MLDS Datahack 2025

Problem Overview

This year's competition: Now that the NOAA has been defunded, an insurance company, InsuraCorp, wants consultation on the use of Machine Learning Models to predict outputs of larger climate models. They have provided a dataset from a custom climate/damage model to test groups on their ability to both forecast weather and the damage from weather. Teams are provided a dataset of pressure, temperature, wind speed and direction, along with civilian damage over that time. Students are tasked with:

1) **Producing 5 day forecasts of wind** speed for the simulated city **GANopolis** for **10 different events**.

2) **Pricing insurance** - Setting a price accounting for how much revenue will occur based on price, and how much insurance will have to be paid out in damages.

Teams will be provided demand curves - so they need to translate their wind speed/damage forecasts into profit estimation.

Teams will also perform a presentation demonstrating their modeling approach and recommendations to InsuraCorp on the use of ML/AI for modeling weather.

Teams will be judged on: Creativity, Technical Mastery, Presentation Quality, and scored metrics: profit and Mean Squared Error (MSE) of wind predictions.

Available Data

You are provided with:

data/

training_data.csv - historical training data from the sim

event_1.csv

event_2.csv

…

event_10.csv

Each of these events files holds the data for the first 5 days of a future event. Your job is to forecast the wind speed for GANopolis for the next 5 days, and assign an insurance price. You will *ONLY* be forecasting GANopolis even though each event contains information for all cities.

Data Features

Column information for both events and training data

hour - hour in sim - starts from 1 and counts up

day - day in sim

hour_of_day - hour but resets every 24 hours

city - city in sim

pressure - air pressure

air_temp - air temperature

ground_temp - ground temperature

velocity_x - wind velocity in x direction

velocity_y - wind velocity in y direction

wind speed - magnitude of wind velocity

damage - amount of damage per customer at that hour

Insurance Pricing/ Profit

The functions for # of customers, revenue, costs, and profit are below. For this competition you will be choosing a price to maximize profit. Damage is summed over the 5 days being forecasted.

$$\text{demand} = (10000 - (\frac{10000}{500})\text{price})$$

$$\text{unit\_contribution\_margin} = (\text{price} - \text{damage})$$

$$\text{profit} = (\text{demand})(\text{unit\_contribution\_margin})$$

We've provided a handy interactive to see how these equations interact.

https://www.desmos.com/calculator/31ibga1bix

Submission

Please update, commit, and push the following files to your GitHub Classroom repo:

Due at 3:30 pm:

> **submission/team_info.json**: Team name, number, and members
>
> **submission/submission.csv**: Prices and wind speed predictions for each event in GANopolis. You are predicting wind speeds for the second 120 hours of each event. See the github
>
> Note: you are only submitting for GANopolis, all other locations are solely features.

Due at 4:00 pm:

> **submission/slides.pptx:** Presentation slides
>
> Code: Please submit what you used to analyze the data and generate your submission/submission.csv file. This should be well commented/readable.
>
> Note: your repo will lock at 4:00 pm, so please push your final slides/code before then!

Cities

City locations are as provided:

```
CITIES = [
    dict(name="Sparseville", x=63, y=35),
    dict(name="Tensorburg", x=214, y=378),
    dict(name="Bayes Bay", x=160, y=262),
    dict(name="ReLU Ridge", x=413, y=23),
    dict(name="GANopolis", x=318, y=132),
    dict(name="Gradient Grove", x=468, y=158),
    dict(name="Offshore A", x=502, y=356),
    dict(name="Offshore B", x=660, y=184),
]
```

InsuraCorp Weather Simulation

The weather simulation is a lattice-boltzmann thermo fluid simulation. Fluid simulation is nasty work, but if you're curious, we have provided a description of how the simulation works in the google drive, titled: weather_sim_description.pdf.

We highly recommend teams don't attempt to match the provided data using another fluid solver as that is a highly daunting task. If any teams successfully accomplish this, please contact one of the officers. We will get your result published in a top tier journal publication.