

UBDA Python Example - text analysis using nltk for sentiment prediction - Aug 2020

Project Title

An example of text analysis using nltk for sentiment prediction

Reference

<https://adataanalyst.com/scikit-learn/countvectorizer-sklearn-example/>

Written by

* **UBDA ** - Mar 4, 2020

* **UBDA ** - Aug 23, 2021 Updated

Sources

All the program(s) and control file are available at
</ubda/apps/examples/python/text>.

Python program(s)

</ubda/apps/examples/python/text/CV.py>

PBS (control file)

</ubda/apps/examples/python/cvxopt/python.pbs>

Data

The datasets are available at www.imdb.com/interfaces/.

Create a directory for program and data:

```
cd $HOME
```

```
mkdir text
```

```
cd text
```

UBDA Python Example - text analysis using nltk for sentiment prediction - Aug 2020

After downloading one selected dataset, there are placed the data in
./text/IMDBDatasetSample-1.csv

```
## Executing the program
```

```
cd $HOME
```

```
cd text
```

```
cp -pr /ubda/apps/examples/python/text/* .
```

```
qsub python.pbs
```

```
### Environment setting
```

The python.pbs is include the proper environment settings, including:

```
module load anaconda3-5.2.0
```

```
source activate demo_py36
```

```
** packages in environment **
```

Name	Version	Build	Channel
_libgcc_mutex	0.1	main	
_tflow_select	2.1.0	gpu	
absl-py	0.8.0	py36_0	
astor	0.8.0	py36_0	
blas	1.0	mk1	
c-ares	1.15.0	h7b6447c_1001	
ca-certificates	2019.10.16	0	anaconda
certifi	2019.9.11	py36_0	anaconda
cloudpickle	1.2.2	py_0	anaconda
cuda-toolkit	9.0	h13b8566_0	
cuda-nn	7.6.0	cuda9.0_0	
cupti	9.0.176	0	
cycler	0.10.0	py_2	conda-forge
cytoolz	0.10.0	py36h7b6447c_0	anaconda

UBDA Python Example - text analysis using nltk for sentiment prediction - Aug 2020

dask-core	2.6.0	py_0	anaconda
dbus	1.13.6	he372182_0	conda-forge
decorator	4.4.0	py36_1	anaconda
expat	2.2.5	he1b5a44_1004	conda-forge
fontconfig	2.13.1	he4413a7_1000	conda-forge
freetype	2.10.0	he983fc9_1	conda-forge
gast	0.3.2	py_0	
gettext	0.19.8.1	hc5be6a0_1002	conda-forge
glib	2.58.3	h6f030ca_1002	conda-forge
grpcio	1.16.1	py36hf8bcb03_1	
gst-plugins-base	1.14.5	h0935bb2_0	conda-forge
gstreamer	1.14.5	h36ae1b5_0	conda-forge
icu	58.2	hf484d3e_1000	conda-forge
imageio	2.6.1	py36_0	anaconda
intel-openmp	2019.4	243	
joblib	0.14.0	py_0	anaconda
jpeg	9c	h14c3975_1001	conda-forge
kiwisolver	1.1.0	py36hc9558a2_0	conda-forge
libedit	3.1.20181209	hc058e9b_0	
libffi	3.2.1	hd88cf55_4	
libgcc-ng	9.1.0	hdf63c60_0	
libgfortran-ng	7.3.0	hdf63c60_0	
libiconv	1.15	h516909a_1005	conda-forge
libpng	1.6.37	hed695b0_0	conda-forge
libprotobuf	3.9.2	hd408876_0	
libstdcxx-ng	9.1.0	hdf63c60_0	
libtiff	4.0.10	h2733197_2	anaconda
libuuid	2.32.1	h14c3975_1000	conda-forge
libxcb	1.13	h14c3975_1002	conda-forge
libxml2	2.9.9	h13577e0_2	conda-forge
markdown	3.1.1	py36_0	
matplotlib	3.1.1	py36h5429711_0	
mkl	2019.4	243	
mkl-service	2.3.0	py36he904b0f_0	

UBDA Python Example - text analysis using nltk for sentiment prediction - Aug 2020

mkl_fft	1.0.14	py36ha843d7b_0	
mkl_random	1.1.0	py36hd6b4f25_0	
ncurses	6.1	he6710b0_1	
networkx	2.4	py_0	anaconda
nltk	3.4.5	py36_0	anaconda
numpy	1.17.2	py36haad9e8e_0	
numpy-base	1.17.2	py36hde5b4d6_0	
olefile	0.46	py36_0	anaconda
openssl	1.1.1	h7b6447c_0	anaconda
pandas	0.25.2	py36he6710b0_0	anaconda
patsy	0.5.1	py36_0	anaconda
pcre	8.43	he1b5a44_0	conda-forge
pillow	6.2.0	py36h34e0f95_0	anaconda
pip	19.3.1	py36_0	
protobuf	3.9.2	py36he6710b0_0	
pthread-stubs	0.4	h14c3975_1001	conda-forge
pyparsing	2.4.2	py_0	conda-forge
pyqt	5.9.2	py36hcca6a23_4	conda-forge
python	3.6.9	h265db76_0	
python-dateutil	2.8.0	py_0	conda-forge
pytz	2019.3	py_0	conda-forge
pywavelets	1.1.1	py36h7b6447c_0	anaconda
qt	5.9.7	h52cfd70_2	conda-forge
readline	7.0	h7b6447c_5	
scikit-image	0.15.0	py36he6710b0_0	anaconda
scikit-learn	0.21.3	py36hd81dba3_0	anaconda
scipy	1.3.1	py36h7c811a0_0	anaconda
seaborn	0.9.0	py36_0	anaconda
setuptools	41.4.0	py36_0	
sip	4.19.8	py36hf484d3e_0	
six	1.12.0	py36_0	
sqlite	3.30.1	h7b6447c_0	
statsmodels	0.10.1	py36hdd07704_0	anaconda
stop-words	2018.7.23	pypi_0	pypi

UBDA Python Example - text analysis using nltk for sentiment prediction - Aug 2020

tensorboard	1.9.0	py36hf484d3e_0	
tensorflow	1.9.0	gpu_py36h02c5d5e_1	
tensorflow-base	1.9.0	gpu_py36h6ecc378_0	
tensorflow-gpu	1.9.0	hf154084_0	
termcolor	1.1.0	py36_1	
tk	8.6.8	hbc83047_0	
toolz	0.10.0	py_0	anaconda
tornado	6.0.3	py36h516909a_0	conda-forge
werkzeug	0.16.0	py_0	
wheel	0.33.6	py36_0	
xorg-libxau	1.0.9	h14c3975_0	conda-forge
xorg-libxdmcp	1.1.3	h516909a_0	conda-forge
xz	5.2.4	h14c3975_4	
zlib	1.2.11	h7b6447c_3	
zstd	1.3.7	h0b5b093_0	anaconda