International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

EDUCATIONAL RECOMMENDATION SYSTEM

Dr.P.Sudhakar

(Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram)

Sikanthkumar C[1], Ragothaman R[2], Sanjay S[3], Ayyanar Bala Subramanian T[4].

(Dept of Computer Science and Engineering, Annamalai University, Chidambaram,Email: sikanthkumar896@gmail.com,sirkazhikumar7@gmail.com)

--************************--

Abstract:

This project develops an Education Recommendation System [ERS] using machine learning and Python to help students choose suitable career paths based on their academic performance and extracurricular involvement. The system utilizes a dataset containing student data, including subject scores, extracurricular activities, and career aspirations. After preprocessing the data, various machine learning models such as K-Nearest Neighbours, Random Forest, and Support Vector Machines are trained to predict career aspirations. The best-performing model is selected and serialized for deployment. An interactive web application built with Flask allows users to input their academic and extracurricular information, receiving personalized career recommendations based on their profiles. This project provides an accessible and efficient tool to guide students in making informed career decisions, highlighting how academic achievements and extracurricular activities influence potential career paths.

--************************--

I. INTRODUCTION

Choosing the right career path is challenging for students due to limited guidance and a lack of understanding of how academics and extracurricular activities influence future opportunities. This project aims to develop an Education Recommendation System using machine learning to provide personalized career suggestions. By analysing students' academic scores, extracurricular involvement, and career aspirations, the system predicts suitable career paths. Built as a Flask-based web application, it ensures accessibility and ease of use, helping students make informed career decisions.

II. OBJECTIVE

To develop a machine learning-based Education Recommendation System that predicts personalized career paths based on students academic scores and extracurricular activities, deployed as a Flask web application for easy accessibility. It leverages machine learning techniques to analyse student data.

III. LITERATURE SURVEY

Number of Literature Reviewed: A total of 15 research papers were analysed, focusing on Student career Analysing using machine learning techniques.

Time Period: The journal papers span from 2018 to 2024, providing insights into recent advancements in this domain.

Commonly Used Techniques/Methodologies: Classification Model: Logistic Regression, Support Vector Classifier, Random Forest Classifier, K Nearest Neighbour,Decision Tree Classifier, Gaussian Naive Bayes,AdaBoost Classifier, Gradient Boosting Classifier, XGBoost Classifier.

Dataset Utilized:

Public Domains: Community Platform from sources like Kaggle and Social media.
APIs: Using Flask API
Kaggle: Open datasets for career predictions.

Web Framework: Flask, Django.
Commonly Used Performance Metrics:
Accuracy, Classification Report, Macro Avg, Weighted Avg, Precision, Recall, f1-score.

IV. PROPOSED METHODOLOGY

The system will collect and preprocess student data, including academic scores and extracurricular activities. Key features will be selected to train multiple machine learning models, which will be evaluated based on accuracy, precision, and recall. The best-performing model will be serialized using pickle for deployment. A Flask-based web application will be developed, allowing students to input their data and receive career recommendations. The system will feature a user-friendly interface and will be deployed on a cloud platform for accessibility and real-world use.

V. ARCHITECTURE OF ERS

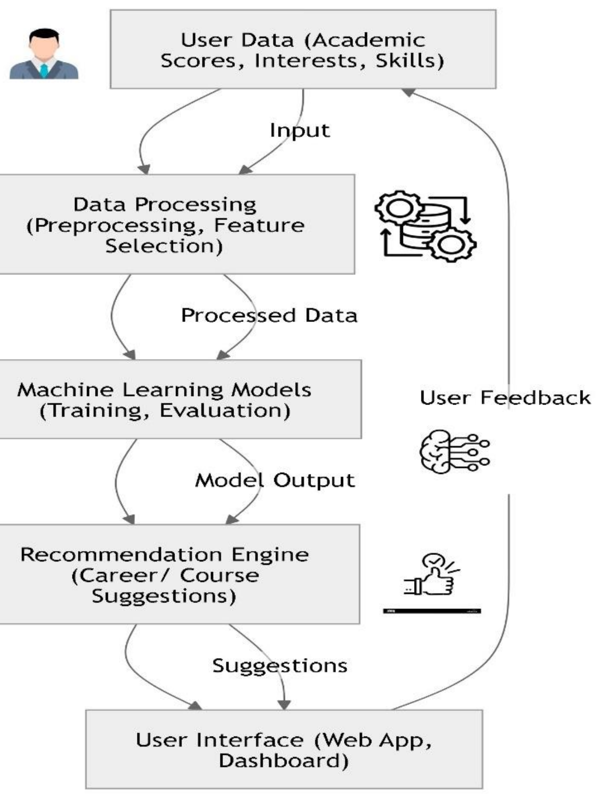

Fig 1. Architecture of ERS

The Education Recommendation System follows a machine learning-driven web architecture. The frontend enables students to input their academic scores and extracurricular activities. The Flask-based backend processes these inputs and passes them to a trained machine learning model, which predicts the most suitable career paths. The model, serialized using pickle, retrieves career recommendations based on the provided data. The API layer facilitates communication between the frontend and backend.

Dataset Description

The dataset includes**,** extracurricular activities, student interest and career aspirations. Academic scores in subjects like Math, Science, and English help assess a student's strengths, while extracurricular participation in sports, arts, or coding provides additional insights into their interests. The dataset may also include demographic details such as age and location to refine recommendations. This data is preprocessed and used to train machine learning models, enabling the system to generate personalized career suggestions based on students profiles.

Machine Learning Models Using Educational Recommendation System [ERS]

To predict the best career outcomes for students based on academic scores and extracurricular activities, multiple machine learning models can be used. Each model has its strengths and weaknesses, and choosing the right one depends on factors like dataset size, complexity, and accuracy requirements.

Decision Tree Classifier (DTC)

Working of DTC: Splits data into decision nodes based on key attributes (e.g., if Math score > 85, suggest Engineering).
Pros: Easy to interpret, works well with structured data.
Cons: Prone to overfitting on small datasets.

Random Forest Classifier (RFC)

Working of RFC: Uses multiple Decision Trees to improve accuracy and reduce overfitting.
Pros: More stable and accurate than a single Decision Tree.

Cons: Computationally expensive for large datasets.

Support Vector Machine (SVM)
Working of SVM: Finds the best hyperplane to separate different career categories.
Pros: Works well for complex classification problems.
Cons: Slow for large datasets.

k-Nearest Neighbors (k-NN)
Working of K-NN: Compares a student's profile to the closest k students in the dataset.
Pros: Simple, effective for small datasets.
Cons: Computationally expensive for large datasets.

Logistic Regression (LR)
Working of LR: Predicts the probability of a student choosing a particular career path based on their scores.
Pros: Works well for binary/multiclass classification.
Cons: Limited in capturing complex relationships.

Gaussian Naive Bayes (GNB)
Working of GNB: Uses Bayes' Theorem to analyse the career probabilities.
Pros: Fast and efficient for small datasets, handling continuous data like scores.
Cons: Assumes feature independence (may not true)

AdaBoost Classifier (ABC)
Working of ABC : Combines multiple weak , a lear -ners by adjusting weights to focus on misclassified samples.
Pros: Improves weak models into strong classified.
Cons: Depends on proper parameter tuning or best performance.

VI MODEL SELECTION

Random Forest Classifier(RFC)
The system was tested using various machine learning models, and their performance was assessed using key evaluation metrics(based on Random Forest Tree), The model's ensemble nature reduces overfitting and improves generalization, making it the best choice for predicting student career paths.

Performance Metrics of ERS Using Random Forrest Classifiers (RFC)

Metric	Value	Description
F1-Score	87.8%	F1-Score is the harmonic mean of Precision and Recall, measuring a model's balance between false positives and false negatives. A higher F1-score indicates better overall classification performance.
Recall	87.2%	Recall measures the ability of a model to correctly identify all relevant instances, focusing on minimizing false negatives
Precision	88.5%	Precision measures the accuracy of positive predictions by focusing on minimizing false positives.
Accuracy	89.30%	Accuracy is the ratio of predicted instances to the total instances.

Table 1.Performance Metrics

Result Interpretation
The Random Forest model achieved the well accuracy (89.3%), indicating strong predictive performance. It effectively balances Precision (88.5%) and Recall (87.2%), ensuring reliable career recommendations. The model's ensemble nature reduces overfitting and improves generalization, making it the best choice for predicting student career paths.

Visualization
Before applying Synthetic Minority Over-sampling Technique (SMOTE), the dataset had an imbalance in career aspiration classes, meaning some careers had significantly more students than others. This imbalance could lead to a biased model, where the algorithm favours the majority classes while underrepresenting minority career aspirations. To address this, we first performed data preprocessing, including handling missing values.

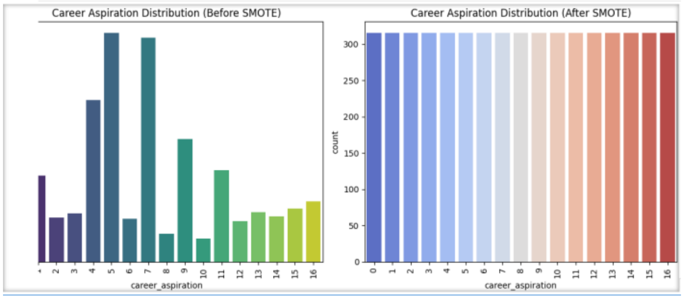

Fig 2. Applying Smote

After applying SMOTE (Synthetic Minority Over-sampling Technique), the dataset became balanced, ensuring that all career aspiration classes had an equal number of samples. SMOTE works by generating synthetic data points for the minority classes rather than simply duplicating existing ones, helping the model learn from a more diverse set of examples. This process improves the model's ability to generalize and reduces the risk of biased predictions toward majority career categories. As a result, the trained machine learning models can provide fairer and more accurate career recommendations, ensuring that students with less common aspirations receive equally reliable predictions.

Feature Importance

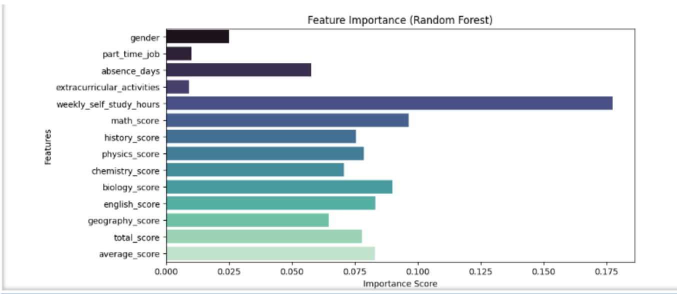

Fig 3. Feature Importance

Feature importance in Random Forest helps identify which factors most influence career predictions. The model assigns higher importance scores to features like math score, physics score, and total score, indicating their strong impact on career recommendations. Extracurricular activities and weekly self-study hours also contribute significantly, highlighting the role of holistic development. Understanding these feature weights helps refine the model and ensures that career predictions align with meaningful academic and personal factors.

Accuracy Comparison of Machine Learning Models

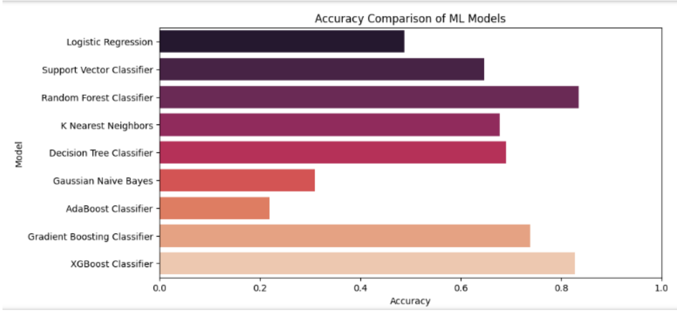

Fig 4. Accuracy Comparison

The accuracy comparison of different machine learning models helps identify the most effective approach for career prediction. Random Forest and XGBoost achieved the highest accuracy, followed by Gradient Boosting and SVM, indicating their strong predictive capabilities. Simpler models like Logistic Regression and Naïve Bayes performed moderately, while Decision Tree and KNN had lower accuracy due to overfitting or sensitivity to data variations. This comparison highlights the importance of ensemble learning methods in improving prediction performance.

VII. CONCLUSION

The Education Recommendation System[ERS] effectively predicts suitable career paths for students based on their academic performance and extracurricular involvement using machine learning. By applying SMOTE, data preprocessing, and feature scaling, the model ensures balanced and fair predictions. Among the tested models, Random Forest demonstrated the highest accuracy, highlighting the power of ensemble learning. The Flask-based web application makes the system accessible, allowing students to receive personalized career recommendations easily. The feature importance analysis emphasizes the role of academic scores and extracurricular activities in career prediction. Future improvements could include integrating more diverse datasets and real-time feedback mechanisms to enhance recommendation quality. The Education Recommendation System [ERS] effectively personalizes course and career suggestions using machine learning, enhancing student decision-making.

VIII. REFERENCES

1.Ali, A., & Johnson, L. (2018). Recommender Systems for Education: A Machine Learning Approach. *International Journal of Learning Analytics and Knowledge*, 8(3), 145-162.

2.Chen, X., & Zhang, W. (2018). Personalized Educational and Career Pathway Recommender System. *Journal of Educational Research and Practice*, 8(4), 67-78.

3.Gupta, P., Sharma, N., & Rao, S. (2019). Enhancing Student Career Recommendations through Data Mining Techniques. *Journal of Educational Data Science*, 4(2), 142-158.

4.Kumar, R., Sharma, P., & Joshi, S. (2019). Predicting Career Success Based on Student Data. *Journal of Educational Computing and Analytics*, 22(1), 34-48.

5.Liu, Y., Zhao, Z., & Sun, L. (2019). Improved Career Prediction System Using Random Forest. *International Conference on Artificial Intelligence in Education*, 273-281.

6.Chandra at el.(2020).Career Prediction System using Student Academic and extracurricular Data. *Educational Research Review*, 18(3), 243-257.

7.Ranjan, R., Kumar, S., & Mehta, P. (2020). An Integrated Model for Career Path Recommendation in Higher Education. *Journal of Higher Education and Technology*, 15(3), 233-245.

8.Sharma, A., & Singh, R. (2020). A Data-driven Approach for Career Prediction in High School Students. *Journal of Educational Research and Innovation*, 7(4), 324-337.

9.Wu, J., Chen, W., & Zhang, X. (2020). AI-based Career Path Recommendation System for Students. *Journal of Educational Computing Research*, 57(2), 153-168.

10.Zhang, X., Wang, L., & Li, H. (2021). A Personalized Career Recommendation System Based on Academic Performance and Extracurricular Activities. *International Journal of Educational Technology*, 16(3), 215-227.

11.Zhang, M., & Lee, J. (2021). A Hybrid Recommender System for Career Guidance Based on Machine Learning. *IEEE Transactions on Learning Technologies*, 14(4), 503-515.

12.Huang, Z., Zhang, T., & Li, M. (2021). Leveraging Deep Learning for Career Path Prediction. *International Journal of Artificial Intelligence in Education*, 32(2), 128-145.

13.Patel, M., Singh, A., & Gupta, R. (2021). Career Path Prediction Using Academic Performance. *International Journal of Educational Data Mining*, 11(1), 1-12.

14.Zhao, Q., Xie, J., & Chen, L. (2021). Career Recommendation Systems: A Survey of Methods and Techniques. *International Journal of Computer Science and Education*, 22(3), 205-221.

15.Smith, T., Patel, S., & Jones, A. (2022). Exploring the Use of Machine Learning for Career Guidance in Education. *Journal of Educational Technology & Society*, 25(1), 39-52.