

Impact of development indicators on CO2 Emission

Dependencies

- Python 3.8.10
 - requests 2.22.0, For importing data from world bank api
 - mlxtend 0.19.0, For association rule mining
 - numpy
 - pandas
 - pearsonr from scipy.stats.stats, For calculating pearson correlation coefficient.

Folder Structure

- *data* contains all the data files that are used as inputs
- *src* contains the python source code
- *config* contains any configuration related files.
- *Output* contains the csv output files(this will include any charts or images)
- *scripts* contains shell scripts for individual modules
- *reports* contains the final report/presentation
- *temp* may contain any intermediate files generated which cannot be classified under input/output
- *run.sh* is the main shell file which will be used to execute the whole project.

How to run

- Download data using [world bank api](#).
 - In config/indicators.csv specify the indicators to use by setting *selected* column to true. Specify the directory they go into using *type* column.
 - run the following shell script to download the data

```
sh ./scripts/getIndicators.sh
```
- Find Association Rules
 - The association requires its own preprocessing which is done by the *scripts/associationPrep.sh*
 - Then we run the association rule mining using *scripts/association.sh*
 - Finally the association analysis is done using *scripts/associationAnalysis.sh*

- All of this is combined into one and can be run using the following

```
sh ./scripts/associationOverall.sh
```

- Find correlation between two indicators
 - In config/features_for_correlation.csv add path of location of indicators files and their name. Also add output path with file_name.csv.
 - python file src/mainFeatureCorrelation.py is use to correlation for each indicators in the config file and save those in output/Correlation. Seprate csv file is genrated for every entry in config file.
 - Run mainFeatureCorrelation.py to genrate correlation result. It is saved in output/Correlation/

```
python3 ./src/mainFeatureCorrelation.py
```
 - Code for findind correlation is in pearsonCorrelation.py in ./src. This file is used by mainFeatureCorrelation.py to calculate R and P_value for each entry in config file

- For running everying use the followins script.

```
sh ./scripts/run.sh
```

Links & Refrences

- [world bank api](#)
- [Association Rule Mining - FP Growth](#)