

Artificial Intelligence and Data Mining in Predictive Analytics Based on Service Evaluation Based on Academic Library Book Data

¹HANAN MOURAD, ¹BHARTI GAWALI, ²VAISHALI KHAPARDE

¹Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, INDIA

²Department of Library and IS, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, INDIA

Abstract: - Due to the tremendous progress, we are witnessing today in the age of digital information, data mining technology has been widely used in libraries for its beneficial effect in the context of big data and its role in libraries, and artificial intelligence (AI) is considered an extension of human intelligence and has captured various sectors. Technological advances can stimulate many human capabilities such as Arithmetic, reading, speaking, understanding, remembering, judging and interactive learning How to efficiently extract big data, extract features and provide users with highquality personalized service are important issues to be resolved in the current application of university library big data. We discuss the algorithm used for classification and its application to data mining. Classification is basically a technique for classifying data into separate categories based on constraints. The C4.5 genetic algorithm, the Naive Bayes algorithm, and others are examples of classification algorithms.

Key-Words: - AI , Genetic algorithm , Naive Bayes algorithm, Regression.

Received: July 7, 2021. Revised: June 25, 2022. Accepted: July 21, 2022. Published: September 27, 2022.

1 Introduction

There is a need for university libraries to demonstrate quantitatively their value to their institutions (Luther, 2008). Consequently, studies such as that of Lonsdale (2003) was conducted and revealed that a strong library program that is adequately staffed, resourced and funded can lead to higher student academic performance. Similarly, Wells’ (1995) study established a positive relationship between academic achievement and the use of a number of different library resources or services. Brown and Malenfant (2015) also affirm that university libraries contribute to student learning and success. In addition, Ida (2016) demonstrated that students from schools with libraries with adequate relevant academic materials perform better than those from schools with no libraries. In light of these views, higher education institutions are encouraged to advance and refine strategies that focus on the library’s contributions to universities’ missions when it comes to student success (Brown & Malenfant, 2015). The current study examines relationships between factors such as age, gender, library usage, and satisfaction and student academic performance. The expected research output is to produce a model that university managements can use to determine the value of the university libraries to its primary customers: students Historically, data mining has been mostly used for applications in the use of Artificial Intelligence in libraries can be seen as a collection of technologies enabling machines to sense, comprehend, act and learn and can perform administrative functions and have provided cutting edge technologies for libraries. Librarianship is known as a profes-

sion known for integrating cutting edge technologies not just for information dissemination but in terms of technology as well. Artificial Intelligence has become the new emerging trend for libraries. Artificial intelligence has proven to be a breakthrough for information sectors such as law, which has had significant impacts (Smith, 2016; Chen and Neary, 2017).[2]

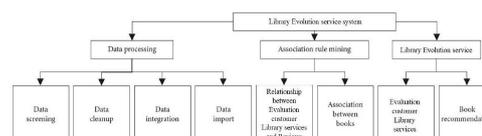


Figure 1: Library Evaluation service system.

2 Comparison Of Different Models

Table 1: Comparison Of Different Models

MODELS	COMPLEXITY & ACCURACY	ADVANTAGED	DIS ADVANTAGES
ANFIS	1. Computationally more expensive than ARIMA 2. Better accuracy than ARIMA at the cost of computation complexity	1. Faster convergence than typical feed-forward neural networks. 2. Smaller size training set. 3. Compact model. 4. Automatic parameter tuning	1. Surface oscillations around points caused due to high partitions 2. Exponential complexity 3. Coefficient signs not always consistent due to monotonic relations
SVM REGRESSION	1. Complexity depends on dimensions of datasets. 2. Time complexity is higher when training set is large. 3. Accuracy is higher in high dimensions datasets	1. Flexibility in the choice of the form of the threshold. 2. good out-of-sample generalization, 4. SVMs deliver a unique solution, since the optimality problem is convex	lack of transparency of results
LINEAR & MULTIPLE REGRESSION ARIMA	Should exist some linear relationship between dependent and independent variable. these models are fitted to time series data either to better understand the data or to predict future points in the series	1. The estimates of the unknown parameters obtained from linear least square regression are optimal 2. It uses data efficiently 3. Good results obtained from relatively small datasets 4. Easy to interpret	1. Outputs can lie outside the range 2. Extrapolation properties will be poor 3. Sensitive to outliers
MULTILAYER PERCEPTRON or NEURAL NETWORK	Space complexity is linear 2. Depends on number of hidden neurons and output neurons. 3. Accuracy depends on number of hidden units	1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience. 2. One of the preferred techniques for gesture recognition. 3. They yield the required decision function directly via training. 4. Neural networks are capable of generalisation, 5. Neural networks are highly fault tolerant Performs better than ARIMA model.	1. Computationally expensive learning process then others 2. scalability

3 Regression

The regression might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable. Typical questions are what is the strength of relationship between dose and effect, sales and marketing spending, or age and income. Second, it can be used to forecast effects or impact of changes. That is, the regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables.

3.1 Linear Regression

Linear regression is probably one of the most important and widely used regression techniques. It's among the simplest regression methods. One of its main advantages is the ease of interpreting results. When implementing linear regression of some dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \dots, x_r)$, where r is the number of predictors, you assume a linear relationship between y and \mathbf{x} : $y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r + \epsilon$. This equation is the

regression equation. $\beta_0, \beta_1, \dots, \beta_r$ are the regression coefficients, and ε is the random error.

Linear regression calculates the estimators of the regression coefficients or simply the predicted weights, denoted with b_0, b_1, \dots, b_r . They define the estimated regression function $f(\mathbf{x}) = b_0 + b_1x_1 + \dots + b_r x_r$. This function should capture the dependencies between the inputs and output sufficiently well.

The estimated or predicted response, $f(\mathbf{x}_i)$, for each observation $i = 1, \dots, n$, should be as close as possible to the corresponding actual response y_i . The differences $y_i - f(\mathbf{x}_i)$ for all observations $i = 1, \dots, n$, are called the residuals. Regression is about determining the best predicted weights, that is the weights corresponding to the smallest residuals. To get the best weights, you usually minimize the sum of squared residuals (SSR) for all observations $i = 1, \dots, n$: $SSR = \sum_i (y_i - f(\mathbf{x}_i))^2$. This approach is called the method of ordinary least squares.

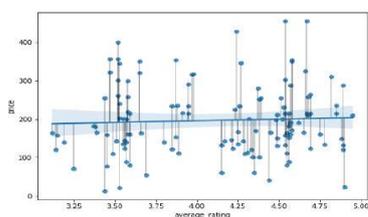


Figure 2: The type of book, average rating with price used are analyzed.

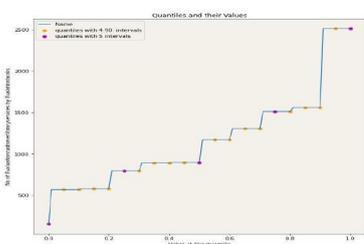


Figure 3: Figure 3 The account value at the quantity No of Evaluation customer library services by Evaluation books

3.2 Multiple Linear Regression

Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable. It is an important regression algorithm that models the linear relationship between a single dependent continuous variable and more than one independent variable. It uses two or more independent variables to predict a dependent variable by fitting a best linear relationship. It has two or more independent variables (X) and one dependent variable (Y), where Y is the value

to be predicted. Thus, it is an approach for predicting a quantitative response using multiple features.

$$\text{Equation: } Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + e$$

Y = Dependent variable / Target variable

β_0 = Intercept of the regression line

$\beta_1, \beta_2, \beta_3, \dots, \beta_n$ = Slope of the regression line

which tells whether the line is increasing or decreasing
 $X_1, X_2, X_3, \dots, X_n$ = Independent variable / Predictor variable
 e = Error

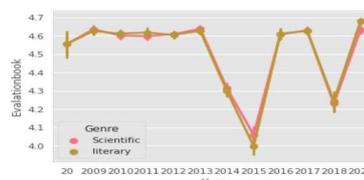


Figure 4: Relation between years of publication book with Evaluation books (good, bad)

4 Analysis of book reviews based on customer reviews to evaluate library services

Classification In this work, we have analyzed and investigated book reviews Based on the following, such as the author, the publishing house, the number of revisions, the number of copies, the year of publication, whether it is a scientific or literary book, as well as the type and number of evaluators. academic degree. Number of years of experience Service evaluators Evaluation of types of books and books. Ranking of reviews Describes the most important written assessments that are rated according to specific criteria and variables. Also, according to the number of reviewers, each book type is calculated to check the ratio Rate books by types of books. the basic How the type of book, type of book, references and year used are analyzed in Figure 5. Figure 6. Book ratings data (maximum to minimum). Figure 7. Analysis of book classifications according to customers' evaluation of services provided in the library The next displays book rating records Based on frequency of reviewers (maximum to minimum). Furthermore, we investigated the data Based on the gender of the rented user, such as male and female. It is clear that users are the most borrowed of books compared to the male user category. We also investigated the data records based on the book Ocean. Each book type is checked and visualized According to the gender of the paid user, i.e., the ratio of males to females Rate. The following figure 8 shows the data in the data set by language like English or British English, Obviously, the book

genre "literature" got a high score Frequency of rental books compared to all other listed books Ocean. While the "general" book type has a low rent book Frequency compared to all other types of books.

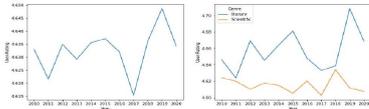


Figure 5: The type of book, reviewers and year used are analyzed.

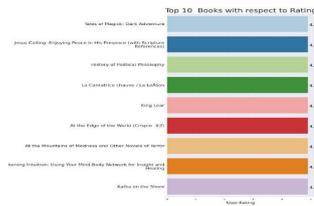


Figure 6: Figure 6. Book ratings data (maximum to minimum).

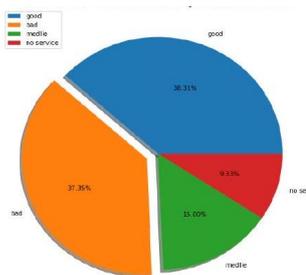


Figure 7: Analysis of book classifications according to customers' evaluation of services provided in the library

5 Data analysis

is the process of acquiring, exploring, cleaning, transforming, and modeling data to discover useful information and present results [3]. Data analysis is the process of examining and modeling a wide range of different data formats, such as structured and unstructured. It is very useful to analyze and examine the huge volume of data to be extracted. Valuable information and hidden patterns can play an important role in making future business decisions. Allow huge volume of data to be analyzed and evaluated through analysis and logical reasoning to generate useful information for better future decisions. The following steps are performed to create useful information and hidden patterns from big data such as data collection

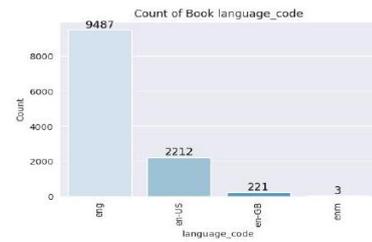


Figure 8: Shows the data in the data set by language like English or British English

or data acquisition. Explore to visualize data and display data distributions of data set variables, clean the obtained data in order to remove irrelevant information and all other outliers, transform from clean data to get standardization in order to increase data reliability and build models with analytical tools to get some results. Big data analytics is a complex process of examining large and varied sizes of a data set that includes structured and unstructured data sets from different in order to discover useful information, such as hidden patterns, market trends, correlations, user interests, and other insights [4]. Big data analytics is a challenge for most people in business organizations [5]. Because the huge volume of data of different formats, such as structured and unstructured data, is stored in different places and different systems, common organizations cannot process and analyze the data through Table 1 . The most important model used for library values. traditional systems. Data is also scattered, and high dimensional matrices are found in big data analytics. Thus, it is necessary to predict missing values on known values of this type of matrices [6, 7]. Two of the specialization challenges facing an existential business are the sheer volume of data. The first challenge is to crack a large amount of data into all the enterprise data, as it is stored in different places and then perform data analysis for hidden patterns and useful information to improve future business decision-making. The second challenge is to provide a platform that can transform a huge volume of unstructured material data into a structured form of knowledge extraction to improve business strategies for traditional databases and systems that are unable to efficiently process large and diverse sets of data. We collected a dataset of library books to examine and analyze large rental book data to extract hidden information and knowledge to improve library services and the operational efficiency of library management. The data in the library lease book statistics are collected in Table 1. Data mining techniques are applied for the pre-processing of the data set to clean up the lease book data. Dealing with missing values and selecting appropriate and useful attributes to in-

Table 2: Library data features and description.

FEATURES	DESCRIPTION
ID	k6-S9-S3...
name	Name of book
author	name of author
publisher	Wiley, Springer..
year	2010...2020
language_code	En – en – us
quantity	3 – 5 – 2
User- Rating	3-4.4-4.5...5
Reviews	1...1000
Price	1 RS...500 Rs
Genre	Scientific - literary
num_pages	50 – 50000
Evaluationlibraryservices	Good-bad-no service
Education	(PhD, MSc....)
ServiceProviderRating	0 -1
Experience	1,2,3,...15
Numberbooks	borrow 1...90000
Reviews	1...11000

crease consistency and effectiveness and reduce data cost. Data storage as well as analysis cost. After pre-processing the library rent book data, we set up the demand data analysis technology to find and discover the hidden information, which is one of the most important pre-processed data for library management to make better future decisions, such as effective management of future orders for books, quality of services for library users, and improve the operation efficiency as well. Data visualization techniques are also used to visualize and display patterns and trends extracted in some graphic formats, such as charts, graphs, etc. These techniques are very useful in analyzing and verifying the huge volume of information discovered from library data to evaluate library resources and services to support and make data-driven decisions to improve the efficiency of operations and provide better services to library users. The following steps are performed to examine and analyze the library book data to discover hidden information about it. Make better future decisions such as effective management of future orders for books, and better services to library users to enhance the operational efficiency of the library. the system.

1)Collection of library rental book data.

3. Data Preprocessing
 - a) Data cleaning
 - b) Handling missing values
 - c) Selection of relevant attributes
4. Data Mining & Visualization
5. Discovered Features

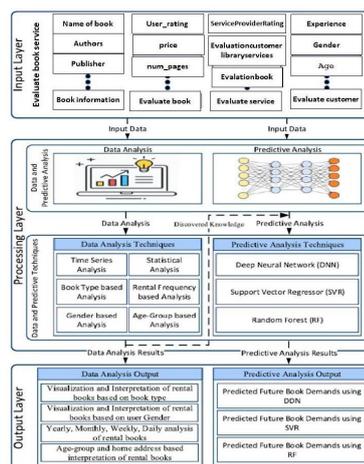


Figure 9: Block diagram of the proposed data and predictive analyses model based on library rental book data.

5.1 COLLECTION OF LIBRARY RENTAL BOOK DATA

In this work, we collected a real dataset from the library of jazan University, the dataset contains 11942 total number of records along with 78 parameters. There is 173,671 total number of unique book records along with 57,203 unique number of users and 173671 unique books. The collected book data is between 2010 to 2020 . The metadata for the data set collected above are given in Table 2 .

5.2 DATA PREPROCESSING

After the book data is acquired, the pre-acquired data set must be processed and outliers removed. In this subsection, data preprocessing technology is used

To convert the raw data set into a format suitable for data analysis and forecasting process [8]. In this work, we implemented the following steps for preprocessing library service library data to increase data set reliability and reduce storage cost.

1. It turns out that there are some duplicate records in the data of the six names. In our data set, we detected 1,814 duplicate records in the book data. Therefore, we remove all duplicate book records from the data set to increase the reliability of the data set.
2. Also, all these records are selected and removed from the dataset that do not contain the book type, gender, reviews, year and price values.
3. Data reduction technology is used to select only relevant and useful data attributes to increase the consistency and efficiency of data and reduce data

Storage cost as well as analysis cost.

5. All other outliers and irrelevant data that the device cannot process are efficient and are removed to increase the reliability and consistency of the library's rented book data. Finally, we only considered relevant and useful features to perform data mining to discover hidden patterns and useful information in order to optimize library resources. In this work, the following data features (parameters) shown in Table 2 are considered for library assessment and prediction book analysis. In the preprocessing data set, the large feature space was reduced to a small number of data features by removing all the features of irrelevant, fixed and missing values. The following table 2 shows the brief features along with their descriptions.

5.3 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 evaluation book data to enable library management to make better future decisions

[9]. 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.

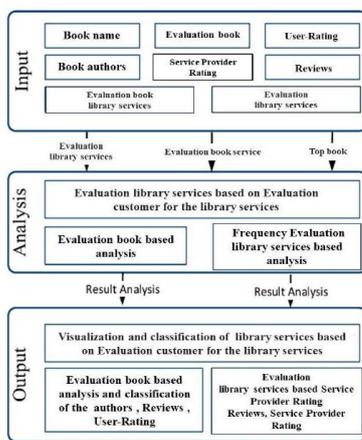


Figure 10: Block diagram for rental book analysis based on rental book frequency, type, and gender.

1. Evaluation service book analysis based on:
 - a) Frequency based analysis of the book name
 - b) Book authors
 - c) Rental user's gender
3. Evaluation library service Analysis

- a) Reviews
 - b) user-rating
 - c) Service Provider Rating
 - d) number-page of book
 - e) language code
4. Evaluation customer service book analysis based on. Reviewers, user-rating- years. 4) Evaluation top book analysis based on authors, language

6 Analysis of the library evaluation according to the years of available books

Time series analysis by years is used to generate valuable information for long-term library staff decisions. In this work, he analyzes the data of the most frequently borrowed book using By Years. The series analysis uses publication years for books to provide a summary of the most important recent books by publication years and ancient books to generate valuable information for long-term library staff decisions. in this work, Analyze the best book evaluation data using an analysis of publication years, number of reviewers, and customer evaluation of services to generate valuable information for library staff in the long run Resolutions [9]. In this work, the data of re-

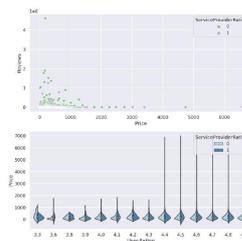


Figure 11: Shows the book data according to as for the number of reviewers and customer ratings based on price

cent books and the number of copies available (from 2010 to 2020) using book ratings analysis (by number of reviewers and customer ratings by price) to generate valuable information regarding book repeatability. The following figure 8 shows the book data according to as for the number of reviewers and customer ratings based on price variance. In a customer feedback library system, string analysis can give an early indication of the total Typical business cycle direction. Figure 9. Analysis of books by years of publication based on the number of reviewers and user ratings (2010-2020).

Figure 10 . Analysis of books by reviewers It can be seen that the total frequency Books are categorized into two groups based on User gender, such as male and female groups. Based on "Best Books" ranks the book reviewed by users in the category of "Female" compared to it highly Users are "male". The

following Figure 8 is used for description and visualization The office services are evaluated according to the number of auditors and the type of clients. Classification of the book is divided into two groups based on the gender of the library clients. In this the work, can be noticed Users have borrowed a large number of books such as Compared to male users we analyzed and investigated rents Written based on the user's gender, to find the relationship Between two groups such as male and female. Based on the results of the daily analysis show that the average frequency The number of books rented by users is high compared to borrowed books for male users.

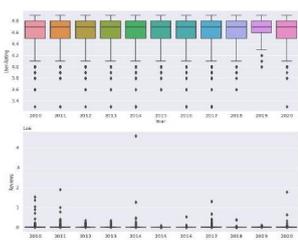


Figure 12: Analysis of books by years of publication based on the number of reviewers and user ratings (2010-2020)

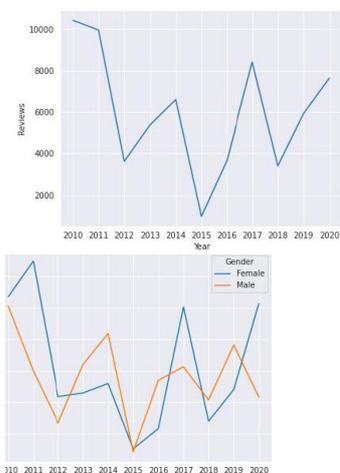


Figure 13: The description and visualization library services are evaluated according to the number of review and the type of customers.

6 Conclusion

Data preparation and processing is used today by data analysts for guidance Discover quality knowledge and help develop effective, high-performance data analytics application systems. In data mining, data preparation and data purification are the first step in the setup responsible for determining quality data from data provided by data preprocessing systems. In fact, data preparation is very important It is easy for us to deal with data easily, even if it is huge Because: (1) real world data is impure; (2) High performance mining systems require quality data; and (3) the quality of the data yields focused patterns. Use data model and forecast analytics to discover patterns and knowledge hidden in a file The set of data, patterns detected, and useful information Used to predict future demands of library books for Effectively plan library resources and provide better services for end users. Library data Analysis and forecasting units. First, we performed and Applied data mining techniques for relevant and basic extraction information, which is important and useful for the library Management to effectively plan and manage library resources. This work used various data analysis techniques, such as Such as analyzing books by years of publication and types of books and Analysis based on gender and age group Analytics. These data analysis techniques are used for analysis Investigate hidden patterns in book rating data to achieve the research objectives. Based on data analysis The results, we ranked each book type according to its own loan book I found rumors that the book of the 'literature' genre has a large volume The number of rental transactions increased by 43% compared to all other transactions types of books. Through the research, we found that the female category is more borrowed than the male category.

References:

- [1] Booyesen Sabeho Tubulingane, Namibia University of Science and Technology, Windhoek, NamibiaUniversity "Library Services and Student Academic Performance" International Journal of Library and Information Services Volume 10 - Issue 2 - July-December 2021.
- [2] Garima Gujral J Shivarama Puttaraj A Choukmath" Perceptions and Prospects of Artificial Intelligence Technologies for Academic Libraries: An Overview of Global Trends" Conference Paper • January 2020
- [3] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," Appl. Artif. Intell., vol. 17, nos. 5_6, pp. 375_381, 2003.
- [4] P. Russom, "Big data analytics," TDWI Best Practices Rep., vol. 19, no. 4, pp. 1_34, 2011.

- [5] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Nat. Sci. Rev.*, vol. 1, no. 2, pp. 293_314, 2014.
- [6] X. Luo, M. Zhou, S. Li, L. Hu, and M. Shang, "Non-negativity constrained missing data estimation for highdimensional and sparse matrices from industrial applications," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1844_1855, May 2020.
- [7] 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.
- [8] D. Tanasa and B. Trousse, "Advanced data preprocessing for intersitesWeb usage mining," *IEEE Intell. Syst.*, vol. 19, no. 2, pp. 59_65, Mar. 2004.
- [9] Q. Luo, "Advancing knowledge discovery and data mining," in *Proc. 1st Int. Workshop Knowl. Discovery Data Mining (WKDD)*, Jan. 2008, pp. 3_5.
- [10] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.
- [11] K. Baba, T. Minami, and T. Nakatoh, "Predicting book use in university libraries by synchronous obsolescence," in *Proc. KES*, Jan. 2016, pp. 395_402.
- [12] 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.
- [13] Y. Shi and Y. Zhu, "Book loan quantity prediction using time series data mining," in *Proc. Int. Conf. Electromech. Control Technol. Transp.*, Nov. 2015.
- [14] G. E. Evans, "Book selection and book collection usage in academic libraries," *Library Quart.*, vol. 40, no. 3, pp. 297_308, Jul. 1970.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] A. Ntakaris, M. Magris, J. Kannianen, M. Gabbouj, and A. Iosiadis, "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.
- [21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436_444, 2015.
- [22] 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.
- [23] C. Gravelines, "Deep learning via stacked sparse autoencoders for automated voxel-wise brain parcellation based on functional connectivity," *Tech. Rep.*, 2014.
- [24] 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.
- [25] 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.