

Dati

Oltre alle 11 feature degli esperimenti preliminari ("FixedAssets", "CurrAssets", "Debtors", "Cash", "Capital", "Ltdebt", "CurrLiab", "WorkingCap", "CurrRatio", "LiqRatio", "Turnover") i dati sono rappresentati mediante nuove feature, ovvero;

- "EBIT"
- "WorkCap_Turn_ratio" (rapporto tra WorkingCap e Turnover)
- "Turn_FixAs_ratio" (rapporto tra Turnover e FixedAssets)
- "EBIT_Turn_ratio" (rapporto tra EBIT e Turnover)

Code

Il codice è aggiornato per recuperare i dati descritti in precedenza, in particolare il modulo "utils" e lo script "read_dataset". Notare che quest'ultimo è modificato aggiungendo un parametro di cut-off anche per la variabile FixedAssets, per evitare divisioni per 0, e quindi valori non numerici. Per completezza, è aggiornata anche la lista delle feature.

Task

Di seguito gli obiettivi di questo step:

- stessi esperimenti di regressione (a sto punto magari escludendo i regressori che hanno visto le performance peggiori, quelli in rosso nel report precedente)
- esperimenti di predizione non solo della variabile "Turnover", ma almeno sulle nuove aggiunte ("EBIT", "WorkCap_Turn_ratio", "Turn_FixAs_ratio", "EBIT_Turn_ratio"), le quali sono le più importanti per un analista.
- Esperimenti di classificazione "a valle" del Turnover: considerando le predizioni effettuate dal regressore, procedere secondo i seguenti step
 - *binarizzazione* della variabile di output: assegnare valore 1 se la differenza tra il Turnover e il valore dell'anno precedente è maggiore o uguale a 0, altrimenti assegnare il valore 0.
 - *valutazione*: la binarizzazione va effettuata sia per il Turnover predetto (\bar{y}) che per quello reale (y) utilizzare le metriche di valutazione classiche (*accuracy*, *precision*, *recall*, eventualmente anche *AUC*) fornite da sk-learn
- Eventuale: classificazione anche sulle altre variabili predette