Ross Lewis, Chunxiao Jing, Yiting Jiang

425

Topic Proposal

Purpose:

We would like to utilize Google Trends data, the Economic Freedom Index, and Alcohol consumption by state per year since 2004 to predict unemployment. Our y value (unemployment) is a numeric value, so this will be a regression problem. Alcoholism is one of the best predictors of unemployment, but we hope to find additional features to make better predictions. Seth Stephens-Davidowitz has research important Google Trends indicators that we can use to enrich our data (http://sethsd.com/research/)

Machine Learning Methods:

We would like to perform descriptive and predictive methods. A neural network will be used as our (most likely) best predictive model. However, we can use linear regression and random forest regression to describe the impact each feature has, and which features are most important.

The Data:

Economic Freedom by state can be retrieved from the CATO institute (https://www.cato.org/economic-freedom-world). Google Trends data can be accessed through their API (<https://trends.google.com/trends/?geo=US>). Unemployment data can be found at the Rhode Island Department of Labor and Training (<http://www.dlt.ri.gov/lmi/laus/us/annavg.htm>). Finally, The National Institute on Alcohol Abuse and Alcoholism has data on alcoholism (https://pubs.niaaa.nih.gov).