

Classifying Nonprofits Using 990 Forms

1. Project background

There are over millions of nonprofit organizations in the US, and a central task of managing these nonprofits is to appropriately classify these organizations into different categories (e.g., human service, environmental protection, and international aid). Because the number of these organizations is huge, classifying nonprofits is almost an onerous and impossible task for human coders.

All nonprofits above a certain size are required to register with and report to the Internal Revenue Service (IRS) annually. An annual report (i.e., 990 Forms) includes extensive information about a nonprofit organization, for example, mission statement, board of directors, program descriptions, and financial status. This challenge asks you to classify the US nonprofits into different categories (i.e., the National Taxonomy of Exempt Entities; NTEE) using information from the 990 Forms.

2. Dataset information

2.1. 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/>

2.2. The National Taxonomy of Exempt Entities (NTEE) classification

The National Taxonomy of Exempt Entities (NTEE) developed by the National Center for Charitable Statistics (NCCS) is the most widely used classification system. There are 26 *major groups* in NTEE which can be grouped into 10 *broad categories*.¹ The NCCS has human-coded about 200 thousand nonprofit organizations, the coded datasets and data dictionaries can be accessed here: <https://nccs-data.urban.org/index.php>

Another useful website is the Nonprofit Mission Classifiers: <http://bit.ly/2KHDHwh>. They will hold an online contest in 2019 fall, and you will have a chance to win a nationally recognized award.

3. Expectations and deliverables

The team is expected to have the following deliverables:

- a. A merged dataset generated from the 990 Forms and NCCS datasets.
- b. Best machine-learning models: one for classifying a nonprofit organization into 10 broad categories, and one for the 25 major groups (i.e., excluding “Z unknown”).² Reporting accuracy, precision, recall, and F1 is necessary.
- c. A standardized workflow and detailed report.

4. Contact

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

¹ Different sources confuse the use of these two terms sometimes.

² Can be the same model.