American Electric Power: Safety Observation Challenge

Imagine This

It's the night before a huge exam. You're in your dorm room, books and notes spread out, deep in study mode. Suddenly, a storm rolls through, and the power goes out. The room goes dark, and your Wi-Fi cuts off. As you scramble to find a flashlight, somewhere out there, unseen and unknown heroes are heading into the storm. These are AEP's field and line workers, braving dangerous conditions to restore power and ensure your safety and comfort.

About American Electric Power (AEP)

AEP is one of the largest electric utilities in the United States, delivering reliable electricity and innovative energy solutions to millions of customers. With a commitment to sustainability and safety, AEP is dedicated to ensuring the well-being of our employees, customers, and communities.

The Challenge

Safety is a major strategic focus for AEP, particularly for our field and line workers who operate in environments that can pose significant risks. We collect vast amounts of safety observations from our workers in the field, with comments ranging from minor concerns to critical hazards that could result in serious injury.

Your Mission

We are excited to invite you to participate in our challenge! Your task is to leverage the power of Large Language Models (LLMs) to identify the highest value comments within our extensive list of safety observations. High value comments are those related to hazards with the potential to cause serious injury or damage to property or the environment. By isolating these critical observations, you will help us take proactive measures to keep our essential workers safe.

Why Participate?

- **Real-World Impact**: The Safety & Health and Data Science teams at AEP are currently working on this very problem. Your efforts could directly contribute to enhancing the safety of AEP's field and line workers.
- **Innovative Problem-Solving**: Use cutting-edge technology to tackle a real-world problem.
- **Recognition and Prizes**: Gain recognition for your innovative solutions and compete for the following:
    - 1st Place - (4) $250 Amazon Gift Cards
    - 2nd Place - (4) $100 Amazon Gift Cards
    - 3rd Place - (4) $50 Amazon Gift Cards

Dataset Details

You will be provided with a CSV file containing approximately 20K records. Each record represents a safety observation and includes the following fields:

- Date: The date when the observation was recorded.
- Observation Type: The safety category or topic addressed in the observation.
- Comments: Freeform text comment provided by the observing supervisor. This is the primary target for your analysis.

The safety observations are made during "CORE visits", a common practice in the utility industry where crew supervisors monitor specific jobs or tasks and note any safe or risky conditions they observe. These comments vary in length and detail, ranging from brief notes to lengthy descriptions. Your challenge is to use open-source LLMs to analyze these comments and identify high-value observations related to serious hazards.

Important: Data privacy and security are paramount. Under no circumstances should the comments be input into public GenAI services like ChatGPT or Gemini, nor should you leverage APIs for such services where data privacy and security cannot be guaranteed. You are encouraged to use LLMs that can be run locally or in a secure, private environment. Do not share the data with anyone outside the HackOHI/O event and delete the data from your environment when the event is over.

We will also provide background information on the types of comments that could be considered high value to aid in your analysis.

Expected Outcomes

Your results can be shared in a format of your choosing, which could include:

- **Code and Output Data**: Provide the code you used for analysis along with the output data highlighting the high-value comments.
- **Dashboard with Visualizations**: Create a dashboard that visualizes your findings, offering insights into the identified high-value observations.

We encourage you to experiment with various prompt engineering and entity extraction methods to enhance your analysis.

Judging Criteria

- **Accuracy**:
  - How accurately does the solution identify high-value comments related to serious hazards?
  - Are the identified comments relevant and correctly prioritized?
- **Innovation**:
  - How creative and innovative is the approach used to analyze the comments?
  - Are there any unique methods or techniques employed that stand out?
- **Implementation**:
  - How well is the solution implemented?
  - Is the code clean, well-documented, and easy to follow?
  - Does the solution run efficiently on the provided dataset?
- **Usability**:

- o How user-friendly is the final product?
- o If a dashboard or visualization is provided, is it intuitive and easy to understand?
- **Impact**:
  - o How impactful is the solution in terms of enhancing the safety of AEP's field and line workers?
  - o Does the solution offer actionable insights that can be practically applied?
- **Adherence to Guidelines**:
  - o Does the solution adhere to the data privacy and security guidelines provided?
  - o Were open-source LLMs used appropriately without resorting to public GenAI services?

We look forward to seeing your innovative solutions and how they can make a difference in the safety of our workforce. Let's work together to create a safer environment for everyone.

Good luck, and we can't wait to see what you come up with!