



# Enhancing irony detection with affective information

Relatore: *Prof.ssa Elisabetta Fersini*

Co-Relatore: *Dott.ssa Debora Nozza*

Presentazione della prova finale di:

*Gianluca Giudice*

*Matricola 830694*

# Sentiment analysis e ironia

---



# Cos'è la sentiment analysis

La sentiment analysis è un campo del NLP che si occupa di costruire sistemi per l'identificazione ed estrazione di opinioni dal testo.



My experience  
so far has been  
fantastic!

POSITIVE



The product is  
ok I guess

NEUTRAL



Your support team  
is useless

NEGATIVE



Un esempio di fonte di espressione sono i social network. Estrarre l'opinione contenuta in questi dati non strutturati generati da utenti permette di conoscere opinioni di consumo, le preferenze, le idee.



# Perché è importante riconoscere l'ironia

- Strumento utilizzato nella comunicazione
- Un individuo esprime l'opposto di ciò che realmente intende
- L'uso dell'ironia in una frase è in grado di cambiare completamente la polarità del sentiment di ciò che viene detto
- Diventa di fondamentale importanza sviluppare dei sistemi di sentiment analysis consapevoli dell'esistenza del fenomeno e in grado di riconoscerlo

# Approccio al problema

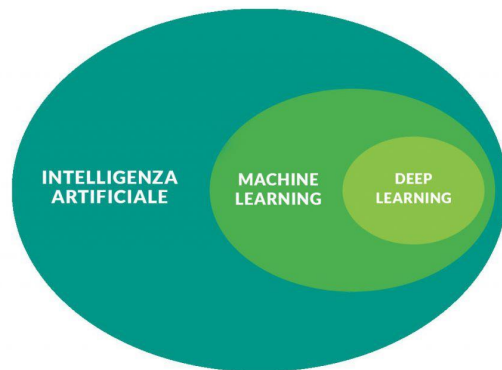
---

# Modelli supervisionati di machine learning

Avendo a disposizione un dataset etichettato composto da 40.000 tweets in lingua inglese, vengono creati dei modelli supervisionati di machine learning utilizzando diversi algoritmi di apprendimento e considerando più features.

Si seguono quindi i seguenti macro-step:

1. Data pre-processing
2. Feature extraction
3. Training dei modelli
4. Valutazione delle performance



---

# Features utilizzate



# Rappresentazione del testo

Necessaria una rappresentazione numerica del testo da dare in input agli algoritmi di apprendimento

- Boolean bag-of-words (BOW)
  - Ogni documento è rappresentato da un multiinsieme
  - Viene costruito un dizionario considerando i tokens univoci
  - Matrice booleana composta da bow sulle righe e parole del dizionario sulle colonne
- BERT (e Sentence-BERT)
  - Rete neurale creata da Google
  - Vengono generati embeddings a partire dai tokens del documento
  - Viene tenuto conto della semantica che il token assume all'interno della frase





# Feature expansion

Si considerano altre features oltre al testo

- Particelle pragmatiche
  - Emoticons
  - Abbreviazioni (slang)
  - Espressioni onomatopeiche
  - Punteggiatura
- POS tags
  - Ad ogni token si associa: Nome; aggettivo; pronome; ecc...
- Emotional features
  - Utilizzo di lessici per associare emozioni relative: Gioia; rabbia; paura; ecc...

Polarità

Frequenza

---

# Algoritmi di apprendimento supervisionato



# Algoritmi utilizzati

- Decision Tree
- Multinomial Naive Bayes
- Support Vector Machine - Kernel lineare
- Bayesian Network

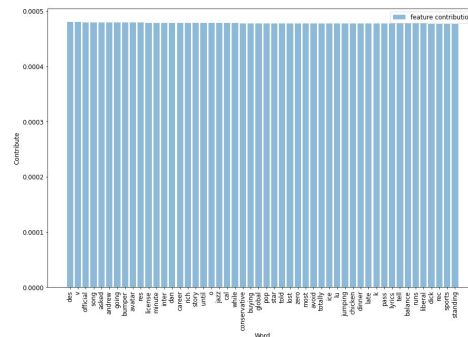
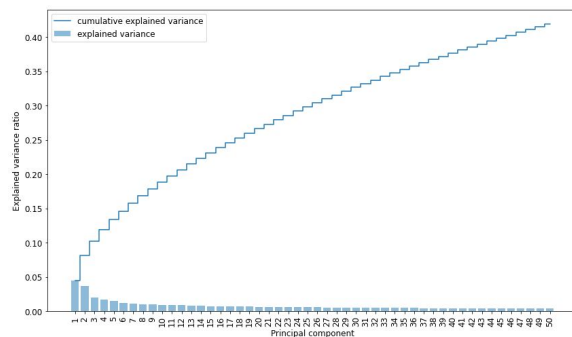
# Esperimenti

---

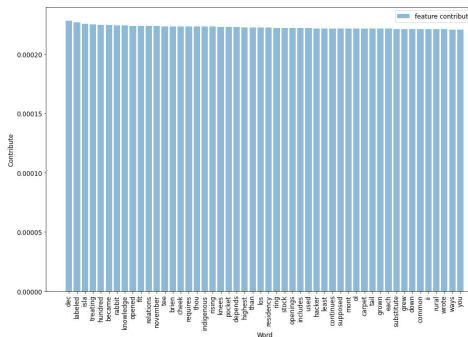
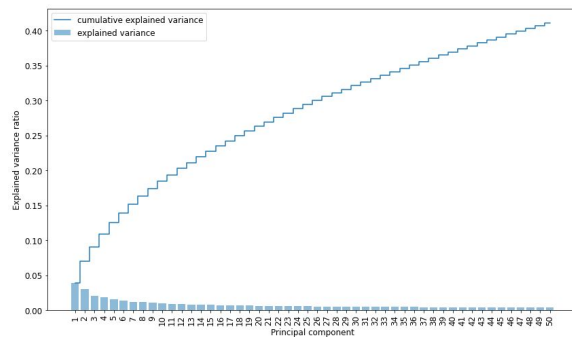


# PCA - Associazione tra lessico e ironia

Dataset  
ironico



Dataset  
non ironico



# Conclusioni

---



# Conclusioni e sviluppi futuri

- Il riconoscimento dell'ironia è un task complesso nell'ambito NLP
- Anche i modelli allo stato dell'arte sono meno performanti rispetto agli umani
- Dai dati emerge che considerare features relative alla sfera emotiva migliora le performance dei modelli
- In letteratura si suggerisce l'importanza del contesto, pertanto considerare questo aspetto in sviluppi futuri