

Received March 24, 2020, accepted April 24, 2020, date of publication April 27, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2990765

# Toward Effective Planning and Management Using Predictive Analytics Based on Rental Book Data of Academic Libraries

NAEEM IQBAL<sup>ID</sup>, FAISAL JAMIL<sup>ID</sup>, SHABIR AHMAD<sup>ID</sup>, AND DOHYEUN KIM<sup>ID</sup>

Computer Engineering Department, Jeju National University, Jeju 63243, South Korea

Corresponding author: Dohyeun Kim (kimdh@jeju.ac.kr)

This work was supported in part by the Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT under Grant 2019M3F2A1073387, and in part by the Institute for Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korean Government (MSIT) (AutoMaTa: Autonomous Management framework based on artificial intelligent Technology for adaptive and disposable IoT) under Grant 2019-0-01456.

**ABSTRACT** Large scale data and predictive analytics are the most challenging tasks in the field of academic data mining. Academic libraries are a great source of information and knowledge to provide a wide range of services to meet end-user requirements. Due to the rapid changes in the educational environment and availability of huge library rental book data, it is required to utilize data mining and machine learning techniques in the context of the academic library to extract and analyze underlying knowledge from rental book data, which is important to facilitate library administration to drive better future decisions to improve and manage library resources effectively. These are the following resources, such as managing future demands of the library books, selection and arrangement of the books, operational efficiency, and also improve the quality of interaction between the library and end-users, etc. This work uses and analyzes a real dataset collected from the library of Jeju National University, the Republic of Korea. The dataset contains 2,211,413 rental book records including 173671 unique book records, 57203 unique number of the rental user, and 78 data parameters. In this paper, we propose a novel model to analyze and predict library rental book data to facilitates library administration in order to plan and manage library resources effectively and provide better services to end-users. The proposed model consists of two different modules; library data analysis and prediction modules. Firstly, we use data mining techniques to analyze and extract useful underlying patterns from library rental book data, which can lead to plan and manage library resources effectively. Secondly, a novel prediction model is proposed based on Deep Neural Network (DNN), Support Vector Regressor (SVR), and Random Forest (RF) to predict future usage of the academic libraries rental books. The performance results of the implemented regression models are evaluated in terms of MAE, MSE, and RMSE. In this paper, it is found that the DNN model performs significantly better than SVR and RF. The experimentation results show that the proposed model improves the future usage of library books to facilitate library administration to plan and manage library resources effectively. Based on the proposed model results, the academic library administration can easily plan and manage resources effectively to provide quality services to end-users.

**INDEX TERMS** Data mining, academic libraries, big data, machine learning, predictive analysis.

## I. INTRODUCTION

In the recent past, collection management processes of academic libraries such as selection, the arrangement of the books, issue/return books, and disposal of selected books are conducted using some static rules-based techniques or

The associate editor coordinating the review of this manuscript and approving it for publication was Xin Luo<sup>ID</sup>.

the intuitions of the library staff [1]. All these processes of the library should be improved based on the actual use of book data to meet the future demands of the academic library and improve operational efficiency in order to provide better services to the academic library users. The statistical methods of traditional libraries have become discommodious and low-efficiency in assisting library users [2]. Due to the huge volume of student's book-loan data, it is required to analyze

and predict rental book data for both library users and staff to make better future decisions like managing future demands of books effectively, provide better services to the library users to enhance the operational efficiency of the library system, etc. The purpose of analyzing library rental book data is to enhance the quality of the mutual effect between the academic library and its users [3].

There are different methods used to improve collection management processes of the library to predict future book use based on past book information. The prediction is the process that aims to predict future use of the given books, such as the number of book loans for the given books data by analyzing historical data. C. Silverstein *et al.* [2] proposed a model to solve the problem of the selection of books based on the frequency of use for the selected books. The authors also used and applied the model to the arrangement process of library books. G. E. Evans *et al.* [4] also proposed a model to investigate the use of library books based on the personal history of selected books in several libraries. S.Pritchard *et al.* [5] also investigated the future use of the library books based on the manner of selection. All of the previous work is based on the classification book-loan records according to various attributes. Therefore, rental book data is a reasonable choice of a resource for the academic libraries to plan and develop their services to meet the future demands of the library and its users effectively. The formation of library services can better assist based on examining and analyzing rental book data using data analytics techniques.

Big Data analytics is one of the complex processes of examining large and diverse volumes of the dataset that include structured and unstructured datasets from different sources in order to discover useful information, such as hidden patterns, market trends, correlations, user interest, and other insights [6]. It can be helpful for the academic library management to understand the hidden information of the large rental book dataset to make better future business decisions. There are different specialized tools available to analyze and examine large datasets, such as predictive analytics, data optimization, and data mining, to extract useful knowledge. These tools enable organizations to analyze and process diverse sets of data to determine the relevant and valuable patterns that are most important to drive better future business decisions. Big Data Analytics provides various types of business benefits, such as better customer service, enhanced operational efficiency, create new opportunities to increase business revenue, and potential advantages over rivals due to effective marketing trends, etc. There are different machine learning (ML) algorithms are available to analyze the huge volume of data to extract hidden information, such as predictive analytics, data and text mining, etc.

With the advancement in ML, it is quite possible to process and extract hidden patterns from rich information provided by big data and to build predictive models to drive a conclusion [7]. The prediction algorithm is the essential module of every prediction system and directly impact the performance and prediction results of the system.

Nowadays, deep neural network (DNN) has been used in various fields [2] of computer science, computer vision, and speech recognition, etc [8]–[10]. DNN is an emerging field of machine learning research [11], which is used to solve complex learning problems of multiple areas [12]–[14]. DNN model is a robust and intelligent to automate data representation in the learning (training) process as compared to the traditional models. Therefore, it is often considered to replace the manual process of features construction by automatic data representation process. It is an artificial neural network with several hidden layers between the input and output layers. DNN model is very robust and efficient to deals with a large number of machine computations and big data by enabling the machine to learn by increasing the number of hidden layers [15]. Many researchers using DNN to design prediction systems, but most of these systems use additional algorithms like text mining, data mining, and other audio information, etc. to enhance the performance of the system. Similarly, a support vector regressor (SVR) is a regression model used for prediction. The SVR is a well-known learning algorithm and works similarly as the support vector Machine (SVM) for classification with some minor differences [16], [17]. The main advantage of SVM/SVR is that it is not model dependent, and its guarantees to give the optimal prediction [18]. Random Forest (RF) is another supervised learning algorithm for the regression model. It uses ensemble learning techniques to train the model by constructing multiple decision trees (DTs). The main idea is to combine multiple DTs to determine the output for the trained model, which is more robust and efficient than relying on a single DT [19], [20].

The usage of academic library books has also shown a significant contribution to the academic success and retention process of students [21], [22]. The results of past researches indicate that students at different performance levels borrow the different numbers of books from the library. Academic libraries are known as the hub of useful information and knowledge to provide several services for its users. Based on historical data, it found that the quality of the library evaluated based on its vast collections. The huge collection of the academic library means to provide plentiful materials for users for the self-learning process. Therefore, many researchers and data scientists are focused on managing academic library resources effectively to provide better services to both academic library staff and end-users; however, processing big data and extracting useful information for better future direction and prediction is still a challenging task in academic data mining.

In this paper, we propose a novel Data and Predictive Analysis based on a Library Rental Book Data (DPA-LRBD) model to utilize data mining and machine learning techniques in the context of the academic library. Our DPA-LRBD model is mainly focused on unearthing useful information and underlying patterns to predict future usage of library books, which aims to facilitate library administration and educationalists to plan and manage library resources

effectively and provide better quality services to end-users. This work analyzed and evaluated a real library rental book dataset acquired from the Jeju National University, the Republic of Korea. The historical dataset contains 2,211,413 rental records including 173671 unique book records, 57203 unique numbers of the rental user, and 78 data parameters. Our proposed DPA-LRBD model consists of two modules, such as data analytics and predictive analytics models. Firstly, we performed and applied data mining techniques to extract hidden patterns, relevant information, and knowledge, which is important to make better future decisions in order to plan and manage library resources effectively, i.e., future demands of the library books, operational efficiency and provide better services to the end-users. This work used different analysis techniques, i.e., statistical analysis, time series analysis, frequency ranking, user-gender based analysis, book type-based analysis, user's age-group, and home address based analysis, etc., to process and investigate the book-loan data to achieve the research objectives. Secondly, a novel prediction model is proposed based on Deep Neural Network (DNN), Support Vector Regressor (SVR), and Random Forest (RF) to predict future usage of the academic libraries rental books to improve the digital library processes such as selection, the arrangement of the books, issue/return books, and disposal of the library books. The proposed DPA-LRBD model entails the following steps, such as the collection of rental books data, data preprocessing, features constructions, splitting the dataset into subsets such as training and testing sets, training/testing machine learning algorithms, and performance evaluation. Our proposed DPA-LRBD model used the following performance evaluation measures, such as MAE, MSE, and RMSE, to assess the performance of all implemented regression models. The experimentation results show that predicting the future use of the library books based on historical rental book data improves and manages the library resources effectively, and also improves the availability of the library books for end-users, such as students and other book readers. Based on the proposed DPA-LRBD model results, library management can easily plan and manage library resources in an effective way to avoid any shortage of books and provide better services to the library end-users.

The main contributions of this study are followed as:

- The main objective of this study is to utilize data mining and machine learning techniques in the context of the academic library to facilitate library administration and educationalists to plan and manage library resources effectively.
- The proposed DPA-LRBD model analyzed and investigated real data of the rental book of the Jeju National University to facilitate library administration.
- The proposed DPA-LRBD model applied data mining techniques to unearth hidden knowledge from the rental book dataset, which is important for library administration to better plan and manage library resources effectively.

- Our work employed the following analysis techniques, such as time series analysis, statistical analysis, book types-based analysis, gender-based analysis, age group, and home address based analysis.
- Based on discovered knowledge, our proposed DPA-LRBD model applied predictive analytics techniques to predict future usage of library books.
- Finally, we demonstrate the effectiveness of the implemented DPA-LRBD model using the following performance measures, such as MAE, MSE, and RMSE. Our proposed model results indicate that the DNN model performed accurately in the prediction process as compared to SVR and RF algorithms.

The rest of this paper is organized as follows. In Section II, Literature review is described briefly. Existing research work in the relevant field is highlighted. Section III presents the proposed data analytics model to analyze and investigate the library rental book data. Section IV presents the predictive analytics model based on machine learning algorithms. Section V presents the experimental environment and results achieved during the experiment. Finally, the paper is concluded in Section VI.

## II. LITERATURE REVIEW

This section presents the literature review and background study for the academic library data analysis and prediction model based on library rental book data. Nowadays, an academic library faces different challenges due to processing and analyzing a large volume of data to drive better future decisions. The first challenge is to process, analyze and investigate a massive volume of data to perform data analysis to discover valuable information and underlying patterns to drive better future decisions. The second challenge is to develop an ML-based platform, which can use to process and extract hidden patterns from large data to enhance business strategies because traditional systems are not capable of processing and analyzing large and diverse sets of data effectively. Currently, there are different research models have been developed based on academic library data, such as student performance prediction model [23], which used the consecutive three years of student's book-loan dataset to predict student performance using a supervised content-aware matrix factorization model. There are some book recommendation models based on user's behavior patterns and profile information [24]–[32]. The other models include a probabilistic keyword model (PKM) [26] based on the collaborative filtering (CF) approach to enhance the performance of the book recommendation systems, a rating based book recommendation system [30] which construct the rating based on user borrowing history, a map-reduced based book recommendation system [31] that recommends books based on users interest includes user statistics and search history data, a weighted graph model [32] for book recommendation systems that is also similar to the association rule method, etc. In [4], G. E. Evans *et al.* proposed a model to investigate the use of library books based on the personal history of selected books

in several libraries. In [5], S.Pritchard *et al.* investigated the future use of the library books based on the manner of selection but the authors used traditional approaches to investigate future usage of books, which takes too much time and cost to process data manually.

Nevertheless, none of the afore-mentioned research models approach the challenge of processing, analyzing and investigating a large volume of the rental book data to utilize data mining in the context of a library to facilitate library administration. These models are not useful to extract unearthen information and valuable patterns from rental book data to help library administration in order to plan and manage library resources such as future book demands to avoid any shortage of books. Moreover, all the mentioned models used the library data to facilitate book readers, whereas our work is not only specific to book readers but also facilitates library administration to plan and manage academic library resources effectively in order to provide quality services to end-users. Our work is based on real data of the rental books collected from Jeju National University to perform pattern mining and prediction to facilitate the library administration, while there are different existing systems dealing with a general mining and prediction challenges. The academic library data is a valuable source for the library administration and data mining to extract underlying patterns and useful information from large rental book data [33]. Therefore, it is vital for the researchers to utilize data mining and machine learning techniques in the context of an academic library to help library administration to plan resources effectively.

Data mining techniques are widely used to analyze and extract hidden relationships and trends from a large amount of data. In [34], the authors proposed a latent factor based model to extract hidden knowledge from high dimensional and sparse matrices. According to the study [35], every organization with a large volume of data provides a platform to utilized data mining techniques to discover hidden knowledge to enhance work progress. There are different pattern mining techniques developed to facilitate academic library system search based on keyword mapping and similarity between the book and user preference [24], [27], [28], [31], [32], [36]. S.T. Yang *et al.* [24] proposed a book-acquisition recommendation model for the university libraries to satisfy borrowers' demands based on historical book inquiry data. This work used text mining and internet technologies to provide suggestions for book acquisition. In this work, the authors collected data from keywords used by borrowers where they have not found particular books within the university library system. Thus, the authors used these extracted keywords and matched with the books database of the bookseller database to obtain the recommended books list, and then the librarians finalize order for new books based on recommended book lists. In [27], the authors utilized the data mining approach to extract user's behavior patterns from the book-loan data log to predict and recommend library books to the students. This work used association rules and mainly focused on knowledge discovery of different majors. The study presented

in [28] also evaluated the book loan dataset using association rules to extract hidden knowledge to predict and recommend books. All these mentioned methods utilized the data mining approach to focus and enhance meta and content search in academic library systems whereas our work utilized following data mining techniques, such as time series analysis, statistical analyses, book types, and gender-based analysis, age group and home address based analysis to discover the useful knowledge from rental book data. Finally, machine learning (ML) techniques are used to analyze discovered knowledge to predict future book demands, which is most important for library administration to drive better future decisions.

There are various machine learning (ML) techniques available to analyze and process discovered patterns from large data automatically in order to build a robust predictive model [15], [37]. Machine learning techniques are commonly used in predicting the future outcome of the resources [38]. D. Lian *et al.* [23] proposed a supervised content-aware matrix factorization model to predict student performance and predict the right books for the students. In this work, the authors used the consecutive three years of student's book-loan dataset for the evaluation of the proposed model, which includes 13,047 undergraduate students, cumulative grade point (CGP) of three consecutive years, and 676,757 records. The dataset consists of the following data attributes like the total number of students, the total number of books, total records of book loans, and the total number of books borrowed. This work achieved an accuracy of 64.38% and used a 5-fold cross-validation method to validate prediction results. In [39], M. Kitajima *et al.* investigated the process of book selection in an academic library based on student's book-loan records. In this work, the authors used Nippon Decimal Classification (NDC) technique to classify the library books and also the turn over rate for each classified category investigated. This work used the book-loan dataset of the academic library of Kyushu University (Japan). The dataset consists of the book accession records from April 2000 to March 2013 and student's borrowed books record from April 2012 to March 2013. The evaluation results of this work shown two different viewpoints for the books selection process in university libraries such as (i) the relationship between the turnover rate of each classified category and the books accession rate and (ii) the relationship between the turnover rate of each classified category and the lab housing rate. In [40], the authors proposed a prediction model to investigate library lending, the total number of readers, and the collection of the university library in China. In this work, the author considered the university library data from August 1993 to September 2010 to predict library lending, the number of readers, and library collection. The authors conducted two evaluation tests, i.e., the f-test and t-test, to found that a strong relationship exists between library lending, readers, and library collection, which collect at the end of every academic year. According to this study, a change in library lending depends on the change in the total number



of library readers and the collection of libraries. All these ML-based models utilized library book data to enhance academic performance and book selection for students whereas our model utilizes ML techniques to predict future demands of books for library stakeholders to improve planning and managing of the library resources.

To best of our knowledge, all the existing models based on academic libraries data are used to facilitate and help students and book readers, whereas our DPA-LRBD model is specific to utilize data mining techniques in the context of the academic library to facilitate library administration. Our work utilized data mining techniques to process and unearth useful information and underlying patterns from a large rental book data collected from Jeju National University, Jeju self-governing province, the Republic of Korea. Finally, we utilized ML techniques to analyze extracted knowledge in order to build a predictive model to predict future demands of academic library books. This predictive model is used to improve the planning and managing of the library resources in order to provide better quality services to end-users, such as book readers. To best of the author's knowledge, it is a novel idea for researchers to utilize data mining and machine learning techniques in the context of the academic library to facilitate library administration.

### III. DATA ANALYSIS OF ACADEMIC LIBRARY RENTAL DATA

Data analysis is the process of acquiring, exploring, cleaning, transforming, and modeling data to discover useful information and present results [41]. Data analysis is the process of examining and modeling a large set of data of different formats, such as structured and unstructured. It is very useful to analyze and investigate the huge volume of data to extract valuable information and hidden patterns, which can play an important role in making business future decisions. It allows to analyze and evaluate the huge volume of data through analytical and logical reasoning to generate useful information to make better future decisions. The following steps are carried out to generate useful information and hidden patterns from the large data such as collection or acquisition of data, data exploration to visualize data and view the data distributions of the dataset variables, cleaning acquired data in order to remove irrelevant information and all other outliers, transformation of cleaned data to get uniformity in order to increase reliability of data, and building model using analytical tools to get some results.

Big Data analytics is one of the complex processes of examining large and diverse volumes of the dataset that include structured and unstructured datasets from different sources in order to discover useful information, such as hidden patterns, market trends, correlations, user interest, and other insights [42]. Big Data analytics is a challenge for most of the business organizations [43]. Because a huge volume of data having different formats, such as structured and unstructured data, stored in different places and different systems across the organization is unable to process and analyze by

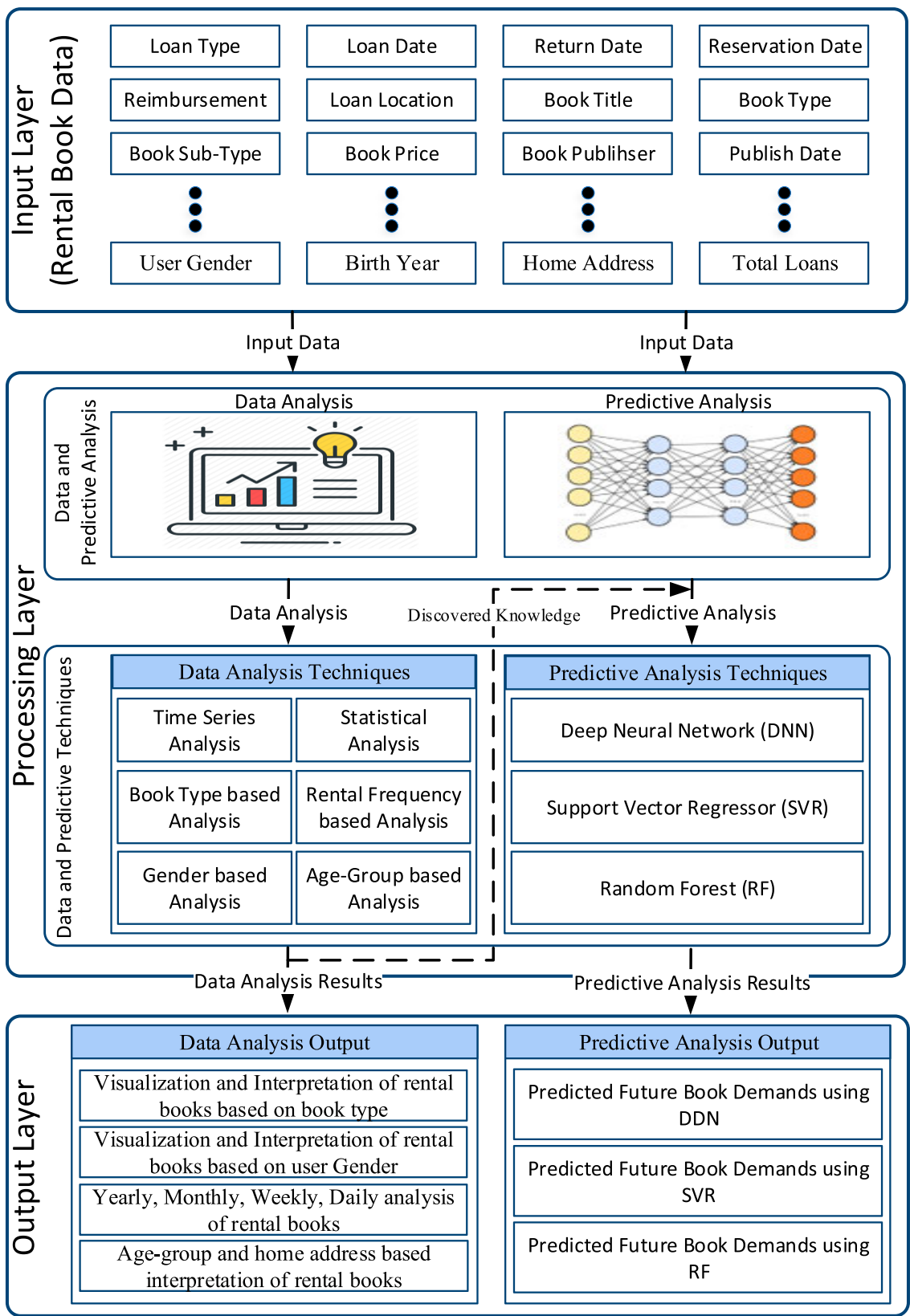
**TABLE 1. Data distribution of library rental book raw data.**

Data Distribution Parameters	Total Records
Total Number of Rental Records	2,211,413
Unique Books	173,671
Unique Users	57, 203
Time Span	2007 to 2019

the traditional systems. Also, data sparsity and high dimensional matrices are also found in big data analytics. Therefore, it is a vital issue to predict missing values on known values for such type of matrices [44], [45]. There are two major challenges faced by the business organization having a huge volume of data. The first challenge is to breaking down the large volume of data to access all data of the organization stored in different places and then perform data analysis to find hidden patterns and useful information to drive a better future business decision. The second challenge is to provide a platform that can transform a massive volume of unstructured data into a structured form to extract knowledge to enhance business strategies because traditional systems and databases are not capable of processing large and diverse sets of data effectively.

In this paper, we collected a library rental book dataset from Jeju National University, Jeju province, the Republic of Korea, to examine and analyze large rental book data to extract hidden information and knowledge to improve library services and operational efficiency of library management. The statistics of the collected library rental book data given in Table 1. Data mining techniques are applied to preprocess the dataset in order to clean rental book data and handle missing values, selection of relevant and useful attributes to increase the consistency and efficiency of the data, and reduce data storage cost as well as analysis cost. After preprocessing of the library rental book data, we applied data analysis techniques to find and discover hidden information from the preprocessed data that is most important to the library management to make better future decisions, such as to manage future demands of books effectively, provide quality services to the library users and also improve operational efficiency as well. Data visualization techniques are also used to visualize and display extracted patterns and trends in some graphical formats, such as charts and graphs, etc. These techniques are very useful to analyze and investigate discovered information from a massive volume of the library data to assess library resources and services to support and make data-driven decisions to improve operational efficiency and provide better services to the library users. The basic flow of the proposed model is shown below in figure 1.

The following steps are carried out to examine and analyze the library rental book data to discover hidden information to make better future decisions like managing future demands of books effectively, provide better services to the library users to enhance the operational efficiency of the library system, etc.



**FIGURE 1.** Block diagram of the proposed data and predictive analyses model based on library rental book data.

**TABLE 2.** Library data features and description.

#	Feature	Description
1	Loan type	It represents the type of loan such as Paperback
2	Reservation date	It represents book reservation status that will be reserved for a default period of 270 days.
3	Loan Date	Issue data of loan book
4	Return Date	Return data of borrowed book
5	Reimbursement	Reimbursement provides lowans with equal access to library resources
6	Return delay	Return a borrowed book late
7	Loan location	It represents book loan location
8	Return place	It represents the place where a user return book
9	Reservation	The date on which reservation is booked.
10	Reservation cancellation type	It is an amount of time if a user failed to return a book within a specified amount of time, then his/her reservation will be canceled.
11	Book title	Title of a book such as Harry Potter
12	Book author	Author of Book
13	Book publisher	Represents publisher of the books
14	Book publish year	Year of the Book publication
15	Book publish date	Date of the Book Publication
16	Import category	Category of the book
17	Media classification type	Media classification, i.e., as print media
18	Type classification	Classification of book data
19	Book price	Price of the Book in KRW.
20	Book type	Category of the Book, i.e., Literature
21	Book sub type	Subcategory of the Book, i.e., Korean Literature
22	User gender	It represents the rental user gender, i.e., male and female.
23	Birth year	Year of birth of the rental user
24	Home address	The home address of the rental user
25	Last loan date	It represents the last date at which the rental user borrows a book.
26	Total loans	Total number of books borrowed by the rental user
27	Loan position	The amount for book security.
28	Reason for bad membership	The reason why the membership of the rental user is not good.

- 1) Collection of library rental book data.
- 2) Data Preprocessing
  - a) Data cleaning
  - b) Handling missing values
  - c) Selection of relevant attributes
- 3) Data Mining & Visualization
- 4) Discovered Features

### A. COLLECTION OF LIBRARY RENTAL BOOK DATA

In this work, we collected a real dataset from the library of Jeju National University, Jeju Province, the Republic of Korea. The dataset contains 2,211,413 total number of rental records along with 78 parameters. There is 173,671 total number of unique book records along with 57,203 unique number of rental users and 173671 unique books. The collected rental book data is between 2007 to 2019. The metadata of the collected dataset is given below in Table 1.

### B. DATA PREPROCESSING

After the acquisition of rental books data, it is required to preprocess the acquired dataset to find and remove outliers. In this subsection, the data preprocessing technique is used to transform the raw dataset into an appropriate form for the data analysis and prediction process [46]. In this work, we performed the following steps to preprocess the library rental book data to increase the reliability of dataset and reduce the storage cost.

- 1) It is found that some duplicate records exist in rental books data. In our dataset, we identified that there are

1,814 duplicate records exist in book data. Therefore, we removed all identified duplicate book records from the dataset to increase the reliability of the dataset.

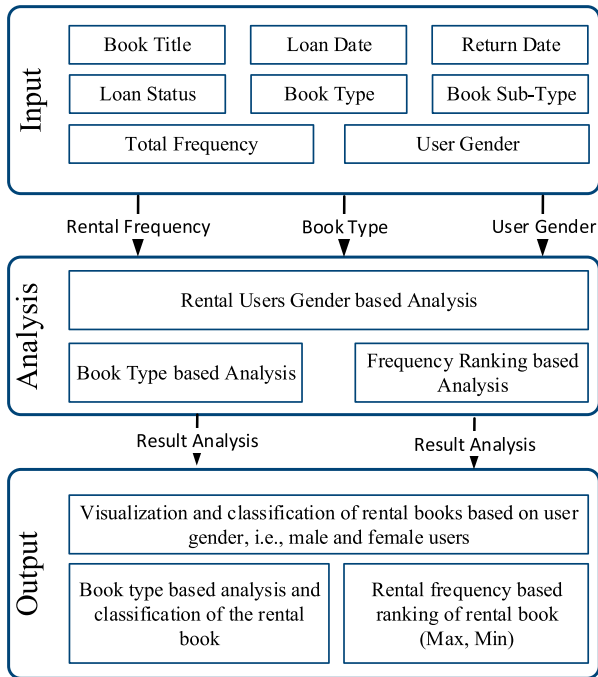
- 2) Also, All those rental records are identified and removed from the dataset that don't have book type, gender, loan date, birth year, and home address values.
- 3) The data reduction technique is used to select only relevant and useful data attributes to increase the consistency and efficiency of the data, and reduce data storage cost as well as analysis cost.
- 4) All other outliers and irrelevant data that cannot process by machine are identified and removed to increase the reliability and consistency of the library rental book data.

Finally, we considered only relevant and useful attributes to perform data mining to discover hidden patterns and useful information in order to improve the library resources. In this work, the following data features (parameters) shown in Table 2 are considered to analyze and predict the library rental book.

In the preprocessing dataset, a large feature space reduced to a small number of data feature by removing all irrelevant, static, and missing values features. The following Table 2 shown the reduced features along with their descriptions.

### C. DATA MINING & VISUALIZATION

In this work, we performed data analysis based on data mining techniques to extract and discover hidden patterns, trends and useful information from the library rental book



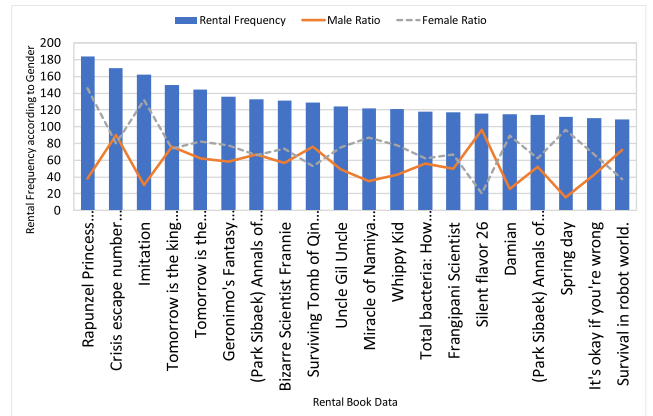
**FIGURE 2. Block diagram for rental book analysis based on rental book frequency, type, and gender.**

data to enable library management to make better future decisions [47]. In essence, data analysis allows for the evaluation of data through analytical and logical reasoning to lead to some outcome or conclusion in some context. We performed the following type of data analysis to get results from the existing dataset.

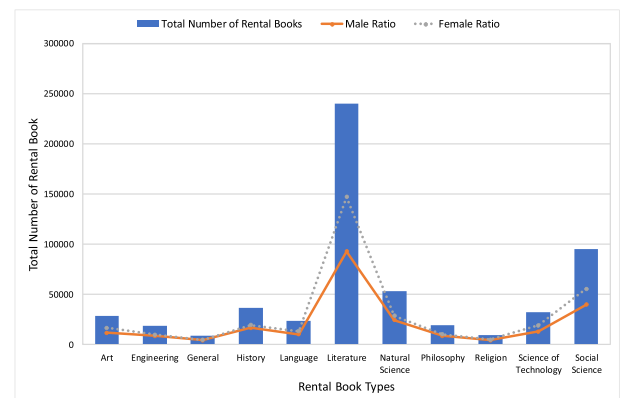
- 1) Rental book analysis based on:
  - a) Frequency based analysis of the rental book
  - b) Book type
  - c) Rental users gender
- 2) Time Series Analysis
  - a) Yearly
  - b) Monthly
  - c) Weekly
  - d) Daily
  - e) Seasonal
- 3) Rental book analysis based on rental user home address
- 4) Rental book analysis based on user age-group

**1) RENTAL BOOK ANALYSIS BASED ON FREQUENCY RANKING**

In this work, we analyzed and investigated the rental book records based on the following, such as frequency ranking of rental book and book types. The frequency ranking of the rental book describes the total count of the rental book from maximum to minimum rank. Also, the rental frequency for each book type is calculated to investigate the ratio of the rental book frequency according to book types. The basic flow of the rental book analysis based on rental frequency, book type and user gender is shown in figure 2.



**FIGURE 3. Rental book frequency ranking (maximum to minimum) data.**



**FIGURE 4. Rental book analysis according to rental book types and gender.**

The following figure 3 presents the rental book records based on the rental frequency (Maximum to Minimum). It is evident that the following rental book “Rapunzel Princess Adventure” has the highest rental frequency among all presented books. While the book which title “Survival Robot World” has the lowest rental frequency among presented books.

Furthermore, we investigated the rental book frequency based on the rental user gender, such as male and female. It is evident that female users borrowed the highest number of books as compared to the male users group.

We also investigated the rental book records based on book types. Each book type is further investigated and visualized according to rental user gender, i.e., male ratio and female ratio. The following figure 4 presents the rental book analysis according to book types, such as Art, Engineering, General, etc. It is evident that the book type “Literature” has a high rental book frequency as compared to all other listed book types. While the book type “General” has a low rental book frequency as compared to all other book types.

Furthermore, we investigated the rental book according to rental user gender based on book types. The rental book



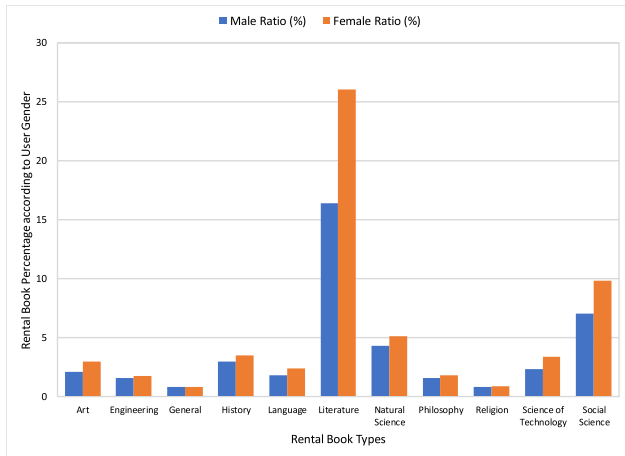


FIGURE 5. Rental book analysis according to rental book types.

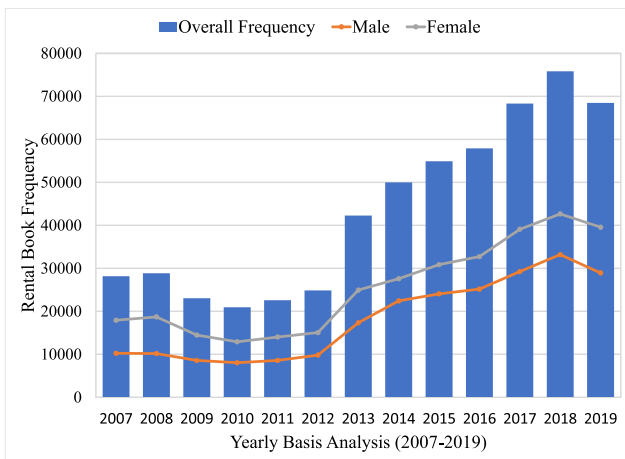


FIGURE 6. Time series analysis of rental book (Yearly basis).

frequency has analyzed for each book type based on rental user gender, i.e., male and female. It is examined that the large number of rental books borrowed by female users for all the listed book types as compared to male users.

The following figure 5 presents the rental book percentage (%) based on user gender for each book type.

## 2) TIME SERIES ANALYSIS

The time series analysis is used to generate valuable information for long-term library staff decisions. In this work, the rental books data analyzed using time series analysis to generate valuable information for long-term library staff decisions [48]. In this work, the rental book data (from 2007 to 2019) are analyzed using time series analysis (i.e., yearly, monthly, weekly, daily, and seasonal basis) to generate valuable information related to rental book frequency.

The following figure 6 presents rental book data according to user gender on a yearly basis. In the Library system, time series analysis can even give an early indication of the overall direction of a typical business cycle.

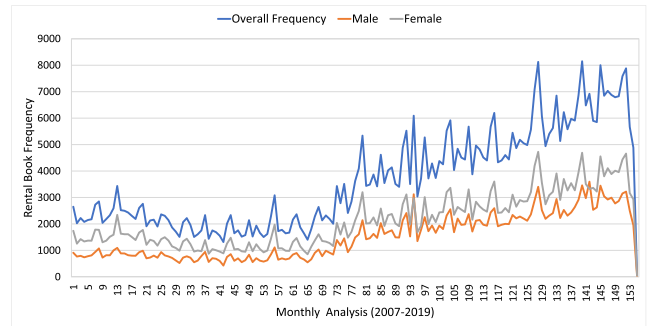


FIGURE 7. Time series analysis of rental book (Monthly basis).

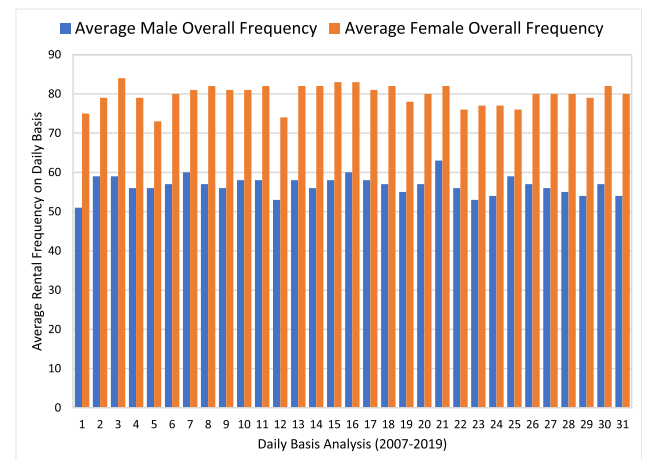
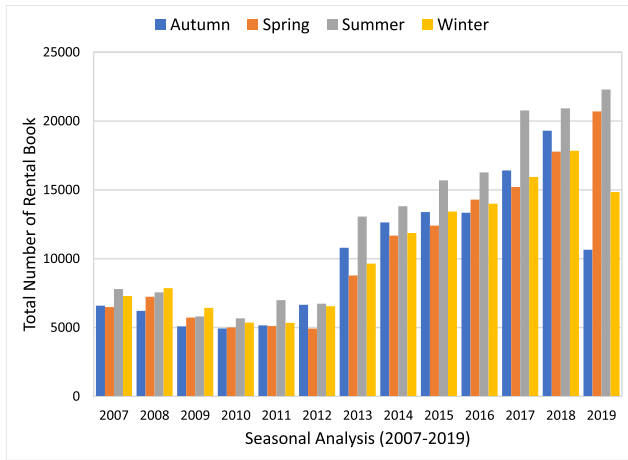


FIGURE 8. Time series analysis of rental book (Daily basis).

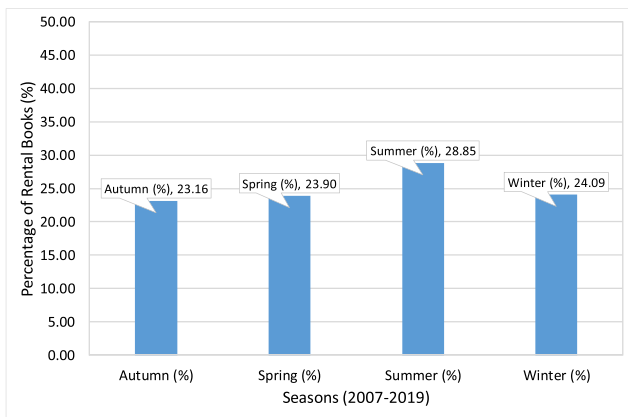
The following figure 7 is used to visualize and interpret rental book (2007-2019) based on rental frequency. In figure 7, it can be observed that the overall frequency of the rental books is classified into two groups based on user gender, such as male and female groups. Based on the monthly analysis of rental books, it is evident that the rental book borrowed by “Female” users has high as compared to “Male” users.

The following figure 8 is used to describes and visualizes the rental book (2007-2019) on a daily basis. In figure 8, it can be observed that the overall frequency of the books is divided into two groups based on the rental user’s gender. In this work, we analyzed rental book frequency based on rental user gender on a daily basis (2007-2019). It can be observed that female users have borrowed a high number of books as compared to male users. We analyzed and investigated rental books based on the user’s gender, to find the relationship between two groups, such as male and female. Based on the daily analysis results, it is evident that the average frequency of rental books borrowed by female users is high as compared to male users.

In this work, we also performed seasons based analysis to examine and investigate rental book frequency. The rental book data are analyzed based on season to get relevant information such as book frequency for each season



**FIGURE 9.** Time series analysis of rental book (Seasonal basis).



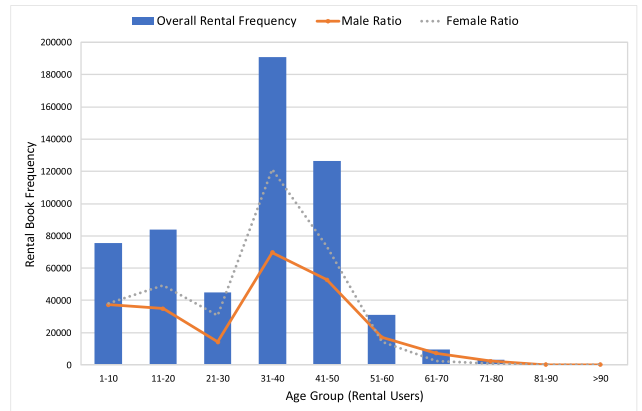
**FIGURE 10.** Percentage analysis of the library rental book based on seasons.

on a yearly basis. The following figure 9 describes rental book data according to seasons such as Autumn (September, October, November), Spring (March, April, May), Summer (June, July, August) and Winter (December, January, February). Based on the seasonal analysis, it is evident that the season known as “Summer” has the highest rental frequency (2007-2019) as compared to all other seasons, i.e., Autumn, Spring, and Winter.

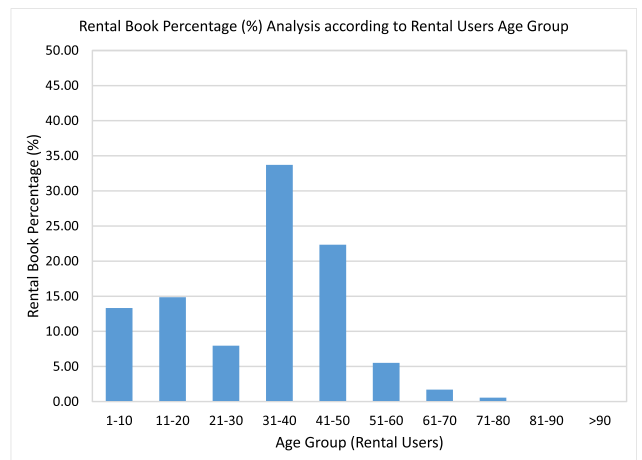
Based on seasonal analysis results, it can be observed in figure 10 that the season known as “Summer” has the highest rental frequency percentage of 28.85% as compared to all other seasons, i.e., Autumn, Spring, and Winter of 23.16%, 23.90%, and 24.09%, respectively.

### 3) RENTAL BOOK ANALYSIS BASED ON RENTAL USER AGE GROUP

The frequency of rental books is also analyzed based on the age group of rental users. In this work, we used the birth year of the rental users to investigate the frequency of books according to defined age groups. The rental book data are investigated and visualized based on user age to



**FIGURE 11.** Rental book analysis according to user gender based on age group.



**FIGURE 12.** Percentage analysis of rental book based on user age group (%).

generate valuable information related to rental users age. So, we extracted user age and categorized rental book data into 10-year age groups shown in figure 11 to presents the frequency of rental books enumerated in each age group.

In the following figure 11, it can be observed that the age group (31-40) has the highest frequency ratio as compared to other defined age groups. Furthermore, defined age groups are classified based on rental user gender, i.e., male and female. Based on the age-group analysis result, it is found that female users have the highest frequency based on defined age-group as compared to male users.

The following figure 12 presents the percentage (%) analysis of rental books based on defined user groups. It is evident that the age group (31-40) has the highest percentage (%) value of 33.71% as compared to all other defined age groups. It can also be observed that the age group (41-50) has the second-highest rental frequency percentage (%) value of 22.34%.

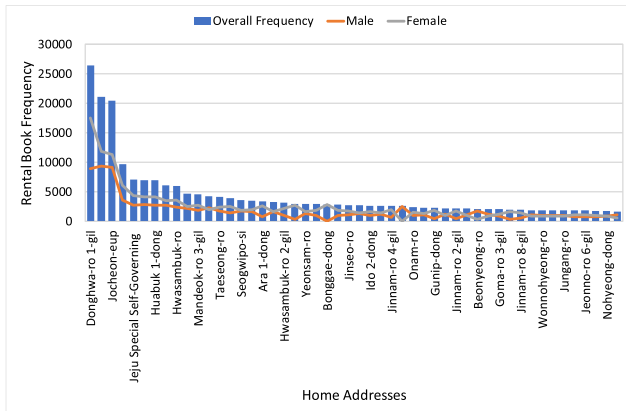


FIGURE 13. Rental book analysis according to gender based on home address.

4) RENTAL BOOK ANALYSIS BASED ON RENTAL USER HOME ADDRESS

In this paper, a home address based analysis is also performed to analyze the rental books data. To visualize rental books based on home address, location titles are extracted from the home address text to use the following location label to visualize a rental book frequency. The following parameters are used as input parameters such as book title, home Address, and return place to analyze and visualize rental book data based on rental user home address.

In this work, a home address parameter is processed to extract town or street information. After that, we used pre-processed data to perform analysis based on the user’s home address. The following figure 13 describes rental book frequency from maximum to minimum based on the rental user’s home address. Based on home address analysis, it is evident that a large number of books borrowed from the following home address “Donghwa-ro 1-gil” as compared to all other home addresses. While a small number of books borrowed by users from the following home address “Nohyeong-dong”.

Furthermore, analysis results of the home address parameter are classified into two groups, such as the total number of male rental users and the total number of female rental users. It can be observed in figure 13 that female users’ frequency is high as compared to male users. It is evident that the frequency of borrowed books by female users is high as compared to male users.

5) DISCOVERED PATTERNS AND FEATURES

Based on data mining techniques, we discovered and extracted the following hidden patterns and features listed in Table 3 from the collected dataset. Based on these discovered patterns and features, we train prediction models in order to perform predictive analysis to make better future decisions to plan library resources effectively and provide quality services to the end-users.

Based on data analysis results, it is evident that a large number of books borrowed by female users as compared to

TABLE 3. Discovered features list.

#	Features Set	Description
1	Book Type	The following features are discovered based on the rental ratio of various book types, such as art, engineering, general, history, language, literature, natural sciences, philosophy, religion, the science of technology, and social science. Also, each book type is further investigated to find a distinct rental ratio for male and female users.
2	Time Series	We used time-series based techniques to extract the following features from the library rental book, such as average rental score on a yearly, monthly, weekly and daily basis. The seasonally based features are also constructed from the library rental book data, such as the average rental ratio of the following seasons, i.e., Autumn, Spring, Summer, and Winter. For each season, we also analyzed rental book frequency based on the user’s gender in order to find a distinct ratio for male and female users.
3	Gender	We analyzed and investigated the rental book data to find rental book ratio based on the rental users gender, i.e., rental ratio of male users and rental ratio of female users.
4	Age Group	The rental book data are also evaluated based on the rental user’s age group to discover features related to different defined age groups. First, the rental book data are categorized into 10-year age groups. Second, we computed rental book ratio for each defined age-group in order to investigate the rental book frequency based on the rental user’s age.
5	Users Home Address	We analyzed and investigated the library rental book data according to the rental user’s home address to extract the rental book ratio for each unique home address. Also, each home address is further processed to find and analyze the rental book relationship between male and female users.
6	Statistical Features	The statistical features are also constructed from the raw dataset to examine the rental book frequency statistically, such as average (mean), standard deviation, histogram and correlation etc.

male users. The rental book transactions are analyzed and evaluated based on book types and found that book type “Literature” is the large number of rental transactions of 43% as compared to all other book types, i.e., Art, Language, History, and Social Sciences, etc. Also, the rental books are analyzed based on user gender for each book type. The results show that the following book type “Literature” is a large number of rental transactions for male and female users of 16.40% and 26.02%, respectively. The rental book percentage (%) of the following book type “General Books” is 1.63%, which is low percentage value as compared to all listed book types. Based on the time series analysis, such as seasonal analysis, it shows that a large number of books borrowed by users in the summer season. The rental frequency percentage of the summer season is 28.85%, which is high percentage value as compared to all other seasons, i.e., Autumn, Spring, and Winter of 23.16%, 23.9%, and 24.09%, respectively. We also investigated the relationship between rental book and the rental user’s age and found that the age-group (31-40) is borrowed the large number of books of 33.71% as compared to other defined user age-groups. The rental book analysis based on the user home address summarizes that the home address “Donghwa-ro 1-gil” is a high rental book frequency of 26,457 as compared

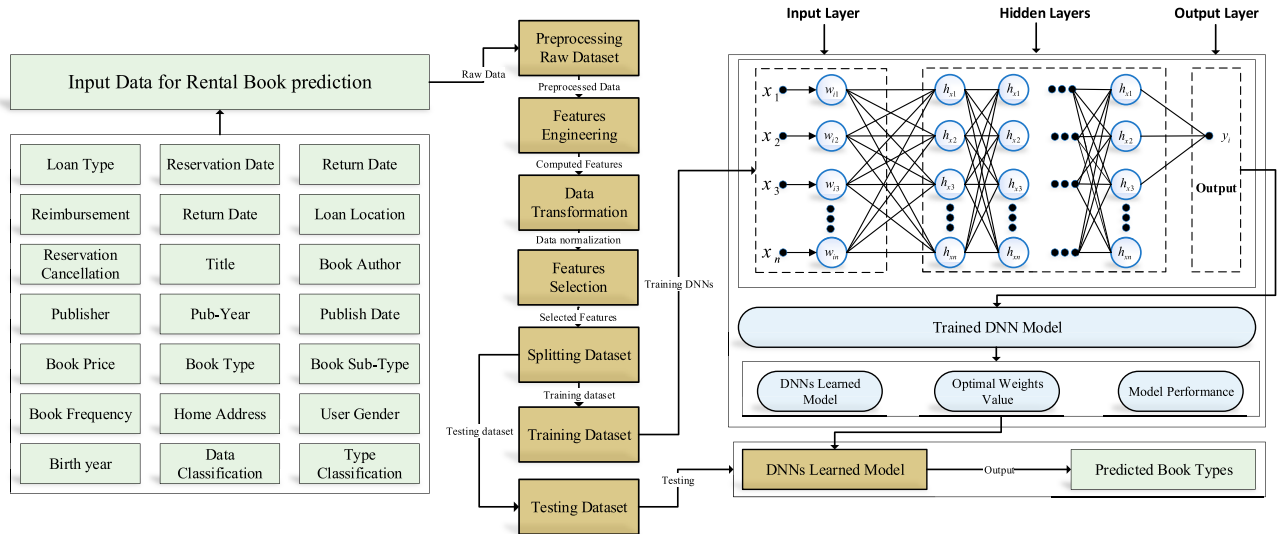


FIGURE 14. Architectural diagram of the proposed predictive analysis model.

to other listed home addresses of the rental users. Based on the data analytics process, it is quite possible for the library administrations to plan resources to meet user’s requirements and provide better services to end-users.

#### IV. PREDICTIVE ANALYSIS OF RENTAL BOOK USING MACHINE LEARNING ALGORITHMS

This section presents the architecture of our proposed model to predict rental book based on knowledge discovered in section III. The proposed architecture model of the library rental book is illustrated in figure 14. In this paper, the following ML regression models, such as DNN, SVR, and RF, have implemented in python to predict and recommend future book details [11], [49], [50]. In this section, we survey the most generally utilized metrics for the regression problem and the basic working flow of the regression models used in this paper.

##### A. DEEP NEURAL NETWORKS

Deep Learning is one of the advanced and robust machine learning strategies which is getting popular. In this network, the data moves in just a single direction, forward, from the input nodes, through the hidden nodes, and to the output nodes using the forward propagation algorithm. Every neuron in one layer has direct associations with the neurons in the resulting layers. It contains an information layer, various hidden layers, and an output layer. The deep neural network (DNN) has several inputs with many hidden layers of neurons. Each neuron has an ‘n’ number of weights for each of its input, each weight is multiplied with each input into the neurons and are added and forms output after having fed through an activation function. The input layer has the activation function given in equation 1, while the hidden and output layers have the SoftMax activation function. Additionally, a multi-layer neuron doesn’t have a direct actuation work in

the entirety of its neurons. A portion of its neurons may have a nonlinear activation work.

$$x_1 = activation(W_1x_0 + b_{k1}). \tag{1}$$

In our proposed DPA-LRBD model, we choose ReLU as the activation function because it is more effective as compared to others, such as sigmoid and tanh [51]. Also, it is used by almost all deep learning models, and it can optimize easier as compared to other activation functions [52]. It is similar to a linear function and allows the model to converge very quickly.

The following equation 2 is used to obtain output for the  $k^{th}$  hidden layer.

$$y_k = ReLU(W_kx_{k-1} + b_k). \tag{2}$$

where  $W_k$  represents the weight matrix between the input and hidden layers,  $x_{k-1}$  represents the input matrix,  $b_k$  is the bias vector, and ReLU indicates the activation function of the proposed DNN model. In this paper, we used the softmax method to obtain and transform the output results in order to obtain prediction value for the actual value  $y$  by using the following equation 3

$$\hat{y} = softmax(W_ox_h + b_o). \tag{3}$$

where  $\hat{y}$  represents the prediction output corresponding to the actual  $y$ ,  $h$  represents the number of hidden layers,  $x_h$  represents the output of the last hidden layer  $h$ , and  $W_o$  and  $b_o$  indicates the output and bias values of the output layer, respectively. Finally, we used the cross entropy method (CEM) to evaluate the proposed model by calculating the difference between predicted and actual values. The CEM works on statistical principles to solve complex problems and to adaptively find efficient estimators for rare event

probabilities [53].

$$CEM = - \sum_{i=1}^n (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)) \quad (4)$$

where  $n$  represents the dimensions of the  $y$  vector, which is equal to the total number of neurons in the output layer.

### B. SUPPORT VECTOR REGRESSION

SVM can also be utilized as a regression method, keeping up all the principle features that characterize the calculation (maximal edge). SVR uses the same standards from the SVM for classification, with just a couple of minor differences. It is one of the fastest developing strategies in machine learning because of its excellent generalization capacity and great assembly execution. SVR is unique concerning other Regression models. It utilizes the Support Vector Machine algorithm for predicting a continuous variable. While other linear regression models attempt to limit the error between the actual and the expected value, Support Vector Regression tries to fit the best line inside a predefined or threshold error value.

### C. RANDOM FOREST

Random Forest classifier (RF) is one of the best outfit learning methods which have been demonstrated to be an exceptionally prominent and powerful technique in machine learning for high-dimensional classification and skewed issues. The basic idea is to combine multiple DTs in order to make robust model to determine the final output rather than relying on a single DT. It is a combination of tree indicators such that each tree relies upon the estimations of an arbitrary vector sampled independently and with a similar conveyance for all trees in the forest. It can be viewed as one classifier which contains a few classification techniques or one strategy yet different parameters of work. For regression problems, the random forest is given by the unweighted average over the collection.

The RF algorithm consists of following three steps:

- 1) The bootstrap random sampling approach is applied to the training set to retrieve  $K$  training set from the original dataset with a size of properties similar to the original training set.
- 2) Develop a model RF for regression problem using a regression tree algorithm for either of the bootstrap training set in order to form a forest with the set of  $K$  decision trees. The RF algorithm uses a random approach to select a subset of features from the original set of features  $m \leq M$  for each  $K$  decision tree.
- 3) Finally, we combine all independent  $K$  decision trees to make a robust model to increase the efficiency of the RF model. The output of each independent DT in RF is determined based on a simple voting method. In our proposed DPA-LRBD model, we combine all subsets of DTs and take an average of all the independent

TABLE 4. Experimentation and implementation setup.

System Component	Description
Operating System	Microsoft Windows 10 (64-bits)
CPU	Intel®Core™i3-2130 CPU at 3.40 GHz
Primary Memory	16 GB RAM
Programming Language	Python
IDE (Platform)	PyCharm
Data Storage	MySQL and MS Excel

DTs as an output, which significantly improves the efficiency of the proposed model.

The following equation 5 is used as a RF regression model.

$$h(x) = \frac{1}{k} \sum_{k=1}^n h(x; \theta_k) \quad (5)$$

## V. PREDICTION RESULTS OF RENTAL BOOK IN LIBRARY

This section presents experimentation and implementation setup, prediction results and performance measures for the proposed predictive analysis model.

### A. EXPERIMENTAL ENVIRONMENT

Table 4 presents the implementation and experimentation setup for the proposed model.

### B. PREDICTION RESULTS

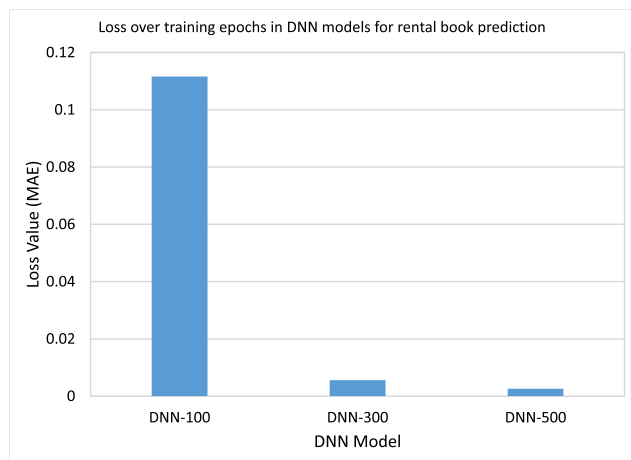
In this subsection, we present the experimental results obtained using DNN, SVR, and RF for the proposed predictive model. In this work, we listed only the top 10 predicted book details based on the input parameters for the users and library administrators. In this paper, the following machine learning (ML) algorithms have implemented in Python, DNN with a linear function, SVR and RF. The following performance measures have utilized to validate the performance of the implemented regression models, such as MAE, MSE, and RMSE, etc.

In this work, we utilized the DNN algorithm for training and testing the proposed predictive model. The DNN algorithm consists of two operational phases, such as the training phase and testing phase. Therefore, we split the pre-processed dataset into two subsets such as training (for training phase) and testing set (for testing phase). There are various statistical sampling techniques are used to split the dataset into subsets. In this paper, we used the sklearn module (Python library) to split the dataset into training and testing with an 70-30 split. Thus, we used 70% of our data for training purposes, and the remaining instances of 30% are used for testing purposes.

$$Data\_Split\_Ratio = \begin{cases} 70, & Training_{set} \\ 30, & Testing_{set} \end{cases} \quad (6)$$

The following figure 20 presents top  $N$  ( $N = 10$ ) predicted books for the library management to plan library resources in an efficient way to provide quality services to library users, i.e., students and faculty members. All the predicted books are presented according to their predicted





**FIGURE 15.** Loss over training epochs in DNN model for rental book prediction.

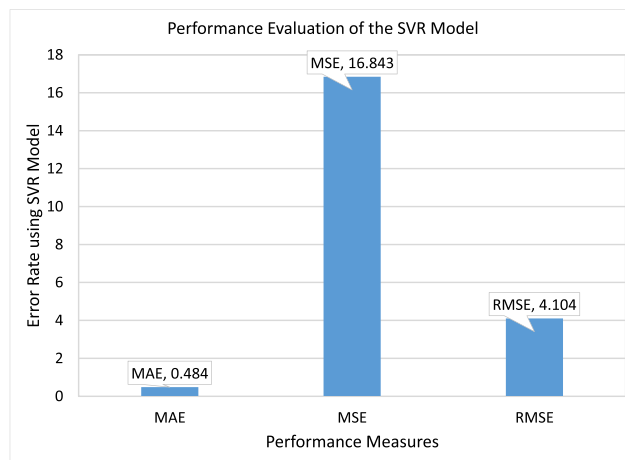
score and popularity based on the extracted hidden features. The predicted book “Rapunzel Princess Adventure” is a high predicted score among all other predicted books.

The following figure 15 describes the loss over training epochs in the DNN model for rental book prediction. It is used to measure the distance between the predicted and actual output during the training phase. It is evident that as the epochs size increase, the loss value (error rate) decreases, and prediction accuracy also improved and increased. In the initial step (DNN-100), the loss value is about 0.11, after the second step (DNN-300), the loss is decreasing gradually. Finally, we achieved a low error rate (loss value) of 0.00257, which indicates that the DNN model produced the most promising prediction results.

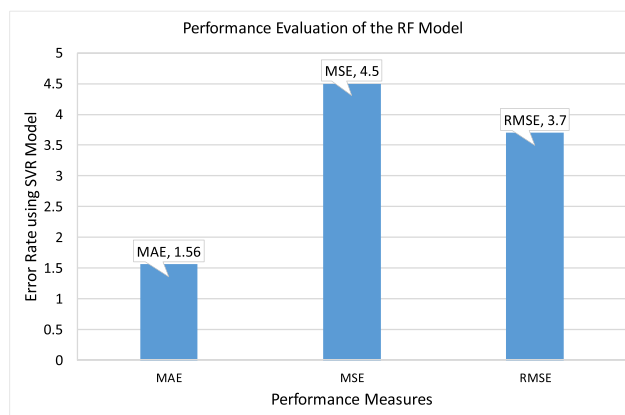
The following figure 16 is used to evaluate the SVR model in terms of MAE, MSE, and RMSE. The prediction performance of the SVR model in terms of MAE, MSE, and RMSE of 0.484, 16.843, and 4.104, respectively. The SVR model also performed well in terms of MAE. Overall, the DNN model performed well in the prediction process as compared to SVR. The range of the performance measures values is between 0 and 100. The ‘0’ indicates that our model performed best in the prediction process, while ‘100’ indicates that our model is performed worst in the prediction process.

The following figure 17 is used to evaluate the RF model in terms of MAE, MSE, and RMSE. The prediction performance of the RF model in terms of MAE, MSE, and RMSE of 1.56, 4.5, and 3.7, respectively. The RF model also performed well in terms of MSE and RMSE as compared to SVR. Overall, the DNN model performed well in the prediction process as compared to SVR and RF models.

The following figure 18 describe and visualize the relationship between actual and predicted book score (rental frequency) using DNN model with 100, 300, and 500 epochs training time. In this work, we configured the DNN model with 100, 300, and 500 epochs timing in the training process. As the number of epoch increases, the accuracy of the proposed model increases, while the error rate (loss)



**FIGURE 16.** Performance evaluation of the SVR model.

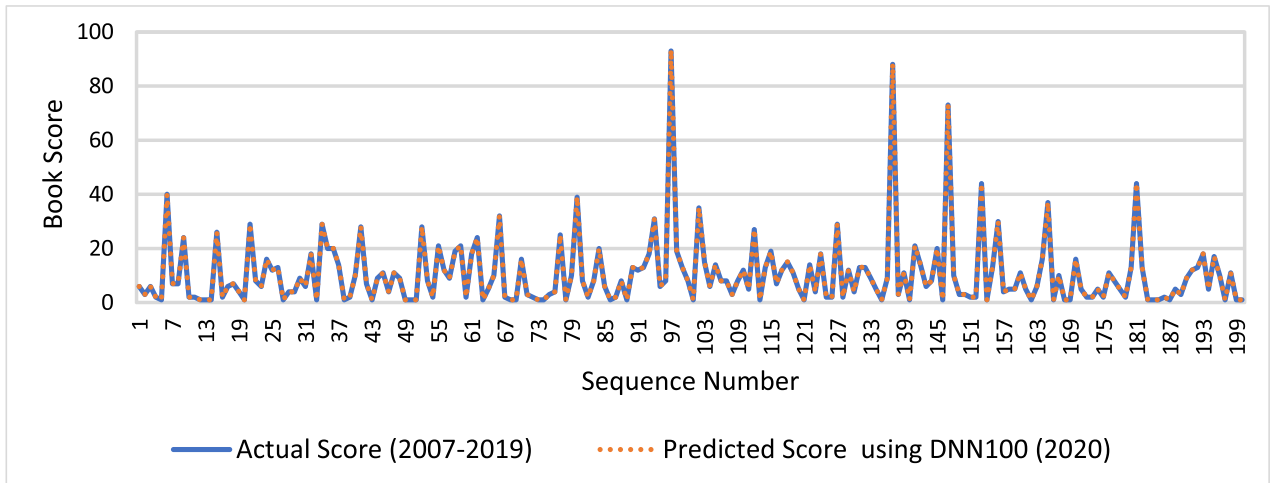


**FIGURE 17.** Performance evaluation of the RF model.

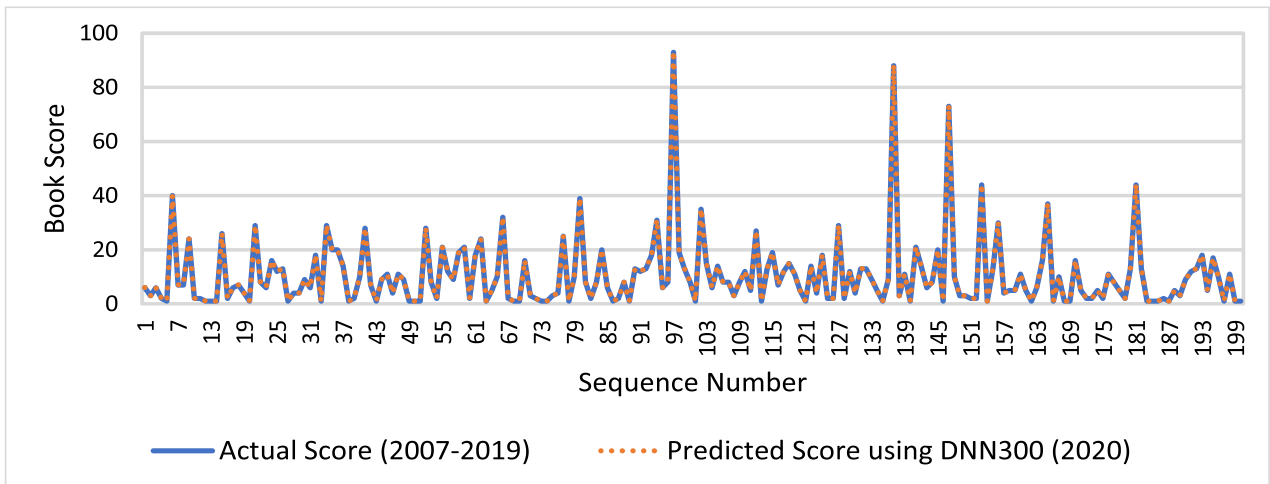
decreases. In each epoch, the loss value reduced, and the accuracy of prediction process increased. In each period, all training instances are passed through the DNN model simultaneously before updating the values of weights. As the number of epochs increases, the loss value decreases. It can be seen that as the epochs size increase, the loss value (error rate) decreases, and prediction accuracy also improved and increased. In the initial step (DNN-100), the loss value is about 0.11, after the second step (DNN-300), the loss is decreasing gradually. Finally, we achieved a low error rate (loss value) of 0.00257, which indicates that the DNN model produced the most promising prediction results with a low error rate.

The following figure 19 (a) presents and visualizes results obtained using the SVR model. The performance of the SVR model in the prediction process is not good as compared to the DNN model. It is evident that the error rate (loss value) of the SVR model is high as compared to the DNN model.

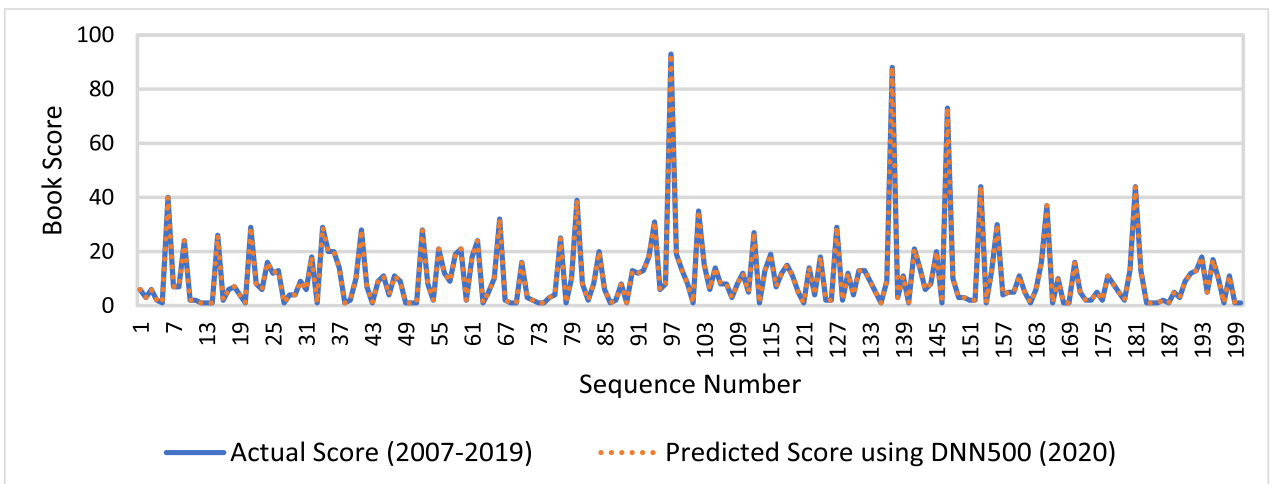
The following figure 19 (b) presents prediction results obtained using the RF model. The RF model also performed well in the prediction process with a low error rate. It can be observed that the error rate (loss value) of the RF model is



(a) Comparison of actual and predicted book score using DNN-100

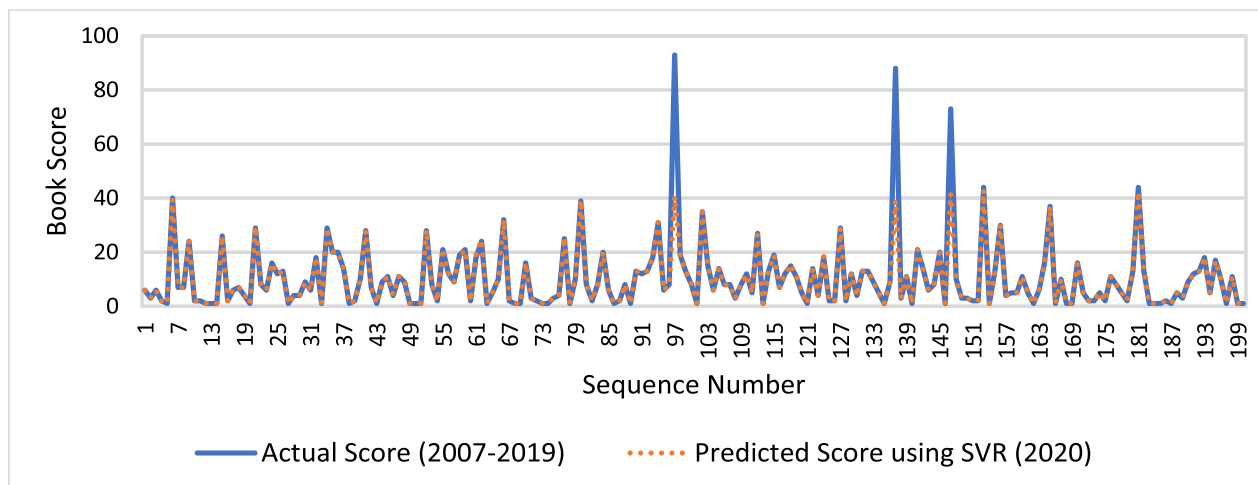


(b) Comparison of actual and predicted book score using DNN-300

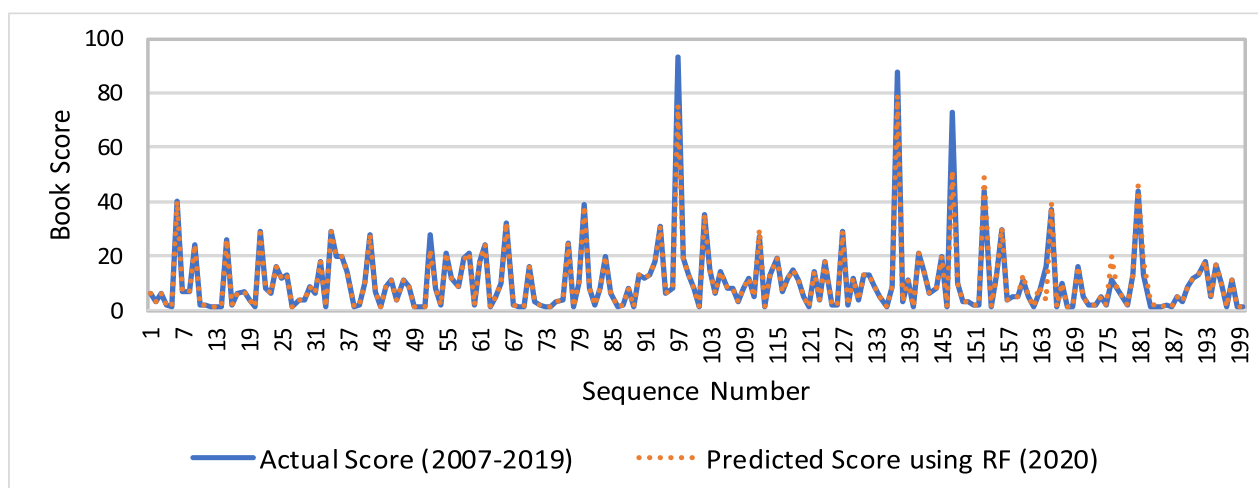


(c) Comparison of actual and predicted book score using DNN-500

**FIGURE 18.** Comparison of actual (2007-2019) and predicted (2020) book score using DNNs 100, 300, 500 epochs.



(a) Comparison of actual and predicted book score using SVR



(b) Comparison of actual and predicted book score using RF

**FIGURE 19.** Comparison of actual (2007-2019) and predicted (2020) book score using SVR and RF.

high as compared to the DNN model. Hence, the DNN model outperformed both SVR and RF to predict future usages of the library book to facilitate library administration in order to drive better future decisions.

The following figure 20 presents the actual and predicted rental frequency using DNN model with 500 epochs in training phase. The following predicted book “Rapunzel Princess Adventure” is achieved the highest rental frequency among all the predicted books. Furthermore, the results shown that the DNN (with epochs 500) performed significantly well as compared to SVR and RF.

The following figure 21 presents, analyze, and compare actual and predicted rental frequency for each book type, such as Literature, Philosophy, and Science of Technology, according to the total number of the rental book. The following results show that the future demands of the library books will be increased.

Based on the following book types expected results, the library administration can plan and manage library resources effectively, such as future book purchases.

### C. PERFORMANCE EVALUATION

Different performance measures are utilized for performance evaluation for distinctive regression problems [26]. In this paper, we used these statistical evaluation measures to determine the effectiveness of our model, for example, Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) and so forth [54].

#### 1) Mean Absolute Error (MAE)

Mean Absolute Error is utilized for evaluation, which measures the deviation between actual values and predicted values. It is obtained by taking the difference between the predicted and target values. It is given by

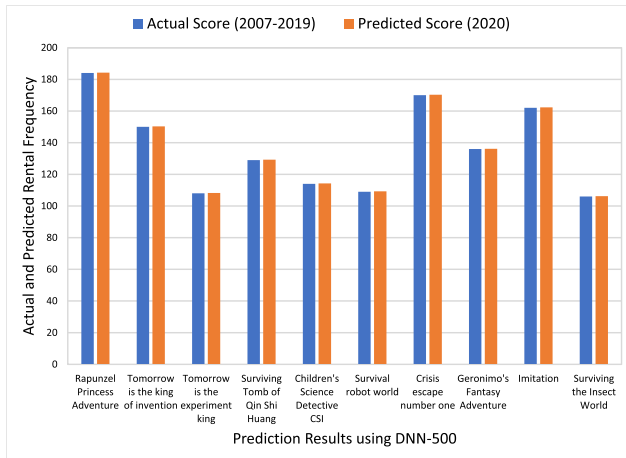


FIGURE 20. Comparison of actual and predicted book score using DNN-500.

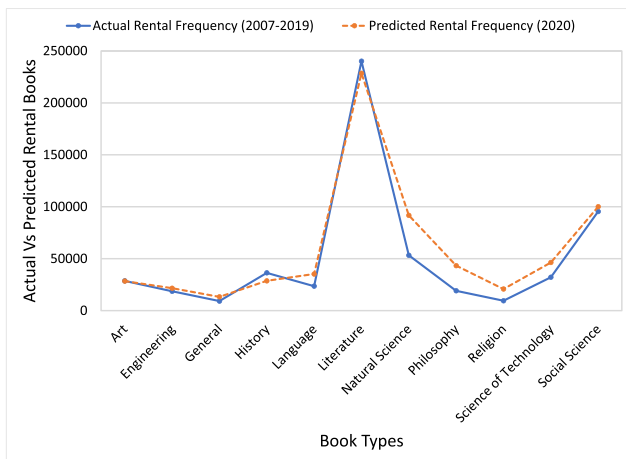


FIGURE 21. Comparison of actual (2007-2019) and predicted (2020) book type.

equation 7:

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \tag{7}$$

2) Mean Square Error (MSE)

It measures the average of the squared difference between the estimated and actual values; the square is taken to remove any negative values; the equation is given by the following equation 8:

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \tag{8}$$

3) Root Mean Square Error (RMSE)

It is utilized to discover the error rate of the regression model and to check the size of the errors to be the same as the size of targets. It is determined by taking the square root of MSE that is given by the following equation 9:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \tag{9}$$

TABLE 5. Statistical performance evaluation of the implemented ML algorithms.

ML Model	MAE	MSE	RMSE
DNN-100	4.00	0.40	6.40
DNN-300	2.70	0.20	4.30
DNN-500	1.40	0.10	2.70
SVR	0.49	16.84	4.11
RF	1.56	4.50	3.70

where  $n$  is the total number of instances in the testing dataset,  $Y_i$  is the actual book frequency value for  $i^{th}$  an instance in the dataset and  $\hat{Y}_i$  is the corresponding predicted book frequency. These statistical measures give us a single number to quantify prediction accuracy. If prediction results are more closed to the actual score of books, then corresponding values for these statistical measures will also be small. An algorithm having the smallest values for these statistical measures will be the best. Table 5 is used to summarize the performance of the implemented algorithms in terms of MAE, MSE, and RMSE.

In this paper, we utilized the following statistical measures to evaluate the performance of the implemented regression models. The DNN model predicts the books accurately based on discovered knowledge from the rental book data, and it has a low error rate in the prediction process in terms of MAE, MSE, and RMSE of 1.40, 0.10, and 2.7, respectively. While the SVR and RF models achieved the highest error rate in the prediction process as compared to DNN model.

VI. CONCLUSION

The proposed DPA-LRBD model employed data and predictive analytics to discover hidden patterns and knowledge in a dataset collected from Jeju National University, the Republic of Korea. The discovered patterns and useful information are used to predict future demands of the library books to plan library resources effectively and provide better services to the end-users. The proposed research findings are based on a detailed analysis of 2,211,413 rental book records, 173,671 unique books, and 57,203 unique users over a period of 13 years (2007-2019). Our proposed DPA-LRBD model consists of two different experimental models; library data analysis and prediction modules. Firstly, we performed and applied data mining techniques to extract relevant and underlying information, which is important and helpful for library management to plan and manage library resources effectively. This work employed different data analysis techniques, such as time series analysis, statistical analyses, book types, and gender-based analysis, age group and home address based analysis. These data analysis techniques are used to analyze and investigate hidden patterns of the rental book data to achieve the research objectives. Based on data analysis results, we ranked each book type according to its rental book frequency and found that book type ‘‘Literature’’ has a large number of rental transactions of 43% as compared to all other book types, i.e., Art, Language, History, and Social Sciences, etc. Based on the time series analysis results, such as seasonal analysis, it is evident that a large number of books borrowed by students in the ‘‘Summer’’ season. The analysis based

on the rental user's gender summarizes that most number of books borrowed by female users of 58.37% as compared to male users of 41.63%. The rental frequency percentage of the "Summer" season is 28.85%, which is the high percentage of the rental book transactions as compared to all other seasons, i.e., Autumn, Spring, and Winter of 23.16%, 23.90%, and 24.09%, respectively. We also investigated the relationship between rental book and the rental user's age and found that most of the books borrowed by the following age-group (31-40). The percentage of the rental book of the age-group(31-40) is 33.71%, which is a high rental percentage as compared to all other defined user age-groups. The rental book analysis based on the user home address summarizes that the home address "Donghwa-ro 1-gil" is a high rental book frequency of 26,457 (no's of rental books) as compared to other listed user's home addresses. Secondly, a novel prediction model is proposed based on DNN, SVR, and RF to predict future book demands, such as book purchases, etc. The proposed DPA-LRBD model used the following performance evaluation measures, such as MAE, MSE, and RMSE, to validated the performance of all implemented regression models. The experimental results shown that the DNN model performed well in the prediction process in terms of MAE, MSE, and RMSE of 1.40, 0.10, and 2.7, respectively. Overall, our proposed DPA-LRBD model results indicate that the DNN model performed well and accurately in the prediction process as compared to SVR and RF algorithms. All other implemented regression models performed well and effectively predict the future book demands with low error rates in terms of MAE, MSE, and RMSE. Based on prediction results, it is found that the future demands of the following listed book types will be increased as compared to the actual rental frequency. Therefore, the library administration can manage library resources in advance to provide better services to library end-users, e.g., students and faculty members. Also, they can easily plan and manage future resources for the library to avoid any shortage of books. Furthermore, our proposed model could be an effective approach for managing future demands of the academic library books and also improving the quality of the academic library services to the end-users. The proposed DPA-LRBD model results empirically demonstrate that data and predictive analytics are feasible solutions to manage academic library resources effectively and to improve the quality of services to fulfill end-user requirements.

Furthermore, the performance of our proposed DPA-LRBD model can be enhanced by analyzing more data to extract and unearth valuable knowledge like sentiment and semantic analysis on rental user reviews data. It can also be enhanced to predict missing values based on known values in high dimensional and sparse matrices, which are usually seen in big data analytics.

### CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

### REFERENCES

- [1] K. Baba, T. Minami, and T. Nakatoh, "Predicting book use in university libraries by synchronous obsolescence," in *Proc. KES*, Jan. 2016, pp. 395–402.
- [2] S. M. Shieber, "Predicting individual book use for off-site storage using decision trees," *Library Quart.*, vol. 66, no. 3, pp. 266–293, Jul. 1996.
- [3] Y. Shi and Y. Zhu, "Book loan quantity prediction using time series data mining," in *Proc. Int. Conf. Electromech. Control Technol. Transp.*, Nov. 2015.
- [4] G. E. Evans, "Book selection and book collection usage in academic libraries," *Library Quart.*, vol. 40, no. 3, pp. 297–308, Jul. 1970.
- [5] S. J. Pritchard, "Purchase and use of monographs originally requested on interlibrary loan in a medical school library," *Library Acquisitions, Pract. Theory*, vol. 4, no. 2, pp. 135–139, Jan. 1980.
- [6] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *J. Parallel Distrib. Comput.*, vol. 74, no. 7, pp. 2561–2573, 2014.
- [7] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, May 2017.
- [8] T. Gao, X. Li, Y. Chai, and Y. Tang, "Deep learning with consumer preferences for recommender system," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Aug. 2016, pp. 1556–1561.
- [9] S. Maharjan, J. Arevalo, M. Montes, F. A. González, and T. Solorio, "A multi-task approach to predict likability of books," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics, Long Papers*, vol. 1, Apr. 2017, pp. 1217–1227.
- [10] A. Ntakaris, M. Magris, J. Kanninen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *J. Forecasting*, vol. 37, no. 8, pp. 852–866, Dec. 2018.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] A.-R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 14–22, Jan. 2012.
- [13] C. Gravelines, "Deep learning via stacked sparse autoencoders for automated voxel-wise brain parcellation based on functional connectivity," *Tech. Rep.*, 2014.
- [14] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [15] Q. Wu, Y. Liu, Q. Li, S. Jin, and F. Li, "The application of deep learning in computer vision," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 6522–6527.
- [16] C.-N. Ko and C.-M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413–422, Jan. 2013.
- [17] N. I. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 24–38, May 2009.
- [18] Imran, S. Ahmad, and D. H. Kim, "Quantum GIS based descriptive and predictive data analysis for effective planning of waste management," *IEEE Access*, vol. 8, pp. 46193–46205, 2020.
- [19] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," *J. Petroleum Sci. Eng.*, vol. 174, pp. 776–789, Mar. 2019.
- [20] C. E. G. Moreta, M. R. C. Acosta, and I. Koo, "Prediction of digital terrestrial television coverage using machine learning regression," *IEEE Trans. Broadcast.*, vol. 65, no. 4, pp. 702–712, Dec. 2019.
- [21] K. De Jager, "Impacts and outcomes: Searching for the most elusive indicators of academic library performance," in *Proc. 4th Northumbria Int. Conf. Perform. Meas. Libraries Inf. Services*. Washington, DC, USA: Association of Research Libraries, 2002, pp. 291–297.
- [22] E. M. Mezick, "Return on investment: Libraries and student retention," *J. Academic Librarianship*, vol. 33, no. 5, pp. 561–566, Sep. 2007.
- [23] D. Lian, Y. Ye, W. Zhu, Q. Liu, X. Xie, and H. Xiong, "Mutual reinforcement of academic performance prediction and library book recommendation," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1023–1028.
- [24] S.-T. Yang and M.-C. Hung, "A model for book inquiry history analysis and book-acquisition recommendation of libraries," *Library Collections, Acquisitions, Tech. Services*, vol. 36, no. 3–4, pp. 127–142, Jan. 2012.



- [25] K. Tsuji, N. Takizawa, S. Sato, U. Ikeuchi, A. Ikeuchi, F. Yoshikane, and H. Itsumura, "Book recommendation based on library loan records and bibliographic information," *Procedia Social Behav. Sci.*, vol. 147, pp. 478–486, Aug. 2014.
- [26] N. Ifada, I. Syachrudin, M. K. Sophan, and S. Wahyuni, "Enhancing the performance of library book recommendation system by employing the probabilistic-keyword model on a collaborative filtering approach," *Procedia Comput. Sci.*, vol. 157, pp. 345–352, 2019.
- [27] C. Chen, L. Zhang, H. Qiao, S. Wang, Y. Liu, and X. Qiu, "Book recommendation based on book-loan logs," in *Proc. Int. Conf. Asian Digit. Libraries*. Berlin, Germany: Springer, 2012, pp. 269–278.
- [28] F. Yan, M. Zhang, J. Tang, T. Sun, Z. Deng, and L. Xiao, "Users' book-loan behaviors analysis and knowledge dependency mining," in *Proc. Int. Conf. Web-Age Inf. Manage.* Berlin, Germany: Springer, 2010, pp. 206–217.
- [29] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in *Proc. 4th Ed. Int. Conf. Innov. Comput. Technol. (INTECH)*, Aug. 2014, pp. 130–134.
- [30] L. Xin, E. Haihong, S. Junde, S. Meina, and T. Junjie, "Collaborative book recommendation based on Readers' borrowing records," in *Proc. Int. Conf. Adv. Cloud Big Data*, Dec. 2013, pp. 159–163.
- [31] J.-M. Gil, J. Lim, and D.-M. Seo, "Design and implementation of mapreduce-based book recommendation system by analysis of large-scale book-rental data," in *Advanced Multimedia and Ubiquitous Engineering*. Singapore: Springer, 2016, pp. 713–719.
- [32] U. Rohini and V. Ambati, "A collaborative filtering based re-ranking strategy for search in digital libraries," in *Proc. Int. Conf. Asian Digital Libraries*. Berlin, Germany: Springer, 2005, pp. 194–203.
- [33] V. Uppal and G. Chindwani, "An empirical study of application of data mining techniques in library system," *Int. J. Comput. Appl.*, vol. 74, no. 11, pp. 42–46, 2013.
- [34] M. Shang, X. Luo, Z. Liu, J. Chen, Y. Yuan, and M. Zhou, "Randomized latent factor model for high-dimensional and sparse matrices from industrial applications," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 1, pp. 131–141, Jan. 2019.
- [35] B. Dumouchel and J. Demaine, "Knowledge discovery in the digital library: Access tools for mining science," *Inf. Services Use*, vol. 26, no. 1, pp. 39–44, Aug. 2006.
- [36] S. Nicholson, "Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 54, no. 12, pp. 1081–1090, Oct. 2003.
- [37] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning," *Arch. Comput. Methods Eng.*, pp. 1–22, Jun. 2019.
- [38] P. Nousi, A. Tsantekidis, N. Passalis, A. Ntakaris, J. Kannianen, A. Tefas, M. Gabbouj, and A. Iosifidis, "Machine learning for forecasting mid-price movements using limit order book data," *IEEE Access*, vol. 7, pp. 64722–64736, 2019.
- [39] M. Kitajima, K. Baba, and T. Minami, "An evaluation of book selection in a university library by loan record analysis," *Int. J. Inf. Edu. Technol.*, vol. 5, no. 10, pp. 728–731, 2015.
- [40] B. Sun, "A study about the prediction of university library lending based on multiple regression analysis," in *Advances in Automation and Robotics*, vol. 1. Berlin, Germany: Springer, 2011, pp. 525–532.
- [41] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Appl. Artif. Intell.*, vol. 17, nos. 5–6, pp. 375–381, 2003.
- [42] P. Russom, "Big data analytics," *TDWI Best Practices Rep.*, vol. 19, no. 4, pp. 1–34, 2011.
- [43] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Nat. Sci. Rev.*, vol. 1, no. 2, pp. 293–314, 2014.
- [44] X. Luo, M. Zhou, S. Li, L. Hu, and M. Shang, "Non-negativity constrained missing data estimation for high-dimensional and sparse matrices from industrial applications," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1844–1855, May 2020.
- [45] X. Luo, H. Wu, H. Yuan, and M. Zhou, "Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1798–1809, May 2020.
- [46] D. Tanasa and B. Trousse, "Advanced data preprocessing for intersites Web usage mining," *IEEE Intell. Syst.*, vol. 19, no. 2, pp. 59–65, Mar. 2004.
- [47] Q. Luo, "Advancing knowledge discovery and data mining," in *Proc. 1st Int. Workshop Knowl. Discovery Data Mining (WKDD)*, Jan. 2008, pp. 3–5.
- [48] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.
- [49] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [50] N. Ahmad, L. Han, K. Iqbal, R. Ahmad, M. A. Abid, and N. Iqbal, "SARM: Salah activities recognition model based on smartphone," *Electronics*, vol. 8, no. 8, p. 881, 2019.
- [51] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2146–2153.
- [52] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [53] S. Panchapagesan, M. Sun, A. Khare, S. Matsoukas, A. Mandal, B. Hoffmeister, and S. Vitaladevuni, "Multi-task learning and weighted cross-entropy for DNN-based keyword spotting," in *Proc. Interspeech*, Sep. 2016, pp. 760–764.
- [54] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan, "Assessing the performance of prediction models: A framework for traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, Jan. 2010.



mainly focused on machine learning, big data, AI-based intelligent systems, analysis of optimization algorithms, and information retrieval.



mainly focused on the Internet of Things application, blockchain application, energy optimization and prediction intelligent service, and mobile computing.



cyber-physical systems, and intelligent systems.



intelligent service, and mobile computing.

• • •