



**Università degli Studi di Camerino**

---

**SCUOLA DI SCIENZE E TECNOLOGIE**

**Corso di Laurea in Informatica (Classe L-31)**

**Cyberbullismo ed Intelligenza Artificiale:  
uno studio basato sul Natural Language  
Processing**

Laureando  
**Matteo Caggiano**

**Matricola 114945**

Relatore  
**Barbara Re**

Correlatore  
**Ivan Compagnucci**

---

A.A. 2022/2023



# Abstract

Il cyberbullismo, nell'era della digitalizzazione e della connettività globale, rappresenta un fenomeno sempre più crescente e preoccupante. A differenza del bullismo tradizionale, che si manifesta attraverso contatti fisici o verbali, il cyberbullismo ha la capacità di raggiungere le vittime in ogni luogo e momento della giornata. Questa caratteristica amplifica notevolmente le potenziali ripercussioni negative sulle vittime, le quali possono sentirsi minacciate in maniera continua, anche quando si trovano in un ambiente che dovrebbe essere percepito come sicuro.

Questa tesi si focalizza principalmente sullo studio del Natural Language Processing (NLP) con l'obiettivo di sviluppare una libreria dedicata alla rilevazione e prevenzione del cyberbullismo. L'interesse verso l'NLP nasce dalla sua peculiare capacità di analizzare e interpretare il linguaggio umano, offrendo così la possibilità di identificare pattern linguistici che potrebbero segnalare episodi di bullismo online.

L'analisi è stata condotta utilizzando il linguaggio di programmazione Python, noto per essere particolarmente adatto alla gestione e all'elaborazione di grandi volumi di dati. Nel corso del mio studio, ho esaminato diverse librerie di NLP, mettendo in luce le loro caratteristiche principali, dalla tokenizzazione alla rilevazione del sentimento. Questa indagine mi ha fornito le basi necessarie per sviluppare una libreria su misura, appositamente progettata per le esigenze di questo studio.

I risultati ottenuti evidenziano che, benché non esista una soluzione unica e definitiva, la combinazione di tecnologia e comprensione del linguaggio umano costituisce un'efficace risposta alla problematica del cyberbullismo. Implementando tali soluzioni su piattaforme digitali e social, si possono non solo rilevare in tempo reale comportamenti inappropriati, ma anche educare gli utenti riguardo l'importanza e le potenziali ripercussioni delle loro parole.



# Indice

<b>Abstract</b>	<b>3</b>
<b>1 Introduzione</b>	<b>7</b>
1.1 Scopo e obiettivo della tesi . . . . .	7
1.2 Struttura della tesi . . . . .	8
<b>2 Bullismo e Cyberbullismo</b>	<b>9</b>
2.1 L'Emergenza del bullismo . . . . .	9
2.2 Sviluppo tecnologico e nascita del Cyberbullismo . . . . .	10
2.3 Impatti del Cyberbullismo nell'era digitale . . . . .	10
<b>3 Natural Language Processing (NLP)</b>	<b>11</b>
3.1 Concetto di NLP . . . . .	11
3.2 Il ruolo cruciale del NLP nella comprensione del linguaggio umano . . .	12
3.3 Approcci NLP per affrontare il fenomeno del Cyberbullismo . . . . .	12
3.4 Tecniche avanzate di analisi del linguaggio impiegate . . . . .	13
<b>4 Tecnologie Utilizzate</b>	<b>15</b>
4.1 L'Incorporazione delle tecnologie nell'analisi del Cyberbullismo . . . . .	15
4.2 Il ruolo chiave di Python nell'analisi dei dati . . . . .	16
4.3 Librerie e pacchetti essenziali utilizzati per la ricerca . . . . .	17
<b>5 Analisi, Confronto e Scelta della Libreria Ottimale</b>	<b>19</b>
5.1 Raccolta e Pre-elaborazione dei dati . . . . .	19
5.2 Metodologie di analisi basate su diverse librerie di NLP . . . . .	20
5.3 Analisi critica delle prestazioni . . . . .	21
5.4 Vantaggi e svantaggi delle librerie scelte . . . . .	25
5.5 Riflessioni sull'accuratezza e l'efficienza . . . . .	26
<b>6 Contributi e Accessibilità</b>	<b>29</b>
6.1 Come la tecnologia può fare la differenza nel contrastare il Cyberbullismo	29
6.2 Condivisione su piattaforme facilmente accessibili . . . . .	30
<b>7 Conclusioni</b>	<b>31</b>
7.1 Sintesi dei principali risultati . . . . .	31
7.2 Potenziali sviluppi futuri . . . . .	32



# 1. Introduzione

Viviamo in un'epoca in cui la tecnologia influisce su ogni aspetto della vita quotidiana, in particolare, Internet ha trasformato il modo in cui interagiamo, lavoriamo, apprendiamo ma soprattutto comunichiamo. Tuttavia, con l'evoluzione di queste nuove modalità di interazione, emergono sfide e problematiche, una delle più pressanti è il cyberbullismo, che è attualmente in continuo sviluppo nel contesto digitale.

La finalità principale della mia tesi è indagare le potenzialità dell'intelligenza artificiale, e in particolare del Natural Language Processing (NLP), come strumento efficace nella rilevazione e nella lotta contro il cyberbullismo. Ma non mi fermo alla pura analisi teorica: mi impegno attivamente nella creazione di uno strumento utile, una specifica libreria software, progettata per intervenire concretamente sul fenomeno e offrire soluzioni pratiche per affrontarlo.

## 1.1 Scopo e obiettivo della tesi

Lo scopo di questa tesi, come anticipato nelle sezioni precedenti, è quello di sviluppare, attraverso l'integrazioni di librerie NLP, uno strumento accessibile a tutti, che permetta di realizzare una rapida e attenta analisi su testi, commenti, o messaggi scambiati nei social network. Studiare e confrontare i diversi pacchetti disponibili diventa cruciale, soprattutto quando si ha la necessità di identificare quali strumenti forniscono risultati più accurati e affidabili. Investigare su come l'intelligenza artificiale, analizzando e riconoscendo pattern linguistici che indicano comportamenti di bullismo online, possa fornire una risposta efficace al problema.

Parallelamente, si intende esplorare in profondità il fenomeno del cyberbullismo, affrontando una panoramica che va dalle sue origini storiche e sociologiche alla sua diffusione nell'era digitale. Si analizzerà il modo in cui questo si manifesta, tenendo conto delle diverse piattaforme online e dei cambiamenti nel comportamento umano dovuti all'anonimato e alla distanza che il mondo digitale può offrire e di conseguenza comprendendo anche i potenziali problemi psicologici, emotivi e sociali per chi ne è vittima, mettendo in luce la gravità del problema e l'urgenza di trovare soluzioni efficaci.

L'obiettivo primario è quindi andare oltre una semplice analisi teorica: si punta a fornire una soluzione pratica e gratuita, in particolare una libreria software con l'unico scopo di rilevare, in tempo reale, possibili comportamenti di cyberbullismo su piattaforme online, offrendo agli sviluppatori e ai gestori delle piattaforme un mezzo per contrastare tempestivamente il fenomeno.

## 1.2 Struttura della tesi

La tesi è articolata in maniera chiara e sistematica, ripartendo l'intera discussione in tre parti distintive per garantire una completa comprensione del problema e delle soluzioni proposte.

### Parte I: Studio del Fenomeno

Questa sezione delinea la genesi e l'evoluzione del cyberbullismo nell'era digitale, affrontando le problematiche sottostanti e le sfide uniche che presenta.

- **Capitolo 2:** Il fulcro di questo capitolo riguarda l'intersezione tra bullismo tradizionale e cyberbullismo. Qui si esplora la metamorfosi del fenomeno con la crescente dipendenza dalla tecnologia e si discute di come esso influenzi vari aspetti della società contemporanea.

### Parte II: Tecnologie Utilizzate e Studiate

La parte centrale della tesi è dedicata all'analisi delle metodologie e degli strumenti tecnologici che possono contribuire a contrastare il cyberbullismo.

- **Capitolo 3:** Questo capitolo si immerge nel mondo del Natural Language Processing (NLP), mettendo in evidenza come questo possa servire come chiave per decifrare e interpretare il linguaggio umano in contesti digitali.
- **Capitolo 4:** Qui l'attenzione si sposta sugli specifici strumenti e librerie utilizzati per implementare soluzioni basate sull'NLP, evidenziando il ruolo predominante di Python e delle sue librerie nell'ambito del Natural Language Processing.
- **Capitolo 5:** L'ultimo capitolo di questa parte si concentra sulla comparazione delle diverse librerie disponibili, fornendo una panoramica critica delle potenzialità e delle limitazioni di ciascuna.

### Parte III: Risultati e Accessibilità

Questa sezione finale si propone di presentare non solo i risultati ottenuti, ma anche di sottolineare l'importanza di rendere tali soluzioni ampiamente disponibili.

- **Capitolo 6:** Qui si discute dell'importanza di garantire un accesso aperto e trasparente alle soluzioni proposte, riflettendo sul potenziale impatto reale nel combattere il cyberbullismo e sull'integrazione di tali soluzioni nelle piattaforme online.
- **Capitolo 7:** In conclusione, questo capitolo fa il punto sui risultati raggiunti, mettendo in luce gli obiettivi conseguiti e proponendo una visione per futuri sviluppi e raffinamenti nell'area.



## 2. Bullismo e Cyberbullismo

Il termine "bullismo" deriva dal termine inglese "bully"[13], che originariamente aveva un significato di "amico", ha assunto, con il tempo, una connotazione negativa e ad oggi fa riferimento a comportamenti ripetuti nel tempo volti a danneggiare fisicamente, verbalmente o psicologicamente un'altra persona. Inizialmente, il bullismo era percepito come un problema legato all'ambiente scolastico, dove alcuni studenti cercavano di affermare la propria superiorità a discapito degli altri, tuttavia, con l'espansione di Internet e dei social media, questi comportamenti si sono trasferiti online, dando vita al cyberbullismo, che a differenza del bullismo tradizionale, può avere un impatto ancora più devastante, in quanto le sue azioni possono raggiungere un pubblico molto più vasto in tempi brevissimi, e inoltre, la natura permanente del contenuto online rende le vittime costantemente esposte.

### 2.1 L'Emergenza del bullismo

In un articolo [22] di recente pubblicazione, dove si parla di questo fenomeno, troviamo la seguente spiegazione:

*Il bullismo è definito come l'aggressione o la molestia reiterate, da parte di una singola persona o di un gruppo di persone, in danno di un minore o di un gruppo di minori, idonee a provocare sentimenti di ansia, di timore, di isolamento o di emarginazione, attraverso atti o comportamenti vessatori, pressioni o violenze fisiche o psicologiche, istigazione al suicidio o all'autolesionismo, minacce o ricatti, furti o danneggiamenti, offese o derisioni*

Le motivazioni alla base del bullismo sono varie e spesso intricate, originando da un mix di dinamiche interne ed esterne, alcune persone, ad esempio, possono vederlo come un mezzo per affermare la propria autorità o per esercitare un senso di controllo sugli altri e questo può nascere dal desiderio di mascherare le proprie vulnerabilità. In altri casi, le ragioni possono essere più radicate in insicurezze personali, infatti alcuni bulli potrebbero sentirsi insicuri a riguardo della propria identità, del proprio aspetto fisico o delle proprie capacità, e prendersela con gli altri può sembrare un modo per elevare il proprio status o autostima. Inoltre, ci sono questioni sociali più ampie che possono contribuire al fenomeno, come la pressione dei coetanei, la ricerca di appartenenza o il desiderio di conformarsi alle norme sociali, che inducono alcuni giovani a comportarsi in modo intimidatorio o aggressivo verso gli altri. Per evidenziare ulteriormente il problema, è opportuno ricordare che:

*Il bullismo non è una ragazzata*

come ricordato in un recente articolo pubblicato su la Repubblica [19]

## **2.2 Sviluppo tecnologico e nascita del Cyberbullismo**

L'integrazione delle tecnologie nella nostra vita quotidiana ha rivoluzionato il modo in cui interagiamo tra di noi, tuttavia, come con qualsiasi progresso, vi sono sia vantaggi che svantaggi.

Negli ultimi anni, la diffusione dei dispositivi digitali tra la popolazione più giovane è stata fenomenale, le piattaforme social sono diventate gli equivalenti digitali dei cortili delle scuole, dove i giovani possono interagire, condividere e crescere.

Questa interconnessione ha portato a nuovi modi di apprendimento, condivisione e espressione di sé, ma come ogni medaglia, ha il suo rovescio, l'espansione di tali piattaforme ha dato anche spazio a nuovi comportamenti nocivi, tra cui il cyberbullismo.

Durante la pandemia da COVID-19, la dipendenza dalla tecnologia è diventata ancora più evidente, la scuola, luogo tradizionalmente fisico, è stata sostituita da aule virtuali, comportando una maggiore esposizione online e offrendo al contempo ulteriori opportunità per comportamenti malevoli.

Mentre gli adolescenti sono una fascia demografica particolarmente esposta, non dobbiamo dimenticare i bambini più piccoli, la familiarità precoce con tablet e smartphone, se non guidata da un'adeguata formazione su un uso sicuro del web, può metterli a rischio.

Il cyberbullismo, così come il bullismo tradizionale, nasce da un desiderio di dominio e controllo, la differenza principale è che il cyberbullismo si svolge in uno spazio digitale, dove l'anonimato può rendere i cyber bully più audaci e le vittime più isolate.

Mentre la tecnologia continua a progredire, è fondamentale che la società riconosca e affronti queste sfide emergenti, l'educazione digitale [5], che comprende la navigazione sicura online, la consapevolezza dei rischi del cyberbullismo e l'importanza dell'empatia nell'era digitale, dovrebbe essere integrata nei programmi scolastici.

*Educare i ragazzi a questi principi significa gettare le basi per poter contare domani su una rete più sana è giusto mostrare alle nuove generazioni che in rete ci sono pericoli come anche grandi opportunità e che il modo di comunicare in quel contesto ha un peso, soprattutto che si può comunicare senza essere secchi, duri o maleducati*

Solo attraverso la consapevolezza e l'istruzione possiamo sperare di navigare con successo in questo nuovo ambiente digitale.

## **2.3 Impatti del Cyberbullismo nell'era digitale**

Il Ministero dell'Istruzione e del Merito ha evidenziato la gravità del cyberbullismo con una dichiarazione che colpisce profondamente [18]:

*Oggi la tecnologia consente ai bulli di infiltrarsi nelle case delle vittime, di materializzarsi in ogni momento della loro vita, perseguitandole con messaggi, immagini, video offensivi inviati tramite smartphone o pubblicati sui siti web tramite Internet.*

Nella realtà virtuale, il concetto di un "rifugio" sicuro dal cyberbullismo sembra svanire. La costanza con cui il cyberbullismo colpisce le vittime evidenzia una delle sue nature più pericolose e traumatizzanti.

## 3. Natural Language Processing (NLP)

Il Natural Language Processing (NLP) rappresenta una delle branche più avvincenti e rilevanti dell'intelligenza artificiale e dell'informatica. Il suo obiettivo principale è quello di creare sistemi che non solo "capiscano", ma anche generino e rispondano al linguaggio in maniera fluida, quasi umana.

Questo capitolo ha l'obiettivo di fornire una visione d'insieme della libreria, evidenziando la sua rilevanza nell'interpretazione del linguaggio umano e la sua applicazione specifica, infatti NLP non si limita a semplici chatbot o assistenti virtuali, la sua applicazione si estende a molteplici sfere, come la traduzione automatica, la sintesi vocale e l'analisi del sentimento. In questo capitolo, ci concentreremo non solo sulle basi dell'NLP e sulle sue incredibili potenzialità, ma anche su come può diventare una risorsa preziosa nella lotta contro il cyberbullismo.

### 3.1 Concetto di NLP

Che cos'è il Natural Language Processing? [14]

*Per Natural Language Processing o elaborazione del linguaggio naturale si intendono algoritmi di intelligenza artificiale in grado di analizzare, rappresentare e quindi comprendere il linguaggio naturale. Le finalità possono variare dalla comprensione del contenuto, alla traduzione, fino alla produzione di testo in modo autonomo a partire da dati o documenti forniti in input*

Il suo punto di forza, come anticipato, è quello di far sì che i computer "capiscano" il nostro modo di parlare, ma non stiamo parlando solo di riconoscere parole, è come se insegnassimo alle macchine a leggere tra le righe, per comprendere quello che diciamo, il come e il perché.

Si immagini di raccontare una storia a un amico: non si vuole solamente che ascolti le parole, ma che comprenda i sentimenti, le emozioni e le intenzioni dietro di esse.

Questo fa capire che, quando si tratta di argomenti delicati come il cyberbullismo, la capacità di comprendere, di NLP, può fare una grande differenza.

## 3.2 Il ruolo cruciale del NLP nella comprensione del linguaggio umano

Il linguaggio è il principale mezzo di comunicazione tra gli esseri umani, esso ci consente di esprimere idee, emozioni, desideri e conoscenze. Il Natural Language Processing è una disciplina fondamentale in questo contesto, servendo da ponte tra l'uomo e la capacità delle macchine di interpretare il suo parlato. Attraverso questa tecnologia, quest'ultime sono ora capaci di rilevare i sentimenti, identificare temi centrali nei testi e perfino creare contenuti che suonano naturali e fluidi come se fossero stati scritti da una persona.

Il perfezionamento delle tecniche di NLP ha portato a una sua adozione sempre più diffusa in numerosi ambiti della nostra vita quotidiana. Si pensi ai chatbot, capaci di dialogare con gli utenti in maniera sempre più naturale, ai sistemi di raccomandazione che scrutano e interpretano le opinioni degli utenti per offrire suggerimenti mirati, o ancora agli avanzati strumenti medici che analizzano con precisione le annotazioni cliniche. Questi sono solo alcuni esempi che sottolineano l'importanza crescente e l'indispensabilità del NLP nel panorama tecnologico contemporaneo.

In sintesi, questa libreria offre strumenti e tecniche che permettono alle macchine di immergersi nel complesso mondo del linguaggio umano, rendendo possibile una cooperazione più fluida tra uomo e macchina.

## 3.3 Approcci NLP per affrontare il fenomeno del Cyberbullismo

Gli sviluppi nell'epoca contemporanea hanno visto il NLP passare da semplici operazioni di parsing e ricerca di parole chiave a sofisticati algoritmi in grado di generare testo, tradurre lingue in tempo reale e partecipare a conversazioni con gli esseri umani. Questi progressi, tuttavia, non sono meramente teorici o limitati a settori come la traduzione o la generazione di contenuti, essi trovano applicazioni concrete in sfide sociali come la lotta contro il cyberbullismo.

NLP entra in gioco in questo scenario fornendo strumenti capaci di:

- **Identificare automaticamente i contenuti offensivi:** Attraverso la classificazione di testi e l'analisi semantica, esso può scovare messaggi, commenti o post che potrebbero avere un tono o contenuto molesti.
- **Analizzare le tendenze:** La capacità di processare grandi quantità di dati in breve tempo permette di rilevare schemi o tendenze emergenti nel cyberbullismo, facilitando gli interventi di prevenzione.
- **Assistenza alle vittime:** Sistemi basati su NLP possono essere integrati in piattaforme social o app di messaggistica per fornire supporto immediato alle vittime, guidandole su come agire o semplicemente offrendo un primo supporto.
- **Educazione e consapevolezza:** Gli algoritmi di NLP possono essere utilizzati per creare contenuti educativi mirati, aiutando gli utenti a comprendere la portata e le implicazioni del cyberbullismo, fornendo consigli su come proteggersi e reagire.

### 3.4 Tecniche avanzate di analisi del linguaggio impiegate

Nell'ampio panorama dell'elaborazione del linguaggio naturale (NLP), le metodologie e le tecniche si sono evolute rapidamente, al passo con i progressi tecnologici. Alcune di queste, considerate "avanzate", rappresentano la frontiera dell'analisi del linguaggio, offrendo soluzioni più raffinate e precise. Questa sezione si propone di esplorare e illustrare alcune di queste metodologie delineando come esse hanno rivoluzionato il modo in cui le macchine comprendono e interagiscono con il linguaggio umano

- **Modelli di Word Embedding:** Questi modelli, come Word2Vec[29] e GloVe, consentono di rappresentare parole in vettori in uno spazio multidimensionale, dove le parole semanticamente simili sono vicine tra loro. Questa rappresentazione vettoriale facilita l'identificazione di termini e frasi offensivi basati sulla loro somiglianza semantica con termini noti.
- **Modelli di apprendimento profondo:** Reti neurali come le Reti Neurali Ricorrenti (RNN[4]) e le Transformer, come BERT, sono addestrate su enormi quantità di testo per comprendere il contesto e la semantica delle frasi. Questi modelli possono identificare molteplici sfumature nel linguaggio, rendendoli strumenti potenti per identificare il cyberbullismo, anche quando viene utilizzato un linguaggio codificato o implicito, tipico dei social. Un esempio di queste può essere visto nella Figura 3.1.
- **Analisi dei sentimenti[15]:** Questa tecnica impiega modelli di NLP per valutare il tono e l'emozione dietro una frase o un testo. È utile per rilevare commenti o messaggi con intenti negativi o maliziosi.
- **Classificazione del testo:** Usando algoritmi supervisionati, i modelli vengono addestrati su dati etichettati per classificare nuovi testi in categorie predefinite, come "offensivo", "neutro" o "positivo". Questo può essere utilizzato per filtrare automaticamente i contenuti dannosi.
- **Tecniche di clustering[21]:** Queste tecniche non supervisionate raggruppano testi simili tra loro, permettendo di identificare temi o argomenti comuni. Un esempio di questo può essere visto nella Figura 3.2.

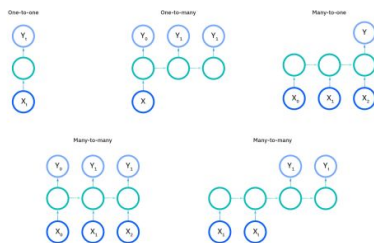


Figura 3.1: Immagine raffigurante le Reti Neurali Ricorrenti

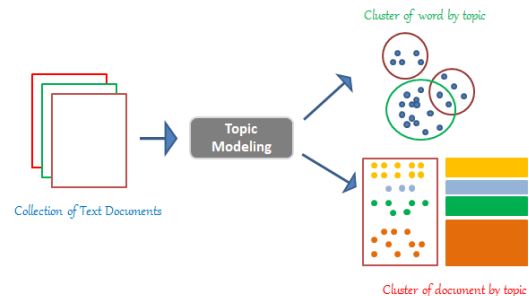


Figura 3.2: Immagine raffigurante il clustering



## 4. Tecnologie Utilizzate

Nell'odierno panorama tecnologico, la selezione degli strumenti giusti è essenziale per garantire l'efficacia e la precisione di qualsiasi progetto. In questo capitolo, ci concentreremo nell'analisi delle tecnologie che abbiamo adottato nel corso della nostra ricerca, dalla versatilità di Visual Studio alle potenti capacità di Python. Discuteremo in dettaglio l'importanza di un codice ben strutturato e come esso ha giocato un ruolo determinante nel massimizzare i benefici derivati dall'uso di queste tecnologie avanzate, nel contesto specifico della nostra analisi.

### 4.1 L'Incorporazione delle tecnologie nell'analisi del Cyberbullismo

La complessità e la natura mutevole del cyberbullismo richiedono un'attenta selezione di strumenti tecnologici che vadano oltre la semplice capacità di raccogliere e elaborare dati. Infatti, questi strumenti devono essere all'avanguardia e, allo stesso tempo, intuitivi per poter svolgere un'analisi profonda ed efficace.

La mia scelta è ricaduta su Visual Studio Code [27], logo in figura 4.1, una piattaforma che combina una serie di funzionalità avanzate con un'interfaccia chiara e diretta, rendendo l'intero processo di programmazione più agile e meno complesso. La sua vasta gamma di funzionalità, insieme alla possibilità di integrare facilmente diverse librerie, fa di Visual Studio un ambiente di sviluppo indispensabile per affrontare un'indagine così importante.

Tuttavia, quando parliamo di analisi dati, Python[25], logo in figura 4.2. emerge come il vero leader. Questo linguaggio, con le sue numerose librerie dedicate, offre soluzioni per ogni fase dell'analisi, dalla pulizia dei dati all'implementazione di modelli avanzati. La flessibilità e la potenza di Python ne fanno uno strumento insostituibile, fondamentale per qualsiasi ricerca che intenda sondare a fondo il fenomeno del cyberbullismo.

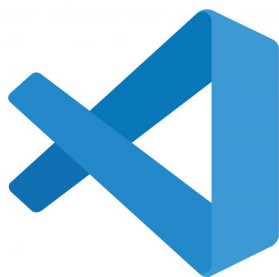


Figura 4.1: Logo Visual Studio Code



Figura 4.2: Logo Python

## 4.2 Il ruolo chiave di Python nell'analisi dei dati

Il mondo dell'analisi dei dati ha subito una metamorfosi straordinaria negli ultimi anni. Questa trasformazione, spinta dall'esigenza di interpretare volumi di dati sempre più grandi e complessi, ha trovato nel linguaggio di programmazione Python il suo protagonista indiscusso.

La rapidità con cui Python è emerso come leader in questo ambito può sorprendere, ma le ragioni della sua ascesa sono chiare e molteplici.

- **Una breve storia di Python:** Python non è nuovo nel panorama della programmazione. Creato alla fine degli anni '80 da Guido van Rossum, è stato concepito per essere un linguaggio di facile lettura, con una sintassi chiara e comprensibile. Questa filosofia ha reso Python una delle scelte principali per chi si avvicina al mondo della programmazione per la prima volta, ma anche per chi, avendo già esperienza, cerca un linguaggio flessibile e potente.
- **Versatilità e Intuibilità:** Python è noto per la sua versatilità. Può essere utilizzato in una vasta gamma di applicazioni, dall'analisi web all'elaborazione delle immagini, e questa flessibilità ha reso facile per gli sviluppatori adottarlo e adattarlo alle loro specifiche esigenze.

In secondo luogo, la sua intuitività. Python è stato progettato con l'obiettivo di essere leggibile e di facile apprendimento. Questo rende la scrittura, la lettura e la manutenzione del codice molto più semplice rispetto ad altri linguaggi. Per chiunque voglia avvicinarsi al mondo dell'analisi dei dati, la curva di apprendimento di Python è notevolmente più dolce rispetto a quella di altri strumenti

- **Librerie e strumenti per l'analisi dei dati:** Il più grande vantaggio di Python risiede nel suo ricco ecosistema di librerie e strumenti, che è in continua espansione. La comunità di sviluppatori di Python è attiva e appassionata, e contribuisce regolarmente con nuove librerie e strumenti che ampliano le capacità del linguaggio. Che si tratti di manipolazione dei dati con Pandas, visualizzazione con Matplotlib, o apprendimento automatico con TensorFlow, c'è una libreria Python per quasi ogni esigenza nell'ambito dell'analisi dei dati.
- **Il ruolo di Python nel NLP:** Quando si tratta di Natural Language Processing, Python risulta ancora una volta essenziale. Librerie come NLTK e spaCy permettono di analizzare, manipolare e interpretare il linguaggio naturale, rendendo possibile studiare fenomeni come il cyberbullismo in modo approfondito e basato su dati.
- **Facilità di integrazione:** Oltre alle sue capacità inerenti l'analisi dei dati, Python si integra facilmente con altri strumenti e tecnologie, rendendolo ideale per progetti multidisciplinari e per la creazione di soluzioni end-to-end.

In sintesi, la popolarità di Python nell'analisi dei dati non è un fenomeno casuale, ma il risultato di una combinazione vincente di versatilità, accessibilità e un ecosistema in costante crescita e innovazione.



### 4.3 Librerie e pacchetti essenziali utilizzati per la ricerca

Python, nella sua essenza, è un linguaggio di programmazione estremamente potente e flessibile. Tuttavia, ciò che lo eleva a un livello superiore e lo rende particolarmente interessante per gli sviluppatori è il suo ricco ecosistema di librerie e pacchetti. Questa vasta collezione di strumenti, costantemente arricchita e migliorata dalla sua dinamica community, è stata ideata per rispondere a svariate esigenze e sfide.

Ciascuna libreria o pacchetto ha l'obiettivo di semplificare o automatizzare specifici compiti o operazioni, che senza di essi sarebbero laboriosi o addirittura proibitivi. Che si tratti di analisi dei dati, visualizzazione, machine learning o elaborazione del linguaggio naturale, esiste quasi certamente una libreria Python che può facilitare notevolmente il lavoro. In questo senso, mentre il linguaggio Python fornisce le fondamenta, sono queste librerie e pacchetti a costruire l'edificio, offrendo soluzioni chiavi in mano per svariati problemi e applicazioni.

Nel contesto della nostra ricerca sul cyberbullismo, l'utilizzo di alcune di queste librerie è stato fondamentale. Non solo hanno accelerato il processo di analisi, ma hanno anche offerto metodi e tecniche all'avanguardia per decifrare e interpretare i dati raccolti. Nel proseguo di questa sezione, approfondiremo le specifiche librerie utilizzate e come hanno contribuito al successo del nostro studio.

- **Pandas[26]:** È molto più di una semplice libreria: rappresenta un vero e proprio ambiente di lavoro per dati strutturati. La sua capacità di gestire grandi set di dati, combinata con funzionalità per la pulizia, l'aggregazione e la visualizzazione, rende Pandas uno strumento indispensabile per l'analista moderno.
- **Scikit-learn[24]:** Questa libreria rappresenta il cuore pulsante del machine learning in Python. Con una gamma impressionante di algoritmi e funzionalità, Scikit-learn ha il potere di decifrare e svelare le intricate relazioni nei dati. Nell'ambito della ricerca sul cyberbullismo, ha permesso di identificare e comprendere schemi e tendenze che sarebbero altrimenti rimasti nascosti.
- **NLTK[28] (Natural Language Toolkit):** Quando si parla di linguaggio e testo, NLTK è la chiave. Sviluppato specificamente per affrontare le sfide uniche del Natural Language Processing, questo toolkit non solo offre funzionalità per l'elaborazione del linguaggio, ma rappresenta anche un prezioso alleato per chi opera nel campo della linguistica computazionale.
- **TensorFlow[3] e Keras[12]:** Questi due strumenti rappresentano l'avanguardia dell'apprendimento profondo. TensorFlow fornisce le fondamenta robuste e scalabili su cui gli algoritmi di apprendimento sono costruiti, mentre Keras rende il processo di costruzione di reti neurali complesse sorprendentemente semplice e intuitivo, fungendo da ponte tra il programmatore e le potenzialità di TensorFlow.
- **Matplotlib[16] e Seaborn[17]:** Entrambe librerie incentrate sulla visualizzazione dei dati. Matplotlib offre la libertà e la flessibilità di creare una vasta gamma di grafici, mentre Seaborn porta la visualizzazione ad un livello superiore, integrando analisi statistiche avanzate e producendo visualizzazioni informative.

Nel campo specifico dell'analisi del cyberbullismo, NLTK si è rivelato insostituibile, fornendo gli strumenti chiave per l'elaborazione e l'analisi dettagliata del linguaggio naturale, facilitando così l'indagine su testi e interazioni online pervase da questa problematica.



## 5. Analisi, Confronto e Scelta della Libreria Ottimale

La scelta della libreria giusta può determinare la qualità e l'efficacia di quest'ultima. Durante la redazione di questa tesi, ho spesso citato diverse librerie e pacchetti integrabili nell'ecosistema Python, ora intendo illustrare i criteri e i motivi che hanno guidato le mie scelte nell'adottarne certe piuttosto che altre.

### 5.1 Raccolta e Pre-elaborazione dei dati

In questo primo step, ho deciso di non addentrarmi troppo all'interno delle caratteristiche di ognuna, ma ho dedicato del tempo ad analizzare i contesti nelle quali queste dovrebbero essere impiegate. I primi due punti dove mi sono focalizzato sono i seguenti:

- **Versatilità:** Affrontare il cyberbullismo richiede strumenti capaci di adattarsi ai molteplici canali in cui si manifesta, in particolare sui social. Una libreria che gestisca non solo il testo ma anche Emoji ed Emoticon, facilita grandemente la comprensione e l'esplorazione di tali ambienti virtuali.
- **Portabilità:** Una libreria deve interfacciarsi facilmente con altri strumenti e piattaforme, garantendo così una maggiore flessibilità nello sviluppo di soluzioni integrate.

Successivamente la mia attenzione si è spostata, verso le modalità di analisi e gestione di grandi set di dati, da questa esigenza nascono infatti:

- **Performance e Scalabilità:** Oltre alla capacità di base, la velocità e l'efficienza con cui una libreria può processare grandi quantità di dati diventano criteri decisivi. Soprattutto nel contesto dei social media, dove le quantità di dati sono enormi, è imperativo che la performance non venga compromessa.
- **Facilità d'uso:** Nella NLP, è fondamentale una libreria che unisca intuitività per chi si avvicina a funzionalità avanzate per gli esperti. Una buona interfaccia e documentazione chiara possono fare una grande differenza.

Infine, per concludere questa sezione, ho rivolto uno sguardo anche a tutto quello che "circonda" questa realtà, estrapolando:

- **Comunità e Supporto:** Una libreria di successo nel campo della NLP si fonda su una comunità solida e attiva: la presenza di sviluppatori dedicati che offrono supporto e tutorial è fondamentale. Allo stesso tempo, la manutenzione costante e gli aggiornamenti regolari assicurano che la libreria rimanga al passo con le evoluzioni nel settore, garantendo affidabilità e innovazione nel lungo termine.

## 5.2 Metodologie di analisi basate su diverse librerie di NLP

Una volta definite le linee guida, a seguito di ulteriori approfondimenti, ho definito tutti i criteri che rispecchiano le esigenze uniche della mia ricerca, ecco una spiegazione dettagliata di ognuno:

- **Lingua:** Una libreria che supporta molteplici lingue garantisce una maggiore copertura e una comprensione più profonda del fenomeno, considerando anche le diversità culturali e linguistiche.
- **Tokenizzazione[1]:** Questa fase, che suddivide il testo in parole o "token", è fondamentale per l'analisi testuale. Una buona tokenizzazione può influire significativamente sulla qualità dell'analisi successiva.
- **Part-of-Speech Tagging[2] (POS):** Identificare la parte del discorso di ogni parola è essenziale per comprendere il contesto e la struttura grammaticale dei dati testuali, fornendo informazioni preziose sul modo in cui le persone comunicano online.
- **Lemmatizzazione[20]:** Questa funzione riduce le parole alla loro forma di base, facilitando l'analisi e la categorizzazione del contenuto. Ad esempio, parole come "correndo", "corre", "corso" potrebbero essere tutte ridotte al lemma "correre".
- **Parsing Sintattico:** Comprendere la struttura delle frasi e le relazioni tra le parole è cruciale per analizzare il tono e l'intenzione dell'autore, aspetti fondamentali nell'identificazione del cyberbullismo.
- **Sentiment Analysis:** Data la natura emotiva del cyberbullismo, identificare e comprendere il sentimento dietro ogni messaggio è essenziale per valutare la gravità e l'intenzione dell'autore.
- **Stopword Removal:** Eliminando le parole comuni che non aggiungono significato significativo al testo, si può focalizzare l'analisi sulle parole chiave pertinenti.
- **Named Entity Recognition[23] (NER):** Identificare nomi di persone, luoghi o entità può fornire insight preziosi sulle dinamiche del cyberbullismo e sugli individui o gruppi più comunemente bersagliati.
- **Metodologia di Analisi:** Una libreria con diverse tecniche e algoritmi di analisi garantisce flessibilità e una maggiore profondità nell'analisi dei dati.
- **Licenza:** A seconda dell'ambito e delle finalità della ricerca, è importante valutare la licenza associata a ogni libreria. Alcune possono avere restrizioni sull'uso commerciale o sulla redistribuzione, mentre altre possono essere completamente open source.

Questi criteri sono stati fondamentali nella mia decisione finale, poiché hanno fornito una chiara roadmap per valutare le molteplici opzioni disponibili e selezionare quelle più in linea con i risultati che si vogliono ottenere.

### 5.3 Analisi critica delle prestazioni

In questa sezione, presenterò le librerie scelte, spiegando per ognuna quali caratteristiche la contraddistinguono dalle altre. Successivamente fornirò delle tabelle, da me utilizzate, per analizzare al dettaglio e confrontare performance/affidabilità di ognuna di esse:

- **NLTK[7] (Natural Language Toolkit)** E' ad oggi un pilastro nel mondo del trattamento del linguaggio naturale. Nasce come progetto accademico ed è cresciuto fino a diventare una delle librerie di riferimento per gli studiosi e gli appassionati di linguistica computazionale. Offre la possibilità, all'utilizzatore, di costruire modelli di sentiment analysis personalizzati, dispone di un'ampia gamma di risorse linguistiche, come corpora e lemmatizzatori ed infine ha tutt'ora una vasta comunità di sviluppatori attiva.
- **TextBlob[10]** Si presenta come una libreria di NLP per Python che fa da ponte tra la semplicità e le funzionalità avanzate. Fondamentalmente, TextBlob può essere visto come un involucro intorno a NLTK, il che rende molto più semplici e intuitivi diversi compiti di NLP. Uno dei principali vantaggi è la sua semplicità, infatti un utente può eseguire operazioni come la correzione ortografica, la traduzione e l'analisi del sentimento con poche righe di codice.
- **spaCy[8]** E' una delle librerie di NLP più avanzate e ad alte prestazioni per Python. Oltre alle funzionalità di base del NLP, spaCy offre moduli per il riconoscimento di entità nominate (NER), la classificazione del testo e la segmentazione delle frasi. Il suo focus, basato sull'efficienza e l'estensibilità, è particolarmente adatto per progetti che richiedono un'elaborazione del linguaggio naturale di alto livello.
- **Vader[11]** E' uno strumento di analisi del sentimento progettato per essere particolarmente efficace nel riconoscere le emozioni in testi come quelli trovati sui social media, recensioni e altri tipi di testi informali. Questa libreria utilizza un insieme di regole e un dizionario di parole predefinite per interpretare la valenza emotiva del testo.
- **Text2Emotion[9]** E' una libreria Python che, come suggerisce il nome, è progettata per estrarre le emozioni dal testo. A differenza delle classiche tecniche di analisi del sentimento, che spesso classificano il testo come positivo, neutro o negativo, Text2Emotion cerca di identificare e quantificare specifiche emozioni come felicità, tristezza, rabbia, paura e sorpresa.

Dopo aver introdotto le librerie e i pacchetti essenziali alla nostra ricerca, è il momento di metterli a confronto. Utilizzeremo i principi e criteri menzionati nella sezione precedente come metriche di valutazione. Questo confronto ci consentirà non solo di comprenderne le potenzialità, ma anche di individuare eventuali limiti di ciascun strumento rispetto all'analisi del cyberbullismo nei social media.

Tabella 5.1: Tabella di Panoramica Generale

Libreria	Lingua	Metodologia di Analisi	Licenza
NLTK	Ottimizzata per l'inglese	Machine Learning e Analisi Lessicale	Apache License 2.0
TextBlob	Ottimizzata per l'inglese	Analisi lessicale	MIT License
spaCy	Multi-lingua	Deep learning	MIT License
Vader	Inglese	Analisi lessicale	MIT License
Text2Emotion	Inglese	Parole chiave e Espressioni	MIT License

Quando si tratta di supporto linguistico, spaCy si distingue tra le librerie in esame per la sua capacità di gestire diverse lingue, consolidandosi come uno strumento potente per l'elaborazione del linguaggio naturale in vari contesti. Anche se NLTK e TextBlob sono principalmente orientati verso l'inglese, offrono la versatilità di adattarsi e interagire con altre lingue attraverso estensioni e risorse aggiuntive. D'altro canto, VADER e Text2Emotion sono specificatamente progettati per l'inglese, con un'attenzione particolare all'analisi di testi informali, come quelli sui social media. In tali ambienti, dove prevale il linguaggio colloquiale e le espressioni emotive sono cruciali, la specializzazione di queste librerie è un asset di valore inestimabile.

La metodologia di analisi varia notevolmente tra le diverse librerie. Mentre sia NLTK che TextBlob sono basati su metodi di analisi lessicale, NLTK offre anche strumenti di machine learning, rendendola una soluzione più flessibile per diversi scenari di NLP. SpaCy, d'altro canto, fa uso del deep learning, offrendo capacità avanzate di elaborazione e maggiore accuratezza, soprattutto per compiti di NLP di livello superiore. Vader e Text2Emotion si concentrano rispettivamente sull'analisi lessicale e sull'identificazione delle emozioni attraverso parole chiave ed espressioni. Quest'ultimo approccio è particolarmente efficace per rilevare le sfumature emotive in testi brevi come quelli dei social media.

Quando si tratta di licenze, tre delle librerie (TextBlob, spaCy e Vader) utilizzano la licenza MIT, che è una licenza permissiva riconosciuta e ben consolidata nel mondo open-source. Questo significa che gli sviluppatori possono utilizzarle, modificarle e redistribuirle con poche restrizioni. NLTK, d'altra parte, adotta l'Apache License 2.0, un'altra licenza permissiva che, oltre alla licenza MIT, offre una protezione esplicita contro l'appropriazione di brevetti. Entrambe le licenze sono ideali per chi cerca flessibilità nell'uso delle librerie senza restrizioni onerose.

Tabella 5.2: Tabella di Base NLP

Libreria	Tokenizzazione	P.O.S.	Lemmatizzazione	Stopword Removal
NLTK	★★★★★	★★★★★	★★★★★	★★★★★
TextBlob	★★★★★	★★★★★	★★★★★	★★★★★
spaCy	★★★★★	★★★★★	★★★★★	★★★★★
Vader	★★★★★	★★★★★	★★★★★	★★★★★
Text2Emotion	★★★★★	★★★★★	★★★★★	★★★★★

Per quanto riguarda la tokenizzazione, spaCy si distingue come leader, essendo veloce, accurato e particolarmente robusto, soprattutto grazie ai suoi modelli pre-addestrati. La sua abilità di gestire eccezioni ed affrontare situazioni intricate lo rende unico nel suo genere. A seguire, NLTK mostra una notevole flessibilità ed è personalizzabile. Tuttavia, potrebbe necessitare di una configurazione aggiuntiva e, in determinati scenari, potrebbe non raggiungere l'accuratezza e la velocità di spaCy. TextBlob, ereditando caratteristiche da NLTK e Pattern, offre una tokenizzazione soddisfacente, ma senza la stessa flessibilità e precisione delle opzioni precedenti. Alla fine della lista, troviamo Vader e Text2Emotion. Concentrandosi principalmente sull'analisi del sentimento e delle emozioni, le loro capacità di tokenizzazione non possono competere in dettaglio e flessibilità con librerie specializzate come NLTK o spaCy.

Per quanto riguarda il POS tagging, spaCy si posiziona in testa: utilizza modelli di apprendimento profondo allenati su vaste quantità di dati, assicurando precisione e velocità. Questa efficienza è ulteriormente evidenziata dal suo pipeline di elaborazione ottimizzato. NLTK, d'altro canto, offre molteplici opzioni e flessibilità, permettendo agli utenti di scegliere tra tagger pre-addestrati o di addestrarne di propri. TextBlob, che combina gli approcci di NLTK e Pattern, è user-friendly ma potrebbe non avere la precisione di spaCy. Infine, è importante sottolineare che Vader e Text2Emotion sono focalizzate principalmente sull'analisi del sentimento, quindi le loro funzionalità di POS tagging potrebbero necessitare di integrazioni esterne per una completa efficienza.

Per quanto riguarda la lemmatizzazione, spaCy eccelle sia in precisione che in velocità, rendendolo una scelta ottimale per molti progetti. NLTK e TextBlob seguono da vicino, mentre entrambi offrono una certa flessibilità, potrebbero necessitare di una configurazione più approfondita per ottenere risultati ottimali. Va notato che Vader e Text2Emotion, essendo focalizzate sull'analisi delle emozioni e del sentimento, non sono state concepite per funzioni di lemmatizzazione.

La rimozione delle "stopword" è una fase fondamentale nella pre-elaborazione del testo nelle attività di NLP. Queste parole, spesso considerate "di riempimento", possono non avere grande rilevanza in certe analisi, come parole come "and", "the" o "is" in inglese. In questo contesto, spaCy si distingue per il suo approccio innovativo e versatile. NLTK, essendo una delle librerie storiche di NLP, presenta una lista di "stopword" per molte lingue e, grazie alla sua esperienza, ha un set di "stopword" ampiamente riconosciuto. Di conseguenza, TextBlob, che si basa su NLTK, offre prestazioni simili. D'altra parte, VADER non privilegia la rimozione delle "stopword", dato che in analisi del sentimento, il contesto e le parole di riempimento possono essere essenziali. Infine, Text2Emotion, essendo focalizzato principalmente sull'identificazione delle emozioni nel testo, presenta capacità limitate in questo ambito.

Tabella 5.3: Tabella di Analisi Sintattica e Semantica

Libreria	Parsing Sintattico	Named Entity Recognition (NER)
NLTK	★★★★★	★★★★★
TextBlob	★★★★★	★★★★★
spaCy	★★★★★	★★★★★
Vader	★★★★★	★★★★★
Text2Emotion	★★★★★	★★★★★

Il parsing sintattico si riferisce all'analisi della struttura grammaticale di una frase, determinando le relazioni tra le parole, qui spaCy utilizzando modelli di apprendimento profondo fornisce alberi di dipendenza facili da navigare e utilizzare risultando veloce e preciso. NLTK invece, fornisce strumenti per il parsing, ma che non hanno le prestazioni di metodi più moderni e richiedono delle configurazioni aggiuntive. Allo stesso modo TextBlob, ha funzionalità di base, ma per analisi dettagliate e complesse, necessita di integrazioni. Infine, troviamo Vader e Text2Emotion che non sono progettate per il parsing sintattico e di conseguenza le loro funzionalità sono limitate nel contesto.

La Named Entity Recognition (NER) è il processo di identificare e classificare le entità nominate (come persone, luoghi, date, organizzazioni) all'interno di un testo, in questo spaCy è uno dei leader nel campo grazie alla sua velocità, precisione e facilità d'uso, segue subito NLTK, che utilizzando algoritmi basati su tecniche tradizionali non ha la stessa precisione degli approcci basati sul deep learning, questa libreria inoltre viene integrata da TextBlob, in quanto al suo interno non abbiamo un modulo specifico per la NER. Seguono Vader e Text2Emotion, che non hanno e non forniscono alcuna funzionalità a riguardo

Tabella 5.4: Tabella di Analisi delle emozioni

Libreria	Sentiment Analysis
NLTK	★★★★★
TextBlob	★★★★★
spaCy	★★★★★
Vader	★★★★★
Text2Emotion	★★★★★

Come suggerisce il nome, Text2Emotion è specificamente progettato per identificare e classificare le emozioni presenti nel testo, infatti, riconosce emozioni come felicità, tristezza, rabbia, paura e sorpresa. Segue Vader, che però è più focalizzato nel rilevare la polarità delle frasi analizzate. Successivamente, troviamo spaCy che non ha funzionalità specifiche a riguardo e per essere utilizzato con tale scopo richiede un certo livello di personalizzazione. Si termina con NLTK e Textblob, che non offrono strumenti specifici, ma resistituiscono un'analisi del sentimento di base (polarità).



## 5.4 Vantaggi e svantaggi delle librerie scelte

NLTK è una delle pietre miliari nel campo del NLP all'interno dell'ecosistema Python. Grazie alla sua storia e tradizione, offre una vasta gamma di strumenti e corpora. La sua abbondante documentazione, unita a una comunità di utenti molto attiva lo rende un punto di riferimento imprescindibile per chi si addentra nel mondo del NLP. Detto ciò, la sua ricchezza, che rappresenta uno dei suoi principali vantaggi, può occasionalmente diventare un ostacolo, poiché alcune attività possono richiedere più codice rispetto ad altre soluzioni più moderne. In termini di pura performance, NLTK potrebbe non sempre garantire la massima velocità o efficienza rispetto ad altre librerie.

TextBlob si distingue per la sua innegabile semplicità d'uso. Progettato con un'interfaccia intuitiva, facilita la realizzazione di operazioni, come l'analisi del sentimento o il part-of-speech tagging. Questa caratteristica lo rende particolarmente attraente per chi si avvicina per la prima volta al NLP. Tuttavia, ogni medaglia ha il suo rovescio: la semplicità di TextBlob può tradursi in limitazioni per compiti NLP di livello avanzato e, nei momenti in cui la precisione è cruciale, potrebbe non essere l'opzione più adeguata.

SpaCy si posiziona come un riferimento nel campo dell'NLP, combinando una straordinaria velocità con modelli avanzati basati sul deep learning. La sua architettura ottimizzata assicura risultati precisi, e il suo supporto multilingua lo rende versatile per vari compiti. Inoltre, la sua community attiva è sempre pronta ad aiutare e condividere nuove risorse. Tuttavia, bisogna tenere presente che la sua avanzata tecnologia necessita di modelli di notevole dimensione. Questo potrebbe rendere spaCy meno pratico in ambienti con limitazioni di memoria o potenza di calcolo.

Vader si distingue nel panorama delle librerie NLP per il suo focus determinato sull'analisi del sentimento, risultando particolarmente efficace in contesti come i social media. La sua capacità di identificare non solo la polarità ma anche l'intensità di un sentimento lo rende un alleato prezioso in questo ambito. Tuttavia, la sua specializzazione potrebbe limitarne l'utilizzo ad ambiti specifici, rendendolo meno adatto a una vasta gamma di applicazioni NLP al di fuori dell'analisi del sentimento.

Text2Emotion è dedicata al rilevamento delle emozioni, un campo che si estende ben oltre l'analisi del sentimento in termini di polarità positiva o negativa. Per coloro che cercano di sondare le profondità emotive di un testo, Text2Emotion si rivela uno strumento ineguagliabile. Tuttavia, proprio come VADER, il suo raggio d'azione specializzato lo rende meno adatto per una vasta serie di applicazioni generali nell'NLP.

Ogni libreria ha il suo posto unico nel vasto mondo del NLP. La decisione su quale utilizzare spesso dipenderà dalle specifiche esigenze del progetto in questione. Se l'obiettivo è un'analisi del sentimento rapida e semplice, VADER o TextBlob potrebbero essere adatti. Se si necessita di una soluzione di NLP completa e performante, spaCy potrebbe essere la scelta migliore. E, come sempre, è fondamentale considerare il contesto e le esigenze specifiche.

## 5.5 Riflessioni sull'accuratezza e l'efficienza

Nelle fasi di analisi del testo, molteplici tecniche sono state implementate per garantire che la libreria funzionasse al meglio delle sue potenzialità. Di seguito, esploreremo in dettaglio tutte le tecniche utilizzate, fornendo una visione chiara delle scelte effettuate e del loro impatto sulle performance complessive della libreria. Attraverso questa riflessione, spero di offrire una panoramica dettagliata di come "Human Behavior Variable" gestisce e analizza i dati testuali per ottenere risultati ottimali.

Per testare le varie librerie ho deciso di analizzare la frase:

*"Sono davvero arrabbiato con Marco. Anche se gli avevo chiesto più volte di non farlo, ha rivelato a Lucia il segreto che volevo condividere solo con lui durante la cena di sabato."*

Si parte con il Traduttore, importando la libreria *googletrans*

```
from googletrans import Translator

#Function to translate text to English
def translate_to_english(text):
    translator = Translator()
    detected_lang = translator.detect(text).lang
    translated = translator.translate(text, src=detected_lang, dest='en').text
    return detected_lang, translated

#Translate the phrase
detected_language, translated_phrase = translate_to_english(input_phrase)
```

Eseguendo questa parte di codice si ottiene il risultato in figura 5.1

```
Frase in input: Sono davvero arrabbiato con Marco.
                Nonostante gli avessi chiesto più volte di non farlo,
                ha rivelato a Lucia il segreto che volevo condividere
                solo con lui durante la cena di sabato.
Lingua rilevata: it
Frase tradotta: I am really angry with Marco.
                Although I had asked him several times not to do it,
                he revealed to Lucia the secret I wanted to share
                only with him during Saturday dinner.
```

Figura 5.1: Risultato della traduzione

Nell'analisi svolta, come mostrato nel codice fornito, è stata effettuata una scelta strategica. Invece di fornire direttamente l'italiano come lingua di input, si è sfruttato il potenziale della libreria per riconoscere automaticamente la lingua del testo tramite il `detect`. Questo permette di garantire una maggiore flessibilità nell'analisi di testi in lingue diverse.

Eseguita questa parte, si procede con la Tokenizzazione, ecco la sezione della libreria dedicata:

```
import nltk
from nltk.tokenize import word_tokenize
from textblob import TextBlob
import spacy

#Carico il modello di lingua inglese di spaCy
nlp = spacy.load("en_core_web_sm")

#Tokenizzazione
spacy_tokens = [token.text for token in nlp(translated_phrase)]
nltk_tokens = word_tokenize(translated_phrase)
textblob_tokens = [word for word in TextBlob(translated_phrase).words]
```

In particolare, osservando la segmentazione del testo attorno al punto ".", in figura 5.2

Tokenizzazione con spaCy:	Tokenizzazione con NLTK:	Tokenizzazione con TextBlob:
I	I	I
am	am	am
really	really	really
angry	angry	angry
with	with	with
Marco	Marco.Even	Marco.Even
.	if	if
Even	I	I
if	had	had
I	asked	asked
had	him	him
asked	several	several
him	times	times
several	not	not
times	to	to
not	do	do
to	it	it
do	,	he
it	he	revealed
,	revealed	to
he	to	Lucia
revealed	Lucia	the
to	the	secret
Lucia	secret	I
the	I	wanted
secret	wanted	to
I	to	share
wanted	share	only
to	only	with
share	with	him
only	him	during
with	during	Saturday
him	Saturday	dinner
during	dinner	.
Saturday	.	
dinner		
.		

Figura 5.2: Risultato della Tokenizzazione

SpaCy è stato in grado di separare correttamente le parole "Marco" ed "Even", garantendo una tokenizzazione precisa. Al contrario, sia NLTK che TextBlob hanno mostrato delle difficoltà, interpretando erroneamente la punteggiatura e compromettendo l'accuratezza di quest'ultima.

Si passa al Part of Speech Tagging, ecco il codice utilizzato all'interno della libreria:

```
import spacy
from nltk.tokenize import word_tokenize
from nltk import pos_tag
from textblob import TextBlob

#Carico il modello di linguaggio di spaCy
nlp = spacy.load("en_core_web_sm")
#Definisco una funzione per il POS tagging con spaCy
def pos_spacy(text):
    doc = nlp(text)
    return [(token.text, token.pos_) for token in doc]
#Definisco una funzione per il POS tagging con NLTK
def pos_nltk(text):
    words = word_tokenize(text)
    return pos_tag(words)
#Definisco una funzione per il POS tagging con TextBlob
def pos_textblob(text):
    blob = TextBlob(text)
    return blob.tags
#POS tagging
spacy_tags = pos_spacy(translated_phrase)
nltk_tags = pos_nltk(translated_phrase)
textblob_tags = pos_textblob(translated_phrase)
```

Alla luce dei risultati, in figura 5.3

POS tagging with spaCy:	POS tagging with NLTK:	POS tagging with TextBlob:
I: PRON	I: PRP	I: PRP
am: AUX	am: VBP	am: VBP
really: ADV	really: RB	really: RB
angry: ADJ	angry: JJ	angry: JJ
with: ADP	with: IN	with: IN
Marco: PROP	Marco.Although: NNP	Marco.Although: NNP
..: PUNCT	I: PRP	I: PRP
Although: SCONJ	had: VBD	had: VBD
I: PRON	asked: VBN	asked: VBN
had: AUX	him: PRP	him: PRP
asked: VERB	several: JJ	several: JJ
him: PRON	times: NNS	times: NNS
several: ADJ	not: RB	not: RB
times: NOUN	to: TO	to: TO
not: PART	do: VB	do: VB
to: PART	it: PRP	it: PRP
do: VERB	,: ,	he: PRP
it: PRON	he: PRP	revealed: VBD
,: PUNCT	revealed: VBD	to: TO
he: PRON	to: TO	Lucia: VB
revealed: VERB	Lucia: VB	the: DT
to: ADP	the: DT	secret: JJ
Lucia: PROP	secret: JJ	I: PRP
the: DET	I: PRP	wanted: VBD
secret: NOUN	wanted: VBD	to: TO
I: PRON	to: TO	share: NN
wanted: VERB	share: NN	only: RB
to: PART	only: RB	with: IN
share: VERB	with: IN	him: PRP
only: ADV	him: PRP	during: IN
with: ADP	during: IN	Saturday: NNP
him: PRON	Saturday: NNP	dinner: NN
during: ADP	dinner: NN	
Saturday: PROP	..: .	
dinner: NOUN		
..: PUNCT		

Figura 5.3: Risultato del Part Of Speech Tagging

La superiorità di spaCy nella categorizzazione è evidente: mentre ha identificato "Marco" come un nome proprio (PROP), sia NLTK che TextBlob hanno fornito risultati meno accurati, classificando "Marco.Although" come un singolo token (NNP).

## 6. Contributi e Accessibilità

Nell'era digitale attuale, l'analisi del linguaggio naturale sta aprendo nuove strade e fornendo strumenti innovativi per affrontare sfide complesse. In questo contesto, la libreria "Human Behavior Variable" rappresenta un progetto sperimentale, aspirando a sfruttare la tecnologia per affrontare problemi sociali come il cyberbullismo. Pur non pretendendo di essere una soluzione definitiva, la libreria mette in evidenza il potenziale dell'intelligenza artificiale in tale ambito. Questo capitolo esplorerà le ambizioni di "Human Behavior Variable" nella lotta contro il cyberbullismo e illustrerà come intendo renderla accessibile al più vasto pubblico attraverso piattaforme aperte e facilmente raggiungibili.

### 6.1 Come la tecnologia può fare la differenza nel contrastare il Cyberbullismo

Il cyberbullismo è un fenomeno che ha guadagnato terreno nell'era digitale, spingendo gli esperti e i tecnologi a cercare soluzioni innovative per identificarlo e contrastarlo. Le tecnologie di analisi del linguaggio naturale (NLP) si sono rivelate particolarmente promettenti in questo contesto. Utilizzando modelli avanzati di apprendimento automatico e deep learning, è possibile analizzare enormi quantità di dati testuali alla ricerca di schemi e comportamenti che segnalano il cyberbullismo. Queste tecniche sono essenziali per identificare discorsi d'odio, insulti e commenti denigratori, consentendo interventi proattivi per prevenire o rispondere a tali comportamenti. La tecnologia offre anche soluzioni innovative come chatbot e assistenti virtuali progettati specificamente per fornire supporto alle vittime di questo fenomeno.

Nell'ambito dell'innovazione tecnologica nel settore del NLP, "Human Behavior Variable" emerge come un tentativo sperimentale, ma anche come una sfida, di integrare diverse tecniche per fornire una soluzione più completa. Questa libreria sfrutta metodi avanzati, combinando algoritmi di ultima generazione con tecniche consolidate, per identificare e categorizzare comportamenti potenzialmente nocivi manifestatisi nel linguaggio online. "Human Behavior Variable" è flessibile nella sua architettura e combina diversi approcci per offrire una panoramica più profonda e ampia del fenomeno del cyberbullismo. Attraverso l'analisi approfondita dei modelli linguistici, sofisticate tecniche di analisi del sentimento e la decodifica delle emozioni espresse nei contenuti testuali, la libreria mira a diventare uno strumento essenziale nella lotta contro questo fenomeno, adeguandosi ai molteplici contesti in cui si manifesta.

## 6.2 Condivisione su piattaforme facilmente accessibili

L'importanza di strumenti efficaci nel contrasto al cyberbullismo è indiscutibile. Tuttavia, affinché questi strumenti abbiano un impatto reale, devono essere facilmente accessibili e condivisi con una vasta comunità. Proprio in quest'ottica, "Human Behavior Variable" è stata progettata per essere condivisa.

L'obiettivo principale va oltre la semplice identificazione e comprensione del cyberbullismo. Si tratta di offrire soluzioni concrete per contrastarlo e prevenirlo. E per raggiungere questo obiettivo, è essenziale che la libreria sia alla portata di tutti. Ecco perché "Human Behavior Variable" sarà disponibile come repository open-source su GitHub[6], , logo in figura 6.1. Questa piattaforma, ampiamente riconosciuta nel mondo dello sviluppo software, permette a sviluppatori, ricercatori e appassionati di accedere facilmente al codice, modificarlo e migliorarlo secondo le proprie esigenze.



Figura 6.1: Logo GitHub

L'adozione di un approccio open-source per "Human Behavior Variable" va oltre la semplice trasparenza e accessibilità; rappresenta un impegno verso un principio di collaborazione aperta. Si prevede che, mano a mano che il progetto cresce e si diffonde, la comunità di sviluppatori e ricercatori possa contribuire con il proprio know-how, proponendo suggerimenti, apportando correzioni e migliorando la base di codice esistente. Questa dinamica collaborativa ha il potenziale di portare a innovazioni e adattamenti imprevisti, potenziando ulteriormente l'efficacia di "Human Behavior Variable".

L'aspirazione principale è quella di vedere sorgere una comunità attiva e dedicata intorno a questa libreria. Una rete di professionisti, studenti, e appassionati di tecnologia, uniti da una visione comune: combattere il cyberbullismo e rendere il web un luogo più sicuro. Attraverso la condivisione di idee, l'apporto di diverse prospettive e la continua ricerca di soluzioni innovative, la tecnologia può passare da un ruolo di semplice rilevatore di problemi a catalizzatore di soluzioni, offrendo risposte efficaci alle sfide del mondo digitale.

La libreria è già disponibile e la si può trovare nel seguente repository:

`https://github.com/caggicaggi/HumanBehaviourVariable.git`

## 7. Conclusioni

Mentre si avvicina la conclusione di questo percorso di ricerca, è fondamentale soffermarsi e riflettere sull'importanza delle scoperte fatte e sulle prospettive future. La sfida del cyberbullismo, profondamente radicata nell'era digitale, ha richiesto un'analisi approfondita e la creazione di soluzioni innovative, come "Human Behavior Variable". Questo capitolo intende riepilogare i principali risultati ottenuti, evidenziando l'efficacia e i limiti dell'approccio adottato. Inoltre, si guarderà al futuro, delineando possibili direzioni e sviluppi che potrebbero rafforzare ulteriormente la lotta contro il cyberbullismo.

### 7.1 Sintesi dei principali risultati

Attraverso l'implementazione e l'analisi di "Human Behavior Variable", sono emersi vari punti di forza e aree di miglioramento. L'approccio combinato ha mostrato:

- **Riconoscimento dei pattern di cyberbullismo:** La libreria ha dimostrato una notevole capacità nel riconoscere schemi linguistici associati al cyberbullismo. Questo è particolarmente rilevante quando si considera la sfida di discernere tra interazioni online innocue e quelle potenzialmente dannose.
- **Versatilità metodologica:** L'integrazione di diversi metodi di NLP ha permesso di catturare un'ampia gamma di comportamenti e intenzioni, dal sarcasmo sottile agli attacchi diretti. Questa combinazione ha reso "Human Behavior Variable" un framework versatile, capace di operare in vari contesti e con diversi tipi di dati testuali.
- **Interattività e adattabilità:** Oltre a rilevare il cyberbullismo, la libreria ha offerto strumenti per interagire con gli utenti, fornendo feedback immediato e aiutandoli a comprendere l'impatto delle loro parole. Questa funzionalità ha il potenziale di educare gli utenti e ridurre comportamenti nocivi.
- **Valore dell'Open Source:** L'approccio open-source ha stimolato la partecipazione attiva della comunità. Molti contributori esterni hanno suggerito miglioramenti e adattamenti, evidenziando la versatilità e la rilevanza della libreria in vari scenari.
- **Implicazioni per la sicurezza online:** L'applicazione pratica di "Human Behavior Variable" in piattaforme e forum online ha rivelato il potenziale della tecnologia nell'assistere moderatori e amministratori nel mantenere un ambiente digitale sano e rispettoso.

## 7.2 Potenziali sviluppi futuri

Il cammino intrapreso con "Human Behavior Variable" rappresenta solo l'inizio di un viaggio alla scoperta delle potenzialità offerte dall'analisi del linguaggio naturale nel contrastare il cyberbullismo. Ecco alcune direzioni promettenti per il futuro sviluppo della libreria e delle sue applicazioni:

- **Collaborazione comunitaria:** Considerando l'approccio open-source di "Human Behavior Variable", ci si aspetta una forte partecipazione della comunità di sviluppatori, ricercatori e professionisti. Questa collaborazione può portare a una rapida evoluzione e adattamento della libreria, sfruttando le diverse competenze e prospettive di ogni contributore. Ciò potrebbe risultare in una maggiore copertura di casi d'uso, un affinamento degli algoritmi e l'inclusione di nuove funzionalità richieste dalla comunità.
- **Applicazione di Diario Personale:** Una potenziale evoluzione potrebbe vedere la creazione di un'applicazione di diario personale che integri "Human Behavior Variable". Questo tool permetterebbe agli utenti di scrivere e riflettere sui propri pensieri e sentimenti, mentre la libreria analizza il testo per fornire feedback. Questa applicazione potrebbe aiutare a identificare tratti o comportamenti di autolesionismo, depressione o ansia, offrendo così un supporto proattivo.
- **Tool per professionisti della salute mentale:** Espandendo l'idea precedente, l'applicazione potrebbe diventare uno strumento prezioso per gli psicologi. Con il consenso del paziente, gli specialisti potrebbero analizzare il contenuto del diario per identificare modelli comportamentali, tratti ricorrenti o tematiche problematiche. La capacità di "Human Behavior Variable" di distinguere tra interazioni "buone" e "cattive" potrebbe offrire agli psicologi una visione più chiara delle dinamiche relazionali dei loro pazienti, facilitando diagnosi e interventi.

Queste sono solo alcune delle potenzialità evoluzioni di "Human Behavior Variable". Con l'evolversi della tecnologia e la crescente consapevolezza dell'importanza della salute mentale nell'era digitale, le opportunità per questa libreria potrebbero estendersi ben oltre quanto immaginato oggi.



# Elenco delle tabelle

5.1	Tabella di Panoramica Generale . . . . .	22
5.2	Tabella di Base NLP . . . . .	23
5.3	Tabella di Analisi Sintattica e Semantica . . . . .	24
5.4	Tabella di Analisi delle emozioni . . . . .	24



# Elenco delle figure

3.1	Immagine raffigurante le Reti Neurali Ricorrenti . . . . .	13
3.2	Immagine raffigurante il clustering . . . . .	13
4.1	Logo Visual Studio Code . . . . .	15
4.2	Logo Python . . . . .	15
5.1	Risultato della traduzione . . . . .	26
5.2	Risultato della Tokenizzazione . . . . .	27
5.3	Risultato del Part Of Speech Tagging . . . . .	28
6.1	Logo GitHub . . . . .	30



# Elenco dei codici

Codice/Traduttore.py . . . . .	26
Codice/Tokenizzazione.py . . . . .	27
Codice/PartOfSpeechTagging.py . . . . .	28



# Bibliografia

- [1] Aruba. *Python per il Natural Language Processing*. <https://www.cloud.it/tutorials/python-language-processing.aspx>. 2021.
- [2] Carmel. *Comprendere il Part-of-speech Tagging*. <https://comeaprire.com/definizioni/comprendere-il-part-of-speech-tagging/>. 2023.
- [3] Developer Resource Center. *Che cos'è TensorFlow: una guida completa*. <https://developer.oracle.com/it/learn/technical-articles/1481879245870-120-what-is-tensorflow>. 2022.
- [4] *Cosa sono le reti neurali ricorrenti?* <https://www.ibm.com/it-it/topics/recurrent-neural-networks>.
- [5] Francesco Gastaldi. «Adolescenti, web e Parole O\_Stili: come educare i ragazzi a comunicare con rispetto». In: *Corriere della sera* (2022).
- [6] *GitHub*. <https://github.com/>.
- [7] *GitHub NLTK*. <https://github.com/nltk/nltk.git>.
- [8] *GitHub spaCy*. <https://github.com/explosion/spaCy.git>.
- [9] *GitHub Text2emotion*. <https://github.com/aman2656/text2emotion-library.git>.
- [10] *GitHub TextBlob*. <https://github.com/sloria/TextBlob.git>.
- [11] *GitHub vaderSentiment*. <https://github.com/cjhutto/vaderSentiment.git>.
- [12] Alessandro Grossi. *Keras, cos'è, a cosa serve, come si usa nel deep learning*. <https://www.ai4business.it/intelligenza-artificiale/keras-cose-a-cosa-serve-come-si-usa-nel-deep-learning/>. 2020.
- [13] Harper. *"Bully"*. Online Etymology Dictionary, 2001.
- [14] Redazione Osservatori Digital Innovation. «Natural Language Processing (NLP): come funziona l'elaborazione del linguaggio naturale». In: (2023).
- [15] *La cassetta degli attrezzi NLP: Analisi del sentimento per semplificare il feedback dei clienti*. <https://sharethis.com/it/data-topics/2022/08/nlp-sentiment-analysis/>.
- [16] Giuseppe Maggi. *Matplotlib*. <https://www.html.it/pag/404701/matplotlib/>. 2010.
- [17] Giuseppe Maggi. *Seaborn*. <https://www.html.it/pag/405220/seaborn/>. 2021.
- [18] Università e Ricerca (MIUR) Ministero dell'Istruzione. *Bullismo e cyberbullismo*. <https://www.miur.gov.it/bullismo-e-cyberbullismo>. 2023.

- [19] Giulia Moretti. «“Il bullismo non è una ragazzata”. Il vademecum dell’esperta per un ritorno a scuola in sicurezza». In: *la Repubblica* (2023).
- [20] Luca Naso. *Lemmatization – NLP di base in Python*. <https://lucanaso.it/2021/01/11/lemmatization-nlp-di-base-in-python-3-di-9/>. 2021.
- [21] *Natural Language Processing (NLP), i modelli per l’elaborazione del linguaggio naturale*. <https://tech4future.info/modelli-nlp-natural-language-processing/>.
- [22] Giovanni Innamorati e Cecilia Ferrara Paola Tamborlini. «L’allarme cyberbullismo, vittime e bulli sono sempre più piccoli». In: *ANSA* (2023).
- [23] pulplearning. *Named Entity Recognition (NER)*. <https://pulplearning.altervista.org/named-entity-recognition-ner/>. 2022.
- [24] pulplearning. *Scikit-learn: librerie Python per il Machine Learning*. <https://pulplearning.altervista.org/librerie-python-per-il-machine-learning-scikit-learn/>. 2021.
- [25] *Python*. <https://www.python.org/>.
- [26] Federico Trotta. *L’analisi dei dati resa semplice: padroneggia Pandas per ottenere risultati dai tuoi dati*. <https://www.codemotion.com/magazine/it/data-science-it/analisi-dei-dati-pandas/>. 2023.
- [27] *Visual Studio Code*. <https://code.visualstudio.com/docs>.
- [28] Webmaster. *NLTK (Native Language Tool Kit) la libreria Python per il Language Processing and Analysis*. <https://www.meccanismocomplesso.org/nltk-native-language-tool-kit-la-libreria-python-per-il-language-processing-and-analysis/>. 2020.
- [29] *Word2vec: modelli addestrabili per la rappresentazione distribuita delle parole*. <https://enricogiannini.com/8/word2vec-modelli-addestrabili-per-la-rappresentazione-distribuita-delle-parole/>. 2019.