

Hackathon Topic: Forecasting Daily Emissions with Environmental Factors

Problem Statement:

Participants are invited to utilize historical Continuous Emissions Monitoring System (CEMS) data in conjunction with El Paso weather data to develop a predictive model for emissions levels at the Montana Units from January 1, 2024, to October 31, 2024. This exercise aims to elucidate the correlation between environmental conditions and emissions output, thereby enhancing participants' understanding of how weather impacts daily emissions and honing their forecasting abilities.

The hackathon is structured into three tiers based on the participants' experience levels:

- **Beginner:** Teams will forecast emissions/load for a single day during the Peak Season (May through August).
- **Intermediate:** Teams will predict emissions for three days: one day during the Peak Season (May through August) and two days during the Off-Peak Season (January through April and September through December).
- **Advanced:** Teams will forecast emissions for three days for each unit: one day during the Peak Season (May through August) and two days during the Off-Peak Season (January through April and September through December). Additionally, teams will analyze the data to determine which unit operates more efficiently.

Challenge Details:

1. Data Provided:

- **Time Period:** CEMS and weather data covering October 28, 2021, to December 31, 2023.
- **Weather Variables:** Daily weather metrics including average, minimum, and maximum temperatures (tavg, tmin, tmax), precipitation (prcp), snowfall (snow), wind direction (wdir), wind speed (wspd), and pressure (pres).
- **CEMS Emissions and Operational Metrics:** Parameters include UNITONBT, SO2TONS, HEATINBA, NH3TONS, UNITONBA, HEAT_QA, HEATINBT, GFLOW_BA, NOXTONS, LOADMWBA, COTONS, and LOADMWBT.

2. Objective:

- **Build a Predictive Model:** Create a model that forecasts daily emissions levels, specifically for key parameters like SO2TONS, NOXTONS, and COTONS, using both historical emissions data and weather data.
- **Identify Influencing Factors:** Analyze which weather variables (e.g., temperature or wind speed) have the strongest correlations with emissions, providing insights into environmental impact on operational output.

- **Explain Predictions:** Provide explanations or a feature importance ranking showing which variables most influence emissions levels, highlighting the role of external weather conditions in emissions forecasts.

3. Expected Deliverables:

- **Predictive Model Output:** Predictions for emissions from January 1, 2024, to October 31, 2024, along with an evaluation of the model's accuracy.
- **Feature Importance Analysis:** Report on which weather and operational parameters most influence emissions forecasts.
- **Visualization:** Graphs or charts that demonstrate trends, predictions, and feature importance, including how weather patterns relate to emissions.

4. Presentation and Data Storytelling:

- **Visualizations:** Create clear, insightful visualizations to illustrate trends, predictions, and key findings. Students should aim to show the relationship between weather conditions and emissions levels and highlight the most significant factors influencing emissions predictions.
- **Storytelling:** Develop a narrative that explains their approach, findings, and recommendations. This should include an overview of how they selected features, built their model, and the insights they uncovered about emissions patterns.
- **Real-World Application:** Explain the potential impact of their findings in a real-world context. For example, they could discuss how their model could help operational teams anticipate emissions spikes and make proactive adjustments.
- **Clarity and Engagement:** Present in a clear and engaging way, practicing professional presentation skills. This includes using visuals effectively, simplifying technical details, and focusing on the “why” behind their insights to make the analysis accessible to a broader audience.

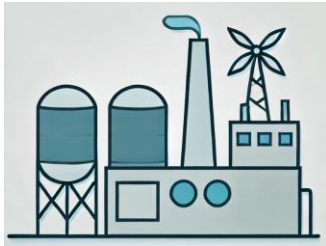
5. Learning Objectives:

- **Predictive Modeling with External Factors:** Apply data science techniques to forecast emissions, integrating multi-variable datasets.
- **Correlative and Feature Importance Analysis:** Understand how external conditions, such as weather, influence emissions and gain experience in identifying key predictors.
- **Environmental and Operational Insight:** Develop awareness of how emissions are impacted by both operational and environmental factors, fostering skills in data-driven environmental decision-making.
- **Effective Communication:** Practice explaining complex analyses and models in a way that non-technical stakeholders can understand and find value in.

- **Data Storytelling Techniques:** Learn how to structure a narrative that turns data into insights, making it relevant and actionable.
- **Visualization Proficiency:** Enhance skills in creating visualizations that not only represent data accurately but also support a compelling story.

This presentation component reinforces the importance of communication in analytics, giving students a well-rounded experience that mirrors real-world expectations.

Lake Power Station



Capacity: 353 Net Megawatts (Peak Period) (88.5 per Unit)

Location: El Paso, Texas

Lake consists of four advanced simple cycle aero-derivative combustion turbines fueled by natural gas, the cleanest fossil fuel available (LAKE-1, LAKE-2, LAKE-3, & LAKE-4). The units feature fuel-efficient turbine generation technology and are designed with best available environmental control technology.