

SENTIMENT ANALYSIS ON ITALIAN PATIENT INFORMATION LEAFLETS

Text Mining Project

Nicola Santillo

Corso di laurea “Tecniche informatiche per la gestione dei dati”

a.a. 2023/2024

Indice

Text Mining Project	1
1. Introduzione	4
1.1. Obiettivo	5
2. Dataset	5
2.1. Estrazione	5
2.2. Tokenizzazione e Stopwords	6
2.3. Preparazione dati per Topic Modelling	6
2.4. Esplorazione e analisi dei dati	7
2.4.1. Morfina	7
2.4.1.1. Bigrammi	9
2.4.1.2. Trigrammi	10
2.4.2. Bisolvon	10
2.4.2.1. Bigrammi	13
2.4.2.2. Trigrammi	14
2.4.3. Borocillina	15
2.4.3.1. Bigrammi	16
2.4.3.2. Trigrammi	17
2.4.4. Dicloream	18
2.4.4.1. Bigrammi	20
2.4.4.2. Trigrammi	21
2.4.5. Fentanyl	22
2.4.5.1. Bigrammi	24
2.4.5.2. Tigrammi	25
3. Modelli	26
3.1. Vader	26
3.2. umBERTo	26
3.3. sentITA	26
3.4. FEEL-IT	27
3.5. Latent Dirichlet Allocation (LDA)	27
3.6. Metriche	27
4. Risultati	29
4.1. Morfina	29
4.1.1. Vader	29
4.1.2. umBERTo	30

4.1.3.	SentITA.....	31
4.1.4.	Confronto risultati tra i modelli umBERTo e sentITA.....	33
4.1.5.	FEEL-IT	34
4.2.	Bisolvon	36
4.2.1.	Vader	36
4.2.2.	umBERTo	37
4.2.3.	SentITA.....	38
4.2.4.	Confronto risultati tra i modelli umBERTo e sentITA.....	40
4.2.5.	FEEL-IT	41
4.3.	Borocillina.....	43
4.3.1.	Vader	43
4.3.2.	umBERTo	44
4.3.3.	SentITA.....	45
4.3.4.	Confronto risultati tra i modelli umBERTo e sentITA.....	47
4.3.5.	FEEL-IT	48
4.4.	Diclorem	50
4.4.1.	Vader	50
4.4.2.	umBERTo	51
4.4.3.	SentITA.....	52
4.4.4.	Confronto risultati tra i modelli umBERTo e sentITA.....	54
4.4.5.	FEEL-IT	55
4.5.	Fentanyl.....	57
4.5.1.	Vader	57
4.5.2.	umBERTo	58
4.5.3.	SentITA.....	59
4.5.4.	Confronto risultati tra i modelli umBERTo e sentITA.....	61
4.5.5.	FEEL-IT	62
4.6.	Topic Modelling	63
4.6.1.	Latent Dirichlet Allocation (LDA).....	63
5.	Conclusioni	67
6.	Sitografia	69
7.	Data Cards	70

Abstract — Questo studio presenta diversi approcci di analisi del sentiment e delle emozioni implementati per migliorare la comprensione e la percezione dei bugiardini dei farmaci. Il focus del progetto è analizzare come il linguaggio utilizzato nei bugiardini influenzi le emozioni e i sentimenti dei lettori, al fine di identificare e correggere le parti del testo che possano risultare allarmanti o difficili da comprendere. Sono stati utilizzati vari modelli di elaborazione del linguaggio naturale, come Vader, SentITA, UmBERTo per l'analisi del sentiment e FEEL-it per il rilevamento delle emozioni. Inoltre, è stato svolto il topic modelling tramite LDA per identificare i temi principali presenti nei testi. I risultati forniscono una base importante per migliorare la comunicazione nei bugiardini, riducendo l'ansia dei pazienti e promuovendo un uso più sicuro e consapevole dei farmaci.

1. Introduzione

L'analisi del sentiment e delle emozioni è una branca emergente dell'intelligenza artificiale e della linguistica computazionale che si concentra sull'estrazione e l'interpretazione di emozioni e opinioni da testi scritti. Questa tecnologia ha trovato applicazioni in una varietà di campi, dall'analisi delle recensioni dei consumatori ai social media e recentemente ha iniziato a essere esplorata anche nel contesto della salute e della medicina, anche nell'analisi dei bugiardini dei farmaci. I bugiardini dei farmaci sono documenti fondamentali che accompagnano ogni medicinale, contenendo informazioni essenziali sull'uso corretto, le dosi, le controindicazioni e i possibili effetti collaterali. La comprensione accurata di queste informazioni è cruciale per i pazienti e gli operatori sanitari, ma spesso il linguaggio utilizzato può risultare complesso e tecnico, creando difficoltà per i non esperti. Queste analisi possono contribuire significativamente a migliorare la comunicazione tra produttori di farmaci e pazienti. Comprendere come le informazioni vengono percepite dai lettori può aiutare a identificare le parti del testo che potrebbero essere allarmanti o difficili da comprendere, permettendo di rendere le informazioni più accessibili e meno intimidatorie. Questo processo può facilitare la riscrittura di frasi in modo più chiaro e rassicurante, riducendo l'ansia e il timore associati all'uso dei farmaci. L'analisi del sentiment può anche aiutare a identificare problemi di comunicazione, consentendo alle aziende farmaceutiche di migliorare la redazione dei bugiardini per rispondere meglio alle esigenze dei pazienti. Un linguaggio più chiaro e rassicurante può favorire una maggiore fiducia nei farmaci e nei loro produttori, contribuendo a un uso più sicuro e consapevole dei medicinali.

Le tecniche utilizzate per l'analisi del sentiment e delle emozioni nei bugiardini dei farmaci includono il rilevamento delle emozioni, l'analisi del sentiment e i modelli di deep learning. Queste tecniche di elaborazione del linguaggio naturale (NLP) permettono di comprendere il contesto e il significato del testo in modo più accurato e sofisticato. Inoltre, l'uso del topic modelling può essere estremamente utile per identificare i temi principali presenti nei testi dei bugiardini. Questo approccio consente di scomporre il testo in argomenti chiave, facilitando l'individuazione delle aree che potrebbero necessitare di una maggiore chiarezza o di una revisione linguistica.

In conclusione, l'analisi del sentiment e delle emozioni nei bugiardini dei farmaci rappresenta un progresso significativo verso una comunicazione sanitaria più efficace e centrata sul paziente. L'uso di tecnologie avanzate di NLP consente di migliorare la comprensione, ridurre l'ansia e fornire informazioni più chiare e rassicuranti, contribuendo a un uso più sicuro e consapevole dei farmaci. Questo approccio non solo migliora l'esperienza del paziente, ma può anche aumentare la fiducia nel sistema sanitario e nei prodotti farmaceutici.

1.1. Obiettivo

L'obiettivo è analizzare il sentiment e le emozioni dei bugiardini analizzando separatamente le sezioni che li compongono.

2. Dataset

Il primo passo è quello di individuare quali bugiardini, tra i diversi tipi di farmaci presenti sul mercato, scegliere. Il criterio di selezione si è basato principalmente sui loro principi attivi. Infatti, sono stati utilizzati medicinali che hanno componenti chimiche, modalità d'uso e scopi differenti. I medicinali scelti sono stati:

- Morfina
- Bisolvon
- Borocillina
- Dicloremum
- Fentanyl

Dopo aver individuato i farmaci, tramite il sito dell'AIFA, i bugiardini sono stati scaricati manualmente in formato PDF. A questo punto, si è provveduto all'estrazione dei contenuti dei vari bugiardini.

2.1. Estrazione

Per l'estrazione del contenuto dai bugiardini in formato PDF, è stato utilizzato il tool **pdfplumber**. Questo strumento si è rivelato particolarmente utile per la sua capacità di gestire vari formati di PDF, consentendo un'estrazione precisa e dettagliata dei testi contenuti nei documenti.

Dopo aver studiato la struttura dei bugiardini, rivelatasi simile, si è passati all'estrazione dei contenuti di interesse. Tramite una funzione creata appositamente, sono stati memorizzati all'interno di un dataframe solo i contenuti dei sei paragrafi che compongono il documento. La struttura del dataframe iniziale è composta da due colonne: 'Paragrafo' e 'Contenuto'. Il dataframe relativo a ciascun farmaco è stato rinominato come segue: *df_morfina*, *df_bisolvon*, *df_borocillina*, *df_dicloremum*, *df_fentanyl*.

Di seguito un esempio di come appare *df_morfina*:

Paragrafo	Contenuto
1. Che cos'è MORFINA CLORIDRATO S.A.L.F. e a cosa serve	MORFINA CLORIDRATO S.A.L.F. contiene il principio...
2. Cosa deve sapere prima di usare MORFINA CLORIDRATO S.A.L.F	Non usi MORFINA CLORIDRATO S.A.L.F. - se è allergico...
3. Come usare MORFINA CLORIDRATO S.A.L.F	Questo medicinale le sarà somministrato da un ...
4. Possibili effetti indesiderati	Come tutti i medicinali, questo medicinale può...
5. Come conservare MORFINA CLORIDRATO S.A.L.F.	Conservi questo medicinale fuori dalla portata.
6. Contenuto della confezione e altre informazioni	Cosa contiene MORFINA CLORIDRATO S.A.L.F. MORF.

2.2. Tokenizzazione e Stopwords

Prima di procedere con l'analisi effettiva del testo, è stata eseguita la tokenizzazione e la rimozione delle stopwords. Inizialmente, sulla colonna 'Contenuto' dei vari dataframe, sono stati rimossi tutti i caratteri non alfabetici, utilizzando la *libreria clean-text*, mentre per la tokenizzazione è stata impiegata la funzione 'word_tokenize'. Invece, per la rimozione delle stopwords in italiano è stato precedentemente scaricato un file, *stopwords-it.txt*, da Github. Quest'ultima operazione consente di ottenere una lista di parole significative, escludendo le parole comuni che non apportano informazioni rilevanti.

Successivamente, tutti i valori sono stati convertiti in minuscolo per uniformare il testo.

Infine, è stata creata una terza colonna 'Numero', che identifica il numero del paragrafo e rinominate le due precedenti in 'Sezione' e 'Contenuto'.

Tutti questi passaggi sono stati implementati in una funzione per essere richiamati per ogni bugiardino da analizzare. Considerando che alcuni nomi dei farmaci includono il nome dell'azienda produttrice, la funzione accetta anche come argomento una stringa contenente il nome dell'azienda, in modo che ogni sua occorrenza possa essere eliminata dal dataset.

La composizione del dataset *df_morfina* appare come segue:

Numero	Sezione	Contenuto
1	che cos'è morfina cloridrato s.a.l.f. e a cosa...	morfina cloridrato contiene principio attivo m...
2	cosa deve sapere prima di usare morfina cloridrato...	usi morfina cloridrato allergico morfina clori...
3	come usare morfina cloridrato s.a.l.f.	medicinale somministrato medico infermiere pel...
4	possibili effetti indesiderati	medicinali medicinale causare effetti indeside...
5	come conservare morfina cloridrato s.a.l.f	conservi medicinale portata vista bambini medi...
6	contenuto della confezione e altre informazioni	contiene morfina cloridrato morfina cloridrato...

2.3. Preparazione dati per Topic Modelling

Per eseguire il topic modelling sui bugiardini dei farmaci, sono stati seguiti una serie di passaggi. Inizialmente, ogni bugiardino è stato rappresentato come un dataset composto da tre colonne: 'Numero', 'Sezione' e 'Contenuto'. Poi si sono unite le righe della colonna 'Contenuto' per ciascun dataset, ottenendo così un unico testo per ogni bugiardino. Questa operazione è stata ripetuta per tutti i bugiardini, creando una stringa unica per ciascun documento, da cui le stopwords erano già state eliminate in precedenza (vedi paragrafo 2.2).

A questo punto, i contenuti di tutti i bugiardini sono stati raggruppati, formando una lista di stringhe, ciascuna rappresentante un bugiardino. Questa lista è stata quindi sottoposta a una nuova tokenizzazione, che ha suddiviso il testo in unità più piccole, come parole o frasi, pronte per l'analisi successiva.

Con i testi adeguatamente preparati, è stato applicato il modello di topic modelling LDA (Latent Dirichlet Allocation) in modo da identificare i temi principali presenti nei testi, raggruppando le parole che frequentemente co-occorrono insieme in base alla probabilità.

2.4. Esplorazione e analisi dei dati

Per ogni bugiardino è stata fatta un'analisi descrittiva divisa in tre parti:

- Nella prima è stato creato un grafico a barre che mostra le 20 parole utilizzate più frequentemente
- Nella seconda è stata creata una WordCloud, strumento grafico che rappresenta visivamente le parole più frequenti. Utilizzando dimensioni proporzionali alla frequenza delle parole, essa facilita la visualizzazione dei temi principali e delle tendenze, rendendo l'analisi dei dati più intuitiva e immediata.
- Nella terza un grafico a barre basato su TF-IDF (Term Frequency-Inverse Document Frequency) visualizza le parole chiave di un documento in modo proporzionale alla loro importanza. Questo tipo di grafico aiuta a identificare rapidamente i temi principali e le parole rilevanti specifiche di un documento rispetto agli altri, migliorando la comprensione e l'analisi dei dati testuali.

Successivamente sono stati generati n-grammi, in particolare bigrammi e trigrammi, per ogni bugiardino. Un n-gram è una sequenza di 'n' parole consecutive in un testo. Gli n-grammi vengono utilizzati per analizzare la struttura del testo, identificando combinazioni ricorrenti di parole che possono fornire una migliore comprensione del contesto e delle associazioni tra termini. Sono stati scelti bigrammi (sequenze di due parole) e trigrammi (sequenze di tre parole) per catturare le combinazioni di parole più comuni e significative. Questo approccio consente di individuare frasi o espressioni che potrebbero essere rilevanti per l'analisi, migliorando l'interpretazione dei contenuti e la rilevazione di pattern linguistici specifici.

Di seguito l'analisi descrittiva dei cinque bugiardini.

2.4.1. Morfina

Tra le parole più frequenti che troviamo all'interno di un bugiardino, oltre al principio attivo del farmaco, c'è tutto ciò che riguarda la terapia (es. dose, mg) e i problemi che potrebbe causare (es. depressione, dipendenza). Da questo primo grafico possiamo notare come questo farmaco possa essere, se preso in maniera inconsapevole, poco salutare.

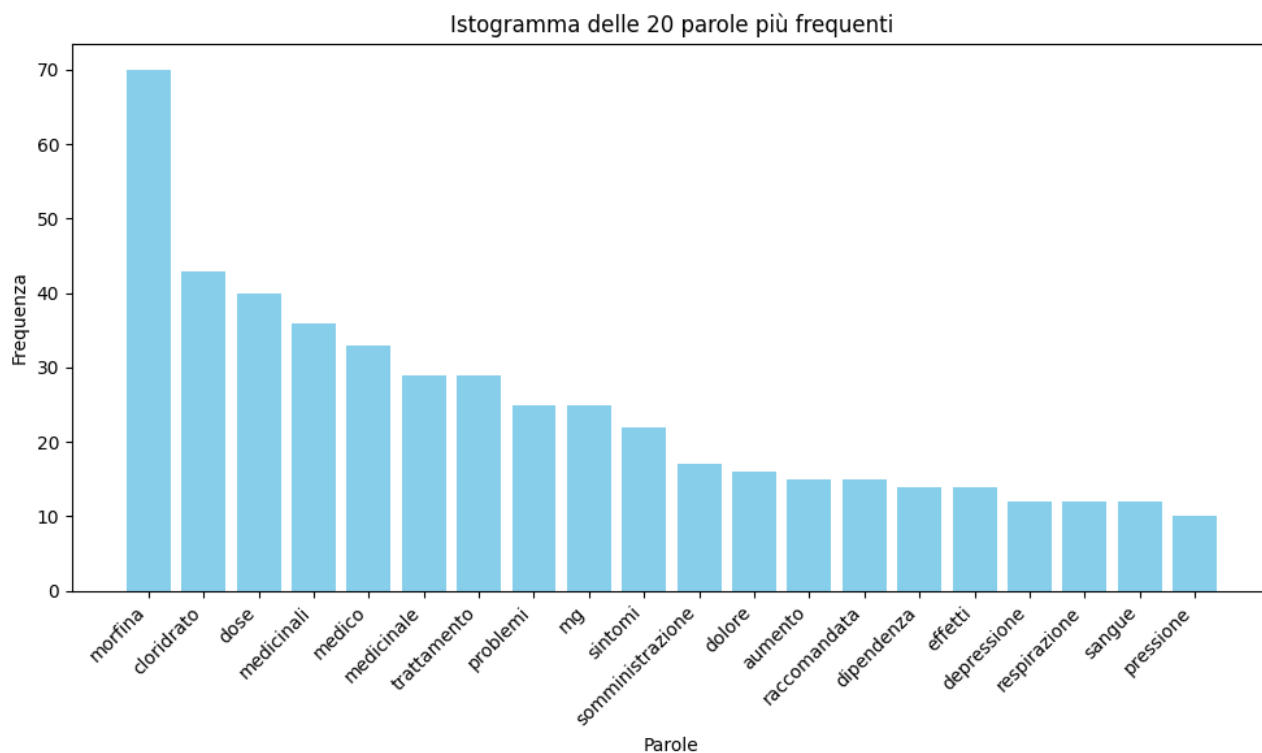


Figura 1 - Istogramma con le 20 parole più frequenti di Morfina



Figura 2 - WordCloud di Morfina

Esaminando il WordCloud, possiamo notare che le parole più grandi corrispondono alle stesse parole menzionate in precedenza. Ancora una volta, queste parole chiave sono prevalentemente di allerta verso chi ne fa uso.

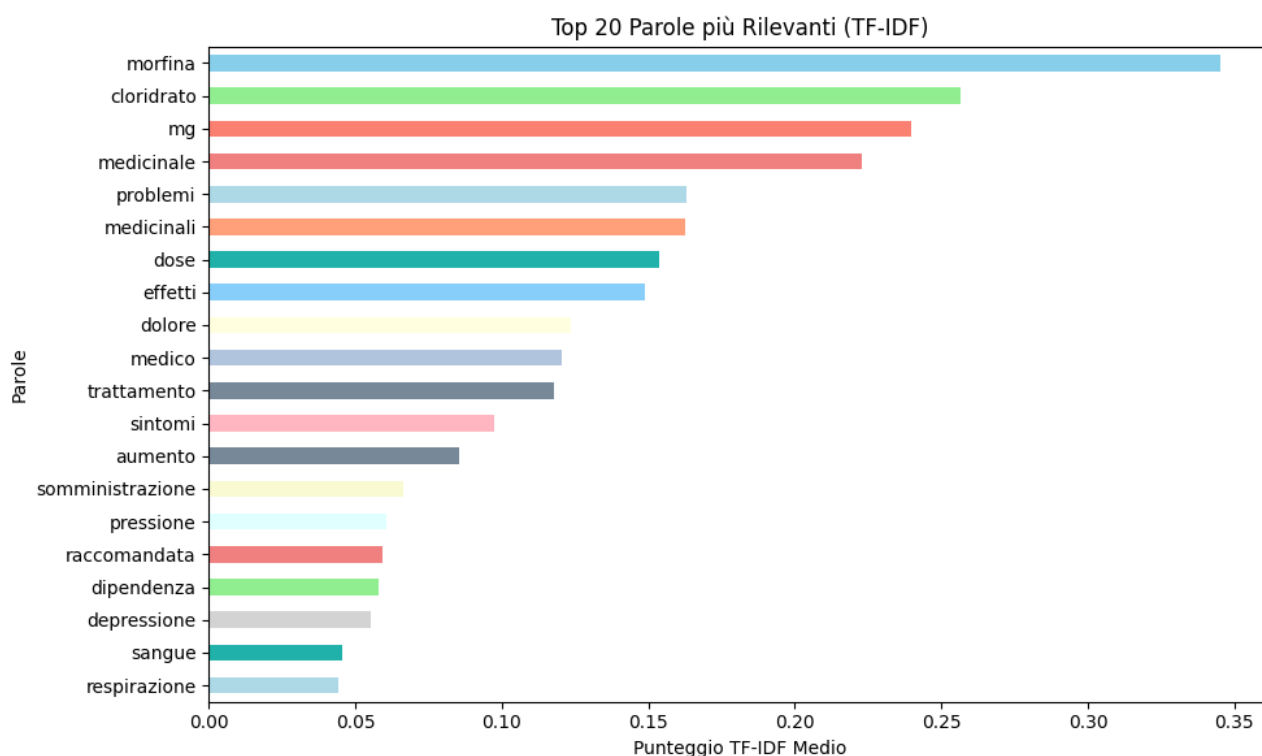


Figura 3 - Istogramma parole più rilevanti di Morfina

Nella *Figura 3* vengono rappresentate le parole più rilevanti tramite TF-IDF. È importante notare come il quantitativo di dose da somministrare ('mg') sia altamente rilevante all'interno del bugiardino. Questo fa capire quanto questo farmaco debba essere preso con estrema cautela per non cadere nella dipendenza.

2.4.1.1. Bigrammi

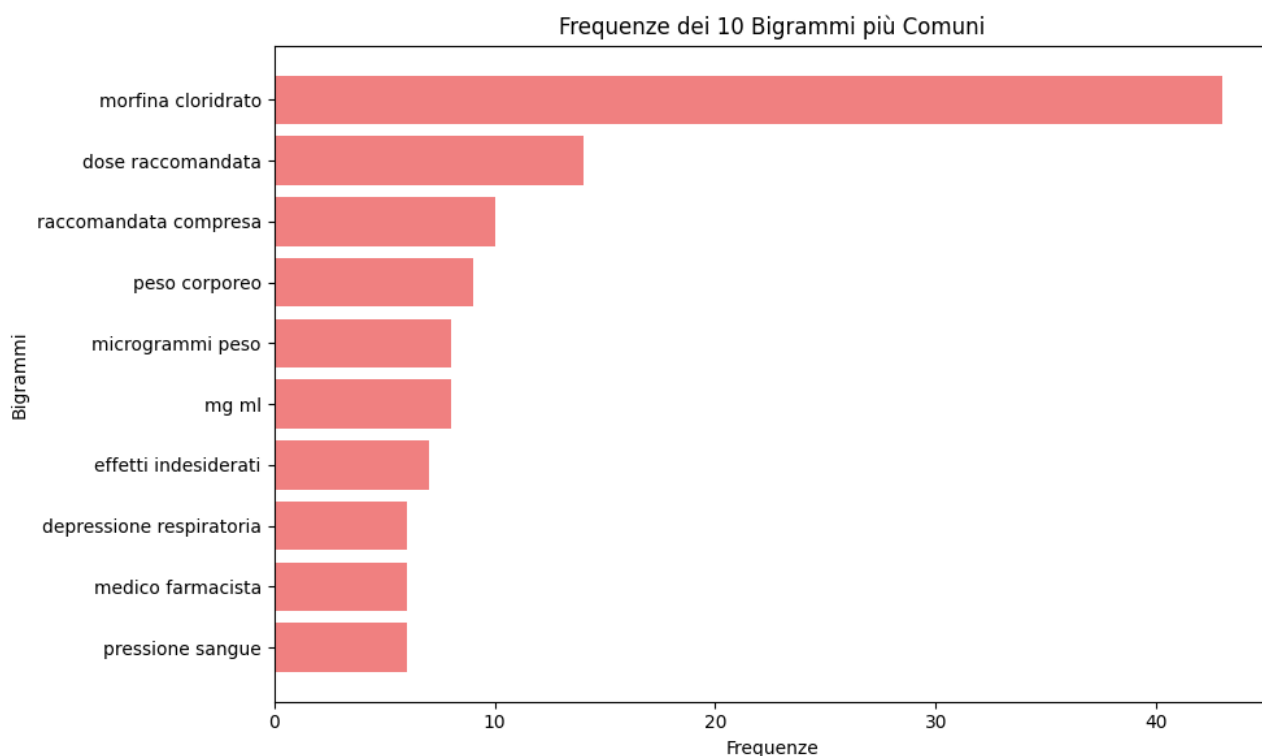


Figura 4 - 10 bigrammi più frequenti nel dataset di Morfina

Il bigramma "morfina cloridrato" risulta essere il più frequente, con oltre 40 occorrenze, seguito da "dose raccomandata" e "raccomandata compresa". Altri bigrammi significativi includono "peso corporeo", "microgrammi peso", "mg ml", "effetti indesiderati" e "depressione respiratoria". Questi bigrammi evidenziano termini chiave legati al dosaggio, agli effetti e alle raccomandazioni mediche dei farmaci.

2.4.1.2. Trigrammi

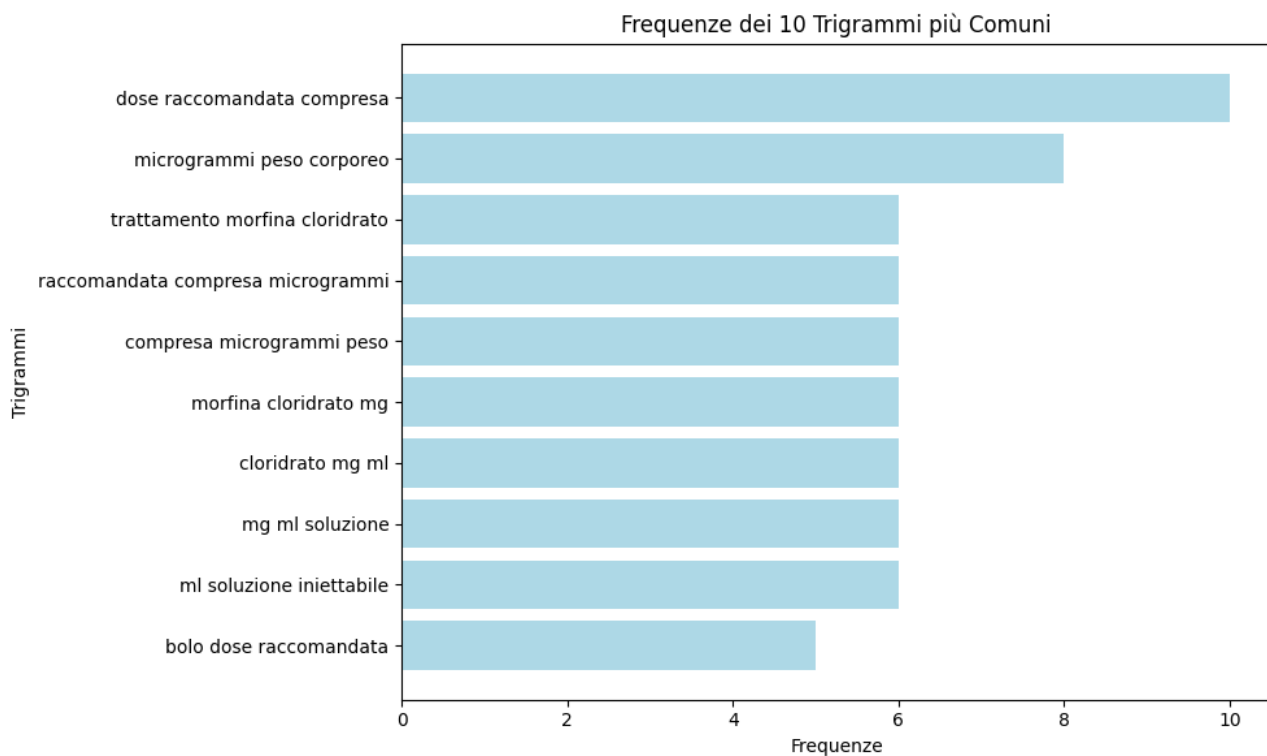


Figura 5 - 10 trigrammi più frequenti nel dataset di Morfina

"Dose raccomandata compresa" è il trigramma più frequente, con oltre 10 occorrenze, indicando spesso le raccomandazioni relative al dosaggio. Segue "microgrammi peso corporeo", che sottolinea l'importanza di dosaggi basati sul peso del paziente. "Trattamento morfina cloridrato" evidenzia l'uso della morfina per specifici trattamenti.

Altri trigrammi rilevanti includono "raccomandata compresa microgrammi", "compresa microgrammi peso" e "morfina cloridrato mg", che spesso compaiono nei contesti di dosaggio e somministrazione. "Cloridrato mg ml" e "mg ml soluzione" si riferiscono alle concentrazioni e alle forme delle soluzioni. Infine, "ml soluzione iniettabile" e "bolo dose raccomandata" indicano specifiche modalità di somministrazione e raccomandazioni di dosaggio.

2.4.2. Bisolvon

Dalla Figura 6 si può vedere che la parola più frequente è relativa ai *ml/mg* (in base all'assunzione del farmaco). Questa prima osservazione ci fa capire come, nonostante non sia un farmaco nocivo, il produttore metta in guardia il consumatore su quanto poterne assumere. Questa ipotesi trova conferma dato che vengono riportate tra le più frequenti anche parole come *'effetti'*, *'reazioni'* ecc.

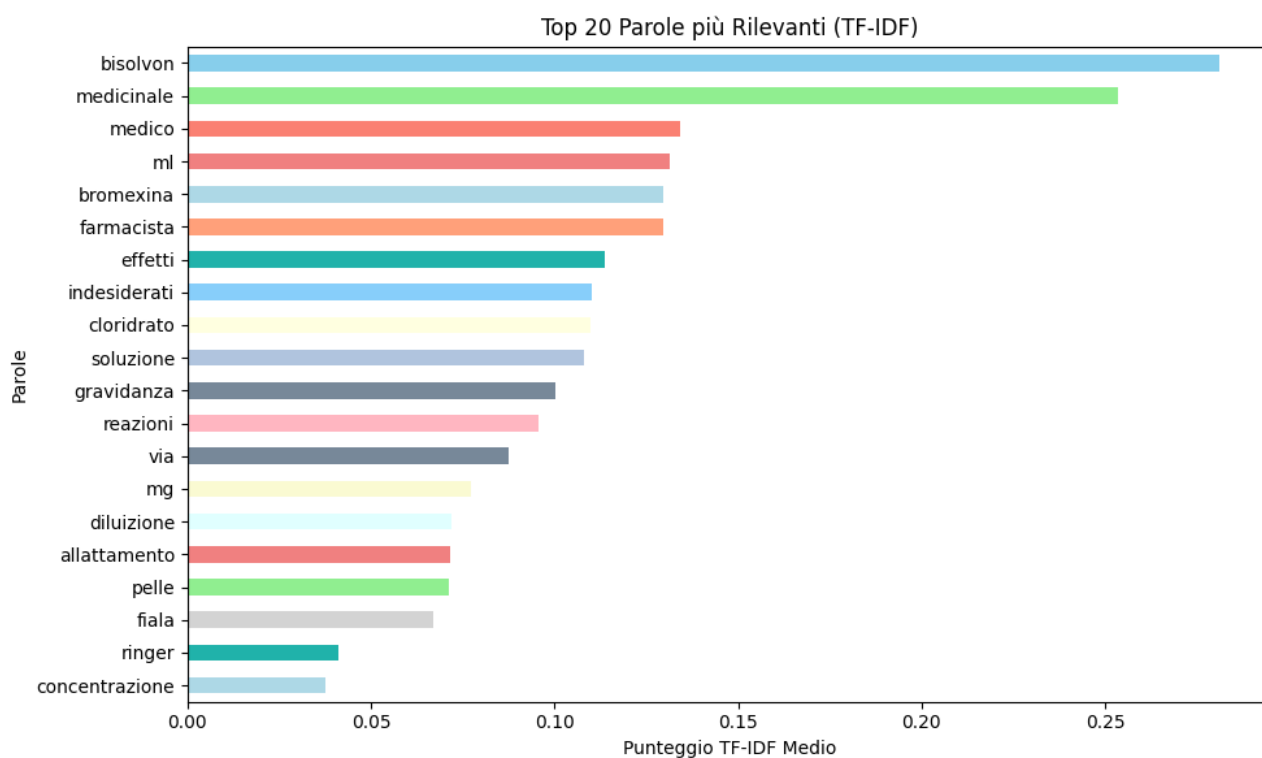


Figura 8 - Parole più rilevanti di Bisolvon

Nella *Figura 8* possiamo confermare quanto evidenziato dalla *Figura 6*. Infatti, tra le parole più rilevanti ci sono '*medico*' e '*farmacista*', entrambe figure che possono consigliare l'uso di questo farmaco nella maniera migliore possibile.

2.4.2.1. Bigrammi

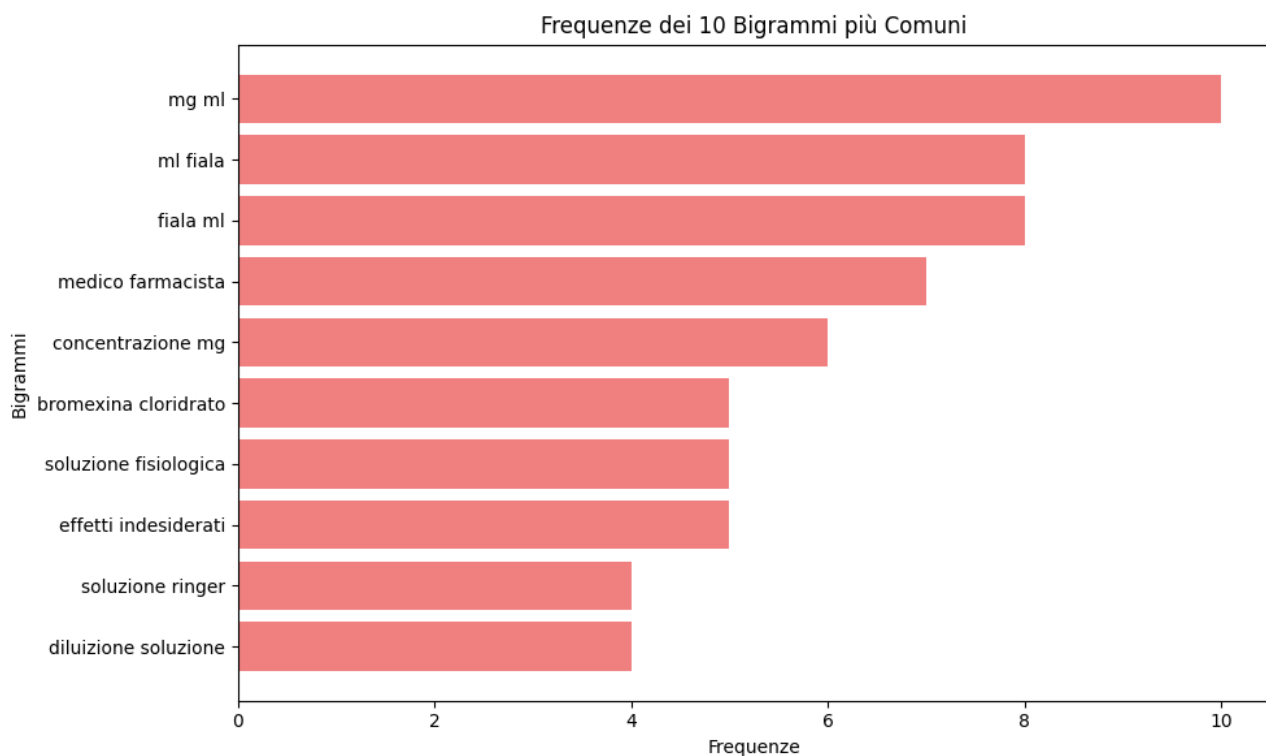


Figura 9 - 10 bigrammi più frequenti nel dataset di Bisolvon

"Mg ml" è il bigramma più frequente, seguito da "ml fiala" e "fiala ml", indicando termini tecnici legati alle dosi e volumi delle soluzioni. "Medico farmacista" è anch'esso comune, sottolineando l'importanza della consulenza medica. Altri bigrammi frequenti includono "concentrazione mg", "bromexina cloridrato" e "soluzione fisiologica". "Effetti indesiderati" e "soluzione ringer" evidenziano aspetti legati agli effetti collaterali e ai tipi di soluzioni utilizzate, mentre "diluizione soluzione" si riferisce ai processi di preparazione.

2.4.2.2. Trigrammi

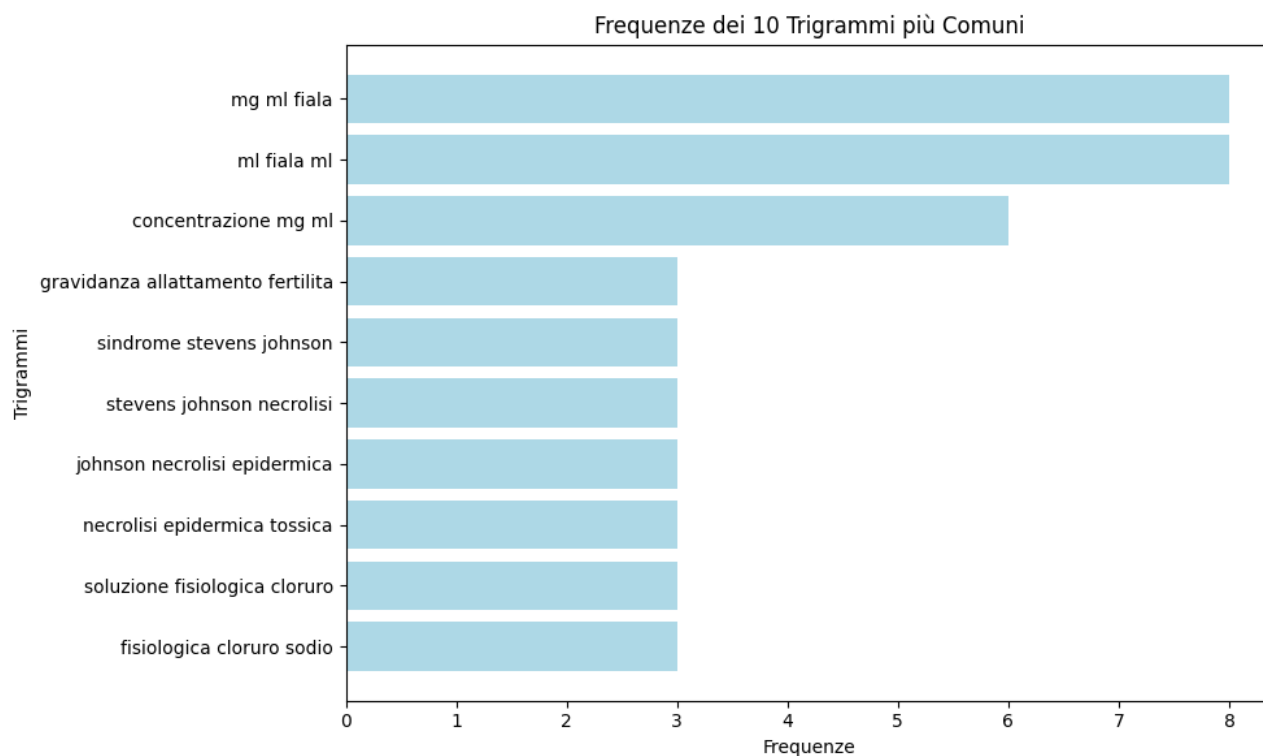


Figura 10 - 10 trigrammi più frequenti nel dataset di Bisolvon

I trigrammi comuni come "mg ml fiala" e "concentrazione mg ml" si riferiscono alle dosi e alle modalità di somministrazione del farmaco. Termini come "gravidanza allattamento fertilità" indicano sezioni dedicate alle avvertenze per l'uso in queste condizioni. La presenza di "sindrome Stevens Johnson" e "necrolisi epidermica tossica" suggerisce che il bugiardino include anche avvertenze su potenziali effetti collaterali gravi.

2.4.3. Borocillina

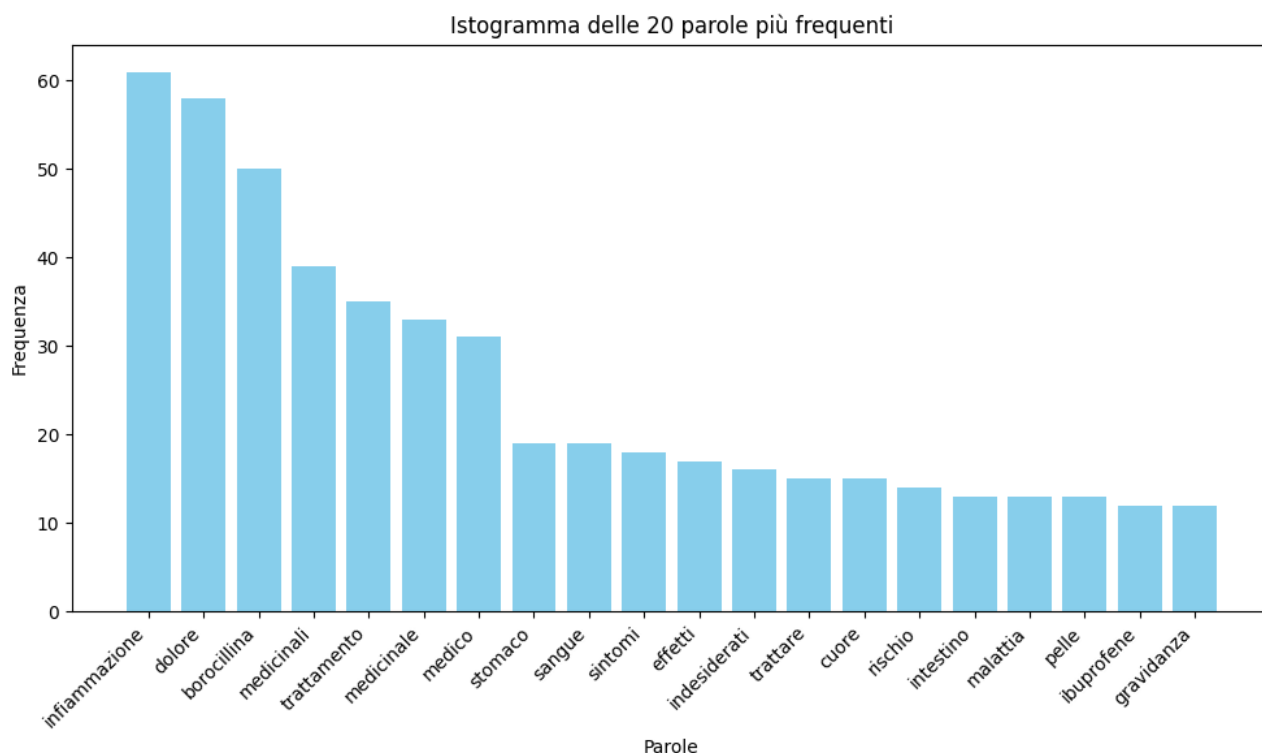


Figura 11 - Istogramma con le 20 parole più frequenti

Nella Figura 11 possiamo notare come tra le parole frequenti c'è il motivo per cui assumere il farmaco in questione, 'infiammazione', da chi dev'essere consigliato ('medico') e il trattamento da seguire per eliminare il sintomo.



Figura 12 - WordCloud di Borocillina

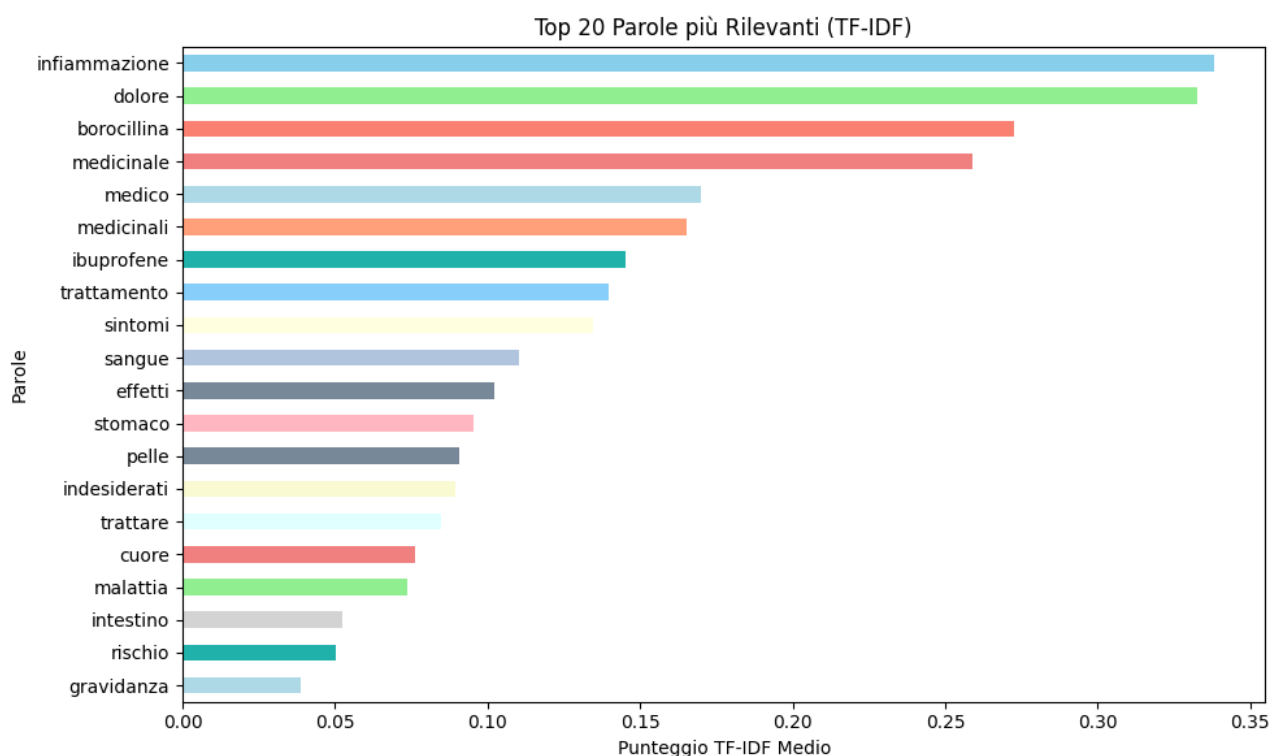


Figura 13 - Parole più rilevanti di Borocillina

L'*ibuprofene*, pur essendo penultima in termini di frequenza nel bugiardino, è ritenuta tra le parole più rilevanti in quanto essendo il principio attivo della Borocillina è colei che provvede alla risoluzione dell'infiammazione.

2.4.3.1. Bigrammi

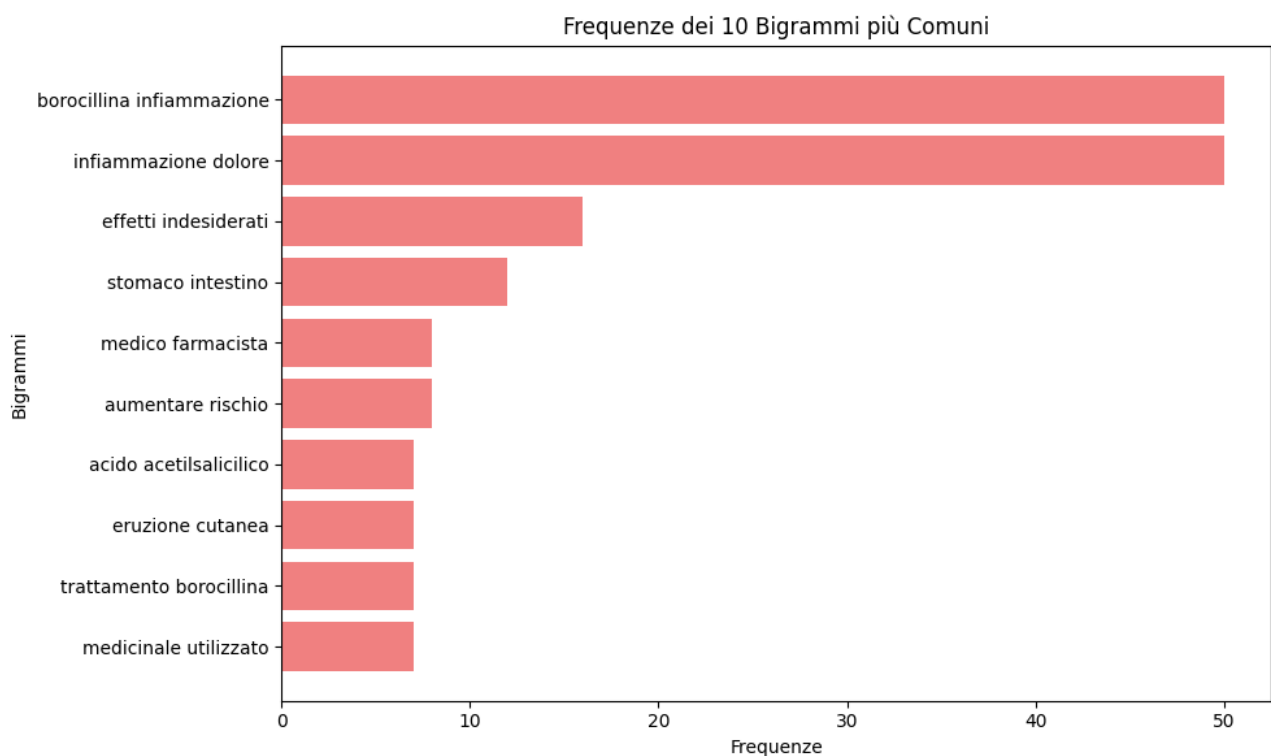


Figura 14 - 10 bigrammi più frequenti nel dataset di Borocillina

Il grafico mostra, con 50 occorrenze, termini legati all'uso e agli effetti della Borocillina, come "borocillina infiammazione" e "infiammazione dolore", evidenziando il suo utilizzo principale nel trattare infiammazioni e dolore. La frequenza di "effetti indesiderati" e "eruzione cutanea" indica l'importanza delle informazioni sugli effetti collaterali. Termini come "medico farmacista" e "aumentare rischio" sottolineano la necessità di consultare professionisti sanitari e avvertono dei rischi potenziali. L'attenzione a "stomaco intestino" riflette possibili effetti sul sistema gastrointestinale.

2.4.3.2. Trigrammi

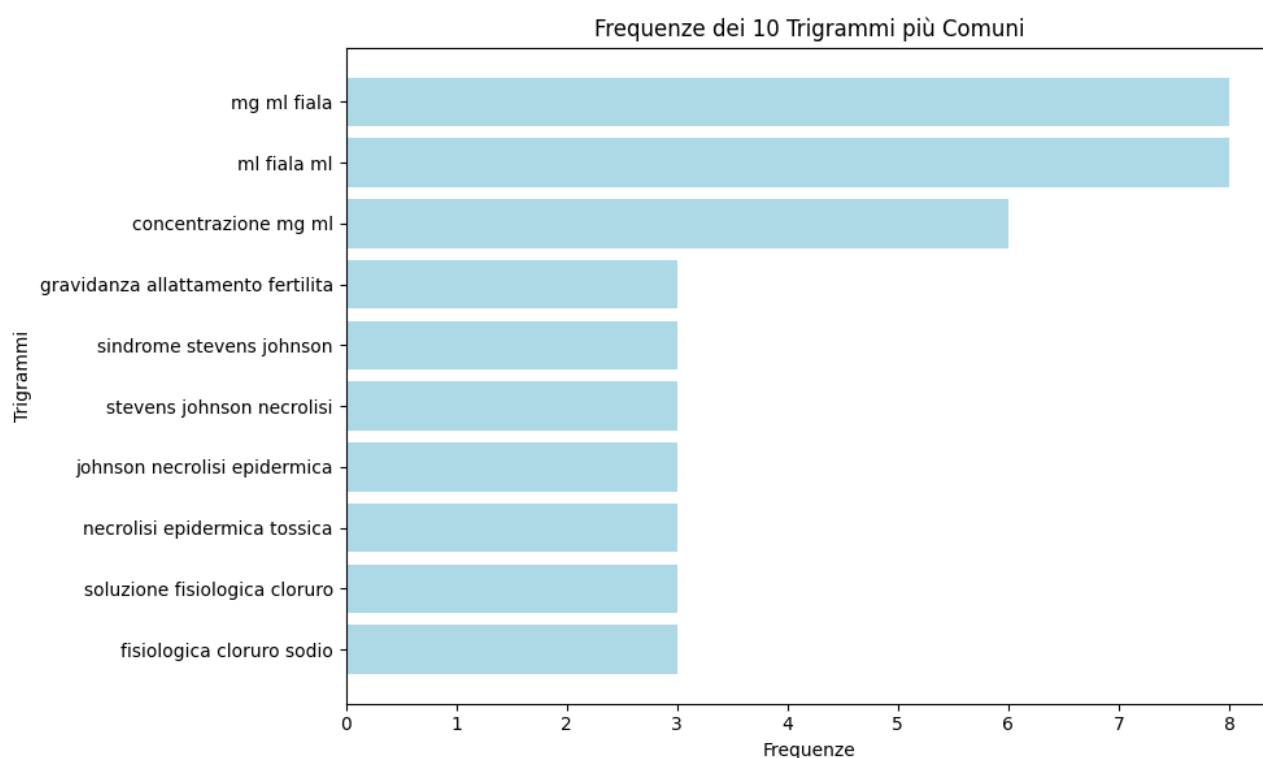


Figura 15 - 10 trigrammi più frequenti nel dataset di Borocillina

L'alta frequenza dei trigrammi relativi alle misurazioni del dosaggio (mg, ml, fiala) suggerisce un focus sulle istruzioni o prescrizioni dei farmaci. La presenza di termini come "gravidanza allattamento fertilita" (gravidanza, allattamento al seno, fertilità) e "sindrome di Stevens Johnson" indica che questo bugiardino probabilmente copre una serie di argomenti medici, tra cui la salute delle donne e condizioni mediche specifiche.

2.4.4. Dicloreeum

Come nei precedenti, anche in questo caso (*Figura 16*) la parola con frequenza maggiore è il nome del farmaco stesso preceduto solo dal sintomo principale per cui viene utilizzato. Pertanto, da questo primo grafico non si evidenziano particolari differenze sul contenuto del bugiardino.

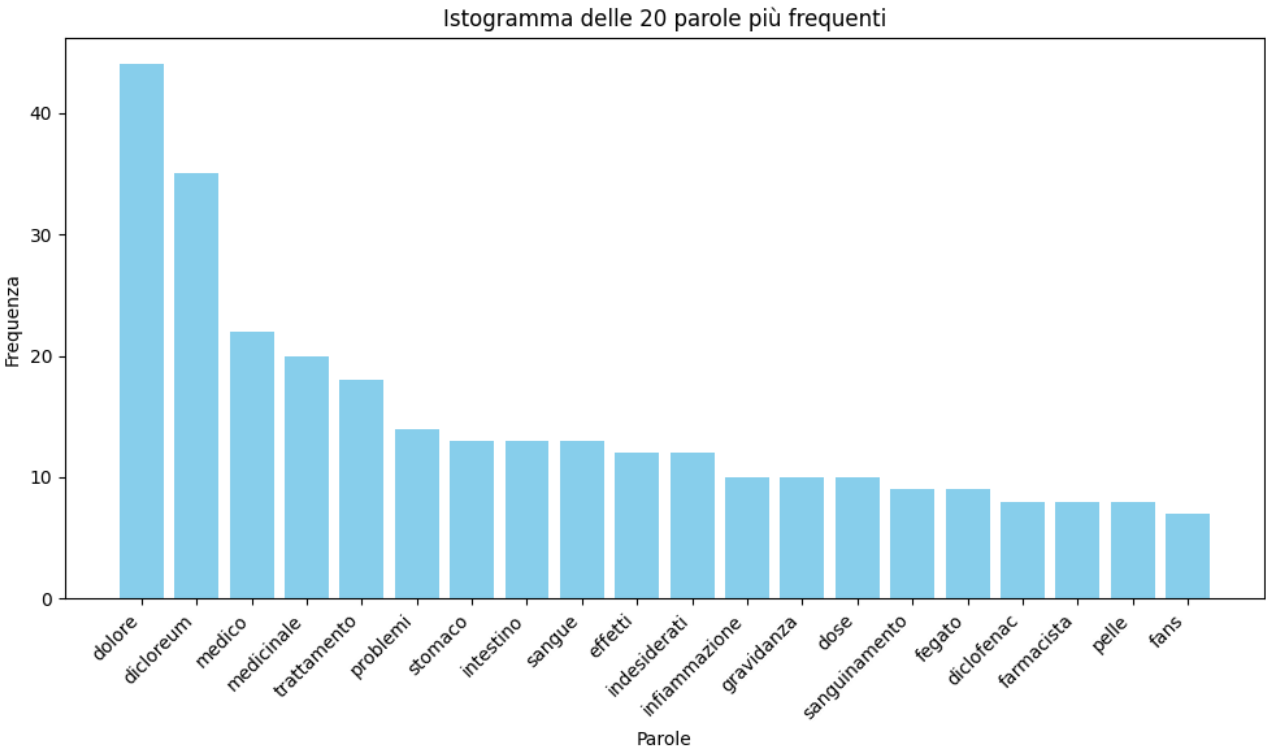


Figura 16 - Istogramma con le 20 parole più frequenti di Dicloream



Figura 17 - WordCloud di Dicloream

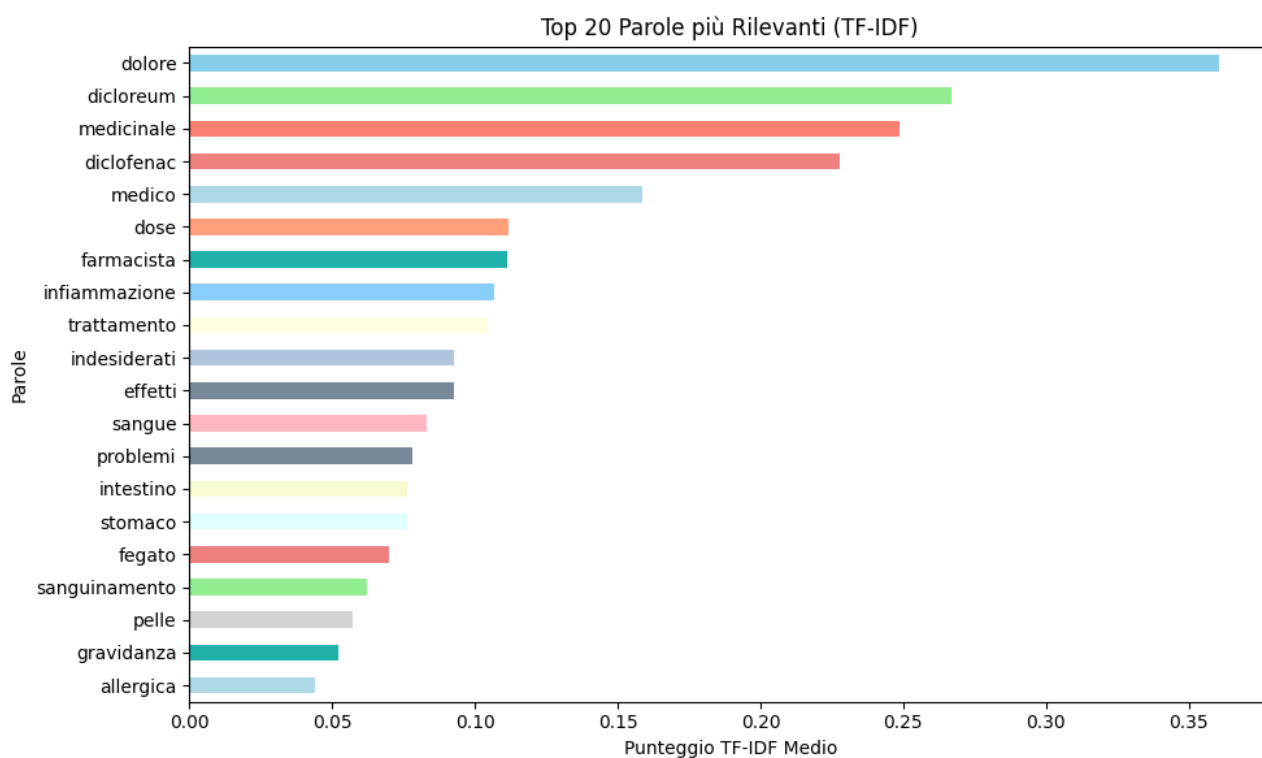


Figura 18 - Parole più rilevanti di Diclorem

Come nella *Figura 15*, anche qui il principio attivo è ritenuto molto rilevante rispetto alla frequenza con cui è stato trovato.

2.4.4.1. Bigrammi

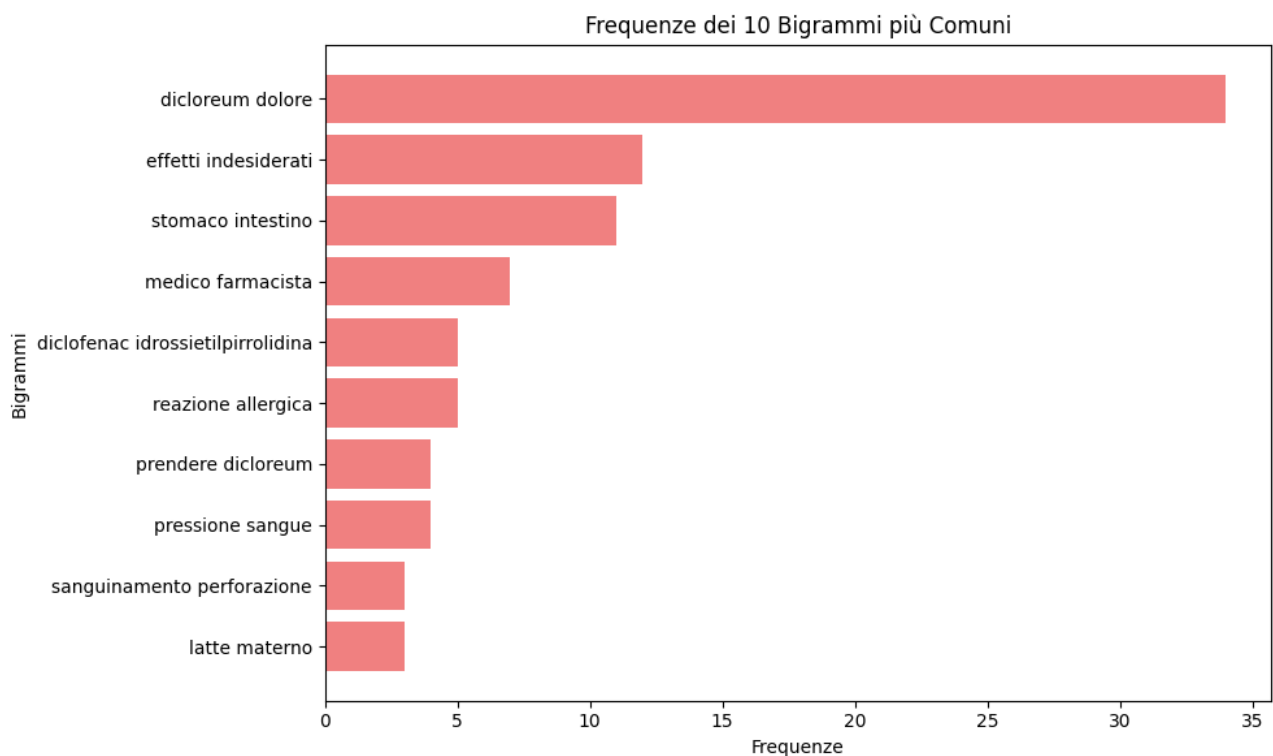


Figura 19 - 10 bigrammi più frequenti nel dataset Dicloream

"Dicloream dolore" è il bigramma più frequente, con oltre 30 occorrenze, indicando l'uso del farmaco per il trattamento del dolore. Seguono "effetti indesiderati" e "stomaco intestino", che evidenziano i possibili effetti collaterali e l'area di impatto del farmaco. "Medico farmacista" è anch'esso comune, sottolineando l'importanza della consulenza medica. Altri bigrammi frequenti includono "diclofenac idrossietilpirrolidina", "reazione allergica" e "prendere dicloream", riferendosi a specifiche sostanze attive, reazioni avverse e istruzioni d'uso. Infine, "pressione sangue", "sanguinamento perforazione" e "latte materno" indicano preoccupazioni specifiche e avvertenze importanti relative al farmaco.

2.4.4.2. Trigrammi

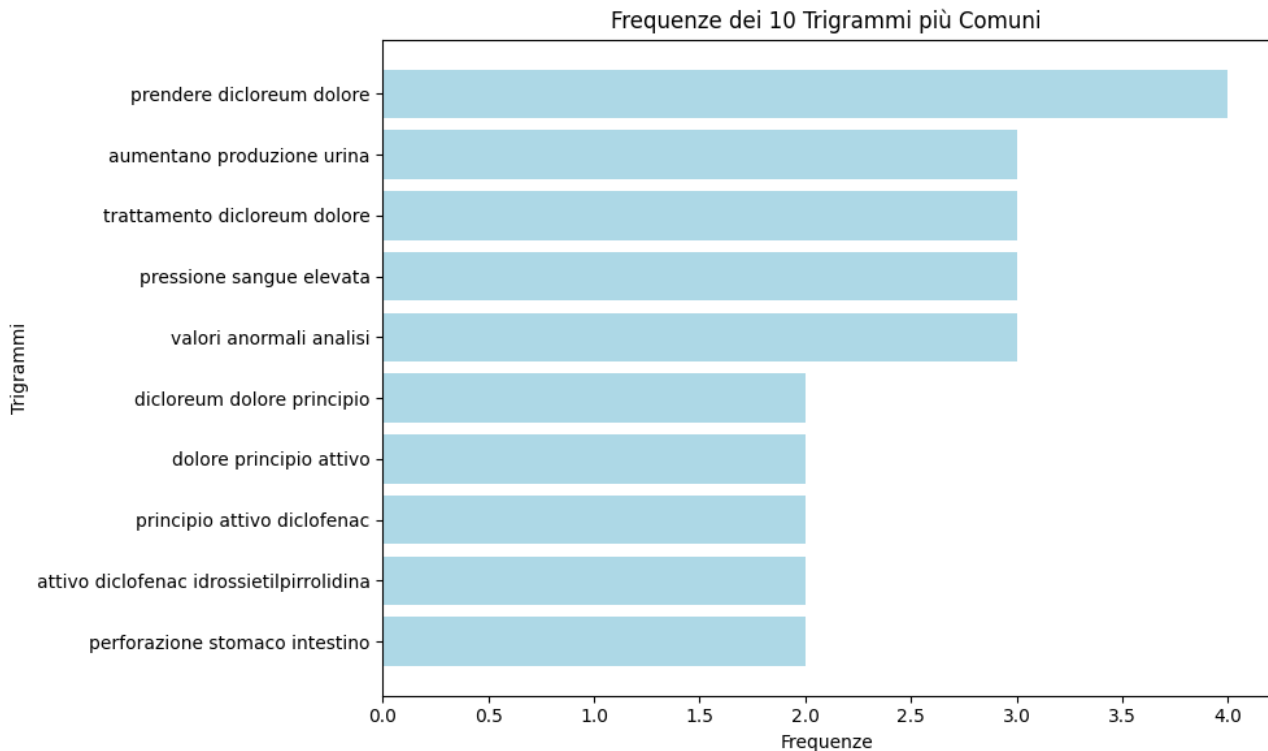


Figura 20 - 10 trigrammi più frequenti nel dataset di Dicloream

"Prendere dicloream dolore" è il trigramma più ricorrente, seguito da "aumentano produzione urina" e "trattamento dicloream dolore". "Pressione sangue elevata" e "valori anormali analisi" sono anch'essi frequenti, indicando preoccupazioni mediche specifiche. Altri trigrammi importanti includono "dicloream dolore principio" e "dolore principio attivo", riferendosi alla composizione e all'uso del farmaco. "Principio attivo diclofenac" e "attivo diclofenac idrossietilpirrolidina" sottolineano i componenti chimici rilevanti, mentre "perforazione stomaco intestino" segnala potenziali gravi effetti collaterali.

2.4.5. Fentanyl

Il farmaco in questione, così come la Morfina, è estremamente nocivo per la salute se preso senza cautele e la prescrizione e supervisione di un medico. Non è un caso, infatti, che tra le parole con frequenza maggiore ci siano proprio *'microgrammi'* e *'dose'* seguiti da *'medico'*.

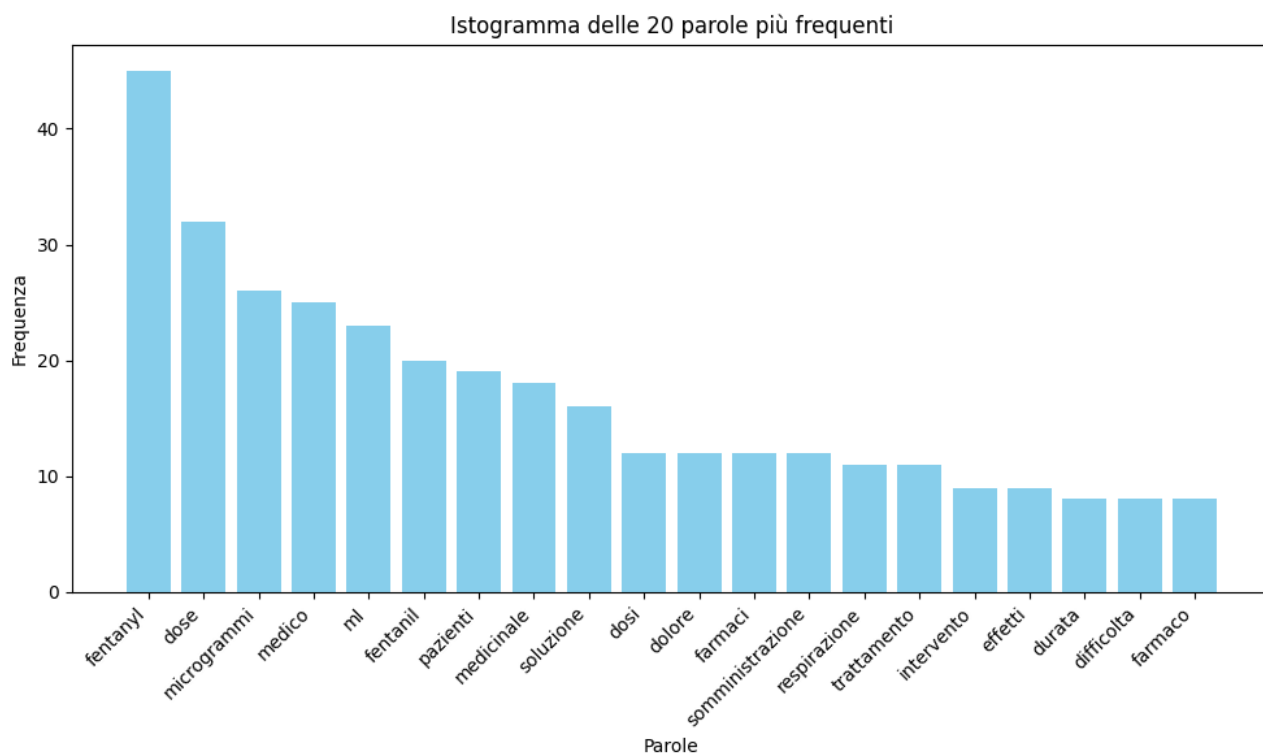


Figura 21- Istogramma con le 20 parole più frequenti di Fentanyl



Figura 22 - WordCloud di Fentanyl

Da questo WordCloud, oltre a quanto già detto, possiamo capire anche in che modo questo farmaco viene somministrato e da cosa sia composto (*'oppioidi'*).

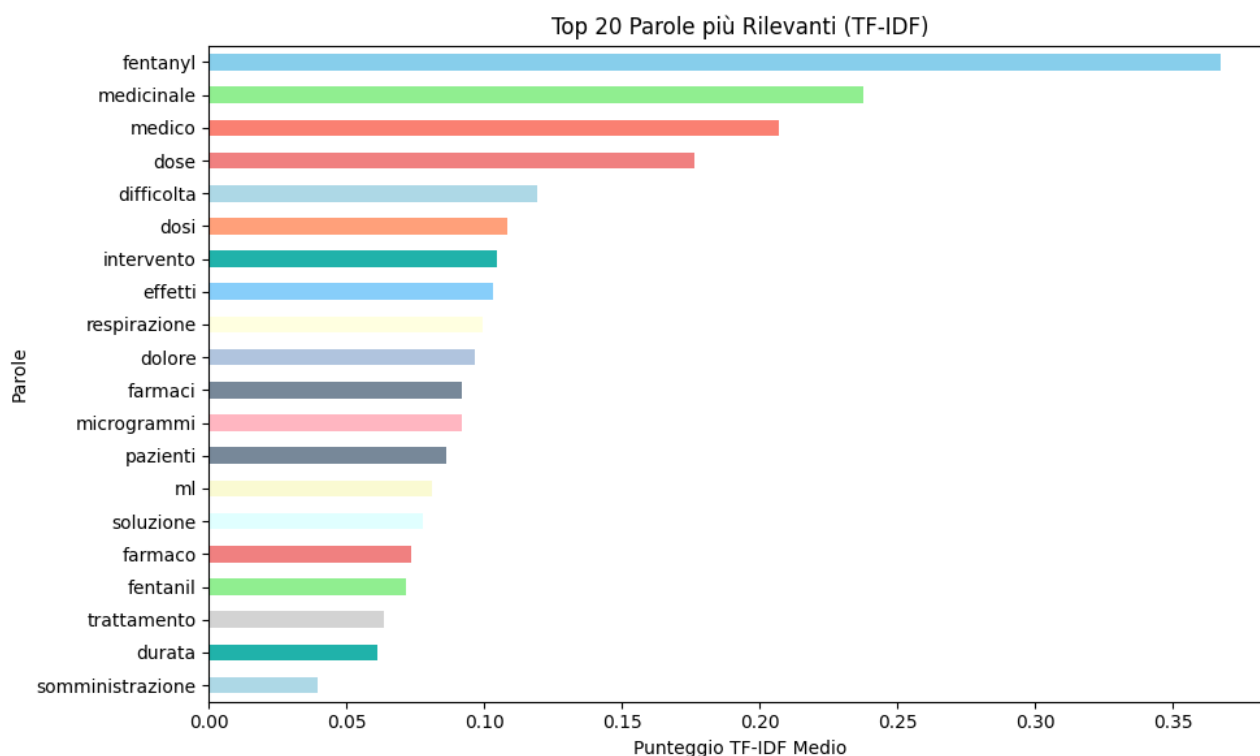


Figura 23 - Parole più rilevanti di Fentanyl

Tra le parole più rilevanti, come per la *Figura 21*, c'è proprio la parola '*medico*' seguito da '*dose*'. Questa osservazione conferma come questo farmaco abbia bisogno di una prescrizione medica e un'attenzione elevata al quantitativo da somministrare. Infatti, la parola '*effetti*' è ritenuta più rilevante di altre nonostante sia meno frequente. Ciò sta a significare che, se usato con sregolatezza, questo farmaco può diventare nocivo per la salute del paziente.

2.4.5.1. Bigrammi

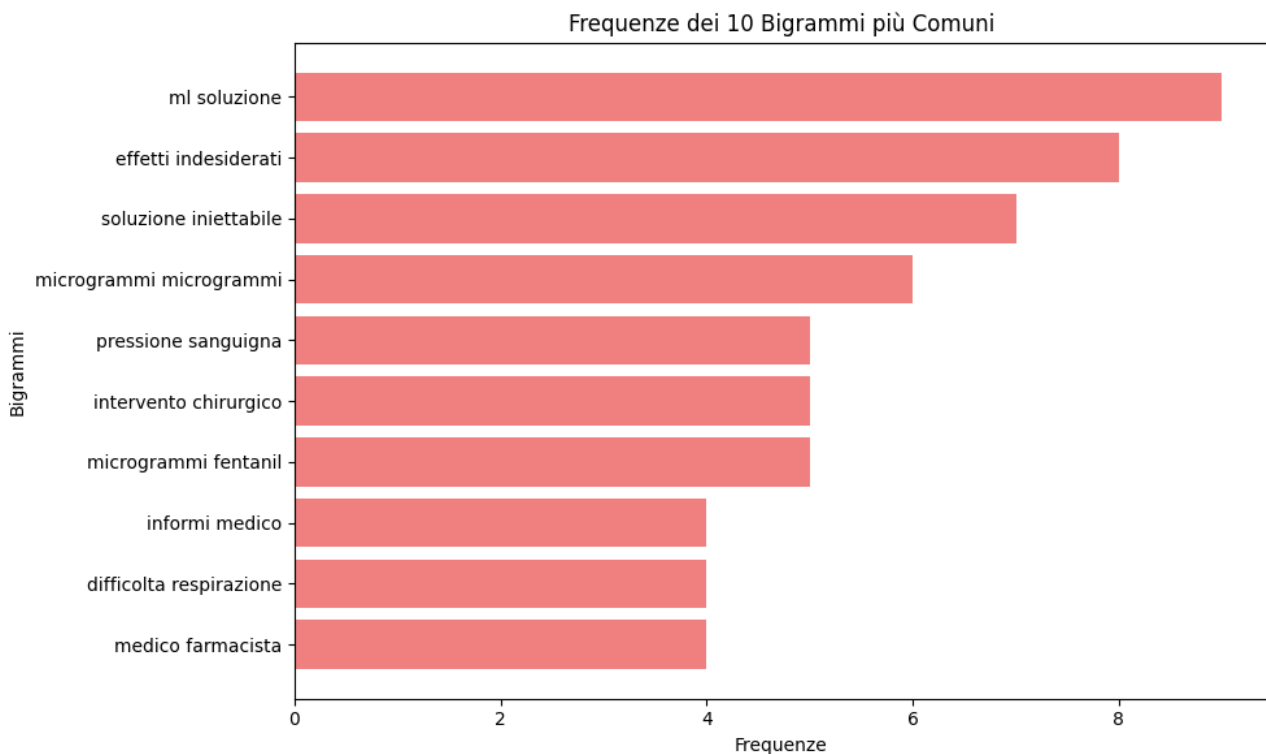


Figura 24 - 10 bigrammi più frequenti nel dataset di Fentanyl

"ml soluzione" è il bigramma più frequente, seguito da "effetti indesiderati" e "soluzione iniettabile", indicando l'uso di soluzioni e le preoccupazioni relative agli effetti collaterali. "Microgrammi microgrammi" e "pressione sanguigna" sono anch'essi comuni, evidenziando le dosi e gli effetti sul sistema cardiovascolare. Altri bigrammi frequenti includono "intervento chirurgico" e "microgrammi fentanil", riferendosi a procedure mediche e dosaggi specifici. "Informi medico" e "difficoltà respirazione" sottolineano la necessità di consultare un medico e i potenziali problemi respiratori, mentre "medico farmacista" rimarca l'importanza della consulenza professionale.

2.4.5.2. Trigrammi

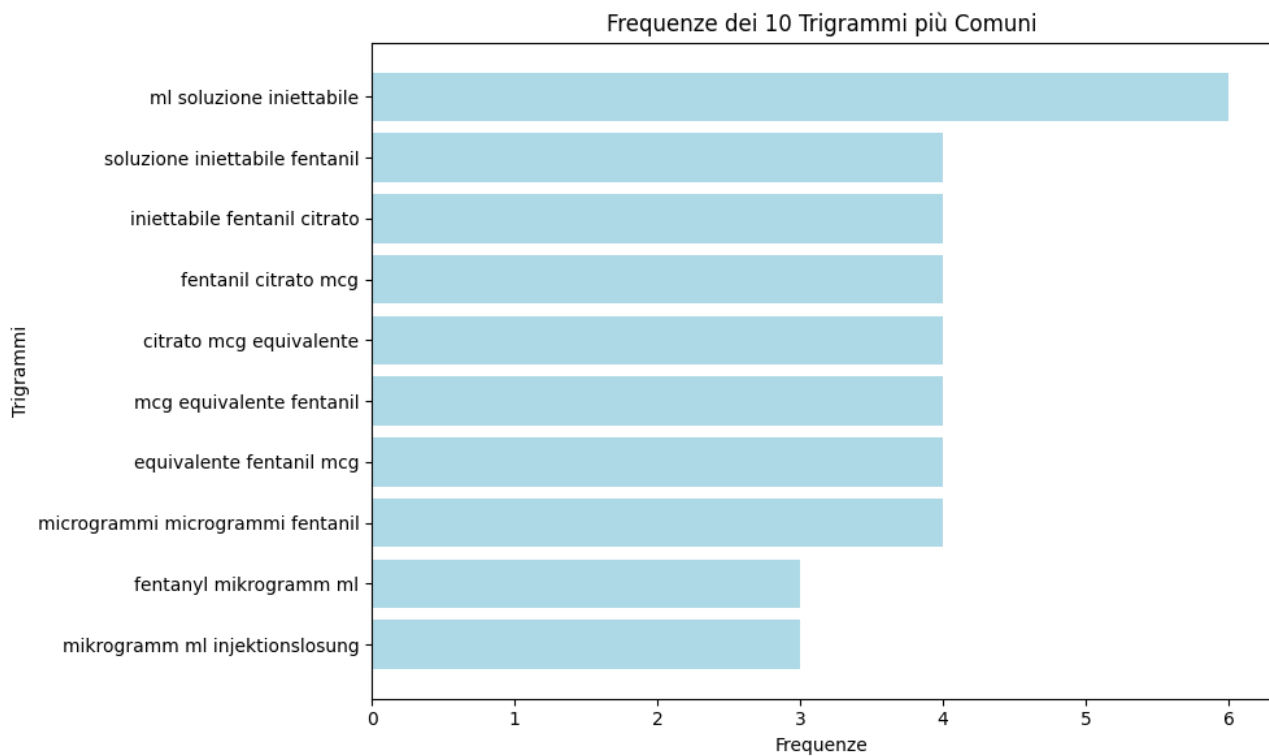


Figura 25 - 10 trigrammi più frequenti nel dataset di Fentanyl

I trigrammi più frequenti sono prevalentemente termini tecnici legati alla somministrazione ed al dosaggio del farmaco, come "ml soluzione iniettabile" e "soluzione iniettabile fentanyl". Questi termini indicano che il contesto principale del bugiardino riguarda le istruzioni per l'uso del farmaco in forma iniettabile. Inoltre, la presenza di trigrammi in altre lingue, come "mikrogramm ml Injektionslösung", suggerisce che il documento è rivolto a un pubblico internazionale. Infine, la ripetizione di parole come "microgrammi" evidenzia l'importanza di specificare le dosi e le equivalenze per un uso sicuro ed efficace del fentanyl.

3. Modelli

I modelli utilizzati per condurre lo studio sui cinque bugiardini sono stati cinque: tre per la Sentiment Analysis, uno per l'Emotion Analysis e uno per il Topic Modelling. La difficoltà principale è stata quella di trovare dei modelli che si adattassero al lessico dei bugiardini selezionati, l'italiano. Di seguito l'elenco dei modelli.

Sentiment Analysis, effettuata tramite i modelli:

- Vader
- umBERTo
- SentITA

Emotion Analysis, effettuata con:

- FEEL-IT

Topic modelling, effettuata con:

- LDA (Latent Dirichlet Allocation)

3.1. Vader

VADER è un modello di Sentiment Analysis basato su regole e lessico, originariamente sviluppato per l'inglese e successivamente adattato all'italiano. Utilizza un dizionario di parole comuni e frasi con punteggi di sentiment predeterminati che indicano positività, negatività o neutralità.

È particolarmente efficace per analizzare testi brevi e informali, ma può essere adattato anche a contesti più formali. La semplicità del modello lo rende veloce ed efficiente, fornendo risultati immediati senza la necessità di un grande volume di dati per l'addestramento.

3.2. umBERTo

UmBERTo è un modello di linguaggio basato su transformer, sviluppato da Musixmatch, pre-addestrato su una vasta quantità di testo italiano. La sua architettura è derivata da roBERTa.

Questo modello è in grado di comprendere contesti complessi e sfumature linguistiche grazie alla sua architettura avanzata, che consente di catturare le dipendenze a lungo termine tra le parole.

La capacità di umBERTo di analizzare testi complessi e di lungo respiro lo rende adatto per applicazioni in cui è necessario un alto livello di comprensione del contesto. È efficace nel gestire linguaggi tecnici e formali, grazie alla sua formazione su una vasta gamma di testi italiani.

UmBERTo ha una capacità massima di 514 token, quindi è stato limitato il contenuto delle sezioni a 255 parole per garantire la compatibilità.

3.3. sentITA

SentITA è un modello di Sentiment Analysis sviluppato specificamente per l'italiano. Utilizza un lessico di parole italiane con punteggi di sentiment associati, derivati da un'analisi linguistica

approfondita. Il modello assegna un valore positivo o negativo a ciascuna parola in base al suo significato emotivo.

SentITA è costruito per essere altamente accurato nella valutazione dei sentimenti espressi in italiano, includendo parole e frasi specifiche della cultura e del contesto italiano. È particolarmente utile per analizzare testi in cui è importante comprendere le sfumature linguistiche locali.

3.4. FEEL-IT

Feel-it è un modello di Emotion Analysis per l'italiano che sfrutta tecniche di deep learning. Addestrato su un ampio corpus di testi italiani, Feel-it utilizza reti neurali per riconoscere e interpretare le emozioni con alta precisione. È progettato per comprendere le espressioni linguistiche e le emozioni in vari contesti.

La potenza del deep learning permette a Feel-it di captare una vasta gamma di emozioni e toni, anche in testi complessi e articolati. Questo modello è estremamente accurato nel rilevare sfumature emotive, rendendolo uno strumento versatile per analisi dettagliate e approfondite delle emotion.

3.5. Latent Dirichlet Allocation (LDA)

LDA (Latent Dirichlet Allocation) è un modello statistico utilizzato per scoprire i temi nascosti presenti in un insieme di documenti. Funziona assegnando a ciascun documento una distribuzione di probabilità su un insieme di argomenti e ad ogni parola una distribuzione di probabilità su un insieme di argomenti. L'obiettivo è trovare una combinazione di argomenti che spieghi la distribuzione di parole osservata nei documenti. LDA assume che ciascun documento contenga un mix di argomenti e ciascun argomento abbia una distribuzione sulle parole. Utilizzando l'approccio di Bayes, LDA inferisce i parametri del modello per massimizzare la verosimiglianza dei dati osservati, consentendo di identificare automaticamente i temi principali presenti nei documenti senza supervisione esplicita.

3.6. Metriche

Per valutare il sentiment del contenuto all'interno dei vari paragrafi sono state utilizzati i *punteggi di polarità*:

- *Negativo (neg)*: Indica che il testo esprime principalmente sentiment negativi, come insoddisfazione, critica o preoccupazione.
- *Neutro (neu)*: Indica che il testo non esprime alcun sentiment particolare o che non è possibile determinare un sentiment chiaro o rilevante.
- *Positivo (pos)*: Identifica contenuti che sono apprezzati o ben accolti.
- *Compound (per VADER)*: Nel contesto del modello VADER, il termine "compound" rappresenta il sentiment complessivo di un testo, calcolato come una combinazione della polarità delle parole presenti nel testo. Questo punteggio composto tiene conto dell'intensità e della combinazione di parole positive e negative nel testo. Il suo valore varia tra -1 (sentiment estremamente negativo) e +1 (sentiment estremamente positivo).

Per valutare la coesione semantica degli argomenti estratti tramite modelli di topic modelling (nel nostro caso LDA) si è utilizzato il **metodo UMass**:

Esso si basa sull'idea che i topic coesi contengano parole che tendono a co-occorrere frequentemente all'interno dei documenti. Per calcolare la coerenza UMass, si considerano le co-occorrenze delle parole all'interno di un singolo documento, valutando quanto queste co-occorrenze supportino l'idea di un topic univoco. Questo metodo non richiede conoscenze esterne come corpora di testo aggiuntivi, ma si limita all'analisi dei documenti stessi per determinare la coesione dei topic. I punteggi più alti di coerenza UMass indicano una maggiore coesione semantica all'interno dei topic estratti.

Lo score è sempre negativo. Pertanto, in valore assoluto, minore è lo score calcolato più coerenza c'è.

4. Risultati

Sono stati calcolati i punteggi di polarità per tutte le sezioni presenti nei vari dataframe, ottenendo valori per la negatività, la neutralità, la positività e il punteggio complessivo di ciascuna (nel caso di Vader). Successivamente, si sono combinati questi dati con il dataframe originale, creando un nuovo dataframe che contiene i punteggi di polarità per le varie sezioni.

Nell'analisi del sentiment all'interno delle sezioni di ogni bugiardino, sono stati confrontati i risultati ottenuti utilizzando i modelli di elaborazione del linguaggio naturale umBERTo e sentITA. Questo confronto ci ha permesso di valutare le differenze e le somiglianze nei punteggi di polarità e sentiment generati da ciascun modello, offrendo un quadro dettagliato sull'affidabilità e l'accuratezza delle due tecnologie nel contesto specifico del testo analizzato.

4.1. Morfina

4.1.1. Vader

Applicando il modello Vader a *df_morfina*, i risultati del sentiment sul contenuto delle varie sezioni del bugiardino sono le seguenti:

Numero	neg	neu	pos	compound	Sezione
1	0.189	0.736	0.075	-0.8316	che cos'è morfina cloridrato s.a.l.f. e a cosa serve
2	0.210	0.707	0.083	-0.9994	cosa deve sapere prima di usare morfina cloridrato
3	0.182	0.725	0.092	-0.9952	come usare morfina cloridrato s.a.l.f
4	0.261	0.620	0.119	-0.9896	possibili effetti indesiderati
5	0.090	0.653	0.257	0.8750	come conservare morfina cloridrato s.a.l.f.
6	0.000	0.863	0.137	0.9231	contenuto della confezione e altre informazioni

Da questi primi risultati tutti i paragrafi risultano essere in prevalenza neutri ma i negativi sono maggiori dei positivi nei primi quattro paragrafi e viceversa negli ultimi due.

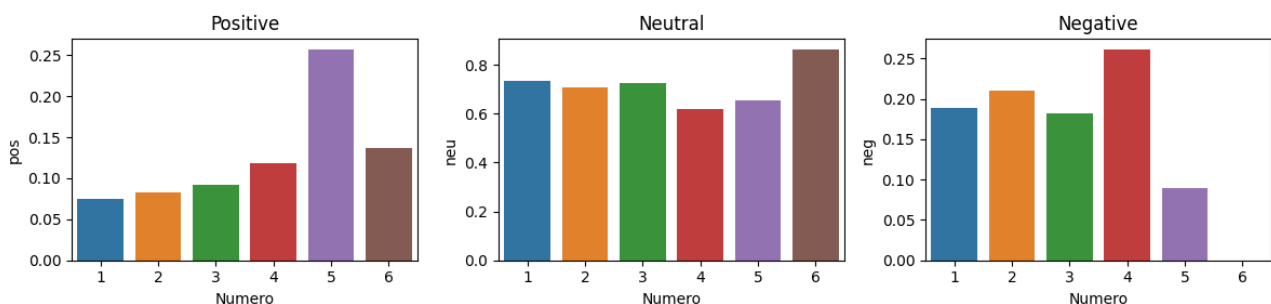


Figura 26 - Sentiment positivo, neutro, negativo con Vader su dataset di Morfina

Il compound score delinea un sentiment estremamente positivo negli ultimi due paragrafi che riguardano il come conservare il prodotto e informazioni aggiuntive, pertanto nulla che provochi negatività. D'altra parte, i primi quattro, che descrivono da come prendere il prodotto a quali rischi si corrono nel prendere quel determinato farmaco, hanno un sentiment estremamente negativo. Questo risultato è conseguenza dell'uso di parole 'negative' per mettere in guardia il consumatore su eventuali rischi.

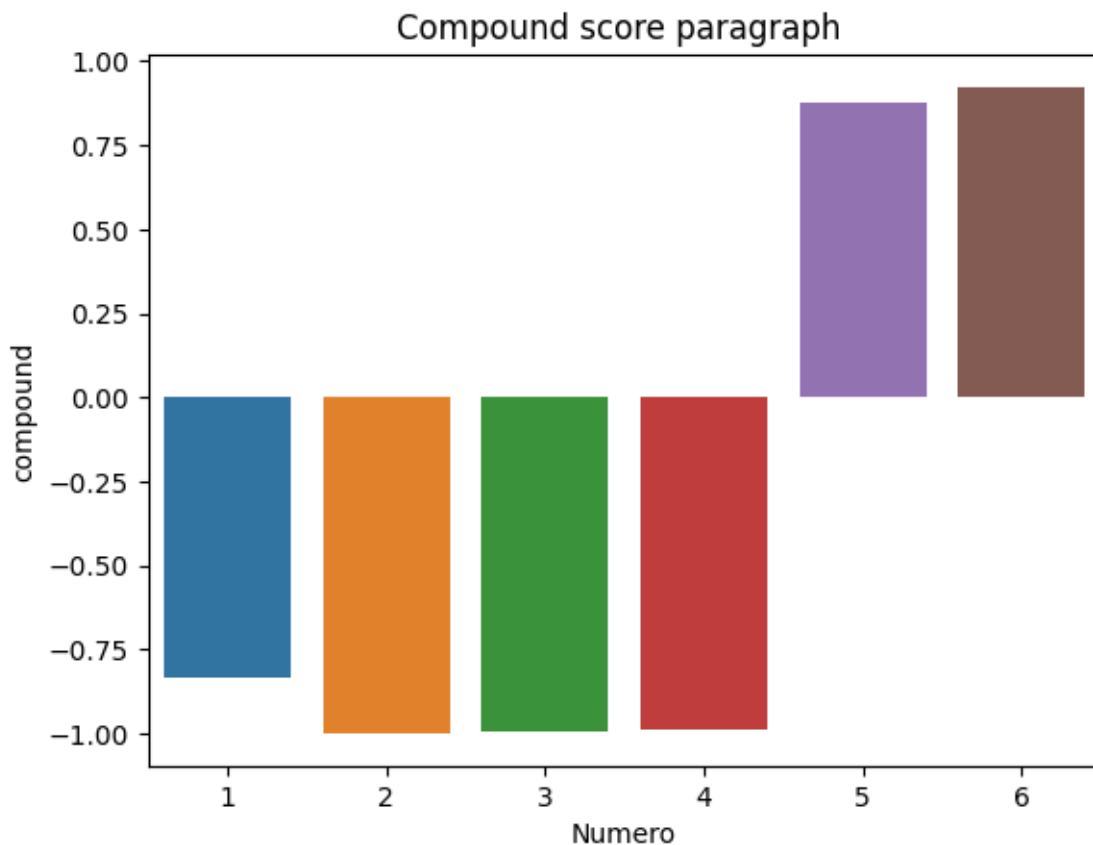


Figura 27 - Compound Score con Vader su dataset di Morfina

4.1.2. umBERTo

Applicando il modello umBERTo a *df_morfina*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.5093	0.4906	che cos'è morfina cloridrato s.a.l.f. e a cosa serve
2	0.4971	0.5028	cosa deve sapere prima di usare morfina cloridrato
3	0.5024	0.4975	come usare morfina cloridrato s.a.l.f
4	0.5024	0.4975	possibili effetti indesiderati
5	0.5049	0.4950	come conservare morfina

			cloridrato s.a.l.f.
6	0.5024	0.4957	contenuto della confezione e altre informazioni

Per questo modello è stato possibile prendere solo le categorie di polarità 'neg' e 'pos'.

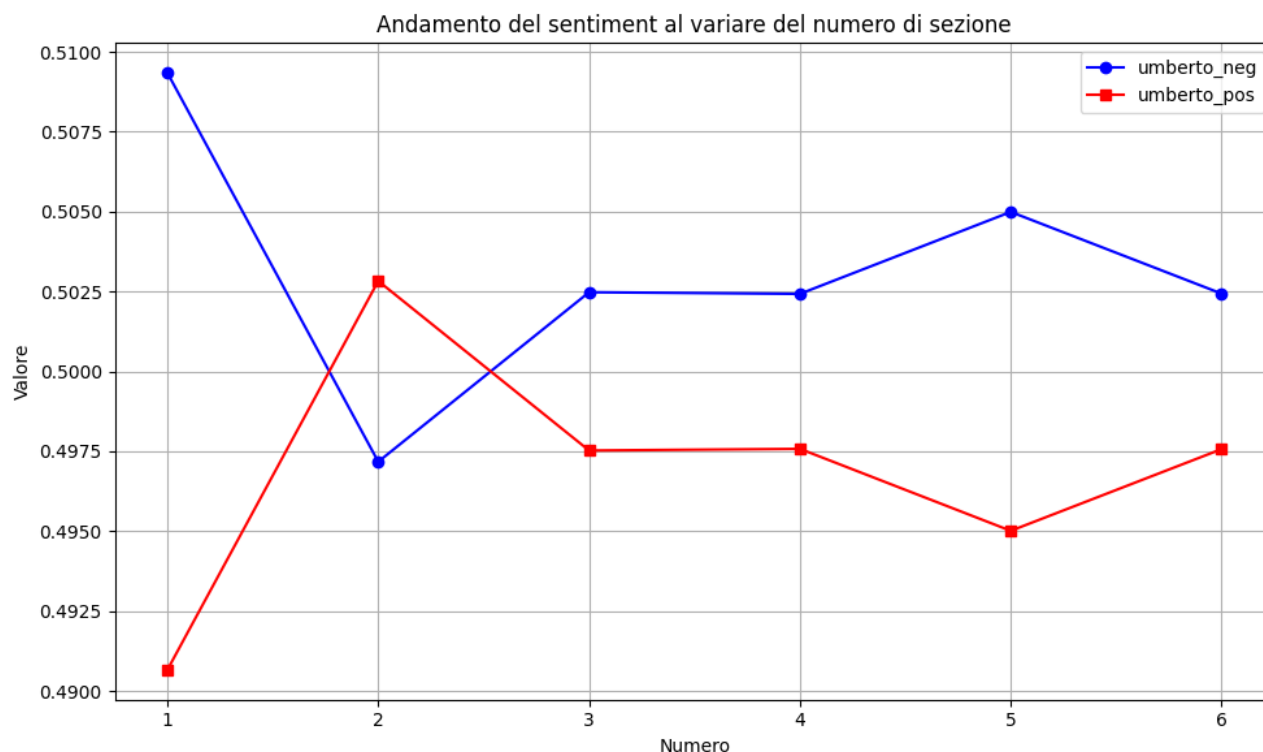


Figura 28 - Variazione del sentiment con umBERTo sul dataset di Morfina

In questo caso il sentiment è negativo tranne che per la sezione due che è leggermente positivo. È importante specificare però che non ci sono discrepanze evidenti tra il sentiment negativo e positivo, anzi. Questo sta ad indicare che in tutte le sezioni non c'è una categoria di sentiment che predomina sull'altra.

4.1.3. SentITA

Applicando il modello SentITA a *df_morfina*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.4702	0.5297	che cos'è morfina cloridrato s.a.l.f. e a cosa serve
2	0.4999	0.5000	cosa deve sapere prima di usare morfina cloridrato
3	0.4556	0.5443	come usare morfina cloridrato s.a.l.f
4	0.4998	0.5001	possibili effetti indesiderati
5	0.4999	0.5000	come conservare morfina cloridrato s.a.l.f.
6	0.4999	0.5000	contenuto della confezione e

			altre informazioni
--	--	--	--------------------

Anche per questo modello è stato possibile prendere solo le categorie di polarità *'neg'* e *'pos'*.

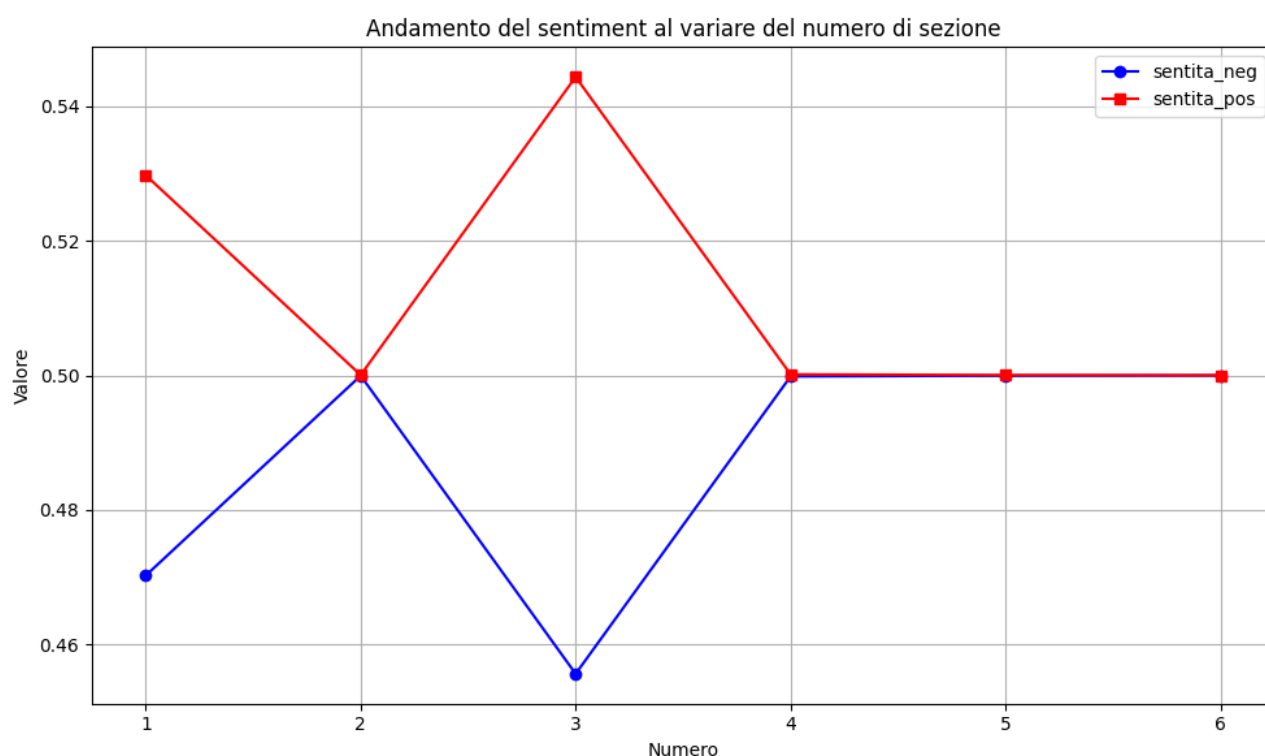


Figura 29 - Variazione del sentiment con SentITA sul dataset di Morfina

I risultati del sentiment tramite questo modello rappresentano, in parte, quello visto nella *Figura 26* con il modello Vader. Infatti, la maggior parte delle sezioni risulta essere neutra in quanto il sentiment negativo e positivo si attesta su 50-50.

4.1.4. Confronto risultati tra i modelli umBERTo e sentiTA

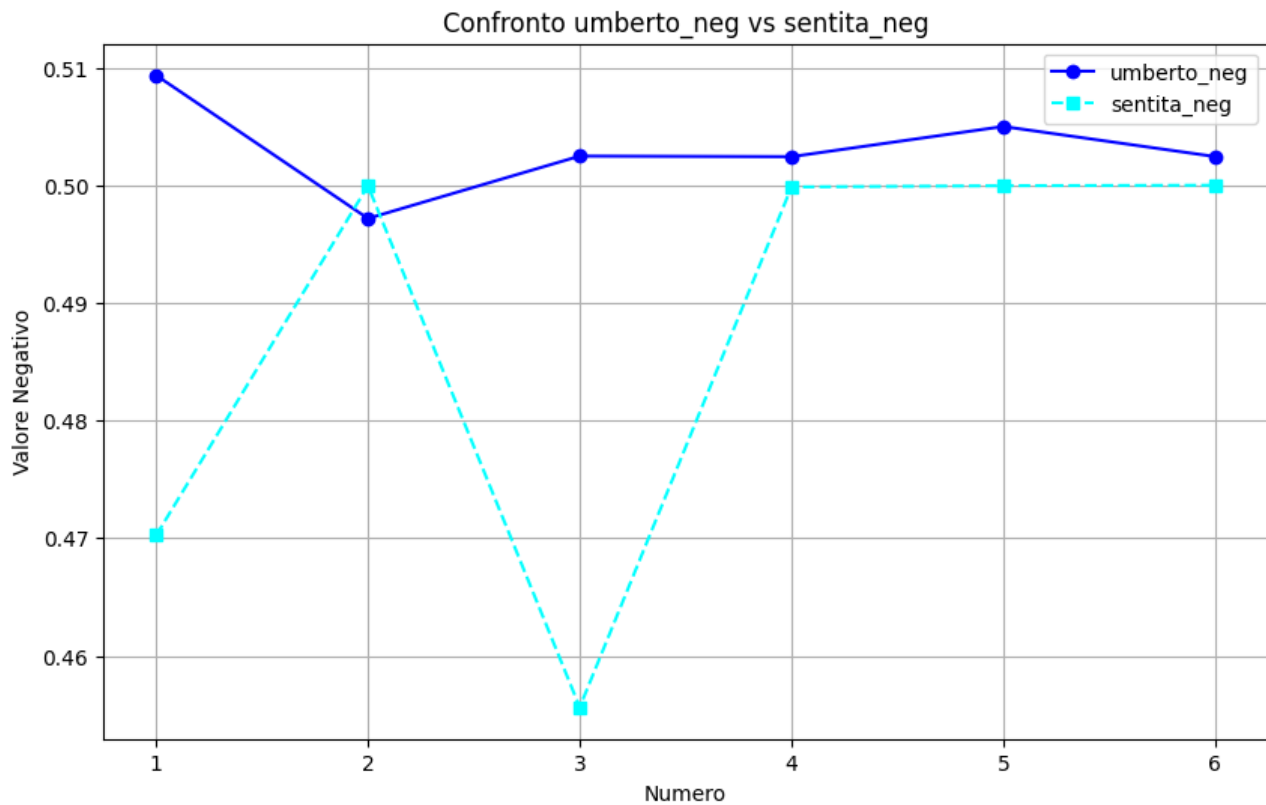


Figura 30 - Confronto sentiment 'neg' tra modelli sentiTA e umBERTo sul dataset di Morfina

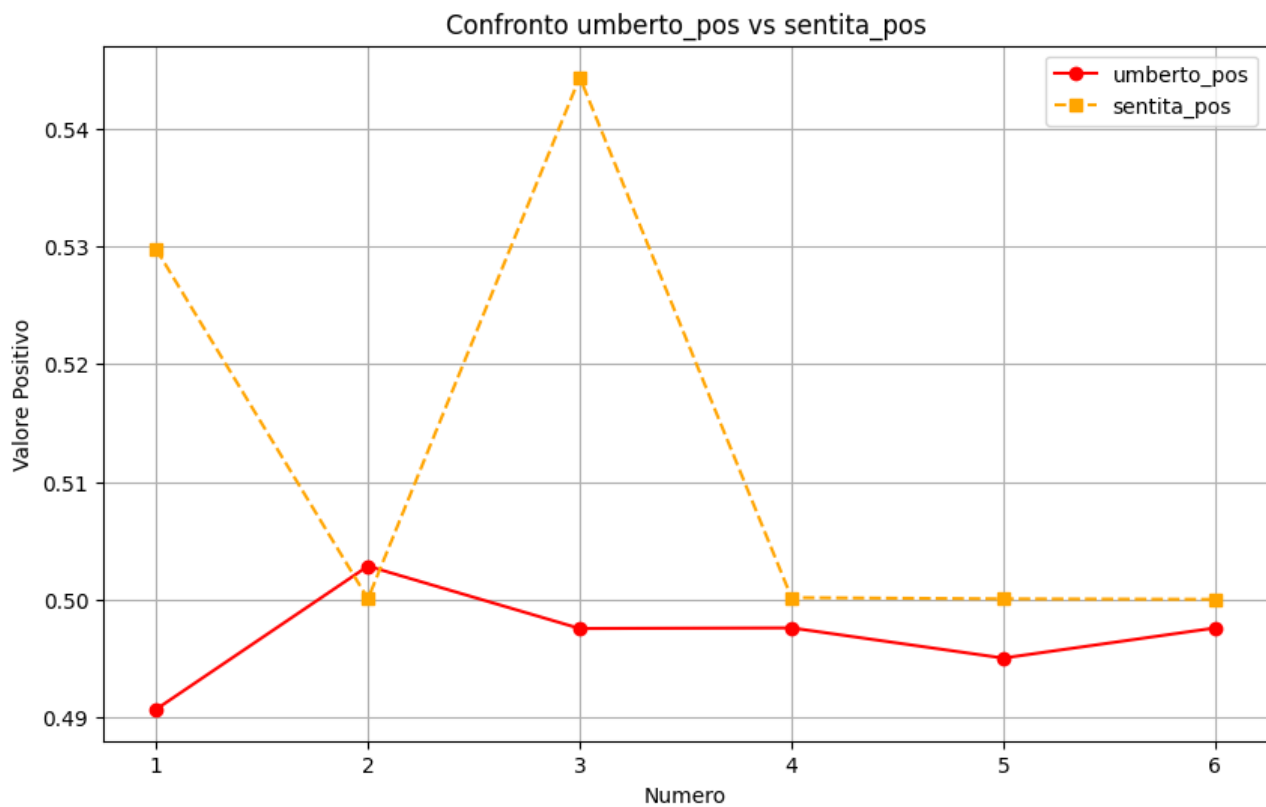


Figura 31 - Confronto sentiment 'pos' tra modelli sentiTA e umBERTo sul dataset di Morfina

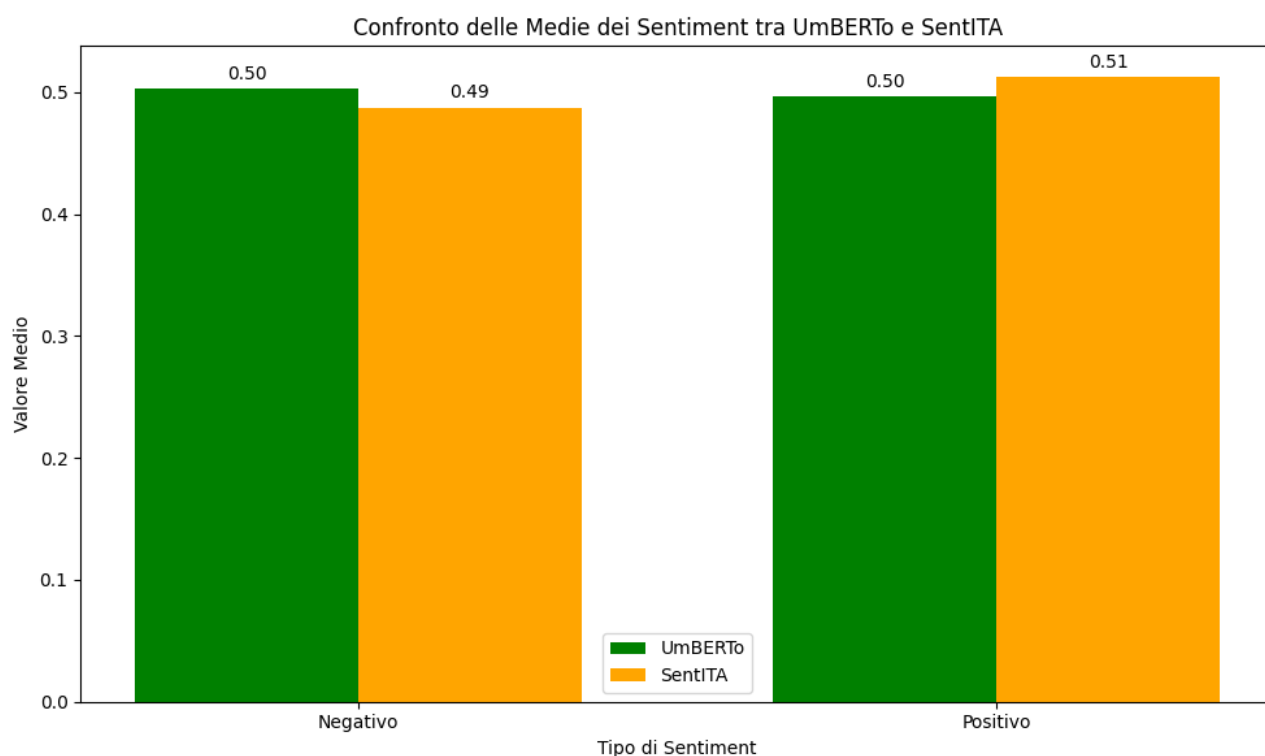


Figura 32 - Media pesata dei risultati del sentiment 'neg' e 'pos' dei modelli sentITA e umBERTo sul dataset di Morfina

Dalla *Figura 30* possiamo notare come il modello umBERTo classifichi 'più negativamente' rispetto al modello sentITA. D'altra parte, nella *Figura 31*, si verifica il comportamento opposto. In particolare, la sezione 3 (dedicata al come usare il farmaco) mostra un picco di sentiment negativo per SentITA che l'altro modello non evidenzia.

In conclusione, dalla *Figura 32*, notiamo come che il modello umBERTo risulti imparziale nella classificazione del sentiment mentre sentITA risulta essere classificare leggermente di più (51%) il sentiment positivo (in particolare a causa delle sezioni 1 e 3).

4.1.5. FEEL-IT

Applicando il modello Feel-it a *df_morfina*, i risultati del Emotion sul contenuto delle varie sezioni del bugiardino sono i seguenti:

Numero	Emotions	Sezione
1	Fear	che cos'è morfina cloridrato s.a.l.f. e a cosa serve
2	Fear	cosa deve sapere prima di usare morfina cloridrato
3	Fear	come usare morfina cloridrato s.a.l.f
4	Fear	possibili effetti indesiderati
5	Fear	come conservare morfina cloridrato s.a.l.f.
6	Fear	contenuto della confezione e altre informazioni

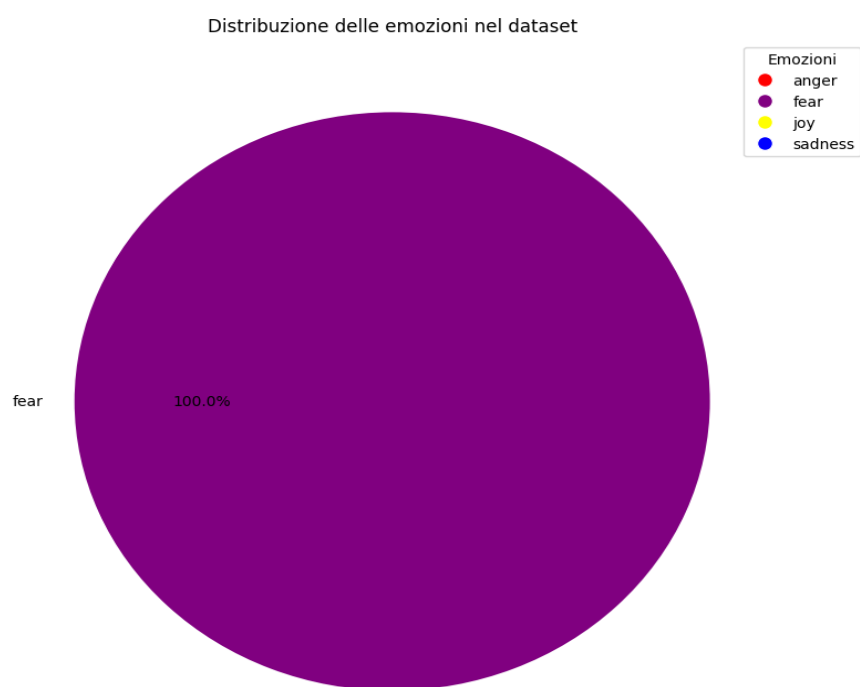


Figura 33 - Modello Feel-it sul dataset di Morfina

Secondo il modello Feel-it, tra tutte quelle possibili (anger, fear, joy, sadness) l'emozione predominante è di 'fear' (paura) in tutte e sei le sezioni. Questo significa che il contenuto analizzato nel bugiardino suscita principalmente un sentimento di paura o preoccupazione tra i lettori. Tale risultato potrebbe essere dovuto alla natura del testo, che è prescrittivo e mette in guardia rispetto alle possibili complicazioni del farmaco.

4.2. Bisolvon

4.2.1. Vader

Applicando il modello Vader a *df_bisolvon*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	neu	pos	compound	Sezione
1	0.000	1.000	0.000	0.000	che cos'è Bisolvon e a cosa serve
2	0.159	0.801	0.041	-0.9682	cosa deve sapere prima di usare Bisolvon
3	0.125	0.787	0.089	-0.3612	come usare Bisolvon
4	0.156	0.804	0.040	-0.9313	possibili effetti indesiderati
5	0.128	0.664	0.208	0.3612	come conservare Bisolvon
6	0.000	0.856	0.144	0.643	contenuto della confezione e altre informazioni

Da questi primi risultati, prendendo in considerazione i punteggi 'neg', 'neu' e 'pos' risulta che il bugiardino in questione non susciti sentimenti né positivi né negativi, mantenendo una neutralità più o meno simile in ogni sezione, tranne che nella cinque.

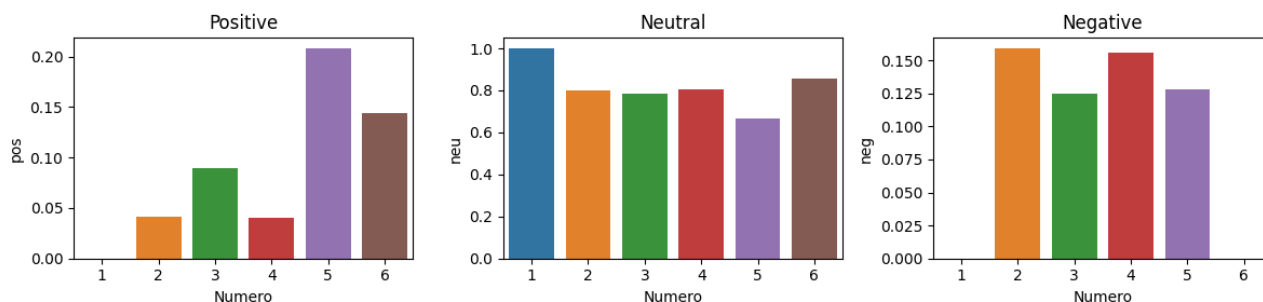


Figura 34 - Sentiment positivo, neutro, negativo con Vader su dataset di Bisolvon

Andando a vedere il *Compound Score* la neutralità vista in *Figura 34* viene confermata solo nella prima sezione. Mentre per quanto riguarda dalla sezione due alla sezione quattro il sentiment è estremamente negativo, soprattutto per la 2 e per la 4. Invece, per le restanti due sezioni il sentiment non è estremamente positivo (non supera il 60%), ma ciò può esser dato dal fatto che in queste due sezioni non ci sia nulla di estremamente 'positivo' in quanto si parla semplicemente di come conservare il farmaco e altre informazioni che riguardano l'azienda che lo produce.

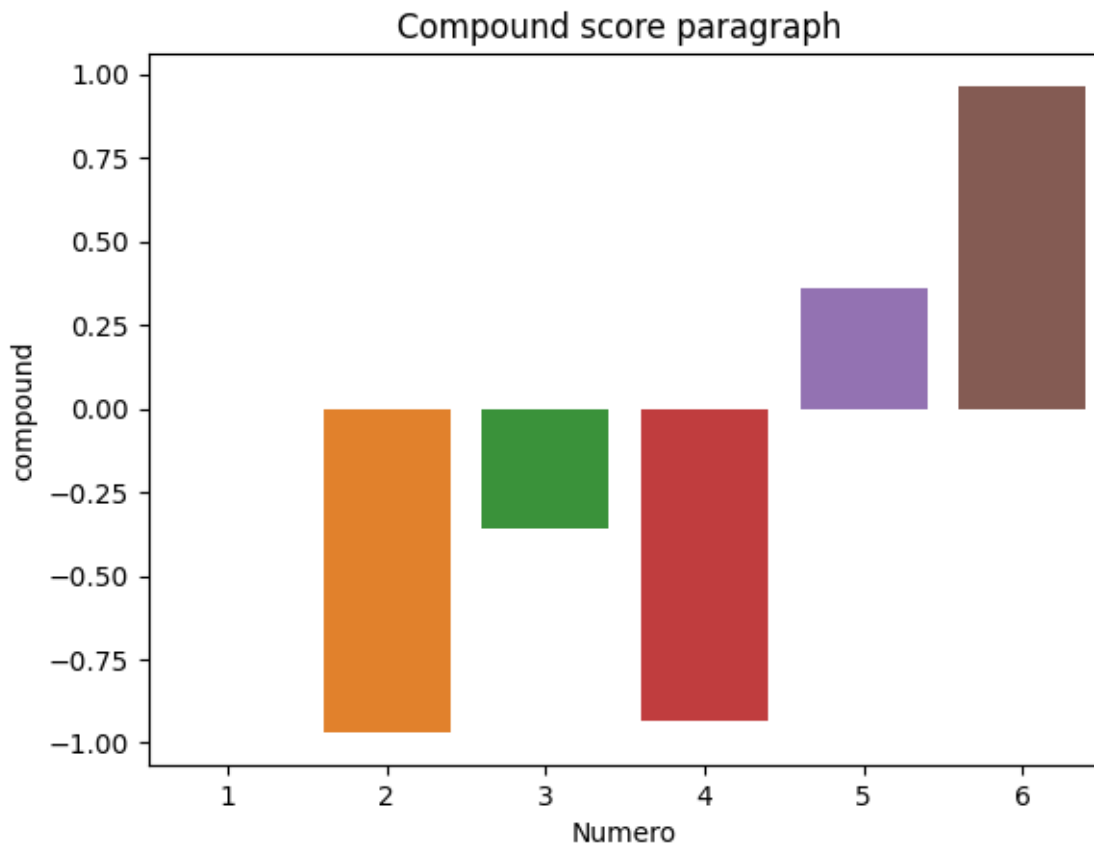


Figura 35 - Compound Score con Vader su dataset di Bisolvon

4.2.2. umBERTo

Applicando il modello umBERTo a *df_bisolvon*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.5121	0.4878	che cos'è Bisolvon e a cosa serve
2	0.5034	0.4965	cosa deve sapere prima di usare Bisolvon
3	0.5141	0.4858	come usare Bisolvon
4	0.4893	0.5106	possibili effetti indesiderati
5	0.5049	0.4950	come conservare Bisolvon
6	0.5069	0.4930	contenuto della confezione e altre informazioni

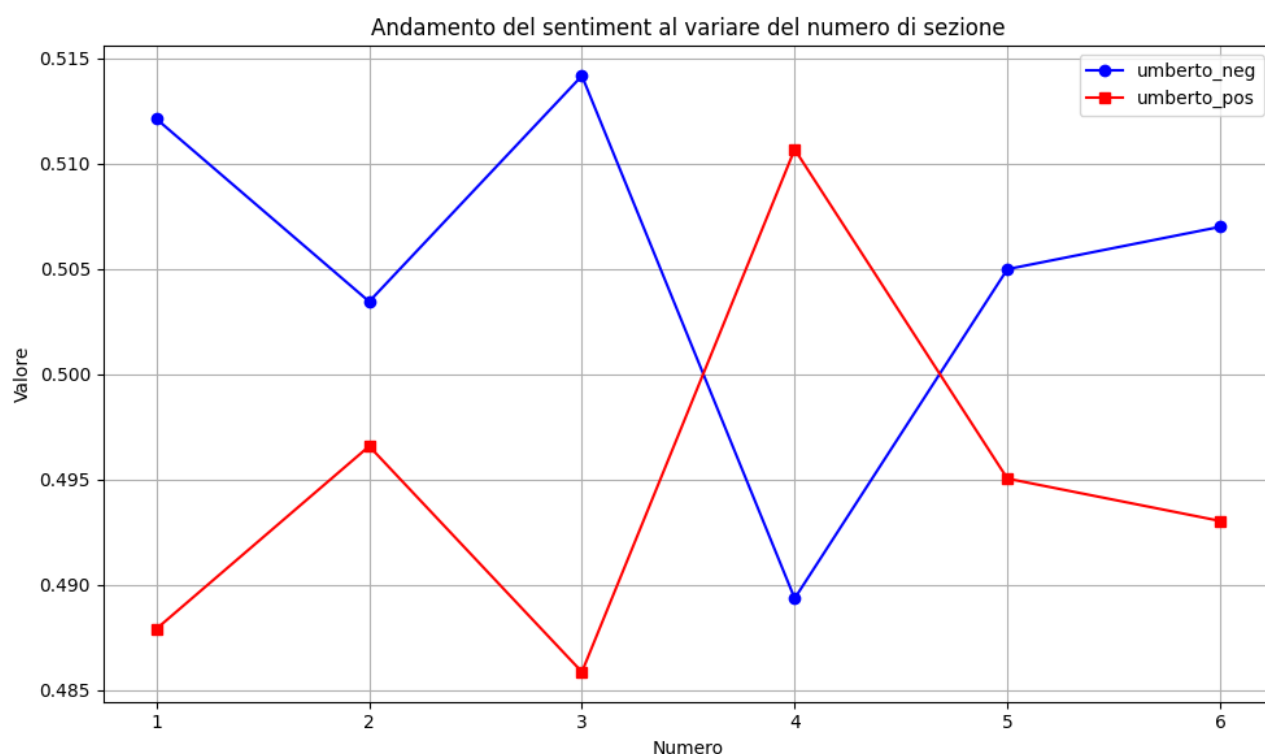


Figura 36 - Variazione del sentiment con umbERTO sul dataset di Bisolvon

I risultati ottenuti, a parte la sezione quattro, rappresentano un sentiment leggermente più negativo. Nonostante questo, i punteggi di polarità per ogni sezione non sono distanti tra di loro; il massimo è nella sezione 3, in cui il sentiment negativo è del 51% contro il 48% positivo. Questo sta a indicare che c'è un equilibrio costante nelle sezioni e che, pertanto, all'interno di questo bugiardino vige la neutralità.

4.2.3. SentITA

Applicando il modello sentiITA a *df_bisolvon*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.4552	0.5447	che cos'è Bisolvon e a cosa serve
2	0.4999	0.5000	cosa deve sapere prima di usare Bisolvon
3	0.4917	0.5082	come usare Bisolvon
4	0.4996	0.5000	possibili effetti indesiderati
5	0.4998	0.5000	come conservare Bisolvon
6	0.4990	0.5004	contenuto della confezione e altre informazioni

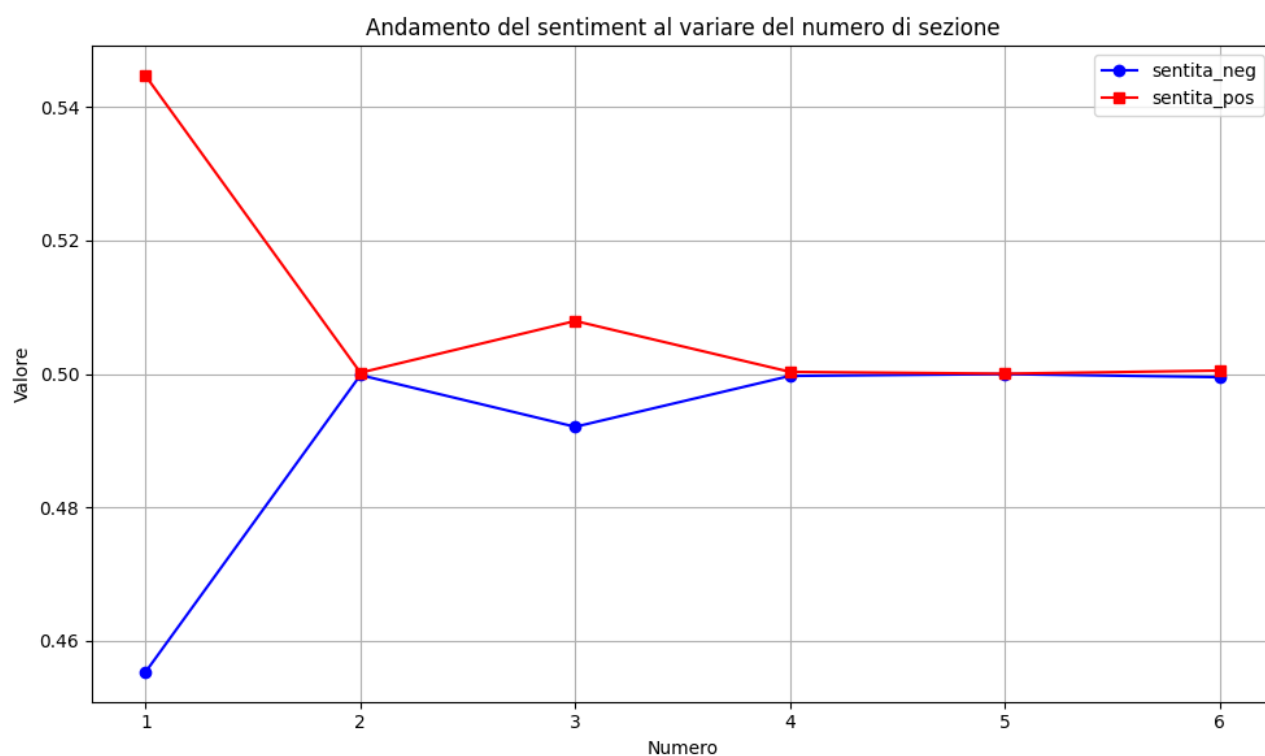


Figura 37 - Variazione del sentiment con SentITA sul dataset di Bisolvon

Come si può vedere in *Figura 37*, questo modello classifica il sentiment come estremamente neutro. Solamente nella prima sezione il sentiment è leggermente positivo rispetto alle altre. In questa sezione, infatti, troviamo la descrizione del farmaco e a cosa serve. Probabilmente il sentiment positivo si spiega con il suo contenuto, che spiega tutti i benefici che il farmaco può dare tramite il suo utilizzo. Eccezion fatta per la sezione 1, tutte le altre hanno la stessa percentuale di sentiment negativo e positivo.

4.2.4. Confronto risultati tra i modelli umBERTo e sentiTA

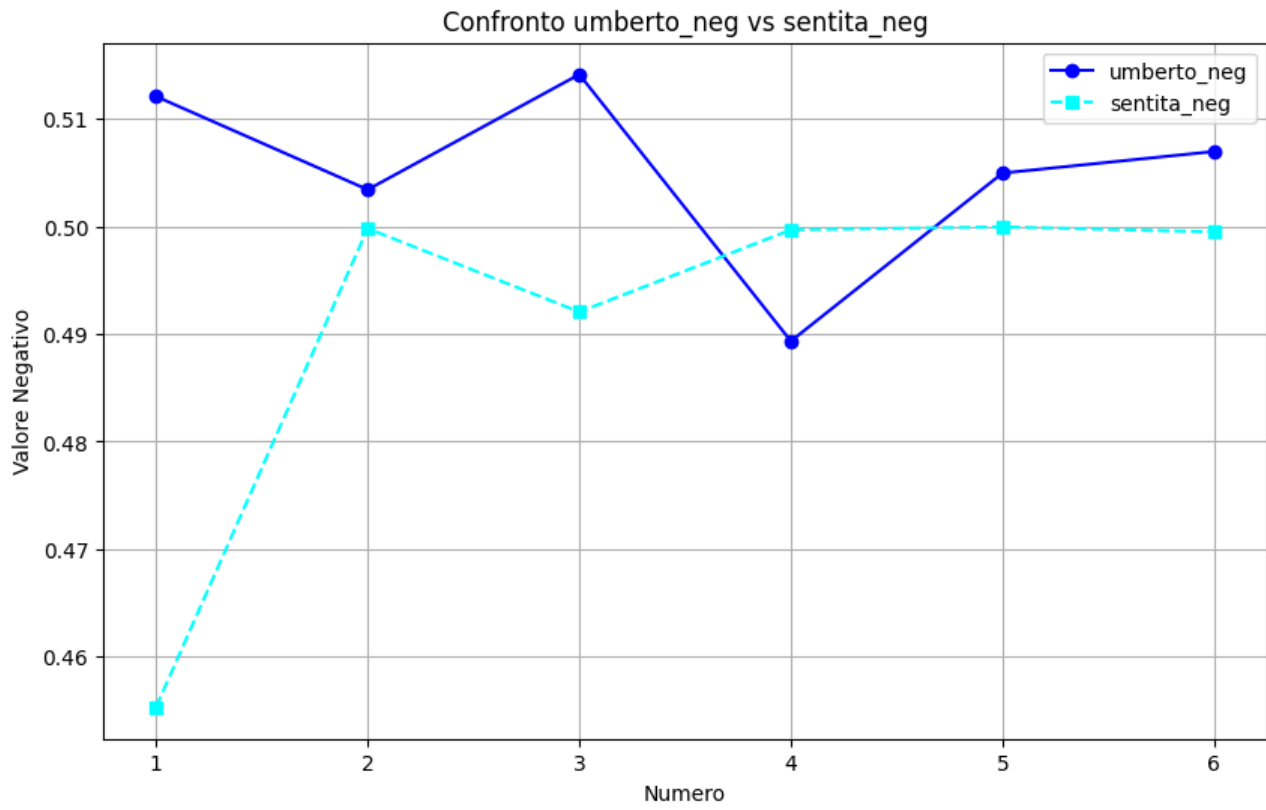


Figura 38 - Confronto sentiment 'neg' tra modelli sentiTA e umBERTo sul dataset di Bisolvon

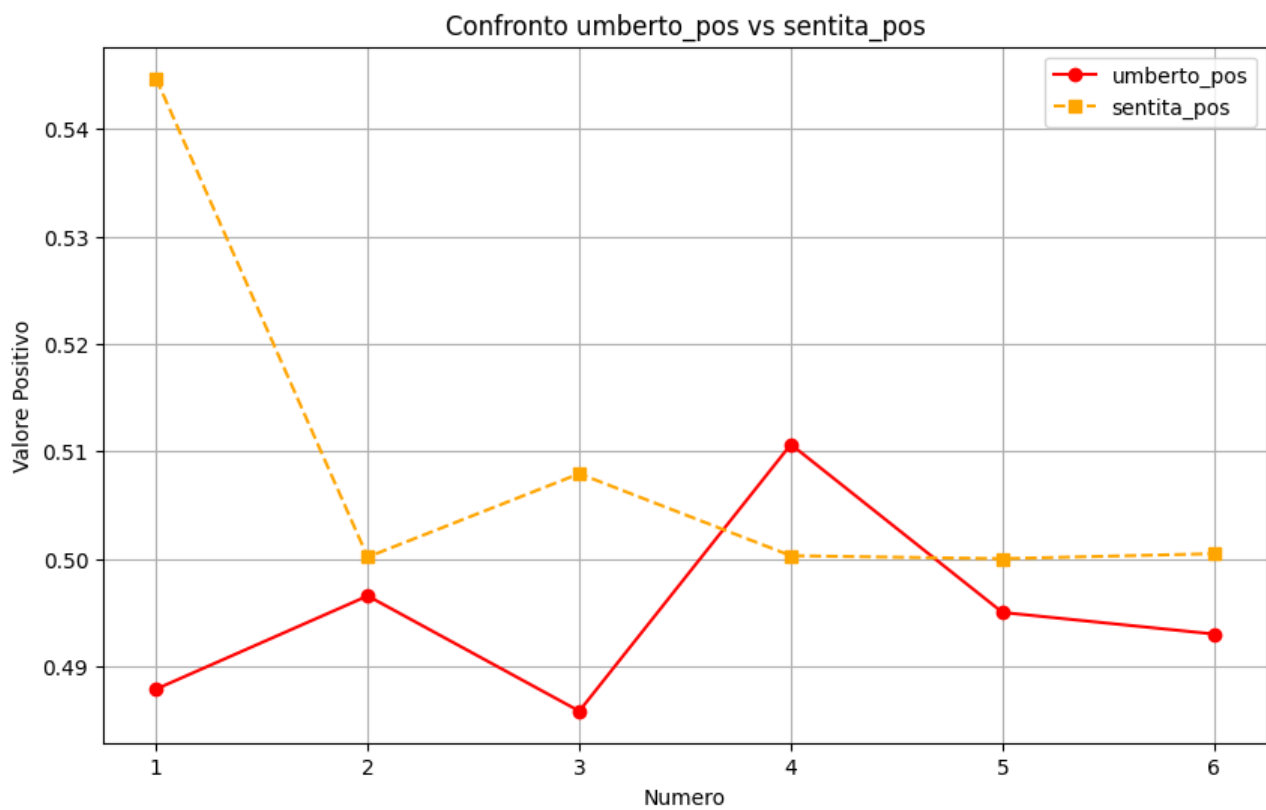


Figura 39 - Confronto sentiment 'pos' tra modelli sentITA e umBERTo sul dataset di Bisolvon

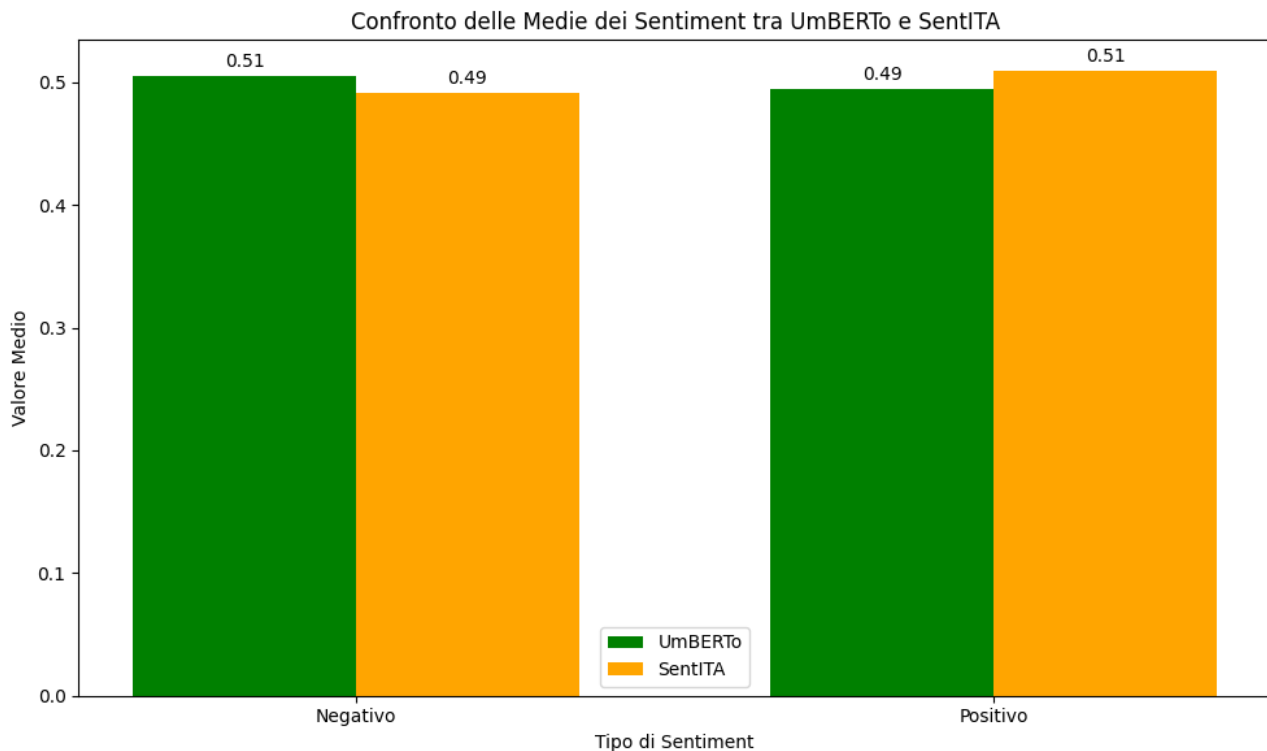


Figura 40 - Media pesata dei risultati del sentiment 'neg' e 'pos' dei modelli sentITA e umBERTo sul dataset di Bisolvon

Dalle Figure 38 e 39 possiamo notare come i due modelli classifichino in maniera diversa le varie sezioni. Infatti, seppur le differenze sono minime, il modello umBERTo classifica il sentiment più negativamente rispetto al modello sentITA. A conferma di quanto detto, nella Figura 40, notiamo come umBERTo classifica il 51% delle volte sentiment negativo rispetto al 49% di sentITA e viceversa per il sentiment positivo.

4.2.5. FEEL-IT

Applicando il modello Feel-it a *df_bisolvon*, i risultati del Emotion delle varie sezioni del bugiardino sono i seguenti:

Numero	Emotion	Sezione
1	Fear	che cos'è Bisolvon e a cosa serve
2	Fear	cosa deve sapere prima di usare Bisolvon
3	Fear	come usare Bisolvon
4	Fear	possibili effetti indesiderati
5	Fear	come conservare Bisolvon
6	Fear	contenuto della confezione e altre informazioni

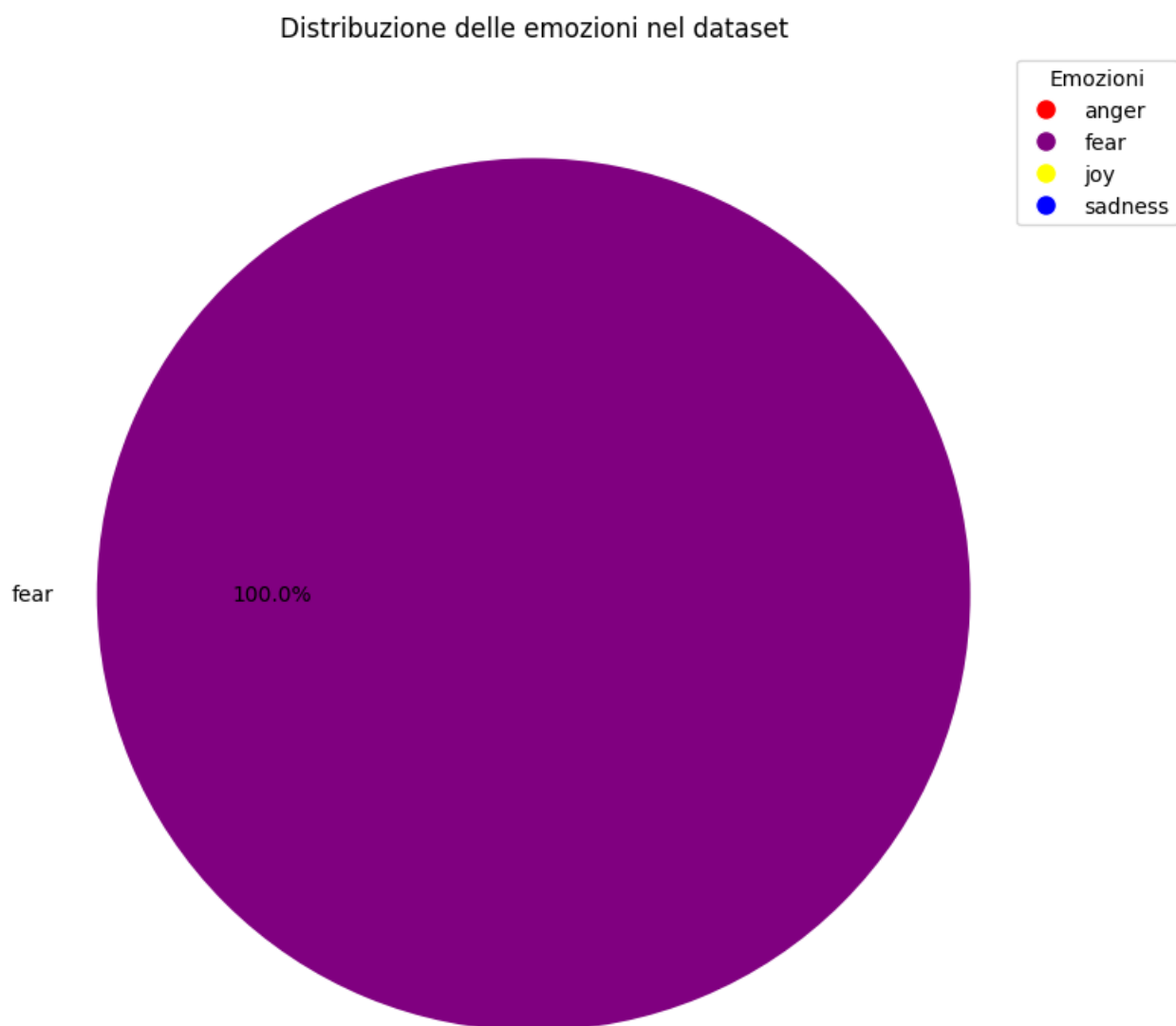


Figura 41 - Modello Feel-it sul dataset di Bisolvon

Analogamente alla Morfina e per analoghe ragioni, il contenuto del bugiardino suscita principalmente un sentimento di paura o preoccupazione.

4.3. Borocillina

4.3.1. Vader

Applicando il modello Vader a *df_borocillina*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	neu	pos	compound	Sezione
1	0.318	0.629	0.052	-0.9788	che cos'è Neo Borocillina infiammazione e dolore e a cosa serve
2	0.253	0.685	0.061	-0.9998	cosa deve sapere prima di prendere Neo Borocillina infiammazione e dolore
3	0.309	0.613	0.078	-0.9976	come prendere Neo Borocillina infiammazione e dolore
4	0.222	0.742	0.036	-0.9990	possibili effetti indesiderati
5	0.152	0.705	0.144	-0.077	come conservare Neo Borocillina infiammazione e dolore
6	0.132	0.801	0.067	-0.709	contenuto della confezione e altre informazioni

Prendendo in considerazione i punteggi 'neg', 'neu' e 'pos' risulta che il bugiardino in questione sia per lo più neutrale, ma meno rispetto ai bugiardini studiati precedentemente. Infatti, soprattutto nelle prime quattro sezioni vi è un sentiment negativo pari al 30% rispetto a quello positivo che arriva al 14% solo in una sezione.

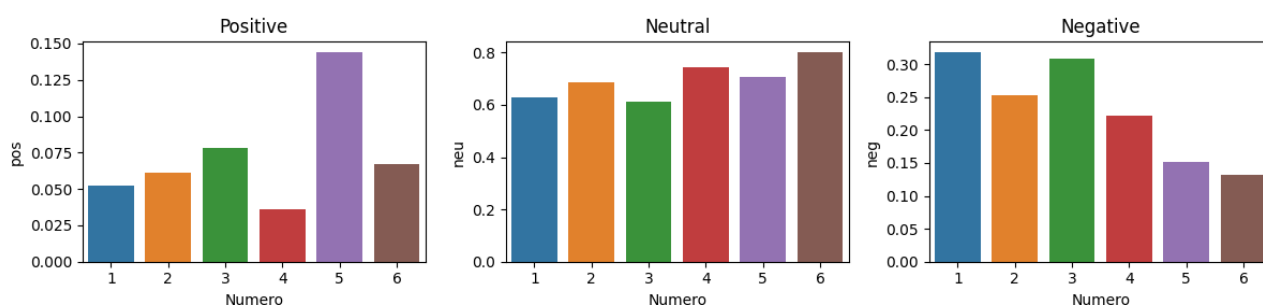


Figura 42 - Sentiment positivo, neutro, negativo con Vader su dataset di Borocillina

Andando ad analizzare il *Compound Score* ogni sezione (tranne la 5, praticamente neutra) risulta essere estremamente negativa, toccando il 99% nelle sezioni 2,3,4. Il sentiment di questo bugiardino però potrebbe essere amplificato dalla presenza di parole come 'dolore' e 'infiammazione' che potrebbero ricondurre ad un sentiment negativo.

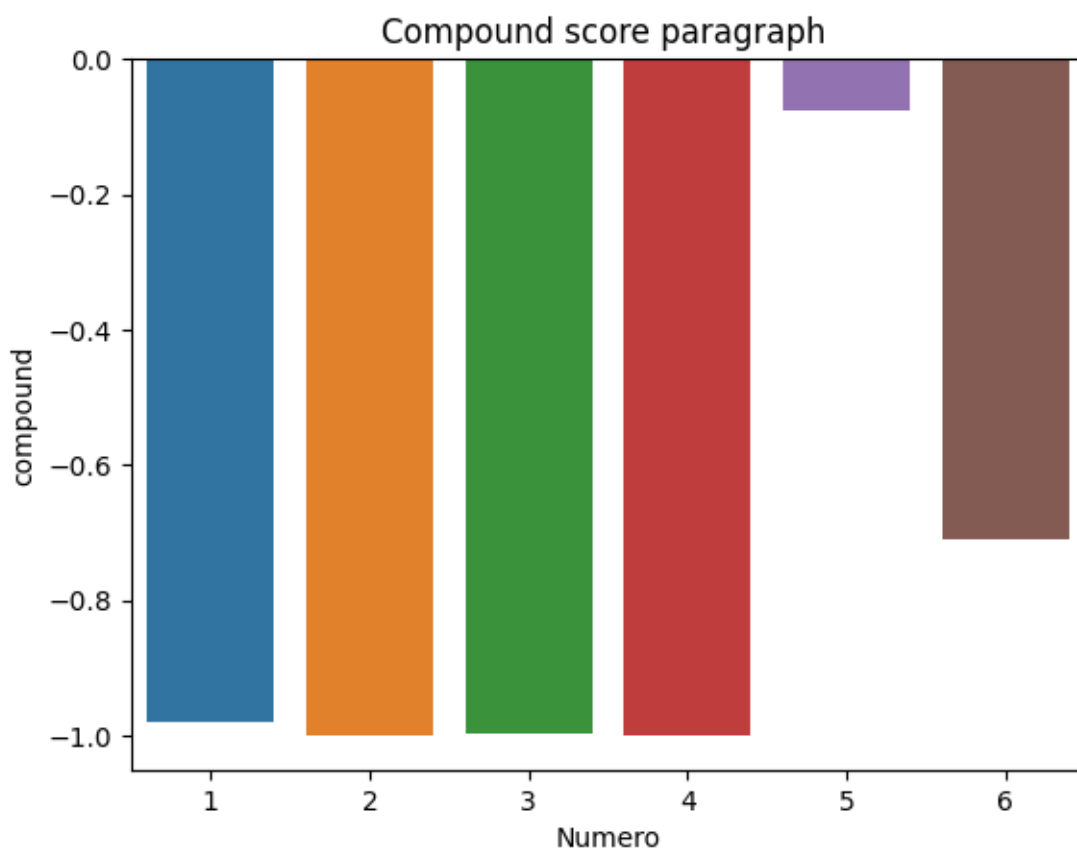


Figura 43 - Compound Score con Vader su dataset di Borocillina

4.3.2. umBERTo

Applicando il modello umBERTo a *df_borocillina*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.4905	0.5094	che cos'è Neo Borocillina infiammazione e dolore e a cosa serve
2	0.4927	0.5072	cosa deve sapere prima di prendere Neo Borocillina infiammazione e dolore
3	0.4973	0.5026	come prendere Neo Borocillina infiammazione e dolore
4	0.4950	0.5049	possibili effetti indesiderati
5	0.4774	0.5255	come conservare Neo Borocillina infiammazione e dolore
6	0.4854	0.5145	contenuto della confezione e altre informazioni

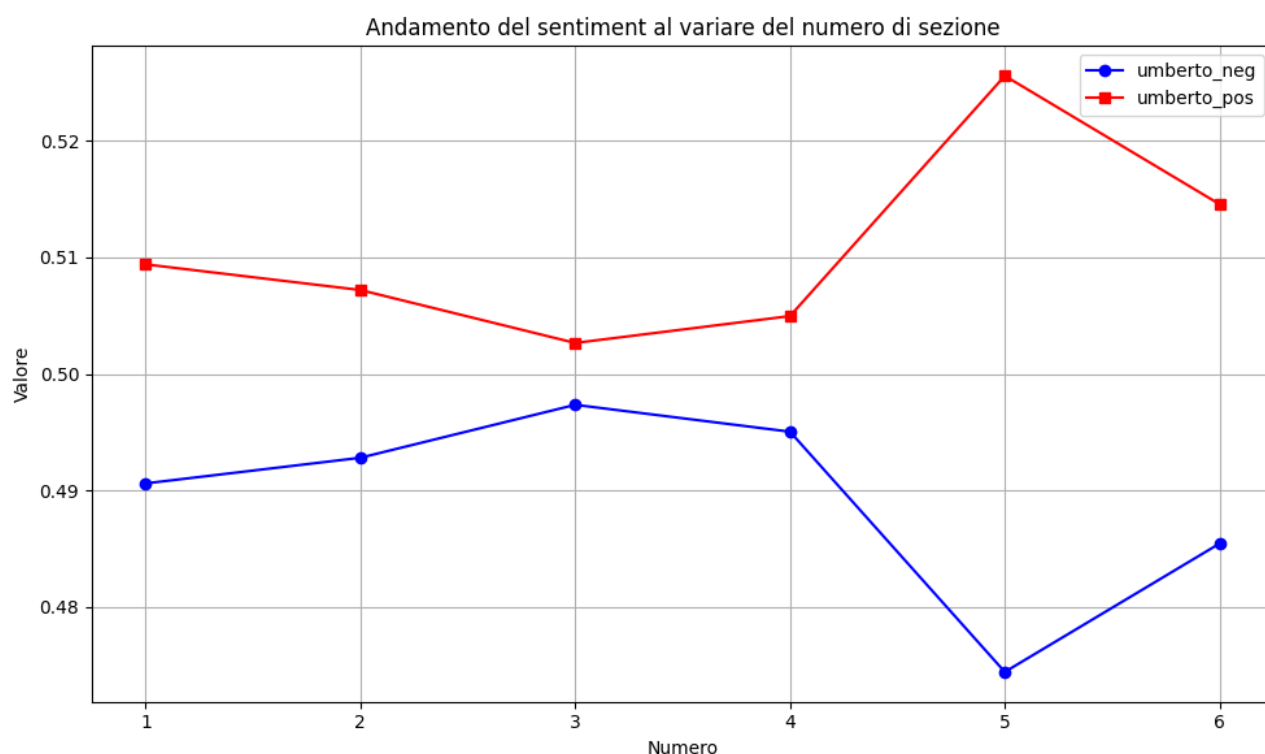


Figura 44 - Variazione del sentiment con umbERTO sul dataset di Borocillina

I risultati ottenuti, a differenza dei dataset precedenti, risultano avere (utilizzando questo modello) un sentiment più positivo. Infatti, anche se di poco, ogni sezione è classificata con un sentiment positivo superiore al 50%, ma le discrepanze tra i sentiment non sono così ampie. Pertanto si tratta di un sentiment neutrale anche in questo caso.

4.3.3. SentITA

Applicando il modello sentITA a df_borocillina, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.42261	0.5777	che cos'è Neo Borocillina infiammazione e dolore e a cosa serve
2	0.4999	0.5000	cosa deve sapere prima di prendere Neo Borocillina infiammazione e dolore
3	0.4637	0.5362	come prendere Neo Borocillina infiammazione e dolore
4	0.4999	0.5000	possibili effetti indesiderati
5	0.4997	0.5000	come conservare Neo Borocillina infiammazione e dolore
6	0.4999	0.5000	contenuto della confezione e altre informazioni

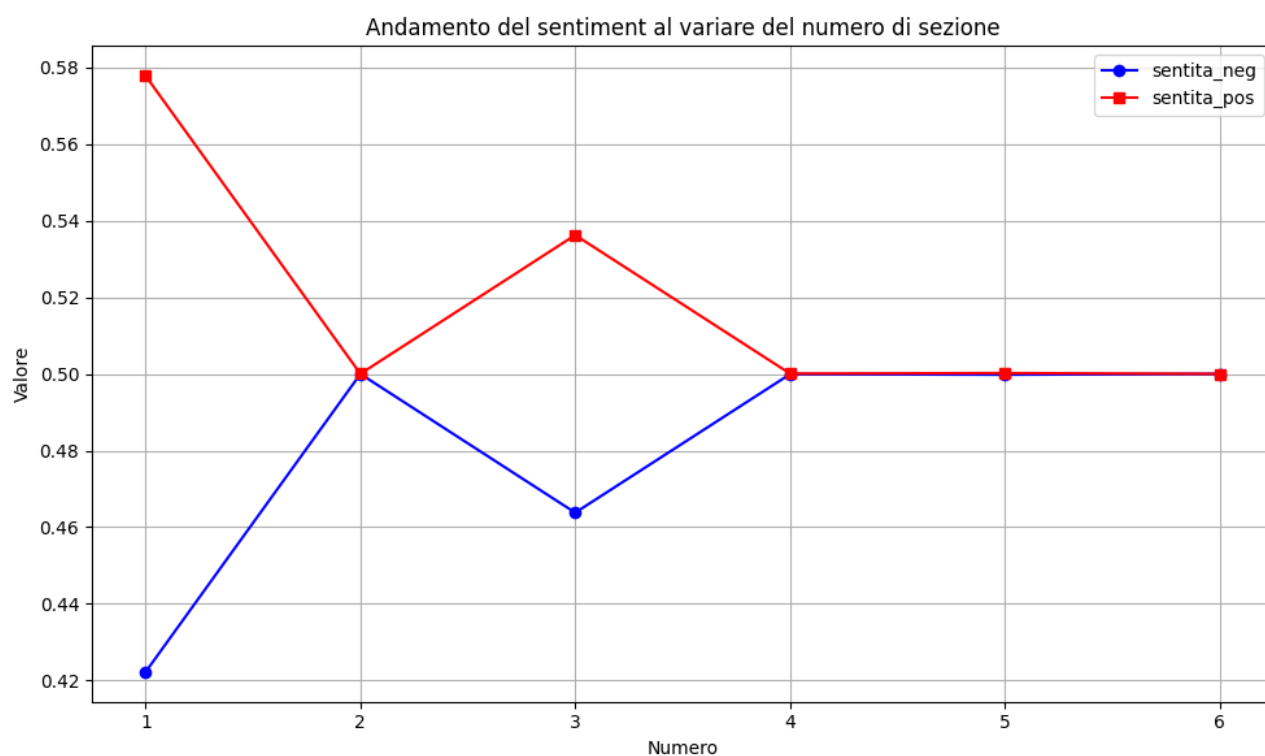


Figura 45 - Variazione del sentiment con SentITA sul dataset di Borocillina

Come in *Figura 37*, questo modello classifica il sentiment di questo bugiardino come estremamente neutro. Soprattutto nella prima sezione il sentiment è leggermente positivo rispetto alle altre. In questa sezione, infatti, troviamo la descrizione del farmaco e a cosa serve e probabilmente all'interno vengono spiegati tutti i benefici che il farmaco può dare tramite il suo utilizzo. Eccezion fatta per le sezioni 1 e 3, tutte le altre hanno la stessa percentuale di sentiment negativo e positivo.

4.3.4. Confronto risultati tra i modelli umBERTo e sentiTA

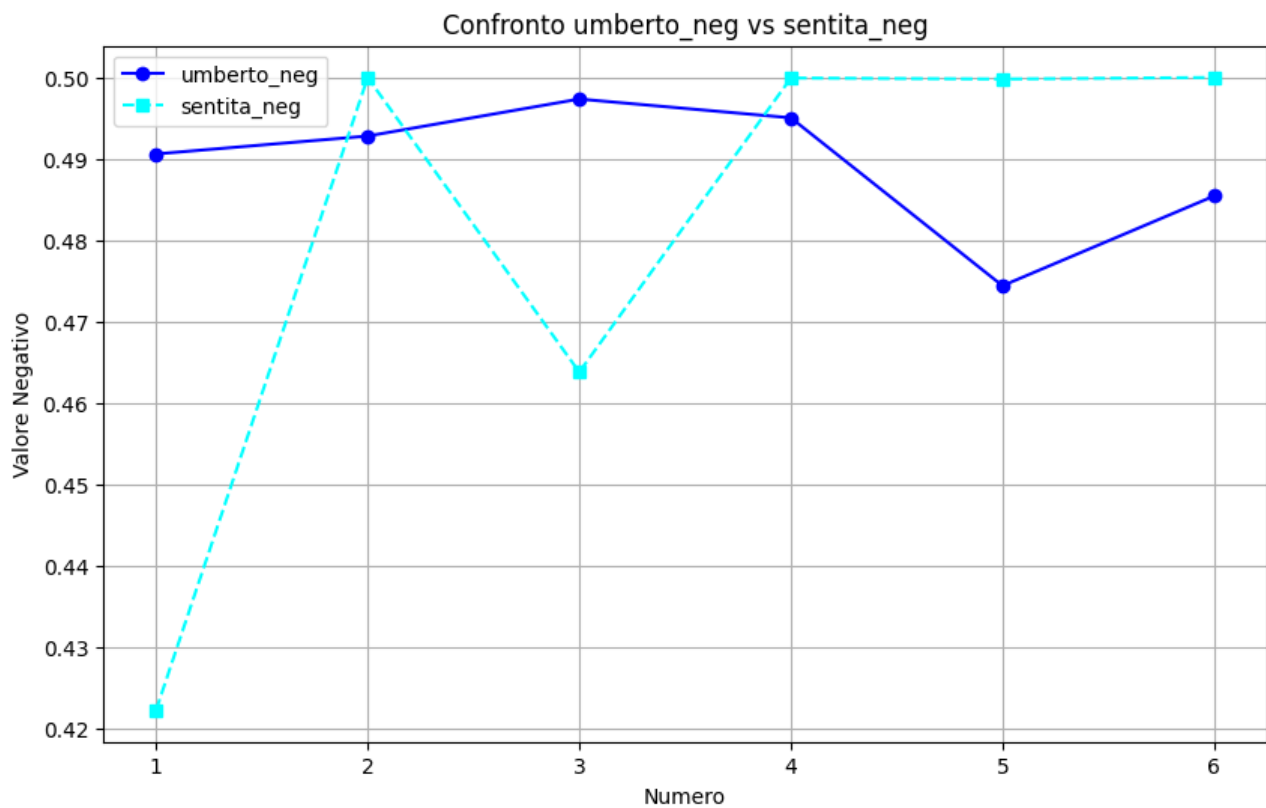


Figura 46 - Confronto sentiment 'neg' tra modelli sentiTA e umBERTo sul dataset di Borocillina

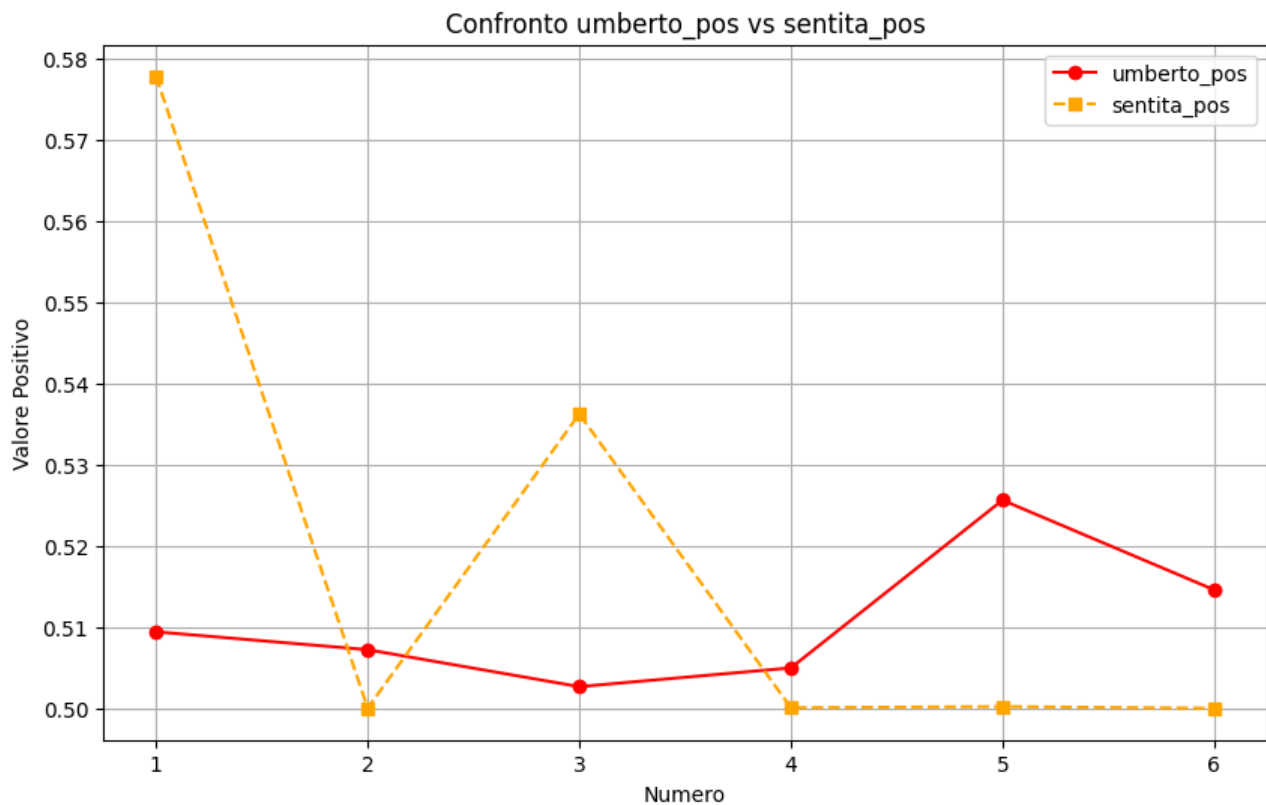


Figura 47 - Confronto sentiment 'pos' tra modelli sentiTA e umBERTo sul dataset Borocillina

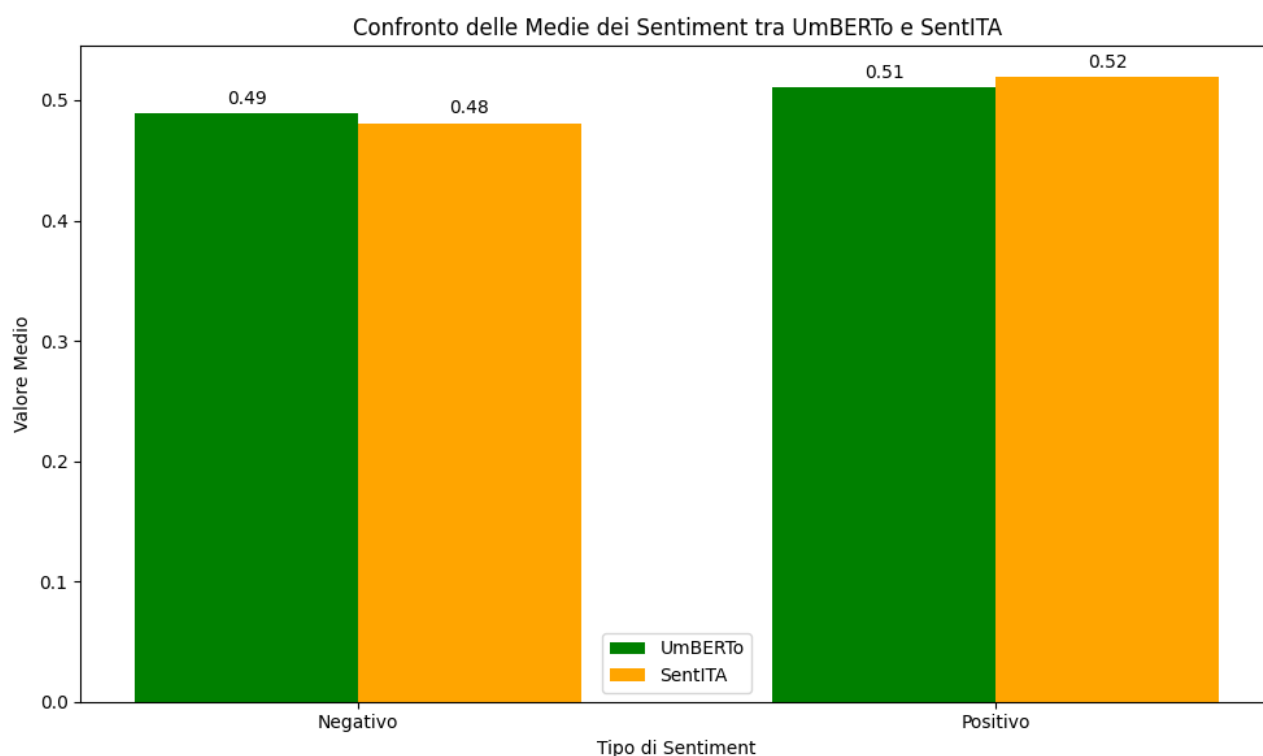


Figura 48 - Media pesata dei risultati del sentiment 'neg' e 'pos' dei modelli sentITA e umBERTo sul dataset di Borocillina

Dalle Figure 46 e 47 non si notano particolari differenze su come i due modelli classifichino la polarità del bugiardino. Infatti, rispetto ai documenti precedentemente studiati, in questo caso entrambi i modelli classificano (in media) nello stesso modo come si può notare in Figura 48. Infatti, sia sentITA che umBERTo tendono a classificare mediamente in maniera positiva le varie sezioni del bugiardino di Borocillina infiammazione e dolore anche se quello che ha classificato in maniera più positiva i vari paragrafi è stato il modello sentITA, con il 52%.

4.3.5. FEEL-IT

Applicando il modello Feel-it a *df_borocillina*, i risultati del Emotion delle varie sezioni del bugiardino sono i seguenti:

Numero	Emotion	Sezione
1	Fear	che cos'è Neo Borocillina infiammazione e dolore e a cosa serve
2	Fear	cosa deve sapere prima di prendere Neo Borocillina infiammazione e dolore
3	Fear	come prendere Neo Borocillina infiammazione e dolore
4	Fear	possibili effetti indesiderati
5	Fear	come conservare Neo Borocillina infiammazione e dolore
6	Fear	contenuto della confezione e altre informazioni

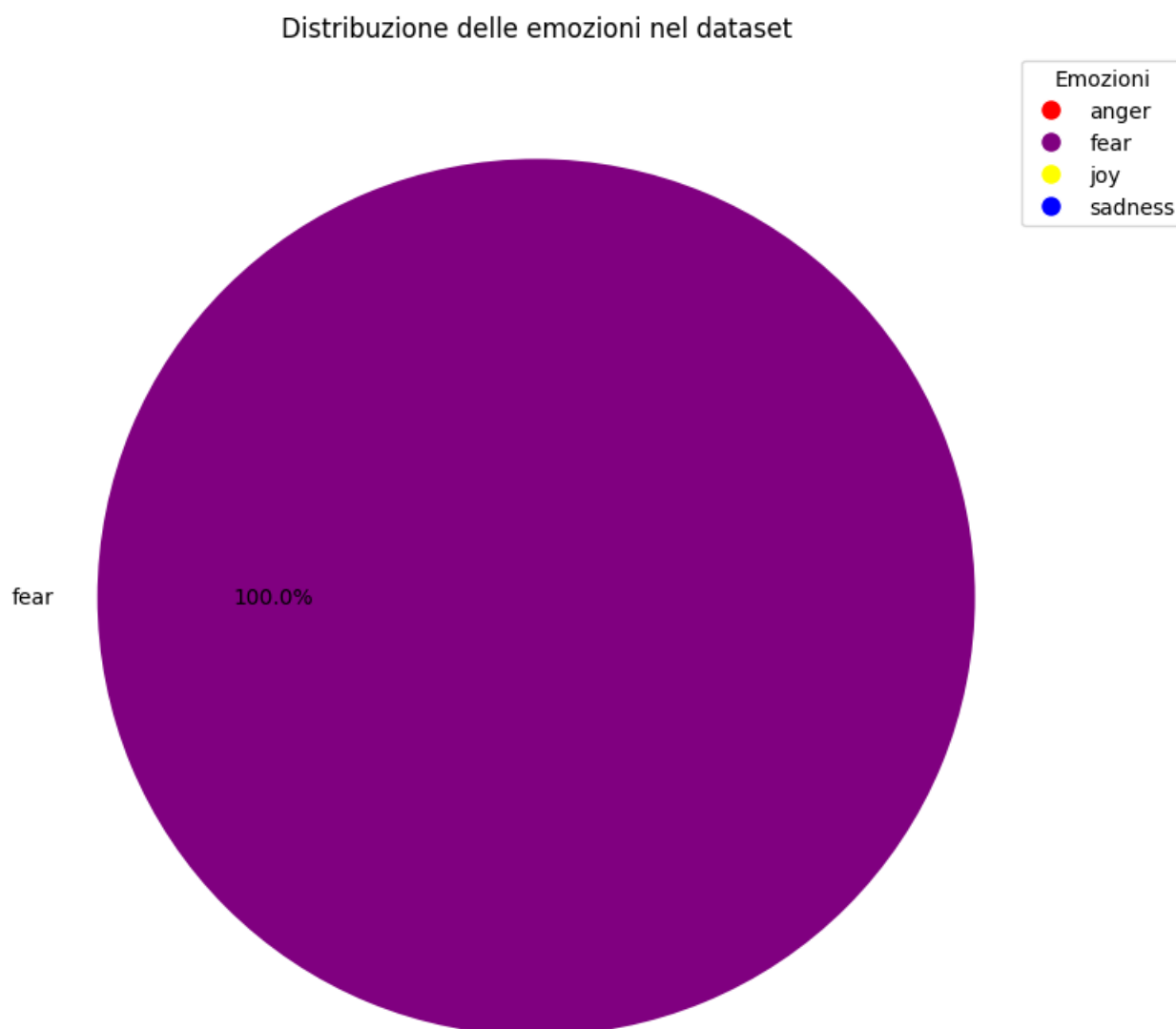


Figura 49 - Modello Feel-it sul dataset Borocillina

Dai risultati ottenuti le emozioni scaturite, secondo il modello Feel-it, tra tutte quelle possibili (joy, anger, sadness) è di 'fear' (paura) in tutte e sei le sezioni. Questo significa che il contenuto analizzato nel bugiardino suscita principalmente un sentimento di paura o preoccupazione tra i lettori. Tale risultato potrebbe essere dovuto alla natura del testo.

Analogamente alla Morfina e per analoghe ragioni, il contenuto del bugiardino suscita principalmente un sentimento di paura o preoccupazione.

4.4. Dicloream

4.4.1. Vader

Applicando il modello Vader a *df_dicloream*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	neu	pos	compound	Sezione
1	0.438	0.442	0.119	-0.9371	che cos'è Dicloream dolore e a cosa serve
2	0.260	0.696	0.044	-0.9997	cosa deve sapere prima di prendere Dicloream dolore
3	0.214	0.750	0.036	-0.9442	come prendere Dicloream dolore
4	0.234	0.708	0.058	-0.9979	possibili effetti indesiderati
5	0.173	0.686	0.142	-0.3400	come conservare Dicloream dolore
6	0.093	0.836	0.071	-0.3818	contenuto della confezione e altre informazioni

Prendendo in considerazione i punteggi 'neg', 'neu' e 'pos' risulta che il bugiardino in questione sia il meno neutrale tra quelli studiati finora. Nella prima sezione troviamo il sentiment negativo più alto, pari al 40%. Ciò significa che in questo paragrafo vi è un lessico che suggerisce che una parte significativa del testo ha una connotazione negativa, confermando la presenza di elementi che potrebbero creare ansia o timore tra chi legge.

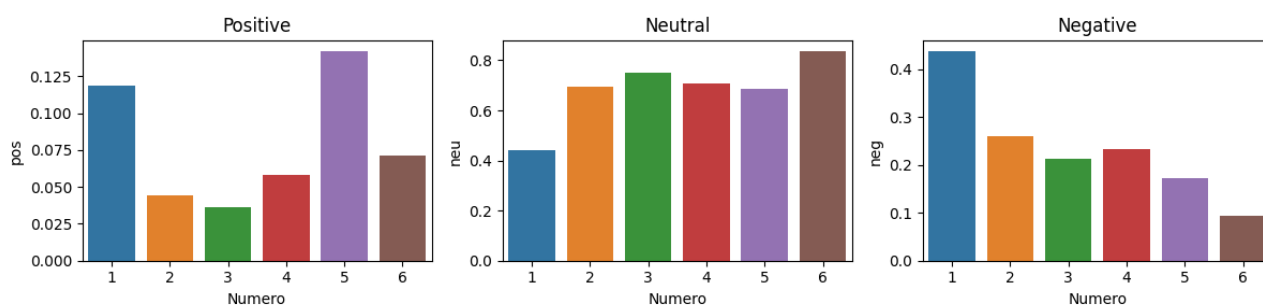


Figura 50 - Sentiment positivo, neutro, negativo con Vader su dataset di Dicloream

Andando ad analizzare il *Compound Score* ogni sezione (tranne la 5 e la 6, praticamente neutre) risulta essere estremamente negativa. Il sentiment di questo bugiardino però potrebbe essere amplificato dalla presenza di alcune parole come ‘dolore’. Questo è spesso comune nei bugiardini, che devono dettagliatamente elencare potenziali effetti collaterali e avvertenze per informare adeguatamente i consumatori sui rischi associati al medicinale.

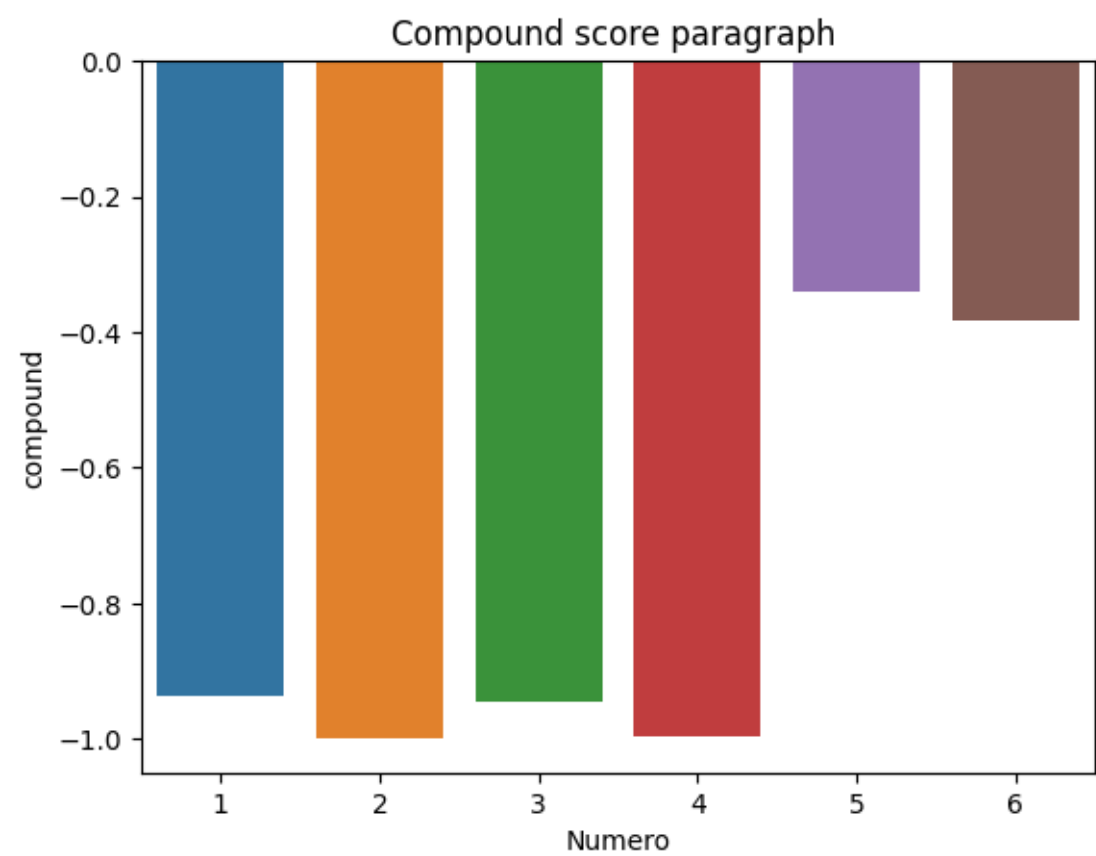


Figura 51 - Compound Score con Vader sul dataset di Dicloream

4.4.2. umBERTo

Applicando il modello umBERTo a df_dicloream, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.5227	0.477	che cos'è Dicloream dolore e a cosa serve
2	0.5300	0.4699	cosa deve sapere prima di prendere Dicloream dolore
3	0.5249	0.4750	come prendere Dicloream dolore
4	0.5249	0.4750	possibili effetti indesiderati
5	0.5224	0.4775	come conservare Dicloream dolore
6	0.5163	0.4836	contenuto della confezione e altre informazioni

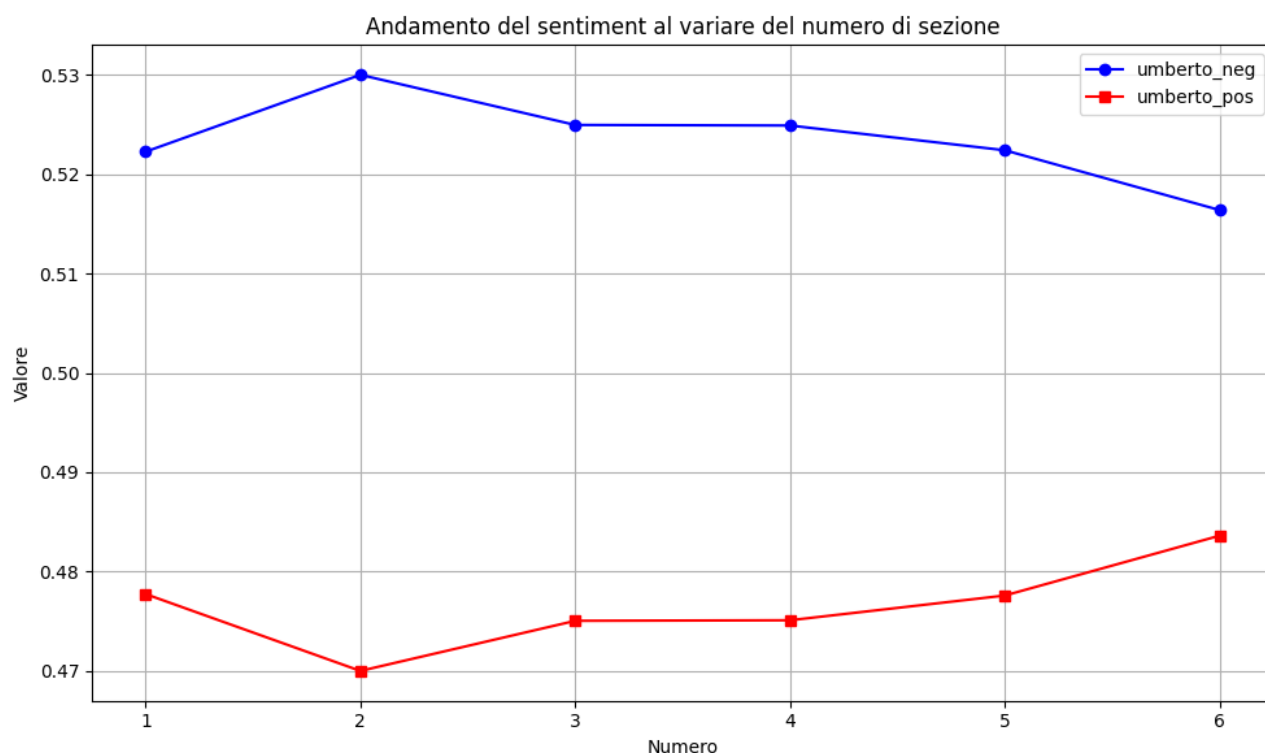


Figura 52 - Variazione del sentiment con umbERTO sul dataset di Dicloream

Il modello, come possiamo vedere in Figura 52, classifica ogni sezione con un sentiment negativo, fino al 53% per il paragrafo due dove viene spiegato come prendere e cosa sapere prima di assumere Dicloream.

4.4.3. SentITA

Applicando il modello sentITA a df_dicloream, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.4996	0.5003	che cos'è Dicloream dolore e a cosa serve
2	0.4999	0.5001	cosa deve sapere prima di prendere Dicloream dolore
3	0.3057	0.6942	come prendere Dicloream dolore
4	0.4999	0.5000	possibili effetti indesiderati
5	0.4999	0.5000	come conservare Dicloream dolore
6	0.5000	0.4999	contenuto della confezione e altre informazioni

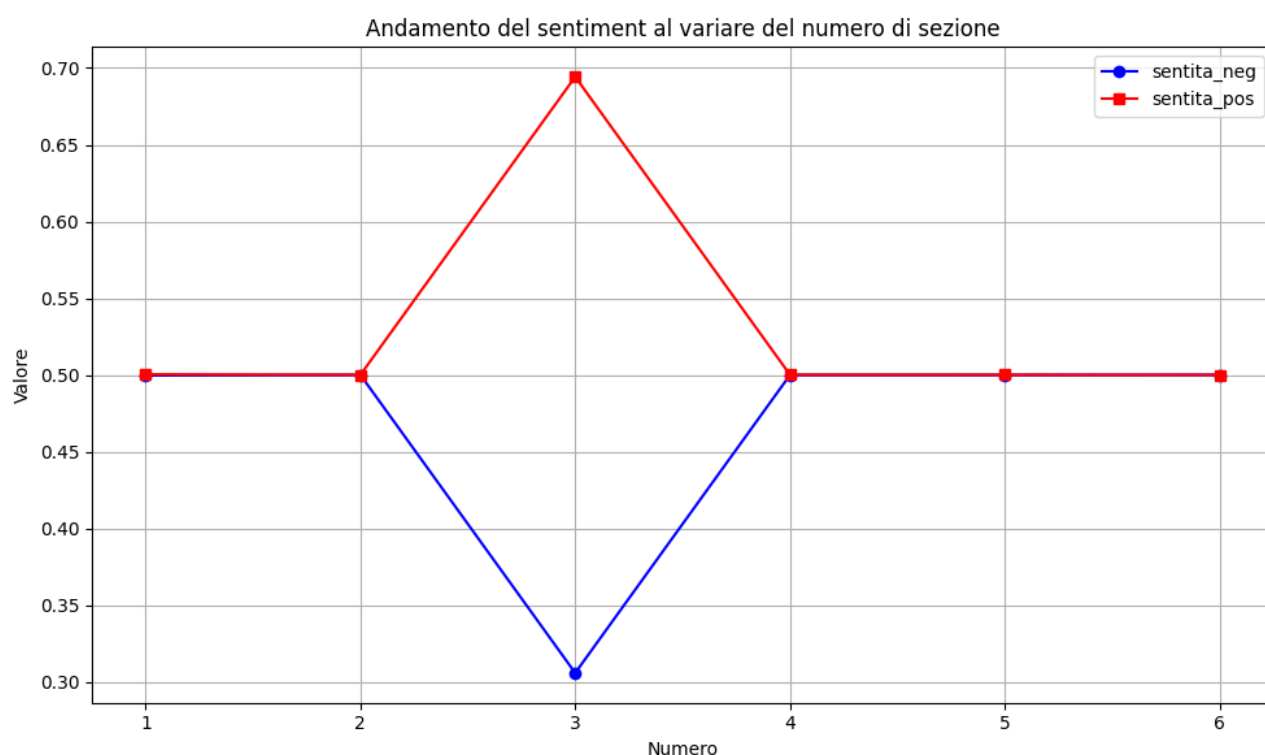


Figura 53 - Variazione del sentiment con SentITA sul dataset Dicloream

Come in *Figura 37 e 45*, questo modello classifica il sentiment di questo bugiardino come estremamente neutro. Nella terza sezione il sentiment è molto positivo. In questa sezione, infatti, troviamo la descrizione di come prendere il farmaco e quindi probabilmente ciò non ha connotazioni negative. Tra tutti i bugiardini studiati finora è il livello di sentiment positivo più alto mai calcolato. Eccezion fatta per la sezione 3, tutte le altre hanno la stessa percentuale di sentiment negativo e positivo.

4.4.4. Confronto risultati tra i modelli umBERTo e sentiTA

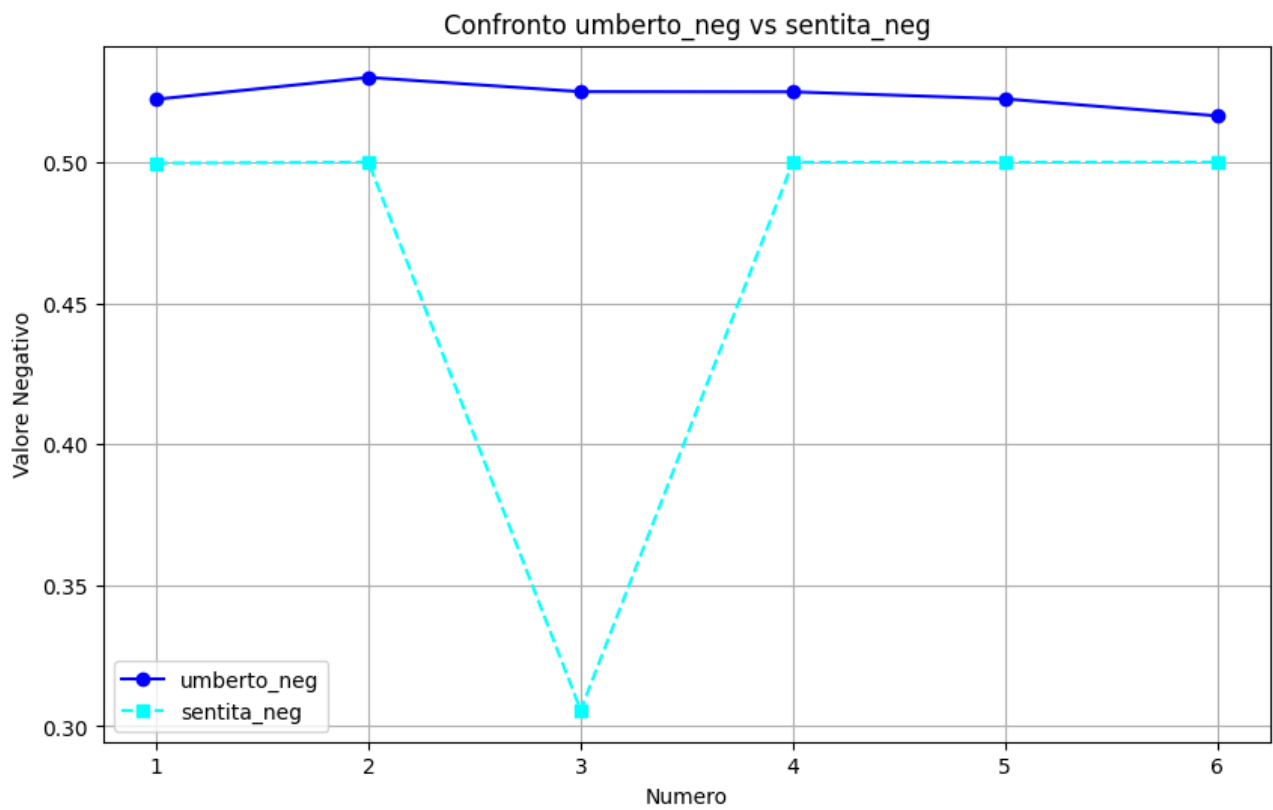


Figura 54 - Confronto sentiment 'neg' tra modelli sentiTA e umBERTo sul dataset di Dicloream

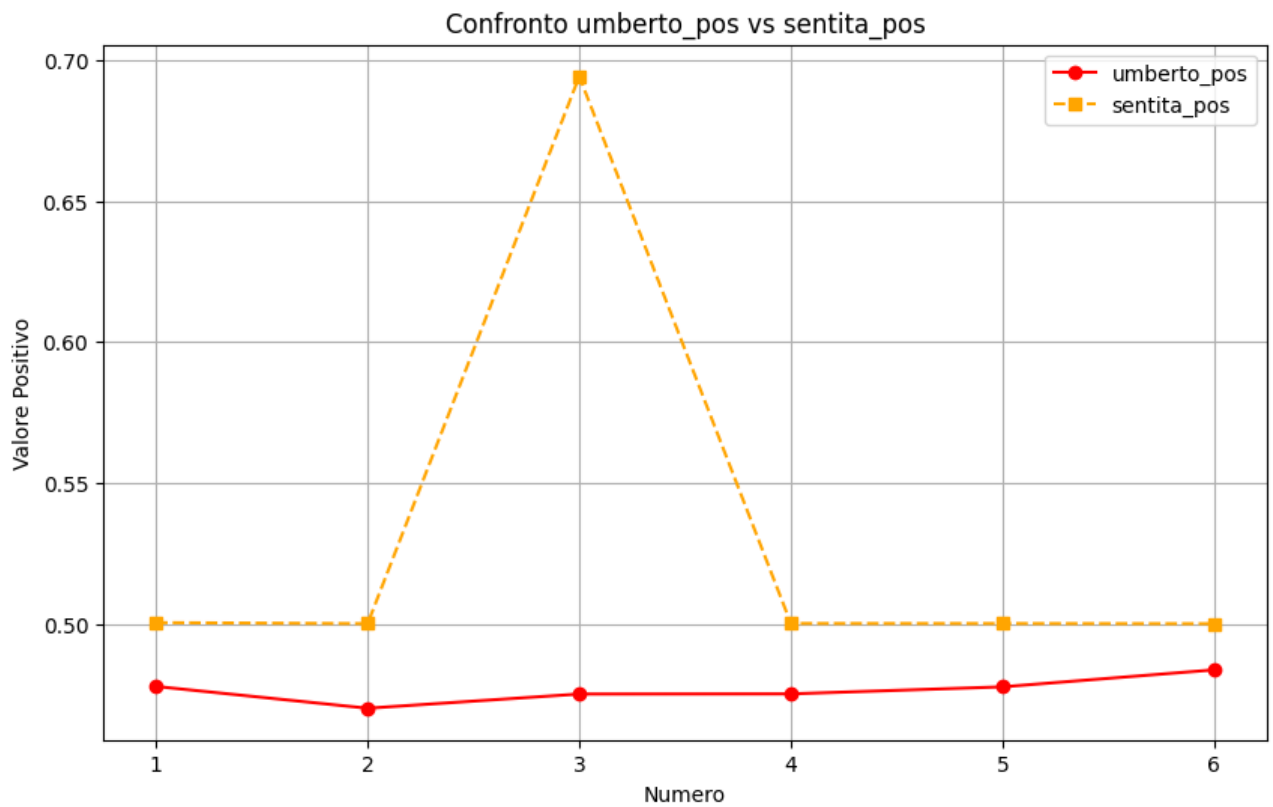


Figura 55 - Confronto sentiment 'pos' tra modelli sentiTA e umBERTo sul dataset di Dicloream

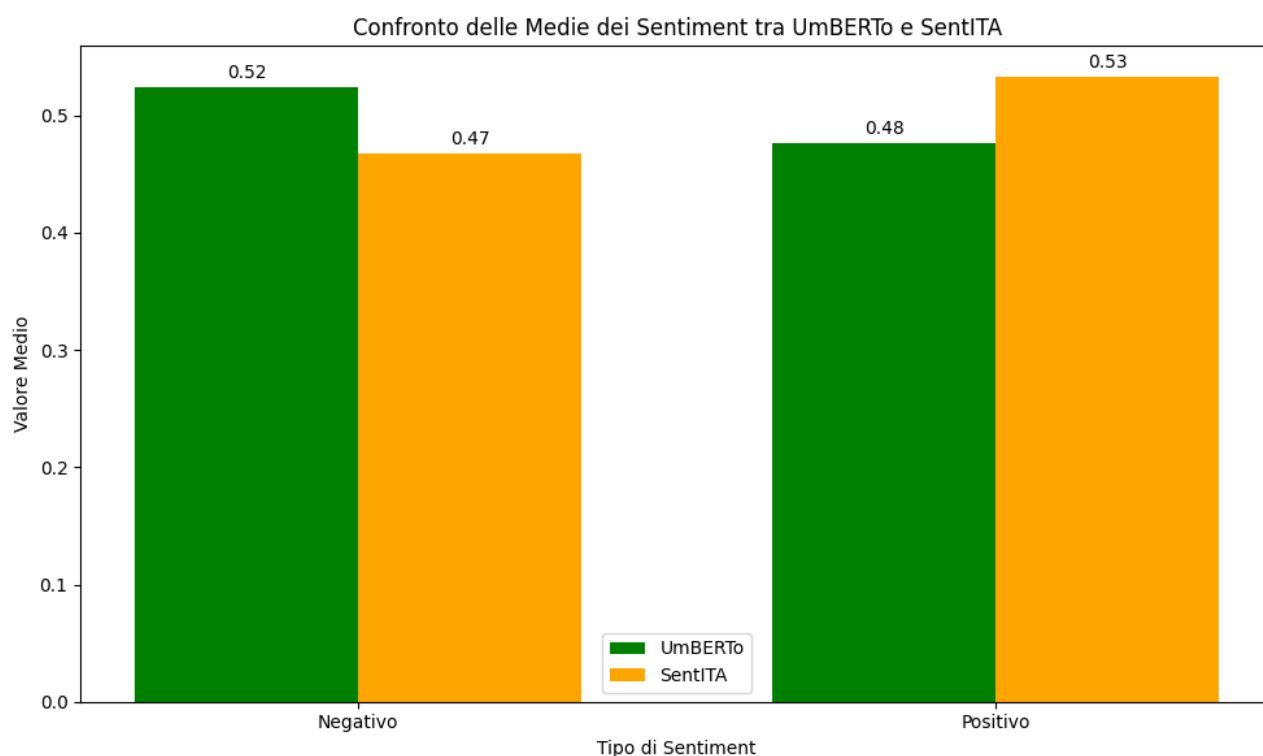


Figura 56 - Media pesata dei risultati del sentiment 'neg' e 'pos' dei modelli sentITA e umBERTo sul dataset di Dicloream

Dalle Figure 54 e 55, possiamo notare come i due modelli classifichino in maniera opposta le varie sezioni del bugiardino. Infatti, umBERTo classifica i vari paragrafi con più sentiment negativo rispetto a sentITA che invece li classifica maggiormente con sentiment positivo. A confermare questa ipotesi c'è la Figura 56, in cui possiamo vedere come il 52% delle volte umBERTo ha classificato le istanze in maniera negativa contro il 53% in cui sentITA ha classificato le istanze in maniera positiva (in media).

4.4.5. FEEL-IT

Applicando il modello Feel-it a *df_dicloream*, i risultati del Emotion delle varie sezioni del bugiardino sono i seguenti:

Numero	Emotion	Sezione
1	Fear	che cos'è Dicloream dolore e a cosa serve
2	Fear	cosa deve sapere prima di prendere Dicloream dolore
3	Fear	come prendere Dicloream dolore
4	Fear	possibili effetti indesiderati
5	Fear	come conservare Dicloream dolore
6	Fear	contenuto della confezione e altre informazioni

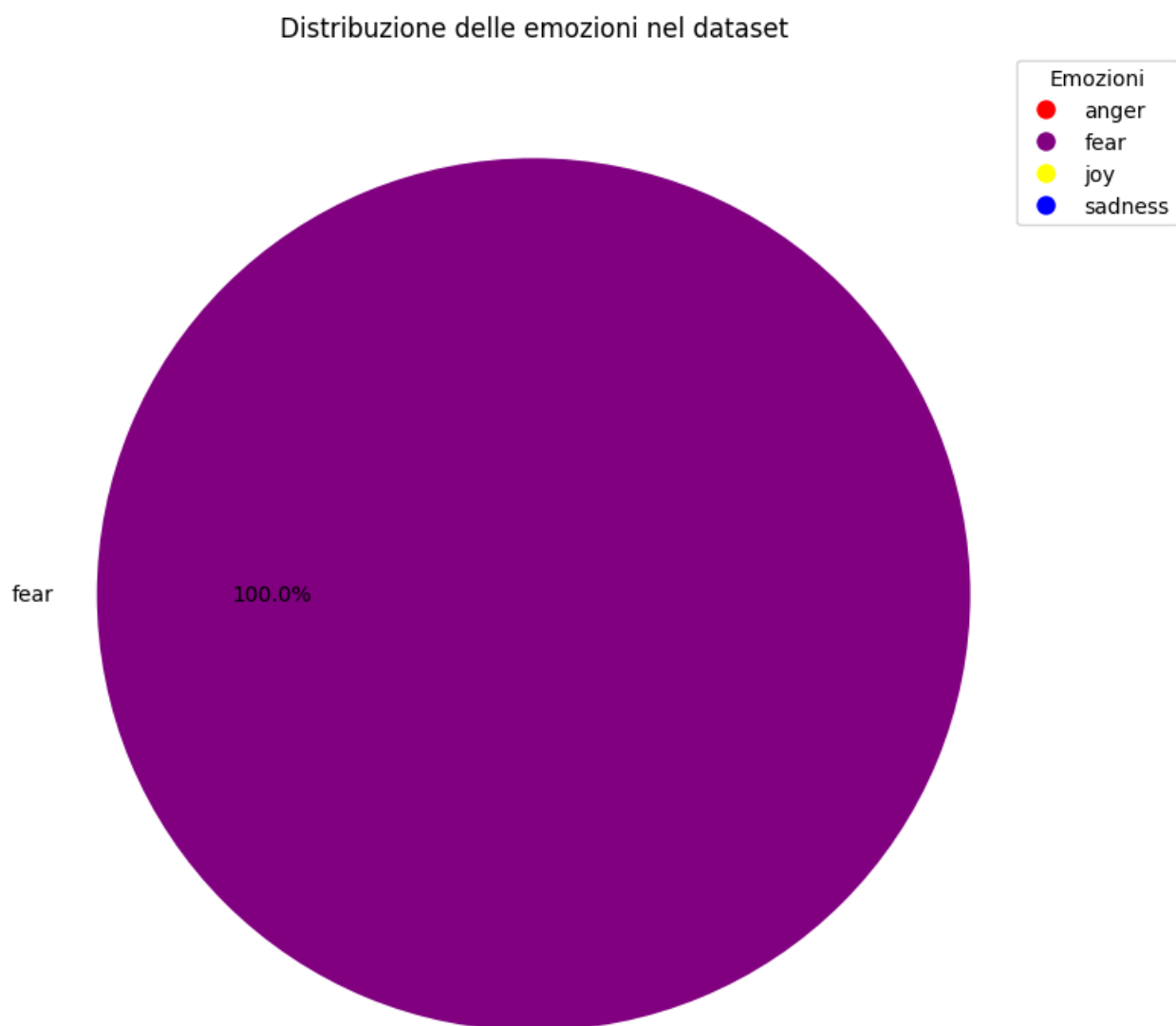


Figura 57 - Modello Feel-it sul dataset di Dicloream

Analogamente alla Morfina e per analoghe ragioni, il contenuto del bugiardino suscita principalmente un sentimento di paura o preoccupazione.

4.5. Fentanyl

4.5.1. Vader

Applicando il modello Vader a *df_fentanyl* i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	neu	pos	compound	Sezione
1	0.143	0.698	0.158	0.2023	che cos'è Fentanyl-Hameln a cosa serve
2	0.192	0.724	0.085	-0.9966	cosa deve sapere prima di usare Fentanyl-Hameln
3	0.085	0.915	0.000	-0.6705	come usare Fentanyl-Hameln
4	0.264	0.649	0.087	-0.9939	possibili effetti indesiderati
5	0.062	0.760	0.178	0.8225	come conservare Fentanyl-Hameln
6	0.061	0.855	0.084	0.8316	contenuto della confezione e altre informazioni

Prendendo in considerazione i punteggi 'neg', 'neu' e 'pos' risulta che il bugiardino in questione per lo più è neutrale.

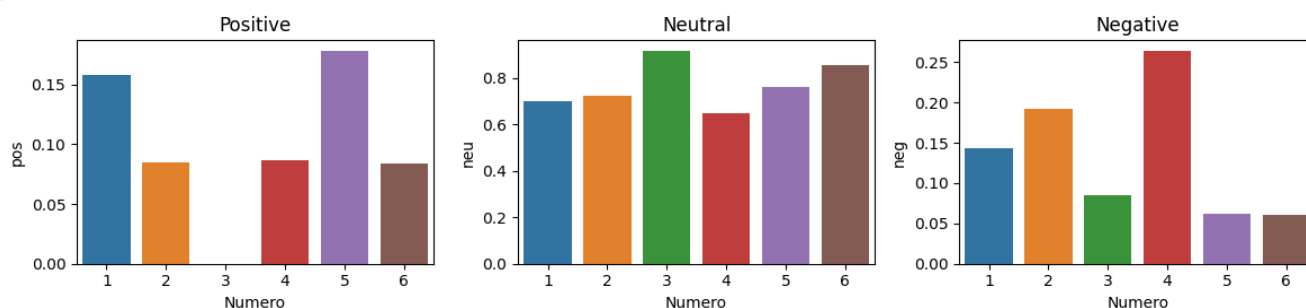


Figura 58 - Sentiment positivo, neutro, negativo con Vader su dataset di Fentanyl

Andando ad analizzare il *Compound Score* le sezioni 2,3,4 risultano essere estremamente negative, toccando il 99% nelle sezioni 2 e 4. Questi paragrafi sono quelli che elencano dettagliatamente potenziali effetti collaterali e avvertenze per informare adeguatamente i consumatori sui rischi associati al medicinale.

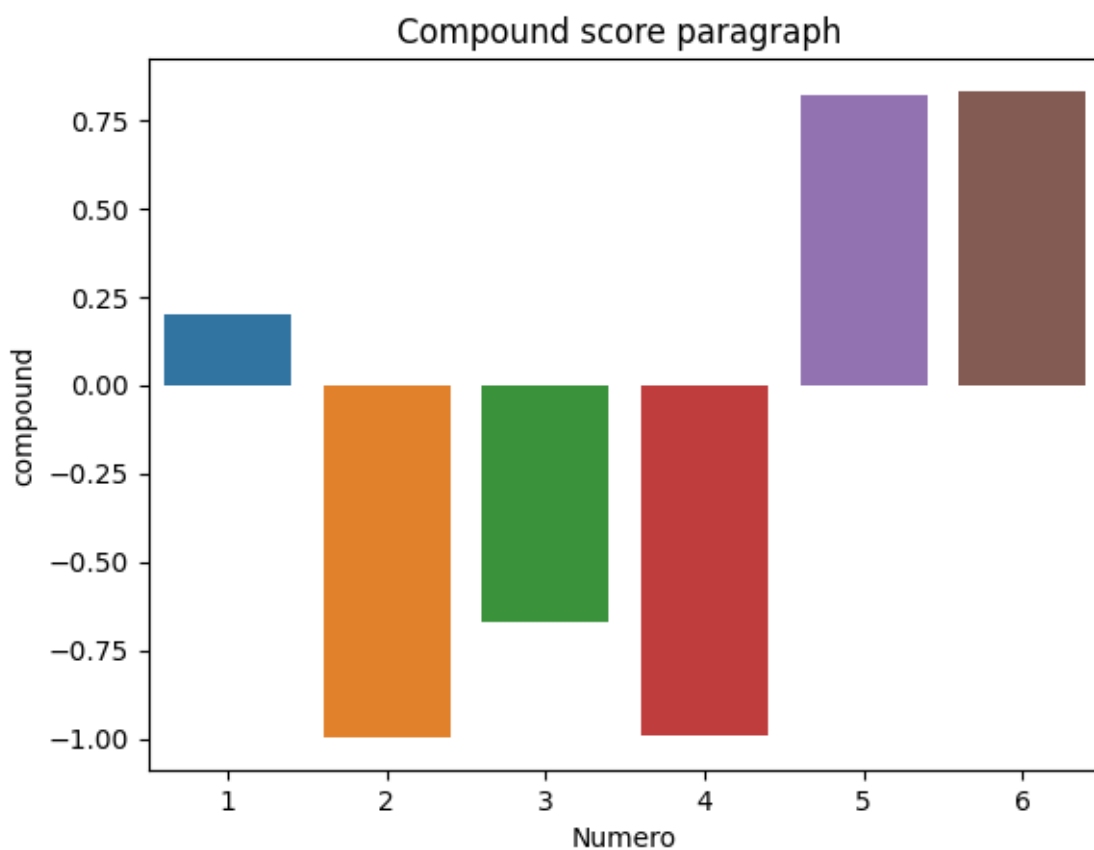


Figura 59 - Compound Score con Vader su dataset di Fentanyl

4.5.2. umBERTo

Applicando il modello umBERTo a *df_fentanyl*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.5469	0.4510	che cos'è Fentanyl-Hameln a cosa serve
2	0.5416	0.4583	cosa deve sapere prima di usare Fentanyl-Hameln
3	0.5365	0.4634	come usare Fentanyl-Hameln
4	0.5293	0.4706	possibili effetti indesiderati
5	0.5339	0.4660	come conservare Fentanyl-Hameln
6	0.5202	0.4797	contenuto della confezione e altre informazioni

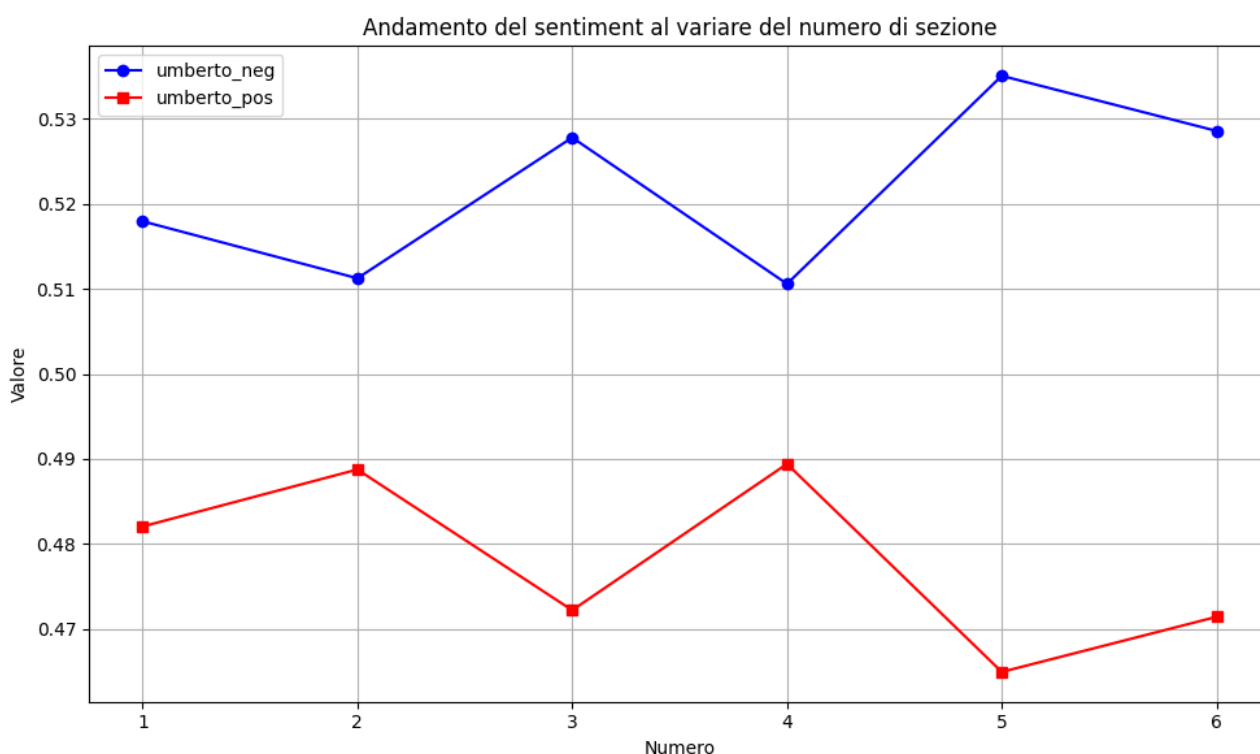


Figura 60 - Variazione del sentiment con umbERTO sul dataset di Fentanyl

Come nel caso del *df_diclorem*, il modello, come possiamo vedere in *Figura 60*, classifica ogni sezione con un sentiment prevalentemente negativo, fino al 53% per il paragrafo cinque, dove viene spiegato come conservare il farmaco. Ciò può essere dovuto al fatto che all'interno di questa sezione ci sono istruzioni poco chiare o tecniche, requisiti di conservazione difficili da seguire, o preoccupazioni riguardo la stabilità e l'efficacia del farmaco.

4.5.3. SentITA

Applicando il modello sentITA a *df_fentanyl*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	neg	pos	Sezione
1	0.4999	0.5000	che cos'è Fentanyl-Hameln a cosa serve
2	0.4999	0.5000	cosa deve sapere prima di usare Fentanyl-Hameln
3	0.3712	0.6287	come usare Fentanyl-Hameln
4	0.4997	0.5000	possibili effetti indesiderati
5	0.4992	0.5000	come conservare Fentanyl-Hameln
6	0.4999	0.5000	contenuto della confezione e altre informazioni

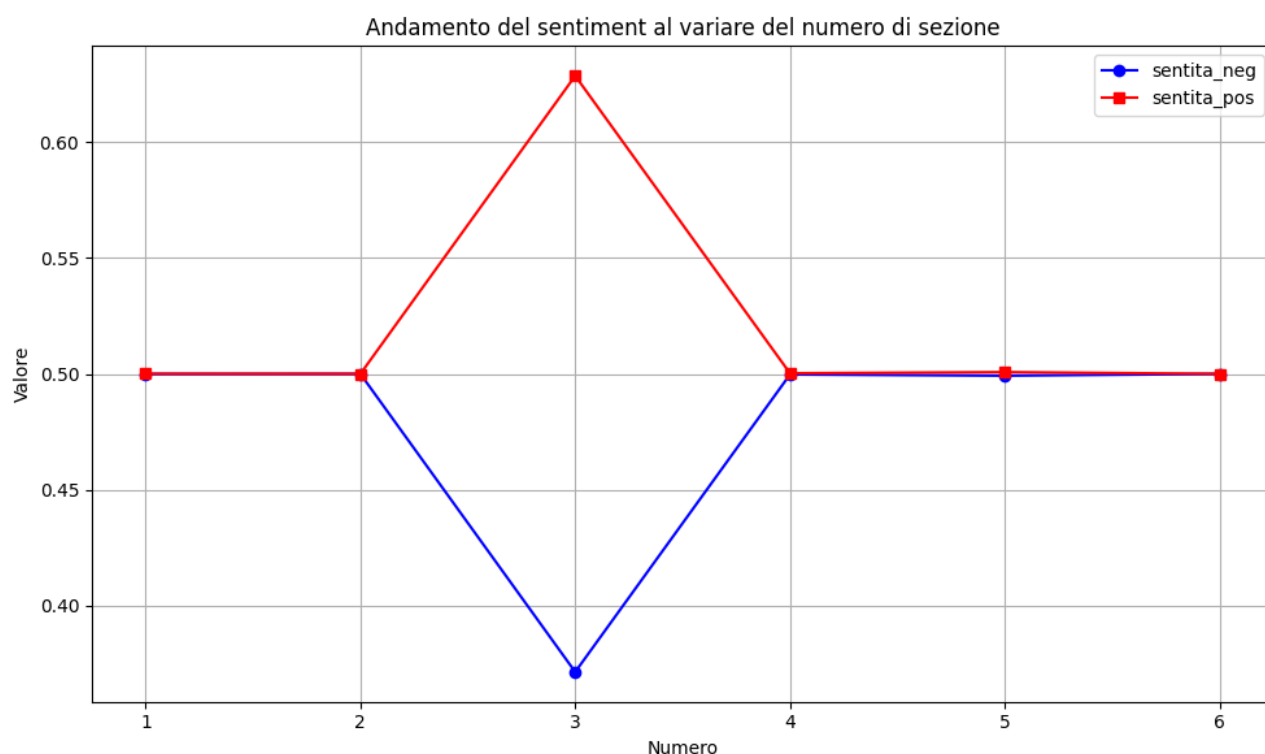


Figura 61 - Variazione del sentiment con SentITA sul dataset Fentanyl

Come in *Figura 45 e 53*, questo modello classifica il sentiment di questo bugiardino come perlopiù neutro. Il sentiment è molto positivo nella terza sezione dove troviamo le istruzioni come usare il farmaco. Ciò potrebbe essere dovuto a informazioni chiare e facili da seguire o una percezione di efficacia del farmaco che, essendo molto potente, ha bisogno di essere diretto e senza fraintendimenti. Eccezion fatta per la sezione 3, tutte le altre hanno la stessa percentuale di sentiment negativo e positivo.

4.5.4. Confronto risultati tra i modelli umBERTo e sentITA

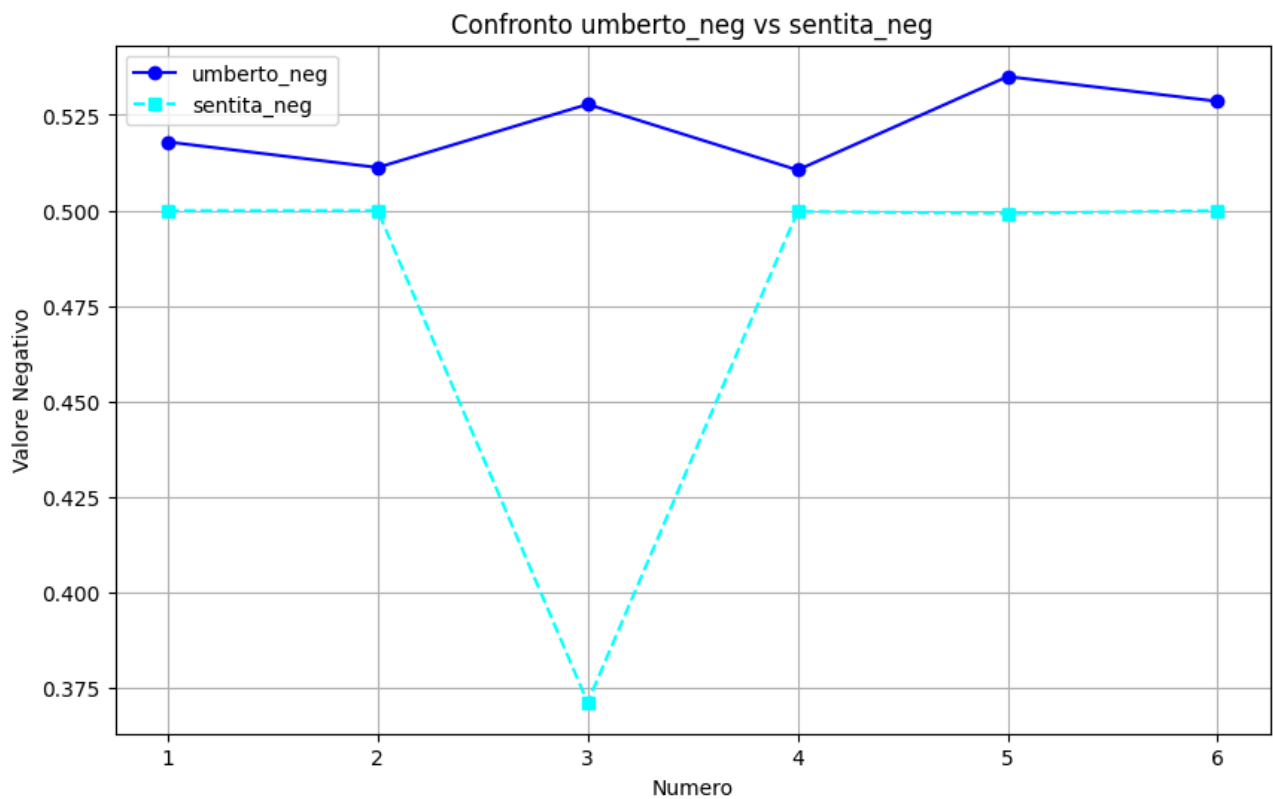


Figura 62 - Confronto sentiment 'neg' tra modelli sentITA e umBERTo sul dataset di Fentanyl

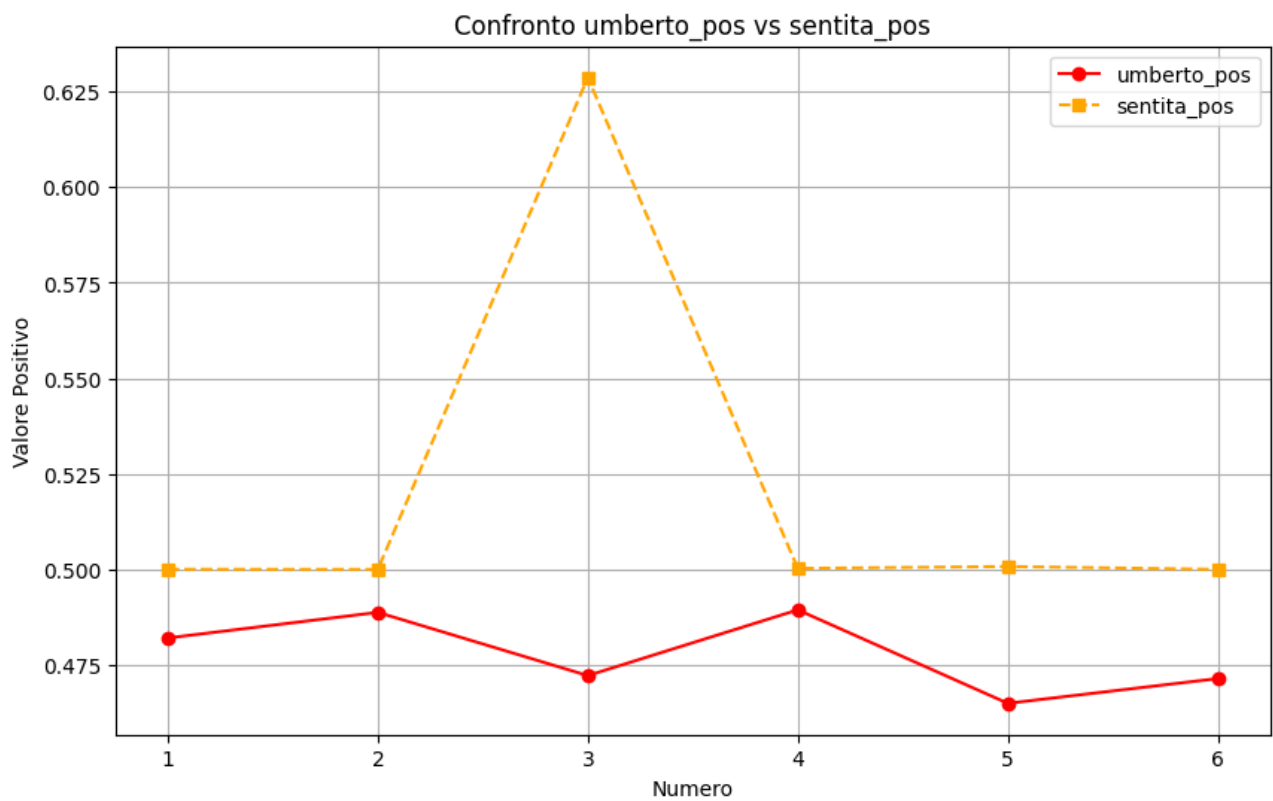


Figura 63 - Confronto sentiment 'pos' tra modelli sentITA e umBERTo sul dataset di Fentanyl

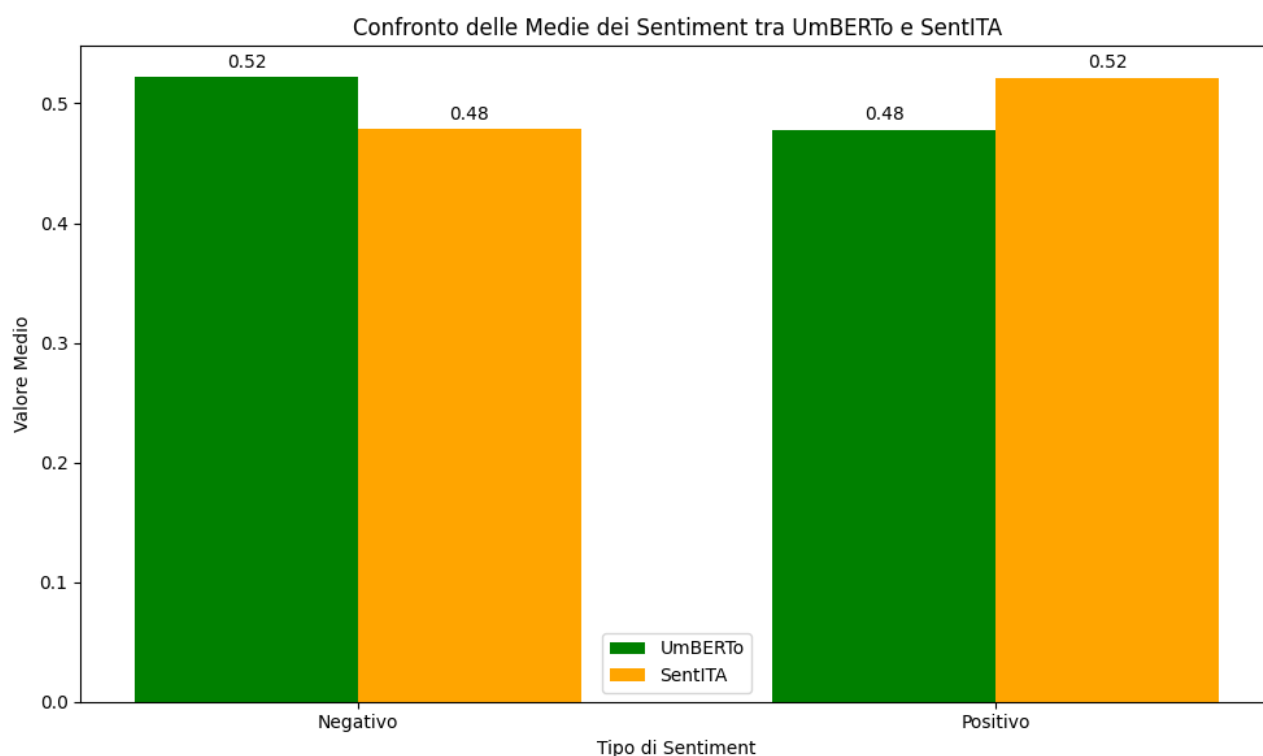


Figura 64 - Media pesata dei risultati del sentiment 'neg' e 'pos' dei modelli sentITA e umBERTo sul dataset Fentanyl

Dalle Figure 63 e 64, possiamo notare come i due modelli (anche se di poco) classifichino in maniera opposta le varie sezioni del bugiardino. Infatti, umBERTo classifica di più i vari paragrafi con sentiment negativo rispetto a sentITA che invece li classifica maggiormente con sentiment positivo. A confermare questa ipotesi c'è la Figura 64, in cui possiamo vedere come il 52% delle volte umBERTo ha classificato le istanze in maniera negativa contro il 52% in cui sentITA ha classificato le istanze in maniera positiva (in media).

4.5.5. FEEL-IT

Applicando il modello Feel-it a *df_fentanyl*, i risultati del sentiment delle varie sezioni del bugiardino sono i seguenti:

Numero	Emotion	Sezione
1	Fear	che cos'è Fentanyl-Hameln a cosa serve
2	Fear	cosa deve sapere prima di usare Fentanyl-Hameln
3	Fear	come usare Fentanyl-Hameln
4	Fear	possibili effetti indesiderati
5	Fear	come conservare Fentanyl-Hameln
6	Fear	contenuto della confezione e altre informazioni

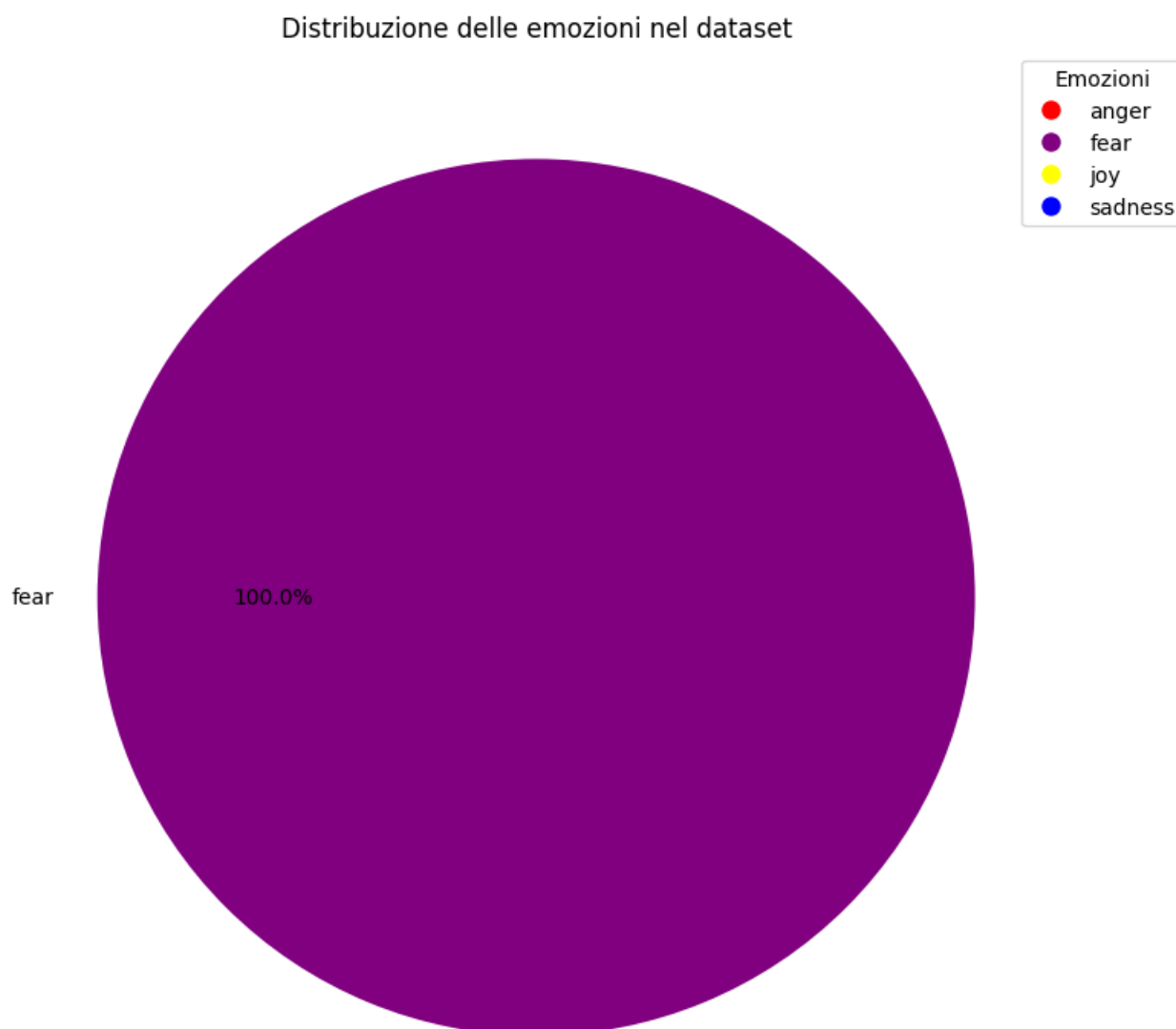


Figura 65 - Modello Feel-it sul dataset Fentanyl

Analogamente alla Morfina e per analoghe ragioni, il contenuto del bugiardino suscita principalmente un sentimento di paura o preoccupazione.

4.6. Topic Modelling

4.6.1. Latent Dirichlet Allocation (LDA)

I risultati ottenuti applicando il modello LDA, dopo aver scelto 5 come numero di topic, sono i seguenti:

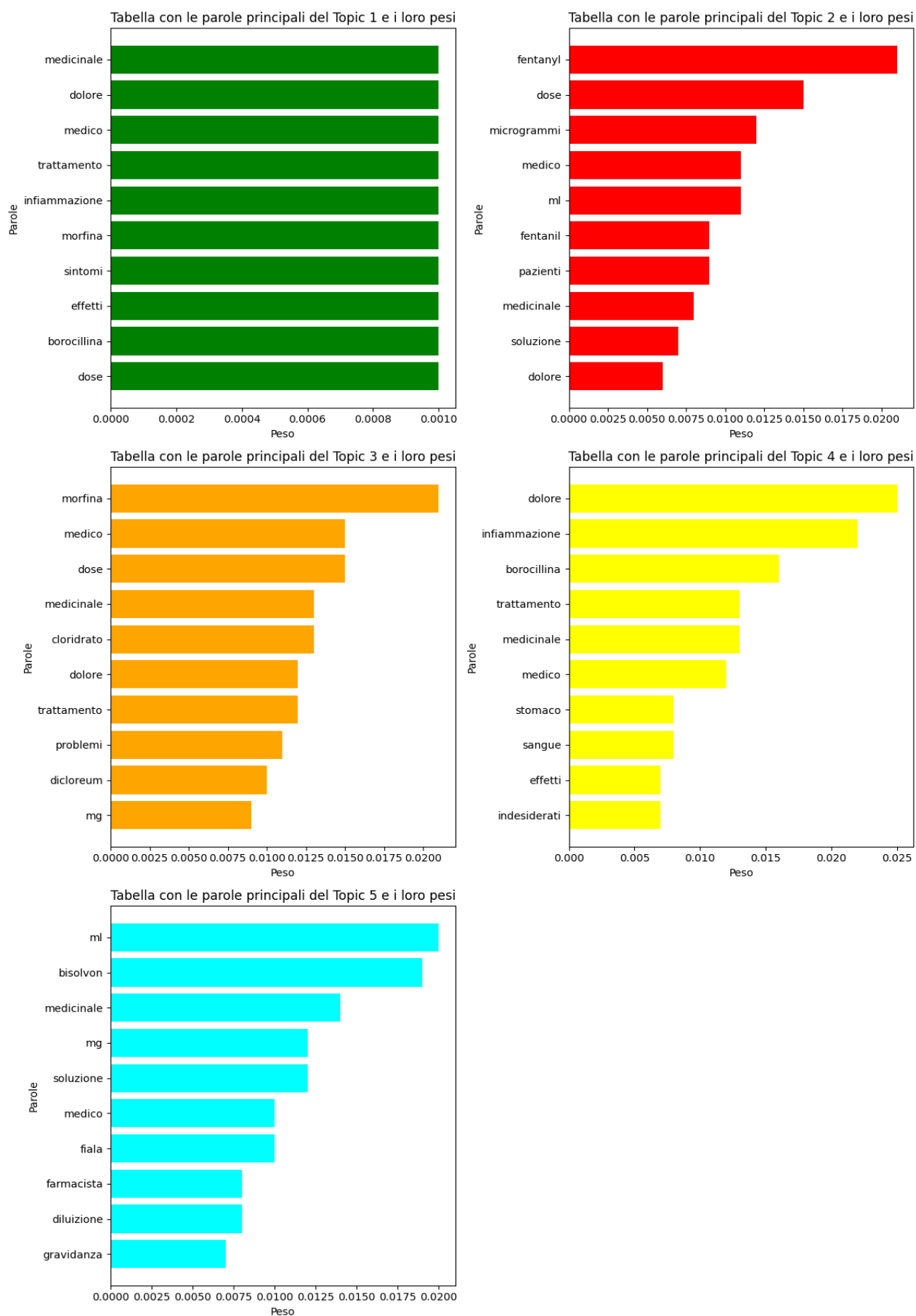


Figura 66 - Istogrammi delle parole e dei loro pesi per ogni topic

I topic ottenuti tramite LDA (Latent Dirichlet Allocation) forniscono un'idea dei temi principali trattati nei testi analizzati. Ogni topic è composto da un insieme di parole chiave, ognuna con un peso che rappresenta la sua importanza relativa all'interno del topic. Ecco una possibile spiegazione per ciascun topic:

- **Topic 1:** Questo topic include termini generici associati ai trattamenti medici. Le parole come "medicinale", "dolore", "medico" e "trattamento" indicano una discussione generale su come i medicinali sono utilizzati per trattare il dolore e altri sintomi. La presenza di parole come "infiammazione", "morfina" e "borocillina" suggerisce che si trattano anche specifici farmaci e condizioni mediche.
- **Topic 2:** Questo topic è fortemente focalizzato sui dosaggi e sulle sostanze specifiche come il fentanyl (fentanil). Le parole "dose", "microgrammi" e "ml" indicano una discussione dettagliata sui dosaggi e le unità di misura. Questo suggerisce un focus sui farmaci in forma liquida o in dosaggi specifici, rilevante per il trattamento di pazienti.
- **Topic 3:** Questo topic si concentra principalmente sulla morfina e sul suo uso nel trattamento del dolore. Parole come "morfina", "cloridrato" e "dose" indicano discussioni su specifiche forme e dosaggi di morfina. La presenza di "dicloream" e "problemi" suggerisce anche considerazioni su effetti collaterali e problematiche correlate.
- **Topic 4:** Questo topic è centrato sui trattamenti per il dolore e l'infiammazione, con una menzione specifica della borocillina. Le parole "dolore", "infiammazione" e "trattamento" suggeriscono discussioni su come questi farmaci vengono utilizzati per trattare tali condizioni. Termini come "stomaco", "sangue", "effetti" e "indesiderati" indicano anche un'attenzione agli effetti collaterali e alle reazioni avverse.
- **Topic 5:** Questo topic è incentrato sui trattamenti liquidi e specifici, come il Bisolvon. Le parole "ml", "soluzione", "fiala" e "diluizione" suggeriscono discussioni su forme liquide di farmaci e dosaggi. La presenza di "medico", "farmacista" e "gravidanza" indica anche considerazioni sul consiglio medico e le precauzioni durante la gravidanza.

Dopo l'identificazione dei topic principali, è stata esaminata la distribuzione dei singoli farmaci nei vari topic per comprendere quale tema fosse predominante per ciascun farmaco, basandosi sulla probabilità di appartenenza. Di seguito i risultati ottenuti:

Farmaco	Topic	Probabilità
Morfina	3	0.9996
Bisolvon	5	0.9987
Borocillina	4	0.9996
Dicloream	3	0.5822
Fentanyl	2	0.9995

I farmaci analizzati mostrano una forte associazione con i topic previsti, indicando che il modello LDA ha efficacemente catturato le principali tematiche legate a ciascun farmaco.

La maggior parte dei farmaci ha una probabilità molto alta (> 99%) di appartenenza al rispettivo topic predominante, con l'eccezione del Dicloream, che ha una probabilità più moderata (58.22%). Questo potrebbe indicare che il Dicloream è discusso in un contesto leggermente più variegato rispetto agli altri farmaci.

Infine, è stata calcolato il punteggio di coerenza UMass per ogni topic. La coerenza UMass misura quanto bene le parole in un topic sono semanticamente correlate tra loro. Di seguito i risultati ottenuti:

Topic	Coerenza Umass
1	-0.65
2	-0.13
3	-0.75
4	-0.22
5	-0.28

Interpretazione dei risultati:

- **Topic 1:** La coerenza di -0.65 indica una moderata correlazione tra le parole del Topic 1. Questo suggerisce che le parole come "medicinale", "dolore", "medico" e "trattamento" sono abbastanza coerenti tra loro, ma ci potrebbero essere alcune parole meno rilevanti che abbassano la coerenza complessiva.
- **Topic 2:** Con una coerenza di -0.13, il Topic 2 ha il miglior punteggio di coerenza tra tutti i topic. Questo indica che le parole come "fentanyl", "dose", "microgrammi" e "ml" sono fortemente correlate e trattano un argomento specifico in modo coerente.
- **Topic 3:** La coerenza di -0.75 indica una correlazione più bassa tra le parole del Topic 3 rispetto agli altri topic. Sebbene includa termini strettamente legati alla morfina e al trattamento del dolore, la presenza di parole come "problemi" e "diclorem" potrebbe introdurre variabilità tematica che riduce la coerenza.
- **Topic 4:** La coerenza di -0.22 per il Topic 4 suggerisce una buona correlazione tra le parole del topic. Termini come "dolore", "infiammazione", "borocillina" e "trattamento" sono strettamente legati, rendendo questo topic relativamente coerente.
- **Topic 5:** Con una coerenza di -0.28, il Topic 5 mostra una buona correlazione tra le parole, anche se non così alta come il Topic 2. Parole come "ml", "bisolvon", "medicinale" e "soluzione" indicano un focus su trattamenti liquidi e specifici.

5. Conclusioni

Lo studio del sentiment dei bugiardini di *Morfina*, *Bisolvon*, *Borocillina*, *Dicloream* e *Fentanyl* ha offerto una visione dettagliata delle sfumature emotive presenti in ciascuna sezione di questi foglietti illustrativi. Analizzando farmaci con diversi principi attivi ha permesso di comprendere se e come il contenuto specifico di ogni bugiardino influenzi il sentiment percepito.

Utilizzando **modelli** di analisi del sentiment addestrati su un lessico italiano, come **VADER**, **umBERTo**, **SentITA** e **Feel-it**, sono stati analizzati i differenti paragrafi di ogni bugiardino. I risultati del modello VADER hanno mostrato un sentiment generalmente neutro per la maggior parte delle sezioni dei bugiardini. Tuttavia, *Borocillina* e, in particolare, *Dicloream*, hanno evidenziato una neutralità significativamente inferiore. Per *Dicloream*, la sezione 1 ha mostrato il sentiment negativo maggiore, suggerendo che le informazioni introduttive o le avvertenze iniziali vengano percepite come più negative. Questo potrebbe essere attribuito alla natura dettagliata e potenzialmente allarmante delle avvertenze sugli effetti collaterali comuni nei farmaci antinfiammatori come *Borocillina* e *Dicloream*.

Il modello SentITA ha teso a classificare il contenuto dei bugiardini con un sentiment mediamente più positivo rispetto ad altri modelli, mentre umBERTo ha mostrato una maggiore propensione verso il sentiment negativo. In particolare, *Borocillina* è stata classificata positivamente da entrambi i modelli, indicando che il linguaggio utilizzato potrebbe essere percepito come rassicurante o meno allarmante. *Dicloream*, invece, ha evidenziato una forte distinzione: umBERTo ha classificato il 53% delle sezioni come negative, mentre SentITA ha assegnato un 52% di sentiment positivo, suggerendo una percezione polarizzata del farmaco. Questa polarizzazione potrebbe derivare dalla combinazione di un linguaggio tecnico dettagliato e delle specifiche avvertenze sui rischi associati all'uso di *Dicloream*.

Il modello Feel-it ha classificato ogni sezione di tutti i bugiardini con l'emozione 'fear'. Questa uniformità potrebbe essere attribuita al linguaggio tecnico e alle avvertenze dettagliate comuni nei foglietti illustrativi, che descrivono potenziali rischi ed effetti collaterali in un modo che può suscitare una risposta di paura nei lettori. La necessità di informare i pazienti sui possibili effetti avversi, pur mantenendo un linguaggio tecnico e preciso, potrebbe spiegare perché l'emozione 'fear' sia prevalente.

Una delle ipotesi derivanti da questi risultati è che la presentazione delle informazioni e il linguaggio utilizzato nei bugiardini giocano un ruolo cruciale nel determinare il sentiment. Anche **se il principio attivo non sembra essere un fattore determinante**, la natura del farmaco e la necessità di descrivere potenziali rischi in dettaglio possono influenzare negativamente il sentiment percepito dai pazienti. Per esempio, i farmaci antinfiammatori come *Borocillina* e *Dicloream* tendono ad avere avvertenze più dettagliate sui potenziali effetti collaterali, il che può contribuire a un sentiment più negativo.

Un'altra ipotesi è che il pubblico destinatario e il contesto di utilizzo del farmaco influenzino la percezione emotiva del bugiardino. Farmaci destinati a trattamenti più complessi o condizioni gravi potrebbero richiedere un linguaggio più cauto e dettagliato, generando così un sentiment più negativo o una maggiore paura tra i lettori.

Un aspetto curioso emerso dallo studio è che, nonostante *Morfina* e *Fentanyl* siano farmaci altamente potenti e delicati, il sentiment negativo non è risultato più marcato rispetto ad altri farmaci analizzati. Una possibile spiegazione è che i bugiardini di questi farmaci, pur trattando sostanze di grande impatto, siano redatti con un linguaggio particolarmente attento e professionale, mirato a rassicurare

i pazienti e a spiegare in modo chiaro ma non allarmante i potenziali rischi e benefici. Inoltre, il pubblico destinatario di *Morfina e Fentanyl* potrebbe essere già consapevole della natura e dei rischi associati a questi farmaci, il che riduce l'impatto negativo del contenuto informativo del bugiardino.

L'analisi del sentiment dei bugiardini suggerisce che il tono del linguaggio è influenzato principalmente dalle modalità di comunicazione e dalle avvertenze generali piuttosto che dal principio attivo del farmaco. I bugiardini di farmaci antinfiammatori come *Borocillina e Dicloream* tendono a generare sentiment più negativi a causa delle avvertenze dettagliate sui potenziali effetti collaterali. La differenza tra i modelli di analisi del sentiment, con SentITA più positivo e umBERTo più negativo, sottolinea l'importanza di utilizzare approcci diversi per una valutazione completa. Inoltre, l'emozione predominante di 'fear' rilevata da Feel-it in tutte le sezioni evidenzia che il linguaggio precauzionale nei bugiardini tende a suscitare preoccupazioni universali. Questo studio mette in luce la necessità di bilanciare l'accuratezza delle informazioni con un tono che non alimenti eccessivamente le paure dei pazienti. Sarebbe bene, in futuro, esplorare strategie di comunicazione che possano ridurre l'ansia dei pazienti senza compromettere la chiarezza e la completezza delle informazioni potrebbe migliorare l'efficacia dei bugiardini.

L'analisi della distribuzione dei vari topic nei bugiardini ha mostrato che il modello LDA è stato efficace nel catturare le tematiche principali di ogni farmaco. Questa analisi ha rivelato come ogni farmaco sia descritto con un linguaggio specifico che si riflette nei rispettivi topic, confermando che il contenuto dei bugiardini è strettamente correlato alle caratteristiche e agli usi dei farmaci.

In conclusione, la comunicazione nei bugiardini richiede un equilibrio delicato: fornire tutte le informazioni necessarie per l'uso sicuro ed efficace del farmaco, senza creare inutili allarmismi. Un linguaggio troppo tecnico o eccessivamente dettagliato sui potenziali effetti negativi può spaventare i pazienti, mentre un approccio troppo superficiale potrebbe non fornire le necessarie indicazioni di sicurezza.

Questo studio evidenzia che il principio attivo non è il principale determinante del sentiment. Le modalità con cui le informazioni vengono presentate giocano un ruolo cruciale nel modo in cui i pazienti percepiscono i rischi e i benefici dei farmaci.

6. Sitografia

1. <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/home>
2. <https://github.com/NicGian/SentITA>
3. <https://stackoverflow.com/questions/62284611/sentiment-analysis-of-italian-sentences>
4. <https://github.com/brunneis/vader-multi>
5. <https://github.com/MilaNLProc/feel-it>
6. <https://towardsdatascience.com/sentiment-analysis-and-emotion-recognition-in-italian-using-bert-92f5c8fe8a2>
7. <https://huggingface.co/MilaNLProc/feel-it-italian-sentiment>
8. <https://huggingface.co/Musixmatch/umberto-commoncrawl-cased-v1>
9. <https://github.com/musixmatchresearch/umberto/>

7. Data Cards