

Project Design Writeup

Craig T. Wood

11.3.16

- **Requirements:**

- Well-articulated problem statement with "specific aim" and hypothesis, based on your lightning talk
 - Problem Statement:
 - Reliable POWER and INTERNET ACCESS are difficult to come by in some low- and middle-income nations in Africa. I aim to prove that investment in one or both of these sectors can produce significant positive ROI for governments.
 - Hypotheses:
 - The following contribute to GDP growth:
 - Greater power access
 - Greater Internet access
 - Greater power reliability (which also increases greater Internet reliability)
 - Greater investment in energy leads to positive GDP returns (*this would prove government action has impact*)
 - Nations that increased capacity saw greater GDP growth than those that didn't
- An outline of any potential methods and models
 - Linear Regression
 - Multiple Linear Regression (power and connectivity predictors)
 - Cross validation - I want to see if I can train my data set well enough to predict results for a new country's efforts, since that's the goal of this study

- Then a grid search to identify my strongest-performing model
- Detailed explanation of extant data available (ie: build a data dictionary or link to pre-built data dictionaries)
 - *[Data dictionary inserted in Google Slides lightning talk deck]*
 - <https://goo.gl/oP6FPA>
- Describe any outstanding questions, assumptions, risks, caveats
 - Each country comes with its own idiosyncrasies and may prevent me from drawing satisfying conclusions across state borders
 - GDP may be affected by so *many* factors that isolating 1-2 predictors may lead to a situation establishing stronger correlation than causation (nonetheless, how strong of a correlation can I establish?)
 - Private monies invested in my target two sectors may influence GDP in ways not shown by the public monies data set I'm using
 - These private investment monies are difficult to track down, or consider reliable
- Demonstrate domain knowledge, including specific features or relevant benchmarks from similar projects
 - A 2016 study done by Khobai, Abel and Leroux published in the International Journal of Energy Economics and Policy took a look at the causal relationship between electric and GDP in South Africa, in both directions. They found a long-term one-way causal relationship, but no causal relationship in the short term in either direction - electricity impacting GDP, or GDP impacting energy consumption.
 - Other studies have been done for individual countries in Africa; multiple countries in Africa; multiple OPEC countries in Africa; high, upper

middle, lower middle and low income countries in Africa; results have differed widely. Some proved a one-way causal relationship between energy and GDP; some found no relationship at all.

- The South African study also utilized power outage and employment variables, finding that power outages did impact GDP.
- Define your goals and criteria, in order to explain what success looks like
 - Goal: prove a positive relationship between investment in energy/internet sector(s) and GDP growth
 - Goal: tell this story in a narrative form palatable to budget-owning government officials
 - Goal: familiarize myself with data set focused on developing world and global economic measures
 - Success shall be defined by the following criteria:
 - Project proves a positive relationship between investment in energy/internet sector(s) and GDP growth
 - Final presentation in a form viable for presentation to sitting government officials

- **Bonus:**

- Consider alternative hypotheses: if your project is a regression problem, is it possible to rewrite it as a classification problem?
 - Good question.
- "Convert" your goal metric from a statistical one (like Mean Squared Error) and tie it to something non-data people can understand, like a cost/benefit analysis, etc.
 - Yes, this is part of the goals I'm setting for myself already.