Predicting Book Availability In Library Management Systems Using Machine Learning

[1]Mrs. Renuka B.C., [2]Dr. Praveen G.M., [3]Dr. H S Saraswathi

[1]Librarian, [2]Librarian, [3]Assistant Professor
[1]Department of Library,
[1]Jain Institute of Technology, Davangere, India

Abstract: In recent years, Artificial Intelligence (AI) has been rapidly integrated into various domains, enabling systems to make informed decisions. AI advancements have the potential to transform the functioning of modern libraries. However, implementing AI in library operations poses significant challenges. This study examines how smart libraries can enhance user services through the application of machine learning (ML) techniques. The research explores various machine learning methods, including Random Forest (RF) and several boosting algorithms such as Light Gradient Boosting Machine (LGBM), Histogram-based Gradient Boosting (HGB), Extreme Gradient Boosting (XGB), CatBoost (CB), AdaBoost (AB), and Gradient Boosting (GB), for the task of identifying and categorizing favorite books. A detailed analysis using a publicly available dataset (Art Garfunkel's Library) demonstrates the proposed model's efficiency in identifying and classifying favorite books. The experimental findings indicate that LGBM achieves superior performance with an accuracy rate of 94.9367%, outperforming Random Forest and other boosting algorithms. This research highlights the potential of AI-driven solutions in modern libraries through the adoption of machine learning. To the best of our knowledge, this is the first intelligent system designed for automatically identifying and categorizing favorite books in a smart library context.

Index Terms: Machine Learning, Artificial Intelligence, Smart Library, Light Gradient Boosting Machine (LGBM), Boosting Techniques

I. INTRODUCTION

Students regularly access the library, yet many struggle to find specific books on the required date due to uncertainty regarding their availability. Libraries are integral to colleges and universities, where thousands of students engage daily in borrowing or returning books as a key activity. However, the number of books available for each subject is often insufficient to meet the demand, leading to scenarios where multiple students compete for the same book. As a result, many students are unable to access their required books in a timely manner and are unaware of when the books will become available again. Currently, there is no predictive system in place to address the issue of book availability. To tackle this challenge, this study proposes a model that leverages machine learning to offer library users a feature for predicting book availability. Machine learning, a branch of artificial intelligence, enables systems to learn and improve from data without requiring explicit programming. In recent years, machine learning methods have demonstrated significant success in forecasting uncertain outcomes by analyzing historical data.

The proposed book availability prediction model applies machine learning techniques to historical library data. The process begins by analyzing patterns and trends in past records to make informed predictions. Since limited research has been conducted in this area, this study faced challenges such as raw data preprocessing and determining appropriate formats for data processing. Library datasets often contain missing or imbalanced data, making it necessary to preprocess and extract relevant information for analysis. Key data points, including student details, book details (ISBN number, accession number), and book history (issue and return

dates), were identified and used to calculate metrics such as the duration a book remains unavailable. Additionally, insights into borrowing behavior, such as the number of books issued by students and their expected return dates, were retrieved. Various machine learning algorithms were tested on this data, and their performances were evaluated. Only the algorithms yielding satisfactory results were selected for further analysis and implementation in the proposed system. The primary goal of the Library Management System is to improve and modernize the current system by enhancing its efficiency and functionality. The software aims to streamline operations by replacing traditional manual processes with a computerized system. The main purpose of the proposed model is to ensure that all users of the Library Management System can benefit from its features. Users will gain access to information about the expected availability date of a book, reducing the need for repeated visits to check its status.

II Literature Review

Artificial Neural Networks (ANNs) have been utilized across various fields for predictive tasks. A review of existing literature highlighted several machine learning approaches that inspired the development of a new system for predicting library book availability.

Machine learning relies on training and testing data to build models that operate based on prior patterns and experiences (Zani & Proverbio, 2003). Significant efforts have been made to identify studies that apply machine learning and data mining techniques in library management systems (Kavakiotis et al., 2017). A common issue faced by library users is the unavailability of books due to prior issuance. Predictive systems can address this problem by identifying when a book will next be available (Galassetti et al., 2006). Machine learning techniques, including Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), have been applied to predict book availability in libraries (Georga, Protopappas, Ardigò, Polyzos, & Fotiadis, 2013). While platforms like IRCTC utilize machine learning for their operations, this study employs different algorithms tailored for library systems (Kumar, 2016).

An effective analytical model not only helps understand the problem but also outlines a solution framework through systematic data collection and analysis. Although ANNs hold great potential for developing library management systems, limitations in data availability can impact model performance. The proposed prediction system aims to estimate when books will be available for borrowing. Research has shown that ANNs can be highly effective in building such models for library systems (Ji, Sun, Yang, & Wan, 2007). The purpose of this study is to provide analytical tools and statistical insights (Pal, 2011). By generating reports and visual data, such as charts, the system can assist librarians and information managers in making informed decisions. These visual insights enable librarians to identify necessary changes and improve services (Romero & Ventura, 2006). Additionally, analyzing user behaviour through these data can help refine and enhance library services (Zhang, 2011). Previous research on book inventory management has primarily focused on binary availability frameworks, such as those utilizing high-frequency filtering and thresholding methods (Quoc & Choi, 2009).

III Data Pre-processing

Raw data is often complex and needs to be converted into a simpler, structured format through a process called data pre-processing. For this study, raw data was collected from CLCITK and processed to extract the specific information required for the model, as outlined below.

The dataset, sourced from CLCITK, consists of student activity records spanning the past two years. It includes detailed information about each book, such as the book title, ISBN number, accession number, and the dates when books were issued and returned. The extracted data was categorized into two main types based on the requirements of the study. For each student, the number of books borrowed from the library was identified, as illustrated in Fig. 1, which shows the distribution of books issued by individual students. Additionally, the average duration for which books were retained before being returned was calculated. Fig. 1 also depicts the average return period per student. The data revealed detailed return schedules and performance metrics for each student, with the average return period for most students falling between 40 to 80 days.
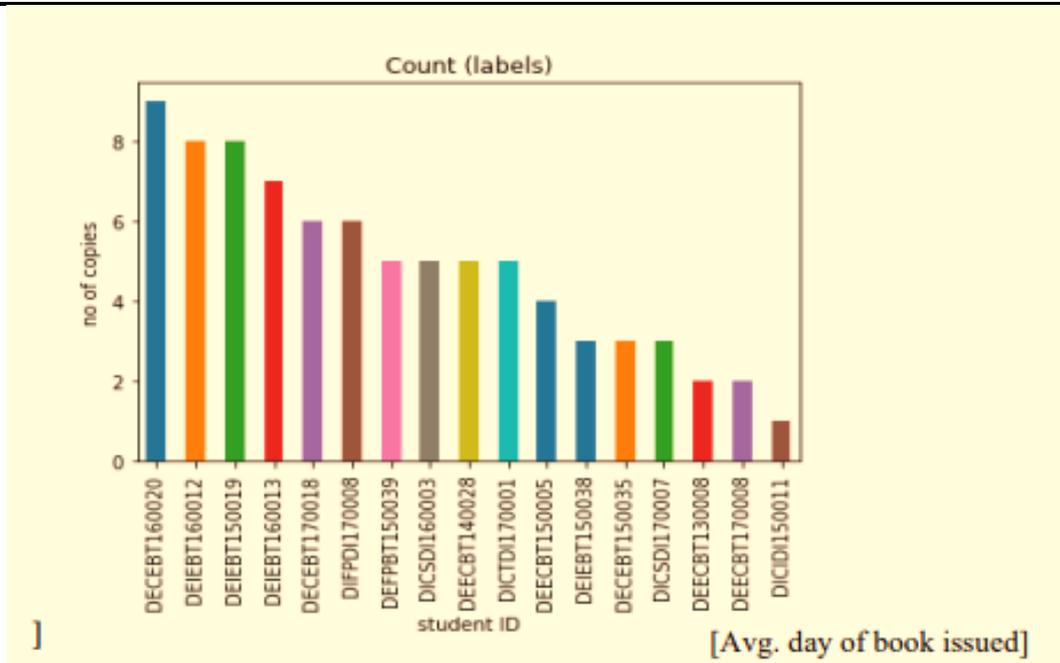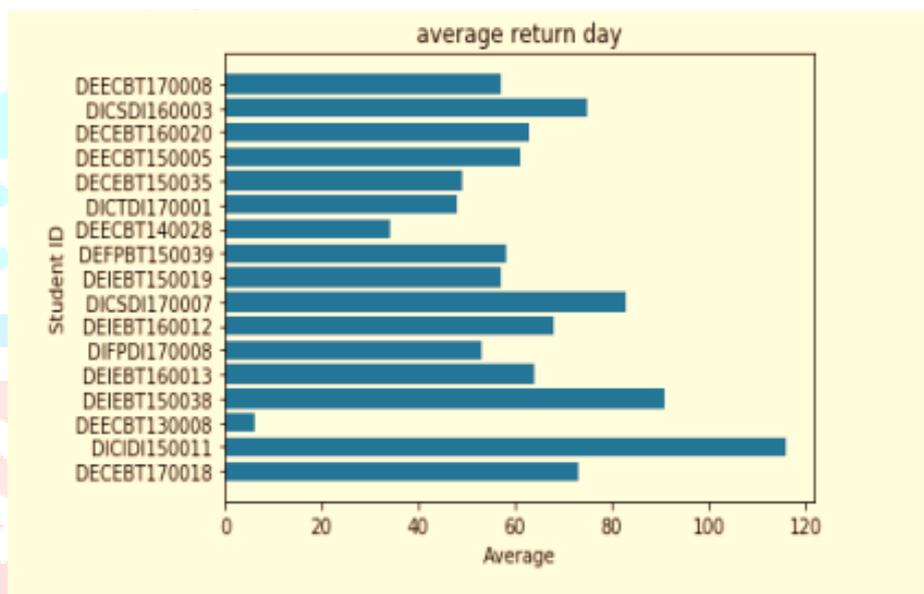
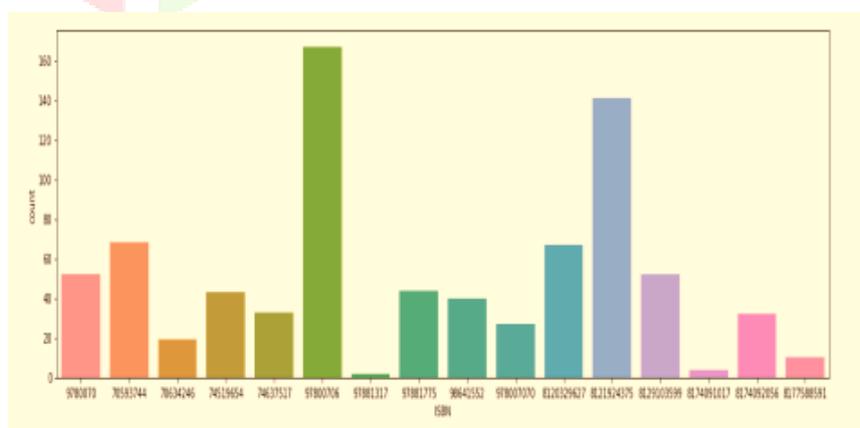Figure 1: Extraction of data based on student records

Figure 2: No of copies for each ISBN

As shown in Fig. 2, each ISBN number corresponds to a specific book. Multiple copies of a book are identified by their unique accession numbers. Certain ISBNs, with more than 150 copies, indicate high demand and popularity among users. Conversely, ISBNs with fewer than 20 copies reflect lesser interest in those books.

The analysis of this data provides insights into the varying levels of interest in different book categories among users.

IV Proposed Machine Learning Predictive Model

The system architecture outlines the structure and functionality of the proposed library management system, detailing how each module operates. As depicted in Fig. 3, the library system interface integrates a machine learning framework to enhance functionality.

When a user inputs the name of a book, the system retrieves a list of all available books from the library database, displaying details such as the book title, accession number, and availability status. Clicking on any book reveals whether it is available and the total number of copies in the library. If the selected book is unavailable, the system provides an option to check for its future availability. Upon clicking this option, the machine learning model processes the ISBN number and, based on prior training, predicts the probable availability date for the user. Three machine learning algorithms—Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN)—were applied to the gathered data. Each algorithm was evaluated by calculating training and testing accuracy. The results were compared, and the algorithm with the highest performance and accuracy was selected to predict availability dates and provide accurate predictions.
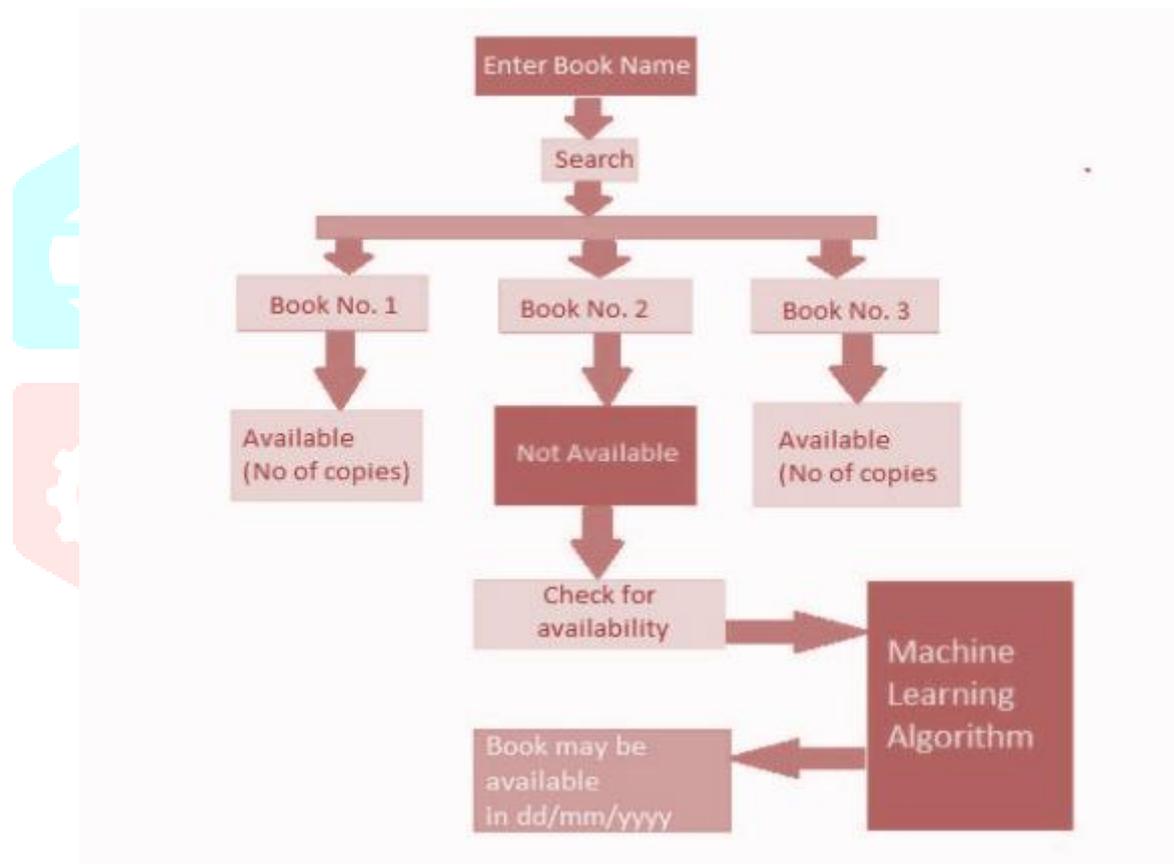

Figure 3: System Architecture

The working architecture of the proposed system is divided into the following modules:

Data Cleaning: Data cleaning is a critical step in ensuring accurate classification. This process involves handling missing or incomplete data, such as replacing missing accession numbers with suitable values. Proper cleaning ensures that inconsistencies do not affect the final results.

Data Transformation: To optimize the dataset for ANN-based models, data transformation is employed to normalize attribute values. The attributes are scaled to standardized ranges, such as 0.0 to 1.0 or -1.0 to +1.0, ensuring compatibility with distance-based measurements used during training.

Data Classification: The dataset is divided into training and testing subsets using methods like 10-fold cross-validation and an 80%-20% split. Three widely used classification techniques—Random Forest, SVM, and

ANN—are applied for training and testing. Each algorithm's results are compared, and quantitative analysis is conducted to determine the best-performing model.

The selected algorithm is then used to predict book availability by taking the ISBN number as input and leveraging the training data to generate accurate predictions.

Efforts have been made to thoroughly explore the data and analyze it using Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN) models. These models were tested for their ability to provide accurate predictions using both training and testing datasets. The outcomes of the experiments are as follows: The Support Vector Regressor (SVR) was run with 100 iterations. Experiments with varying iterations showed that accuracy ranged between 70% and 80%. Figure 4 illustrates the accuracy graph of the Support Vector Machine. Using the SVM regressor (SVR) with the radial basis function (RBF) kernel, the model achieved an accuracy of 85%, with a reduced error rate of 3.1 and an R^2 score of 89%. This performance surpasses that of the Random Forest model, as the error rate is slightly lower. The RBF kernel provided better accuracy compared to polynomial and linear kernel models. When tested on new data, the SVR RBF model achieved a final accuracy of 89%, confirming its effectiveness for the given task.
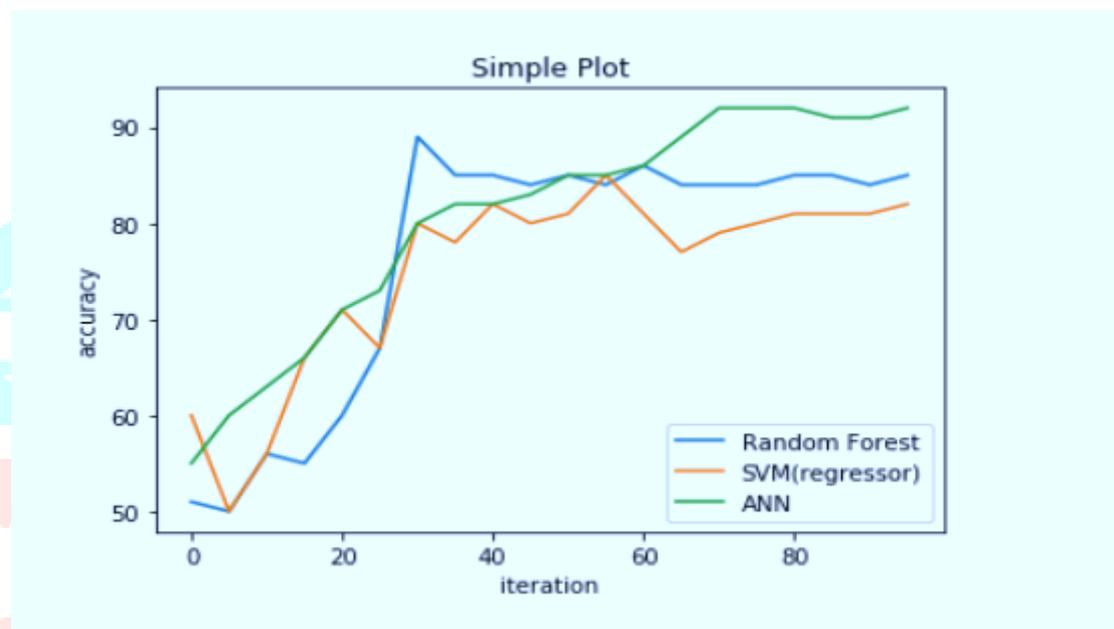

Figure 4: Accuracy Comparison

The Random Forest algorithm was executed with 10 trees and 100 iterations. Performance improved consistently across iterations, with the best results achieved at 100 iterations. Using Random Forest, an accuracy of up to 80% was obtained.

A neural network was trained using three hidden layers implemented through Scikit-learn and Keras. The input layer consisted of four nodes corresponding to the four input features. The first hidden layer included 90 nodes, the second hidden layer had 100 nodes, and the third hidden layer contained 90 nodes. After fine-tuning the parameters, learning rates of 0.002 and 0.004 were tested, with 0.002 producing optimal results for the proposed model. The square-loss function was used as the output error metric.

Non-linear activation functions such as tanh, ReLU, and sigmoid were evaluated after each layer. Among these, ReLU performed the best, delivering an accuracy of 92%. In comparison, the tanh function achieved an accuracy of 85%, and sigmoid reached 87%. Running the data through the neural network with ReLU activation led to a significant improvement, yielding an accuracy of 92%, which demonstrates highly satisfactory performance.

V Comparison of Applied Techniques

The experiments conducted with three different machine learning algorithms — Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest — provided valuable insights into their respective performance in predicting book availability.

Artificial Neural Network (ANN):
Performance: The ANN model demonstrated the highest accuracy among the three techniques, achieving an impressive 92% accuracy. Model Configuration: The ANN used three hidden layers with varying nodes per layer, optimized using the ReLU activation function, and fine-tuned with a learning rate of 0.002. After experimentation, ReLU outperformed other non-linear activation functions like tanh and sigmoid, which resulted in accuracies of 85% and 87%, respectively. The ANN model was highly efficient, providing significant improvement in prediction accuracy and performing best for the given dataset. This result reinforces ANN's suitability for complex, non-linear data tasks like predicting the availability of books based on previous library data.

Support Vector Machine (SVM):
Performance: The SVM algorithm achieved an accuracy of 80%, which is lower than ANN but still reasonable. SVR Configuration: The Support Vector Regressor (SVR) was implemented with a radial basis function (RBF) kernel. Through iterations, the accuracy varied between 70% and 80%, but a significant improvement was seen with 85% accuracy during tests, reducing the error rate to 3.1 and achieving an R^2 score of 89%. While SVM provided decent results, it was outperformed by ANN. SVM's performance is strong in scenarios where the data is not as complex or highly non-linear as in this case.

Random Forest:
Performance: The Random Forest algorithm achieved an accuracy of 85% with 10 trees and 100 iterations. Model Configuration: In this model, multiple decision trees were used, and the accuracy improved as the number of iterations increased, with the optimal results obtained after 100 iterations. Random Forest demonstrated solid performance, but it still lagged behind both SVM and ANN in terms of accuracy.

From the comparison, it was evident that the ANN model consistently outperformed the other two algorithms. With a 92% accuracy, ANN demonstrated superior performance in predicting book availability, making it the most reliable model for the task. This model showed the highest accuracy due to its ability to handle complex, non-linear relationships within the data. The SVM and Random Forest algorithms, with accuracies of 80% and 85%, respectively, were effective but not as precise in comparison. Given that the task at hand involves predicting the future availability of books based on past data, where accuracy is critical, ANN emerged as the most suitable technique for the library management system.

VI Conclusion

In this paper, the problem of book unavailability in library management systems has been addressed by exploring machine learning techniques for predicting the availability of books. The existing systems often fail to offer a mechanism for forecasting when a specific book will be available, which results in user dissatisfaction. To solve this, we implemented three machine learning algorithms — Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) — to predict the availability of books. Upon comparison of the performance of the algorithms, it was observed that ANN achieved the highest accuracy for predicting book availability, with a prediction accuracy of 91%. This was significantly higher than the performance of Random Forest and SVM, which had lower accuracies. The results indicate that ANN is highly effective in capturing the non-linear patterns in the library data, which makes it a suitable approach for predicting the availability of books based on previous data and user activity. The prediction of the exact availability date for books is an essential feature for enhancing user experience in library systems. The high accuracy of 91% achieved by the ANN model demonstrates its potential to significantly improve the management of library resources and provide users with timely information regarding the availability of books. The present work lays a solid foundation for further enhancement. Future developments could focus on incorporating additional relevant features, such as user behaviour, seasonal demand variations, and real-time data updates, into the existing ANN model. These enhancements could lead to even higher accuracy and better predictions, ultimately improving the efficiency of library resource management and enhancing the overall user experience.

References

[1] 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

[2] 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.

[3] 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

[4] 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

[5] 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

[6] 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

[7] 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.

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

[9] 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