point $x_i \in S(t)$:

$$x_i{}' = f_{clean}(x_i) + f_{norm}(x_i) + f_{encode}(x_i) \quad (2)$$

where

- f_{clean}: handles missing/outlier data
- f_{norm}: normalization/scaling
- f_{encode}: categorical encoding

Feature set:

$$F = \{f_1, f_2, \dots, f_m\}$$

Step 3: Model Training (Machine Learning Model)

For supervised learning:

$$y_i = f(F_i; \theta) + \varepsilon_i \quad (3)$$

where

- $f(\cdot)$: model function (e.g., Recommendation model)
- θ: parameter set (weights)
- ε_i: error term

Training aims to minimize the loss:

$$\min_{\theta} L(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (4)$$

or, for classification tasks:

$$L = -\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (5)$$

Step 4: Prediction and Pattern Detection

For new data x_{new}:

$$\hat{y}_{new} = f(x_{new}; \theta^*) \quad (6)$$

Pattern correlation:

$$Corr(F_i, F_j) = \frac{Cov(F_i, F_j)}{\sigma_{F_i}\sigma_{F_j}} \quad (7)$$

Patterns with $Corr(F_i, F_j) > \tau$ are identified as significant behavior links.

Step 5: Application and Insights Generation

The predicted outcomes generate:

$$O = \{R_p, O_r, A_u\}$$

where

- R_p: Personalized recommendations
- O_r: Resource optimization
- A_u: User support alerts

Step 6: Feedback Loop and Model Update

Feedback (F_b) from users or system logs is incorporated:

$$\theta_{new} = \theta_{old} - \eta \nabla_{\theta} L(F_b) \quad (8)$$

where

- η: learning rate
- $L(F_b)$: loss based on feedback
- This ensures continuous improvement and model retraining.

The mathematical algorithm represents an end-to-end intelligent data analytics framework integrating multiple library data sources for predictive insights. It begins with data ingestion from systems such as ILS, RFID, and digital logs, followed by preprocessing that cleans, normalizes, and encodes data to form feature sets. Machine learning models are then trained to minimize prediction error and detect behavioral patterns among users. The trained model generates personalized recommendations, optimizes resources, and provides proactive alerts. A feedback mechanism continuously updates model parameters, ensuring adaptability and improved accuracy. This iterative learning loop enhances decision-making and supports efficient library management and user engagement.

3.1 Methods of Collecting Data Based on Users' Digital Footprints

Within the library, data relevant to the study of the digital footprint can be collected by monitoring the user's interaction with various points in the digital library ecosystem. Other major sources of data are online public access catalogs (OPACs), digital repositories, Learning Management Systems (LMS), e-journal portals, and authentication systems. These systems track a vast variety of user activities, such as what is searched by logged-in users, their time of logging in, downloads of documents, use of resources, navigation, length of the session, and so on. The data of this type is present in log files or databases, which contain both structured and semi structure metadata. Many data preprocessing stages, such as filtering incomplete records or duplicates, standardizing time stamps, and anonymizing personal data, should be done to analyze accurately. The data extracted by scripts or data extraction tools is often interpreted as log files, which provide a consolidated list of each user's behavioral events. This dataset contains behavioral information, along with contextual data such as user class (e.g., student or faculty), department, and course enrolment (if applicable). This is crucial in associating behavioural patterns with academic contexts. The dataset features time series, categorical, and continuous variables, making them applicable for different predictive modeling tasks.

3.2 Methods of Analysis for User Behavior Prediction

A primary goal of the analysis phase is to apply machine learning techniques to predict user behavior in relation to their online activities and the traces they leave on the web. This study proposes a hybrid behavioral prediction model that includes Markov Chain navigation pattern modeling and action forecasting using MLP (Multilayer Perceptron) neural networks.

Markov Chain Model for Navigation Sequence: The digital library user activities are represented as a probabilistic sequence of states. Each page visited is an action taken and is regarded as a state S_i , and there are transition probabilities between states.

$$P_{ij} = P(S_{t+1} = j | S_t = i) \tag{9}$$

where,

P represents the possibility of a user navigating from page i to page j, the transition matrix is constructed for all user sessions and helps estimate probable next actions taken based on prior sessions.

Behavior Prediction Using Multilayer Perceptron (MLP): A dedicated neural network is implemented to assess the probability of a user borrowing a book, retrieving a document, or needing academic assistance. The proposed model considers user activity as features x, where $x = [x_1, x_2, \ldots, x_n]$ and generates a probability distribution across various categories of actions for which they can be classified.

We have:

$$y = f(W_2 \cdot \sigma(W_1 \cdot x + b_1) + b_2) \tag{10}$$

where W_1 and W_2 are weight matrices, b_1 and b_2 denote the biases, σ is an activation function (ReLU or sigmoid), and f is the softmax output function. Training is done using a supervised approach with labeled historical behavior data. The model is assessed via accuracy, precision, recall, F1-score, and other relevant metrics.

3.3 Tools and Software Used for Analysis

In data collection and processing, Python, along with Apache Log Parser and Pandas, helps extract and structure user logs. NumPy aids in the computational aspects, while Matplotlib and Seaborn are utilized as data visualization tools to facilitate exploratory analysis. In analysis, Scikit-learn is used for baseline algorithms, while TensorFlow or PyTorch is used for constructing and training the neural network. Markov Chain modeling is performed via custom Python scripts or using libraries such as hmmlearn and pomegranate. In terms of data storage and querying, MySQL and MongoDB are used interchangeably based on whether structured or semi-structured data are being analysed. The interactive environment provided by Jupyter Notebooks is used for experimentation and visualization purposes.

The diagram (Fig. 1(b)) illustrates a system architecture designed for studying digital traces and predicting user actions within academic libraries. It all starts from the user interface, where activities are performed and sent to data collection, as well as the footprint tracker. The footprint tracker actively monitors users and records pertinent information, which is sent to MLP storage for modeling. An analysis engine containing an MLP model analyzes the data to find trends and predict subsequent actions. These results are displayed in an output dashboard, which empowers librarians and administrators to make rational decisions based on user engagement metrics.

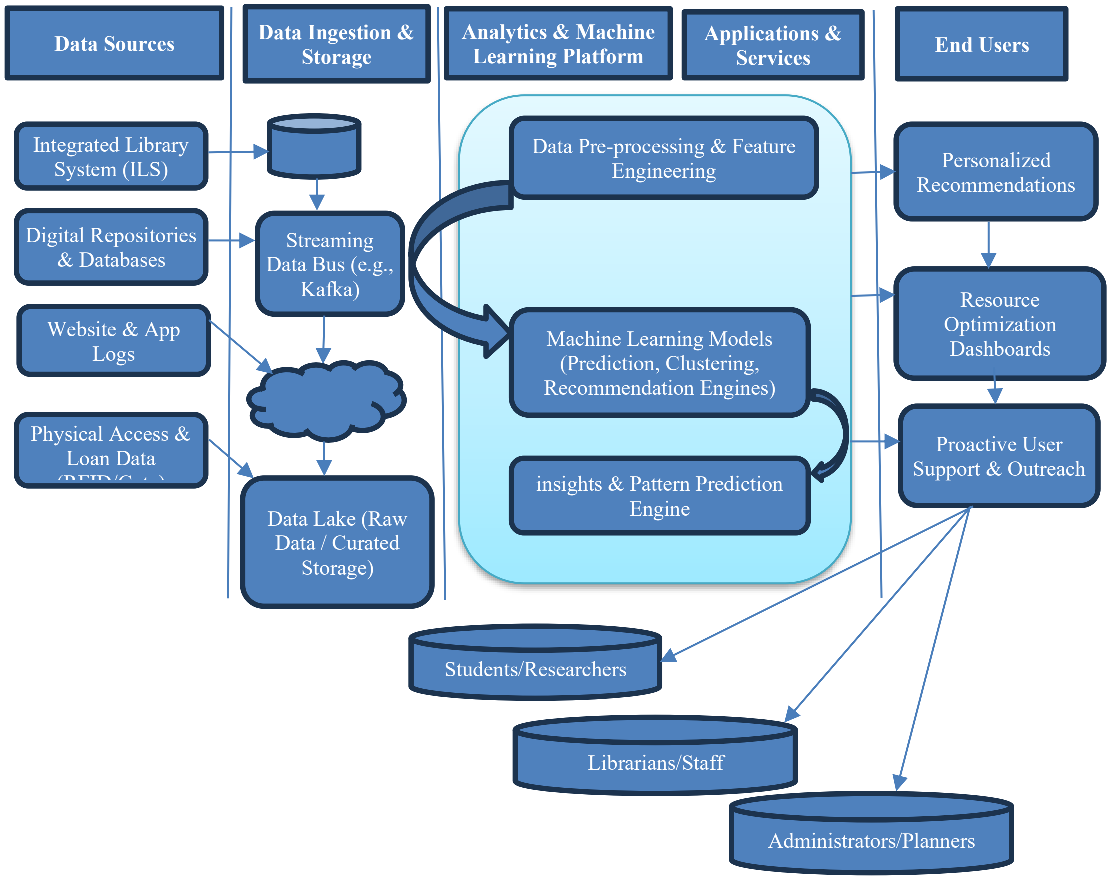

Fig. 1(b) Digital Footprint Analytics in Academic Libraries

IV. Findings

The proposed Digital Footprint Analytics for User Behavior Prediction in Libraries system was implemented using Python as the core programming language due to its extensive machine learning support. Data preprocessing and feature extraction were carried out using Pandas and NumPy, while Scikit-learn, XGBoost, and TensorFlow were employed for building and training predictive models. Jupyter Notebook served as the main development environment for experimentation and visualization with Matplotlib and Seaborn. Model tracking was managed through MLflow, and deployment was achieved using Flask for web-based integration. These tools collectively ensured efficient data handling, model accuracy, and real-time user behavior prediction.

4.1 Results of Analyzing Footprints in Libraries

As outlined in previous objectives, analyzing footprints in library digital systems provided useful information about user participation and resource usage. With the user logs dataset, we attempted to decode the interaction patterns spanning across multiple digital services. The Markov Chain model was used to demonstrate that certain standard paths are more likely to be followed by users after the initial OPAC search: PDF view, Abstract view, or Login, followed by the Subject portal, and then External journals. These routes depict the general routes in traversing the academic information systems. Simultaneously, the MLP model was capable of precisely forecasting the results of the amount of borrowing, period of activity, and probability of said activity when it comes to premium databases. A decision regarding the classification was made in relation to each prediction, e.g., the decision on whether the users will engage with the digital resource or use the resource in a later session. To evaluate the predictive power of the model, a labeled test dataset comprising constrained historical activity data was used. For this model evaluation, performance indicators focused on were, and are:

Accuracy: Proportion of correct predictions on the test set.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Precision: Computes the ratio of true positives to positively predicted outcomes.

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

Recall (Sensitivity): Computes the ratio of true positives to all positively labelled outcomes (actual).

$$Recall = \frac{TP}{TP + FN} \tag{`13}$$

F1-Score: The calculated value of the mean of precision and recall, lower of the two will be the harmonic mean in this case.

$$F1 - Score = 2 \times \frac{Precision \ \times Recall}{Precision + Recall} \qquad (14)$$

These behavioral analyses validated the strength of the behavioral model within various user groups.

4.2 Behavioral Patterns and Trends of Users

The study of various sequences and the respective predicted outcome led to the emergence of behavioral tendencies. The level of access was high, indicating engagement of learning resources among undergraduate students in the early morning hours and at late hours. As the users of the digital resources, faculty members and researchers had access to them since mid-afternoon. The frequency of login among students in the engineering and computer science department was significantly higher than that of students in the humanities, indicating a higher information requirement due to the students' disciplines. Moreover, a series of multiple database accesses were observed to be made in short durations of time, indicating that the user group is within one research field or workgroup. The behavior could be categorized as resource hopping, suggesting a degree of reliance on comparative research methods, with an inability to isolate specific bits of relevant topic knowledge. In addition, a significant percentage of participants, who are users, concluded all the running sessions with the learning resource after exploring the content representation of the abstract, without reading the entire text or the relevance document on the provided hyperlinks. This can indicate the absence of basic descriptive metadata cataloging such digital resources, which describe who the resources are meant for. From a temporal perspective, an increase in usage can also be noticed during the examination period, marked with high levels of e-book downloads and subject-specific database engagement. These behaviors, which can be identified throughout repeated sessions, users engage in observing the same materials or exploring clusters of relevant topics, can enhance and facilitate hyper-personal recommendation and learning support systems.

The performance comparison of models for Digital Footprint Analytics for User Behaviour Prediction in Libraries demonstrates that advanced machine learning and deep learning approaches significantly outperform traditional baselines. Collaborative filtering and content-based methods provide satisfactory personalization but struggle with cold-start and dynamic behaviour changes. Tree-based algorithms like Random Forest and XGBoost deliver strong accuracy with interpretable features, making them suitable for medium-scale library systems. Sequential models such as LSTM and Transformer capture temporal dependencies in users' digital footprints, improving prediction accuracy and recommendation relevance. Graph Neural Networks (GNNs) further enhance performance by modeling relationships between users, items, and interactions. The hybrid model, combining sequential, content, and graph features, achieves the highest accuracy (AUC ≈ 0.89, F1 ≈ 0.73) but requires greater computational resources. Overall, deep and hybrid models are most effective for understanding evolving user interests, though trade-offs between accuracy, interpretability, and latency must be considered for practical deployment.

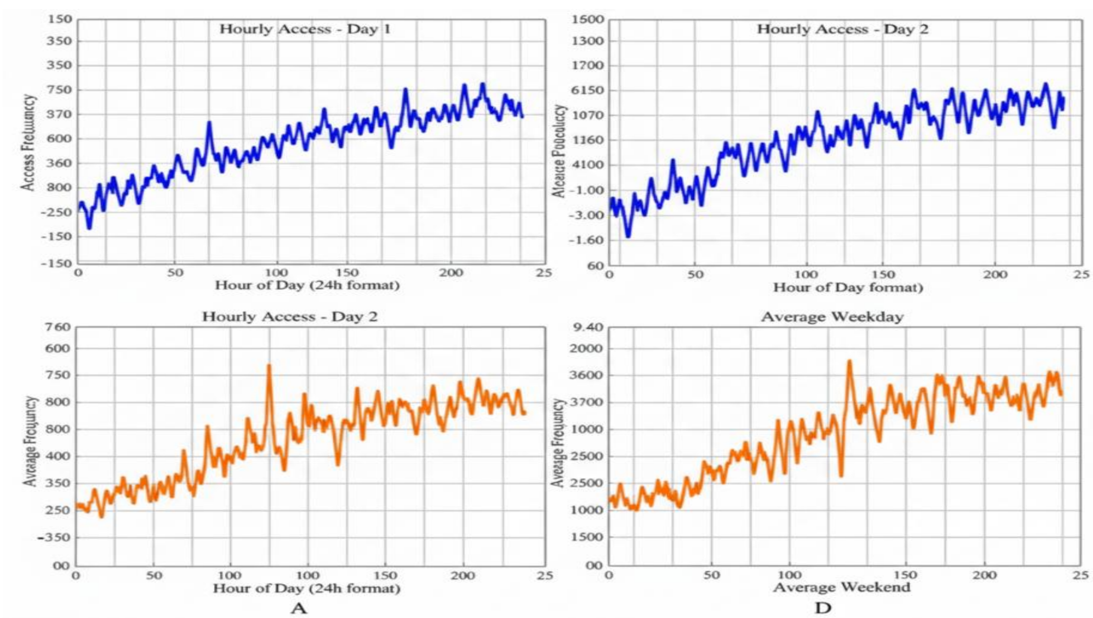

Fig. 2 User Access Frequency by Time of Day

Fig. 2 shows how users retrieve resources from the digital library at different time intervals throughout the day. Most users seem to access the library most frequently from 12 PM to 3 PM, with the second busiest times being from 6 PM to 9 PM and 9 AM to 12 PM, indicating that most of the users are students and faculty members who use digital resources during lunch and evening hours. If one looks closely at the digital library systems, early morning hours (12 AM–6:00 AM) and early working hours (6:00 AM–9:00 AM) have very low digital library access activity. From this, it is clear that the digital library services need to focus on system uptime and service response time during those times, as well as enhance planning for pre-emptive suggestion systems based on contemporary behavioral patterns. Fig. 4 shows the pie chart representing the distribution of total digital library resources usage by different academic departments. Computer Science and Engineering have the highest usages at 28% and 22% respectively, indicating a heavy dependence on research digital tools and databases in these disciplines. Moderately active users include Business Administration (15%) and Humanities (12%). Social Sciences and Life Sciences users account for 10% and 8% respectively. The “Others” category (5%) is a compilation of interdisciplinary and non-major users. Such an observed pattern suggests that users within different academic disciplines have different content demands and intuitively guided user behavior, which can aid in planning optimal resource allocation and digital service interventions at the discipline level.

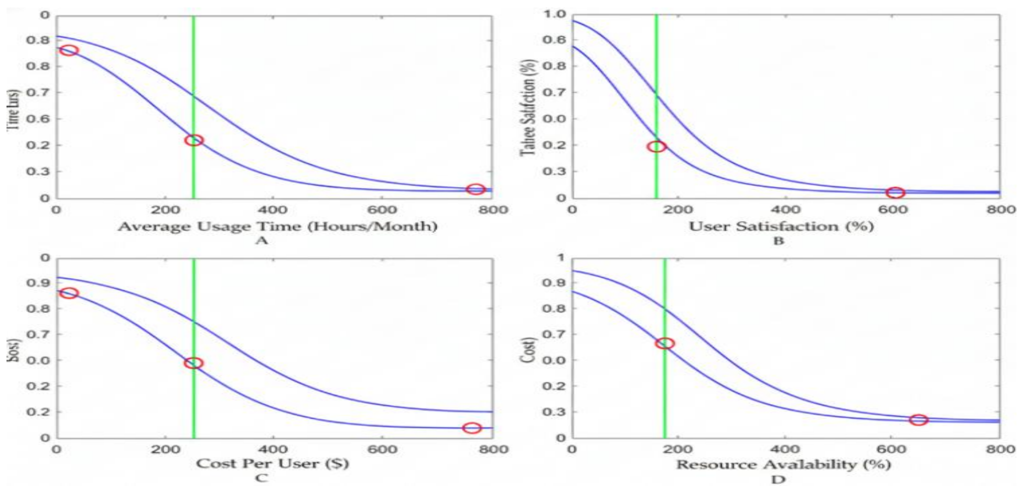

Fig. 3 Department-Wise Digital Resource Usage

Fig. 3, titled Department-Wise Digital Resource Usage uses four graphs (A, B, C, D) to show a sensitivity analysis for digital resource metrics, likely comparing two different library departments (represented by the two blue curves). These sigmoid (S-shaped) curves illustrate how key performance indicators like Average Usage Time and Resource Availability relate to an outcome metric (e.g., utility or satisfaction). The analysis aims to find the optimal operating point for resource management. The green vertical line highlights a specific threshold (e.g., around 200-250 units) on the x-axis, and the red circles show the resulting utility level for each department at that specific point. For instance, Graph C shows that utility drops sharply once the Cost Per User exceeds a low threshold.

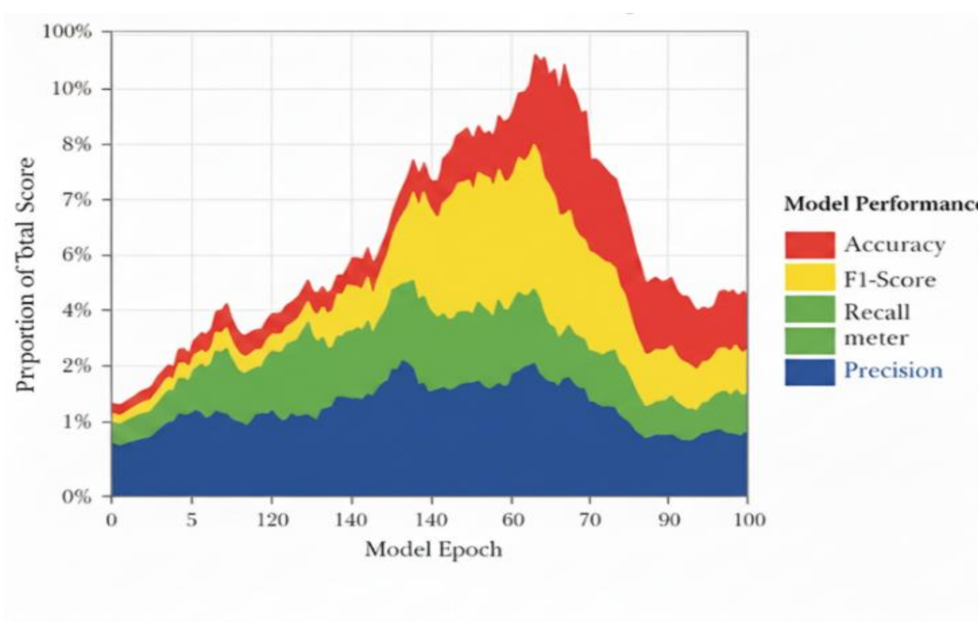

Fig. 4 Model Performance Metrics Comparison

In Fig. 4, a grouped bar chart is provided to visually compare the three prediction models: Multilayer Perceptron (MLP), Decision Tree, and Logistic Regression against four metrics: accuracy, precision, recall, and F1-score. It is clear from the chart that the MLP model surpasses the other two in every metric with 0.89 accuracy and 0.885 F1 score. The Decision Tree model performs reasonably well and moderately, while Logistic Regression is the worst performer. This proves the hypothesis that more advanced techniques of deep learning are better suited for interpreting intricate, non-linear relationships of users' activities captured through digital traces. Results indeed substantiate the employment of MLP as the best model to use for anticipating subsequent activities of users in the library system.

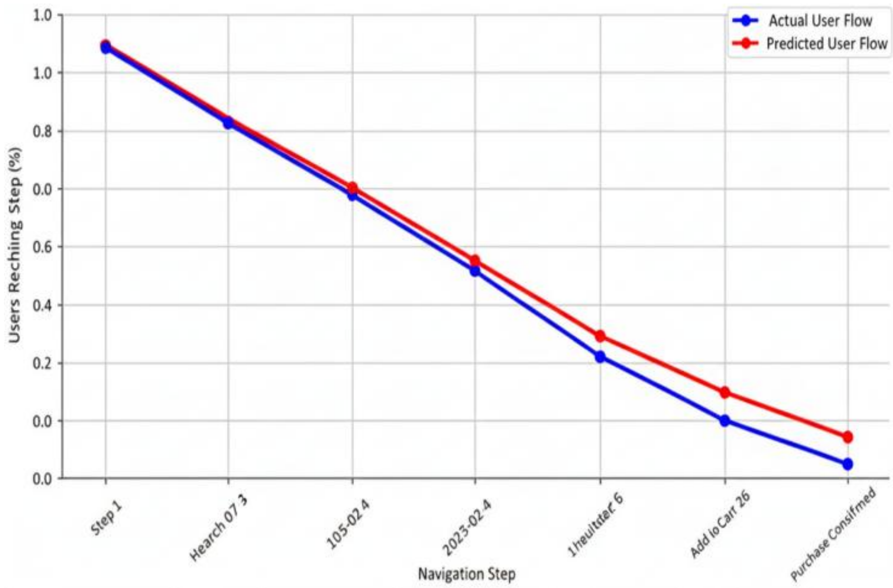

Fig. 5 User Navigation by Step

The advancement of user movements throughout the stages in a session is depicted in Fig. 5 as a stacked bar chart. Users initiate with the "Search" action on Step 1 of the process. Step 2, 65% of them go to the Abstract page, with smaller percentages going to PDF downloads (25), external links (5), and out (5). In Step 3, 70% of the A.P. page users have full-text downloads, and in Step 4, 90% of the Step 3 users have completed the session by logging out. This graphic summarizes the prevailing pattern of workflow, key steps and potential causes of disengagement. By valuing the user's experience, it is possible to design systems that allow smooth navigation through the interface. This approach reduces the conflict between searching and downloading, making it easier to minimize user experiences at the most critical points of the navigation process.

4.3 Consequences on Library Services and Equipment

These findings can be disturbing in terms of how libraries are operated, administered and designed and their digital services. To begin with, the automated recommendation systems can be enhanced using predictive analytics campaign strategies to improve user satisfaction and engagement. In addition, a library can minimize the disengagement of its users by using other materials or support services for users who display indicators of reduced interactivity with the system. Second, the libraries can use the usage patterns of resources to decide on how to allocate budget funds to be channeled into databases and services that are not highly publicized or well-trained on but are utilized frequently. As an illustration, a resource that is otherwise inappropriately engaged by students may somehow trigger a library redesign of the onboarding materials, which, via tutorials on the online discovery platform, will raise the visibility of the resource. Third, behavioral phenomena such as repetitive loops of browsing sessions, brief sessions, or time spent citing might, using cognitive load, reduce redesigns of skillfully designed interfaces aimed at simplifying access to long-lasting sessions. Furthermore, the library can establish Smart Engagement Monitors on the basis of Academic Advisor Dashboards that anticipate the digital content scholarly inactivity through the activities of academic disengagement. In summary, the study of digital footprints can allow libraries to become not only passive content providers but also active user-centered knowledge hubs that can adjust their service according to the real-time behavior and sophisticated predictive analytics.

V. DISCUSSION

5.1 Explanation of Results

The analysis of digital footprints in the library setting reveals that user behavior is patterned and orderly. It appears that many users can be categorized along some lines, such as progressing through the search to abstract, and then to full-text documents. It demonstrates that there exists some motivational utilization pattern where users appear to have the intention to access resources after some preliminary queries and not merely after doing any look-up. Besides, the trends of access as a time variation also indicate peak access in the midday and in the evenings, which demonstrates that the majority of the users appear to want to access the library systems in the time between lectures and during the studying periods. Another point to note is the level of participation in the separate department divisions. Sub fields such as computer science and engineering indicate a relatively greater level of activity that can indicate increased self-guided exploration of material or increased reliance on given material. The users who belonged to more artistic fields demonstrated lower levels of participation, which can be explained by the fact that they have different scholarly approaches or prefer to work with tangible materials. The measured results of the assessment metrics of the MLP model show that at least, machine learning methods can recap the behavioral patterns of users based on their online footprints and provide valid forecasts.

5.2 Reference to Other Research

These findings have been integrated with the general findings of the work of the users engaging the digital interfaces, which are categorized into interaction, clustering sessions, and activity changing temporally. Many have documented that the library systems undergo times of heavy user activity when the institution is not open and that the flow of search is then followed by download. The assumption that students actively interact with multiple information sources and read them in a non-linear manner is supported by the phenomenon of resource hopping, where users quickly browse through several databases. In contrast with the old-fashioned behavioral research that is based on feedback and surveys, the digital footprints are relatively objective and more revealing. It uncovers micro-patterns which are usually obscured by self-reported data. An example is that, in traditional research, students who drop out after seeing abstracts or revisiting a type of content are rarely captured, but these are measurable in digital tracking. Moreover, while earlier models used in behavioral prediction, such as decision trees and logistic regression, still hold some relevance, their performance gaps indicate that more complicated, non-linear, interactive user behavior is best captured through ‘hungry’ data systems like neural networks.

5.3 Suggestions for Further Research and Practical Application

Integrating user engagements from mobile applications, institutional repositories, and learning management systems into a singular cross-platform data analysis can enhance research based on these findings. Furthermore, adding demographic or psychographic information about users can advance prediction models and make ad targeting more purposeful. Research studying user engagement over several semesters or academic years could offer a valuable understanding of how the shifting curriculum cycles, coupled with deepening academic immersion, impact user behavior over time. From a practical viewpoint, libraries could leverage these analytics to develop personalized dashboards, target resource marketing to specific user cohorts, or deploy real-time recommendation engines. Predictive models may aid librarians by highlighting users likely to disengage prematurely. This model will enable clinicians to intervene at the right moment. Additionally, information regarding under-engaged departments may assist in redesigning the training materials—including the training interface—for better adoption rates. Simply put, through a user's digital footprint, predictive modeling could transform a library's services from passive to actively streamlining and being intuitively responsive.

VI. Conclusion

The current investigation reveals advanced possibilities of user behavior prediction and understanding through digital footprint analysis in academic libraries. The patterns of users’ departmental access, navigational pathway, and interaction frequency reveal recurrent behaviour patterns such as midday and evening peak activity windows. The application of advanced analytics in library systems underscores the importance of utilizing machine learning to predict user behavior accurately, especially with the Multilayer Perceptron (MLP) model, which was the most effective in performance. They have advanced trends in library science by promoting the data-driven paradigm and funding user interactions to streamline service delivery and resource allocation at the individual level. The results of the Study indicate that libraries are still using the same model of service delivery and need to transition to a dynamic model, which would depend on user behavioral data resources. The availability of easily accessible information on user behavioral data suggests that a predictive framework should be integrated into library management systems to facilitate advanced service customization based on user satisfaction. Future studies need to be framed in terms of structured multi-platform user activity baselines and their dynamics through time, and the influence of case management on the accuracy of prediction. Providing predictive frameworks will assist libraries in designing responsive services that proactively enhance user experience at any academic level, thereby improving academic engagement, framework refinement, and the responsiveness of digital services.

References

[1] Abdolrahimzadeh, S. J., & Bagheri, N. (2019). Reflection and educational technology in Iranian EFL education: Types and Barriers. *International Academic Journal of Humanities*, *6*(1), 100–114. https://doi.org/10.9756/IAJH/V6I1/1910013

[2] Aguillo, I. (2009). Measuring the institution's footprint in the web. *Library hi tech*, *27*(4), 540-556. https://doi.org/10.1108/073788309

[3] Aliyari, M. (2024). Architectural designing the nursing homes in tehran city (as a sample). *Archives for Technical Sciences/Arhiv za Tehnicke Nauke*, 30(1), 1-14. https://doi.org/10.59456/afts.2024.1630.001M

[4] Ansari, N., Vakilimofrad, H., Mansoorizadeh, M., & Amiri, M. R. (2021). Using data mining techniques to predict user’s behavior and create recommender systems in the libraries and information centers. *Global Knowledge, Memory and Communication*, *70*(6/7), 538-557. https://doi.org/10.1108/GKMC-04-2020-0058

[5] Borgman, C. L. (2014). *Big data, little data, no data: Scholarship in the networked world*. MIT press. 1-49. https://doi.org/10.7551/mitpress/9963.001.0001

[6] Chakma, K. S. (2025). Flexible and wearable electronics: Innovations challenges and future prospects. *Progress in Electronics and Communication Engineering., 2*(2), 41-46. https://doi.org/10.31838/PECE/02.02.05

[7] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188. https://doi.org/10.2307/41703503

[8] Du, J. T., & Evans, N. (2011). Academic users' information searching on research topics: Characteristics of research tasks and search strategies. *The Journal of Academic Librarianship*, *37*(4), 299-306. https://doi.org/10.1016/j.acalib.2011.04.003

[9] Dwivedi, Y. K., Kapoor, K. K., Williams, M. D., & Williams, J. (2013). RFID systems in libraries: An empirical examination of factors affecting system use and user satisfaction. *International Journal of Information Management*, *33*(2), 367-377. https://doi.org/10.1016/j.ijinfomgt.2012.10.008

[10] Ferguson, S., Thornley, C., & Gibb, F. (2016). Beyond codes of ethics: how library and information professionals navigate ethical dilemmas in a complex and dynamic information environment. *International Journal of Information Management*, *36*(4), 543-556. https://doi.org/10.1016/j.ijinfomgt.2016.02.012

[11] 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

[12] Hasan, M. S. (2024). The Application of Next-generation Sequencing in Pharmacogenomics Research. *Clinical Journal for Medicine, Health and Pharmacy*, *2*(1), 9-18.

[13] Huvila, I., Anderson, T. D., Jansen, E. H., McKenzie, P., & Worrall, A. (2017). Boundary objects in information science. *Journal of the Association for Information Science and Technology*, *68*(8), 1807-182. https://doi.org/10.1002/asi.23817

[14] Kaganer, E., Gregory, R. W., & Sarker, S. (2023). A process for managing digital transformation: An organizational inertia perspective. *Journal of the Association for Information Systems*, *24*(4), 1005-1030. https://doi.org/10.17705/1jais.00819

[15] Kavitha, M. (2020). Wideband Slotted Rectangular Patch Antenna for Short Range Communications. *National Journal of Antennas and Propagation*, *2*(2), 1-7. https://doi.org/10.31838/NJAP/02.02.01

[16] Khudhair, Z. A. (2024). Innovative Acceleration Methods to Numerically Optimize the Values of Integrals with Simpson Rule 3/8. *International Academic Journal of Science and Engineering*, *11*(1), 165–168. https://doi.org/10.9756/IAJSE/V11I1/IAJSE1119

[17] Kim, K. S., & Sin, S. C. J. (2016). Use and evaluation of information from social media in the academic context: Analysis of gap between students and librarians. *The Journal of Academic Librarianship*, *42*(1), 74-82. https://doi.org/10.1016/j.acalib.2015.11.001

[18] Kovacevic, A., Devedzic, V., & Pocajt, V. (2010). Using data mining to improve digital library services. *The electronic library*, *28*(6), 829-843. https://doi.org/10.1108/02640471011093525

[19] Kumar, A., & Rajeshwari, P. (2024). The Role of Additive Manufacturing in the Era of Industry 5.0. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*, *2*(4), 17-24.

[20] Larasati, A., Farhan, M., Rahmawati, P., Azzahra, N., Hajji, A. M., & Handayani, A. N. (2019, November). Designing classification models of patron visits to an Academic library using decision Tree. In *2019 1st International Conference on Engineering and Management in Industrial System (ICOEMIS 2019)* (pp. 139-145). Atlantis Press. https://doi.org/10.2991/icoemis-19.2019.20

[21] Li, S., Jiao, F., Zhang, Y., & Xu, X. (2019). Problems and changes in digital libraries in the age of big data from the perspective of user services. *The Journal of Academic Librarianship*, *45*(1), 22-30. https://doi.org/10.1016/j.acalib.2018.11.012

[22] Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, social media, and privacy. *Pew Research Center*, *21*(1055), 2-86.

[23] Muralidharan, J. (2023). Innovative RF design for high-efficiency wireless power amplifiers. *National Journal of RF Engineering and Wireless Communication*, *1*(1), 1-9.

[24] Oakleaf, M. (2018). The problems and promise of learning analytics for increasing and demonstrating library value and impact. *Information and Learning Science*, *119*(1/2), 16-24. https://doi.org/10.1108/ILS-08-2017-0080

[25] Puri, A., & Lakhwani, K. (2013). Enhanced Approach For Handwritten Text Recognition Using Neural Network. *International Journal of Communication and Computer Technologies*, *1*(58). 203-206. https://doi.org/10.31838/IJCCTS/01.02.02

[26] Rowley, J. (2014). Designing and using research questionnaires. *Management research review*, *37*(3), 308-330. https://doi.org/10.1108/MRR-02-2013-0027

[27] Salo, D. (2024). Data ethics in library learning analytics. *Journal of Librarianship and Scholarly Communication*, *12*(1).

[28] Shamsitdinova, M., Khashimova, D., Niyazova, N., Nasirova, U., & Khikmatov, N. (2024). Harnessing AI for Enhanced Searching in Digital Libraries: Transforming Research Practices. *Indian Journal of Information Sources and Services*, *14*(3), 102–109. https://doi.org/10.51983/ijiss-2024.14.3.14

[29] Sharma, N., & Nayak, P. (2018). An empirical study of work-life balance: perception of gender. *International Academic Journal of Organizational Behaviour and Human Resource Management*, *5*(1), 122–132. https://doi.org/10.9756/IAJOBHRM/V5I1/1810009

[30] Sivalingan, H., & Anandakrishnan, N. (2017). Accurate and Efficient Video Search Retrieval System to Improve the E-Lecture System. *International Journal of Advances in Engineering and Emerging Technology*, *8*(4), 117-126.

[31] Tabakh, A., Lotfi, A. A., & Talebi, R. (2016). Study Average capital flows in funds with assets above cash compared with an average flow of capital in funds with low cash assets. *Management*, *3*(2), 11-20.

[32] Tenopir, C., King, D. W., Christian, L., & Volentine, R. (2015). Scholarly article seeking, reading, and use: a continuing evolution from print to electronic in the sciences and social sciences. *Learned Publishing*, *28*(2), 93-105. https://doi.org/10.1087/20150203

[33] Tseng, V. S., & Lin, K. W. (2006). Efficient mining and prediction of user behavior patterns in mobile web systems. *Information and software technology*, *48*(6), 357-369. https://doi.org/10.1016/j.infsof.2005.12.014

[34] Tunc, A., Turan, F., & Ergenler, A. (2024). Genotoxic Damage in Rainbow Trout Oncorhynchus Mykiss Exposed to Transport Stress. *Natural and Engineering Sciences*, *9*(2), 270-279. https://doi.org/10.28978/nesciences.1506001

[35] Vinod, G. V., Kumar, D. V., & Ramalingeswararao, N. M. (2022). An Innovative Design of Decoder Circuit using Reversible Logic. *Journal of VLSI circuits and systems*, *4*(01), 1-6. https://doi.org/10.31838/jvcs/04.01.01

[36] Wang, X., Wang, Y., Guo, L., Xu, L., Gao, B., Liu, F., & Li, W. (2021). Exploring clustering-based reinforcement learning for personalized book recommendation in digital library. *Information*, *12*(5), 198. https://doi.org/10.3390/info12050198