



UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Tesi di Laurea in Ingegneria Informatica

**Esplorazione dello stato mentale tramite
tecniche di text mining nelle community
online**

Relatore:

Dott. Francesco Pistolesi

Candidato:

Lorenzo Valtriani

ANNO ACCADEMICO 2023/2024

Indice

1	Introduzione	5
1.1	Disturbi depressivi	5
1.2	Lo scopo dello studio	6
2	Background	9
2.1	Web Scraping	9
2.2	Text Mining	9
2.2.1	Fase di pre-elaborazione	9
2.2.2	Tecniche di Text Mining	10
3	Dati	11
3.1	La comunità di BeyondBlue	11
3.2	Discussioni, post e risposte	12
3.3	Il database	12
3.4	Ottenere i dati	13
3.5	I risultati dello scraping	14
4	Metodo	15
4.1	Generazione delle parole target	15
4.2	Fase di Text Mining	16
4.3	Data Visualization	18
5	Discussione dei risultati	21
5.1	Analisi del primo utente: Daniel12	21
5.2	Analisi del secondo utente: Mares73	23
6	Conclusioni	29
A	Appendice	31
A.1	Script per la fase di scraping	31
A.2	Script per la fase generazione dei punteggi	32
A.2.1	Classe Analizzatore	32
A.2.2	Main	34

Capitolo 1

Introduzione

1.1 Disturbi depressivi

Secondo il Manuale Diagnostico e Statistico dei Disordini Mentali [1], i disturbi depressivi rappresentano un gruppo di condizioni mentali caratterizzate da persistente tristezza, perdita di interesse per le attività quotidiane e una serie di sintomi fisici, cognitivi ed emotivi che compromettono il funzionamento personale e sociale. Questi disturbi, che vanno oltre la normale tristezza, influenzano profondamente la qualità della vita e richiedono un trattamento adeguato.

Secondo i dati dell'Organizzazione Mondiale della Sanità (2023), circa il 5% della popolazione mondiale adulta soffre di disturbi depressivi, con una prevalenza tra le donne rispetto agli uomini. Inoltre, le persone con disturbi depressivi maggiori o persistenti presentano un rischio significativamente più alto di suicidio, con stime che indicano che *circa il 60% dei suicidi sia legato a diagnosi di disturbi dell'umore*. Tra i disturbi depressivi principali si includono:

- **Disturbo depressivo maggiore (MDD)**: caratterizzato da episodi di depressione grave, perdita di interesse, alterazioni del sonno, dell'appetito e del funzionamento cognitivo.
- **Disturbo depressivo persistente (PDD)**: noto anche come distimia, presenta sintomi meno intensi rispetto all'MDD, ma persistenti nel tempo (almeno due anni).
- **Disturbo disforico premestruale**: colpisce il 3-9% delle donne in età fertile, con sintomi psicologici invalidanti.
- **Disturbo depressivo indotto da sostanze/farmaci**: causato dall'uso di sostanze o farmaci psicoattivi.

Questa panoramica introduttiva sui disturbi depressivi permette di inquadrare il contesto dello studio oggetto di questa tesi, volto a monitorare lo stato mentale degli utenti di una community online, utilizzando tecniche di text mining per prevenire

i rischi legati a disturbi mentali.

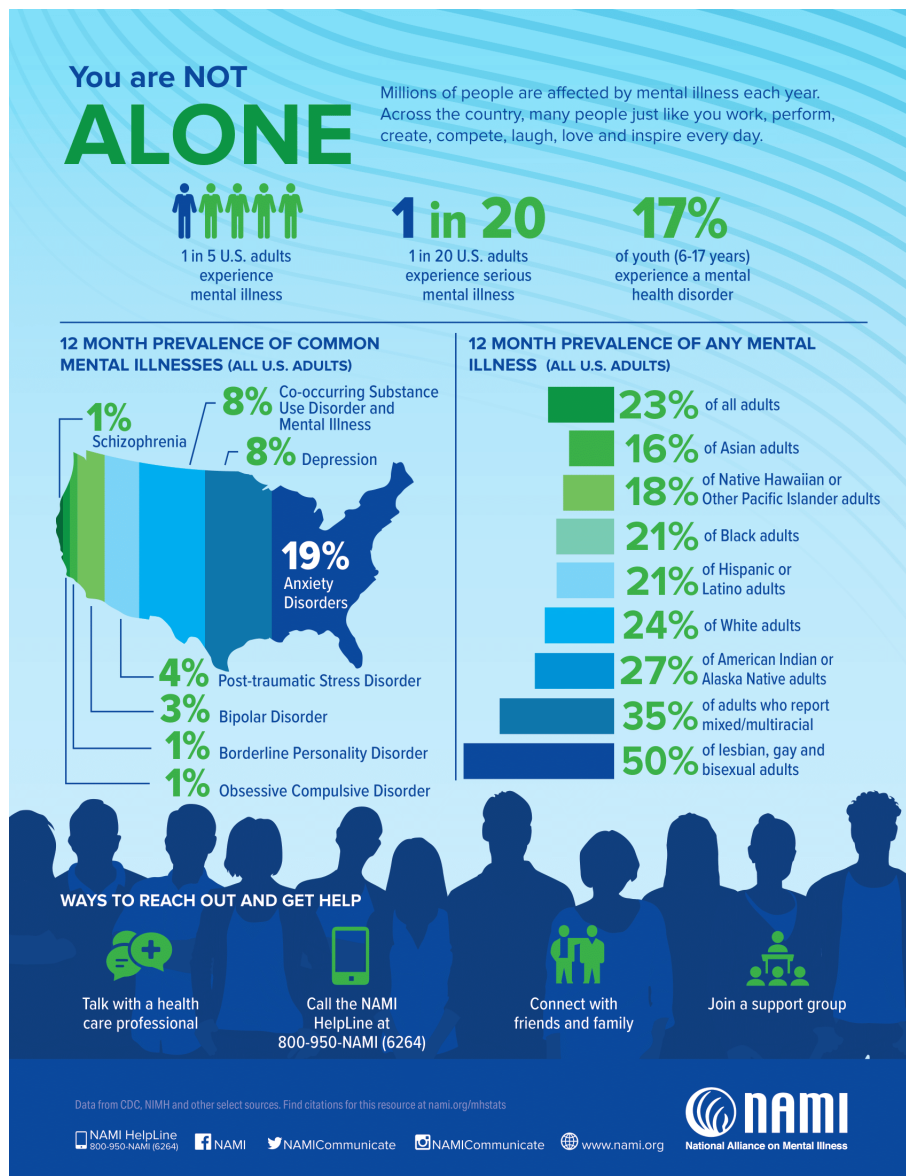


Figura 1.1: Come notiamo dall'infografica sovrastante, in America, la depressione è la malattia mentale più diffusa negli adulti, seconda solo all'ansia; da notare poi che all'interno della comunità LGBTQ+, una persona su due soffre di un disturbo mentale [4]

1.2 Lo scopo dello studio

Lo scopo dello studio è realizzare uno strumento per il monitoraggio dello stato mentale degli utenti di una community online denominata BeyondBlue [2], la quale si prefigge l'obiettivo di combattere lo stigma legato alla salute mentale. All'interno di questa comunità, si sviluppano conversazioni tra utenti che affrontano quotidianamente problematiche legate a vari disturbi mentali, quali ansia e depressione.

Il fulcro della ricerca consiste nell'analizzare i post pubblicati dagli utenti, applicando tecniche di text mining per studiare i termini utilizzati. L'obiettivo è assegnare un punteggio a ciascun post in base al livello di rischio potenziale che l'utente rappresenta per sé stesso, partendo dalle parole che usa. Attraverso tale approccio, si intende costruire un profilo dell'andamento dello stato mentale dell'utente nel tempo. Questo monitoraggio potrebbe consentire di individuare tempestivamente situazioni critiche in cui il disturbo mentale supera una determinata soglia critica, consentendo così di intervenire velocemente e in modo mirato.

L'utilizzo di tecniche avanzate di analisi del linguaggio e l'assegnazione di punteggi di rischio costituiscono un approccio innovativo di sfruttare le potenzialità del text mining per la prevenzione e il supporto nel contesto della salute mentale.

Capitolo 2

Background

Questa sezione illustra le principali tecniche applicate nello studio: web scraping e text mining.

2.1 Web Scraping

Il web scraping è una tecnica usata per estrarre dati da siti web in modo automatizzato. I dati estratti dal web scraping possono essere immagini, video, o testo, come nel caso dell'applicazione oggetto di questo studio. Una volta estratti, i dati sono memorizzati in un database, per essere più facilmente utilizzati in seguito.

2.2 Text Mining

Il text mining è il processo di estrazione di informazioni utili e significative da dati testuali non strutturati. Poiché circa l'80% dei dati nel mondo ha un formato non strutturato, il text mining è una pratica estremamente preziosa. Questo processo si basa su tecniche di elaborazione del linguaggio naturale (NLP), che permettono ai computer di comprendere e interpretare il linguaggio umano dopo aver effettuato una pre-elaborazione del testo [3].

2.2.1 Fase di pre-elaborazione

La pre-elaborazione dei testi nel text mining è infatti fondamentale per preparare i dati testuali prima che possano essere analizzati. Questa fase preliminare è composta dalla **rimozione delle stop words**: articoli, congiunzioni o altre parole comuni che però non aggiungono valore semantico, **tokenizzazione**: divisione del testo in unità più piccole chiamate token, per esempio parole, **lemmatizzazione**: riduzione di una parola alla sua forma base o lemma, per esempio *studio* diventa *studiare*, **lowercasing**: tutte le lettere vengono convertite in minuscolo per facilitare il confronto fra termini, **rimozione di punteggiatura e simboli**: perché non sono portatori di informazione rilevante.

2.2.2 Tecniche di Text Mining

Una volta completata la pre-elaborazione, si susseguono diverse fasi che dipendono dal tipo di applicativo di cui abbiamo bisogno:

- **Tecniche di estrazione di informazioni**

Se ci sono alcune parole target che ci interessano, in questo caso potrebbe essere di nostro interesse controllare quali termini gli utenti utilizzano e se questi possono essere ricondotti al disturbo depressivo.

- **Tecniche di analisi del sentiment**

Ci sono poi tecniche di analisi del sentiment il cui obiettivo è determinare il tono emotivo del testo (positivo o negativo). Per esempio, controllando se una parola è seguita da una negazione che ribalta completamente la semantica della frase, oppure avverbi come “molto” o “poco” che modificano significativamente il significato della frase. Nel nostro caso di studio, utilizziamo questa tecnica per determinare quanto una parola sia l’espressione di un determinato livello di gravità del sentimento depressivo.

Capitolo 3

Dati

3.1 La comunità di BeyondBlue

BeyondBlue è una comunità online australiana che si dedica al supporto e alla sensibilizzazione dei temi della salute mentale, dando particolare importanza in particolare modo la depressione, l'ansia e altri disturbi correlati.

Attraverso discussioni dedicate, gli utenti della community possono condividere il loro stato emotivo e le loro esperienze, trovando supporto e persone con cui confrontarsi.

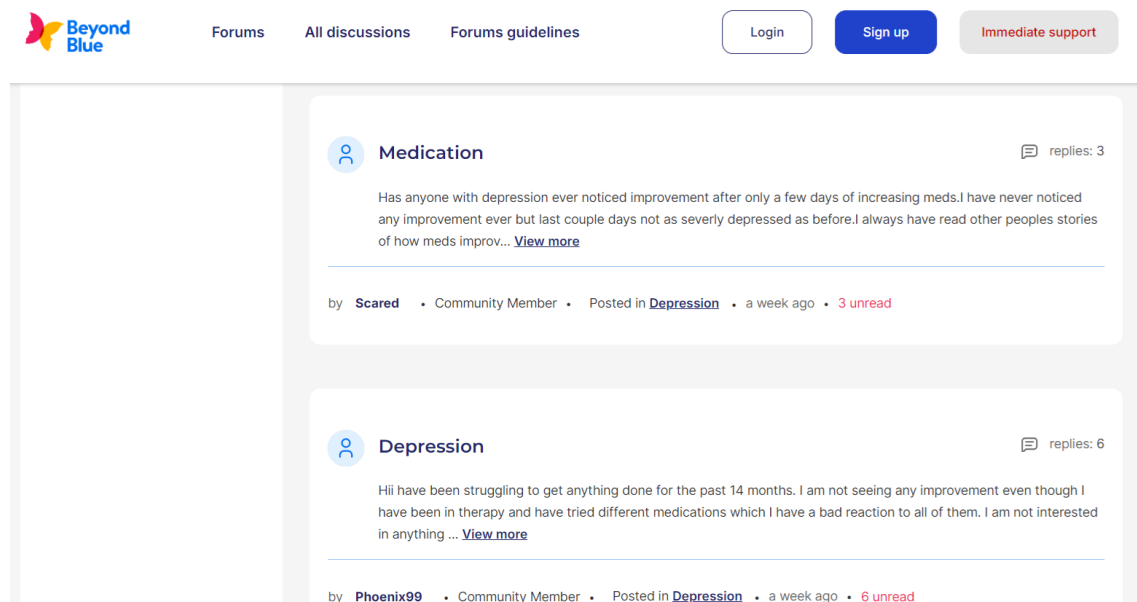


Figura 3.1: Esempio di due discussioni dalla community Beyond Blue in cui notiamo la possibilità di fornire risposte attraverso la pubblicazione di post relativi ad una specifica discussione.

3.2 Discussioni, post e risposte

La prima fase del progetto è consistita nell'ottenere i dati da analizzare. Per ogni discussione, sono stati analizzati tutti i post con un potenziale valore nell'analisi emotiva dell'utente che li ha scritti.

Considerata una discussione, non sono stati memorizzati tutti i suoi post, ma solo quelli scritti dall'autore della discussione, in quanto utili ad analizzare l'evoluzione del suo stato emotivo, tralasciando quindi le risposte di altri utenti. Sebbene le risposte contengano informazioni utili, non costituiscono valore aggiunto per lo scopo di questo lavoro di tesi.

3.3 Il database

Per memorizzare i dati è stato progettato un database realizzato secondo il modello relazionale, caratterizzato dal seguente diagramma E-R:

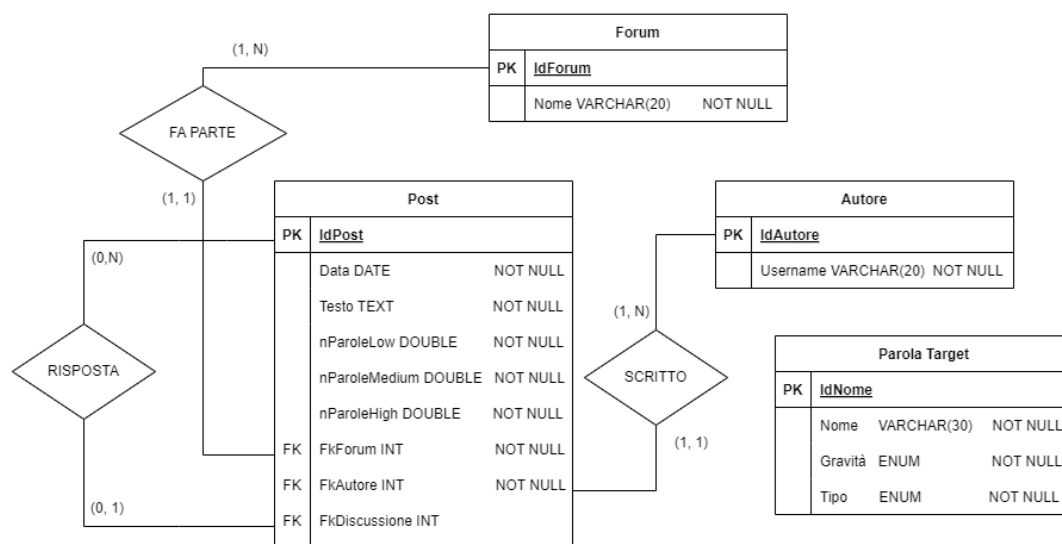


Figura 3.2: Diagramma E-R del database realizzato.

Procediamo con la spiegazione di questo:

- **Forum:** Questa entità rappresenta il forum in cui gli utenti pubblicano i propri post, nel nostro studio è stato esaminato quello relativo al tema della depressione. Un forum può contenere molteplici post.
- **Autore:** Questa entità rappresenta l'autore del post, il quale può scrivere più post nel tempo.
- **Post:** Questa entità rappresenta il post che un utente ha scritto. Il post può aprire una discussione oppure può essere una risposta a una di queste (in questo caso la chiave esterna `FkDiscussione` si riferirà al post che apre quella

discussione. Un post può essere la risposta di un solo post, mentre un post di avvio discussione può avere anche più post in risposta. Altri attributi da nominare sono: `nParoleLow`, `nParoleMedium`, `nParoleHigh`. Questi tre attributi conterranno il numero di parole target che il post contiene, quelle parole che descriveranno l'andamento dello stato emotivo dell'autore del post. **Parole Target**: Questa entità non è relazionata a nessun'altra e contiene la parole target che andremo a ricercare all'interno dei vari post. Queste parole sono un allarme rispetto al fatto che l'utente soffra di depressione e servono a valutare il livello di gravità del suo disturbo. Infatti, a ogni parola è associato un livello di gravità (basso, medio, alto), indicato nell'attributo **Gravità**. L'attributo **Tipo** invece indica la categoria sintattica di cui fa parte la parola (nome, aggettivo, avverbio), oppure se questa è formata da più termini come nel caso dei modi di dire; in questo caso l'attributo conterrà ('multipla').

3.4 Ottenere i dati

Questa sezione descrive come sono stati ottenuti i dati. Come descritto nel capitolo precedente, la tecnica che si utilizza per ottenere dati da un sito web si chiama web scraping. In questo progetto è stata implementata in linguaggio PHP.

L'idea alla base del web scraping è ottenere il codice sorgente della pagina web che vogliamo "scrapare", a partire dal suo URL, e andare a ottenere tutte quelle informazioni che ci interessano, nel nostro caso i post target. Il programma di scraping partirà dalla prima pagina delle discussioni relative al forum sulla depressione. Nella pagina, prenderà in esame tutte le discussioni sotto "All discussions".

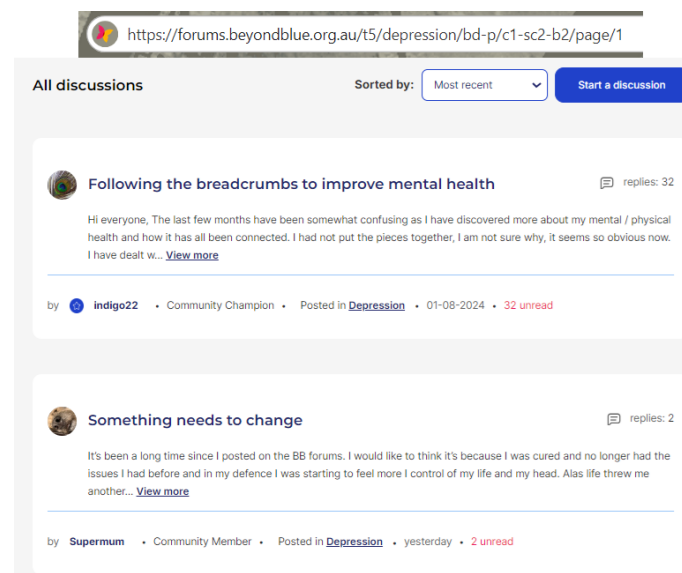


Figura 3.3: Porzione della pagina con URL sovrastante da cui inizia il programma di scraping.

Per ogni discussione, otteniamo l'URL della relativa pagina per ricavare informazioni su tutti i post. Per ogni post pubblicato dall'utente che ha dato inizio alla

discussione, andiamo a scrapare la singola discussione con l'obiettivo di ricavare i seguenti dati:

il testo del post: che contiene l'informazione principale a cui facciamo riferimento per analizzare l'andamento dello stato mentale dell'utente; **la data di pubblicazione:** ci serve poiché l'andamento dello stato mentale dovrà essere temporale; **lo username dell'autore del post.**

Questo procedimento verrà effettuato per ogni discussione della pagina, e quando la pagina sarà finita, si passerà alla successiva.

3.5 I risultati dello scraping

Al termine della sua esecuzione, lo scraper ha memorizzato nel database circa 17.800 post relativi al tema della depressione, su un totale di 46.653 post presenti. Questo implica che i restanti post, che rappresentano meno di due terzi del totale, sono risposte di supporto indirizzate a coloro che hanno avviato discussioni nel forum.

IdPost	Data	Testo	FkAutore	FkForum	FkDiscussio
1	2024-09-12	Hi I am really struggling today everything is getting on top of me financially under pressure hardly eating lack of work housing uncertainty no family talk to etc	1	1	1
2	2024-09-13	Im not doing well. My meds are not working and Im at times worried i am going out of my mind. I pathetically check beyond blue forums all day waiting for a notification. Anything to sto...	2	1	1
3	2024-09-13	This depression is like a nightmare I cant wake from.	2	1	2
4	2024-09-13	Replying to posts at first was automatic with no particular motive accept to offer some assistance. Then my selfesteem somehow got involved in which my reply had to be the only one t...	2	1	2
5	2024-08-01	Hi everyone, The last few months have been somewhat confusing as I have discovered more about my mental / physical health and how it has all been connected. I had not put the pi...	3	1	2
6	2024-08-02	Hi ER, Thank you for your kind words of encouragement. I worked out when I was going through talk therapy how, why and when the dysthymia and major depression began, these ...	3	1	5
7	2024-08-03	Hi ER, I am really glad to hear you are getting better at feeling others negative energy. That is an intuitive sense I have always had, but not always listened to, but have learned, like...	3	1	5
8	2024-08-04	Hi charlebrown, Thank you for your response and encouragement, it sounds like you have some experience in these matters yourself. Eagle Ray and I have been communicating on d...	3	1	5
9	2024-08-05	Hi therising, I hope you have been keeping well. I'm not sure if you have read through the entire thread so far but I have no problem feeling other peoples energy, it's very instructu...	3	1	5
11	2024-08-05	Hi therising, I totally get the memory thing, I think it comes with the territory, I have blocks of memory missing from my younger years. I assume you know when I say spirituality, tha...	3	1	5
12	2024-08-09	Hi ER, I have not replied to your latest thread because it is quite a specific subject and I want to allow someone a chance to respond who may have had that type of experience. I am ...	3	1	5
13	2024-08-12	Hi ER, I reported the missing post over the weekend and it has shown up now. Wow, you have made some seriously good progress recently. You are so much more sure of yourself a...	3	1	5
14	2024-08-31	Hi ER, I just wanted to check in with you to see how things are going. I have been a bit busy answering posts and realised we haven't connected for a couple of weeks. Are you still...	3	1	5
16	2024-09-13	Hi ER, Sorry for the delay in getting back to you. I've been like your situation is leading the way with the healing work you are doing at the moment. You are making such a lot of progr...	2	1	6

Figura 3.4: Immagine di alcuni post memorizzati all'interno del database.

Capitolo 4

Metodo

4.1 Generazione delle parole target

Questa fase dello studio ha costituito un passaggio cruciale, il cui obiettivo è stato ottenere un dataset composto da parole, quali nomi e aggettivi, da associare semanticamente al disturbo della depressione. A ciascuna di queste parole è stato inoltre attribuito un livello di gravità, corrispondente alla severità del disturbo manifestato dall'utente. È quindi essenziale assegnare accuratamente a ogni termine il relativo livello di gravità.

La lista di nomi e aggettivi è stata generata utilizzando modelli generativi, come ChatGPT, per semplificare il processo. Il prompt utilizzato richiedeva la creazione di termini che potessero essere impiegati da un individuo per descrivere la propria esperienza con la depressione. Successivamente, è stato richiesto di ampliare la lista includendo parole comunemente utilizzate dagli adolescenti e nei social network. Le parole sono poi state classificate in tre livelli di gravità (basso, medio, alto), e per ciascuna di esse sono stati generati i sinonimi.

Inoltre, sono stati aggiunti al dataset anche alcuni avverbi di quantità che, qualora presenti nel testo in riferimento a un aggettivo target o a un verbo derivato da un nome target, modificano il livello di gravità di quest'ultimo. Un esempio è rappresentato dall'espressione "very depressive", che aumenta la gravità del termine "depressive".

IdNome	Nome	Gravita	Tipo
165	crisis	high	nome
69	cry	low	nome
185	cut	high	nome
241	damage	high	aggettivo
223	dark	medium	aggettivo
140	darkness	medium	nome
11	dead	high	nome
129	deadness	medium	nome
189	death	high	nome
173	defeat	high	nome
252	defenseless	low	aggettivo
160	defenselessness	medium	nome
100	delay	low	nome
209	delicate	low	aggettivo
174	denial	high	nome
192	departure	high	nome
179	dependence	medium	nome
195	deplete	high	aggettivo
371	depression	medium	nome
171	deprivation	high	nome
322	desolate	high	aggettivo

Figura 4.1: Immagine di alcune parole target che sono state generate.

4.2 Fase di Text Mining

Questa sezione descrive la fase che ha avuto lo scopo di assegnare punteggi ad ogni post analizzato, in funzione delle parole che si trovano all'interno.

L'obiettivo è contare il numero di parole target all'interno di ogni testo, per ogni livello di gravità.

Questo processo è stato implementato ed eseguito utilizzando il linguaggio Python: di seguito il codice del main.

```

1 from classi.Analizzatore import Analizzatore
2
3 # Diamo il valore alle gravit per gli avverbi di quantità passandole come parametri
4 an = Analizzatore(0.5, 2)
5
6 # Vengono inizializzati i dizionari dei nomi, aggettivi e avverbi e le loro gravit
7 an.inizializzaParoleTarget()
8
9 # Adesso cominciamo ad analizzare tutti i post degli utenti, utente per utente.
10 # Dando ad ogni post un punteggio relativo alla sua gravit in relazione all'utente che lo ha pubblicato
11 an.analizzaPosts()

```

Il codice utilizza la classe Analizzatore ([A.2.1](#)), come si osserva alla quarta riga, dove viene istanziata una variabile di tale classe. Al costruttore della classe sono passati due parametri che definiscono le modalità di modifica del valore associato a una parola target, applicate agli avverbi che indicano una quantità ridotta o elevata. La classe utilizza due librerie: NLKT, per scaricare il pacchetto delle stopwords, e spaCy, per impiegare un modello di elaborazione del linguaggio naturale.

Il costruttore dichiara le strutture dati necessarie, come i dizionari per i nomi, aggettivi, avverbi, verbi e parole multiple (come per esempio modi di dire composti da più parole) target. Questi dizionari associano ad ogni termine il relativo livello

di gravità (basso, medio, alto). Contiene anche il dizionario per i modificatori degli avverbi che contiene due chiavi:

- **low**: indica quanto considerare una parola preceduta da un avverbio di quantità ridotta, nell'esempio considerando che viene passato 0.5 come parametro, verrà considerata la metà.
- **high**: indica quanto considerare una parola preceduta da un avverbio di quantità aumentata, nell'esempio considerando che viene passato 2 come parametro, verrà considerata il doppio.

Infine, alla fine del costruttore, si trova anche la definizione di un ulteriore dizionario che conterrà per ogni livello di gravità il numero di parole target trovate all'interno di un singolo post.

Alla riga 7, viene invocato il metodo `inizializzaParoleTarget`, il cui scopo è inizializzare i dizionari dichiarati nel costruttore con le parole target (nomi, aggettivi, avverbi, e parole multiple) prelevate dalla tabella specifica del database. Il dizionario dei verbi non viene inizializzato poiché i verbi non sono stati inseriti nel database.

Infine, alla riga 11, viene invocato il metodo `analizzaPosts`, il quale esamina, per ogni utente, tutti i post da lui scritti in ordine cronologico, contando per ciascuno le parole target che contiene. I primi termini che vengono cercati sono le parole multiple, le quali, se vengono trovate, oltre ad essere contate, saranno cancellate dal testo, così che le singole parole che le formano non siano poi contate due volte nella ricerca successiva.

Prima di ricercare all'interno del testo le singole parole target, ogni post viene sottoposto a una fase di pre-elaborazione, durante la quale la punteggiatura viene rimossa, le stop words eliminate, e le parole sottoposte a lemmatizzazione.

Successivamente, il post viene tokenizzato, ciascun token viene analizzato per estrarne la relativa categoria grammaticale (nome, aggettivo, avverbio, verbo). Per ogni categoria grammaticale verranno svolte operazioni differenti:

- **nome/aggettivo**: si verifica la sua presenza nel dizionario delle parole target e, se presente, viene considerato aggiornando la relativa entrata dell'apposito dizionario.
- **verbo**: se il dizionario dei verbi è vuoto, viene verificato se il termine può essere derivato da un nome presente nel dizionario corrispondente. In tal caso, il termine viene inserito nel dizionario dei verbi, così che nelle analisi successive verrà riconosciuto subito come target; se il verbo non è target, viene inserito in un dizionario apposito per i verbi non target, così che non sia necessario confrontarlo nuovamente con tutti i nomi target nelle analisi future, ottimizzando i tempi di esecuzione.
- **avverbio**: la libreria `spaCy` consente di ottenere il termine associato a quell'avverbio nel contesto del post. Questo termine, che può essere un aggettivo o un

verbo, viene inserito in un dizionario che associa a ciascun termine il modificatore corrispondente all'avverbio. In questo modo, se il termine è un verbo o un aggettivo, prima di considerarlo nel contatore, se risulta target, si verificherà la sua presenza anche in questo ulteriore dizionario applicando quindi se serve il modificatore adatto.

Una volta analizzato tutto il post, il numero di parole trovate per ogni livello di gravità viene memorizzato nel database, all'interno della tabella dei post.

La figura seguente mostra un testo estratto dalla community online, a cui facciamo riferimento per questo studio, con le parole target evidenziate con un colore diverso in funzione della gravità. Il giallo usato per la gravità bassa, arancione per quella media, e rosso per quella alta.

Thank you so much for replying, I was really **unsure** about posting, cause it makes my **depression** feel so much more real. I have been living in **denial** of these feelings for so long now, trying to just fake happiness for the sake of my husband and kids. I am failing I just can't keep faking. (...) We have two children a 13yr old and a 10yr old. I am 34. My kids now being older are being affected by my **depression**, they don't understand why I don't do anything. I don't know what to tell them that will help them to understand. I feel **defeated**. I have had to be so strong and look after myself from such a young age and I just can't do it anymore. Honestly I would really like to just slip away to whatever comes after this life. I feel it would be so much better for them to not have to witness this disease **defeating** me. I have been taking AD for most of my life, I have swapped a few times when I felt there was no improvement. I practice meditation and hypnosis daily, and it helps in the moment. I honestly just feel so **empty** all of the time, nothing I do shifts the uneasiness and **weight** I feel crushing my spirit. My husband tries to be understanding but quickly loses his temper with me, I don't blame him though it must be so awful being around me all of the time. I know that I need help, I **fear** that I won't make it if I don't seek help again. I just don't know where to turn as every doctor or counsellor I have seen always blames my childhood for my **depression** and yes I had an awful up bringing my life was full of **violence**, hatred, abuse, alcohol and drugs. I just don't think that's my problem. I want to get better, I just don't know how? Everything that I have tried has failed. So now, I don't go out anywhere, I don't speak about it, I have no friends or family. I stay within the **confines** of my house, my prison. I'm **exhausted**, I understand that I must make the move to seek help, but it just seems like the effort required is so much more than I am.

Figura 4.2: Testo di un utente della community con parole target evidenziate per gravità.

4.3 Data Visualization

Questa fase ha avuto l'obiettivo di rappresentare i dati raccolti attraverso grafici intuitivi, al fine di agevolare la comprensione dei risultati ottenuti. In particolare, miriamo a visualizzare la variazione nel tempo del numero di parole target digitate in ciascun post degli utenti, suddivise per livello di gravità del disturbo.

La visualizzazione scelta è una heatmap suddivisa in tre righe, ciascuna corrispon-

dente a un livello di gravità (basso, medio e alto), mentre le colonne rappresentano i diversi post. Il colore di ciascuna cella riflette il numero di parole target presenti nel post relativo, con colori più freddi che indicano una minore presenza di parole target e colori più caldi che indicano un maggior numero di tali parole. Inoltre, il numero di parole necessario affinché una cella assuma un colore più caldo varia a seconda del livello di gravità della riga corrispondente.

Le regole adottate per la rappresentazione dei valori nelle celle sono le seguenti:

- **Prima riga:** rappresenta la numerosità delle parole di gravità bassa. Il valore mappato nel colore è il numero di parole di gravità bassa diviso per 3.
- **Seconda riga:** rappresenta la numerosità delle parole di gravità media. Il valore mappato nel colore è il numero di parole di gravità media diviso per 2.
- **Terza riga:** rappresenta la numerosità delle parole di gravità alta. Il valore mappato nel colore corrisponde esattamente al numero di parole di gravità alta.

Questo differente schema di pesatura è stato pensato poiché lo stesso numero di parole target associate a un livello di gravità alto rappresenta una maggiore criticità rispetto a un uguale numero di parole target di gravità media o bassa.



Figura 4.3: Esempio di heatmap prodotta.

Lo scopo di questa rappresentazione è identificare eventuali pattern che possano indicare una fase di peggioramento o miglioramento del disturbo depressivo degli utenti nel tempo.

Capitolo 5

Discussione dei risultati

Questo capitolo contiene una discussione dei risultati dello studio effettuata analizzando i pattern di alcune heatmap ritenute più interessanti. Le heatmap sono state confrontate con i corrispondenti post, analizzando se il pattern di colore possa rappresentare correttamente la dinamica dello stato emotivo dell'utente.

In particolare, a titolo di esempio, sono state analizzate due specifiche heatmap, di due utenti diversi che si sono contraddistinti per il loro elevato numero di post e per la loro alta attività all'interno della community.

5.1 Analisi del primo utente: Daniel12

Il primo utente corrisponde al nickname **Daniel12** all'interno della community. L'utente, come racconta nei suoi 265 post, soffre di depressione, ansia e gravi problemi di autostima. Le colonne evidenziate nella heatmap in Fig. 5.1 fanno riferimento ad alcuni post scritti dall'utente in un arco temporale che va dal 26 Giugno 2022 (per il primo post preso in esame) fino al 6 Luglio 2022, quindi l'utente fa passare alcuni giorni prima di scrivere un nuovo post.

Analizzando visivamente le colonne evidenziate, possiamo notare come, nei primi giorni, lo stato dell'utente sembrerebbe peggiorare dato che i colori della prima colonna si spostano su colori ancora più caldi come il viola, relativo alle parole di gravità alta. Nei giorni ancora successivi la situazione sembra invece opposta, infatti si può notare come i colori mutino verso toni più freddi: l'utente arriva persino a non utilizzare alcun termine di gravità alta.



Figura 5.1: Heatmap di Daniel12

Per confermare questa situazione, sono stati letti direttamente i post relativi alle colonne, così da controllare se effettivamente vi fosse un riscontro tra i colori visua-

lizzati sulla heatmap e il contenuto dei testi.

Hi therising, I think I need to, with help from the psychologist and those around me, understand who I am better and develop some comfort in who I am in relation to what you say because at the moment there is none of it at all and hasn't been for a very long time even if I have been able to manage it better in the past. I let different aspects of myself get way out of control for sure, for example the **anxious** side has filled me with **regret** that I haven't backed myself to feel comfortable in myself to trust myself to be "normal" around others and in relationships, some people say the anxiousness in relationships is probably because that person isn't the right person and you know it deep down but I don't feel that way. I feel that it's just me, it's part of who I am to be **fearful** to let myself fully go and not **doubt** who I am or what I can offer. I've retreated inside myself more and more and I don't feel like even trying to get out of myself anymore, i really do see no light at the **end** of the tunnel at the **end** of the tunnel at the moment. Not sure if you had similar experiences but I constantly feel like I am getting the life sucked out of me, like a balloon deflating. Any little thing triggers this as well at the moment and I actually feel a physical unwillingness to get out of it, like mentally I want to try and pep myself up but physically I can't do it. Feels like event attempting it is like moving a mountain at the moment.

Figura 5.2: Post del 26 Luglio 2022 di Daniel12

Il post in Fig. 5.2 riflette un evidente stato di profondo disorientamento emotivo e frustrazione personale. L'utente esprime una difficoltà nel sentirsi a proprio agio con se stesso e sente che vari aspetti della sua personalità (in particolare l'ansia) sono diventati incontrollabili. Questa ansia gli causa rimpianti e lo porta a non avere fiducia nelle proprie capacità, soprattutto nelle relazioni interpersonali. Nel tempo, si è chiuso sempre di più in se stesso e ora fatica persino a voler uscire da questo stato, arrivando a non vedere alcuna luce in fondo al tunnel: "*do see no light at the end of the tunnel*".

Nel post successivo, guardando ai colori, l'utente mostra un peggioramento del suo stato emotivo. Proseguendo con l'analisi del relativo post, troviamo il seguente testo:

Hi therising. Sorry for not getting back to you earlier *it's been a whirlwind week*, just feel like I've had very little effort to even drag myself out of bed feels like my spirit so to speak is so low it's hard to really move past any of it. (...) For me it's a case of holding myself back because I'm **afraid** of being ultimately **rejected** which probably stems from severe self esteem issues, so I'm effect my comfort and peace comes from holding back and not showing that extra 10-20% of myself so as not to be exposed. Does that make sense at all? (...) It feels like a huge relief to me, one because I don't need to go and then feel like everyone around me is better than me or more relaxed and fun and two it eliminates my **fear** of being judged so I'm comfortable in staying home even though staying home means also feeling this **dark depression**. I don't know if you can relate or whether this sounds silly but it feels like at the moment if I go out, whether that's to the office for work or out on weekend, it slowly makes me feel worse because I feel like I'm reminded by people around (not by what they say but just observing everyone) how **bad** I am and how unflattering or not relaxed as a person I am. Then on the flip side of that if I stay home I can't **escape** from feeling like I'm in a bottomless pit of **depression** and constantly replaying past events in my head or **negative** dialogue as you've mentioned before, is this normal? I can't even function anywhere properly it feels. The only **small** respite I am getting is from reading at the moment, which I picked up more recently to try and **escape** my head. It works most of the time but sometimes I read a chapter and I don't even know what I've read because I haven't actually absorbed any of it. (...)

Figura 5.3: Post del 1° Luglio 2022 di Daniel12

Come è possibile notare, il post mostra un chiaro peggioramento dello stato emotivo dell'utente rispetto al precedente. L'utente descrive una settimana estremamente difficile, "*it's been a whirlwind week*", in cui la mancanza di energia è diventata così pronunciata da rendere difficile persino alzarsi dal letto. Il senso di depressione sembra aver raggiunto livelli più profondi, con una sensazione di "spirito" molto basso, bloccato in un ciclo di immobilità. Notiamo quindi che l'aumento delle parole target trovate all'interno del testo, in questo caso, implichi anche un peggioramento dello stato emotivo.

Dopo alcuni giorni, precisamente cinque, l'utente torna attivo sulla community scrivendo un ulteriore post, che dal colore della colonna sembrerebbe annunciare un buon miglioramento rispetto al post del 1° Luglio. Quest'ultimo post era caratterizzato dal seguente contenuto:

Hi Hanna, Hope you're well and thanks for both posts and the video recommendation. I watched it and it does make sense to me also....I am trying to start doing some sort of housekeeping thinks each day/throughout the week to try help. Finding things to calm me down has been a **struggle** but I guess I need to start **small**, sounds silly but for some reason detailing and cleaning my car seems to do this for a short period. I mentioned to my psychologist that the inner voice suffocates me and we are working on where these voices come from and what experiences have caused this sort of downward spiral I feel like I am in. Thanks again for all your continued advice and support I genuinely appreciate everything you've said to me!

Figura 5.4: Post del 6 Luglio 2022 di Daniel12

La caratteristica più evidente è senz'altro il numero esiguo di parole target trovate all'interno, infatti ne figurano solamente due, nessuna di queste di gravità alta. In questo post, l'utente mostra segni di miglioramento rispetto ai precedenti. Nonostante continui a riconoscere le sue difficoltà emotive, come il senso di soffocamento causato dalla sua "voce interiore", l'utente sta attivamente cercando modi per affrontare il proprio stato mentale. Ha iniziato a svolgere piccole attività quotidiane, come la pulizia della sua auto, per cercare di ritrovare un equilibrio e rilassarsi.

5.2 Analisi del secondo utente: Mares73

Adesso passiamo a un altro utente che è stato ritenuto interessante per la sua situazione costantemente grave: **Mares73**.

Come possiamo leggere dai suoi post, **Mares73** è una madre con seri problemi di ansia e depressione, oltre ciò, il marito è un uomo poco paziente, che non capisce la sua condizione e che si è trovato ammalato di leucemia. In giovane età, **Mares73** ha subito violenza sessuale da parte del prete della sua città, del quale si fidava. Dal contesto presentato, si capisce come la condizione di **Mares73** non può cambiare nel giro di settimane o mesi, poiché è ben radicata nella sua vita. Ciò si può notare anche dalla heatmap in Fig. 5.5. Prendiamo in considerazione tre post, relativi alla stessa discussione iniziata il 10 Febbraio 2015 e conclusa il 26 Luglio dello stesso

anno. Come notiamo dalle date, l'utente scrive di rado: solamente 13 post nel lasso di tempo sopracitato.



Figura 5.5: Heatmap di Mares73

Come fatto con l'utente precedente, sono stati analizzati alcuni post, indicati in verde, per verificare se la situazione emotiva di **Mares73** è realmente costante come la heatmap ci suggerisce. Il primo post è datato 11 Febbraio, di seguito il testo:

Dear Neil & Mary thank you for your care & support. Mary I know I will be relying on lists each day to get me through. The **saddest** part for me is I'd felt **distant** from my husband for quite a while-him coping with **pressures** of work & me trying to manage my **depression** & the trauma of the royal commission etc. Then a few weeks ago we talked for hours & opened up & both felt the love that's endured through 20 years of marriage. So both of us & the kids had a :vow renewal" ceremony where him & I renewed our vows & told kids how much our family means to us. It was so **raw** & intimate & special. We made plans for this year. My daughter is 15 & my son is only 10. We were so close until the news that ripped apart our hopes & dreams. My son is very **sensitive** & always been closer to his Dad. The only emotion he is displaying is **anger**. He won't talk to me & it's very hard. (...) My daughter has just started year 10 & I'm watching her closely as she's always been so very mature & responsible. But I've caught her drinking twice so I'm so **worried** for her. I don't even know how we will survive on life insurance as no sick leave covers 9 months. The specialist is top in his field. He said without immediate treatment the leukemia would travel very fast. He said 2 things could kill my husband along the journey too much chemo so it's possible to die from infection & then he doesn't respond to the chemo which he has too for a bone marrow transplant. And if he makes it to that stage his body has to respond & not **reject** the new bone marrow. It's acute lymphoblastic leukemia. As for me I know I'll have to function on auto. The kids need one parent whose ok. I'll have to find the strength. I wish I didn't have the trauma of the legal case against the priest but I'm going to do it cause my husband said it was important to him I get justice. Justice? I want my life back. (...) Overwhelmed at how I feel, conscious of being positive around my husband & **worried** about my kids. We both have tried to give our kids a life full of love instilling confidence in them. Never wanting them to feel or experience the trauma we both went through as kids. But life's dealt a hand I cannot control they are traumatized in a different way. How will I survive all this & keep mentally strong. I can't afford **depression** or **anxiety** I'm the anchor.

Figura 5.6: Post dell'11 Febbraio di Mares73

Dal post traspare che **Mares73** sta attraversando una situazione estremamente difficile, provando un misto di tristezza, preoccupazione e senso di responsabilità. Esprime dolore per la distanza emotiva col marito, al quale è stata recentemente diagnosticata una leucemia linfoblastica acuta. **Mares73** manifesta inoltre profonda preoccupazione per i figli, in particolare per il figlio, che mostra solo rabbia, e per la figlia che ha iniziato a bere. Si sente sopraffatta e si preoccupa di dover mantenere la stabilità emotiva per il bene della famiglia, nonostante la sua stessa lotta con la depressione e il trauma del passato.

Il successivo post analizzato è datato 19 Febbraio, quindi oltre una settimana dopo il precedente. Dalla heatmap, si nota come il colore relativo alle parole di gravità

alta diventa viola, sintomo del fatto che l'utente ha usato un numero più elevato di queste parole, escludendo parole di gravità bassa.

Dear friends well within 10 days we've gone from being a family with hopes & plans to my husband now hospitalized fighting for his life. The Dr & social worker spoke to me yesterday. They said my husband is now in **isolation** undergoing 5 days of intensive chemo. (..) I'm putting on the mask for everyone yet I'm literally a bundle of nerves underneath. I really need advice/support on dealing with extreme **anxiety** that I wake up with, experience all day & night. It's awful & without the **anxiety** I think I'd be coping better inside myself. Any suggestions welcome from those who have lived on pure adrenalin. Other weird thing is I spent all last yr looking for volunteery work I'd enjoy. Then now I've had a call that debt housing needs a site manager on on of the big tower estates in Sydney to be site manager & there are funds for programs I think would be worth running. Lots of autonomy given a volunteer role & it's exactly what I was looking for but **bad** timing. I'm meeting the management staff today. One part of me thinks it's way too much to take on right now. Other part of me thinks if I could do a few hours each morning it may keep me focused on something else & have opportunities to not be so isolated. My family says a very ill husband, 2 kids & a court case is enough especially given I'd be working with very socially disadvantaged people. Not sure what to do. Selfishly I want to **escape** the pain sometimes as its all my life is full off. My son is so hateful towards me, hasn't spoken to me with anything but spite & **anger** telling me I'm a **hopeless**. He hates me & just wants his Dad. My daughter is closer & is wonderful although her **anger** is starting to show by blaming me & criticizing me for **smallest** things. I'm just so tired waking at 3rd & can't sleep again. I feel so **alone**.

Figura 5.7: Post dell'19 Febbraio di Mares73

Come ci si aspettava, **Mares73** descrive una situazione che rimane estremamente negativa e costante, con un carico emotivo sempre più pesante. Il marito è ora in isolamento per cure intensive contro la leucemia, e **Mares73** si sente sopraffatta da un'ansia costante, che la accompagna giorno e notte. Si sente sempre più isolata, stanca e incapace di trovare sollievo, svegliandosi di notte e non riuscendo più a dormire, con un profondo senso di solitudine e disperazione che pervade ogni aspetto della sua vita, sentimenti espressi anche attraverso i termini "alone" e "hopeless". L'analisi emotiva di Mares73 è stata conclusa analizzando l'ultimo post in evidenza, aspettandoci che la situazione non cambi rispetto alle precedenti, infatti anche guardando la heatmap rimane in linea con i precedenti post.

Dear friends it's been a while but I've often thought of each of you & wondered how your going. My life is a roller coaster. Husband with acute leukemia whose mostly been in intensive care due to infections. We were told a bone marrow transplant was his only hope yet he's too sick to have one. He's only 40. Priest who drugged & sexually assaulted me when I was 13. It's affecting all intimate relationships & I now have to decide the evidence what I feel of about disclosing to him. That **alone** is huge. Thing is I can cope really well in accrisis-my whole life I've been the mother figure. When my husband was first diagnosed my mother's first comment was "oh mary This Is Your Time To shone, your made for dealing with **crises** so it's best it happened to you rather than someone who doesn't cope with a **crisis**". That sums up my whole relationship & my role growing up as the eldest in a family where my father abused me & was always drunk. My mother will only say she loves me if I give her money. My father committed **suicide** by alcohol & medication overdose 3 yrs ago. I have no family support. Before **depression** hit me I was a functioning high achiever & always had been. Yet when I was serially assaulted by a stranger around the time of my father's **death** -I **fell** apart. And so the last 6 yrs I've lost all friends & have become very **introverted** & isolated, **afraid** to go out & so **sad** at my **emptiness**. Yet 7 years ago I was a different woman. Now I **struggle** to find things I enjoy & I feel so **alone**. Until last week I was coping with visiting my husband, caring for kids etc. Then 2 days ago I crashed & it was as though **depression** had overtaken me without me seeing it coming. I rarely get time to myself as im meeting others needs yet I had 2 hrs to myself & I couldn't enjoy reading -my favorite hobby. I could only stare into space. I know the **darkness** is starting & I don't know how to stop it especially when my husband's supposed to come home tomorrow & I can't even get the house sorted. Overwhelmed Wish we knew when it would hit & what to do.

Figura 5.8: Post del 24 Maggio di Mares73

In questo post, datato 24 Maggio, **Mares73** descrive una continua e intensa sofferenza emotiva e isolamento, con la sensazione di essere sopraffatta dalle responsabilità. Il marito, malato di leucemia, è troppo debole per un trapianto di midollo osseo. **Mares73** deve affrontare anche il trauma passato di abusi sessuali e la mancanza di supporto familiare, in particolare da parte della madre. Nonostante sia sempre stata vista come una persona capace di gestire le crisi, ora si sente profondamente vulnerabile, con la depressione che l'ha travolta improvvisamente, rendendola incapace di godersi anche le attività che un tempo amava, come la lettura. Si sente sola, sopraffatta e incapace di fermare il "buio" che sta iniziando a prenderla, soprattutto con il ritorno a casa imminente del marito e la pressione di mantenere tutto sotto controllo.

Come possiamo notare, il lasso di tempo tra gli ultimi due post analizzati supera i tre mesi, ma la situazione negativa persiste nella vita di **Mares73**.

Concludendo, analizzando questi post, è stata confermata un'effettiva correlazione tra il numero di termini target utilizzato dagli utenti e la gravità del loro stato emotivo.

Il sistema di text mining sviluppato in questa tesi può quindi essere di aiuto per monitorare l'andamento temporale dello stato emotivo di persone che esprimono il loro disagio scrivendo su community online. Il grande vantaggio risiede nella trasformazione del testo in heatmap che rendono immediatamente individuabili pattern di miglioramento o peggioramento, consentendo per esempio lo sviluppo di tecniche più sofisticate per l'analisi automatica delle heatmap, con conseguente allertamento

dei soccorsi qualora si verificano situazioni di profonda gravità. Un simile approccio di monitoraggio potrebbe aiutare a dare supporto mirato alle persone che soffrono di disturbi dell'umore, salvando anche le loro vite.

Capitolo 6

Conclusioni

Lo studio condotto nell'ambito di questa tesi ha riguardato la progettazione e sviluppo di un metodo di text mining e data visualization per la rilevazione pattern caratterizzati da miglioramento o peggioramento dello stato emotivo di utenti che scrivono in community online. Il sistema sviluppato ha mostrato una notevole efficacia, trasformando i post in heatmap dalle quali poter facilmente individuare il cambio di lessico utilizzato dagli utenti, usando un dizionario di parole generate a partire da termini caratteristici riportati nel Manuale Diagnostico e Statistico dei Disturbi Mentali (DSM-5).

Sviluppi futuri includono l'impiego di dizionari contenenti più termini e l'adozione di tecniche più sofisticate che consentano di analizzarne il contesto di un post.

Un ulteriore sviluppo di questo studio consiste nella generazione di un ampio dataset di heatmap per l'addestramento di modelli di intelligenza artificiale volti a identificare e prevedere potenziali segnali di peggioramento dello stato emotivo dell'utente, consentendo un intervento sempre più tempestivo.

Appendice

```

1 <?php
2 include_once 'lib/parser/simple_html_dom.php';
3
4 require_once ( __DIR__."/lib/Autore.php");
5 require_once ( __DIR__."/lib/Forum.php");
6 require_once ( __DIR__."/lib/Post.php");
7
8 set_time_limit(0);
9
10 $page = 1;
11 while($html = file_get_html("https://forums.beyondblue.org.au/t5/depression/bd-p/c1-sc2-b2/page/$page")){
12     $patherDIV = $html->find('div.custom-message-list.all-discussions')[0];
13     if(!is_null($patherDIV))
14         $articles = $patherDIV->find("article.custom-message-tile");
15
16     $url = "";
17     foreach ($articles as $article) {
18         $aTAGs = $article->find('a[title]');
19         foreach ($aTAGs as $aTAG) {
20             if($aTAG->title != "View profile"){
21                 $titolo = $aTAG->title;
22                 $url = $aTAG->href;
23             }
24         }
25
26         $page_ = 1;
27         $reply_ = -1;
28         $AutorePrincipale = "";
29         while($html_ = file_get_html("https://forums.beyondblue.org.au/$url/page/$page_")){
30             $replsDIV = $html_>find("div.MessageView");
31             $num_reply = $html_>find("span[class='lia-hidden']")[0]->innertext;
32             if($reply_ >= intval($num_reply))
33                 break;
34
35             $parole_bandite_data = array("<span class='local-date'\>", "</span>", "&lrn;");
36             $parole_bandite_testo = array("<div class='lia-message-body-content'\>");
37             $pattern_emoji = '/[\x{1F600}-\x{1F64F}\x{1F300}-\x{1FAFF}\x{2600}-\x{27BF}]/u';
38
39             foreach ($replsDIV as $replyDIV){
40                 $nameUser = $replyDIV->find("img[class='lia-user-avatar-message']")[0]->title;
41                 if(is_null($replyDIV->find("span[class='local-friendly-date']")[0]))
42                     $date = $replyDIV->find("span[class='local-date']")[0]->outertext;
43                 else
44                     $date = $replyDIV->find("span[class='local-friendly-date']")[0]->title;
45
46                 $date = preg_replace('/[-\x20-\x7E]/', '', str_replace($parole_bandite_data, "", trim($date)));
47                 $messaggio = preg_replace('/<[~]*>/', ' ', str_replace($parole_bandite_testo, " ", $replyDIV->find(
48                     "div[class='lia-message-body-content']")[0]->outertext));
49                 $messaggio = preg_replace($pattern_emoji, '', $messaggio);
50                 $messaggio = html_entity_decode($messaggio);
51
52                 try {
53                     $Autore = new Autore();
54                     $Autore->setUsername($nameUser);
55
56                     $data = $date;
57                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
58                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
59                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
60                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
61                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
62                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
63                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
64                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
65                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
66                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
67                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
68                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
69                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
70                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
71                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
72                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
73                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
74                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
75                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
76                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
77                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
78                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
79                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
80                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
81                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
82                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
83                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
84                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
85                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
86                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
87                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
88                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
89                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
90                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
91                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
92                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
93                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
94                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
95                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
96                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
97                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
98                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
99                     $data = preg_replace('/\x20-\x7E/', ' ', $data);
100                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
101                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
102                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
103                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
104                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
105                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
106                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
107                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
108                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
109                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
110                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
111                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
112                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
113                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
114                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
115                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
116                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
117                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
118                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
119                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
120                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
121                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
122                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
123                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
124                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
125                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
126                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
127                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
128                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
129                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
130                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
131                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
132                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
133                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
134                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
135                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
136                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
137                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
138                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
139                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
140                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
141                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
142                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
143                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
144                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
145                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
146                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
147                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
148                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
149                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
150                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
151                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
152                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
153                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
154                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
155                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
156                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
157                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
158                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
159                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
160                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
161                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
162                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
163                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
164                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
165                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
166                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
167                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
168                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
169                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
170                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
171                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
172                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
173                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
174                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
175                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
176                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
177                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
178                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
179                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
180                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
181                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
182                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
183                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
184                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
185                    $data = preg_replace('/\x20-\x7E/', ' ', $data);
186                    $data = preg_replace('/\x20-\x7E/', ' ',
```

```

54
55     $Forum = new Forum();
56     $Forum->setNome("depression");
57
58     $Post = new Post();
59     $datetime = new DateTime($date);
60     $date = $datetime->format("Y-m-d");
61     $Post->setData($date);
62     $Post->setTesto($messaggio);
63     $Post->setForum($Forum);
64     $Post->setAutore($Autore);
65
66     if($reply_ != -1)
67         $Post->setDiscussione(new Post($IdDiscussione));
68
69     if($AutorePrincipale == "" || $AutorePrincipale == $nameUser)
70         $Post->save($nameUser, "depression");
71
72     } catch (InvalidArgumentException $e) {
73         echo $e;
74         return;
75     } catch (mysqli_sql_exception $e) {
76         echo $e;
77         return;
78     }
79
80     if($reply_ == -1){
81         $IdDiscussione = $Post->getId();
82         $AutorePrincipale = $nameUser;
83     }
84     $reply_ ++;
85 }
86 $page++;
87 }
88 }
89 $page++;
90 }
91 ?>

```

A.2 Script per la fase generazione dei punteggi

A.2.1 Classe Analizzatore

```

1  import string
2  import nltk
3  from nltk.corpus import stopwords
4  import spacy
5  from classi.Database import Database
6
7  class Analizzatore:
8      def __init__(self, v_little, v_very):
9          nltk.download('stopwords')
10         self.nlp = spacy.load('en_core_web_lg')
11
12         self.nounce = {}
13         self.adjective = {}
14         self.adverb = {}
15         self.multiple = {}
16         self.verb = {}
17         self.verb_no_target = {}
18
19         self.gravity_adv = {
20             "low": v_little,
21             "high": v_very
22         }
23
24         self.numero_parole = {
25             "low": 0,
26             "medium": 0,
27             "high": 0
28         }
29
30     def inizializzaParoleTarget(self):
31         db = Database()
32         query = """
33             SELECT Nome, Gravita, Tipo
34             FROM parola_target
35             """

```



```

36     rows = db.myExecute(query)
37     for row in rows:
38         name = row[0]
39         gravity = row[1]
40         type = row[2]
41
42         if type == "nome":
43             self.nounce[name] = gravity
44         elif type == "aggettivo":
45             self.adjective[name] = gravity
46         elif type == "avverbio":
47             self.adverb[name] = self.gravity_adv[gravity]
48         elif type == "multipla":
49             self.multiple[name] = gravity
50
51     db.close()
52
53     def pre_elaborazione(self):
54         self.post_attuale = self.post_attuale.lower()
55         self.post_attuale = self.post_attuale.translate(str.maketrans('', '', string.punctuation))
56         self.post_attuale = ''.join([i for i in self.post_attuale if not i.isdigit()])
57
58         parole = self.post_attuale.split()
59         stop_words = set(stopwords.words('english'))
60         self.post_attuale = ' '.join([parola for parola in parole if parola not in stop_words])
61
62         doc = self.nlp(self.post_attuale)
63         testo = ' '.join([token.lemma_ for token in doc])
64         testo = ' '.join(testo.split())
65         return testo
66
67     def derivate_from_nounce(self, verb):
68         suffissi = ["", "er", "ing", "ed", "ion", "ment", "ance", "ence"]
69         for nome, gravita in self.nounce.items():
70             for suffisso in suffissi:
71                 if (verb + suffisso == nome):
72                     return gravita
73         return False
74
75     def calcolaPunteggioParola(self, parola, gravita, parole_post_adv = False):
76         if (parole_post_adv == False or parola not in parole_post_adv):
77             self.numero_parole[gravita] += 1
78         else:
79             self.numero_parole[gravita] = self.numero_parole[gravita] + parole_post_adv[parola]
80
81     def calcolaPunteggioPost(self, testo):
82         doc = self.nlp(testo)
83         parole_post_adv = {}
84         for token in doc:
85             parola = token.text
86             type = token.pos_
87             if type == "VERB":
88                 if parola in self.verb:
89                     self.calcolaPunteggioParola(parola, self.verb[parola], parole_post_adv)
90                 elif parola not in self.verb_no_target:
91                     gravity = self.derivate_from_nounce(parola)
92                     if (gravity != False):
93                         self.verb[parola] = gravity
94                         self.calcolaPunteggioParola(parola, self.verb[parola], parole_post_adv)
95                 else:
96                     self.verb_no_target[parola] = -1
97
98                 if parola in parole_post_adv:
99                     del parole_post_adv[parola]
100
101             elif type == "NOUN":
102                 if parola in self.nounce:
103                     self.calcolaPunteggioParola(parola, self.nounce[parola])
104
105             elif type == "ADJ":
106                 if parola in self.adjective:
107                     self.calcolaPunteggioParola(parola, self.adjective[parola], parole_post_adv)
108
109                 if parola in parole_post_adv:
110                     del parole_post_adv[parola]
111
112             elif type == "ADV":
113                 if parola in self.adverb:
114                     parole_post_adv[token.head.text] = self.adverb[parola]
115
116     def punteggiamentoMultiple(self):
117         self.post_attuale = self.post_attuale.lower()
118         for frase, gravita in self.multiple.items():
119             if frase in self.post_attuale:
120                 self.numero_parole[gravita] += 1
121                 self.post_attuale = self.post_attuale.replace(frase, "")

```

```

122
123 def analizzaPosts(self):
124     db = Database()
125     query = """
126         SELECT IdAutore
127         FROM autore
128         """
129     autori = db.myExecute(query)
130     for autore in autori:
131         query = """
132             SELECT Testo, IdPost
133             FROM post
134             WHERE FkAutore = %s
135             GROUP BY Testo, Data
136             ORDER BY Data, IdPost
137             """
138     posts = db.myExecute(query, [autore[0]])
139     for post in posts:
140         self.id = post[1]
141         if self.id in self.post_target_numeric or self.id in self.post_target_string:
142             print(str(self.id)+"\n")
143
144         self.post_attuale = post[0]
145         self.punteggioParoleMultiple()
146         post_elaborato = self.pre_elaborazione()
147         self.calcolaPunteggioPost(post_elaborato)
148         query = """
149             UPDATE post
150             SET nParoleLow = %s, nParoleMedium = %s, nParoleHigh = %s
151             WHERE IdPost = %s
152             """
153         nLow, self.numero_parole["low"] = self.numero_parole["low"], 0
154         nMedium, self.numero_parole["medium"] = self.numero_parole["medium"], 0
155         nHigh, self.numero_parole["high"] = self.numero_parole["high"], 0
156         db.myExecute(query, [nLow, nMedium, nHigh, post[1]], False)

```

A.2.2 Main

```

1 from classi.Analizzatore import Analizzatore
2
3 # Diamo il valore alle gravit per gli avverbi di quantita passandole come parametri
4 an = Analizzatore(0.5, 2)
5
6 # Vengono inizializzati i dizionari dei nomi, aggettivi e avverbi e le loro gravit
7 an.inizializzaParoleTarget()
8
9 # Adesso cominciamo ad analizzare tutti i post degli utenti, utente per utente.
10 # Dando ad ogni post un punteggio relativo alla sua gravit in relazione all'utente che lo ha pubblicato
11 an.analizzaPosts()

```

Bibliografia

- [1] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing, 5th edition, 2013.
- [2] Beyond Blue. Beyond blue online community: Depression and mental health support, 2024. Ottenere i dati relativi ai post degli utenti per il forum sulla depressione.
- [3] IBM. Cos'è il text mining? Ultimo accesso: 13 settembre 2024.
- [4] National Alliance on Mental Illness (NAMI). You are not alone: Mental health in america - 2023, 2023. Ultimo accesso: 14 settembre 2024.