Brendan Hill’s Long-term Unemployment Capstone proposal – 7/27/2016

Can we consistently predict who is long-term unemployed?

Testing the viability of targeted policy to decrease the recent unemployment trends

* 1. Problem to solve:
     1. Since the Great Depression, long-term unemployment is at historic highs. Current policy has done little to decrease this trend. Could a new policy that direct targets the long-term (LT) unemployed be a viable alternative? My capstone project aims to determine whether a new targeted policy is a viable alternative to current ineffective federal programs. That policy could be viable if it could consistently target the same factors over several years. If not, policy would have to be regularly updated targeting current factors of LT unemployment, which is extremely difficult if not impossible to do.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

Policy makers and any advocate group that influences labor policy (e.g. Labor department researchers, think tanks, and interest groups) would be ideal candidates. A candidate designing a specific policy to reverse the LT unemployment trend would inevitably want to know what type of policy could be effective for several years. If my analysis shows specific factors consistently influence LT unemployment, they can design policies that most effectively target/address these factors. If my analysis shows these factors are inconsistent, instead of trying to design a targeted policy that isn’t viable, they could use their time more efficiently by working on other potentially effective policies.

* 1. Data I’ll use for this analysis.

The data I’ll use is the Current Population Survey (CPS), specifically the March supplement or the Annual Social and Economic Supplement (ASES). The standard CPS collects data to produce frequently sited labor market statistics (e.g. the unemployment rate). In addition to these statistics, this survey produces other important economic indicators including the poverty rate, the percent covered by health insurance, and median family income. The underlying data used to produce these figures cover a significant number of factors that potentially influence long-term unemployment; making it a great source for my analysis.

The survey is produced by the Bureau of Labor Statistics (BLS) and the Census Bureau. On their website, they make the survey’s microdata available to the public: <http://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpsmarch>

* 1. In brief, outline your approach to solving this problem (knowing that this might change later).

Using variables from the ASES, I will build a model that predicts who currently has been unemployed for 27 weeks or longer (the definition of long-term unemployment). The predictive equation used in the model will be built and optimized using machine learning algorithms, possibly a logistic regression algorithm, given the binary nature of the outcome variable.

After building this model, I will test if it can accurately predict who is LT unemployed in previous years, or if certain factors are more influential in one year than the other. To determine if the odds of factors are significantly different over the two time periods, I will most likely use t-test results from an ANOVA.

* 1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

 My deliverables will include a written report with graphics describing the model used to predict LT unemployment and the results of my analysis testing whether or not it’s consistentent across time. I will also deliver the programs used in the analysis in IPython notebooks on my GitHub repository. Using the paper and the graphics produced by my code, I will deliver a powerpoint presentation and use it to give a presentation to the class.