I tried to participate in this competition starting in April. Below are the brief steps I took that led to my first submission:

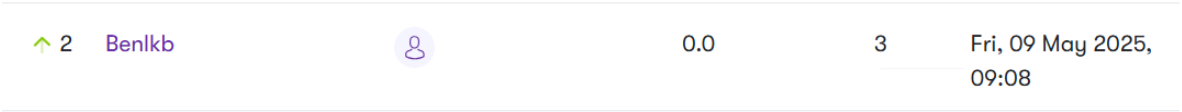
1. Look at youtube tutorials about how to set up my environment
    a. installed data libraries such as pandas, numpy
    b. also imported matplotlib, seaborn
    c. and Ml toolkit such as scikit-learn, light gbm
2. gathered all raw weather readings
    a. pulled together daily measurement values of each region such as temperature, wind speed, pressure.
3. However, sometimes sensores go quiet(in the competitions case, the values are missing) so I filled these missing values with the most reasonable guess, which is the respected data's mean(computed with pandas).
4. Since raw data are not enough to observe a clear weather trend, I created extra clues from the basic data
    a. Ex: Made raw wind speed values more meaningful by converting them into the familiar Beaufort scale("light breeze," "strong breeze," etc.)
    b. Helped algorithm recognize patterns more easily
5. After rearranging the data into more specific labels, I turned the multi-site readings into one unified snapshot per day
    a. I basically rearranged the data so that each row represented a single date, and each column held that day's reading from one town(all readings such as wind speed, temperature, pressure…etc)
6. Then, I finally started training using a powerful algorithm called a gradient boosted tree model) which showed my past days' snapshots, along with known weather labels for each day.
    a. over time it learned which combinations of numbers tend to lead to which labels.
    b. chose this algorithm initially since it's the most recommended algorithm for starters(according to the internet at least)
7. To make sure it wasn't just memorizing, I held back a portion of my known days as a mini test, and asked the model to predict those labels– and measured its accuracy, seeing if it succeeded and where it stumbled.
8. Finally, when the accuracy was about 0.78, I fed in the unlabeled snapshots from the test period.
9. Although the trained model succeeded in having a near 80% accuracy, my submission didn't receive any points as my portal's file requirement is weird, and my form wasn't correct.
    a. I uploaded three more times across a week and it still didn't work.

Reflection:

Through this initiative, I now have some basic understanding about how to create a model and how to train one. Although through the process I had to heavily rely on youtube videos and chat gpt, I still managed to develop one model. The process of developing one was really hard, especially when I got stuck in a threshold(I was stuck at a 0.68 accuracy one day). I tried other algorithms such as light gbm, or re-arranged data labels, which finally resulted in slight improvements.

Honestly, half the time during this initiative I am actually confused and only have a slight idea on what I am doing, but the process of trying out different algorithms, training the model, and rearranging dates based on my own observation and educated guess is nevertheless interesting. It felt like solving a puzzle without a definite answer.

However, I still didn't figure out how to properly submit the files and then didn't proceed to try to submit new submissions as I was preparing for Ap tests and then sat til the end of the semester, therefore, I plan to train another model this summer.



picture of my failed submissions.