

Misurare l'efficacia dei messaggi pubblicitari sui social media. Il caso della pandemia in Italia

Angela Maria D'Uggento¹, Claudia Marin^{2*}, Albino Biafora³
Margaret Antonicelli⁴, Barbara Cafarelli⁵

¹Dipartimento di Economia e Finanza, Università degli Studi di Bari Aldo Moro,
²Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli
Studi di Bari Aldo Moro, ⁴Department of Humanistic Studies, Faculty of Communication,
IULM University, Milano, ⁵Dipartimento di Economia, Management e Territorio,
Università di Foggia.

Riassunto: Negli ultimi anni i ricercatori si sono sempre più interessati ai social media come preziosa fonte di dati per rilevare e analizzare le opinioni delle persone su vari argomenti. In occasione di eventi estremi, ad esempio, gli utenti della rete possono diffondere rapidamente informazioni e opinioni online, generando un impatto che può avere implicazioni sia positive che negative sulla percezione del fenomeno collettivo. Twitter è una delle piattaforme che consente alle persone di informarsi su eventi e fatti in tutto il mondo, comprese le epidemie, ed esprimere le proprie reazioni. Pertanto, i social media possono essere utilizzati da governi, aziende e altre organizzazioni per conoscere il sentiment della collettività e interagire con messaggi istituzionali o supporto emotivo. Durante la pandemia di Covid-19, molte aziende hanno utilizzato spot televisivi per infondere fiducia e speranza negli individui, in particolare nelle famiglie, promuovendo solidarietà e fratellanza, mantenendo il marchio in secondo piano. Questo articolo si propone di definire una metrica di valutazione per misurare l'efficacia del cambiamento pubblicitario di grandi marchi, utilizzando un corpus di 20.982 tweet e indicatori creati sulla base dei risultati ottenuti dall'analisi del sentiment. I nostri risultati suggeriscono che gli utenti di internet hanno sperimentato un senso di appartenenza e apprezzamento durante il lockdown, oltre a un atteggiamento positivo verso la pubblicità televisiva, precedentemente percepita come invasiva.

Keywords: analisi del sentiment, consapevolezza del brand, pandemia di Covid-19, social media.

* Autore corrispondente: claudia.marin@uniba.it

1. Introduzione

Nel 2020, il mondo ha vissuto una pandemia senza precedenti che ha cambiato radicalmente il modo in cui le persone vivevano, lavoravano, studiavano e trascorrevano il tempo libero. L'impatto economico e sociale dell'epidemia è stato devastante, causando perdita di posti di lavoro, aumento delle disuguaglianze e disagi notevoli in molti settori. Le aziende sono state costrette ad adottare nuove strategie per sopravvivere, modificando i loro processi lavorativi, riorganizzando i canali di comunicazione e anche offrendo servizi logistici alternativi.

L'epidemia di Coronavirus ha creato una situazione critica non solo dal punto di vista medico, ma anche sociale, poiché la pandemia ha avuto, fra l'altro, un impatto significativo sull'isolamento delle persone. I social media sono diventati parte integrante della vita quotidiana, offrendo nuovi modi per connettersi e comunicare [Basile et al., 2022]. In Italia, l'uso dei social media è aumentato del 21% dopo l'inizio del lockdown nel 2020 rispetto alla fine del 2019 [Basile et al., 2022], posizionando il Paese al secondo posto tra gli utilizzatori dopo la Spagna, dove l'aumento è stato del 48% [Pang and Lee, 2008].

Con l'obiettivo comune di rispettare la sensibilità di chi ha perso improvvisamente una persona cara e aumentare l'empatia con i clienti in un momento di dolore collettivo, le aziende hanno adottato diverse strategie di marketing a seconda del tipo di prodotti venduti. Le imprese, che solitamente investivano ingenti somme nella pubblicità per rafforzare l'immagine e la consapevolezza del marchio, hanno modificato il contenuto degli spot televisivi per trasmettere messaggi di fratellanza e speranza.

La consapevolezza pubblicitaria di un marchio riguarda le tracce che rimangono nella memoria dei consumatori [Tsytarau and Palpanas, 2012]. In particolare, essa si riferisce alla capacità dei consumatori di ricordare o riconoscere un marchio [Hu et al., 2013], o alla semplice conoscenza dell'esistenza di un brand. Inoltre, è legata all'identificazione con un'azienda e influisce sull'impatto della pubblicità sulle reazioni e percezioni dei consumatori [Bhattacharya and Sen, 2003].

I nuovi paradigmi degli spot televisivi introdotti durante la pandemia miravano a suscitare sia risposte affettive che cognitive nei destinatari. Le prime riguardano l'attaccamento sentimentale a un prodotto, le seconde la percezione e la memorizzazione del contenuto pubblicitario associato al marchio. La risposta cognitiva può essere utilizzata per misurare l'efficacia di una campagna pubblicitaria, valutando la sua capacità di impressionare il pubblico, essere compresa e ricordata [15].

Gli indicatori più importanti per misurare l'efficacia della comunicazione includono il riconoscimento, il richiamo e la consapevolezza. Maggiore è la capacità di un marchio di essere immediatamente menzionato, più forte sarà la sua presenza nella memoria dei consumatori, massimizzando il “brand lift”, ovvero l'aumento della percezione e consapevolezza di un prodotto o marchio nel tempo.

Durante la pandemia, il comportamento quotidiano dei cittadini di tutto il mondo è cambiato, così come il rapporto tra consumatori e brand. I consumatori hanno modificato le loro abitudini di acquisto, preferendo lo shopping online a causa del lockdown: le cene al ristorante sono state sostituite da consegne a domicilio, gli acquisti nei negozi da frequenti acquisti su e-commerce e i rapporti faccia a faccia da videochiamate e interazioni sui social media.

Le aziende e i marchi hanno dovuto ripensare sia le loro tecniche di vendita che il modo di fare pubblicità per attrarre il pubblico in un contesto economico così critico. Durante il lockdown, messaggi di incoraggiamento e vicinanza sono stati trasmessi in tutte le comunicazioni, anche pubblicitarie, enfatizzando valori come l'amore, la famiglia e la solidarietà tra le persone.

Alcune note aziende italiane hanno lanciato campagne pubblicitarie mirate al momento sociale corrente, trasmettendo messaggi di connessione tramite videochiamate, internet e intrattenimento a distanza. Anche le case automobilistiche hanno seguito questa direzione, come FCA con lo spot “Inno alla casa” e Volkswagen con l'hashtag #IoRestoInGarage. Brand di cosmetici e make-up hanno introdotto hashtag positivi come “#torneràilsorriso”, accompagnati da canzoni nostalgiche e immagini di momenti felici pre-pandemia.

Questo studio analizza l'impatto del cambiamento del paradigma pubblicitario di alcuni noti brand nei settori alimentare e dei beni di consumo durante il lockdown italiano iniziale (febbraio-maggio 2020). Per raggiungere questo obiettivo sono state utilizzate tecniche di analisi del sentiment [Aggarwal and Zhai, 2002] per analizzare un vasto numero di tweet italiani raccolti online, integrando i metodi tradizionali con tecniche di analisi avanzate per valutare la capacità della pubblicità di catturare l'attenzione dei clienti.

2. Materiali e metodi

2.1 L'analisi del sentiment

L'analisi del sentiment (Sentiment Analysis, SA) è una tecnica di text mining che mira a estrarre informazioni basate sulle opinioni e a rilevare il “sentimento” pre-

sente nei testi [Pang and Lee, 2008]. Comprende una vasta gamma di tecniche, tra cui l'analisi testuale, l'estrazione di pattern e cifre stilistiche, la classificazione dei documenti, il clustering, la visualizzazione, il data mining e il machine learning [Dang and Ahmad, 2015; Aggarwal and Zhai, 2012]. In particolare, la SA è un'area di ricerca attiva nel campo dell'elaborazione del linguaggio naturale (Natural Language Processing, NLP), che utilizza pensieri, opinioni, emozioni, valutazioni, impressioni e atteggiamenti delle persone che possono essere inserite come variabili nei modelli predittivi. Grazie alla rilevanza dei social media come piattaforma di espressione, la SA trova applicazioni in ambiti economici, sociali, aziendali e di ricerca, analizzando l'enorme quantità di dati testuali disponibili online.

La SA può essere applicata su qualsiasi tipo di testo trovato in rete, sia esso strutturato o non strutturato, utilizzando approcci lessicografici o algoritmi di machine learning. L'approccio lessicografico si basa sull'analisi del punteggio di sentiment di un documento attraverso un lessico predefinito, mentre gli algoritmi di machine learning, che utilizzano metodi di classificazione supervisionata e modelli di deep learning, identificano contenuti emotivi in modo più complesso. Ricercatori autori di interessanti studi, utili per approfondire i metodi della SA, sono Liu e Zhang (2012), Tsytsarau e Palpanas (2012) e Giachanou e Crestani (2016).

2.2 Principali approcci per l'analisi del sentiment

L'approccio lessicografico, basato su tecniche di classificazione non supervisionata, analizza la polarità semantica delle parole per determinare il sentiment globale di frasi e documenti. Ogni parola nel lessico riceve un indice positivo o negativo, combinato, poi, per produrre un indice di sentiment complessivo. Questo approccio è particolarmente utile per testi tradizionali (blog, forum, recensioni di prodotti), ma meno efficace con i contenuti brevi e colloquiali di Twitter. Tra gli strumenti più utilizzati troviamo SentiStrength, migliorato da Thelwall et al. nel 2010 e 2012, e SentiCircles [Saif et al., 2016], che integra l'analisi del contesto delle parole.

Gli approcci di machine learning, distinti in apprendimento supervisionato e non supervisionato, utilizzano algoritmi di intelligenza artificiale per risolvere problemi di classificazione. Nel caso supervisionato, il sistema apprende da un set di dati etichettati; nell'approccio non supervisionato, il sistema identifica modelli comuni nei dati non etichettati.

L'approccio basato su grafi utilizza la co-occorrenza degli hashtag per classificare i sentimenti associati. È utile in assenza di dati etichettati e si basa sull'assunzione che le persone siano in grado di influenzare i sentimenti collettivi.

Le tecniche di propagazione delle etichette sono utilizzate per costruire grafi di connessione a partire da tweet, n-gram e hashtag.

Indipendentemente dall'approccio scelto, la SA permette alle aziende di comprendere il sentiment dei clienti verso prodotti, marchi o servizi, espresso nei loro feedback. Queste opinioni, raccolte tramite tweet, recensioni, email, chat e altre interazioni, rappresentano un prezioso patrimonio di conoscenza che può essere trasformato in un vantaggio competitivo per i proprietari del dato. Analizzando correttamente tali dati, le aziende possono adattare prodotti, strutture o strategie pubblicitarie alle esigenze dei consumatori.

Modelli statistici di SA valutano popolarità (positiva, negativa o neutrale), emozioni (gioia, rabbia, tristezza, ecc.) e livelli di interesse nel tempo, osservando l'effetto di nuovi prodotti o comunicati stampa sul sentiment. Inoltre, analisi in tempo reale permettono di identificare rapidamente problemi sui social media e di reagire immediatamente alle emozioni dei clienti, migliorando il rapporto tra consumatori e brand.

2.3. Fonte dei dati e preparazione

Dal suo lancio nel 2006, Twitter ha registrato una crescita esponenziale nella sua base di utenti, grazie alla sua facilità di accesso e alla capacità di caricare, pubblicare e scaricare rapidamente commenti degli utenti. È diventato uno dei più grandi archivi di dati generati dagli utenti, alimentato dall'uso di Internet e dei social network come principali strumenti di comunicazione nella vita quotidiana.

In questo studio, ci proponiamo di analizzare le opinioni degli utenti e di evidenziarne i temi principali.

Inoltre, la Twitter Sentiment Analysis (TSA), una sottocategoria dell'Analisi del Sentiment, è stata sviluppata per valutare gli stati emotivi espressi nei tweet.

Tuttavia, i tweet presentano spesso un linguaggio più informale rispetto a pagine web, blog o articoli di notizie, caratterizzato da maiuscole, parole allungate, abbreviazioni, gergo e neologismi. Tutte queste peculiarità devono essere considerate durante il processo di tokenizzazione, una fase di pulizia o filtraggio che mira a eliminare elementi irrilevanti per la classificazione. Questo passaggio riduce gli errori e rende il processo più efficiente. La tokenizzazione consiste nello "scomporre" un testo in unità più piccole, chiamate token. In questa fase, è fondamentale eliminare i cosiddetti "elementi inquinanti" senza alterare il significato del testo. Tra questi elementi rientrano: URL (ad esempio, <http://...>), etichette come RT (retweet), menzioni e hashtag (@, #), punteggiatura, spazi extra e numeri.

Per identificare il sentiment nei tweet raccolti, sono stati creati due distinti lessici: uno per i termini positivi e uno per quelli negativi [Kiritchenko et al., 2014]. Per l'acquisizione dei tweet, la formazione del dataset e il successivo pre-processing è stato utilizzato il linguaggio di programmazione Python.

Sebbene esistano diverse librerie Python per l'estrazione diretta di tweet, non tutte consentono di recuperare tweet più datati. A tal fine, è stata impiegata la libreria `GetOldTweets3`, che ha permesso di estrarre tweet dal 1° gennaio 2020 al 30 giugno 2020, utilizzando una lista di parole chiave correlate allo scopo dello studio come query di ricerca.

La lista di parole chiave è stata generata includendo nomi di marchi presenti nelle pubblicità, hashtag associati e termini più generali come “spot TV” o “spot pubblicitario coronavirus” (vedi Tabella 1 per l'elenco completo).

Sono stati raccolti i seguenti campi: data del tweet, ID del tweet, contenuto originale, parola chiave utilizzata per l'estrazione e tag di riferimento.

Il dataset finale, codificato in formato .csv con UTF-8, includeva 20.982 tweet in lingua italiana, etichettati in base a tre intervalli temporali predefiniti:

- prima del lockdown (01/01/2020 - 08/03/2020);
- durante il lockdown (09/03/2020 - 03/05/2020, “fase 1”);
- dopo il lockdown (04/05/2020 - 30/06/2020).

I tweet sono stati sottoposti a pre-processing per rimuovere stop word e altri elementi irrilevanti. Il testo pulito, combinato con i metadati (ID, data e ora, parola chiave, tag di riferimento e periodo temporale), ha formato il corpus di testo utilizzato per lo studio.

Per quantificare il sentiment nei commenti dei consumatori sugli spot televisivi dei brand su Twitter, è stato utilizzato il **lessico openER** [Russo et al., 2016]. Questo strumento assegna a ciascun termine una polarità (negativa, positiva o neutra) e un indice associato. Il processo di polarizzazione dei tweet avviene tramite una funzione automatizzata denominata “**search**”, che, dato un argomento di ricerca come input (ad esempio un nome, un hashtag o un utente), genera automaticamente un dataset e fornisce i seguenti output:

1. **File .csv (_stack.csv):** include tutti i dati originali del dataset di partenza, preservandoli per eventuali analisi future.
2. **File .csv (_score.csv):** attribuisce un punteggio a ciascun tweet; tali punteggi sono utili nella fase di analisi delle correlazioni.
3. **File .csv (_opin.csv):** assegna una polarità (positiva/negativa/neutral) a ciascun tweet.

4. **Grafico scatterplot:** visualizza la polarità dei tweet in relazione a un trend storico di riferimento.

La classificazione dei tweet si basa su un sistema di pesatura o di assegnazione di punteggi. Ogni tweet, dopo essere stato sottoposto al processo di pulizia, viene suddiviso in un vettore di singole parole, ognuna delle quali rappresenta un'entità indipendente.

Per avviare questo tipo di analisi, è fondamentale definire un **lessico**, ovvero un tipo di dizionario che include termini della lingua italiana associati a un livello di positività, negatività o neutralità. Questo elenco è particolarmente utile poiché la maggior parte dei programmi di analisi del sentiment è progettata per analizzare testi in lingua inglese.

Durante l'analisi del sentiment basata su un lessico, vengono generalmente applicate funzioni di combinazione, come somma o media, per ciascun documento, tenendo conto del contesto locale delle parole, ad esempio la presenza di negazioni o intensificazioni.

I dati sono stati elaborati con i linguaggi di programmazione Python e R Studio.

3. Risultati

Per comprendere come sia effettivamente cambiata la percezione dei messaggi televisivi è stato effettuato uno studio sulla frequenza dei tweet nel periodo di tempo analizzato.

Un grafico molto intuitivo comunemente utilizzato in tali casi è la wordcloud, che consente una immediata esplorazione delle parole presenti con maggior frequenza nel testo analizzato.

La Figura 1 illustra la frequenza dei 100 termini più utilizzati nei Tweet di commento alle pubblicità durante la pandemia.

Figura 1. Wordcloud dei 100 termini più frequentemente utilizzati nel periodo di osservazione



La Figura 2 è, invece, riferita ai tweet del giorno 7 maggio 2020, in cui sono state scritte esattamente 32.496 parole, l'89% delle quali sono state categorizzate come relative al brand Lavazza. L'hashtag #Lavazza quel giorno ha raggiunto infatti la terza posizione tra le tendenze su Twitter in Italia (c.d. Trending Topic). La word-cloud in Fig. XX mostra le 50 parole più frequenti associate al brand Lavazza.

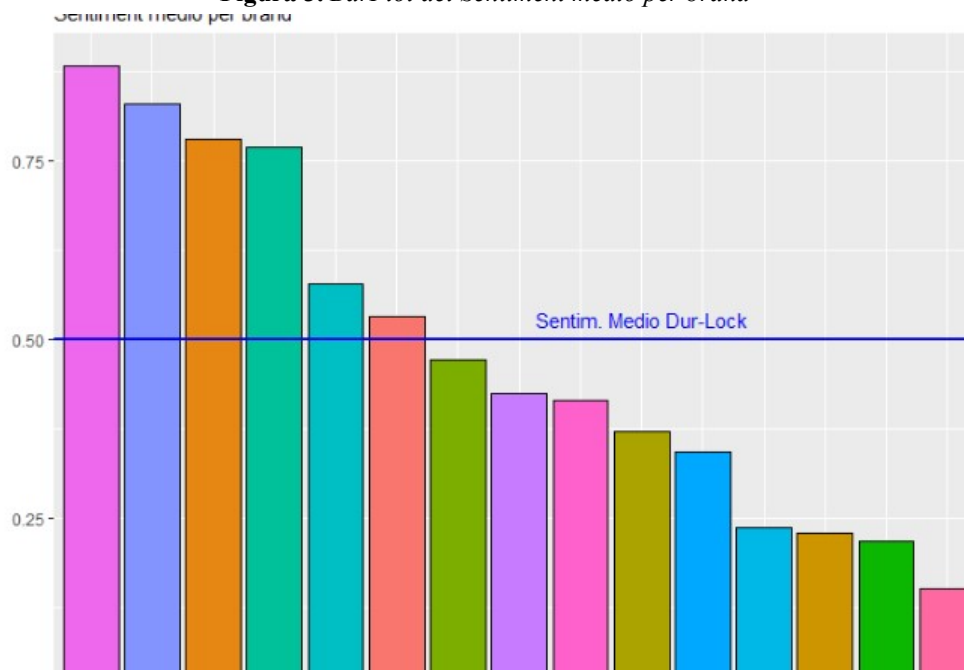
Figura 2. WordCloud spot Lavazza nei tweet del 7 maggio 2020



Le parole più utilizzate riguardano molti aggettivi positivi come: “bello/bella”, “grande” e “bellissimo” ed altre parole di polarità negativa come “fascisti”. La parola più legata al Brand è Charlie Chaplin, protagonista assoluto dello spot con il suo discorso sull'umanità, tratto dal celebre film “Il Grande Dittatore”. Per comprendere la wordcloud è utile sintetizzare il contenuto dello spot Lavazza. Tale pubblicità ha suscitato ampio dibattito sui media, riscuotendo sia apprezzamento, in quanto ha suscitato sentimenti positivi per la comunità durante la crisi, ma attirando anche molte critiche perchè considerato cinicamente opportunist. In particolare, alcuni tweet hanno anche evidenziato il passaggio di Lavazza dalla clip comica nazional-popolare al messaggio emozionale sul progresso dell'umanità, evidenziando proprio il cambio di paradigma nella struttura del messaggio.

L'analisi che prende in considerazione i marchi più “twittati” può risultare interessante per comprendere quale brand abbia generato opinioni positive o negative durante il periodo del lockdown. A tale scopo è stato sommato lo score di ogni tweet riferito ad un determinato brand per poi dividere il risultato per il numero di tweet di quel brand. Si è così ottenuto il sentiment medio per brand, rappresentato in Figura 3.

Figura 3. BarPlot del Sentiment medio per brand



È evidente come nessun brand abbia ricevuto mediamente opinioni negative, mentre quelle positive mostrano punteggi differenziati, da quelli relativi agli spot Esselunga e WindTre, a quelli degli spot Toyota, MulinoBianco, Bauli e Famila che hanno raggiunto uno score medio di 0.75. In cima alla lista si collocano, quindi, i brand che, attraverso una scelta strategica di cambiamento del metodo di approccio negli spot pubblicitari televisivi, hanno creato maggiormente opinioni positive sul social network Twitter.

4. Conclusioni

I primi risultati di questo studio evidenziano come il ruolo assunto dai social media nella quotidianità delle persone li abbia, di fatto, inseriti tra gli strumenti con i quali è possibile misurare l'efficacia di una campagna pubblicitaria, integrando, o in alcuni casi sostituendo, quelli tradizionali.

La pandemia è stato un evento devastante, sconosciuto e del tutto inaspettato. Durante il lockdown, con le stesse immagini terrificanti che provenivano da tutto il mondo, è stato fondamentale contrastare lo sconcerto e trasmettere fiducia e speranza. Il messaggio delle istituzioni italiane “andrà tutto bene” era onnipresente.

Le aziende hanno fatto la loro parte in quanto operatori del sistema economico.

I format tradizionali degli spot sono stati inevitabilmente adattati, suscitando reazioni nei consumatori che, in questo studio, si è cercato di misurare attraverso l'aumento del numero di tweet, il loro contenuto e le attività digitali quotidiane. I risultati delle analisi confermano che i consumatori hanno notato e apprezzato il fatto che i grandi marchi abbiano modificato la loro strategia comunicativa durante la pandemia, contribuendo, con i loro spot televisivi, a diffondere messaggi di speranza e sui principi di solidarietà.

Le tecniche dell'analisi testuale consente ai ricercatori di gestire l'esigenza di automatizzare la comprensione di dati testuali non strutturati, non solo provenienti da libri e giornali, ma anche da poster, blog, recensioni dei clienti su Internet e post sui social media.

Il presente studio si concentra su un corpus specifico di dati non strutturati, trasformandoli opportunamente per effettuare analisi di text mining e identificare pattern.

Dopo una fase accurata di estrazione e pre-elaborazione dei tweet con algoritmi di elaborazione del linguaggio naturale, è stato creato un dataset composto da

20.982 tweet analizzati in tre periodi distinti: prima del lockdown, durante il lockdown e dopo il lockdown.

La novità degli approcci pubblicitari dei brand sembra aver evocato, soprattutto durante il lockdown, un senso di gratitudine e appartenenza, meno evidente negli altri due periodi. In questi ultimi, infatti, i tweet si sono concentrati principalmente sui prodotti dei brand o su elementi percepiti come fastidiosi.

Per le ricerche future, ci si propone di affrontare i limiti imposti dai dizionari lessicali e dalle librerie computazionali (sia in R che in Python), abbondanti per la lingua inglese ma spesso inadeguate o incomplete per l'italiano.

La traduzione dei corpora italiani in inglese non è un compito semplice: richiede un lavoro preparatorio complesso, tra cui lo sviluppo di una funzione specifica di somiglianza tra stringhe, basata su un lessico composto esclusivamente da forme regolari delle parole. Inoltre, la coniugazione dei verbi rappresenta una discrepanza significativa.

L'analisi dei bigrammi e delle co-occorrenze potrebbe migliorare la costruzione di un lessico italiano adeguato e, alla luce degli obiettivi di questo studio, facilitare l'identificazione degli spot più efficaci e delle loro caratteristiche distintive.

Riferimenti bibliografici

- Aggarwal C.C., Zhai C. (2012). *Mining Text Data*. Springer Nature. doi:10.1007/978-1-4614-3223-4.
- Basile, V. Misuraca M., Spano M. (2022). Thematic analysis on online education issues during COVID-19. In: SIS2022 51st Scientific Meeting of the Italian Statistical Society. Book of the Short Papers, Balzanella A., Bini M., Cavicchia, C. Verde, R. (eds) pp. 437-445. Pearson.
- Bhattacharya C.B., Sen S. (2003). Consumer– company identification: A framework for understanding consumers' relationships with companies. *Journal of Marketing*, 67(2), 76-88.
- Dang S., Ahmad P.H. (2015). A Review of Text Mining Techniques Associated with Various Application Areas. *Int J Sci Res (IJSR)*, 4(2), 2461-2466.
- Giachanou A., Crestani F. (2016). Like it or not: a survey of Twitter Sentiment analysis methods. *ACM Comput. Surv.*, 49(2):Article 28. doi:10.1145/2938640.
- Hu X., Tang J., Gao H., Liu H. (2013). Unsupervised sentiment analysis with emotional signals. In *Proceedings of the International World Wide Web Conference*, pages 607–618.
- Kiritchenko S., Zhu X., Mohammad S. M. (2014). Sentiment analysis of short in-

- formal texts. *J. Artif. Intell. Res.*, 50, 723–762. doi:10.1613/jair.4272.
- Liu, B., Zhang L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In Aggarwal C. and Zhai C. editors, *Mining Text Data*, Springer Nature, pp. 415-463.
- Pang B., Lee L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, 2(1-2), 1-135, doi:10.1561/15000000011.
- Russo I. et al. (2016). OpeNER Sentiment Lexicon Italian - LMF, ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics “A. Zampolli”, National Research Council (CNR), Pisa. <http://hdl.handle.net/20.500.11752/ILC-73>.
- Saif H., Bashevoy M., Taylor S., Fernandez M., Alani H. (2016). A Platform for Contextual and Conceptual Sentiment Analysis. *Extended Semantic Web Conference*.
- Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas, A. (2010), Sentiment strength detection in short informal text. *J. Am. Soc. Inf. Sci.*, 61: 2544-2558. <https://doi.org/10.1002/asi.21416>.
- Thelwall M., Buckley K., Paltoglou G. (2012), Sentiment strength detection for the social web1. *Computer Science*.
- Tsytsarau M., Palpanas T. (2012). Survey on mining subjective data on the web. *Data Min Knowl Disc* 24(3):478-514. doi:10.1007/s10618-011-0238-6.