  
  
  
  
  
Improving the Air Quality Index of Ecovillage Boekel with EcohomeAI

***Project Plan***

Version 1.0

March 10th, 2021

Team 05

**HBO-ICT**

|  |  |
| --- | --- |
| Data Student | |
| Family name, initials | Shukla, Akshara  Ludick, Juan Jacques  Stefanova, Yanislava  Nguyen, Phat Thien Phuc  Ouwens, Mel  Zeiss, Dylan |
| Project period (from-until) | Feb 2021 – May 2021 |
| Data Company | |
| Name company/institution | Ecovillage Boekel |
| Department | Eco Village |
| Address | Boekel, Netherlands |
| Company tutor | |
| Family name, initials | Vlems, Ad |
| Position | Client |
| Contact | [ad@ecodorpboekell.nl](mailto:ad@ecodorpboekell.nl) |
| University tutor | |
| Family name, initials | Schellekens, Casper C.H. |
| Project Plan | |
| Title | Improving the Air Quality Index of Ecovillage Boekel with EcohomeAI |
| Date | 10/03/2021 |
| Version | **1.0** |

**Table of Contents**

[**I. Project Statement 3**](#_heading=h.gjdgxs)

[1.1. The Company 3](#_heading=h.30j0zll)

[1.2. Project Lead & Contacts 4](#_heading=h.1fob9te)

[1.3. Problem Description 4](#_heading=h.3znysh7)

[1.4. Requirement Analysis 5](#_heading=h.2et92p0)

[1.5 Feasibility Analysis 6](#_heading=h.tyjcwt)

[1.6. Project Goal 6](#_heading=h.3dy6vkm)

[1.6. Project Deliverables 8](#_heading=h.1t3h5sf)

[1.7. Project Constraints 8](#_heading=h.4d34og8)

[1.8. Project Risks 8](#_heading=h.2s8eyo1)

[**II. Project Planning 9**](#_heading=h.17dp8vu)

[2.1. Project Phasing 9](#_heading=h.3rdcrjn)

[2.2. Communication Plan 10](#_heading=h.26in1rg)

[**III. Main Research Question 11**](#_heading=h.lnxbz9)

[**IV. Sub-Questions 11**](#_heading=h.35nkun2)

[**V. Project Methodology 11**](#_heading=h.1ksv4uv)

[5.1 Software Engineering 11](#_heading=h.44sinio)

[5.1.1 Deployment of Software Solution 11](#_heading=h.2jxsxqh)

[5.1.2 Software Solution 12](#_heading=h.z337ya)

[5.1.3 Basic Flowchart 14](#_heading=h.3j2qqm3)

[5.1.4 System Environment Diagram 15](#_heading=h.1y810tw)

[5.2 Artificial Intelligence 16](#_heading=h.4i7ojhp)

[5.2.1 SGN Goal Implementation with AI 17](#_heading=h.2xcytpi)

[5.2.2 Performance Measurement 18](#_heading=h.1ci93xb)

[5.2.3 Possible Algorithms 22](#_heading=h.3whwml4)

[5.2.4 Data Privacy and Data Fluency 24](#_heading=h.2bn6wsx)

[5.3 Systems and Networks 25](#_heading=h.qsh70q)

[5.4 Cyber Security 25](#_heading=h.3as4poj)

[**VI. Solution Evaluation 25**](#_heading=h.1pxezwc)

[6.1 Minimum Viable Product & GUI Design 25](#_heading=h.49x2ik5)

[**VII. Conclusions and Recommendations 29**](#_heading=h.2p2csry)

[**VIII. Acknowledgments 29**](#_heading=h.3o7alnk)

[**IX. References 29**](#_heading=h.147n2zr)

# I. Project Statement

As Perter Drucker said: “If you cannot measure it, you cannot solve it”. This saying corresponds to our topic of concern in this document. With the sharp increase in economic and technological development of cities and planning towards a more technological driven society, environmental issues continue to rise mainly water, noise, and air pollution. The World Health Organization (WHO) reports that 9 out of 10 people around the world breathe polluted air. As humans, we are the most abundant source contributing solely towards air pollution by using energy to drive our vehicles, power our houses, run our data centres and many more.

Ecovillage Boekel believes in their vision of taking care of the earth, receiving its gifts thankfully and living in harmony with all other lifeforms and works on catering their needs in a very sustainable way by working with all the seventeen sustainable development goals.

The team aims at building the **EcohomeAI** which shows you the air quality in your area and alerts you when levels are ‘bad’. By tracking all the key pollution markers; particulate matter PM2.5, NO2, O3, SO2 and CO gases, you can see a complete and accurate picture of the air you breathe.

## 1.1. The Company

An ecovillage can be defined as “human-scale full-featured settlement in which human activities are harmlessly integrated into the natural world in a way that is supportive of healthy human development and can be successfully continued into the indefinite future.” In short: A small group of houses where a group of people lives in harmony with each other and the world around them.

The ecovillage is in Boekel, Netherlands and comprises of

1. About 30 houses made out of industrial hemp and wood.
2. A community building to gather and celebrate together.
3. Working space with professional tools to keep on building on the village.
4. A food forest to produce their own food.

## 1.2. Project Lead & Contacts

The project client is Mr. Ad Vlems, the founder and the head of constructing the Eco Village. His contact details can be found as below:

Email: [ad@ecodorpboekell.nl](mailto:ad@ecodorpboekell.nl)

The project tutor is Mr. Casper Schellekens, he will be accompanying and providing necessary feedback for our work on different sprints. His contact details can be found as below:

Email: [c.schellekens@fontys.nl](mailto:c.schellekens@fontys.nl)

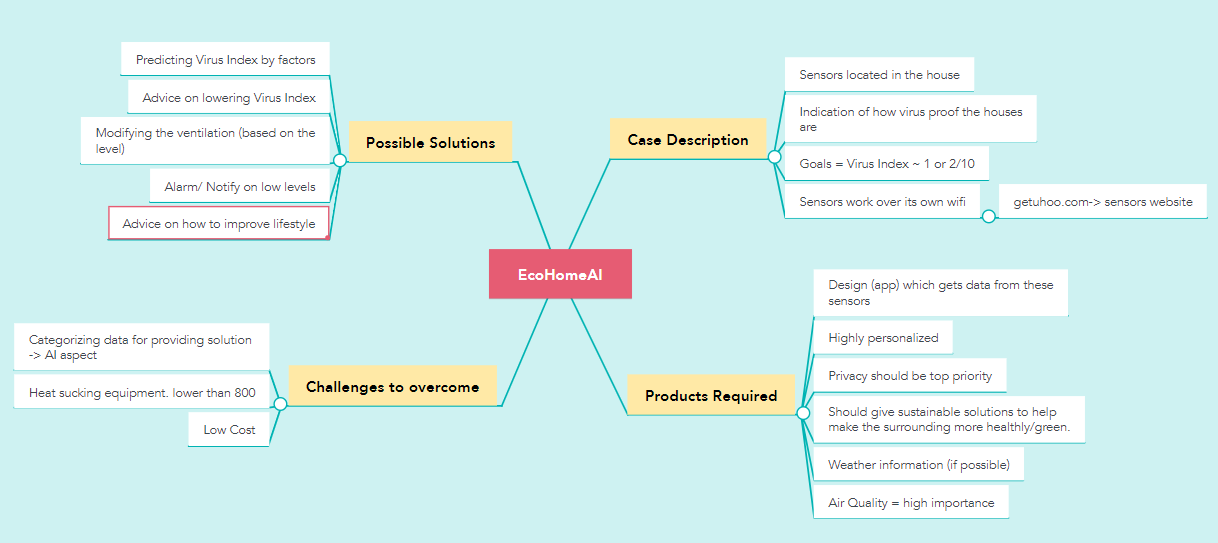
|  |  |  |
| --- | --- | --- |
| **Team Members** | **Role** | **Contact** |
| Juan Jacques Ludick | Software Engineering | [576504@student.belgiumcampus.ac.za](mailto:576504@student.belgiumcampus.ac.za) |
| Stefanova Yanislava | Cyber Security | [y.stefanova@student.fontys.nl](mailto:y.stefanova@student.fontys.nl) |
| Nguyen Phat Thien Phuc | Artificial Intelligence | [p.phatthienphuc@student.fontys.nl](mailto:p.phatthienphuc@student.fontys.nl) |
| Akshara Shukla | Artificial Intelligence | [440692@student.fontys.nl](mailto:440692@student.fontys.nl) |
| Mel Ouwens | Systems and Network | [mel.ouwens@student.pxl.be](mailto:mel.ouwens@student.pxl.be) |
| Dylan Zeiss | Software Engineering | [575965@student.belgiumcampus.ac.za](mailto:575965@student.belgiumcampus.ac.za) |

## 1.3. Problem Description

After analysing the situation of the Eco Village in Boekel for implementing several sensors on different rooms and their aim to provide their residents with the best air purified homes, the following identified issues are listed which provided the team a clear overview of the current situation.

1. **Data from devices lack coordination with outside data:** There is a device that collects massive from the indoor condition. Are linked with outdoor conditions to provide detailed information such as weather, house, virus warning, and global data.
2. **Device data are not fully interactive:** The data shown in the software of the devices is just explicit raw basic data visualization. IT will not show any science or deep understanding of what is happening in the area.
3. **Device sensor cannot bring any right prediction or health suggestion to the customer:** They do have some functions in the application which help customers benefit from that. However, it is still not actively developed by giving any good insight into where they are living.
4. **The software was not implemented with AI:** The current software mainly shows the static exposed and raw data. And AI has not detailed to give any particular action for the user.

Additionally, a mind map was developed for a better classification of different areas which will be covered.



## 1.4. Requirement Analysis

##### System Request

###### Business need:

1. The need is for the system to control, monitor and improve the quality and humidity of the air in the houses in the ecovillage.
2. To allow the residents of the ecovillage to be able to access and manage this data efficiently and securely.
3. The product needs to have a predictive AI system that can make weather predictions into the future and suggestions based on the environmental data collected by the sensors.

###### Business requirements:

1. It should produce management reports.
2. It should give clients an easier way to manage the sensors in their home.
3. It should provide a safer and cleaner environment for the Boekel Ecovillage residents.
4. It should receive data from sensors in the homes in the Boekel Ecovillage and from weather reports in the area.
5. It should be able to automatically recommend potential courses of action for any event that may arise such as the air being too dry in a home or in multiple homes.

###### Business Value:

1. Increase in air quality.
2. It would increase the moral of the residents who live in the Boekel Ecovillage.
3. Increased worker efficiency.

## 1.5 Feasibility Analysis

#### Technical feasibility

* Boekel Ecovillage should have all the technical resources to implement the AI system and maintain it.
* They must have the technical resources improve the current air quality and comfort.
* They need a storage space for all the data that has been collected.
* The product should have a way to operate offline.
* The product should be able to be installed in the designated round living complexes and be able to connect to uHoo air quality sensors independently.

#### Economic feasibility

|  |  |
| --- | --- |
| **Cost** | **Benefits** |
| Technological costs increase. | Improvement in air quality |
| Database set up costs. | Improvement in resident happiness |
| Increase in electricity consumption. | It could attract more investors. |

#### Organizational Feasibility:

The residents of Boekel Ecovillage would prefer a system that is convenient for them as they will be the ones operating the product system and they will be living in the area where the product is active. The product must be simple to understand and maintain as well as smart enough to implement a predictive AI system that can make weather predictions into the future and suggestions based on the environmental data collected by the sensors.

#### Legal Feasibility

1. All user information needs to be kept confidential.
2. All staff that will work with the system would have to sign a confidentiality agreement.
3. Data would have to be encrypted and stored on a secure database on site or on a secure database using a trusted cloud service provider. The terms of this agreement would be outlined within the SLA.
4. Only Authorised personnel should have access to the system itself and the system should keep track of who is using it.
5. Only high-ranking staff should be able to perform maintenance.
6. Agreements with providers must be kept in a secure location.
7. Clients must have multiple ways of proving it is in fact them trying to login.

## 1.6. Project Goal

The goal of the Boekel Ecovillage **EcohomeAI** solution is that it should be a system that improves the air quality inside of the houses of the ecovillage and it should achieve this by collecting data from multiple indoor sensors, weather reports and the user. The system should act upon the data that it collected by recommending solutions to any problems that may arise such as dry air conditions. By tracking all the key pollution markers; particulate matter PM2.5, NO2, O3, SO2 and CO gases, you can see a complete and accurate picture of the air you breathe and where improvements could be made.

The strategic goal of the team involves the below mentioned factors.

1. **Power to the user health:** Recognize virus index to Manage the health of worker and employee in coronavirus risk mitigation. Apply machine learning to measure the combination of index and giving advices.
2. **Artificial Intelligent in your eyes**: Short term event prediction and filter the noise.
3. **AI on Air quality**: Air Quality Data Analytic combined with prediction.
4. **Holistic and calibrated data**: based on the professional monitor devices sensors from uHoo.

We Decided to use a combination of rapid throwaway prototyping, Evolutionary prototyping and Interviewing the stakeholders.

Throwaway prototyping

*Benefits:*

1. You reduce risk in a project by quickly stepping through the initial development phases.
2. You are selecting (or at least you better be) a prototyping language or framework that allows you to rapidly develop your application and meet your prototyping goals.
3. You can usually select a programming language that lends itself to being thrown away – further reducing the incentive to try and “make it work’. Yes, programmers and companies can easily be “hoarders” if they are not careful.
4. The last primary benefit, is that it is in-itself an agile development exercise for the participants and stakeholders in a project that helps define goals, realize potential pitfalls, and generate better risk assessment or the project.

*Risks:*

1. You must commit time and resources to undertake this effort.
2. You may not be using the prototyping time to maximum efficiency; however, throwaway prototyping does not mean you throw and bunch of things at a wall (metaphorically speaking) and see if they stick. This is not a reason to not have a well thought out plan and execute according to the goals you want to achieve.

##### Evolutionary prototyping

*Benefits:*

1. The delivery of the system is sped up.
2. The user engages with the system.
3. The system is more likely to meet the user requirements.

*Risks:*

1. Knowing when it is necessary to stop tweaking the system and finish the development.

The advantages of interview studies, are noted below:

1. It provides flexibility to the interviewers.
2. The interview has a better response rate than mailed questions, and the people who cannot read and write can also answer the questions.
3. The interviewer can judge the non-verbal behaviour of the respondent.
4. The interviewer can decide the place for an interview in a private and silent place, unlike the ones conducted through emails which can have a completely different environment.
5. The interviewer can control over the order of the question, as in the questionnaire, and can judge the spontaneity of the respondent as well.

None of the disadvantages of interviews apply to our project.

## 1.6. Project Deliverables

Following are the list of the deliverables we promise to deliver our client:

1. Project proposal document
2. An online website (holds higher preference)
3. An Android application
4. Exploratory Data Analysis Jupyter notebook
5. Python notebook of entire solution
6. Implementation plan
7. Design document

## 1.7. Project Constraints

Following are the constraints which shall be strictly followed:

1. The project must be completed in 12 weeks.
2. We will be using Python Programming language.
3. Delivering documents on their assigned deadlines.
4. Completing assigned work in every sprint and asking for feedback from the tutor.
5. Implementing agile methodology throughout.
6. Planning regular meetings with the client to discuss our progress.

## 1.8. Project Risks

Following is the risk analysis performed for being aware of the possible risks that may arise and how we can appropriately avoid them.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Id** | **Risk** | **Probability** | **Impact** | **Mitigation** |
| 01 | The dataset provided includes biasedness towards one specific air quality measuring parameter. | Possible | High | Applying a machine learning algorithm which will help in making the data fairer and more trusted. |
| 02 | Lack of variance in model. | Possible | High | Reanalyse the machine learning model and exercise the different data types of present. |
| 03 | Low accuracy rate | Probable | Extreme | Opting for another algorithm which provides better accuracy percentage. |
| 04 | The data of the user’s profile not being secure. | Probable | High | Make sure our database is secure. |
| 05 | Not being able to deliver our technology on time. | Low | High | To have better time management skills and communicate with the client, tutor and amongst ourselves weekly. |

# II. Project Planning

# 2.1. Project Phasing

Since we are applying agile methodology hence, we will be having 3 sprints of 4 weeks each making up to our 12 weeks duration of the entire project. Following mentioned is the plan for the two sprints.



## 2.2. Communication Plan

Below mentioned, is the communication plan designed by the team.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Purpose** | **Medium** | **Frequency** | **Audience** |
| Project team meetings | * Review status of the project * Brainstorm further steps | MS Teams | Weekly Tuesday 12:00 PM CET | All members |
| Software, Cyber Security, System Review | * Brainstorm ideas * Formulate plan | Discord  MS Teams | Once a week | Ludick, Juan Jacques  Stefanova, Yanislava  Zeiss, Dylan  Ouwens, Mel |
| AI review | * Plan further steps * Brainstorm ideas | MS Teams | Once a week | Shukla, Akshara  Nguyen Phat Thien Phuc |
| Tutor Meetings | * Feedback on the progress and documents | MS Teams | Twice per month | All members  Schellekens, Casper C.H. |
| Client Meetings | * Discuss the progress * Ask questions | MS Teams | Once a month | All members  Schellekens, Casper C.H  Vlems, Ad |

# III. Main Research Question

*How can we personalize EcohomeAI technology with AI and Software in augmenting the air quality index at Ecovillage Boekel?*

# IV. Sub-Questions

1. Which machine learning algorithm or techniques are used to predict air quality in the eco village domain?
2. How do the proposed methods handle or predict the different sustainable outcomes and into what categories?
3. What resolutions were made with the proposed techniques?
4. What steps are taken if the data collected from the sensors are not appropriate?

# V. Project Methodology

Since this project includes participation of different domains of ICT mainly Software Engineering, Artificial Intelligence, Cyber Security and Networks and Security, henceforth, this chapter deals with the approach and relevance of each of the domains towards formulating the final solution of our EcohomeAI.

## 5.1 Software Engineering

### 5.1.1 Deployment of Software Solution

1. Built According to your Requirements.

Custom software creation ensures that your company's apps and services can be fully adapted to your needs. It's simple to use and can be applied through your entire business. When it comes to tech, each organization is different, and it's difficult to find a one-size-fits-all solution for everyone. Working with personalized goods not only provides you with a lot of potential for expansion, but it also shows that you support your company's development and success.

1. Custom Software are more Secure.

External hacking risks are minimized when software is designed specifically for your business.

With a personalized software solution, you can encrypt all of your company's data and rest assured that it is safe. Since your product is not available for use by other businesses, the improved protection provided by a custom designed application is unrivalled by any other commercial software.

1. You Can be part of the Design Process.

Nobody knows the industry as you do. You will participate in the process by recommending appropriate functionalities if you invest in custom software creation.

1. Lower Costs

Frequently, the basic software license necessitates the procurement of additional hardware in order to work properly, which may add up to a considerable cost. Custom software creation is very useful and important for companies because it needs less investment than fixed or readily available software. One of the most significant considerations that a business seeks is Return on Investment. A software solution speeds up your workflow, increasing your return on investment.

1. An Advantage Over Your Competition

The standard software solution's first aim is to make your company more effective and usable, giving you an edge over your competitors. Pre-existing apps would most likely be similar to or equivalent to what the rivals are already using. When you have the same resources at your side, you're less likely to be able to successfully distinguish yourself and progress up the market. You will achieve dominance if you have an effective service as a result of the business processes being streamlined.

1. Automation of day-to-day activities

Some tasks in an organization are replicated over and over. Any of them can be automated, freeing up time that can be used to find potential clients, improve employee training, or add new services to what you already sell. If you want to automate your day-to-day business activities, custom software can be a valuable resource.

1. Decrease human error.

When the organization is run by hand, the chances of human error are far higher. You can avoid those errors from being fatal to your company by using a software solution.

1. Integration with other software.

Pre-made software isn't always compatible with your hardware. Custom software creation provides an atmosphere in which it is easier to integrate the product with other software already in use. Custom software removes the need for integration and can seamlessly integrate into any business's software environment, allowing for seamless integration with other software packages.

1. Safety technical support.

A major benefit is that you will have access to a professional support team that was active in the application creation process, ensuring that all of the concerns are easily resolved.

1. Custom Software License Agreement.

When you build custom software for your business, you own the software as well as the code that supports it, giving you more power.

### 5.1.2 Software Solution

We plan on creating an application that can be accessed via any web browser that would allow the user to interact with the data that the uHoo collects through its sensors to help aid in increasing the standard of living for any user who has the product. This application can be expanded on and adapted into an app that can be available for users to download through the app store on their mobile device.

Our software solution will work as follows; data the sensors capture are sent every minute to a database that is connected to our application our application will then make use of AI to create an eco ‘best friend’ that will provide the users with solutions to problems they may face in their household when it comes to air quality.

Security plays an important part in our software solution, passing data through fire walls and ‘clean’ coding will allow us to maintain a standard that can keep the users important data safe and secure.

The solution will be comprised of standard programming languages but the most prevalent one would be Python as all team members are comfortable with the language and we feel as if it would best encapsulate all the aspects from AI to storge.

### 5.1.3 Basic Flowchart

A picture containing diagram

Description automatically generated

The software flowchart for our software solution is rather simple as the entire solution is basically a IFTTT (If Not This Then That). It starts with the sensors on the uHoo device collecting all the data it can from the air, whether that be air pressure, humidity or COVID index level. This data is then processed into useful information that is stored on a database which can either be the cloud or a physical solution such as a server or Arduino.

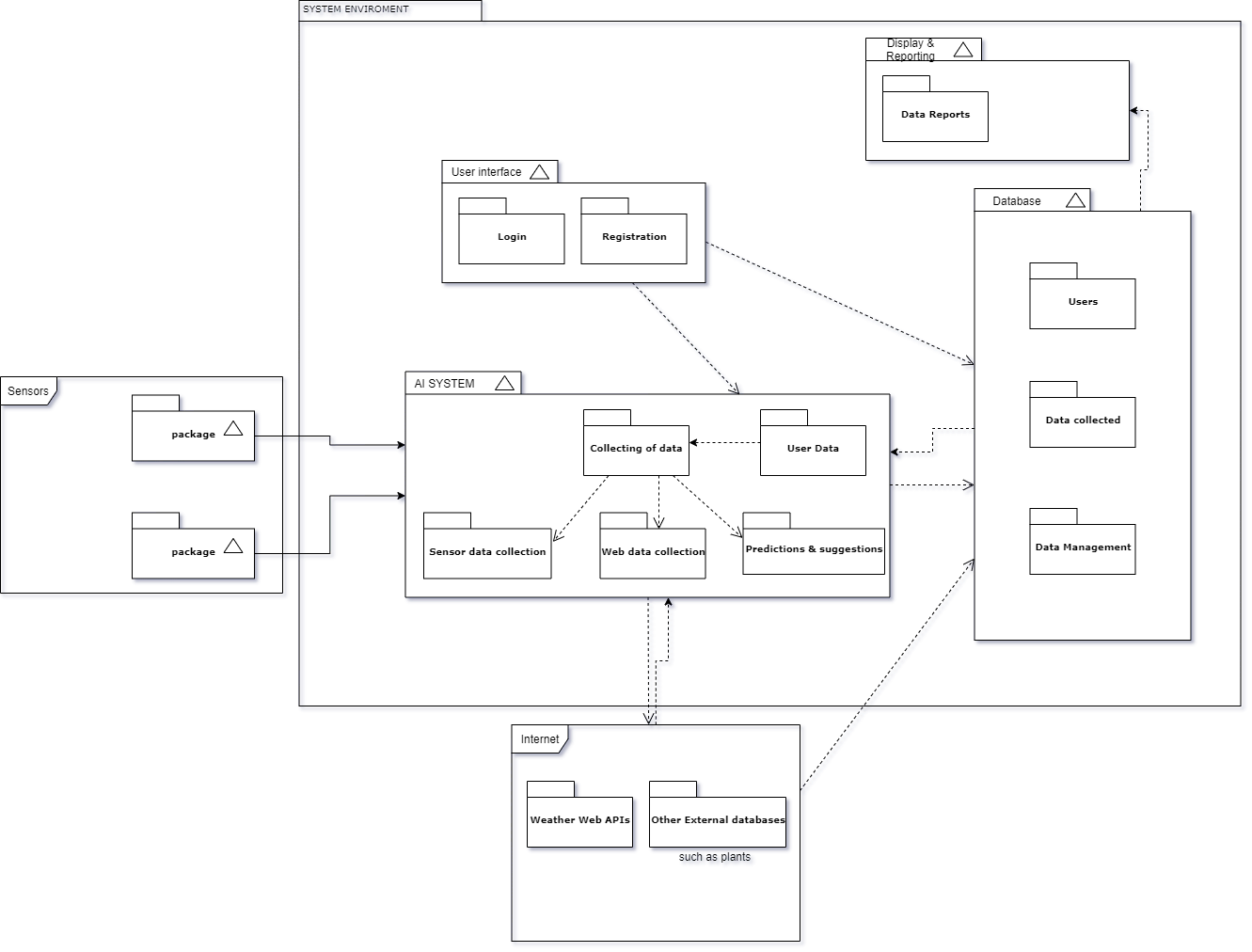
Once stored the information is then shared with the program that we are developing to help improve the air quality and thus improve the overall standard of living, this program will make use of AI to help aid in providing the best possible solution to the user, this solution is based on several factors whether that be financial status, time of day or past solutions accepted by the user.

Once a suitable suggestion is decided by the AI it presents it to the user which can either be feasible and accepted by the user or not feasible and declined by the user, if the suggestion were to be declined the AI would go back and come up with the next best solution that the user is more likely to accept, this would continue in a ‘loop’ until the user id finally happy with the suggestion.

When the user accepts the suggestion, it is thought that the user would commence with what they were asked and an increase in air quality would be seen and collected by the sensors. Depending on what suggestion the user accepted, the answer would be stored and sent back to the database to help improve the AI’s decision-making next time.

Although basic it is thought that this process would occur often as it's important for the AI to make less ‘mistakes’ in its suggestions to minimize the number of times the user will not accept the suggestion it provides.

### 5.1.4 System Environment Diagram



## 5.2 Artificial Intelligence

###### Technical apply on AI process:

The technical aim includes performing analysis with Jupiter Notebook in Python, Tableau, and co-ordination into the software after all AI method apply to bring the simple user interface for end-user.

*Task Defined in teamwork:*

|  |  |
| --- | --- |
| **Data Engineer** | **Machine Learning** |
| Akshara will handle the main part of Data EDA analysis and preparing data. Mark will handle the feature function. | Mark and Akshara both developing the concept of AI solution into the Device and software. |

### 5.2.1 SGN Goal Implementation with AI

This section discusses how our solution is going to be corresponding to the different sustainable development goals (SDGs).

The project would have multi-discipline, and the core of AI had the care of sustainability product. It perpetuates how we can educate humans and users about climate. Also, with the implementation, we will suggest an eco-friendly user interface. Hence, the mindfulness of different factors such as goal 12, responsible consumption, and production. The ability to use the product with easy-to-read and adaptability with green environment. CO2 emission data is another key to magnifying and identifying how the manager will construct the house and check progress.

The function of AI on the device can broaden more to goal 11 Sustainable cities and communities. AI solutions elicit and inspire how a user can impact to eco-system. Knowing how to modify the house system, manage the virus index from the house to breathe the healthy air by advice from the software indifference area of the village. In short, the team helps eco-village technical solutions interact with the problem that the globe is facing. Making the excellent quality to satisfied level for each goal is our mission.

#### Implementation of Agile Approach

Since the approach is highly going to be working in an agile environment to ensure the product we are delivering is fully up to the mark of our client, we have broken down the measures we will be taking into consideration when applying various machine learning algorithms in developing our Artificial Intelligence.

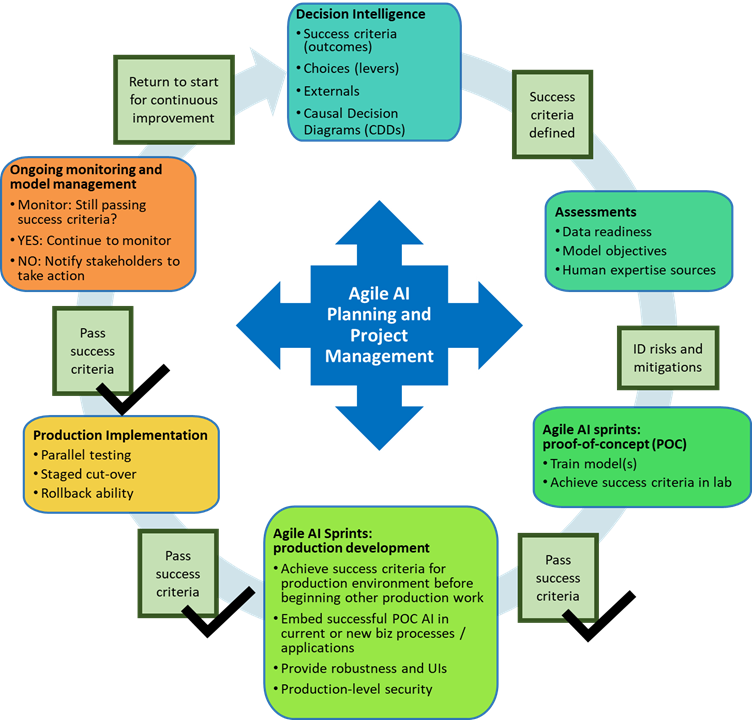


Figure 1. Agile AI method for ECO village in Artificial Intelligent

### 5.2.2 Performance Measurement

#### Dataset

We have been provided with the dataset of minute-by-minute record of the key air quality parameters in the bedroom and living room from our client for the month of January 2021. Both datasets include approximately 34485 rows.

Both datasets data includes minute by minute record of the necessary air quality measuring variables which include:

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of Data** | **Definition** | **Unit of Measurement** | **Solution Approach.** |
| Date and Time | The time frame for the month of January. | Minutes, Seconds | Prediction and accuracy percentage. |
| Temperature | The temperature of the rooms with installed sensors | Celsius | Predictive and coordinate with index outside |
| Relative Humidity | The measurement of the amount of water vapors in the air. | Percentage | Air Quality Prediction |
| PM 2.5 | Particulate Matter 2.5, an air pollutant which reduces visibility and causes air to be hazy when levels are elevated and is prone of lungs and heart. | Micrograms per cubic meter (μg/m3) | Air Quality Analytic |
| CO2 | The natural greenhouse gas, carbon dioxide. | Parts-per-million (ppm) | Air Quality Analytic |
| TVOC | Total Volatile Organic Compounds (tVOC) is the total concentration of all volatile organic compounds such as benzene, ethylene, glycol etc. present in the air. | Micrograms per cubic meter (μg/m3) | Air Quality Prediction |
| Air Pressure | Measured by a device called barometer is the pressure within the atmosphere within the atmosphere of Earth. | Kilopascals | Air Quality Analytic |
| CO | The odorless and colorless gas, carbon monoxide. | Parts-per-million (ppm) | Air Quality Prediction |
| Ozone | Ozone (O3) is the highly reactive gas corresponding directly to causing chest pain, shortness of breath and throat irritation when inhaled. | Dobson Unit | Air Quality Analytic |
| NO2 | Nitrogen Oxide primarily gets in the air from the burning of various fossil fuels. It can cause reduced lung function, increased asthma attacks when present in abundance. | Microgram (µg) | Build Air Quality Prediction |
| Virus Index | These levels are determined based on scientific analysis of five air quality factors which includes temperature, relative humidity, PM2.5, Carbon Dioxide, and Nitrogen Dioxide. | Integer | Apply for virus prediction index. |

The organization providing the sensors is uHoo and they essentially came up with their own virus index by measuring the above listed variables which is mentioned below. This index will fundamentally be used in classifying whether a specific room is healthy or not.



In the process of performing our exploratory data analysis, both datasets will be divided into separate training and testing data samples. The goal would be finding if there exists any biases and to implement an algorithm in the model to decrease variance, bias and improve predictions. Additionally, the size of the training set will be greater than the testing set to avoid overfitting and gain better accuracy percentage.

#### Modelling





|  |  |
| --- | --- |
| **RATING** | **MEANING** |
|  |  |
| **Excellent** | According to current research, negative impacts on ecosystems are unlikely. |
| **Fine** | All values are under the legal health protection limits. Effects on ecosystems can no longer be ruled out. |
| **Moderate** | The health protection limits are mostly still met. Effects on ecosystems are increasingly possible. |
| **Poor** | The measured values are at the level of health protection limit values. Health impairments of sensitive persons may occur sporadically. |
| **Very Poor** | The health protection limits have been exceeded. Health impairments of sensitive persons are possible. The population is increasingly informed about the pollutant situation. |
| **Severe** | The measured values are at alarming levels. The health protection thresholds are clearly exceeded. Health impairments of all persons are possible. |

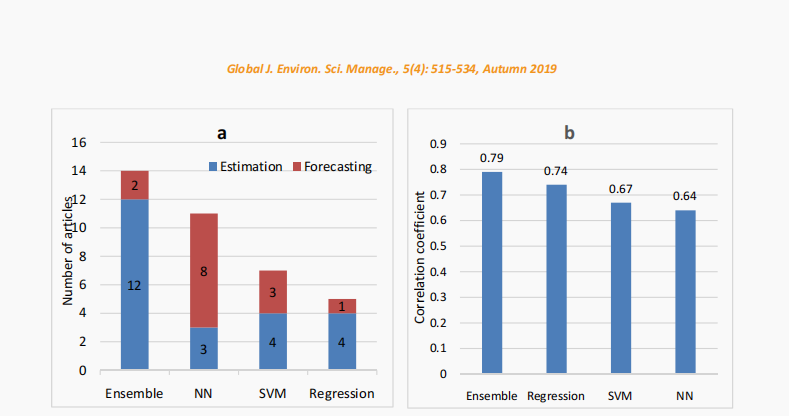
First,we are going to be working with the data collected from the device called GitUhoo, it might give massive data involving different factors. During the device-organized process, they might have some errors or missing values that we must deal with. Therefore, dealing with data preparation such as cleaning and visualizing on how the data will be captured and appear in software based on our software and infrastructure will develop in the next two months. After giving the conclusion or any specific user advice, the modelling of data is picked up. The project will be dealing with big data; the more the house gets old, the more data collected from the device. Mainly a data server will be developed to store the data.

Air quality will follow the method of combination with different data, we calculate the data training set and give the statistic to measure the accuracy and make prediction from modelling. Our air quality index is calculated based on the averages of all pollutant concentrations measures in a full hour, a full 8 hours. Move for the predictive on AI part for next 8hr in a day, next 24 hr. (e.g., between 9:00Am and 10:00 AM)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rating** | **O3 (ΜG/M³)** | **PM2.5 (ΜG/M³)** | **PM2.5 (ΜG/M³)** | **NO2 (ΜG/M³)** | **CO (MG/M³)** | **CO (MG/M³)** | **NH3 (ΜG/M³)** | **VOC (INDEX)** |
| **Timeframe** | 1h / 24h | 1h | 24h | 1h / 24h | 1h | 8h | 1h / 24h | 1h / 24h |
| **Calculation method** | max. hourly value | average | average | max. hourly value | average | average | max. hourly value | max. hourly value |
| **Excellent** | 0 - 33 | 0 - 7 | 0 - 5 | 0 - 25 | 0 - 2 | 0 - 1 | 0 - 3 | 0 - 50 |
| **Fine** | 33 - 65 | 7 - 15 | 5 - 10 | 25 - 50 | 2 - 4 | 1 - 2 | 3 - 7.5 | 51 - 100 |
| **Moderate** | 65 - 120 | 15 - 30 | 10 - 20 | 50 - 100 | 4 - 8 | 2 - 4 | 7.5 - 37.5 | 101 - 150 |
| **Poor** | 120 - 180 | 30 - 55 | 20 - 25 | 100 - 200 | 8 - 30 | 4 - 10 | 37.5 - 15,000 | 151 - 200 |
| **Very Poor** | 180 - 240 | 55 - 110 | 25 - 60 | 200 - 400 | 30 - 100 | 10 - 30 | 15,000 - 150,000 | 201 - 300 |
| **Severe** | from 240 | from 110 | from 60 | from 400 | from 100 | from 30 | from 150,000 | 301 - 500 |

### 5.2.3 Possible Algorithms

According to the report of machine learning algorithms in air quality modelling which was published in 2019 from Global J.Envision.



From the above research graph, we can come out with some scientific thought and some comparisons. The algorithm to apply into the notebook and prediction should stay on a high level of the correlation coefficient.

Therefore, the possible algorithms to be implemented will be the  ***Ensemble*** method, which covers multiple techniques:

1. **Bagging:** Bagging is the ensemble method behind powerful machine learning algorithms such as random forests. In this part, we can utilize the theory behind this technique and build our bagging models using the sci-kit-learn library. Split it in Voting part, and Average.

* **Voting** will combine majority model and apply classification.
* **Averaging** will combine the average mean of a model and apply classification on difference index NO2, Virus, Temperature. Some another ensemble method might use as a call of Heterogeneous.

1. **Boosting:** Boosting is a class of ensemble learning algorithms that includes award-winning models such as AdaBoost. In this part, use it to predict the temperature index of this dataset! And apply gradient boosting algorithms such as CatBoost and Boost. Some of the benefits might mention reducing variance, avoiding overfitting, and more stability and robustness.

* This one will be used for estimation of the data of PM2.5, temperature, air pressure, and CO2, which relate to people's health. The progress of ensemble packing with multiple method algorithms (Regression, Decision Tree etc.). With 79% accuracy level based of the correlation coefficient. It is a strongly fraction of F\_1 score to bring better accuracy after modeling prediction.

1. **Stacking:**it is the technique where we can know the model, such as where is strength and weakness which can be adjusted as an effective combiner model. With the aim after that to define tasks, and delegate responsibilities. Some of the functions will take part and participation. For this technique, we can build a different layer and apply Gaussian Naïve Bayes Classifier, with KNN using “ ball tree” algorithm. And the second layer of meta estimator. For all of them will stack the ensemble together for the final prediction step.

* As the diagram, it shows that NN (K-nearest-neighbored) show that they achieved 64% percent. And following by that is SVM, it is called a support vector machine. Following the visualization graph, we can see that NN helps us forecast the virus index and temperature, humidity with 8/4 percent.

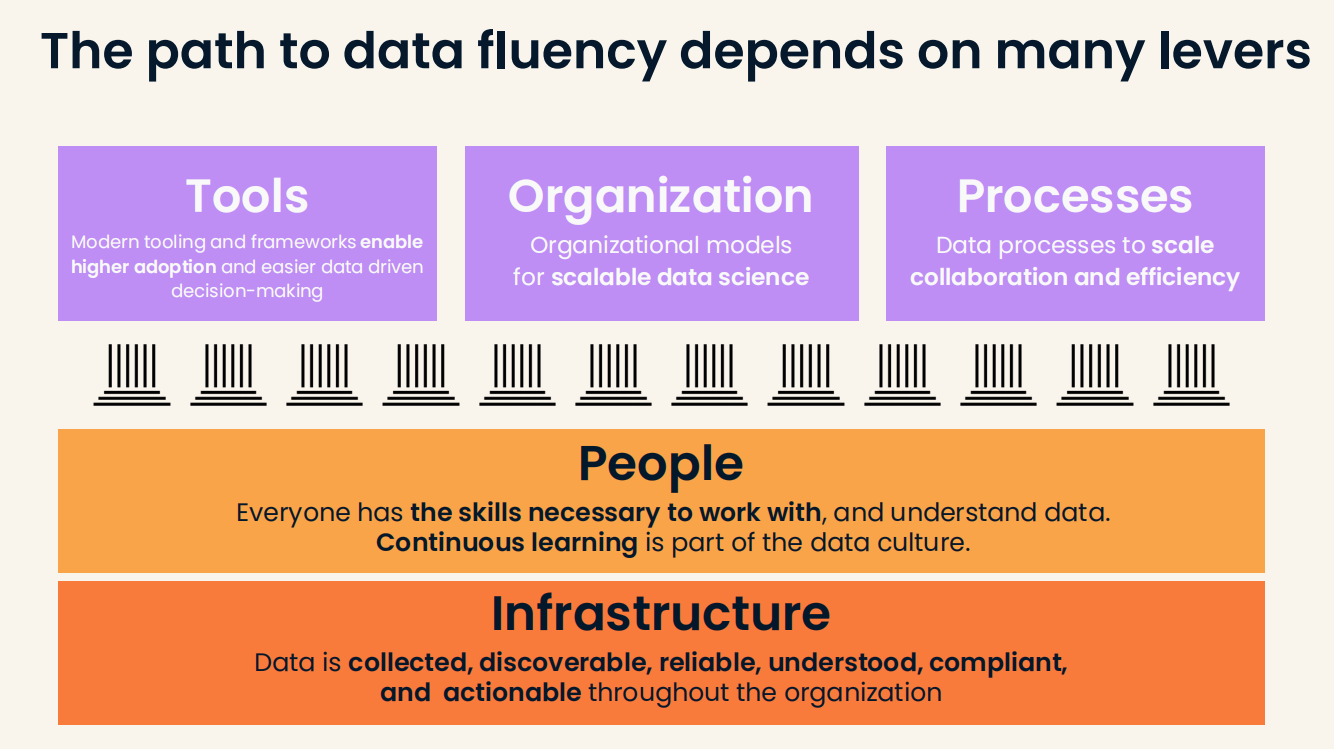
### 5.2.4 Data Privacy and Data Fluency

##### a) Data Privacy

1. According to EUROPEAN DATA PROTECTION SUPERVISOR: apply from machine learning due to the interaction within software applications and the sensor's datasets. It would have considered using privacy technique service to save our private data from the outside business. (it can be able to share)
2. Would apply DATA wall in the application, allow strict permission level.
3. We are making the privacy user in the cloud technique, which relates to databases.
4. Only manager and Eco-village permission user are able to interact with the AI predictive output excels file.

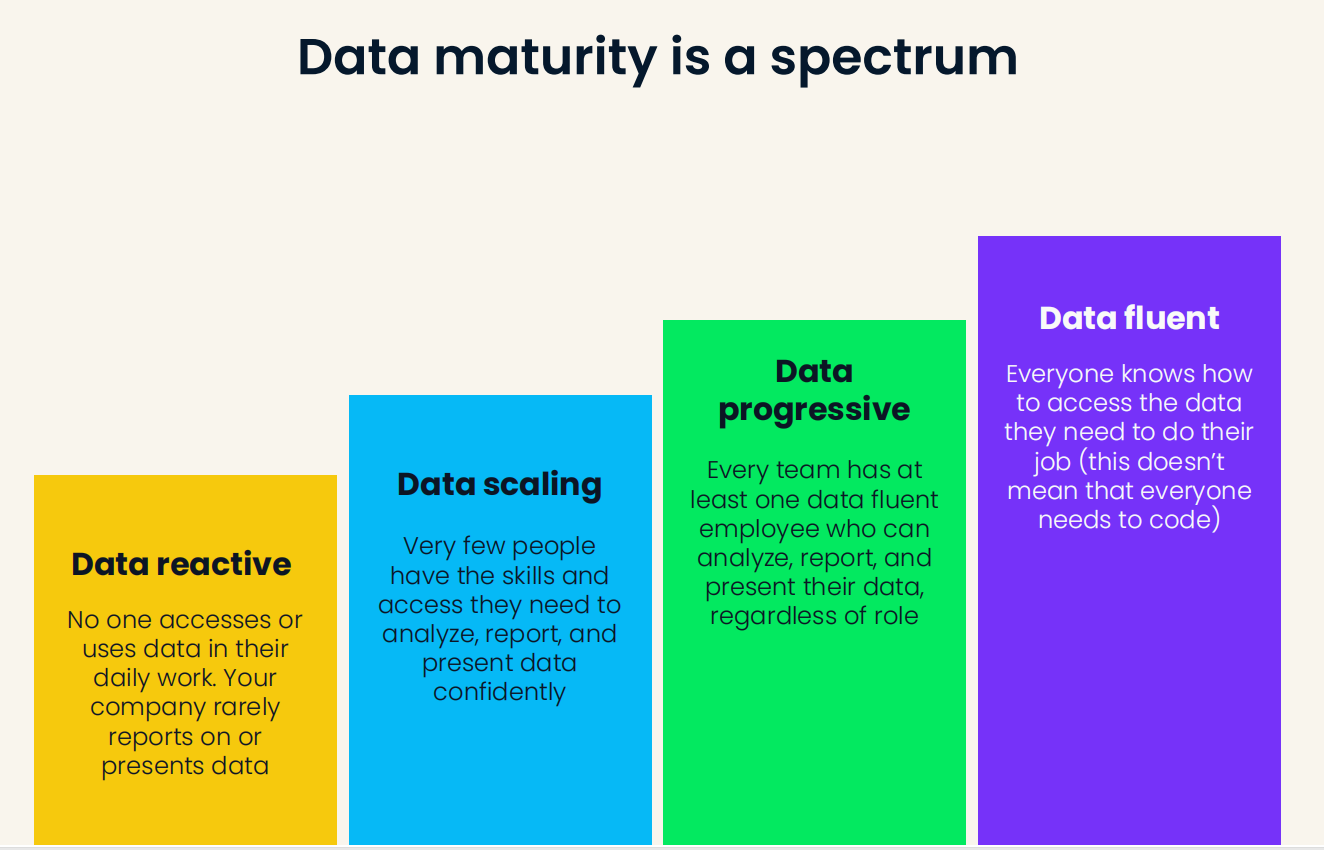
##### b) Data Fluency

* To manage the Data Fluency for this project would need:



Tool, organize, and process in this project we have multiple stakeholder such as manager,

Tool, organize, and process in this project we have multiple stakeholders such as manager, developer, team software. That is why data fluency should be clear on this point following the responsibility of the AI team. It would make everyone understand the works behind the software and AI product. Some of result criteria would be mentioned for the outcome understood, compliant, and actionable for the Eco-village product.



**Figure 5 DATA SPECTRUM**

The picture explains for every step into Data Fluent from this project. From reactive to scaling and progressive and data fluent, there for software team, cyber security, and infrastructure are all able to work together with AI project and approach the Data Maturity spectrum.

## 5.3 Systems and Networks

The view on this project is to develop and do research on an indoor air quality notification service that notifies the users about ways to improve the current air quality and comfort. And can make predictions into the future to give those alerts proactively.

Also due to the ecovillage's autonomous nature of wanting to provide water, electricity, and heat independently. It would be fair allow this service to work independently and potentially offline as well if not available, making weather predictions with Artificial Intelligence and able to be installed in the designated round living complexes and able to connect to uHoo air quality sensors independently.

## 5.4 Cyber Security

Cybersecurity is an important part of our software solution. We are going to identify the vulnerabilities and risks that we might encounter by doing risk analysis and find out how to prevent them by designing, implementing and testing a secure product.

We should make sure our database is secure to protect the personal information of the users and the data from the sensors that the application uses.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Probability** | **Impact** | **How to prevent** |
| SQL Injection | Probable | High | Firewalls  Encrypt the password hashes |
| XSS  (Cross-site scripting) | Probable | High | Escaping user input  Encode potentially harmful characters |
| CSRF  (Cross-site request forgery) | Probable | High |  |
| Password cracking | Possible | High | Require strong passwords  Accounts being locked after  a few failed login attempts |

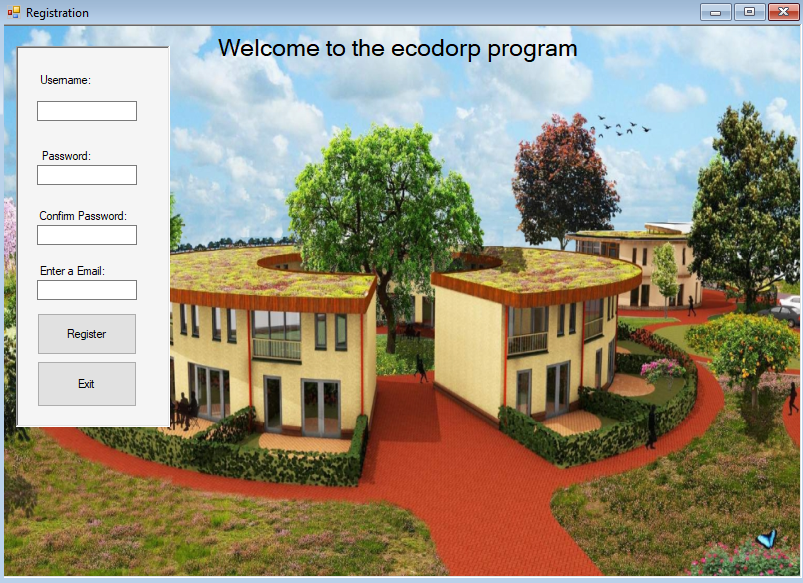
# VI. Solution Evaluation

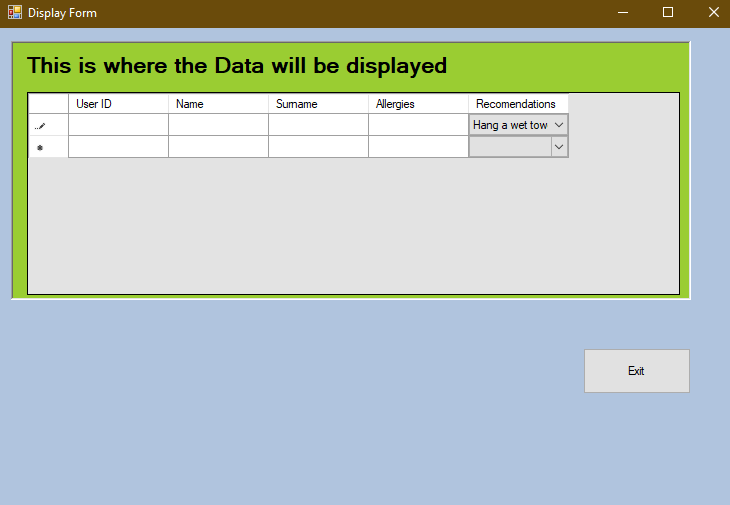
## 6.1 Minimum Viable Product & GUI Design

From the algorithm prediction we can bring out some of advice and solution for current activity follow 4 categoricals which give the quality advice for the customer in a specific moment.

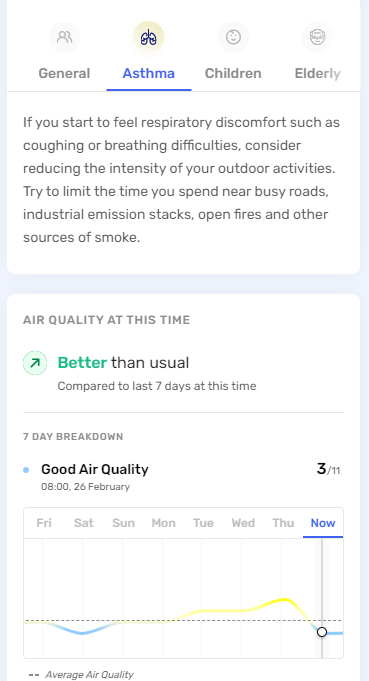
##### Web Application GUI Designs

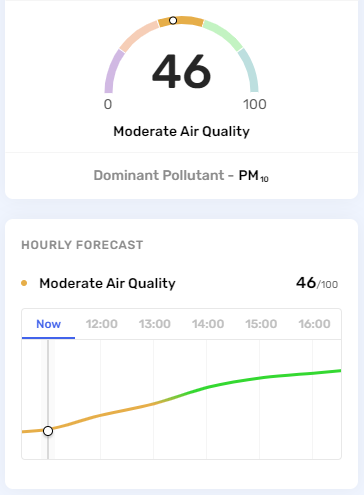




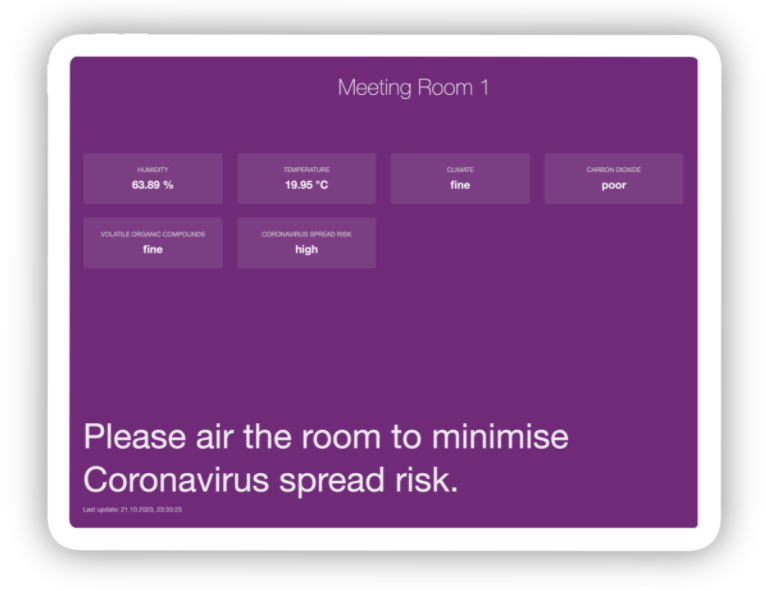


##### Android Application GUI

***General – Asthma – Children-Elderly***







# VII. Conclusions and Recommendations

To conclude the work distribution of sprint 1 which consists of 4 weeks was completed on time. Additionally, weekly meetings with the tutor and between the team members helped the team to think outside the box for producing a fair and new solution for improving the air quality index in the ecovillage of Boekel. For the next sprint, the brainstorming of our next actions may follow with the focus relying mostly on comparing algorithms, wireframes of software solution and constructing a database.

# VIII. References

Bailey, K. (. (n.d.). *www.sociologygroup.com*. Retrieved from sociologygroup.com: https://www.sociologygroup.com/advantages-disadvantages-interview-research/

Department of Health in the US. (n.d.). *www.health.state.mn.us*. Retrieved from health.state.mn.us: https://www.health.state.mn.us/communities/environment/air/toxins/voc.htm

Ecodorp Boekel. (n.d.). *www.ecovillageboekel.nl*. Retrieved from ecovillageboekel.nl: https://www.ecovillageboekel.nl/how-build-ecovillage-go-with-flow/

FDGweb. (2021). *www.fdgweb.com*. Retrieved from fdgweb.com: https://www.fdgweb.com/throwaway-prototyping/#:~:text=The%20main%20benefits%20of%20throwaway,and%20meet%20your%20prototyping%20goals.

Mauro Castelli, F. M. (2020). *www.hindawi.com*. Retrieved from hindawi.com: https://doi.org/10.1155/2020/8049504

teach-ict.com. (n.d.). *www.teach-ict.com*. Retrieved from teach-ict.com: https://www.teach-ict.com/as\_a2\_ict\_new/ocr/A2\_G063/331\_systems\_cycle/prototyping\_RAD/miniweb/pg3.htm

USA.gov. (n.d.). *www3.epa.gov*. Retrieved from epa.gov: https://www3.epa.gov/region1/airquality/pm-human-health.html