Earvin Sahayathasan

500709422

2020-01-27

Ryerson Capstone: Dataset on Tweets for Nuclear Energy

Problem and Theme

1. The problem I wish to tackle is to predict what causes a person to agree or disagree with nuclear energy using specific words in tweets as indicators to show how an individual may sway to make a predictive analysis of how the general public view nuclear energy in relation to what they believe in.

Research Questions

1. What are the most common words to provoke positive opinion? Negative opinion?

What problems are shown to most impact negative opinion?

Is there any tweets that do not pertain to nuclear energy such as outliers that will not help with the general consensus?

Dataset

1. The data source I am using comes from the general website that was provided from the links provided called <https://www.figure-eight.com/data-for-everyone/>:

The title of the dataset is called: Judge emotions about nuclear energy from Twitter

https://d1p17r2m4rzlbo.cloudfront.net/wp-content/uploads/2016/03/1377191648\_sentiment\_nuclear\_power.csv

Techniques and Tools

1. The tools that are expected to be used is R data program and Weka for both predictive and classifying analysis on the dataset.

The techniques that will be practiced are the Text Classification and Sentiment Analysis because the text must be analysed to determine the reason behind the results and since the results are textual simple Classification and Regression cannot be used. Anomaly Detection will be used to remove any statements that are irrelevant and may sway predictive results.