Melissa Mock

Springboard

Project Ideas

**What problem do you want to solve?**

According to several reports, such as a UNESCO paper on [Language Vitality and Endangerment](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf), by the end of the 20th century 90% of languages will be replaced by dominant languages in most regions in the world. As globalization increases and modern development spreads, smaller communities are in danger of being absorbed by the dominant culture and losing their unique identity. Language captures and carries a lot of aspects of a culture so as they extinct other elements of a people’s heritage disappear too. Can certain development indicators can predict language endangerment? How can we use this information to take protect languages that are likely to become endangered from development?

**Who is your client and why do they care about this problem?**

UNESCO has invested a lot of time and money in research and programs to prevent endangered languages from becoming extinct. This organization works primarily to research and provide recommendations for NGOs, governments, language communities and language professionals to work together towards preserving endangered languages. UNESCO has limited resources so they would benefit from a model that predicts what languages are more likely to be endangered based on development indicators. They can more strategically target communities and take preventative measures to limit endangerment at an early stage.

**Which data are you going to use for this? How will you acquire this data?**

I will use the [World Bank World Development Indicators](http://data.worldbank.org/data-catalog/world-development-indicators) dataset that details development indicators labeled by type and year. It is available to download off the World Bank website. In addition, I requested the UNESCO’s Languages in Danger data set from a form on their website.

**Outline your approach to solving this problem.**

1. Data collection
2. Data cleaning (merge data sets,
3. Exploratory Data Analysis
4. Machine learning models

**What are your deliverables?**

1. Notebooks for:
   1. Data cleaning
   2. Exploratory data analysis
   3. Machine learning models
2. Capstone report
3. Capstone presentation