Profiling Environmental Nonprofits in the State of Texas

1. Project background

In the next 30 years, Texas will experience more frequent and more extreme weather events. At the same time, Texas is experiencing explosive growth, and the state's population is projected to *double* by 2050. A fundamental challenge to ensuring Texas's future resilience is understanding current impacts, and projecting the effects, of climate change and population growth on the environment, infrastructure, society, and human health. Yet, understanding the threats is only the first step. Connecting and implementing solutions with a broad set of stakeholders working who work on the ground to address these challenges is critical.

One mechanism is working with nonprofits. There are nearly 50,000 nonprofits working in Texas, many of which are relevant for Planet Texas 2050 areas of focus. This challenge asks you to classify nonprofits in Texas into predefined Planet Texas 2050-relevant categories and describe the profile of these nonprofits.

2. Dataset information

2.1. Uncategorized organization list. About 14,000 lines of records include "EIN" and "Name_Mission_Program".

2.2. Human-coded organization list. A thousand lines of records include "EIN", "Name_Mission_Program", "Service area", and "Function". The "Service area" and "Function" are human-coded labels.

Table 1 Description of columns

Columns	Datatype	Description
EIN	String	Employer Identification Number of an organization. The number is unique to each organization.
Name_Mission_Program	String	Combined text including organization name, mission, and program description of an organization.
Service area	Categorical	Six categories describing an organization's service area: "Water", "Urbanization", "Energy", "Ecosystem", "Other-E", and "Other-NE". (E: environment-related but does not belong to the above categories; NE: not environmental nonprofits.)
Function	Categorical	Three categories describing an organization's role in policy: "planning", "policy implementation", and "communication."

2.3. 990 Forms. The machine-readable data, documentation, and usage examples of 990 Forms from 2013 to present can be accessed here: https://registry.opendata.aws/irs990/. You can use this dataset to profile all the environmental nonprofits in Texas.

3. Task Description

- Task 1: Train a classifier using the human-coded organization list.
- Task 2: Classify the uncategorized organizations into different service domains and policy functions using your trained classifier.
- Task 3: A throughout description of all environmental nonprofits in Texas.

4. Expectations and deliverables

a. An organization list including EIN, predicted results of service area and policy function.
b. A machine-learning classifier: Report accuracy, precision, recall, and F1score.
c. A standardized workflow and detailed documentation explaining:
    - Data preprocessing, for example, text processing (tokenizing, spellchecking …).
    - Feature representation and extraction methods (e.g., bag-of-words, word embedding …).
    - Model training and decision making (e.g., dataset resampling, hyper-parameter stochastic searching, and grid searching…)
    - Other important procedures for producing a robust classifier.
d. Illustration (tables, figures, GIS mapping, etc.) of the characteristics of all TX environmental nonprofits.

5. Contact

Please contact Yuntao Lu (yuntaolu@utexas.edu) regarding your team preparation or other questions.