Udacity Machine Learning Nanodegree Capstone Project

Using news websites to generate a geographically based behavioral analysis

Definition

This project is designed to generate a geographically coordinated behavioral analysis using stories from local news websites all over the United States. Using a web scraper such as the python library Scrapy, I intend to manually select a sample of websites from each state or region that represent their local news outlets. From this set of data, using supervised learning techniques similar to those used in spam email filters, I intend to create an algorithm which will detect general social attitudes, specifically in the realm of politics, for each region.