JOURNAL

Karandeep Singh Mann

Team KDB

x23295945

MSCAI

Programming for AI

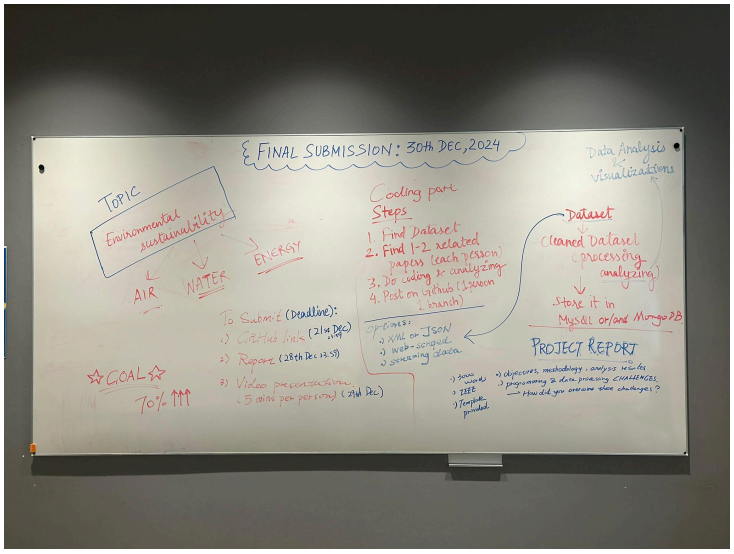
The following journal discusses my contribution for the work that we have done for successful completion of our Programming for AI project. The approximate time spent is mentioned for each section along with the challenges I faced during my project.

10th Dec, 2024: Finalizing a topic for the project

- First Group meeting held to discuss and finalize the topic of the project.
- Taking from the "Data Governance and Ethics" module we have studied sustainability for AI, hence we chose sustainability as our main topic.
- Under Sustainability, we finalized the topic: "Data Driven Sustainability: Tracking Air, Water and Energy for a Greener Future" where each person will handle one aspect of sustainability.
- I chose to do the "Air" part
- Time Spent: 1-1.5 hrs

11th Dec, 2024: Finalizing Work Distribution and Deadlines

- Second Group meeting happened to discuss in detail about the tasks to be done for our project.



- Deadline was finalized for final submission of our project (30th December, 2024)
- Deadlines were decided for individual components of the project as well including coding part (21st December, 2024), report writing (28th December, 2024) and video recording (29th December, 2024)

UPDATE: Due to some unforeseen circumstances, the work got delayed and the deadline for submission got shifted to 3rd January, 2025.

- Before making the final report, it was decided that everyone will be writing some portion of the report, i.e. Abstract, Introduction, Related Work etc for their part (for me it was Air quality) so that final report making will be easier to do. Deadline for this work was 21st December, 2024.
- For the final report work,, I chose the following sections to work on:
    a. Related Work (50%)
    b. Methodology
    c. Bibliography
    d. Final Report Collation, Editing and Formatting
- Time Spent: 2 hrs
- Challenges: Conflict of interest to choose which part of the report will be done and deciding on deadlines because of other module's work involved.

13th Dec, 2024: Paper Research for finalizing approach to the Air Part

- Explored various papers on google scholar to formulate the approach that I can use within my project and select a relevant dataset on the basis of the work to be done.
- Found one paper by M.Rahman et. Al. that focussed on key pollutants and understanding the impact of them which formed the base of my study for this project.
- Time Spent: 1-2 hrs
- Challenges: Struggled to find the relevant paper because some papers which I felt would be rightful by reading the abstract, I couldn't access them through institutional login.

14th Dec, 2024: Relevant Dataset search

- My research to find a dataset started. Various Sites were explored such as UCI Machine Learning Repository, Kaggle, OpenWeatherMAP, OpenAQ, EPA etc to find the right dataset.
- Suitable dataset is found on Open Meteo API which contains information about key pollutants in the air like PM2.5, Carbon Monoxide, Ozone etc, for one city, New Delhi in India. This dataset felt perfect for my air quality part of the project.

- Time Spent: 2-3 hrs
- Challenges: Access restriction because of API keys, data not available in JSON/XML format, incomplete information and non relevant data and some APIs not returning correct results

16th Dec, 2024 - 21st Dec 2024: Coding for the dataset

- This time period involved the coding part.
- Initially I did all the necessary data processing and transformation on the dataset. After transformation, the dataset was saved within MySQL.
- Next step was performing visualizations which gave me various analytic results and insights about the air quality trends and patterns in New Delhi.
- To make it more impactful, I developed machine learning models to predict PM2.5 levels in the air alongside a Flask & Streamlit based web application to show real-time monitoring of pollutants.
- Uploaded my code for my air section on github.
- Time Spent: 5-6 days (a few hours each day)
- Challenges:
  i) Choosing the correct plots to find relevant insights within the data
  ii) Carefully analyze if we could remove the outliers from the dataset.
  iii) Handling various SQL syntax errors such as: "1054 (42S22): Unknown column 'NaN' in 'field list'" etc. while saving data within MySQL database.
  iv) The graphs in the Streamlit application were changing rapidly hence a slider has to be implemented to adjust the timings.
  v) "SQL_PASSWORD" environment variable wasn't working properly because my SQL password has a special character which needed to be encoded.

23rd Dec, 2024 - 25th Dec 2024: Doing my part for final version of report

- Combined each teammate's contribution to their component for Methodology to create a final full fledged Methodology section containing all three components: Air, Water and Energy alongside writing 50% of the related work section.
- I created flow charts on draw.io describing the workflow and process within each component (air, water, energy).
- Time Spent: 3 days (a few hours each day)

- Challenges: To fully understand the work done by teammates so as to create a correct flow chart diagram and explain the steps performed by each teammate.

27th Dec, 2024: Preparing Presentation Slides for video demonstration

- For better understanding, I created PPT slides for the "air" section of the project, using which I created my part of the video for 5 minutes to be done.
- I created a 10 pages PPT slideshow which briefly explains all the work that I have done for the air section.
- Time Spent: 2 hrs

30th Dec, 2024: Video Shoot and Report Compilation

- I shot the video for the Air part and sent it to the other team member for compilation.
- I compiled the report section into one final report document and sent it for approval by other teammates which got approved within one go.
- Time Spent: 30 mins for video shoot, 1-1.5 hours for report compilation on the given template in project description