

# CAPITOLO 4: Analisi del sentiment di un dataset di articoli di giornale riguardanti l'IA ed il suo impatto sul mondo del lavoro

## 4.1 Confronto tra termini a valenza positiva e negativa in un'analisi esplorativa

Inizialmente, per formare un'idea sui principali sottotemi trattati, si disegna la wordcloud con i principali termini, divisi tra positivi e negativi, che compaiono nei testi degli articoli considerati. Per determinare la natura positiva o negativa dei termini, è stato utilizzato il dizionario Bing.

```
library(tidyverse)

## — Attaching core tidyverse packages — tidyverse 2.0.0
##
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats   1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
## ✓ lubridate 1.9.4      ✓ tidyr      1.3.1
## ✓ purrr     1.0.2
## — Conflicts — tidyverse_conflicts()
##
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
## conflicts to become errors

library(tidytext)
library(tm)

## Caricamento del pacchetto richiesto: NLP
##
## Caricamento pacchetto: 'NLP'
##
## Il seguente oggetto è mascherato da 'package:ggplot2':
##
##      annotate

library(wordcloud)

## Caricamento del pacchetto richiesto: RColorBrewer

library(dplyr)
library(readxl)
gruppo_articoli<-
read_excel("C:/Users/1jaco/OneDrive/Desktop/SMA/Progetto_SMA/Dataset_finale/Datase
tEsteso.xlsx")
```

*##estrazione dei testi degli articoli*

```
v_text <- data.frame(text=gruppo_articoli$testo)
```

*#trasformazione dei dati in modo da avere liste positive e negative*

```
positive_senti <- get_sentiments("bing") %>% #utilizzo del dizionario di bing per l'analisi
```

```
  filter(sentiment == "positive") #utilizzo filtro per termini positivi; con il complemento si
```

*#trova la lista di parole negative*

```
bing <- get_sentiments("bing") #data frame con parole e relativi sentiment (secondo dizionario bing)
```

```
v_tidy_data <- v_text %>%
```

```
  mutate(linenumero = row_number()) %>%
```

```
  ungroup() %>%
```

```
  unnest_tokens(word, text)
```

*#unnest\_tokens(): Fa parte del pacchetto tidytext e serve a suddividere un testo in unità più piccole, solitamente parole (tokenizzazione).*

*#word: Nome della nuova colonna che conterrà i token (cioè le parole estratte).*

*#text: Colonna originale che contiene il testo completo da tokenizzare.*

*#esclusione dei termini tematici*

```
theme_words <- c("ai", "artificial", "intelligence", "work", "like")
```

*# Filtrare con subset*

```
v_tidy_data <- subset(v_tidy_data, !(word %in% theme_words))
```

```
library(reshape2)
```

```
##
```

```
## Caricamento pacchetto: 'reshape2'
```

```
##
```

```
## Il seguente oggetto è mascherato da 'package:tidyr':
```

```
##
```

```
##      smiths
```

```
library(wordcloud)
```

```
v_tidy_data %>%
```

```
  inner_join(bing) %>% # Effettua un join interno tra v_tidy_data e bing, restituendo solo le righe in cui le parole di v_tidy_data sono presenti anche nel dizionario "bing".
```

```
  count(word, sentiment, sort = TRUE) %>% #Conta quante volte ogni parola appare per ciascun tipo di sentimento.
```

```
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
```

*#Fa parte del pacchetto reshape2 ed è utilizzato per trasformare i dati in una tabella incrociata (pivot table).*

*#word ~ sentiment: Le parole diventano righe e i sentimenti (positivo e negativo) diventano colonne.*

*#value.var = "n": I valori della tabella sono i conteggi delle parole.*

```
  comparison.cloud(colors = c("red", "dark green"),  
                    max.words = 100)
```

```
## Joining with `by = join_by(word)`
```



La wordcloud mostrata in questo formato permette di evidenziare un dibattito acceso sugli impatti dell'intelligenza artificiale sul lavoro. Le parole come “benefit”, “gains”, “better” e “skilled” suggeriscono una percezione positiva di alcune opportunità che l'intelligenza artificiale può offrire, come il miglioramento delle competenze, dei processi lavorativi e dei benefici economici. Al contrario, termini come “risk” e “loss” evidenziano preoccupazioni legate a disoccupazione tecnologica, perdita di posti di lavoro e, in generale, i lati negativi di una cattiva gestione della nuova tecnologia. Ulteriori temi che emergono dall'analisi sono quelli della complessità e dell'incertezza. Parole come “complex”, “critical”, “problem” indicano che il tema dell'IA nel lavoro è percepito come difficile da comprendere o da affrontare, con implicazioni che richiedono soluzioni ben ponderate e consapevoli. Termini come “trust” e “fear” sottolineano l'importanza di costruire fiducia nei confronti dell'IA, contrastando la paura delle sue conseguenze negative.

Da sottolineare la presenza di ambiguità in alcuni dei termini evidenziati dalla wordcloud. Tra queste, la parola “cloud” potrebbe riferirsi a tecnologie specifiche, come il cloud computing, che sono direttamente collegate all'IA e alla trasformazione digitale, dunque, probabilmente nella realtà non possiede la connotazione negativa che qui gli viene data.

In conclusione, la tensione tra opportunità (riqualificazione, miglioramento) e rischi (disoccupazione, complessità) sembra di primaria importanza.

Ora si vogliono comparare globalmente le quantità di termini riconosciuti come positivi e negativi negli articoli, per avere un'idea sulla sensazione predominante tra le due.

Si effettua il conteggio termini positivi e negativi secondo il dictionary\_LSD2015, il dizionario fornito dalla library quanteda.

```
library(quanteda)

## Package version: 4.1.0
## Unicode version: 15.1
## ICU version: 74.1

## Parallel computing: 12 of 12 threads used.

## See https://quanteda.io for tutorials and examples.

##
## Caricamento pacchetto: 'quanteda'

## Il seguente oggetto è mascherato da 'package:tm':
##
## stopwords

## I seguenti oggetti sono mascherati da 'package:NLP':
##
## meta, meta<-

toks_news<-quanteda::tokens(gruppo_articoli$testo)
toks_news_lsd <- quanteda::tokens_lookup(toks_news,
dictionary=data_dictionary_LSD2015[1:2]) #classificazione dei termini nei
differenti testi in pos e neg sulla base del dizionario fornito da quanteda
head(toks_news_lsd, 2)

## Tokens consisting of 2 documents.
## text1 :
## [1] "positive" "positive" "negative" "positive" "positive" "positive"
## [7] "negative" "positive" "positive" "positive" "positive" "positive"
## [ ... and 47 more ]
##
## text2 :
## [1] "positive" "positive" "negative" "negative" "positive" "negative"
## [7] "positive" "positive" "negative" "positive" "positive" "negative"
## [ ... and 57 more ]

dfmat_news_lsd <- dfm(toks_news_lsd) #create a document-feature matrix, per ogni
documento, presenti numero di termini positivi e negativi
head(dfmat_news_lsd, 2)

## Document-feature matrix of: 2 documents, 2 features (0.00% sparse) and 0
docvars.
##           features
## docs    negative positive
## text1         11      48
## text2         22      47
```

```
#evidenziazione di un sentiment positivo
```

```
dfmat_news_lsd<-as.matrix(dfmat_news_lsd)  
rownames(dfmat_news_lsd) <- gruppo_articoli$giornale
```

```
apply(dfmat_news_lsd, 2, sum) #somma delle colonne della matrice -> tot. termini  
positivi e negativi
```

```
## negative positive  
##      1886      3722
```

Ciò che si nota in generale è la maggioranza di termini positivi rispetto ai negativi.

## 4.2 Analisi del sentiment degli articoli dal punto di vista cronologico

Si utilizza la libreria syuzhet per il calcolo del sentiment su ciascuno dei testi degli articoli presenti; comparazione dell'andamento del sentiment nel tempo sulla base delle date a cui risalgono gli articoli selezionati.

```
library(syuzhet)  
library(ggplot2)  
library(zoo)
```

```
##  
## Caricamento pacchetto: 'zoo'  
  
## Il seguente oggetto è mascherato da 'package:quanteda':  
##  
##      index
```

```
## I seguenti oggetti sono mascherati da 'package:base':  
##  
##      as.Date, as.Date.numeric
```

```
gruppo_articoli$sentiment <- get_sentiment(gruppo_articoli$testo)
```

```
#calcolo della media mobile per smussare il grafico così da renderlo maggiormente  
leggibile
```

```
gruppo_articoli$sentiment_ma <- rollmean(gruppo_articoli$sentiment, k = 8, fill =  
NA, align = "left")
```

```
gruppo_articoli$data <- as.Date(gruppo_articoli$data)
```

```
# Creiamo il grafico
```

```
ggplot(gruppo_articoli, aes(x = data, y = sentiment_ma)) +  
  geom_line(color = "blue") +   # Linea  
  geom_point(color = "red") +   # Punti  
  labs(  
    title = "Il sentiment degli articoli sull'IA nel tempo",
```

```

  x = "Data",
  y = "Sentiment"
) +
theme_minimal()

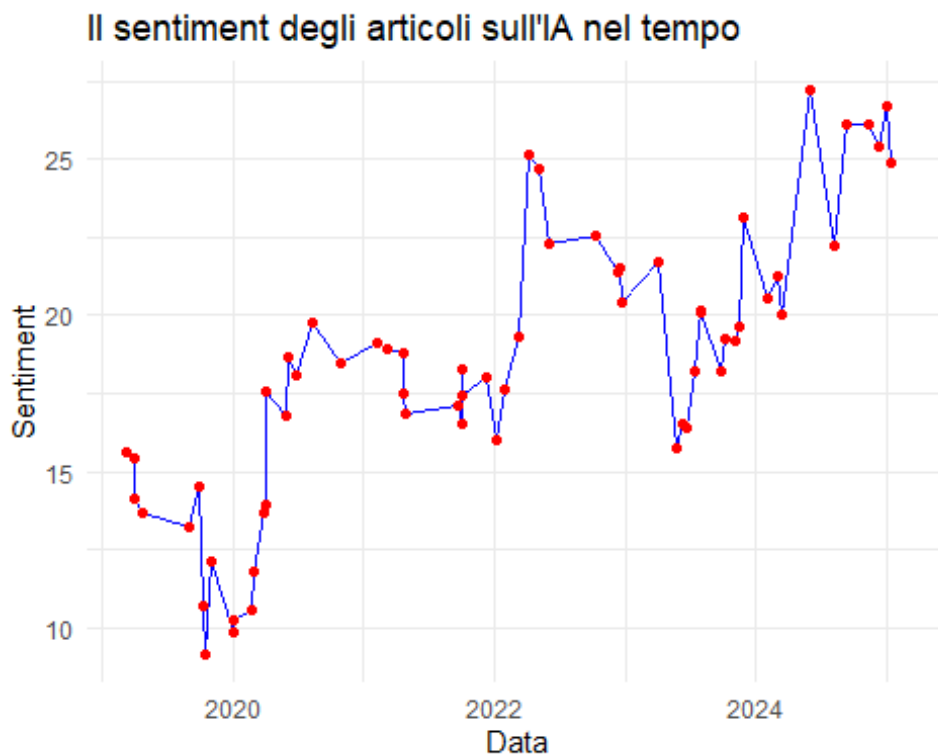
```

## Warning: Removed 7 rows containing missing values or values outside the scale range

## (`geom\_line()`).

## Warning: Removed 7 rows containing missing values or values outside the scale range

## (`geom\_point()`).



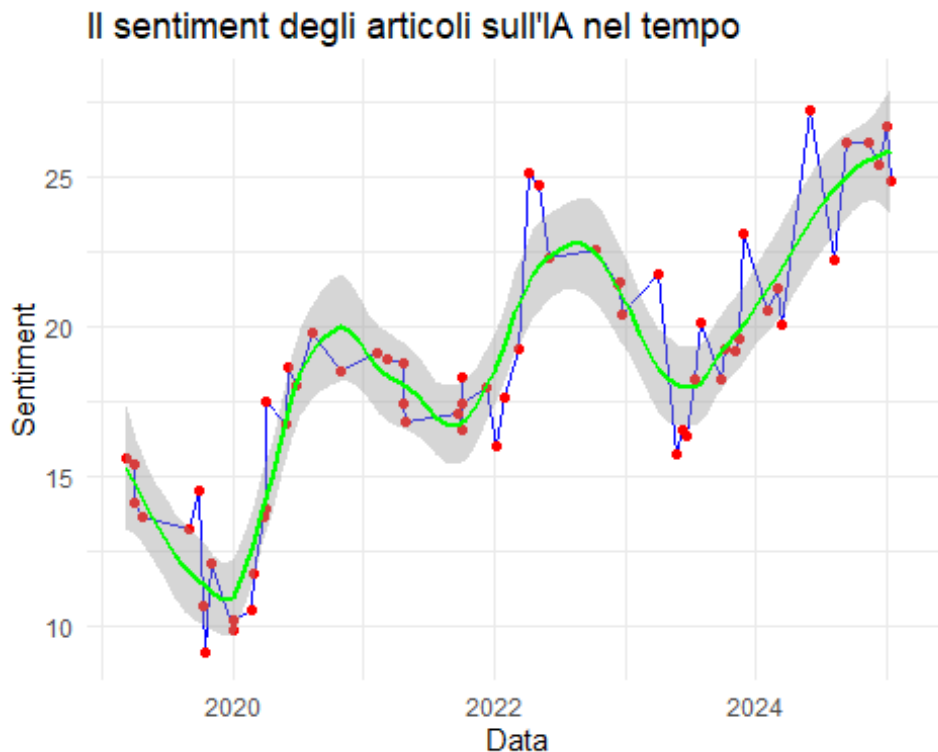
*#poichè il risultato non appare ancora chiaramente interpretabili si utilizza il metodo Loess per far sì che si abbia un andamento più chiaro.*

```

ggplot(gruppo_articoli, aes(x = data, y = sentiment_ma)) +
  geom_line(color = "blue") + # Linea
  geom_point(color = "red") + # Punti
  geom_smooth(method = "loess", color = "green", span = 0.3) + # Aggiungi
smoothing LOESS
  labs(
    title = "Il sentiment degli articoli sull'IA nel tempo",
    x = "Data",
    y = "Sentiment"
  ) +
theme_minimal()

```

## `geom\_smooth()` using formula = 'y ~ x'



L'andamento del sentiment dei testi degli articoli è estremamente variabile nel tempo. Ciò può dipendere da due principali fattori: In primis, si tratta di un tema sul quale si è dibattuto a partire dalla sua introduzione e sul quale ancora coesistono visioni ed opinioni fra loro contrastanti. Nell'analisi effettuata, le ampie fluttuazioni del sentiment presenti tra articoli temporalmente ravvicinati può essere ascritta alla differenza di vedute dei diversi autori. In secondo luogo, l'intelligenza artificiale, specialmente riguardo il suo utilizzo nel mondo del lavoro, è una tecnologia in continua evoluzione e ciò la rende suscettibile, contemporaneamente, a previsioni sia fiduciose sia scettiche sugli effetti del suo sviluppo. In generale, da questo punto di vista sembra che vi sia un trend crescente nel tempo, seppur caratterizzato da ampie fluttuazioni. Il fatto che dagli articoli più recenti traspaia un sentiment mediamente più positivo rispetto ad articoli del passato può essere dovuto ad un maggiore livello di consapevolezza in materia.

### 4.3 Studio delle emozioni primarie emergenti dai testi degli articoli

In questa fase si utilizza il dizionario di sentiment NRC per far emergere il livello in cui le otto emozioni primarie sono presenti negli articoli.

```
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")
library("quanteda")
library(lubridate)
library(data.table)
```

```
##
## Caricamento pacchetto: 'data.table'

## I seguenti oggetti sono mascherati da 'package:zoo':
##
##     yearmon, yearqtr

## I seguenti oggetti sono mascherati da 'package:reshape2':
##
##     dcast, melt

## I seguenti oggetti sono mascherati da 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year

## I seguenti oggetti sono mascherati da 'package:dplyr':
##
##     between, first, last

## Il seguente oggetto è mascherato da 'package:purrr':
##
##     transpose

library(tm)

corpus<-corpus(gruppo_articoli$testo)
f<-get_nrc_sentiment(corpus)
dim(f)

## [1] 70 10

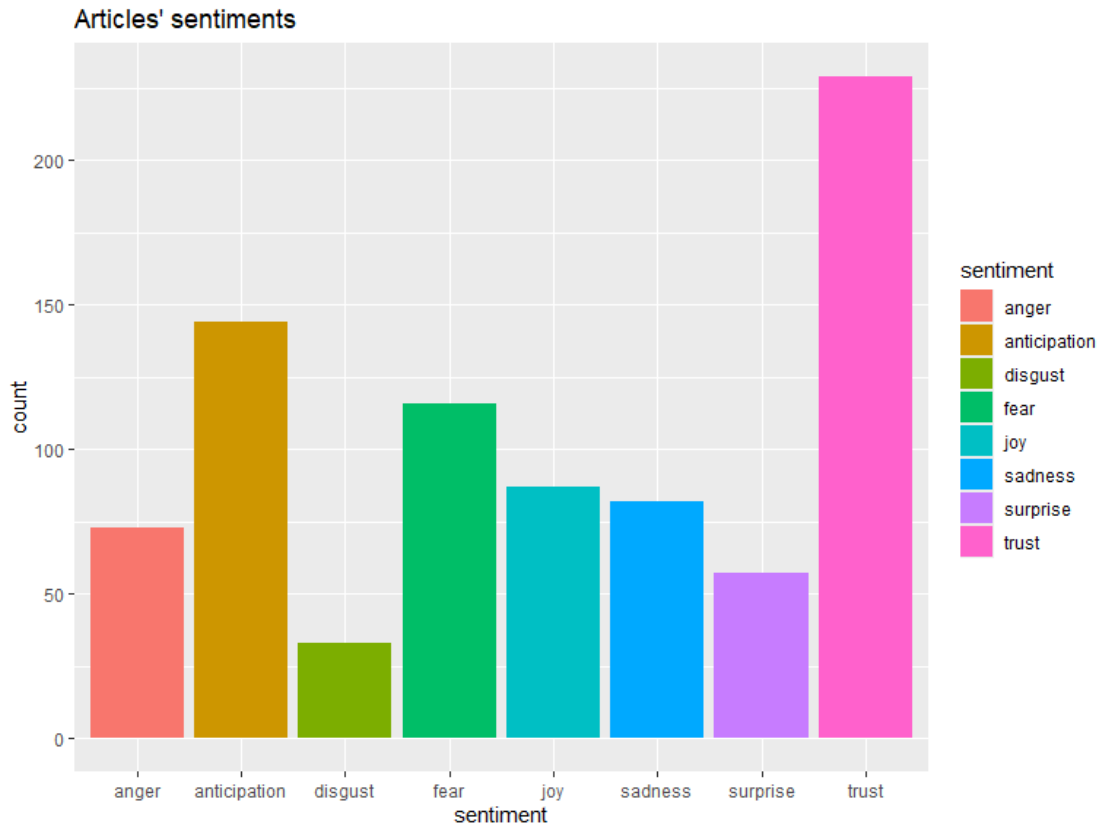
f1<-f[,1:8] #eliminazione colonne relative a emozioni "positivo" e "negativo"
dim(f1)

## [1] 70  8

td<-data.frame(t(f1)) #emozioni sulle righe, articoli sulle colonne
td_new<- data.frame(rowSums(td[1:8])) #somma valori emozioni sui documenti
#Transformation and cleaning
names(td_new)[1] <- "count"
td_new<- cbind("sentiment" = rownames(td_new),td_new)
rownames(td_new) <- NULL
# td_new2<-td_new[1:8,] se volessi escludere sentimenti negativi e positivi
#Plot One - count of words associated with each sentiment
quickplot(sentiment, data=td_new,
weight=count,geom="bar",fill=sentiment,ylab="count")+ggtitle("Articles'
sentiments")

## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```





L'emozione complessivamente più presente è la fiducia nei confronti delle nuove tecnologie ed in particolare dell'IA. Questo può essere ricollegato a quanto si diceva nel paragrafo 4.1 circa la percezione delle potenzialità dell'IA riguardo il miglioramento dei processi lavorativi e, dunque , delle performance aziendali. Molto rilevante è anche la presenza dell'anticipazione, che attesta come i risvolti dell'applicazione dell'IA non siano completamente definibili, questo può portare ad eccitazione ma anche ad ansia riguardo il futuro. La terza emozione maggiormente presente è la paura, che, in contrasto rispetto alla prima, indica la presenza di posizioni scettiche riguardo le nuove tecnologie. Infine, i livelli simili di presenza di gioia e tristezza sottolineano la diversità delle vedute riguardo il tema.