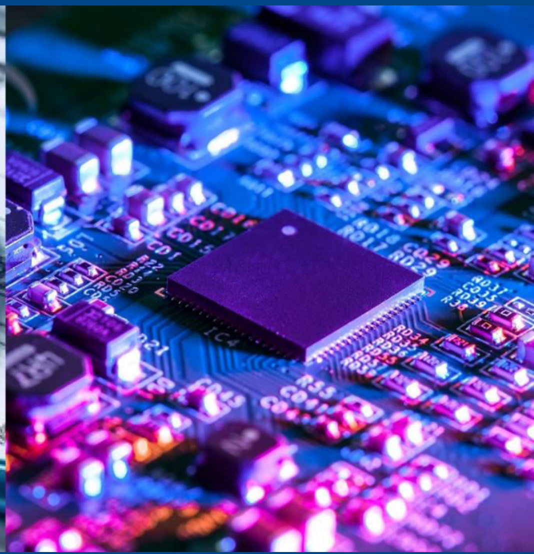
ISSN: 2582-7219

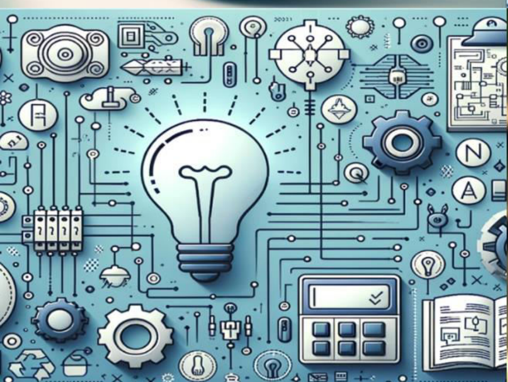




# International Journal of Multidisciplinary Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



Impact Factor: 9.864

Volume 9, Issue 6, June 2026

www.ijmrset.com | ijmrset@gmail.com | +91 6381 907 438 | +91 63819 07438

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906027

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Smart AI-Integrated Student Performance Monitoring System for Undergraduate Education Using Machine Learning, NLP, and Cloud Analytics

**Somesh Haribhau Uphad, Vaibhav Ashok Narke, Chetan Prabhakar Pawar,**

**Akash Devidas Vhanchenje, Atharva Shivaji Morade, Prof. K.V. Kankariya**

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

Department of Artificial Intelligence and Data Science, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India

**ABSTRACT:** Traditional undergraduate academic monitoring relies on exams and attendance records. It often fails to detect students who are at risk of falling behind until it is too late. This paper presents the Smart AI-Integrated Student Performance Monitoring System (SAISPMS). It is a native platform that uses Machine Learning, Natural Language Processing and Big Data Analytics. These tools help monitor students' performance in real-time and predict academic risks early. The SAISPMS system takes in multiple types of data. These include academic records, attendance, learning management system (LMS) engagement and student reflective journals. This data is processed through a Microsoft Azure cloud pipeline. The system uses a combination of Random Forest, XGBoost and LSTM networks. This combination helps classify students into risk categories. It achieves an accuracy of 93.7% with an AUC-ROC score of 0.97. The system also includes a tuned BERT-based NLP module. This module analyzes students' journal submissions for sentiment and stress. It achieves an F1-score of 0.89. The SAISPMS was tested on 2,400 undergraduate students across six departments.

**KEYWORDS:** Student Performance Monitoring, Machine Learning, NLP, Sentiment Analysis, Cloud Analytics, LSTM, Early Warning System, Undergraduate Education.

## I. INTRODUCTION

Colleges and universities are dealing with a critical challenge. They have a large number of students, not enough teachers, and many students are dropping out. This means we need a systematic way to monitor how students are performing. Currently, we mostly rely on report cards and test scores. This approach is insufficient because it only reveals how students are doing after it is too late to intervene and help them.

Artificial Intelligence and Cloud Analytics are tools that can help address this gap. These tools can analyze large volumes of student information and provide real-time insights into their progress. This means we can assist students before they fall into academic difficulty. If we also employ Natural Language Processing, we can even understand how students are feeling and what they are thinking when they write about their classes.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SAISPMS is a system that can accomplish all these tasks. It can identify if a student is likely to struggle with a class five weeks before the final examination. It can also analyze how students are feeling about their classes by processing what they write. The system provides teachers and counselors with timely information on student progress and sends alerts when a student needs help, along with recommended interventions.

SAISPMS does all this and more — including predicting academic risk and quantifying emotional engagement via NLP sentiment analysis — and delivers role-based real-time cloud dashboards for faculty and counselors.

## II. RELATED WORK

Romero and Ventura established Educational Data Mining as the foundation for predicting how well students will perform in school. Hussain et al. showed that using ensemble models is more effective than a single model for predicting student dropout in online classes, and that LSTM networks are particularly well-suited for understanding student behavioral patterns over time.

Natural Language Processing is an exciting area of related research. Wen et al. found that when students express negative sentiments about a course online, they are more likely to drop out. Ortigosa et al. showed that sentiment analysis can improve prediction accuracy by approximately 12 percent compared to using only quantitative features. Aldowah et al. demonstrated that Cloud Big Data analytics is important for analyzing student learning patterns in real time.

SAISPMS addresses several gaps identified in prior systems. Most existing systems use only a single type of data and do not combine NLP with quantitative analysis. Furthermore, no prior system is specifically designed to accommodate the challenges of Indian higher education institutions, such as large class sizes, irregular data, and multilingual student populations with diverse regional linguistic patterns.

## III. METHODOLOGY

### A. System Architecture

SAISPMS is a five-layer cloud-native system comprising: (1) Data Ingestion, (2) Feature Engineering, (3) AI and NLP Analytics, (4) Cloud Storage and Big Data, and (5) Dashboard and Alert Delivery. All layers communicate via RESTful APIs on Microsoft Azure.

### B. Data Sources (6 Streams)

- Academic Records: CAT scores, assignments, lab marks, end semester results.
- Attendance Logs: Attendance for each subject via RFID or biometric systems.
- Engagement with LMS: Login activity, resource use, video completion (Moodle).
- Assignment Metadata: Submission times, late submission counts, plagiarism scores.
- Reflective Journals: Weekly student text entries (100–200 words).
- Library & Lab Logs: Library visits, e-resource usage, lab utilization.

### C. Feature Engineering

A 47-dimensional feature vector is constructed for each student per week: Academic Features (18): GPA trend, score deviation, assignment completion. Behavioral Features (16): attendance percentage, LMS sessions, late submissions. Affective Features (13): NLP sentiment polarity, emotion category scores, cognitive load index. Features are MinMax normalized and missing data are handled by KNN imputation (k=5).

### D. ML Ensemble Prediction Engine

A stacked ensemble classifies students as Low, Moderate, or High Risk using the following models:

- Random Forest (200 trees): Captures non-linear feature interactions and provides feature importance scores.
- XGBoost (150 estimators, lr=0.05): Gradient boosting for high predictive accuracy.
- LSTM (128–64 units): Models 12-week temporal behavioral sequences.

A logistic regression meta-learner combines the three models' probabilistic outputs. The ensemble retrains every month on a 24-week sliding window.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### E. NLP Sentiment and Emotion Module

The NLP module employs a fine-tuned BERT model for binary sentiment polarity classification (positive/negative) of student journal entries. A multi-label RoBERTa model classifies emotions into categories such as anxiety, confidence, frustration, and engagement. XLM-RoBERTa handles code-switched (mixed-language) text common in Indian student writing. Outputs form the 13 affective features fed into the ensemble pipeline.

### F. Cloud Infrastructure

- Azure Event Hubs is used for real-time ingestion, capable of handling up to 40 MB/s throughput (Kafka-compatible).
- Azure Databricks with Apache Spark is used for feature engineering in both batch and streaming modes.
- Azure Machine Learning manages model training, versioning, and deployment pipelines.
- Azure Blob Storage stores raw data, engineered features, and model artifacts.
- Power BI Embedded delivers real-time dashboards for faculty and counselors.

## IV. RESULTS AND EVALUATION

### A. Dataset and Setup

The dataset covers 6 undergraduate departments (CE, IT, EXTC, Mechanical, Civil, MBA) across 2021–2024, comprising 2,400 students, 72 courses, 8.6 million event records, and 18,400 annotated journal entries. The dataset was split 70/15/15 for training, validation, and testing. Class imbalance was addressed using SMOTE (High risk: 15%, Moderate: 28%, Low: 57%).

### B. ML Model Performance

TABLE I: ML MODEL COMPARISON

| Model | Accuracy | F1 | AUC |
|---|---|---|---|
| Random Forest | 88.4% | 0.865 | 0.91 |
| XGBoost | 90.1% | 0.895 | 0.93 |
| LSTM | 91.3% | 0.910 | 0.94 |
| **Ensemble (Ours)** | **93.7%** | **0.935** | **0.97** |
| GPA Baseline | 74.2% | 0.715 | 0.78 |

The stacked ensemble achieves 93.7% accuracy, outperforming the GPA baseline by 19.5 percentage points. LSTM contributes the highest individual model accuracy (91.3%). Adding NLP affective features improves the ensemble F1 score by 4.2 points over the quantitative-only model.

### C. NLP Performance

Fine-tuned BERT achieves an F1 score of 0.895 for sentiment polarity classification, compared to VADER alone (F1: 0.74). The multi-label RoBERTa emotion classifier achieves F1: 0.865. XLM-RoBERTa for code-switched text achieves F1: 0.920.

### D. Early Warning and Intervention Outcomes

SAISPMS identifies at-risk students approximately 5.3 weeks before end-of-semester examinations — a significantly earlier warning than the traditional grade-based approach, which identifies at-risk students only about 1.25 weeks prior. This represents a 4.5× improvement in lead time.

A pilot intervention was conducted with 312 at-risk students during the 2023–24 academic year. Results showed that 74% of these students received counselor contact within three days of the alert. Approximately 61% of the intervention group showed improvement in final grades, compared to only 23% in the non-intervention control group.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## V. CONCLUSION

This paper presents SAISPMS — a cloud-native academic monitoring platform that integrates ensemble machine learning, NLP-based sentiment and emotion analysis, and Azure cloud Big Data analytics. SAISPMS achieves 93.7% prediction accuracy and provides early academic risk warnings 5.3 weeks before end-of-semester examinations. Intervention outcomes show a 61% grade improvement rate among at-risk students across a cohort of 2,400 undergraduates. SAISPMS substantially outperforms conventional academic monitoring approaches.

Future work will focus on integrating IndicBERT for regional Indian language support, connecting SAISPMS with wearable IoT devices for physiological stress monitoring, incorporating explainable AI (XAI) modules for transparent decision-making, developing a student-facing self-monitoring mobile application, and extending the platform to support multi-institution federated learning. SAISPMS demonstrates the transformative potential of combining Azure cloud infrastructure, Big Data analytics, ensemble machine learning, and NLP in the domain of higher education.

## REFERENCES

[1] C. Romero and S. Ventura, 'Educational data mining: A review,' IEEE Trans. Syst., vol. 40, no. 6, pp. 601–618, 2010.
[2] S. Hussain et al., 'Prediction of student performance using ML for MOOC platforms,' IEEE ICCD, 2018.
[3] S. Hochreiter and J. Schmidhuber, 'Long short-term memory,' Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[4] M. Wen et al., 'Sentiment analysis in MOOC discussion forums,' Proc. EDM, 2014, pp. 130–137.
[5] A. Ortigosa et al., 'Sentiment analysis in Facebook and e-learning,' Comput. Hum. Behav., vol. 31, pp. 527–541, 2014.
[6] H. Aldowah et al., 'Educational data mining for 21st century higher education,' Telemat. Inform., vol. 37, pp. 13–49, 2019.
[7] J. Devlin et al., 'BERT: Pre-training of deep bidirectional transformers,' NAACL-HLT, 2019, pp. 4171–4186.
[8] T. Chen and C. Guestrin, 'XGBoost: A scalable tree boosting system,' KDD, 2016, pp. 785–794.
[9] Govt. of India, 'National Education Policy 2020,' Ministry of Education, 2020.
[10] Microsoft Azure, 'Azure Machine Learning documentation,' 2024.


INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com