Projecting Analytical Views for Energy Transition products to determine Proper classification of Company Industry to correct ‘Other’ Category

7/5/2021

# Overview

## Project Background and Description

|  |  |
| --- | --- |
|  | Energy Transition has been a central topic in the news for the last 6 months and many companies are beginning to see the change in culture and policies and preparing for the position themselves properly. A certain energy data company would like to know how they can improve their data collection process and retroactively correct thousands of data points for companies. These data points classify a company’s industry as “other” rather than their actual industry. The goal of this project will be to use the company views to cluster the company in the proper industry type. This will allow for improved analytics and future targeting of company to solution products. |

## Dataset

|  |  |
| --- | --- |
|  | As the products were recently deployed, the last 3 years of data will be extracted from the user activity database. The data will be solely encompassing the analytical views in respect to their company, industry, and document name. |

## Data Science Approach

|  |  |
| --- | --- |
|  | 1. Clean the data and make sure all values are numerical. A zero is a good data point as these are all measurements of views. 2. Explore the data for trends, counts, distributions, and shares of the views. 3. Determine if any standardization/transformation needs to be done to the data and choose 3 clustering models to model the expected industry. 4. Determine accuracy of models. |

## Deliverables

|  |  |
| --- | --- |
|  | Deliverable will be a Jupyter notebook of the code, a written report, and presentation slide deck. |