

Secondo Report sull'analisi della violenza contro le donne veicolata dai social media. Periodo 1 Dicembre 2022 – 31 Agosto 2024

Dal 2020 L'Istituto Nazionale di Statistica, in collaborazione con il Dipartimento per le Pari Opportunità (DPO), ha iniziato ad analizzare il fenomeno della violenza di genere nell'ambito dei *social media*, al fine di osservare come questo fenomeno viene rappresentato e analizzare come gli stereotipi di genere sono veicolati nello spazio virtuale.

In particolare, è stato promosso un progetto di analisi del *sentiment*¹ ed *emotion*² sulle interazioni generate dai *social media* (*Twitter* - X, pagine pubbliche di *Instagram* e *Facebook*, *Webnews*) volto a osservare come i *social* producano e/o riproducano stereotipi di genere, amplifichino il linguaggio violento oppure generino indignazione e quali nuove forme di violenza di genere possono generarsi *online* (come ad esempio la cyber-violenza).

In questo report si presentano i dati raccolti nel periodo compreso tra il 1° Dicembre 2022 e il 31 Agosto 2024, con i quali si aggiornano le informazioni già rilasciate nel primo report (<https://www.istat.it/statistiche-per-temi/focus/violenza-sulle-donne/il-fenomeno/stereotipi/>).

La violenza sulle donne raccontata dai *social*

L'attenzione alla violenza sulle donne facilitata dalla tecnologia costituisce un tema centrale nel dibattito internazionale. UN WOMEN definisce la violenza facilitata dalla tecnologia (Technology Facilitated Gender Based Violence -TFGBV) come “*any act, that is committed, assisted, aggravated or amplified by the use of ICTs or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms*”³. La violenza online trova le stesse origini culturali e sociali della violenza contro le donne off-line. La TFGBV è radicata e resa possibile da norme discriminatorie di genere che si intersecano con altre forme di discriminazione basate su razza, etnia, identità di genere, orientamento sessuale e abilità, come altri fattori intersezionali che vanno considerati quanto si osserva questo nuovo fenomeno connesso alla diffusione della comunicazione online e all'uso dei *social media*. Diversi sono i metodi per studiare l'incidenza del fenomeno delle molestie online (in primis [l'indagine sulla sicurezza dei cittadini](#)); osservare come questo tema viene trattato nei *social* costituisce una fonte di informazione rilevante. L'Istat ha continuato a monitorare l'attività

¹ La *Sentiment Analysis* (SA), definita come lo studio delle opinioni e dei sentimenti espressi da dati testuali (Liu, 2012), è una tecnica in rapida crescita nell'ambito della ricerca del NLP, grazie anche all'ampia gamma di applicazioni effettuate in diversi campi di analisi. Le tecniche di SA vanno da regole e metodi relativamente semplici a procedure avanzate di *deep learning* (per una rassegna dettagliata si veda: Liu 2012; Medhat, Hassan e Korashy 2014). La *sentiment analysis* (anche conosciuta come *opinion mining*) è lo studio computazionale delle opinioni, delle valutazioni, degli atteggiamenti e delle emozioni delle persone nei confronti di entità, individui, questioni, eventi, argomenti e dei loro attributi (Liu, B., Zhang, L. (2012). *A Survey of Opinion Mining and Sentiment Analysis*. In: Aggarwal, C., Zhai, C. (eds) *Mining Text Data*. Springer, Boston, MA, (pag. 415. Liu, B., Zhang, L. (2012). *A Survey of Opinion Mining and Sentiment Analysis*. In: Aggarwal, C., Zhai, C. (eds) *Mining Text Data*. Springer, Boston, MA, (pag. 415). I risultati dello studio descrivono la prevalenza di sentiment positivo, negativo e neutro. Si veda a questo proposito la nota metodologica.

² L'*emotion detection*, similmente alla *sentiment analysis*, di cui condivide metodologie e tecniche di analisi, descrive la prevalenza di 6 emozioni (Amore, Gioia, Neutro, Tristezza, Paura, Rabbia), con cui le persone si esprimono sui *social*. Sulla scelta delle emozioni e sul processo di estrazione delle informazioni dai *social* si veda la nota metodologica.

³ UN Women, 2023, Technology Facilitated Violence against Women-Report of the Foundational Meeting of the Expert Group

dei *social*, osservando come i temi della violenza sulle donne, degli stereotipi e l'immagine sociale che si sviluppa ne esalti gli aspetti di indignazione oppure se esiste un'amplificazione del linguaggio violento che rafforza i processi di vittimizzazione.

L'analisi dei picchi registrati nel tempo consente di individuare quali temi generino maggiormente le discussioni e quali siano le modalità con cui gli utenti dei *social* reagiscono alla violenza contro le donne e/o generano discussioni intorno ad essa. Il metodo adottato aiuta a capire quali sono i messaggi che scatenano la discussione collettiva, fornendo una mappa degli argomenti (come ad esempio il *body-shaming*, il femminicidio e lo stupro), che sono rilevati attraverso la *sentiment* e l'*emotion analysis* al fine di osservare le polarità (positive, neutre e negative) e lo spettro delle emozioni associate. Come meglio spiegato nella nota metodologica, il metodo sperimentato può aiutare anche a monitorare le nuove forme di violenza *online* contro le donne e ad aprire una nuova strada volta a integrare diverse fonti di dati (Van den Brakel J., Söhler E., Daas P., Buelens B., 2017; Daas P., Puts M., 2014), arricchendo in tal modo il quadro informativo predisposto dall'Istat in costante aggiornamento sulla [Violenza sulle donne](#).

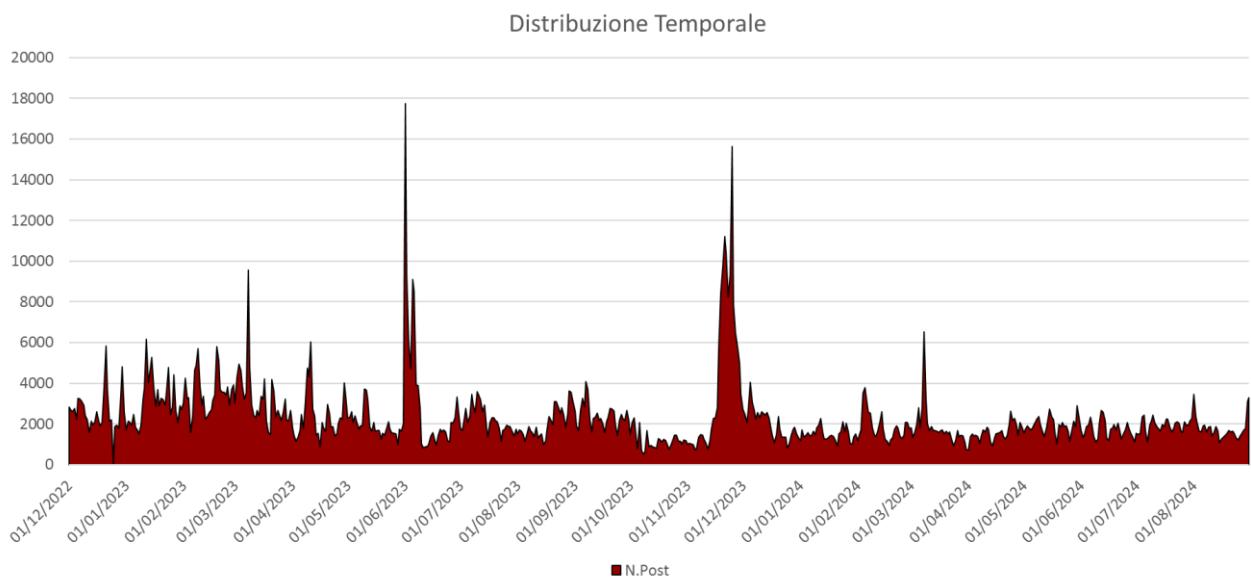
Analizzare il modo in cui la violenza di genere e gli stereotipi di genere vengono rappresentati nei *social* costituisce una nuova frontiera di studio del fenomeno. Occorre infatti osservare come muta la percezione stessa e la narrativa della violenza sulle donne nella realtà virtuali. Attraverso un processo di cattura dei contenuti veicolati dai *social*, sulla base della presenza di almeno una parola appartenente ad un insieme di parole filtro, (predisposto da esperti di dominio), è possibile raccogliere ed osservare le opinioni e i contenuti di "post". Inoltre, l'analisi consente di osservare come tali messaggi siano utilizzati per contrastare, condannare o isolare la cultura dello stereotipo e della violenza di genere, oppure se, al contrario, essi ne esaltino, come detto, il lato negativo, amplificando la portata di odio e di violenza. I dati riportati consentono quindi di:

1. Osservare i volumi della produzione sui *social* cogliendone gli andamenti nel tempo. Sulla base di questo dato è possibile individuare gli eventi/temi che hanno particolarmente caratterizzato la produzione di messaggi *social* sul tema. Tipicamente si tratta di eventi commemorativi come la Giornata internazionale per l'eliminazione della violenza contro le donne del 25 novembre, ma i picchi si registrano anche in coincidenza di fatti di cronaca o legati a provvedimenti giuridici o discussioni animate da eventi di varia natura (sportivi, culturali, ecc.) che generano reazioni contrastanti tra gli utenti.
2. Rappresentare come tali conversazioni utilizzino un linguaggio violento o di indignazione nei confronti dei temi scatenanti la discussione.
3. Analizzare le emozioni connesse agli atteggiamenti di indignazione e di violenza espressi.
4. Effettuare una *topic analysis* sui messaggi per evidenziare quali siano gli argomenti trattati nell'ambiente virtuale.

I dati dal 31 dicembre 2022 al 31 agosto 2024 hanno riguardato un totale di **1.467.035** **contenuti** sulla violenza di genere: si tratta di cui **684.504** prodotti da X (*twitter*), seguiti da **262.525** *Facebook* e **131.252** *Instagram*. Da segnalare anche i **340.264** prodotti dal *WEB* e

48.104 da rassegna stampa, che, tuttavia non sono oggetto di analisi del *sentiment* e dell'*emotion*, perché sostanzialmente con contenuti informativi. I *tweets* di X assorbono circa la metà delle conversazioni monitorate (46,6%).

Fig. 1 – Andamento temporale dei post generati sui canali social (X, Facebook, Instagram, pagine web e rassegne stampa). Periodo di riferimento 1 dicembre 2022, 31 agosto 2024).
Valori assoluti



Fonte: Indagine Istat, 2024

La metodologia adottata consente di controllare se i messaggi sono di natura prettamente informativa e giornalistica oppure se siano solo espressione di opinioni/giudizi. In questo caso, attraverso l'adozione di un classificatore binario si possono identificare due tipi molto diversi di contenuto e di realizzare l'analisi solo su *post* non strettamente informativi (si veda nota metodologica allegata) al fine di rilevare con maggior evidenza quale sia il sentimento che emerge dai *social* su determinati *topic* (temi).

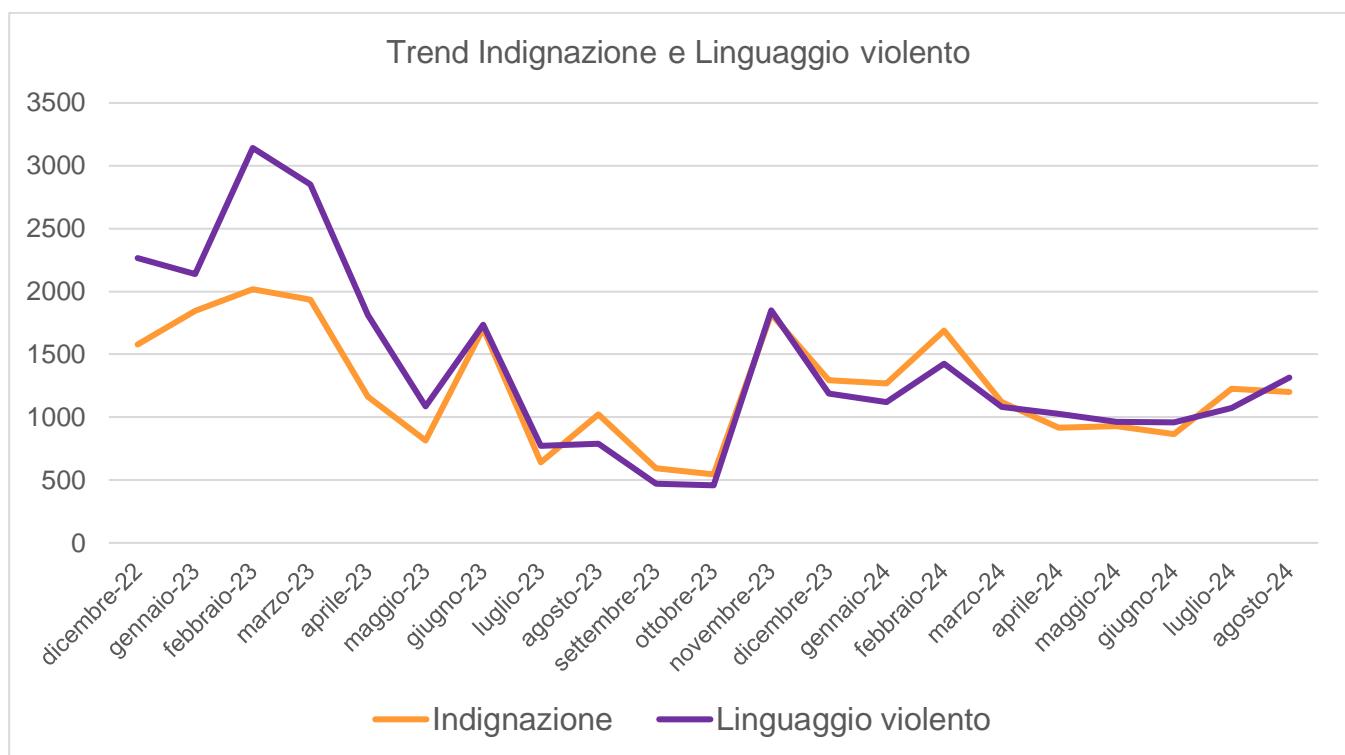
Il focus di seguito presentato si concentra sul modo con cui “si parla” *online*: esso è stato effettuato sulle reazioni di odio, di aggressività e di violenza che gli eventi e/o fatti di cronaca generano nelle conversazioni, e sul senso di indignazione che gli stessi eventi provocano tra gli *user* dei *social* e del *web*. Tale focus è stato realizzato attraverso la creazione di un *dataset* selezionato di contenuti “non informativi” estratto sulla base di un sotto insieme di *keyword* volte a cogliere le opposte categorie di odio e indignazione.

Va specificato che il totale dei contenuti informativi è pari a **906.422** che rappresenta il 61,8% dei contenuti totali. Per la maggior parte si tratta di post di X dove il contenuto informativo è maggiormente presente.

Considerando il sotto-insieme dei contenuti non-informativi (pari a **560.613**), la figura 2 riporta l’andamento, nel periodo considerato, di entrambe le reazioni ed evidenzia una leggera predominanza dei contenuti violenti: in termini quantitativi sono **23.928** i contenuti di indignazione contro **27.292** contenenti linguaggi violenti. Questo aspetto appare in

leggera controtendenza rispetto a quanto rilevato nel precedente report (che osservava il periodo dal 1° novembre 2021 al 30 novembre 2022⁴), dove invece si è registrata la presenza, tra gli utenti *social*, di una elevata consapevolezza nel voler contrastare la violenza basata sul genere e sugli stereotipi, attraverso contenuti di indignazione. Se l'indignazione, la consapevolezza emergono soprattutto in corrispondenza di eventi che attirano maggiormente la reazione dei *social* (come, ad esempio, il 25 novembre, oppure in caso di femminicidi efferati come quello che ha scatenato il picco del 25 giugno 2023), altri eventi collegati a donne che prendono parola nella sfera pubblica (ad esempio interventi in occasioni di eventi televisivi) sono connotati da contenuti di violenza online. I dati dunque evidenziano che, quanto più le figure femminili note intervengono sul tema della violenza sulle donne, tanto più esse sono soggette ad attacchi d'odio.

Figura 2 – Andamento giornaliero sui social del linguaggio violento e di indignazione sui temi di violenza di genere (Periodo 1 dicembre 2022 – 31 agosto 2024). Valori assoluti



Fonte: Indagine Istat, 2024

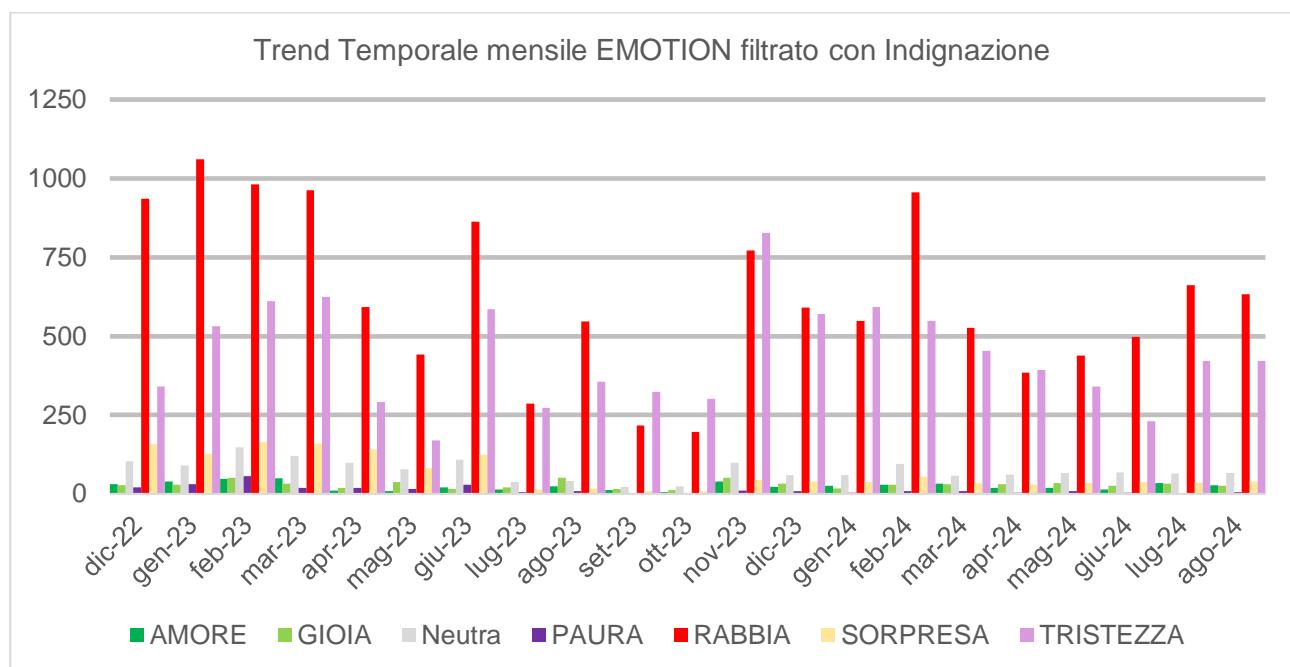
Il linguaggio violento, evidenzia quanto nel dibattito pubblico sia ancora molto diffusa una cultura volta a rafforzare lo stereotipo di genere, che costituisce la radice socioculturale della violenza contro le donne.

La metodologia adottata consente di andare ancora più in profondità nell'analisi per capire che tipo di emozioni si associano alle reazioni di violenza/odio e quali a quelli di indignazione in concomitanza dei picchi (eventi/episodi). Le Figure 3 e 4 riportano tali emozioni

⁴ <https://www.istat.it/statistiche-per-temi/focus/violenza-sulle-donne/il-fenomeno/stereotipi/>

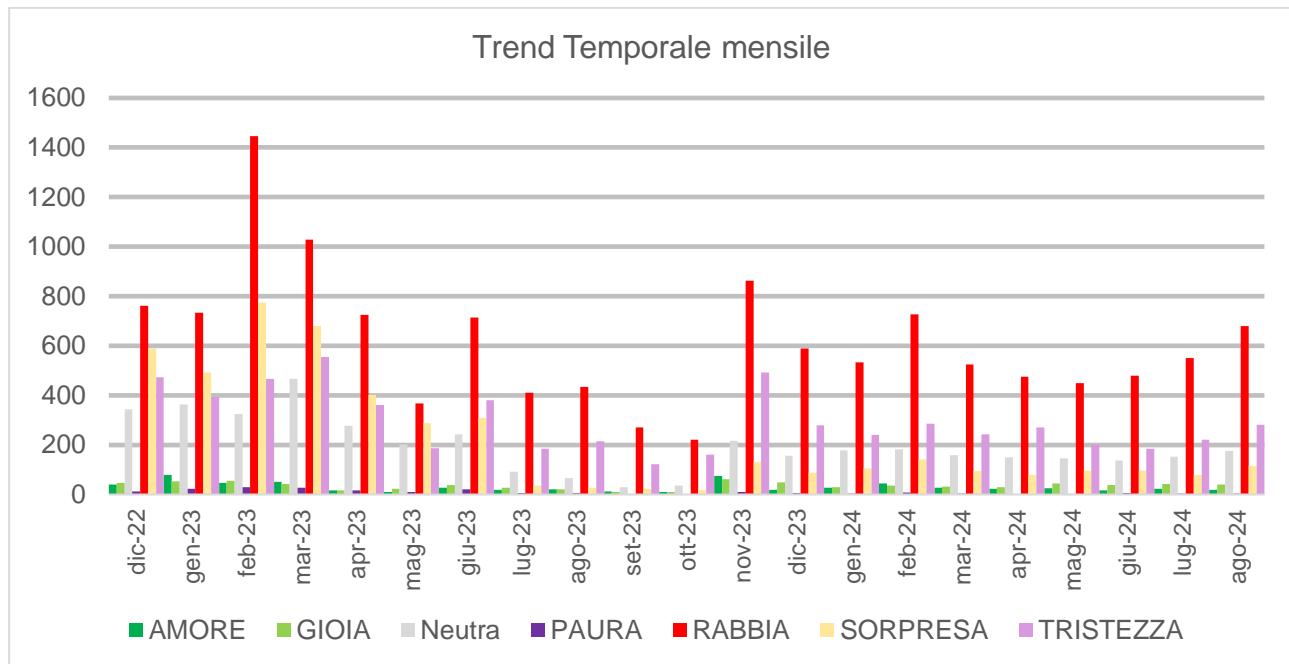
rispettivamente per il linguaggio violento e per l'indignazione. Come è evidente sono soprattutto la rabbia e la sorpresa ad animare vistosamente il tenore delle discussioni violente, mentre è ancora la rabbia ma questa volta accompagnata dalla tristezza a caratterizzare i contenuti di indignazione.

Figura 3 - Emotion detection registrata sui messaggi con reazioni di violenza (Periodo 1 dicembre 2022 – 31 agosto 2024). Valori assoluti sul numero di post totali di indignazione



Fonte: Indagine Istat, 2024

Figura 3 - Emotion detection registrata sui messaggi con reazioni di indignazione (Periodo 1 dicembre 2022 – 31 agosto 2024). Valori assoluti sul numero di post totali di linguaggio violento.



Fonte: Indagine Istat, 2024

I risultati della *topic analysis*

Accanto al trend temporale dei contenuti *social*, ISTAT ha condotto un'analisi dei *topic* utilizzando la metodologia del *topic modelling* (Vayansky et al., 2020; Kherwa et al., 2019; Blei et al., 2009). Si tratta di un modello statistico utilizzato per l'analisi di temi (o argomenti) nascosti (o latenti) in un insieme di documenti. Si tratta di un algoritmo di apprendimento non supervisionato, che permette di scoprire in modo automatico i temi principali presenti in un corpus o collezione di testi. Nello specifico è un modello probabilistico generativo che tramite una distribuzione probabilistica considera ogni documento come una combinazione di temi (o *topic*), dove ogni tema è a sua volta una distribuzione di probabilità su parole. Il modello assume che ci siano K argomenti (il cui numero deve essere predefinito) e che ogni documento sia generato da una mistura di questi argomenti.

Scorrendo i grafici interattivi proposti per i tre canali *social* analizzati nello stesso periodo di riferimento (pubblicati sulla pagina “[utilizzo dei social](#)”), si osserva che, sebbene con un approccio analitico diverso (il metodo non supervisionato), i risultati sui temi emergenti sono fortemente coerenti con quelli rilevati attraverso il modello supervisionato (per una spiegazione dettagliata si veda la nota metodologica allegata).

Inoltre tale analisi ci consente di osservare come talune tematiche trovino spazi di discussioni diversi per canale e non sono influenzati dalla incidenza dei contenuti di un canale sull'altro.

Così ad esempio sulla piattaforma **X** si ripropongono temi connessi ad eventi che hanno richiamato la reazione collettiva sui casi di femminicidio registrati in Italia nel corso del 2023 e 2024. I *topic* riportano parole come insegnare, patriarcato, misoginia, domestica, a sottolineare come il dibattito si concentri sulla necessità di intervenire sul piano culturale.

Su *Facebook* il tema della denuncia è particolarmente evidente nel *topic* 4, ma altre occorenze di parole mostrano l'incidenza avuta da altri eventi sul tema della violenza sulle donne (come la guerra israelo-palestinese e l'evento sportivo delle Olimpiadi).

Su *Instagram* è doveroso ricordare che sono gli *hashtag* ad essere catturati accanto alle parole chiave. Questo comporta anche il richiamo di commenti e reazioni in lingua straniera (vista la presenza di parole-chiave in inglese come *#bodyshaming*, *#stalking* ecc.). Si è deciso di mantenere quindi nei *topic* le parole in inglese, escludendo quelle in altre lingue straniere. Dall'analisi emerge una maggiore attenzione ai fenomeni di cronaca nera: il femminicidio (*topic* 1), il tema della violenza psicologica (*topic* 4), il *bodyshaming* (*topic* 5), i casi di discriminazione sportiva (*topic* 5).