

# Emozioni pubbliche in tempo di guerra: analisi del sentiment nei tweet sui conflitti Israele-Palestina e Ucraina-Russia

1<sup>st</sup> Ruggiero Matteo

Dipartimento di Informatica

Università degli Studi di Salerno

Fisciano (SA)

m.ruggiero97@studenti.unisa.it

**Abstract**—Questo studio esplora l’analisi dei sentimenti e delle emozioni nei testi pubblicati sui social network, focalizzato su eventi geopolitici e crisi internazionali. Attraverso l’applicazione di modelli linguistici preaddestrati basati su architetture Transformer (disponibili su Hugging Face), vengono analizzati due dataset composti, rispettivamente, da tweet relativi ai conflitti Russia–Ucraina e Israele–Palestina. Utilizzando tecniche di sentiment analysis a più classi (emotion classification) e polarizzazione binaria (positivo, negativo, neutro), lo studio confronta le risposte emotive e discorsive nei due scenari, evidenziando analogie e differenze nei pattern emotivi prevalenti. I risultati mostrano una rilevante prevalenza di emozioni quali paura e rabbia in entrambi i casi, sebbene i meccanismi di diffusione e il tono delle narrazioni appaiano legati al contesto socioculturale di riferimento. L’analisi evidenzia inoltre possibili limitazioni dovute alla comunicazione ironica o sarcastica tipica dei social, nonché alla disparità dimensionale dei dataset.

**Index Terms**—NLP, Hugging Face, BERT, Sentiment Analysis, Emotional Analysis

## I. INTRODUZIONE

Al giorno d’oggi, i conflitti armati hanno trovato una nuova area di rappresentazione e confronto: i social network. In particolare Twitter (attualmente X) si è affermato come strumento fondamentale per analizzare in tempo reale l’opinione pubblica globale riguardo gli avvenimenti geopolitici e non solo. I messaggi pubblicati dagli utenti, spesso brevi, sarcastici e intensamente emotivi, costituiscono una fonte preziosa per comprendere le dinamiche di coinvolgimento emotivo e diffusione delle informazioni durante la crisi.

In questo progetto, lo scopo è studiare e confrontare il sentiment/emotion espresso su Twitter in relazione a due dei conflitti più rilevanti e mediaticamente seguiti dell’ultimo decennio: **la guerra tra Israele e Palestina e il conflitto tra Russia e Ucraina**.

Utilizzando tecniche di Natural Language Processing (NLP) e modelli preaddestrati per l’analisi del sentiment e dell’emotion, questo studio si propone di analizzare le principali tendenze emotive manifestate dagli utenti in entrambi i contesti di conflitto. L’intento è analizzare le somiglianze e le differenze nella percezione della società, mettendo in

risalto le dinamiche della comunicazione, le narrazioni dominanti e i potenziali fenomeni di polarizzazione. L’analisi ha l’obiettivo di aiutare a capire l’importanza dei social media nella formazione dell’opinione pubblica in momenti di crisi internazionale. Ogni sezione ha lo scopo di studiare e rispondere a diverse research question, che sono riassunte nella Tabella I sottostante:

|            |  |
|------------|--|
| <b>RQ1</b> | È possibile capire cosa è successo un determinato giorno grazie all’utilizzo di tecniche di NLP (BERTopic) sui tweet?                                      |
| <b>RQ2</b> | Qual è l’emozione dominante nei vari topic rilevati da BERTopic?   |
| <b>RQ3</b> | Scegliendo un periodo “soddisfacente”, è possibile osservare una predominanza di emozioni positive anche durante un evento tragico come la guerra?         |
| <b>RQ4</b> | Qual è il sentimento degli utenti di Twitter relativo a determinate occorrenze?  |
| <b>RQ5</b> | Qual è la prevalenza del sentimento degli utenti in questo caso?   |
| <b>RQ6</b> | Quali differenze emergono nell’emozione trasmessa dai tweet relativi ai due conflitti messi a confronto?   |
| <b>RQ7</b> | Un evento uguale (bombardamento) produce una risposta discorsiva più coesa e polarizzata nel caso israelo-palestinese rispetto al conflitto russo-ucraino? |

TABLE I: Domande di ricerca (RQs)

## II. RELATED WORK

In [1] gli autori propongono un’analisi quantitativa e del tono dei tweet relativi alla guerra russo-ucraina pubblicati da leader politici di Stati Uniti, Regno Unito, Germania, Francia e Polonia. Utilizzando il software LIWC-22, le autrici analizzano 111 tweet, suddividendo il lavoro in tre fasi principali: (1) analisi del tono (positivo/negativo/neutro), (2) studio della frequenza dei termini, e (3) previsione dell’andamento emotivo nei tweet futuri.

I risultati mostrano un’evoluzione significativa del tono nel tempo: prima dell’inizio della guerra i tweet erano tendenzialmente positivi o neutri, mentre nel primo mese di conflitto

si è osservato un aumento sia del tono positivo (soprattutto nei messaggi di sostegno) sia del tono negativo (legato alla condanna dell'aggressione). L'analisi semantica evidenzia l'uso ricorrente di parole come Ukraine, war, Russia, support, e freedom, spesso accompagnate da espressioni retoriche e marcatori politici come slava Ukraini o sham referendums.

Infine, gli autori utilizzano un modello di regressione quadratica per prevedere il tono dei prossimi tweet dei leader analizzati. I risultati indicano una prevalente crescita del tono negativo nei tweet futuri, soprattutto da parte del Cancelliere tedesco e del Presidente statunitense, mentre per il Presidente polacco si prevede un andamento più positivo.

In [2] gli autori propongono un'analisi del sentiment su larga scala dei tweet relativi al conflitto Russia-Ucraina, focalizzandosi sul periodo tra gennaio e marzo 2022. Lo studio si basa su un corpus di oltre 600.000 tweet in inglese e circa 1.600 in russo, raccolti tramite hashtag tematici connessi alla guerra. Per la classificazione dei tweet, gli autori impiegano modelli pre-addestrati disponibili su Hugging Face: nello specifico, DistilRoBERTa-base per i tweet in inglese e XLM-RoBERTa-base per quelli in lingua russa. I tweet in inglese sono stati suddivisi in sette categorie emozionali (rabbia, disgusto, paura, gioia, tristezza, sorpresa, neutro), mentre quelli in russo sono stati classificati in polarità positiva, negativa o neutra.

I risultati mostrano che nei tweet in inglese dominano la paura (32%) e la rabbia (15%), mentre nei tweet in russo circa l'87% è caratterizzato da sentiment negativo. Questi dati suggeriscono che la percezione pubblica della guerra, nelle prime fasi del conflitto, sia stata fortemente associata a emozioni di insicurezza, condanna e allarme. Le espressioni linguistiche più ricorrenti riflettono solidarietà verso l'Ucraina, condanna dell'invasione e preoccupazione per le vittime civili e le conseguenze geopolitiche del conflitto.

In [3] gli autori analizzano la propagazione spaziale e temporale dei sentimenti espressi su X/Twitter riguardo al conflitto israelo-palestinese esploso nell'ottobre 2023. Il dataset include 151.406 tweet geolocalizzati, raccolti da 146 Paesi tra il 7 ottobre e il 24 novembre 2023. Gli autori utilizzano un modello BERT pre-addestrato e fine-tuned (in particolare, bert-base-cased da Hugging Face) per classificare ogni tweet in sei categorie emotive: anger, fear, joy, love, sadness e surprise.

L'analisi si articola in tre fasi: (1) classificazione dei tweet per emozione, (2) costruzione di una rete di propagazione basata sui reply tra utenti, e (3) rilevamento delle comunità di diffusione del sentiment tramite algoritmi di clustering. I risultati mostrano che le emozioni più espresse nel corpus sono rabbia e tristezza, con picchi significativi di fear e surprise in corrispondenza degli attacchi iniziali e dei bombardamenti più intensi. L'emozione "love" appare raramente ma aumenta leggermente in concomitanza con appelli alla pace e mobilitazioni umanitarie.

### III. BACKGROUND

#### A. BERTopic

BERTopic è un metodo di topic modeling basato su transformer, per generare rappresentazioni semantiche profonde dei

testi. BERTopic opera direttamente sugli embedding testuali prodotti da modelli preaddestrati. Il procedimento inizia con la creazione di embed densità per ogni documento, seguito da una riduzione delle dimensioni attraverso UMAP o PCA, che permette di mantenere le relazioni strutturali tra i dati. In seguito, i punti vengono uniti utilizzando algoritmi di clustering densi, quali HDBSCAN o KMeans. Per ciascun cluster, BERTopic adotta un metodo chiamato class-based TF-IDF (c-TF-IDF) per individuare le parole chiave che descrivono il contenuto in modo comprensibile. In questo progetto, il modello è stato impiegato per individuare e analizzare i principali argomenti discussi su Twitter in relazione ai conflitti Russia-Ucraina in relazione ad un determinato giorno offrendo una base tematica utile per l'interpretazione del emotion dei tweets raccolti.

#### B. j-hartmann/emotion-english-distilroberta-base [4]

È un modello preaddestrato per il riconoscimento delle emozioni nei testi in lingua inglese. Basato sull'architettura DistilRoBERTa, il modello è stato fine-tuned specificamente per la classificazione multi-classe delle emozioni in frasi brevi o singole espressioni testuali. Il modello è stato addestrato su una versione annotata del dataset GoEmotions, un ampio corpus sviluppato da Google contenente oltre 58.000 frasi etichettate con 28 emozioni distinte più una classe neutra. Tuttavia, nella versione fornita da J. Hartmann, l'output è semplificato in 6 emozioni primarie più la classe neutra: *joy, sadness, anger, fear, surprise, disgust, e neutral*.

#### C. Davlan/bert-base-multilingual-cased-ner-hrl [5]

È un modello preaddestrato per il Named Entity Recognition (NER) in più lingue, basato sull'architettura BERT. Il modello è stato fine-tuned su dati NER multilingue ed è in grado di riconoscere entità come persone (PER), organizzazioni (ORG), luoghi (LOC) e altri tipi di nomi propri in testi scritti in diverse lingue. Il modello mantiene la distinzione tra lettere maiuscole e minuscole e supporta alfabeti diversi.

#### D. cardiffnlp/twitter-roberta-base-sentiment [6]

È un modello preaddestrato per la sentiment analysis ottimizzato specificamente per il linguaggio di Twitter. Si basa sull'architettura RoBERTa-base ed è stato ulteriormente fine-tuned su un ampio corpus di tweet annotati con etichette di sentiment, rendendolo particolarmente adatto a catturare le sfumature linguistiche, gergali e informali tipiche dei social media inclusivi di abbreviazioni, emoji, hashtag e slang. Il modello classifica ciascun testo in tre classi principali: *positive, neutral e negative*.

#### E. sentence-transformers/all-mpnet-base-v2 [7]

È un modello di sentence embedding creato per produrre rappresentazioni dense e ricche di significato di frasi o brevi testi, mantenendo le relazioni semantiche anche in testi complessi o grammaticalmente imprecisi.

## IV. DATASETS E PREPROCESSING

### A. Datasets

Sono stati utilizzati due dataset separati per la realizzazione di questo progetto.

Per quanto riguarda la guerra Russia-Ucraina è stato utilizzato Ukraine-Russian Crisis Twitter Dataset [8], reperibile da Kaggle. Il dataset è composto da oltre 1,2 milioni di tweets relativi al conflitto. I dati sono raccolti giornalmente tramite l'API di Twitter. Le feature presenti nel dataset sono: **text** (contenuto testuale del tweet), **user\_location** (posizione dichiarata dall'utente), **user\_created** (data di creazione dell'account), **retweet\_count**, **reply\_count**, **like\_count** (metriche di interazione).

Per lo studio della guerra Israele-Palestina è stato utilizzato Israel-Palestine Conflict Tweets Dataset [9], anche esso reperibile su Kaggle. Il dataset è composto da numero di Tweets totali pari a 15,478 compresi tra il 17 Ottobre 2023 e il 17 Dicembre 2023, ciascuno identificato da un campo **id** univoco e accompagnato da un **link** diretto (link) al post originale su Twitter. Il contenuto testuale del tweet è riportato nella colonna **text**, mentre la **data** e l'**ora** di pubblicazione sono indicate nel campo **date**.

### B. Preprocessing

Le tecniche di preprocessing adottate per la pulizia e la normalizzazione del testo sono state le medesime per entrambi i dataset analizzati.

I tweet originali contenevano frequentemente elementi come emoticon, menzioni (*username*), link esterni e hashtag. Le informazioni considerate non rilevanti ai fini dell'analisi, ovvero emoticon, menzioni e collegamenti ipertestuali, sono state rimosse in fase di pulizia.

Al contrario, gli hashtag sono stati trattati come componenti informative rilevanti, in quanto spesso racchiudono parole chiave significative legate al contenuto e al contesto del tweet. Per questo motivo, si è proceduto alla loro segmentazione tramite tecniche di word segmentation, con l'obiettivo di suddividerli in sequenze di parole semanticamente interpretabili. A tal fine, è stata utilizzata la libreria **wordsegment**, integrata con **NLTK** per il supporto linguistico di base. **wordsegment** si basa su un modello probabilistico costruito su un ampio corpus di lingua inglese e consente la segmentazione di stringhe continue secondo la massima probabilità lessicale. Una volta segmentati, gli hashtag sono stati privati del simbolo #, mantenendo però le parole estratte nel testo, trattandole come parte integrante del messaggio da analizzare. Inoltre sono stati rimossi i possibili duplicati e, dove possibile, i tweet sono stati filtrati per lingua inglese.

### C. NER Preprocessing

Per migliorare l'efficacia del processo di Named Entity Recognition (NER), è stata implementata una nuova fase di normalizzazione semantica su un elenco di entità pertinenti al conflitto israelo-palestinese. Questa fase prevede la sostituzione automatica di nomi ed espressioni conosciute con

versioni correttamente capitalizzate, conformi alle aspettative del modello NER. La procedura si avvale di un dizionario di entità riconosciute e di una mappa personalizzata che mantiene inalterate sigle e nomi propri (idf diventa IDF, unrwa diventa UNRWA, e united nations diventa United Nations), realizzando la sostituzione in modo prioritario per le frasi più lunghe, per prevenire conflitti. Questo intervento ha portato a un incremento della precisione nel riconoscimento delle entità, preservando la struttura linguistica originale del tweet.

## V. RICONOSCIMENTO DEGLI AVVENTIMENTI MEDIANTE BERTOPIC CONFLITTO RUSSO-UCRIANO

Il primo studio condotto è stato finalizzato a rispondere alla prima Research Question (RQ), "*È possibile riuscire a capire cosa è successo un determinato giorno grazie all'utilizzo di tecniche di NLP (BERTopic) sui tweet?*"

A tal fine, è stato selezionato un giorno casuale all'interno del periodo osservato e applicato il modello BERTopic al corpus dei tweet pubblicati in quella data. L'obiettivo era verificare se i temi emersi dal topic modeling fossero sufficientemente informativi da ricostruire gli eventi salienti della giornata, attraverso l'analisi delle parole chiave e dei tweet rappresentativi per ciascun topic.

Successivamente, è stata affrontata la seconda Research Question: "*Qual è l'emozione dominante nei vari topic rilevati da BERTopic?*". In questo caso, l'attenzione si è spostata sull'aspetto emotivo dei contenuti, con l'obiettivo di associare a ciascun topic un'emozione prevalente, sfruttando modelli di sentiment analysis fine-tuned per la classificazione emotiva.

Infine, sulla base dei risultati emersi da questa fase di analisi, è stata formulata una terza Research Question: "*Scegliendo un periodo "soddisfacente", è possibile osservare una predominanza di emozioni positive anche durante un evento tragico come la guerra?*".

### A. Procedimento e discussione sui risultati ottenuti

Al fine di rispondere a questa domanda, è stato selezionato un giorno casuale all'interno del dataset, corrispondente al **28 maggio 2023**. Dall'analisi temporale preliminare è emerso che la maggior parte dei tweet è stata pubblicata in due intervalli distinti: tra le 11:30 e le 12:05 e tra le 16:15 e le 16:50. Per questo motivo, i dati utilizzati in questo studio sono stati ulteriormente filtrati anche in base all'orario, al fine di focalizzarsi sui picchi di attività più rilevanti della giornata. In seguito all'applicazione delle tecniche di preprocessing, descritte in IV-B, è stato possibile applicare il modello BERTopic alla colonna "text" del dataset. Il processo ha previsto il calcolo degli embedding e la tokenizzazione automatica mediante il modello preaddestrato *paraphrase-multilingual-MiniLM-L12-v2*. Gli embedding così ottenuti sono stati successivamente passati a UMAP per la riduzione dimensionale e a HDBSCAN per il clustering non supervisionato. I risultati ottenuti da BERTopic sono riportati nella Tabella II:

TABLE II: Riepilogo dei topic generati da BERTopic

| Topic | Count | Name                         |
|-------|-------|------------------------------|
| -1    | 122   | _ukraine_war_the_russia      |
| 0     | 336   | _ukraine_the_war_russia      |
| 1     | 31    | _ukraine_war_drones_drone    |
| 2     | 19    | _graham_money_russians_spent |
| 3     | 13    | _media_the_russia_in         |

TABLE III: Representative Documents per Topic

| Topic | Representative Document                                     |
|-------|---|
| 0     | British intelligence Ukraine war map as of today...         |
| 1     | Russian airborne troops somewhere in Ukraine...             |
| 2     | Russian army launched a massive drone attack...             |
| 3     | US Senator Lindsey Graham during a meeting with Zelensky... |
| 4     | The British propaganda newspaper The Times has...           |

I risultati possono essere interpretati nel seguente modo:  
**Topic -1 e 0**

risultano esserci riferimenti frequenti a Kyiv, Bakhmut, e all'escalation militare in corso. I tweet parlano di attacchi aerei, difesa antiaerea attiva e forte tensione nella capitale ucraina, inclusa la percezione di esplosioni simili a un terremoto. C'è inoltre una componente fortemente patriottica e identitaria, con hashtag come stand with Ukraine.

#### Topic 1

Questo topic si concentra su un evento specifico: un attacco notturno con droni kamikaze Shahed. Diversi tweet parlano di 52 droni abbattuti su 54, suggerendo che si è trattato di uno degli attacchi più intensi fino a quel momento. Le città prese di mira includono Kyiv e altre aree dell'Ucraina, confermando un picco di attività militare russa proprio in quella data

#### Topic 2

Questo topic si concentra sulle dichiarazioni del senatore statunitense Lindsey Graham. Queste affermazioni hanno generato forti reazioni online, con accuse di cinismo, militarismo e propaganda da parte di alcuni utenti, e sostegno patriottico da altri.

#### Topic 3

evidenzia una riflessione sulla copertura mediatica occidentale del conflitto. I tweet mostrano scetticismo verso i media e una critica all'evoluzione della narrazione strategica della guerra, ad esempio sulla città di Bakhmut.

Con una breve e intuitiva ricerca su internet è stato possibile effettuare un confronto tra i risultati ottenuti da BERTopic e le notizie effettive pubblicate sulle testate giornalistiche che seguono gli avvenimenti della guerra giorno per giorno [10].

#### Abbattuti 52 droni su 54, alla vigilia del Kiev day

L'esercito ucraino questa notte ha abbattuto 52 droni di fabbricazione iraniana nello spazio aereo del Paese sui 54 lanciati dall'esercito del Cremlino proprio alla vigilia del Kiev Day, quando la capitale celebra l'anniversario della sua fondazione di 1541 anni fa con concerti, fiere e mostre. L'Aeronautica militare ucraina ha reso noto l'abbattimento degli Shaded lanciati dalle regioni russe di Bryansk e Krasnodar indicando che l'attacco è stato effettuato a più ondate per 5 ore, come riporta Rbc-Ucraina. Secondo informazioni non definitive, un uomo di 40 anni è rimasto ucciso in un distretto di Kiev e una donna è stata ferita.

#### (a) Notizia catturata da Topic 1

##### Kiev: bombardate città a est e ovest dell'Ucraina

I raid aerei dell'esercito russo hanno colpito questa mattina alle cinque Zhytomyr, in Ucraina occidentale, come riferisce l'emittente pubblica Suspilne. Nella notte è stata bombardata anche Nikopol, nella regione orientale di Dnepropetrovsk, ha reso noto il capo militare regionale Sergey Lysak: «Sei abitazioni e due edifici sono stati distrutti. Il gasdotto e le linee elettriche sono stati danneggiati. Non ci sono stati morti o feriti: questa è la cosa principale», ha scritto su Telegram.

#### (b) Notizia catturata da Topic -1 e 0

Ucraina: Cremlino, 'parole senatore Graham su dei russi vergogna per Usa'

#### (c) Notizia catturata da Topic 3

Fig. 1: Notizie relative al 28 Maggio 2023 (La Stampa).

Come mostrato nella Figura 1, le notizie emerse tramite BERTopic risultano coerenti con le principali informazioni reperibili online e pubblicate da una delle testate giornalistiche più autorevoli. Questo confronto consente di rispondere positivamente alla prima Research Question, confermando che l'utilizzo di BERTopic sui tweet permette di ricostruire con buona precisione gli eventi salienti di una giornata specifica.

### B. Emotional Analysis

Al fine di rispondere alla seconda Research Question, è stata condotta un'analisi emozionale su due livelli. In un primo momento, l'analisi è stata applicata separatamente a ciascun topic individuato da BERTopic, al fine di identificare l'emozione dominante all'interno di ogni gruppo tematico. Successivamente, è stato eseguito uno studio complessivo sull'intero corpus, utilizzando il modello preaddestrato *j-hartmann/emotion-english-distilroberta-base(III-B)*, specificamente ottimizzato per la classificazione delle emozioni nei testi in lingua inglese.

I risultati derivanti da entrambe le analisi sono riportati in Tabella IV e Tabella V.

TABLE IV: Distribuzione delle emozioni per topic

| Emozione | Topic -1 | Topic 0 | Topic 1 | Topic 2 | Topic 3 |
|----------|----------|---------|---------|---------|---------|
| Anger    | 24       | 101     | 12      | 2       | 1       |
| Fear     | 19       | 63      | 10      | 2       | 2       |
| Joy      | 10       | 25      | 3       | 3       | 0       |
| Neutral  | 53       | 91      | 2       | 1       | 9       |
| Sadness  | 13       | 46      | 1       | 6       | 1       |
| Surprise | 3        | 8       | 3       | 2       | 0       |
| Disgust  | 0        | 2       | 0       | 0       | 0       |

Dall'osservazione dei dati emerge chiaramente che le emozioni predominanti sono la **rabbia** e la **paura**, a testimonianza della forte carica emotiva con cui vengono commentati

TABLE V: Distribuzione delle emozioni nei tweet analizzati

| Emozione | Numero di tweet |
|----------|-----------------|
| Neutral  | 156             |
| Anger    | 154             |
| Fear     | 95              |
| Sadness  | 67              |
| Joy      | 41              |
| Surprise | 16              |
| Disgust  | 2               |

gli eventi bellici. Si registra inoltre una significativa presenza del sentimento neutrale, che può essere attribuita al fatto che Twitter viene spesso utilizzato anche per la diffusione di informazioni e aggiornamenti oggettivi, privi di una marcata componente soggettiva o emotiva.

### C. Liberazione di Kherson per studio comparativo.

Al fine di rispondere alla terza RQ specificata in V, è stato considerato un giorno "felice" per l'opinione pubblica, **la liberazione di Kherson**, al fine di osservare se a predominare nell'emozione dei sentimenti fossero sentimenti positivi e paragonare i due avvenimenti da un punto di vista emozionale

1) *Procedimento e discussione sui risultati.*: Per questo studio sono stati aggregati in un unico dataset i tweet pubblicati tra il 10 e il 13 novembre 2022, con l'obiettivo di analizzare non soltanto il giorno della liberazione di Kherson, ma anche i giorni immediatamente precedenti e successivi.

Le tecniche di preprocessing adottate sono le stesse descritte nella Sezione IV-B, e hanno consentito di ottenere una base dati significativamente più ampia rispetto allo studio precedente, per un totale di 174.484 tweet validi.

Rispetto all'implementazione iniziale di BERTopic, sono state introdotte alcune modifiche metodologiche: il modello per il calcolo degli embedding è rimasto invariato (*paraphrase-multilingual-MiniLM-L12-v2*), ma per la fase di riduzione dimensionale e clustering sono stati impiegati rispettivamente la Principal Component Analysis (PCA) e l'algoritmo KMeans, al posto di UMAP e HDBSCAN.

I risultati ottenuti con questa configurazione sono stati molto soddisfacenti e sono presentati nella Tabella VI e VII.

TABLE VI: Riepilogo dei topic più frequenti

| Topic | Count  | Name                             |
|-------|--------|----------------------------------|
| -1    | 15,613 | canada_united_states_kherson     |
| 0     | 45,597 | ukraine_russia_the_of            |
| 1     | 3,158  | election_day_2022_midterms       |
| 2     | 2,279  | canada_cdn_poli_canadian         |
| 3     | 492    | canada_newfoundland_shipping_ous |
| 4     | 480    | missile_military_the_us          |
| 5     | 457    | remembrance_forget_let_we        |
| 6     | 440    | raz_hot_time_celeste             |
| 7     | 438    | drone_election_states_united     |
| 8     | 417    | essay_homework_class_assignment  |

TABLE VII: Esempi di tweet rappresentativi per ciascun topic

| Topic | Documento rappresentativo                              |
|-------|--|
| 0     | Why is Russia retreating from Kherson, Canada...       |
| 1     | After order, Russia army to withdraw from Kherson...   |
| 2     | Another Democrat win elections 2022, election...       |
| 3     | AC Air Canada, CDN poli, Canada CDN news, CDN media... |
| 4     | Wedding photographer, weddings journal, notebook...    |
| 5     | The object of war is not to die for your country...    |
| 6     | On this day we honour those who have served...         |
| 7     | It is now time to accept all surrendering Rus...       |
| 8     | Russia Ukraine war drone footage, Ukrainian dr...      |
| 9     | Get online class, homework, assignment and exam...     |

I risultati possono essere interpretati nel seguente modo:

- Il Topic 0 (il più frequente), dominato da termini come ukraine, russia, war e kherson, rappresenta il cuore della narrazione relativa alla guerra, con riferimento diretto alla ritirata russa e all'ingresso delle forze ucraine nella città. I tweet di esempio confermano la presenza di dichiarazioni ufficiali, aggiornamenti strategici e riflessioni sul significato simbolico della liberazione.
- Il Topic 4 integra la narrazione militare con riferimenti a missili, NATO e difesa, e include tweet che esprimono cautela sulla possibilità che la ritirata russa potesse essere una trappola. Questo suggerisce un discorso più tecnico o strategico sulla situazione a Kherson.
- Il Topic 7, invece, si concentra sugli attacchi tramite droni, riportando numerosi tweet che descrivono offensive ucraine con droni modificati contro truppe russe proprio nell'area di Kherson. L'associazione con termini come election day e united states può indicare un'interferenza tra eventi mediatici paralleli, dovuta a co-occorrenza temporale o retweet misti.
- Anche il Topic 6 fa riferimento a Kherson, sebbene in un contesto più frammentato, includendo tweet di opinione, speculazioni su rischi tattici e riflessioni geopolitiche. L'elevata frequenza di termini apparentemente non coerenti potrebbe indicare rumore o co-occorrenze non tematiche, a riprova della complessità del clustering automatico su dati reali e rumorosi come i tweet.
- Il Topic -1, pur includendo Kherson tra le parole chiave, risulta semanticamente più dispersivo e caratterizzato da contenuti generici o promozionali legati a Canada, United States, China, ecc. Questo suggerisce la presenza di tweet non rilevanti oppure marginali rispetto al focus della giornata, raccolti in un cluster residuale.

2) *Emotional Analysis e conclusioni*: I risultati ottenuti dalla successiva analisi emozionale, applicata sia a livello di singoli topic che sull'intero corpus di tweet, sono riportati rispettivamente in Tabella VIII e Tabella IX

TABLE VIII: Distribuzione delle emozioni per topic

| Emozione | Topic -1 | Topic 0 | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 |
|----------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Anger    | 1830     | 8802    | 377     | 299     | 0       | 91      | 0       | 415     | 220     | 0       |
| Fear     | 1645     | 7064    | 191     | 116     | 0       | 72      | 0       | 0       | 149     | 0       |
| Joy      | 3260     | 8218    | 537     | 437     | 250     | 97      | 179     | 1       | 0       | 33      |
| Neutral  | 5601     | 13049   | 1239    | 960     | 168     | 140     | 79      | 0       | 48      | 369     |
| Sadness  | 1927     | 5682    | 289     | 255     | 0       | 47      | 177     | 0       | 0       | 0       |
| Surprise | 1285     | 2488    | 499     | 180     | 47      | 30      | 0       | 0       | 0       | 0       |
| Disgust  | 65       | 294     | 26      | 32      | 0       | 3       | 1       | 0       | 0       | 0       |

TABLE IX: Distribuzione delle emozioni nei tweet analizzati

| Emozione | Numero di tweet |
|----------|-----------------|
| Neutral  | 23402           |
| Joy      | 14482           |
| Anger    | 12745           |
| Fear     | 10144           |
| Sadness  | 9232            |
| Surprise | 5078            |
| Disgust  | 429             |

L'analisi evidenzia come, in questo caso, l'emozione predominante sia quella neutrale, seguita da un'elevata incidenza del sentimento di gioia. Questo risultato consente di rispondere alla terza Research Question, suggerendo che, a fronte di un evento drammaticamente connotato come la guerra, l'episodio della liberazione di Kherson sia stato percepito in modo prevalentemente positivo dall'opinione pubblica sui social network. La presenza significativa di contenuti neutri può essere attribuita alla diffusione di aggiornamenti informativi e dichiarazioni ufficiali, mentre la componente di gioia riflette la valenza simbolica e strategica dell'evento nella narrativa della resistenza ucraina.

## VI. GUERRA ISRAELE-PALESTINA

In questo contesto, il primo studio condotto è stato finalizzato a rispondere alla prima Research Question (RQ): *"Qual è il sentimento degli utenti di Twitter relativo a determinate occorrenze?"*. Le occorrenze in questione sono state individuate attraverso tecniche di Named Entity Recognition (NER) applicate al corpus testuale con l'obiettivo di analizzare come varia il sentimento associato a ciascuna di queste entità nelle conversazioni su Twitter.

Successivamente, l'analisi si è concentrata sul sentimento espresso dagli utenti in relazione a un avvenimento percepito come **"positivo"** con l'intento di rispondere all'ultima RQ: *"Qual è la prevalenza del sentimento degli utenti in questo caso?"*.

### A. Procedimento e discussione sui risultati

Per rispondere alla prima Research Question è stata condotta un'analisi sull'intero dataset, al fine di identificare i principali soggetti menzionati nei tweet e il sentimento ad essi associato.

Le tecniche di preprocessing adottate per i tweet relativi al conflitto israelo-palestinese sono le stesse descritte nella Sottosezione IV-B, con l'obiettivo di rimuovere elementi non

rilevanti per l'analisi, quali menzioni (@), emoticon e collegamenti a risorse esterne.

Anche in questo caso, gli hashtag non sono stati eliminati completamente, in quanto ritenuti informativamente utili. Al contrario, sono stati segmentati in parole di senso compiuto tramite la libreria `wordsegment`, integrata con il supporto linguistico di NLTK, così da conservarne il contenuto semantico all'interno del testo analizzato.

Per il rilevamento delle occorrenze è stato utilizzato il modello preaddestrato *Davlan/bert-base-multilingual-cased-ner-hrl*, descritto in precedenza nella Sezione III-C. Al fine di migliorare l'efficacia del processo di Named Entity Recognition (NER), è stata applicata una metodologia di preprocessing specifica, anch'essa illustrata nella Sezione IV-C, che ha permesso di normalizzare e capitalizzare correttamente le entità più rilevanti legate al conflitto, favorendo un riconoscimento più accurato da parte del modello.

I risultati ottenuti sono visibili in Tabella X

TABLE X: NER Israele-Palestina

| Entità         | Etichetta | Frequenza |
|----------------|-----------|-----------|
| Hamas          | ORG       | 12168     |
| Israel         | LOC       | 5600      |
| Gaza           | LOC       | 3199      |
| Palestine      | LOC       | 1166      |
| IDF            | ORG       | 621       |
| Biden          | PER       | 344       |
| Netanyahu      | PER       | 307       |
| Qatar          | LOC       | 263       |
| Free Palestine | LOC       | 178       |
| Israel         | ORG       | 163       |
| West Bank      | LOC       | 135       |
| Hezbollah      | ORG       | 133       |
| UNRWA          | ORG       | 95        |
| netanyahou     | PER       | 93        |
| Un             | ORG       | 72        |
| Egypt          | LOC       | 68        |
| Jerusalem      | LOC       | 68        |
| Lebanon        | LOC       | 66        |
| Bi             | PER       | 64        |
| United Nations | ORG       | 63        |

L'entità più frequentemente menzionata è "Hamas" (ORG), con oltre 12.000 occorrenze, seguita da "Israel" (LOC), "Gaza" e "Palestine". Questo indica una forte centralità del gruppo militante e delle aree geografiche direttamente coinvolte nel conflitto. La presenza rilevante di termini come

”IDF” (Israeli Defense Forces)<sup>1</sup>, ”Netanyahu”, ”Biden”, e ”Hezbollah” riflette l’attenzione su attori politici e militari chiave, sia regionali che internazionali.

È interessante notare anche la comparsa di entità geopolitiche e istituzionali come ”Qatar”, ”Egypt”, ”UN-RWA<sup>2</sup>”, e ”United Nations”, che suggerisce una narrazione che va oltre lo scontro diretto, includendo dinamiche diplomatiche e umanitarie.

La prima analisi sulle occorrenze per capire quali sono le parole che compaiono più frequentemente nello stesso tweet, non in maniera consecutiva, ottenendo i risultati ottenuti nella Tabella XI

TABLE XI: Co-occorrenze tra le entità più frequenti nel corpus Israele–Palestina

| Count | Entità 1       | Entità 2  |
|-------|----------------|-----------|
| 2720  | Hamas          | Israel    |
| 1694  | Gaza           | Hamas     |
| 1212  | Gaza           | Israel    |
| 463   | Hamas          | IDF       |
| 366   | Israel         | Palestine |
| 364   | Hamas          | Palestine |
| 264   | Gaza           | Palestine |
| 216   | Hamas          | Qatar     |
| 179   | Hamas          | Netanyahu |
| 172   | Gaza           | IDF       |
| 161   | IDF            | Israel    |
| 137   | Biden          | Hamas     |
| 119   | Biden          | Israel    |
| 113   | Israel         | Netanyahu |
| 108   | Biden          | Gaza      |
| 100   | Hamas          | West Bank |
| 94    | Hamas          | Hezbollah |
| 85    | Gaza           | Netanyahu |
| 77    | Free Palestine | Israel    |
| 70    | Free Palestine | Hamas     |

La relazione più frequente è tra Hamas e Israel (2720 occorrenze), seguita da Gaza con Hamas e Israel, a conferma del forte legame semantico tra questi soggetti nei discorsi online. Emergono anche co-occorrenze rilevanti con attori istituzionali o internazionali come IDF, Biden e Qatar, indicando un coinvolgimento percepito non solo locale ma anche geopolitico.

#### B. Sentiment Analysis sulle occorrenze.

L’analisi del sentimento applicata alle occorrenze rilevate, i cui risultati sono riportati in Tabella XII, ha evidenziato una netta predominanza di sentimenti negativi e neutri. Su un totale di 14.854 tweet analizzati, soltanto 62 sono stati classificati come positivi. Questo dato suggerisce che le conversazioni online relative alle entità considerate riflettano principalmente emozioni di preoccupazione, tensione o neutralità informativa, coerentemente con la gravità del contesto trattato. Tra i 62

tweet identificati con sentimento positivo, 20 risultano contenere esplicitamente le occorrenze “Hamas” e “Israel” nel corpus complessivo, senza alcun filtro relativo alla data.

TABLE XII: Distribuzione sentiment tra coppie di entità

| Entity1        | Entity2   | Positive | Neutral | Negative | Total |
|----------------|-----------|----------|---------|----------|-------|
| Hamas          | Israel    | 20       | 911     | 1789     | 2720  |
| Gaza           | Hamas     | 12       | 682     | 1000     | 1694  |
| Gaza           | Israel    | 5        | 456     | 751      | 1212  |
| Hamas          | IDF       | 8        | 116     | 339      | 463   |
| Israel         | Palestine | 3        | 115     | 248      | 366   |
| Hamas          | Palestine | 5        | 97      | 262      | 364   |
| Gaza           | Palestine | 3        | 96      | 165      | 264   |
| Hamas          | Qatar     | 0        | 116     | 100      | 216   |
| Hamas          | Netanyahu | 0        | 62      | 117      | 179   |
| Gaza           | IDF       | 4        | 53      | 115      | 172   |
| IDF            | Israel    | 1        | 45      | 115      | 161   |
| Biden          | Hamas     | 2        | 41      | 94       | 137   |
| Biden          | Israel    | 0        | 34      | 85       | 119   |
| Israel         | Netanyahu | 0        | 49      | 64       | 113   |
| Biden          | Gaza      | 0        | 24      | 84       | 108   |
| Hamas          | West Bank | 1        | 19      | 80       | 100   |
| Hamas          | Hezbollah | 0        | 36      | 58       | 94    |
| Gaza           | Netanyahu | 0        | 28      | 57       | 85    |
| Free Palestine | Israel    | 0        | 26      | 51       | 77    |
| Free Palestine | Hamas     | 0        | 33      | 37       | 70    |

#### C. Analisi occorrenze su subset temporale

Per rispondere alla seconda Research Question, il dataset è stato filtrato considerando il periodo compreso tra il 15 novembre e il 15 dicembre 2023. Questo intervallo temporale è stato contraddistinto da una serie di tregue parziali, durante le quali si sono verificati scambi di ostaggi, una temporanea sospensione dei bombardamenti e l’ingresso di aiuti umanitari nella Striscia di Gaza. La tregua si è ufficialmente conclusa il 1° dicembre, data in cui Israele ha ripreso le operazioni militari, avviando una massiccia offensiva via terra e via aerea nella zona di Khan Younis.

1) *Occorrenze, Bigram e Sentiment Analysis*: Le tecniche di preprocessing utilizzate sono le medesime descritte precedentemente.

In questo caso, le occorrenze sono state estratte utilizzando il modello preaddestrato per la Named Entity Recognition sul dataset filtrato, come riportato in Tabella XIII.

Si osserva che, nonostante una leggera diminuzione nei conteggi assoluti, il profilo delle relazioni semantiche tra entità rimane sostanzialmente invariato. Ciò evidenzia come la struttura del discorso online sia fortemente polarizzata e si concentri stabilmente attorno a pochi nuclei concettuali chiave: **Hamas, Israele, Gaza, IDF** e alcune figure politiche. Questo rafforza la validità dell’analisi condotta e ne conferma la ripetibilità su differenti campioni del dataset.

Successivamente, è stata analizzata la frequenza delle coppie di parole consecutive (bigrammi), i cui risultati sono riportati in Tabella XIV.

<sup>1</sup>Forze di difesa israeliane

<sup>2</sup>E’ un’agenzia delle Nazioni Unite che fornisce assistenza, protezione e servizi essenziali ai rifugiati palestinesi che vivono in Giordania, Libano, Siria, Cisgiordania e Striscia di Gaza



TABLE XIII: Frequenza delle entità nominate nel dataset

| Entity         | Label | Count |
|----------------|-------|-------|
| Hamas          | ORG   | 12026 |
| Israel         | LOC   | 5501  |
| Gaza           | LOC   | 3153  |
| Palestine      | LOC   | 1137  |
| IDF            | ORG   | 616   |
| Biden          | PER   | 343   |
| Netanyahu      | PER   | 303   |
| Qatar          | LOC   | 253   |
| Free Palestine | LOC   | 175   |
| Israel         | ORG   | 163   |
| West Bank      | LOC   | 133   |
| Hezbollah      | ORG   | 128   |
| UNRWA          | ORG   | 93    |
| netanyahou     | PER   | 92    |
| Un             | ORG   | 71    |
| Egypt          | LOC   | 66    |
| Lebanon        | LOC   | 66    |
| Bi             | PER   | 64    |
| Jerusalem      | LOC   | 63    |
| United Nations | ORG   | 6     |

TABLE XIV: Frequenza dei bigrammi di entità

| Entity1   | Entity2   | Count |
|-----------|-----------|-------|
| Israel    | Hamas     | 1510  |
| Hamas     | Israel    | 1399  |
| Hamas     | Gaza      | 909   |
| Gaza      | Hamas     | 773   |
| Israel    | Gaza      | 599   |
| Gaza      | Israel    | 450   |
| IDF       | Hamas     | 245   |
| Hamas     | IDF       | 209   |
| Israel    | Palestine | 170   |
| Palestine | Hamas     | 151   |

L'analisi dei bigrammi conferma che la retorica dominante nei tweet è fortemente binaria e conflittuale, con un lessico orientato a confronti diretti tra i principali attori del conflitto: Hamas, Israel, Gaza e l'IDF. Le frequenze quasi simmetriche dei bigrammi inversi (es. *Hamas-Gaza* vs *Gaza-Hamas*) suggeriscono una rappresentazione bidirezionale e paritetica delle relazioni tra queste entità.

La sentiment analysis condotta sui bigrammi, visibile in Tabella XV, non ha evidenziato cambiamenti significativi nonostante il periodo analizzato fosse caratterizzato da eventi relativamente positivi.

I bigrammi *Israel-Hamas* e *Hamas-Israel* risultano essere i più frequenti, con una netta prevalenza di sentimenti negativi (oltre 990 ciascuno), riflettendo una percezione fortemente polarizzata del conflitto. Allo stesso modo, le combinazioni *Hamas-Gaza* e *Israel-Gaza* mantengono lo stesso schema, con il sentimento negativo che supera nettamente quello neutro e positivo. I sentimenti positivi sono estremamente rari: ad esempio, solo 14 tweet positivi su 1.510 per *Israel-Hamas*. In

alcuni casi, come *Hamas-Qatar* o *Qatar-Hamas*, il sentimento prevalente risulta neutrale, suggerendo una narrazione più descrittiva che emotivamente coinvolta

TABLE XV: Distribuzione del sentiment per bigrammi di entità

| Bigram             | Positive | Neutral | Negative | Total |
|--------------------|----------|---------|----------|-------|
| Israel - Hamas     | 14       | 502     | 994      | 1510  |
| Hamas - Israel     | 5        | 397     | 997      | 1399  |
| Hamas - Gaza       | 7        | 361     | 541      | 909   |
| Gaza - Hamas       | 2        | 290     | 481      | 773   |
| Israel - Gaza      | 1        | 219     | 379      | 599   |
| Gaza - Israel      | 1        | 149     | 300      | 450   |
| IDF - Hamas        | 4        | 76      | 165      | 245   |
| Hamas - IDF        | 2        | 37      | 170      | 209   |
| Israel - Palestine | 1        | 51      | 118      | 170   |
| Palestine - Hamas  | 2        | 31      | 118      | 151   |
| Hamas - Palestine  | 3        | 35      | 95       | 133   |
| Hamas - Qatar      | 0        | 63      | 66       | 129   |
| Palestine - Israel | 1        | 39      | 82       | 122   |
| Palestine - Gaza   | 2        | 36      | 64       | 102   |
| Gaza - Palestine   | 3        | 31      | 64       | 98    |
| Qatar - Hamas      | 0        | 53      | 37       | 90    |
| Netanyahu - Hamas  | 0        | 33      | 53       | 86    |
| Hamas - Hezbollah  | 0        | 28      | 37       | 65    |
| Hamas - Netanyahu  | 0        | 21      | 41       | 62    |
| IDF - Gaza         | 0        | 20      | 41       | 61    |

## VII. CONFRONTO FINALE TRA LE DUE GUERRE

Questo confronto è stato effettuato per poter rispondere alla RQ: *“Quali differenze emergono nell’emozione trasmessa dai tweet relativi ai due conflitti messi a confronto?”*

### A. Procedimento e risultati ottenuti

Per rendere possibile un confronto equo, e in considerazione della significativa differenza di dimensione tra i due dataset, è stato effettuato un campionamento casuale: dal corpus relativo al conflitto russo-ucraino è stato selezionato il 5% dei dati provenienti da dieci dataset distinti, successivamente unificati in un unico insieme. In questo modo, entrambi i dataset risultano composti da un numero equivalente di tweet (14.854). Anche in questo caso, sono state adottate le stesse tecniche di preprocessing precedentemente illustrate per entrambi i dataset (IV-B).

L'analisi delle emozioni è stata condotta mediante il modello preaddestrato *j-hartmann/emotion-english-distilroberta-base*, descritto nella sezione III-B. I risultati ottenuti sono riportati in Tabella XVI. Dall'analisi emergono due differenze particolarmente significative: la distribuzione dei tweet classificati come *joy* e *anger*. Nel dataset relativo al conflitto russo-ucraino, l'emozione della gioia è sensibilmente più presente rispetto al dataset sul conflitto israelo-palestinese. Al contrario, in quest'ultimo, si osserva una netta predominanza di tweet associati alla rabbia. Questo contrasto suggerisce che, mentre il conflitto russo-ucraino lascia spazio a momenti di speranza, resistenza o soddisfazione per eventi percepiti come positivi,



il conflitto israelo-palestinese è percepito online in termini fortemente polarizzati e carichi di tensione emotiva, con una narrazione dominata dalla collera e dall'indignazione.

TABLE XVI: Conteggio delle emozioni nei dataset Russia e Palestina

| Emozione | Russia | Palestina |
|----------|--------|-----------|
| Neutral  | 4479   | 5035      |
| Anger    | 3081   | 4807      |
| Joy      | 2328   | 844       |
| Fear     | 2168   | 2021      |
| Sadness  | 1736   | 1030      |
| Surprise | 942    | 716       |
| Disgust  | 120    | 401       |

## VIII. OVERLAP DI EVENTI, CLUSTERING E CORRELAZIONI (IOT DATA ANALYTICS)

Tra gli eventi presi in esame nel seguente studio figurano due attacchi particolarmente gravi a danno di civili, avvenuti in contesti di guerra differenti ma accomunati dall'elevato impatto umanitario e mediatico.

Nel contesto del conflitto israelo-palestinese, tra il 4 e il 6 dicembre 2023, l'Ufficio delle Nazioni Unite per il Coordinamento degli Affari Umanitari (OCHA) ha segnalato una serie di bombardamenti intensi da parte delle forze israeliane su Gaza, condotti via terra, mare e aria. Gli attacchi hanno colpito in particolare i settori orientali di Gaza City, il campo profughi di Jabalia e l'area a est di Khan Younis. Secondo quanto riportato, decine di civili palestinesi sono stati uccisi o feriti, molti dei quali si trovavano in zone precedentemente designate dai militari israeliani come aree sicure.

Nel caso del conflitto russo-ucraino, l'evento considerato è il bombardamento del 14 gennaio 2023 sulla città di Dnipro, in Ucraina. In quell'occasione, un missile Kh-22 lanciato dalle forze russe ha colpito un grattacielo residenziale, provocando il crollo di un'intera sezione dell'edificio. Il bilancio è stato particolarmente grave: 46 morti (tra cui 6 bambini) e circa 80 feriti, con centinaia di residenti rimasti senza casa. Si è trattato del raid più letale contro edifici civili registrato in Ucraina nei sei mesi precedenti.

Entrambi gli eventi rappresentano casi emblematici di attacchi ad alta intensità contro aree civili densamente popolate, nonostante siano caratterizzati da due periodi temporali completamente diversi.

Lo studio che verrà spiegato successivamente mira a rispondere alla seguente RQ: *"Un evento uguale (bombardamento) produce una risposta discorsiva più coesa nel caso israelo-palestinese rispetto al conflitto russo-ucraino?"*

### A. Procedimento

Per il seguente studio, le tecniche di preprocessing adottate sono le medesime descritte nella sottosezione relativa al preprocessing (IV-B). In particolare, sono state rimosse menzioni, emoticon e punteggiatura, mentre gli hashtag sono stati segmentati e considerati come testo informativo rilevante.

Successivamente, sono stati calcolati gli embeddings utilizzando il modello preaddestrato *sentence-transformers/all-mpnet-base-v2*, precedentemente descritto nella Sezione III-E. I due dataset oggetto dell'analisi sono stati unificati, e ad ogni tweet è stata associata una colonna *Origine*, al fine di consentire un duplice approccio allo studio: da un lato un'analisi complessiva dell'insieme unificato, e dall'altro un confronto differenziato tra i due contesti di origine. In entrambe le situazioni è stata utilizzata la PCA, sugli embeddings precedentemente calcolati, come metodologia per la riduzione della dimensionalità, utilizzando come *n\_components* 2, ottimale per la rappresentazione 2D dei cluster. Successivamente mediante **elbow method** è stato definito il numero ottimale di cluster e infine clusterizzato utilizzando il KMeans. La visualizzazione dei cluster ottenuti è stata fatta come UMAP

1) *Risultati ottenuti da unione dei due dataset*: Considerando l'unione dei due dataset, la clusterizzazione risultante è riportata in Figura 2.

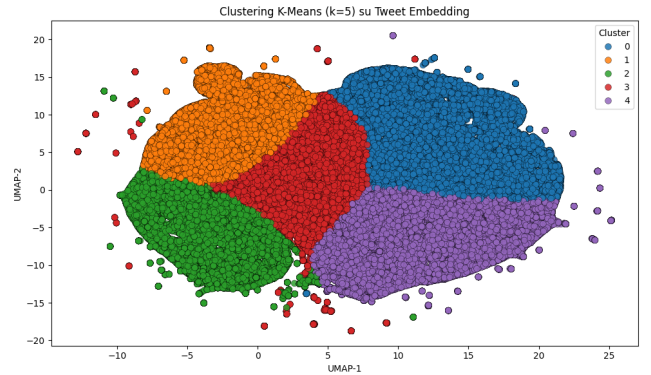


Fig. 2: Cluster su dataset unito

Per ciascun cluster sono state assegnate delle label rappresentative, ottenute mediante l'applicazione del metodo TF-IDF sui termini presenti nei tweet associati. Le parole chiave emerse per ciascun cluster sono le seguenti:

#### TF-IDF su dataset unito

**Cluster 0:** ukraine, war, russia, russian, germany, ukrainian

**Cluster 1:** miss, universe, canada, universe2022, ukraine, germany

**Cluster 2:** irgc, hamas, terrorists, israel, iran, terrorist

**Cluster 3:** ukraine, russia, war, people, russian, putin

**Cluster 4:** russia, ukraine, state, war, terrorist, russian

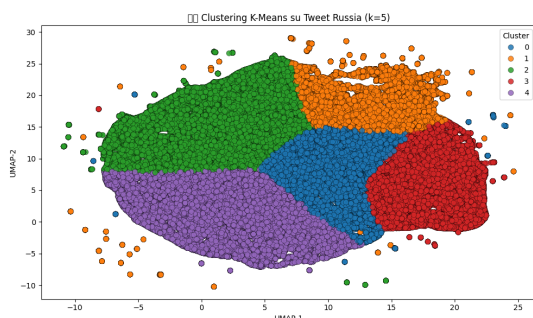
Possiamo osservare come i cluster riflettano chiaramente due narrazioni conflittuali distinte, Russia-Ucraina e Israele-Palestina, con l'aggiunta di un cluster fuori tema, potenzialmente legato a contenuti rumorosi o non pertinenti. Questo risultato suggerisce che la tecnica di clustering abbia identificato correttamente i nuclei concettuali principali presenti nel dataset unificato.

È importante sottolineare che, a causa della marcata differenza di dimensionalità tra i due dataset, solo un cluster è stato associato al conflitto Israele–Palestina, mentre la maggior parte della struttura è dominata dai contenuti relativi al conflitto Russia–Ucraina.

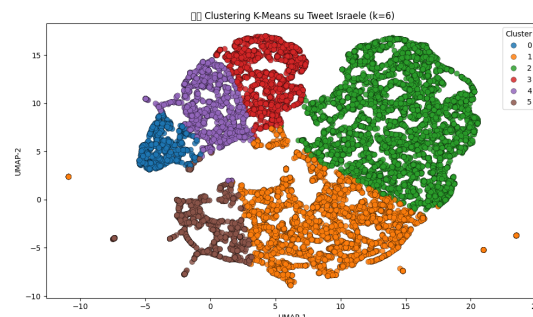
La validità complessiva del clustering è ulteriormente confermata da un valore di Silhouette Score pari a 0.436, che indica una buona separazione tra i cluster e una coerenza interna accettabile nella distribuzione dei dati.

2) *Risultati e discussione studio separato.*: Anche per lo studio dei due avvenimenti in maniera separata è stato seguito lo stesso procedimento descritto in precedenza: è stato utilizzato l’elbow method per determinare il numero ottimale di cluster e successivamente applicato il metodo KMeans per effettuare la clusterizzazione, visualizzata tramite UMAP.

I risultati ottenuti sono riportati in Figura 3, che mostra i cluster distinti per i due conflitti analizzati.



(a) Cluster relativi al conflitto Russia–Ucraina



(b) Cluster relativi al conflitto Israele–Palestina

Fig. 3: Distribuzione dei cluster per ciascun conflitto analizzato separatamente tramite UMAP

Anche in questo caso, ad ogni cluster sono state assegnate delle etichette rappresentative utilizzando TF-IDF:

#### TF-IDF su dataset relativo al conflitto Russia-Ucraina

**Cluster 0:** ukraine, russia, war, russian, slava, nato  
**Cluster 1:** irgc, terrorists, iran, russia, people, terrorist  
**Cluster 2:** russia, ukraine, war, state, terrorist, russian  
**Cluster 3:** miss, universe, canada, universe2022, germany, ukraine  
**Cluster 4:** ukraine, war, russia, russian, germany, bakhmut

#### TF-IDF su dataset relativo al conflitto Israele-Palestina

**Cluster 0:** netanyahou, le, la, et, les, que  
**Cluster 1:** israel, hamas, gaza, palestine, war, hou  
**Cluster 2:** hamas, israel, gaza, war, hostages, ceasefire  
**Cluster 3:** hamas, que, la, en, israel, el  
**Cluster 4:** le, la, les, hamas, et, israel  
**Cluster 5:** war, hou, netanya, biden, terrorists, criminals

L’analisi dei cluster generati separatamente per i due conflitti mostra risultati interessanti e coerenti con le rispettive narrazioni dominanti.

Per quanto riguarda il conflitto Russia–Ucraina, i cluster emersi riflettono tematiche centrali come la guerra in corso, l’identità nazionale, il coinvolgimento della NATO e la città di Bakhmut, frequentemente menzionata nel contesto bellico. È presente anche un cluster legato al concorso di Miss Universo, che rappresenta un contenuto fuori tema, verosimilmente derivante da rumore nei dati. Il valore di Silhouette Score pari a 0.417 indica una discreta coesione interna tra i cluster e una moderata separazione tra le classi.

Nel caso del conflitto Israele–Palestina, i cluster risultano fortemente polarizzati e centrati sui protagonisti chiave del conflitto, come Hamas, Israele, Gaza e concetti associati alla ceasefire e agli ostaggi. Alcuni cluster includono anche termini in francese o spagnolo, suggerendo la presenza di tweet multilingua e un contesto comunicativo globale. Uno dei cluster appare più rumoroso, con termini generici o duplicati, ma nel complesso la separazione tra i temi è ben delineata. Questo è confermato da un Silhouette Score di 0.448, leggermente superiore rispetto al caso ucraino, a indicare una maggiore compattezza interna e distinzione tra i cluster tematici.

3) *Considerazioni finali:* I risultati finali sono visibili nelle tabelle: XVIII, XVII e XIX

Successivamente a questa analisi finale è possibile rispondere alla RQ, posta precedentemente.

I risultati indicano che, a fronte di un evento analogo (bombardamento su civili), la risposta discorsiva osservata nel dataset israelo-palestinese risulta più coesa e polarizzata rispetto a quella del conflitto russo-ucraino. La cohesione intra-cluster calcolata mediante la distanza media degli embedding mostra valori inferiori per i cluster del dataset israelo-palestina (es. mean distance minima = 0.0727) rispetto a quelli del dataset russo (minima = 0.1036), segnalando una maggiore densità

semantica. Inoltre, il Silhouette Score globale risulta maggiore per Israele-Palestina (0.448) rispetto alla Russia-Ucraina (0.417), suggerendo una separabilità più netta tra i cluster. La correlazione di Spearman tra la dimensione dei cluster e la loro distanza media è negativa in entrambi i casi (Israele:  $\rho = -0,60$ , Russia:  $\rho = -0,70$ ), ma la coesione nei cluster israeliani si mantiene più alta nonostante la minore numerosità, confermando un discorso social più compatto e polarizzato nel contesto israelo-palestinese.

#### Russia – Cohesion dei Cluster

TABLE XVII: Cohesion dei cluster nel dataset Russia

| Cluster | Mean Distance | Std Distance | Size  |
|---------|---------------|--------------|-------|
| 0       | 0.123022      | 0.047474     | 19292 |
| 1       | 0.151634      | 0.062137     | 15074 |
| 2       | 0.123687      | 0.055084     | 30141 |
| 3       | 0.112031      | 0.056082     | 21579 |
| 4       | 0.103622      | 0.045840     | 30474 |

#### Israele – Cohesion dei Cluster

TABLE XVIII: Cohesion dei cluster nel dataset Israele

| Cluster | Mean Distance | Std Distance | Size |
|---------|---------------|--------------|------|
| 0       | 0.114408      | 0.056041     | 963  |
| 1       | 0.086950      | 0.042895     | 4002 |
| 2       | 0.072741      | 0.036410     | 5888 |
| 3       | 0.117433      | 0.048711     | 1852 |
| 4       | 0.131334      | 0.052711     | 1492 |
| 5       | 0.119781      | 0.056032     | 1301 |

#### Correlazione di Spearman

TABLE XIX: Correlazione di Spearman tra grandezza del cluster e coesione

| Paese   | Spearman | p-value |
|---------|----------|---------|
| Russia  | -0.700   | 0.188   |
| Israele | -0.600   | 0.208   |

#### IX. LIMITAZIONI DEL LAVORO

Una prima possibile limitazione del presente lavoro riguarda la natura comunicativa tipica dei social network, dove il contenuto testuale è spesso caratterizzato da toni sarcastici, ironici o implicitamente ambigui. Questi elementi stilistici risultano difficilmente riconoscibili da parte dei modelli di sentiment analysis utilizzati, i quali, in diversi casi, hanno assegnato un'emozione o un tono polarizzato non coerente con il significato reale del messaggio.

Una seconda criticità è legata all'asimmetria dimensionale tra i due dataset analizzati, che rende complesso stabilire

un confronto oggettivo e bilanciato tra i due contesti. La maggiore numerosità di uno dei corpora può infatti introdurre effetti di distorsione nell'analisi comparativa, influenzando la distribuzione dei topic, dei sentimenti e delle emozioni emerse.

#### X. CONCLUSIONI

L'implementazione del progetto ha permesso di rispondere alle diverse Research Questions inizialmente formulate, offrendo una panoramica comparativa tra i due conflitti oggetto di studio. I risultati evidenziano come le guerre analizzate rappresentino, attualmente, un terreno privilegiato di dibattito e narrazione all'interno dei social network, dove opinioni, emozioni e prese di posizione si diffondono rapidamente a livello globale.

Lo studio ha confermato l'efficacia di BERTopic nell'individuare correttamente gli argomenti più discussi in giornate specifiche, specialmente in corrispondenza di eventi connotati negativamente. L'analisi delle emozioni associate ai topic ha mostrato una predominanza di rabbia e neutralità, quest'ultima probabilmente legata alla diffusione di informazioni senza un chiaro coinvolgimento emotivo. Al contrario, durante un periodo considerato relativamente positivo, i tweet si sono caratterizzati principalmente per un tono neutro accompagnato da una maggiore presenza di emozioni positive, come la gioia.

Nel caso del conflitto israelo-palestinese, i contenuti analizzati mostrano una tendenza emotiva più negativa rispetto a quelli riferiti alla guerra russo-ucraina. In riferimento all'ultima domanda di ricerca, l'analisi dei cluster semantici suggerisce che, a parità di evento (bombardamento), la discussione su Twitter risulta più coesa e verosimilmente più polarizzata nel contesto israelo-palestinese. Questa osservazione si basa sui valori di distanza media inferiori registrati nei cluster relativi a Israele, che indicano una maggiore coesione semantica tra i tweet. Sebbene nel dataset russo emerga una correlazione negativa più marcata tra la dimensione del cluster e la sua coesione ( $\rho = -0,700$ ), essa non risulta statisticamente significativa, così come nel caso di Israele ( $\rho = -0,600$ ). Tuttavia, la distribuzione dei cluster israeliani appare nel complesso più compatta e meno dispersa, suggerendo una narrazione più unificata e coerente.

#### REFERENCES

- [1] R. Nazarchuk and S. Albota, "Tweets about ukraine during the russian-ukrainian war: Quantitative characteristics and sentiment analysis." in *MoMLeT+ DS*, 2023, pp. 551–560.
- [2] L. Ramos, "Sentiment analysis of russia-ukraine conflict tweets using roberta," *Uniciencia*, vol. 37, 06 2023.
- [3] D. Wang, S. Yue, Y. Wen, K. Wu, T. Zhong, M. Chen, Z. Yu, L. Yuan, and G. Lü, "Unveiling the spatiotemporal propagation patterns of sentiments regarding the israeli-palestinian military conflict," *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1–17, 2025.
- [4] J. Hartmann, "Emotion english distilroberta-base," <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/>, 2022.
- [5] Davlan, "bert-base-multilingual-cased-ner-hrl," <https://huggingface.co/Davlan/bert-base-multilingual-cased-ner-hrl/>.

- [6] F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, and L. Neves, "TweetEval: Unified benchmark and comparative evaluation for tweet classification," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1644–1650. [Online]. Available: <https://aclanthology.org/2020.findings-emnlp.148>
- [7] "sentence-transformers/all-mpnet-base-v2," <https://huggingface.co/sentence-transformers/all-mpnet-base-v2/>.
- [8] "Ukraine conflict twitter dataset," <https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows/>.
- [9] "Israel-palestine conflict tweets dataset," <https://www.kaggle.com/datasets/mehyarmlaweh/israel-palestine-conflict-tweets-dataset/>.
- [10] "Avvenimenti conflitto 28 maggio 2023," [https://www.lastampa.it/esteri/2023/05/28/diretta/russia\\_u\\_craina\\_guerra\\_newsoggi-12828830/](https://www.lastampa.it/esteri/2023/05/28/diretta/russia_u_craina_guerra_newsoggi-12828830/).