DataAnalyticsProject

Installation

Install Python3 and R

Python Virtual environment preparation

It is suggested (not mandatory) to use a virtual environment:

- create it

```
# linux
python3 -m venv /path/to/myenv

# windows
python -m venv c:\path\to\myenv
```

- activate it

```
# linux
source /path/to/myenv/bin/activate # bash
source /path/to/myenv/bin/activate.fish # fish

# windows
c:\path\to\myenv\Scripts\activate.bat # cmd
c:\path\to\myenv\Scripts\Activate.ps1 # powershell
```

look at Python3 venv for more.

Requirements installation

```
pip install -r requirements.txt
```

Spacy italian

Run this command (see https://spacy.io/models/it for more)

```
python -m spacy download it_core_news_sm
```

N.B: some requirements may need a C++ compiler installed in the machine (e.g. spacy)

R dependency

Install `sentix` from https://github.com/valeriobasile/sentixR and its dependencies:

- Download `sentix_0.0.0.9000.tar.gz` from github (e.g `wget https://github.com/valeriobasile/sentixR/raw/master/sentix_0.0.0.9000.tar.gz`)
- Install R dependencies (from R shell):

```
install.packages(c("udpipe", "dplyr"))
```

- Install `sentix` from downloaded archive (seems `--no-staged-install` is required due to hardcoded paths).
  Run this from `bash` / `cmd`:

```
R CMD INSTALL --no-staged-install sentix_0.0.0.9000.tar.gz
```

Data

Put `products.json` and `reviews.json` datasets inside `data` folder

Usage

Activate first the virtual environment if any

Notebooks

We used `.Rmd` notebook with the help of `jupytext` in order to be able to version them on git

- Run `jupyter notebook`
- Manually open jupyter webapp if needed (probably at http://localhost:8888/)
- Open a notebook from `src` folder:
    - `Network.Rmd` : contains network analysis over products
    - `Sentiment.Rmd` : contains sentiment analysis over reviews

Dash app

- Enter `src` folder (e.g. `cd src`)
- Run the app:

  ```
  python webapp/app.py
  ```

  or for windows

  ```
  python .\webapp\app.py
  ```

- Go to http://127.0.0.1:8050/. This link is also showed in previous command output