

UNIVERSITA' DEGLI STUDI DI SALERNO

Corso di laurea in Informatica



# SENTIMENT ANALYSIS FOR FAKE NEWS

**Relatore:**

Prof.ssa Dajana Conte

**Candidato:**

Veronica Marcantuono

Matricola 05121/05426

Anno Accademico

**2020/2021**

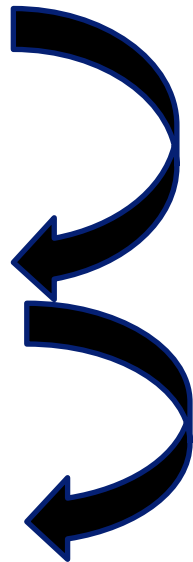
# Motivazioni

L'introduzione del mondo di internet ha fatto emergere delle problematiche legate alla diffusione delle fake news sul web:

**1** Disinformazione elevata;

**2** Diffusione rapida;

**3** Influenzamento del comportamento e delle opinioni di chi legge;



# Obiettivi



---

Sviluppo di **indici** sintattici per le fake news, tecniche semantiche con NLP e modelli di elaborazione del testo.



---

Sviluppo per la **rilevazione del sentimento** su testi per la lingua inglese ed italiana

# Indici sintattici

1

Numero di caratteri

2

Flesch: Per la leggibilità di un testo in inglese:

$$F = 206,835(84,6 * S) - (1,015 * P)$$

S=il numero medio di sillabe per parola P=il numero medio di parole per frase

3

Gunning Fog: Misura la facilità di lettura e la comprensione di un testo in inglese (tra 6 e 9) o (tra 14 e 17)

$$GI = 0.4 * \left[ \left( \frac{n^{\circ} \text{ parole}}{n^{\circ} \text{ frasi}} \right) + 100 \left( \frac{n^{\circ} \text{ parole complesse}}{n^{\circ} \text{ parole}} \right) \right]$$

# Semantica con NLP

La semantica con NLP(Natural Language Processing) si occupa di studiare il significato naturale della lingua.

PASSAGGIO CHIAVE:

PRE-ELABORAZIONE

NORMALIZZAZIONE



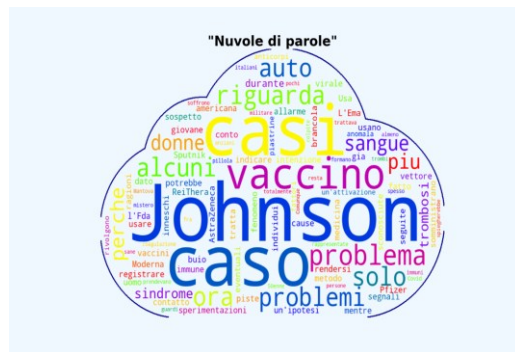
# Elaborazione del testo

Tecnica oppure insieme di programmi in grado di effettuare diverse azioni su testi e documenti, modifiche al testo, elementi grafici.

1

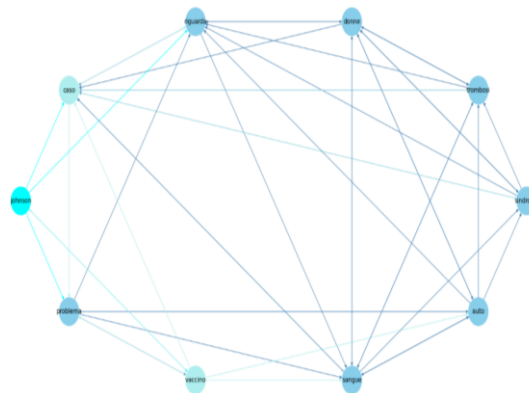
## WORDCLOUD

Modo ideale per estrarre le parti più pertinenti in un testo



2

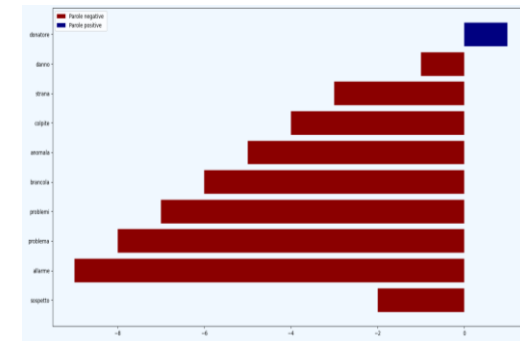
## OCCORRENZE & CO-OCCORRENZE



3

## POLARITA'

Punteggio assegnato agli aspetti positivi o negativi di un documento di testo



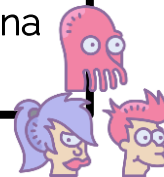
# Sentiment Analysis

1

La Sentiment Analysis o opinion mining, è un'analisi procedurale di calcolo dei sentimenti e delle opinioni espresse nei documenti testuali.

2

I soggetti che producono questi “giudizi” vengono definiti opinion holder- portatori di opinione e il loro commento può essere su una scala positiva, negativa e neutra



3

Teniche utilizzate per la rilevazione del sentimento:

- Lessico
- Machine e Deep Learning



Negative



Neutral



Positive

# Sentiment con TextBlob e VADER

TextBlob ad oggetti specifici attribuisce un punteggio di polarità (negativo/positivo, da -1 a 1) ed un punteggio di soggettività (oggettivo/soggettivo, da 0 a 1)

	titolo	testo	sentiment	titolo	sentiment	notizia
0	POPULAR CONSERVATIVE SENATOR JEFF SESSIONS Say...	Senator Jess Sessions tough love talk may be ...	positive		positive	
1	New U.S. helicopters mark major change for Afg...	KANDAHAR, Afghanistan (Reuters) - Senior Afgha...	negative		negative	
2	Senator Collins says not committed to tax bill...	WASHINGTON (Reuters) - Republican U.S. Senator...	negative		negative	
3	Disgusted Republicans Unable To Reach UNHINGE...	Donald Trump is already making the Republican ...	negative		negative	
4	New York's political leaders reach deal on 201...	NEW YORK (Reuters) - New York state's governor...	negative		positive	
...	...	...	...	...	...	...
9995	PRESIDENT TRUMP TO CUBA: Send Back Escaped COP...	The inspiration behind the Black Lives Matter...	negative		negative	
9996	Britain's U.S. ambassador discussed Trump retw...	LONDON (Reuters) - Britain's ambassador to the...	negative		negative	
9997	Miami hospitals prepare for surge in births du...	(Reuters) - While many Miami hospitals are shu...	negative		positive	
9998	DEATH PANELS? PRINCETON PROFESSOR WANTS TO KIL...	Hell has a special spot reserved for this horr...	negative		negative	
9999	Trexit? Online bettors play early end to Trump...	NEW YORK/LONDON (Reuters) - Bettors on both si...	positive		negative	

VADER Valence Aware Dictionari e sEntiment Reasoner, basata su regole

VALORI SOGLIA:

Positivo: punteggio composto  $\geq 0.0533$

Neutro:  $-0.05 < \text{punteggio composto} < 0.05$

Negativo: punteggio composto  $\leq -0.05$

	testo	neg	neu	pos	compound
	Senator Jess Sessions tough love talk may be ...	0.032	0.829	0.139	0.9943
	KANDAHAR, Afghanistan (Reuters) - Senior Afgha...	0.043	0.874	0.084	0.9332
	WASHINGTON (Reuters) - Republican U.S. Senator...	0.074	0.820	0.107	0.9245
	Donald Trump is already making the Republican ...	0.048	0.889	0.064	0.3394
	NEW YORK (Reuters) - New York state's governor...	0.017	0.844	0.139	0.8591
	...	...	...	...	...
	The inspiration behind the Black Lives Matter...	0.132	0.746	0.122	-0.8943
	LONDON (Reuters) - Britain's ambassador to the...	0.070	0.898	0.032	-0.5106
	(Reuters) - While many Miami hospitals are shu...	0.025	0.934	0.040	0.7543
	Hell has a special spot reserved for this horr...	0.089	0.832	0.079	-0.5412
	NEW YORK/LONDON (Reuters) - Bettors on both si...	0.019	0.968	0.013	-0.2838



# Sentiment con FEEL-IT e SENTITA

FEEL-IT è una libreria, per la classificazione del sentimento in italiano. Usa modelli di machine learning sulla classificazione cioè in output i sentimenti basati su positivo o negativo



	text	emozione	sentimento
Bill Gates ha predetto l'epidemia da Covid-19 ...		fear	negative
L'Ibuprofene peggiora la malattia da Covid-19\n		fear	negative
Pensioni ridotte del 50%\n		sadness	negative
La vitamina C favorisce la guarigione da Covid...		fear	positive
L'audio sulle teorie del complotto,\n		anger	negative
La foto delle numerose bare in fila a Bergamo,\n		sadness	negative
La delibera sullo stato d'emergenza tenuta nas...		anger	negative
Bere piu acqua per "spazzare via il virus",\n		anger	negative
Non sara una guerra nei prossimi decenni a cau...	text	fear	negative
L'unico vaccino che potra richiedere la licenz...	emozione	fear	negative

SENTITA Sentiment Italian un modello per estrarre l'analisi del sentimento per testi in lingua italiana. Due segnali in output compresi tra 0 ed 1, più vicina all'uno sentimento positivo più vicina allo zero sentimento negativo. Che possono essere attivati entrambi oppure soltanto uno.



bill gates ha predetto l'epidemia da covid-19 in un intervento del 2015,	- opos: 0.0006030798	- oneg: 0.0008339286
l'ibuprofene peggiora la malattia da covid-19	- opos: 0.061067373	- oneg: 0.54970276
pensioni ridotte del 50%,	- opos: 0.08978844	- oneg: 0.5020303
la vitamina c favorisce la guarigione da covid-19,	- opos: 0.09887245	- oneg: 0.042767406
l'audio sulle teorie del complotto,	- opos: 0.009593755	- oneg: 0.018769324
la foto delle numerose bare in fila a bergamo,	- opos: 0.018445253	- oneg: 0.0016320944
la delibera sullo stato d'emergenza tenuta nascosta,	- opos: 0.0058143735	- oneg: 0.0060275197
bere piu acqua per "spazzare via il virus",	- opos: 0.035314113	- oneg: 0.92394483
stipendi ridotti per le alte cariche dello stato a causa dell'emergenza,		

# Dataset con i dati ottenuti

1

RIFERIMENTO: dataset isot fake news detection

NEWS: 21.417 Real-News 23.481 Fake-News

ETICHETTE: label (true/false), titolo, testo, soggetto

[Fake News Detection Datasets - University of Victoria \(uvic.ca\)](https://uvic.ca/fake-news-detection-datasets)



2

COSTRUZIONE DATASET: 50% Real-News 50% Fake-news

ETICHETTE: label(true/false), titolo, testo, sentimento titolo e notizia, complessità

label	titolo	testo	sentimento titolo	sentimento notizia	complessità
False	POPULAR CONSERVATIVE SENATOR JEFF SESSIONS Say...	Senator Jess Sessions tough love talk may be ...	positive	positive	>700
True	New U.S. helicopters mark major change for Afg...	KANDAHAR, Afghanistan (Reuters) - Senior Afgha...	negative	negative	>700
True	Senator Collins says not committed to tax bill...	WASHINGTON (Reuters) - Republican U.S. Senator...	negative	negative	>700
False	Disgusted Republicans Unable To Reach UNHINGE...	Donald Trump is already making the Republican ...	negative	negative	>700
True	New York's political leaders reach deal on 201...	NEW YORK (Reuters) - New York state's governor...	negative	positive	>=300 e <=500
...	...	...	...	...	...
False	PRESIDENT TRUMP TO CUBA: Send Back Escaped COP...	The inspiration behind the Black Lives Matter...	negative	negative	>700
True	Britain's U.S. ambassador discussed Trump retw...	LONDON (Reuters) - Britain's ambassador to the...	negative	negative	>=500 e <=700
True	Miami hospitals prepare for surge in births du...	(Reuters) - While many Miami hospitals are shu...	negative	positive	>700
False	DEATH PANELS? PRINCETON PROFESSOR WANTS TO KIL...	Hell has a special spot reserved for this horr...	negative	negative	>700
True	Trexit? Online bettors play early end to Trump...	NEW YORK/LONDON (Reuters) - Bettors on both si...	positive	negative	>700

# Conclusioni



Realizzazione  
automatica leggibilità  
del testo ed  
individuazione fake  
news



Ricerca entità  
principale in un testo  
con tecniche di  
elaborazione del testo



Rilevazione del  
sentimento per la  
lingua italiana ed  
inglese

# Grazie per l'attenzione



Negative



Neutral



Positive