2022

# UTILIZING THE POTENTIALS OF BIG DATA IN LIBRARY ENVIRONMENTS IN NIGERIAN FOR RECOMMENDER SERVICES

Job Onekutu Sani
*Federal University, Lokoja, Nigeria.*, sani.job4real@gmail.com

Nneka Angela Oseji
*Federal University, Lokoja, Nigeria*, nnekaosejiangela@gmail.com

# UTILIZING THE POTENTIALS OF BIG DATA IN LIBRARY ENVIRONMENTS IN NIGERIAN FOR RECOMMENDER SERVICES

**Oseji, Nneka Angela**
*Department of Library and Information Science*
*Federal University Lokoja*
nnekaosejiangela@gmail.com,

**Sani, Job Onekutu**
*Department of Library and Information Science*
*Federal University Lokoja*
sani.job4real@gmail.com, job.sanni@fulokoja.edu.ng

## Abstract

The big data revolution has gained global attention and initiated creative innovations in every field and libraries as engines of access to information have also been affected by this new trend. Libraries in this part of the world have not utilized the amazing potential of big data in library services. In this time, when various terms such as algorithms age, petabytes age, data age, etc. are been used to describe the activities initiated by machine learning, industries and organizations can achieve much by incorporating inspiring and innovative tools to improve services and performance. In this vein libraries in Nigeria are expected against all odds to make their services more interactive, attractive, innovative, and exciting by utilizing cloud technologies and machine learning techniques to create recommender services. This paper titled *"Utilizing the Potentials of Big Data in Nigeria Library Environments by Recommender Services",* focuses on the concept and characteristics of big data and its importance in complementing traditional library services, areas for applying big data systems in libraries, the concept of recommender systems and how it works, adopting recommender systems in libraries for maximum benefits, tools, and techniques for setting up big data recommender systems in libraries, challenges of big data recommender systems in libraries in Nigeria and strategies for overcoming big data challenges in library systems. The paper is based on a contextual analysis of literature from various scholarly works. The paper will also proffer recommendations based on the study.

**Keywords:** *Big Data, Library, Nigerian Library Environment, Recommender services*

## Introduction

Libraries have been known as the bedrock of knowledge since ancient times, focusing on physical collections of books, journals, periodicals, manuscripts, and research papers, as well as an unconventional method of presenting these materials by arranging them on shelves, and thus,

the search for information is based on the indexed authors, titles, subjects, and the physical proximity of relevant content (Bamgbade et al., 2015). The quality and amount of relevant and current material held in a library is what makes it significant and relevant. It is worth noting that libraries, regardless of their kind (academic, public, school, or special), are all service-providers, with all of their operations aimed toward meeting the requirements of varied users.

Libraries have long been recognized as service-oriented institutions dedicated to providing relevant information resources and quality services to its clients. As a result, Adeniran (2011) observed that clients are portrayed as the cause for the library's existence. To meet the information demands of library users, a healthy collection of recent and relevant items that will be valuable to the users must be maintained.

However, the world's computerization is accelerating, resulting in the development of massive amounts of data that cannot be efficiently managed using the old technique. Furthermore, Hoy (2014) observed that modern life creates data at an astounding rate that shows no signs of slowing down. This circumstance gives legitimate grounds to assess these elements, as it is noted that the massive expansion of data is referred to as "Big Data" (Djafri, Bensaber & Adjoudj, 2018).

**Big Data Concept and Characteristics**

Big data refers to data sets that are so massive and complicated that typical data processing applications are incapable of performing analysis, collection, storage, data curation, sharing, and transmission. To put it in another way, big data may be defined as a large volume, high velocity, and diverse set of information assets that necessitate cost-effective and new ways of information are processed for improved insight and decision making (Sonawane, 2018).

According to Zhan and Widen (2019), big data refers to enormous data sets with rapid expansion in multiple file formats, which might complicate data handling techniques and expedite the development of technical solutions. According to Cisco and the World Economic Forum, the relationship of billions of people using electronic devices generates a flood of data every day, and those virtual financial transactions, social media traffic, and GPS coordinates now produce over 2.5 quintillion bytes of so-called "big data" every day (Moulding, 2016). The qualities and variety of data sets make it difficult to use and manage data. Data scientists or commercial

businesses can organize data that has been collected in a uniform way .Some data types, however, are existing in an unstructured form obtained from diverse resources such as e-mails and internet gathered data (Wang, Xu, Chen & Chen, 2016).

Ishwarappa and Anuradha (2015) defined big data using four ideas. He underlined that big data is a data asset with a high volume, velocity, and diversity that requires new technologies and analytical approaches to convert to value. In addition to Ishwarappa and Anuradha's four big data concepts, De Mauro, Greco and Grimaldi (2016) introduced the idea of veracity, emphasizing that veracity is a critical quality of big data that influences the validity of the analysis. According to Kitchin and McArdle (2015), the distinguishing characteristics of big data are velocity and exhaustively, as it has taken many libraries in developing nations several years for half of their collection to be available in digital format.

From the aforementioned definitions, five (5) big data qualities were identified: volume, velocity, variety, value, and veracity, sometimes known as the (5v's) which are discussed as follows.

**Volume:** The magnitude of a data collection is referred to as its volume (typically terabytes and petabytes).The data might be so large that it cannot be evaluated on a single system, similar to how gigabytes of data can be handled by computer software. This required the application of big data analysis technology.

**Velocity:** The term velocity relates to data in motion, especially the rate at which data is generated, processed, and evaluated. In today's algorithm age, blogs and microblogs generate data at a quicker rate. Twitter, for example, generates over 6,000 tweets every second, in addition to over 350,000 tweets sent per minute, 500 million tweets per day, and approximately 200 billion tweets per year.

**Variety:** Variety is concerned with handling the complexity and heterogeneity of numerous datasets, including structured, semi-structured, and unstructured data; data types include textual data (e.g., blogs and text messages) and non-textual data (e.g., videos, images, and audio recordings). Ordinary data can be loaded into Microsoft Excel sheets, however big data comprises data extensions (such as blogs, tweets, visited websites, brief messages, chat discussions, and e-mails) that cannot be inserted into an MS Excel sheet, according to Erevelles,

Fukawa and Swayne (2016). As a result, maintaining and interpreting massive data necessitates a diverse set of data analysis sources and measurement methods.

**Value:** Two large data values were found by Ishwarappa and Anuradha (2015). First, considering the cost of processing large data, the IT infrastructure needed to manage it, and the turnover value, keeping in mind that businesses invest considerably in establishing IT infrastructures for big data storage and processing. The second form of value is the benefits gained from data processing.

**Veracity:** The accuracy and importance of large data are linked to its veracity. Inappropriate data in volume, velocity, and diversity may exist, raising processing and infrastructure costs while decreasing turnover value. However, data validity defines its accuracy and objectivity.

To describe the idea and features of big data, it is critical to know that large data sources are broadly classified into three categories.

**Data streams:** They are created or processed by computer, and mobile-mediated data such as log files, sensor data, position tracking, and processor generated data.

**Library Social Networks:** such as Google+, Facebook, YouTube, Twitter, Linked In, blogs, WhatsApp, Instagram, and Pinterest, are examples of social media platforms.

**Public Domains:** which include data that is publicly available on the internet

**The Importance of Using Big Data in Library Services**

Libraries have traditionally been responsible for managing information, which includes acquiring, organizing, and making collections available. The fundamental problem with conventional libraries is their incapacity to track down materials that have been removed or missing from shelves, effectively keep track of a large number of book records, and to cope with late deliveries (Tarique & Rani, 2017). As a result, conventional libraries can no longer deliver their customary services because they are overloaded with day-to-day operations.

Recent technology advancements in the twenty-first century have forced libraries and librarians to incorporate ICT applications into their day-to-day operations in order to effectively and efficiently deliver relevant services to their users. According to Preater (2014), with the advent of

Information and Communication Technology (ICT), the information industry became highly competitive, to the point where information has become a commercial commodity, with various actors outside of libraries (e.g., mega-bookstores, online book dealers, publishers and aggregators, and even the internet) competing for patrons' attention. Their job is to provide patrons with a private route of information (both physical and intangible information products) at the request of individuals/organizations.

Furthermore, according to Cox (2018), libraries have to compete with other departments within their companies or institutions for attention, recognition, staff resources, technology, cooperation, partnerships, possibilities, and so on. Librarians should also capitalize on these aspects in order to better position themselves for improved and innovative service delivery today and in the future. Kościelniak and Puto. (2015) and Chang (2016) demonstrated how big data has been used to improve business decision making, anticipate future healthcare trends, and analyze customer service satisfaction. This has made Big Data a promising area for future improvements in economic and societal value, as well as a potential source of competitive advantage for institutions/organizations in the long run. Hence, librarians must be fully aware of Big Data and the need of expanding data-related activities in libraries as one of the essential technical breakthroughs.

With the advent of big data, libraries and librarians must now understand how to make large datasets more usable, visible, and accessible by developing taxonomies, devising metadata systems, and systematizing retrieval techniques. This has compelled librarians, regardless of library type, to begin acquiring basic understanding of big data, its impact on the nature of their profession, and how big data may be used efficiently for various objectives. Sengupta (2016) stated that big data is not just about increasing the amount of data, but also about improving the tools for storing, aggregating, combining, and analyzing the data.

As a result, corporate librarians are now obliged to learn how firms store large data, how data mining offers them a competitive edge, and how they may assist users in overcoming problems in accessing and utilizing massive data sets in their job. Furthermore, librarians at Scientific Research Libraries are required to understand the distinction between big data and scientific data as well as the influence of new software and technology utilized for big data analysis. Similarly,

humanities and social science librarians are obliged to acknowledge that big data is increasingly ubiquitous in other fields as well, and is not limited to science and industry. Digital archivists, data curators, and other librarians are increasingly expected to advise their organization/faculty on the storage and accessibility of large data collections.

As a result, Sengupta emphasized that it is the sole responsibility of library administration to examine the types of big data sets their library could be gathering and analyzing using big data tools, while library professionals consistently engage in the discussion on new methods of data analytics using big data tools. Big data may be used by libraries and librarians to build library collections, track the usage of library resources by users, and so on. It may also be utilized for library management, user happiness, and information retrieval in the shortest possible time.

Nowadays, library work encompasses a wide range of unstructured and structured data and data types, including texts, metadata, photos, audio files, videos, research data, 3D digital copies, and software. Referencing, indexing, and presenting these contents, thus, libraries will only succeed in keeping with the times if cutting-edge indexing techniques from the field of big data are utilized instead of 19th-century library practices. The concepts of library accuracy as well as a sad reflection on previous obligations continue to result in enormous and detailed metadata being developed and digital items, such as analogue books, being catalogued (Ball, 2019).

According to Adetayo, Adeniran and Gbotosho (2021), the following are specific reasons for applying big data to library services:

**Effective Management of information:** Huge volumes of data are being created in every sector of the economy be it in the field of genetic engineering, health sector, market research surveys, industries, etc. Academics are attempting to use the collections to correctly analyze data and arrange information. As a result, many libraries are now involved in the big data sector for some time, albeit sometimes unknowingly. Because of their digitized holdings, extensive libraries, in particular, contain an almost unmanageable quantity of data.

**For Libraries to Effectively Analyze their Extensive data holdings/Services:** Libraries are directly affected by big data since they may utilize big data techniques to evaluate their enormous data holdings, such as better understanding their patrons and so delivering new or enhanced services.

**For Increase in Reliability and Access to Big Data for Research:** Academics at universities will increasingly rely on big data for creative and ground-breaking research, which has direct and indirect impact on the society and libraries as center of knowledge.

**Economic Reasons:** Big data can also enhance cost savings, automation, and faster and better decision making by individuals and organizations.

**Provide Other Opportunities** According to Regalado (2013), analyzing big data is critical and provides opportunities. Currently, just 0.5 percent of all data in the globe has been evaluated, implying that there are still lots of opportunities to participate in big data initiatives.

All of these factors should drive libraries as custodians of knowledge and information, to engage in big data management, since big data is proliferating everywhere. As a result, the use of digital technology to improve big data services will turn libraries into smart libraries, giving smart services to smart users (Schöpfel, 2018).

Libraries now provide online resources and services. Libraries are also promoting their services and activities using social media platforms such as Facebook and Instagram. Librarians may collect more online data and analyze it to add value to their services with the aid of rising technologies and tools such as analytics software. As a result, libraries may utilize big data to make better decisions about collection development, public space updates, and tracking the usage of library assets (Sonawane, 2018).

**The Use of Big data in Library**

Big data technology could be used in library. Main process involves collection selection, organization, description and modeling, storage, presentation or visualization. Of course data analysis is also important. In addition, the amount of storage and processing has grown the complexity of the library data and the challenges of working with it have also accelerated. In the library two aspects of data mining could be achieved: one is using data stored in the library and another is using the data collected during the process when users use the library service. Hence, Affelt (2015) and Kaladhar, Naicka and Rao (2018) discussed some of the uses of library big data as follows:

**Decision Making:** Data-driven approach, which takes the data as the basis, to make decision or recommendation, is a common method used in many areas. Based on the data, the decision could be more useful. For example, the library could use collaborative data mining techniques and text analytics to optimize the collections (books or journals) to generate better search results and to make recommendation for the books. At the end, this approach would improve the user satisfaction by providing better service, and efficient usage of library resources.

**Develop New Data Format:** Resource Sharing is one of the important goals of library. For example, OCLC (Online Computer Library Center) has been working to produce a Google-like "knowledge card" based on the reformatted library data and the card can be linked to from the outside. Library data could become linked data in order to achieve interoperability on the Web. Another example is that British Library studied the linked data of the library collections and modeled the people, events, places which are related to holdings in the library.

**Data Standardization and Data Modeling:** From a single work, like a research paper, or a book, the relationships from co-authors, citations, geo-location, dates, named entities, subject classification, institution affiliations, publishers and historical circulation information could be easily extracted.

**User Behavior Study:** The information of library collections could be mined through big data technology. On the other hand, it is possible to record and track library user's activity and to store that data in large-scale data storage, and then conduct data analysis. The result could then be used to potentially improve the overall user experience, and user satisfactory of library service.

## Areas for Applying Big Data Systems in Libraries

Big data-driven library technologies offer customized, remote, real-time, and virtualized services to library customers. The fast expansion in the volume, veracity, velocity and variety of library data generated by various library instruments opens up new avenues for investigating interactions with library users (Nicholson & Bennett, 2016). Big data technologies might be leveraged to create intelligent services. Hence, it has become critical for libraries to use big data to absorb information and improve services. Big data and smart libraries have the potential to improve technical and reader services.

**Collection Development and Preservation of Data Sets:** As more users become interested in working with big data, they will need guidance and material to work with. In order to support research, some governmental authorities have mandated agencies to open their data vaults for researchers. Unfortunately, this mandated for public access to these data sets does not provide for preservation or create any user-centered services for the information. Hence, libraries and librarians are well-positioned to help users understand how and where to find these data sets and to preserve them for future users.

**Technical Services:** RFID (Radio Frequency Identification Gadget) is a device that aids in the identification and tracking of data from objects (Pujar & Satyanarayana, 2015). RFID can be used to manage library collections. This is achieved by adding an RFID tag to each item in the procurement department that contains bibliographic information, transaction logs, and virtual representations. This will make each title distinct. Cataloguing may benefit from the use of cloud computing. Cloud computing is a type of Internet-based computing that provides computers and other devices with on-demand computational services and data. It is a model for acquiring, sharing, on-demand access to a pool of customizable computer resources (Khuntia et al., 2016). Libraries may use cloud computing to store all of their catalogue entries, which can then be harvested automatically by other libraries, similar to how Mendeley harvests citations.

**Readers' services:** RFID might be used by the circulating device to charge and discharge books. By combining the library card with RFID tags, it is possible to link catalogue circulation, overdue items, and penalties. As a result, libraries will be able to deploy smart technology to inform clients about overdue products and make it possible for them to pay penalties online (Addepalli & Addepalli, 2014). Furthermore, libraries will be able to sell library assets more effectively by utilizing smart shelves based on user behavior in the library and transaction data. This may be performed by telling the customer about new arrivals in a topic field that they were seeking for during previous library visits (Pujar & Satyanarayana, 2015).

If a library user discovers a book in the collection but it is marked as borrowed by another user, the second user may reserve it. When a reserved book is returned via the automated station, it is stored in a separate reserved book container. At this point, the device would send the user a message or an e-mail to confirm the availability of the book (Purnik, 2019). Users can utilize library smartphone applications to find resources, reserve books and research spaces, stay up to

date on library activities, and take part in user education efforts (Guo et al., 2018; Kerr & Pennington, 2018).

Furthermore, Sonawane (2018) added that Big data analytics is applicable in the following areas of the library services:

**Enhancing superior search results:** Data mining and text analytics on the past loan records and book bibliographies could enhance search results and recommendations.

**Demand Analysis:** It would help in forecasting demand for new existing titles.

**Planning library collection:** The technology used would optimize to plan the category mix in the collection by taking into consideration the space and budget constraints.

**Cataloguing Services:** Big data enables the library to become more user-friendly by providing tailored and intelligent services. The two categories of library data are cataloguing and process/transactional data. Catalogue data is the inherent data and knowledge of library files, whereas process data is developed or provided by library and service administration. The first kind of data consists mostly of documental, bibliographical, and financial data, whereas the second type primarily of log, user, and record data. Library big data analytics allow extraordinary digital library innovations such as personalized recommendation services and library user behavior/habit analysis, which add considerable value and insights for librarians, users, and services.

The improvements and benefits provided to librarians by intelligent big data analytical tools and digital management procedures that enable the library to provide competitive goods and services at a low cost demonstrate the value of librarians. Improving the user experience and enjoyment in the library is filled with user value. By analyzing library big data in its many forms, service value in terms of service and process quality and efficiency may be increased.

In order to apply big data techniques in library, Kaladhar, Naick and Rao (2018) noted some works to be done such as, but not limited to:

1. Central data repositories, where data are stored, maintained, and cataloged
2. Data standards, to which collected data should follow

3. Data communities, which collect, maintain, and curate data
4. Analysis tool

Big data provides libraries with opportunities to frequently reinvent the service patterns that they previously employed to satisfy patrons through recommender services.

**Recommender systems**

Recommender System (RS) is a tool, which gives recommendations to the user according to their needs or query. Basically, the main feature or function of the Recommender System is to predict the user interest by relating the history, information, profile and queries used, searched, created and expressed by the user. Recommender systems are software tools and techniques providing suggestions for items to be of use to a user. For instance recommender systems may recommend news on a news portal, or products in an online shop, or even services. The recommendations are usually tailored to a given user or a given group of users. Since recommendations are personalized, they may vary from one user to another or from one user group to another (Kabore, 2012). According to Isinkaye, Folajimi and Ojokoh (2015) recommender systems is define as systems which support the social process of collecting, aggregating, and communicating recommendations for a social group. For universities, recommender systems in scientific digital libraries hold the promise of supporting , consulting and communication processes of research and education, and thus improves teaching and research processes.

Recommender Systems (RSs) are automated tools and strategies that allow recommendations to users about items that may be of importance to them. Recommender systems indicate users' items of interest depending on their direct and indirect interests, other users' interests, and user and item features. Any of the practical implementations that use such devices may involve (recommending books, products, videos, jobs, music).The aim of the Recommender System is to handle the information overload problem and improve the customer relationship management. It can play a huge role in classifying the information for bibliographic control and information retrieval mechanism. Implementation of a recommender system provides lots of benefits to the users like identifying user interest, locating relevant information and helping in dynamic research activity (Isinkaye, Folajimi & Ojokoh 2015).

Recommender systems are helpful in searching contents since they help users to find information which can't be found anywhere else. However, recommender systems are often applied in many search engines to index and find non-conventional data. Recommender systems (RS) was introduced in the mid 90's and it's was first alluded by Jussi Karlgren at Columbia University in a specialized report as a "Digital Bookshelf" in 1990. There has been significance contribution made by various experts since then. The paradigm shift of RS has gone through lots of technological advancement (Gupta & Pandey, 2019). The goal of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them

Recommender systems usually make use of either or both collaborative filtering and content-based filtering (also known as the personality-based approach), as well as other systems such as knowledge-based systems. Collaborative filtering approaches build a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete, pre-tagged characteristics of an item in order to recommend additional items with similar properties (Melville & Sindhwani, 2012).

According to Gupta and Pandey (2019), a recommendation system helps an organization to create loyal customers and build trust by providing their desired products and services for which they came on your site. The recommendation system today are so powerful that they can handle the new customer too who has visited the site for the first time. They recommend the products which are currently trending or highly rated and they can also recommend the products which bring maximum profit to the company. Library services are well suited for the adoption of recommender systems, especially services that support the user in the search for literature in the catalog. From a librarian's point of view, a recommendation system can be seen as a form of catalog enrichment or as a substitute for traditional subject classification. The recommendation systems assist librarians in setting up and keeping up-to-date the library's holdings. From a more technical viewpoint, a recommendation system may be seen as a Web 2.0 application, since the user indirectly generates the data used by the system, and the integration into the catalog is being done by means of a mashup – two aspects of Web 2.0 applications.

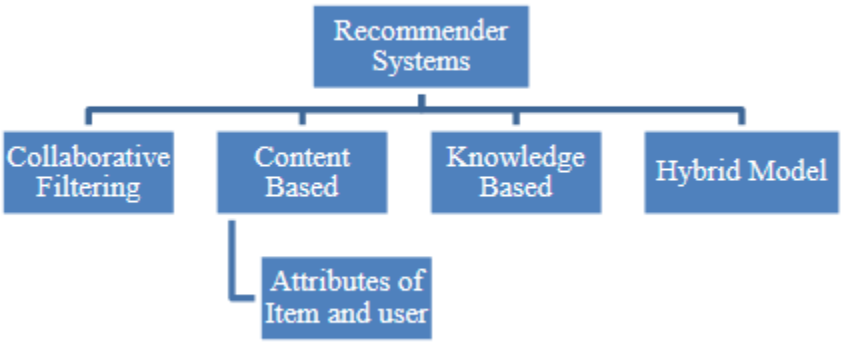
**How does a Recommendation System Work?**

Techlabs (2021) emphasized that recommendation systems use specialized algorithms and machine learning solutions. Driven by the automated configuration, coordination, and management of machine learning predictive analytics algorithms, the recommendation system can wisely select which filters to apply to a particular user's specific situation. It facilitates marketers to maximize conversions and average order value. Recommender systems can forecast user ratings, even before they have provided one, making them an effective tool. Mainly, a recommendation system processes data through four phases as follows: -

- **Collection**: Data collected can be explicit (ratings and comments on products) or implicit (page views, order history, etc.).
- **Storing**: The type of data used to create recommendations can help you decide the kind of storage you should use- NoSQL database, object storage, or standard SQL database.
- **Analyzing**: The recommender system finds items with similar user engagement data after analysis.
- **Filtering**: This is the last step where data gets filtered to access the relevant information required to provide recommendations to the user. To enable this, you will need to choose an algorithm suiting the recommendation system.

**Types of Recommender Systems**

The recommendation techniques can be broadly categorized into collaborative filtering content-based filtering, knowledge-based system, hybrid systems etc.




**Content Based Recommendation:-** Content-based recommendation systems find out items of interest for users by analyzing item descriptions. These systems generate list of item profiles for the users, based on data provided by users. It uses a metric of term-frequency (TF) and inverse document frequency (IDF). The product of TF*IDF is used to identify the importance of the item. Term-frequency determines how many times the item is occurring in a document. IDF identifies the importance of the item. It is log of the ratio of total number of documents and number of documents containing the item. After calculating TF*IDF, vector space model is used to identify preference of a user to an item. In vector space model, vector is created for the item and corresponding attributes. Users' vector is also created based on his preference on the attributes of the items. Finally cosine angle between these vectors are calculated to identify the similarity between user and item vectors. Correlation based approach can also be used for content based recommendation.

**Collaborative filtering based recommendation:** These recommendations are based upon preference of similar users. Item will be recommended to the user based on the preference of other similar users for the same item. If set of users have strongest correlation in the past, they will be identified as 'nearest neighbor'. Score of the new items will be predicted based upon the scores of nearest neighbor. Many collaborative filtering algorithms exist for data sets where numbers of items are less than number of users. In collaborative filtering Pearson correlation or Log-likelihood ratio can be used to identify preferred items for the user.

**Knowledge Based Recommendation System**: In these systems knowledge about the user's need, his preferences etc. are used for recommendation. These systems are based upon the

knowledge of user's need for a particular item and can therefore reason about the relationship between a need and a possible recommendation.

**Hybrid Recommendation System:** The features of content based, collaborative filtering and knowledge based systems can be combined together to improve the recommendation accuracy. Burke describes hybrid recommender systems based on the hybridization methods. The weighted hybridization method combines the scores of several recommendation techniques together. Switching hybrid system switches to different recommendation techniques based on the current situation. Feature combination combines features from different recommendation data sources into a single recommendation algorithm. Cascade recommendation uses series of recommendations; the recommendations given by one recommender will be refined by the other. Feature augmentation uses output from one technique as an input feature to another. In meta-level system, model learned by one recommender acts as input to another.

**Examples of Recommendation Engines Put to Work**

Recommendation engines are at the front and center of projecting library service and resources. The key point is that they can be utilized in almost every library to optimize and improve patrons' experience. Some examples of recommendation engines relevant for libraries are:

***Personalized recommendations:*** Such engines help understand the preferences and intent of each visitor and show the most relevant recommendation type and resources in real-time. Recommendations improve as the engine learns more about each visitor.

***Website personalization:*** It allows Increase in patronage and exchanges by segmenting and targeting visitors with real-time personalized messages and offers.

***Real-time notifications:*** Recommender systems help libraries build trust with their patrons and create a sense of presence and urgency while showing real-time notifications of patrons' activities on the website.

**Tools and Techniques for Setting up Big Data Recommendation Systems in Libraries**

Dwivedi and Roshni (2017) describes few of the big data tools that can be used to set up a recommendation system to include tools like Hadoop, Mahout, MLlib, R, Python – Crab among others.

***Hadoop:*** this is an Apache project or framework which allows the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. The main components of Hadoop are as following: Hadoop Common which supports the other Hadoop modules; Hadoop Distributed File System (HDFS) which provides access to application data and high-throughput; Hadoop YARN which is a framework for scheduling the job resource management of the hadoop cluster; Hadoop MapReduce which is based on YARN (Yet Another Resource Negotiator) system and provides parallel processing for large data sets.

***Apache Mahout:*** This is an open source, scalable machine learning library from Apache Software foundation which implements many machine learning algorithms in different categories like collaborative filtering, clustering, classification and dimensionally reduction. This tool also makes use of Hadoop for processing big data.

***MLlib:*** This tool is developed as part of the Apache Spark project as an engine for big data processing which has two packages (spark.mllib and spark.ml). MLlib as a machine learning (ML) library includes algorithms such as classification, regression, clustering, collaborative filtering, and dimensionality reduction. It is new as compared to other data mining and processing tools. The data processing using MLlib is very fast as the processing takes place in memory.

***Crab:*** Crab-scikits.recommender is a recommender engine for Python. Crab is a Python framework for building recommender engines integrated with the scientific Python packages (numpy, scipy, matplotlib). It has support for recommender algorithms like user-based and item-based collaborative filtering. The recommender interfaces can be easily combined with more than 10 different metrics like cosine, tanimoto, pearson, euclidean using, Scipy and Numpy based optimized functions.

***R:*** R is an integrated suite of software facilities for data manipulation, calculation and graphical display under GNU General Public License (GPL). R is the most popular tools as per KD nuggets software poll for Data Mining for year 2016, which provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering etc.) and graphical techniques. It also supports numerous data mining algorithm and can also be easily integrated with Hadoop for processing big data. Furthermore, it can be used for recommendation using association rules, collaborative techniques etc.

**Challenges of Big Data Application in Libraries**

Affelt (2015) and Kaladhar, Naicka and Rao (2018) noted some issues which are common to both library data and big data as listed below

***Lack of Data Analysts*:** The key issue is that data analysts need not only the skills of statistics and computer science, but also skills of domain knowledge and collaboration ability. Therefore, the challenges faced by librarians are the ability to manage the information of big data. It seems that short-course training might not be sufficient.

***Lack of Requisite Facilities:*** Successful implementation of Big data recommender systems in libraries is dependent on the availability of requisite facilities such as ICTs, internet and network services, Softwares, constant power supply, etc. which are mostly not available in some libraries in Nigeria, most especially in some public, special and school libraries.

***Budget Issues:*** Although more and more people understand the great benefit of using big data analysis, the IT investment such as analytics servers, high-performance computing servers are needed. It seems that most of library administrations have not yet placed big data on the table because of shrinking budgets as well. Research data managed by projects are paid less attention due to the challenge of human resources. Moreover, a lot of research data which were produced ten year ago is still analogue, digitizing these resources is not a simple task, which need a lot of time and personnel resource.

***Technical Challenges:*** Big data involves techniques such as capturing, storing, processing and presenting data. Data in the library have different types and might be in various statues. Some data might be waiting for digitalization. On the other hand, a large set of data often contains

some dirty or false data. Therefore, correctly removing those noises needs some work. Due to heterogeneous types and formats of research data, integrating them become a very tough job. Many types of research data are considerably less usable when they are in their raw state than after they have had filters or algorithms or other processing performed on them. Those work need budget to build tools and provide other supports as well.

***Privacy:*** Big data is mining the data and discovering knowledge and therefore, has privacy issue. New risks of system intrusions might arise due to the accessibility of a large amount of data. In order to receive the most accurate and correct recommendation, the system must acquire the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Recommender systems raise privacy concerns as the behavior data they utilize could potentially sacrifice users' privacy. Naturally, the question of reliability, security and confidentiality of the given information arises. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs. Data security issues have not been well considered for library big data concept (Wanaskar, Vij & Mukhopadhyay, 2013).

***Cold-Start:*** Cold start problem refers to the situation when a new user or item just enters the system. There are three kinds of cold start problems: new user problem, new item problem and new system problem (Sharma & Mann, 2013). In such cases, it is really very difficult to provide recommendation as in case of new user, there is very less information about user that is available and also for a new item, no ratings are usually available and thus collaborative filtering cannot make useful recommendations as in the case where we have new item as well as new user. However, content-based methods can provide recommendations if there is a new item as they do not depend on any previous rating information of other users. These problems can be solved using the hybrid approach.

***Scalability:*** Scalability is the property of a system which indicates its ability to handle growing amount of information in a graceful manner. With enormous growth in information over internet, it is obvious that the recommender systems are having an explosion of data and thus it is a great challenge to handle this continuously growing demand. Some of the recommender system algorithms deal with the computations which increase with growing number of users and items. In CF computations grow exponentially and get expensive, sometimes leading to inaccurate

results. Methods proposed for handling this scalability problem and speeding up recommendation formulation are based on approximation mechanisms. Even if they improve performance, most of the time they result in accuracy reduction (Tiwalola & Asafe, 2015)

***Over Specialization Problem:*** Users are restricted to getting recommendations which resemble those already known or defined in their profiles in some cases, and it is termed as over specialization problem (Chen, Wu, Xie & Guo, 2011). It prevents user from discovering new items and other available options. However, diversity of recommendations is a desirable feature of all recommendation systems. After solving the problem using genetic algorithms, user will be provided with a set of different and a wide range of alternatives.

**Conclusion**

In today's digital world, including recommendation systems is an investment worth making because users are constantly dependent on these innovative services for informed decision making. Recommender systems will not only enhance the user experience and engagement but also generate more revenue and accord more recognition to library profession. However, developing a recommendation system takes a significant understanding of data. Your recommendation system is only as effective as it is built to be.

**Way Forward**

**Adequate training and constant updates:** Library profession should ensure that they have members who specialize in data analytics and are constantly following the trends in data science.

**Provision of Adequate facilities**: The government and the management of library should make availability the necessary facilities to enable the practice of recommender systems in the library thereby improve their services

**Addressing Privacy-Related Concerns with Big Data:** Datafication model and Resource Description and Access (RDA) are two way privacy issues can be addressed. The datafication paradigm aids in illuminating new privacy issues. It provides a fresh perspective on modeling privacy and the privacy effects of massive data. The library's catalogs can be improved via Resource Description and Access (RDA) and linked data, while there is presently debate in the

field of library and information science about these strategies. Compared to big data, RDA and linked data are more accurately reflective of the more established cataloging system. Using RDA, catalogers can more easily create and encode bibliographic associations.

**Overcoming technical issues associated with big data:** Most data come in an unstructured format and needs mining and cleaning in order to bring out meaning therefore librarians should major in machine learning and also collaborate with system experts to build recommender systems in the library.

## References


Adetayo, A. J., Adeniran, P. O. & Gbotosho, A. O. (2021). Augmenting traditional library services: role of smart library technologies and big data. *Library Philosophy and Practice (e-journal)*. 6164. https://digitalcommons.unl.edu/libphilprac/6164

Addepalli, S. L.,& Addepalli, S. G. (2014). Library management system using RFID technology. *International Journal of Computer Science and Information Technologies,* 5(6), 6932–6935. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.657.7931&rep=rep1&type=pdf

Adeniran, P. (2011). User satisfaction with academic libraries services: Academic staff and students perspectives. *International Journal of Library and Information Scienc.* 3(10): 209-216.

Affelt, A., (2015). *The accidental data scientist: big data applications and opportunities for librarians and information professionals*. Medford, New Jersey.

Ball, R. (2019) Big Data and Their Impact on Libraries. *American Journal of Information Science and Technology*. 3(1), 1-9. doi: 10.11648/j.ajist.20190301.11

Bamgbade, B. J., Akintola, B. A., Agbenu, D. O., Ayeni, C. O., Fagbami, O. O., & Abubakar, H. O. (2015). Comparative analysis and benefits of digital library over traditional library. World Scientific News, 24, 1–7. http://psjd.icm.edu.pl/psjd/element/bwmeta1.element.psjd-12f68008-735d-4b63-af6f-9748f444c5d4/c/WSN_24__2015__1-7.pdf

Chang, A. C. (2016). Big data in medicine: The upcoming artificial intelligence. *Progress in Pediatric Cardiology*. 100(43), 91–94. https://www.infona.pl/resource/bwmeta1.element.elsevier-88d9ca39-ebb9-3567-80b3-387f45aa131d

Chen, Y., Wu, C., Xie, M., & Guo, X. (2011). Solving the sparsity problem in recommender systems using association retrieval. *Journal of Computers.* 6(9),

Cox, J. (2018). Positioning the academic library within the institution: a literature review. *New Review of Academic Librarianship*. 24(3–4), 219–243. https://doi.org/10.1080/13614533.2018.1466342

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*. 65(3), 122–135. https://doi.org/10.1108/LR-06-2015-0061

Djafri, L., Bensaber, D. A., & Adjoudj, R. (2018). Big Data analytics for prediction: parallel processing of the big learning base with the possibility of improving the final result of the prediction. *Information Discovery and Delivery*. 46(3), 147–160. https://doi.org/10.1108/IDD-02-2018-0002

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research.* 69(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001

Frederick, D. E. (2016). Data, Open Science and libraries – The Data Deluge Column. *Library Hi Tech News*. 33(8), 11–16. https://doi.org/10.1108/LHTN-09-2016-0040

Gupta, V. and Pandey, S. (2019). Recommender systems for digital libraries: a review of concepts and concerns. *Library Philosophy and Practice (e-journal).* 2417. https://digitalcommons.unl.edu/libphilprac/2417

Hoy, M. B. (2014). Big Data: An Introduction for Librarians. *Medical Reference Services Quarterly.* 33(3), 320–326. https://doi.org/10.1080/02763869.2014.925709

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273

Ishwarappa, & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and hadoop technology. International Conference on Computer, Communication and Convergence (ICCC 2015), 48(C), 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Kabore, S. C. (2012). Design and implementation of a recommender system as a module for Liferay portal. Master Thesis: Barcelona School of Computing (FIB), University Polytechnic of Catalunya (UPC) Master in Information Technologies.

Kaladhar, A., Naick, B. R. D. & Rao, K. S. (2018). Application of Big Data Technology to Library data:A review. *International Journal of Library and Information Studies.* 8(2). 25-30.

Khuntia, S. K., Mishra, M., Ramesh, D. B., Librarian, J., librarian, A., & Librarian, C. (2016). Applicability of information technology in libraries with a step ahead to smart library in 21st century. Indian Journal of Library Science and Information Technology, 1(1), 22–25. https://www.ipinnovative.com/media/journals/IJLSIT_1(1)_22-25.pdf

Kościelniak, H., & Puto, A. (2015). Big data in decision making processes of enterprises. Part of Special Issue: International Conference on Communications, Management, and Information Technology (ICCMIT'2015), 65, 1052–1058. https://doi.org/10.1016/j.procs.2015.09.053

Liu, S., & Shen, X. L. (2018). Library management and innovation in the Big Data Era. *Library Hi Tech*. 36(3), 374–377. https://doi.org/10.1108/LHT-09-2018-272

Moulding, J. (2016, February 29). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020. https://www.v-net.tv/2016/02/29/cisco-visual-networking-index-global-mobile-data-traffic-forecast-update-2015–2020/

Nicholson, S. W., & Bennett, T. B. (2016). Dissemination and discovery of diverse data: do libraries promote their unique research data collections? *International Information and Library Review*. 48(2), 85–93. https://doi.org/10.1080/10572317.2016.1176448

Preater, A. (2014). Information as a commodity. https://www.preater.com/2014/06/03/information-as-a-commodity/

Pujar, S. M., & Satyanarayana, K. V. (2015). Internet of Things and libraries. *Annals of Library and Information Studies (ALIS)*. 62(3), 186–190. http://op.niscair.res.in/index.php/ALIS/article/view/9800

Rahhali, M., Oughdir, L., & Jedidi, Y. (2021). E-Learning Recommendation System for Big Data Based on Cloud Computing. *International Journal of Emerging Technologies in Learning (iJET), 16*(21), 177–192. https://doi.org/10.3991/ijet.v16i21.25191

Schöpfel, J. (2018). Smart Libraries. *Infrastructures,* 3(4), 43. https://doi.org/10.3390/infrastructures3040043

Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*. 51(3), 771–780. https://doi.org/10.1177/0961000617742465

Sengupta, S. (2016). *Big Data: The Next Big Opportunity for Librarians.*, 2016 . In ICCLIST 2016, New Delhi,India, 15.1.2016. [Conference paper]. http://eprints.rclis.org/34349/1/final-proceeding-iicclist16-2s-view-50-51.pdf

Sharma, M., & Mann, S. (2013). A survey of Recommender Systems: Approaches and Limitations. *International Journals of Innovation in Engineering and Technology*, ISSN:2319-1058

Tarique, M., & Rani, V. P. (2017). Implementation of RFID in library management system based on Internet of Things (IOT). *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2(3), 315–321.

Techlabs, M. (2021). Types of Recommendation Systems & Their Use Cases. https://medium.com/mlearning-ai/what-are-the-types-of-recommendation-systems-3487cbafa7c9

Tiwalola, A. B. & Asafe, Y. N. (2015). A comprehensive study of recommender systems: prospects and challenges. *International Journal of Scientific & Engineering Research.* 6(8). 699-714

Wanaskar, U. H., Vij, S. R., & Mukhopadhyay, D. (2013). A Hybrid Web Recommendation System based on the Improved Association Rule Mining Algorithm. Department of Computer Engineering, Pune, India.

Wang, C., Xu, S., Chen, L., & Chen, X. (2016). Exposing library data with big data technology: A review. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science, ICIS 2016 - Proceedings, 1–6. https://doi.org/10.1109/ICIS.2016.7550937

Zhan, M., & Widén, G. (2019). Understanding big data in librarianship. *Journal of Librarianship and Information Science*. 51(2), 561–576. https://doi.org/10.1177/0961000617742451