Information Research - Vol. 30 No. CoLIS (2025)

Advancing library operations with AI: data-driven insights for academic resource management

Néstor A. Nova, Hernán Morales, Juan Pájaro, Andrea González

DOI: https://doi.org/10.47989/ir30CoLIS52261

Abstract

Introduction. The rapid evolution of artificial intelligence technologies affords opportunities and challenges for libraries. The study analyses the application of artificial intelligence tools for business intelligence purposes in a university library.

Method. This study used artificial neural networks to extract metadata from a syllabus corpus and then applied a string-matching model to integrate the extracted data with the loan library database. Finally clustering algorithms were employed to analyse the results, providing valuable insights into resource usage patterns. Data was extracted from faculty databases from one university in Colombia.

Results. This study identified the potential of integrating artificial intelligence with business intelligence tools to enhance resource management and optimise university library operations, facilitating a better alignment between academic syllabi and available materials.

Conclusions. The study found that artificial intelligence tools are valuable for university libraries in optimising processes based on data analysis. This suggests that libraries should design and implement business intelligence initiatives to automate manual tasks, providing valuable information to managers and academic directors for decision-making in administrative and academic contexts.

Introduction

Artificial intelligence (AI) has rapidly become an integral part of our digital lives. Its capacity to simulate human intelligence, particularly in pattern recognition and decision-making, has transformed various sectors such as business, management, medicine, military, education, gaming, libraries, etc. The idea of creating AI systems in libraries dates back to 1990. These intelligent library systems provide knowledge-based services to both the library staff and patrons (Asemi and Asemi, 2018). While AI's origins can be traced back to computer science research focused on mimicking human behaviour, its applications have expanded exponentially. Today, AI powers everything from smartphones to complex industrial systems. Libraries, traditionally at the heart of information dissemination, are also undergoing a transformation driven by AI. Library management involves tasks that are often repetitive and time intensive. To enhance efficiency and effectiveness, many libraries are adopting automation for their operations (Dwivedi et al., 2013).

The growing demand for information and the rapid evolution of digital technologies have necessitated a corresponding adaptation in library services. As part of ongoing technological advancements, libraries are continuing to explore AI applications to enhance operations and user experiences, building on a long history of automation in the field. The gap between the vast amount of data and information they manage and their ability to harness it effectively using advanced technologies presents a significant challenge for librarians. Therefore, libraries risk becoming obsolete in this digital era unless they adopt modern technologies and innovate in the delivery of their information and services (Lund et al., 2020). This study is grounded in business intelligence (BI) as a theoretical framework, which conceptualises how data analytics can inform decision-making in organisational contexts (Hamad et al., 2021). BI integrates principles such as data collection, integration, analysis and visualisation to transform raw information into actionable insights. These principles rely on key variables, including data quality, accessibility, analytical models and performance metrics, which enable organisations to optimise operations and strategic planning. In university libraries, applying BI involves leveraging structured and unstructured data – such as circulation records, digital access logs and user preferences – to enhance resource management and align services with academic needs. By harnessing the potential of BI, libraries can enhance their services, optimise resource allocation and demonstrate their value to the academic community (Okunlaya et al., 2022).

There are several reasons why it is relevant to study the support of BI and IA to library services optimisation. In the context of university libraries, the task of classifying and organising bibliographic references is both complex and highly significant within the business intelligence and operational management activities conducted by each institution (Hamad et al., 2021). The demand for data analytics in libraries has increased to support data-driven decision-making processes, such as acquisition, curation and updating of bibliographic resources, budgeting, investments and more (Adetayo et al., 2021). These decisions are correlated with academic management activities, including the design, monitoring, updating and/or evaluation of syllabi.

This study aims to optimise the management of bibliographic resources in university libraries by leveraging AI and BI, ensuring alignment and consistency between course syllabi and available library resources, thereby improving resource and academic management efficiency. However, this relationship is intricated, due to the vast amount of data associated with bibliographic resources (Zhang et al., 2024), and it is hampered by the lack of automation in the process of identifying, extracting and structuring bibliographic references within each syllabus, which are often written in unstructured format and without following a specific format, such as APA, IEEE and others.

To address the above gaps, the present study follows the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology to develop two IA algorithms. One enables the agile and efficient automation of data extraction activities, reducing operational costs and the lag between updating information in syllabi and decision-making in library operational processes. The other employs a clustering model to categorise bibliographic resources utilised by two faculties of one university based on the library's loan data, aiming at optimising resource allocation and informing effective resource management.

This research examines how AI tools can support education management initiatives, including quality assurance, accreditation processes, curricular innovation and the internationalisation of curricula, among others. But also libraries can also take advantage of this technologies for ensuring the availability of resources for students and professors and for the optimisation of process and services delivered, for instance, in budgeting activities which implies negotiation with suppliers and editorials, or making group purchases for several faculties using the same resources, or maybe for gathering better discounts.

To achieve these objectives, we employ a case study research design focusing on two faculties of one university in Colombia. These faculties were selected due to their high concentration of academic resources and volume of resource loans at the library, making them pivotal hubs of academic activity within the whole institution. By analysing and comparing the characteristics of the two faculties, we can elucidate the specific factors that influence the data-based decision-making processes at the academic and business extent.

The theoretical framework, including previous research findings and theoretical perspectives, is discussed in detail. The methodology employed in this study is described thoroughly for each algorithm. The results and their implications are presented, followed by a critical discussion of the findings and their limitations. Finally, the paper concludes with recommendations for decision makers at library and academic programmes to enhance the education management activities.

Theoretical background

Strategic, tactical and operational use of AI in university libraries

Artificial intelligence refers to the ability of machines to simulate human intelligence, enabling them to perform tasks such as learning, problem-solving, and pattern recognition (Kotu and Deshpande, 2018). Rooted in computational science, AI encompasses a range of techniques, including machine learning, natural language processing (NLP) and computer vision, among others, which collectively empower systems to process complex datasets and make informed decisions. Its applications have permeated industries like healthcare, education and library sciences, where data-centric operations are crucial.

AI is increasingly used in university libraries to support decision-making across strategic, tactical and operational levels (Adetayo et al., 2021; Hamad et al., 2021; Jayavadivel et al., 2024; Okunlaya et al., 2022). The integration of AI into library operations not only enhances resource management but also optimises service delivery, aligning library functions with the broader goals of higher education institutions. The implementation of AI algorithms in libraries has revolutionised decision-making processes, enabling data-driven strategies to enhance both operational efficiency and user experience. These algorithms offer libraries robust tools to address complex challenges and adapt to the evolving demands of academic communities.

At the strategic level, AI can help librarians to support long-term planning and decision-making by analysing vast datasets to identify trends in information consumption, user preferences and research needs (Meesad and Mingkhwan, 2024). Predictive analytics enables libraries to anticipate shifts in academic disciplines and allocate resources accordingly. By identifying patterns in borrowing histories, digital access logs and user demographics, libraries can anticipate future needs and make informed decisions about acquisitions, budget allocations, space utilisation, underused areas and staffing requirements, among others (Pence, 2022). For instance, AI algorithms can forecast demands for specific collections (Barsha and Munshi, 2023), facilitating data-driven budget allocation and the development of future-oriented digital repositories. Furthermore, AI aids in aligning library goals with institutional priorities, such as enhancing student success (Al Fraidan, 2024). By grouping users with similar behaviours or needs, clustering algorithms help libraries to segment their audience for targeted services (Wang et al., 2011). For example, libraries can identify clusters of undergraduate students who require foundational resources versus faculty members who need specialised materials. Classification algorithms, on the other hand, assist in automated content categorisation, enabling more efficient organisation and retrieval of information (Echedom and Okuonghae, 2021).

On the tactical level, libraries can use AI-driven tools for advanced cataloguing, metadata enrichment and automated indexing, ensuring seamless integration of physical and digital resources (Ahmed et al., 2023; Das and Islam, 2021) and generating more accurate and comprehensive descriptive information, improving discoverability and retrieval (Barsha and Munshi, 2023). These tools enable libraries to optimise workflows, enhance interoperability with other academic systems and improve the discoverability of resources. Additionally, AI systems can analyse user feedback and interaction data to refine service offerings, such as personalised recommendations and adaptive learning platforms tailored to the needs of diverse user groups (Deo et al., 2020).

At the operational level, AI can potentially enhance the day-to-day functioning of library services. Automated systems manage routine tasks such as inventory management, overdue notifications and real-time updates of book availability, thereby reducing the workload on library staff (Jayawardena et al., 2021; Pence, 2022). Chatbots and virtual assistants powered by NLP provide instant support to users, addressing queries related to research assistance, database access or citation guidance (Bagchi, 2020). Furthermore, machine learning models jointly with the Internet of Things (IoT) can be employed to detect patterns of resource misuse or optimise space utilisation within library facilities (Khan et al., 2022).

This multidimensional application of AI highlights its potential to not only modernise but also redefine the role of university libraries in the knowledge economy (Barsha and Munshi, 2023; Das and Islam, 2021; Pence, 2022; Priya and Ramya, 2024). By integrating AI into workflows, libraries can become hubs of innovation, enhancing research capabilities, enabling interdisciplinary collaboration and advancing digital literacy. This requires coordinated efforts among professionals, researchers and policymakers to ensure ethical, equitable and sustainable implementation.

Despite its potential, AI adoption in university libraries presents critical challenges. A key concern is data bias, as AI models rely on datasets that may reinforce inequalities in information access (Khan et al., 2023). Automation also risks diminishing human expertise in collection development and user support, potentially affecting service quality (Pence, 2022). Moreover, AI's predictive capabilities are often overestimated – it can identify trends but lacks contextual understanding of academic needs, limiting its role in decision-making (Harisanty et al., 2023). Ethical concerns around data privacy further complicate its implementation, requiring strict regulatory compliance. These limitations highlight that while AI can optimise processes, it must complement rather than replace human judgement in academic environments.

The value of business intelligence in library operations and services

The evolution of ICT tools and services is reshaping the processes through which organisational decisions are made, significantly enhancing strategic planning capabilities, and improving organisational efficiency (Gauzelin and Bentz, 2017). This transformation also necessitates that libraries and information centres innovate and adapt to align with the dynamic demands of a rapidly evolving digital society. In this context, libraries and information science professionals are embracing change and acquiring new knowledge to maintain their relevance and effectiveness

(Vijayalatha, 2023). The integration of data and analytical methodologies is redefining the roles and functions of libraries in this shifting landscape.

Academic libraries, as the heart of the university and main pillars of educational and research activities, use database technologies to facilitate services (Long and Wu, 2012). The amount of library data is growing exponentially, encompassing structured, semi-structured and unstructured formats. This vast volume exceeds the processing capacity of traditional databases, necessitating innovative approaches. BI has emerged as a critical field, analysing large datasets to uncover hidden patterns, correlations, and insights, transforming raw data into actionable information (Balachandran and Prasad, 2017). Building on the transformative potential of AI in university libraries, BI enacts a complementary framework to enhance data-based decision-making processes across operational and service-oriented domains (Adetayo et al., 2021; Asemi and Asemi, 2018; Hamad et al., 2021).

BI consists of analytics tools and methods that can be applied to any organisational context for gathering and analysing its information to provide support for decision-making processes (Hamad et al., 2021). By leveraging data integration, analytics and visualisation tools, BI provides actionable insights that empower libraries to optimise their operations and deliver superior services tailored to user needs. In operational processes, BI tools enable libraries to manage resources more effectively through the integration and analysis of data from various systems, such as circulation records, digital repositories and user management platforms (Adetayo et al., 2021; Hamad et al., 2022). For instance, when analysing book circulation, data mining can provide scientific and managerial data to enhance the service quality of the library (Long and Wu, 2012). In addition, dashboards and reporting tools provide library staff with real-time metrics on resource utilisation, workflow efficiency and inventory trends (Briggs, 2020). These insights help to streamline activities such as procurement, shelving and space allocation, improving operational efficiency while considering constraints related to data sharing, storage and tracking, which can limit AI's full potential.

The correlation between library operational decisions and academic management activities – such as the design, monitoring, updating and evaluation of curricula – underscores the library's role as a strategic partner in academic governance (Sandhu, 2018; Sant-Geronikolou and Martínez-Ávila, 2019). BI tools enable libraries to provide academic departments with insights into how students and faculty use resources, allowing for data-informed adjustments to curriculum design (Nguyen et al., 2020; Romero and Ventura, 2020). For example, usage patterns of specific resources can highlight emerging fields of interest, prompting updates in course content to align with academic and industry trends. This symbiosis fosters a cohesive approach to academic excellence, where libraries directly influence the relevance and quality of educational programmes.

The use of business intelligence for data-driven insights contributes to understanding the role of libraries in the academic ecosystem. For example, analysing peak service hours and user demographics can help refine staffing schedules or prioritise investment in digital resources that support learning (Khademizadeh et al., 2022). Additionally, BI enables libraries to generate evidence-based reports that display their contributions to student success and faculty productivity, strengthening their position as indispensable academic partners. For instance, BI empowers university libraries to analyse students' reading habits, enabling the development of personalised recommendations that promote engagement with diverse resources (Dogan et al., 2024). These insights can support academic programmes in strengthening students' reading, comprehension and writing skills, thereby enriching their overall learning experience.

The use of data mining techniques in academic libraries helps uncover hidden patterns in users' information-seeking behaviours, enabling collection development that aligns with users' actual informational needs (Ansari et al., 2021). Through the analysis of usage patterns, libraries can segment their audience based on needs, such as undergraduate students requiring basic research support or faculty members needing access to advanced scholarly databases. This segmentation allows for personalised services, such as targeted resource recommendations, customised training sessions, or tailored communication strategies that improve user satisfaction and engagement (Dogan et al., 2024; Paul, 2022).

Research method

Research approach

In this research, we use the Design Science Research (DSR) approach (Hevner, 2007) to design and evaluate the AI model, aligning with the library's operational needs. DSR is particularly suited for this study because the research problem involves developing and implementing an artefact - a data-driven AI model – to address a practical challenge: automating the extraction and analysis of bibliographic resources. Unlike purely empirical or theoretical approaches, DSR enables iterative testing and refinement of a solution within a real-world context (Gregor and Hevner, 2013). Following Gregor and Hevner's (2014) framework, this study qualifies as '*improvement research*,' addressing a known context where existing solutions are lacking or inefficient. Prior studies have successfully applied DSR in library settings to develop book recommendation systems (Nykänen, 2021) and automated classification frameworks (Kragelj and Kljajić Borštnar, 2021), demonstrating its relevance for designing AI-driven tools in academic environments.

We based our research on a staged process that allowed for multiple iterations between rigour, relevance and design cycles (Hevner, 2007). Rigour cycles involved research about connecting the IA-based solutions design process with literature related to library operational processes. Relevance cycles included a deep exploration of the site investigation, eliciting design requirements and evaluating the prototype. Design cycles involved the core activities of building and evaluating the IA algorithm as well as the design process itself. A prototype of two IA algorithms was designed to assess its value to offer BI in practice. Prototyping reduces cost by addressing technical issues before any instantiation or implementation of the algorithm in the library processes. In the design cycles, the CRISP-DM methodology was applied. This methodology consists of six stages for developing data-driven projects: business understanding, data understanding, data preparation, modelling, evaluation and deployment (Wirth and Hipp, 2000).

Case selection

The Alfonso Borrero Cabal Library at Pontificia Universidad Javeriana is one of Colombia's largest academic libraries, with a collection of over 1,000,000 volumes across diverse disciplines. The library manages an extensive catalogue, including more than 600,000 e-books and access to numerous academic databases, supporting advanced research and education. In addition to traditional library resources, the library oversees the university's course catalogue, which comprises approximately 9,000 courses and grows by about 400 new courses annually. This oversight includes ensuring that curriculum-related resources are accessible within the library's collection. While the library holds valuable databases, BI initiatives remain limited due to challenges in data standardisation and the lack of digital skills among librarians, particularly in the applied use of AI and data science for library services. These limitations hinder the automation and strengthening of internal processes, leaving untapped potential for data-driven enhancements in resource management and service delivery.

Business understanding

The library is facing data challenges due to the lack of standardisation in the way the professors and administrative staff present and consolidate the bibliographic information of the syllabus of each subject. Currently, the university holds around 9,000 subjects. Some subject syllabi can have just five references, but others can hold more than fifty. Each professor designs, builds and registers those references in their own way, and this is because they do not follow a standard such as APA or IEEE to upload the references. This means that a single cell in an Excel file contains the full list of bibliographic references for a course syllabus in an unstructured format, making the information challenging to process. In 2020, due to the pandemic situation, 25 people conducted a manual extraction process to consolidate a database of more than 25,000 resources identified among the different university programmes. However, this dataset only reflects the state of bibliographic resources up to 2020. Due to the cost and time required, the library lacks an automated method to update this information each semester.

Additionally, having bibliographic references in a structured format only partially addresses the issue, as interoperability problems between the library and university information systems currently prevent this information from being integrated with the library's loan database. This limitation results in minimal business intelligence capabilities, reducing the potential for gaining insights into demand patterns, usage of bibliographic resources, most relevant and popular knowledge areas, collection obsolescence and course internationalisation levels, among others. As a result, challenges remain in optimising collection availability, loan processes, procurement and negotiations with resource suppliers. These limitations also affect strategic actions such as programme accreditation, self-assessment, curriculum updates, disciplinary alignment and curriculum internationalisation.

Data collection and analysis

For this study, we collected data from the library systems. The model training uses a bibliographic resource database manually curated by the library in 2020. An initial review of the data revealed that 43 per cent of titles were duplicates, leaving 4,778 unique titles for the training dataset. We also analysed the library's initial loan database containing 88,480 records, of which 23,819 are unique. These are registered across 18 faculties and 63 departments, covering a total of 250 programmes. As part of the CRISP-DM approach, the research included business understanding, pre-processing, processing, post-processing and deployment. This cycle was executed twice: first, to develop a named entity recognition (NER) algorithm using machine learning techniques to automatically and structurally extract bibliographic data from the syllabus catalogue; and second, to create a clustering model for analysing bibliographic resources with the goal of generating insights that enable efficient management of these resources.

Results

The findings of this study are based on previous research that developed algorithms for the automatic extraction of bibliographic reference texts (Daza and Lopez, 2023) and the clustering of the extracted information (Espinosa and Tapiero, 2024), in the context of the digital transformation of the Alfonso Borrero Cabal Library.

The extraction processes

Pre-processing: to create a dataset suitable for training a NER model, a Python script was developed to generate synthetic citations in the specified order of fields defined in the mapping file and the references style. The AnyStyle dataset (Fenton et al., 2022) was selected to supplement the base dataset, because it provides a labelled dataset with a good diversity of citation styles and fields used. The final dataset contains the original field of bibliographic references in the database and the corresponding synthetic citations as labels.

Processing: to face these issues, authors developed a NER algorithm with supervised machine learning (ML) techniques which allows to extract the bibliographic data from the syllabus in a structured and automatic way. To build the model, authors performed five activities to create the automatic extraction process:

a. The first step was to design and develop a ML model that recognises bibliographic metadata in the corpus. We used Bidirectional Encoder Representations from Transformers – BERT as an NLP model which can be fine-tuned for tasks like recognising named entities. The model was trained with the bibliographic reference dataset manually extracted by the library. The training was based on the DistilBERT-base-multilingual-cased (Sanh, 2019) because it covers up to 104 languages and, with 134 million parameters, is significantly faster than mBERT-base, which has 177 million parameters. Its design distinguishes between uppercase and lowercase characters, which is useful for identifying authors' initials and titles within the text.

Serie	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)
epoch 1 | batch 8	75.9	76.2	76.1	84.2
epoch 2 | batch 16	82.7	83.4	83.0	88.5
epoch 1 | batch 16	55.9	50.3	52.9	71.9
epoch 3 | batch 8	90.9	91.5	91.2	93.5

Table 1. Evaluation of BERT models.

b. The second step was to run and fit the model to identify the metadata of a bibliographic citation: author, title, year, editor and others. In this case, we ran experiments with different epochs and batches as shown in Table 1, so that we found that by running three epochs and eight batches we got a better performance in the extraction in terms of the precision, recall, F1 and accuracy parameters.

Post-processing: the third step was to evaluate the algorithm with the subject catalogue of two faculties: communication and languages, and engineering. Figure 1 presents a proof of concept, where all bibliographic references from the faculty's courses were processed in a structured manner, despite being initially unorganised. The model extracts metadata from these references, assigning each reference's information to the course from which it was sourced. The model successfully identifies metadata such as author, year, title and publisher. It also processes references in multiple languages, regardless of citation style or format.

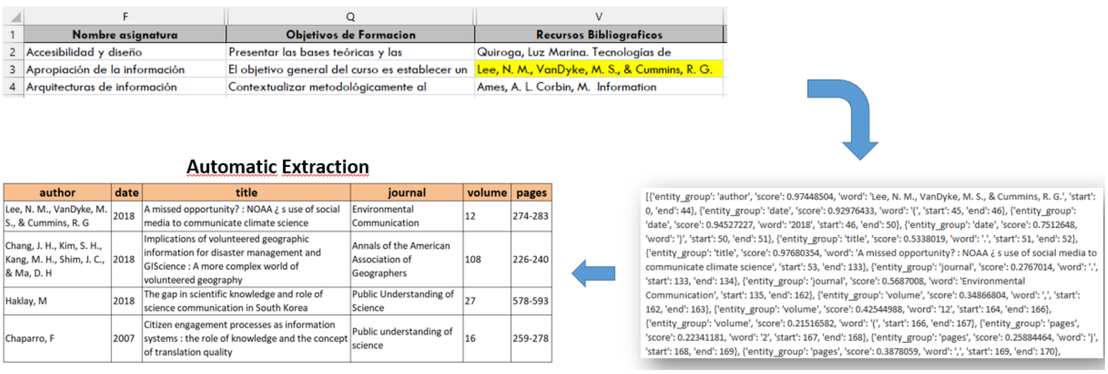

Figure 1. Proof of concept for the extraction process.

The business intelligence process

Pre-processing: this task involved collecting and preparing the bibliographic data required for clustering. This included data cleaning, outlier removal, lemmatisation, variable normalisation and vectorisation. The result was the generation of bag of words and TF-IDF matrices, along with corresponding features for each vectorisation technique, enabling more efficient and meaningful analysis and processing of the data.

Additionally, relevant features for clustering, such as authors, topics and keywords, were selected. Subsequently, the need to integrate fundamental information sources – the library loan database and faculty syllabi – was identified. However, this process was challenging due to the lack of standardisation between the two databases (syllabi and loans) and the absence of a key to link records between the datasets. Therefore, various techniques were explored to create a key that would facilitate the integration of these databases.

Cosine similarity was tested to measure the degree of similarity between two vectors (Kocher and Savoy, 2017), as well as Siamese neural network for evaluating sentence similarity (Chi and Zhang, 2018), and fuzzy string matching, a technique used to find strings that approximately match a specific pattern (Rao et al., 2022). Fuzzy string matching was chosen, as it more effectively captures the semantics of the text and achieves higher similarity scores between features.

Processing: cluster analysis is a common technique in data exploration aimed at discovering hidden patterns and groupings within datasets (Lund and Ma, 2021). This process relies on unsupervised machine learning, meaning it identifies relationships within the data without direct user intervention (Kettenring, 2006). For modelling purposes, the authors used an unsupervised K-means clustering algorithm (Wang et al., 2011), applying clusters to the data in a simple and effective manner by optimising the sum of the distances between each object and the centroid of its cluster. For clustering, a total of eight variables were considered, of which three are numerical and the remaining five are categorical. These variables include the number of occurrences (number of times the book appears in the catalogue of a syllabus), number of times borrowed, loan recency, date, location, publisher, most frequent programme and most frequent profile.

Model	Silhouette coefficient	Davies-Bouldin index
K-means	0.08	2.64
GMM	0.05	2.22
LDA + K-means	0.56	0.86
LDA + GMM	0.67	0.47

Table 2. Evaluation of clustering models.

Four different clustering models were evaluated to identify the most suitable approach for the dataset: K-means, Gaussian Mixture Model (GMM), Latent Dirichlet Allocation (LDA) combined with K-means, and LDA combined with GMM. These models were evaluated through metrics such as the Silhouette coefficient and the Davies-Bouldin index to identify the structure and dispersion of the clusters. Table 2 presents the evaluation results. The K-means model was selected because it allows for more accurate differentiation between clusters (Q=4), avoids cluster overlap and provides greater precision in identifying the distribution of each cluster.

Post-processing: the next step was to identify the characteristics of each cluster (see Table 3).

Cluster	Number of occurrences	Number of times borrowed	Loan recency (months)
1	2.1	2	6
2	1.2	6	5
3	1	2	9
4	1	2	3

Table 3. Cluster parameters for selected variables.

Cluster 1 consists of resources used moderately in the library. They are borrowed regularly and have been utilised by a considerable number of people. However, they are neither the most in-demand nor the most recent, with a recency of 6 months, possibly indicating use in courses offered annually.

Cluster 2 consists of the most in-demand resources in the library. They are frequently borrowed and have been used by a large number of people. Additionally, these are not the most recently borrowed resources and may refer to subjects taught once a year.

Cluster 3 consists of resources that are used infrequently in the library. Although they have been borrowed a considerable number of times, they are not as in-demand as the resources in cluster 2. Additionally, the most recent loan has an average of 9 months.

Cluster 4 consists of the most recent resources in the library. They are not as in-demand as the resources in cluster 2, but they have been borrowed a considerable number of times recently.

On the other hand, it was observed that there is a number of texts with infrequent queries and high recency, which may indicate that these texts are outdated or have low usage in reference classes, an important aspect to analyse.

The last step was to deploy the model in an information system that could be used by library staff and academic programme directors, to make the necessary administrative, operational and academic decisions to ensure the optimal performance of the library service in terms of bibliographic resource management. This was achieved using a dashboard in Microsoft Power BI (see Figure 2), which enables the visualisation of different layers of the clustering variables for library and academic management.

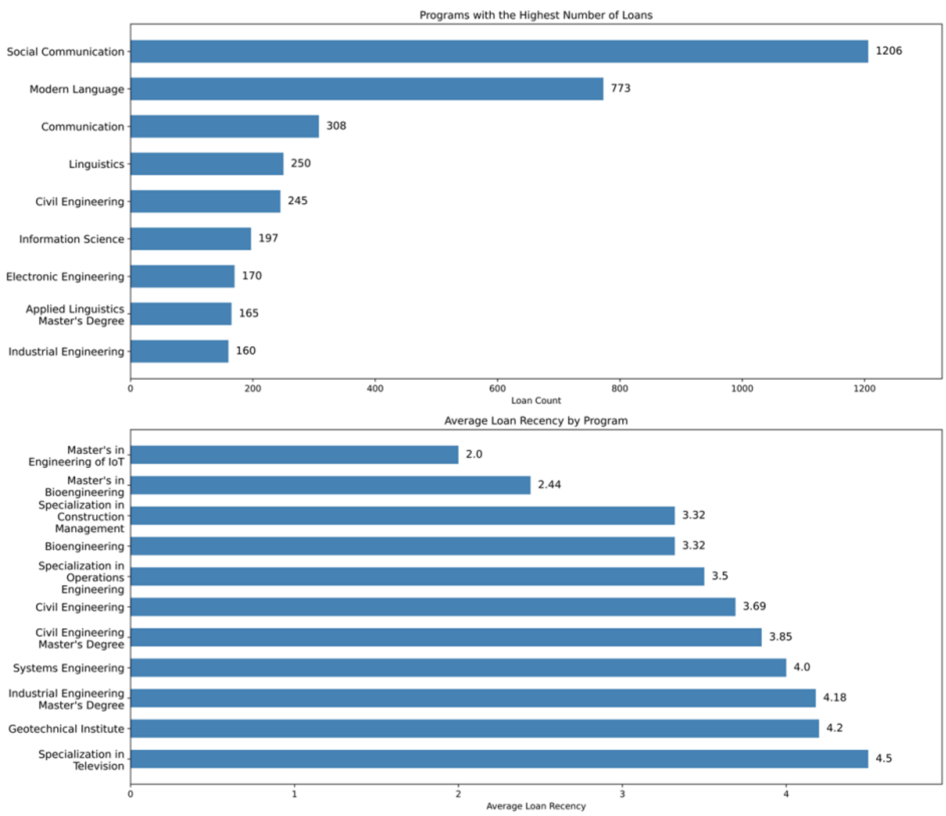

Figure 2. Visualisation of clustering variables.

Discussion and conclusions

The integration of AI into university library resource management offers opportunities to support operational and academic decision-making (Das and Islam, 2021; Harisanty et al., 2023), though efficiency gains depend on how institutions define priorities, allocate resources and address ethical concerns in data governance. By using BI tools supported by AI, this study explored methods to improve bibliographic resource allocation and automate processes such as data extraction from syllabi. These findings align with BI principles, which emphasise data-driven decision-making through integration, analysis and visualisation. By structuring and analysing bibliographic data, the study illustrates how BI can enhance resource distribution and align academic syllabi with available materials (Khan et al., 2023). The BI-driven approach provides a model for more efficient library management, supporting both operational and academic decisions while optimising processes like budgeting and resource acquisition.

This work faced several challenges in achieving its stated aim, including managing and integrating databases, creating clustering models and transforming technical results into valuable information for decision-makers at both administrative and academic levels (Okunlaya et al., 2022). Each of these tasks represents a key process in data science projects, aimed at delivering valuable outcomes that enable end users to achieve tangible benefits, optimise operations and effectively leverage the processed information.

There are several reasons for these challenges: first, libraries possess vast amounts of data, but not all of it is properly structured for processing by AI algorithms. This implies significant effort in data cleaning, normalisation and enrichment to make it suitable for effective analysis and utilisation in AI-driven applications. Second, university libraries typically use a variety of systems and platforms to manage their collections, services and users. Integrating AI into this heterogeneous environment requires substantial development and configuration effort. Third, AI models can be computationally demanding, especially when applied to large datasets. Scalable and efficient technological infrastructures are essential to ensure the performance of AI applications. And fourth, to design and implement AI-based solutions, libraries staff should become digitally savvy in topics such as data management, AI knowledge and technical expertise. Additionally, strong digital leadership skills are crucial to guide the planning process, align stakeholders and ensure successful deployment and integration of these solutions into the organisational systems of libraries (Lund et al., 2020).

This raises the question of whether university libraries have the necessary infrastructure and expertise to effectively implement AI solutions. The integration of AI requires not only technological infrastructure but also specialised skills, strategic planning and organisational commitment to fully capitalise on its benefits (Priya and Ramya, 2024). Therefore, a digital transformation roadmap (Okunlaya et al., 2022) for university libraries could foster the realisation of a value-adding AI ecosystem that facilitates the effective integration of artificial intelligence into library and university activities, aligning with and supporting the institution’s mission and vision.

The findings of this study contribute to the understanding of the potential of BI in improving operational and academic management efficiency in universities. In particular, the results suggest that AI techniques can streamline resource allocation, enhance decision-making processes through data-driven insights and optimise the classification and accessibility of bibliographic resources, supporting academic excellence and institutional effectiveness.

While this study demonstrates the effectiveness of AI-driven BI tools in optimising library resource management, further discussion is warranted on the practical consequences of these findings. Our case study provided an ideal environment to explore how AI can enhance decision-making in a university setting. By automating processes such as syllabus data extraction and bibliographic resource allocation, AI-enhanced BI systems facilitated more informed decision-making in budgeting, acquisitions and resource distribution. For instance, library administrators leveraged these insights to anticipate demand, optimise spending and align collections with evolving academic needs. However, successful implementation requires ongoing evaluation and adaptation to ensure the tool's outputs remain pertinent and actionable. Future studies should explore real-world adoption by library staff at Javeriana and similar institutions, assessing how these tools influence daily workflows and strategic decision-making across different academic disciplines.

The design process and design product test also suggest that IA solutions should be used as an auxiliary tool in conjunction with human BI efforts to ensure the highest level of accuracy and precision. Regular monitoring and evaluation of the models are essential to identify and address potential biases or limitations in the input and output data. These factors should be considered when making decisions about implementing AI-driven solutions in university libraries, ensuring that the necessary technological infrastructure, data readiness, digital capabilities and leadership support are in place to maximise the benefits and address potential challenges effectively.

The study's results should be taken in light of their limitations. Due to the nature of the research and computational infrastructure, the paper is limited to using bibliographic data of two faculties (communication and engineering). Because of the initial state of the datasets, extensive data cleaning, curation and preparation steps were required prior to the modelling and analysis tasks, which consumed a significant percentage of the time allocated for this research.

Future research could test these algorithms in faculties such as medicine or design, where language and citation practices may differ from those previously analysed. Additionally, it could explore deploying extraction and clustering models within existing library information systems, assessing the adoption of this technology by library staff and integrating these data-driven decision models into academic programmes. Such integration would ensure that syllabi remain continuously updated, aligned with institutional goals and supported by library resources.

Acknowledgements

We are grateful to Alfonso Borrero Cabal Library at Pontificia Universidad Javeriana for providing the data and information for this study. Our sincere appreciation goes to the Faculty of Communication and Language for their support of this initiative, and to Cristian Lopez, Andres Daza, Nathalya Espinosa and David Tapiero for their valuable contributions to this work. We also thank the anonymous reviewers for their helpful comments.

About the authors

Néstor A. Nova is Assistant Professor in the Department of Information Science at Pontificia Universidad Javeriana. He received his PhD from Universidad Javeriana. His research interests focus on IA-based knowledge management, interaction between ICTs and society for real-world problem-solving and digital transformation. He can be contacted at novanestor@javeriana.edu.co

Hernan Morales is Adjunct Professor in the Department of Information Science and Head of Collections Development at the Alfonso Borrero Cabal Library of Pontificia Universidad Javeriana. He holds a Master's in Strategic Management of Information and Knowledge from Universidad Pontificia Bolivariana. He can be contacted at hmorales@javeriana.edu.co

Juan Pájaro is Assistant Professor in the Department of Clinical Epidemiology and Biostatistics and a doctoral student at the Faculty of Engineering at Pontificia Universidad Javeriana. His work focuses on artificial intelligence projects for the development of health services and information management. He can be contacted at juanpajaro@javeriana.edu.co

Andrea González is Adjunct Professor, Department of Information Science, Pontificia Universidad Javeriana. She is a doctoral student at the School of Economics and Administrative Sciences, Universidad Nacional de Colombia. Her work focuses on data analysis and information studies for academic management in universities. She can be contacted at agonzalezsa@unal.edu.co

References

Adetayo, A. J., Asiru, M. A., & Omolabi, I. B. (2021). Building ambidexterity in libraries: role of competitive intelligence. *Library Philosophy and Practice* (e-Journal), 5763.

Ahmed, M., Mukhopadhyay, M., & Mukhopadhyay, P. (2023). Automated knowledge organization AI ML based subject indexing system for libraries. *DESIDOC Journal of Library & Information Technology*. https://doi.org/10.14429/djlit.43.01.18619

Al Fraidan, A. (2024). Integrating AI-powered library systems to enhance research and learning in English language departments: a case study of faculty and student perceptions. *Library Progress International*. 44(3), 12058-12065. https://doi.org/10.48165/bapas.2024.44.2.1

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.

Asemi, A., & Asemi, A. (2018). Artificial intelligence (AI) application in library systems in Iran: a taxonomy study. *Library Philosophy and Practice*, 2(2).https://digitalcommons.unl.edu/libphilprac/1840/

Bagchi, M. (2020). Conceptualising a library chatbot using open source conversational artificial intelligence. *DESIDOC Journal of Library & Information Technology*, 40(06), 329-333. https://doi.org/10.14429/djlit.40.06.15611

Balachandran, B. M., & Prasad, S. (2017). Challenges and benefits of deploying big data analytics in the cloud for business intelligence. *Procedia Computer Science*, 112, 1112-1122.

Barsha, S., & Munshi, S. A. (2023). Implementing artificial intelligence in library services: A review of current prospects and challenges of developing countries. *Library Hi Tech News*, 41(1), 7-10.

Briggs, T. (2020). Using data visualization and dashboards to support the decision making of public library branch managers. Doctoral Dissertation, Wilmington University.

Chi, Z., & Zhang, B. (2018). A sentence similarity estimation method based on improved Siamese network. *Journal of Intelligent Learning Systems and Applications*, 10(4), 121-134.

Das, R. K., & Islam, M. S. U. (2021). Application of artificial intelligence and machine learning in libraries: a systematic review. *arXiv*:2112.04573. https://doi.org/10.48550/arXiv.2112.04573

Daza, A., & Lopez, C. (2023). *Prueba de concepto para la extracción de recursos bibliográficos desde los catálogos de asignaturas de la Facultad de Comunicación y Lenguaje de la Pontificia Universidad Javeriana.* http://repository.javeriana.edu.co/handle/10554/67393

Deo, G. S., Mishra, A., Jalaluddin, Z. M., & Mahamuni, C. V. (2020). Predictive analysis of resource usage data in academic libraries using the Vader Sentiment algorithm. In *Proceedings of the*

12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 221-228). https://doi.org/10.1109/CICN49253.2020.9242575

Dogan, O., Yalcin, E., & Areta, O. (2024). Digitalization for enhancing reading habits: the improved hybrid book recommendation system with genre-oriented profiles. *Library Management*, 45. https://doi.org/10.1108/LM-03-2024-0030

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

Echedom, A. U., & Okuonghae, O. (2021). Transforming academic library operations in Africa with artificial intelligence: opportunities and challenges: A review paper. *New Review of Academic Librarianship*, 27(2), 243–255. https://doi.org/10.1080/13614533.2021.1906715

Espinosa, N., & Tapiero, D. (2024). *Modelo de clusterización para la gestión eficiente de recursos bibliográficos en la Biblioteca Alfonso Borrero Cabal y la optimización de procesos académicos en las facultades de Comunicación y Lenguaje e ingeniería de la Pontificia Universidad Javeriana*. Pontificia Universidad Javeriana.

Fenton, A., Keil, S., Krtek, J., & Ilja, S. (2022). *AnyStyle* [Code repository]. AnyStyle. https://github.com/inukshuk/anystyle (Archived by the Internet Archive at https://web.archive.org/web/20250327125258/https://github.com/inukshuk/anystyle)

Gauzelin, S., & Bentz, H. (2017). An examination of the impact of business intelligence systems on organizational decision making and performance: the case of France. *Journal of Intelligence Studies in Business*, 7, 40-50. https://doi.org/10.37380/jisib.v7i2.238

Gregor, S., & Hevner, A. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–355.

Hamad, F., Al-Aamr, R., Jabbar, S. A., & Fakhuri, H. (2021). Business intelligence in academic libraries in Jordan: opportunities and challenges. *IFLA Journal*, 47(1), 37–50. https://doi.org/10.1177/0340035220931882

Hamad, F., Fakhuri, H., & Abdel Jabbar, S. (2022). Big data opportunities and challenges for analytics strategies in Jordanian academic libraries. *New Review of Academic Librarianship*, 28(1), 37–60. https://doi.org/10.1080/13614533.2020.1764071

Harisanty, D., Anna, N. E. V., Putri, T. E., Firdaus, A. A., & Noor Azizi, N. A. (2023). Is adopting artificial intelligence in libraries urgency or a buzzword? A systematic literature review. *Journal of Information Science*, 01655515221141034. https://doi.org/10.1177/01655515221141034

Hevner, A. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2), 4.

Jayavadivel, R., Arunachalam, M., Nagarajan, G., Shankar, B. P., Viji, C., Rajkumar, N., & Senthilkumar, K. R. (2024). Historical overview of AI adoption in libraries. In *AI-assisted library reconstruction* (pp. 267-289). IGI Global.

Jayawardena, C., Reyal, S., Kekirideniya, K. R., Wijayawardhana, G. H. T., Rupasinghe, D., & Lakranda, S. (2021). Artificial intelligence based Smart library management system. In *Proceedings of the 6th IEEE International Conference on Recent Advances and Innovations in Engineering* (ICRAIE), (6, pp, 1-6). https://doi.org/10.1109/ICRAIE52900.2021.9703998

Kettenring, J. R. (2006). The practice of cluster analysis. *Journal of Classification*, 23(1), 3-30.

Khademizadeh, S., Nematollahi, Z., & Danesh, F. (2022). Analysis of book circulation data and a book recommendation system in academic libraries using data mining techniques. *Library & Information Science Research*, 44(4), 101191. https://doi.org/10.1016/j.lisr.2022.101191

Khan, A. U., Zhang, Z., Chohan, S. R., & Rafique, W. (2022). Factors fostering the success of IoT services in academic libraries: a study built to enhance the library performance. *Library Hi Tech*, 40(6), 1976-1995.

Khan, R., Gupta, N., Sinhababu, A., & Chakravarty, R. (2023). Impact of conversational and generative AI systems on libraries: a use case Large Language Model (LLM). *Science & Technology Libraries*, 43(4), 1-15. https://doi.org/10.1080/0194262X.2023.2254814

Kocher, M., & Savoy, J. (2017). Distance measures in author profiling. *Information Processing & Management*, 53(5), 1103-1119.

Kotu, V., & Deshpande, B. (2018). *Data science: concepts and practice*. Morgan Kaufmann.

Kragelj, M., & Kljajić Borštnar, M. (2021). Automatic classification of older electronic texts into the Universal Decimal Classification–UDC. *Journal of Documentation*, 77(3), 755-776.

Long, X., & Wu, Y. (2012). Borrowing data mining based on association rules. In *Proceedings of the International Conference on Computer Science and Electronics Engineering* (2, pp. 239-242). https://doi.org/10.1109/ICCSEE.2012.179

Lund, B. D., Omame, I., Tijani, S., & Agbaji, D. (2020). Perceptions toward artificial intelligence among academic library employees and alignment with the diffusion of innovations' adopter categories. *College & Research Libraries*, 81(5), Article 5. https://doi.org/10.5860/crl.81.5.865

Lund, B., & Ma, J. (2021). A review of cluster analysis techniques and their uses in library and information science research: K-means and K-medoids clustering. *Performance Measurement and Metrics*, 22(3), 161-173.

Meesad, P., & Mingkhwan, A. (2024). Data-driven library management: from data to insights. In P. Meesad & A. Mingkhwan, *Libraries in transformation* (Vol. 157, pp. 169–209). Springer Nature. https://doi.org/10.1007/978-3-031-69216-1_6

Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education: an integrated view. *Journal of Information Systems Education*, 31(1), 61.

Nykänen, A. (2021). A novel hybrid recommendation system for library book selection. Master's Thesis, Oulu University. https://oulurepo.oulu.fi/handle/10024/18684 (Archived by the Internet Archive at https://web.archive.org/web/20250327130140/https://oulurepo.oulu.fi/favicon.ico)

Okunlaya, R. O., Syed Abdullah, N., & Alias, R. A. (2022). Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*, 40(6), 1869–1892.

Paul, M. (2022). Segmentation of the reader and library user population in Poland. *The Library Quarterly*, 92(3), 274–295. https://doi.org/10.1086/719910

Pence, H. E. (2022). Future of artificial intelligence in libraries. *The Reference Librarian*, 63(4), 133–143. https://doi.org/10.1080/02763877.2022.2140741

Priya, S., & Ramya, R. (2024). Future trends and emerging technologies in AI and libraries. In Applications of artificial intelligence in libraries (pp. 245-271). IGI Global.

Rao, P. J., Rao, K. N., & Gokuruboyina, S. (2022). An experimental study with fuzzy-wuzzy (partial ratio) for identifying the similarity between English and French languages for plagiarism detection. *International Journal of Advanced Computer Science and Applications*, 13(10), 393-401.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. https://doi.org/10.1002/widm.1355

Sandhu, G. (2018). The role of academic libraries in the digital transformation of the universities. In *Proceedings of the 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services* (ETTLIS) (pp. 292-296). https://doi.org/10.1109/ETTLIS.2018.8485258

Sanh, V. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper, and lighter. *arXiv Preprint* arXiv:1910.01108.

Sant-Geronikolou, S., & Martínez-Ávila, D. (2019). Prospects of library use data integration in campus information systems: a glocalized perspective. *El Profesional de La Información*, 28. https://doi.org/10.3145/epi.2019.jul.10

Vijayalatha, C. (2023). Enhanced and changing role of library and information professionals in digital era. In *Proceedings of the National Conference on Exploring the Past, Present, and Future of Library and Information Science* (p. 130).https://shorturl.at/egtCu

Wang, R., Tang, Y., Liu, G., & Li, Y. (2011). K-means clustering algorithm application in university libraries. In *Proceedings of the IEEE 10th International Conference on Cognitive Informatics and Cognitive Computing* (ICCI-CC'11) (pp. 419-422). https://doi.ieeecomputersociety.org/10.1109/COGINF.2011.6016175

Wirth, R., & Hipp, J. (2000). CRISP-DM: towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1, (pp. 29-39).https://cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

Zhang, X., Zang, L., Bai, W., & Liu, H. (2024). Big data analysis and application of library information resources. *Applied Mathematics and Nonlinear Sciences*, 9. https://doi.org/10.2478/amns-2024-1212