International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

AI-Powered Student Performance & Career Guidance System

Shivam Sharma[1], Shivangi Sharma[1], Suchita Singh[1], Yash Lawaniya[1], Anshika Singh[2], Ayushi Sharma[2], Diwakar Shrivastava[2]

[1,2] Department of Computer Science & Engineering, Hindustan College of Science and Technology, Farah, Mathura, Uttar Pradesh, India

Corresponding author(s). E-mail(s): shivamsharma1197@gmail.com, shivangisharma1503@gmail.com, suchtasingh223@gmail.com, yashlawaniya2804@gmail.com

---***---

How to Cite this Article:

Sharma, S., Sharma, S., Singh, S., Lawaniya, Y., Singh, A., Sharma, A. & Shrivastava, D. (2026). AI-Powered Student Performance & Career Guidance System. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04). https://doi.org/10.55041/ijcope.v2i4.710

https://doi.org/10.55041/ijcope.v2i4.710

Abstract - Predicting student performance from lifestyle habits and delivering actionable career guidance remain critical yet largely disjoint challenges in engineering education. Existing approaches typically address these problems independently and often rely solely on academic indicators, lacking deployable, student-centric interfaces. This study presents a unified, end-to-end web platform that integrates performance prediction with intelligent career guidance. The system is trained on a 1,000-record Kaggle lifestyle dataset (ages 17-24) comprising 14 features, including sleep patterns, social media usage, dietary habits, and mental health indicators, along with an augmented dataset of 5,000 records. Multiple machine learning models were evaluated, including Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, XGBoost for classification, and Multiple Linear Regression (MLR), Ridge, and Lasso for regression, across three train-test splits (70:30, 80:20, and 90:10). Results demonstrate that Random Forest and XGBoost achieved perfect classification performance with 100% accuracy and ROC-AUC of 1.000 across all datasets and splits. For regression, MLR achieved an R2 of 0.8955 with an RMSE of 5.41 on the augmented dataset, outperforming comparable approaches without requiring complex feature engineering. The platform further incorporates a nine-agent GPT-4o-mini-based career intelligence pipeline, enhanced with live job-market insights via the Adzuna API and deterministic fallback mechanisms. The proposed system uniquely combines real-time predictive analytics, adaptive career recommendations, and a fully deployed three-tier architecture, making it a comprehensive solution for modern student support systems.

Keywords: Student Performance Prediction; Lifestyle Analytics; Wellness Score; Machine Learning; Multiple Linear Regression; Random Forest; XGBoost; Career Intelligence; Multi-Agent Systems; GPT-4o-mini; Adzuna API; scikit-learn; FastAPI; React; Educational Data Mining

1. INTRODUCTION

Academic performance prediction and structured career guidance remain two persistent and largely disconnected challenges in higher engineering education. While students routinely engage in activities such as managing sleep schedules, social media usage, dietary habits, and mental well-being, they often lack a clear understanding of how these lifestyle factors quantitatively influence their academic outcomes. Even when predictive insights are available, there is typically no integrated mechanism to translate these predictions into informed and actionable career decisions.

Over the past decade, machine learning approaches have been widely explored for student performance prediction. However, most existing systems exhibit one or more critical limitations. First, a significant proportion of studies rely exclusively on academic indicators such as grades, attendance, and prior scores, thereby neglecting lifestyle and behavioural factors that play a crucial role in student performance. Second, many approaches operate at an aggregate or institutional level, focusing on identifying at-risk student cohorts rather than providing real-time, individualised predictions. Third, a majority of these systems remain confined to offline experimental environments, such as Jupyter notebooks, without offering deployable, student-facing interfaces that enable practical usage. Similarly, existing career guidance systems fall short in several respects. Many lack integration with real-time labour market data, resulting in outdated or generic recommendations. Others depend heavily on large language model (LLM) APIs without incorporating fallback mechanisms, leading to reliability issues in real-world deployments. Furthermore, several systems require prior inputs such as detailed resumes, which limits their applicability for early-stage undergraduate students who may not yet possess formal professional profiles.

To address these gaps, this paper proposes a unified framework titled the AI-Powered Student Performance and Career Guidance System. The system integrates predictive analytics with intelligent career recommendation within a single platform. It accepts 14 lifestyle and academic input features and generates both a continuous wellness score and a three-class performance prediction. Building on these outputs, the system dynamically guides the student through an adaptive assessment process and a nine-agent decision pipeline. This pipeline incorporates real-time job market data via the Adzuna API, evaluates automation risk, constructs a personalised six-month learning roadmap, and produces a composite career suitability score. The proposed system is implemented as a fully functional three-tier web application using React for the frontend, Node.js for middleware orchestration, and FastAPI for machine learning services. The platform has been validated across multiple modern browsers, ensuring accessibility and usability in real-world scenarios.

The key contributions of this work lie in the development of a comprehensive and practically deployable framework that integrates machine learning-based performance prediction with intelligent career guidance. First, an extensive evaluation of four classification models and multiple regression techniques is conducted across two dataset sizes and three different train-test splits, resulting in consistently high performance, including perfect classification accuracy. Second, the study introduces three engineered composite features that effectively reduce the original 14-feature input space while preserving approximately 98% of the predictive capability, thereby improving model efficiency without significant loss of information. Third, a novel nine-agent directed acyclic graph (DAG) pipeline is designed, in which each agent operates with a deterministic fallback mechanism to ensure system robustness and reliability under varying conditions. Fourth, the system incorporates real-time job market intelligence through integration with the Adzuna API, enabling dynamic and context-aware career recommendations. Finally, the entire framework is implemented as a fully functional, scalable web application, validated across major browsers, providing real-time interaction and decision support for students in a user-friendly environment.

2. LITERATURE REVIEW

A total of twelve research papers published between 2023 and 2025 were systematically reviewed to evaluate the current state of student performance prediction and career guidance systems. To ensure fairness and accuracy in comparison, dataset characteristics, evaluation metrics, and reported results were verified directly from the original publications. A consolidated comparison is presented in Table 1.

2.1 Student Performance Prediction

Recent studies in student performance prediction demonstrate a strong emphasis on machine learning techniques, though with varying data sources, feature sets, and evaluation methodologies.

Yadav [1] utilised the same 1,000-record Kaggle dataset employed in this study, consisting of students aged 17-24 with features such as study hours, sleep patterns, social media usage, and mental health indicators. The study achieved its best performance using Ridge Regression with polynomial feature engineering, reporting an R2 value of 0.9015 and RMSE of 5.03. Additionally, K-Means clustering was applied to identify three distinct student groups. However, the work remained limited to offline analysis without any deployment or real-time prediction capability.

Chen et al. [2] proposed a novel RMBN (Relation Matrix-Based Network) data processing framework, which transforms student records into a bipartite relational structure and applies Louvain community detection for clustering. Classification was performed using Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). The dataset consisted solely of academic records from a Chinese university, excluding lifestyle variables. The study focused on clustering quality metrics such as modularity rather than predictive accuracy and did not include any deployment.

Ahmed [3] analysed Open University data from Ethiopia using features such as CGPA, demographic attributes, and course-related information. The study achieved a classification accuracy of 93% using Random Forest for binary pass/fail prediction and reported a minimum RMSE of 1.13 for regression tasks. However, the dataset was restricted to academic and demographic features, limiting its generalisability to lifestyle-based prediction.

W. Ahmed et al. [4] utilised a 10,000-record dataset with six features, including study hours, previous scores, and sleep duration. Their ensemble Voting Regressor model achieved a high R2 of 0.9883 with a very low RMSE of 0.1050. However, this low error is attributable to the normalised 0-1 scale of the target variable, making direct comparison with real-score datasets inappropriate. The study also incorporated explainability techniques such as SHAP and LIME but lacked deployment.

Wang and Yu [5] examined behavioural engagement data from 300 students enrolled in MOOC courses over a ten-week period. Using clickstream-derived indicators, they applied logistic regression to predict binary academic outcomes. The study focused primarily on correlation analysis rather than predictive performance metrics and is limited in applicability due to its reliance on online learning behaviour rather than general lifestyle factors.

2.2 Career Guidance and Skill-Gap Studies

Career guidance systems have evolved with the integration of artificial intelligence and natural language processing, yet several limitations persist.

Mohammad et al. [6] developed CareerMAS using multiple large language models, including LLaMA and Mistral, trained on over 5,500 job postings. However, the system lacked deployment and did not incorporate academic performance data.

Lai et al. [7] conducted a survey-based study involving 143 students to analyse perceived skill gaps using clustering techniques. While informative, the work was descriptive in nature and did not provide predictive or actionable outputs.

Chumwatana and Hpone [8] utilised natural language processing to extract in-demand skills from IT job postings, identifying trends such as cloud computing and DevOps. However, the approach lacked personalisation at the student level.

Babu et al. [9] proposed a system combining machine learning with the Gemini API for career recommendations. Although partially real-time, the system did not include fallback mechanisms, making it vulnerable to API failures.

Soni et al. [10] developed a hybrid course recommendation system that requires detailed resume inputs, limiting usability for early-stage students.

Kulugh et al. [11] adopted a Waterfall-based development approach using LLaMA, restricting iterative improvements and adaptability.

Spahic-Bogdanovic et al. [12] integrated Neo4j with Retrieval-Augmented Generation (RAG) and Claude 3 Sonnet for academic planning. While technically advanced, the system's scalability depends heavily on ontology quality and remains unvalidated at scale.

2.3 Comparison Between Existing Work and Proposed System

Table 1 presents a comprehensive comparison of all reviewed studies alongside the proposed system, highlighting key differences in datasets, methodologies, evaluation metrics, and deployment status.

While Yadav [1] achieved a slightly higher R2 value (0.9015) compared to our regression model, this was accomplished using polynomial feature engineering and without incorporating a classification framework or deployment. Similarly, W. Ahmed et al. [4] reported superior regression metrics; however, their results are based on a simplified dataset with a normalised target variable, rendering direct comparison inappropriate.

In contrast, the proposed system demonstrates strong performance across both classification and regression tasks on a more complex, multi-feature lifestyle dataset. More importantly, it uniquely integrates predictive modelling with real-time career intelligence within a fully deployed web platform. None of the reviewed studies simultaneously provide (i) multi-class performance prediction, (ii) lifestyle-based modelling, (iii) real-time job market integration, and (iv) a production-ready user interface.

Table 1: Comparison of All 12 Reviewed Papers vs. Our System

Author & Year	Dataset	Size	Features / Type	Best Model	Best Metric
Yadav (2025)	Same Kaggle dataset	1,000	16 lifestyle/habit	Ridge Reg. (poly.)	R2=0.9015, RMSE =5.03
Chen et al. (2023)	Chinese univ. academic	N/R	Academic GPA only	RMBN+ SVM/KNN	Modularity Q-score (no acc.)
Ahmed (2024)	Ethiopian univ. CGPA	N/R	Academic+ demog.	RF (binary)	Acc=93%, RMSE =1.13
W. Ahmed et al. (2025)	Kaggle 6-feat. norm.	10,000	6 features (0-1 scale)	Voting Reg. (ensem.)	R2=0.9883, RMSE =0.1050*
Wang & Yu (2025)	MOOC clicklog (China)	300	11 MOOC behaviours	Logistic Reg.	Correlation analysis (binary)
Mohammad et al. (2025)	Bangladesh job postings	5,500+	Job post text	Multi-agent LLM	No numerical metric
Lai et al. (2025)	Survey, 1 university	143	Self-reported survey	Clustering+k-NN	Descriptive only
Chumwatana & Hpone (2025)	IT job postings, scraped	N/R	Job post NLP	NLP+LLM mining	Skill frequency counts
Babu et al. (2025)	Single Indian univ.	N/R	Academic + API	RF + KNN + Gemini	Partial real-time placement
Soni et al. (2025)	Resume + course data	N/R	Resume-based	Hybrid collab. filt.	Recommendation quality
Kulugh et al. (2025)	N/R	N/R	User input text	LLaMA (Waterfall)	No numerical metric
Spahic et al. (2025)	ESCO ontology	N/R	Knowledge graph	Neo4j+RAG+Claude 3	No numerical metric
Our System (1000 RF)	Kaggle lifestyle	1,000	14 feat. + 3 eng.	Random Forest / XGB	Acc=100%, ROC-AUC=1.0000
Our System (5000 RF)	Augmented lifestyle	5,000	14 feat. + 3 eng.	Random Forest / LR	Acc=100%, ROC-AUC=1.0000
Our System (5000) MLR	Augmented lifestyle	5,000	14 feat.	MLR (all features)	R2=0.8955, RMSE =5.41

N/R denotes results that were not reported in the corresponding published paper. The RMSE value of 0.1050 reported by W. Ahmed et al. (2025) is calculated on a normalised Performance Index ranging from 0 to 1, whereas the RMSE of 5.41 in this study is based on a real-world examination score scale of 0 to 100. Due to this difference in target variable scaling, the two RMSE values are not directly comparable.

Similarly, the apparent difference in regression performance between Yadav (2025) and the Multiple Linear Regression (MLR) model presented in this work ($R2 = 0.9015$ versus $R2 = 0.8706$) can be attributed to methodological differences rather than inherent model superiority. Specifically, Yadav employed polynomial feature engineering on the same dataset, introducing higher-order feature interactions that naturally improve model fit. In contrast, the present study utilises standard linear regression without polynomial expansion to maintain model simplicity, interpretability, and deployment efficiency. Therefore, this work does not claim superiority in regression performance but instead emphasises a balanced trade-off between accuracy, generalisability, and real-world applicability within an integrated system.

2.4 Research Gap

A critical analysis of the reviewed literature reveals five consistent and significant research gaps. First, the majority of existing student performance prediction systems rely heavily on academic and institutional records, such as grades, attendance, and historical scores, while largely ignoring lifestyle and behavioural factors. This dependence makes such systems impractical for early-stage students, particularly those in their first semester, who lack sufficient academic history for meaningful prediction.

Second, despite promising experimental results, none of the reviewed studies provides a fully deployed, real-time system that enables individual students to interact with predictive models directly. Most implementations remain confined to offline environments, limiting their practical usability.

Third, there is a clear disconnect between performance prediction and career guidance in existing research. No system in the reviewed literature integrates both functionalities into a single unified platform, thereby missing the opportunity to translate predictive insights into actionable career decisions.

Fourth, current career guidance systems exhibit limitations in robustness and scalability. Many rely on expensive large language model APIs without incorporating fallback mechanisms, making them unreliable in production environments. Others are constrained by geographic, institutional, or domain-specific datasets, which restrict their general applicability.

Fifth, and most notably, none of the reviewed systems incorporates live external job market data through APIs. As a result, their recommendations are often static, outdated, or disconnected from real-world demand trends.

These gaps collectively highlight the need for an integrated, scalable, and real-time system that combines lifestyle-based performance prediction with dynamic, data-driven career guidance—an objective addressed by the proposed framework.

3. SYSTEM DESIGN AND METHODOLOGY

3.1 System Architecture

The proposed system is designed as a scalable three-tier web application architecture, consisting of a frontend presentation layer, a backend application layer, and a dedicated machine learning service layer. This modular design ensures separation of concerns, scalability, and ease of deployment. The frontend is implemented using the React framework, bundled with Vite for optimized performance. The user interface is styled using Tailwind CSS, while Recharts is used for data visualization and Framer Motion enhances interactivity through animations. The frontend provides multiple functional routes, including a student dashboard, adaptive quiz interface, deep analytics view, progress tracking module, session history viewer, and a real-time prediction interface. All routes are secured using JSON Web Token (JWT)-based authentication. Client-side state management is handled using Zustand, a lightweight flux-based store that ensures efficient state updates and minimal overhead.

The backend layer is developed using Node.js with the Express.js framework, and MongoDB for persistent storage. This layer manages user authentication, quiz generation and evaluation, orchestration of the nine-agent pipeline using a Directed Acyclic Graph (DAG), and storage of user sessions and progress data. Real-time communication between client and server, particularly for quiz progression and agent execution updates, is implemented using Server-Sent Events (SSE), enabling continuous streaming of results without requiring repeated client polling. The machine learning layer is implemented using FastAPI, which hosts trained scikit-learn pipeline models. This service exposes RESTful endpoints, including a /predict endpoint, which accepts input features and returns classification and regression outputs in real time. The separation of ML services into an independent microservice ensures flexibility in model updates and scalability in production environments.

3.2 Dataset

The primary dataset used in this study is the student_habits_performance dataset obtained from Kaggle. It contains 1,000 records representing undergraduate students aged 17-24, covering diverse lifestyle, behavioural, and academic attributes. The dataset includes 14 input features: age, gender, study hours per day, social media usage, Netflix consumption, part-time job status, attendance percentage, sleep duration, diet quality, exercise frequency, parental education level, internet quality, mental health rating, and extracurricular participation. The target variable, originally labelled exam_score (continuous range: 0-100, mean: 69.6, standard deviation: 16.9), is redefined in this study as wellness_score to reflect a broader interpretation of student performance influenced by lifestyle factors. It is important to note that the dataset is synthetic and not tied to a specific institution, which enhances generalisability but also requires cautious interpretation of extremely high predictive performance results. To improve robustness and enable comparative evaluation, an augmented dataset of 5,000 records was generated using statistical resampling techniques while preserving the original feature distributions. Data preprocessing included mode imputation for missing categorical values (e.g., parental education level), label encoding for ordinal features, one-hot encoding for nominal variables, and normalization of numerical features using StandardScaler. These preprocessing steps ensure consistency and optimal performance across different machine learning models.

3.3 Feature Engineering

To enhance predictive capability and reduce feature complexity, three composite features were engineered based on domain insights into student behaviour. The Productivity Ratio captures the balance between focused academic effort and digital distraction, defined as the ratio of study hours to the combined time spent on social media and streaming platforms. The addition of a constant term in the denominator prevents division by zero and stabilises the metric. The Health Score represents an aggregated measure of physical and mental well-being, computed as a weighted combination of sleep duration, exercise frequency, and mental health rating. The weights are assigned based on their relative importance in influencing cognitive performance and overall productivity.

The Distraction Score quantifies total time spent on non-academic digital activities, serving as a direct indicator of potential productivity loss. A reduced-feature model was also developed using a subset of six inputs: the three engineered features along with study hours, attendance percentage, and part-time job status. This reduced model retains approximately 98% of the predictive performance of the full feature set, demonstrating the effectiveness of feature engineering in simplifying model inputs without significant accuracy loss.

3.4 Machine Learning Pipeline

All machine learning models were implemented using pipeline constructs from scikit-learn to ensure consistent preprocessing, transformation, and prediction workflows. For classification tasks, four models were evaluated: Logistic Regression, K-Nearest Neighbors (k = 5), Random Forest (with 100 decision trees), and XGBoost. For regression analysis, Multiple Linear Regression, Ridge Regression, and Lasso Regression were implemented to predict continuous wellness scores. Model evaluation was conducted using three different train-test splits: 70:30, 80:20, and 90:10, to assess performance stability across varying training data sizes. The 80:20 split is reported as the primary configuration for consistency with standard evaluation practices. Performance metrics included accuracy and ROC-AUC for classification, and R2 and RMSE for regression.

3.5 Career Intelligence Module: Nine-Agent Pipeline

The career guidance component is implemented as a multi-stage intelligent pipeline consisting of nine specialised agents organised in a Directed Acyclic Graph (DAG) structure. The process begins with an adaptive quiz powered by GPT-4o-mini, comprising 20 questions distributed across five dimensions: technical knowledge, problem-solving ability, domain awareness, learning aptitude, and career clarity. The quiz dynamically adjusts its difficulty based on user responses using an acceleration mechanism, ensuring personalised assessment. To eliminate positional bias in answer options, the Fisher-Yates shuffle algorithm is applied.

Following the quiz, the pipeline executes in two stages. Stage 1 consists of a Skill Analysis Agent that evaluates user competencies. Stage 2 involves seven agents executed in four parallel waves: Wave 1 (Market Intelligence and Automation Risk), Wave 2 (Demand Modelling and Learning Path Generation), Wave 3 (Suitability Scoring and Resource Recommendation), and Wave 4 (Career Trajectory Prediction). While five agents utilise GPT-4o-mini for reasoning and content generation, two agents—Demand Modelling and Suitability Scoring—compute their outputs deterministically using weighted mathematical formulations based on market data and skill alignment metrics. GPT is used in these agents only for generating explanatory narratives. To ensure robustness, all agents are executed using asynchronous Promise.allSettled logic, and each LLM-dependent component includes a deterministic fallback mechanism that guarantees structured output even in the event of API failure. The final Composite Suitability Score is computed as a weighted aggregation of multiple factors, including skill alignment, job demand, automation safety, geographic distribution, and demand trends, enabling comprehensive and data-driven career recommendations.

3.6 Learning Path Tracking

The system incorporates a persistent learning path tracking mechanism to support continuous skill development. Recommended skills generated by the Learning Path

Agent are stored in MongoDB along with metadata such as status (e.g., to-do, in progress, completed) and session identifiers. Students can update their progress through a RESTful API endpoint (PATCH /api/v1/student/skill-status), enabling dynamic tracking of skill acquisition. The frontend visualises this data through progress dashboards, displaying metrics such as total skills assigned, number of completed tasks, and completion percentage. Additionally, session-wise performance trends are plotted to provide longitudinal insights into student improvement over time. This mechanism transforms the system from a one-time recommendation tool into a continuous learning support platform, enabling iterative improvement and long-term engagement.

4. RESULTS AND DISCUSSION

4.1 Regression Results

The regression performance of the proposed system is evaluated using Multiple Linear Regression (MLR) across different dataset sizes and feature configurations. The results for the 80:20 train-test split are presented in Table 2.

The findings indicate that on the augmented dataset of 5,000 records, the full-feature model achieves an R2 value of 0.8955 with an RMSE of 5.41, demonstrating strong predictive capability on a continuous 0-100 scale. In comparison, the reduced model with six selected features achieves an R2 of 0.8758, thereby retaining approximately 97.8% of the predictive performance while significantly reducing input dimensionality.

A closer inspection of Table 2 further reveals that the model exhibits consistent performance across both dataset sizes, with only marginal variation in error metrics. This consistency suggests that the model generalises well and is not sensitive to dataset scaling. Additionally, the small gap between training and testing performance indicates that overfitting is minimal.

Table 2: Regression Performance - Multiple Linear Regression (80:20 Split)

Configuration	Dataset Size	R2	MAE	RMSE	Train R2	Train RMSE
All 14 Features	1,000	0.8706	4.677	5.762	0.8757	6.039
All 14 Features	5,000	0.8955	4.269	5.410	0.8994	5.358
6 Engineered Features	1,000	0.8771	4.595	5.614	0.8845	5.807
6 Engineered Features	5,000	0.8758	4.774	5.898	0.8826	5.788

Note: All metrics are computed on the 0-100 wellness_score scale.

4.2 Classification Results

The classification performance is evaluated across multiple models, dataset sizes, and feature configurations. The results for the 1,000-record dataset with all features are shown in Table 3. It can be observed from Table 3 that ensemble-based models, particularly Random Forest and XGBoost, consistently achieve perfect classification performance across all train-test splits. Logistic Regression also demonstrates strong performance, suggesting that the class boundaries are largely linearly separable. In contrast, KNN exhibits slightly lower accuracy, especially in the 70:30 and 80:20 splits, due to its sensitivity to local data distribution.

Table 3: Classification Results - 1,000 Records (19 Features)

Split	Model	Accuracy	Precision	Recall	F1-Score
70:30	Logistic Regression	99.00%	99.33%	97.99%	98.63%
70:30	KNN (k=5)	98.00%	97.82%	97.94%	97.88%
70:30	Random Forest	100.00%	100.00%	100.00%	100.00%
70:30	XGBoost	100.00%	100.00%	100.00%	100.00%
80:20	Logistic Regression	95.00%	95.50%	93.59%	94.53%
80:20	KNN (k=5)	95.00%	96.19%	95.51%	95.85%
80:20	Random Forest	100.00%	100.00%	100.00%	100.00%
80:20	XGBoost	100.00%	100.00%	100.00%	100.00%
90:10	Logistic Regression	98.00%	97.62%	96.15%	96.87%

Split	Model	Accuracy	Precision	Recall	F1-Score
90:10	KNN (k=5)	100.00%	100.00%	100.00%	100.00%
90:10	Random Forest	100.00%	100.00%	100.00%	100.00%
90:10	XGBoost	100.00%	100.00%	100.00%	100.00%

To further analyse the impact of feature engineering, the same experiments were conducted by excluding engineered features. The results are presented in Table 4.

Interestingly, Table 4 shows that classification performance remains largely unaffected even after removing engineered features, indicating that the original features already contain strong predictive signals.

Table 4: Classification Results - 1,000 Records (16 Features)

Split	Model	Accuracy	Precision	Recall	F1-Score
70:30	Logistic Regression	99.00%	99.33%	97.99%	98.63%
70:30	KNN	97.67%	97.54%	97.72%	97.62%
80:20	Logistic Regression	99.50%	99.66%	98.72%	99.18%
80:20	KNN	99.50%	99.57%	99.66%	99.61%
90:10	Logistic Regression	98.00%	97.62%	96.15%	96.87%
90:10	KNN	100.00%	100.00%	100.00%	100.00%

The experiments were then extended to the augmented 5,000-record dataset. The results with all features are shown in Table 5.

As observed in Table 5, model performance improves slightly with increased dataset size, particularly for Logistic Regression and KNN. However, Random Forest maintains perfect performance across all splits, demonstrating robustness to dataset scaling.

Table 5: Classification Results - 5,000 Records (19 Features)

Split	Model	Accuracy	Precision	Recall	F1-Score
70:30	Logistic Reg.	99.93%	99.94%	99.95%	99.95%
70:30	KNN	99.00%	99.13%	98.87%	98.99%
70:30	Random Forest	100.00%	100.00%	100.00%	100.00%
80:20	Logistic Reg.	100.00%	100.00%	100.00%	100.00%
80:20	KNN	98.60%	98.54%	98.54%	98.54%
80:20	Random Forest	100.00%	100.00%	100.00%	100.00%
90:10	Logistic Reg.	100.00%	100.00%	100.00%	100.00%
90:10	KNN	99.40%	99.39%	99.33%	99.36%
90:10	Random Forest	100.00%	100.00%	100.00%	100.00%

4.3 Feature Comparison

To summarise the effect of feature selection, Table 7 compares the best-performing model (Random Forest) across all configurations.

Table 7: Random Forest Performance Across Configurations (80:20 Split)

Configuration	Dataset	Features Used	Accuracy	Macro F1
All Features	1,000	19	100.00%	1.0000
Reduced Features	1,000	16	100.00%	1.0000
All Features	5,000	19	100.00%	1.0000
Reduced Features	5,000	16	100.00%	1.0000

These results confirm that classification performance is invariant to feature reduction, which has important implications for deployment efficiency and data collection simplicity.

4.4 Comparison with Prior Work

To position the proposed system within existing research, Tables 8 and 9 provide a direct comparison with prior studies.

Table 8: Regression Comparison with Existing Studies

Study	Dataset	Scale	Model	R2	RMSE
Yadav (2025)	Same	0-100	Ridge	0.9015	5.03
W. Ahmed et al.	Kaggle	0-1	Voting	0.9883	0.105
Our System (5K)	Same	0-100	MLR	0.8955	5.41

The comparison shows that while some prior studies report higher R2 values, these improvements are often due to feature transformations or differences in target scale. In contrast, the proposed system prioritises interpretability and deployability.

Table 9: Classification Comparison with Existing Studies

Study	Task	Model	Accuracy
Ahmed (2024)	Binary	RF	93%
Our System	3-class	RF	100%

This highlights that the proposed system not only achieves superior accuracy but also addresses a more complex multi-class classification problem.

4.5 Feature Importance

Feature importance analysis provides insight into model decision-making. The results are presented in Table 10.

Table 10: Feature Importance (Random Forest, 5K Dataset)

Rank	Feature	Importance
1	wellness_score	0.5813
2	study_hours	0.1633
3	productivity_ratio	0.0557
4	health_score	0.0388
5	distraction_score	0.0360

As shown in Table 10, both behavioural and engineered features contribute significantly, validating the hybrid feature design approach.

4.6 Career Intelligence Module Results

The effectiveness of the career intelligence module is demonstrated through representative outputs shown in Tables 11 and 12.

Table 11: Sample Quiz Output

Dimension	Score
Technical Knowledge	75%
Problem Solving	50%
Learning Aptitude	100%
Overall	70%

The quiz results form the input to the nine-agent pipeline, whose execution summary is presented in Table 12.

Table 12: Pipeline Execution Summary

Stage	Output Example	Time
Skill Analysis	Backend Dev: 77%	<3s
Market Data	450 jobs	3-6s
Suitability	74/100	<2s

These results demonstrate that the system can generate meaningful, real-time career recommendations.

4.7 System Performance Metrics

Finally, system-level performance is evaluated and summarised in Table 13.

Table 13: System Performance Metrics

Metric	Value
Page Load Time	< 2 sec
SSE Latency	< 200 ms
Pipeline Time	15-35 sec
Compatibility	Chrome, Firefox, Edge

The results confirm that the system is efficient, responsive, and suitable for real-world deployment.

5. CONCLUSION

This study presents a unified AI-powered system for student performance prediction and intelligent career guidance, addressing two traditionally separate challenges within a single, deployable framework. The proposed system leverages 14 lifestyle and academic features to predict both a continuous wellness score and a three-class performance category, and subsequently integrates these predictions into a nine-agent career intelligence pipeline.

The experimental results demonstrate strong and consistent model performance. Random Forest and XGBoost classifiers achieved perfect classification accuracy and ROC-AUC across all dataset sizes and train-test splits, indicating clear separability of performance categories within the feature space. In regression, the Multiple Linear Regression model achieved an R2 value of 0.8955 on the augmented dataset, demonstrating reliable predictive capability while maintaining interpretability.

A key contribution of this work lies beyond predictive accuracy. The system successfully integrates machine learning models with real-time career guidance using live job market data, adaptive assessment, and personalised learning path generation. Unlike prior studies, the proposed framework is fully implemented as a scalable web application, enabling real-time, individual-level interaction. Additionally, the engineered feature set demonstrates that reduced input complexity can retain nearly the same predictive performance, supporting efficient deployment.

Overall, the proposed system bridges the gap between predictive analytics and actionable decision-making, offering a practical solution for enhancing student awareness, performance, and career readiness.

6. LIMITATIONS AND FUTURE WORK

Despite its contributions, the proposed system has certain limitations that must be acknowledged. First, the dataset used in this study is synthetic and sourced from Kaggle rather than collected from real institutional environments. While this improves generalisability across contexts, it may not fully capture the variability, noise, and complexity present in real-world student data. As a result, the observed high classification performance should be interpreted with caution.

Second, the classification labels are derived from the continuous wellness score, which inherently simplifies the classification task and contributes to near-perfect separability. In real-world scenarios, performance categories may not follow such clean boundaries.

Third, the current implementation does not include advanced model interpretability techniques such as SHAP or LIME. Although linear regression provides some level of interpretability, ensemble models used for classification remain relatively opaque to end users.

Fourth, the career intelligence module relies partially on external APIs and large language models. While deterministic fallback mechanisms ensure robustness, the quality and depth of recommendations may vary depending on external data availability and API performance.

Future work will focus on addressing these limitations and extending the system in several directions. A primary objective is the integration of real institutional datasets to validate model performance under realistic conditions and improve generalisability. Additionally, incorporating explainability techniques such as SHAP and LIME will enhance transparency and user trust by providing feature-level insights into predictions.

From a system perspective, the platform will be extended to mobile environments using cross-platform frameworks, enabling wider accessibility. Furthermore, the career module will be enhanced with continuous job market monitoring through scheduled API polling, allowing dynamic updates and proactive career recommendations. Finally, future research may explore hybrid models that balance interpretability and non-linearity, further improving predictive performance without compromising usability.

REFERENCES

[1] A. Yadav, "Predicting student academic performance: A machine learning analysis of study habits and lifestyle factors," Int. J. Adv. Res., vol. 13, no. 9, pp. 244-252, Sep. 2025.

[2] Z. Chen, G. Cen, Y. Wei, and Z. Li, "Student performance prediction approach based on educational data mining," IEEE Access, vol. 11, pp. 131260-131272, 2023.

[3] E. Ahmed, "Student performance prediction using machine learning algorithms," Appl. Comput. Intell. Soft Comput., vol. 2024, Art. no. 4067721, 2024.

[4] W. Ahmed et al., "Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions," Sci. Rep., vol. 15, Art. no. 26879, Jul. 2025.

[5] J. Wang and Y. Yu, "Machine learning approach to student performance prediction of online learning," PLOS ONE, vol. 20, no. 1, Art. no. e0299018, Jan. 2025.

[6] S. Mohammad et al., "CareerMAS: A multi-agent LLM framework for career guidance and job placement," in Proc. Int. Conf. AI Appl., 2025.

[7] J. W. Lai et al., "Learning analytics for bridging the skills gap: A data-driven study of undergraduate aspirations and skills awareness," J. Learn. Anal., 2025.

[8] T. Chumwatana and A. K. K. Hpone, "Bridging the IT skill gap with industry demands: An AI-driven text mining approach," in Proc. Int. Conf. AI Data Min., 2025.

[9] S. Babu et al., "Career Compass: AI-driven placement prediction and personalized career development," Int. J. Comput. Appl., 2025.

[10] P. Soni et al., "Improving graduate outcomes by identifying skills gaps and recommending courses based on career interests," J. Educ. Technol. Syst., 2025.

[11] N. Kulugh et al., "Artificial intelligence-powered personalised career guidance system," Int. J. Artif. Intell. Educ., 2025.

[12] M. Spahic-Bogdanovic et al., "iModuleBuddy: A hybrid AI-based academic planning system," in Proc. Conf. Intell. Educ. Syst., 2025.

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York, NY, USA: Springer, 2009.

[14] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[15] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.

[16] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.

[17] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189-1232, 2001.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD), 2016, pp. 785-794.

[19] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825-2830, 2011.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016, pp. 1135-1144.

[22] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans. Syst., Man, Cybern. C, vol. 40, no. 6, pp. 601-618, Nov. 2010.

[23] R. S. J. d. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," J. Educ. Data Min., vol. 1, no. 1, pp. 3-17, 2009.

[24] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in Proc. ACM SIGIR, 1999, pp. 230-237.

[25] X. Amatriain and J. Basilico, "Recommender systems in industry: A Netflix case study," in Recommender Systems Handbook, Springer, 2015.

[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993-1022, 2003.

[27] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[28] OpenAI, "GPT-4 technical report," 2023.

[29] Adzuna, "Adzuna job search API documentation," 2025. [Online]. Available: https://developer.adzuna.com

[30] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2021.