Career Path Forecasting

ABHISHEK G DANGI   
 *Department of CSE (Data Science)*  
 *Dayananda Sagar College of Engineering*Bengaluru  
abhishekgdangi033@gmail.com

SAMBHRAM S NAREGAL  
*Department of CSE (Data Science)*  
 *Dayananda Sagar College of Engineering*Bengaluru  
sambhramnaregal22@gmail.com AMARNATH V  
*Department of CSE (Data Science)*  
 *Dayananda Sagar College of Engineering*Bengaluru  
amarnath29052004@gmail.com

lK YUVRAJ SUNDARAM  
*Department of CSE (Data Science)*  
 *Dayananda Sagar College of Engineering*Bengaluru  
yuvrajsundram18@gmail.com

*Abstract*—Contemporary education requires a general assessment of students' potential along with grades. In this regard, an Academic Performance Indicator (API) score between 0 and 10 on the basis of machine learning, taking into account students as a whole in academics, co-curricular activities, internship, and online certifications, is suggested. The system is trained on a database of more than 3000 students and incorporates Random Forest model to generate API scores and suggest appropriate career paths (job, higher education, or entrepreneurship). The result illustrates the capability of the model in generating strong API scores and career path predictions and introduces an efficient tool to be scaled to institutional analytics as well as to advise students. Our prototype system, trained on real world dataset shows promising results in accurate profiling and personalized career recommendations.

Keywords—Academic performance, API Score, machine learning, career prediction, student profiling, Random Forest,

# INTRODUCTION

The conventional education assessment system has been dominated by a single performance parameter—scholastic marks for ages. In the majority of institutions, Grade Point Average (GPA) or percentage marks are the only parameters to measure a student's capability, intelligence, and potential. But as the needs of the contemporary industries change and as education becomes diversified, such limited performance parameters have increasingly often been found wanting. Students today participate in extracurricular activities such as co-curricular competitions, internships, and online self-learning, all of which play a critical role in their overall development. These aspects are not generally taken into account in conventional grading systems, resulting in a huge gap in student assessment and career readiness evaluation.

Institutions of learning and employers are finally getting the concept that exam scores are insufficient to judge the actual potential or future success of a student. A student participating in hackathons, industry internships, or online certifications may be more industry-ready and job-ready than an exam-smart but otherwise unskilled student. Moreover, students participating in leadership activities, entrepreneurial ventures, or socially responsible initiatives are likely to demonstrate innovation, grit, and initiative that are impossible to quantify with exam scores.

To meet this increasing gap, this paper introduces a new measure, the Academic Performance Indicator (API) score, a normalized composite score from 0 to 10 that takes into account four major pillars of student development: academic performance, co-curricular activities, internship experience, and online course completions. The API score is intended to provide a comprehensive and data-driven picture of student performance. In contrast to rigid traditional approaches, the API framework is dynamic, transparent, and extensible, and therefore more responsive to the demands of contemporary education. Today's globalized and increasingly competitive labour market leaves students in dire uncertainty about what to do after graduation. Do they enter employment, pursue further studies, or seek entrepreneurship? The decision is critical and must be made after having a precise notion of their very own strengths, experiences, and abilities. Thus, this research contributes an additional step further by using machine learning algorithms such as Random Forest not only to calculate the API score but also to predict a student's best post-academic path—either employment, further studies, or entrepreneurship.

The system is trained and tested on a dataset of more than 3000 student records covering a wide range of educational, co-curricular, and professional activities. This hybrid model of student evaluation and counselling is not only useful to the students but also to placement cells, recruiters, and institutions. It facilitates data-driven decision-making and assists in the identification of students with balanced skills. By shifting away from one-dimensional evaluation and adopting a holistic, AI-driven model, this study paves the way for the next generation of academic analytics and student success prediction.

# LITERATURE STUDY

In the study [1], supervised learning techniques such as regression and decision tree classifiers were implemented to predict students’ academic performance. Feature selection was optimized using a Genetic Algorithm (GA), and the models were validated through cross-validation to ensure reliability.

In the study [2], a systematic review and meta-analysis were conducted to examine the association between emotional intelligence and personality traits—including the Big Five, self-esteem, self-efficacy, optimism, and proactive personality—as well as career adaptability. Heterogeneity was evaluated using subgroup analysis, while publication bias was assessed using Egger’s regression and Orwin’s fail-safe N.

In the study [3], data were gathered from students at Al Akhawayn University and subjected to preprocessing to clean and structure academic records. Clustering, regression, and machine learning techniques were then applied using Python libraries such as NumPy, Pandas, and Scikit-learn to analyze recent trends in student performance and project future academic outcomes.

In the study [4], the dataset underwent preprocessing involving array reshaping, elimination of irrelevant elements, and sorting based on GPA, credits earned, and birth year. Clustering and regression techniques, facilitated by Python libraries like NumPy, Pandas, and Scikit-learn, were employed to identify learning trends and forecast academic performance.

In the study [5], marks were predicted using a regression model, while grades were classified via a decision tree classifier. Feature selection was optimized through a Genetic Algorithm (GA), and cross-validation methods were used to assess model performance. The GA-enhanced decision tree classifier demonstrated superior accuracy compared to other models.

In the study [6], a systematic literature review was conducted to explore various data-driven approaches aimed at predicting student performance. The process involved identifying and analyzing relevant research articles, categorizing prediction strategies into machine learning and statistical methods, and evaluating them based on accuracy, feature sets, and application domains.

In the study [7], a web application was developed to estimate learners' academic achievement using machine learning algorithms. Data collection, cleaning, encoding, and scaling were carried out, and multiple models were evaluated. The Random Forest algorithm exhibited the best performance. The application highlights significant academic variables and enables score prediction via an intuitive interface.

In the study [8], a backpropagation neural network using a multi-layer perceptron architecture was applied, utilizing Tanh and SoftMax activation functions. Data preprocessing included cleaning, integration, and normalization. K-fold cross-validation was used for training, and results were assessed using accuracy, precision, recall, and AUC metrics. Academic performance was categorized into eight hierarchical levels based on the evaluation.

In the study [9], a real university dataset was collected and preprocessed through cleaning, encoding, discretization, and feature scaling. To address class imbalance, resampling techniques such as SMOTE and ADASYN were employed. Both deep neural networks and traditional classifiers were trained and validated using an independent test set and stratified 5-fold cross-validation.

In the study [10], data mining techniques such as Decision Trees, Support Vector Machines (SVM), and Naive Bayes classifiers were utilized to forecast academic performance based on pre-admission scores (HSGA, SAAT, GAT). Data preprocessing included record cleansing and transformation. Models were evaluated using accuracy, precision, recall, and F1-score, with ANN yielding the highest performance. The study's outcomes influenced changes in admission requirements to enhance student success.

# METHODOLOGY

The process is a well-planned machine learning pipeline that forecasts a student's API score and recommends a relevant career. It consists of several steps like data collection, preprocessing, computation of API score, application of machine learning models, evaluation, and system integration. The architecture is designed in a way to make it scalable and flexible to accommodate changing datasets.

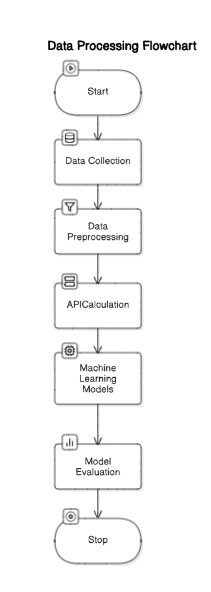


Fig.1 Data Flow Diagram

## Data Collection

The data set consists of student records, obtained from academic portals and validated self-reported profiles. The distinguishing features are:

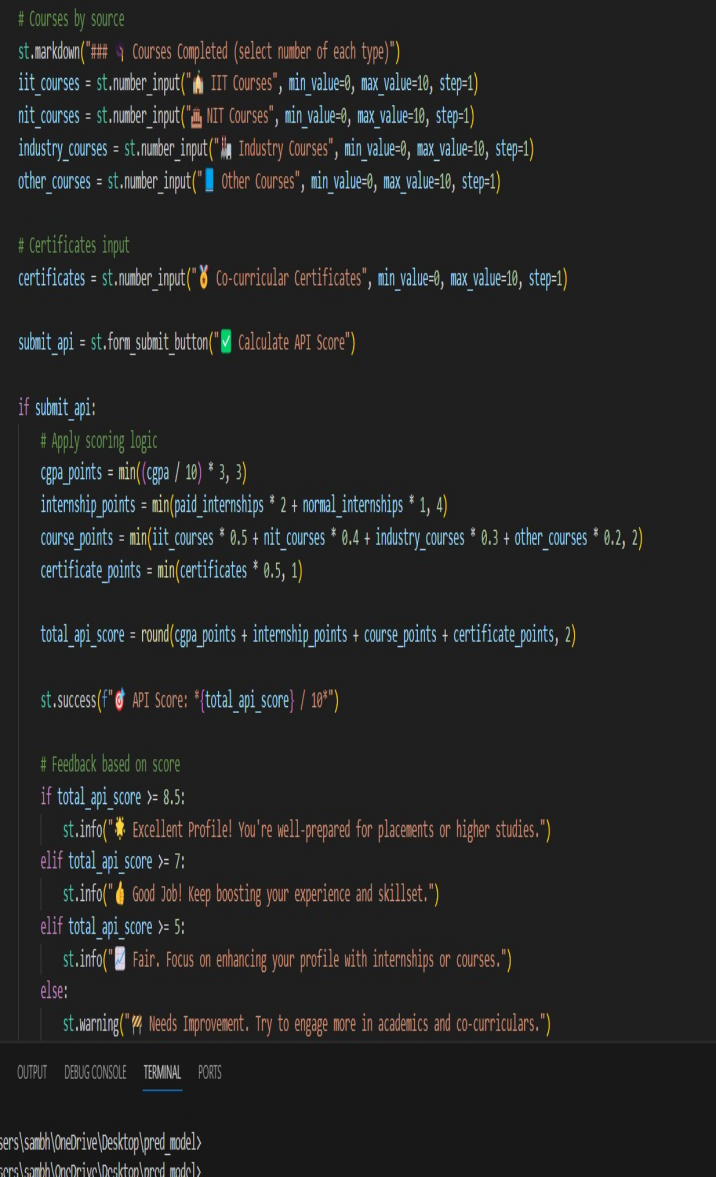
* Academic marks (semester-wise GPA)
* Number of co-curricular activities participated or won
* Internship details (duration, size of firm, relevance)
* Quantity and nature of online certifications (from platforms like Coursera, Udemy, etc.)
* Career outcome (employed, higher education, business start-up)

## Data Preprocessing

* Missing Value Handling: Mean/median imputation for numerical data.
* Normalization: All the numeric features are normalized between the range of 0–1.
* Categorical Encoding: Last career path labels are converted to numerical form.
* Feature Aggregation: Weighted indices are constructed to measure co-curricular engagement, quality of internships, and intensity of online learning.

## Academic Performance Indicator (API) Computation

Academic Performance Indicator (API) is a composite score of 0-10, intended to gauge the overall academic and professional readiness of a student. It combines various elements of a student's profile—academic performance, co-curricular activities, internship exposure, and online course participation. It is a cumulative score that indicates not only excellence in the class but also field exposure, professionalism, and self-improvement.



Each of these indices is normalized to a common scale (typically 0–1) to ensure consistency and comparability.

## Machine Learning Models

1. Random Forest Regressor: Provides the API score as a result of a feature vector from all normalized inputs. Chosen due to its robustness and capability to handle mixed data types.
2. Decision Tree Classifier: It classifies the most probable career path using the detailed student profile with API score and individual measures.

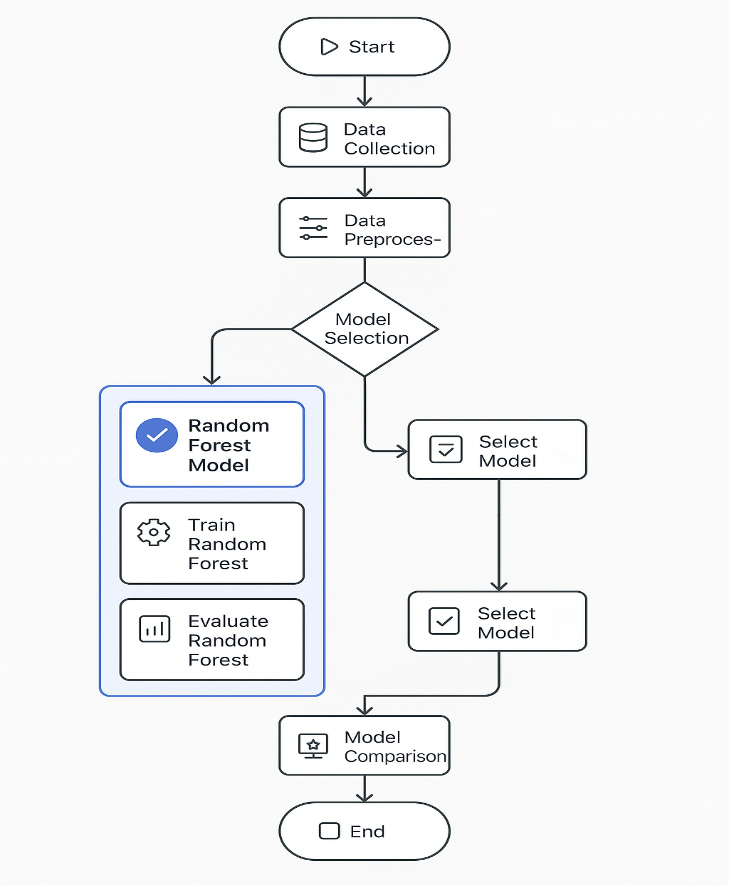


Fig.2 Machine Learning Model

## Model Evaluation

We have employed two kinds of evaluation techniques:

1. For API Score Prediction (Regression Task):

We apply models such as Random Forest to forecast the API score. Here, we quantify how well the predicted score approximates the true score in terms of these measures:

* RMSE (Root Mean Squared Error): Shows how far predictions are from actual scores on average. Lower is better.
* MAE (Mean Absolute Error): The average of the difference between the predicted and actual scores. Also, the lower the better.

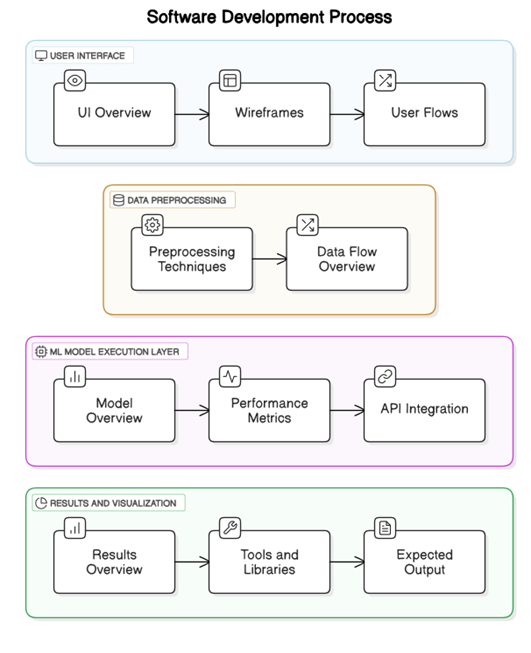
These steps help us to calibrate the model to be as accurate as possible in grading students.

2. For Career Path Prediction (Classification Task)

We use models like Random Forest to predict what type of career a student will pursue. We confirm this with:

* Accuracy: The proportion of correct predictions out of all predictions.
* Precision: Out of all the students predicted to choose certain career path, how many actually did.
* Recall: Out of all the students who actually choose a certain career path, how many were correctly predicted.

## System Architecture



# RESULTS AND DISCUSSION

The performance achieved through training and testing the designed machine learning models is both revealing and promising. The analysis was performed over a dataset of more than 3000 student profiles, and the performance of each model was measured in terms of the appropriate metrics like accuracy, precision, recall, RMSE (Root Mean Square Error), and MAE (Mean Absolute Error).

## API Score Prediction Results

The Random Forest Regressor model was used to make the API score prediction. Following hyperparameter tuning and 10-fold cross-validation:

RMSE: 0.48

MAE: 0.37

R² Score: 0.89

This reflects a high predictive ability, with little error between predicted and actual API scores. Random Forest was used because it can efficiently work with both numerical and categorical data and is also resistant to overfitting.

## Career Path Classification Results

Decision Tree Classifier was employed to classify students into one of the three career paths: Job, Higher Studies, or Startup. The model provided the following metrics:

Accuracy: 86.3%

Precision: 84.7%

Recall: 85.1%

F1 Score: 84.9%

The high accuracy indicates that the model is successful in leveraging the API score and other characteristics (e.g., GPA, internships, certifications) to suggest a career path.

# CONCLUSION

This study proposes a novel and integrated method for student assessment through the creation of an Academic Performance Indicator (API) score by employing machine learning algorithms. Unlike the conventional system that is heavily dependent on scholastic marks, the novel system integrates various facets of the student's life—academics, co-curricular activities, internships, and online certifications—to produce one API score. The method is reflective of the changing demands of the job market as well as the education sector, which now seek well-rounded student profiles rather than individual scholastic brilliance.

In total, this project not only redefines student evaluation in a move towards a more equitable and realistic system of grading but also enables educational institutions to provide customized career guidance.

##### ACKNOWLEDGMENT

I would like to extend my sincerest appreciation to Dr. Aarathi S and our HOD Dr. Rashmi S whose brilliant advice, know-how, and motivation played a pivotal role in guiding this research. I am incredibly grateful to Dayananda Sagar College of Engineering for granting the required infrastructure and premises that made it possible to conduct this study smoothly.

##### REFERENCES

1. Abdallah Namoun and Abdullah Alshanqiti, “Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review”, 2020, <https://doi.org/10.3390/app11010237>
2. Kalpana R A, Karunya S and Rashmi R, “Student Academic Analyser and Career Guidance System Using Data Analytics and Visualization Techniques”, 2023, <https://doi.org/10.1109/ICCEBS58601.2023.10448522>
3. Kathleen M. Quinlan and K. Ann Renninger, “Rethinking employability: how students build on interest in a subject to plan a career”, 2022, <https://doi.org/10.1007/s10734-021-00804-6>
4. Shah Hussain and Muhammad Qasim Khan, “Student‑Performulator: Predicting Students’ Academic Performance at Secondary and Intermediate Level Using Machine Learning”, 2021, <https://doi.org/10.1007/s40745-021-00341-0>
5. Sakshi Vashisht, Poonam Kaushal and Ravi Vashisht, “Emotional intelligence, Personality Variables and Career Adaptability: A Systematic Review and Meta-analysis”, 2023, DOI: 10.1177/0972262921989877
6. Maria Ezziani, Salima Benhayoun and Yousra Chtouki, “Data Analysis of Student Academic Performance Using Social and Academic Indicators”, 2022, DOI: 10.1109/ITHET56107.2022.10031673
7. Nabila Kamal, Farhana Sarker, Arifur Rahman, Sazzad Hossain and Khondaker A. Mamun, “Recommender System in Academic Choices of Higher Education: A Systematic Review”, 2024, DOI: 10.1109/ACCESS.2024.3368058
8. Kathleen M. Quinlan and K. Ann Renninger, “Rethinking employability: how students build on interest in a subject to plan a career”, 2022, <https://doi.org/10.1007/s10734-021-00804-6>
9. Yifang Wang, Tai-Quan Peng, Huihua Lu, Haoren Wang, Xiao Xie and Huamin Qu, “Seek for Success: A Visualization Approach for Understanding the Dynamics of Academic Careers”, 2022, <https://doi.org/10.1109/TVCG.2021.3114790>
10. Edmund F. Agyemang, Joseph A. Mensah, Obu-Amoah Ampomah, Louis Agyekum and Justice Akuoko-Frimpong, “Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application”, 2024, DOI: 10.7176/CEIS/15-1-09