Problem Statement and Goals
Software Engineering

Team 21, Alkalytics
Sumanya Gulati - gulats10
Kate Min - mink9
Jennifer Ye - yej52
Jason Tran - tranj78

Table 1: Revision History

Date	Developer(s)	Change
09-19-2024	Jennifer Ye	Inital Rough Work
Date2	Name(s)	Description of changes
...	...	...

1 Problem Statement

This project aims to aid in the data management and analysis of an ocean alkalinity enhancement experiment process.

1.1 Background

The goal is to develop a scalable process to capture CO_2 using a combination of electric fields and membranes. The experiment process creates a very dilute base using a very efficient generation method. Although this process is thermodynamically efficient, the low concentration limits practical applications. As earth's temperature rises, so does the ocean's, affects its ability to absorb CO_2. However, the resulting product of this process can be used to increase the CO_2 absorption by adding the very dilute base from the experiment back into the ocean. This could be a great additional solution to the battle of climate change but will require a massive volume of solution to have a significant impact.

1.2 Problem

As this process is still on a small scale, it is currently being perfected which means many more experiments must be done to be able to bring it to a global

scale. The process has many niche technical challenges which requires high precision for success. Optimization of data is critical to improve process efficiency. This is a big data software problem requiring the ability to find and fine-tune specific parameters.

The team aims to provide a solution to store and manage large amounts of data being collected in the ongoing experimental system. The data must be labelled, organized and have the ability to search and compare across datasets for details efficiently.

1.3 Inputs and Outputs

Currently the data is being stored across Excel sheets. However, the data can be exported to a Comma Separated Values (CSV) file. This will be the main input for the Alkalytic project. The data collection device will also export the data as a CSV file. There are no other input file types.

There are two possible outputs from this project. The exact choice will depend on the users' needs. The output could be an exportable CSV file of the desired data. It could also be a visual plot of desired data. Both are outputs of this project, but which one will be produced is decided by the user.

1.4 Stakeholders

Since this experiment study is still on-going, it is of small scale.
The main stakeholders currently are:

- Dr. Charles De Lannoy
  The project supervisor. He is the one who discovered the alkalinity enhancement experiment process and is working on expanding the scale.
- Bassel Abdelkader
  A Doctor of Philosophy student who is working along side Dr. Charles De Lannoy in running the experiments and analysis.
- Current students/members of the lab working on the study
  There may be new students getting involved with the study project. However, they may be there for a short period of time. Although they will still be interacting with the experimental data, they may not be a primary stakeholder.
- The founder of the study
  The founder of the study is the one who is currently funding the research project. They will be interested in any new tools being used to further enhance the analysis or data management process.

1.5 Environment

This software will be compatible with any browser on a computer that is connected to the internet. This software will also be compatible with Windows 10 and 11. There is no hardware component.

2 Goals

This project has a total of six goals.

2.1 Consolidate Organized Accurate Data with Proper Labels

The raw data input for this project is a CSV file exported from Excel. With hundreds of experiments, where each has ten or more attributes will account for thousands of data points. As the study gets larger, the current data collection format is not sustainable. As such it is a goal to be able to migrate all the existing data to another solution where the solution is not only scalable but also extendable. To migrate all the data also means that the data is properly labelled as it had been originally. This goal can be measured by a percentage of data properly imported and labelled.

2.2 Querying from Multiple Sets

The experiment has multiple pages of Excel sheets that data is recorded on. Although the data is separated and organised across sheets, in order to analyse the data, queries will need to be made. This project should be able to pull data across sheets quickly and efficiently. Currently, retrieving the data is manual, slow and tedious. This project strives to make data retrieval easy and efficient. This goal can be measured by the time it takes for a query to complete.

2.3 Inter-Parameter Comparability

To further aid the data analysis of the experiment data, the product should be able to compare multiple parameters. This could be a simple comparison of two attributes or more, but also narrowing down the search windows. One of the goals for this project is to facilitate control over the data to be seen, as well as the ranges for that data. This goal can be measured by the number of parameters that can be search/compared at once.

2.4 Web Interface

A web interface allows the user to easily interact with the backend processes. This interface should be user friendly and displays all the functions that are needed to input data, and return the appropriate data. The returned data should be given in a readable format. This goal can be measured by how many

of the required tasks it can successfully perform, and if they are clear for the user.

2.5 Flexibility in Expansion (i.e. adding additional parameters)

Although a big portion of this project is to migrate the existing data into a better solution, the product should also be built with the possibility of expansion. As the experiment grows, there could be a new parameter to be recorded. There could also be the possibility for this tool to be used in other experiments that do not follow the same format. The product should still be able to handle those new changes, among others. It is our goal to be able to create a tool to be used as a general data recording tool. This goal can be measured by how well the application adapts to unconvential data.

2.6 Visual Plot of Specified Parameters and Results

The product should be able to use the data it retrieves and interpret it in a visual manner. This product must be able to plot the data in the desired format of the user, which can then be exported. The data will not be useful if it is not readable or understandable to the user. This goal can be measured by the quality of the generated visualizations.

3 Stretch Goals

Stretch goals show what additional features that could be implemented as the main goals of the project are achieved. There is a total of four stretch goals for this project.

3.1 Automatically Download Data from Data Collection Device

The current state of the experiment requires a data collection device. This device only holds a certain number of experiment trials and the data must also be manually exported at the end of each day. If the recorded data is not exported within a week, all the data will be overwritten. As a stretch goal, the product could be connected to this device and upload the data either when it finishes collecting data per experiment or once it hits a predetermined limit. This is a stretch goal as it requires the team to gain a deeper understanding of how the device records the data. This goal can be measured by how often data is successfully automatically downloaded, or by how little data is missing.

3.2 Dynamic Dashboard to Generate Comparison Reports

As an extension of one of the main project goals, the visual plot can be in the form of a dashboard with a small written conclusion of the data requested instead of just the plotted graph. The choice of visual plot can also be given to the user. The dashboard could dynamically adapt to additional information, user interactions and removals without having to regenerate the entire plot from scratch. This goal can be measured by how well the plots changes dynamically.

3.3 Mobile Development and Accessibility

This project is intended to have a website interface. As a stretch goal, the project could be extended to a mobile app for easy access and reachability. The website and mobile app could also incorporate more accessibility features for those with disabilities. This may be but not limited to colour contrasts, screen readers, big fonts and responsiveness to those who have trouble seeing. This can be measured by how many of these extra features are successfully implemented.

3.4 Machine Learning Analysis and Projections

One major stretch goal is to incorporate artificial intelligence into the data analysis. In hopes to make data analysis easier for the user and more adaptable to any possible analysis. By introducing a machine learning algorithm, it would be able to automatically query the data, customise that query, analyse the data and create a customizable dashboard board as well based on any sort of prompt. The machine learning aspect can also automatically return conclusions based on prompts.

4 Challenge Level and Extras

The assigned challange level for this project is general. As this project does not have any direct need for research on the development side, the features and abilities needed to develop may still be more complicated than aniticipated.
There are two extra components the team has chosen to take up for this project.

- Usability Testing
- User documentation

This is list of extras is subject to change. However the team has decided that these extras will not only benefit us as a team with ensuring the project is continuously moving forward but also ensuring that the stakeholders are aware of how the project is progressing.

Appendix — Reflection

1. What went well while writing this deliverable?

   When writing this deliverable, it was simple in a sense where many of the topics and ideas presented here were talked about before hand with both the team and the supervisors. It was more of a matter of expanding on ideas and topics in further detail to ensure that the problem and goals were aligned with previous conversations. This allowed the writing process to be more time consuming than actually being difficult.

2. What pain points did you experience during this deliverable, and how did you resolve them?

   The biggest pain point during this deliverable was figuring out how to properly explain the problem and goals. Many if not all ideas presented were simply ideas. It was quite difficult to be able to not only explain the goals but to also connect it back to why they are important to have in this project. Displaying the understanding of how pieces fit together and explaining why certain things are in scope was the most challenging part of this deliverable.

3. How did you and your team adjust the scope of your goals to ensure they are suitable for a Capstone project (not overly ambitious but also of appropriate complexity for a senior design project)?

   As a team, this was very much a concern in the beginning. This resulted in our team setting up strong measures of communication. Before even starting on this deliverable we made sure to go through each section with the supervisor. We had discussions throughout the meeting where the team brings a more technical point of view to aid in the supervisor's needs and wants. It was through this process that we did not have to adjust the scope of the groups goals much as we started with the a strong foundation and understanding with all parties involved.