



# FellTrack

Analizzare il sentiment dei post sui diversi social media e  
confrontarli tra loro

Mattia Castiello 0512120487  
Paolillo valentina 0512114820

FeelTrack®

# Punti fondamentali

Sentiment  
analysis

Obbiettivi

Dati

Modelli di  
machine  
learning  
utilizzati

Analisi  
Esplorativa  
dei Dati

Conclusioni

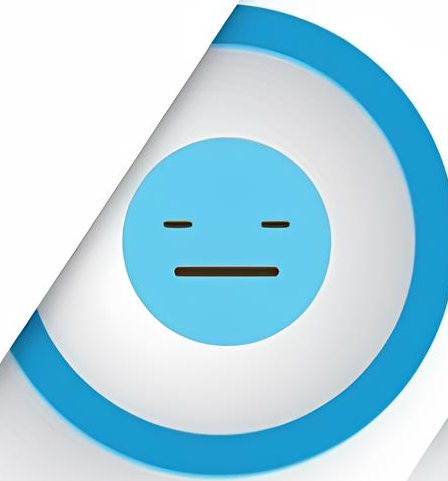
# SENTIMENT ANALYSIS

---



## NEGATIVE

Totally dissatisfied with the service. Worst customer care ever.



## NEUTRAL

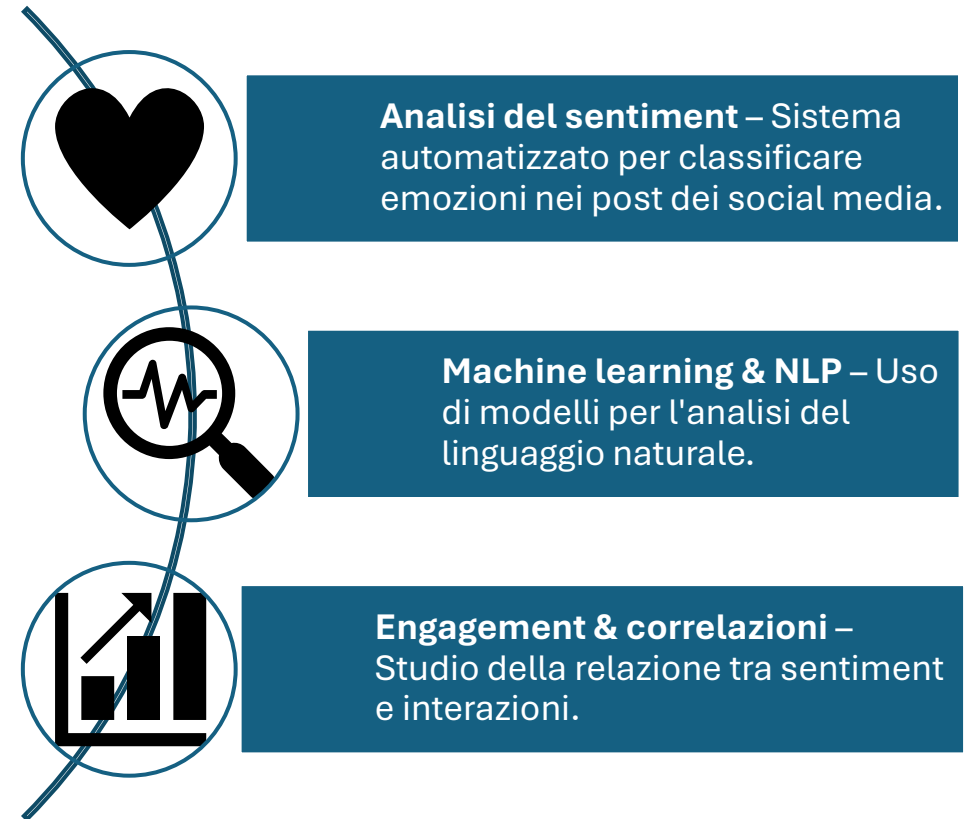
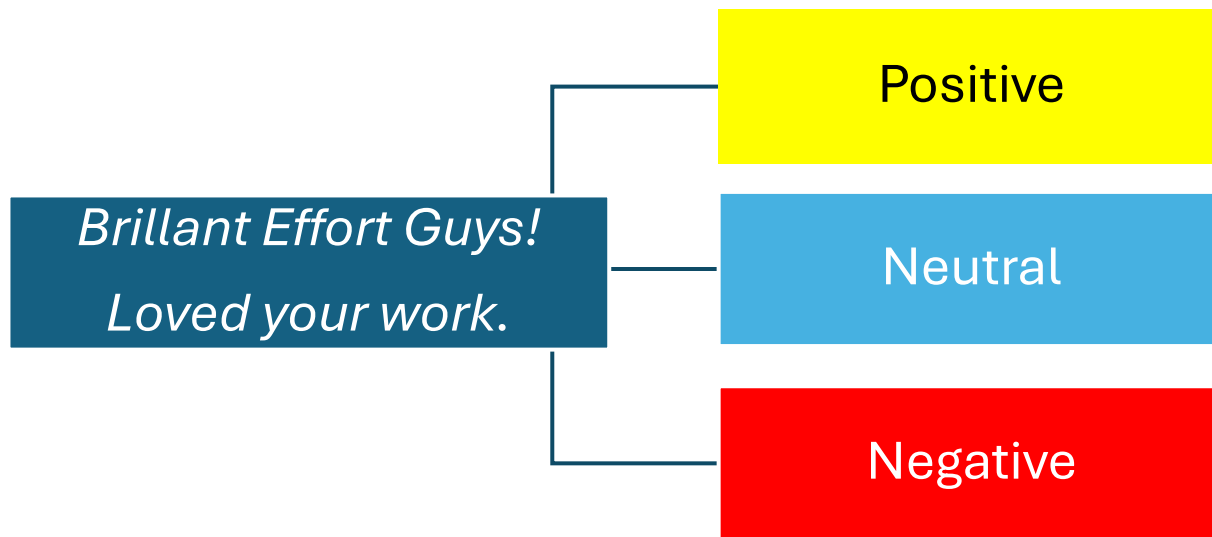
Good Job but I will expect a lot more in future.



## POSITIVE

Brilliant effort guys! Loved Your Work.

# Obiettivi



# Dati

I dataset

## **Social Media Sentiment Analysis: Test**

- User
- Platform
- Post
- Hashtag
- Sentiment
- Likes
- Shares
- Comment

**Test set**

**Training set**

## **Social Media Sentiment Analysis: Train**

- Id
- Body
- Sentiment Type

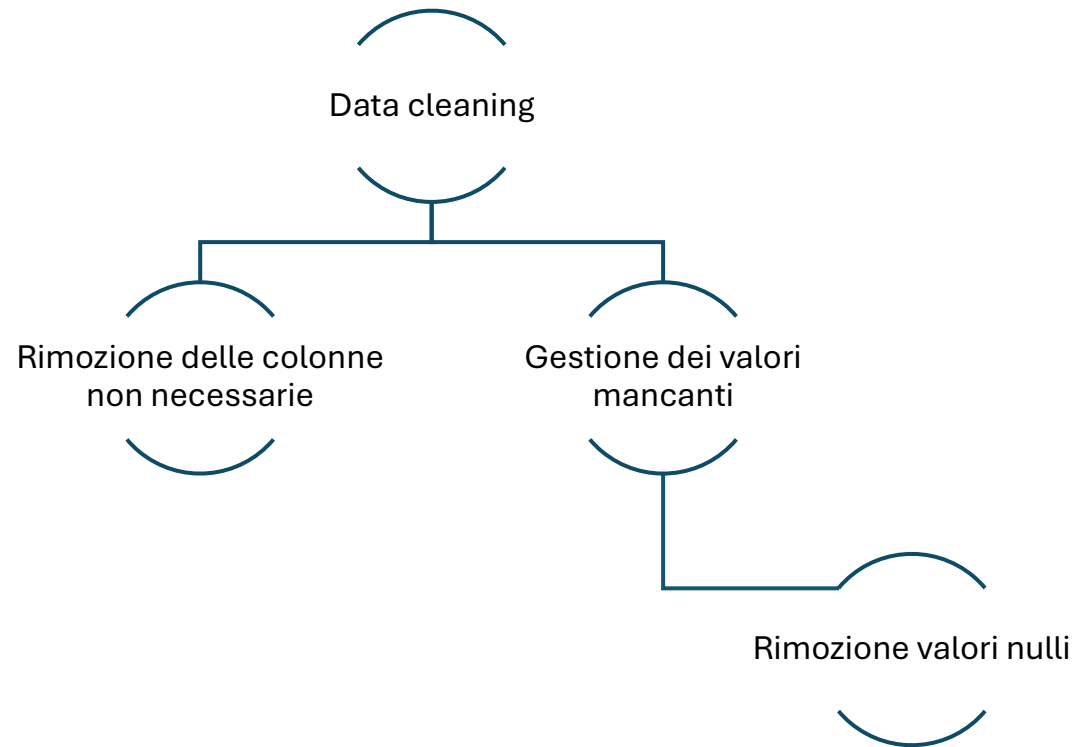
+

## **Twitter Sentiment Dataset**

- clean\_text
- category

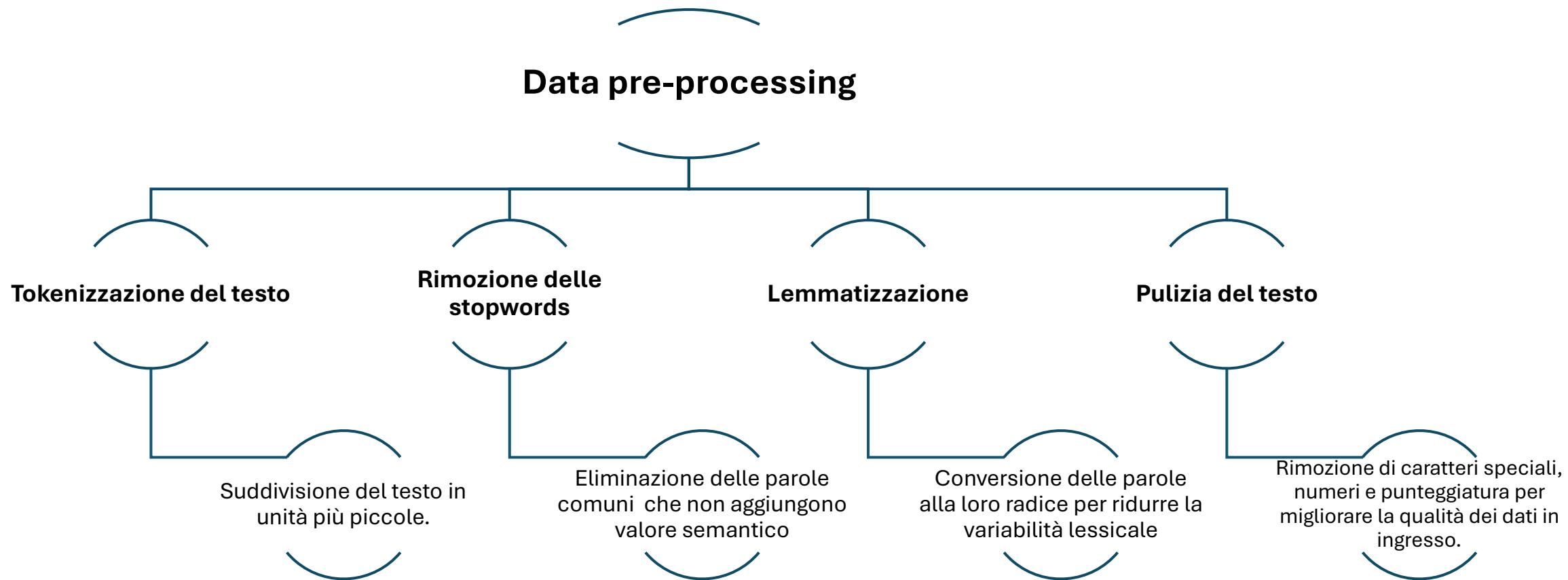
# Dati

## Data cleaning



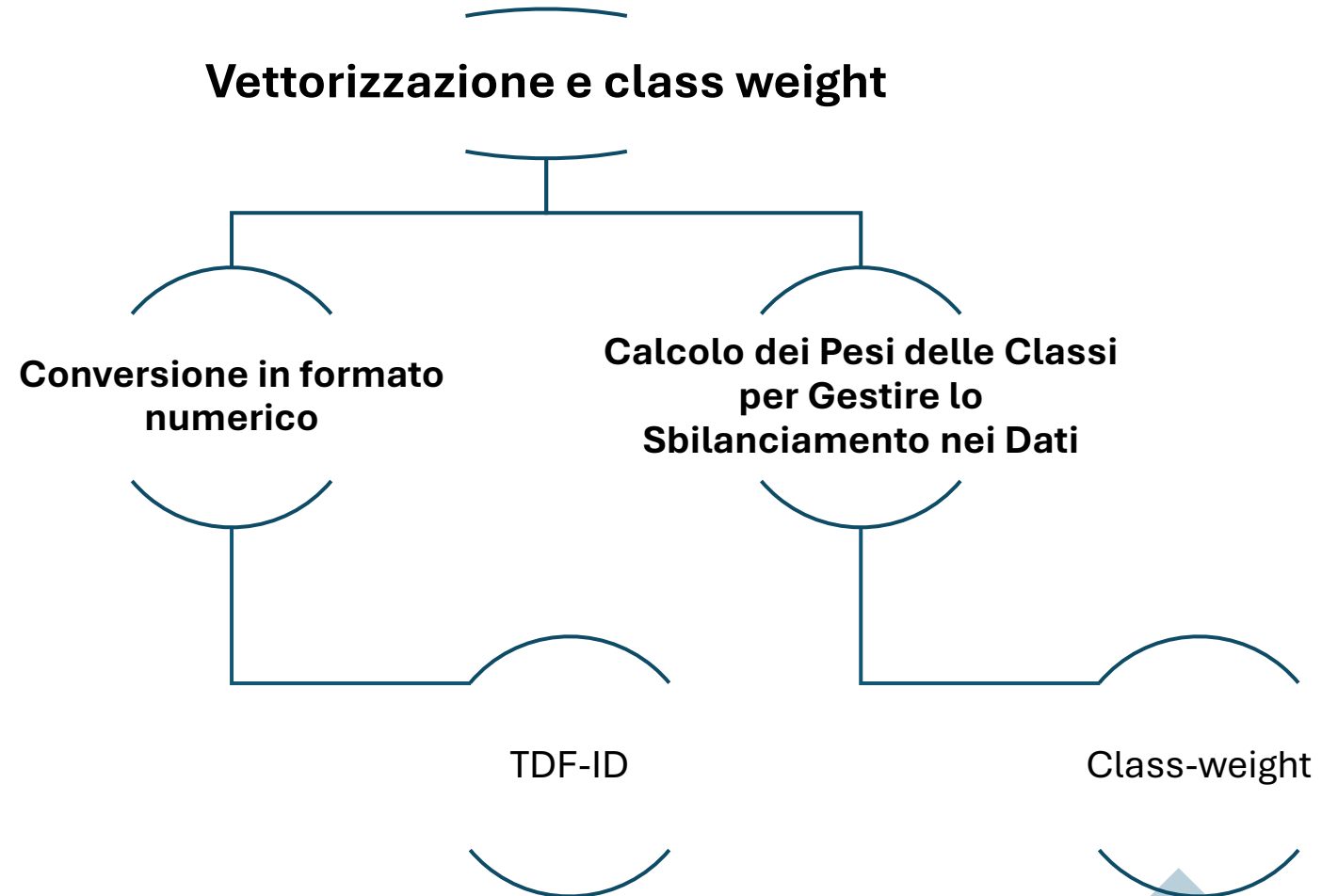
# Dati

## Data pre-processing



# Dati

Vettorizzazione e class weight





# Modelli di machine learning utilizzati



**Logistic  
Regression**

A 3D plot showing a non-linear decision boundary (a curved surface) separating two classes of data points (red and blue) in a 3D space.



**Random  
Forest  
Classifier**

A diagram illustrating a Random Forest Classifier, showing a tree structure with nodes and branches, representing the ensemble of decision trees.



**Linear SVC**

A 3D plot showing a linear decision boundary (a flat plane) separating two classes of data points (red and blue) in a 3D space. The plot is labeled with "SUPPORT VECTORS" and "LINEAR SEPARATION".

(Support Vector machines  
Classifier)

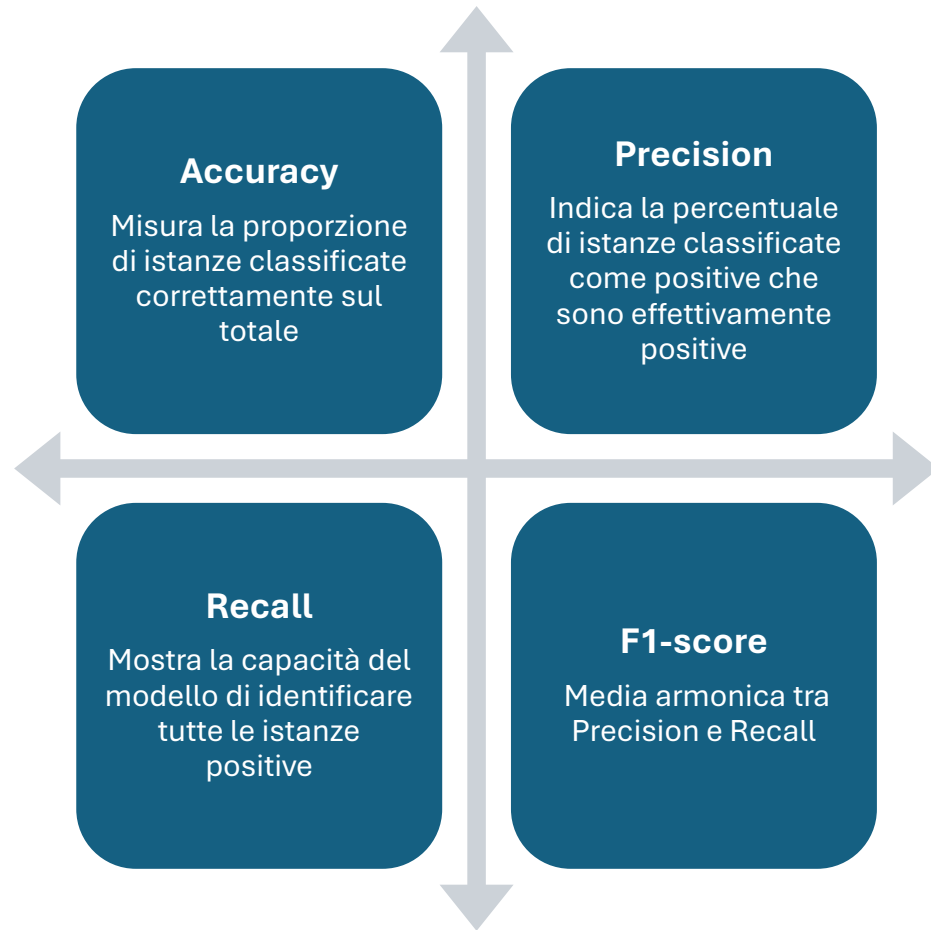


**MLPClassifier**

A diagram illustrating a Multilayer Perceptron (MLP) Classifier, showing a neural network structure with an Input Layer, Hidden Layer, and Output Layer.

# Modelli di machine learning utilizzati

Metriche di valutazione utilizzate



Matrice di confusione		
	Istanze realmente positive	Istanze realmente negative
Istanze predette come positive	Veri positivi	Falsi positivi
Istanze predette come negative	Falsi negativi	Veri negativi

# Modelli di machine learning utilizzati

## Logistic Regression

Assegna una **probabilità** a ciascuna classe

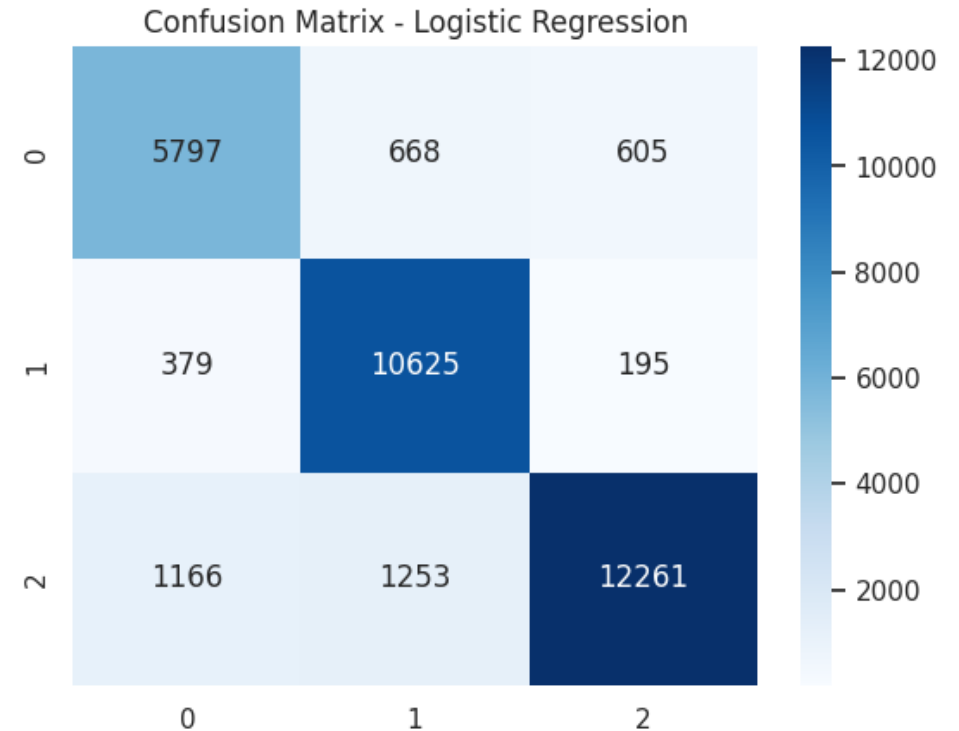


Garantisce che la somma delle probabilità sia **1**, permettendo di identificare la classe più probabile



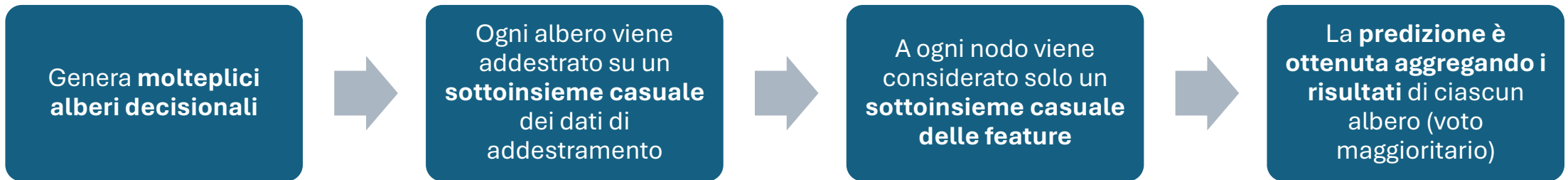
il modello sceglie la **classe con probabilità più alta**

	precision	recall	f1-score	support
negative	0.79	0.82	0.80	7070
neutral	0.85	0.95	0.89	11199
positive	0.94	0.84	0.88	14680
accuracy			0.87	32949
macro avg	0.86	0.87	0.86	32949
weighted avg	0.88	0.87	0.87	32949

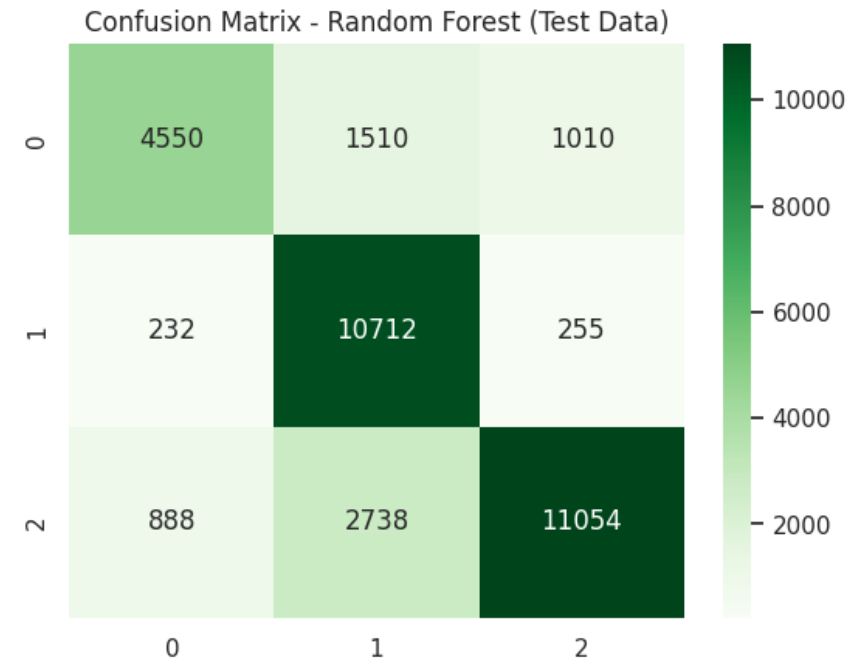


# Modelli di machine learning utilizzati

## Random Forest Classifier



	precision	recall	f1-score	support
negative	0.80	0.64	0.71	7070
neutral	0.72	0.96	0.82	11199
positive	0.90	0.75	0.82	14680
accuracy			0.80	32949
macro avg	0.81	0.78	0.78	32949
weighted avg	0.82	0.80	0.80	32949



# Modelli di machine learning utilizzati

## Linear SVC

Trova il **miglior iperpiano** per separare i dati in uno spazio **n-dimensionale**

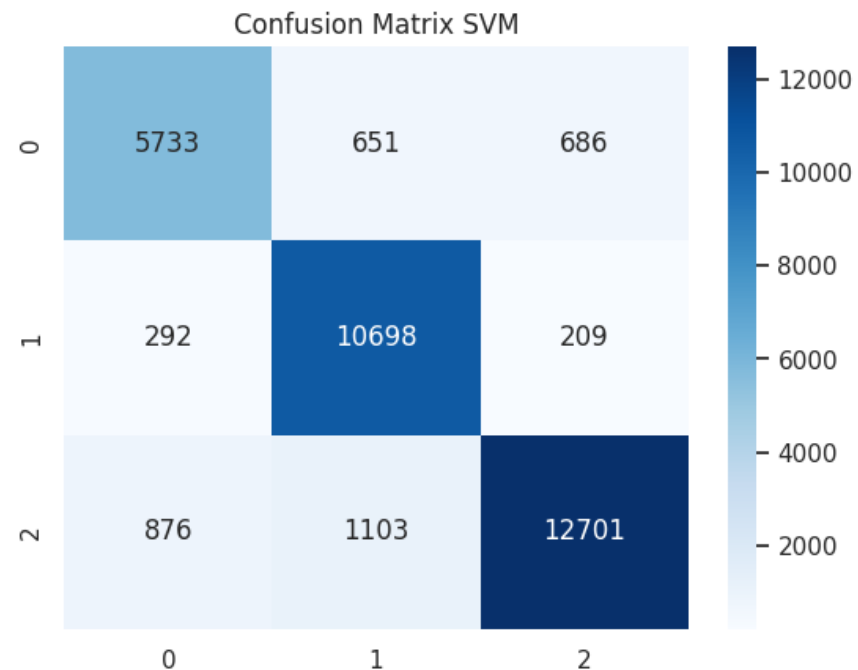


**Massimizzare il margine** tra le classi, ovvero la distanza tra l'iperpiano e i punti dati più vicini



I dati vengono classificati in base alla loro **posizione rispetto all'iperpiano**

	precision	recall	f1-score	support
negative	0.83	0.81	0.82	7070
neutral	0.86	0.96	0.90	11199
positive	0.93	0.87	0.90	14680
accuracy			0.88	32949
macro avg	0.87	0.88	0.87	32949
weighted avg	0.89	0.88	0.88	32949



# Modelli di machine learning utilizzati

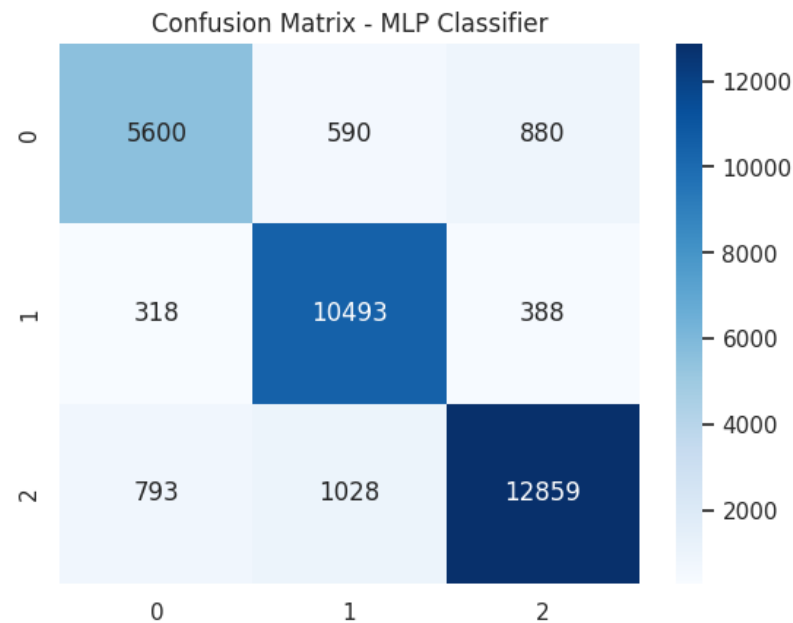
MLPClassifier - la rete neurale artificiale

Composto da **input layer**,  
**hidden layers** e **output layer**

Ogni neurone elabora i dati  
con una combinazione lineare  
tra gli input e i pesi

L'algoritmo **Backpropagation**  
aggiorna i pesi per migliorare  
le predizioni

	precision	recall	f1-score	support
negative	0.83	0.79	0.81	7070
neutral	0.87	0.94	0.90	11199
positive	0.91	0.88	0.89	14680
accuracy			0.88	32949
macro avg	0.87	0.87	0.87	32949
weighted avg	0.88	0.88	0.88	32949

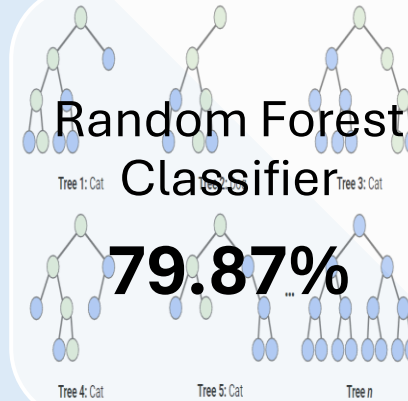


# Modelli di machine learning utilizzati

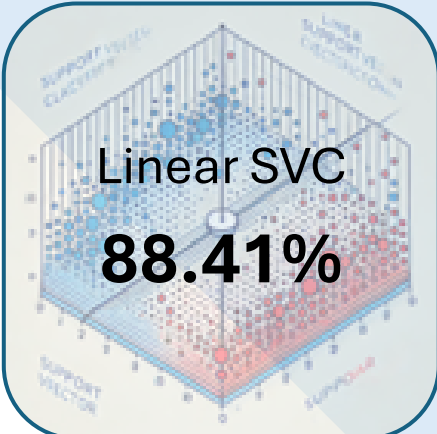
Performance – Accuracy



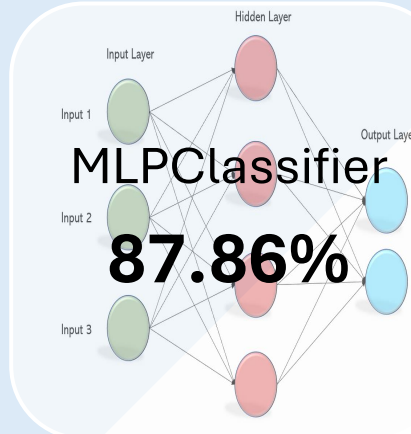
Logistic  
Regression  
**87.05%**



Random Forest  
Classifier  
**79.87%**

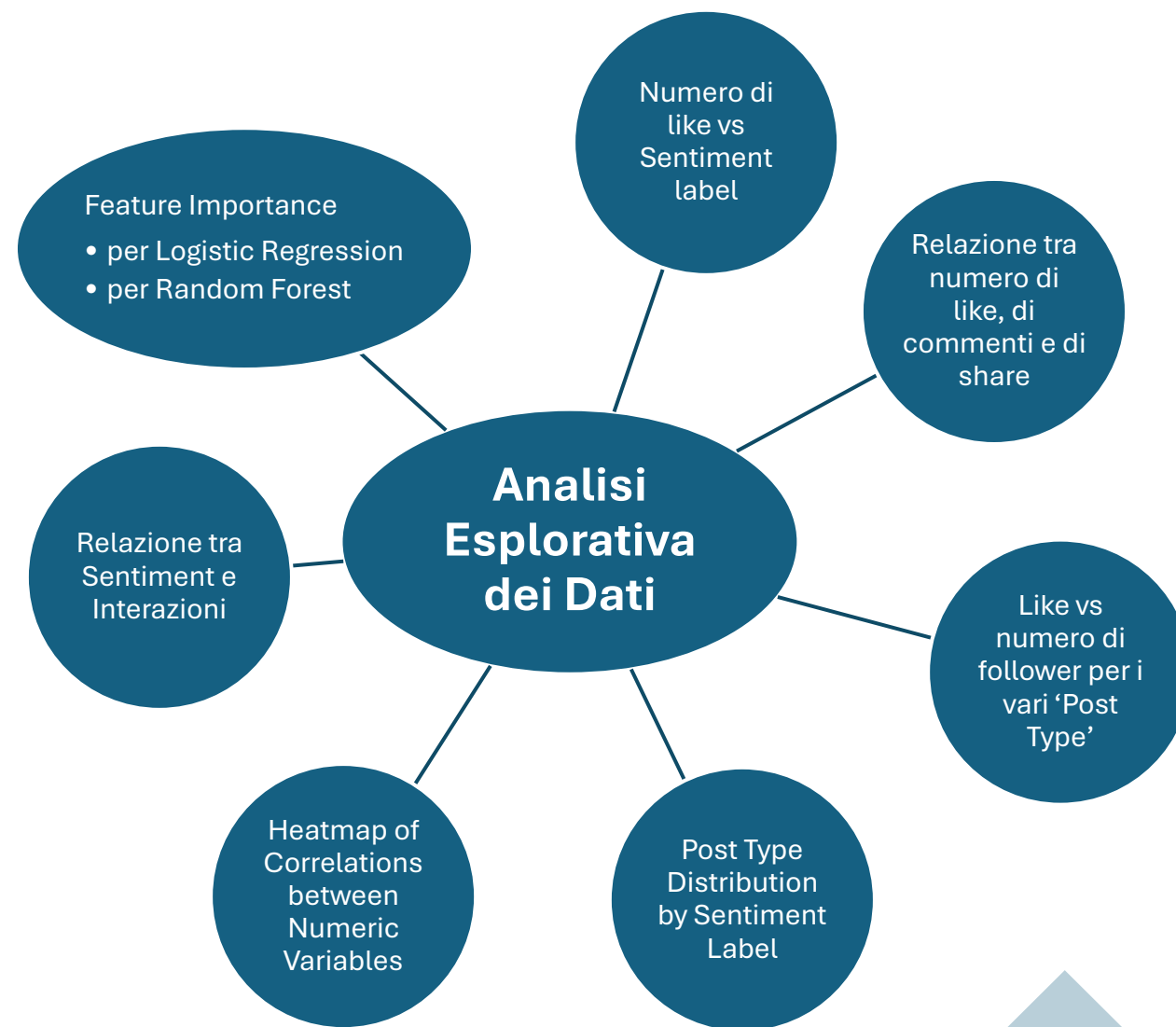


Linear SVC  
**88.41%**



MLPClassifier  
**87.86%**

# Analisi Esplorativa dei Dati

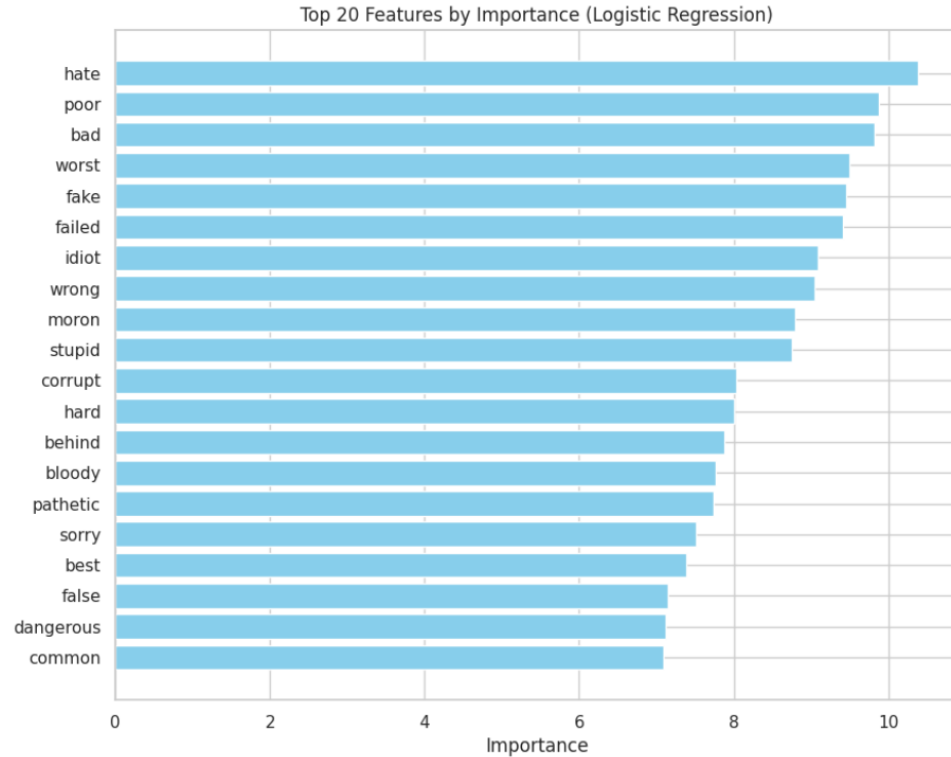




# Analisi Esplorativa dei Dati

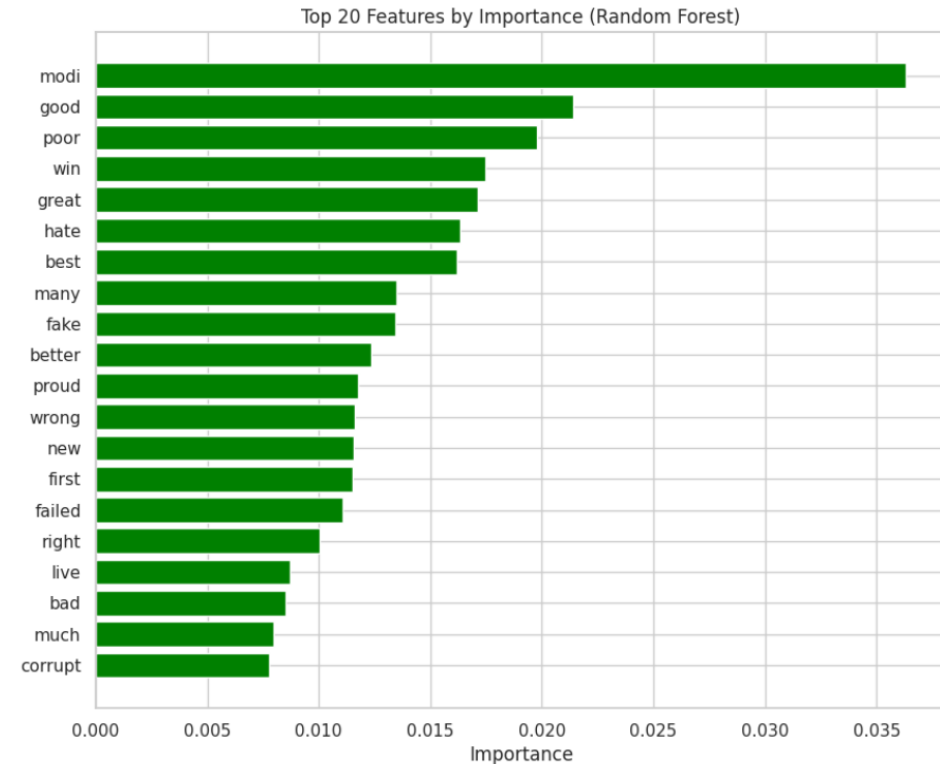
## Feature Importance

### Logistic regression



Hate, poor, bad...

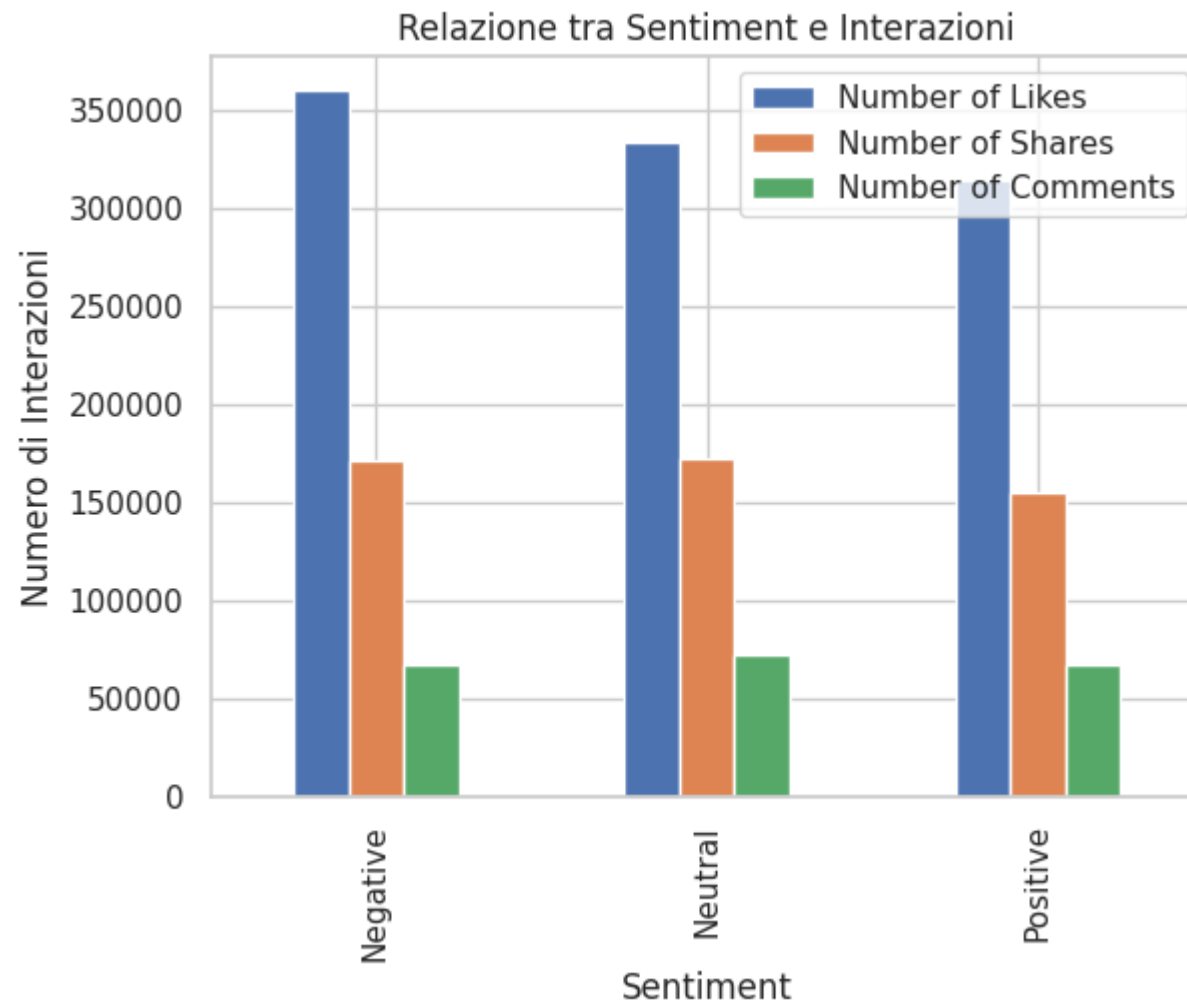
### Random Forest



Modi, good, poor...

# Analisi Esplorativa dei Dati

Relazione tra sentiment e Interazioni



# Conclusioni

## Efficacia dell'analisi del sentiment

FeelTrack ha dimostrato la validità del machine learning nell'analisi del sentiment sui social media

## Performance dei modelli

SVC Linear, MLP Classifier e Logistic regression risultano equiparabili in termini di accuratezza e più performanti di Random Forest Classifier

## Correlazione sentiment-engagement

Forte relazione tra emozioni nei post e interazioni degli utenti

## Feature importance

Identificazione dei termini chiave che influenzano maggiormente la classificazione del sentiment

## Applicazioni pratiche

FeelTrack fornisce insight utili per aziende, ricercatori e analisti nello studio del comportamento e delle dinamiche di engagement sui social media