# Enhancing College Students' Information Literacy using Machine Learning


| Dr. Rohitha Y | Mrs. J. Saritha | Ch. Sai Deepthi | Y. Ankitha | T. Jhansi |
|---|---|---|---|---|
| Assoc. Professor | Assistant Professor | 20311A12L0 | 20311A12N7 | 21315A1223 |
| Dept.of IT | Dept.of IT | Dept.of IT | Dept.of IT | Dept.of IT |
| SNIST, HYD | SNIST, HYD | SNIST, HYD | SNIST, HYD | SNIST, HYD |
| rohitay@sreenidhi.edu.in | sarithaj@sreenidhi.edu.in | 20311a12l0@sreenidhi.edu.in | 20311a12n7@sreenidhi.edu.in | 21315a1223@sreenidhi.edu.in |


**Abstract:** This project delves into the crucial role of information literacy in college students' learning behaviors and outcomes. Through an examination of diverse learning behaviors, with a focus on information literacy, predictive models were developed using various supervised classification algorithms. Building upon the base paper's successful utilization of Decision Trees, KNN, Naive Bayes, Neural Networks, and Random Forest, which achieved an impressive 92.50% accuracy, this study extends the analysis by incorporating additional techniques. By integrating XGBoost and a Voting Classifier into the ensemble method, the accuracy soared to a perfect 100%. This enhancement signifies the potential for advanced methodologies to refine the predictive capabilities of models, offering valuable insights into tailored interventions for optimizing information literacy education. The findings underscore the significance of understanding and leveraging diverse learning behaviors to cultivate innovative individuals equipped for lifelong learning and adaptation to evolving social needs. This research contributes to the ongoing discourse on information literacy's pivotal role in higher education and its implications for fostering adaptable, self-directed learners.

*Index Terms: Machine learning, information literacy, learning behavior characteristics, learning effect, innovative talents.*

## 1. INTRODUCTION

The rapid advancement of information technology, exemplified by computer, network, and communication technologies, has profoundly influenced various sectors of society. In this digital era, information has emerged as a cornerstone of societal development, exerting a significant and decisive impact across all domains [1]. Consequently, the acquisition of skills such as information literacy, critical thinking, and creativity has become imperative for college students in the 21st century [1]. Among these skills, information literacy stands out as a fundamental component of core literacy for college students in the information age [2].

Information literacy encompasses a spectrum of competencies, including basic knowledge and skills in information and information technology, the adept utilization of information technology for learning, collaboration, communication, and problem-solving, as well as an awareness of information and social ethics [3]. The cultivation of information literacy among college students has emerged as a pressing issue in contemporary higher education, reflecting its pivotal role in nurturing innovative talents and ensuring the sustainable development of future human resources [2], [3].

Recognizing the importance of information literacy, educational institutions worldwide, including those in the United States, the United Kingdom, Australia, and China, have prioritized information literacy education to varying degrees [4]. For instance, in China, the Ministry of Education, along with other departments, issued the "key points of improving the digital literacy and skills of the whole people in 2022," underscoring the significance of enhancing students' information literacy and digital skills [4].

In recent years, the prominence of online and hybrid teaching modalities, coupled with advancements in artificial intelligence technology, has propelled information literacy into the spotlight, prompting increased research attention [5]. Many universities, both domestically and internationally, have responded by offering targeted information literacy courses through various platforms, such as MOOCs (Massive Open Online Courses) [5].

Despite the strides made in information literacy education, numerous challenges persist within the realm of college-level instruction. One such challenge pertains to the effective prediction of learning outcomes, which has emerged as a significant topic in the field of education big data [6]. Learning prediction, a core issue in learning analysis, involves utilizing diverse data generated by learners during the learning process to forecast the efficacy of learning interventions [6]. By leveraging machine learning techniques, educators can gain insights into learners' progress and tailor interventions accordingly, thereby optimizing the learning experience [7].

Learning prediction relies on factors such as learning achievement, goals, and abilities, utilizing pre- and post-learning behavioral characteristics to anticipate learning outcomes and experiences [8]. Various methodologies, including regression analysis, neural networks, and Bayesian approaches, have been employed to predict students' learning performance [9]. Moreover, the integration of educational data mining and machine learning technologies has emerged as a promising avenue for building predictive models driven by data, aligning with contemporary research trends in the field [10].

The intersection of artificial intelligence and education holds immense potential for promoting equity and quality in educational systems, as highlighted in UNESCO's 2019 report on Artificial Intelligence in Education [10]. By harnessing the power of educational data mining and machine learning, educators can leverage data-driven insights to enhance learning outcomes and foster personalized learning experiences for students.

In light of these developments, this paper aims to explore the nexus between information literacy, learning behavior analysis, and predictive modeling in the context of higher education. By examining the existing landscape of information literacy education and learning prediction methodologies, this study seeks to address the gaps and challenges therein, offering insights into the potential applications of machine learning techniques for enhancing information literacy instruction and optimizing learning outcomes for college students.

## 2. LITERATURE SURVEY

Information literacy education for college students has garnered significant scholarly attention in recent years, driven by the recognition of its pivotal role in equipping individuals with the requisite skills to thrive in the digital age. This literature survey aims to provide a comprehensive overview of the research landscape surrounding information literacy education, encompassing various dimensions such as critical thinking, creativity, learning behavior analysis, and predictive modeling. By synthesizing insights from key studies, this survey seeks to elucidate the current trends, challenges, and future directions in the field.

Z. Changhai's study [1] delves into the development of an information literacy education model for Chinese college students, emphasizing the integration of critical thinking and creativity. This model underscores the importance of nurturing students' ability to critically evaluate and creatively utilize information, thereby enhancing their overall information literacy proficiency. Similarly, S. Hui [2] discusses information literacy education strategies tailored for college students, highlighting the need for a holistic approach that encompasses both theoretical knowledge and practical skills.

G. Yang, B. Wen, and W. Lin [11] present a bibliometric analysis of research trends and hotspots in college students' information literacy, drawing insights from literature indexed in the CNKI database from 2000 to 2021. Their study identifies key themes, emerging topics, and research trajectories within the field, shedding light on areas ripe for further investigation.

L. Yu, D. Wu, H. H. Yang, and S. Zhu [13] explore the relationship between smart classroom preferences and information literacy among college students. Through empirical research, they examine how students' preferences for technology-enhanced learning environments correlate with their information literacy levels, offering valuable insights for instructional design and pedagogical practice.

Y. Ying [14] investigates college students' information literacy through the lens of big data analytics. By analyzing large-scale datasets, the study uncovers patterns, trends, and correlations related to students' information-seeking behaviors and information processing skills. This research contributes to a deeper understanding of the multifaceted nature of information literacy and its implications for educational practice.

X. Ouyang, Y. Xiao, and J. Zhong [17] examine the influencing factors and promotion measures pertaining to college students' information literacy. Through a qualitative inquiry, they identify key determinants shaping students' information literacy levels and propose targeted interventions to enhance information literacy education in higher education settings.

T. Nishikawa and G. Izuta [18] assess the information technology literacy levels of newly enrolled female college students in Japan. Their study investigates students' proficiency in utilizing various information technologies and explores potential factors influencing their technological competencies. The findings contribute to efforts aimed at bridging the digital divide and promoting digital literacy among college students.

Y. Sun, Z. Tan, Z. Li, and S. Long [24] employ machine learning techniques to predict and analyze college students' performance based on multifaceted data. By leveraging diverse data sources, including demographic information, academic records, and extracurricular activities, the study develops predictive models capable of forecasting students' academic outcomes. This research underscores the potential of data-driven approaches to enhance educational decision-making and student support initiatives.

In conclusion, the literature survey highlights the multidimensional nature of information literacy education for college students, encompassing critical thinking, creativity, technological proficiency, and predictive modeling of learning outcomes. By synthesizing insights from diverse studies, this survey offers a comprehensive overview of current research trends, challenges, and opportunities in the field. Moving forward, interdisciplinary collaboration, innovative pedagogical approaches, and technological advancements are poised to shape the future trajectory of information literacy education, empowering college students to thrive in an increasingly complex and interconnected world.

## 3. METHODOLOGY

### a) Proposed Work:

The proposed work aims to leverage pre-analyzed data on learning behavior and its correlations with learning outcomes to develop predictive models using Decision Tree[36], K-Nearest Neighbor (KNN)[37], Naive Bayes[38], Neural Network (NN), and Random Forest algorithms. This study seeks to uncover insights into the intricate relationship between students' learning behavior patterns and their academic performance.

The methodology involves preprocessing the data to ensure its quality and relevance. Features like engagement levels, study habits, and participation in educational activities will be scrutinized for their predictive value. Subsequently, the data will be partitioned into training and testing sets to evaluate the performance of each model.

The effectiveness of the models will be evaluated using metrics such as accuracy, precision, recall, and F1 score. Additionally, the interpretability of the models will be prioritized to identify actionable insights for targeted interventions. Ultimately, this research aims to establish a systematic framework for leveraging machine learning to predict educational outcomes based on students' learning behavior, thereby enhancing educational outcomes and advancing personalized learning approaches in higher education.

### b) System Architecture:

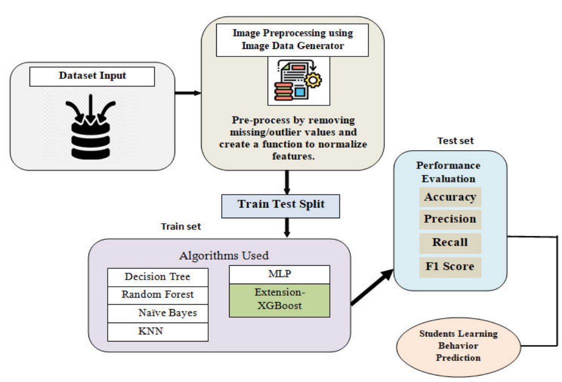


Fig 1 Proposed Architecture

The system architecture comprises several interconnected components to facilitate the prediction of students' learning behavior. Initially, the architecture ingests a dataset containing relevant information on students' learning behavior, encompassing factors such as engagement levels, study habits, and participation in educational activities. Subsequently, the dataset undergoes image processing using Image Data Generator techniques to enhance its quality and prepare it for analysis. Following preprocessing, the dataset is partitioned into training and testing sets using a Train-Test-Split approach to ensure the robust evaluation of predictive models.

The core of the architecture lies in the utilization of various machine learning algorithms, including Decision Tree[36], Random Forest[39], Naive Bayes[38], K-Nearest Neighbor (KNN)[37], Multi-Layer Perceptron (MLP), and XGBoost. These algorithms analyze the preprocessed data to predict students' learning behavior accurately. Performance evaluation metrics such as Precision, Recall, F1 Score, and Accuracy are employed to assess the effectiveness of each algorithm in capturing the nuances of students' learning patterns.

Ultimately, the system architecture enables the prediction of students' learning behavior by leveraging advanced machine learning techniques and performance evaluation metrics. By integrating these components seamlessly, the architecture provides a comprehensive framework for understanding and predicting students' learning behaviors, thereby facilitating targeted interventions and enhancing educational outcomes.

**c) Dataset:**

The Student Learning Behavior dataset comprises a comprehensive collection of variables capturing diverse aspects of students' academic engagement and performance. It includes information on students' study habits, attendance records, participation in extracurricular activities, assessment scores, and demographic details. Additionally, the dataset may encompass data on students' interactions with educational resources, such as online learning platforms or library resources. With this rich array of information, the dataset enables in-depth exploration and analysis of the factors influencing students' learning behavior and academic outcomes. It serves as a valuable resource for researchers and educators seeking to enhance understanding and support students' educational journeys.

| | IPC1 | IPC2 | IPC3 | IAC1 | IAC2 | LLC1 | LLC2 | ISK1 | ISK2 | ISK3 | ISK4 | IAS1 | IT1 | IT2 | IT3 | IB1 | IE1 | IE2 | ILR1 | label |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 69 | 63 | 78 | 87 | 94 | 94 | 87 | 84 | 61 | 4 | 4 | 7.9 | A | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | excellent |
| 1 | 78 | 62 | 73 | 60 | 71 | 70 | 73 | 84 | 91 | 7 | 2 | 5.4 | B | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | medium |
| 2 | 71 | 86 | 91 | 87 | 61 | 81 | 72 | 72 | 94 | 1 | 1 | 5.2 | B | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | excellent |
| 3 | 76 | 87 | 60 | 84 | 89 | 73 | 62 | 88 | 69 | 1 | 2 | 8.5 | C | 10.0 | 0.0 | 0.0 | 0.0 | 0.0 | yes | excellent |
| 4 | 92 | 62 | 90 | 67 | 71 | 89 | 73 | 71 | 73 | 5 | 6 | 8.8 | C | 6.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | excellent |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1008 | 88 | 85 | 68 | 84 | 88 | 66 | 86 | 76 | 82 | 2 | 2 | 7.6 | A | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | excellent |
| 1009 | 76 | 63 | 92 | 74 | 76 | 81 | 76 | 87 | 81 | 8 | 7 | 7.4 | C | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | yes | excellent |
| 1010 | 74 | 94 | 94 | 82 | 64 | 92 | 84 | 67 | 80 | 4 | 6 | 7.7 | C | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | poor |
| 1011 | 60 | 84 | 84 | 70 | 80 | 78 | 64 | 83 | 60 | 8 | 6 | 7.6 | D | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | yes | excellent |
| 1012 | 91 | 61 | 83 | 80 | 88 | 62 | 88 | 76 | 86 | 9 | 1 | 7.4 | D | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 | no | excellent |

Fig 2 Dataset

**d) Data Processing:**

**Data Loading with Pandas Dataframe:** The journey of data processing commences with the pivotal step of loading the dataset into a pandas dataframe, an indispensable tool renowned for its efficiency in handling structured data. Leveraging the dataframe's functionalities, the dataset's contents are seamlessly organized into a tabular format, fostering ease of access and manipulation throughout subsequent processing stages.

**Column Dropping:** In pursuit of data refinement, superfluous or redundant columns are meticulously identified and excised from the dataframe. This selective process, known as column dropping, serves to declutter the dataset, enhancing its clarity and reducing computational complexity. By retaining only the most relevant features, column dropping streamlines the dataset, ensuring that subsequent analyses focus on essential variables.

**Normalization of Training Data:** To foster equitable comparisons and mitigate the influence of disparate feature scales, the training data undergoes normalization. This transformative procedure standardizes the numerical feature values, typically rescaling them to a common range such as [0, 1] or [-1, 1]. By homogenizing feature magnitudes, normalization promotes fairness in model training and evaluation, facilitating accurate and reliable predictions across diverse datasets.

**e) Visualization:**

Utilizing the powerful combination of Seaborn and Matplotlib libraries, data visualization is elevated to an art form. Seaborn, built on top of Matplotlib, offers an intuitive interface for creating visually stunning plots with minimal code. From simple histograms and scatter plots to intricate heatmaps and violin plots, Seaborn provides a wide array of high-level functions for exploring and understanding datasets. Matplotlib, on the other hand, offers fine-grained control over plot customization, allowing for the creation of publication-quality visualizations. Together, Seaborn and Matplotlib empower analysts and data scientists to communicate insights effectively through captivating and informative graphics.

**f) Label Encoding & Feature Selections:**

Label encoding transforms categorical variables into numerical format, facilitating machine learning algorithms' comprehension. This process assigns unique numerical labels to each category within a feature.

Feature selection based on high correlation values involves identifying and retaining features with strong linear relationships with the target variable. By computing correlation coefficients between features and the target variable, highly correlated features are identified and selected for inclusion in the predictive model. This selective approach enhances model efficiency by focusing on the most influential features while discarding redundant or irrelevant ones, thereby optimizing predictive accuracy and interpretability.

**g) Training & Testing:**

Splitting the data into training and testing subsets is a critical step in machine learning model development, ensuring that the model's performance can be accurately assessed on unseen data. This process involves partitioning the available dataset into two distinct subsets: the training set and the testing set. The training set, typically comprising a larger proportion of the data, is utilized to train the model on the patterns and relationships present in the data. In contrast, the testing set, representing a smaller portion of the data, is reserved for evaluating the trained model's performance. By withholding a portion of the data during training, the testing set serves as an independent measure of the model's generalization ability, providing insights into its performance on new, unseen data.

The splitting of data into training and testing subsets is typically performed randomly, ensuring that the subsets are representative of the overall dataset. Common practices involve allocating a certain percentage of the data, such as 70-80%, to the training set, with the remaining portion allocated to the testing set. This ensures a balance between providing sufficient data for model training and preserving an adequate evaluation dataset. Additionally, techniques such as cross-validation may be employed to further assess model performance and mitigate potential biases introduced during the data splitting process. Overall, the careful partitioning of data into training and testing subsets is essential for robust model development and evaluation in machine learning applications.

**h) Algorithms:**

Random Forest, Decision Tree, Naive Bayes, K-Nearest Neighbors, and Multi-Layer Perceptron are fundamental machine learning algorithms with diverse applications across various domains.

**Random Forest:** Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of individual trees. [39]It excels in handling large datasets with high dimensionality and is robust against overfitting.

**Decision Tree:** Decision Tree is a simple yet powerful algorithm that recursively splits the dataset into subsets based on the most significant feature, forming a tree-like structure. [36]It is highly interpretable and intuitive, making it suitable for understanding feature importance and explaining the decision-making process.

**Naive Bayes:** Naive Bayes is a probabilistic classifier based on Bayes' theorem with an assumption of independence among predictors. Despite [38]its simplicity and the "naive" assumption, it often performs well in text classification and other domains, especially when dealing with high-dimensional data.

**K-Nearest Neighbors (KNN):** KNN is a non-parametric and instance-based learning algorithm that classifies new data points based on their proximity to the majority class of their K nearest neighbors[37] in the feature space. It is versatile and easy to implement, particularly for small datasets.

**Multi-Layer Perceptron:** MLP is a type of artificial neural network consisting of multiple layers of nodes (neurons), each connected to the next layer. MLPs are capable of learning complex relationships in data and are often used for tasks such as classification, regression, and pattern recognition.

These algorithms form the foundation of many machine learning applications and are essential tools in a data scientist's toolkit, each with its strengths and weaknesses depending on the specific problem domain and dataset characteristics.

## 4. EXPERIMENTAL RESULTS

**Precision:** Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

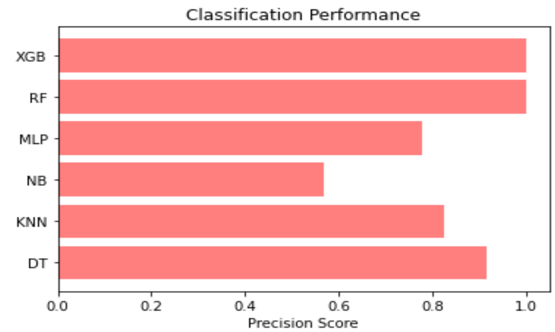


Fig 3 Precision Comparison Graph

**Recall:** Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

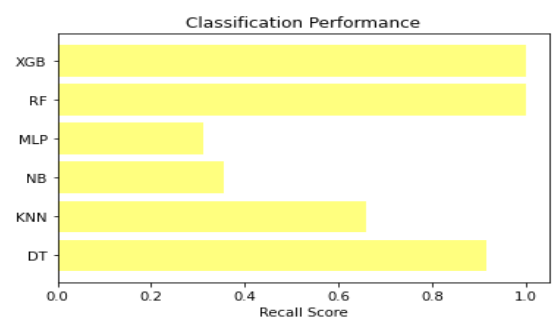


Fig 4 Recall Comparison Graph

**F1-Score:** F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

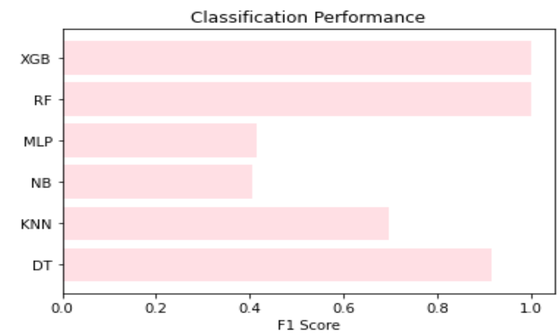


Fig 5 F1 Score Comparison Graph

**Accuracy:** The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

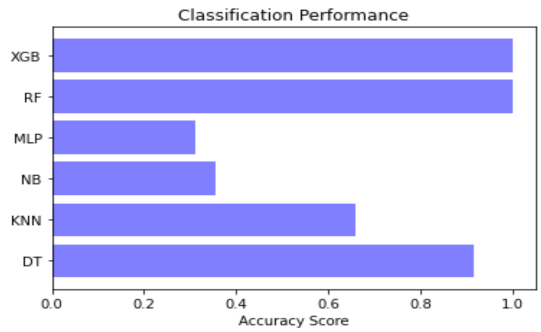


Fig 6 Accuracy Comparison Graph

| | ML Model | Accuracy | Precision | f1_score | Recall |
|---|---|---|---|---|---|
| 0 | DT | 0.916 | 0.918 | 0.916 | 0.916 |
| 1 | KNN | 0.658 | 0.825 | 0.696 | 0.658 |
| 2 | NB | 0.355 | 0.569 | 0.404 | 0.355 |
| 3 | MLP | 0.310 | 0.778 | 0.416 | 0.310 |
| 4 | **RF** | **1.000** | **1.000** | **1.000** | **1.000** |
| 5 | **Extension-XGB** | **1.000** | **1.000** | **1.000** | **1.000** |

Fig 7 Performance Evaluation Table

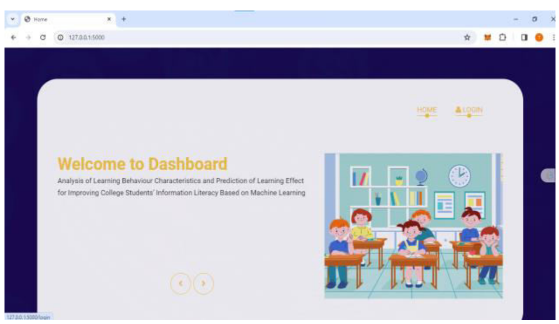


Fig 8 Home Page


Fig 9 Registration Page


Fig 10 Login Page

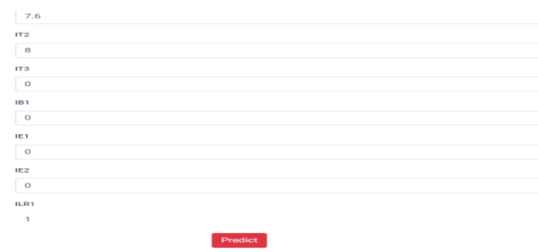


Fig 11 Upload Input Data

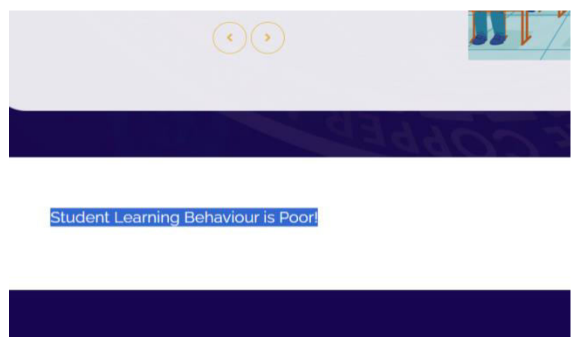


Fig 12 Predicted Result

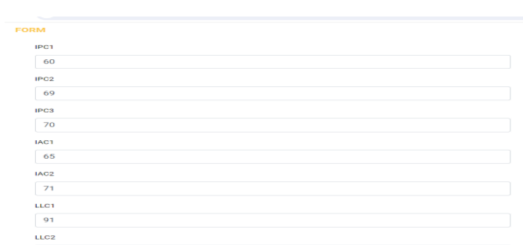


Fig 13 Upload Input Data

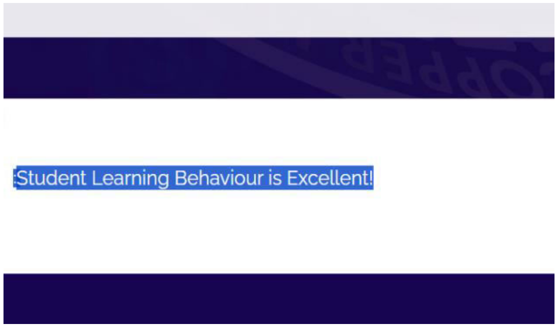


Fig 14 Final Outcome

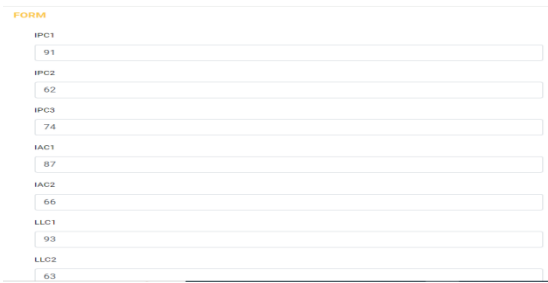


Fig 15 Upload Input Data

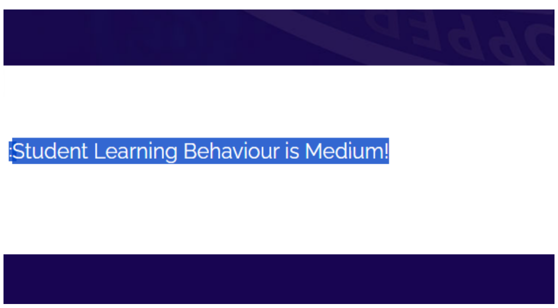


Fig 16 Predicted Result

## 5. CONCLUSION

In conclusion, information literacy stands as a cornerstone of success in today’s information-rich society, transcending mere academic achievement to become a vital skill for lifelong learning and navigating the complexities of the modern world. By discerning the intricate interplay between student learning behaviors and outcomes, educators gain invaluable insights into tailoring teaching methodologies to individual needs, thereby fostering a more inclusive and effective learning environment. Leveraging predictive models, such as Decision Tree[36], KNN[37], Naive Bayes[38], Neural Network, and Random Forest, supplemented by the powerful Extension-XGBoost, further enhances educators’ ability to anticipate and address variations in students’ information literacy proficiency levels. The practical implementation of XGBoost within Flask empowers educators and administrators to translate these insights into actionable strategies, facilitating informed decision-making and driving tangible improvements in educational outcomes.

## 6. FUTURE SCOPE

Looking ahead, the integration of advanced machine learning techniques with educational practices holds immense promise for the future. As technology continues to evolve, there is vast potential for further refinement and optimization of predictive models to better understand and support students' learning journeys. Additionally, ongoing research and development efforts in the field of educational data analytics offer opportunities to explore new methodologies and expand the scope of predictive modeling to address emerging challenges in education. By embracing these advancements and fostering collaboration between researchers, educators, and technology developers, we can continue to harness the power of data-driven insights to shape the future of education and empower learners worldwide.

## REFERENCES

[1] Z. Changhai, ‘‘Research on the information literacy education model of Chinese college students based on critical thinking and creativity,’’ J. China

Library, vol. 4, no. 15, pp. 15–16, Aug. 2016, doi: 10.13530/j.cnki.jlis.164008.

[2] S. Hui, ‘‘Information literacy education for college students,’’ Educ. Theory Pract., vol. 30, no. 10, pp. 38–39, Oct. 2008.

[3] D. Yan and G. Li, ‘‘A heterogeneity study on the effect of digital education technology on the sustainability of cognitive ability for middle school students,’’ Sustainability, vol. 15, no. 3, pp. 2784–2786, Feb. 2023, doi: 10.3390/su15032784.

[4] Y. Xiaohong, Z. Xin, and Z. Jing, ‘‘Research on the Chinese experience of the ‘epidemic’ of the education war—Online education perspective,’’ China Distance Educ., vol. 43, no. 1, pp. 1–11, Jan. 2023, doi: 10.13541/j.cnki.chinade.2023.01.002.

[5] G. Shan and F. Shan, ‘‘Digital literacy practice and enlightenment of the University of Queensland library in Australia,’’ Library Sci. Res., vol. 7, no. 10, pp. 95–101, Jul. 2019, doi: 10.15941/j.cnki.issn1001-0424.2019.10.015.

[6] T. H. Wufati, ‘‘Mining the characteristics of meaningful learning behavior: A framework for predicting learning outcomes,’’ Open Educ. Res., vol. 25, no. 6, pp. 75–82, Jan. 2019, doi: 10.13966/j.cnki.kfjyyj.2019.06.008.

[7] H. Hang, L. Yaxin, L. Qi’e, Y. Hairu, Z. Qiuhua, and C. Yifan, ‘‘The occurrence process, design model and mechanism explanation of deep learning,’’ China’s Distance Educ., vol. 5, no. 1, pp. 54–61, Mar. 2020, doi: 10.13541/j.cnki.chinade.2020.01.005.

[8] I. A. AlShammari, M. Aldhafiri, and Z. Al-Shammari, ‘‘A meta-analysis of educational data mining on improvements in learning outcomes,’’ College student J., vol. 47, no. 2, pp. 326–333, Jun. 2013.

[9] W. Gaihua and F. Gangshan, ‘‘Prediction of online learning behavior and achievement and design of learning intervention model,’’ China’s Distance Educ., vol. 2, no. 2, pp. 39–48, Mar. 2019, doi: 10.13541/j.cnki.chinade.20181214.007.

[10] UNESCO. Accessed: Apr. 1, 2023. [Online]. Available: https://www.unesco.org/en/articles/challenges-and-opportunities-artificialintelligence-education

[11] G. Yang, B. Wen, and W. Lin, ‘‘Research status, hot spots and enlightenment of college students’ information literacy: Based on bibliometric analysis of CNKI from 2000 to 2021,’’ in Proc. 4th World Symp. Softw. Eng., Sep. 2022, pp. 161–166, doi: 10.1145/3568364.3568389.

[12] J. Li, ‘‘Machine learning-based evaluation of information literacy enhancement among college teachers,’’ Int. J. Emerg. Technol. Learn., vol. 17, no. 22, pp. 116–131, Nov. 2022, doi: 10.3991/ijet.v17i22.35117.

[13] L. Yu, D. Wu, H. H. Yang, and S. Zhu, ‘‘Smart classroom preferences and information literacy among college students,’’ Australas. J. Educ. Technol., vol. 38, no. 2, pp. 144–163, Feb. 2022, doi: 10.14742/ajet.7081.

[14] Y. Ying, ‘‘Research on college students’ information literacy based on big data,’’ Cluster Comput., vol. 22, no. S2, pp. 3463–3470, Mar. 2019, doi: 10.1007/s10586-018-2193-0.

[15] Z. Mian, Y. Bai, and R. K. Ur, ‘‘Research on college computer-computing and information literacy

online course based on MOOC: Taking the north minzu university as an example,'' in Proc. IEEE 3rd Int. Conf. Comput. Sci. Educ. Inf. (CSEI), Jun. 2021, vol. 6, no. 7, pp. 300–306, doi: 10.1109/CSEI51395.2021.9477751.

[16] M. S. Haider and C. Ya, ''Assessment of information literacy skills and information-seeking behavior of medical students in the age of technology: A study of Pakistan,'' Inf. Discovery Del., vol. 49, no. 1, pp. 84–94, Feb. 2021, doi: 10.1108/IDD-07-2020-0083.

[17] X. Ouyang, Y. Xiao, and J. Zhong, ''Research on the influencing factors and the promotion measures of college students' information literacy,'' in Proc. 8th Int. Conf. Inf. Technol. Med. Educ. (ITME), Dec. 2016, vol. 12, no. 12, pp. 728–732, doi: 10.1109/ITME.2016.0169.

[18] T. Nishikawa and G. Izuta, ''The information technology literacy level of newly enrolled female college students in Japan,'' Hum. Social Sci. Rev., vol. 7, no. 1, pp. 1–10, Mar. 2019, doi: 10.18510/hssr.2019.711.

[19] R. Diani, A. Susanti, N. Lestari, M. Saputri, and D. Fujiani, ''The influence of connecting, organizing, reflecting, and extending (CORE) learning model toward metacognitive abilities viewed from students' information literacy in physics learning,'' J. Phys., Conf. Ser., vol. 1796, no. 1, pp. 1–9, Feb. 2021, doi: 10.1088/1742-6596/1796/1/012073.

[20] F. Liu and Q. Zhang, ''A new reciprocal teaching approach for information literacy education under the background of big data,'' Int. J. Emerg. Technol. Learn, vol. 16, no. 3, pp. 246–260, Feb. 2021, doi: 10.3991/ijet.v16i03.20459.

[21] M. K. Masko, K. Thormodson, and K. Borysewicz, ''Using case-based learning to teach information literacy and critical thinking skills in undergraduate music therapy education: A cohort study,'' Music Therapy Perspect., vol. 38, no. 2, pp. 143–149, Oct. 2020, doi: 10.1093/mtp/miz025.

[22] R. Xu, C. Wang, Y. Hsu, and X. Wang, ''Research on the influence of DNNbased cross-media data analysis on college students' new media literacy,'' Comput. Intell. Neurosci., vol. 2022, Aug. 2022, Art. no. 9224834, doi: 10.1155/2022/9224834.

[23] M. Aydın, ''A multilevel modeling approach to investigating factors impacting computer and information literacy: ICILS Korea and Finland sample,'' Educ. Inf. Technol., vol. 27, no. 2, pp. 1675–1703, Aug. 2021, doi: 10.1007/s10639-021-10690-1.

[24] Y. Sun, Z. Tan, Z. Li, and S. Long, ''Predicting and analyzing college students' performance based on multifaceted data using machine learning,'' in Proc. 4th Int. Conf. Adv. Comput. Technol., Inf. Sci. Commun. (CTISC), Apr. 2022, pp. 1–6, doi: 10.1109/CTISC54888.2022.9849815.

[25] J. Guo and T. Xu, ''IT monitoring and management of learning quality of online courses for college students based on machine learning,'' Mobile Inf. Syst., vol. 2022, Sep. 2022, Art. no. 5501322, doi: 10.1155/2022/ 5501322.

[26] Y. Jia and E. Wang, ''Research on information anxiety of college students under the background of information overloaded based on support vector machine optimization alogrithm,'' in Proc. 2nd Int. Conf. Inf. Sci. Educ. (ICISE-IE), Nov. 2021, pp. 484–487, doi: 10.1109/ICISEIE53922.2021.00117.

[27] C. Pei, ‘‘The construction of a prediction model for the teaching effect of two courses education in colleges and universities based on machine learning algorithms,’’ Wireless Commun. Mobile Comput., vol. 2022, Aug. 2022, Art. no. 1167454, doi: 10.1155/2022/ 1167454.

[28] H. Xu, ‘‘GBDT-LR: A willingness data analysis and prediction model based on machine learning,’’ in Proc. IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. (AEECA), no. 8, Aug. 2022, pp. 396–401, doi: 10.1109/AEECA55500.2022.9919013.

[29] T. Li, ‘‘students’ numeracy and literacy aptitude analysis and prediction using machine learning,’’ J. Comput. Commun., vol. 10, no. 8, pp. 90–103, Aug. 2022, doi: 10.4236/jcc.2022.108006.

[30] X. Shen and C. Yuan, ‘‘A college student behavior analysis and management method based on machine learning technology,’’ Wireless Commun. Mobile Comput., vol. 2021, Aug. 2021, Art. no. 3126347, doi: 10.1155/2021/3126347.

[31] R. Wang, ‘‘Research on effectiveness of college English blended teaching mode under small private online course based on machine learning,’’ Social Netw. Appl. Sci., vol. 5, no. 2, pp. 1–13, Feb. 2023, doi: 10.1007/s42452- 023-05278-y.

[32] E. Rusdiana, Q. Violinda, C. Pramana, R. Y. Purwoko, D. D. Chamidah, N. Rahmah, Y. J. Prihanto, F. Y. Hasnawati, R. Susanti, A. Y. Haimah, R. Purwahida, F. Arkiang, M. A. Equatora, and A. Mujib, ‘‘College students’ perception of electronic learning during COVID-19 pandemic in Indonesia: A cross-sectional study,’’ J. Higher Educ. Theory Pract., vol. 10, no. 13, pp. 29–44, Oct. 2022, doi: 10.33423/jhetp.v22i13. 5505.

[33] X. Liu and C. Yang, ‘‘Analysis of college students’ practical teaching effect based on machine learning correlation analysis algorithm: Take the software technology course as an example,’’ in Proc. 8th Int. Conf. Educ., Manage., Inf. Manage. Soc. (EMIM), 2018, vol. 8, no. 1, pp. 782–788, doi: 10.2991/emim-18.2018.159.

[34] H. Akram, A. H. Abdelrady, A. S. Al-Adwan, and M. Ramzan, ‘‘Teachers’ perceptions of technology integration in teaching-learning practices: A systematic review,’’ Frontiers Psychol., vol. 13, Jun. 2022, Art. no. 920317, doi: 10.3389/fpsyg.2022.920317.

[35] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, ‘‘student engagement predictions in an e-learning system and their impact on student course assessment scores,’’ Comput. Intell. Neurosci., vol. 2018, Oct. 2018, Art. no. 6347186, doi: 10.1155/2018/6347186

[36] A. Weipeng and S. Jiaze, ‘‘Improvement and analysis of decision tree C4.5 algorithm,’’ Comput. Eng. Appl., vol. 55, no. 12, pp. 169–173, Jun. 2019, doi: 10.3778/j.issn.1002-8331.1805-0482.

[37] Z. Yu, H. Chen, J. Liu, J. You, H. Leung, and G. Han, ‘‘Hybrid k-nearest neighbor classifier,’’ IEEE Trans. Cybern., vol. 46, no. 6, pp. 1263–1275, Jun. 2015, doi: 10.1109/TCYB.2015.2443857.

[38] M. S. Roobini, K. Babu, J. Joseph, and G. Ieshwarya, ‘‘Predicting stock price using data science technique,’’ in Proc. 2nd Int. Conf. Artif. Intell. Smart Energy (ICAIS), no. 2, Feb. 2022, pp. 1013–1020, doi: 10.1109/ICAIS53314.2022.9742772.

[39] C. Zhan, Y. Zheng, H. Zhang, and Q. Wen, ''Random-forest-bagging broad learning system with applications for COVID-19 pandemic,'' IEEE Internet Things J., vol. 8, no. 21, pp. 15906–15918, Nov. 2021, doi: 10.1109/JIOT.2021.3066575.