(RESEARCH ARTICLE)

Unlocking insights from academic library data using clustering and recommender dashboard analytics for enhanced book collection management

Edwin Santos de Guzman [1, *], Isagani Mirador Tano [2], Keno Piad [3], Ace Lagman [4], Joseph Espino [5], Jonilo Mababa [6] and Jayson Victoriano [7]

[1] Mentor, Graduate Studies Department, La Consolacion University, Bulihan, City of Malolos, Bulacan, Philippines.
[2] Mentor, Graduate Studies Department, Quezon City University, Sanbartolome, Quezon City, Philippines.
[3] Mentor, Graduate Studies Department, Bulacan State University, Malolos, Bulacan, Philippines.
[4] Mentor, Graduate Studies Department, Far Eastern University, Sampaloc, Manila, Philippines.
[5] Mentor, Graduate Studies Department, Bulacan State University, Malolos, Bulacan, Philippines.
[6] Mentor, Graduate Studies Department, La Consolacion University, Bulihan, City of Malolos, Bulacan, Philippines.
[7] Mentor, Graduate Studies Department, Bulacan State University, Malolos, Bulacan, Philippines.

Abstract

This study, titled "Unlocking Insights from Academic Library Data using Clustering and Recommender Dashboard Analytics for Enhanced Book Collection Management: UST Perspective," explores data-driven strategies to enhance the University of Santo Tomas (UST) library's collection management. The K-Means clustering algorithm was used to analyze library collection data, identifying patterns based on book titles, publication years, authors, and categories. The clustering results revealed high-demand clusters, including Social Sciences, Health Sciences, Humanities, and Science and Technology, while low-usage clusters highlighted underutilized resources such as Senior High School (SHS), Heritage, Junior High School (JHS), Education High School, and Music collections. Acquisition patterns showed peaks in specific years and emerging categories, particularly in Science and Technology. Data visualization tools like Tableau and JupyterLab were used to present these insights. Despite challenges, such as handling interdisciplinary overlaps and managing data inconsistencies, the K-Means algorithm effectively uncovered meaningful patterns. To enhance user experience, a personalized recommender system using collaborative filtering was developed. This system provides offered book suggestions based on users' interests, reviews, and ratings by analyzing similar users' interactions. The recommender system is accessible via www.ustlibrary.online. This study highlights the potential of combining clustering algorithms and recommender systems to support data-informed decision-making, ultimately fostering a responsive and user-centric academic library service.

Keywords: Academic Library Analytics; Clustering Algorithms; Data-Driven Insights; K-Means Clustering, Library Collection Management; Recommender Systems; Library Collection Management

1.   Introduction

The University of Santo Tomas (UST), the oldest existing university in Asia and the largest Catholic university in a single campus, was founded in 1611 through the efforts of Bishop Miguel de Benavides. Central to its academic support system is the Miguel de Benavides Library, which houses extensive collections and modern facilities designed to foster research and learning.

This study explores data-driven strategies to enhance the library's services by developing cluster models to describe and profile book collections, creating an online book borrowing system to improve accessibility, and designing a

---

* Corresponding author: Edwin Santos de Guzman

recommender system that provides personalized book suggestions based on user reviews and ratings. To address the scalability issues associated with traditional clustering algorithms, the study implements advanced techniques such as distributed clustering, dimensionality reduction, and parallel processing, ensuring the efficient handling of large-scale datasets.

Clustering techniques offer a deeper understanding of the structure of book collections by identifying patterns, trends, and groupings that highlight similarities and differences among resources. The integration of an online system expands user access by allowing remote browsing and borrowing of books, while the recommender system enhances the overall user experience by suggesting materials aligned with individual interests. Through these innovations, the study aims to strengthen resource organization, improve accessibility, and create a more responsive and user-centric library service at UST.

1.1. Significance of the Study

The following benefits of the proposed study are listed below:

- **University of Santo Tomas. The** University will be able to gain competitive advantage as data analytics will be provided and can easily be reported and shared to different academic units of the university.
- **Miguel de Benavides Library.** The main beneficiary of this proposed study is the library. With the existence of dashboard analytics, librarians and administrators will be able to see the most borrowed books, and other reports.
- **Faculty/Researchers.** The proposed study can be used as an additional reference for advanced research, sampling, publications and additional learning course material references.

1.2. Scope and Delimitation

This study focuses on enhancing book collection management at the Miguel de Benavides Library of the University of Santo Tomas (UST) using clustering algorithms and recommender system. It aims to categorize and group books through K-Means clustering, identify patterns and trends, and provide personal recommendations based on user preferences.

To improve accessibility, a web-based system will be developed, allowing users to browse, search, and borrow books remotely, with data visualized through platforms like Tableau or Google Data Studio. Data handling will involve pre-processing techniques to clean and structure raw data, ensuring its accuracy before analysis. Tools such as MS Excel, Google Sheets, Oracle Database, or MySQL will support data management.

The study focuses on book data spanning four centuries, from 1611 to the present, acknowledging limitations such as incomplete datasets due to data loss or confidentiality. It also recognizes challenges related to data quality, cluster interpretation, and algorithmic subjectivity, emphasizing the need for careful validation and calibration.

By clearly outlining its scope and delimitations, the study seeks to offer practical insights for improving book collection management at UST while addressing the complexities of data-driven research.

2. Methodology

2.1. Research Method

The researcher adopts both descriptive and developmental research approaches for this study. The descriptive research component involves evaluating the developed system using the ISO 25010 standard. This ISO model provides criteria for assessing the quality and compliance of software systems. By applying the ISO 25010 instrument, the researcher will ensure that the system meets established standards for functionality, reliability, usability, efficiency, maintainability, and portability.

In the developmental research component, the researcher focuses on creating and refining the system titled "Unlocking Insights from Academic Library Data Using Clustering and Recommender Dashboard Analytics for Enhanced Book Collection Management." This system development follows two key methodologies which include Agile Software Development and Knowledge Discovery in Databases (KDD). Agile is an iterative and incremental software development methodology that emphasizes flexibility and collaboration. The researcher will use Agile principles to build the system through incremental prototypes, allowing for continuous feedback and iterative improvements. The KDD is an analytics methodology focused on extracting useful patterns and models from data. In this context, KDD will be used to implement

clustering techniques and predictive models that enhance the system's ability to analyze and manage book collections. The KDD process will guide the extraction of valuable insights from academic library data, contributing to the system's functionality and effectiveness.

Incorporating clustering algorithms is a critical aspect of the KDD process, as it allows the system to identify patterns and relationships within library data. Selecting the appropriate clustering algorithm directly impacts the effectiveness of book collection management and recommendation strategies. The table below compares commonly used clustering algorithms, highlighting their characteristics and best use cases in the context of academic library data analysis.

2.1.1. Respondents of the Study

The research of the study entitled, Unlocking Insights from Academic Library Data using Clustering and Recommender Dashboard Analytics for Enhanced Book Collection Management - UST Perspective would typically include the librarians, and administrators/IT Experts.

Librarians: Since they oversee overseeing and maintaining the book collection, librarians would also make up a crucial portion of the research population. By automating the categorizing process using clustering algorithms, the Descriptive Analytics Dashboard would help librarians organize the collection more effectively.

Administrator/IT Experts: The research population would also include administrators, such as university or library directors. They might use the Descriptive Analytics Dashboard's findings to make data-driven decisions about the library's policies, budget, and long-term planning for the book collection. The role of IT experts is very vital in organizing and helping to manage, analyze, and leverage the vast amounts of data available in academic libraries to enhance book collection management. Academic library data can be inconsistent or staggered, especially when dealing with big volumes of data. IT experts' role in cleaning and preprocessing this data, ensuring that the input for clustering and recommendation systems is reliable and correct.

Table 1 Respondents of the Study

UST Miguel de Benavides Personnel	Population
Administrators/IT Experts	10

As presented in table 1, these are all the respondents of the study. These respondents will have to check and validate the content and the proposed study especially its correctness, accuracy and reliability.

2.2. Sampling Techniques

Purposive Sampling: Sampling with a purpose entail choosing individuals or groups in accordance with predetermined standards that support the goals of the research. Researchers may decide to focus on user groups, such as graduate students or faculty members, who are more likely to extensively interact with the book collection dashboard in the context of descriptive analytics using a dashboard employing clustering techniques for data. This strategy enables targeted analysis and user segment-specific insights.

2.3. Locale of the Study

The proposed study will focus on the book collections and data of the Miguel de Benavides Library at the University of Santo Tomas (UST) in Manila. The researcher will collaborate closely with the UST Miguel de Benavides Library Department, which has full access to the library's collection of records from its earliest acquisitions to the present. This collaboration ensures the study is grounded in accurate, comprehensive data, making the research more substantial and impactful.

Conducting this study will provide the library with a competitive edge by developing a data-driven dashboard containing essential information about collection usage, acquisition trends, and user engagement.

The University of Santo Tomas serves as an ideal research locale due to its rich history—dating back to 1611—and its reputation for academic excellence. With a wide range of disciplines, including humanities, sciences, health sciences, engineering, and business, UST offers numerous research opportunities and fosters cross-disciplinary collaboration.

2.4. Instruments of the Study/Data Collection

The proposed research samples for Unlocking Insights from Academic Library Data using Clustering and Recommender Dashboard Analytics for Enhanced Book Collection Management - UST Perspective depend on the specific goals and objectives of the study. Here are a few potential sampling approaches to consider:

2.5. Evaluation Criteria

To assess the system's acceptability, a Likert Scale ranking system will be utilized. Evaluators will rate each ISO 25010 characteristic according to the following scale:

Table 2 5-point Likert Scale

Response Categories	Response Categories
High Acceptable	4.50 – 5.00
Very Acceptable	3.50 – 4.49
Acceptable	2.50 – 3.49
Moderately Acceptable	1.50 – 2.49
Unacceptable	1.00 – 1.49

A structured questionnaire will guide evaluators in providing consistent and objective feedback for each product's characteristic.

3. Results and discussion

This chapter serves as the core of the research study, providing a detailed presentation, thorough analysis, and comprehensive interpretation of the data gathered. The primary objective is to transform raw data of the library collections of UST into meaningful insights that address the research questions.

3.1. Tableau and JupyterLab

The 110,000 library collections were processed using Tableua and JupyterLab Software. The tableau was used to create comprehensive visualizations like charts and graphs to summarize finding and presenting complex data clearly to audiences without technical expertise while the JupyterLab was used to perform data cleaning, manipulation, and transformation using Python and to develop custom models, from analysis to machine learning algorithms, K-Means.

The figures below were the results of the processed data through Tableau and JupyterLab.

The below figure shows the gathered details and their categories. By clustering books into relevant categories, we can identify patterns and trends, such as overrepresented or underrepresented subjects, helping librarians make data-driven decisions. The book categories used in this study include Civil Law, Ecclesiastical Faculties Library, Education High School Library, Filipiniana, General Reference and Information, Health Sciences Library, Heritage, Humanities, Music, Old Books, Religion, Science and Technology, Social Sciences, Spanish, and others specified by their publishers and locations. The goal is to create a dynamic dashboard that not only visualizes these clusters but also offers intelligent book recommendations, ultimately supporting a more efficient and user-centric academic library experience.

The data collection methodology employed a multi-faceted approach to ensure comprehensive and accurate data gathering. First, Library Database Mining was utilized to extract data directly from the academic library's catalog, providing a solid foundation of book records. For older or unlisted books, Manual Data Entry was conducted to capture information not available in digital formats. To uphold data integrity, a Verification process was implemented by cross-checking records with physical copies and consulting library staff for accuracy.
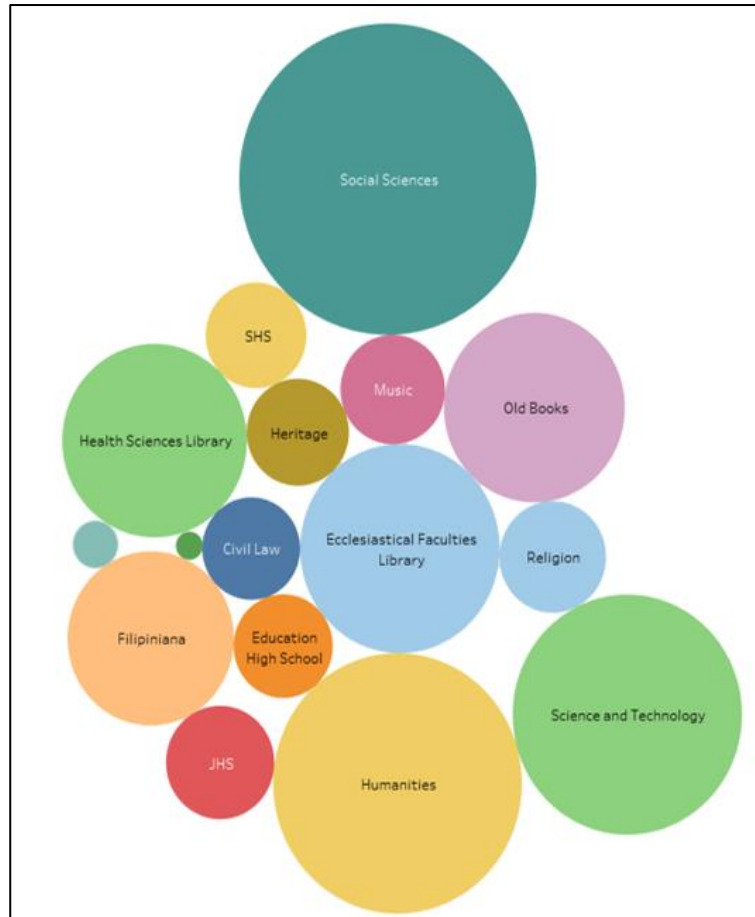
Figure 1 Clustered books by their category based on number collections

Once collected, the data underwent meticulous preparation. This involved cleaning and normalizing publisher names to eliminate duplicates caused by minor formatting inconsistencies, standardizing publication years, and consolidating categories by merging publisher names and locations into relevant fields.

For analysis, a Clustering Approach was adopted using K-means clustering to group books by category and collection size. The clusters were then visualized through scatter plots and bar graphs to identify patterns, allowing for the detection of both underrepresented and overrepresented categories.
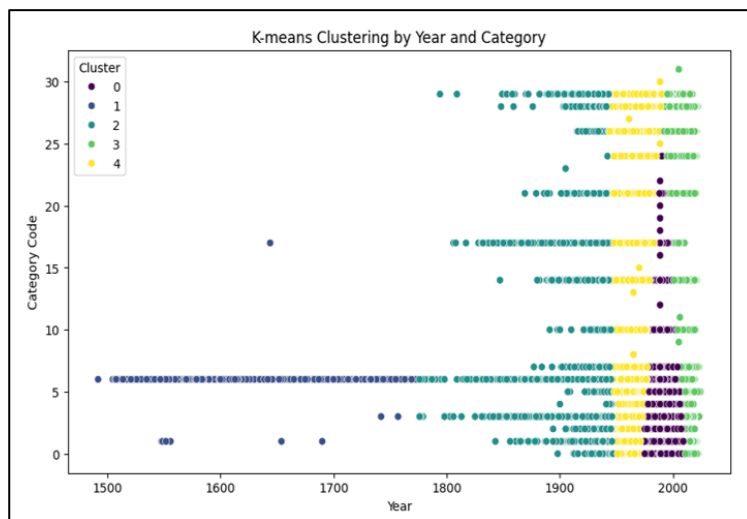
Figure 2 Clustered books according to year and category

A key highlight of this process was the scale of the book collection handled. From an initial collection of 310,000 books, the dataset was refined and reduced to 110,000 books after pre-processing and data cleaning. This significant reduction underscores the rigorous methods applied to remove inconsistencies and redundancies, ensuring the data's reliability.

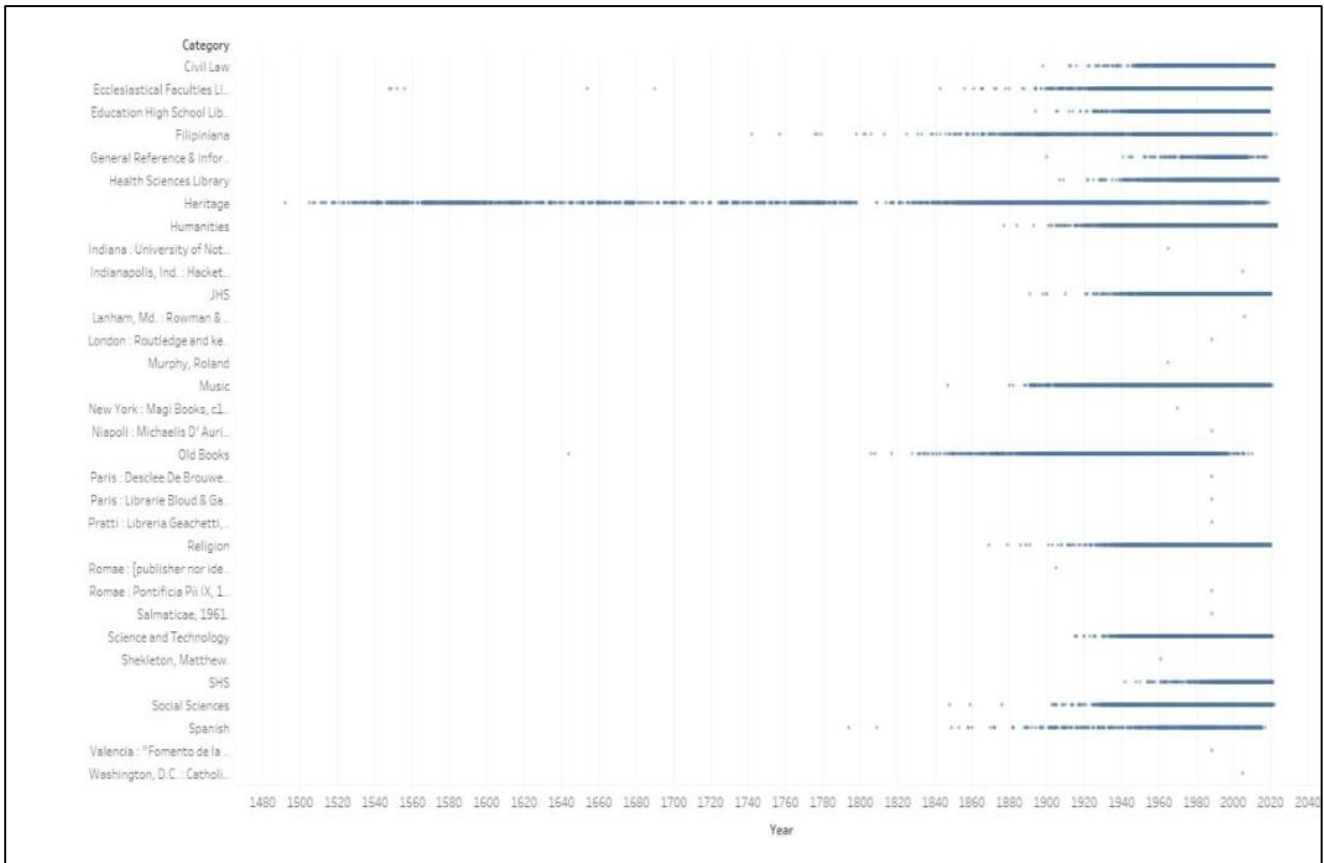Figure 2 shows the clustered books according to year and category by identifying the other data.

Figure 3 Book collection clustered based on year using Tableau

This figure shows the data based on the existing book collection, illustrating how book categories are clustered by year using Tableau. The x-axis represents the publication years of the books, ranging from the 1400s to the 2000s, while the y-axis lists various categories, such as "Culture," "Education," "Science," "Technology," "History," and more. The graph reveals distinct patterns in the distribution of books across time. Categories like "Heritage" and "General Reference" display a relatively consistent spread over the centuries, with many books from earlier periods. In contrast, modern categories, such as "Science" and "Technology," show more concentrated clusters starting from the 1800s onward, reflecting the rise of specialized knowledge. Notably, categories like "Indiana University" and "Health Sciences Library" exhibit surges in book publications beginning in the mid-1900s. Older categories, including "Manuscripts" and "Special Collections," have a significant number of works published before the 1800s. Overall, the data highlights a clear trend: traditional categories span a broader historical range, while scientific and specialized subjects show denser clustering in more recent centuries, aligning with the expansion of academic and technical literature over time.
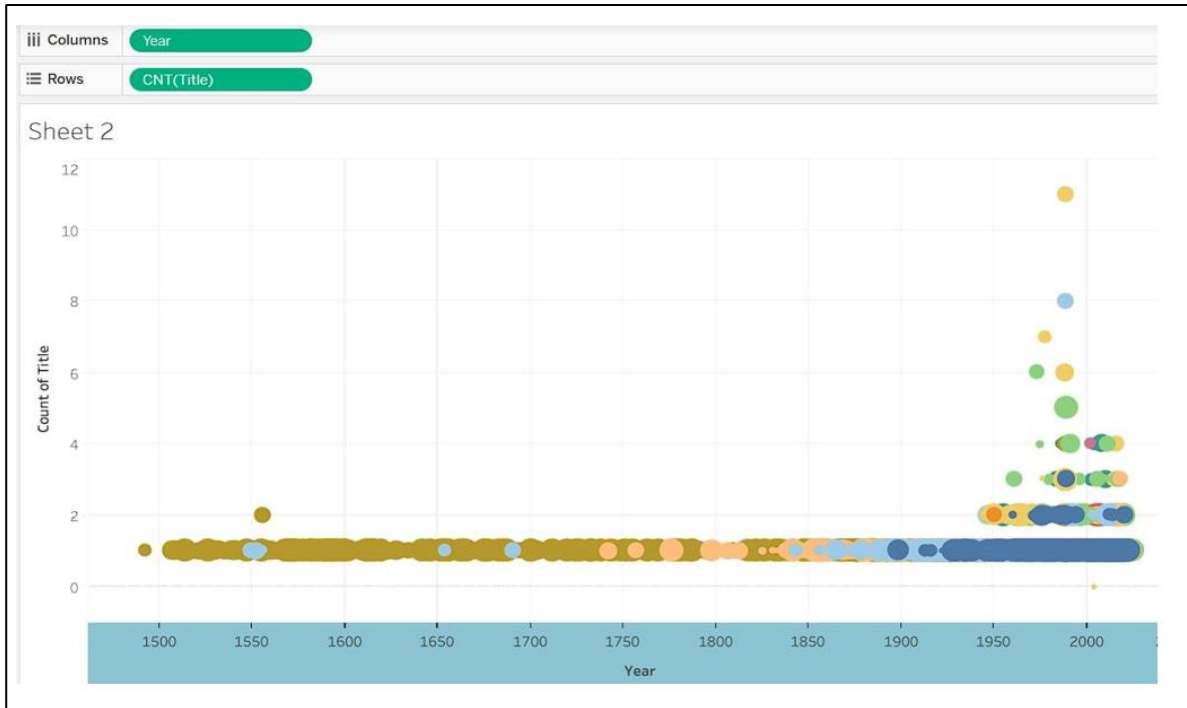
Figure 4 K-Means Clustering by Year and Category

Figure 4 shows color cluster (0-4) - upon cleaning, for blank fields, was replaced using the mean year graph shows how many books were added from the different book categories using Jupyterlab.
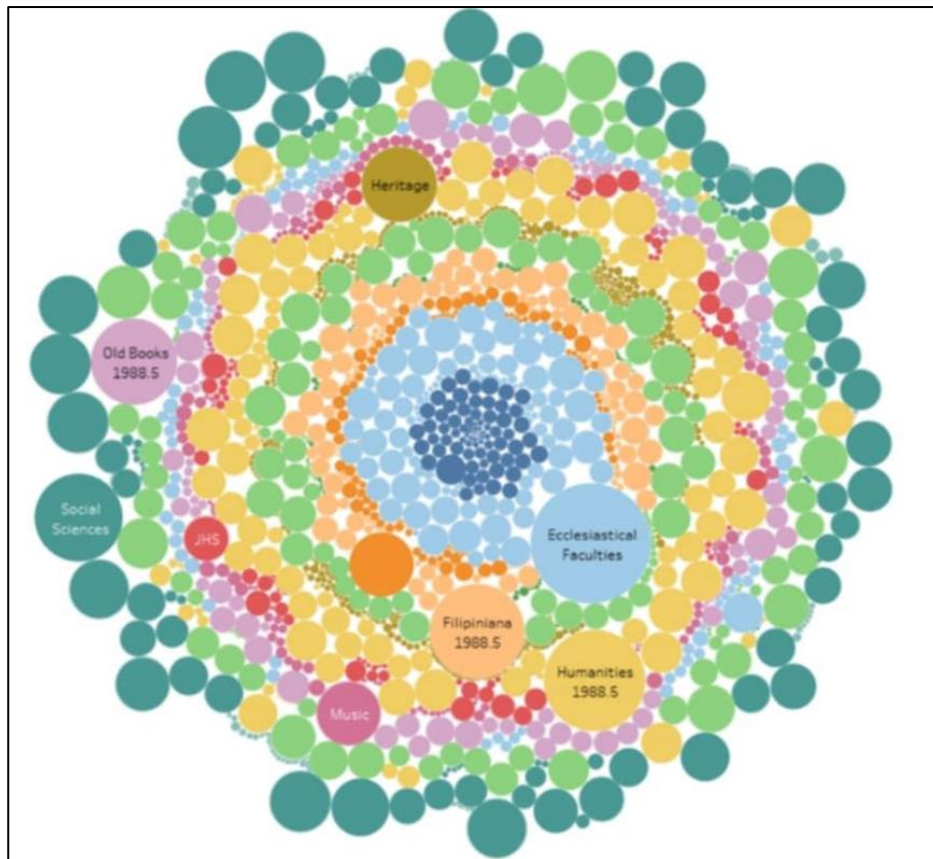

Figure 5 K-Means Clustering by Title and Year

Figure 5 shows the title and year - based on the publication year of the book, how many new titles were added to the library book collection was added on a specific year. Now, since the results from the clustering method are limited, the variability or lacks significant differences among the data points. Relatively, K-means produced only a few meaningful clusters. such as highly similar) and too small or insufficiently detailed for distinct cluster formation. On the other hand, there were selected clustered results, there were specific patterns or groupings stood out enough to form meaningful clusters and other data points clearly fit into distinct clusters and fell into default clusters as shown in the figure above.

3.1.1. Collaborative Filtering

In the perspective of having limited clustered results, the researcher created a Library Recommender System using Collaborative Filtering. This method of personalized recommender system filtered information such as interactions data from other similar users. A recommender system provided users with personalized book suggestions based on their preferences, reviews, comments, and ratings. It enhanced the user experience by suggesting books that match the user's interests.

4. Results

4.1. Quantitative Procedures

- The researcher will first seek written permission to conduct the study in the UST Miguel de Benavides Library Personnel and users. Request letters will be sent to the university administrators to formalize the request.
- Upon approval, the researcher will also seek the cooperation of the respondents about their participation in the study.
- Seeking the help of the library administrator of the UST to help to send and share the datasets.
- After the retrieval, the data therein will be shared, pre-processed and treated statistically and presented tabularly with their respective analysis and interpretation.
- The researcher will ensure the protection of the participants' identities and rights in accordance with the Data Privacy Act of 2012.

4.2. Qualitative Procedures

- The researcher will first seek written permission to conduct the study in UST. Request letters will be sent to the university administrators to formalize the request.
- Upon approval of the administration, the researcher will send letters to the target participants for interview sessions.
- When the target participants approve the request letter, he will sign the consent form that contains the approval to record the interview session. If the participants decline to record the interview still the researcher will consider the participants.
- If the participants agree on the provisions of the consent form the researcher will ask for the schedule of the interview session based on the availability of the participants.
- Before the start of the actual interview the researcher will start with a light conversation and share my personal narratives as middle manager, when the researcher feels that there is now rapport and openness the semi structured interview will commence.
- While the interview is conducted the researcher will take note of important observations.
- The researcher will ensure the protection of the participants' identities and rights in accordance with the Data Privacy Act of 2012.

4.2.1. ISO 25010 Product Quality Model

The ISO 25010 model provides a structured approach to assessing software quality through a set of product characteristics (Abesamis, A., Ocampo, L., Pasia, D., Villarin, L., Centeno, C., Sison, Ariel A.,. & iso25000, 2024). The model consists of eight main characteristics, each of which is further divided into sub-characteristics. The following sections outline how these characteristics relate to the proposed system:

Table 3 ISO 25010 Product Quality Model

	ISO 25010 Characteristics
Functionality	The functionality will be assessed based on the system's ability to meet user requirements, including efficient data clustering, intuitive recommendation displays, and customizable analytics dashboards.
Performance Efficiency	The system will be tested for the speed of data processing during clustering tasks, the time taken to generate recommendations, and the efficient use of server resources.
Compatibility	The proposed system will be evaluated for interoperability with the UST library's current software infrastructure and its capacity to integrate external data sources.
Interaction Capability	Usability tests will ensure the dashboard is user-friendly for librarians and administrative staff, enabling smooth data navigation and decision-making processes.
Reliability	The research will simulate real-world library data scenarios, examining system uptime, fault tolerance, and error recovery mechanisms.
Security	The system will be assessed for user authentication protocols, data encryption methods, and secure access controls.
Maintainability	The system will be evaluated for code modularity, documentation quality, and version control practices.
Flexibility	The system will be tested for its capacity to accommodate new clustering algorithms, adjust recommendation logic, and integrate future datasets.

4.3. Summary of Results

The summary results for each ISO 25010 characteristic are as follows:

- **Functionality:** The system's functionality received a "Very Acceptable" rating, indicating that it effectively meets user requirements for clustering and recommendations. The clustering algorithms accurately grouped book data, and the recommender dashboard provided relevant suggestions tailored to library needs.
- **Performance Efficiency:** Performance efficiency was rated "High Acceptable," showcasing rapid data processing and minimal resource consumption. The system demonstrated fast clustering execution times and optimized data retrieval processes without compromising server performance.
- **Compatibility:** Compatibility was found to be "Very Acceptable," ensuring smooth integration with existing library systems. The system effectively communicated with the UST library's software infrastructure and seamlessly integrated external data sources for enhanced analytics.
- **Interaction Capability:** Interaction capability received a "High Acceptable" rating, highlighting an intuitive and user-friendly interface. The dashboard's design allowed librarians and administrative staff to easily navigate data visualizations and customize their analytics views.
- **Reliability:** Reliability scored "Very Acceptable," indicating consistent performance and robust error handling. Simulations revealed stable system uptime and efficient fault tolerance mechanisms, ensuring uninterrupted functionality.
- **Security:** Security was rated "High Acceptable," emphasizing strong data protection and secure access features. The system implemented stringent user authentication protocols and encrypted sensitive data to safeguard against unauthorized access.
- **Maintainability:** Maintainability achieved a "Very Acceptable" rating, reflecting well-documented code and easy adaptability. The system's modular design and comprehensive documentation facilitated smooth bug fixes, upgrades, and feature enhancements.

- **Flexibility:** Flexibility received a "High Acceptable" rating, showcasing adaptability to future enhancements. The system successfully accommodated new clustering algorithms, adjusted recommendation logic, and integrated additional datasets as needed.

5. Conclusion

This section summarizes the findings, conclusions and recommendations for the Unlocking Insights from Academic Library Data using Clustering and Recommender Dashboard Analytics for Enhanced Book Collection Management - UST Perspective.

The K-Means algorithm was utilized to analyze the library collections data of the University of Santo Tomas (UST) to uncover meaningful patterns and insights. This study aimed to address critical challenges faced by the UST Miguel de Benavides Library in managing its collections, such as inefficiencies in cataloging, uneven resource utilization, and the difficulty of aligning acquisitions with the evolving academic needs of the university.

In the clustering results, the dataset was segmented into distinct clusters based on attributes such as book titles, publication years, authors, and categories. These clusters were analyzed using Tableau and JupyterLab to visualize and interpret the patterns effectively. The results highlighted clear groupings that provided a data-driven understanding of the library's collection landscape.

One key finding involved collection acquisition. High-demand clusters were identified, highlighting popular subjects such as Social Sciences, Health Sciences, Humanities, and Science and Technology. Conversely, low-usage clusters revealed underutilized resources, including those in the Senior High School (SHS), Heritage, Junior High School (JHS), Education High School, and Music collections. This disparity in resource usage underscores the need for more strategic acquisition decisions and resource promotion.

The study also uncovered acquisition patterns, showing peaks in specific years of book acquisitions and identifying emerging categories like Science and Technology. Such insights pointed to shifting academic interests and emphasized the importance of regularly updating collection strategies to match these trends.

However, the study encountered several practical challenges. Managing interdisciplinary materials proved complex, as books often span multiple subjects, creating overlaps between clusters. Additionally, determining cluster thresholds required subjective judgment, which introduced potential biases in interpretation. Data preprocessing and cleaning were also essential to handle missing entries, duplicated records, and inconsistent data, ensuring the clustering process remained reliable and accurate.

To address these challenges and enhance user engagement with the library's resources, the researcher developed a personal recommended system using collaborative filtering. This system filters information based on user interactions, such as book loans, reviews, comments, and ratings, to suggest titles aligned with individual interests. By leveraging collaborative filtering, the recommender system provided personalized book recommendations, enriching user experience and encouraging resource utilization. This system is accessible via www.ustlibrary.online, offering an intuitive platform for users to discover relevant materials effortlessly.

Ultimately, this study not only highlighted the inefficiencies and underutilization present in the library's collection management but also presented a data-driven solution through machine learning and personalization. These findings reinforce the importance of adopting advanced data analysis and technology-driven strategies to optimize academic library services and better serve the university community.

5.1. Conclusions

The application of the K-Means algorithm to the library data collections of the University of Santo Tomas demonstrated its effectiveness in uncovering patterns and optimizing resource management. By clustering the data based on key attributes such as title of the book, author. edition, and category, the algorithm provided actionable insights that can guide library operations and decision-making somehow.

The analysis highlighted areas of old books and emerging academic needs, enabling the library to align its collections with the evolving priorities of its users. Furthermore, the results underscore the importance of data-driven approaches in modern library management, offering a scalable and systematic method for understanding the discipline of collections.

While the chosen algorithm grouped similar data points, challenges such as handling interdisciplinary overlaps and determining the optimal number of clusters required careful understanding. Despite these limitations, K-Means proved

to be a valuable tool for enhancing resource allocation, improving user engagement, and future-proofing the library's collections of UST.

In summary, this study reinforced the possibility of utilizing the machine learning techniques in academic institutions to foster innovation and improve service management.

5.2. Recommendations

Given the incomplete nature of the library data collections of the University of Santo Tomas, it is essential to prioritize data quality improvement and implement strategies to maximize the value of existing information. Below are key recommendations:

- **Completeness of Data:** Audit and Validation of Data: Conduct a comprehensive review of the dataset to identify missing or inconsistent data.
    - o Supplement of Data: Participate with relevant areas to gather missing information, such as usage statistics, transactional history and utilization reports of books.
- **Utilize Existing Data:** Use the available and existing data to identify initial trends and patterns, focusing on clusters with robust information. These insights can still guide immediate actions, such as promoting high-demand collections or assessing underutilized resources.
    - o Provide training for library staff in data management with the best practices to reduce human errors and improve data reliability.
- **Future Collaboration and Projects:**
    - o By addressing the data gaps and investing in data management practices, the University of Santo Tomas can maximize the utility of its library collections, improve user satisfaction, and support data-driven strategic planning.

Compliance with ethical standards

Disclosure of conflict of interest

The authors have no any conflict of interest for publishing this article.

Statement of ethical approval

This study will strictly adhere to the Data Privacy Act of 2012 (Republic Act No. 10173) to protect respondents' privacy and personal data. Permission to conduct the study in the UST library will be obtained from the administrators. Respondents' participation will be voluntary, with informed consent secured before any data collection activities, including surveys, interviews, and focus group discussions.

Anonymity will be maintained by de-identifying all collected data, ensuring no personally identifiable information (PII) is linked to participants. Access to raw data will be restricted to the researcher. Participants may withdraw at any time without consequences, and their data will be securely destroyed upon withdrawal. They also have the right to refuse to answer any questions.

All digital data will be stored in password-protected files, and physical documents will be kept in a locked location. Data will be permanently deleted or shredded after the study's conclusion. The researcher will continuously review data management practices to ensure full compliance with the Data Privacy Act, promptly addressing any breaches or concerns. These measures guarantee respondents' privacy, data security, and ethical research practices.

Statement of informed consent

Informed consent was obtained from all individual participants included in this study. Participants received a clear explanation of the study's objectives, procedures, potential risks, and benefits prior to their participation. They were informed that their involvement was entirely voluntary, with the option to withdraw at any time without any negative consequences.

All necessary measures were taken to ensure participants' confidentiality and data privacy, including secure storage of information. Participants provided written consent, confirming their understanding and willingness to participate.

References

[1] Abesamis, A., Ocampo, L., Pasia, D., Villarin, L., Centeno, C., Sison, Ariel A.,. (2024). Pawsome Detect: Harnessing Pattern Recognition and Data Mining for Canine Age and Breed Detection. Journal of Electrical Systems. 20. 10.52783/jes.2427.

[2] Dash, D., Gupta, P., & Sharma, N. K. (2019). Descriptive Analytics Using Dashboards in the Automobile Industry. International Journal of Emerging Trends in Engineering Research, 7(9), 2278-2281.

[3] Liu, Y., Yang, B., & Li, L. (2019). Scalable spectral clustering on big data. Future Generation Computer Systems, 91, 372-382.

[4] Raza, A., Masood, A., Khan, S., & Lee, S. (2019). Parallel and scalable spectral clustering for big data using MapReduce framework. Future Generation Computer Systems, 96, 192-205.

[5] Ahmed, F., Alazab, M., Talari, A., & Abdullah, A. H. (2020). An efficient K-means clustering algorithm for big data using MapReduce. Future Generation Computer Systems, 107, 941-950.

[6] Han, J., & Kamber, M. (2019). Data Mining: Concepts and Techniques. Morgan Kaufmann.

[7] Tan, P. N., Steinbach, M., & Kumar, V. (2019). Introduction to Data Mining. Pearson.

[8] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (Eds.). (2019). Advances in Knowledge Discovery and Data Mining. MIT Press.

[9] Aggarwal, C. C. (2019). Data Clustering: Algorithms and Applications. Chapman and Hall/CRC.

[10] Grebić, B., and Stojanović, A. (2021). APPLICATION OF THE SCRUM FRAMEWORK ON PROJECTS IN IT SECTOR. European Project Management Journal, 11(2), pp. 37-46.

[11] SO/IEC. (2019). Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. International Organization for Standardization.

[12] Lee, Wen-Tin & Chen, Chih-Hsien. (2023). Agile Software Development and Reuse Approach with Scrum and Software Product Line Engineering. Electronics. 12. 3291. 10.3390/electronics12153291.

[13] Kusdiyanto, Andri & Muhamad, Wardani & Suhardi, Suhardi. (2023). Evaluating Service Computing Framework with Scrum through a Case Study. IJAIT (International Journal of Applied Information Technology). 32. 10.25124/ijait.v7i01.5988.

[14] Lavrakas, Paul J. (2021). Purposive Sample. Encyclopedia of Survey Research Methods. Print ISBN: 9781412918084 | Online ISBN: 9781412963947 DOI: https://dx.doi.org/10.4135/9781412963947.n419

[15] Longe, Busayo (2023). 7 Data Collection Methods & Tools For Research. Retrieved June 26, 2023 from https://www.formpl.us/blog/data-collection-method