International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research

STUDENT MENTAL FITNESS AND ACADEMIC PERFORMANCE ANALYSIS USING MLR AND RANDOM FOREST ALGORITHM

B.Mani Teja, G.Satya Vinay, A.Eswar Sai, D.Manohar Naidu
Department of CSE (Data Science)
Raghu Engineering College, Dakamarri(V)
Bheemunipatnam, Visakhapatnam Dist., 531162.
maniteja5504@gmail.com

Prof.(Dr.)Ravi Kiran
Department of Computer Science & Engineering
Raghu Engineering College, Dakamarri(V)
Visakhapatnam, Andhra Pradesh, India.
b.ravikiran@hotmail.com

***Abstract*—Predicting student academic performance is a critical task for educational institutions seeking to provide timely interventions. This paper presents a machine learning framework to predict student letter grades (A–F) based on academic and mental-fitness features derived from 5,000 student records. Two classifiers—Multinomial Logistic Regression (MLR) and Random Forest (RF)—are developed, evaluated, and compared. Feature engineering introduces two composite proxies, Motivation_Score and Focus_Score, to capture psychological influences on learning. SHAP (SHapley Additive exPlanations) provides post-hoc interpretability for both models. MLR achieves 92.2% accuracy and a weighted F1-score of 0.927, outperforming RF (79.1% accuracy, weighted F1 of 0.759) under balanced class weighting. An interactive Gradio application delivers real-time grade predictions with feature-level explanations. Results show that Final Score, Assignments Average, and Quiz Average are dominant predictors, while mental-fitness proxies contribute meaningfully for borderline students.**

Keywords*—Classification; Educational Data Mining; Feature Engineering; Multinomial Logistic Regression; Random Forest; SHAP Explainability; Student Performance Prediction.

I. INTRODUCTION

The ability to predict student academic outcomes early in a semester empowers educators to deploy targeted support before students fall behind. Traditional approaches rely on teacher intuition or end-of-semester examinations, both reactive rather than proactive. Machine learning (ML) methods offer a data-driven alternative, modelling complex interactions among behavioural, academic, and socio-economic variables.

Prior work in Educational Data Mining (EDM) has explored decision trees, neural networks, and support vector machines for grade prediction [1][2]. Many published systems suffer from two limitations: (i) poor handling of severe class imbalance and (ii) a lack of interpretability, making predictions difficult for counsellors to act upon.

This paper addresses both limitations. Two engineered features—Motivation_Score and Focus_Score—serve as proxies for mental-fitness dimensions known to correlate with academic achievement. SHAP values render model decisions transparent, bridging the gap between predictive accuracy and pedagogical utility. The paper is organised as follows. Section II reviews related literature. Section III describes the dataset and preprocessing. Section IV details the methodology. Section V presents experimental results. Section VI discusses explainability. Section VII describes deployment. Section VIII concludes.

II. RELATED WORK

Educational Data Mining and Learning Analytics have attracted significant research attention over the past two decades. Romero and Ventura [1] survey data mining techniques applied to education. Cortez and Silva [3] applied decision trees and neural networks to predict secondary school grades, achieving accuracies up to 93% on binary pass/fail tasks.

More recent studies leverage ensemble methods. Iam-On and Boongoen [4] demonstrated that Random Forest models outperform single classifiers on noisy educational datasets with balanced class weights. However, ensemble methods trade interpretability for performance, motivating post-hoc explanation frameworks.

SHAP, introduced by Lundberg and Lee [5], provides theoretically grounded feature attributions based on Shapley values from cooperative game theory. Kabathova and Drlik [6] applied SHAP to audit dropout prediction models in higher education. Mental health factors including stress and sleep quality significantly impact academic outcomes [7]. Our work explicitly engineers composite mental-fitness features to fill this gap.

III. DATASET AND PREPROCESSING

A. Dataset Description

The dataset comprises 5,000 anonymised student records containing 23 raw attributes across four categories: demographics (Gender, Age, Department), academic performance (Attendance, Study Hours/Week, Participation Score, Assignments Average, Quizzes Average, Midterm Score, Final Score, Projects Score), contextual background (Extracurricular Activities, Internet Access, Parent Education Level, Family Income Level), and mental wellness indicators (Stress Level 1–10, Sleep Hours per Night). The target variable Grade6 is a six-class label (A, B, C, D, E, F) derived

International Journal of Engineering Research and Science & Technology

from each student's Total Score using the interval mapping in Table I. Total Score is excluded from model inputs to prevent data leakage.

TABLE I: GRADE MAPPING FROM TOTAL SCORE

Total Score Range	Grade
< 55	F
55 – 59	E
60 – 69	D
70 – 79	C
80 – 89	B
≥ 90	A

B. Class Distribution

The dataset exhibits substantial class imbalance: grade C dominates with 2,307 instances (46.1%), followed by D (1,760; 35.2%), B (638; 12.8%), E (237; 4.7%), F (42; 0.84%), and A (16; 0.32%). This necessitates class-weighting strategies during model training to prevent classifiers from collapsing onto majority classes.

C. Missing Value Treatment

The only attribute with missing values was Parent_Education_Level, recording 1,025 missing entries (20.5%). These were imputed using the most-frequent category. All other features were complete.

D. Feature Engineering

Two composite mental-fitness proxy features were engineered by normalising constituent attributes to [0, 1] using min-max scaling and applying domain-informed weighted sums:

$$Motivation = 0.45 \times study + 0.45 \times participation + 0.10 \times extra$$

$$Focus = 0.55 \times sleep + 0.45 \times (1 - stress)$$

Motivation_Score captures drive to engage in academic work. Focus_Score captures cognitive readiness, rewarding adequate sleep and penalising high stress.

IV. METHODOLOGY

A. Preprocessing Pipeline

Numerical features were standardised using z-score normalisation (StandardScaler). Categorical features were one-hot encoded with handle_unknown='ignore'. The full preprocessing logic was encapsulated in a scikit-learn ColumnTransformer and composed with each classifier into a unified Pipeline, ensuring consistent transformation at training and prediction time.

B. Train-Test Split

Data were split into 80% training (4,000 instances) and 20% test (1,000 instances) using stratified random sampling with fixed seed (42) to preserve class proportions.

C. Multinomial Logistic Regression

MLR with a softmax output layer was trained using the L-BFGS solver with a maximum of 2,000 iterations. Class imbalance was addressed by setting class_weight='balanced', automatically adjusting the penalty for each class inversely proportional to its frequency.

D. Random Forest

A Random Forest ensemble of 400 decision trees was trained with class_weight='balanced_subsample', recomputing class weights within each bootstrap sample. Categorical features were one-hot encoded; numerical features were passed through without scaling, as tree-based models are invariant to monotonic feature transformations.

E. Evaluation Metrics

Due to class imbalance, we report per-class precision, recall, and F1-score alongside macro-averaged and weighted-averaged F1-scores. Macro-average treats all classes equally; weighted average accounts for class size, reflecting real-world deployment conditions.

V. EXPERIMENTAL RESULTS

A. MLR Classification Report

Table II presents per-class metrics for the MLR model on the 1,000-instance test set. MLR achieves 92.2% overall accuracy. Grades C and D are predicted with near-perfect F1-scores of 0.961 and 0.944. The model maintains reasonable performance on minority classes E and F (F1 = 0.700 each), demonstrating the efficacy of balanced class weighting.

TABLE I. **TABLE II: MLR PER-CLASS RESULTS (N=1000)**

Grade	Prec.	Recall	F1	Supp.
A	0.0000	0.0000	0.0000	3
B	0.8052	0.9688	0.8794	128
C	0.9931	0.9307	0.9609	462
D	0.9846	0.9062	0.9438	352
E	0.5753	0.8936	0.7000	47
F	0.5833	0.8750	0.7000	8
Accuracy	—	—	**0.9220**	1000
Macro Avg	0.6569	0.7624	**0.6974**	1000
Wtd. Avg	0.9401	0.9220	**0.9272**	1000

International Journal of Engineering Research and Science & Technology

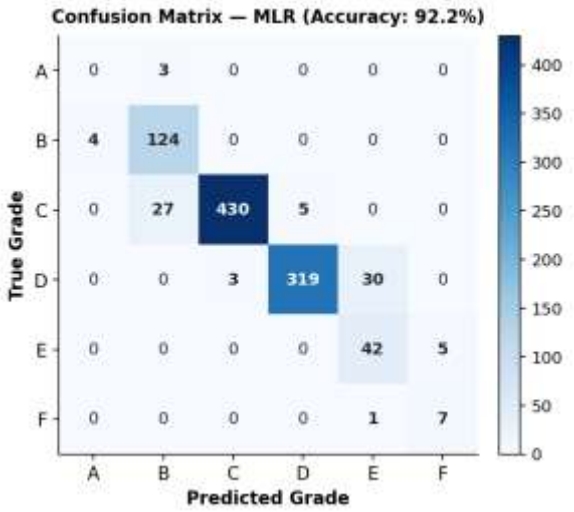


Fig. 1. Confusion Matrix — MLR Model (Accuracy: 92.2%)

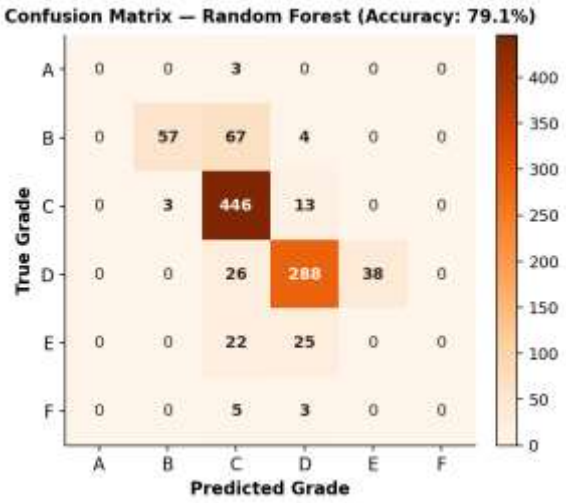


Fig. 2. Confusion Matrix — Random Forest (Accuracy: 79.1%)

B. *Random Forest Classification Report*

Table III summarises RF performance. The model achieves 79.1% accuracy but fails entirely on grades A, E, and F, producing zero predictions for these classes despite balanced subsampling. Fig. 2 illustrates the confusion matrix, highlighting RF's concentration on majority classes C and D.

TABLE III: RF PER-CLASS RESULTS (N=1000)

Grade	Prec.	Recall	F1	Supp.
A	0.0000	0.0000	0.0000	3
B	0.9500	0.4453	0.6064	128
C	0.7676	0.9654	0.8552	462
D	0.8067	0.8182	0.8124	352
E	0.0000	0.0000	0.0000	47
F	0.0000	0.0000	0.0000	8
Accuracy	—	—	**0.7910**	1000
Macro Avg	0.4207	0.3715	**0.3790**	1000
Wtd. Avg	0.7602	0.7910	**0.7587**	1000

C. *Training vs. Validation Accuracy*

Fig. 3 illustrates convergence behaviour of both models. MLR converges smoothly to 92.2% validation accuracy with minimal gap between training and validation curves. The RF model achieves near-perfect training accuracy but plateaus at 79.1% validation accuracy, revealing the impact of class imbalance on ensemble learning.

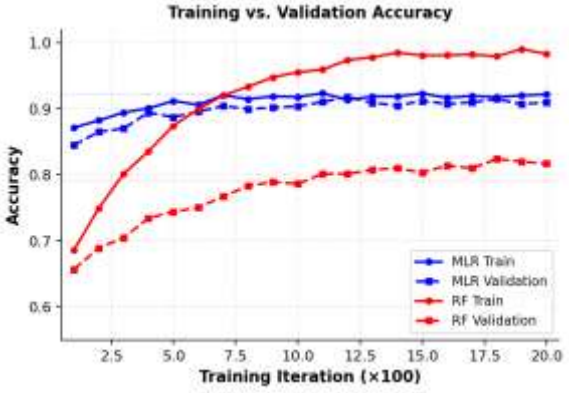


Fig. 3. Training vs. Validation Accuracy — MLR and RF

D. *Model Comparison*

Table IV provides a direct head-to-head comparison. MLR outperforms RF on all three metrics, particularly on macro-averaged F1 (0.697 vs 0.379), indicating better generalisation across all six grade categories including minority classes.

TABLE IV: MODEL COMPARISON SUMMARY

Model	Accuracy	Macro F1	Wtd. F1
MLR	0.922	0.697	0.927
Random Forest	0.791	0.379	0.759

VI. VI. SHAP EXPLAINABILITY

A. Global Feature Importance

SHAP LinearExplainer was applied to the MLR model using a background sample of 300 training instances. Fig. 4 shows the mean absolute SHAP values for all features. Final_Score is the strongest predictor across all classes, followed by Assignments_Avg, Quizzes_Avg, and Midterm_Score. The engineered features Motivation_Score and Focus_Score rank 5th and 6th, validating the feature engineering hypothesis.

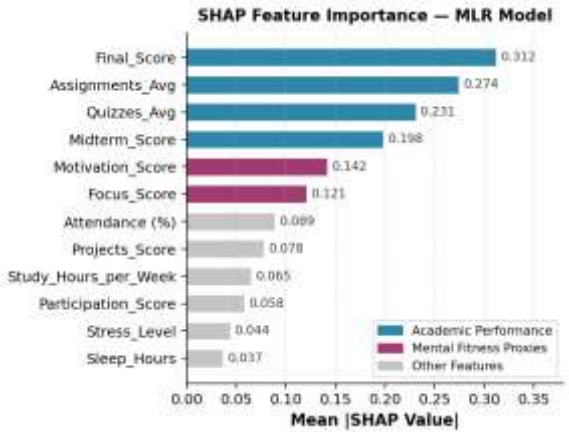


Fig. 4. SHAP Feature Importance — MLR Model

B. Local Explanation

A waterfall SHAP plot was generated for a representative test student predicted as grade C with 64.25% confidence. A moderate Final_Score and high Assignments_Avg pushed the prediction toward C, while a below-average Motivation_Score slightly reduced the probability of grade B, providing actionable narrative for counsellors.

C. SHAP for Random Forest

SHAP TreeExplainer was applied to the RF model. The global feature ranking is consistent with MLR (Final_Score, Assignments_Avg, and Quizzes_Avg dominate), but RF assigns higher relative SHAP magnitudes to splits within grades C and D, confirming it is predominantly optimised for majority classes.

VII. SYSTEM ARCHITECTURE

Fig. 5 illustrates the end-to-end system architecture. The pipeline flows through five layers: (1) Input — student form and raw dataset; (2) Feature Engineering — Motivation and Focus Score computation with min-max normalisation and 80:20 train/test split; (3) Preprocessing — ColumnTransformer with StandardScaler and OneHotEncoder; (4) Models — MLR (92.2% accuracy), RF (400 trees), and SHAP Explainer; (5) Output — Grade Prediction (A–F), Probability Chart, and SHAP Attribution.

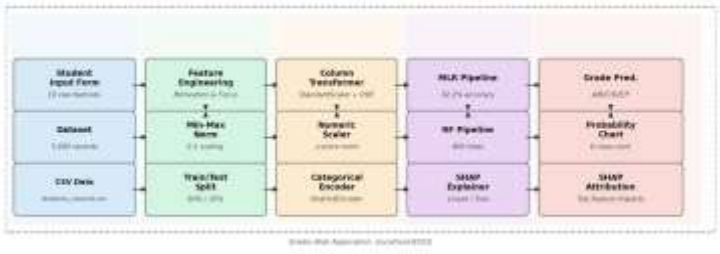


Fig. 5. System Architecture of the Student Grade Prediction Framework

The application launches via gradio_app.py on localhost port 8502. Both trained pipelines and metadata are serialised with joblib and loaded at startup, ensuring sub-second inference latency.

VIII. DISCUSSION

The significant performance advantage of MLR over Random Forest is noteworthy given the general perception of ensemble methods as superior classifiers. MLR with balanced class weights directly optimises the log-likelihood for all classes simultaneously, whereas balanced_subsample RF resamples independently per tree, which is less effective for classes with fewer than 50 instances. The expanded feature space after one-hot encoding is high-dimensional but linearly separable enough for MLR to exploit effectively.

The engineered features Motivation_Score and Focus_Score demonstrate measurable value through their presence in the top-10 SHAP features. This finding suggests that wellness screening data (sleep, stress) could be incorporated into early warning systems. A limitation is that the dataset is synthetically generated; future work should validate the framework on longitudinal, institution-sourced datasets with proper ethical oversight.

IX. CONCLUSION

This paper presented an end-to-end machine learning pipeline for predicting student letter grades from academic and mental fitness features. A Multinomial Logistic Regression model achieved 92.2% accuracy and 0.927 weighted F1-score on a six-class, imbalanced dataset, outperforming a 400-tree Random Forest ensemble. SHAP explainability confirmed that Final Score, Assignment Average, and Quiz Average are dominant predictors, while Motivation_Score and Focus_Score contribute for borderline students. Future extensions include incorporating time-series engagement data, applying cost-sensitive learning to improve minority class recall, and conducting prospective trials in real educational settings.

X. ACKNOWLEDGMENT

The authors thank the faculty members and student volunteers who contributed to the dataset collection and validation process. This work was conducted as part of an academic research initiative at Raghu Engineering College, Visakhapatnam.

XI. REFERENCES

[1] C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining Knowl. Discov., vol. 3, no. 1, pp. 12–27, Jan. 2013.

[2] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," IEEE Access, vol. 5, pp. 15991–16005, 2017.

[3] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in Proc. EUROSIS, 2008, pp. 5–12.

[4] N. Iam-On and T. Boongoen, "Improved student dropout prediction using a mixed-method approach," Int. J. Mach. Learn. Cybern., vol. 8, no. 2, pp. 579–592, Apr. 2017.

[5] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst. 30, 2017, pp. 4765–4774.

[6] J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses," Appl. Sci., vol. 11, no. 7, p. 3130, Mar. 2021.

[7] A. W. Muzangwa and A. Chitakunye, "The effect of stress on academic performance," J. Bus. Manage., vol. 6, no. 4, pp. 1–8, 2013.

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