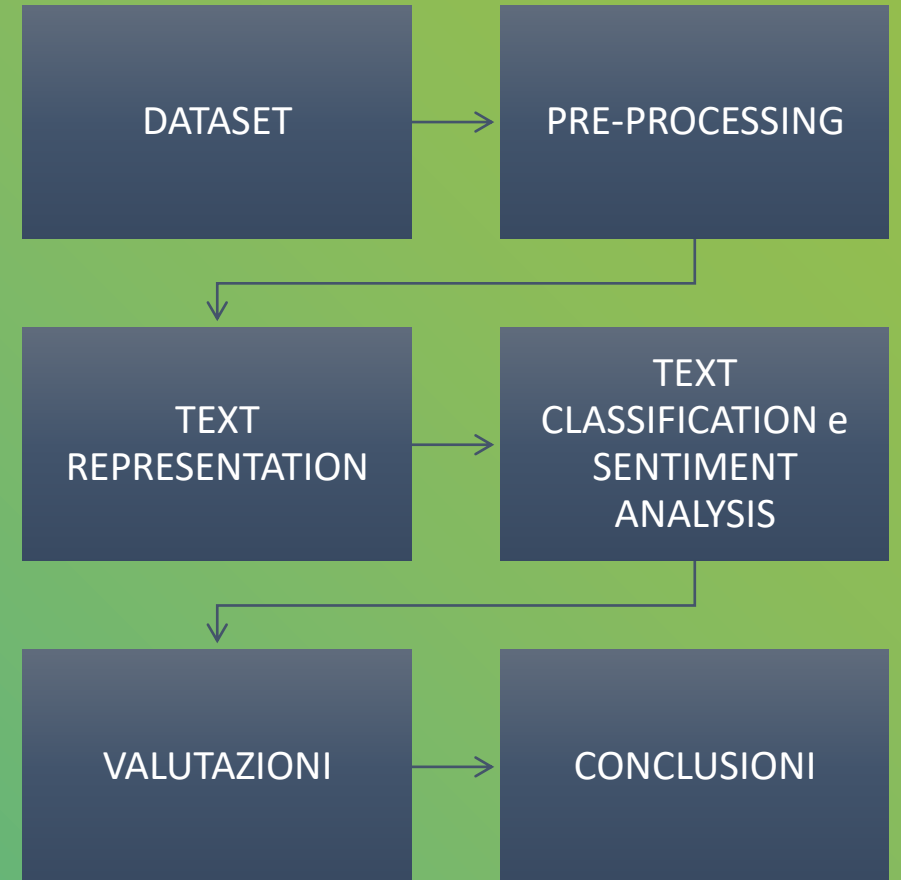


Sentiment Analysis e Text Classification delle recensioni dei farmaci


- Della Mura Dario, matricola 793751
- Doci David, matricola 799647
- Filip Sara, matricola 852864

Pipeline

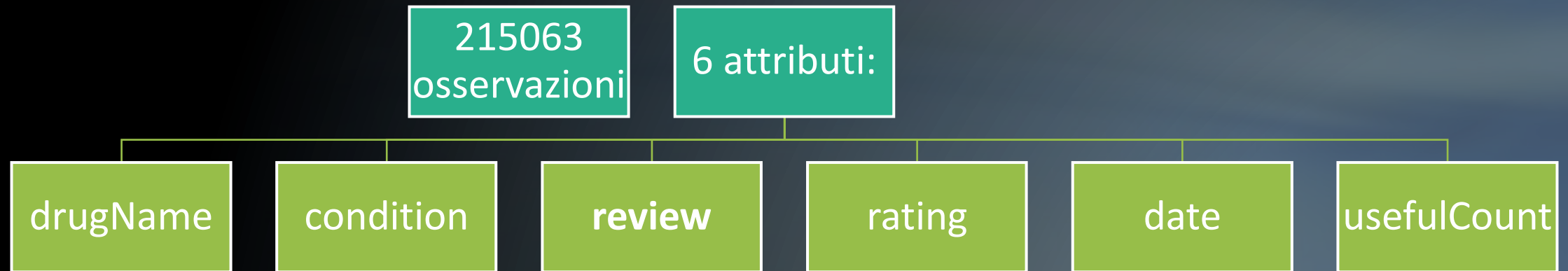




Obiettivi

- Eseguire una Text Classification per classificare le recensioni in positive o negative, confrontando diversi approcci metodologici
 - Eseguire la Sentiment Analysis sulle recensioni
 - Eseguire una classificazione usando un nuovo algoritmo di Google per classificare recensioni positive, negative, neutre
- 

Dataset



Pre-processing

Tokenizzazione

- Creazione di token

Normalizzazione

- Trasformazione in lettere minuscole
- Rimozione spazi bianchi
- Rimozione caratteri speciali

Stemming

- Snowball

Rimozione stopwords

- Rimozione di parole comuni e non comuni con poco significato semantico

Text Representation

+

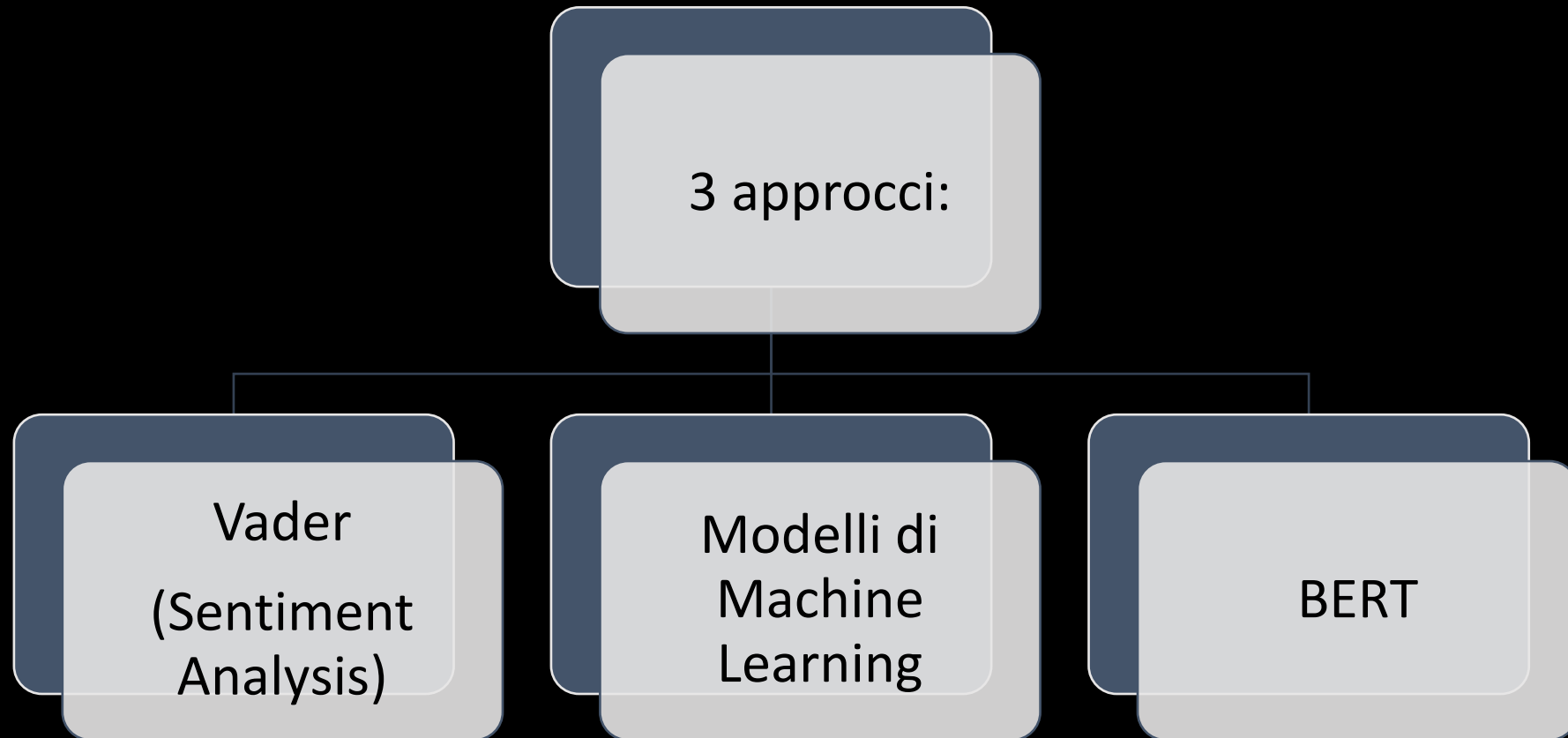
•

○

Modello Bag-of-Word

Matrice TF-IDF

Text Classification - Sentiment Analysis





Vader

- basato sul lessico
 - coglie sia la polarità che l'intensità del sentimento
 - punteggio composto (*compound*)
-

Valutazione

- Vader

ACCURACY	0.63
----------	------

Modelli di Machine Learning

Modelli:

Support Vector
Machine (SVM)

Regressione
logistica

Random Forest

Decision Tree

XGBoost

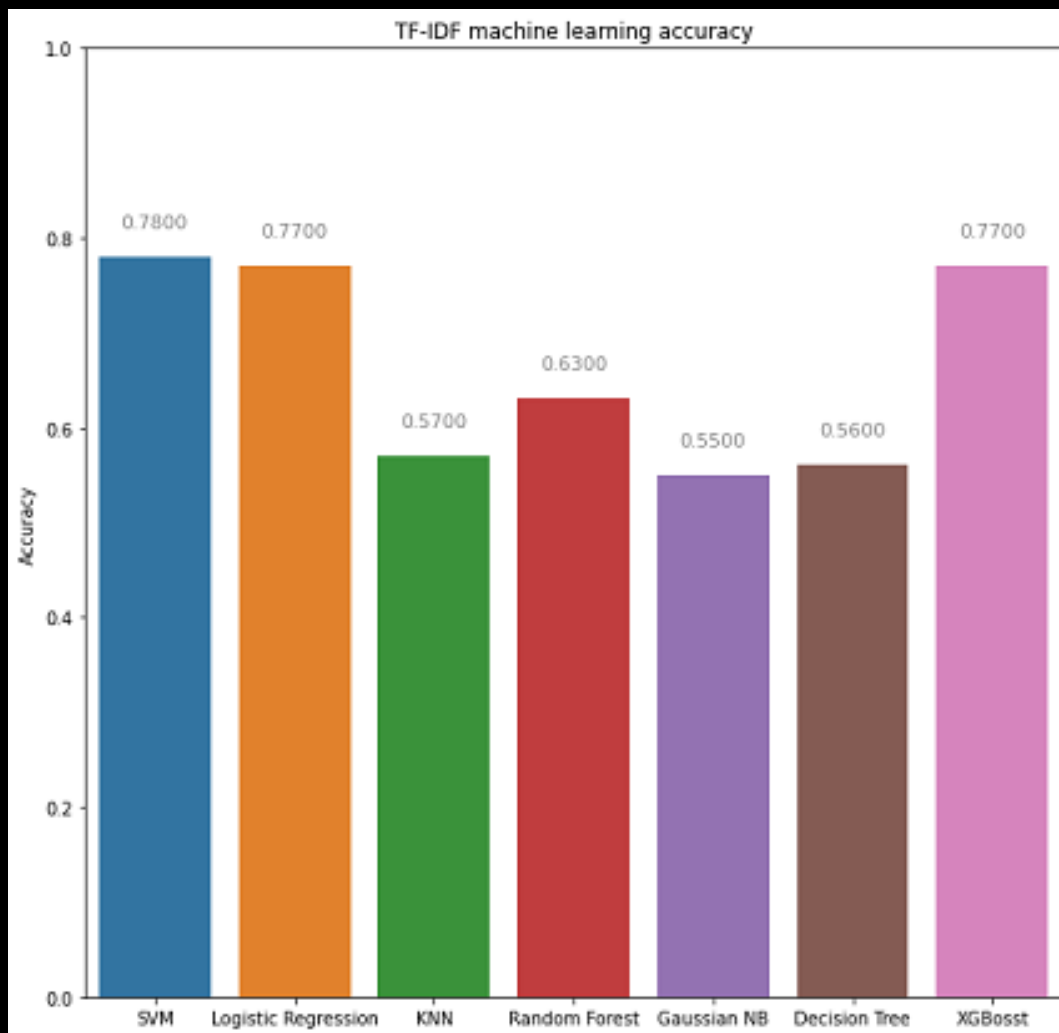
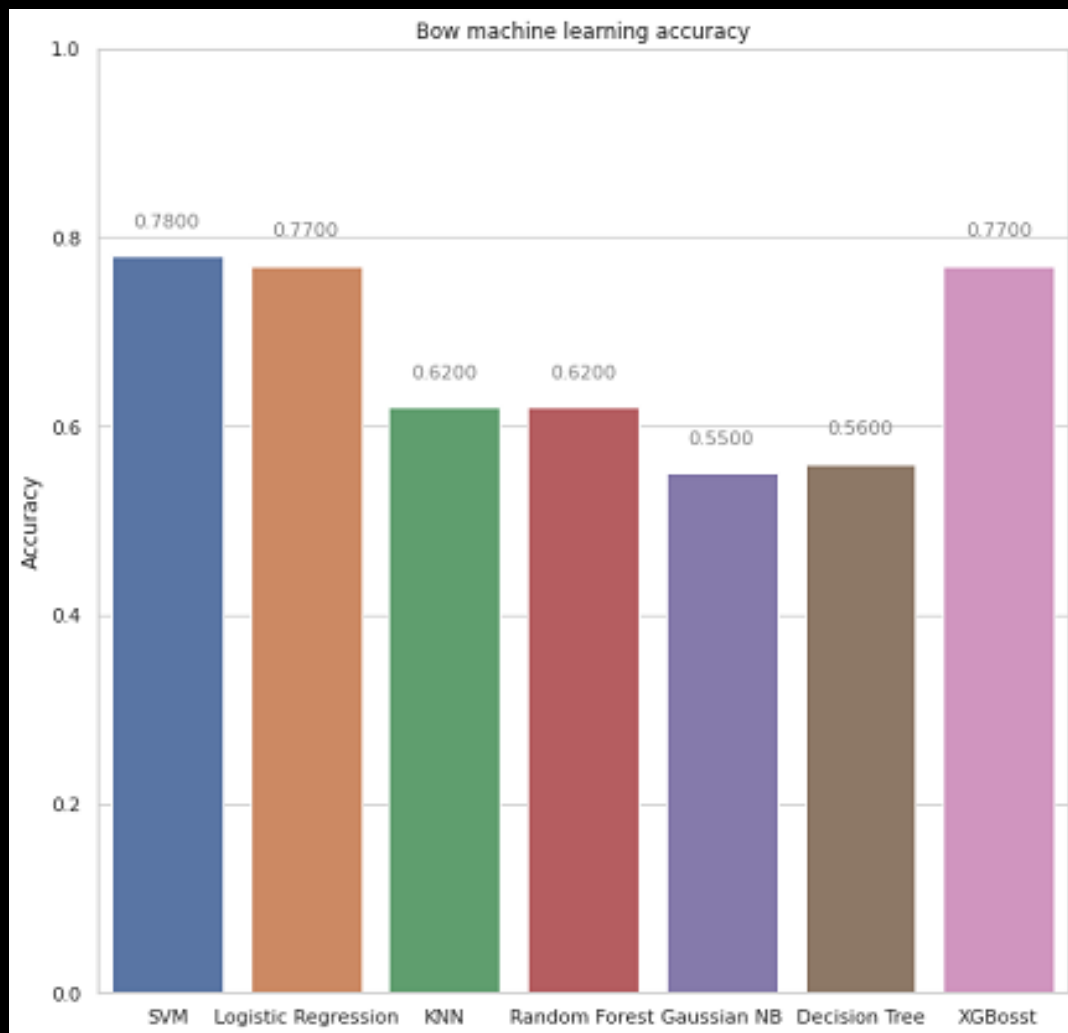
Guassian Naive
Bayes

KNeighbors

Tecniche di riduzione della dimensionalità:

- kBest
- PCA

Valutazioni modelli ML





BERT

- Utilizza Transformer
- È uno stack di encoders
- Crea rappresentazioni della parola
- Prevede parole mascherate in una frase


Valutazioni

BERT

	Precision	Recall	F1 score	Support
Positive	0.82	0.74	0.78	265
Neutral	0.71	0.67	0.69	265
Negative	0.59	0.68	0.63	265
Accuracy			0.70	795
Macro avg	0.71	0.70	0.70	795
Weighted avg	0.71	0.70	0.70	795



Conclusioni

- SVM è risultato il miglior modello di ML indipendentemente dalla *text representation* utilizzata
 - Non risulta corretto il confronto tra i tre approcci poiché si basano su fondamenti teorici diversi
 - Con tutti gli approcci implementati si sono ottenuti risultati soddisfacenti
- 

Grazie per l'attenzione!