International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

Academic Performance Prediction on Multisource, Multifeature Behavioral Data

Gnanasekar V[1], Maheshwari S[2], Ragini K P[3], Vidhya V[4], Nathiya S[5], Jaishree K[6]

[1] *Head of the department (HOD) , Department of Artificial Intelligence and Data Science , Hindusthan College of Technology Salem, India*

[2] *Assistant Professor,Department of Artificial Intelligence and Data Science , Hindusthan College of Technology Salem, India*

[3,4,5,6] *Department of Artificial Intelligence and Data Science Hindusthan Gandhi College of Technology Salem, India*

Abstract—*This work presents an intelligent academic performance prediction framework that analyzes multisource, multifeature behavioral data collected from digital learning environments, campus activity records, classroom interactions, and assessment histories. Traditional prediction systems frequently depend on a narrow set of variables such as examination marks or attendance alone, which limits their ability to capture the complex behavioral factors that influence student outcomes. The proposed system integrates heterogeneous learning traces and transforms them into a unified representation for early, data-driven academic risk estimation.*

The system operates through four major modules. First, a Data Acquisition Module gathers academic, temporal, engagement, and behavioral records from multiple institutional sources. Second, a Feature Engineering Module extracts indicators such as learning regularity, assignment consistency, attendance stability, interaction frequency, and performance progression. Third, a Predictive Analytics Module applies machine learning and sequence-aware deep learning models to estimate final academic performance and identify at-risk students. Finally, a Visualization and Intervention Module presents interpretable alerts, dashboards, and personalized recommendations that support timely academic counseling and targeted remediation.

I. INTRODUCTION

In modern educational institutions, students generate large volumes of digital traces through Learning Management Systems (LMS), online assessments, classroom participation platforms, smart attendance systems, library access, and other campus services. These records reflect not only what students achieve, but also how consistently they engage with learning activities. However, many institutions still evaluate academic risk using limited static indicators, making it difficult to identify subtle signs of disengagement before poor results become visible.

Academic performance is influenced by multiple factors, including prior marks, attendance behavior, assignment submission habits, online content access, discussion participation, temporal learning patterns, and the stability of engagement over time. A single-source dataset cannot fully represent these factors. Consequently, prediction models built from only grades or attendance often overlook important behavioral signals. Multisource, multifeature data fusion provides a stronger analytical foundation because it combines academic evidence with behavioral patterns observed across both online and offline learning contexts.

To address these limitations, this paper proposes a comprehensive prediction framework for academic performance based on multisource, multifeature behavioral data. The model emphasizes early warning, holistic student profiling, and actionable feedback. It supports preprocessing of heterogeneous records, feature construction, predictive modeling, and intervention dashboards. The final objective is not merely to classify students by performance level, but to provide educators with a practical decision-support system capable of identifying at-risk learners early and guiding evidence-based academic support.

II. LITERATURE SURVEY

The literature on educational data mining and learning analytics has increasingly focused on predicting student performance using behavioral and temporal data. Existing studies show that predictive accuracy improves when systems combine academic records with learning traces, interaction data, and engagement features. This section reviews the major research directions relevant to the proposed framework: single-source performance prediction, multisource behavioral modeling, deep learning approaches, explainable prediction, and the remaining research gap.

A. Traditional Academic Performance Prediction Systems

Early academic prediction systems primarily relied on static indicators such as internal marks, demographic attributes, semester grades, and attendance. Classical algorithms including Decision Trees, Naive Bayes, Logistic Regression, Support Vector Machines, and Random Forests were used to estimate pass/fail status or

© 2026 The Author(s). Published by IJCOPE Journal. Website: https://ijcope.org/

grade categories. While these approaches are useful for baseline prediction, they do not adequately capture the evolving nature of learning behavior. Their dependence on limited features reduces their effectiveness for early intervention, especially when academic deterioration begins as a behavioral pattern before it appears in formal assessments.

B. Multisource and Multifeature Behavioral Data Analytics

Recent research emphasizes combining multiple data sources to obtain a more complete view of student learning. The AugmentED approach proposed for academic performance prediction integrates behavioral data from digital campus environments, online and offline learning interactions, and temporal feature patterns. Related work on multimodal data fusion has also shown that combining theory classes, practical sessions, Moodle interactions, and examination records can improve prediction quality. These findings motivate the use of feature groups such as access frequency, study regularity, attendance trends, assignment punctuality, and engagement stability.

C. Sequence Learning and Deep Predictive Models

Behavioral data are often sequential because student actions evolve week by week. Deep learning architectures such as Long Short-Term Memory (LSTM), Bidirectional LSTM, attention-based networks, and hybrid temporal models are increasingly used to learn trends from activity histories. Sequence-aware models can capture delayed effects, repeated behavioral fluctuations, and progressive disengagement. Studies using VLE clickstream data and engagement sequences report strong early-prediction capability, indicating that temporal information is critical when designing a system for semester-level academic forecasting.

D. Explainable Learning Analytics and Intervention Support

Prediction alone is insufficient unless educators can understand why a student is flagged as at risk. Explainable approaches such as feature importance ranking, SHAP-based interpretation, and visual analytics help convert model output into actionable support. Institutions require dashboards that reveal which factors contribute to low predicted performance, such as declining attendance, reduced platform engagement, irregular submissions, or a sudden reduction in practice activity. Interpretable prediction helps academic mentors plan targeted follow-up rather than relying on generic remedial measures.

E. Research Gap

Although previous studies demonstrate the value of behavioral analytics, many proposed models remain either dataset-specific or limited to a single educational platform. Some systems emphasize prediction accuracy but do not provide a clear intervention workflow, while others use multimodal inputs without presenting a practical architecture for deployment in institutional settings. The proposed framework addresses this gap by integrating multisource data ingestion, feature engineering, performance prediction, and intervention visualization into one coherent model suitable for academic decision support.

III. EXISTING SYSTEM

Existing student performance monitoring systems frequently depend on examination scores, internal assessments, and attendance percentages. These variables provide useful summary information, but they offer only a partial picture of learning progression. A student may show declining academic readiness through delayed submissions, irregular login behavior, reduced quiz attempts, or unstable learning routines long before final marks fall. Systems that ignore these behavioral indicators are often late in detecting academic risk.

Many current Learning Management Systems provide raw activity logs, yet they do not transform these logs into meaningful predictive indicators. Faculty members may view counts of logins or assignment uploads, but these are rarely integrated with classroom attendance, historical academic trends, assessment performance, and engagement consistency. The absence of unified data fusion results in fragmented analytics and weak personalization of academic support.

Another limitation is that traditional models often treat student records as static snapshots rather than evolving sequences. This reduces the ability to detect gradual behavior change, such as decreasing engagement over several weeks. In addition, prediction outputs are commonly presented as opaque risk labels without sufficient explanation of the underlying causes, making it difficult for teachers to design targeted interventions.

The proposed work responds to these challenges by shifting from isolated score-based monitoring to a multisource behavioral prediction ecosystem. It combines heterogeneous institutional data, extracts meaningful temporal and behavioral indicators, applies predictive models, and generates explainable student-level insights. This approach enables a more proactive academic support process focused on early detection, intervention prioritization, and continuous monitoring.

IV. PROPOSED SYSTEM

The proposed system is a comprehensive analytics framework for predicting student academic performance using multisource, multifeature behavioral data. Its central objective is to identify students who may be academically vulnerable before final outcomes are known. The system collects data from LMS usage, attendance records, internal assessments, assignment submissions, quiz behavior, classroom participation indicators, and optional campus interaction logs. These sources are consolidated into a structured student profile for prediction.

A dedicated Feature Engineering Layer converts raw traces into interpretable variables. Examples include weekly login frequency, content access duration, assignment punctuality, attendance regularity, quiz retry behavior, discussion activity, laboratory participation, and trend-based change features. Both aggregate indicators and temporal sequences are retained, allowing the predictive engine to learn stable patterns as well as week-to-week fluctuations. This design improves sensitivity to behavioral decline that may not be reflected in marks alone.

The Predictive Analytics Layer combines classical machine learning models with deep sequential models. Baseline models such as Random Forest and Gradient Boosting can provide efficient institutional benchmarking, while LSTM or attention-enabled sequence networks can capture dynamic engagement trajectories. The output is presented through risk bands, performance probability estimates, and intervention priority levels. By synthesizing student behavior into actionable predictions, the system supports advisors, faculty, and administrators in delivering early and personalized academic assistance.

Technical Vector:

The Predictive Intelligence Vector acts as the analytical core of the system. It transforms heterogeneous educational data into a structured decision-support pipeline by combining data cleaning, statistical feature extraction, temporal behavior modeling, and supervised learning. The vector measures not only current performance but also the evolution of student engagement across the semester. It enables the prediction engine to differentiate between temporary variation and persistent academic risk, thereby producing more reliable forecasts for early intervention.

Fundamental Vector

- The framework evaluates four foundational dimensions: Academic Readiness, Engagement Intensity, Learning Regularity, and Behavioral Stability. Academic Readiness reflects marks, quiz results, and assessment trends. Engagement Intensity represents LMS usage, content access, and participation frequency. Learning Regularity captures attendance consistency and submission punctuality. Behavioral Stability measures whether a student maintains or loses engagement over time. Together, these vectors create a multidimensional student representation that is more informative than single-score monitoring.

Advantages of the Proposed System:

- Multisource Student Profiling: The system unifies academic, behavioral, temporal, and engagement data instead of depending on a single record type. This creates a richer understanding of how learning behavior influences performance.
- Early Academic Risk Detection: By analyzing weekly trends and behavioral changes, the model can identify at-risk students before final examination performance becomes visible, enabling timely support.
- Explainable Decision Support: The framework provides interpretable feature-level insights so that faculty can understand whether risk is driven by attendance decline, submission delay, weak assessment scores, or reduced engagement.
- Scalable Institutional Analytics: The architecture can support departments, colleges, and digital learning platforms by converting existing academic databases and learning logs into practical intervention dashboards.

SYSTEM REQUIREMENTS

Hardware Specifications: Although the proposed framework can be implemented as a web-based analytics application, the infrastructure must support multisource data ingestion, preprocessing, model training, and dashboard reporting. The recommended configuration below ensures reliable experimentation, institutional deployment, and responsive prediction services.

Processor (CPU): A modern multi-core processor such as Intel Core i5/i7 or AMD Ryzen 5/7 is recommended. The system performs data preprocessing, feature aggregation, and model inference on potentially large student records. Multi-core computation supports faster batch processing and concurrent analytics requests.

Memory (RAM): A minimum of 8 GB RAM is suitable for prototype implementation, while 16 GB or more is preferable for institutional datasets and deep learning experiments. Sequence modeling, dataframe transformations, and feature matrix construction can become memory-intensive when multiple semesters of behavior logs are considered.

Storage Environment: A 256 GB SSD or larger is recommended to store datasets, transformed features, trained models, visualization exports, and application files. Fast storage improves read/write speed during preprocessing and supports efficient retrieval of historical behavioral records.

Network Capabilities: A stable internet or campus intranet connection is required when data are synchronized from LMS servers, attendance services, or cloud databases. Low-latency connectivity improves dashboard responsiveness and ensures that risk indicators are updated in near real time.

Software Requirements:

- Python: Python serves as the primary data science and machine learning environment. It supports preprocessing, exploratory analysis, feature generation, model training, and evaluation using libraries such as Pandas, NumPy, Scikit-learn, TensorFlow, or PyTorch.
- Backend Framework (Node.js / FastAPI / Flask): A lightweight backend service can be used to expose prediction APIs, manage student profile requests, and connect analytics outputs to the user interface. This layer enables deployment of trained models inside an institutional web platform.
- Database Management (PostgreSQL / MySQL / MongoDB): The system stores academic records, behavioral logs, feature snapshots, model outputs, and intervention histories. Relational databases are suitable for structured academic data, while document databases may be used for flexible event logs.
- Authentication & Security: Secure login, role-based access control, encrypted transmission, and controlled data visibility are required because the application processes student-level behavioral and academic information. Faculty and administrators should only access records relevant to their responsibilities.
- Cloud Storage & File Management: Optional cloud storage may be used to retain data exports, model reports, trained weights, and institutional documents. Access control policies should prevent unauthorized exposure of student performance records.
- Frontend Environment (React.js & Tailwind CSS): A responsive dashboard interface can display performance predictions, trend graphs, at-risk alerts, and intervention summaries. Faculty can filter by department, semester, course, or risk group to prioritize academic support.

Data Processing and Mathematical Modeling

The proposed system applies a structured analytical pipeline to convert raw behavioral observations into meaningful prediction features. The workflow includes preprocessing, feature normalization, sequence encoding, predictive modeling, and model evaluation.

.

- 1. Data Cleaning and Normalization: Raw inputs may contain missing values, inconsistent timestamps, duplicate events, and features measured on different scales. The system applies cleaning rules, missing-value treatment, and normalization so that variables such as attendance rate, quiz score, login count, and time-on-task can be modeled consistently.
- 2. Multifeature Behavioral Representation: The system constructs features for participation frequency, assignment punctuality, assessment progression, attendance regularity, weekly engagement, and behavior change rate. These indicators are grouped into academic, engagement, regularity, and temporal vectors that represent the learner comprehensively.
- 3. Predictive Modeling with Machine Learning and LSTM: Traditional classifiers and regressors establish baseline performance, while LSTM-based models capture temporal learning dynamics. The target may be final grade category, probability of course success, or risk of low performance. Performance is measured using accuracy, precision, recall, F1-score, ROC-AUC, and mean error metrics where appropriate.
- 4. Risk Stratification and Feedback Generation: Prediction results are translated into interpretable categories such as low risk, moderate risk, and high risk. The system highlights dominant risk factors and supports personalized feedback, including increased practice recommendations, attendance counseling, assignment planning, or mentor follow-up.

Data Ingestion, Storage, and Communication

The communication architecture of the platform is designed as a unified educational analytics bus. It collects data from heterogeneous sources, stores normalized student records, delivers them to the predictive engine, and returns explainable outputs to dashboards used by educators and administrators.

- Multisource Data Synchronization: The ingestion layer receives records from LMS logs, assessment databases, attendance systems, and optional campus interaction sources. Scheduled data synchronization ensures that the prediction pipeline reflects recent student behavior while maintaining a consistent institutional data schema.

- Secure Identity and Access Propagation: Every student record is associated with a protected identity key. Role-based permissions ensure that students, faculty advisors, and administrators view only the information relevant to their responsibilities. Secure APIs and token-based authentication help protect academic privacy.

- Model Service Orchestration: Prediction services can be deployed through secure backend endpoints. The trained model receives current feature vectors, computes performance probabilities or risk labels, and returns structured outputs for visualization. This design separates the application layer from the model engine and supports future model upgrades.

- Persistent State Management: Historical feature snapshots, predictions, intervention notes, and outcome labels are stored for longitudinal evaluation. This enables the institution to compare forecast quality across semesters and to retrain models as student behavior patterns evolve.
- Dashboard and Reporting Layer: The visualization module converts outputs into student trend charts, risk summaries, cohort dashboards, and faculty-level reports. It supports evidence-based academic monitoring and provides a feedback loop for evaluating whether interventions improve student outcomes.

V ADVANTAGES

The proposed system offers a significant improvement over conventional score-based monitoring by integrating multisource behavioral analytics with predictive modeling. It enables institutions to identify academic difficulty as a pattern of behavior rather than as a final consequence revealed only through examination failure.

A major advantage is the use of multidimensional evidence. Since the model analyzes attendance, LMS interaction, submissions, quizzes, and time-dependent engagement behavior together, it produces more holistic risk estimates than systems based on any single academic variable. This enhances the reliability of early alerts and supports better-informed faculty decisions.

The framework also strengthens intervention planning. Instead of issuing a generic warning, the system can point to the factors associated with performance decline, helping educators distinguish between concept difficulty, poor attendance, delayed submissions, and disengagement. This supports personalized academic assistance and efficient allocation of mentoring resources.

Finally, the architecture is adaptable and scalable. It can be applied in colleges, universities, online learning platforms, and blended learning environments. As more historical data become available, institutions can retrain the model, evaluate fairness, and refine their intervention policies without redesigning the entire system.

APPLICATION

The proposed project can be applied across several educational settings where timely understanding of student performance is valuable. Its combination of data fusion, predictive modeling, and visual analytics supports both academic management and learner-centered intervention.

- Institutional Early Warning Systems: Colleges and universities can use the platform to identify students whose learning behaviors indicate a rising risk of poor performance, allowing advisors to intervene before final examinations.
- Learning Management System Analytics: Online and blended learning platforms can integrate the prediction engine with activity logs, quiz attempts, and submission patterns to generate dynamic academic dashboards.
- Faculty Mentoring and Remedial Planning: Departments can prioritize counseling, bridge courses, and personalized remedial tasks based on the specific behavioral factors contributing to each student's risk score.
- Academic Quality Assurance and Policy Support: Administrators can analyze cohort-level trends, evaluate how engagement relates to performance, and design data-informed retention strategies for different programs and semesters.

VII CONCLUSION

The development of an academic performance prediction system based on multisource, multifeature behavioral data represents an important advancement in educational analytics. By moving beyond isolated grade-based monitoring, the framework provides a more complete understanding of how student engagement, regularity, and learning activity patterns relate to final academic outcomes.

A key contribution of the proposed system is its ability to combine heterogeneous evidence into a unified prediction pipeline. LMS logs, attendance, assignments, assessments, and behavioral sequences are processed together to produce richer representations of student learning. This integration is particularly valuable for identifying early signs of academic risk that are not yet visible in formal examination records.

The framework also addresses the practical need for timely intervention. Through interpretable risk indicators, feature-level explanations, and dashboard visualizations, educators can move from reactive result analysis to proactive learner support. Students can receive more personalized guidance, while institutions can prioritize mentoring resources more effectively.

From a modeling perspective, the use of machine learning and temporal deep learning supports both efficient baseline prediction and advanced sequence-aware analysis. Behavioral trajectories across weeks or modules can reveal patterns of irregular engagement, delayed participation, or gradual academic disengagement. These patterns are critical for building reliable early warning systems.

From a technical perspective, the proposed architecture supports data ingestion, preprocessing, prediction services, secure storage, and visual reporting within one deployable framework. This makes the system suitable for institutional adoption and future integration with existing educational software environments.

The model also opens opportunities for responsible educational decision-making. Prediction should assist human judgment rather than replace it. Therefore, explainability, privacy safeguards, and fairness evaluation should remain central when deploying student analytics systems in real academic settings.

In conclusion, academic performance prediction using multisource, multifeature behavioral data offers a scalable and meaningful path toward personalized education. It enables early identification of vulnerable learners, supports targeted intervention, and helps institutions transform educational data into actionable academic intelligence.

VIII. FUTURE SCOPE

Future work can extend the system through multimodal and real-time analytics. Additional signals such as text feedback, discussion sentiment, classroom interaction markers, and mobile learning activity may enrich behavioral representation. Advanced fusion models can be studied to combine structured, sequential, and textual educational data.

Another promising direction is the integration of explainable and fairness-aware artificial intelligence. Future models can include bias monitoring across demographic or program groups, confidence estimation for predictions, and transparent explanations that help teachers validate whether a system-generated alert is pedagogically meaningful.

The framework can also evolve into an adaptive intervention platform. Instead of only predicting risk, future systems may recommend personalized learning plans, automatically assign revision resources, track response to interventions, and measure whether predicted risk decreases over time. This would convert the model from a passive forecasting tool into an active academic support ecosystem

IX. REFERENCE

The following references support the development of multisource behavioral analytics, student performance prediction, temporal learning models, and intervention-oriented educational data mining.

[1] L. Zhao, K. Chen, J. Song, X. Zhu, J. Sun, B. Caulfield, and B. Mac Namee, "Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data," IEEE Access, vol. 9, pp. 5453–5465, 2021, doi: 10.1109/ACCESS.2020.3002791.https://www.researchgate.net/publication/393631049_Gamification_in_Education_Enhancing_the_Students_Engagement_and_Learning_Achievements_through_the_Integration_Use_of_Game_based_Learning

[2] W. Chango, R. Cerezo, and C. Romero, "Multi-source and multimodal data fusion for predicting academic performance in blended learning environments," Computers & Electrical Engineering, vol. 89, 2021.https://www.researchgate.net/publication/394844305_SMART-Scholar_A_Scholarship_Management_and_Tracking_System

[3] X. Li, T. Zhang, and H. Ogata, "Student Academic Performance Prediction Using Deep Multi-Source Behavior Data," IEEE Access / educational analytics literature, 2020.https://www.sciencedirect.com/science/article/pii/S2772662225000724

[4] Y. Liu, R. Sim, and colleagues, "Predicting Student Performance Using Clickstream Data and Machine Learning," Education Sciences, vol. 13, no. 1, 2023.https://ieeexplore.ieee.org/document/11271208

[5] M. Fazil et al., "A Novel Deep Learning Model for Student Performance Prediction Using Engagement Data," Journal of Learning Analytics, 2024.https://ijsrem.com/volume09issue01january2025/

[6] W. Dai et al., "Learning Analytics for Early Identification of At-Risk Students and Feedback Intervention," Journal of Learning Analytics, 2025.https://pmc.ncbi.nlm.nih.gov/articles/PMC4063894/

[7] H. Acosta et al., "Multimodal Learning Analytics for Predicting Student Collaboration Satisfaction," Proceedings of Educational Data Mining, 2024.doi.org

[8] W. Zou et al., "Prediction of Student Academic Performance Utilizing a Practical Machine Learning Framework," Applied Sciences, vol. 15, no. 7, 2025.doi.org

[9] Y. Wu et al., "Research on Prediction Algorithm of College Students' Academic Performance Based on Multimodal Data Fusion," 2025.doi.org

[10] M. R. Borna et al., "Analyzing Click Data with AI: Implications for Student Academic Achievement Prediction," Frontiers in Education, 2024.https://www.sciencedirect.com/topics/computer-science/deep-learning

[11] M. Adnan et al., "Predicting At-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models," IEEE Access, vol. 9, pp. 7519–7539, 2021.doi.org

[12] H. Sahlaoui et al., "Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations," IEEE Access, vol. 9, pp. 152688–152703, 2021.ijsred.com

[13] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics Dataset," Scientific Data, vol. 4, 2017.https://link.springer.com/article/10.1186/s43093-026-00739-3

[14] R. Parkavi, P. Karthikeyan, and A. Sheik Abdullah, "Predicting Academic Performance of Learners with the Three Domains of Learning Data Using Neuro-Fuzzy Model and Machine Learning Algorithms," Journal of Engineering Research, 2024.https://ieeexplore.ieee.org/document/10974121

[15] S. Boujmiraz et al., "Predicting Student Performance: A Comprehensive Review of Machine Learning, Deep Learning, and Explainable AI," 2026.https://www.sciencedirect.com/science/article/abs/pii/S0040162524001264

[16] L. Yu et al., "Interpreting Predictive Learning Sequences in a College Learning Environment," Journal of Learning Analytics, 2025.doi.org

[17] H. Guo et al., "A Human-Centred AI Approach to Contextualizing Test Performance Through Process Data," Journal of Learning Analytics, 2024.doi.org

[18] C. Herodotou et al., "Facing Dashboard Design for Online and Distance Education Predictive Analytics," Journal of Learning Analytics, 2025.doi.org

[19] U. T. Van, B. H. Tieu, and D. H. Tran, "A Multimodal Learning Analytics Model Based on Deep Learning for PredictingStudentPerformanceUsingTabularandTimeSeriesDataFusion,"DiscoverArtificialIntelligence,2026.doi.org

[20] P. V. Sargent et al., "Beyond Time on Task: Behavioral Indicators for Learning Analytics and Student Performance Prediction," Journal of Learning Analytics, 2026.https://www.mdpi.com/2079-9292/13/17/3396