

Milestone 3: Handing on the Baton

Gerald John M. Manipon

Indiana University

Author Note

D590 Data Science in Practice – Spring 2019 – Milestone 3 – Due April 30, 2019

SWOT Analysis and Next Steps	3
---	----------

References	5
-------------------------	----------

SWOT Analysis and Next Steps

The bulk of my contributions this semester was as a member of the Disaster Resilience project (<https://3.basecamp.com/3947469/projects/10567608>) with Logan Paul as the investigator. Because this project only got bootstrapped into the DSIP program this semester (Spring 2019), we decided that an important contribution I could make are a set of Jupyter notebooks that would help onboard new team members to get them familiar with the project and the datasets. As a result, I've create 3 notebooks. The first notebook sets the stage to help new team members get familiar with the basis of the project (Cutter et al. "Disaster Resilience Indicators for Benchmarking Baseline Conditions"). In the second notebook, we start delving into the meat of project by walking the user through recreating a few figures from the paper. In the third notebook, we start exploring the dataset to determine if there are any relationships amongst the indicators and to determine if the presence of one or more indicators have any sort of association with the presence or value of other indicators. The jupyter notebooks I developed are located at the following IU GHE repo: <https://github.iu.edu/INFODER/cutter>.

To close out my work with this project, I've performed the following SWOT analysis to identify key strengths, weaknesses, opportunities, and threats:

		Helpful to achieving the objective	Harmful to achieving the objective
Internal origin (attributes of the system)		<i>Strengths</i>	<i>Weaknesses</i>
		<ul style="list-style-type: none"> theoretical basis established by the Cutter paper case studies available availability of data sources 	<ul style="list-style-type: none"> completeness of data for all indicator variables and counties data sources for disaster costs by county needs exploration and perhaps data wrangling
External origin (attributes of the environment)		<i>Opportunities</i>	<i>Threats</i>
		<ul style="list-style-type: none"> untapped data sources publicly available potential DSIP students to perform analysis exploration of machine learning techniques 	<ul style="list-style-type: none"> unavailability of public data sources for all indicator variables and counties maintaining momentum with subsequent team members

In terms of strengths, the key asset for the project is that it is building off of the study performed by Cutter et al. which provides a blueprint and foundation for proceeding with the project. The study also provides case studies which can serve to help verify the project's

implementation as well as provide examples of different approaches to use during analysis. In terms of weaknesses, the main issue currently is that the set of resilience indicator variables for the institutional subcomponent is incomplete, lacking data for the percent population covered by a recent hazard mitigation plan, flood coverage, previous disaster experience, mitigation and social connectivity, and percent population in storm ready communities. With regards to opportunities for the project, I feel that there are still untapped data sources that may supplement the main dataset, aside from the missing institutional variables, to add additional dimensions to the study space. Finally, in terms of external threats to the project, I feel that we need to keep the momentum going by ensuring DSIP students in the following semesters carry on the work that the team has done so far.

In terms of next steps, we can focus our attention to the weaknesses, opportunities, and threats that we identified in our SWOT analysis above. First of all, the project should continue to flesh out the rest of the resilience index datasets, particularly the missing variables for the institutional subcomponent. As that effort continues, it is also imperative to attain data for all US counties across a consistent temporal range. Some datasets are limited to specific US regions and thus analysis and any potential machine learning applications will be limited regionally. In the same vein, some datasets are only available for certain years and so any analysis would have to account for that temporal variability. There is still a bit of data exploration and data wrangling that needs to be done however that doesn't preclude any efforts to enhance the datasets with supplementary and ancillary datasets. In particular, I feel that disaster datasets such as the IBTrACS storm track dataset (<https://www.ncdc.noaa.gov/ibtracs>) can add an additional dimension to the analysis of the resilience index dataset. The EM-DAT database (<https://www.emdat.be>) which maintains a database of all natural disasters across the globe could also expand the resilience index analysis and machine learning by providing information on disaster types which typically occur based on regions.

That being said, the current set of resilience index datasets are sufficient to perform exploratory analysis and machine learning at least on specific regions and time spans. The 3rd ipython notebook I provided, "Indicator Relationships", attempts to explore and assess if there are any relationships amongst the indicators and provides an initial assessment of the correlation between the subcomponent variables. Future DSIP students should be able to perform additional and alternate analysis to help the project gain further insight into any underlying relationships.

In summary, the Disaster Resilience project for the most part is in the first phase of data science: preparation. There is much work to be done to get the project's datasets wrangled into a form where it can proceed fully with the building phase of the project. However, as stated above, that doesn't preclude the project from exploring the dataset to get a feel for the data. After all, as data scientists it is our duty to stay close to the data. Personally, I feel that the goals of this project and any outcomes that come out of it will be of great societal benefit and I will certainly keep an eye on it as it further develops with future DSIP students' contributions.

References

Cutter, Susan L.; Burton, Christopher G.; and Emrich, Christopher T. (2010) "Disaster Resilience Indicators for Benchmarking Baseline Conditions", Journal of Homeland Security and Emergency Management: Vol. 7: Iss. 1, Article 51, DOI: 10.2202/1547-7355.1732.