Winter 1-20-2021

A Machine Learning Based Book Availability Prediction Model for Library Management System

Dersin Daimari
d.daimari@cit.ac.in

Mwnthai Narzary
Central Institute of Technology Kokrajhar

Nabajyoti Mazumdar
Central Institute of Technology Kokrajhar, n.jyoti@cit.ac.in

Amitava Nag
Central Institute of Technology Kokrajhar, amitava.nag@cit.ac.in

A Machine Learning Based Book Availability Prediction Model for Library Management System

Dersin Daimari [*]   *Mwnthai Narzary**   *Nabajyoti Mazumdar**   *Amitava Nag**
*Central Institute of Technology Kokrajhar, Deemed to be University.

Abstract

The Library Management System supports numerous users every day and yet many users cannot avail books in real time for their use. For a user, it will be very beneficial to have a system which can predict the possible availability of the issued books. In this paper, Machine learning is used on the data obtained from the library to predict the date for book availability. Random forest, support vector and neural network are used and the result trend are compared using keres and SKlearn. From the study, the result shows that it is possible to know and govern the availability of the books issued. The learned model can then be used to predict the availability of the book. However, the analysis accuracy is reduced when the quality of library data is incomplete. In this study, streamline machine learning algorithms for effective prediction of books in library system is used. The experiment of the modified prediction models over real-life library data collected from Central library of Central Institute of Technology Kokrajhar (CLCITK) was used. To overcome the difficulty of incomplete data, a latent factor model to reconstruct the missing data was used. This study is a proposal for a new model using different machine learning method and to compare performance among them and to identify the more suitable method for the prediction system of book availability in libraries.

Keywords: Library, Library Management System, Library Data, Machine Learning and Artificial Intelligence.

1. Introduction

Students use the Library every day yet most of students are not able to find the book on a fixed date due to uncertainty associated with books availability. Library system is one of the major parts of a college, university where daily thousands of students use the system for issuing or returning books as a major activity. There may not be sufficient volume of books available for each subject to satisfy the overall student's requirement. For a subject there may be n number of books available but there may be more than n students demanding the same book. Thus, many students may not get the book in time and they have no idea about the possible availability of the book in the library. Currently there is no

such system to predict the book availability and thus we have aimed to propose a model that will use the machine learning method to facilitate the library user with this added feature. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Now a days, machine learning techniques have shown substantial progress in predicting the output of an uncertain system using past records. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the provided examples. In the proposed model of book availability prediction method, Machine learning technique is applied using previous years library data.

Considering the limited work in this field this study met various challenges like data pre-processing and how could data be pre-processed from the raw data. There is no format available in which data should be processed. To build the model the main task was to identify the set of data. The data set have been taken from the library that contains lots of missing data and data are not balanced. The study have worked on the data and based on requirement, the relevant information from the data that were needed for analysis was extracted. From the set of data, all student information and book information like ISBN number, accession number, and the book history like issue date, return date etc. were extracted. From history, one can identify the number of days that the book is not present in the library. The information that, in how many days a student would return a book and how many books he or she has issued from library was retrieved. Different machine learning algorithms are applied and experimented with data. The result was different performance with different algorithms. Finally, based on the algorithms' performance only those algorithms that provided suitable results for the analysis were selected.

2. Objective

The main objective of the Library management system is to enhance and upgrade the existing system by increasing its efficiency and effectiveness. The software improves the working methods by replacing the existing manual system with the computer-based system. The main objective of the model is that each and every user of the Library Management System are benefited. The user would acquaint the knowledge on which date the book will be available so that the user might not have to repeatedly check in for the book availability.

3. Literature Review

Artificial Neural Network has been applied in various field for the prediction purpose. Literature review highlighted different existing machine learning approaches that inspired to take up to develop a new library book availability prediction system.

As one can know that machine learning works is based on training and testing data, so one can generate a model such that the machine works based on its past experiments (Zani & Proverbio, 2003). Extensive efforts were made to identify articles employing machine learning and data mining techniques on library management (Kavakiotis, et al., 2017). For a library user, a major issue arises when books become unavailable for issuing because it has been issued by the previous user hence the user can issue if the book is available (Galassetti, et al., 2006). Machine learning methods, such as Random Forest, support vector machines (SVM), Artificial Neural Network (ANN) (Sudharsan, Peeples, & Shomali, 2014) are applied to predict the availability of book in the library (Georga, Protopappas, Ardig`o, Polyzos, & Fotiadis, 2013). Though IRCTC have machine learning algorithm for their model whereas different machine learning algorithm is used for this study (Kumar, 2016). A good analysis model should not only provide the mechanism of the problem understanding but also the frame work of the solution, then it should be studied throughout by collecting data about the system. ANN can be used for development of library management system; however, limited data leads to the down fall of the model. The prediction will determine when the book will be available for issuing. ANN has the highest potential for developing the model of Library Management System (Ji, Sun, Yang, & Wan, 2007). The purpose of the study is to provide statistics and analysis software (Pal, 2011). It will offer librarians or Information manager's sufficient information to make right decisions. The statistical data or chart produced by the program will easily make librarians sense what changes need to be made (Romero & Ventura, 2006). A research would also be conducted to study the reader's behaviour from those statistical data or charts and then improve or modify the library services (Zhang, 2011). Previous work on book inventory management has typically focused on book available and not available such as the framework in (Quoc & Choi, 2009) using high-frequency filtering and thresholding.

4. Data Collection and Pre-processing

As one know that raw data are in complex form, extracting data from a complex form to certain simplified format is known as data pre-processing. Like the collected raw data from CLCITK and then extracted the required data for the model which are discussed below-

4.1. Data Processing

The majority of the data from CLCITK, which contains the student history record for last two years. The data contains the details of each of the book like book title, ISBN number and accession number and issued/return date.

4.2. Feature Engineering

The extracted data are into two types based on the requirements of set of data.

4.2.1. Extraction of data based on student records

For each student it was identified how many books a particular student have taken from the library as given in the Fig: 1 refers number of copies taken by each student. Again it was also found that the average days that a particular student has to return the book. Fig: 1 refers average days of a student issued a book. It was found all the student information regarding the return date for a particular student and the students' performance. It was considered that the average student those returning is in between 40-80.

4.2.2. Extraction of data based on ISBN Number

From Fig: 2, ISBN number of the book identify the name of the book. For each ISBN number there are many numbers of copies which is represented by its accession number. Some ISBN number have more than 150 copies and its issue reflects the importance of that the book is more in demand. Again, some ISBN number have less than 20 copies and its issue means less interest of that book. From the figure a conclusion can be sorted about the interest of users for different types of books.

Count (labels)
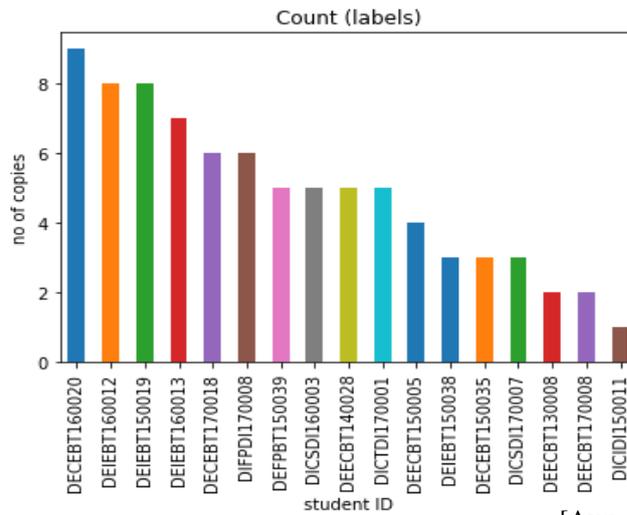
[#copies taken by a student   ]   [Avg. day of book issued]
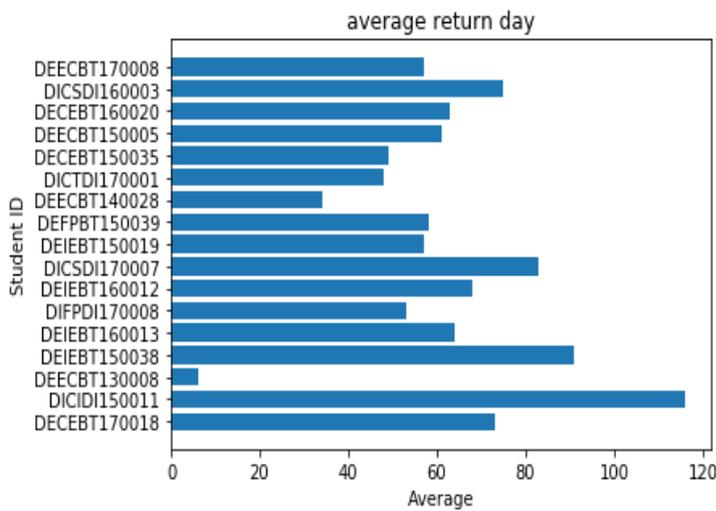
average return day

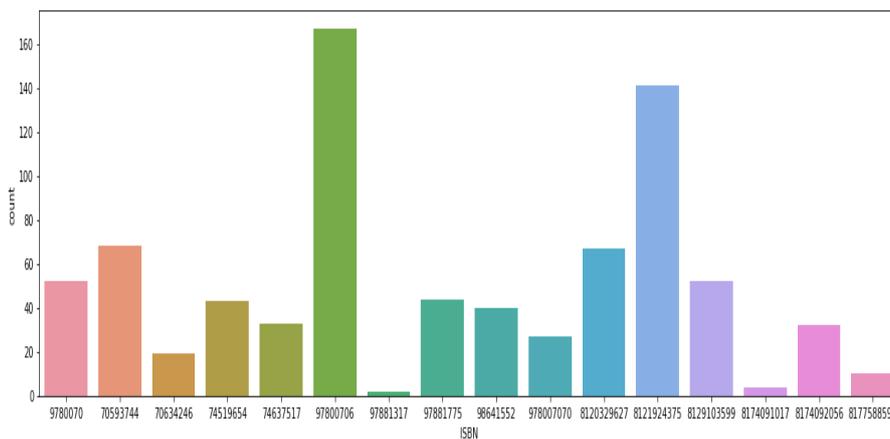Figure 1: Extraction of data based on student records

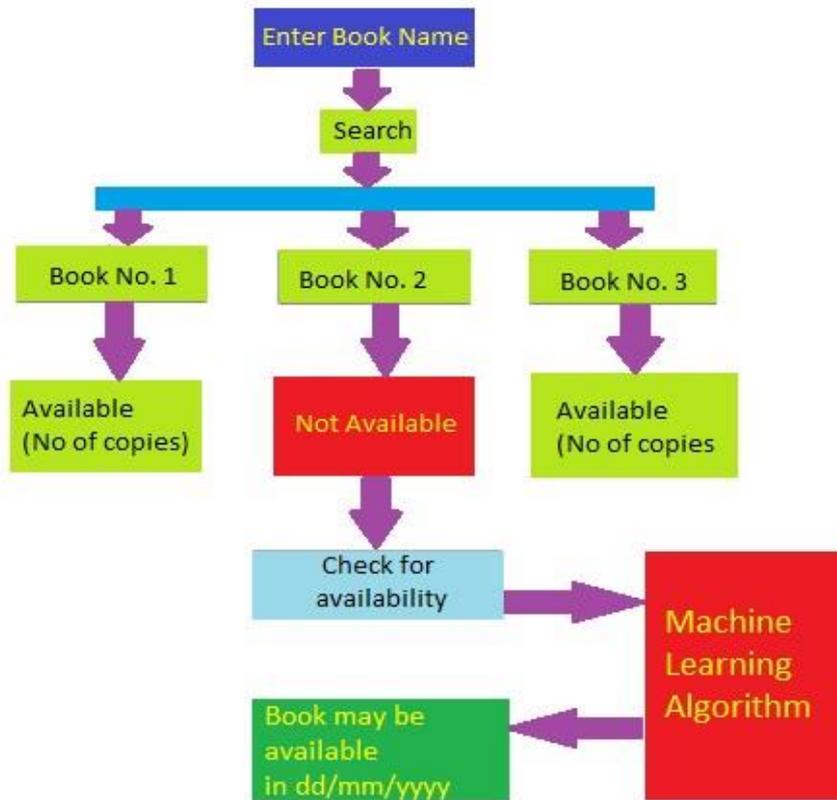
Figure 2: No of copies for each ISBN

Figure 3: User interface model

5. Proposed Approach

In this section the proposed architecture of library management system is discussed.

5.1. System Architecture

In system architecture, it is shown how the proposed system will look and the function of each model is discussed. Fig: 3 shows the interface of the library system where the primary concern is to build the machine learning technique in that interface.

Supposed in library the book name is entered and then it will retrieve all the book that are present in library from library database Book No.1, Book No. 2 and Book No. 3 and so on which indicate the name of the book having accession number separately. If any one of the book is clicked it will show the status of the book whether it is available or not and also the number of copies in the library. Now if a book is not available in the library as shown in the figure: 2, book is not available it will give one button to check for availability. If that button is clicked then it will directly enter through the model

that have been designed and the model will take input as ISBN number and based on previous training experience it will give predicted availability date to the user.

Machine learning technique have been applied on the gathered data. Here, three machine learning algorithm namely, Random forest, Support vector machine (SVM), and Artificial Neural Network (ANN) are used. Initially, all the algorithm was experimented using the data and it was calculated the test and training accuracy after which the performance of all the algorithm was compared. Finally, a particular algorithm was decided which provided the best results and better performance to perform probable prediction date and prediction accuracy as well. During the prediction of probable date, the book ISBN number was taken as input and based on the training set it gave predicted result.

5.2. Working Architecture

The working architect of the propose model is further divided into different modules as discussed below.

5.2.1. Data Cleaning

Data cleaning is one of the most important steps to be considered while considering classification of the set of data. This study considers two main pre-processing filters: replacing missing values and replacing N/A elements with suitable values. In the set of data taken some of the accession number are missing so it is replaced by appropriate values so that it does not conflict with the final result.

5.2.2. Data Transformation

The procedure normalizes the set of data as because ANN based techniques require distance measurements in the training phase. It converts attribute values to small-scale ranges like 0.0 to 1.0 or -1.0 to +1.0.

5.2.3. Data Classification

The set of data is distributed into two disjoint sub-sets, namely the training set and the test set. Basically, two different techniques for distributing the training and test sets of data separately. They are namely 10-fold cross-validation and the 80%-20% distribution among the training and test sets of data. In the present work, three well-known classification techniques namely Random forest, SVM and ANN for training and testing purposes were employed.

Finally, the results are compared and generated by individual classifiers for quantitative analysis.

6. Result and Discussion

Here, an effort is completed in order to be familiar with the prefecture detailed and it is processed by implementing together a variety of Support Vector Machine, Random forest methods and ANN. These models were experimented in the good opinion of the accurate prediction of training and test of data. The estimated outcome of the models are discussed below–

6.1. Result Using Support Vector Machine

SVR with 100 iteration is performed. The experiment using different iteration was performed but the accuracy varies in between 70-80. Fig: 4 refers the accuracy graph of Support Vector Machine. From the support vector machine, svm regress was used or from the fig one can observed that the accuracy of svr is 85% and it has reduced the error rate of 3.1 and the r2-score is 89%. This accuracy is better than random forest. Here error rate reduces small amount. This is the result found from svr rbf. When the data was tested with svr polynomial and linear model, svr rbf gives the best accuracy among those two. The new test data is obtained by the accuracy of 89%.
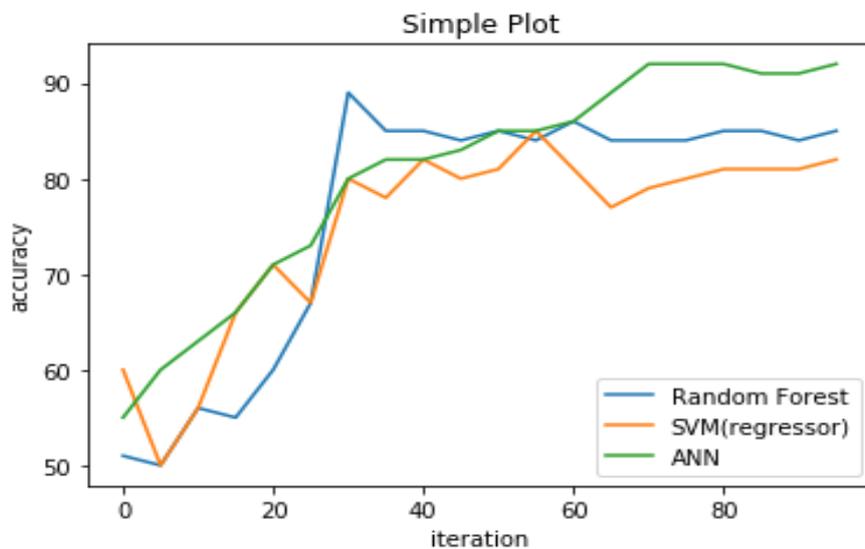

Figure 4: Accuracy Comparison

6.2. Result using Random Forest

Random forest was performed with 10 tree and 100 iteration and in each iteration the performance increases. When varying iteration is performed there was better result with 100 iteration. Using Random Forest accuracy up to 80% was achieved.

6.3. Result Using Neural Network

A neural network was trained with three hidden layers using sklearn and keras. The 4 node of the input layer corresponds to the 4 input features. For the first hidden layer 90 nodes was used, for second hidden layer 100 nodes was used and for third hidden layer 90 nodes were used and so on. After parameter tuning the learning rate of 0.002 and 0.004 produced good result for the proposed model. 0.002 was used as the learning rate and square-loss as the output error function. It was evaluated with tanh, ReLU and sigmoid non-linearity after every layer and found reLU to perform best. Using tanh, the accuracy was 85 per, with ReLU the accuracy was 92 per and using sigmoid the accuracy was 87 per. Then, the data was run in neural network and it gave a substantial improvement. 92% of accuracy which is a very satisfactory performance.

6.4. Comparison of Used Techniques

From the above experiment it was observed that ANN performance which is 91% accuracy represented in the Fig: 4 is better than the other algorithms, SVM accuracy 80% and Random forest accuracy 85%. From that it is decided that the testing of ANN is better than other two algorithms. Because during experiment the set of data was divided into training and test set and operation was performed as the main task is to predict new data which will be given externally.

7. Conclusion and Future Work

The user of a library system often complains about the unavailability of books and most existing library management system does not support any mechanism to predict the possible availability of the desired book. In this paper, we have applied different machine learning methods like Random Forest, SVM, ANN and compared their output for the desired book availability prediction. It is observed that the ANN have shown substantial prediction accuracy over the Random Forest and SVM methods. It is realised that the prediction of availability of the book on a specific day is an important

feature and the accuracy obtained is 91% with the implementation of ANN. In future, there is a scope of adding other relevant features to the existing ANN model to enhance the accuracy of the system.

References

Galassetti, P., Tate, D., Neill, R., Richardson, A., Leu, S., & Davis, S. N. (2006). Effect of differing antecedent hypoglycemia on counterregulatory responses to exercise in type 1 diabetes. *American Journal of PhysiologyEndocrinology and Metabolism*, E1109-E1117. doi:https://doi.org/10.1152/ajpendo.00244.2005

Georga, E., Protopappas, V., Ardig`o, D., Polyzos, D., & Fotiadis, D. (2013). A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. *Diabetes Technology & Therapeutics*, 634-643.

Ji, B., Sun, Y., Yang, S., & Wan, J. (2007). Artificial Neural Networks for Rice Yield Prediction in Mountainous Regions. *The Journal of Agricultural Science*, 249-261. doi: 10.1017/S0021859606006691

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal*, 104-116. doi:https://doi.org/10.1016/j.csbj.2016.12.005

Kumar, D. A. (2016). Hospitality & Public Health: A case study of Indian Railway Catering and Tourism Corporation (IRCTC). *International Journal of Social Science and Economics Invention*, 421-429. Retrieved from https://doi.org/10.23958/ijssei/vol02-i10/02

Pal, J. K. (2011). Usefulness and applications of data mining in extracting information from different perspectives. *Annals of Library and Information Studies*, 7-16. Retrieved from http://nopr.niscair.res.in/handle/123456789/11552

Quoc, N., & Choi, W. (2009). A framework for recognition books on bookshelves. *International Conference on Intelligent Computing* (pp. 386-395). Springer.

Romero, C., & Ventura, S. (2006). *Data Mining in e-learning.* WIT Press.

Sudharsan, B., Peeples, M., & Shomali, M. (2014). Hypoglycemia prediction using machine learning models for patients with type 2 Diabetes. *Journal of Diabetes Science and Technology*, 86-90.

Zani, A., & Proverbio, A. M. (2003). *Cognitive electrophysiology of mind and brain.* Academic Press (An Imprint of Elsevier Science).

Zhang, M. (2011). Application of data mining technology in digital library. *Journal of Computers*, 761-768. Retrieved from https://pdfs.semanticscholar.org/30e1/49d089975dae658c2fec9b62a77f4cf81316.pdf