

Progetto Data Mining A.A. 2016/2017

Membri: gaspa,....

Tipologia: progetto personalizzato (early group).

Descrizione:

Backblaze [link](#) è una azienda che fornisce servizi di backup online. Dal 2013 pubblica e rende scaricabili [link](#) report e statistiche dei dischi fissi che vengono usati per questo servizio.

In particolare il dataset scelto sarà quello dell'anno 2016 (l'ultimo a disposizione, oltre ad essere quello più grande).

Il dataset è composto da 366 file .csv (uno per ogni giorno dell'anno). Ogni file .csv contiene al suo interno una riga per ogni disco monitorato, ogni riga contiene il numero seriale del disco, il modello e i dati S.M.A.R.T. di quel giorno (e qualche altro campo in più).

All'inizio dell'anno 2016 i dischi registrati nel dataset erano 57546, alla fine del 2016 i dischi monitorati risultano essere 73653.

S.M.A.R.T. [link](#) è un sistema di monitoraggio incluso in ogni disco fisso. Esso rileva e riporta vari indicatori riguardanti soprattutto l'affidabilità del disco con lo scopo di permettere il rilevamento anticipato di guasti.

Alcuni dei dati che S.M.A.R.T. rileva sono: settori riallocati, conteggio di tentativi di avvio di rotazione dei piatti, rate di errori di lettura, ecc. Rileva però anche altri valori come ad esempio la temperatura, il numero di ore in cui il disco è stato in rotazione, cicli di start/stop, ecc.

Riguardo all'utilizzo di questi dati Backblaze impone 3 requisiti chiavi sullo scaricamento e l'uso di questi dati.

Si richiede che [link](#):

- si citi Backblaze come sorgente dei dati.
- Si accetta di essere l'unico responsabile per il come i dati vengono usati.
- questi dati non vengano venduti.

Tutti i requisiti ne permettono l'uso in questo progetto.