

Barometro dell'odio Elezioni europee 2019



Nota metodologica

Contesto operativo

Quanto e come i candidati al Parlamento europeo parlano di diritti? Quanto e come ricorrono al linguaggio dell'odio nel farlo? Quali sono le reazioni degli utenti del web? Amnesty International Italia monitora per **40 giorni (dal 15 aprile fino alla conclusione della campagna elettorale, il 24 maggio)** i profili **Facebook e Twitter dei candidati al Parlamento europeo** più attivi online e dei **leader di partito** ai quali fanno riferimento. Osserverà, inoltre, le **reazioni e risposte degli utenti**, per rilevare le eventuali correlazioni tra toni e messaggi veicolati dalla politica e sentimento delle persone rispetto a determinati temi.

Il monitoraggio dell'*hate speech* online realizzato nell'ambito del "**Barometro dell'odio - Elezioni europee 2019**" per le elezioni europee 2019 completa un percorso di analisi del fenomeno del discorso d'odio avviato nel 2018, in occasione delle elezioni politiche.

Con obiettivi diversi e modalità aggiornate abbiamo deciso di proseguire il lavoro di osservazione avviato per monitorare il dibattito sui social media durante la campagna elettorale che conduce all'appuntamento alle urne del 26 maggio 2019. Nel clima di costante campagna elettorale che caratterizza il dibattito politico in Italia, infatti, l'*hate speech* è costantemente diffuso e raggiunge picchi di intensità in prossimità degli appuntamenti elettorali.

La **procedura di monitoraggio** dei social media sviluppata nell'ambito del progetto "Barometro dell'odio – Elezioni europee 2019" si compone di diverse fasi, descritte nelle pagine che seguono: la raccolta dei contenuti, la campionatura, la visualizzazione e valutazione, i controlli incrociati, l'analisi.

Raccolta dei dati

I contenuti sono raccolti da Twitter e Facebook per mezzo di algoritmi.

La lista di politici monitorati elaborata da Amnesty International Italia include i **candidati delle principali liste** (Europa Verde, Forza Italia Fratelli d'Italia, Lega, Movimento 5 Stelle, Più Europa, Sinistra Insieme) ai seggi italiani del Parlamento europeo; l'elenco include altresì i leader di partito non in corsa ai quali i candidati monitorati fanno riferimento.

I contenuti raccolti sono divisi in due macro-categorie: **tweet/post dei politici e relative/i risposte/commenti degli utenti generici**.

Risposte/commenti degli utenti generici includono anche le repliche alle stesse/i risposte/commenti. Per consentire l'osservazione delle interazioni tra utenti generici, infatti, è rilevato - nel caso di Facebook - fino al quarto livello di risposta (es.: la replica alla replica al commento a un post).

Le due tipologie di contenuto (tweet/post dei politici e relative/i risposte/commenti degli utenti generici) subiscono un trattamento diverso nel processo di monitoraggio.

Tweet/post dei politici sono raccolti a partire **dal 15 aprile 2019 fino al 24 maggio 2019**. Risposte/commenti degli utenti generici sono raccolti a partire dal 15 aprile 2019 fino al 31 maggio 2019, a una settimana dalla chiusura della raccolta dei tweet/post dei politici, per garantire che la maggior parte delle risposte/commenti a tali contenuti venga raccolta.

Campionamento

Sulla base della lista dei candidati al Parlamento europeo, sono valutati **i feed raccolti dai 40 politici più attivi [1] su Twitter, dai 40 più attivi su Facebook e dai 40 complessivamente più attivi**. Per assicurare un'equa rappresentazione degli schieramenti e del territorio, sono valutati i feed raccolti dai 4 politici complessivamente più attivi per ogni lista e il feed di almeno un rappresentante di ogni circoscrizione per lista. Per ogni lista valuteremo il feed raccolto da almeno una donna e da almeno un uomo.

Nel caso dei dati preliminari rilasciati prima del 15 maggio, valutiamo i feed dei 40 politici più attivi nelle due settimane che precedono la data di disseminazione. Per quanto riguarda i dati rilasciati il 10 maggio, quindi, i 40 politici più attivi sono determinati sulla base dell'attività registrata a partire dal 24 aprile.

Una volta che la lista definitiva dei politici è determinata tutti i tweet/post di quei politici sono valutati. Poiché il volume delle risposte/commenti ai tweet/post è ampio, raccoglieremo un **campione di risposte/commenti in modo causale**. L'obiettivo è determinare la qualità delle risposte/commenti sulla base della qualità dei tweet/post, rilevando l'eventuale correlazione. In considerazione del fatto che è difficile determinare in anticipo tali elementi, raccoglieremo un campione di risposte/commenti da ogni tweet/post (dove disponibili) per raggiungere una media di 1.000 risposte/commenti valutati per politico; tale tetto sale a 2.000-3.000 per i politici che ricevono oltre 100.000 risposte/commenti.

[1] L'attività sui social media è definita come la semplice somma del numero di tweet/post pubblicati dal politico con il numero di risposte/commenti ricevuti. Monitoreremo l'attività sui social media per determinare la lista finale dei politici i cui feed sono valutati per un mese, fino al 15 maggio (incluso).

Cercheremo, inoltre, di raccogliere almeno 4 risposte/commenti per post (quando disponibili); ciò significa, nel caso di un numero molto ristretto di politici che registrano un'attività particolarmente intensa, che vi sarà un numero di risposte/commenti da valutare che potrà superare significativamente i 4.000.

Il campione di risposte/commenti raccolti con questo metodo è disponibile per la valutazione non è rappresentativo delle reali proporzioni ed è, per tanto, sottoposto ad appropriata ponderazione nella fase finale di analisi.

L'applicazione dei coefficienti di ponderazione (pesi) avviene attraverso l'assegnazione per politico e per tweet/post.

Visualizzazione e valutazione dei contenuti

Circa **150 attivisti** di Amnesty International Italia sono coinvolti nella fase di valutazione dei contenuti.

Tutti i contenuti da sottoporre a valutazione sono automaticamente "impacchettati" in file da circa 50 contenuti (tra tweet/post e risposte/commenti). Questi pacchetti sono caricati in modo automatico all'interno di un'applicazione che consiste in un'interfaccia per la valutazione.

I campi di valutazione in dettaglio

Per ogni contenuto sono indicati dal valutatore: **tema** (donne, lgbti, disabilità, migranti rifugiati e persone con background migratorio, rom, minoranze religiose, solidarietà, povertà socio-economica, altro); **accezione** (negativa, positiva); se negativa la **tipologia** (non problematico, problematico, hate speech, ambiguo); se problematico o hate speech il **target** (il politico autore del contenuto, un altro politico, l'autore del commento/risposta precedente, un singolo individuo o un gruppo perché riconducibile a una categoria soggetta a discriminazione, altro); categoria del target (donne, lgbti, persone con disabilità, migranti rifugiati e persone con background migratorio, rom, musulmani, ebrei, un singolo o un gruppo per lo svolgimento di attività di tipo umanitario e/o solidaristico, persone in condizione di povertà socio-economica, non riconducibile ad alcuna categoria/altro).

Per la **definizione di hate speech** ci atteniamo a quella contenuta nella *Raccomandazione di politica generale n.15 dell'ECRI relativa alla lotta contro il discorso d'odio* (adottata l'8 dicembre 2015).

L'interfaccia mostra le seguenti informazioni all'attivista-valutatore:

- il contenuto;
- il nome del politico dal cui feed proviene il contenuto;
- la distinzione tra tweet/post e risposta/commento;
- nel caso di risposta/commento, è mostrato il tweet/post in replica al quale il contenuto è stato pubblicato.

Acquisite queste informazioni l'attivista valuta il contenuto, spuntando le opzioni contenute negli appositi campi. I campi di valutazione sono (in dettaglio nel box alla pagina precedente): il tema; l'accezione; in caso di accezione negativa la distinzione tra contenuto non problematico/problematico/hate speech; in caso di contenuto problematico/hate speech la tipologia di target; il gruppo sociale al quale il target è riconducibile (se presente).

Cross-checking

Poiché può essere difficile per i valutatori essere perfettamente allineati rispetto all'individuazione del livello di offensività di un contenuto, per generare valutazioni coerenti e convergenti è stato elaborato un meccanismo che consente di tenere sotto controllo il margine di incoerenza/divergenza.

Tutti i contenuti sono valutati da 3 diversi valutatori, selezionati in modo casuale. I dati relativi ai contenuti per i quali tutte e 3 le valutazioni sono allineate sono accettati come definitivi.

Quando non vi è allineamento, il contenuto è sottoposto all'ulteriore valutazione (definitiva) di un gruppo di esperti.

Analisi quantitativa

L'analisi quantitativa (dato complessivo e per singolo politico) sarà focalizzata, tra gli altri, sui seguenti punti:

- eventuale correlazione tra tono del tweet/post del politico e risposte/commenti problematici/hate speech ricevuti dagli utenti generici;

l'ipotesi è che i tweet/post negativi scatenino una maggiore quantità di commenti negativi (indicazione di questa tendenza ci è già arrivata nel monitoraggio pilota [2])

- eventuale correlazione tra tema e quantità di contenuti negativi/problematici/hate speech; l'ipotesi è che alcuni temi scatenino contenuti problematici/hate speech più di altri, a prescindere dai toni utilizzati dal politico;
- identificazione delle categorie prese di mira con maggiore frequenza dai tweet/post negativi dei politici e nelle risposte/commenti degli utenti generici;
- in che modo si inquadrano i temi sui quali ci soffermiamo all'interno del frame narrativo "Europa".

Come menzionato, il campionamento - così come strutturato nell'ambito del "Barometro dell'odio - Elezioni europee 2019" - è finalizzato a garantire un data set che includa sufficienti esempi di ogni tipologia di tweet/post e risposte/commenti a ognuno di questi tweet/post (dove disponibili). Il dataset così ottenuto, tuttavia, non riflette le reali proporzioni dei contenuti sui social media monitorati e, per tanto, deve essere sottoposto a ponderazione. La ponderazione è basata sullo schema descritto dal seguente esempio.

Privacy

- **Nome e ID dell'utente** - Tali dati sono raccolti attraverso le API (Application Programming Interface) di Twitter e Facebook. Nel caso di Facebook il nome e l'ID dell'utente generico non sono in alcun modo raccolti (è la stessa API a non consentirlo). Ciò non vale per le pagine pubbliche dei politici monitorati. Nel caso di Twitter, esso consente di ottenere nome e ID dell'utente generico, i quali tuttavia sono rimossi dall'algoritmo che utilizziamo e non vengono, dunque, salvati (sono salvati, invece, quelli dei politici monitorati).
- **ID del contenuto** - Nel caso di Facebook è necessario, solo nella fase iniziale, conservare l'ID del contenuto raccolto, al fine di individuare eventuali risposte/commenti attraverso la Facebook Graph API. Una volta individuate/i queste/i ultime/i, l'algoritmo rimuove l'ID del contenuto. Nel caso di Twitter l'ID del contenuto è rimosso immediatamente dall'algoritmo e non salvato (sono salvati, invece, quelli dei politici monitorati).

I **dati grezzi** (prima della valutazione) sono salvati su una piattaforma alla quale si accede solo tramite autorizzazione, che è ristretta agli sviluppatori. Ogni dato contenente l'ID di un contenuto è cancellato entro una settimana dal termine del progetto.

I **dati già sottoposti a valutazione** sono archiviati sulla stessa piattaforma (tutti gli elementi che possono ricondurre all'identità dell'utente generico - nome e ID - sono rimossi prima dell'archiviazione; sono mantenuti nel caso dei politici monitorati), con le stesse restrizioni relative all'accesso. Tali dati sono accessibili a ricercatori accreditati su richiesta.

[2] La procedura di monitoraggio sviluppata per il "Barometro dell'odio - Elezioni europee 2019" è stata testata nel corso di un monitoraggio pilota che ha avuto luogo tra novembre 2018 e gennaio 2019.

Supponiamo che vi siano solamente 2 politici attivi sui social media: il politico A, che riceve 500 risposte/commenti e il politico B, che ne riceve 2.000.

Ricorrendo alla valutazione di risposte/commenti selezionati casualmente dall'algoritmo per la valutazione, supponiamo di rilevare che il politico A riceve il 20% di risposte/commenti negativi, mentre il politico B ne riceve il 10%.

Per calcolare la proporzione totale di risposte/commenti negativi, iniziamo stabilendo che i commenti totali sono 2.500 (500+2.000). Sul totale delle risposte/commenti il politico A ha ricevuto risposte/commenti nella misura $500/2.500 = 1/5$; il politico B ne ha ricevuti $4/5$ ($2.000/2.500 = 4/5$).

Sulla base di questi due coefficienti (politico A $1/5$ e politico B $4/5$) l'incidenza totale risulta: $(1/5) \times 20\% + (4/5) \times 10\% = 12\%$.

Un simile ragionamento è applicato nella ponderazione delle risposte/commenti degli utenti generici ai tweet/post dei politici (cosicché, per esempio, un tweet/post che riceve 5.000 risposte/commenti abbia un peso pari a 5 volte il peso di un tweet/post che riceve 1.000 risposte/commenti).

Per calcolare gli errori standard sono, inoltre, utilizzate tecniche di bootstrap per il ricampionamento.

Analisi qualitativa

Parallelamente all'analisi quantitativa, un gruppo di ricercatori e esperti condurrà un'analisi di tipo qualitativo sui dati con focus che spaziano dall'analisi testuale a quella giuridica.

Gli esperti

La procedura di monitoraggio è stata ideata e sviluppata da Amnesty International Italia. Il "Barometro dell'odio - Elezioni europee" è stato reso possibile dal supporto di Rania Wazir, data scientist che ha elaborato gli algoritmi necessari allo svolgimento dell'intera procedura.

Hanno contribuito alla definizione dei campi di valutazione e dell'impostazione dell'analisi qualitativa, inoltre, gli esperti del Tavolo odio, spazio di confronto sui discorsi d'odio

promosso da Amnesty International Italia a partire da aprile 2018, che mette insieme che mette in rete ricercatori e organizzazioni della società civile impegnati nello studio e/o nel contrasto dell'hate speech online.

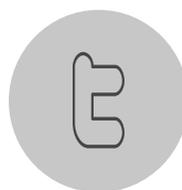
I dati sono disseminati attraverso il sito di Amnesty International Italia a questo indirizzo:
<https://www.amnesty.it/cosa-facciamo/elezioni-europee/>



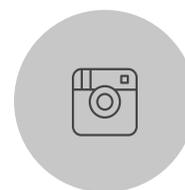
barometro@ai-italy.it
Ufficio stampa:
press@amnesty.it



facebook.com/Amnesty
InternationalItalia



@amnestyitalia



@amnestyitalia

