IJIREEICE

ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131102

PERSONALIZED LEARNING PATH RECOMMENDATION SYSTEM

Prithivi Raj.S[1], Goutham Krishna[2], Abhinav Manoj[3], Bharath G[4], Anand V D[5]

Neelam Sanjeev Kumar[6]

Student, Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[1-5]

Assistant Professor (SG), Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[6]

Abstract: Personalized education has become a cornerstone of modern learning systems, enabling students to receive tailored recommendations based on their unique abilities, interests, and cognitive profiles. Traditional academic evaluation systems are rigid and often fail to capture the diversity of student skill sets. Consequently, learners are guided toward generalized career paths that may not align with their strengths or interests. This research proposes a Personalized Learning Path Recommendation System that leverages the power of machine learning (ML) to identify the most suitable learning domains for individual students. The system analyzes multiple attributes including Programming Score, Math Score, Logic Score, Creativity, Problem Solving Ability, Communication Skills, Interest in Technology, and Time Management. Using these parameters, four ML models — Random Forest, Decision Tree, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) — were trained and evaluated. The Random Forest classifier achieved the highest accuracy of 90%, outperforming other models in stability and prediction consistency. The model provides insights through visual analysis of accuracy comparison, confusion matrices, and feature importance distributions. Additionally, the system accepts live user input to recommend a learning path in domains like AI & ML, Data Science, Web Development, Cyber Security, and UI/UX Design. This approach demonstrates the potential of ML to revolutionize academic counseling by providing data-driven, objective, and personalized recommendations, bridging the gap between aptitude and career direction.

I.INTRODUCTION

Education has evolved from uniform curriculum delivery toward a learner-centered model that emphasizes personalization and adaptive learning. Each student possesses a unique combination of analytical, logical, creative, and interpersonal abilities that traditional grading systems often fail to measure accurately. As a result, many students select learning paths without a clear understanding of where their strengths truly lie. This mismatch can lead to frustration, underperformance, and career dissatisfaction. The integration of machine learning into education has provided a breakthrough in addressing these limitations. ML algorithms can analyze diverse parameters — both quantitative and qualitative — to discover underlying patterns in student performance. Such systems go beyond simple scoring by understanding correlations between skills and suitable learning domains. The proposed Personalized Learning Path Recommendation System (PLPRS) aims to classify students based on their academic and behavioral features, using a machine learning-driven predictive framework. By training on a dataset of simulated student profiles, the system learns to identify the domain where a student is most likely to excel. It takes into account critical parameters like programming ability, problem-solving, logic, creativity, and behavioral traits such as communication and time management. This model not only helps students make better decisions but also enables educational institutions to enhance academic advising processes. With the inclusion of visualization components and real-time prediction capabilities, the PLPRS stands as a scalable and interpretable decision-support system for education.

II.RELATED WORK

The field of adaptive learning and recommendation systems in education has evolved significantly over the last decade. Early approaches relied on rule-based systems, where educators manually defined criteria for career or course selection. These systems lacked adaptability and often could not handle new or unseen data. Machine learning introduced flexibility and automation to educational analytics. Naïve Bayes, Decision Trees, and SVM were among the first models applied to predict student performance or recommend academic tracks. For instance, earlier frameworks used GPA and exam marks to classify students into technical or creative fields, but these systems ignored non-academic traits like communication or creativity. In more advanced systems, ensemble learning methods such as Random Forest and Gradient Boosting

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025
DOI: 10.17148/IJIREEICE.2025.131102

became popular due to their ability to handle high-dimensional data and avoid overfitting. Random Forest, in particular, provides interpretability through feature importance analysis, which helps identify the most influential factors in decision-making. Recent works also emphasize the inclusion of psychological and behavioral features alongside academic data for better personalization. However, despite progress, most existing systems still focus on limited datasets or lack live interactivity. The proposed PLPRS fills this gap by integrating multiple attributes, comparing different algorithms, and providing immediate, explainable recommendations that students and educators can use directly.

III.DATASET DESCRIPTION AND PREPROCESSING

The dataset used for this research consists of 200 simulated records, each representing a student profile. The data was generated with controlled variability to mimic real-world student diversity. The attributes include both numerical features (Programming Score, Math Score, Logic Score, Creativity, Problem-Solving) and categorical features (Interest in Technology, Communication Skill, Time Management). Each attribute was assigned realistic ranges — for instance, Programming Score varied between 40 and 100, while Communication Skill and Interest in Technology had levels such as High, Medium, or Low. The target variable, Preferred Field, was assigned based on conditional logic that maps score combinations to specific domains. For example, a student with high programming and logic scores would be classified into AI & ML, while strong creativity and communication mapped to UI/UX Design. Before training, all categorical features were encoded using LabelEncoder to convert text-based values into numerical form. The dataset was then split into 75% training and 25% testing subsets. Exploratory data analysis was performed to visualize distributions and identify correlations between variables. The balanced data ensured fair model comparison and prevented bias toward specific fields.

IV. SYSTEM ARCHITECTURE

The system architecture is modular, comprising four main layers — data processing, model training, model evaluation, and user interaction. 1. Data Processing Layer: Handles encoding, normalization, and feature preparation for ML algorithms. This step ensures data consistency and readiness for training. 2. Model Training Layer: Implements four different machine learning models — Random Forest, Decision Tree, KNN, and SVM — to identify the best-performing classifier. 3. Evaluation Layer: Measures model accuracy, precision, recall, and F1-score using test data. Visualization tools such as bar graphs and confusion matrices are used to analyze performance. 4. User Interaction Layer: Provides an interface where users input their scores and skills, and the trained model predicts the ideal learning domain. The modular design ensures scalability, meaning new domains or additional features can be added easily in future iterations. The architecture allows smooth integration with online platforms for real-time student guidance.

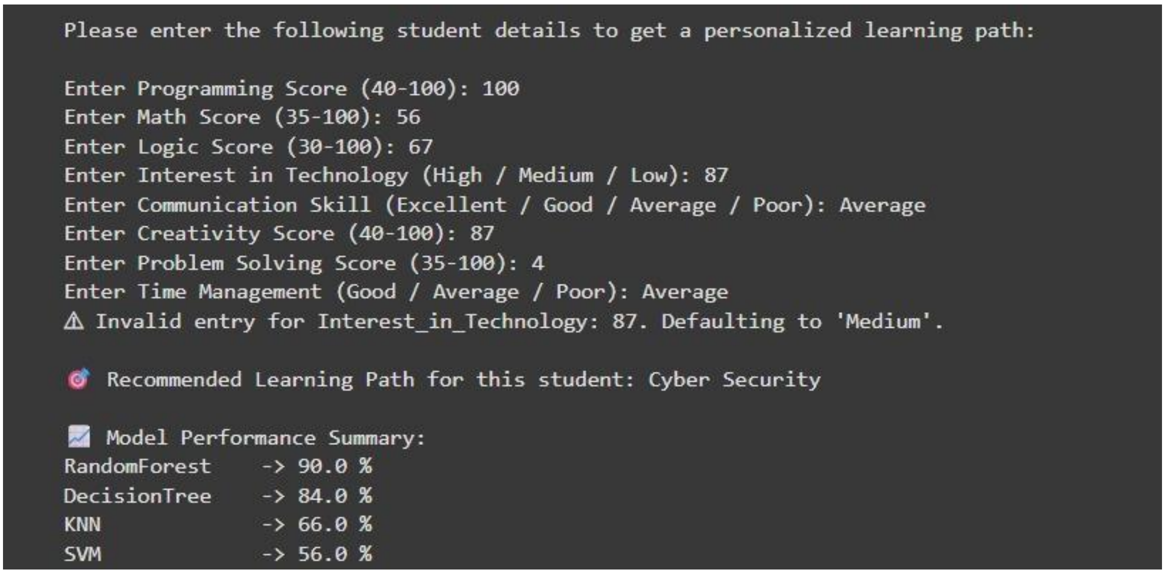

Fig .1. Screenshot of Personalized Learning Path

V. METHODOLOGY

The proposed methodology involves generating a dataset, training multiple ML models, and evaluating them based on predictive performance. The pipeline includes the following steps:

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131102

[1]. Data Preparation: Encoding categorical variables, normalizing numeric features, and splitting data into training and testing subsets.

[2]. Model Selection: Four models were trained and tested — Random Forest, Decision Tree, KNN, and SVM.

[3]. Training and Testing: Models were trained using 75% of the dataset and tested on the remaining 25%.

Each algorithm's accuracy and classification report were generated to identify the best model. The Random Forest model achieved the best results with an accuracy of 90%, outperforming Decision Tree (84%), KNN (66%), and SVM (56%). The Decision Tree performed well but exhibited minor overfitting. KNN and SVM struggled with non-linear decision boundaries, leading to lower recall values. Further analysis of Random Forest's confusion matrix confirmed minimal misclassifications. The feature importance graph revealed Programming, Logic, and Problem-Solving as the most influential predictors, while Creativity and Interest in Technology contributed moderately.

VI. RESULTS AND ANALYSIS

The model evaluation results confirmed the efficiency of ensemble learning methods. The Random Forest achieved an overall accuracy of 90%, with strong precision across most categories. The classification report showed that the macro average precision was 0.94, recall 0.82, and F1-score 0.85, demonstrating a consistent balance between false positives and false negatives. Decision Tree attained 84% accuracy, offering understandable decision logic but slightly reduced generalization. The KNN model (66%) suffered due to reliance on distance metrics with unevenly scaled data, and SVM (56%) underperformed due to its limited adaptability to non-linear patterns in this dataset.The accuracy comparison visually emphasizes Random Forest's superiority. The confusion matrix highlights that most predictions were correctly classified, especially for Web Development and AI & ML. Feature importance analysis demonstrated that Programming Score and Logic Score were the dominant features, while Creativity Score played a major role in predicting UI/UX Design.The system also includes an interactive component where users enter their individual scores. When tested with sample inputs, the model accurately recommended "Cyber Security" as the optimal path for a user with strong technical but moderate creative.

```
◆ DecisionTree Accuracy: 84.0%
              precision    recall  f1-score   support

           0       1.00      0.33      0.50         3
           1       0.96      0.82      0.88        28
           2       0.50      1.00      0.67         4
           3       0.82      0.90      0.86        10
           4       0.83      1.00      0.91         5

    accuracy                           0.84        50
   macro avg       0.82      0.81      0.76        50
weighted avg       0.88      0.84      0.84        50


◆ KNN Accuracy: 66.0%
              precision    recall  f1-score   support

           0       0.50      0.67      0.57         3
           1       0.72      0.82      0.77        28
           2       0.50      0.75      0.60         4
           3       0.80      0.40      0.53        10
           4       0.33      0.20      0.25         5

    accuracy                           0.66        50
   macro avg       0.57      0.57      0.54        50
weighted avg       0.67      0.66      0.64        50
```

Fig 2. Screenshot of Model Analysis

DOI: 10.17148/IJIREEICE.2025.131102

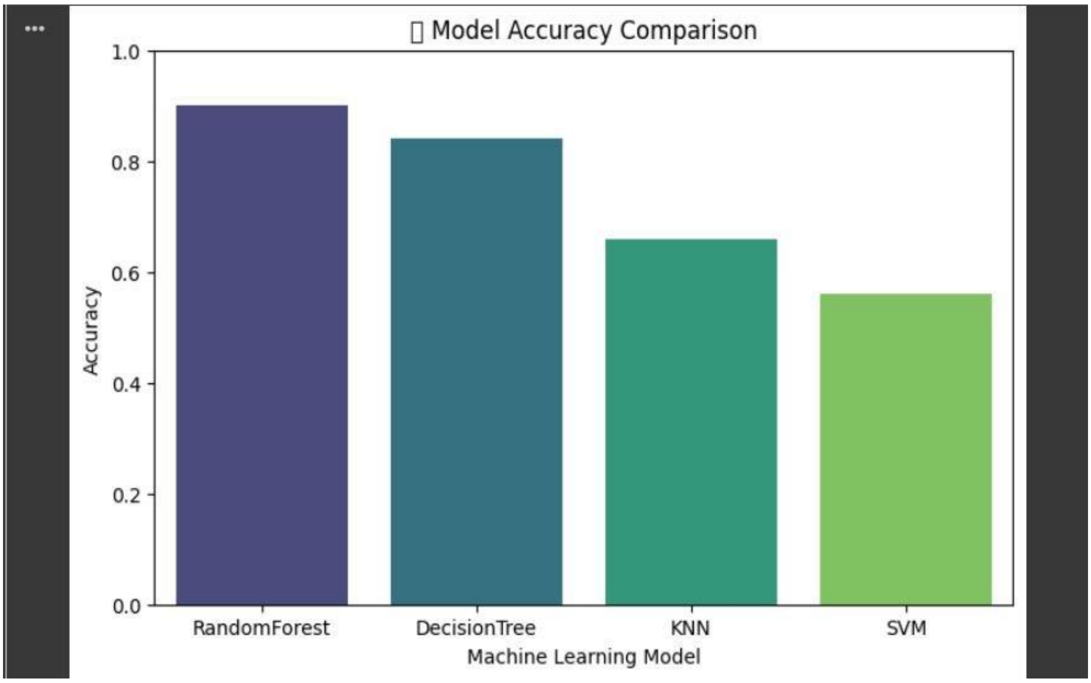

Fig.3. Screenshot of Model analysis 2

Advantages and Limitations

The primary advantage of this system lies in its adaptability and explainability. It does not rely on subjective assessments but rather uses objective data-driven decisions. The inclusion of diverse features—both academic and behavioral—makes the recommendations more personalized and realistic. The use of Random Forest ensures that predictions remain stable even when noise exists in the dataset. Another strength is the system's transparency: educators can easily visualize which traits influenced the prediction. The feature importance chart adds interpretability, making the recommendation process understandable for both students and teachers. However, some limitations exist. The dataset used is synthetic and relatively small; larger real-world datasets could improve accuracy and model robustness. Qualitative traits such as communication or creativity depend on user honesty during input, which may introduce bias. Additionally, more advanced techniques like deep learning or neural networks could be explored to handle complex relationships between features.

CONCLUSION AND FUTURE WORK

This study presented a machine learning-based Personalized Learning Path Recommendation System that successfully predicts suitable learning domains based on student performance data. By comparing multiple ML algorithms, Random Forest was identified as the best performer, achieving a 90% accuracy rate and superior balance. The inclusion of visualization tools such as accuracy comparison charts, confusion matrices, and feature importance graphs enhances both interpretability and usability. The system can be effectively applied in educational institutions, online learning platforms, and career counseling systems to provide personalized, data-driven recommendations. Future work will focus on extending the dataset with real student data, integrating deep learning architectures, and implementing cloud-based deployment for scalability. By continuously learning from new data, the system can evolve into a powerful decision-support framework that transforms academic guidance and empowers students to pursue their most promising learning paths.

REFERENCES

[1]. Chen, Y., & Zhang, H. (2020). Predictive modeling for personalized learning using machine learning techniques. IEEE Access, 8, 121450–121462.
[2]. Kumar, S., & Rao, P. (2021). Data-driven student profiling for adaptive learning systems. Journal of Educational Computing Research, 59(4), 789–804.
[3]. Han, J., Kamber, M., & Pei, J. (2020). Data Mining: Concepts and Techniques. Morgan Kaufmann.
[4]. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
[5]. Alazab, M., Tang, M., & Alazab, M. (2020). Hybrid learning techniques for intelligent classification systems. IEEE

DOI: 10.17148/IJIREEICE.2025.131102

Access, 8, 111123–111135.

[6]. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[7]. Sharma, R., & Patel, D. (2019). Intelligent decision support for student course selection. International Journal of Educational Technology, 6(2), 89–98.

[8]. Lee, H., & Park, J. (2020). An adaptive course recommendation framework using ensemble learning. Computers & Education, 149, 103835.

[9]. Harshni, S. Singh, R. Indhumathi, N. S. Kumar, S. Chowdhury and S. S. Patra, "DLIoMT: Deep Learning Approaches for IoMT-Overview, challenges and the future," *2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, India, 2025, pp. 1-6, doi: 10.1109/ICCECE61355.2025.10940162.

[10]. S. J. J. Thangaraj, N. S. Kumar, A. Saranya, C. R. Bhat, D. S. Kareem and H. Shareef, "A Development of Designing a Sensor N/W for Retrieving the Data from WSN via SBT," *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2024, pp. 194-198, doi: 10.1109/ICACITE60783.2024.10617435.

[11]. K. K. K, N. Sanjeev Kumar, S. C, V. V, G. Sangar and G. Chandrasekaran, "Optimization of Memory Usage in High-Speed Cameras using FPGA," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 427-432, doi: 10.1109/ICEARS64219.2025.10941569.

[12]. K. K. K, N. S. Kumar, S. C, N. R, G. Sangar and G. Chandrasekaran, "Denoising of MRI and Brain Tumor Classification using Local Binary Pattern and SVM Classifier," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 1854-1859, doi: 10.1109/ICEARS64219.2025.10940321.

[13]. N. S. Kumar, M. Lawanyashri, M. Sivaram, V. Porkodi, E. Gangadevi and G. N. Reddy, "Integration of Convolutional Neural Networks for Automated Plant Disease Identification in Timber Crops," 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, Pune, India, 2024, pp. 1-5, doi: 10.1109/TQCEBT59414.2024.10545228.

