

Reading course Syrto: materiale preliminare

Contenuto della cartella compressa `Syrto.rar`

Dataset:

Il dataset è nel formato *parquet*. Il file `data_full_1.3.parquet` contiene i dati relativi all'andamento annuale di milioni di aziende italiane, dal 2009 al 2017. La cartella `GDP` contiene informazioni sul prodotto interno lordo dei vari paesi (in questa fase non è oggetto di approfondimento, ma il reader originale Syrto recupera anche questa informazione).

Codice:

Lo script Python `read_dataset` contiene le funzionalità di lettura del dataset. Il metodo `read` legge un file *parquet* e restituisce un oggetto di tipo *Dataframe* (gestito dalla libreria *pandas*) contenente i dati richiesti. Come da codice commentato, il preambolo seleziona aziende di un determinato settore (in questa fase i parametri relativi rimarranno quelli assegnati nel codice, in quanto al momento abbiamo considerato solo le aziende di un determinato settore). Per completezza, il codice contiene anche una variabile (`targets`) in cui sono riportati gli attributi dei dati. Il primo obiettivo dei regressori/classificatori da implementare sarà comunque la predizione del target "Turnover".

N.B. La cartella `syrto` contiene il modulo in cui sono implementate le funzionalità di lettura e analisi del file contenente il dataset.

Dipendenze:

Per il corretto funzionamento del codice sono necessarie le librerie Python:

- `pandas`
- `torch`
- `pyarrow`

Documenti e articoli:

I documenti utili per il corso si trovano specificati di seguito. I file PDF si trovano nella cartella documentazione:

- Report finale progetto Syrto 1.0 (link): <https://cordis.europa.eu/project/id/320270/reporting/it>
- Presentazione finale Syrto 1.0: `presentazione_syrto_completa.pdf`
- Articolo scientifico 'A survey of time series forecasting from stochastic method to soft computing', di P.Hendikawati: `hendikawati_2020.pdf`
- tutorial su time series forecasting, completi di esempi di implementazione Python (link): <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775>
<https://iopscience.iop.org/article/10.1088/1742-6596/1613/1/012019/pdf>