International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

ScholarShield AI – Safeguarding Student's Academic Futures

Ayan S. Sayyad[1], Om S. Natekar[2], Harsh R. Sangar[3], Sarthak A. Kumbhar[4], Sumit B. Bhanushali[5]
*[1]Professor, Department of Computer Science And Engineering,
Nanasaheb Mahadik College of Engineering*
*[2,3,4,5]Students, Department of Computer Science and Engineering,
Nanasaheb Mahadik College of Engineering*
*Emails: ayansayyad@gmail.com, natekarom1@gmail.com , harshsangar44@gmail.com ,
kumbsarthak04@gmail.com , bhanushalisumit5040@gmail.com*

---***---

Abstract - This project aims to address the growing challenge of student dropouts in educational institutions through the use of artificial intelligence, data analytics, and predictive modeling. It tackles key issues such as the lack of early identification mechanisms, manual monitoring of student performance, and limited data-driven intervention strategies. By collecting and analyzing student-related data including attendance, academic performance, behavioral indicators, and socio-economic factors, the system predicts dropout risk at an early stage and categorizes students into risk levels. The proposed system provides actionable insights and intervention recommendations to teachers and academic administrators, enabling timely support for at-risk students. ScholarSheild AI features an intuitive web-based dashboard that visualizes risk trends, individual student profiles, and institutional analytics. Built using a modern technology stack comprising React, Flask, and machine learning models, the system integrates real-world educational data to support informed decision-making. This paper presents the project's background, objectives, methodology, system architecture, and its strong potential to improve student retention, academic outcomes, and institutional effectiveness in contemporary educational environments.

Keywords: Context-Aware Computing, Smart Scheduling, Productivity Analyt- ics, Generative AI, Model Context Protocol.

1. INTRODUCTION

In today's data-driven educational ecosystem, student success is no longer determined solely by academic performance, but by a combination of behavioral patterns, engagement levels, and timely institutional support. However, many educational institutions continue to face significant challenges in monitoring student progress and identifying early warning signs of potential dropouts. Traditional approaches rely heavily on manual record keeping, isolated data sources, and periodic evaluations, which often delay critical interventions. To address these challenges, ScholarShield AI has been developed as an intelligent, data-centric student retention platform. The system leverages machine learning and educational data analytics to predict dropout risk at an early stage and categorize students based on risk severity. By integrating multiple student data parameters into a single analytical framework, ScholarShield AI provides educators and administrators with actionable insights and evidence-based intervention recommendations. This automated and predictive approach enhances decision-making, supports timely interventions, and strengthens student retention efforts within modern educational environments.

2. OBJECTIVES

The primary objectives of the project are meticulously defined to address key challenges in student retention and early dropout identification:

- Unified Student Insight Platform:
To develop a single integrated interface that connects student academic records, attendance data, behavioral indicators, and performance metrics. This minimizes manual effort and enables educators to easily access all relevant student information in one place.

- Automated Risk Intelligence:
To incorporate machine learning–based predictive mechanisms that analyze student data and generate immediate risk assessments, allowing early identification of potential dropouts without manual evaluation.

- Intelligent Alerts and Recommendations:
To facilitate automated alerts, risk notifications, and intervention suggestions for teachers and administrators, enabling timely actions to support at-risk students and improve retention outcomes.

- Data Security:
To ensure a high level of data security through role- based access control and secure authentication mechanisms, guaranteeing that sensitive student information remains protected.

- Reliability:
To implement the system in a dependable deployment environment that ensures high availability, consistent performance, and reliability for real-time academic monitoring and analysis.

3. PROBLEM DEFINITION

The Traditional student monitoring systems used in educational institutions are often plagued by inefficiencies. They function as static record-keeping tools that store attendance and academic data but lack the intelligence to interpret patterns or predict student outcomes. This results in manual evaluation, disconnected data sources, and the absence of early warning insights. For instance, a system may record a student's low attendance, but it does not identify why the attendance is declining, which subjects are affected, or what intervention is required at the right time. These limitations hinder timely support for students who may be gradually disengaging due to academic, personal, or socio-economic challenges. The ScholarShield AI system is designed to overcome these issues by automating data analysis, identifying dropout risk through predictive modeling, and providing actionable insights to educators. By integrating multiple student data points into a unified framework, the system enhances early intervention, improves decision making, and strengthens communication between educators and academic administrators.

4. PROPOSED SOLUTION

To address the limitations of manual monitoring and delayed interventions, the **ScholarShield AI** system proposes a proactive, intelligence-driven web platform designed to predict student dropout risks before they become critical. Unlike traditional systems that function merely as data repositories, the proposed solution integrates **Predictive Analytics** and **Machine Learning** to transform raw academic and behavioral data into actionable insights.

The solution is architected as a modern web application using a **React-based frontend** for intuitive visualization and a **Flask-based backend** for robust data processing. The system operates by aggregating diverse data points—including attendance logs, academic grades, and socio-economic factors—and processing them through a trained machine learning model. This model classifies students into distinct risk categories (e.g., High, Medium, Low), allowing educators to prioritize their attention effectively. Furthermore, the system moves beyond simple prediction by generating specific **intervention recommendations**, thereby equipping academic administrators with the tools necessary to implement timely, evidence-based support strategies that directly improve student retention rates.

5. LITERATURE REVIEW

1. "Early Warning Systems in Education" – B. Allensworth et al. (2009)
This study investigates how predictive analytics can identify at-risk students in schools. It demonstrates that early identification of warning signs such as declining attendance and grades significantly improves intervention outcomes.
This supports ScholarShield AI's goal of early dropout risk prediction [1]."The Organized Mind" – D. J. Levitin (2015)

2. "Learning Analytics in Higher Education" – G. Siemens (2013)
This research explores the use of data-driven approaches to monitor student engagement and performance. It emphasizes the need for automated systems to handle large datasets and generate actionable insights, validating the project's approach to integrating multiple student data points into a predictive model [2].

3. "Student Engagement and Retention" – T. Tinto (1993)
This study examines how student engagement and institutional support affect retention rates. It shows that fragmented monitoring and delayed interventions lead to higher dropout rates. This reinforces the need for ScholarShield AI's unified dashboard that consolidates academic, attendance, and behavioral data for timely intervention [3].

4. "Predictive Modeling for Student Success" – J. Arnold and
K. Pistilli (2012)
This research illustrates how predictive algorithms can forecast academic risk and provide personalized recommendations. By automating the detection of at the risk students, ScholarShield AI reduces manual monitoring efforts and supports proactive decision-making [4].

5. "Improving Academic Outcomes Through Data-Driven Interventions" – L. Dawson et al. (2014)
This study highlights the economic and social benefits of structured intervention strategies. It emphasizes that timely support increases student retention and academic success. This aligns with ScholarShield AI's core feature of actionable intervention recommendations [5].

6. METHODOLOGY

The development of the ScholarShield AI system follows a structured Data Science Lifecycle integrated with Agile software development practices. This approach ensures that the predictive models are accurate and that the web application is user-friendly and responsive. The methodology is executed in four distinct phases:

6.1 Data Collection and Ingestion The foundation of the system lies in the acquisition of high-quality historical data. Data is aggregated from institutional databases and digital records, focusing on variables known to influence student retention.

- **Academic Records:** Term-end grades, assignment scores, and backlog history.
- **Attendance Logs:** Daily and monthly attendance percentages.
- **Demographics:** Socio-economic status, distance from school, and family background.
- **Behavioral Indicators:** Participation in extracurricular activities and disciplinary records.

6.2 Data Preprocessing and Feature Engineering Raw educational data often contains noise and inconsistencies. In this phase, the data is cleaned to handle missing values and outliers. Numerical features, such as attendance, are normalized to a standard scale, while categorical variables (e.g., gender, socio-economic bracket) are converted into numerical formats using encoding techniques. Feature engineering is then applied to select the most impactful attributes that contribute to the dropout prediction, reducing model complexity and improving performance.

6.3 Model Development and Training Various supervised machine learning algorithms specifically classification models such as Logistic Regression, Random Forest, and Decision Trees are employed to train the system. The dataset is split into training and testing sets to evaluate model accuracy. The model learns to identify patterns and correlations between the input features (e.g., low attendance + low grades) and the target variable (Dropout vs. Retention). The algorithm with the highest accuracy and recall rate is selected for deployment.

6.4 System Integration and Deployment The final phase involves integrating the trained model into the web application. A **Flask (Python)** backend serves as the API provider, receiving user inputs from the **React.js** frontend and passing them to the model. The model processes the data and returns a risk probability score, which is then visualized on the dashboard. This modular integration allows for real-time risk assessment and ensures the system is scalable for future data updates.

Key Points from Methodology

- **Diverse Data Sources:** Integration of academic, attendance, and socio-economic data for a holistic view of the student.
- **Data Cleaning:** Implementation of rigorous preprocessing to handle missing values and normalize data for accurate analysis.
- **Supervised Learning:** Utilization of classification algorithms (e.g., Random Forest, Logistic Regression) to predict binary outcomes (At-Risk vs. Safe).
- **Feature Selection:** Identification of critical factors influencing dropouts to optimize model efficiency.
- **RESTful API Integration:** Seamless connection between the ML model and the user interface using Flask APIs.
- **Real-time Visualization:** Instant rendering of risk scores and trends on a React-based dashboard for immediate educator insight.

7. PROPOSED SYSTEM

The ScholarShield AI system is developed using a systematic, phased approach built on a strong conceptual foundation to ensure comprehensive and effective implementation. It ensures secure, scalable, and efficient handling of student data across multiple modules. The proposed system includes the following steps:

• Secure Authentication: Students' and administrators' access is controlled via secure role-based authentication, ensuring that sensitive educational data is accessed only with explicit permissions.

• Data Collection: The system retrieves student-related data, including attendance records, grades, behavioral indicators, and socio-economic information from institutional databases and management systems.

• AI Processing: The aggregated data is passed to machine learning models that assess dropout risk and categorize students into risk levels. The AI also identifies contributing factors and generates actionable intervention recommendations for each student.

• Result Display: The dashboard updates dynamically, presenting risk scores, intervention suggestions, and analytics visualizations. Educators and administrators can access individual student profiles and overall trends in real time to support informed decision-making.

7.1. SYSTEM ARCHITECTURE

The system relies on a serverless architecture to handle asynchronous requests efficiently. The following architecture outlines the overall working of the system:

Fig 7.1: System Architecture

Key Components:

• User Client: The frontend is built on Next.js, providing a responsive in- terface. The user initiates the process by logging in.

• Database: Registration data, user preferences, and access tokens are encrypted and stored in the Firestore Database (NoSQL).

• Backend API: The Backend services run on Firebase Functions. These functions act as the "glue" between the user interface and external APIs.

• Context Engine: This component handles the logic for the MCP, determining which keywords to search for in the user's connected accounts.

7.2 DATA FLOW DIAGAM

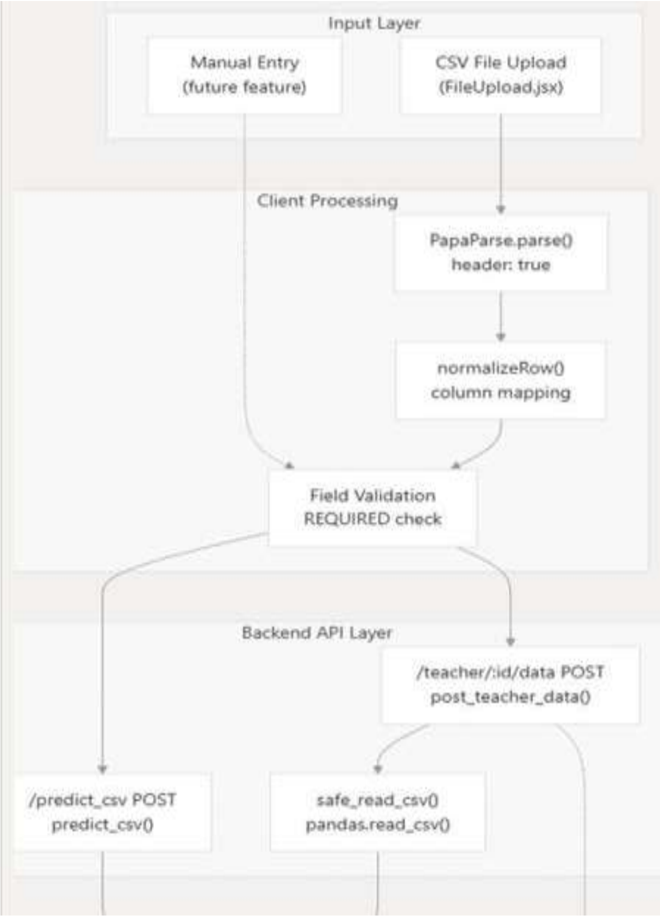

Fig. 7.2: Data Flow Diagram

The flow of data ensures that sensitive student information is processed securely and used only for analysis and decision support without being permanently stored in the predictive model.

Explanation of Flow:
• Step 1: Authorized users (teachers or administrators) log in to the system using secure authentication credentials and role based access control.
• Step 2: The system retrieves student-related data such as attendance records, academic performance, and behavioral indicators from institutional databases or uploaded datasets.
• Step 3: The processed data is passed to the machine learning engine, where dropout risk prediction and risk categorization are performed.
• Step 4: The generated risk scores, insights, and intervention recommendations are sent back to the frontend and rendered dynamically on the educator's dashboard for real-time monitoring and decision-making.

7.3 USE CASE DIAGRAM

The system is designed to serve multiple actors, primarily the End User and the System Admin.

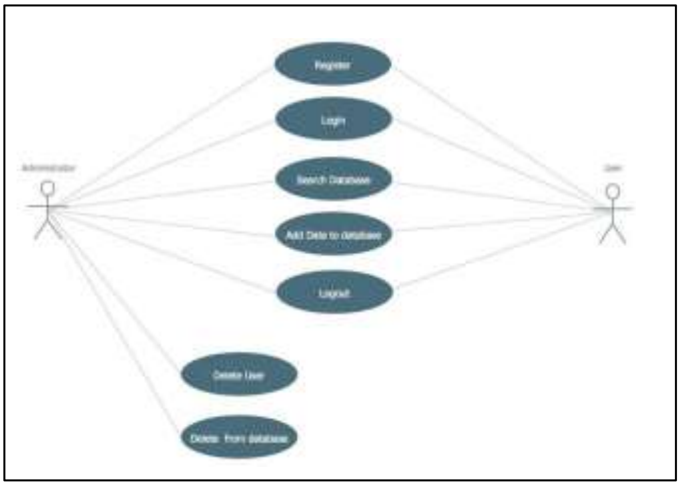

Fig. 7.3: Use Case Diagram

This diagram illustrates the interactions between the actors and the system com- ponents.
Key use cases include ”Login”, ”View Schedule”, ”Generate Context Summary”, ”Sync Tasks”, and ”View Analytics.”

8. RESULTS

Fig. 8.1: Platform Landing Page

Platform Landing Interface: Serves as the central entry point, clearly defining the system as an "AI-Powered Student Retention Platform" with intuitive navigation.
This is the main landing page of the ScholarShield AI system. It features a clean and intuitive interface designed to introduce users to the platform and encourage secure access. The primary call-to-action is the “Get Started” button, which directs users to the authentication process. The page highlights the system’s core objective of identifying at-risk students through AI-driven analytics and supporting educators with actionable insights. It serves as a secure and professional entry point to the application.

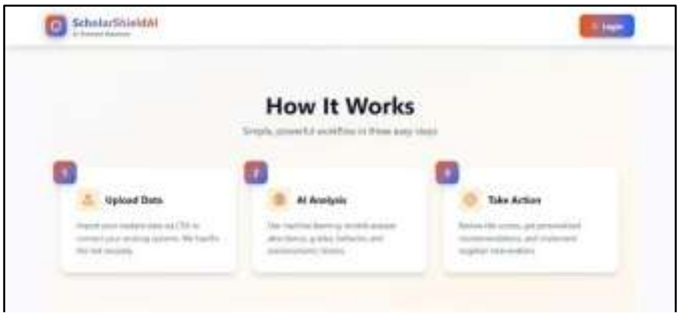

Fig. 8.2: Tutorial Page

Operational Workflow: Visualizes the simplified three- step process—Upload Data, AI Analysis, and Take Action—making the system easy to understand for non-technical users. The workflow diagram explains the system operation in three simple steps:

1. Upload Data – Student data is uploaded via CSV or connected systems.
2. AI Analysis – Machine learning models analyze academic, behavioral, and socio-economic factors.
3. Take Action – Educators review insights and implement personalized interventions.

This structured workflow ensures ease of use and effective adoption by educational institutions.

Fig. 8.3: Login Page

Secure Authentication: A dedicated login portal ensuring role-based access control (Teacher/Principal) to protect sensitive student data.

The Login Page allows authorized users such as teachers or administrators to securely access the system. It contains input fields for username and password along with validation mechanisms to ensure secure authentication.

The page supports a demo login mode, enabling quick access for testing and demonstration purposes. Its minimalistic layout enhances usability while maintaining security standards. Successful authentication redirects the user to the main dashboard.

Fig. 8.4: Teacher Dashboard

Teacher Dashboard: Provides educators with a classroom-level snapshot, immediately highlighting the "Top At-Risk Students" and summarizing class statistics (e.g., 64 High Risk students).

The Teacher Dashboard acts as the central control panel of the system. It provides a summarized view of student risk analytics using visual cards and indicators. The dashboard displays the total number of students analyzed along with categorized dropout risk levels such as High Risk, Medium Risk, and Low Risk.

A refresh option enables real-time updates of data. This dashboard empowers educators to quickly identify problem areas and prioritize intervention strategies.

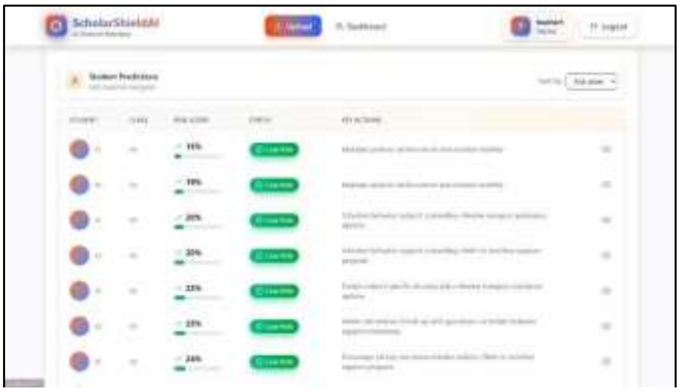

Fig. 8.5: Student Prediction

Cohort Monitoring List: Displays a sortable list of students with their calculated risk scores (e.g., 15% Low Risk) and assigned "Key Actions" for continuous monitoring.

The Student Prediction Table provides a detailed list of individual students along with their risk scores, risk status, and recommended actions. The table supports sorting based on risk level, enabling teachers to focus on high-priority cases. Each row represents a student record analyzed by the AI model, ensuring transparency and interpretability of predictions.

Fig. 8.6: Individual Information & Profile

Individual Risk Profiling: A detailed student profile view that correlates specific metrics—such as 57.8% attendance and family background—to a precise "High Risk" score of 78%.
This screen presents a detailed profile of a selected student. It includes academic and socio-economic parameters such as attendance percentage, grade average, failed subjects, family income, parental education, transport availability, health conditions, and extracurricular involvement.
A dropout risk score is prominently displayed along with a visual progress indicator. This holistic view helps educators understand the underlying factors contributing to student risk.

Fig. 8.7: AI-Powered Recommendation

AI-Powered Recommendations: Translates risk data into actionable advice, offering specific interventions like "Enroll in remedial tutoring" or "Initiate attendance follow-up".
The Student Prediction Table provides a detailed list of individual students along with their risk scores, risk status, and recommended actions. The table supports sorting based on risk level, enabling teachers to focus on high-priority cases. Each row represents a student record analyzed by the AI model, ensuring transparency and interpretability of predictions.

Fig. 8.8: Principal Dashboard

Principal Dashboard: Offers institutional-level analytics aggregating data from 1,204 students, visualizing the overall risk distribution (11% High Risk) to aid administrative decision-making.
This section visually presents the risk distribution among students using numerical indicators and progress bars. Each risk category is color-coded for immediate recognition.
The summary helps educators understand the overall academic health of their class or institution and supports data-driven decision-making by highlighting critical risk segments.

9. SOFTWARE & HARDWARE REQUIREMENTS

To ensure optimal performance and development efficiency, the following requirements were established:

Hardware Requirements

- Intel® Core™ i3/i5 Processor or equivalent for development.
- 8GB RAM (minimum) to handle the local development server and API testing tools.
- Color Monitor, Keyboard, Mouse for interface design.
- High-speed Internet Connectivity for real-time API communication.

Software Requirements

- Operating System: Windows 10 or higher / macOS / Linux.
- Backend Runtime: Python 3.10+ for executing Flask based server-side.
- Cloud Platform: Firebase Functions (Serverless).
- Frontend Framework: ReactJS with Next.js for server side rendering.
- Database: Firestore (NoSQL) for flexible schema management.
- AI Model: scikit-learn and Python ML libraries .
- IDE: Visual Studio Code or PyCharm with linting and code formatting tools for efficient development.

10. TECHNOLOGY USED

React JS: A JavaScript library used for building dynamic and responsive user interfaces. It allows the creation of interactive dashboards and ensures smooth performance even with large student datasets.

Flask: A comprehensive backend-as-a-service platform used to store user data, manage authentication, and run serverless functions. It allows for real-time data synchronization between the database and the client.

Scikit-learn: A Python library used for implementing machine learning models, including classification and risk prediction algorithms. It enables reliable dropout risk scoring and early warning generation.

Tailwind CSS: A utility-first CSS framework used to design the UI quickly and consistently, ensuring that dashboards and reports are mobile-responsive and visually coherent across devices.

11. ADVANTAGES

The ScholarShield AI system offers a range of transformative benefits for educators, administrators, and institutions.

Key advantages include:

• Real-Time Student Monitoring: Unlike static record systems, this platform allows educators to access updated attendance, academic performance, and behavioral indicators at any time. This ensures that student information is always current and actionable.

• AI-Based Personalized Insights: The system generates automated, personalized risk assessments and intervention recommendations for each student. It filters out irrelevant data and highlights the most critical indicators that require attention.

• Early Warning System: Using predictive models and logic-based rules, the system identifies students at risk of dropping out or falling behind academically, allowing educators to intervene proactively before issues escalate.

• Administrative Support: Tasks such as generating student risk reports, summarizing key trends, and producing intervention suggestions are automated. This significantly reduces the manual administrative workload and allows educators to focus on direct student support.

• Secure Data Management: Role-based access and secure authentication ensure the privacy and protection of sensitive student information. All data is processed securely, and no personal data is stored beyond what is necessary for analysis and intervention.

12. Applications:

This system serves a wide array of practical purposes within educational institutions:

- Student Monitoring: Enables educators and administrators to instantly access all relevant academic, attendance, and behavioral information for each student.
- Early Intervention: Provides timely alerts and recommendations to address at risk students before disengagement escalates.
- Institutional Analytics: Offers dashboards and visualizations for tracking overall trends, class performance, and dropout risk across the institution.
- Resource Allocation: Supports administrators in efficiently allocating support resources (mentoring, counseling, remedial programs) based on predictive insights.

13. Future Enhancement/ Scope:

The ScholarShield AI system opens up a broad scope for future enhancements to further benefit educational institutions:

• Enhanced Data Sources: Future versions may integrate additional data streams, such as learning management system (LMS) interactions, online quiz results, and extracurricular activity participation, to provide a more comprehensive risk assessment.

• Predictive Intervention Modeling: Advanced AI models could recommend personalized intervention strategies based on historical success rates, tailoring support plans for individual students.

• Natural Language Processing for Reports: Future iterations may automatically analyze teacher comments, feedback, and student essays to identify early signs of disengagement or academic struggle.

• Mobile and Real-Time Alerts: Integration with mobile platforms could provide educators and administrators with real-time notifications about at the risk students, ensuring timely intervention regardless of location.

14. Conclusion

The ScholarShield AI system represents a significant advancement in educational technology by automating student monitoring, enhancing contextual understanding, and providing access to real-time analytics. This intelligent platform effectively addresses persistent challenges in traditional student management, such as fragmented record-keeping, manual identification of at-risk students, and the absence of proactive, data-driven insights. By leveraging machine learning models and data integration techniques, the system provides a comprehensive framework that delivers actionable intervention recommendations, visualizes risk trends, and simplifies administrative decision making.

ACKNOWLEDGEMENT

The authors sincerely thank the Department of Computer Science & Engineering, Nanasaheb Mahadik College of Engineering, for providing guidance and support throughout this research work.

REFERENCES

[1] B. Allensworth and L. Easton, "What Matters for Staying On-Track and Graduating in Chicago Public Schools," Consortium on Chicago School Research, 2009.G. Siemens, "Learning Analytics: The Emergence of a Discipline," American Behavioral Scientist, vol. 57, no. 10, pp. 1380–1400, 2013.

[2] T. Tinto, Leaving College: Rethinking the Causes and Cures of Student Attrition, 2nd ed., University of Chicago Press, 1993.

[3] J. Arnold and K. Pistilli, "Course Signals at Purdue: Using Learning Analytics to Increase Student Success," Journal of Learning Analytics, vol. 1, no. 1, pp. 2–17, 2012.

[4] L. Dawson et al., "Educational Data Mining and Learning Analytics for Student Retention," Journal of Learning Analytics, vol. 1, no. 3, pp. 59–82, 2014.

[5] scikit-learn Documentation: https://scikit learn.org/stable/documentation.html

[6] React Documentation: https://reactjs.org/docs/getting-started.html

[7] Flask Documentation: https://flask.palletsprojects.com/en/2.3.x

BIOGRAPHIES

Ayan S. Sayyad
He is a Professor in the Department of Computer Science & Engineering at Nanasaheb Mahadik College of Engineering, Peth, Sangli. His research interests include Artificial Intelligence, Machine Learning, and Smart Working AI Models. He has guided multiple undergraduate projects and actively contributes to academic research and innovation activities.