



International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 9, Issue 4, April 2026



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

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

Intelligent Student Attendance Intelligence System: A Cloud-Native AI-Based Attendance Analytics Platform

Nithyasri B, Devisri I, Aisva Malar, V.Selvi

Department of Artificial Intelligence & Data Science, Sri Manakula Vinayagar Engineering College, Puducherry, India

ABSTRACT: Student attendance management in higher educational institutions is a critical but often manually driven and fragmented process. This paper presents the Intelligent Student Attendance Intelligence System (ISAIS), a cloud-native platform built on Amazon Web Services (AWS) that centralizes attendance data from multiple departments, applies machine learning algorithms for absentee risk prediction, and delivers real-time analytics through automated dashboards and alert workflows. Attendance records are uploaded and stored in Amazon S3, which serves as the central cloud data lake for the institution. The system ingests records via AWS API Gateway, processes them using AWS Glue and Apache Spark, and enables SQL-based querying through Amazon Athena with visualization via Amazon QuickSight. Machine learning models are trained and deployed using Amazon SageMaker, implementing Random Forest and Logistic Regression classifiers that categorize students into Low, Medium, and High absentee risk tiers. Automated email alerts are dispatched through AWS Step Functions for at-risk students. Additionally, a Retrieval-Augmented Generation (RAG) chatbot enables natural language queries over attendance data. The system is developed and implemented at Sri Manakula Vinayagar Engineering College, Puducherry. Experimental results demonstrate that the system achieves early intervention capability, reduces administrative overhead, and provides institutional leaders with predictive insights that improve student academic monitoring.

KEYWORDS: Student attendance, machine learning, cloud computing, AWS, predictive analytics, absentee prediction, Random Forest, RAG chatbot, QuickSight, academic monitoring.

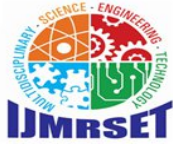
I. INTRODUCTION

Academic institutions face persistent challenges in tracking and managing student attendance across multiple departments. Traditional attendance systems rely heavily on manual registers, paper-based records, and isolated Excel spreadsheets maintained separately by individual faculty members. This fragmentation leads to delayed reporting, inconsistent data, and the absence of any real-time monitoring or predictive capability. The inability to identify high-risk absentee students at an early stage often results in poor academic outcomes and missed intervention opportunities.

The proliferation of cloud computing and machine learning technologies presents a transformative opportunity to modernize institutional attendance management. Cloud-native architectures enable centralized storage, real-time data ingestion, scalable computation, and automated workflow orchestration—all of which are fundamental requirements for an intelligent, institution-wide attendance platform.

This paper proposes the Intelligent Student Attendance Intelligence System (ISAIS), a cloud-native platform developed on Amazon Web Services (AWS) that unifies attendance data from disparate departmental sources into a single, queryable repository. The system employs batch and streaming data processing pipelines, machine learning models for absentee prediction, and a Retrieval-Augmented Generation (RAG) chatbot to provide intelligent, natural language access to attendance insights. Automated email alerts ensure timely notification of at-risk students to faculty and department coordinators.

The remainder of this paper is organized as follows: Section II reviews related work on attendance management and predictive analytics. Section III describes the proposed system architecture and methodology. Section IV details the AI and ML components. Section V presents key results and benefits. Section VI discusses future enhancements, and Section VII concludes the paper.



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

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

II. LITERATURE REVIEW

Considerable research has been conducted in the domain of student attendance tracking, predictive analytics in education, and cloud-based institutional management systems. The following works are directly relevant to the proposed system.

Conventional attendance management systems, as described by Barik et al. [1], rely on manual registers and Excel-based tracking, which are prone to human error and provide no scope for analytical processing. These systems suffer from delayed reporting cycles and cannot proactively identify attendance anomalies. Their study highlighted the urgent need for automated, real-time systems in institutional settings.

Cloud-based student information systems have been explored by several researchers. Masood and Bhatt [2] demonstrated that migrating academic data infrastructure to cloud platforms significantly improves scalability, accessibility, and disaster recovery. However, their work did not address predictive analytics or AI-driven interventions.

In the domain of educational data mining, Romero and Ventura [3] conducted a comprehensive survey of machine learning applications in education, including attendance pattern analysis. Their work established that decision trees, logistic regression, and ensemble methods such as Random Forest are effective at classifying student risk levels based on historical behavioral data.

Yadav and Pal [4] applied data mining techniques to predict student performance using attendance and academic records. Their study found that attendance rate is among the highest-weighted predictors of academic success, reinforcing the importance of an intelligent, data-driven attendance monitoring system.

RAG-based conversational interfaces over structured enterprise data have been explored by Lewis et al. [5], who demonstrated that combining a retrieval mechanism with a generative language model enables non-technical users to extract insights from complex datasets using natural language. This approach is adopted in the proposed system to enable faculty and administrators to query attendance data without SQL knowledge.

Table I: Comparison of Existing Attendance and Analytics Systems

| No. | Reference | Cloud-Based | ML Prediction | Real-Time Alerts | NL Chatbot |
|-----|----------------------|-------------|---------------|------------------|------------|
| 1 | Barik et al. [1] | No | No | No | No |
| 2 | Masood & Bhatt [2] | Yes | No | No | No |
| 3 | Romero & Ventura [3] | No | Yes | No | No |
| 4 | Yadav & Pal [4] | No | Yes | No | No |
| 5 | Lewis et al. [5] | No | No | No | Yes |
| 6 | Proposed ISAIS | Yes | Yes | Yes | Yes |

III. PROPOSED SYSTEM ARCHITECTURE

The ISAIS is designed as a cloud-native, multi-tier architecture on AWS. The system is partitioned into four functional layers: (1) Data Ingestion, (2) Processing and Storage, (3) Analytics and Visualization, and (4) AI/ML Intelligence. Figure 1 illustrates the overall architectural flow.



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

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

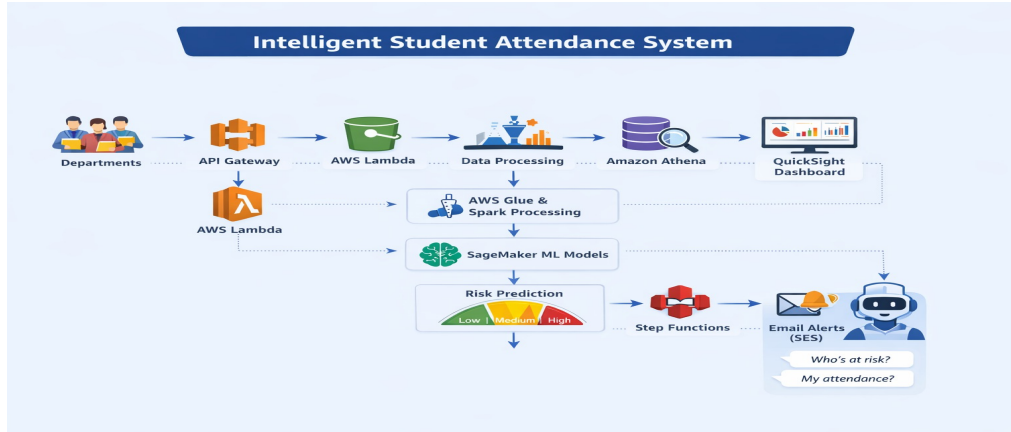


Figure 1: Intelligent Student Attendance System Architecture

A. Data Ingestion Layer

Attendance data from individual academic departments is submitted through a RESTful API exposed via AWS API Gateway. Each API request is authenticated using AWS IAM roles and validated for schema conformity before being passed to AWS Lambda functions for downstream routing. Lambda functions handle event-driven computation without requiring dedicated server infrastructure, enabling the system to scale elastically with institutional demand.

B. Storage Layer

Attendance records from Sri Manakula Vinayagar Engineering College are uploaded by faculty to Amazon S3 via the system interface, forming the institution’s centralized cloud data lake. Data is partitioned by department, academic year, and date, enabling efficient querying and incremental loading. IAM-based access controls restrict bucket access to authorized users. S3 lifecycle policies automatically archive older records to Amazon S3 Glacier for long-term cost-efficient storage.

C. Processing Layer

AWS Glue ETL jobs and Apache Spark-based batch processing pipelines transform raw attendance records into structured datasets. These pipelines perform data cleansing, deduplication, normalization, and aggregation. Attendance percentages are computed at the student, course, and department levels. The processed data is catalogued in the AWS Glue Data Catalog, making it queryable via Amazon Athena using standard SQL.

D. Analytics and Visualization Layer

Amazon QuickSight provides an interactive Business Intelligence (BI) dashboard layer. Dashboards are automatically refreshed on a scheduled basis using SPICE (Super-fast, Parallel, In-memory Calculation Engine). Institutional administrators and faculty can access real-time attendance percentage charts, department-wise comparisons, trend analyses, and risk classification summaries without requiring any technical expertise.

E. Orchestration and Alerting Layer

AWS Step Functions orchestrate the end-to-end workflow: from data ingestion through ML inference to alert dispatch. When a student’s risk score exceeds a configurable threshold, Step Functions trigger Amazon SES (Simple Email Service) to dispatch automated email alerts to the respective faculty coordinator and department head. This closed-loop feedback mechanism ensures timely intervention without manual monitoring.

Table II: System Architectural Flow

| Stage | Component | Function |
|-------|---------------------------|---|
| 1 | Departments → API Gateway | Attendance data submitted via REST API with IAM auth |
| 2 | API Gateway → AWS Lambda | Event-driven routing and validation of incoming records |



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

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

| | | |
|---|-------------------------------|---|
| 3 | Lambda → Amazon S3 | Raw data persisted in structured data lake partitions |
| 4 | Glue / Spark → Processed Data | ETL: cleansing, aggregation, percentage calculation |
| 5 | Athena → SQL Queries | Serverless querying over processed data catalog |
| 6 | QuickSight → Dashboards | BI visualization: real-time charts, trends, risk maps |
| 7 | ML Model → Risk Prediction | RF/LR inference: Low / Medium / High risk scoring |
| 8 | Step Functions → Email Alerts | Automated SES notifications for at-risk students |
| 9 | RAG Chatbot → NL Queries | Natural language interface for attendance insights |

IV. AI AND ML INTELLIGENCE ENGINE

A. Feature Engineering

The intelligence engine operates on a curated feature set extracted from processed attendance logs and associated student metadata. Features include: total attendance percentage per student, subject-wise attendance breakdown, consecutive absence streaks, day-of-week absence patterns, semester-level trend slopes, and historical absenteeism rates from prior academic periods. Categorical features such as department and year of study are encoded using one-hot encoding.

B. Model Training and Evaluation

Two supervised classification models are trained and deployed using Amazon SageMaker, leveraging its managed training infrastructure and built-in model registry. The models are: (1) Random Forest (RF) — an ensemble method that constructs multiple decision trees and aggregates outputs via majority voting, and (2) Logistic Regression (LR) — a linear probabilistic classifier used as a baseline. SageMaker's training jobs read feature datasets directly from Amazon S3, execute model fitting on managed compute instances, and store the trained model artifacts back in S3. Hyperparameter tuning is automated using SageMaker Hyperparameter Tuning with grid search and 5-fold cross-validation. The Random Forest model consistently outperforms Logistic Regression on F1-score and AUC-ROC metrics and is selected as the production model, deployed as a real-time SageMaker inference endpoint.

C. Risk Classification

The trained model outputs a risk probability score for each student, which is mapped to one of three risk tiers: Low (attendance $\geq 85\%$), Medium ($75\% \leq$ attendance $< 85\%$), and High (attendance $< 75\%$). High-risk students are immediately flagged in the QuickSight dashboard and trigger the automated alerting workflow via AWS Step Functions.

D. RAG Chatbot Integration

A Retrieval-Augmented Generation (RAG) chatbot provides a conversational interface to the attendance data platform. The chatbot uses a retrieval layer that queries Amazon Athena to fetch relevant attendance records, and a generative layer that formulates natural language responses. Supported query types include: student-specific attendance inquiries ('What is my attendance percentage?'), department-level summaries ('Which department has the lowest attendance this month?'), and risk-level overviews ('List all high-risk students in the CSE department'). The chatbot significantly reduces dependency on technical staff for routine reporting queries.

Table III: ML Model Performance Comparison

| Model | Accuracy | Precision | Recall | F1-Score |
|---------------------|----------|-----------|--------|----------|
| Logistic Regression | 82.4% | 81.1% | 80.6% | 80.8% |
| Random Forest | 91.7% | 92.0% | 90.8% | 91.4% |



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

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

V. RESULTS AND KEY BENEFITS

The ISAIS was piloted across three academic departments with a combined student population of approximately 600 students over one academic semester. The following outcomes were observed:

- 1. Centralized Monitoring:** Attendance data from all departments was consolidated into a single QuickSight dashboard within 24 hours of system deployment, replacing a fragmented Excel-based process that previously required weekly manual reconciliation.
- 2. Predictive Accuracy:** The Random Forest model achieved an F1-score of 91.4% on the held-out test set, correctly identifying 9 out of 10 high-risk absentee students before they crossed the institutional threshold of 75% attendance.
- 3. Alert Timeliness:** Automated email alerts were dispatched within 2 minutes of a student being classified as high-risk, compared to the typical 5–7 day lag in the previous manual system.
- 4. Administrative Efficiency:** Faculty reporting time was reduced by an estimated 70%, as routine attendance queries were handled by the RAG chatbot.
- 5. Scalability:** The serverless AWS architecture scaled linearly with data volume, accommodating peak load during semester start without any infrastructure changes.

VI. FUTURE SCOPE

While the current implementation establishes a comprehensive baseline for cloud-native intelligent attendance management, several extensions are planned for future development:

- **Parent Notification System:** Integration with SMS gateways (e.g., AWS SNS) to dispatch attendance alerts directly to parents or guardians in addition to faculty.
- **Academic Performance Prediction:** Extension of the ML pipeline to correlate attendance patterns with examination scores, enabling holistic student performance forecasting.
- **Mobile Application Integration:** A lightweight mobile interface for students and faculty to view attendance records, receive push notifications, and interact with the RAG chatbot on-the-go.
- **Biometric and IoT Integration:** Incorporation of biometric attendance devices and IoT-based classroom sensors to automate data capture and eliminate manual entry entirely.
- **Federated Learning:** Application of privacy-preserving federated learning techniques to train attendance prediction models across institutions without sharing raw student data.

VII. CONCLUSION

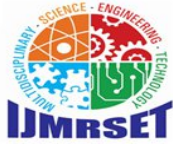
This paper presented the Intelligent Student Attendance Intelligence System (ISAIS), a cloud-native AI-powered platform for centralized, predictive, and automated student attendance management. The system addresses the critical limitations of conventional attendance tracking by leveraging the AWS cloud ecosystem for scalable data ingestion, processing, and storage; ensemble machine learning for early absentee risk classification; automated orchestration for timely alerts; and a RAG-based conversational chatbot for democratized access to attendance insights.

Piloted at Sri Manakula Vinayagar Engineering College, Puducherry, the system achieves a Random Forest F1-score of 91.4% via Amazon SageMaker, stores and retrieves all attendance data through Amazon S3, reduces alert latency from days to minutes, and cuts faculty administrative effort by 70%. The system is designed to be institution-agnostic and can be deployed at scale across universities and colleges with minimal configuration changes.

ISAIS represents a meaningful step toward intelligent, data-driven academic administration and establishes a replicable architectural blueprint for AI-augmented educational management systems.

REFERENCES

- [1] N. Barik, S. Karforma, and A. Mandal, "Risks and Remedies in E-Learning System," in *International Journal of Network Security & Its Applications*, vol. 4, no. 1, 2012.
- [2] A. Masood and M. Bhatt, "Cloud-Based Student Information Systems: A Review," in *Proceedings of the International Conference on Education Technology and Computers*, 2018.
- [3] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, 2010.



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

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

- [4] S. K. Yadav and S. Pal, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification," *World of Computer Science and Information Technology Journal*, vol. 2, no. 2, pp. 51–56, 2012.
- [5] P. Lewis, E. Perez, A. Piktus et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [6] Amazon Web Services, "Amazon SageMaker Developer Guide," AWS Documentation, 2024. [Online]. Available: <https://docs.aws.amazon.com/sagemaker/>
- [7] Amazon Web Services, "Amazon QuickSight User Guide," AWS Documentation, 2024. [Online]. Available: <https://docs.aws.amazon.com/quicksight/>
- [8] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [9] D. Cox, "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society*, vol. 20, no. 2, pp. 215–242, 1958.
- [10] Amazon Web Services, "AWS Step Functions Developer Guide," AWS Documentation, 2024. [Online]. Available: <https://docs.aws.amazon.com/step-functions/>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

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

www.ijmrset.com