FRAUD DETECTION USING ISOLATION FOREST FOR RFID-BASED ATTENDANCE MONITORING SYSTEM

Mark Lister V. Nalupa[1], Jodie Rey D. Fernandez[1], Wencel Jean C. Dacay[1], Miriam M. Bergado[1]
[1]University of Science and Technology of Southern Philippines, C.M. Recto Ave., Lapasan, Cagayan de Oro City, Philippines
*For Correspondence; Tel. +639269681744, Email: marklister.nalupa@ustp.edu.ph

ABSTRACT: *Monitoring students' attendance using RFID technology on a regular basis is of utmost importance in the education sector. However, the use of RFID-based attendance has been heavily criticized for its vulnerability to counterfeit attendance. Anyone holding the RFID tag can mark an attendance to the system. This study aims to design and develop fraud detection on attendance. The system implementation resulted in an RFID-based attendance monitoring system capable of detecting suspicious attendance. The Isolation Forest, an unsupervised machine learning algorithm, was used to identify anomalies within the attendance dataset. The result shows that the machine learning model attained an accuracy score of 95.69%, a precision score of 96.33%, and a recall score of 95.81%. The study concludes that detecting suspicious attendance using an Isolation Forest is possible. It is recommended to be integrated with existing RFID-based attendance monitoring systems to minimize fraud or counterfeit attendance.*

Keywords: fraud detection, RFID-based attendance, isolation forest, attendance monitoring system

INTRODUCTION

Increased efficiency and productivity of the school are opportunities that the school administrators have seen in integrating radio-frequency identification (RFID) technology in monitoring attendance. It is observed and proven in many studies that a well-monitored school will have better discipline and higher academic performance compared to schools without surveillance [5, 6, 8]. Students avoid disciplinary issues when they know that attendance is tracked and communicated to parents. The intervention of the technology improved the discipline of the students resulting in better grades.

The development of the RFID-based student attendance monitoring system cannot miss out that it is a much-needed solution in the education sector for its quickness in identifying students. The report published showed a continuing growth of enrolment in the Philippines [1]. The increasing number of enrolled students, particularly in public schools, implies an increased effort, which may affect other key performance areas of a teacher in monitoring the students. Thus, it is vital to minimize the time consumed in a manual attendance system. Additionally, schools still implementing the manual attendance system may face problems in managing accurate attendance reporting. The outdated practices in monitoring students' attendance cannot assure the veracity of data collected manually. That being said, the RFID-based student attendance monitoring system cuts downtime and improves the accuracy of monitoring student attendance.

However, given these benefits to students and advantages to educational institutions, the use of an RFID-based student attendance monitoring system has been heavily criticized for its drawbacks in terms of reliability. The existing system is vulnerable to counterfeit attendance. A study revealed that RFID technology in the attendance system has low reliability [7]. Anyone holding the RFID tag can mark an attendance to the system. This drawback in the existing system affects the integrity of the system and the data collected.

In RFID systems, various methods were put into place to guarantee data integrity. Facial recognition [4] [12], mobile applications [1], and Biometrics [1] [2] [13] are a few examples of multi-factor verification techniques. Although these systems provide alternatives to robust existing RFID systems, they are unable to identify fraud.

In recent years, machine learning models have been heavily incorporated into electronic systems. It enables electronic systems to learn autonomously from existing data and use this acquired knowledge to make assessments, predictions, and decisions independently. Significant developments that incorporate machine learning algorithms in various organizations have been introduced. One key benefit machine learning renders to many firms is to prevent them from fraudulent acts. Machine learning in fraud detection is aiding different organizations in tackling security issues efficiently. Fraud detection and prevention are the main concerns in many organizations. One typical example which addresses fraud and applies a machine learning model is the biometric attendance system. The system is trained with inputs of biometric identity like the thumb, iris, or face [9] [13]. Since it is trained, it can validate future input and quickly identify the person. Aside from biometrics, another approach presented is to prevent fraudulent acts in attendance [9]. This approach uses the recorded entry time of a person as the basis of training the machine learning model, which predicts the sign-in time. The pattern of entry time is usually recurrent in individuals. If the actual sign-in time is not near what is predicted, the attendance will be marked as suspicious. Integration of detecting false positive readings using classifiers based on logistic regression, support vector machine, and decision tree is suggested [11]. While another study introduced an approach using anomaly detection algorithms like One-Class K-means with Randomly-projected features Algorithm (OCKRA), Isolation Forest, and One-Class Support Vector Machines (SVM) [15]. These algorithms were found to have a good performance in terms of the overall Area under the Curve (AUC) [15]. Additionally, the Isolation Forest algorithm works well in high dimensional problems which have a large number of irrelevant attributes and in situations where the training set does not contain any anomalies [10].

Considering the impact of RFID technology in monitoring student attendance in educational institutions and the issue at hand presented, the focal objective of the study is to develop a student attendance monitoring system based on RFID technology with a mitigating mechanism that aims to detect fraud in attendance. In this study, fraud may refer to suspicious or false attendance.

METHODOLOGY

System Requirements

The researchers were able to gather information on the existing process of monitoring attendance in the school. The researchers administered a few interviews and observations to the school personnel. The prepared open-ended and close-ended questions were asked to mainly inquire about the entire process and challenges encountered in monitoring attendance. The researchers collated and analyzed the data collected. From that data, the researchers were able to identify the system requirements. The requirements of the system are shown in Table 1. The table shows the list of system requirements and their corresponding category.

Table 1: System Requirements

Category	System Requirement
Input Requirements	-The system shall allow students to tap the RFID tag on the RFID reader. -The system shall allow students to press their fingers on the fingerprint scanner.
Process Requirements	-The system shall record the date and time of students' entrance and exit to the school. -The system shall detect suspicious attendance in the school.
Output Requirements	-The system shall provide a monitoring interface for the entrance and exit of students in the school. -The system shall provide feedback for the detected suspicious entrance of students into the school. -The system shall generate an attendance report.
Control Requirements	-The system shall only allow authorized users at the server and application levels. -The system shall maintain separate user privileges.
Performance Requirements	-The system shall provide an effective and efficient attendance monitoring system capable of detecting suspicious students' entrance into the school. -The system shall provide real-time data collection on the entrance and exit of students in the school. -The system shall be operational during school hours.

System Design

Figure 1 shows the system architecture of the RFID-based attendance monitoring system. The system employs client-server architecture. There are five major hardware components in this structure: server, RFID terminal for entrance, RFID terminal for exit, RFID registration terminal, client computer, and wireless router. The microcomputer is configured as the server computer that hosts, delivers, and manages the resources and services being requested by the client computer. In the RFID terminal for entrance, the microcomputer is connected to an RFID reader, fingerprint scanner, and buzzer. It is programmed to send the unique identification information of the RFID tag or fingerprint data to the server for time logs in entering the school. On the hand, the RFID terminal for entrance comprises a microcomputer connected to an RFID reader and buzzer. Its purpose is to send the unique identification information of the RFID tag to the server for time logs in exiting the school. The buzzer in the foregoing RFID terminals provides feedback for successful time logs. As illustrated in the figure, the fingerprint scanner is only present in the RFID terminal for entrance since it will be used to verify if a suspicious entrance is detected. The RFID registration terminal comprises a microcomputer connected to an RFID reader, keyboard, and monitor. This hardware component handles the student's corresponding RFID tag registration to the server's database. The client computer connects to the microcomputer server to access the data in the generation of reports and other processes for attendance monitoring. All the components in this structure are connected via a wireless router.

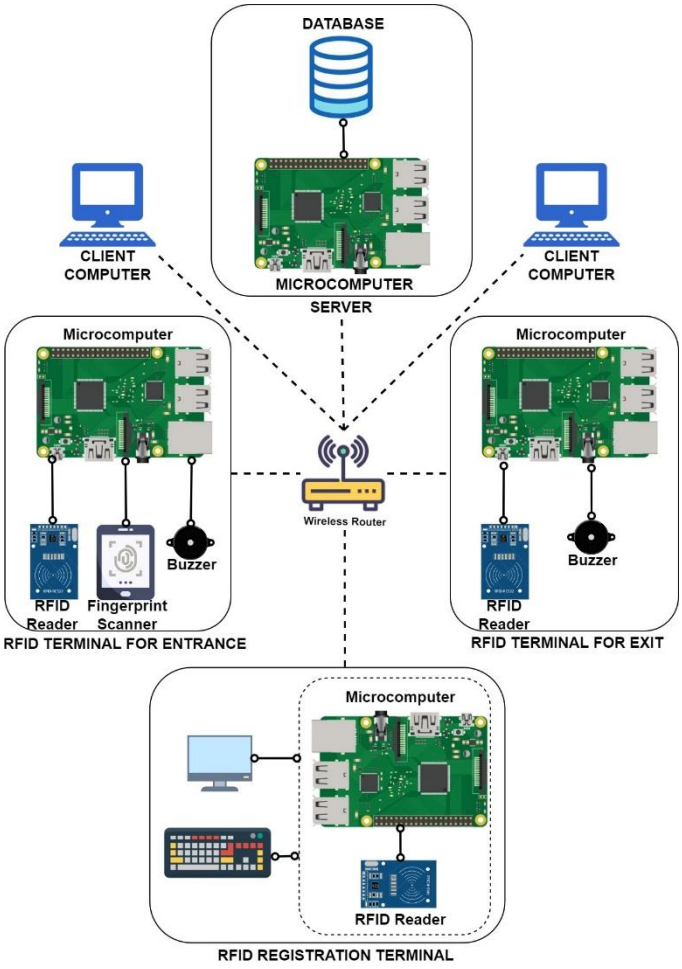

Figure 1: System Architecture

Figure 2a and Figure 2b provide the visualization of the process of detecting suspicious attendance in the school. It starts with the student tapping the RFID tag on the RFID reader at the entrance gate. The RFID reader determines the RFID tag's unique identification number and sends it to the microcomputer server to check if it is registered in the database. If it is not registered, the time log is not recorded. Otherwise, the microcomputer server retrieves its current time and loads the machine learning model afterward. The machine learning model compares the retrieved current server time to the threshold time of the student. The threshold time varies for each student since it is determined based on the historical time logs of the student tapping the RFID terminal for entrance. If the retrieved time from the microcomputer server for the student's time log is within the threshold time, the microcomputer will record the time logs for entrance into the database. Otherwise, verification via a fingerprint scanner will be performed. The student will press his finger on the

fingerprint scanner. Upon pressing the fingerprint scanner, the microcomputer will match the scanned fingerprint to the stored sample fingerprint and retrieve the current server time. If the scanned fingerprint matches the stored sample fingerprint, the microcomputer will save the time log of the student's entrance into the database; otherwise, it will not.

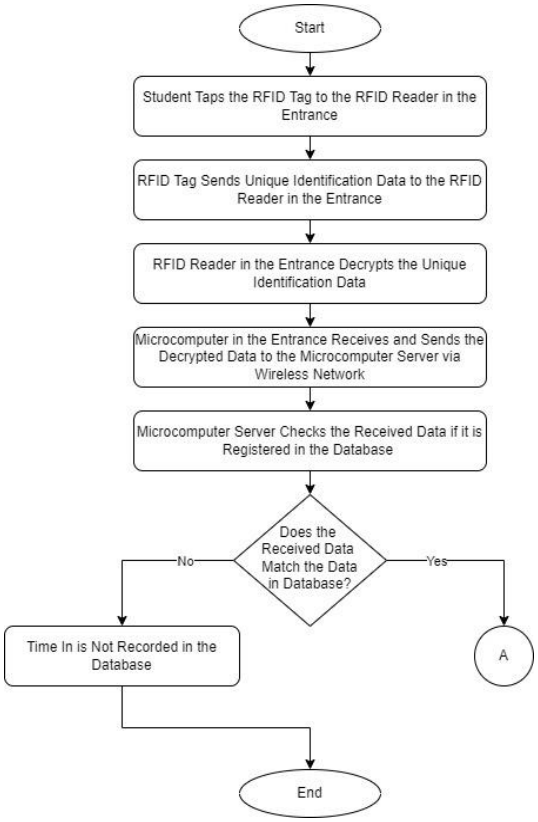

Figure 2a: System Flowchart

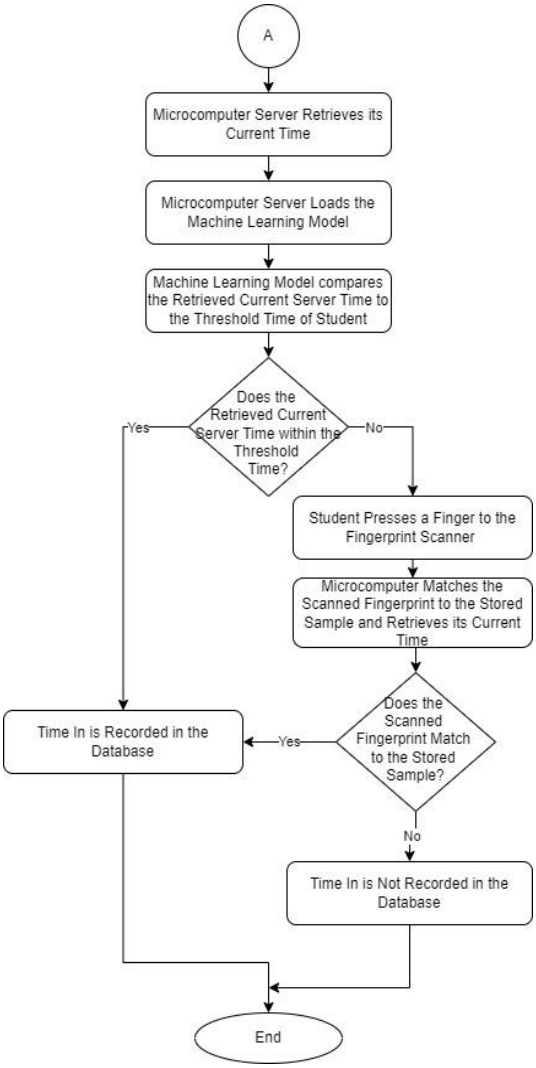

Figure 2b: System Flowchart

Figure 3 shows a use case diagram that presents the high-level function and interaction between the system and its actors. The system has four actors: Admin, Teacher, Student, and Security Personnel. Each actor has corresponding privileges in accessing the system. The admin configures the system by adding RFID terminals and creating user group privileges. In addition to the admin's interaction with the system, the admin can create a class section for each school year, register students' RFID tags, and generate comprehensive attendance reports. The teacher is limited only to registration of students to the class section and generation of attendance reports. The student interactions with the system are to tap the RFID tag to the RFID reader of the terminal for every entrance and exit to the school and press the registered finger on the fingerprint scanner if classified the time logs as suspicious during the entrance. The security personnel receives notification for all detected suspicious attendances. Furthermore, security personnel can view real-time gate logs.

System Implementation The programming and scripting languages were used for the software development of the RFID-based attendance monitoring system. In presenting the layout of the web pages of the system, Tailwind and Vue were used. Both are open-source frameworks for building graphical user interfaces. Tailwind is a Cascading Style Sheets (CSS) framework, while Vue is a front-end JavaScript framework. In implementing the server-side algorithm of the system, Node was employed. It is a back-end JavaScript runtime environment that runs on a JavaScript Engine and executes JavaScript code outside a web browser. In governing the operation of microcomputers, Python programming language was used. It is a high-level, general-purpose programming language. In storing the data collected by the system, MySQL was utilized. It is a relational database management system that uses Structured Query Language (SQL) commands to query and operate the system's database.

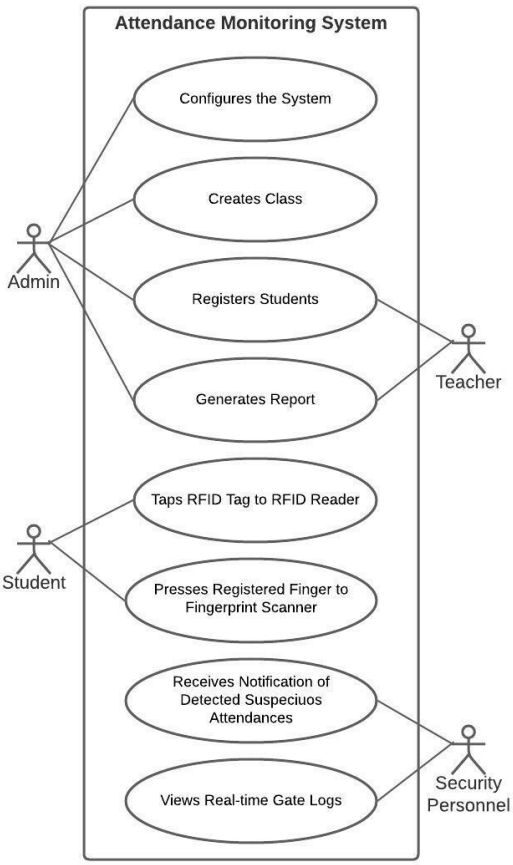

Figure 3: Use Case Diagram

The Isolation Forest was employed to detect suspicious attendances at the gate. It is an unsupervised machine learning algorithm for identifying anomalies within a dataset by isolating anomalies as they are few and different. It focuses on anomalies as a dataset will have fewer than normal data points. Their feature values are very different from those of the normal data points and, hence, easier to find. The Isolation Forest isolates the observation by applying a recursive partition to randomly select a feature and then select a split value between the minimum and maximum values for the selected feature. In this way, random partitioning produces a shorter path for anomalies than for normal data points [10]. The datasets used in detecting the suspicious attendances are the historical time logs of the student tapping the RFID terminal for entrance.

Microcomputers, RFID readers, fingerprint scanner, and buzzers were used for the hardware development of the RFID-based attendance monitoring system, particularly the server and RFID terminals. The following are the specifications of mentioned hardware components: Raspberry Pi 4 8GB RAM 128 ROM for microcomputers, RFID-RC522 model for RFID readers, AS608 Optical Fingerprint Sensor Module for fingerprint scanner, and Piezo Buzzer for buzzers. All microcomputers were statically configured with a unique IP address for server access and identification of RFID terminals for entrance and exit. Furthermore, 3D print enclosures were fabricated to protect all hardware components inside the server and RFID terminals.

System Testing

In order to ensure that the system met the requirements, the system was tested. This entire process aimed to identify all the defects in the system. The system had undergone different levels of testing to eliminate bugs and minimize the possible errors that could affect the system's performance. Unit testing was conducted to examine every system's unit during its development. It ensures that it meets the original requirements and functions as expected. Subsequently, integration testing was done to detect the flaws in the interactions between the units within the system. Next, system testing was executed, wherein the complete system was tested as a whole. This stage verified the system's compliance with the system requirements and overall quality standards. In this testing stage, different performance metrics were performed to evaluate the system's performance. It is to quantify the accuracy of the machine learning component of the system. The accuracy, precision, and recall were determined.

RESULTS & DISCUSSIONS

Web Application

Access to the web application is restricted to authorized users. The user will be directed to the Home Page, as shown in Figure 4, if correct credentials are provided on the Login page. The Home Page displays the latest student who tapped the RFID Terminal at the entrance and exit gate.

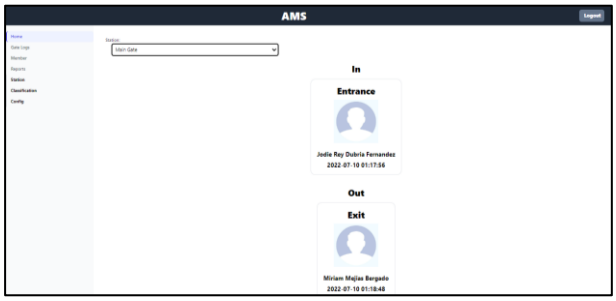

Figure 4: Home Page

Figure 5 shows the Gate Logs page, which displays the list of all students who tapped their RFID tag in the RFID terminal. The data in the table is arranged by date in chronological order from the latest data to the oldest data.

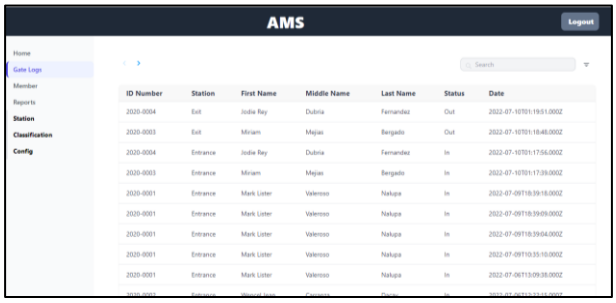

Figure 5: Gate Logs Page

Shown in Figure 6 is the Member page. The authorized user may add, update or delete a student on this page. Furthermore, an authorized user may view this page to check if a student already has an RFID tag.

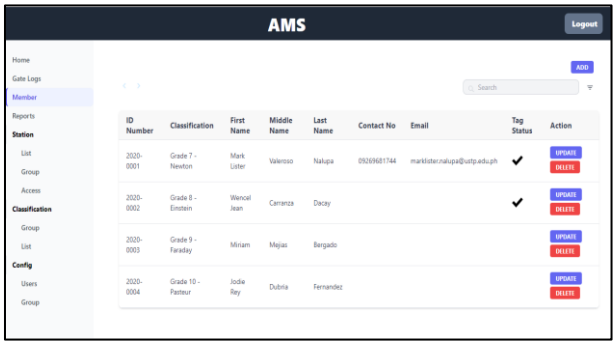

Figure 6: Member Page

Figure 7 presents the Reports Page. This page displays a form that would allow the authorized user to generate a report on students' attendance. The user shall select the details of what data in the report shall be reflected. The report is based on the Philippine Department of Education School Form 2 (SF2).

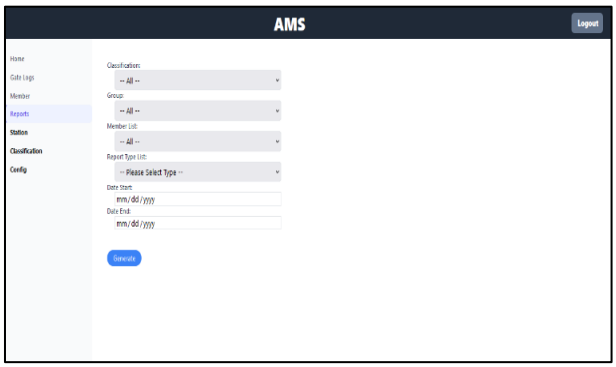

Figure 7: Reports Page

Server and RFID Terminals

Reflected in Figure 8a, Figure 8b, Figure 8c, and Figure 8d are the 3D printed enclosure of the server computer, RFID registration terminal, RFID terminal for entrance, and RFID terminal for exit, respectively. The enclosure box acts as a shield to protect the inside-located microcomputer and RFID reader.

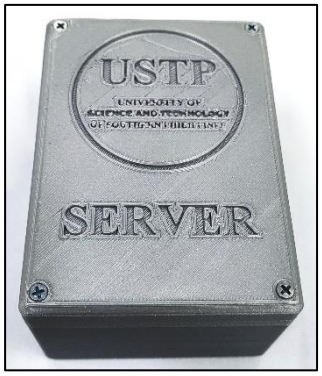

Figure 8a: Server

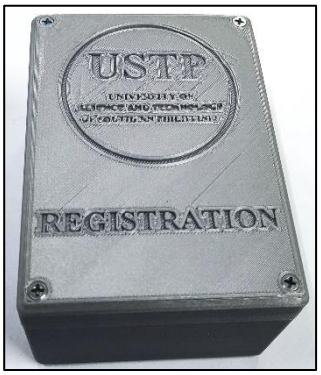

Figure 8b: RFID Registration Terminal

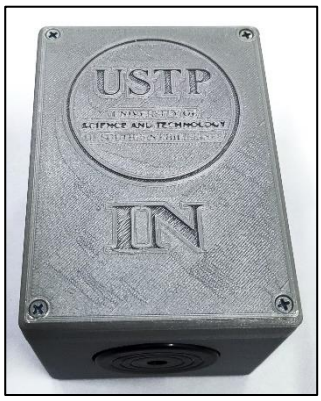

Figure 8c: RFID Terminal for Entrance

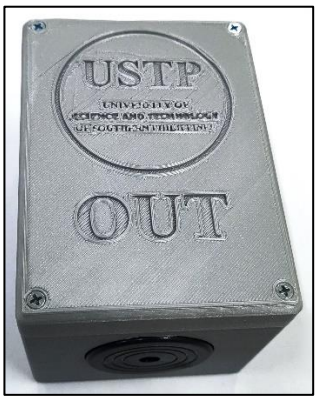

Figure 8d: RFID Terminal for Exit

Training and Validation Result of Machine Learning Model

A dataset of time-in logs from Kaggle was used to train and evaluate the machine learning model. It was utilized to test Isolation Forest in detecting the anomalies within the dataset since there were no existing data on students' time-in to the school.

Before classifying the data in the datasets, if it is an anomaly, date and time were converted into float values to be processed in the model, as shown in Figure 9 and Figure 10.

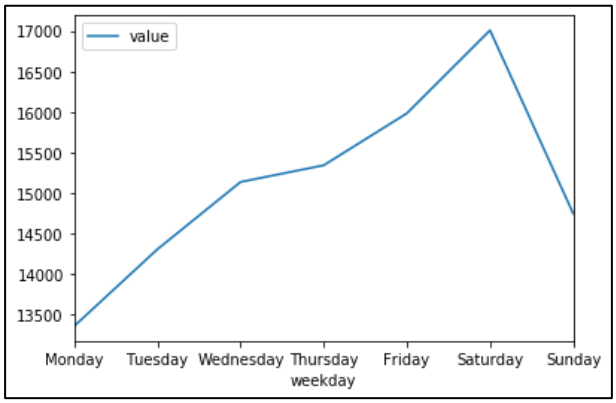

Figure 9: Date Converted to Float

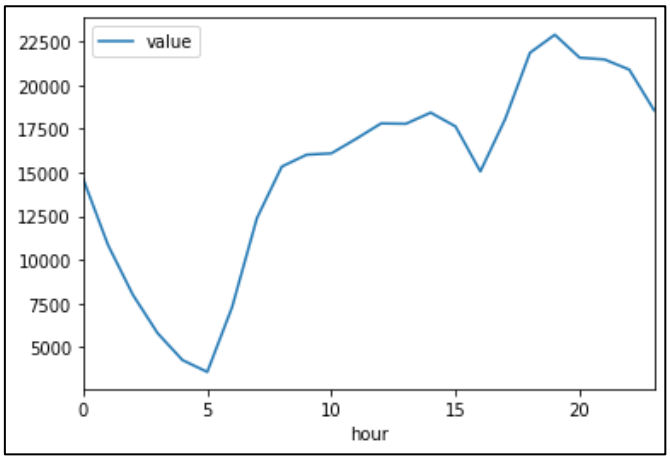

Figure 10: Time Converted to Float

Figure 11 shows all the data plotted in the histogram to show the visualization of the dataset's peaks and lows, indicating the data anomaly.

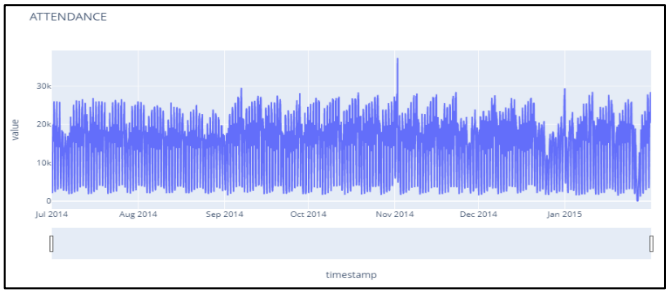

Figure 11: Data in Histogram

Figure 12 shows the classified anomaly from the dataset, indicating inconsistencies and tardy student data. The red dot is the anomaly of the data from the dataset.

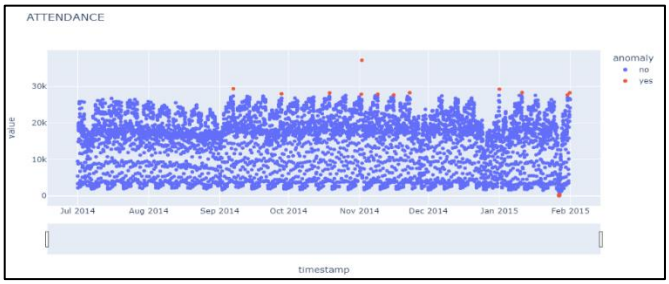

Figure 12: Classified Anomaly from the Dataset

Detected anomalies are presented in Figure 13, which shows the timestamp (time-in of a student), value (float value of the date and time), hour of the specified date, day in the week, and anomaly (yes or no).

	timestamp	value	hour	weekday	anomaly
1631	2014-09-06 23:00:00	29418.5	23	Saturday	yes
2135	2014-09-27 23:00:00	28024.5	23	Saturday	yes
2639	2014-10-18 23:00:00	28253.5	23	Saturday	yes
2971	2014-11-01 19:00:00	27912.0	19	Saturday	yes
2977	2014-11-02 01:00:00	37204.5	1	Sunday	yes
3143	2014-11-08 23:00:00	27926.0	23	Saturday	yes
3311	2014-11-15 23:00:00	27713.0	23	Saturday	yes
3479	2014-11-22 23:00:00	28299.0	23	Saturday	yes
4417	2015-01-01 01:00:00	29292.0	1	Thursday	yes
4655	2015-01-10 23:00:00	28351.0	23	Saturday	yes
4795	2015-01-16 19:00:00	27571.5	19	Friday	yes
5040	2015-01-27 00:00:00	94.5	0	Tuesday	yes
5041	2015-01-27 01:00:00	39.5	1	Tuesday	yes
5042	2015-01-27 02:00:00	29.0	2	Tuesday	yes
5043	2015-01-27 03:00:00	9.5	3	Tuesday	yes
5044	2015-01-27 04:00:00	14.5	4	Tuesday	yes
5045	2015-01-27 05:00:00	29.0	5	Tuesday	yes
5046	2015-01-27 06:00:00	88.0	6	Tuesday	yes
5131	2015-01-30 19:00:00	27698.0	19	Friday	yes
5155	2015-01-31 19:00:00	28288.5	19	Saturday	yes

Figure 13: Detected Anomaly

Table 2 shows the validation result of the anomaly detection model for attendance monitoring. It shows that the Isolation Forest attained an accuracy score of 95.6873%, with a precision score of 96.3297% and a recall score of 95.8134%.

Table 2: Validation Result

Accuracy Score	Precision Score	Recall Score
95.6873%	96.3297%	95.8134%

CONCLUSION & RECOMMENDATION

This study presented the concept of developing a student attendance monitoring system that can detect suspicious attendance in the school. The researchers were able to successfully develop an attendance monitoring system that provides school personnel to monitor the student's attendance. The system was implemented using RFID technology. Furthermore, the researchers were able to integrate the machine learning model in detecting suspicious attendance in the school using Isolation Forest. The detected suspicious attendance was subjected to verification using a fingerprint scanner. The validation result showed that the machine learning model was capable of detecting possible fake attendance based on historical time logs.

The suspicious attendance detection allowed accurate recording of student attendance. In addition, it prevented students from committing fake attendance in school, which is a serious academic offense. Thus, it is recommended to be integrated with existing RFID-based attendance monitoring systems. This system allows teachers and other school personnel to effectively and efficiently monitor the student's attendance in the school.

ACKNOWLEDGMENT

The researchers would like to thank University of Science and Technology of Southern Philippines (USTP) for funding the project and the USTP's Center for Artificial Intelligence for its support in conducting the study.

REFERENCES

[1] Adal, H., Promy, N., Srabanti, S., & Rahman, M. (2018, February). Android based advanced attendance vigilance system using wireless network with fusion of bio-metric fingerprint authentication. In 2018 20th International Conference on Advanced Communication Technology (ICACT) (pp. 217-222). IEEE.

[2] Ahmed, A., Olaniyi, O. M., Kolo, J. G., & Durugo, C. (2016). A multifactor student attendance management system using fingerprint biometrics and RFID techniques.

[3] Albert, J. M. & Raymundo, J. (2016). Trends in out-of-school children and other basic education statistics. Retrieved from https://dirp3.pids.gov.ph/websitecms/CDN/PUBLICATIONS/pidsdps1639.pdf

[4] Alburaiki, M. S. M., Johar, G. M., Helmi, R. A. A., & Alkawaz, M. H. (2021, August). Mobile based attendance system: face recognition and location detection using machine learning. In 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC) (pp. 177-182). IEEE.

[5] Dey, I. (2018). Class attendance and academic performance: A subgroup analysis. International Review of Economics Education, 28, 29–40. doi: 10.1016/j.iree.2018.03.003

[6] Irwin, N., Burnett, K. M., & Mccarron, P. A. (2018). Association between attendance and overall academic performance on a module within a professional pharmacy degree. Currents in Pharmacy Teaching and Learning, 10(3), 396–401. doi: 10.1016/j.cptl.2017.11.008

[7] Kariapper, R., & Razeeth, S. (2019). RFID Based (IoT) Automatic Attendance System: A Survey Analysis. SSRN Electronic Journal. doi: 10.2139/ssrn.3372734

[8] Kieffer, M. J., Marinell, W. H., & Neugebauer, S. R. (2014). Navigating into, through, and beyond the middle grades: The role of middle grades attendance in staying on track for high school graduation. Journal of School Psychology, 52(6), 549–565. doi: 10.1016/j.jsp.2014.09.002

[9] Kumar A. (2019) Android based Cost-Efficient AI Attendance System, International Journal of Engineering Research & Technology (IJERT) Volume 08, Issue 11 (November 2019)

[10] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining. Retrieved from https://doi.org/10.1109/icdm.2008.17

[11] Ma, H., Wang, Y., & Wang, K. (2018). Automatic detection of false positive RFID readings using machine learning algorithms. Expert Systems with Applications, 91, 442-451.

[12] Rathod, H., Ware, Y., Sane, S., Raulo, S., Pakhare, V., & Rizvi, I. A. (2017, January). Automated attendance system using machine learning approach. In 2017 International Conference on Nascent Technologies in Engineering (ICNTE) (pp. 1-5). IEEE.

[13] Reddy, K. N., Alekhya, T., Sushma Manjula, T., & Rashmi, K. AI-Based Attendance Monitoring System.

[14] Shukla, V. K., & Bhandari, N. (2019, February). Conceptual framework for enhancing payroll management and attendance monitoring system through RFID and biometric. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 188-192). IEEE.

[15] Villa-Pérez, M. E., Alvarez-Carmona, M. A., Loyola-González, O., Medina-Pérez, M. A., Velazco-Rossell, J. C., & Choo, K. K. R. (2021). Semi-supervised anomaly detection algorithms: A comparative summary and future research directions. Knowledge-Based Systems, 218, 106878.