

Predicting PRC Influence on Other Nations

Caleb Harding (10/10 Interest)

Description:

Previous analyses have look at measuring and predicting the influence of the People's Republic of China (PRC) on other countries. I would like to take a crack at this problem as well and see if some creativity in feature selection or the model used (beyond just regression, which I believe is most common in these sorts of analyses) could achieve high accuracy in predicting PRC influence on other countries. A commonly used metric for "influence" is the UN voting record, and that may be appropriate here.

Data

Each unit of observation in the data would be a country, with the dependent variable being the UN voting record. Depending on desired complexity, this could be a binary value for whether or not they voted with China on a critical issue, or it could be a percentage of the number of votes over a given time period where they vote the same as the PRC. Features could include physical proximity to China, democracy score, journalistic freedom, whether they participate in China's Belt and Road Initiative, whether Huawei is building their 5G network, whether they have an FTA or treaty with the US, total debt, debt to the PRC, total trade and trade deficit with the PRC, GDP, population, GDP per capita, etc. Some of those variables would likely be harder to collect/scrap than others, and ones that are not crucial could be excluded.

	GDP	Population	GDP pc	Debt (% GDP)	BRI?	PRC Imports (2022)	Democracy	UN w/PRC
Canada	2.09E+12	40097761	52722	106.6	No	42289683755	8.88	38.60%

Data Collection

This task would likely require web-scraping. Much of the data is embedded in tables on webpages (Wikipedia has a lot). Using python libraries requests, BeautifulSoup, and pandas, we should be able to scrape, parse, and then combine the datasets (all of which include country name). The following sites could be used for some of the features mentioned:

- GDP pc: [https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)_per_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita)
- GDP: [https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal))
- Population: [https://en.wikipedia.org/wiki/List_of_countries_by_population_\(United_Nations\)](https://en.wikipedia.org/wiki/List_of_countries_by_population_(United_Nations))
- Total debt (as percentage of GDP): https://en.wikipedia.org/wiki/List_of_countries_by_government_debt
- List of countries participating in the BRI: <https://www.silkroadbriefing.com/news/2020/10/02/chinas-belt-and-road-initiative-all-participating-countries-by-income-group/>
- Imports and Exports to China (and from that we can also calculate trade balance): <http://stats.customs.gov.cn/indexEn> This site generates downloadable CSV files, so no scraping is necessary
- Democracy score: https://en.wikipedia.org/wiki/The_Economist_Democracy_Index
- UN Voting Data: <https://dataviz.yiqinfu.com/unview/> (may need to find newer data)