Universitat Oberta de Catalunya

uoc.edu

Tipologia i cicle de dades · PRACTICA · 2017-2018

EEES · Màster de data Science

Relació i patrons entre moviments de índexs de borsa i cryptomoneda

Nom i Cognoms: Albert Costas Gutiérrez

UOC practica 1. Tipologia i cicle de vida.

Descripció

Datasets per la comparació de moviments i patrons entre els principals indexs borsatils espanyols i les crypto-monedes

Links	Fitxers
Repositori github: https://github.com/acostasg/scraping Repositori kaggle Open data: • https://www.kaggle.com/acostasg/stock-index • https://www.kaggle.com/acostasg/crypto-currencies	• Document PDF amb les respostes de les preguntes i els noms dels components del grup. • Fitxer amb el codi Pyhton per obtenir les dades • Carpeta CSV amb les dades

Estructura

```
scraping
├── pdf
│   └── acostasg-PRACTICA_1.pdf  # Document pdf amb les respostes a les preguntes i els noms del componentes del grup
│
├── csv  # datasets
│   ├── cypto_currencies
│   │   └── ... # fitxers csv
│   └── stock_index
│       └── ... # directoris per data amb els csv
│
├── projects
│   ├── scraping_crypto_currencies.py # scraping url criptomoneda
│   └── scraping_stock_indexs.py # scraping url el economista
│
├── README.md
├── scraping.py # fitxer python incial
└── setup.py
```

Autors

Albert Costas Gutierrez - acostasg@uoc.edu

Llicència

Fonts de dades

- http://www.eleconomista.es
- https://coinmarketcap.com

Les dades de borsa i crypto-moneda estan en última instància sota llicència de les webs respectivament.

Respostes de les preguntes

1. Subtítol del dataset. Agregueu una descripció àgil del vostre conjunt de dades pel vostre subtítol.

En aquest hi ha 2 datasets, però l'objectiu és poder **comparar** en el mateix període de temps si hi ha **relació o es podreixen patrons** comuns entre els **moviments borsatils** dels principals indexes espanyols i els **moviments de les crypto-monedes**.

Subtítol:

«Datasets per la comparació de moviments i patrons entre els principals índexs borsatils espanyols i les crypto-monedes»

2. Imatge. Agregueu una imatge que identifiqui el vostre dataset visualment.

Disposem de 2 datasets, en el primer dels moviments borsatils dels principals *índexs* espanyols:

El segon dels moviments en el mateix període de temps de la cryptomoneda:

3. Context. Quina és la matèria del conjunt de dades?

En aquest cas el context és detectar o preveure els **diferents moviments que es produeixen per una serie factors**, tant de moviment interns (compra-venda), com externs (moviments polítics, econòmics, etc...), en els principals índexs borsatils espanyols i de les crypto-monedes.

Hem seleccionat diferents fonts de dades per generar fitxers «csv», **guardar diferents valors en el mateix període de temps**. És important destacar que ens interessa més les tendències alcistes o baixes, que podem calcular o recuperar en aquests períodes de temps.

4. Contingut. Quins camps inclou? Quin és el període de temps de les dades i com s'ha recollit?

En aquest cas el contingut està format per diferents csv, especialment tenim els fitxers de moviments de **cryptomoneda**, els quals s'ha generat **un fitxer per dia del període de temps estudiat**.

Pel que fa als moviments del principals **índexs borsatils s'ha generat una carpeta per dia del període, en cada directori un fitxer amb cadascun del noms dels índexs.** Degut això s'han comprimit aquests últims abans de publicar-los en el directori de «*open data*» kaggle.com.

Pel que fa als camps, ens **interessà detectar els moviments alcistes i baixistes**, o almenys aquelles que tenen un patró similar en les cryptomonedes i els índexs. Els camps especialment destacats són:

Camps comuns:

- **Nom**: Nom empresa o cryptomoneda;
- **Preu**: Valor en euros d'una acció o una cryptomoneda;
- **Volum:** En euros/volum 24 hores,acumulat de les transaccions diàries en milions d'euros

Crypto-currencies:

- **Simbol**: Símbol o acrònim de la moneda
- **Cap de mercat**: Valor total de totes les monedes en el moment actual
- **Oferta circulant**: Valor en oportunitat de negoci
- **% 1h, % 2h i %7d**, tant per cent del valor la moneda en 1h, 2h o 7d sobre la resta de cyprtomonedes.

Stock Index:

- **Estat**: Estat final en tancament en alta o baixa del dia.
- **Var. Per cent**: Variació en el moment del tancament amb tant per cent respecte el dia anterior
- **Var. En euros**: Variació en el moment del tancament amb euros respect el dia anterior.
- **Capitalització**: Valor de l'empra respecte les seves accions.
- **PER**: La ràtio preu-benefici
- **Rent./Div:** Rendibilitat de l'acció respecte el valor inicial de la acció.

5. Agraïments. Qui és propietari del conjunt de dades? Inclou cites de recerca o anàlisi anteriors.

En aquest cas les fonts de dades que s'han utilitzat per a la realització dels datasets corresponent a:

- http://www.eleconomista.es
- https://coinmarketcap.com

Per aquest fet, les dades de borsa i crypto-moneda estan en **última instància sota llicència de les webs** respectivament.

Pel que fa a la **terminologia financera** podem veure vocabulari en renta4banco. [https://www.r4.com/que-necesitas/formacion/diccionario]

També podem veure **un estudi anterior** on poder tenir primícies de com han enfocat els algoritmes:

- https://arxiv.org/pdf/1410.1231v1.pdf

6. Inspiració. Per què és interessant aquest conjunt de dades? Quines preguntes li agradaria respondre la comunitat?

En aquest cas **el «trading» en cryptomoneda** és relativament nou, força popular per la seva formulació com a mitja digital d'intercanvi, **utilitzant un protocol que garanteix la seguretat, integritat i equilibri** del seu estat de compte per mitjà d'un entramat d'agents.

La comunitat podrà respondre, entre altres preguntes, a:

- Està afectant o hi ha **patrons comuns** en les cotitzacions de cryptomonedes i el mercat de valors principals del país d'Espanya?
- Els efectes o agents **externs afecten per igual a les accions o cryptomonedes**?
- Hi ha **relacions cause efecte** entre les acciones i cryptomonedes?

7. Llicència. Cal que seleccioneu una d'aquestes llicències i cal dir perquè l'heu seleccionada:

Pel que fa a la llicència hem de tenir present les fonts de dades, en aquesta practica **hem de referenciar a les fonts**, i en última instància són aquestes les propietàries de les dades.

Per aquest fet hem seleccionat la llicència:

- Database released under Open Database License, **individual contents** under Database Contents License.

En definitiva, com que hi ha part de les **dades de diferents fonts citarem les fonts de les dades perquè els usuaris de les mateixes puguin veure els propietaris d'aquestes**, i en la llicència seleccionada farem constatar aquest fet.

8. Codi: Cal adjuntar el codi amb el que heu generat el dataset, preferiblement amb R o Python, que us ha ajudat a generar el dataset.

El codi que s'ha programat per generar el dataset està amb codi Python amb el paquet o modul **Beautiful Soup** que encapsula el «*parser*» de les pàgines web. En aquest sentit s'ha generat una estructura on poder agregar noves pàgines i extraure dades, en la **carpeta projectes hi ha els dos fitxers Python que tenen la responsabilitat de quines pàgines i com és realitzat el «scraping».**

Destacar el fitxer que s'encarrega de la pàgina web de l'«El economista» per poder extraure les principals dades dels indicadors de la borsa espanyola, **el qual té un llistat d'aquells índexs que s'han d'extraure les dades**.

Per l'altra banda el resultats es guarden en la **carpeta csv**, on es referencia clarament **en el nom del fitxer o carpeta la data** de quan es van extraure les dades (la data), realitzats diàriament.

```
scraping
├── csv
│   ├── cypto_currencies
│   │   └── ... # fitxers csv
│   └── stock_index
│       └── ... # directoris per data amb els csv
│
├── projects
│   ├── scraping_crypto_currencies.py # scraping url criptomoneda
│   └── scraping_stock_indexs.py # scraping url el economista
│
├── README.md
├── scraping.py # fitxer python incial
└── setup.py
```

S'ha utilitzat com wiki i com a repositori de codi el **github** amb llicència de codi lliure:

https://github.com/acostasg/scraping

9. Dataset: Dataset en format CSV

Els fitxers csv generats que componen el dataset s'han **publicat** en el repositori kaggle.com:

- https://www.kaggle.com/acostasg/stock-index/
- https://www.kaggle.com/acostasg/crypto-currencies

Per una banda, els fitxers els «stock-index» estan comprimits per **carpetes amb la data d'extracció** i cada fitxer amb el nom dels índexs borsatil. De forma diferent, les cryptomonedes aquestes estan **dividides per fitxer on són totes les monedes amb la data d'extracció**.

Referencies

- https://stackoverflow.com/questions/6159900/correct-way-to-write-line-to-file-in-python
- Llibre manual: Richard Lawson. Web Scraping with Python. Packt Publishing Ltd, October 2015. 174 p. ISBN 9781782164371
- https://docs.python.org/2/tutorial/inputoutput.html