Conceptualizing Student Risk: A Dual-Objective Machine Learning Model and System Architecture for Academic and Social Media Outcomes

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 has become deeply embedded in students' daily lives, influencing how they manage time, emotions, and academic responsibilities. However, most research continues to examine academic performance, digital habits, and psychological well-being in isolation. This study introduces a dual-objective machine learning framework designed to jointly predict two outcomes: (i) academic success based on pre-admission records and (ii) negative social media impact using behavioural and mental-health indicators. The data were drawn from a higher education institution. Pre-admission records included demographic details, interview scores, and subject-specific marks, while a structured questionnaire captured perceptions of social media interference with sleep, sleep quality, mood, and stress or anxiety. Binary outcomes were defined for academic success (Successful / Not Successful) and social media impact (Negative Impact / No Impact) using institutional thresholds. After preprocessing with label encoding and median imputation, five classifiers—Random Forest, Support Vector Machine, multi-layer perceptron neural network, Gradient Boosting, and a soft-voting ensemble—were trained with an 80/20 split.Gradient Boosting consistently outperformed others, achieving 100% accuracy for both tasks. Random Forest followed closely, while Support Vector Machine and neural networks struggled with minority classes. Results confirm the effectiveness of gradient boosting and highlight the value of integrating behavioural and mental-health features for AI-driven student risk analytics.

Keywords: Academic Performance; Social Media Impact; Educational Data Mining; Machine Learning; Gradient Boosting; Student Risk Prediction

1. Introduction

Social media platforms such as WhatsApp, Instagram, YouTube and other AI-enhanced tools have become central to the daily experience of students in higher education. These platforms enable rapid communication, facilitate informal peer learning and offer access to a large volume of educational resources. At the same time, they can contribute to fragmented attention, reduced study time, sleep disruption and increased anxiety. The educational system must therefore deal with social media as both an opportunity and a potential source of risk.

Parallel to this development, educational data mining and learning analytics have emerged as key means of understanding and supporting students based on institutional data. Machine learning (ML) and deep learning (DL) models are increasingly used to predict academic performance, detect at-risk students and assist with timely interventions. Recent work emphasises the need to move beyond traditional cognitive indicators towards a more holistic view that incorporates behavioral and socio-emotional information.

Several studies have explored the relationship between social media use and academic outcomes. Nti et al. (2022) modeled the impact of different aspects of social networking site use on student grades using decision trees and random forests and reported that intensive and in-class use tends to depress performance. ML classifiers were applied to study the relationship between social media usage and academic performance, showing that non-linear models uncover insights that go beyond what simple correlations can reveal. Despite this progress, three issues remain:

1. Academic performance and social media impact are often modeled separately rather than as components of a unified analytical framework.
2. Mental-health indicators such as sleep quality, mood disturbance and anxiety are rarely incorporated directly into predictive models of student risk, even though they are strongly associated with social media use in the wider psychological literature.
3. Many studies employ a narrow set of algorithms; systematic comparisons that include powerful ensemble methods, alongside neural networks, are relatively scarce.

This paper addresses these gaps by proposing a dual-objective machine learning framework that predicts academic success from pre-admission and academic variables; Predicts negative social media impact from questionnaire-based behavior and mental-health measures, compares multiple supervised learning algorithms under a single, transparent pipeline; and embeds these models within a system architecture suitable for integration into institutional student management systems.

2. Literature Review

[1] Social media acts as a complex socio-technical environment in which students communicate build identity and access information. Cano (2019) describes how social media data can be combined with machine learning to infer behavioral patterns and user characteristics. Bashiri and Kowsari (2024) highlight the transformative influence of large language models and AI tools on student social media engagement, including how these technologies change personalization, communication efficiency and collaborative learning.

[2] Predicting academic performance from student data is an established area within educational data mining. Khan and Hande (2022) review a wide range of machine learning techniques for performance prediction, including decision trees, random forests, Support Vector Machines and neural networks. Abdrakhmanov et al. (2024) propose a framework for predicting performance in STEM disciplines using machine learning methods, stressing the importance of feature selection and robust evaluation. Bano et al. (2023) construct a neural network–based mathematical model for predicting engineering students' performance, illustrating the potential of relatively simple deep networks.

[3] Lin et al. (2023) survey deep-learning techniques in educational data mining, covering architectures such as convolutional neural networks (CNNs), recurrent networks and attention mechanisms. Abdel-Hamid et al. (2014), although working on speech recognition, show how CNNs can learn robust hierarchies of features, a concept that has inspired later applications in

educational modelling. Mao et al. (2024) argue that many educational phenomena are inherently temporal and advocate time-series analysis for tracking evolving patterns of engagement and performance.

[4] Recently, research has moved beyond pure prediction to consider socio-academic and economic drivers of performance. Hosen et al. (2025) combine machine learning and causal analysis to investigate how socio-academic and economic factors relate to exam-centric evaluation measures, suggesting that predictive models should be complemented by causal reasoning where policy decisions are at stake. Prazeres and Pina (2024) discuss AI-integrated learning approaches for social science educators and consider how AI systems change teaching practice.

[5] Wang (2025) focuses on AI-enabled student management systems, highlighting data pipelines and algorithms for continuous performance monitoring and intervention. AL-anazi et al. (2025) and Li (2023) show how affective and engagement signals can be integrated into such systems to provide more nuanced support.

3. Research Gap and Contribution

While existing studies have explored academic performance and social media impact separately, there is a significant gap in frameworks that integrate these two dimensions within a unified predictive system. Moreover, mental-health indicators such as sleep quality, mood disturbance, and anxiety, which are strongly linked to social media usage, are rarely incorporated directly into predictive models of student risk. Additionally, current research tends to focus on a narrow set of ML algorithms, lacking comprehensive comparative evaluations that include powerful ensemble methods alongside neural networks. This paper addresses these gaps by proposing a dual-objective machine learning framework that jointly predicts academic success and negative social media impact using distinct but interconnected feature sets. The framework is embedded in a three-layer system architecture—data, analytics, and application—designed for seamless integration into institutional student management systems, thereby offering a more holistic and actionable approach to student risk analytics through AI-enabled student support.

4. Conceptual Model and System Architecture

4.1 Conceptual Model

The conceptual model aligns three elements: student attributes, predictive models and risk outcomes.

- **Academic branch**: Pre-admission and academic variables (demographics, interview scores, subject-specific marks and related fields) are used to predict whether a student meets a defined threshold for academic success.
- **Social media branch**: Questionnaire-based indicators of social media interference with sleep, general sleep quality, mood disturbance and anxiety/stress are used to determine whether a student is experiencing negative social media impact.

The two branches share a common ML layer but operate on different feature sets and target variables. Both branches feed into a joint risk profile for each student.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 11, Issue , 10 2022

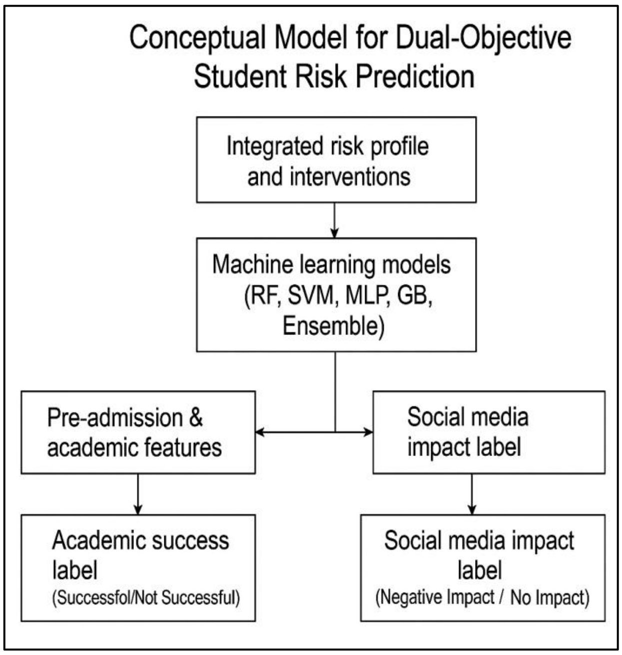

Figure 1. Conceptual model for dual-objective student risk prediction

4.2 System Architecture

The proposed system architecture has three layers: **data layer**, **analytics layer** and **application layer**.

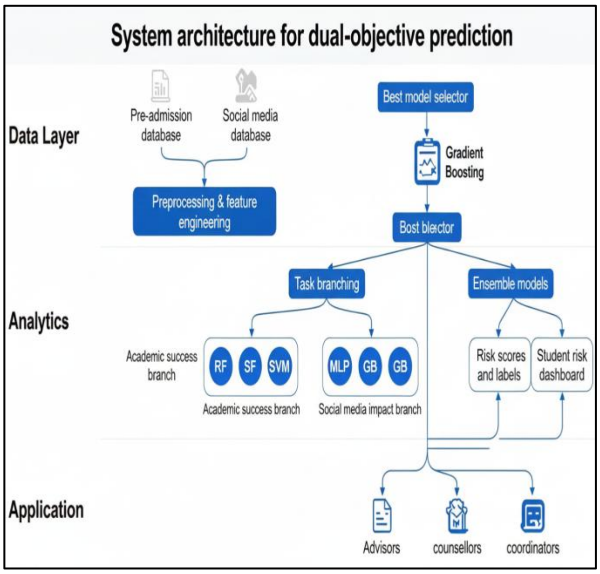

Figure 2. System architecture for dual-objective prediction.

System architecture presents a three-layer framework linking data, analytics, and application. At the bottom **data layer**, pre-admission records and social media/well-being responses flow into a *Preprocessing & feature engineering* block. The **analytics layer** then branches into *Academic success* and *Social media impact* paths, where RF, SVM, MLP, Gradient Boosting, and Ensemble models are evaluated, and a *Best model selector* chooses Gradient Boosting as the primary deployment model, with RF and Ensemble as alternatives. In the **application**

layer, model outputs are converted into risk scores and visualized on a *Student risk dashboard*, which supports advisors, counsellors, and coordinators in planning student interventions.

5. 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 generalise 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.

6. 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*.