

## Sentiment Analysis - Progetto Pratico

*Alessio Castrica*

Questo documento ha lo scopo di commentare il Jupyter notebook necessario per conseguire l'idoneità nel corso di Social Data Analytics.

Il progetto pratico contiene l'analisi di titoli di articoli relativi a notizie di mercato, con l'obiettivo di prevedere l'andamento giornaliero della società menzionata nelle notizie.

In particolare, vengono usati dati dal 2020-06-01 al 2022-06-01, allegati nel file "news\_sentiment.h5" relativi ai titoli degli articoli e al prezzo medio della società Apple (ticker: AAPL).

Come indice del sentiment di mercato viene usato il ritorno giornaliero.

La struttura del notebook è la seguente:

1. Caricamento delle librerie e dei dati;
2. Ricampionamento e pre-elaborazione dei dati;
3. Analisi del sentiment usando prima Vader e poi classificazione basata sul Naive Bayes;
4. Visualizzazione dei risultati.

Inoltre i due metodi di analisi del sentiment sono tra loro molto diversi.

Il metodo di sentiment analysis basato su Vader assegna un polarity score tra -1 ed 1 ai titoli delle notizie, mentre il metodo basato sul teorema di bayes vettorizza i testi e predice la probabilità di avere un certo market sentiment dati i titoli delle notizie di quel giorno.

Quest'ultimo metodo risulta avere un'accuracy del 60% sul dataset di test, riuscendo a prevedere in maniera corretta il sentiment di mercato in più della metà degli esempi.