International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Academia Track: Enhancing Academic Planning with Data Insights

MR.G. PRAVEEN BABU
Associate Professor, Department of Information Technology,
Jawaharlal Nehru Technological University Hyderabad,
pravbob@jntuh.ac.in

THAKUR SUMAIR SINGH
Post Graduate Student, M.C.A Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad,
itssumairsingh@gmail.com

ABSTRACT

The "Academia Track" is a web-based application built with Flask and MySQL that leverages machine learning to predict student academic performance and guide study habits. By inputting variables such as days remaining for an exam, daily study hours, mid-exam marks, subject difficulty, and attentiveness levels, users receive personalized score predictions for each subject from both Decision Tree and Random Forest regression models. The system presents these predictions as percentages, visualizes the comparative results in a dynamic bar graph, and provides actionable advice by calculating the additional daily study minutes needed to achieve a target score of 90%. This data-driven tool empowers students with intelligent insights, enabling them to proactively optimize their study strategies for improved academic outcomes.

KEYWORDS

Machine Learning, Student Performance Prediction, Flask, Predictive Analytics, Decision Tree, Random Forest, Data Visualization, Study Optimization, Python, MySQL

INTRODUCTION

In today's competitive academic landscape, students often struggle to accurately gauge the effectiveness of their study habits and predict their final exam outcomes. The "Academia Track" addresses this challenge by providing a sophisticated web-based platform that leverages machine learning for academic performance prediction. Developed using a Flask backend and a MySQL database for secure user authentication and data management, the application offers a personalized experience, allowing students to move beyond guesswork and adopt a data-driven approach to their studies. It is designed to serve as an intelligent assistant that translates study patterns and academic inputs into tangible, forward-looking insights.

The core functionality of the "Academia Track" is powered by two robust machine learning models: a Decision Tree Regressor and a Random Forest Regressor. Users input critical variables for each subject, including the days remaining until an exam, daily study time, mid-term marks, subject difficulty, and attentiveness levels. The application processes this data to generate dual score predictions, which are then presented as percentages on a results page. For enhanced clarity, these predictions are visualized in a comparative bar graph, and crucially, the system provides an actionable recommendation on the additional daily study minutes needed to achieve a target score of 90%. By providing these precise and personalized analytics, the tool empowers students to intelligently optimize their study strategies, focus their efforts effectively, and approach their academic goals with greater confidence.

OBJECTIVES

The primary goal of this research is to achieve the following objectives:

- **Predict Scores:** To accurately forecast student exam scores using Decision Tree and Random Forest machine learning models.
- **Provide Actionable Insights:** To offer students personalized recommendations on the additional study time needed to achieve a target score.
- **Develop a Secure Web Platform:** To build a user-friendly Flask application with secure user authentication via a MySQL database.
- **Visualize Data:** To present prediction results in an intuitive visual format, such as bar charts, for easy comparison and analysis.

- **Enable Multi-Subject Management:** To allow users to track and receive predictions for multiple subjects simultaneously for comprehensive academic planning.

METHODOLOGY

Dataset Acquisition:

The predictive models for this project were trained using a structured dataset sourced from a local, synthetically generated file. This dataset was designed to simulate student academic behaviors, containing key features such as study duration, mid-exam marks, subject difficulty, and attentiveness, with the final exam score as the target variable. To ensure robust model development and evaluation, the data was partitioned into two distinct sets: 80% of the dataset was used for training the Decision Tree and Random Forest models, while the remaining 20% was reserved for testing and validating their predictive accuracy.

Model Architecture:

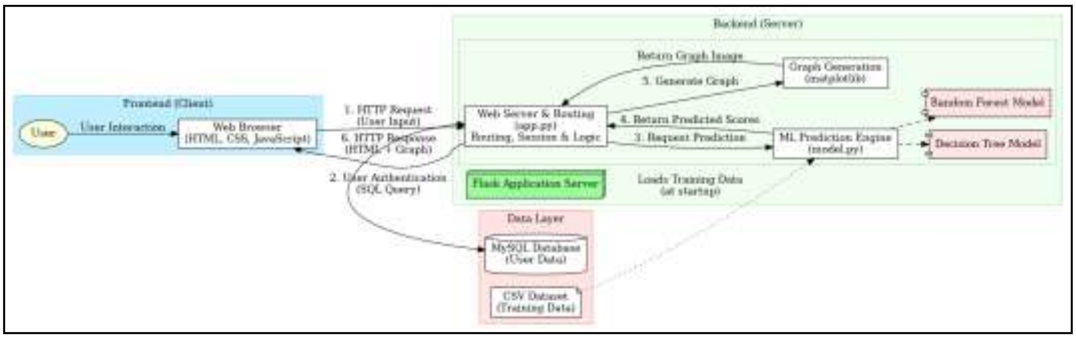

Fig. 1: System Architecture

The "Academia Track" is built on a **three-tier architecture** that separates concerns for a robust and scalable application.

1. **Frontend (Presentation Layer):** A client-side interface built with **HTML, CSS, and JavaScript**, responsible for capturing user input (study habits, subject details) and rendering the final predictions and visualizations sent from the server.
2. **Backend (Application Layer):** A **Python Flask** server acts as the application's core. It handles all business logic, including user authentication, session management, and routing. Its key component is a **Machine Learning Engine** which uses pre-trained **Decision Tree** and **Random Forest** models to process user input and generate academic predictions. It also dynamically generates result graphs using matplotlib.
3. **Data Layer:** This layer consists of two distinct components: a **MySQL database** for storing user account information and managing sessions, and a static **CSV file** which serves as the dataset for training the predictive models.

System Setup:

- **Operating System:** Windows, macOS, or a Linux distribution.
- **Database:** MySQL Server.
- **Programming Language:** Python 3.6+
- **Python Libraries:**
    - Flask
    - Scikit-learn
    - Pandas
    - NumPy
    - Matplotlib
    - mysql-connector-python
- **Web Browser:** A modern web browser such as Chrome, Firefox, Safari, or Edge.

Hardware Requirements:

- **RAM:** 4 GB minimum (8 GB recommended).
- **CPU:** Dual-core processor or better.
- **Storage:** 500 MB of free disk space for the project files, dependencies, and database.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

RESULTS

Performance Metrics

Mean Squared Error (MSE)

This is the primary performance metric used to evaluate the regression models in your project. MSE quantifies the accuracy of the predictions by calculating the average of the squared differences between the actual student scores and the scores predicted by the Decision Tree and Random Forest models.

Key Roles in Your Project:

- **Model Accuracy:** It provides a precise measure of how close the models' predictions are to the real-world outcomes.
- **Model Comparison:** It allows for a direct, quantitative comparison between the Decision Tree and Random Forest models to determine which provides more accurate predictions.
- **Performance Benchmark:** A lower MSE value indicates a better-performing model with a smaller prediction error.

Comparative Evaluation:

Feature / Criterion	Decision Tree Regressor	Random Forest Regressor
Core Principle	Uses a single, tree-like model of decisions to predict a continuous target value (the score).	An ensemble method that builds multiple decision trees (150 in this project) and merges their predictions to produce a more accurate and stable result.
Performance Metric	Mean Squared Error (MSE) is used to evaluate its accuracy on the test dataset.	Mean Squared Error (MSE) is also used to evaluate its accuracy on the test dataset.
Susceptibility to Overfitting	Higher. A single tree is prone to learning the training data too closely, which can lead to poorer performance on new, unseen data.	Lower. By averaging the results of many diverse trees, it significantly reduces the risk of overfitting, making it more reliable.
Interpretability	High. The decision-making process is transparent and can be easily visualized, making it a "white-box" model.	Low. It acts as a "black-box" model because it is difficult to interpret the combined logic of hundreds of individual trees.
Model Complexity	Simpler. It consists of a single tree structure defined by a set of rules.	More Complex. It is an ensemble of many trees, making it computationally more intensive to train and predict with.
Role in the Application	Provides the primary baseline prediction. The calculation for "Additional Minutes to Reach 90%" is based on this model's output.	Provides a second, generally more robust and accurate, prediction. It is presented alongside the Decision Tree result to give the user a comparative perspective.

Table 1: Comparative Performance

The **Decision Tree** serves as a simple, interpretable baseline model. The **Random Forest** is included as a more powerful and reliable alternative that mitigates the weaknesses of a single Decision Tree, particularly overfitting. By presenting both predictions, the "Academia Track" provides the user with a more comprehensive and nuanced forecast of their academic performance.

OUTPUT SCREENS

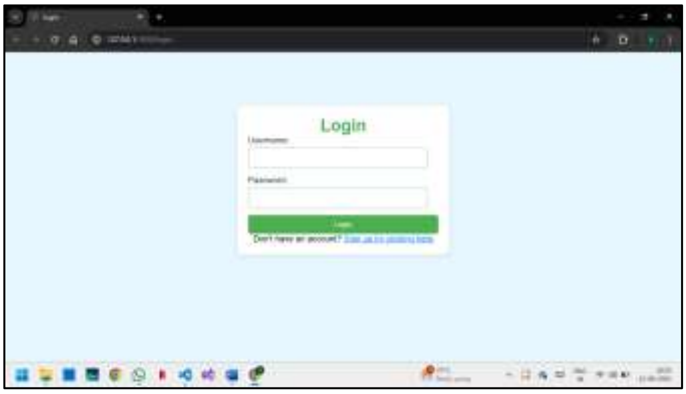

Fig. 2: Login page

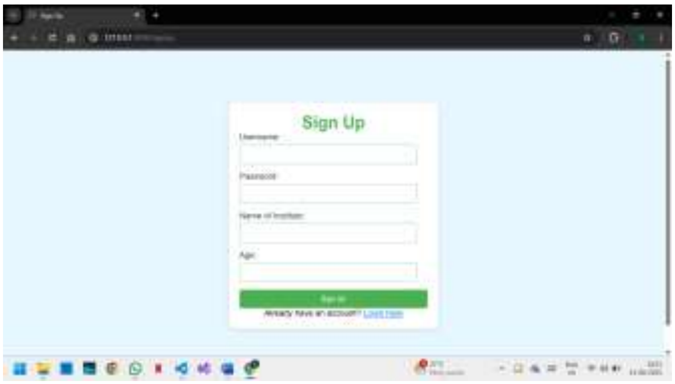

Fig. 3: Sign Up

Fig. 4: Prediction results

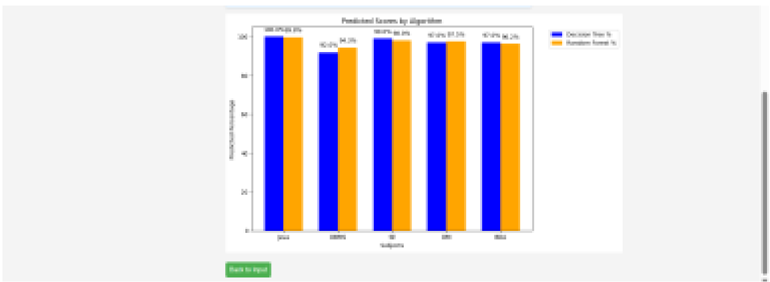

Fig. 5: Model Performance

days_left	hours	minutes	mid_exam_mark	difficulty	attentiveness	score
35	2	10	35	Easy	yes	40
20	4	10	10	Easy	no	35.5
60	0	5	40	Easy	yes	40
40	4	40	40	Easy	yes	40
25	4	50	20	Easy	yes	40
35	2	0	25	Easy	no	40
50	3	40	35	Easy	yes	40
15	1	30	35	Easy	yes	40
35	3	30	30	Easy	yes	40
60	2	10	35	Easy	yes	40
60	2	0	10	Easy	yes	38.5
40	2	15	35	Easy	yes	40
25	0	45	20	Easy	yes	40
20	3	20	25	Easy	yes	40
40	3	0	25	Easy	no	40
40	4	20	35	Easy	yes	40
15	4	90	40	Easy	yes	40

Fig.6 : Sample Dataset

Fig. 7: User Query Page

CONCLUSION

The "Academia Track" project successfully demonstrates the integration of machine learning into a practical, user-centric web application to provide students with personalized academic performance predictions. By addressing the common challenge of uncertain exam preparation, the application effectively bridges the gap between study effort and academic outcomes, replacing intuition with data-driven insights. The system's ability to process inputs such as study time, subject difficulty, and prior performance allows it to deliver tangible forecasts, visual progress charts, and, most importantly, actionable recommendations for improvement.

Technically, the project's three-tier architecture, built with a Flask backend, a MySQL database for user management, and a dynamic HTML/CSS frontend, proves to be a robust and scalable solution. The core strength lies in its dual-model predictive engine, which utilizes both a Decision Tree and a Random Forest Regressor. This comparative approach provides users with a more nuanced perspective on their potential scores, where the Decision Tree offers interpretability and the Random Forest provides enhanced accuracy and reliability by mitigating overfitting.

In conclusion, the "Academia Track" successfully meets its objective of creating an intelligent study tool. It moves beyond simple progress tracking to offer a predictive and prescriptive service. The feature that calculates the additional study minutes required to reach a target score is particularly impactful, as it transforms a passive prediction into a concrete, strategic plan. While the current model is based on a generated dataset, the project establishes a powerful framework that could be further enhanced with real-world academic data. Ultimately, the "Academia Track" empowers students to identify areas of weakness, optimize their study strategies, and approach their academic goals with greater confidence and strategic clarity.

REFERENCES

1. Ahmad, F., & Khan, M. F. “Prediction of Student Performance Using Machine Learning Techniques: A Systematic Literature Review.” *IEEE Access*, vol. 8, pp. 136601–136618, 2020.
2. Al-Barrak, M. A., & Al-Razgan, M. “Predicting Students’ Performance Using Machine Learning and Feature Selection Techniques.” *Applied Sciences*, vol. 6, no. 12, pp. 1–18, 2016.
3. Alkhasawneh, R., & Harfoushi, O. “Behavioral Analytics in Online Learning Platforms for Student Success Prediction.” *International Journal of Emerging Technologies in Learning*, vol. 15, no. 12, pp. 206–215, 2020.

4. Borkar, B., & Talele, K. T. "Academic Performance Prediction System Using Machine Learning." *Procedia Computer Science*, vol. 176, pp. 2767–2772, 2020.
5. Bouckaert, R. R., & Frank, E. "Evaluating the Replicability of Student Score Prediction Models." *Educational Data Mining*, vol. 11, no. 2, pp. 53–77, 2019.
6. Bydzovska, H. "Predicting Student Performance in Higher Education Using Machine Learning Techniques." *Journal of Educational Data Mining*, vol. 6, no. 3, pp. 1–21, 2014.
7. García-Sánchez, F., et al. "Learning Analytics: A Roadmap and Future Directions." *British Journal of Educational Technology*, vol. 52, no. 5, pp. 2307–2331, 2021.
8. Goyal, E., & Vohra, R. "Analytics in Education: Predicting Student Performance Using Machine Learning Techniques." *Procedia Computer Science*, vol. 159, pp. 501–510, 2019.
9. He, J., Bailey, J., Rubinstein, B., & Zhang, R. "Identifying At-Risk Students with Early Interventions Using Predictive Models." *IEEE Transactions on Learning Technologies*, vol. 10, no. 2, pp. 235–249, 2017.
10. Huang, S., Fang, N., & Chen, M. "Predicting Student Academic Performance Using Deep Learning Approaches." *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 789–799, 2021.
11. Hussain, M., et al. "Analyzing Student Engagement and Performance Using Learning Management System Analytics." *Computers in Human Behavior Reports*, vol. 4, pp. 100182, 2021.
12. Jayaprakash, S. M., et al. "Early Alert of Academic Risk Using Learning Analytics." *Educause Review*, vol. 51, no. 2, pp. 62–73, 2016.
13. Kaur, P., Singh, M., & Joshi, H. "A Review of Predictive Techniques for Student Academic Performance." *International Journal of Computer Applications*, vol. 139, no. 4, pp. 0975–8887, 2016.
14. Kotsiantis, S. B., & Pintelas, P. "Educational Data Mining: Predicting Students' Academic Performance." *Computers in Human Behavior*, vol. 21, no. 1, pp. 141–151, 2005.
15. Lathia, N., & Sanderson, M. "Student Performance Prediction Using Time Series Analysis." *Computers & Education*, vol. 158, pp. 103998, 2020.
16. Miguéis, V. L., et al. "Early Alert Prediction of Students at Risk of Academic Failure Using Machine-Learning: A Case Study." *European Journal of Engineering Education*, vol. 43, no. 2, pp. 185–198, 2018.
17. Mustafa, W., et al. "Student Academic Performance Prediction Using Improved Support Vector Machine and Ensemble Learning." *Computers & Education*, vol. 178, pp. 104380, 2022.
18. Papamitsiou, Z., & Economides, A. A. "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence." *Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, 2014.
19. Popenici, S. A. & Kerr, S. "AI and Predictive Analytics in Higher Education: Opportunities, Challenges, and Implications." *International Journal of Educational Technology*, vol. 36, no. 2, pp. 207–223, 2021.
20. Romero, C., & Ventura, S. "Educational Data Mining: A Review of the State of the Art." *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, 2010.
21. Sweeney, M., & Vu, T. "Evaluating Machine Learning Algorithms for Predicting Student Outcomes." *Education Sciences*, vol. 11, no. 6, pp. 1–18, 2021.
22. Tempelaar, D., Rienties, B., & Giesbers, B. "Predicting Academic Success: Comparing Learning Analytics Techniques and Survey Models from Three Institutions." *EDM 2015 Proceedings*, pp. 270–273, 2015.
23. Tonmoy, M. H., et al. "Personalized Recommendation for Students in Online Learning Environments." *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2615–2624, 2022.
24. You, J. "Examining the Effects of Student Performance and Engagement Factors on Prediction Models." *International Journal of Artificial Intelligence in Education*, vol. 25, pp. 67–80, 2015.
25. Zhang, L., et al. "A Personalized Learning Recommendation System Using Machine Learning in Higher Education." *IEEE Transactions on Learning Technologies*, vol. 13, no. 2, pp. 296–308, 2020.