Categorizing NIH Utility Plant Data

Nashita Bhuiyan

College Park Scholars – Science & Global Change Program

Computer Science

nbhuiyan@terpmail.umd.edu

CPSG230

College Park Scholars Academic Showcase, May 5, 2023

Service Site

- **Site Name:** National Institutes of Health, Office of Research Facilities (ORF)
- **Address:** 9000 Rockville Pike, Bethesda, MD 20894
- **Contact Email:** orfresponse@mail.nih.gov
- **URL:** https://orf.od.nih.gov/Pages/default.aspx
- Site Supervisors: Dr. Chenggang Yu
- **Mission:** The ORF maintains and supports NIH's campus facilities. More specifically, the Utilities Engineering Branch (UEB), the division I worked in, aims at optimizing NIH's Central Utility Plant (CUP) operations by collecting data and performing analysis.

The front of the Utilities Engineering Branch building, where my office was located.

Issues Confronting Site

There are around 350 million data points collected daily by the UEB. A major challenge confronting the site is parsing through and making use of all this information. One data source which is particularly difficult to work with is the shift turnover logs. Shift turnover logs are how CUP operators update one another on the status of CUP equipment. The information stored in these logs would be valuable to other departments if search and analysis procedures could be performed on them. However, the current formatting of the logs make these tasks impossible. To resolve this issue, I was tasked with developing a program to label log incidents by the equipment it discusses.

Activities

My main task during this internship was to develop a program to extract categorizable information from shift turnover logs. Towards this goal I performed the following tasks:

- Developed a preliminary program that extracts information based on predefined patterns.
- Created a presentation explaining how the program works and shared it at the annual NIH interns expo.
- Developed the framework for a program that utilizes Natural Language Processing to categorize shift turnover logs.

The interior of a Cogen, a piece of equipment within the central utility plant. Equipment like Cogen are reported on in the shift turnover logs and identified by the program. (https://rb.gy/sa0na)

Impact

The program I developed during my internship targeted the issue of extracting relevant information from shift turnover logs. This program will be used to develop tools to search and analyze shift turnover logs. Through this internship experience I gained basic data science skills, experience in collaborating with people from various disciplines, and perspective on how data can be utilized to optimize performance of real-world tasks.

One of the visualization tools currently used by the site to report on the status of CUP equipment.

Future Work

A major goal of the site is to develop a comprehensive search system for shift turnover logs which can be used to identify failing equipment and causes of failure. The program which I developed resolves the major issue of categorizing logs which is essential to the development of such a system. Personally, this internship helped me develop invaluable skills such as the ability to clearly communicate technical information to those outside of my field of expertise. I will apply the skills in my future jobs and projects.