# A Dual-Objective Machine Learning Framework for Predicting Academic Success and Social Media Impact on Students

**Mr. Amit Vilasrao Tale[1], Dr. Bhaskar Vijayrao Patil[2]**
[1]MAEER's MIT Arts, Commerce and Science College, Pune, India
avtale@mitacsc.ac.in
[2]Bharati Vidyapeeth (Deemed to be University), Institute of Management, Kolhapur, India.
bhaskar.patil@bharatividyapeeth.edu

## Abstract

Social media is increasingly embedded in students' daily lives, influencing their attention, emotional regulation, and academic focus. Yet, most research examines academic success, social media use, and mental health separately. This study proposes a dual-objective machine learning framework that simultaneously predicts academic success and negative social media impact among higher education students. Two datasets from an institution were used: pre-admission records comprising demographic details, interview scores, and subject marks, and a questionnaire assessing social media interference with sleep, mood, and stress. Binary labels were defined for academic success (Successful/Not Successful) and social media impact (Negative Impact/No Impact). After label encoding and median imputation, five classifiers—Random Forest, Support Vector Machine, Multi-layer Perceptron, Gradient Boosting, and a soft-voting ensemble—were trained using an 80/20 split. Gradient Boosting achieved perfect test accuracy in both predictive tasks, followed by the ensemble and Random Forest models with similarly high accuracy. The Support Vector Machine and neural network performed moderately but showed weak detection of minority outcomes. The study underscores the effectiveness of gradient-boosted ensembles for educational data mining and demonstrates how integrating behavioral and mental-health indicators with academic data can enhance AI-enabled systems for student performance analytics and institutional decision-making.

**Keywords:** academic performance; social media impact; educational data mining; machine learning; gradient boosting; student risk prediction

## 1. Introduction

Social media platforms like WhatsApp, Instagram, YouTube, and other AI-enhanced tools have become central to the daily lives of higher education students. These platforms facilitate rapid communication, informal peer learning, and access to vast educational resources. However, they also present challenges, including fragmented attention, reduced study time, sleep disruption, and increased anxiety, positioning social media as both an opportunity and a potential source of risk for educational institutions. Concurrently, educational data mining and learning analytics have risen as crucial tools for understanding and supporting students by leveraging institutional data. Machine learning (ML) and deep learning (DL) models have increasingly been deployed to predict academic performance, identify at-risk students, and enable timely interventions. Recent research underscores the importance of transcending traditional cognitive indicators by integrating behavioral and socio-emotional factors into predictive models.

Several studies have examined the relationship between social media use and academic outcomes. For instance, Nti et al. (2022) used decision trees and random forests to model how different social networking site usage patterns impact student grades, finding that intensive and in-class usage tends to depress academic performance. Various machine learning classifiers were used to analyze this relationship, showing that non-linear models provide deeper insights beyond simple correlations. Yet, despite these advances, three significant gaps persist. First, academic performance and social media impact are often modeled in isolation rather than within a unified analytical framework. Second, mental-health indicators such as sleep quality, mood disturbances, and anxiety—strongly associated with social media are seldom included directly in predictive student risk models. Third, many studies limit themselves to a narrow range of algorithms, lacking systematic comparisons involving powerful ensemble methods alongside neural networks.

This paper addresses these gaps by proposing a dual-objective machine learning framework that jointly predicts academic success using pre-admission and academic variables and negative social media impact from behavioral and mental-health questionnaire data. It compares multiple supervised learning algorithms within a transparent modeling pipeline and situates these predictive models within three-layer system architecture—comprising data, analytics, and application layers—designed for seamless integration into institutional student management systems.

## 2. Research Gap and Contribution

Synthesizing the literature, three gaps emerge:

1. Academic performance and digital life are typically studied separately, despite their intertwined nature.
2. Mental-health proxies linked to social media use are rarely incorporated directly into predictive models of student risk.
3. Comparative evaluations of multiple machine learning algorithms, including ensemble methods, in a unified framework remain limited.

This study contributes by:

- Designing a dual-objective framework that predicts academic success and negative social media impact in the same institutional context.
- Integrating pre-admission variables, social media behavior and mental-health indicators into a coherent modeling pipeline.
- Systematically comparing Random Forest, Support Vector Machine, multi-layer perceptron neural network, Gradient Boosting and a soft-voting ensemble for both tasks.
- Presenting a system architecture that situates these models within AI-enhanced student management and intervention processes.

## 3 Research Methodology

Architecture distinguishes between offline training and online inference.

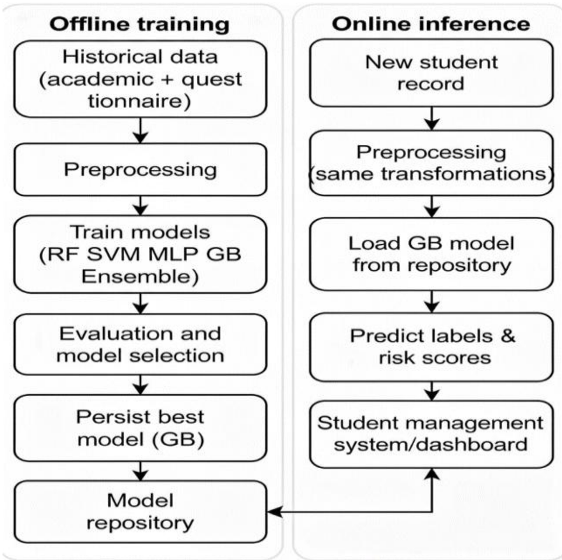


**Figure 1. Training and deployment flow.**

**Offline training:** Historical academic and questionnaire data are cleaned and transformed. Several models (RF, SVM, MLP, GB, Ensemble) are trained and compared, and the best one—typically Gradient Boosting—is stored in a model repository for deployment.

**Online inference:** New student records go through the same preprocessing steps. The stored best model is loaded to generate risk labels and scores for academic success and social media impact. These predictions are displayed on a student management dashboard, and the newly labeled data are added back to the historical dataset to support regular retraining and gradual model improvement.

This architecture can be implemented using conventional institutional databases for the data layer, Python and common ML libraries for the analytics layer, and an internal web dashboard for the application layer.

## 1. Data and Methods

This study employed two datasets from the same institution, each supporting a distinct predictive task. The pre-admission dataset included demographic variables, interview scores (0–100), subject-specific marks, and other academic attributes for 500 students. The social media and well-being dataset, collected via a structured questionnaire, captured social media interference with sleep, self-reported sleep quality, mood disturbance, and frequent anxiety or stress using Likert-scale items. The datasets were linked through STUDENTID for descriptive analysis, though predictive modeling was performed separately.

Academic success was labeled "Successful" when both interview and subject scores were at least 70, yielding 53 Successful and 447 Not Successful students. Negative social media impact was identified when all four conditions—moderate or strong sleep interference, poor or moderate sleep quality, notable mood disturbance, and frequent anxiety or stress—were met, classifying 112 students as Negative Impact and 388 as No Impact.

Data preprocessing for each task included label encoding categorical features and median imputation for missing numeric values. Each dataset was split into training (80%) and test (20%) sets with a fixed random seed. Model evaluation employs accuracy, confusion matrices, precision, recall, and F1-scores, with comparative results summarized across all machine learning models.

## 5. Results

### 5.1 Academic success prediction

The academic success prediction task exhibited strong class imbalance, with only 53 of 500 students (10.6 percent) identified as Successful and 447 as Not Successful. This imbalance highlights the importance of evaluating recall for the minority class. Across 5-fold cross-validation, Gradient Boosting and the soft-voting Ensemble achieved the highest accuracies and macro-F1 scores, closely followed by Random Forest, while Support Vector Machine and the MLP neural network showed weaker performance due to poor detection of Successful students.

On the test set, Gradient Boosting reached perfect accuracy (1.00), correctly classifying all students. The Ensemble achieved 0.99 accuracy with minimal misclassifications, and Random Forest followed at 0.98, missing only a few Successful cases. By contrast, the Support Vector Machine and MLP reported around 0.91 accuracy but failed to identify most successful students, demonstrating high majority bias and negligible recall for the minority class central to academic success analysis.

**Table 1. Test accuracy for academic success prediction**

| Model | Test accuracy | Comment |
|---|---|---|
| Random Forest | 0.98 | Very high accuracy; small number of Successful missed |
| Support Vector Machine | 0.91 | Strong majority bias; fails on Successful class |
| MLP neural network | 0.91 | Similar to SVM; low recall for Successful students |
| Gradient Boosting | 1.00 | Perfect separation on the test set |
| Soft-voting Ensemble | 0.99 | Near-perfect; captures almost all Successful students |

A detailed metric table (not shown here) would report, for each model, precision, recall and F1-score for both classes plus macro-F1. In that table, Gradient Boosting and the Ensemble score close to one on all indicators, Random Forest remains strong with slightly lower minority-class recall, and SVM and MLP show near-zero recall and F1 for the Successful class.

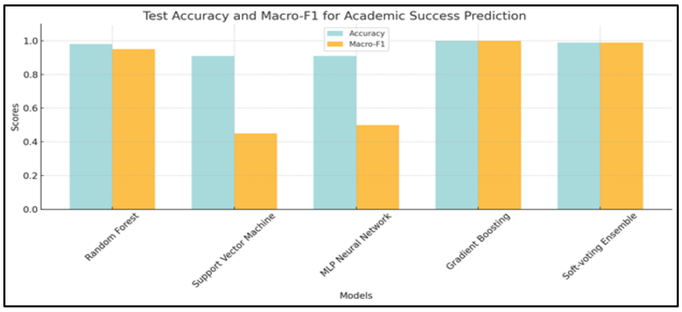


**Figure 2. Test accuracy and macro-F1 for academic success prediction**

### 5.2 Social media impact prediction

The social media impact prediction task showed moderate class imbalance, with 388 students labeled No Impact and 112 as Negative Impact. While less severe than the academic dataset, this imbalance still challenged certain models. Under 5-fold cross-validation, Gradient Boosting and the soft-voting Ensemble achieved the highest mean accuracies and macro-F1 scores, followed closely by Random Forest. The MLP neural network delivered intermediate performance, and Support Vector Machine remained the weakest, often predicting most cases as No Impact.

On the test set, Gradient Boosting achieved perfect classification accuracy (1.00), while the Ensemble recorded 0.99 accuracy, misclassifying only a few cases. Random Forest followed with 0.96 accuracy and strong recall for the minority class. The MLP obtained 0.87 accuracy but lower recall and F1 for Negative Impact, whereas the Support Vector Machine reached 0.78 accuracy and failed to identify any Negative Impact instances, resulting in poor overall balance.

**Table 2. Test accuracy for social media impact prediction**

| Model | Test accuracy | Comment |
|---|---|---|
| Random Forest | 0.96 | High accuracy: few Negative Impact cases missed |
| Support Vector Machine | 0.78 | Dominated by No Impact predictions; misses minority |
| MLP neural network | 0.87 | Moderate; partial detection of Negative Impact |
| Gradient Boosting | 1.00 | Perfect performance on both classes |
| Soft-voting Ensemble | 0.99 | Very close to Gradient Boosting; minimal errors |

Again, a separate metric table would document class-wise precision, recall and F1, and macro-F1. For this task, Gradient Boosting and Ensemble show balanced, near-perfect scores, Random Forest achieves strong but slightly lower recall on Negative Impact, MLP is moderate, and SVM shows zero recall for the minority class.

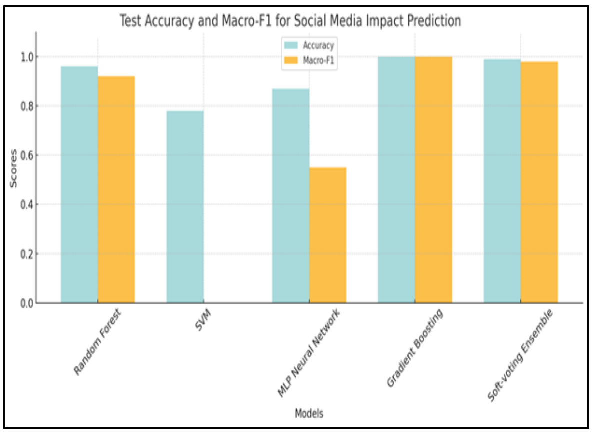


**Figure 3. Test accuracy and macro-F1 for social media impact prediction**

### 6. Discussion

The findings highlight the strength of gradient-boosting algorithms in predicting both academic success and social media impact using tabular educational data. Gradient Boosting achieved perfect test accuracy in both domains, with Random Forest and the soft-voting ensemble also performing competitively, particularly in the social media task. These results support prior research recognizing the superior capability of tree-based ensemble models in educational data mining (Khan & Hande, 2022; Lin et al., 2023; Turkmenbayev et al., 2025).

Support Vector Machine and the neural network showed limited sensitivity to minority classes, achieving high accuracy mainly by overpredicting the majority class. This imbalance leads to practical challenges, as missing Successful students or those experiencing Negative Impact undermines early intervention efforts. Future research could address this through class rebalancing, cost-sensitive training, or optimized threshold calibration (Hosen et al., 2025).

The proposed dual-objective framework extends earlier works (Nti et al., 2022; Shahzadi et al., 2020) by integrating behavioural and mental-health variables—such as sleep, mood, and anxiety—into social media impact prediction. It also complements existing academic performance models (Abdrakhmanov et al., 2024; Fahd & Miah, 2023), aligning with emerging AI-based student management systems (Wang, 2025; Prazeres & Pina, 2024) that support data-driven educational decision-making.

## 7. Limitations and Future Work

This study has some important boundaries. It uses data from a single institution, so the models and cut-off values may not directly generalize to other universities or programmes. The labels for "academic success" and "negative social media impact" are based on fixed rules; more graded categories could describe student realities better. The analysis is cross-sectional, capturing only one point in time and not how behavior and performance change across semesters. In addition, social media and well-being measures are self-reported and therefore vulnerable to recall and social desirability bias.

Future work can broaden and deepen the framework by: testing it in multiple institutions; trying advanced deep-learning and multimodal models; adding time-series and sequence approaches to track trajectories; using explainable AI to clarify which features drive predictions; combining predictive models with causal analysis; and deploying the system inside real AI-enabled student support platforms in collaboration with teachers, mentors and counselors.

## 8. Conclusion

This paper has presented a dual-objective machine learning framework and system architecture for predicting students' academic success and negative social media impact. Using pre-admission data and a social media–well-being questionnaire, five models were trained and compared on each task. Gradient Boosting consistently outperformed other methods, achieving perfect test accuracy across both domains, while Random Forest and a soft-voting ensemble also showed strong performance. Support Vector Machine and a multi-layer perceptron neural network delivered acceptable overall accuracy but failed to reliably detect minority-class students.

The framework illustrates how educational data mining can move beyond purely academic indicators to incorporate digital life and mental-health proxies, yielding a more comprehensive view of student risk. The proposed architecture positions predictive models within an institutional context, supporting dashboards and intervention workflows that can be refined over time. By linking social media impact and academic outcomes within a single analytical framework, the study contributes to ongoing efforts to design AI-enabled, student-centered educational systems that are both data-informed and sensitive to the broader well-being of learners.

## References

1. Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(10), 1533–1545. https://doi.org/10.1109/TASLP.2014.2339736
2. Abdrakhmanov, R., Zhaxanova, A., Karatayeva, M., Niyazova, G. Z., Berkimbayev, K., & Tuimebayev, A. (2024). Development of a framework for predicting students' academic performance in STEM education using machine learning methods. International Journal of Advanced Computer Science and Applications, 15(1). https://doi.org/10.14569/IJACSA.2024.0150105
3. AL-anazi, R. G., Alhammad, N. M., Rehman, M. B., Aljohani, N. J., Almohammade, N., Alsuhaibani, R. S., Alasmari, M., & Darem, A. A. (2025). A hybrid deep learning model with feature engineering technique to enhance teacher emotional support on students' engagement for sustainable education. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-17600-2
4. Bano, F., Serbaya, S. H., Rizwan, A., Shabaz, M., Hasan, F., & Khalifa, H. S. (2023). A neural network based mathematical model predicting students' performance in engineering education. Research Square. https://doi.org/10.21203/rs.3.rs-2643597/v1
5. Bashiri, M., & Kowsari, K. (2024). Transformative influence of LLM and AI tools in student social media engagement: Analyzing personalization, communication efficiency, and collaborative learning. arXiv. https://doi.org/10.48550/arXiv.2407.15012
6. Cano, A. (2019). Social media and machine learning. In IntechOpen eBooks. IntechOpen. https://doi.org/10.5772/intechopen.78089
7. Fahd, K., & Miah, S. J. (2023). Designing and evaluating a big data analytics approach for predicting students' success factors. Journal of Big Data, 10(1). https://doi.org/10.1186/s40537-023-00835-z
8. Hao, N., Meng, C., Zhao, N., Zhang, J., & Zheng, F. (2025). Artificial intelligence-assisted psychological intervention mechanisms for university students in the context of new media technologies: An analysis based on data from the National Institute of Mental Health. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1619818
9. Hosen, M. B., Ahmed, S., Akter, B., & Anannya, M. (2025). Impact, causation and prediction of socio-academic and economic factors in exam-centric student evaluation measures using machine learning and causal analysis. arXiv. https://doi.org/10.48550/arXiv.2506.12030
10. Khan, A., & Hande, K. N. (2022). A survey on prediction and analysis of students' academic performance using machine learning technique. International Journal for Research in Applied Science and Engineering Technology, 10(6), 780–789. https://doi.org/10.22214/ijraset.2022.43192
11. Li, Z. (2023). AI-assisted emotion recognition: Impacts on mental health education and learning motivation. International Journal of Emerging Technologies in Learning (iJET), 18(24), 34–52. https://doi.org/10.3991/ijet.v18i24.45645

12. Lin, Y., Chen, H., Xia, W., Fan, L., Wu, P., Wang, Z., & Li, Y. (2023). A comprehensive survey on deep learning techniques in educational data mining. arXiv. https://doi.org/10.48550/arXiv.2309.04761
13. Mao, S., Zhang, C., Song, Y., Wang, J., Zeng, X., Xu, Z., & Wen, Q. (2024). Time series analysis for education: Methods, applications, and future directions. arXiv. https://doi.org/10.48550/arXiv.2408.13960
14. Prazeres, F. G., & Pina, A. J. da S. (2024). AI integrated learning for higher education social science educators. USDAAD – International Journal of Social Sciences and Humanities Research. Retrieved from https://dergipark.org.tr/tr/pub/usdad/issue/86150/1529088
15. Turkmenbayev, A., Abdykerimova, E., Nurgozhayev, S., Karabassova, G., & Baigozhanova, D. (2025). The application of machine learning in predicting student performance in university engineering programs: A rapid review. Frontiers in Education, 10. https://doi.org/10.3389/feduc.2025.1562586
16. Wang, Y. (2025). Artificial intelligence in student management systems to enhance academic performance monitoring and intervention. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-19159-4
17. Nti, I. K., et al. (2022). Prediction of social media effects on students' academic performance using machine learning algorithms. Journal of Computers in Education. Springer.
18. Shahzadi, E., et al. (2020). Machine learning methods for analysing the impact of social media on students' academic performance. Journal of Advances in Applied & Interdisciplinary Sciences.