National College of Ireland

Master of Science in Artificial Intelligence

Programming for Artificial Intelligence (MSCAI1)

2024-2025

Bintong Chen

23135808

Final Project Journal

Jan 3, 2025

Dec 10, 2024

Today our team had our first meeting to discuss the project. We decided that we would like to work on the topic of sustainability, as we studied sustainability in AI during the Data Governance and Ethics module and now have a good understanding of this subject. We decided on the topic "Data-Driven Sustainability: Tracking Air, Water, and Energy for a Greener Future," and each team member will focus on either air, water, or energy to find a dataset to work on. My chosen subtopic is energy, and the next challenge will be finding a dataset within this domain.

Progress Made today: Topic selection finalized

Dec 11, 2024 – Team Meeting

Today, we spent two hours in a team meeting. We finalized the workflow, set the timeline for each milestone, and defined each member's responsibilities for the project. The following is the meeting notes:

**Goal**: 70 / 100.

**Timeline:**

- Coding & GitHub link: Due before Dec 21, 2024
- Report: Due before Dec 28, 2024
- Video Presentation: Due before Dec 29, 2024
- Final Submission: Dec 30, 2024

**Coding & GitHub Section** (step-by-step):

1. Find an appropriate dataset
2. Find 1~2 related research papers
3. Coding & data analysis
4. Post on GitHub (each team member works on a separate branch)

**Individual Report Components** (before final report development):

1. Dataset explanation
2. Research question (Problem to solve)
3. Solution approach
4. Technical details & implementation
5. Evaluate performance
6. Evaluate impact
7. Future improvement

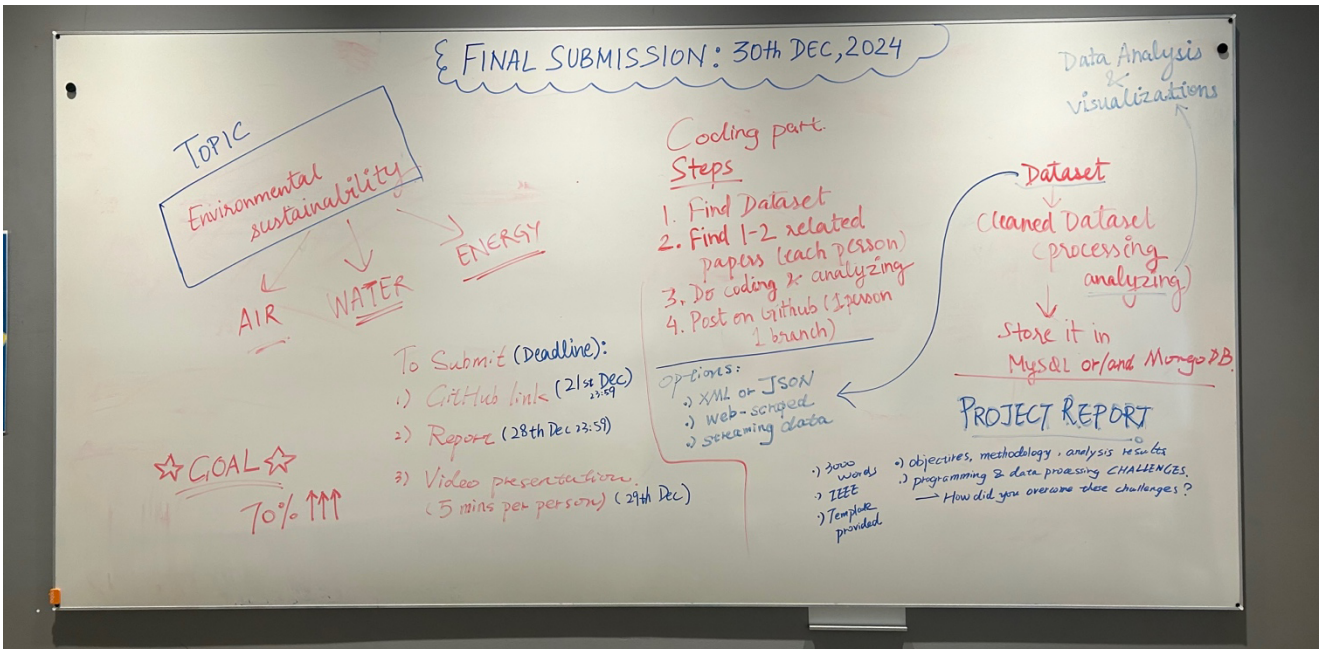
**Report workload Division:**

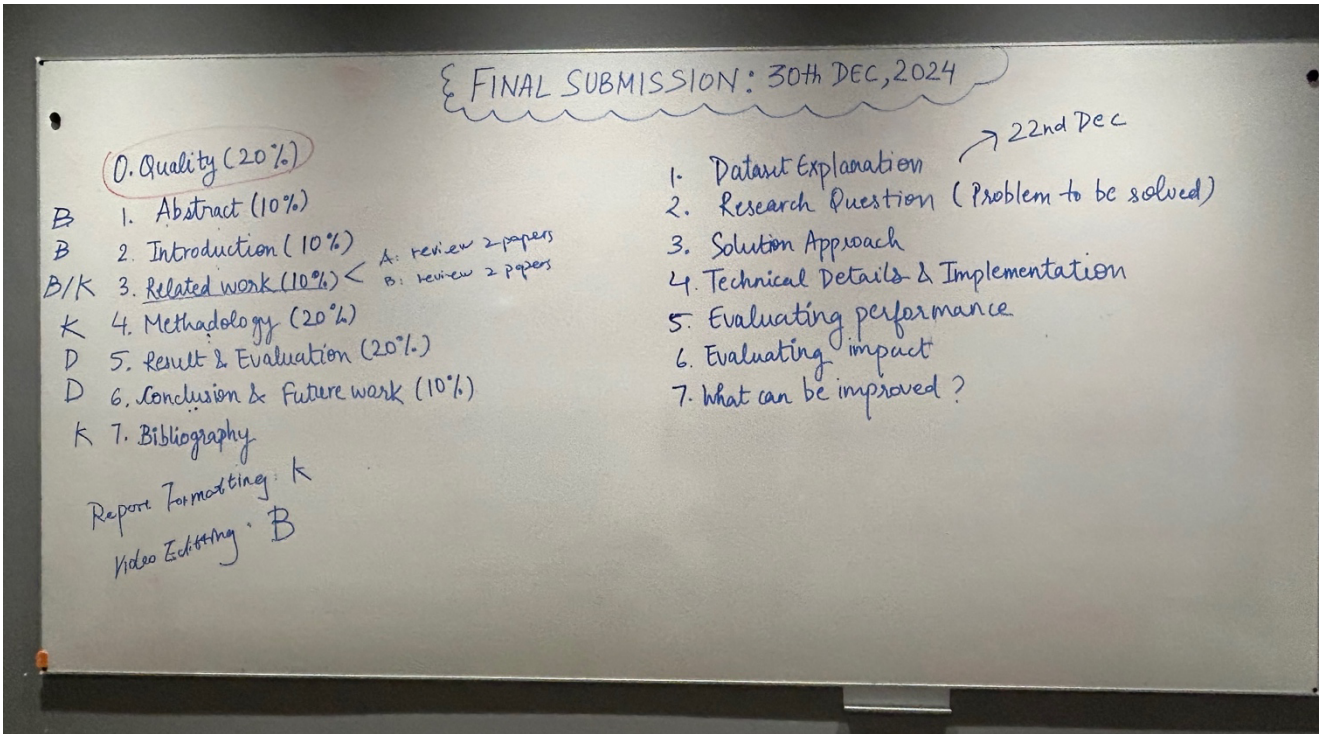
1. Abstract – Bintong
2. Introduction – Bintong
3. Related Work – Bintong & Karandeep
4. Methodology – Karandeep
5. Results & Evaluation – Dhruva
6. Conclusion & Future Work – Dhruva

7. Bibliography – Karandeep

**Additionally:**

- Report formatting: Karandeep
- Presentation video editing: Bintong




Progress Made today: Team meeting; project plan discussed; workload divided.

Dec 12, 2024

Today, I began working on my section of the project and created a more detailed workflow for the coding process. The planned workflow is outlined below:

1. find a semi-structured dataset or use web-scraped/streaming data
2. store the data in MySql (my chosen database management system)
3. read data from MySql and store it as csv
4. analyze the data and generate visualizations

Today's focus was to find an appropriate dataset for this project. Since my section of the project is focused on energy, I researched datasets related to electricity, gas, and solar energy. Eventually, I came across a 'Daily Gas Supply' dataset on Data.Gov.ie, which I believe will be a good choice for this project.

Progress Made today: Coding workflow outlined; research completed; dataset identified.

Dec 13, 2024

My first attempt was to use BeautifulSoup to web scrape the data to retrieve the data from the source website. But even after debugging and checking the code, I still couldn't get the data I needed which made me very frustrated. Later, I realized that the issue was that the website's content was loaded dynamically with JavaScript, so BeautifulSoup was not able to retrieve the data when the page initially loaded. To work around this, I decided to extract the data in JSON format instead. Then I used the code with the BeautifulSoup approach to grab the dataset's metadata.

The next step tomorrow will be to create a database and a table in MySQL Workbench and write the code to store the JSON data in MySQL.

Progress Made today: Dataset retrieval completed.

Dec 14, 2024

Today, I worked on coding two functions: one to create the MySQL connection and another to insert JSON data into MySQL. The first function was straightforward: after creating the connection, I only needed to validate it and handle exceptions. However, the second function took more time as there were many details to consider, such as checking if the data already exists in the database and handling the date format.

I also edited and improved the code I had been working on because I decided to adopt a modular approach. I separated each step into individual functions and called these functions one by one to execute the tasks required in the project brief. I believe this approach makes the code easier to understand and debug later, and I can also reuse these functions for other tasks in the future.

Progress Made today: MySQL connection established; JSON data inserted into MySQL.

Dec 15, 2024

Today, I worked on a function to export data from a MySQL database table into a CSV file. The purpose of this step is to make sure that the data can be turned into a dataframe by using Pandas for further processing, analysis, and visualization. I specifically ensured that the function would overwrite the existing CSV file each time it is called, instead of unintentionally appending data. This was achieved by using the 'write' mode while writing to the CSV file. Additionally, I added a safeguard to check if the data retrieved from the database is empty before proceeding. This helped to avoid creating an empty file.

Progress Made today: function to export data from a MySQL database table into a CSV file.

Dec 16, 2024

Today, I focused on the data pre-processing, transformation, analysis, and visualization sections. Since I've worked on a comprehensive data analytics project before for another module, I was already familiar with the tasks involved in data analytics. However, I did spend some time exploring the dataset to uncover meaningful insights and patterns. I also experimented with various visualization techniques to find the most effective way to present these findings.

Progress Made today: data pre-processing, transformation, analysis, and visualization done.

Dec 17, 2024

Today I worked on the function to store the processed clean data into a MySQL database table. I spent some time debugging the SQL query in my code and eventually realized the issue was caused by not accounting for the new columns added during data processing when creating the table in MySQL Workbench. Another challenge I encountered was converting the date-related columns into the correct format.

Additionally, I improved the function to connect to the MySQL database. Instead of hardcoding the required values for creating the connection, such as the user and password, I created environment variables to securely store this sensitive information. This approach allows other people to replicate the workflow by simply updating the values to match their own MySQL connection settings.

Dec 18, 2024

As I still have some spare time before the agreed-upon deadline with my teammate, I spent today working on predictive modeling with machine learning. While this is not a requirement in the project brief, I wanted to practice what I've learned in other modules. I've previously worked with linear regression and random forest models, but I also explored the ARIMA model for this project, which was new to me. I came across this model during my research and found it particularly suitable for my dataset, as it is designed for time-series forecasting.

Dec 19, 2024

Today, I went through my code a few more times, made some edits, and added a few inline comments to clarify certain sections. I also created and worked on the README.md file. While working on it, I organized the workflow of the code to make it easier to follow, and included step-by-step instructions for replication in the README.md file. Additionally, I created a requirements.txt file so the necessary Python libraries can be installed with just one command.

I believe I've completed the coding portion and am now ready to move on to the report!

Dec 21, 2024

Today is the day our team agreed to complete the coding part of this project. To share it on GitHub, our team decided to create an organization with the three of us as members and host the project under the organization. With this approach, the project doesn't need to rely on any personal account. I reviewed my previous study notes about collaboration on GitHub and then uploaded my code accordingly without encountering any conflicts.

The workflow is shown below:

1. clone the repository
2. create a new branch named 'energy-section'
3. add my file to the repository folder
4. check changes using the command 'git status'
5. stage change using the command 'git add --all'
6. commit changes
7. push the branch
8. Click 'New Pull Request' on the GitHub page
9. Merge the pull request.

Dec 26, 2024

In the past few days, I have been working on the report for this project. Our team decided on the following approach: each team member will create a full report, pretending it is an individual project focusing only on the dataset they have been working on. Later, as we previously discussed, everyone will be responsible for a few sections and will rewrite these sections based on the three individual reports created by the team. We believe this approach is ideal as it gives us a chance to fully understand and become familiar with each other's work.

My responsible sections are Abstract, Introduction, and Related Work. For the Related Work section, I will review academic papers related to water and energy topics, while another team member will review papers on the air topic.

I found writing the report not too difficult, as our team had many discussions previously, and everyone has an individual report for their part with at least 1,000 words. The writing process is very straightforward.

Dec 31, 2024

I just finished recording my part for the video presentation yesterday. I decided to present my code and the data visualizations and verbally explain what I have done, some decision-making processes, and the findings of this project. I chose this approach because the flow of my code naturally aligns with the flow of the project tasks and decision-making.

I am also responsible for editing the final video, which I will do once I receive all the footage from my team members.