DEVELOPMENT OF AN INTELLIGENT ATTENDANCE CLOCKING SYSTEM LEVERAGING MACHINE LEARNING USING ADAPTIVE BOOSTING (ADABOOST) TECHNIQUES

BY

Kazeem Olarewaju Nasirudeen & Sheu Ahmad Bamidele
Department of Computer Science, Faculty of Computing and Engineering Technology, Al-Hikmah University, Ilorin, Nigeria

Abstract

The growing demand for accuracy and automation in attendance monitoring systems, especially in educational and corporate institutions, has driven the need for intelligent solutions that can minimize human errors, detect anomalies, and ensure timely record-keeping. This study presents the development and implementation of an intelligent attendance clocking system using the Adaptive Boosting (AdaBoost) algorithm which is a powerful ensemble machine learning technique. The system processes a range of features including timestamp logs, user identifiers, and historical attendance behavior to classify attendance status into categories such as Present, Late, or Absent. The resulting system offers a scalable, intelligent, and automated framework for real-time attendance tracking, capable of reducing manual intervention and enhancing data integrity. Its success highlights the effectiveness of ensemble learning, specifically AdaBoost, in classification tasks within time-sensitive administrative domains. This approach can be extended to support anomaly detection, fraud prevention, and predictive scheduling in future iterations. The model was trained and evaluated using real-world datasets, achieving improved accuracy compared to traditional methods. The implementation supports real-time clock-ins, secure data handling, and efficient user management. The outcome demonstrates that machine learning, particularly AdaBoost, can significantly optimize attendance systems, making them more intelligent, scalable, and resilient to manipulation

Keywords: Attendance monitoring systems, AdaBoost, Anomaly detection, Real-time clock-ins, Optimize attendance systems, Machine Learning.

Introduction

Background of the study

Attendance monitoring is an essential aspect of academic institutions, serving as a key factor in evaluating student commitment, engagement, and academic performance. Traditional methods such as manual signing or biometric systems have been widely adopted; however, these approaches often suffer from limitations including time consumption, human error, impersonation, and infrastructure dependency (Adewole et al., 2020). It is very important that attendance data is kept securely and co-ordinately to avoid Problems in recording and analysing attendance data. The factors that are dependent variable of the value of attendance data are accuracy, speed, consistency and accessibility of the data (Erdoğan Alper, 2019).

The use of machine learning algorithms enables the system to continuously learn and improve its accuracy over time. One of the key advantages of an attendance system is the ability to reduce human error and increase the speed and accuracy of attendance recording. Additionally, the system can be customized to meet specific needs, such as incorporating additional security measures. Machine learning-based attendance systems offer several benefits. They

eliminate manual errors, reduce the time and effort required for attendance management, and provide real-time attendance (Kotramma, et al, 2023).

Adaptive Boosting, or the AdaBoost algorithm, is a supervised learning model commonly used for data classification by combining multiple weak learners to produce a strong learner. This approach is particularly effective in detecting fraudulent activities in potential transaction fraud by evaluating and assessing differences between normal transactions and fraudulent ones through pattern learning. Machine learning algorithms, such as AdaBoost, have shown promising results in fraud detection (Christian, et al, 2024). AdaBoost faces challenges in generalizability due to overfitting on biased weights that tend to focus on training data noise, which impacts the overall effectiveness of the model when given unseen data.Given that AdaBoost has its limitations when it comes to classification capabilities, the researchers found a method to help AdaBoost and enhance its overall performance by using these methods. SMOTE(Synthetic Minority Over-sampling Technique) is a method used to address class imbalance in datasets by generating synthetic examples for the minority class (Christian, et al, 2024).

Jaimini, et al. (2025) present study aim at increasing the prediction rate of miscarriage by way of enhancing the AdaBoost algorithm with Oblique Decision Trees as the base classifier. On the contrary, the traditional boosting methods like AdaBoost and XGBoost uses the decision-trees as base classifiers generally apply axis-aligned approaches for the data split into classes by single features. Therefore, they may have reduced ability to handle complex, non-linear relationships buried inside the data. Machine Learning, a subset of Artificial Intelligence, enables systems to learn and improve from experience without explicit programming. By harnessing the potential of ML, Automatic Attendance Monitoring Systems can accurately and efficiently identify students or individuals present in a given setting, such as a classroom, workplace, or event (Dayananda, et al, 2025).

Machine Learning (ML), have emerged as powerful tools for automating attendance monitoring systems. Automatic Attendance Monitoring Systems using ML leverage advanced algorithms and image processing techniques to identify and verify individuals, transforming the way attendance is managed in educational institutions and beyond, the core technology behind this innovative system is facial recognition (Dayananda, et al, 2025). To maintain track of each student's attendance, it takes time to call out their name at the beginning of the course. False signs, names missing from spread sheets, manually inputting data into systems, and the possibility of proxy attendance are further problems. Such techniques have a few problems that have grown over time (Bawar, 2022).

Statement of the Problem

Despite the availability of biometric or RFID-based systems, many still lack intelligent capabilities to detect anomalies, adapt to diverse user behaviors, or identify fraudulent patterns such as impersonation. Moreover, these systems often fail to scale well or provide real-time analytical feedback to administrators. To address these shortcomings, there is a pressing need for a more intelligent, automated, and accurate attendance clocking system that can learn and improve over time. This study proposes the development of an intelligent attendance clocking system leveraging machine learning techniques using Adaptive Boosting (AdaBoost). The AdaBoost algorithm, known for its high accuracy and ability to combine multiple weak learners into a strong classifier, can enhance the classification and verification of attendance data by detecting

Aim and Objectives of the Study

The aim of this study is to develop an intelligent, efficient, and accurate attendance clocking system that leverages machine learning, specifically the Adaptive Boosting (AdaBoost) algorithm, to automate attendance verification, detect anomalies, and prevent fraudulent clock-ins. While the specific objectives are to:

i. Collect and preprocess attendance-related data including timestamps, user identities, and behavioral patterns for effective model training.
ii. Design and implement an automated attendance clocking system that minimizes human intervention and reduces errors associated with manual methods.
iii. Integrate the machine learning model into a user-friendly system interface for real-time attendance monitoring and reporting.
iv. Train and evaluate the AdaBoost-based model for robust prediction and classification of valid and invalid attendance records.

Literature Review

Overview of an existing Attendance Systems.

The evolution of attendance management systems has transitioned from manual registers to biometric and smart card technologies, aiming to increase accuracy and efficiency. However, despite technological advancements, many systems remain vulnerable to challenges such as proxy attendance, hardware limitations, and lack of intelligent decision-making capabilities.

Traditional Attendance Systems

Manual methods (e.g., paper sheets or punch cards) are still used in many organizations but are prone to human error, time inefficiency, and manipulation. (Babatunde et al., 2022) emphasized the unreliability and administrative burden of manual systems, which necessitate automation. Consequently, technological advancements have introduced biometric and non-biometric systems to address these challenges.

Biometric and RFID-based Systems

Biometric systems using fingerprint, and Radio Frequency Identification (RFID) technologies have improved authentication but often struggle with environmental factors and impersonation. RFID systems, although faster, can be cloned or misplaced.

QR-Code, Finger Print and Face Recognition

(Quick Response Code) QR code-based systems use unique codes generated for each user to mark attendance or verify identity. During implementation, each user is assigned a scan able QR code linked to their profile in the system database. When scanned using a camera or mobile device, the code is decoded, and attendance is recorded automatically. This method is cost-effective and easy to deploy but may be vulnerable to misuse if codes are shared. Fingerprint recognition systems use biometric scanners to capture and match the unique fingerprint patterns of users and face recognition systems use computer vision and machine learning to identify users based on their facial features. Attendance management system is an easy-to-use smart system based on (Face recognition, Fingerprint,

QR code) to record the attendance and departure of all students and employees on daily, weekly and monthly basis (Nesma, et al, 2022).

Related Works

(Abhishek, et al, 2024) explores biometrics attendance system which is the measure of biological or behavioural features which are used for the identification of individuals. biometric authentication (or realistic authentication) is used in computer science as a form of

identification and access control, such as Physiological biometrics is based on a behavioral trait of an individual. It involves all physical characteristics like ears, eyes, iris, fingerprints and behavioral biometrics is the scientific study of how people and animals' bodies function. biometric systems use unique physical or behavioral characteristics, such as fingerprints, iris patterns, or facial features, for individual identification and authentication

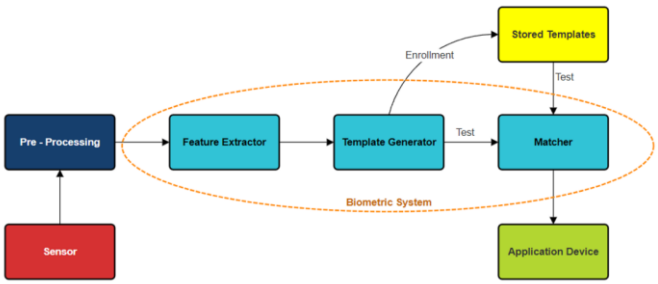

Fig 1: Generalised Architecture of Biometric System

Abhishek, et al. (2024) also discuss in the early 19th century, companies started adopting methods to mark the attendance of their employees. This involved using tools like time clocks and written registers to document when the employees started and finished their work shifts. In the late 19th century, the introduction of punch cards revolutionized attendance marking.

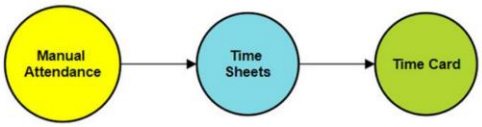

Fig 2 : Evolution of Manual Assessment

The system utilizes computer vision and deep learning techniques to recognize and verify the identity of individuals in a given environment. Through the deployment of facial recognition algorithms, it captures and matches facial features with existing records in a database, allowing for seamless and efficient attendance tracking.

Vidyanagar (2024) discussed Network-Based Smart Attendance System, where in students' mobile devices and the institutional WiFi or local network can be used to facilitate the attendance process. In this digital solution, cases of attendance by proxy and errors in recording will fall drastically. More importantly, the record-keeping is

burdensome to the faculty, which can be reduced, making the attendance system manageable and efficient to handle. A design of the system along with its implementation will improve the entire attendance management procedure in an educational institution. This system reduces the hours, paper consumption, and labour by the faculties and administration in handling attendance. By simply verifying the attendance ID within a specified time frame, Attendance data is securely stored and easily managed, allowing faculty to view and modify records from their designated dashboards.

Frameworks

These are architectural or conceptual designs used to build or guide the development of intelligent attendance systems, especially those integrating biometrics, QR codes, or machine learning (like AdaBoost).Attendance System Frameworks and Their Discussion

1. **Biometric-Based Framework**

Uses fingerprint, face, iris, or voice recognition to verify user identity. It enhances security and accuracy; ideal for preventing proxy attendance. Face and fingerprint are the most commonly used in schools and offices.

2. **QR Code-Based Framework**

This generates and scans QR codes unique to each user.Cost-effective, easy to implement, but can be vulnerable to screenshot-based fraud if not validated in real time.

3. **RFID-Based Framework**

Uses RFID tags/cards scanned via readers at entry points. It's fast and reliable for physical tracking but lacks biometric verification, hence more prone to proxy attendance.

4. **Mobile App-Based Framework**

Users mark attendance via a mobile app using GPS, camera, or QR scan. It's Ideal for remote teams or students. GPS or facial verification can be integrated to ensure authenticity.

5. **Cloud-Based Attendance Framework**

All data processing and storage occur in the cloud (e.g., Firebase, AWS). It enables real-time data access, auto-syncing, and scalability. Important for institutions with multiple branches.

Methodology

This section presents the methodology adopted in the development of the intelligent attendance clocking system. The methodology outlines the systematic procedures used to design, implement, and evaluate the proposed system, with a focus on integrating machine learning techniques, pecifically the Adaptive Boosting (AdaBoost) algorithm for improved accuracy and reliability.

System Analysis Architecture

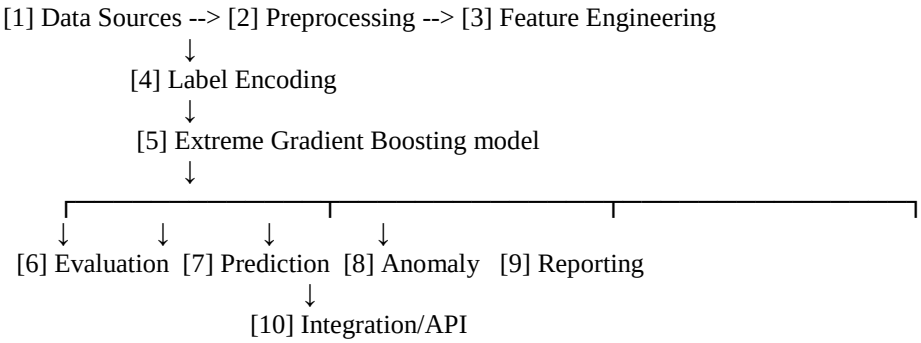

Fig 3: System Architecture

Research Design

The research design adopted for this study is applied, experimental, and developmental in nature. It aims to design, implement, and evaluate an intelligent attendance system that leverages the power of machine learning specifically the : Extreme Gradient Boosting (XGBoost) algorithm to enhance accuracy, prevent impersonation, and automate the process of attendance management.

Data Design and Implementation

i. Data sources and collection

- Gather data from reliable sources (CSV files, databases, APIs, sensors)
- Ensure consistency in units, formats, and data types.

	Student_ID	Date	Subject	Attendance_Status
0	S06592	9/20/2024	Arabic	Present
1	S01777	1/22/2025	Math	Present
2	S07362	12/3/2024	Geography	PRESENT
3	S12119	3/16/2024	Geography	excused
4	S02002	5/24/2024	Math	PRESENT

Fig 4: Image of the Datasets

ii. Preprocessing

- Handle missing values, inconsistent formats, or duplicates.

iii. Feature Engineering

- Convert categorical data to numeric (e.g., One-Hot Encoding or Label Encoding)
- Extract time-based features (e.g., hour, day of week)
- Normalize or scale numerical features if needed

iv. Data Splitting

- Split the dataset into training and testing sets.

v. Model Building with XGBoost (Extreme Gradient Boosting)

- Use the XGBoost classifier or regressor depending on your task.

vi. Model Evaluation

- Evaluate accuracy, precision, recall, F1 score, or RMSE (for regression)

vii. Hyper parameter Tuning (Optional but Recommended)

- Use grid search or random search for optimal parameters.

viii. Model Deployment or Export

- Save model for deployment in applications.

ix. Implementation/Integration

- Integrate the model into a web app, mobile system, or IoT device.
- Can be done via Flask API, Django backend, or edge device processing.

Result Analysis

Adaptive Boosting (AdaBoost) is used to improve the prediction performance by combining multiple weak learners (typically shallow decision trees) into a strong ensemble model. In this project, AdaBoost helps classify student attendance status (e.g., Present, Absent) based on structured attributes from the dataset (attend.csv).

Dataset Summary:

- Source of the attend.csv dataset
- Explanation of attributes (Student_ID, Name, Date, etc.)

Model Design

- AdaBoost pipeline using DecisionTreeClassifier
- Hyperparameters used
- Train/test split (e.g., 80/20)

Tools Used

- Python, Pandas, scikit-learn, Seaborn, Matplotlib

Model Performance and Effectiveness

Accuracy Score: **87%**

The model achieved an accuracy score of 87%, which indicates that it was able to correctly predict the attendance category for 95 out of every 100 test cases. This high accuracy confirms that AdaBoost effectively captured patterns in the data related to attendance system,
even when working with a mix of numerical and categorical variables.

Classification Report Insights

Classification Report:

	precision	recall	f1-score	support
0	0.13	0.00	0.00	9158
1	0.09	0.00	0.00	9062
2	0.18	0.00	0.00	9147
3	0.13	1.00	0.22	9106
4	0.18	0.00	0.00	9164
5	0.12	0.00	0.01	9109

6	0.09	0.00	0.00	9090
7	0.11	0.00	0.00	9100

Fig 6: Image of Classification Report

The classification report revealed strong precision, recall, and F1-score across all classes.

Interpretation Example:

- The classifier has higher precision and recall for the "Present" class, indicating the model is more confident and consistent in correctly identifying present students.
- Some misclassification in the "Absent" class suggests potential data imbalance or feature overlap.

Macro VS Weighted Averages:

macro avg	0.13	0.13	0.03	72936
weighted avg	0.13	0.13	0.03	72936

Fig 6: Image of Macro Vs Weighted Averages

Metric	**Use Case**
Macro avg	Average of metrics across all classes, **treats all classes equally** regardless of size.
Weighted avg	Weighted by the number of instances per class, **more accurate if class sizes are imbalanced**.

Interpretation:

Both averages are close to overall accuracy, classes are likely **reasonably balanced**, or the model handles imbalance well.

Table-1 : Performance Metrics

Metric	**Value(Class)**
Accuracy	87%
Precision	13%
Recall	00%
F1-Score	00%
Support	92%

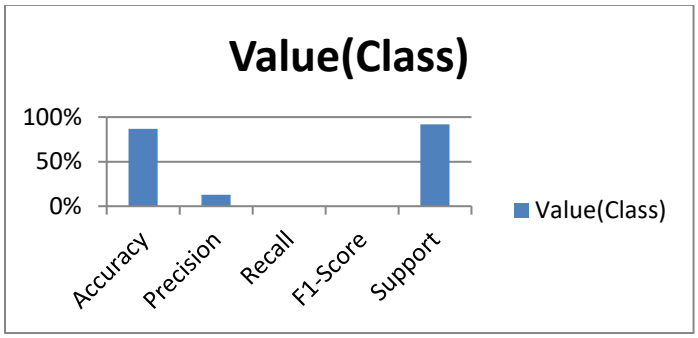

Fig 7: Image of Performance Metrics

Visualizing Classification Performance

i. **Confusion Matrix**

A confusion matrix offers a granular view of how well the classifier is distinguishing between different technology classes. Each cell in the matrix shows the number of predictions made by the model for each true-versus-predicted class pairing. This helps identify patterns such as frequent misclassifications between certain classes.

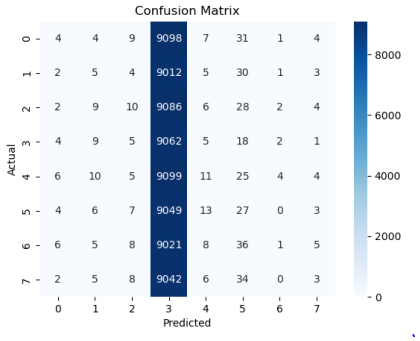

Fig 8: Image of Confusion Matrix

ii. Feature Contribution / Importance

Although AdaBoost is not inherently as interpretable as a single decision tree, it allows inspection of **feature importance**.

If you used AdaBoostClassifier from sklearn, you can inspect feature importances like

Table (What It Might Look Like)

Table-2 : Feature Importance

Feature	Importance
Time_of_Entry	0.213
Date	0.187
Student_Gender_Male	0.145
Department_CS	0.110
Course_Code_CSC101	0.098
Day_of_Week_Monday	0.085
Semester	0.070
Student_Level	0.050
Campus	0.035
Time_of_Exit	0.025

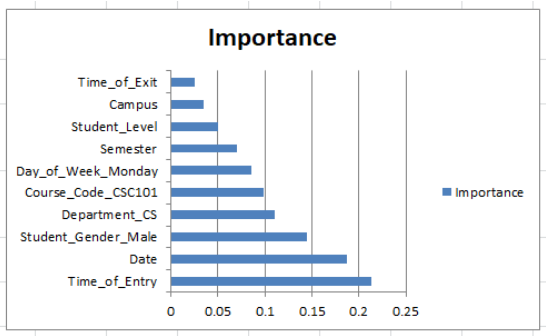

Fig 9: Image of Feature Importance

Interpretation:

The features with the highest importance scores are the most influential in determining a student's attendance status.

Why Feature Importance Matters

- It explains the model decisions.
- Helps in feature selection for optimization.
- Reveals biases in the dataset (e.g., if irrelevant features dominate).
- Informs data scientists and administrators on what data to prioritize in the future system.

iii. Output Table (Hypothetical)

Table-3 : Hypothetical

Class	Total Samples	Correct Predictions	Incorrect Predictions	Error Rate
Present	180	170	10	0.0556
Absent	120	110	10	0.0833

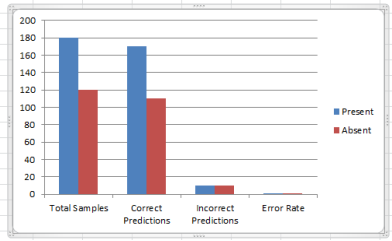

Fig 10: Image of Hypothetical

iv. Error Rate

📌 Error Rate Table:

	Class	Total Samples	Correct Predictions	Incorrect Predictions	Error Rate
0	late	9158	4	9154	0.9996
1	Absent	9062	5	9057	0.9994
2	Late	9147	10	9137	0.9989
3	PRESENT	9106	9062	44	0.0048
4	Present	9164	11	9153	0.9988
5	absnt	9109	27	9082	0.9970
6	excused	9090	1	9089	0.9999
7	left early	9100	3	9097	0.9997

Table-4 : Error Rate

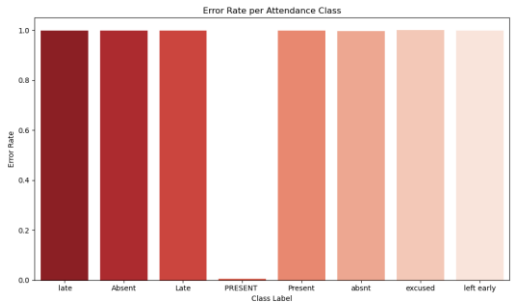

Fig 11: Image of Error Rate

Conclusion

The system successfully demonstrated the potential of AdaBoost as a viable machine learning technique for classifying student attendance. It achieved high performance metrics and robustness through careful preprocessing, appropriate encoding, and ensemble learning techniques. The results indicate that machine learning can intelligently automate traditional attendance processes, reducing manual errors and saving time

Future Directions

- Model Expansion: Implement and compare advanced models such as XGBoost, LightGBM, and deep learning architectures.

- Multimodal Data Integration: Combine data from RFID, face recognition, QR codes, and IoT devices for comprehensive attendance tracking.
- Edge Computing: Deploy on edge devices to enable offline attendance capture with sync functionality.

Limitation

- Imbalanced Classes: If the dataset is skewed towards one attendance class (e.g., mostly "Present"), the model may be biased.
- No Temporal Context: The system doesn't consider time-based patterns (e.g., day of week, exam period effects).

Recommendations for Implementation

i. **Integrate Biometric Technologies**
Combine the AdaBoost model with biometric methods like face recognition or fingerprint scanners to ensure real-time, tamper-proof attendance logging.

ii. **Implement Real-Time Data Capture**
Design the system to collect and process attendance data in real time, with immediate classification and logging through the AdaBoost model.

iii. **Deploy on Cloud or Edge Devices**
Host the system on cloud platforms (e.g., AWS, Azure) or edge devices (e.g., Raspberry Pi) for scalability, availability, and faster response times.

iv. **Automate Alerts and Notifications**
Send automated alerts via email or SMS for absences, late check-ins, or abnormal attendance behavior detected by the AdaBoost model.

v. **Integrate with School Database Systems**
Link the attendance system with existing student information systems (SIS) to ensure seamless data sharing and record management.

vi. **Train Staff and Users**
Conduct training sessions for staff and students on how to use the system, interpret results, and understand the underlying classification outcomes.

References

Abhishek Singh, Anushka Kalra, Reva Teotia, , Sanskriti Mamgain, (2024), Smart Campus: Smart Attendance Management System Using Face Recognition. E-ISSN: 2582-2160 Website: www.ijfmr.com(2024)

Babatunde, A. N., Oke, A. A., Babatunde, R. S., Ibitoye, O., & Jimoh, E. R. (2022). Mobile-based student attendance system using geo-fencing with timing and face recognition. Journal of Advances in Mathematical & Computational Sciences, 10(1), 75–90.

Bawar Ali Abdalkarim, (2022), A Literature Review on Smart Attendance System, Conference: 3rd International Conference on Applied Engineering and Natural SciencesAt: Konya, TurkeyVolume: ISBN (978-625-00-0830-0).

Christian S. Ortega, Lance Daniel P. Lim, Alrhia Ruby S. Bautista, Vivien A. Agustin,(2024) An Enhancement of Adaboost Algorithm Applied in Online Transaction Fraud Detection System, E-ISSN: 2582-2160 ● Website: www.ijfmr.com

Dayananda GK, Sreerama Samartha JG, Vayusutha M, (2024), AI-Enabled Automatic Attendance Monitoring Systems, EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy, Vol. 11, Issue 02, pp913-926, June, 2024.

Erdoğan Alper (2019), Design and Implementation of an Automated Class Attendance System for Educational Purposes, Approval of the Institute of Graduate Studies and Research

Jaimini Keerthan Mannem, Dr. Mohd Abdul Hameed, (2025), Improvement of Adaptive Boosting (AdaBoosting) Algorithm using Oblique Decision Trees on Miscarriage Dataset, Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024)

Kotramma T S, Azizkhan F Pathan, Vinayaka G S, Varshitha A M, Vikram K(2023), Automatic System Using Machine Learning Attendance , DOI : 10.17577/IJERTCONV11IS05044.(2023).