**Modello 1**

Input tweet su diversi hashtags, target l’argomento

**Tipi argomenti:**

* Sport
* Crypto
* Ambientale / Ecologico
* Guerra Russia/Ucraina

Obiettivo: capire il contesto per rindirizzare ad un vocabolario specifico

**Modello 2**

Input stessi dati di prima

Obiettivo creare diversi dizionari specializzando Vader su ogni argomento per fare una analisi più specifica sull’argomento di interesse.

* Logist regression per assegnare i pesi
* Usare ebbedding per parole simili e inserire le parole trovate con lo stesso peso
* Creare prima il vocabolario, successivamente fare un sentiment con vader e vader migliorato da noi con il nostro vocabolario (per fare questo bisogna fare vader extesion)

**Confronto vader**

Valuto i miei testi con vader e aggiungo al mio dataset la variabile conpound, rifaccio tutto con il nostro vader e associo un'altra variabile con i valori di conpound del nostro vader valutando le differenze tra i due con quello di partenza.

**Domande**

* **Modello logistico come valutare le parole più importanti**

Scritto sul codice pesi

* Se usare gli n\_gram o singole parole

Singole parole

* Se usare N-Gram o singole parole nella parte delle similarità

Singole parole

Valutare vader senza n grammi e provare anche con n grammi più avanti

* Se inserire altre parole alla stopwords

Non faccio stopword, non serve

#### **Dataset**

#### Trip Advisor Hotel Reviews

#### Finance

* Amazon Fine Food Reviews
* Drug