BELLEVUE COLLEGE
COMPUTER SCIENCE

CS 481 SENIOR CAPSTONE I (3 Cr.)

Fall 2020

Supervisor:
Industry partner: https://www.watttime.org
Preparation Date: Connor Guest connor@watttime.org

1. Subject/Keyword Classification
Data Science, Algorithms, ETL
2. Title of Project
Collect, Forecast, & Visualize – Global Electricity Generation Data
3. Description of Project
Background: a collaboration between multiple nonprofits, governments, and tech companies is currently building an AI to continuously monitor pollution from every power plant in the world from space, using satellite data. A crucial component of that system is its ability to be trained on a sufficiently large and accurate training dataset. This will be a mix of readily available satellite imagery, combined with accurate ground truth generation data. The purpose of this project is to help build that ground truthing system. Students will implement real-time scrapers for web-based data from power grids around the world and analyze the data for accuracy. A forecasting model will be implemented to predict future generation values and a front-end tool will be created to visualize the collected data.
4. Key Techniques/Technologies/Tools
This project has four components. (1) Write scrapers, likely in Python, to scrape data from power grid websites around the world. (2) Analyze the scraped data for accuracy. (For example, many coal-fired power plants report emissions that are exactly 1,000 times smaller than is possible, which is a clear sign that they are reporting in the wrong units.) (3) Create a forecasting model to predict future generation values (4) Build a tool to visualize the data
5. Project outcome
(1) Python code that will sit within an existing WattTime tool which will automatcially scrape, test, and correct data

(2) Forecasting model
(3) Front-end visualization tool of the data

6. Prerequisites

Python
Familiarity with unit analysis

7. Additional information

Students will be working within a framework provided by WattTime which already includes a robust ETL pipeline, database, and significant support on where to find data and how to interpret it.