Personalized Learning Path for Students Using Different Machine Learning Approach

1st Tejaswini Jaywant Durge
dept. of Information Technology
Vishwakarma Institute of Information Technology
Pune
tejaswini.22210270@viit.ac.in

2nd Arya Jalindar Kadam
dept. of Information Technology
Vishwakarma Institute of Information Technology
Pune
arya.22210766@viit.ac.in

***Abstract*—This paper introduces a machine learning approach to create personalized learning paths for university students based on their academic and career performance. By analyz- ing data such as Cumulative Grade Point Average (CGPA), hackathon participation, internships, Aptitude test scores and student's domain of interest to design a customized road-map for their personal growth. It uses data from previously placed students, including their referred courses and resources, to suggest similar learning materials to students with matching academic and technical profiles.**

The introduced paper aims to assist students in enhancing areas where they require the most support. In addition to incorporating a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis, the primary objective is to pinpoint each student's individual weaknesses and the areas where improve- ment is needed, the platform guides students toward taking real, actionable steps to close their skill gaps, boosting their confidence and making them more competitive in the job market. This approach not only helps students improve academically but also prepares them for better career opportunities.

***Index Terms*—Machine learning, SWOT Analysis, Personalized Road-map, Random Forest, K means**

I. Introduction

In the fast-paced world of higher education, students often struggle to identify the specific steps needed to improve academically and prepare for their careers. While traditional mentorship from seniors and professors can provide guidance, it's not always personalized enough to address each student's unique strengths and weaknesses. Machine learning offers a modern solution by using data-driven insights to create individualized learning paths.

This research introduces a platform designed to offer per- sonalized guidance using data from placed students—those who have already secured jobs. By analyzing their academic records and the courses and resources they found valuable, the platform suggests these same resources to students with similar academic and technical profiles. The system is trained on different models and the best-performing model is then used to make personalized recommendations, including relevant books, online courses, and tools to help students overcome their challenges. Through this personalized guidance, students can gain a deeper understanding of which subjects to prioritize, ways to bridge knowledge gaps, and strategies to enhance readiness for prospective career openings.

II. Literature Review

K-Means clustering, first choose the number of clusters (K) and pick K initial centroids randomly. Each data point is then assigned to the nearest centroid, forming clusters. Then update the centroids by averaging the points in each cluster and repeat the process until the centroids stabilize. The out- come is K clusters where similar data points are grouped. [1] By ensuring the centroids are well-spread from the beginning the enhanced approach helps prevent the algorithm from getting stuck in suboptimal solutions and improves overall clustering accuracy.

[2]studied student learning pathways using network models and deep learning. Their research demonstrated that the learning sequences of successful students have a fractal pattern, and deep learning networks could accurately predict student outcomes.

[3]applied supervised machine learning algorithms to predict student performance. The paper used decision trees, Bayesian networks, and other models to provide early predictions, helping educators identify weaker students and provide targeted interventions.

[4] compared classification results of Random Forest and J48 decision trees on 20 versatile datasets. The study concluded that Random Forest performs better on large datasets, while J48 is more efficient for smaller datasets, demonstrating how dataset size impacts model accuracy.

III. Methodology

A. Traditional Student Pathway Suggestion

The traditional way of finding personalised pathway to success are as follows:

Students collect data from various sources like their peers and teachers to find what suits their interest and what is the perfect path to follow in future. Then the student used to collect the data and analyze it and then follow respectively but even then, there is no chance of getting success, as the student might fail in the middle of the pathway cause of various undiscovered difficulties.

This is not only time-consuming but very confusing and does not provide a lot of assurance which leads to confusion and misleading. Sometimes it's important to know the success of the person suggesting the pathway itself so as to assure it. Moreover, this approach depended heavily on external opin- ions and lacked a structured or data-driven process. Students had to rely on trial and error, often wasting valuable time and effort on paths that turned out to be unfit. The lack of personalized guidance meant that students could end up in pathways that didn't suit their individual learning styles, interests, or strengths. This uncertainty and lack of tailored direction made the traditional pathway gathering inefficient and frustrating, leading to missed opportunities or setbacks when the chosen route proved incompatible with their capabilities or aspirations.

B. *Machine Learning Custom Pathway Suggestion*

The new methodology for student pathway suggestion using machine learning:

This Application uses the earlier student data to train the model and analyze the student's capacity and the reliability with the curriculum. By collecting student data from various students which include the column seen in Table I.

As seen in Table I there are three distinct columns hence there is a use of three different machine learning models in this algorithm.

TABLE I
FEATURES FOR TRAINING MODELS

Test Scores (9 cols)	**Coding and Platform Scores (5 cols)**	**Extra Activities (4 cols)**
These include the test scores from the syllabus and from other similar examinations	This includes other scores like on coding platforms	These include the extra activity scores that the student might have taken part in like hackathons and internships

1) Model 1: Coding and Platform Scores: This part of the data is trained on Unsupervised models so as to make clustering of the data points. To find the effective unsupervised clustering model for this specific data set different tests were done. The results of the test are as follows:

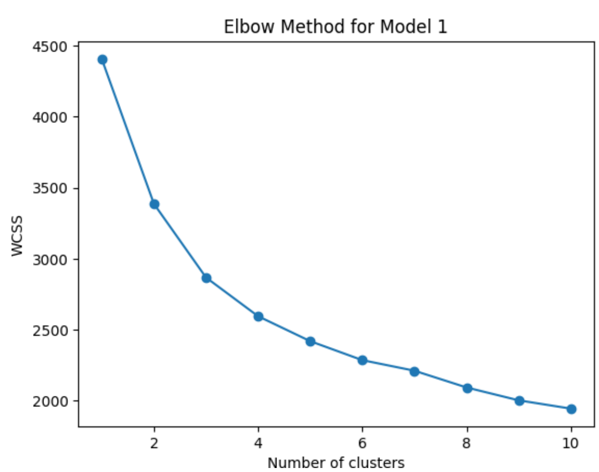

Fig. 1.

Firstly, to decide the number of clusters elbow method for model 1 was taken and from the Fig 1 it was decided to consider the number of clusters to be 4.

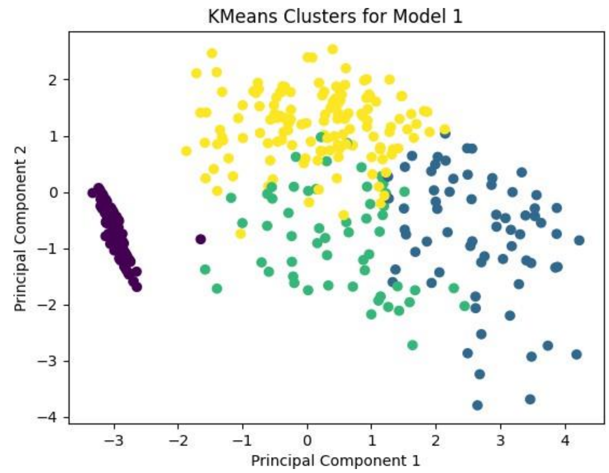

Fig. 2.

From the Fig 2 there are four distinct clusters formed from the Model 1 which can be further used for the prediction

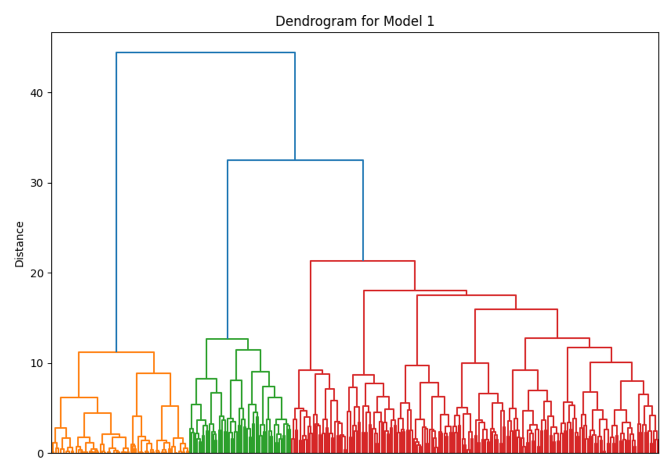

Fig. 3.

From the Hierarchical clustering model, 3 different clusters were identified and were plotted in the form of dendrogram.

From these 2 models the means square error for K means model was found to be less than the hierarchical clustering model.

2) Model 2: Extra Activities: The same type to methodology was proposed as for the *Model 1* part.

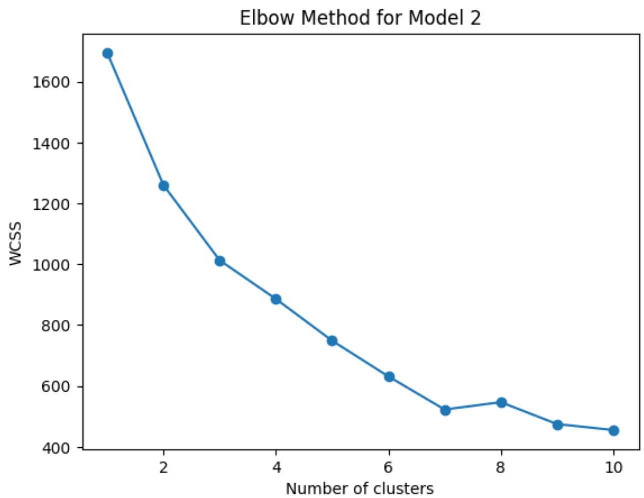

Fig. 4.

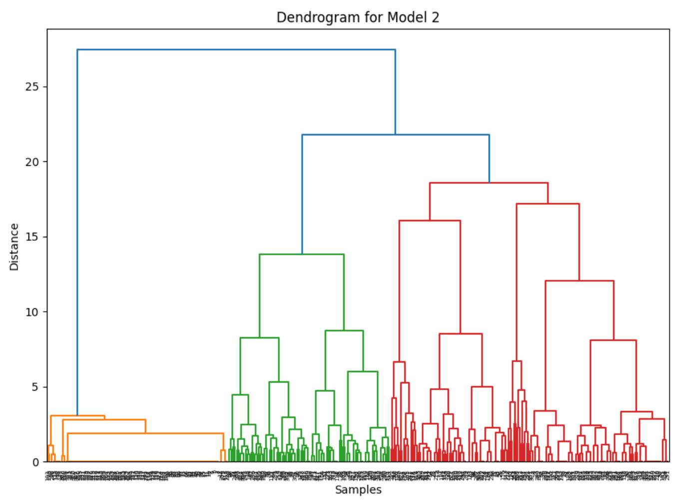

Fig. 5.

From Fig.3 and Fig.4 it is observed that between the 2 models the means square error for K means model was found to be less than the hierarchical clustering model and hence the K means was chosen.

3) Model 3: Test Scores: Several supervised learning models were tested, including Random Forest (RF), Decision Trees (DT), Deep Neural Networks (DNN), and Logistic Regression (LR). The comparison results are as follows:

MSE (Mean Squared Error)

- LR MSE: 86.2151
- RF MSE: 1.101
- DT MSE: 1.651
- DNN MSE: 17.866

From the from these results, it is clear that Random Forest (RF) significantly outperformed the other models, achieving the lowest MSE of 1.101. This indicates that RF provided the most accurate predictions, effectively capturing the underlying patterns in the data. Decision Trees (DT) also performed well, with an MSE of 1.651, though not as robust as RF. Deep Neural Networks (DNN) and Logistic Regression (LR) had notably higher MSE values, with LR being the least effective model for this dataset.

Given the superior performance of Random Forest, it was selected as the most suitable model for further predictions and analysis in this context.

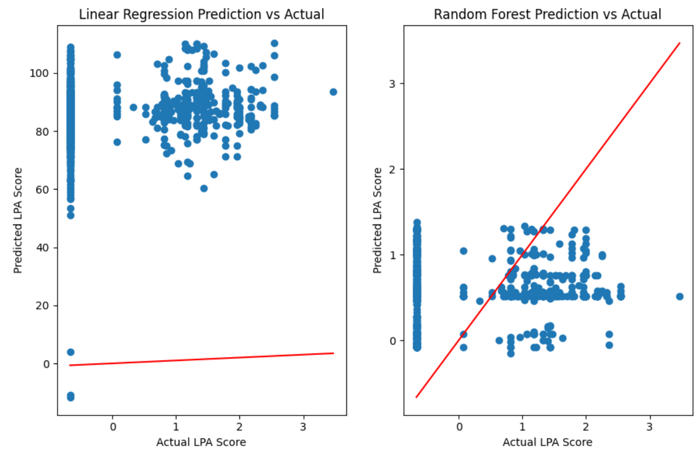

Fig. 6.

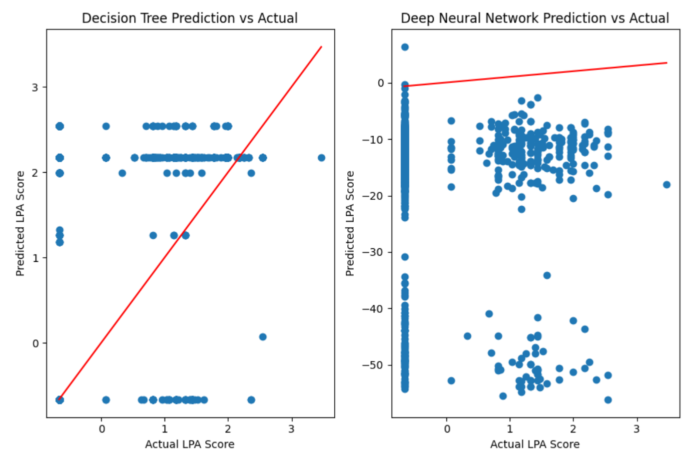

Fig. 7.

Fig.6, Fig.7 shows that Random Forest can best predict the scores.

C. Final Pathway Suggestion

A student data is collected which includes all the described columns in Table I. This student data is collected from senior/alumni/graduates students from the institute. Along with the other data as described earlier it also includes the pathway which that specific student followed during his student life in that institute.

Suppose we need to find the correct pathway for the student ‘A’. The three different models can be used for three different tasks the

- Model 3: This model will be used to authenticate the student data and tell if the student’s suggested pathway is reliable or not, whether it’s up to the mark or not. This model will output a 1-100 score/rating that will decide whether this student has done better in the past or will do better in past. Which alternately tell us if its suggested pathway reliable.

After the student rating for this dataset has been predicted Model 1 and Model 2 can be used to match the similarities of the student ‘A’ with this dataset students (senior/alumni/graduates).

- Model 1 and Model 2: By using the model predict method we could find the models clustering group for this ‘A’ student. We run the same predict method for all the students in the student dataset.
  After this is done we can find the same student that matches the same interest as well as lies in the same cluster that the student ‘A’ lies in and suggests the dataset student’s pathway if the dataset student rating is high enough.

IV. RESULTS

An Interactive Application is made with the help of python and tkinter python-library. The UI of the application window can be seen in Fig.8.

Fig. 8.

By entering the students index number the student can get its personalized pathway according to the machine learning done in the background. This tool provides a streamlined and data-driven alternative to the traditional trial-and-error method, significantly reducing the time and effort needed to find the best personalized pathway for students.

V. COMPARISON WITH OTHER METHODOLOGIES

An Interactive Application is made with the help of python and tkinter python-library. The UI of the application window can be seen in Fig.8.

VI. CONCLUSION

In this paper, we have presented a machine learning-based approach for providing personalized learning pathways to university students. By analyzing data such as academic performance, coding platform scores, extra-curricular activities, and student interests, the platform offers tailored recommendations to guide students on their academic and career journeys. The system leverages advanced models, including Random Forest, K-Means clustering, and Decision Trees, to identify patterns and predict the most suitable pathways.

Compared to traditional methods of seeking advice from peers and teachers, the machine learning approach offers a more data-driven, structured, and personalized solution. It minimizes trial and error, allowing students to focus on areas where they need improvement and prepare more effectively for their future careers. The inclusion of a SWOT analysis further enhances the platform's ability to pinpoint individual weaknesses and recommend actionable steps to address them.

The developed interactive application, which simplifies pathway suggestions, demonstrates the power of combining data science and education to enhance student outcomes. With real-time feedback and tailored recommendations, this platform represents a significant advancement in academic guidance, offering students a clear, focused path toward success and greater preparedness for the job market.

REFERENCES

[1] Y. Li and H. Wu, “A Clustering Method Based on K-Means Algorithm,” Phys Procedia, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.
[2] P. Ortiz-Vilchis and A. Ramirez-Arellano, “Learning Pathways and Students Performance: A Dynamic Complex System,” Entropy, vol. 25, no. 2, Feb. 2023, doi: 10.3390/e25020291.
[3] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, “Student Performance Prediction Model based on Supervised Machine Learning Algorithms,” in IOP Conference Series: Materials Science and Engineering, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1757-899X/928/3/032019.
[4] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, “Random Forests and Decision Trees,” 2012. [Online]. Available: www.IJCSI.org
[5] S. Kumar and S. Kumar Yadav, “Data Mining: A Prediction for Performance Improvement of EngineeringStudents using Classification Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification Saurabh Pal,” 2012. [Online]. Available: https://www.researchgate.net/publication/221710771
[6] Dr. L. A. Deshpande, “Revolutionizing The Education Industry: Swot Analysis And Predictive Modeling Approach To Enhance Students’ Educational Proficiency,” Educational Administration: Theory and Practice, pp. 3758–3765, May 2024, doi: 10.53555/kuey.v30i5.3530.
[7] A. S. Courtner, “Impact of Student Engagement on Academic Performance and Quality of Relationships of Traditional and Nontraditional Students,” International Journal of Education, vol. 6, no. 2, p. 24, May 2014, doi: 10.5296/ije.v6i2.5316.
[8] V. Desai and K. Oza, “Personalized Learning Approach for Outcome Based Distance Education.” [Online]. Available: https://www.researchgate.net/publication/378364655
[9] M. J. Jacobson, “Complex systems in education: Scientific and educational importance and implications for the learning sciences,” 2006. [Online]. Available: https://www.researchgate.net/publication/220040401