A Research Study Presented to the Faculty of

CABANATUAN CITY SENIOR HIGH SCHOOL

Sta. Arcadia, Cabanatuan City

In Partial Fulfillment of the Requirements

for the Subject Capstone

EasyWiz: Utilization and Implementation of RFID Technology in Student Lockers and Machine Learning Models for Face Recognition-based Attendance System in Cabanatuan City Senior High School

Nicole Amber Hennesey

James Aaron Erang

Ezekiel Lopez

Fiona Leigh Pagtama

Raven Gose

Project Authors

Ma'am Jessmia Cacatian

Project Adviser

June 2022

RATIONALE

Regardless of the significance of human life as both argued as important and non-significant by anthropic principle (Weinberg, 1989; Barrow & Tipler, 1988; Schrodinger, 1944) and the laws of nature (Kaku, 1995), respectively. Providing the rights, especially security and privacy is arguably an utmost necessity. However, due to the inefficiency and high vulnerability of currently implemented system mechanisms and the technology itself, it is easily voided. Furthermore, the exponential increase of technological advancements (Kaku, 1995) and the revolution it provides in understanding of the grand design (Katori *et al*., 2022) the progression of existing threats metamorphosed in the realms of technology, such as telemetry and spywares. Hence efforts are made to yield products that would prevent and decrease the progression of threat.

One of the yields by continuous efforts against security and privacy violation and exploitation in technological level is the movement to lean towards free and open source software (FOSS), while increased and improved locking mechanism has been the direction of combat against physical threats, and combination of both has also been proven to be efficient. Nevertheless, both measures have its own advantages and disadvantages, such as FOSS softwares requires the necessity of accessible codebase and algorithm, and physical measures on the other hand ascertain the decreased vulnerability of the mechanism through intensive maintenance.

Due to the accelerating improvement of technologies, it is necessary to keep up with the changes. Since stabilization is not sufficient, implementation is necessary as everyone is affected by changes in macro level due to the interconnectivity (Lewin *et al*., 2018; Lewin *et al*., 2022), a successful implementation of technological advancements would yield a benefit

to the particular society. Thus, the goal of the study is to implement the current technological advancement in Cabanatuan City Senior High School.

STATEMENT OF THE PROBLEM

The study aims to introduce a new system that replaces traditional methods, specifically the attendance checking and accessing container, such as storage room or student lockers via the utilization of RFID-based systems and microcomputers and microcontrollers, particularly arduino and raspberry pi, and determine whether it is more convenient than the traditional way. Hence, in the final analysis the study sought to :

1. Determine if the RFID Identification Card can be used in :

    a.) accessing lockers ;

    b.) and checking attendance.

2. Make a regression model of the cost-effectiveness of RFID-based systems against traditional methods in the time frame of :

    a.) 15 months ;

    b.) 24 months ;

    c.) 60 months.

3. Determine whether the RFID Technology system can replace the traditional checking of attendance and mechanical lock in terms of :

    a.) Convenience ;

    b.) Efficiency ;

    c.) security ;

    d.) and privacy.

OBJECTIVES

It is critical to guarantee the safety and security of students in school. Due to the exponential improvement of technology (Kaku, 1995), adapting to the changes is a necessity; thus, this study aims to provide information on implementation of the technological advancements to deal with and resist the infringement on security and privacy as well as its exploitation. Therefore, the aims of the study are :

a. Implement the current technological advancement for use in long term.
b. To develop *de novo* convenient and efficient system as a replacement of traditional attendance and locker access.
c. Utilize RFID card and machine learning for use in locker and attendance, and facial recognition software, respectively.

THEORETICAL FRAMEWORK

According to Letvina *et al.* (2021) machine learning, also known as computational learning theory, is the ability of a program to improve itself through repeated iteration on previous and new datasets. Avrim Brum, from Carnegie Mellon University, further defined it as an ability of the algorithm to improve itself and learn from new inputs and adapt to changes, which improves the performance through "*experience*". By doing so, machine learning algorithms can predict and simulate phenomena based on the previous dataset or random inputs from curated dataset, moreover, depending on the model type, it is also possible to produce *de novo* outputs from the algorithm (Letvina *et al*., 2021).

Regression algorithms, study the hypothesis(es), $\boldsymbol{h_0}$, based on parameters, θ_n, β_m, and values of the independent variable, $\boldsymbol{x_i}$, given on previous dataset, to predict the value of the

dependent variable y_i. The study modeled the cost efficiency of the RFID based systems over specific timeframes via utilization of simple linear regression algorithm, also known as ordinary least squares regression.

CONCEPTUAL FRAMEWORK

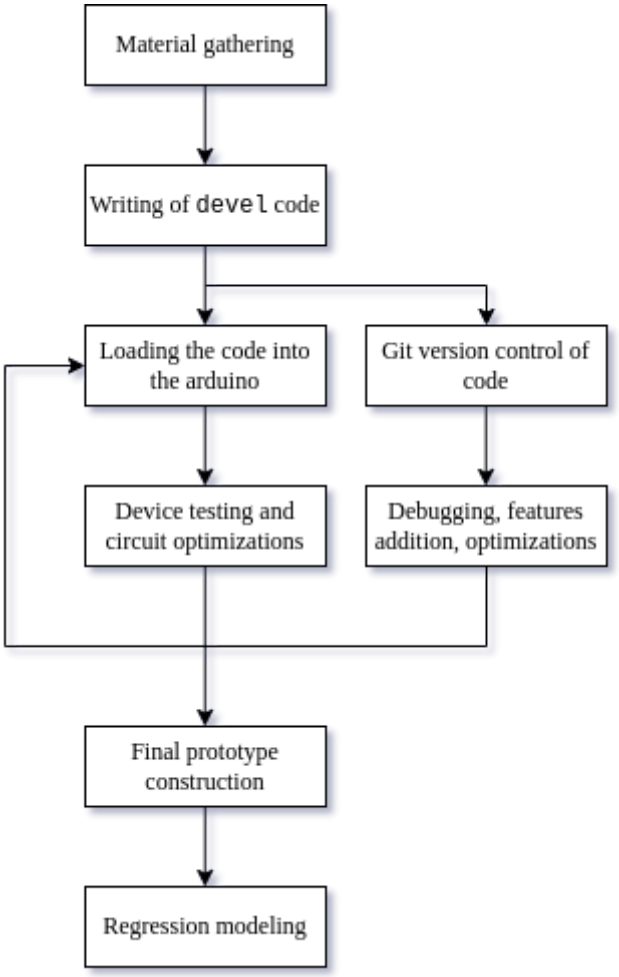


Figure 1: Conceptual Framework of the Study.

Figure 1 presents a brief description of the methodology of the study. The series is broken into 3 phases, the first phase is the preparation, this includes the gathering of the materials needed in the study, and the writing of the *pseudo*-code and its theoretical

implementation. The second phase is the pre-trial, this includes the forking of the main branch of the git version control repository of the main code, and loading of the first version of the main code into the system for testing. The git fork of the code was then further optimized and developed for implementation of new features that will be loaded in the system for next testing. The process was done repeatedly until there were no noticeable bugs to the end user perspective, considerable execution speed, and smooth execution of the program. Finally, the researchers removed the possible conflicts per code branch, and then branches of the code were merged with the main branch.

In the final phase, where the reported and caught bugs, errors and logical error, the final prototype was constructed, and a regression model was devised to show the cost efficiency of the models against traditional methods.

SIGNIFICANCE OF THE STUDY

The interconnectivity of every organism (Lewin *et al*., 2018; Lewin *et al*., 2022) relays the changes and actions of one to another, thus a change in one society will affect another and may affect even larger systems, which is also known as butterfly effect. Hence, making the correct action would probably lead to improvement of the particular society at the very least, while making the opposite will lead to acceleration of depressing consequence or the "*unyielding despair*", once echoed by mathematician and philosopher, Bertrand Russell (Kaku, 1995). The main goal of the study is to address the gap and vulnerability of currently implemented mechanisms and technology of traditional methods in Cabanatuan City Senior High School through utilization and implementation of microcomputers, microcontrollers and RFID technology as replacement of the traditional and inefficient methods currently in use.

School Administrators. This study can help the school administrators in determining the most cost-effective way by introducing this new system while also increasing the convenience for school administrators.

Students. This study can assist students in ensuring their safety and privacy within the school grounds. Using arduino and RFID technology for the school can increase student security and make students feel comfortable and safe inside the school area.

DEFINITION OF TERMS

Branch – is derivative of code base.

Container/Docker/Toolbox – is an isolated *pseudo* operating system (OS) used for virtualization, testing, and debugging of software, as well as simulation.

Container – is any physical storage or space, such as a locker or storage room.

Devel – is a programming technology for development, usually used as a name and package suffix of a branch of main code repositories and binaries, respectively.

Fork/Forking – is a function of the git version control system that replicates the repository of the desired code base

Merge – is the combining of two different branches or forks.

Modules – are external code library files, with a suffix of typically .py, that contains a specific set of functions.

Notebook – is a hosted Jupyter kernel and Linux kernel provided by Google Brain Team of Google Inc., with a sole purpose for use in data science, machine learning, deep learning and artificial intelligence, which can be also used in computational methods such as simulation and astrophysical data analysis.

Push – sending the changes from one repository, either local or other `git` repository, to the specific branch.

Pull – process of fetching the updates from the `git` repository into another repository.

Regression – is an algorithm (computer science) and set of equations (mathematics and statistics) used to predict y_i (dependent variable) using x_i with β and θ based on previous dataset.

`./` – parent directory.

`../` – previous directory.

`CWD` – current working directory.

`$HOME` – main directory of the user under either `/var` or `/` (root).

`$PATH` – directory where the system would search for executables.

CHAPTER II

REVIEW OF RELATED LITERATURE

This section presents the review of related literature and studies regarding the mechanisms and theoretical principle of RFID, electromagnetism, algorithms and the programming languages, and the derivation of the equations presented in the later section.

According to Standard Model of Quantum Physics (Kibble, 2015) the electromagnetic waves carried by quanta of photons, γ, are product of interactions by electrons (e^-) as shown by Feynman diagram below:

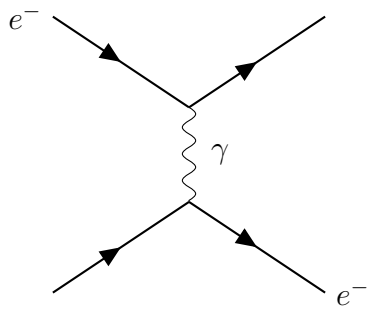


Figure 1: Electron-electron interaction as shown by Feynman's diagram.

Where as the wave function, $\phi(\{\boldsymbol{r_i}\boldsymbol{\sigma_i}\})$, and its energy, E_v^e are determined by the equation:

$$H^e = \sum_{k=1}^{N} - \frac{\hbar}{2m}\nabla_{r_k}^2 + \sum_{k=1}^{N} v(\boldsymbol{r_k}) + \frac{1}{2}\frac{1}{4\pi\epsilon_0}\sum_{\substack{k,k'\\k=1}}^{N} + \frac{e^2}{|\boldsymbol{r_k} - \boldsymbol{r_{k'}}|} \tag{1}$$

Since light is a dual nature, hence the term "wave-particle duality of light" (Giancoli, 2016; Kibble, 2015), which can be shown with Einstein's equation, as energy and mass is interchangeable entity, it can used as a method for efficient information transfer.

Considering an object with mass, m_0, and assuming that it moves under the influence of force, F. According to the Work-Energy theorem, the change in energy, E, is the work done by force, F :

$$E = \int_0^x F\ dx \tag{2}$$

Since $F = ma$:

$$F = \frac{d}{dt_1}(m_1 v_1) = \frac{d}{dt_1}\left(\frac{m_0 v_1}{\sqrt{1 - \frac{v_1^2}{c_1^2}}}\right) \longrightarrow \frac{m_0 a_1}{\left(1 - \frac{v_1^2}{c_1^2}\right)^{3/2}}$$

As if the mass change then force, F, is changed in momentum $F = \frac{d}{dt}(mv)$. Then it follows that:

$$E = \int_0^x F\, dx = \int_0^x \frac{m_0 a_1}{\left(1 - \frac{v_1^2}{c_1^2}\right)^{3/2}}\, dx$$
$$\Longrightarrow m_0 \int_0^x \frac{a_1}{\left(1 - \frac{v_1^2}{c_1^2}\right)^{3/2}}\, dx$$

Since :

$$a_1 = \frac{dv_1}{dt_1} = \frac{dv_1}{dx}\frac{dx}{dt_1} = v_1 \frac{dv_1}{dx}$$

Through substitution E can be rewritten as:

$$E = m_0 \int_0^x \frac{v_1}{\left(1 - \frac{v_1^2}{c^2}\right)^{3/2}} \frac{dv_1}{dx}\, dx$$
$$E = m_0 \int_0^{v_1} \frac{v_1}{\left(1 - \frac{v_1^2}{c^2}\right)^{3/2}}\, dv_1 \tag{3}$$

And finally:

$$E = c^2(m_1 - m_0) \longrightarrow \Delta m c^2 \tag{4}$$

The RFID then utilizes the electromagnetic particles and wave described previously, whereas it carries information based on the wavelength, $\lambda = \frac{c}{f}$, and velocity of the particle. While the matter-energy interaction can further be described using different set of equations, that is deemed not necessary in this review.

While the information sent by the RFID receiver, the processing happens in computer, as the information needs to be further processed for utilization, since in the most basic form it is an entropy which is defined as (McDonnell & Ikeda, 2011):

$$H(X) = -E_{p(X)}[\boldsymbol{log} p(x)] \tag{5}$$

And since information is also a system contained in a universe, it is also assumed to dissipate over time. Zak (2018) mathematically modeled the heat death of universe using the law of thermodynamics that describes the entropy. Madalung (1926) cited by Zak (2018) modeled the hydrodynamics version of the Schrodinger equation through, in which the quantum potential and the classical potential can be seen, last term and F:

$$\frac{\partial S}{\partial t} + \nabla \bullet \left(\frac{\rho}{m}\nabla S\right) = 0 \tag{6}$$

$$\frac{\partial S}{\partial t} + (\nabla S)^2 + F - \frac{\hbar^2 \nabla^2 \sqrt{\rho}}{2m\sqrt{\rho}} = 0 \tag{7}$$

Where ρ and S are the components of wave function, $\Psi = \sqrt{\rho} e^{i\frac{S}{\hbar}}$ and planck constant, $\hbar$ divided by 2π. And by using ansats, Madelung equations can be converted into Schrodinger equation :

$$\sqrt{\rho} = \Psi exp(-i\frac{S}{\hbar})$$

Using Liouville equation:

$$\frac{\partial \rho}{\partial t} + \nabla \bullet (\rho F) = 0 \tag{8}$$

Which was used to generalized the concept of probablity, ρ, generated by system of ordinary differential equations, $\frac{dv}{dt} = F[v_1(t), v_2(t), \ldots v_n(t), t]$:

$$\frac{dv}{dt} = F[p(v)] \tag{9}$$

$$\frac{\partial \rho}{\partial t} + \nabla \bullet \rho F[p(v)] = 0 \tag{10}$$

Equation 9 was generated using Liouville equation generated by Equation 10, which is in contrast to $\frac{\partial \rho}{\partial t} + \nabla \bullet (\rho F) = 0$, nonlinear with respect to the probability density, ρ.

Following the derivation made by Zak (2016), the force, F that plays the role of feedback from Liouville equation, to the:

$$F = c_0 + \frac{1}{2_1}\rho - \frac{c_2}{\rho}\frac{\partial \rho}{\partial v} + \frac{3}{\rho}\frac{\partial^2}{\partial v^2} \tag{11}$$

$$c_0 > 0, c_1 > 0, c_3 > 0$$

Then through reduction, it transforms into :

$$\ddot{v} = c_0 + \frac{1}{2_1}\rho - \frac{c_2}{\rho}\frac{\partial \rho}{\partial v} + \frac{3}{\rho}\frac{\partial^2}{\partial v^2} \tag{12}$$

And then the Liouville equation will transform into :

$$\frac{\partial \rho}{\partial t} + (c_0 + c_1\rho)\frac{\partial p}{\partial V} - c_2\frac{\partial^2}{\partial V^2} + c_3\frac{\partial^3 \rho}{\partial V^3} = 0$$

This equation is known as kdV-Burgers (Korteweg-deVries-Burgers) partial differential equation (PDE) :

$$\int_{-\infty}^{\infty} \rho dV = 1$$

And finally the final change in entropy can be show by assuming that the simplest cases of the system that is $c_0 = 0, c_2 = 0, c_3 = 0, c_1 > 0$:

$$\frac{\partial H}{\partial t} = -\frac{\partial}{\partial t}\int_{-\infty}^{\infty} \rho ln\rho dV = -\int_{-\infty}^{\infty} \dot{\rho}(ln\rho + 1)dV$$
$$= \int_{-\infty}^{\infty} \frac{1}{2}c_1 \frac{\partial}{\partial V}(\rho^2)(ln\rho + 1)dV$$
$$= \frac{1}{2}c_1 \left[|_{-\infty}^{\infty} \rho^2(ln\rho + 1) - \int_{-\infty}^{\infty} \rho dV \right] = \frac{1}{2}c_1 < 0 \tag{13}$$

This difference called *Schrȯdinger paradox* : in a system obeying the second law of thermodynamics, all isolated system, such as this universe, is expected to reach a state of maximum disorder (Zak, 2018).

Thus, to read the information sent by the RFID, computers are necessary, since information is entropy.

Reading the information can be done using programming language, as it is a well established basic fact that kernel interaction is the precedence of programming language or machine language interaction with the input, usually by C and C++ in operating systems, creating a closed loop of: interface $\longrightarrow$ programming langauge response $\longrightarrow$ kernel $\longrightarrow$ hardware $\longleftarrow$ interface $\longrightarrow \ldots$.

Programming languages can be used efficiently with algorithms, which is defined as a set of instructions for a certain task. One exemplification is its utilization in Machine Learning, also known as computational learning theory (Letvina et al., 2021).

One of the other prominent regression algorithm is logistic algorithm which is used for explanation of relation of independent and dependent variable, as shown below:

$$\nabla J(\theta) = \theta + \sum_{i=1}^{m} \frac{exp\left(\left\langle -y_i\hat{\phi}(x_i), \theta \right\rangle\right)}{1 + exp\left(\left\langle -y_i\hat{\phi}(x_i), \theta \right\rangle\right)}(-y_i\hat{\phi}(x_i))$$
$$= \theta + \sum_{i=1}^{m}(p(y_i|x_i, \theta) - 1)y_i\hat{\phi}(x_i)$$

then, the Hessain can be computed :

$$\nabla^2(\theta) = I - \sum_{i=1}^{m} p(y_i|x_i,\theta)(1 - p(y_i|x_i,\theta)\hat{\phi}(x_i)\hat{\phi}(x_i)^\top \tag{14}$$

Simple linear regression on the other hand, which was the regression model the study used, can be derived as follows. Since the direct regression approach minimizes the sum of squares :

$$S(\beta_0, \beta_1) = \sum_{i+1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Then the partial derivations of $S(\beta_0, \beta_1)$ with respect to β_0, and with respect to β_1, can be solved as Equation 2, and 3, respectively :

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_t - \beta_0 - \beta_1 x_i) \tag{15}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_t - \beta_0 - \beta_1 x_i) x_i \tag{16}$$

The solutions of the two equations are called direct regression estimators, or orginary least squares estimators of β_0 and β_1.

$$\frac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_0^2} = -2 \sum_{i=1}^{n} (-1) = 2n$$

$$\frac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_1^2} = 2 \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_0^2 \partial \beta_1} = 2 \sum_{i=1}^{n} x_t = 2n\bar{x}$$

The Hessian matrix, is then given as :

$$H^* = \begin{pmatrix} \dfrac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_0^2} & \dfrac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_0^2 \partial \beta_1} \\ \dfrac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_0^2 \partial \beta_1} & \dfrac{\partial^2 S(\beta_0, \beta_1)}{\partial \beta_1^2} \end{pmatrix} \tag{17}$$

$$= 2 \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \tag{18}$$

$$= 2 \begin{pmatrix} \ell' \\ x' \end{pmatrix} (\ell, x) \tag{19}$$

The matrix H^* is a positive definite if its determinant and the element in the first row and column of H^* are positive. The determinant of H^* is given by:

$$|H^*| = 4\left(n \sum_{i=1}^{n} x_i^2 - n^2 \bar{x}^2 \right) \tag{20}$$

$$= 4n \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{21}$$

$$\leq 0 \tag{22}$$

The the equation or the fitted line or the fitted linear regression model is :

$$y = b_0 + b_i x \tag{23}$$

where β_0 or b_0 is the processed value of dependent variable, and $\theta_1, \theta_2, \ldots, \theta_n$ are the parameters. Based on this, the regression model were defined, $\mathbf{y_i} = m\mathbf{x_i} + b$, where the algorithm will learn the hypothesis using the existing given dataset to predict $\mathbf{y_i}$ using $\mathbf{x_i}$ and

the parameters of the hypothesis, $\theta_1, \theta_2, \ldots, \theta_n$, that can be presented in matrix form :

$$\mathbf{x_i} = \begin{pmatrix} x_{11} & x_{32} & . & . & . & . & x_{1n} \\ x_{21} & x_{32} & . & . & . & . & x_{2n} \\ x_{31} & x_{32} & . & . & . & . & x_{3n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & . & . & x_{mn} \end{pmatrix} \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ . \\ . \\ \theta_j \\ . \\ . \\ \theta_m \\ \theta_n \end{pmatrix}_{n+1,1} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_j \\ . \\ . \\ y_m \\ y_n \end{pmatrix}_{m,1} \tag{24}$$

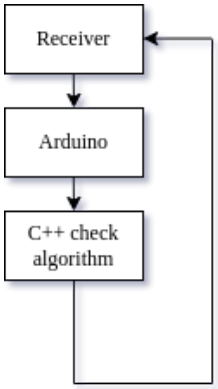
CHAPTER III

This chapter presents the complete description of all procedures in the conduct of the study.

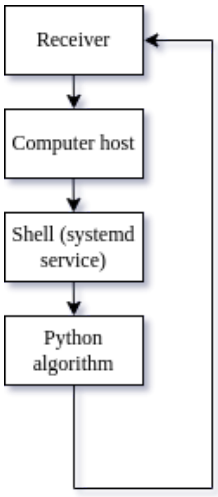
RESEARCH METHODOLOGY

Device Engineering

The lock access and attendance system works similarly to each other in regards of input process. The receiver takes the information or the fingerprinting of the RFID card and relay it to the arduino, while the camera and the RFID information is relayed on to the host system, in case of attendance system.



(a) Mechanism of RFID-based lock access.



(b) Mechanism of the RFID-based attendance device.

Figure 2: Mechanism of arduino based systems. (a.) shows the mechanism of RFID-based lock system, while (b.) shows the mechanism of RFID-based attendance system.

However, it differs in the process of systemically verifying and checking the information received. Figure 1 exhibits the checking proccess of the RFID-based lock access, and Figure 2 exhibits the checking proccess of the RFID-based attendance system. The RFID-based lock access checks the information natively using the code embded in its arduino, written in C++20 with aid of `Arduino.h` and various modules.

On contrary to the RFID-based attendance system, whereas it needs a large database of students and teachers information, it is necessary to use a larger system, which is the host computer. The arduino relays it to the host computer, which would be intercepted by `systemd`

service and relay to the shell, and finally to the checking algorithm written in Python.

RFID-based lock system algorithm

As shown from Figure 1a, the RFID-based locking system uses the native C++20 code inserted to the arduino board, as opposed to the RFID-based attendance that relays the RFID card information to the main system.

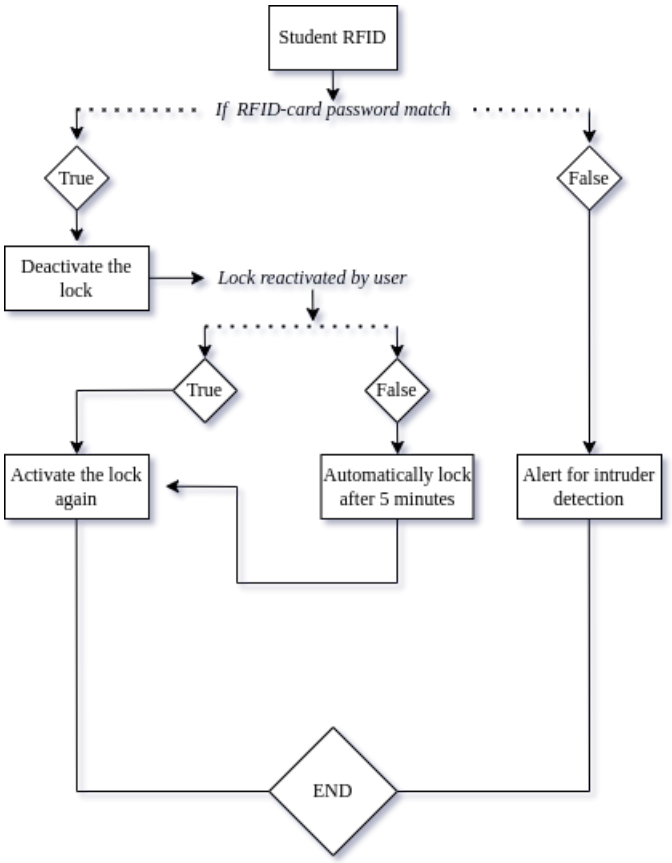


Figure 3: RFID-based lock system algorithm.

RFID-based attendance algorithm

The main algorithm, as shown in the Figure 3 below, of the system will be written in `Python -v 3.9.1` and C++20 with `Arduino.h` with the student database stored offline in JavaScript Object Notation (JSON) file.

There are various third party modules used to aid the python standard library was for successful execution of the function. `pandas.DataFrame()` was used for attendance of the

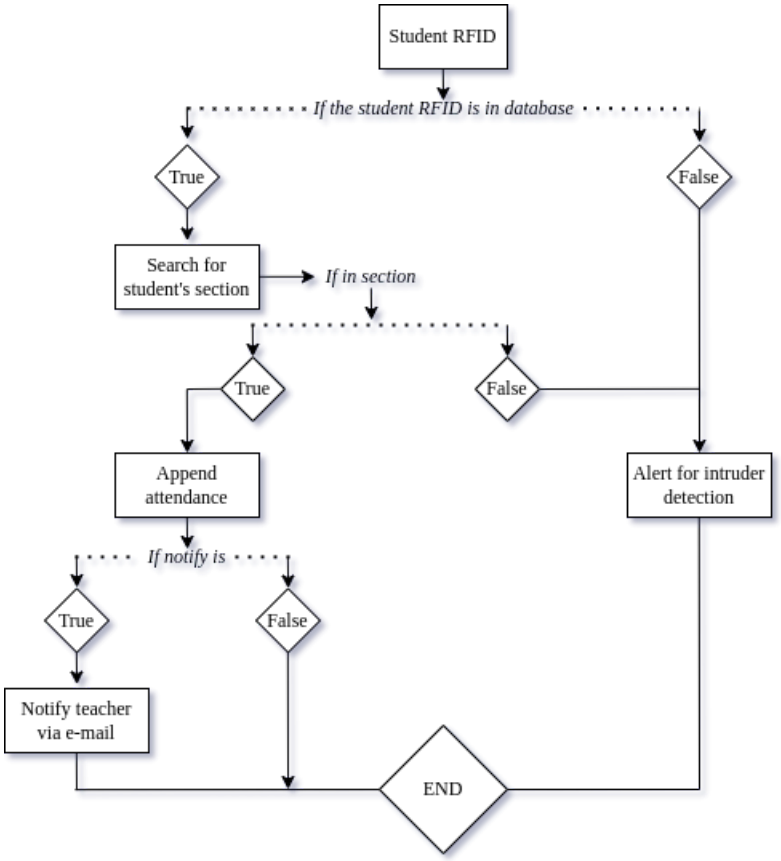


Figure 4: Algorithm of RFID-based Attendance.

students, as well as `pandas.to_excel` function for reporting done to teachers through email everyday. The email was sent using `Python`'s `smtplib` and `ssl`, Simple Mail Transfer Protocol, and Secure Sockets Layer, respectively.

`git` repository was also setup to be able to modify the data through graphical interface, whereas an updated was pulled every $24hours$ by a defined `systemd` service using Linus Torvald's `git`. By iterating over the JSON database of the student ID and teacher's email address to be used for cases of intruders, the checking was performed.

Then an equation was devised to compare two strings. ID comparison system works by comparing the percent difference of two strings using `SequenceMatcher` function provided by `difflib` :

$$\%_{difference} = \frac{|x_{\forall\in\{\theta_1\}} \equiv y_{\forall\in\{\theta_2\}}|}{\sum\theta \mid N_{\theta_1} + N_{\theta_2}} \tag{25}$$

Since the `difflib.SequenceMatcher()` does not make error, the value of $\%_{difference}$

should always be equal to 1. Otherwise, the system would recognize the input as an intruder or unregistered student, which would send an alert to the respective teacher of the student.

STATISTICAL TREATMENT AND ANALYSES

To study the cost efficiency of the RFID-based systems, a regression model was established. Where the algorithm will learn the hypothesis, $\mathbf{h}$ using the existing given data set that it can be used to predict $\mathbf{y_i}$ using $\mathbf{x_i}$ and the parameters of the hypothesis, $\theta_1, \theta_2, \ldots, \theta_n$, in an equation form of : $h_\theta(\mathbf{x_i}) = \theta_n + \theta_o(\mathbf{x_i})$.

The regression model was based on the previous prices of materials used for the device, and traditional tools needed (e.g. Attendance book, padlock, etc.). The data were preprocessed with goal of increasing its sensitivity to be able to capture the precision and accuracy in the regression model :

$$\beta_f = \frac{\beta_i}{\mathbf{ln}(\sqrt{\beta_i})} + |(\theta_1 - \theta_2)| \tag{26}$$