Natural Language Processing of Job Loss Related Tweets During the Pandemic

Chen, Benjamin (School: Paul Laurence Dunbar High School)

Due to the pandemic, this year has seen high levels of unemployment and job loss. This project seeks to use natural language processing techniques to analyze first person tweets disclosing job loss. In particular, the project seeks to answer the question: which states have the most complaints about their unemployment systems? Using topic modeling, a subtopic of tweets was discovered that revolved around complaints about the unemployment system. Then, a variety of techniques were used to determine the US state that these tweets originated from. From that information, a map was created to show which states had the most complaints. Data from the Department of Labor was used to account for different things and provide a more accurate picture. California, Florida, New York, Texas, and Ohio were the 4 worst states in terms of raw complaints. After dividing complaints by initial claims, Wyoming, Florida, Ohio, and Nevada were the worst. Looking at the percentage of initial complaints that were unprocessed, Montana, Minnesota, and Georgia were the least efficient. Dividing complaints by unpaid claims showed that California, western states, and the northeast had more complaints proportional to the amount of unpaid people in those states. States like Florida, California, and Ohio should be further looked at to see why their unemployment system caused complaints. On the other hand, states like Georgia should be looked at for their extremely low amount of complaints despite their inefficient system. This could give leaders insights into preventing citizen frustration in difficult situations. A further step will involve hand labeling a random subset of tweets for location to evaluate the accuracy of the programmatic method.

Awards Won: