

## Siemens Energy



Siemens Energy is the energy subsidiary of Siemens AG, the largest industrial manufacturing company in Europe. Siemens Energy is known for its renewable energy ventures, specifically regarding power, energy, transmission, and wind. Siemens AG employs approximately 293,000 people worldwide and grossed \$69 billion in revenue in 2021 alone.

### Project Description

This project will have three parts. Siemens Energy is mostly looking to **develop analytics that predict and identify power outages based on internet history and weather data**. An HDAG team will attempt to tackle all three goals over the course of the semester, but the client understands that potentially only the first two goals may be attainable in the given time frame:

3. **Create a database of the Google (or other web engine) searches that could be emblematic of a power outage.** Compare this to a known database of outages over the past couple of years.
4. **Create a NLP/ML model that determines the most important keywords to predict an outage.** This model should be able to determine the most relevant keywords by comparing the two databases that were created above.
5. **Correlate weather data to outage history.** Find a suitable weather dataset that is interpretable and powerful enough to predict outage history. Create a ML model that bridges the connection between these two data sources.

**Internal Partners:** Siemens Sensproduct Manager, previous manager of Digital Industry Services (AI & ML based Services)

**Datasets:** TBD. Most of this case will require you to find datasets (Google trends, search engines history, etc.). Siemens may provide an internal outage history dataset, though the details are still in the works.

**Preferred Coding Languages:** Any

### Specific Skills

1. **Research:** Finding suitable databases (searches and weather data). Hypothesizing on which google search keywords are the most suitable
2. **NLP:** analyzing which words are the most powerful in predicting outages.
3. **Machine Learning:** using weather data to predict outages. Time series analysis, regressions, and geospatial neural networks if necessary.
3. **Data Visualization and Interpretability:** Creating useful and interactive visualizations to interpret the data. Communicating the limits of any model(s) to the client

Expected Technical Difficulty: **Intermediate / Advanced**