TIP for Tat: Political Bias in Human Rights Trafficking Reporting

This repository contains replication materials for the paper, *Rachel Harmon, Daniel Arnon, and Baekkwan Park "TIP for Tat: Political Bias in Human Rights Trafficking Reporting"* in *British Journal of Political Science*.

For questions, contact Baekkwan Park (baekkwan.park@gmail.com)

Directory structure and Files

- **code**: All the scripts to run the models and plot figures in the paper.
- **data**: documents and annotated list for the documents
- **outputs_all**: all the outputs from the models
  - *console*: log outputs for models (.txt)
  - *model_output*: model outputs (.csv)
  - *plots*: plots (.pdf)
- **outputs_asia**: all the Asia outputs from the models
  - *console*: log outputs for models (.txt)
  - *model_output*: model outputs (.csv)
  - *plots*: plots (.pdf)

Requirements

- Python3
  - scikit-learn 0.19.1
  - SciPy 1.0.0
  - Numpy 1.13.3
- R 3.6.2

Setup and Executing the Scripts (Estimated Running Time: 8 hours)

- **execute.sh**: this will run all the models and plot figures.
  - After the execute.sh, add the file path to the location of replication folder

```
$ ./execute.sh /path/to/REPLICATION_FILE
```

Directory Structure (REPLICATION_FILE)

1. Folder Structure Only

In the **REPLICATION_FILE** folder, you can find the following sub-folders and instruction files.

```
REPLICATION_FILE
│
├── README.md:
│          This is the file that you are reading right now.
│
├── execute.sh:
│          As explained above, by running this file, you run all
```

```
|          the models and draw plots.
|
├── code
|   |
|   ├── main.py:
|   |       This is the file that runs all the models the paper.
|   |       You do not need to run the rest of the code individually
|   |       because they are all imported and used as modules in this file.
|   |
|   ├── data_processing.py:
|   |       It converts all the raw documents into document term matrices for
|   |       models.
|   |
|   ├── classifier.py:
|   |       It estimates all the models (classifiers) used in the article.
|   |       (Naive Bayes, Logistic Regression, SVM, Random Forests)
|   |
|   ├── model_testing.py
|   |       It runs all the classifiers for rolling window prediction.
|   |
|   ├── voting_classifier.py
|   |       It runs majority vote classifier for rolling window predictions.
|   |
|   ├── bias.py
|   |       It calculates in-window and out-window bias for each model.
|   |
|   ├── output.py
|   |       It writes, save and export all the results from each model in
|   |       *.csv and *.txt format.
|   |
|   └── evaluation.R:
|          It draws line plots for accuracies for all models.
|
├── data
|   |
|   ├── DATA.pkl:
|   |       This file contains about 3,000 txt files from the State
|   |       Department Human Trafficking Reports (2001 - 2017) and TIP ratings.
|   |
|   └── wordcounts2.csv:
|           This file contains data for Figure 2 (median word counts for reports)
|
├── outputs_all
|   |
|   ├── model_output:
|   |     |   This folder contains results from rolling window predictions for all states.
|   |     |
|   |     └── fixed_5_model_evaluation.csv:
|   |            This is 5 year rolling window prediction results.
|   ├── console:
|   |     This directory contains log outputs for all the models. You can
|   |     find all the outputs for cross validation, tuning parameters,
|   |     various forms of evaluation metrics(precision, recall, F1,
|   |     accuracies, confusion matrix, etc.)
|   |
|   └── plots:
|         |    This directory contains all the plots used in the paper.
|         |
|         └── 5_Rolling_All.pdf
|
└── outputs_asia
    |
    ├── model_output:
    |     |   This folder contains results from rolling window predictions for Asian states.
    |     |
    |     └── fixed_5_model_evaluation.csv:
    |            This is 5 year rolling window prediction results.
    ├── console:
    |     This directory contains log outputs for all the models. You can
    |     find all the outputs for cross validation, tuning parameters,
    |     various forms of evaluation metrics(precision, recall, F1,
    |     accuracies, confusion matrix, etc.)
    |
    └──  plots:
```

```
│   This directory contains all the plots used in the paper.
│
└── 5_Rolling_Asia.pdf
```