



Ricerca di Sistema elettrico

Studio delle tecniche di Sentiment Analysis e Data Analytics per l'estrazione di conoscenza dai social media contents di tipologia testuale

B. Di Martino, A. Esposito, F. Marulli, M. Graziano



Università
degli Studi
della Campania
Luigi Vanvitelli

Dipartimento di Ingegneria

STUDIO DELLE TECNICHE DI SENTIMENT ANALYSIS E DATA ANALYTICS PER L'ESTRAZIONE DI CONOSCENZA DAI
SOCIAL MEDIA CONTENTS DI TIPOLOGIA TESTUALE

B. Di Martino, A. Esposito, F. Marulli, M. Graziano (Dipartimento di Ingegneria Università degli studi della
Campania "L. Vanvitelli")

Aprile 2021

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dello Sviluppo Economico - ENEA

Piano Triennale di Realizzazione 2019-2021 - II annualità

Obiettivo: Tecnologie

Progetto: Tecnologie per la penetrazione efficiente del vettore elettrico negli usi finali

Work package: Local Energy District

Linea di attività: Tecniche di Sentiment Analysis e Big Data Analytics per Twitter analysis

Responsabile del Progetto: Claudia Meloni, ENEA

Responsabile del Work package: Claudia Meloni, ENEA

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione "*Tecniche di Sentiment Analysis e Big Data Analytics per Twitter analysis*"

Responsabile scientifico ENEA: Dr. Gregorio D' Agostino

Responsabile scientifico UniCampania - Dip. Ingegneria: Prof. Beniamino Di Martino

Indice

SOMMARIO

1	INTRODUZIONE	
2		6
2.1		7
2.2		9
2.3	TWITTER SENTIMENT ANALYSIS	
2.4	TOOLS PERI SENTIMENT ANALYSIS	
3		36
3.1	COS'È L'ANALISI DEI DATI E PERCHÉ È IMPORTANTE	
3.2	TIPI DI METODI DI ANALISI DEI DATI	
3.3	PROCESSO DI ANALISI DEI DATI	
3.4	TECNICHE DI ANALISI DEI DATI	
3.5	ANALISI DEI DATI IN AMBIENTE BIG DATA	
3.5. 1	COSA SONO I BIG DATA	
3.5. 2	PIATTAFORME DI ELABORAZIONE E RACCOLTA DI BIG DATA	
3.5. 3	TECNICHE DI BIG DATA ANALYTICS	
4	ENERGY COMMUNITIES E SOCIAL ANALYSIS	
5		36

● Sommario

Il lavoro riportato nel presente report si focalizza sullo studio dello stato dell'arte nell'ambito del Natural Language Processing, rivolto in particolare all'analisi dei contenuti testuali prodotti dagli utenti dei moderni Social Networks. Particolare attenzione è stata rivolta allo studio di tool e tecniche correntemente in uso per analizzare piattaforme di microblogging quali Twitter, e nello specifico alla Sentiment Analysis applicata ai contenuti testuali pubblicati dagli utenti. Il report analizza scopi e obiettivi della Sentiment Analysis, e fornisce informazioni specifiche su diversi tool applicati in tale ambito, descrivendone le specifiche funzionalità e capacità e mostrandone l'interfaccia utente. Vengono sottolineati gli obiettivi prettamente di Business della Sentiment Analysis, volti a migliorare le conoscenze delle compagnie relativamente ai loro stessi clienti, e dunque votati ad un miglioramento della Customer Care.

Il report si sofferma inoltre sulle diverse tipologie di Sentiment Analysis che è possibile eseguire sui contenuti testuali, e sulle diverse tecnologie che possono essere utilizzate per ottenere informazioni più o meno dettagliate relativamente ai trend popolari. Nell'esaminare i diversi tool, il report si sofferma sugli specifici ambiti in cui essi possono essere applicati con successo, e sui benefici che essi comportano nell'analizzare l'enorme mole di dati resa disponibile dai Social.

Proprio sul trattamento dei dati è focalizzata la seconda sezione del report, che si concentra sulla Data Analysis e sulle metodologie da sfruttare in tale ambito. In particolare, il report fa riferimento a tecniche di Big Data analysis, diventate fondamentali in seguito all'accrescimento dei volumi di dati dovuti alla digitalizzazione delle attività delle compagnie, al sempre maggiore uso dei social network, e alla nascita di nuove sorgenti di informazioni come l'Internet of Things. Il report si focalizza sulla descrizione di tecniche volte all'analisi di tali moli di dati, identificando le diverse parti che costituiscono gli step di una buona Big Data Pipeline, e indicando tool e strumenti utilizzabili in ciascuna delle fasi individuate.

1 Introduzione

Questa attività è stata finalizzata a condurre uno studio relativo allo stato dell'arte sulle principali tecniche di analisi e inferenza applicabili ai contenuti prodotti dagli utenti sui Social Network, con particolare riferimento ai contenuti di tipologia testuale prodotti nelle piattaforme di microblogging quale quella di Twitter.

L'avvento dei Social Media e la diffusione degli User Generated Content, nati grazie alle tecnologie del Web 2.0, hanno infatti permesso a tutti gli utenti della Rete di esprimere e condividere la propria opinione con il resto del mondo. Questo fenomeno è di rilevante importanza per i knowledge worker, che analizzano i contenuti testuali pubblicati su Internet per ricavare informazioni e conoscenza utilizzabile in fase decisionale. Si sono affermati molteplici metodi di Sentiment Analysis, con lo scopo di identificare e classificare le opinioni e gli umori, espressi da un autore in relazione ad un'entità o un evento, a seconda della loro polarità. Questo tipo di analisi è stato utilizzato con obiettivi diversi in numerosi settori, come la sanità, finanza, politica, sociologia e psicologia. Twitter rappresenta una delle piattaforme di Social Networking più popolari ed utilizzate e la sua produzione testuale (tweets) è oggetto, già da alcuni anni, di analisi, al punto da aver creato un filone di ricerca chiaramente profilato e denominato "twitter analysis". L'obiettivo primario della twitter analysis è quello di estrarre, attraverso opportune tecniche di Big Data Analytics, gli Users' Insights rispetto ad eventi, fenomeni di costume e società, azioni intraprese da privati e/o amministrazioni pubbliche. L'estrazione degli Users' Insights consente di effettuare analisi di profilazione degli utenti ovvero, di profilare il "sentiment" che fenomeni, politiche ed azioni suscitano nella popolazione. Gli utenti tendono ad esprimersi in modalità libera sui social network e questo aspetto di completa disinibizione degli utenti rappresenta un fattore di grande valore quando si vuole caratterizzare, ad esempio, l'indice di gradimento di un fenomeno. Se contenuti prodotti sui social rappresentano "oro" per l'estrazione di conoscenza, il processo stesso di estrazione di conoscenza significativa non è banale e coinvolge metodologie e tecniche di data e text mining non affatto banali, data la brevità dei singoli messaggi, come nel caso dei tweet (tipicamente tra i 120 ed i 150 caratteri).

Sebbene attualmente esistano tecniche molto sofisticate e performanti di Natural Language Processing (NLP) (ad esempio, quelle basate su Deep Learning e Deep Neural Network) non è scontato che esse siano immediatamente applicabili all'analisi e all'estrazione di Insights su testi molto brevi e talvolta anche molto irregolari sotto il profilo lessico-grammaticale. La grande sfida

legata all'analisi e alla estrazione di conoscenza dai dati prodotti dalla User Generation Contents (UGC), sui social network e, più in generale nel web è proprio quella di adattare tecniche e sistemi ben assestati di NLP e text mining a testi brevi ed irregolari, al fine di catturare opinioni e verificare la presenza e la polarità di sentimenti espressi in essi. La possibilità di ricavare informazioni dai contenuti prodotti dagli utenti sui social network ha aperto delle interessanti prospettive sul piano della comunicazione tra parti diverse, (utenti e istituzioni e/o pubbliche amministrazioni (P.A). Le P.A. hanno avuto la possibilità di ridurre il gap con i cittadini, legato all'eccesso di burocrazia, proprio utilizzando canali di comunicazione ed interazioni come Twitter. Questi sistemi di interazione hanno generato flussi informativi che hanno permesso di fotografare l'opinione che gli utenti hanno rispetto all'operato e/o alle scelte della amministrazione e le relazioni che intercorrono tra tutti gli attori che operano nello spazio geografico e territoriale dove si concretizza l'azione amministrativa, quali il trasporto e l'illuminazione pubblica.

● 2 Sentiment Analysis

Con **Sentiment analysis** (anche nota con il nome di **Opinion Mining**) si indica l'insieme delle tecniche e procedure atte allo studio e l'analisi di informazioni testuali, al fine di rilevare valutazioni, opinioni, atteggiamenti ed emozioni relative ad una certa entità (prodotto, persona, argomento, ecc). Essa è il processo di rilevamento del sentimento positivo o negativo nel testo. Questo tipo di analisi ha importanti applicazioni in campo sociale, economico e politico. Ad esempio, un'azienda potrebbe essere interessata a conoscere le opinioni dei consumatori relative ai propri prodotti. Ma anche i potenziali compratori di un determinato prodotto o servizio saranno interessati a conoscere l'opinione e l'esperienza di qualcuno che ha già acquistato o utilizzato il prodotto. D'altro canto anche un personaggio pubblico (politica, spettacolo, sport) potrebbe essere interessato a sapere cosa la gente pensa di lui. Naturalmente esistono già strumenti per la rilevazione dei consensi e delle opinioni (sondaggi e indagini statistiche); ma tramite tecniche di Opinion Mining si hanno costi di rilevazione nettamente inferiori e in molti casi molta più autenticità informativa: infatti, le persone non sono obbligate ad esprimere opinioni (come in sondaggi e interviste), al contrario, queste fluiscono liberamente senza alcuna costrizione.

Gli analisti di dati (Data analysts) utilizzano Sentiment Analysis per estrarre informazioni per ricerche di mercato, monitorare la reputazione del marchio e del prodotto. Questa procedura è anche abbastanza utile per ottenere ciò che i clienti pensano e per agire di conseguenza migliorando la cosiddetta “customer experience”. Negli ultimi anni, vista la crescita esponenziale nell’uso di social media, le persone e le aziende nei loro processi decisionali, utilizzano sempre più le informazioni (opinioni e preferenze) provenienti da mezzi come recensioni, forum, discussioni, blog e social network, tuttavia, la ricerca di opinioni sul Web da parte di un utente o azienda risulta essere un problema molto arduo a causa di alcuni fattori quali:

- La proliferazione di una grande varietà di fonti informative;
- Il grande volume di dati testuali attraverso cui si presentano le informazioni presenti nei siti che non sempre è decifrabile in maniera ottimale (si pensi ai lunghi testi che si trovano nei forum e blog).
- La distorsione a cui è soggetta l’analisi soggettiva delle informazioni testuali (es: le persone tendono a prestare maggiore attenzione e interesse alle opinioni che risultano coerenti alle proprie attitudini e preferenze).

Dall’analisi di tali fattori si evince la necessità di ricorrere all’uso di sistemi automatizzati di Opinion Mining, per superare la soggettività che connota le informazioni rese disponibili dalle recensioni, forum e blog presenti sul web, al fine di ottenere una metodologia di Sentiment analysis che sia il più possibile oggettiva.

La Sentiment analysis è uno dei compiti più difficili nell’elaborazione del linguaggio naturale, perché anche gli umani fanno fatica ad analizzare accuratamente e oggettivamente i sentimenti che connotano un’informazione.

○ 2.1 Applicazione della Sentiment Analysis nel business

La sentiment analysis è estremamente importante perché aiuta le aziende a capire rapidamente le opinioni generali dei loro clienti. Ordinando automaticamente il sentiment dietro le recensioni, le conversazioni sui social media e altro, è possibile prendere decisioni più veloci e accurate.

Le applicazioni della Sentiment analysis sono infinite e possono essere applicate a qualsiasi settore, dalla finanza alla vendita al dettaglio, dall'ospitalità alla tecnologia. Alcune delle applicazioni più popolari della sentiment analysis nel business sono:

- **Social Media Monitoring - Monitoraggio dei social media:** la Sentiment analysis è utilizzata nel monitoraggio dei social media, consentendo alle aziende di ottenere informazioni su come i clienti si sentono riguardo a determinati argomenti, e rilevare problemi urgenti in tempo reale prima che vadano fuori controllo.
- **Brand Monitoring - Monitoraggio del marchio:** la sentiment analysis permette di monitorare automaticamente tutte le opinioni intorno ad un marchio analizzando tutte le informazioni che lo riguardano disponibili sui social media, ma anche su internet, su siti di notizie, blog, forum, recensioni di prodotti e altro.
- **Voice of customer (VoC) - Voce del cliente (VoC):** I social media e il monitoraggio del marchio offrono informazioni immediate, non filtrate e inestimabili sul sentimento/opinione dei clienti, ma si può applicare tale analisi anche su sondaggi e sulle interazioni con il supporto clienti per capire le emozioni e le opinioni dei clienti. Tracciare il sentiment dei clienti nel tempo aggiunge profondità per aiutare a capire come i sentimenti/opinioni verso singoli aspetti di una attività possono essere cambiati.
- **Customer Service - Servizio clienti:** È possibile utilizzare l'analisi del sentiment e la classificazione del testo per organizzare automaticamente le richieste di assistenza in arrivo in base all'argomento e all'urgenza, per indirizzare al reparto corretto e assicurarsi che le più urgenti siano gestite subito. Analizzare le interazioni con l'assistenza clienti per garantire che i vostri dipendenti stiano seguendo il protocollo appropriato. Aumentare l'efficienza, in modo che i clienti non rimangano in attesa del supporto. Diminuire i tassi di abbandono; dopo tutto è meno complicato mantenere i clienti che acquisirne di nuovi.
- **Market Research - Ricerca di mercato:** La Sentiment Analysis permette tutti i tipi di ricerche di mercato e di analisi della concorrenza. Se un'azienda sta esplorando un nuovo mercato, anticipando le tendenze future o cercando un vantaggio sulla concorrenza, la Sentiment Analysis può fare la differenza in quanto è possibile analizzare le recensioni online dei propri prodotti e confrontarle con quelle della concorrenza.

I benefici generali dell'analisi del sentiment includono:

- **Ordinamento dei dati su scala:** Ci sono troppi dati aziendali da elaborare manualmente. La sentiment analysis aiuta le aziende a elaborare enormi quantità di dati in modo efficiente e conveniente.
- **Analisi in tempo reale:** La sentiment analysis può identificare i problemi critici in tempo reale, per esempio una crisi di PR sui social media si sta aggravando? Un cliente arrabbiato sta per abbandonare l'azienda? I modelli di sentiment analysis possono aiutare a identificare immediatamente questo tipo di situazioni, in modo da poter agire subito.
- **Criteri coerenti:** Si stima che le persone siano d'accordo solo per il 60-65% delle volte nel determinare il sentiment di un particolare testo. Etichettare il testo in base al sentiment è altamente soggettivo, influenzato da esperienze personali, pensieri e credenze. Utilizzando un sistema centralizzato di analisi del sentiment, le aziende possono applicare gli stessi criteri a tutti i loro dati, aiutandole a migliorare l'accuratezza e ad ottenere migliori intuizioni.

○ 2.2 Tipologie di Sentiment Analysis

La sentiment analysis è basata sull'impiego di tecniche di elaborazione del linguaggio naturale (NLP) e di algoritmi di apprendimento automatico, per determinare automaticamente il carattere emotivo dietro le conversazioni online.

Gli algoritmi che è possibile adottare per implementare modelli di Sentiment Analysis sono numerosi e si diversificano a seconda della quantità di dati da analizzare ed in base alla accuratezza del modello.

Gli algoritmi di Sentiment Analysis vengono classificati in tre gruppi principali:

- **Basati su regole:** questi sistemi eseguono automaticamente l'analisi del sentiment sulla base di una serie di regole elaborate manualmente. Di solito, un sistema basato su regole utilizza un insieme di regole create dall'uomo per aiutare a identificare la soggettività, la polarità o il soggetto di un'opinione. Queste regole possono includere varie tecniche NLP sviluppate nella linguistica computazionale, come:
 - Stemming, tokenization, part-of-speech tagging e parsing.

- Lessici (cioè liste di parole ed espressioni).
- **Automatico:** i sistemi si basano su tecniche di apprendimento automatico per imparare dai dati. I metodi automatici, contrariamente ai sistemi basati su regole, non si basano su regole create manualmente, ma su tecniche di apprendimento automatico. Un compito di sentiment analysis è solitamente modellato come un problema di classificazione, per cui un classificatore viene alimentato con un testo e restituisce una categoria, ad esempio positiva, negativa o neutra.
- **Sistemi ibridi:** combinano sia approcci basati su regole che automatici. I sistemi ibridi combinano gli elementi desiderabili di tecniche basate su regole e automatiche in un unico sistema. Un enorme vantaggio di questi sistemi è che i risultati sono spesso più accurati.

I modelli di sentiment analysis si concentrano sulla polarità (positivo, negativo, neutro) ma anche su sentimenti ed emozioni (arrabbiato, felice, triste, ecc.), urgenza (urgente, non urgente) e persino intenzioni (interessato o non interessato).

A seconda di come si vogliono interpretare i feedback e le richieste dei clienti, è possibile definire e adattare le proprie categorie per soddisfare le proprie esigenze di sentiment analysis. Alcuni dei tipi più popolari di sentiment analysis sono:

- **Sentiment Analysis a grana fine - Fine-grained Sentiment Analysis:** Se la precisione della polarità è importante per un business, si possono espandere le categorie di polarità per includere:
 - **Molto positivo**
 - **Positivo**
 - **Neutro**
 - **Negativo**
 - **Molto negativo**

Questo è di solito indicato come Sentiment Analysis a grana fine, e potrebbe essere utilizzato per interpretare le valutazioni a 5 stelle in una recensione, per esempio:

Molto positivo = 5 stelle

Molto negativo = 1 stella

- **Rilevamento delle emozioni - Emotion detection:** Questo tipo di sentiment analysis mira a rilevare le emozioni, come felicità, frustrazione, rabbia, tristezza e così via. Molti sistemi di rilevamento delle emozioni utilizzano lessici (cioè liste di parole e le emozioni che trasmettono) o complessi algoritmi di apprendimento automatico. Uno degli aspetti negativi dell'uso dei lessici è che le persone esprimono le emozioni in modi diversi. Alcune parole che tipicamente esprimono rabbia, come male o uccidere potrebbero anche esprimere felicità.
- **Sentiment Analysis basata sugli aspetti - Aspect-based Sentiment Analysis:** Di solito, quando si analizzano i sentimenti dei testi, come le recensioni dei prodotti, si vuole sapere quali aspetti particolari o caratteristiche le persone stanno menzionando in modo positivo, neutro o negativo. È qui che la sentiment analysis basata sugli aspetti può aiutare, per esempio in questo testo: "*La durata della batteria di questa macchina fotografica è troppo breve*", un classificatore basato sugli aspetti sarebbe in grado di determinare che la frase esprime un'opinione negativa sulla caratteristica durata della batteria.
- **Sentiment Analysis multilingue - Multilingual sentiment analysis:** La Sentiment Analysis multilingue può essere difficile. Implica un sacco di pre-elaborazione e di risorse. La maggior parte di queste risorse sono disponibili online (ad esempio i lessici del sentimento), mentre altre devono essere create (ad esempio i corpora tradotti o gli algoritmi di rilevazione del rumore), ma è necessario sapere come codificare per utilizzarli.

Un sistema efficiente di Sentiment Analysis deve basarsi su un'adeguata libreria di sentimento per rilevare correttamente i sentimenti e i punteggi nelle parole o nelle frasi.

Le **librerie di sentimento** sono composte da una collezione di dizionari che includono aggettivi e frasi che sono stati precedentemente valutati manualmente. Questo processo di punteggio deve essere fatto con attenzione in modo che il sistema di Sentiment Analysis possa differenziare, in seguito, tra parole come '*cattivo*' e '*orribile*', tenendo '*orribile*' un significato più negativo. A parte questo, se si ha un motore di Sentiment Analysis multilingue, ci devono essere librerie per ogni lingua supportata. Ogni libreria può essere personalizzata e modificata a seconda delle necessità: aggiungendo o rimuovendo frasi/parole, mettendo a punto i punteggi.

Una volta che questi dizionari sono preparati, una serie di regole deve essere scritta nel software in modo che il computer sia in grado di rilevare correttamente il sentimento espresso verso un particolare argomento, sulla base della sua vicinanza ad altre parole positive e negative. Un chiaro esempio è la frase aggettivale *'un po' deluso'* rispetto a *'totalmente deluso'* che porta un sentimento più negativo.

Per analizzare correttamente una frase per il sentiment, c'è bisogno di scomporla in pezzi che coinvolgono diversi sotto-processi, incluso il POS-tagging, quest'ultimo, consiste nell'identificazione degli elementi di base di un testo, come verbi, nomi, aggettivi e avverbi. Molte lingue seguono alcune regole e schemi in termini di formazione delle parole che possono essere tradotti per un programma informatico per sviluppare un POS-tagger di base. Tuttavia, un sistema affidabile di Sentiment Analysis deve essere costruito su un accurato software di comprensione del linguaggio naturale per ottenere risultati precisi di POS-tagging, che sono cruciali per identificare diverse combinazioni di frasi.

Un sistema di valutazione del sentiment basato su frasi e regole può trarre false conclusioni, in quanto ad esempio poiché vedrà che l'hotel è stato descritto come *"non molto buono"*, assegnerà un punteggio di sentiment negativo e passerà alla recensione successiva. Tuttavia, i lettori umani vedranno chiaramente che questa recensione in realtà racconta una storia diversa. Anche se al cliente non è piaciuto l'hotel in sé, la vista dalla camera contribuisce a migliorare l'esperienza del cliente. Pertanto, un sistema automatizzato di Sentiment Analysis deve prendere in considerazione ogni singola parola e struttura per assegnare punteggi di sentimento adeguati.

Se si ha bisogno di fare un'analisi che rilevi sottili differenze di significato, bisogna cercare uno strumento che impieghi sia tecniche di apprendimento automatico che di elaborazione del linguaggio naturale.

Il ruolo principale delle tecniche di machine learning nella sentiment analysis è quello di automatizzare le funzioni di analisi del testo su cui si basa la sentiment analysis (segmentazione, POS-tagging, estrazione di entità...). Per esempio, quando i data scientist addestrano un modello di apprendimento automatico alimentandolo con un gran numero di documenti di testo contenenti esempi pre-tagati, esso rileverà automaticamente la Sentiment Analysis nei documenti futuri. Questo è possibile grazie a tecniche di apprendimento automatico supervisionato e non supervisionato, come le reti neurali e l'apprendimento profondo. Il machine learning aiuta anche gli

analisti di dati a risolvere i problemi dipendenti dal contesto causati dall'evoluzione del linguaggio naturale.

Alcuni sistemi ibridi di Sentiment Analysis combinano l'apprendimento automatico con tecniche di elaborazione del linguaggio naturale per raggiungere una maggiore precisione. E' importante fare una differenza tra l'elaborazione del linguaggio naturale e l'apprendimento automatico. Da un lato, una Sentiment Analysis basata su NLP diventa uno strumento efficace per costruire una base per il POS-tagging e la Sentiment Analysis. Dall'altro lato, le tecniche di apprendimento automatico possono aiutare a risolvere compiti complessi di elaborazione del linguaggio naturale, come la comprensione dei doppi sensi attraverso l'addestramento automatico.

Una combinazione di tecniche di ML e NLP coprirà, quindi, l'intera procedura di analisi del testo per la Sentiment Analysis, dalla segmentazione di basso livello e l'analisi della sintassi fino alla differenziazione semantica a seconda del contesto in cui appare una parola.

2.3 Twitter Sentiment Analysis

Milioni di persone usano Twitter ed esprimono le loro emozioni come felicità, tristezza, rabbia, ecc. Per lavorare sui dati testuali, cercano di estrarre i dati dalle piattaforme dei social media. Ci sono molti siti di social media come Google Plus, Facebook e Twitter che permettono di esprimere opinioni, punti di vista ed emozioni su determinati argomenti ed eventi. Twitter è stato fondato nel 2006 e attualmente è la piattaforma di microblogging più famosa.

Gli utenti di Twitter usano postare i loro pensieri, emozioni e messaggi sui loro profili, chiamati tweet. Il limite di parole di un singolo tweet è di 140 caratteri. La Sentiment Analysis di Twitter si basa sul campo NLP (elaborazione del linguaggio naturale). Per il testo dei tweet, si usano tecniche NLP come la tokenizzazione delle parole, la rimozione delle parole di arresto come io, me, mio, nostro, tuo, è, era, ecc. L'elaborazione del linguaggio naturale gioca anche una parte nel preprocessamento dei dati attraverso operazioni come pulire il testo e rimuovere i caratteri speciali e i segni di punteggiatura. La Sentiment Analysis è molto importante perché è possibile conoscere le tendenze delle emozioni delle persone su argomenti specifici con i loro tweet.

Ascoltare attentamente la voce dei clienti su Twitter utilizzando la Sentiment Analysis permette alle aziende di capire il loro pubblico, di stare al passo con ciò che viene detto del loro marchio - e dei loro concorrenti - e di scoprire nuove tendenze nel settore.

Eseguire la Sentiment Analysis sui dati di Twitter comporta cinque passi:

1. **Raccogliere i dati rilevanti di Twitter:** È importante che i dati di Twitter siano rappresentativi di ciò che si sta cercando di scoprire perché verranno usati per:

- Addestrare il modello di analisi del sentiment
- Testare le prestazioni del modello sui dati di Twitter

E' necessario considerare il tipo di tweet che si vuole analizzare:

- Tweets attuali: utili per tracciare parole chiave o hashtag in tempo reale.
- Tweet storici: utili per confrontare i sentimenti in periodi diversi.

2. **Pulire i dati utilizzando tecniche di pre-elaborazione:** Una volta raccolti i tweet di cui si necessita per la analisi del sentiment, bisogna preparare i dati. I dati dei social media non sono strutturati e devono essere puliti prima di usarli per addestrare un modello di Sentiment Analysis - dati di buona qualità porteranno a risultati più accurati. La pre-elaborazione di un set di dati di Twitter comporta una serie di compiti come la rimozione di tutti i tipi di informazioni non facilmente trattabili come emoji e/o irrilevanti come caratteri speciali e spazi vuoti extra. Può anche comportare miglioramenti di formato, eliminare i tweet duplicati o i tweet che sono più corti di tre caratteri.
3. **Creare un modello di apprendimento automatico per la Sentiment Analysis:** dopo aver preparato i dati occorre costruire un modello di apprendimento automatico che permetta di implementare la Sentiment Analysis.
4. **Analizzare i dati di Twitter utilizzando il modello di sentiment analysis:** bisogna integrare i dati di Twitter che si vuole analizzare con il modello di Sentiment Analysis creato.

5. **Visualizzare i risultati della vostra sentiment analysis su Twitter:** gli strumenti di visualizzazione dei dati aiutano a spiegare i risultati della sentiment analysis in modo semplice ed efficace. Tra i più noti tool di visualizzazione dei dati si annoverano i seguenti:

- **Google Data Studio:** È possibile utilizzare questa piattaforma gratuita e semplice di Google per creare report interattivi.
- **Looker:** Questa è una piattaforma di analisi dei dati aziendali, creata per gestire tutti i tipi di dati all'interno delle diverse aree di un'azienda. È possibile connettersi con diversi database e creare grafici e tabelle di dati.
- **Tableau:** è un software di business intelligence e analisi, che consente di lavorare con un gran numero di fonti di dati per creare visualizzazioni di dati dinamici convincenti. E' molto facile da usare e non richiede alcuna abilità di codifica.

2.4 Tools per Sentiment Analysis

Uno strumento di Sentiment Analysis è un software che analizza le conversazioni.

Tali software usano varie tecnologie come l'apprendimento automatico, l'elaborazione del linguaggio naturale, l'analisi del testo e l'analisi sociale per eseguire l'analisi del sentiment su grandi collezioni di testo che possono provenire da post di social media, gruppi di discussione online, portali di notizie, recensioni online, ecc.

Ci sono diversi software per la sentiment analysis, di seguito si elencano i principali:

- **HubSpot di Service Hub:** Gli strumenti Service Hub di HubSpot includono uno strumento di feedback dei clienti che può scomporre le risposte qualitative dei sondaggi e valutarle per l'intento positivo o negativo. Utilizza i sondaggi NPS (Net Promoter Score) per chiarire se la recensione di un cliente era buona o cattiva e li organizza in base al loro sentiment. Gli utenti analizzano i risultati guardando una dashboard completa che include grafici e tabelle che forniscono una panoramica della soddisfazione del cliente. Vi è una versione gratuita con funzionalità ridotte e una versione estesa a pagamento.

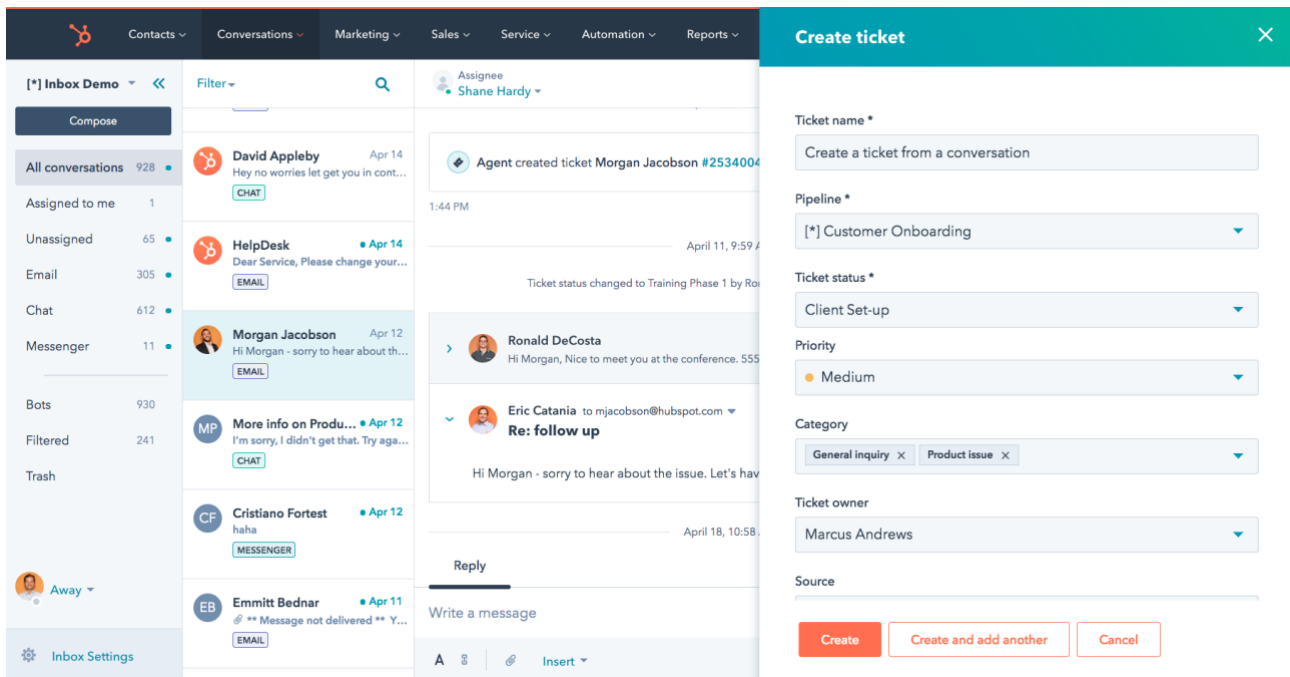


Figura 1. HubSpot di Service Hub

- **Quick Search di Talkwalker:** è un motore di ricerca a pagamento che offre una copertura ampia sui social network, inclusi Facebook, siti di notizie, blog e forum e offre una panoramica immediata sul brand permettendo di avere informazioni sulla sfera social dei concorrenti di un'azienda, permettendo di sfruttare i loro punti di forza e di debolezza per migliorare e coltivare una strategia di marketing vincente. Include le seguenti features:
 - **Panoramica rapida sul brand:** fornisce una panoramica dei KPI (Key Performance Index) che contano come volume, sentiment, demografia e geolocalizzazione;
 - **Trovare trends in tempo reale:** identificare i trends emergenti per migliorare i propri contenuti aziendali;
 - **Ideazione del contenuto:** creare contenuto virale che risulti “popolare” con il pubblico a cui si rivolge un prodotto;
 - **Competitive intelligence:** confronta diversi brand con la concorrenza.

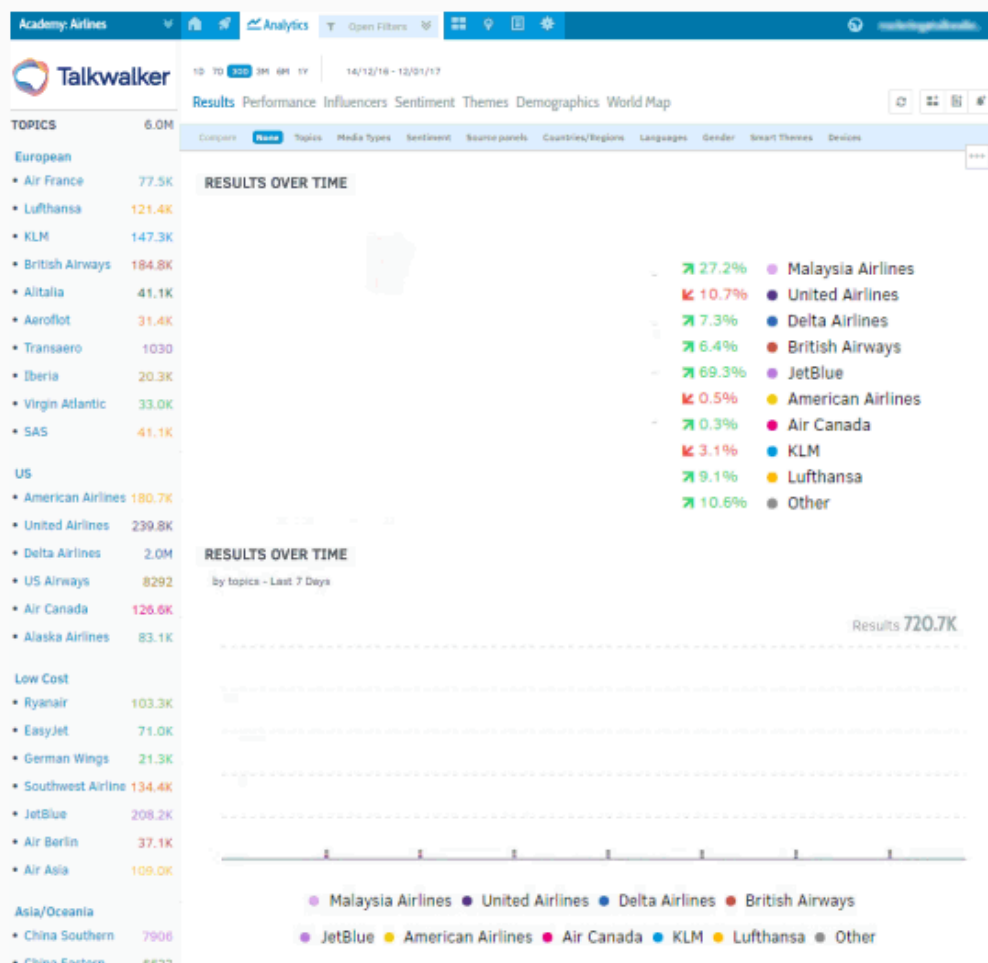


Figura 2. Talwalker



Figura 3. Quick Search di Talkwalker - McDonald's vs Burger King confronto tra i due grandi produttori di hamburger.

- **Hootsuite Insights:** Analizza automaticamente tutte le piattaforme di social media, i siti di notizie, i forum e i blog per rilevare insights che includono influenze, storie, tendenze e sentiment. E' anche possibile controllare le conversazioni online che avvengono su siti di notizie, blog e forum. La funzione Insight può inoltre accedere ai dati in tempo reale ad oltre 100 milioni di fonti in più di 50 lingue. Vi è una versione gratuita di 30 giorni e poi è possibile usare le features offerte a pagamento.

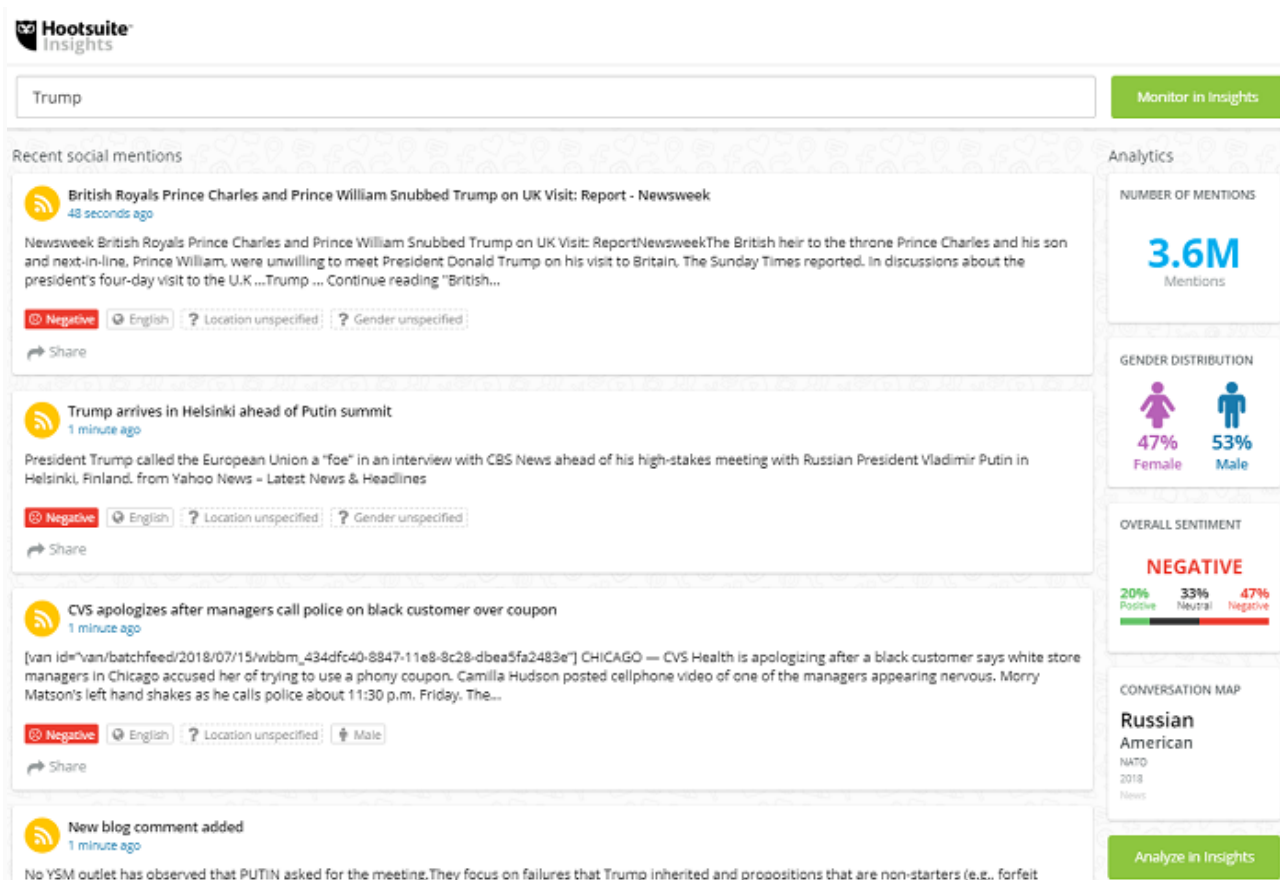


Figura 4. Insights di Hootsuite

- **Rapidminer:** Si tratta di una piattaforma di analisi di science data che fornisce l'analisi del testo per aiutare i brand ad ottenere l'analisi del sentiment. Esso unisce in un solo tool la data preparation, machine learning e predictive model deployment. Con questo tool è possibile analizzare le recensioni online ed i social media post, nonché le pubblicazioni ufficiali ed i documenti. I brand possono identificare gli argomenti di tendenza che circolano tra i consumatori e i clienti, raccogliere feedback sui lanci di prodotti e trovare nuove aree

per l'espansione del business. Vi è una versione gratuita di 30 giorni e poi è possibile usare le features offerte a pagamento.

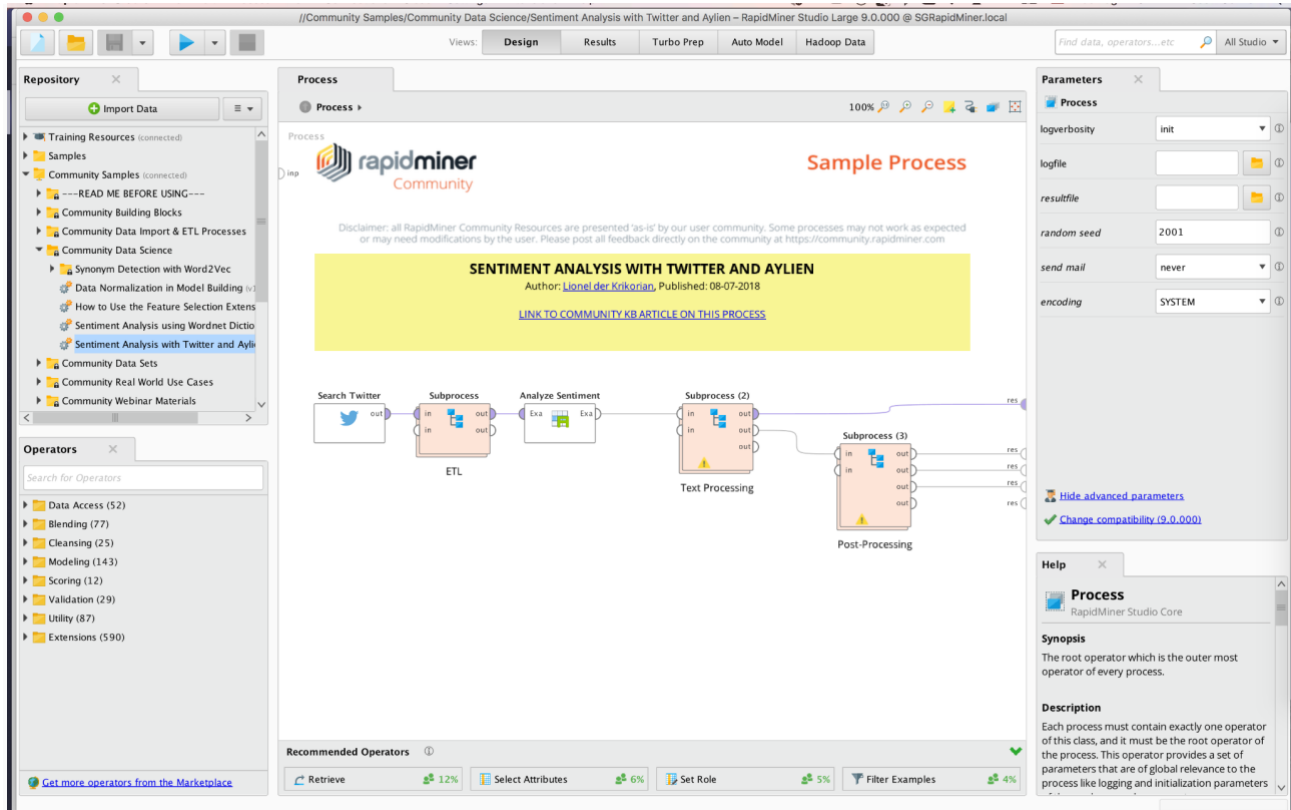


Figura 5. Rapidminer

eaningCloud: si tratta di un API (Application Programming Interface) di sentiment analysis in gradi di implementare un'analisi multilingua e dettagliata di contenuto proveniente da fonti diverse, determinando se mostra un sentiment positivo, negativo o neutrale. Le frasi sono identificate in base alla loro correlazione, il che permette di identificare il valore di popolarità globale del testo. Include le seguenti Feature:

- **Sentiment globale:** opinione generale espressa in tweet, post sul blog, recensioni.
- **Sentiment a livello di attributo:** analizza il sentiment specifico di ogni frase
- **Identificazione di opinioni e fatti:** distingue tra oggettivo e soggettivo.
- **Rilevamento dell'ironia:** identifica i commenti in cui il sentiment è opposto rispetto a quanto scritto.

- **Polarità:** valutata da *“molto negativo”* a *“molto positivo”*.
- **Accordo e disaccordo:** individua opinioni opposte, contraddittorie o ambigue.

Questo strumento di analisi del *sentiment* permette agli utenti di caricare dizionari personalizzati da utilizzare nell'analisi e nella classificazione, e c'è anche un add-on Excel per Windows. Le lingue attualmente incluse sono inglese, spagnolo, francese, italiano, catalano e portoghese. Vi è una versione gratuita di 30 giorni e poi è possibile usare le features offerte a pagamento.

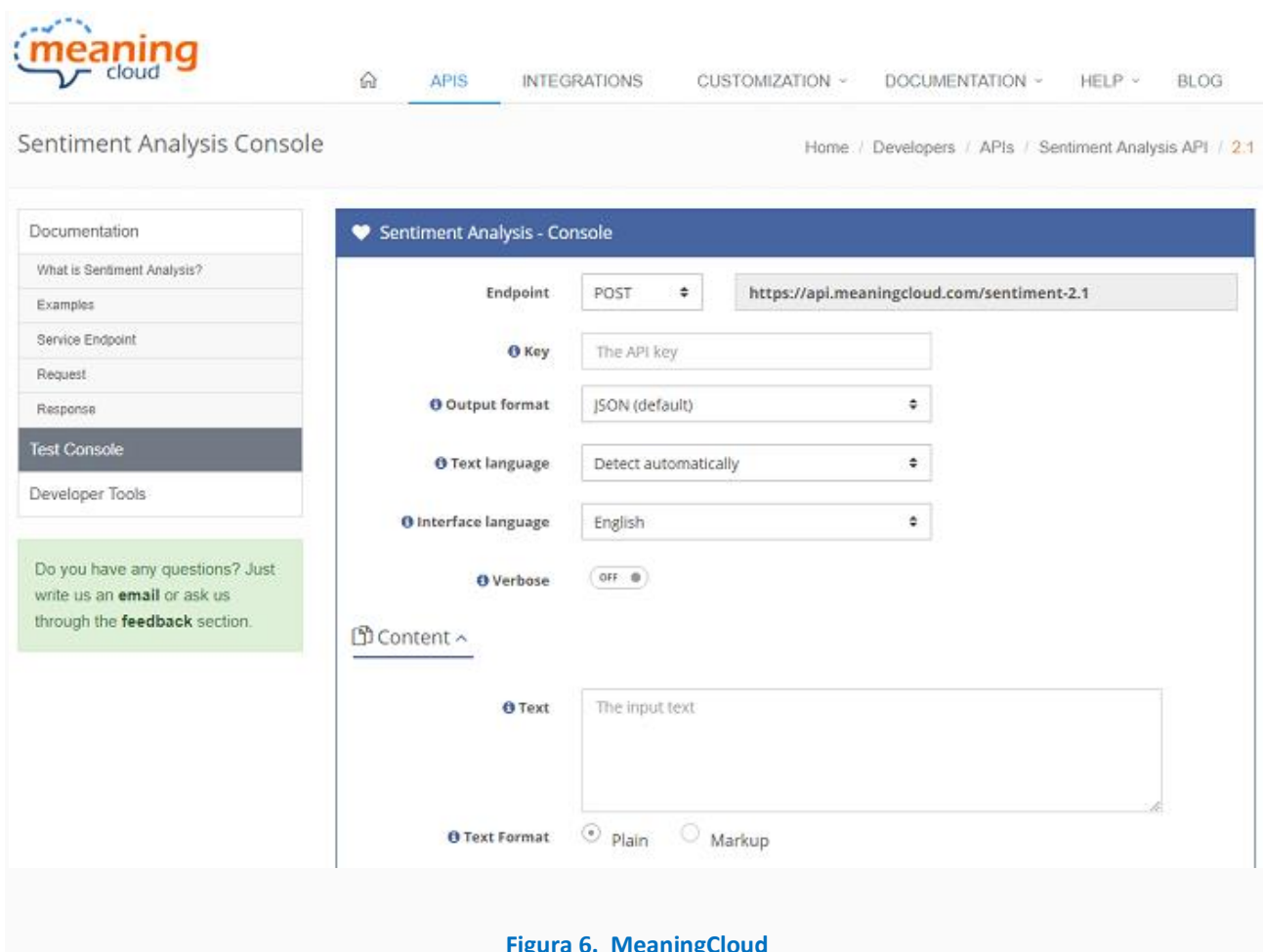


Figura 6. MeaningCloud

● **ocial Mention:** Si tratta di una piattaforma di ricerca in tempo reale che, pur essendo

S

piuttosto base, monitora oltre 100 piattaforme social tra cui Twitter, Facebook, FriendFeed, YouTube e Digg+; oltre che blog e siti di notizie. Aggrega inoltre i contenuti generati dagli utenti, permettendo di monitorare e misurare ciò che i consumatori dicono del tuo *brand*. E' uno strumento gratuito di analisi dei social media facile da usare, in quanto gli utenti non devono creare un account o scaricare un software, bisogna solo navigare sul sito e cercare una parola chiave (il proprio brand, l'industria, il concorrente, etc) come si farebbe con qualsiasi motore di ricerca. Dopo aver inserito la ricerca, Social Mention tira fuori i dati sulla parola chiave da ogni sito di social media e li mette insieme in un report completo in cui classifica i risultati in base al sentiment (positivo, negativo, neutro) alle parole chiave più importanti, ai migliori utenti, ai migliori hashtag e alle fonti. Lo strumento è gratuito. All'interno dello strumento verrà visualizzato un numero limitato di risultati di ricerca e, se si desidera monitorare a fondo il proprio marchio, è possibile avviare un abbonamento completo.

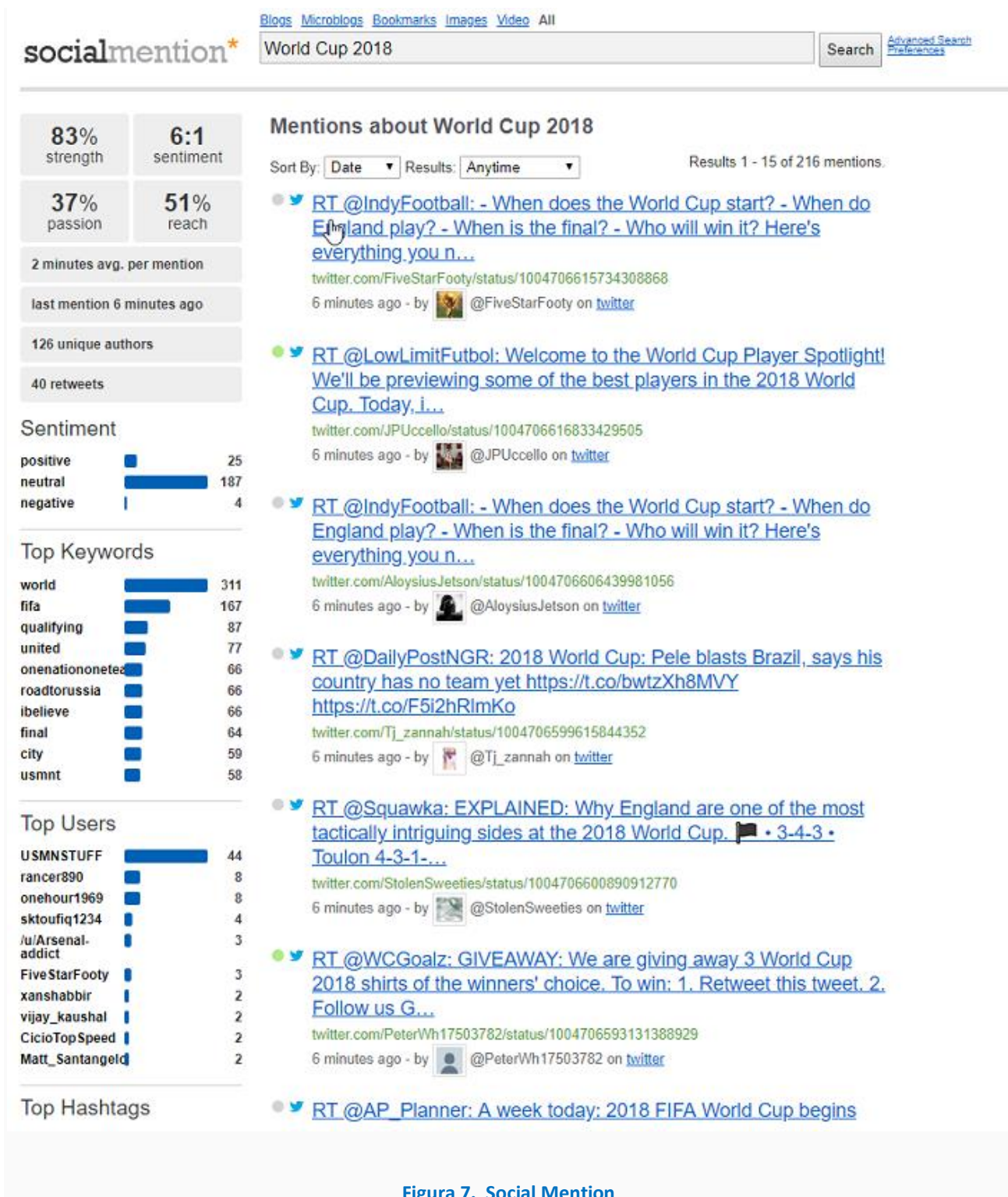


Figura 7. Social Mention

- Sentiment Analyzer:** Questo strumento di Sentiment Analysis utilizza la linguistica computazionale e l'analisi del testo per identificare il sentiment dietro il linguaggio scritto. E' gratuito.

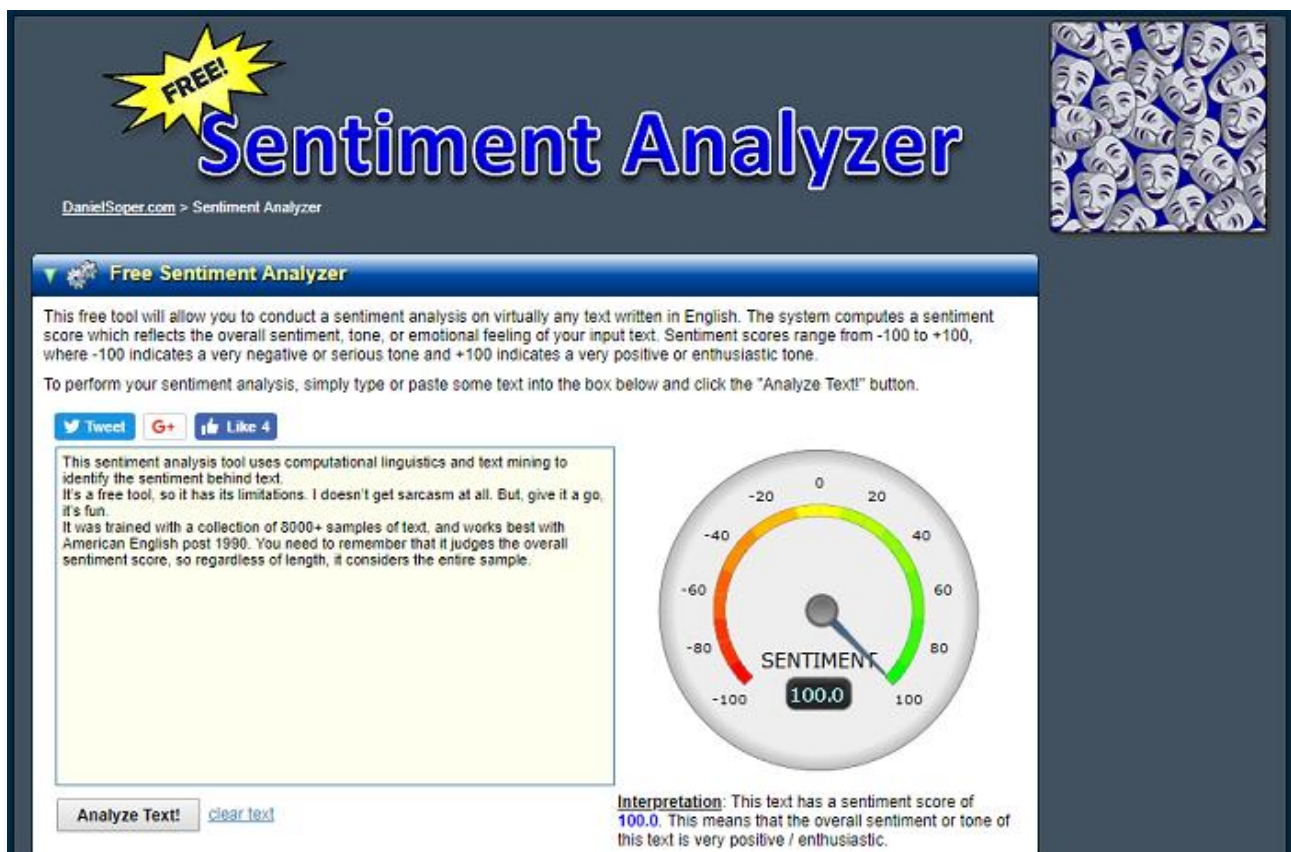


Figura 8. Sentiment Analyzer

• **entiStrength:** è uno strumento gratuito di Sentiment Analysis ideato per la ricerca accademica. Anche se è possibile richiedere la versione commerciale a pagamento Java o la licenza per la versione online. Funziona solo su Windows. Offre un'analisi automatica del *sentiment*. Analizza fino a 16.000 testi al secondo nei *social* e nel web, con un'accuratezza a livello umano per l'inglese.



Quick Tests (English version):

Enter text:

Output: ☒ Dual, ☐ binary, ☐ trinary, ☐ scale

Keyword test:

Enter keywords (comma-separated list, no spaces):

Topic test:

Select domain (broad topic):

Other languages: [Finnish](#), [German](#), [Dutch](#), [Spanish](#), [Russian](#), [Portuguese](#), [French](#), [Arabic](#), [Polish](#), [Persian](#), [Swedish](#), [Greek](#), [Welsh](#), [Italian](#), [Turkish](#).

Figura 9. SentiStrength

• **Sentiment**: si tratta di uno strumento di Sentiment Analysis facile da usare, per i documenti in lingua inglese o blocchi di testo. Attualmente è in fase beta, di conseguenza potrebbe avere dei bug. Analizza velocemente grandi quantità di testo in pochi secondi e calcola il sentiment del contenuto assegnando un valore positivo, negativo o neutro. E' possibile richiedere una demo gratuita.

Sentigem BETA
Sentiment Analysis API

Wow! Easy to use sentiment analysis tool for English language documents or blocks of text. Sentigem is currently in beta phase, so may be buggy. But, it's worth persevering.

**A recent study by the Zurich University of Applied Sciences credited us as one of the world's foremost Sentiment Analysis engines. [[1](#)] [Read the full research paper](#)

© 2013 Sentigem • [About](#) • [Contact](#) • [FAQ](#) • [Developers' API](#)

Figura 10. Sentigem

- **Repustate:** si tratta di uno strumento online per le Sentiment Analysis con una sofisticata API di analisi del testo che valuta accuratamente il sentimento dietro le risposte dei clienti. Fornisce informazioni in 23 lingue diverse. Questo software può anche valutare il sentimento che sta dietro al gergo (come FYI, bc, tbh) e anche emojis per determinare se il sentimento dietro un messaggio è negativo o positivo. Inoltre è possibile personalizzare l'API in modo che identifichi un linguaggio specifico, e a riconoscere i significati alternativi delle parole. E' possibile usare lo strumento con una prova gratuita, poi si inizia con 99 dollari al mese.

SENTIMENT-ANALYSIS

SENTIMENT ANALYSIS

Sentiment

Extract the sentiment from a piece of text. Scores range from **-1** (negative) to **1** (positive) with a score of **0** being "neutral". Emoticons, emoji, and internet short forms are given greater weight in the algorithm.

REQUEST BODY SCHEMA: multipart/form-data

Parameter	Type	Description
text	string	A block of text to analyze. The sentiment score will be for the entire block of text.
lang	string	The two letter code of the language you want to analyze the sentiment in. The default is English (en); you do not need to specify anything if you're just scoring English text.
emoji	integer	If you'd like to ignore emoji when calculating sentiment, pass '0' as a value. Any other value or omitting this argument will result in emoji being considered when determining sentiment.

Responses

- > 200 OK
- > 400 error

Request samples

POST /score.json

Payload CURL

Content type: multipart/form-data

Copy

text=This is a great day to go for

Response samples

200 400

Content type: application/json

Copy Expand all Collapse all

```
{
  "score": 0.9229105562880748,
  "status": "OK"
}
```

Figura 11. Repustate

- **MonkeyLearn:** è uno strumento di Sentiment Analysis che è facile da personalizzare. Basta creare tag di categorizzazione e poi evidenziare manualmente diverse parti del testo per mostrare quale contenuto appartiene ad ogni tag. Con il tempo, il software impara da solo e

può elaborare più file contemporaneamente. MonkeyLearn fornisce anche ai suoi clienti uno strumento gratuito "Word Cloud" che dice loro quali parole sono usate più frequentemente all'interno di ogni tag di categorizzazione. Questo può aiutare le aziende a scoprire i blocchi stradali comuni dei clienti, cercando menzioni ripetute di prodotti o servizi specifici. Se notate che un prodotto è costantemente elencato sotto un tag di categorizzazione negativo, questo suggerisce che c'è un problema con quel prodotto di cui i clienti non sono soddisfatti. Si tratta di una piattaforma di apprendimento automatico di facile utilizzo per l'analisi del testo che utilizza un modello pre-trained sentiment analysis che è possibile testare anche in modalità gratuita. MonkeyLearn può essere integrato con software ed App come Excel, Fogli di Google, Zapier o Zendesk, permettendo così agli utenti di integrare il suo strumento di analisi dei sentimenti con qualsiasi software utilizzato regolarmente. Per chi fosse alla ricerca di un sistema di analisi personalizzato ha la possibilità di scegliere l'algoritmo e anche i parametri che definiscono il modello grazie ad una etichettatura dei testi molto intuitiva e veloce. Utilizzando le statistiche delle prestazioni è possibile eseguire il debug del modello una volta che è pronto, fornendo (anche ai propri clienti) risultati affidabili e precisi in poco tempo. Lo strumento include un piano gratuito. I piani premium partono da 299 dollari al mese.

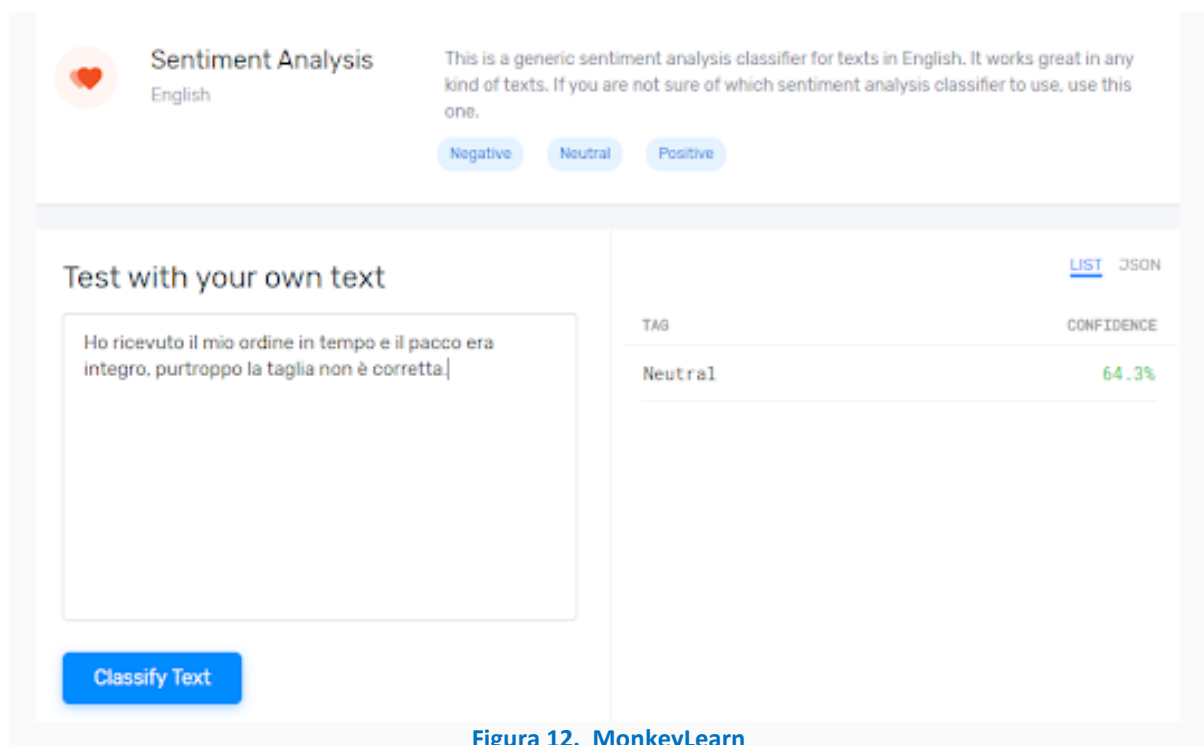


Figura 12. MonkeyLearn

- **IBM Watson:** è una piattaforma multi-cloud che offre molte API per la Sentiment Analysis basata su NLP. Il Watson Tone Analyzer, ad esempio, si concentra sui ticket di supporto e sui sondaggi di soddisfazione dei clienti, permettendo così agli operatori di capire se gli agenti sono educati e desiderosi di aiutare i clienti e se hanno veramente risolto il problema dell'utente finale. IBM Watson ha un approccio completo all'analisi dei testi, abbastanza flessibile da soddisfare le esigenze di qualsiasi cliente, indipendentemente dal settore o dal campo in cui si opera. Lo strumento include un piano gratuito. I piani premium dipendono dal numero di previsioni necessarie.

IBM Watson Developer Cloud



Tone Analyzer

This service uses linguistic analysis to detect joy, fear, sadness, anger, analytical, confident and tentative tones found in text.

*This system is for demonstration purposes only and is not intended to process Personal Data. No Personal Data is to be entered into this system as it may not have the necessary controls in place to meet the requirements of the General Data Protection Regulation (EU) 2016/679.

By using this application, you agree to the [Terms of Use](#)

Resources:

[Documentation](#)

[API Reference](#)

[Fork on Github](#)

[Start for free in IBM Cloud](#)

Sample use cases

Choose an example to learn how you can adjust the tone of your content to change people's perceptions, or improve its effectiveness. [Learn more](#).

☒ Tweets
 ☐ Online Review
 ☐ Email message
 ☐ Product Review in French
 ☐ Your own text

Analyzing Customer Engagement Data? Try out the [Tone Analyzer Customer Engagement Endpoint](#).

I hate // this phone company products, you'd have to torture me to get me to use // this phone.

The emojis in #ThisPhone are stupid.

#ThisPhone is a useless, stupid waste of money.

#ThisPhone is the worst phone I've ever had - ever 😞.

#ThisPhone another ripoff, lost all respect SHAME.

I'm worried my #ThisPhone is going to overheat like my brother's did.

#ThisPhoneCompany really let me down... my new phone won't even turn on.

Analyze

Figura 13. IBM Watson

- **Lexalytics:** offre uno strumento di analisi del testo che si concentra sulla spiegazione del perché un cliente sta rispondendo alla tua azienda in un certo modo. Utilizza l'elaborazione

del linguaggio naturale per analizzare il testo, poi esegue una Sentiment Analysis per determinare l'intento dietro il messaggio del cliente. Infine, Lexalytics conclude il processo compilando le informazioni che ricava in una visualizzazione facile da leggere e da condividere. Mentre la maggior parte degli strumenti di analisi del sentiment dice come si sentono i clienti, Lexalytics si differenzia in quanto comunica perché i clienti si sentono in quel modo eseguendo l'analisi delle emozioni per valutare i sentimenti e gli stati d'animo che si "nascondono" dietro le parole dei clienti. Una volta terminata l'analisi del sentiment, lo strumento fornisce una serie di risultati visivi facilmente interpretabili. Con la demo è possibile testare gratuitamente il servizio, se lo si vuole continuare ad usare è a pagamento.

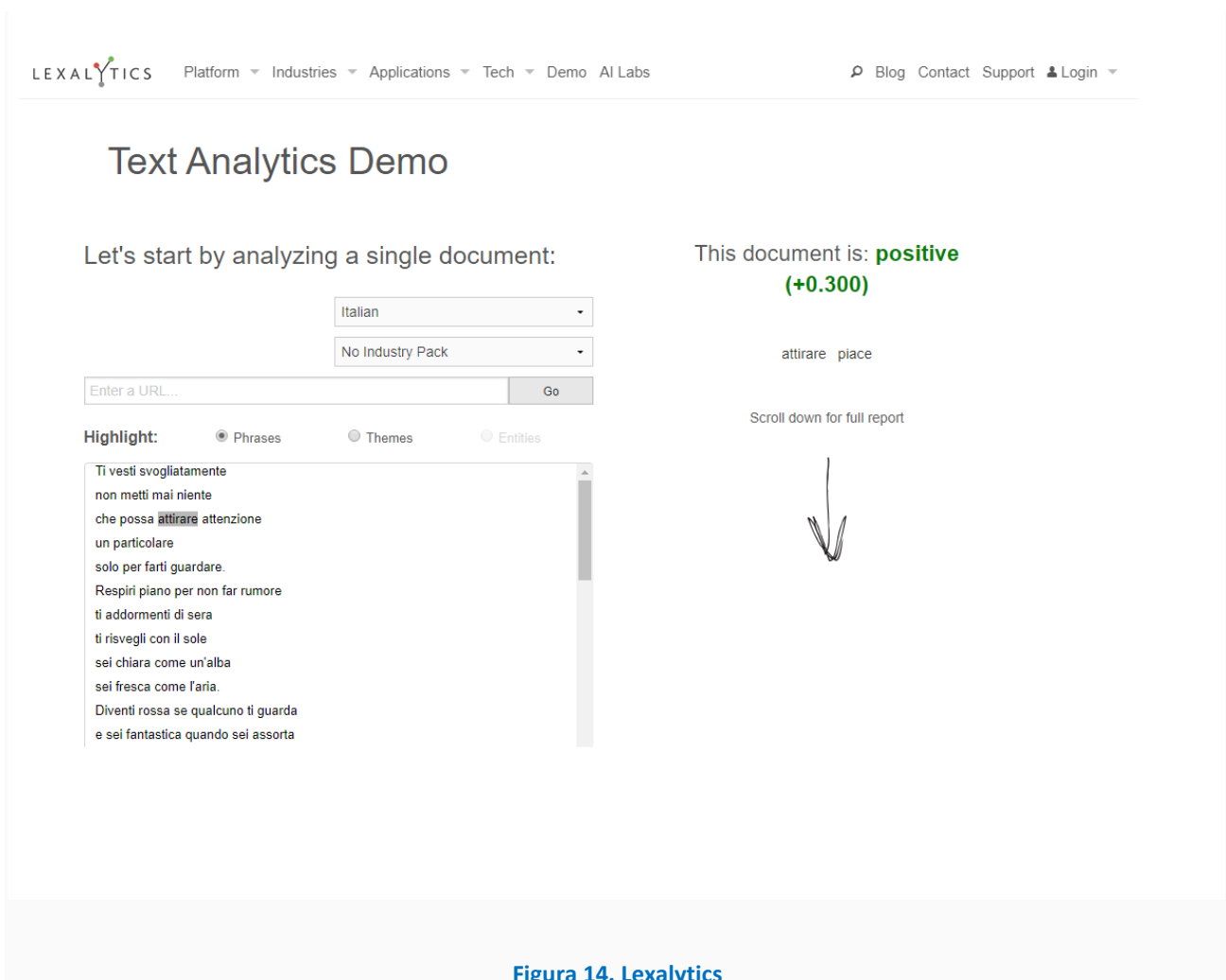


Figura 14. Lexalytics

- **Rosette**: è una API che usa l'IA (intelligenza artificiale) per analizzare il linguaggio naturale. È stata utilizzata dapprima per eseguire l'analisi del sentiment sui social media, ma poi ha trovato la sua naturale evoluzione nell'analisi di interi documenti e di singole entità

menzionate nel testo, ad esempio il sentiment espresso dai clienti quando citano un prodotto, un'azienda o brand o una persona specifica. Rosette è in grado di identificare parti del discorso attraverso l'analisi morfologica e la lemmatizzazione (il raggruppamento di forme di parola inflesse in modo che non vengano analizzate separatamente) andando ad identificare i sentimenti fino a 30 lingue diverse. E' ottimo per le aziende internazionali perché può rivedere i dati basati sul testo in oltre 30 lingue diverse. Questo significa che un'azienda non deve tradurre le conversazioni prima di caricarle, il che non solo è più veloce, ma assicura una maggiore precisione. Poiché la maggior parte dei clienti userà stenografia o gergo, gli strumenti di traduzione di terze parti possono inavvertitamente cambiare il significato del loro testo. Con Rosette, il sistema è costruito per analizzare il testo nella lingua in cui è scritto, quindi non perderai nessun feedback prezioso anche se è scritto in modo informale.

E' a pagamento.

The screenshot displays the Rosette Text Analytics web interface. At the top, the Rosette logo and navigation links (PRODUCT, CUSTOMERS, NLP BLOG, ABOUT, LOGIN) are visible. The main content area shows a sample text in Spanish with sentiment markers (green for positive, red for negative) above certain words. Below the text, a 'Key' section indicates the sentiment colors. To the left, a 'Person' entity list identifies names like Morten Tyldum, Jennifer Lawrence, Michael Sheen, Laurence Fishburne, Real Humanes, Chris Pratt, Andy Garcia, and Thomas Newman. To the right, an 'Organization' entity list identifies 'The Imitation Game'. Below that, a 'URL' entity list shows a link to a film affinity review. At the bottom right, a 'Title' entity list identifies the word 'director'.

Figura 15. Rosette

- **Critical Mention:** è un tool diverso da quelli elencati in precedenza perché analizza le notizie e le altre pubblicazioni che fanno riferimento ad un'azienda. L'analisi del sentiment di Critical Mention si concentra sul sentiment che si trova non nei post sui social media, ma in articoli di notizie e altre fonti di business online, tra cui TV e fonti in onda, una caratteristica rara. Questo avvisa l'organizzazione dei cambiamenti nella percezione del marchio, non solo dai sentimenti dei clienti, ma da quelli degli analisti aziendali e di altri osservatori del settore. Quelle informazioni possono fornire una prospettiva potente che sarebbe di grande valore strategico nella gestione del marchio Critical Mention può persino cercare file video e trasmissioni in diretta, può anche avvisare delle storie che appaiono in televisione. Il suo reporting grafico di alta qualità è eccezionale tra i prodotti della sua categoria. Uno svantaggio è che gli strumenti di analisi del sentiment sono limitati. Lo strumento è a pagamento ed è possibile richiedere una demo gratuita.



Figura 16. Critical Mention

- **Brandwatch:** è uno strumento di Sentiment Analysis che si concentra sulla gestione del marchio infatti fornisce una feature di "image insights" che può identificare le immagini associate ad un marchio. Per esempio, se si carica un'immagine del logo del marchio di un'azienda. Brandwatch naviga nel web per le immagini che includono quel logo. Poi,

compila le immagini in una lista ed evidenzia esattamente dove il logo del marchio caricato appare. Inoltre, il software di Brandwatch fornisce interessanti approfondimenti su ogni immagine che trova. Questo include metriche come il volume delle menzioni, i seguaci aggregati e l'ultima attività. È in grado non solo di scansionare le menzioni di testo di un marchio online, ma anche di immagini del marchio, ad esempio, gli usi del logo e altre opere d'arte associate al marchio. Tracciando queste apparenze, è possibile analizzare come si comporta un marchio con un pubblico target particolare sui social media. Molti osservatori del settore considerano Brandwatch lo strumento principale per il monitoraggio parallelo del marchio su tutti i canali dei social media. Gli svantaggi includono i tempi di ritardo segnalati dagli utenti per query di grandi dimensioni e la mancanza di download di dati in blocco. Lo strumento prevede una prova gratuita di 30 giorni e al seguito è a pagamento.

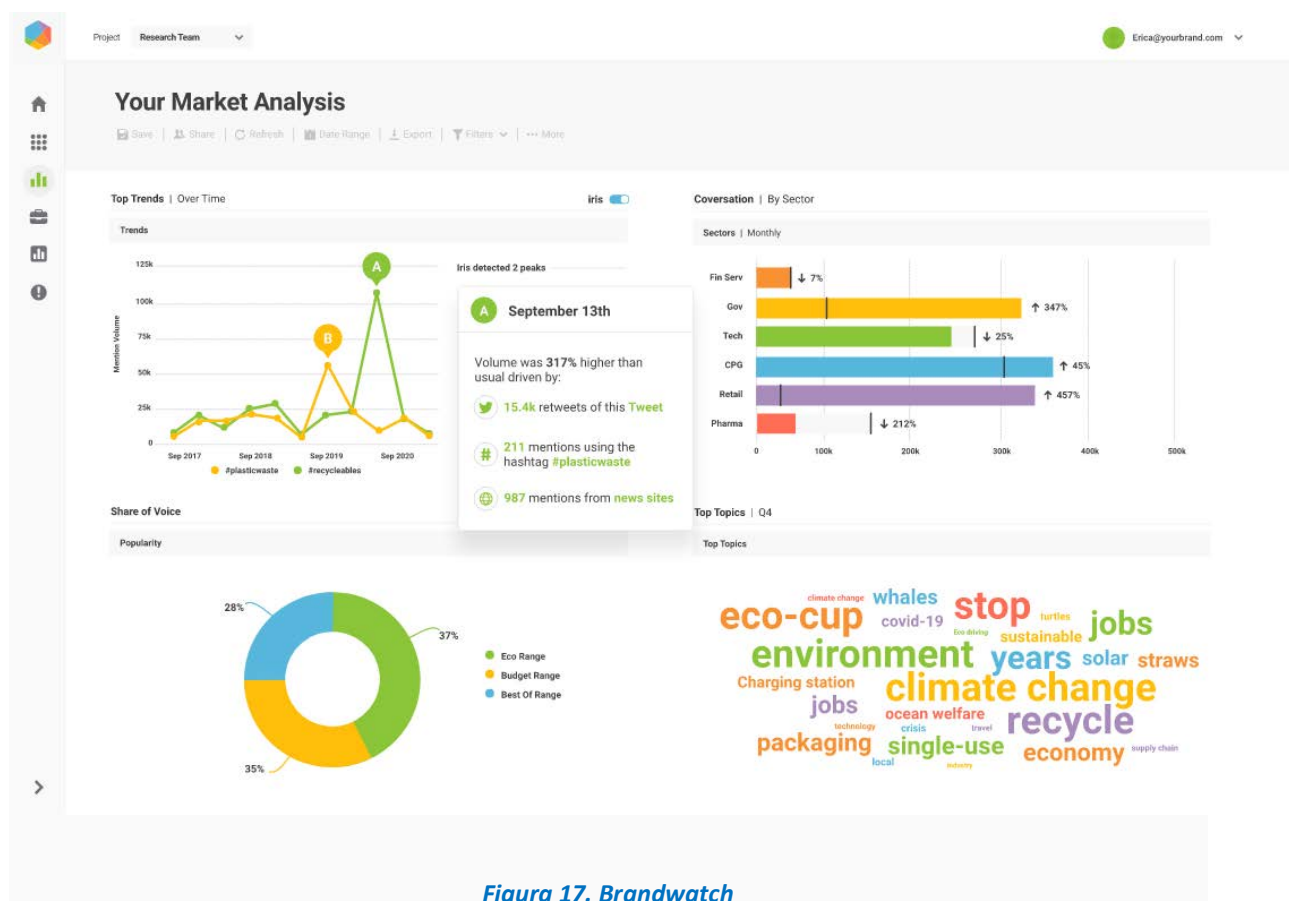


Figura 17. Brandwatch

- **Clarabridge:** è uno strumento di Sentiment Analysis che a differenza di quelli visti in precedenza si concentra anche sulla Speech Analytics, cioè sulla Sentiment Analysis su dati audio. Ciò è particolarmente utile per le aziende che si affidano ai call center come canale

per vendere o fornire assistenza ai clienti. Questo software online non solo analizza il discorso del chiamante, ma anche il suo tono di voce e i sottili spunti per interpretare il sentimento. Calabridge ha una soluzione per la gestione dell'esperienza del cliente che include uno strumento che crea metriche sullo stato d'animo dei clienti analizzando le e-mail, chatbot aziendali e sondaggi. Inoltre, combina approcci lessicali e grammaticali per effettuare la Sentiment Analysis di ogni frase all'interno di un testo. Può essere chiesta una demo e ottenere un preventivo su misura.



Figura 18. Calabridge

- **Aylien:** è un altro strumento online che sblocca il valore nascosto dei testi eseguendo la Sentiment Analysis e classificandoli in Positivo o Negativo, o in Soggettivo e Obiettivo. La piattaforma di analisi testuale permette anche di costruire il proprio modello senza problemi, e non c'è bisogno di sapere molto sull'apprendimento automatico o sulla NLP per iniziare. E' a pagamento.

Alcuni Tool specifici per la Twitter Sentiment Analysis sono:

- **IntenCheck:** è una piattaforma di analisi del testo basata sul cloud che offre una funzione di Sentiment Analysis per analizzare i testi dai post dei social media, e-mail e comunicazioni dei dipendenti su qualsiasi canale. Per eseguire l'analisi del sentiment di Twitter, è sufficiente caricare i tweet e i post nello strumento e sarete in grado di classificare i sentimenti (come sentimenti negativi e positivi) e le emozioni (come la rabbia o il disgusto) e tracciare eventuali insincerità presenti nei tweet. Lo strumento permette anche di approfondire se i tweet parlano del passato, del presente o del futuro, e se sono inclinati a favore o contro il marchio. Può essere chiesta una demo e ottenere un preventivo su misura.

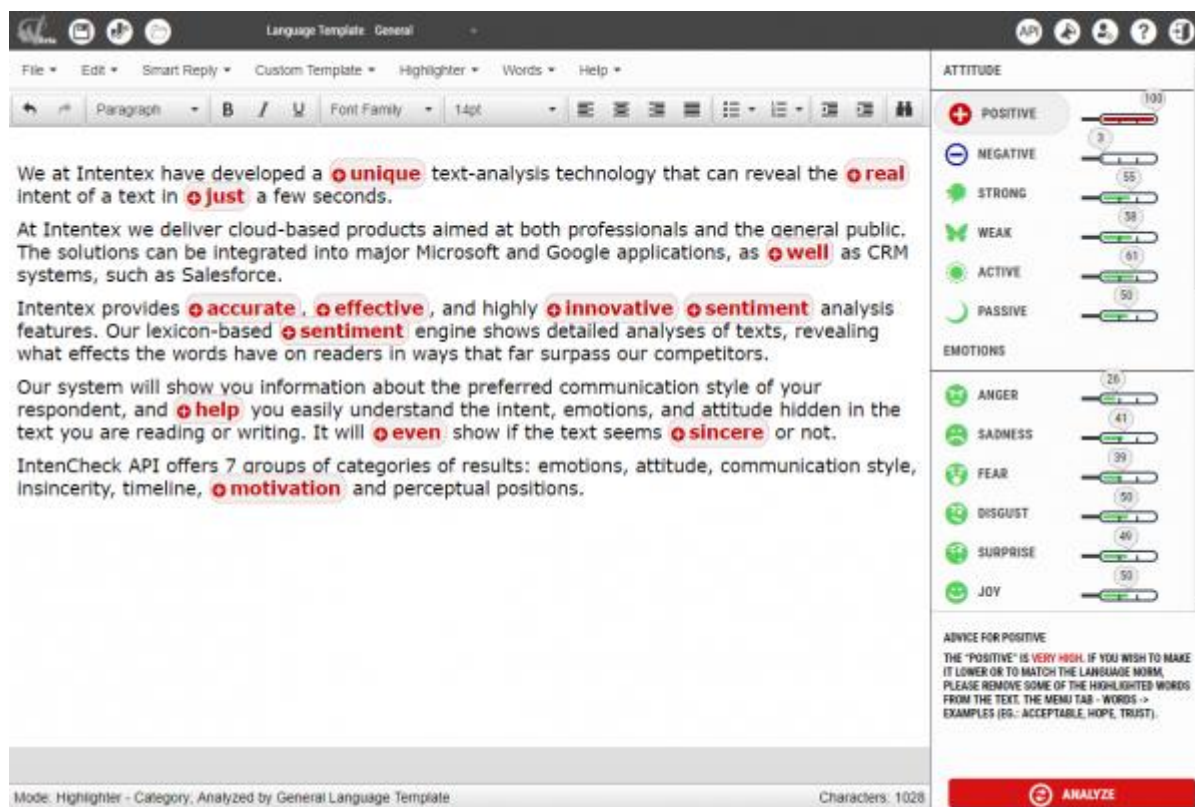


Figura 19. IntenCheck

- **Social Searcher:** è uno strumento gratuito di monitoraggio dei social media che permette di eseguire la Sentiment Analysis e la ricerca in tempo reale di menzioni di particolari parole chiave attraverso le piattaforme di social media (Facebook e Twitter). Basta caricare i dati di Twitter (come le menzioni, i messaggi diretti e i tweet) nello strumento per analizzare i proprio post su Twitter e i tweet dei seguaci o potenziali clienti dell'azienda in esame. Lo strumento aiuta a generare un punteggio di sentiment, a evidenziare i post che stanno ricevendo i sentimenti più positivi o negativi, e a controllare il sentiment popolare verso un

marchio o prodotto cercando il nome di un prodotto o marchio. Ogni risultato presenta un pulsante il cui colore indica il sentiment positivo, negativo o neutro. Inoltre, suddivide i rapporti per piattaforme di social media, in modo da poter vedere esattamente come il vostro marchio si sta comportando attraverso diverse applicazioni e canali. C'è una versione a pagamento (che si può provare gratuitamente per 14 giorni) ed una versione totalmente gratuita che offre 100 ricerche in tempo reale al giorno e due avvisi via *email*. Include filtri come il tipo di post, il canale social e il sentiment. Include le seguenti Features:

- **Storico illimitato:** accesso diretto allo storico delle menzioni sui social. E' possibile catalogare i risultati per pagina, data, popolarità, canale social, *sentiment* e tipo di contenuto
- **Analisi esaustiva:** distribuzione dei post per settimana, giorno, ora, *sentiment*, link, utenti, *keywords* ed *hashtag*.
- **Riconoscimento della lingua:** analisi del *sentiment* in inglese, tedesco, francese, italiano, portoghese, russo, olandese e spagnolo.
- **Alert email e notifiche:** nuove *menzioni* sui social

The screenshot shows the user interface of a social media monitoring tool. At the top, there is a search bar with a magnifying glass icon and a 'SEARCH SETTINGS' button. To the right of the search bar are two toggle switches: 'email alerts' and 'monitoring'. Below the search bar is a section for 'NAME OF SEARCH' with a text input field labeled 'Search Name'. Underneath this are three tabs: 'KEYWORDS' (highlighted in blue), 'SOURCES', and 'MORE'. The 'KEYWORDS' section contains three columns: 'Keywords' (posts should contain at least one of the keywords below), 'Exact keywords' (posts should contain at least one of the exact phrases below), and 'Minus keywords' (posts can't contain any of the keywords below). Each column has a text input field with a green plus button next to it. The 'Keywords' field contains 'sentiment tools', 'Exact keywords' contains 'sentiment analysis', and 'Minus keywords' contains 'Add new word'. Below these fields is a green 'SAVE' button. The 'Filter Search' section follows, with two columns: 'Post Types' and 'Sentiment'. Under 'Post Types' are checkboxes for 'Link', 'Photo', 'Status', and 'Video'. Under 'Sentiment' are checkboxes for 'Positive', 'Negative', and 'Neutral'. At the bottom is the 'Sources' section, which lists various social media platforms with checkboxes: WEB, TWITTER, FACEBOOK, GOOGLE+, YOUTUBE, INSTAGRAM, REDDIT, DAILYMOTION, TUMBLR, VIMEO, VKONTAKTE, and FLICKR.

Figura 20. Social Searcher

- **Text2Data:** è una piattaforma di analisi con capacità di monitoraggio dei social media, Sentiment Analysis e analisi del testo. La funzione di Sentiment Analysis è disponibile come parte della sua Text Analysis Platform. Consente di analizzare i sentimenti dei social media utilizzando un plug-in di Microsoft Excel che aiuta a monitorare i sentimenti in tempo reale. Lo strumento offre una modalità dedicata alla Sentiment Analysis di Twitter per condurre la Sentiment Analysis dei tweet, la classificazione del sentimento, la scoperta di temi nascosti, l'analisi di parole chiave specifiche e il rilevamento dello slang. E' a pagamento.

The image shows the Text2Data Excel plugin interface. The top ribbon includes tabs for FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, and TEXT2DATA. The TEXT2DATA tab contains several groups of icons: My Account, Service Credentials, Service Settings, Blacklisted Words, Analysis (with checkboxes for Show/hide Analysis Pane, Twitter mode, Generate Summary, Detect Entities, Detect Themes, Detect Facets), Analyse selection, and Online Help. Below the ribbon, a table displays the results of the sentiment analysis.

Source text	Sentiment	Sentiment Score	Summ...
Great service.	positive	0.2495	
I am very happy with the service I get, a...	positive	0.21913193135637998	
The service was absolutely superb and ...	positive	0.23936080258442033	
I have made a claim because unfortuna...	neutral	0.1672446526893466	
Well the ladies that I spoke to gave me ...	positive	0.18130362977623685	
I was not happy with the service I recei...	negative	-0.2495	

Below the table, a screenshot of the Microsoft Excel interface shows the data being imported. A context menu is open over the first row of the table, with the 'Analyse | TEXT2DATA' option highlighted. The menu also includes options for Cut, Copy, and Paste Options.

Figura 21. Text2Data

- **NCSU Tweet Visualizer - Sentiment Viz:** Si tratta di uno strumento gratuito per la sentiment analysis di Twitter. Basta inserire una parola chiave e NCSU (North Carolina State University) Tweet Visualizer fornirà i tweet relativi dalla settimana precedente. L'intervallo di tempo è tuttavia più breve per i soggetti più popolari. Posizionando il cursore sopra un punto è possibile visualizzare i singoli tweet degli utenti di Twitter e verificare dove appaiono nello spettro del *sentiment*. Si può scegliere di visualizzare:

- **Sentiment:** l'emozione nel tweet è stimata; il colore blu indica un sentiment negativo, laddove il verde è positivo
- **Argomenti:** permette di raggruppare i tweet nei relativi argomenti utilizzando algoritmi basati sul machine learning
- **Heatmap** - si tratta di una griglia che mappa i tweet; il rosso indica un numero superiore alla media, il blu un numero inferiore

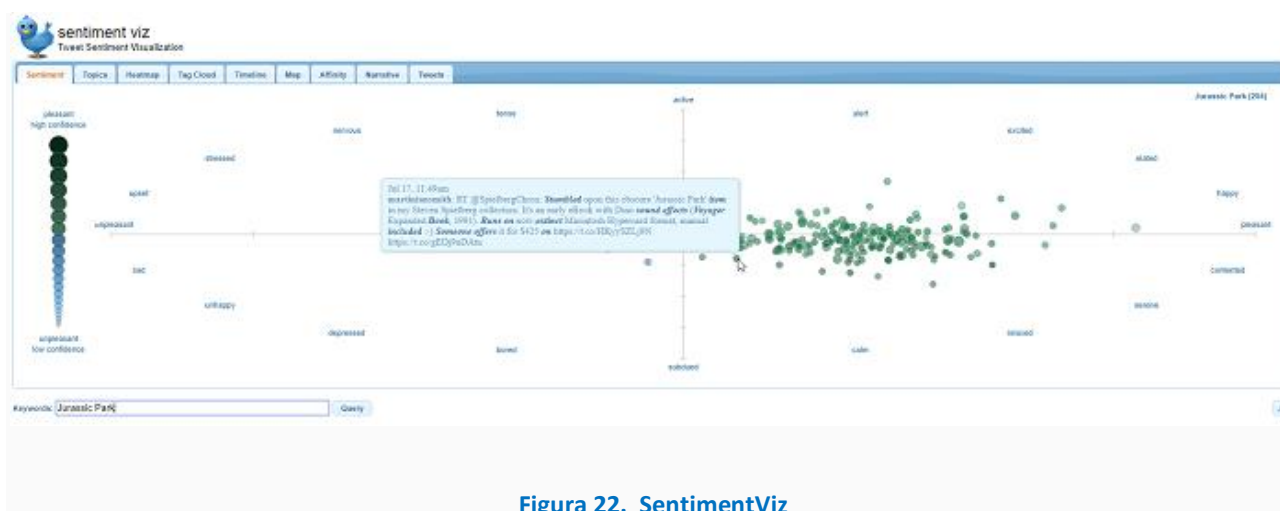


Figura 22. SentimentViz

● 3 Data Analytics

3.1 Che cos'è l'analisi dei dati e perchè è importante

La scienza dei dati, o data science, interessa sempre più ogni settore della nostra società, non ha caso da molti definita ormai come **data-driven**. In misura sempre maggiore, infatti, le organizzazioni utilizzano la data science per trasformare i dati in un vantaggio competitivo, ridefinendo prodotti e servizi e prendendo decisioni in modo mirato.

La Data Analytics è un comparto della data science che consiste nell'adozione e nell'utilizzo di tecniche analitiche avanzate rispetto a data set molto estesi e diversificati (Big Data), che possono includere dati strutturati, semi-strutturati e non strutturati, provenienti da sorgenti eterogenee e di diverse dimensioni.

Big data è un termine consolidato ed utilizzato per riferire data set la cui dimensione o tipologia supera la capacità dei tradizionali database relazionali nel catturare, gestire ed elaborare i dati con latenze basse. I big data possiedono alcune caratteristiche distintive tra cui, elevato volume, elevata velocità o estrema varietà. L'Intelligenza Artificiale (AI), la tecnologia mobile, i social media e l'IoT (Internet of Things) hanno trasformato la complessità dei dati, dando luogo a nuove forme e fonti di dati.

I big data sono generati da sensori, dispositivi, video/audio, reti, file di log, applicazioni transazionali, web e social media e la maggior parte di essi viene generata in tempo reale e su vastissima scala. L'analisi dei big data consente agli analisti, ai ricercatori e agli utenti di business di prendere decisioni in modo più accurato e veloce, utilizzando dati precedentemente inaccessibili o inutilizzabili. Le organizzazioni possono utilizzare tecniche di advanced analytics, quali ad esempio **analytics di testo, machine learning, predictive analytics, data mining, statistiche ed elaborazione del linguaggio naturale**, per ottenere nuovi insight da origini dati precedentemente non sfruttate, in modo indipendente o insieme ai dati aziendali esistenti.

Alcuni casi d'utilizzo dell'applicazione della data analytics dei big data, oltre a quelli citati, sono¹:

- **Rilevamento e riduzione delle frodi:** Monitoraggio delle transazioni in tempo reale, per il riconoscimento preventivo di schemi e comportamenti anomali che possono essere indicatori di attività fraudolente.
- **Incremento di efficienza delle supply chain:** Raccolta ed analisi di big data per seguire il percorso dei prodotti verso la loro destinazione, identificando inefficienze e capire i punti di efficientamento e riduzione di costi e tempo.
- **Migliorare le integrazioni dei clienti ed aumentare la segmentazione della clientela:** l'aggregazione di dati strutturati, semi-strutturati e non strutturati dai touchpoint tramite i quali i clienti interagiscono con l'azienda, per ottenere una vista a 360 gradi del

¹ <https://www.ibm.com/it-it/analytics/hadoop/big-data-analytics>

comportamento e delle motivazioni dei clienti per progettare un marketing personalizzato più efficace. Le origini dati includono dati provenienti dai social media, da sensori, da dispositivi mobili, da opinioni e da registrazioni delle chiamate.

3.2 Tipi di metodi di analisi dei dati

I dati possono essere analizzati in accordo a due principali tipologie di metodi: le analisi **quantitative** e quelle **qualitative**. La prima tipologia riguarda le tecniche di analisi in cui l'informazione è espressa numericamente, e come tale può essere utilizzata per effettuare calcoli di diversa natura; può essere, inoltre, rappresentata in modalità visuale, tramite l'impiego di tabelle o grafici di sintesi. Le analisi quantitative forniscono quindi indicazioni sulle modalità con cui classificare le possibili cause di problemi, o quantificare il loro impatto, ma non suggeriscono indicazioni esplicite e dirette riguardanti in che modo (come) e/o quale problema (cosa) affrontare per primo. La seconda tipologia di analisi, invece, risponde a quesiti del tipo "come, perché, cosa" in forma testuale e offrono la possibilità di definire un problema e delle azioni da intraprendere per affrontare il problema stesso.

I processi di Big Data Analytics, più precisamente, possono essere classificati in accordo a quattro tipologie, che si distinguono in base al livello di maturità delle metodologie utilizzate, e di conseguenza alle informazioni che è possibile estrarre:

- **Analitiche Descrittive (Descriptive):** rappresentano le metodologie che descrivono la situazione passata e attuale dei processi aziendali. Tra le metodologie più diffuse, l'analisi descrittiva o statistica applica tutti i passi per fornire un progetto il più completo possibile, di quello che è accaduto nella storia dei dati raccolti. Questo tipo di analisi permette di avere una visione chiara di quello che è accaduto nel passato, fornendo quindi gli elementi necessari per dare supporto a decisioni tipicamente basate sull'esperienza di chi ne usufruisce. Uno step successivo nell'analisi descrittiva è chiamato analisi diagnostica, la quale fornisce una analisi più approfondita su quelle che sono le ragioni relative all'avvenimento di eventi passati.

- **Analitiche Predittive (Predictive):** includono le tecniche che effettuano l'analisi dei dati per rispondere a domande relative ad eventi futuri. In questo ambito troviamo tecniche come regressione, forecasting, modelli predittivi. È in questo contesto che può entrare in gioco anche il machine learning. Utilizzare le osservazioni ottenute per derivare delle predizioni su quelli che potrebbero essere i comportamenti futuri è quella che viene definita metodologia predittiva di analisi. L'individuazione di trend, pattern, cluster o relazioni di causa-effetto, fatta in maniera ingegnerizzata grazie all'applicazione di algoritmi di Machine Learning, fa sì che venga sfruttata a pieno la mole di dati a disposizione per non lasciare completamente all'esperienza del decisore, al suo gut-feeling, quelle che sono le azioni da intraprendere: fornisce quindi un valido aiuto verso un reale approccio data-driven.
- **Analitiche Prescrittive (Prescriptive):** riguardano modelli in grado di ipotizzare una serie di scenari futuri. Alcuni esempi di applicazione di questa tipologia di analitiche sono forniti dalla ottimizzazione delle supply chain e nella manutenzione predittiva (prevenzione di guasti e fermo macchina). La metodologia in oggetto non si limita a prevedere o descrivere cosa è accaduto (o è probabile che accada); essa suggerisce anche quello che è il corso delle azioni e potenziali implicazioni che queste possono avere, con lo scopo di generare raccomandazioni o decisioni automatizzate. Questa richiede un perimetro ben definito e specifici algoritmi così da fornire le corrette indicazioni.
- **Analitiche Automatiche (Automated):** includono tutti gli strumenti in grado di effettuare autonomamente un'azione sulla base delle analisi di dati effettuate. Esempi di questa tipologia sono rappresentati dal *"dynamic pricing"* dei prodotti su un sito di e-commerce oppure lo smistamento automatico delle pratiche in ambito bancario o assicurativo, con l'obiettivo di identificare le frodi.

Gli **Advanced Analytics**, infine, comprendono le categorie di Predictive, Prescriptive e Automated Analytics. Lo scopo ultimo di queste metodologie è fornire un più ampio supporto ai decisori aziendali, in taluni casi andando ad automatizzare delle specifiche azioni.

3.3 Processo di analisi dei dati

Il processo di analisi e interpretazione dei dati è più iterativo che lineare, e l'aspetto ciclico è un fattore fondamentale, perché soggetto continuamente a verifica, con la possibilità di tornare

indietro per applicare correzioni. In ogni caso, il tipico ciclo di vita di un progetto di data science si può riassumere in cinque fasi, che vanno dalla comprensione del problema alla scelta dei dati più opportuni per risolverlo:

1. **Business understanding:** cioè definire un obiettivo e i risultati potenziali;
2. **Data collection:** questa fase richiede competenza specifica, ma anche strategia e tecnologia opportuna. Innanzi tutto, i dati possono provenire da fonti anche molto eterogenee e possono essere o meno strutturati, inoltre, selezionare in questa fase alcuni dati rispetto ad altri preclude alcune strade sin dal principio;
3. **Data preparation:** in questa fase si valuta se i dati raccolti sono o meno rappresentativi del problema che si vuole risolvere e si implementano tutta una serie di tecniche che vanno dalla gestione dei dati mancanti, alla correzione dei valori scorretti, l'eliminazione dei duplicati, la strutturazione dei dati per l'algoritmo e l'individuazione di caratteristiche specifiche emergenti. Dopo aver preparato i dati, si esegue una prima analisi esplorativa applicando i primi test statistici e le prime tecniche di visualizzazione delle informazioni;
4. **Data modelling:** questa è la fase della scelta e della costruzione del modello da applicare per la risoluzione del problema. Due sono gli aspetti fondamentali: la costruzione del modello e la messa a punto di parametri di controllo e validazione. Il modello viene individuato e attivato dall'algoritmo. Attraverso il machine learning, per esempio, è possibile sviluppare modelli sia descrittivi, che lavorano a livello di raccomandazione, che predittivi, focalizzati sulla previsione di possibili tendenze future: un esempio è quello della regressione lineare per prevedere futuri valori in borsa. Sulla base degli insights ottenuti, si valutano quindi accuratezza e rilevanza del modello, che in generale deve essere tendere alla maggiore semplicità possibile, e si effettuano eventuali calibrazioni necessarie;
5. **Sviluppo e riattivazione:** In questa fase, è necessario confrontare le performance effettive del modello scelto rispetto ai KPI di business fissati all'inizio, anche con il coinvolgimento di coloro che lavorano direttamente sul problema. Risulta inoltre fondamentale la capacità dell'esperto di data science di comunicare con il mondo esterno, vale a dire di rendere comprensibile il progetto e di convincere gli stakeholder coinvolti.

3.4 Tecniche di analisi dei dati

Le metodologie di analisi descritte possono essere applicate utilizzando diverse tecniche e tecnologie. L'attività di analisi dei dati può essere mappata su tre tecniche principali, adottate in maniera singola e/o combinata, per soddisfare le differenti esigenze che si presentano: **Business Intelligence, Data Mining e Data Visualization.**

■ Business Intelligence (BI)

La BI e gli Analytics in senso più ampio sono la principale tecnica attualmente utilizzata per l'analisi dei dati. Si basa sulla strutturazione dei dati aziendali da diverse sorgenti e sulla definizione di un modello semantico di metadati dove vengono applicate le logiche di business e regole di contesto aziendale per fare in modo che i dati grezzi si trasformino in vera e propria informazione a valore. I sistemi di BI sono diventati ormai una commodity in tutte le realtà aziendali, in tutte le industry, e permettono di supportare in maniera descrittiva e diagnostica le decisioni all'interno di un'organizzazione.

Il vantaggio di queste tecniche è quello noto di sistemi di BI che hanno una forte componente di conoscenza e governance centralizzate, fruibili da tutti i livelli organizzativi. Di contro la poca tempestività nei nuovi sviluppi, i limitati insights e i costi di gestione elevati, hanno fatto sì che negli anni siano state valutate alternative per l'analisi dei dati.

■ Data Mining

Il Data Mining (letteralmente "Estrazione Dati") ha l'obiettivo dell'estrazione di pattern e conoscenza dai dati e non l'estrazione dei dati stessa. Essa rappresenta il processo che combina statistica, Machine Learning e tecnologia per la scoperta automatica o semiautomatica di pattern, correlazioni, cluster, classificazioni, profilazioni e regressioni di grandi volumi di dati. Possiamo dire, semplificando, che le tecniche di Data Mining sono quelle centrali nelle moderne iniziative di Data Science e Intelligenza Artificiale, che forniscono la possibilità di andare oltre a una metodologia di analisi semplicemente descrittiva, fornendo della conoscenza ulteriore e supportando ancora più puntualmente gli analisti, aiutandoli a capire quali potrebbero essere comportamenti futuri.

Il grande pro è quello di vedere una reale estrazione di conoscenza, realmente azionabile per portare valore all'azienda. Non sempre però le organizzazioni possiedono le skills adatte

internamente per fare in modo che questo avvenga: avvalersi di figure esterne e formare tecnicamente i propri dipendenti non tuttavia sempre possibile per limiti di tempo e budget.

Data Visualization & Exploration

In queste tecniche, lo sforzo maggiore non è tanto nella fase di processamento e modellazione dei dati, ma nella loro interpretazione via rappresentazione grafica per poter individuare situazioni anomale o potenziali opportunità, sfruttando visualizzazioni avanzate. Questo tipo di tecnica si avvale di strumenti che si basano su una forte predisposizione all'analisi dei dati in modalità self service, dove sono direttamente i professionisti con attitudini più di processo e meno tecniche che applicano la loro conoscenza ed esperienza per cercare di estrarre informazioni utili.

Queste tecnologie sono già molto diffuse ed alcune integrano già in maniera agile, piccole funzionalità derivate dal mondo statistico, come l'integrazione di trendline, outlier detection o clustering.

Grazie a tecniche di questa tipologia si crea uno svincolo dai tempi tecnici del reparto IT, si possono creare importanti ingaggi del business e non sussiste la dipendenza da abilità tecniche, avendo così modo di sfruttare il know-how interno. Dall'altro lato questi approcci, nel lungo termine, esibiscono la mancanza di una "struttura" sottostante che permetta di ingegnerizzare i risultati ottenuti all'interno della value chain aziendale e difficilmente riescono a scalare per farne un utilizzo più esteso.

3.5 Analisi dei dati nell'ambiente Big Data

3.5.1 Cosa sono i Big Data

"Big Data" vuol dire (letteralmente) "Grandi dati", ovvero grandi quantità di dati, che presi insieme occupano molto spazio di archiviazione, nell'ordine dei Terabyte (unità di misura multipla del byte, corrispondente a 2 alla quarantesima byte, ovvero a 1.048.576 megabyte, il simbolo è TB).

Una ricerca su Google, un nostro acquisto al supermercato, una foto, un messaggio vocale, un tweet. Tutti questi sono dati. La maggior parte delle nostre attività quotidiane, oggi, crea dei dati, che possono essere raccolti, analizzati e monetizzati.

Super computer e algoritmi ci permettono di analizzare la sempre crescente mole di dati generati ogni giorno. Le CPU dei computer potrebbero a breve arrivare alla potenza di calcolo del cervello umano. L'intelligenza artificiale potrebbe in poco tempo far sostituire ai robot molti dei lavori che facciamo oggi. Questo è possibile anche grazie alla enorme quantità di dati che oggi vengono generati, che possono essere facilmente analizzati dalle macchine, rivelando percorsi e connessioni tra le molte attività umane, oltre a creare dei profili dettagliati su di noi. Entra in gioco anche un discorso di privacy, che non è per niente banale.

Il fenomeno dei "Big Data", o meglio il fenomeno di immagazzinare, gestire e analizzare grandi quantità di dati non è in realtà un fenomeno recente, anzi è un fenomeno che fa parte della lunga storia dell'evoluzione del genere umano.

Così come è avvenuto per innovazioni quali internet, data center o cellulari, anche i Big Data sono uno step verso un modo diverso di gestire il business e la società. Secondo alcuni studi pare che ognuno di noi generi in media 12 gigabyte di dati ogni giorno, e il numero è in costante crescita. Si pensi ad esempio a tutti i post che vengono creati ogni giorno su Facebook. Un numero incredibile di dati generati dagli utenti, che verranno salvati da qualche parte e, forse, un giorno analizzati.

Tutte le attività che quotidianamente sono svolte sui dispositivi digitali producono dati. Si tratta di una grandissima quantità di informazioni che possono essere raccolte, analizzate e quindi anche valorizzate dal punto di vista economico. È l'era della big data analytics. E il significato di big data e analytics è rispettivamente l'insieme di dati eterogenei, ossia ottenuti da molteplici fonti diverse. E la scoperta, l'interpretazione e la comunicazione di modelli significativi in tali dati al fine di avviare un processo decisionale maggiormente efficace.

Big data è un termine applicato ai data set la cui dimensione o tipo supera la capacità dei database relazionali tradizionali di catturare, gestire ed elaborare i dati con bassa latenza. I big data possiedono una o più delle seguenti caratteristiche: elevato volume, elevata velocità o estrema varietà. L'AI (Artificial Intelligence - Intelligenza Artificiale), la tecnologia mobile, i social media e l'IoT (Internet of Things) stanno portando la complessità dei dati verso nuove forme e fonti di dati. Ad esempio, i big data provengono da sensori, dispositivi, video/audio, reti, file di log, applicazioni

transazionali, web e social media — gran parte di essi viene generata in tempo reale e su vastissima scala.

L'analisi dei big data consente agli analisti, ai ricercatori e agli utenti di business di prendere decisioni in modo più accurato e veloce, utilizzando dati precedentemente inaccessibili o inutilizzabili. Le aziende possono utilizzare tecniche di advanced analytics, quali ad esempio analytics di testo, machine learning, predictive analytics, data mining, statistiche ed elaborazione del linguaggio naturale, per ottenere nuovi insight da origini dati precedentemente non sfruttate, in modo indipendente o insieme ai dati aziendali esistenti.

I Big Data sono una risorsa ormai riconosciuta come la più grande fonte di conoscenza e crescita per ogni business. Senza dati non si sopravvive alla competizione tra le aziende ancora poco digitalizzate. Infatti, le difficoltà insite nella Big Data Analytics, dovute alla natura complessa dei dati, che sono troppi, diversi, veloci e variabili, oggi sono ampiamente superate con le recenti tecnologie per i Big Data.

3.5.2 Piattaforme di raccolta ed elaborazione Big Data

I processi principali che compongono il ciclo di vita dei Big Data possono essere raggruppati in due macro-aree:

- **Big Data Management:** racchiude i processi e le tecnologie per l'acquisizione e la memorizzazione dei Big Data e la preparazione ed il recupero degli stessi;
- **Big Data Analytics:** racchiude i processi utilizzati per analizzare e acquisire informazioni utili da grandi dataset, allo scopo di interpretare e descrivere il passato (descriptive analytics), predire il futuro (predictive analytics) o consigliare azioni (prescriptive analytics).

Analizziamo più nel dettaglio tutte le varie fasi:

1. **Generazione ed Acquisizione del Dato:** i dati possono essere generati da diverse fonti, ad esempio possono essere *“human generated”*, *“machine generated”* o *“business generated”* (dati aziendali). Essi possono essere raccolti in modi diversi, ad esempio accedendo alle API messe a disposizione dai servizi web (ad esempio le Twitter API), oppure utilizzando software

di Web Scraping, che eseguono operazioni di crawling, parsing ed entity extraction per la raccolta automatica di dati da documenti presenti in Internet. Il **framework Apache Tika** automatizza queste informazioni. Altro modo di raccogliere i dati è importarli direttamente da database, per tali operazioni possiamo essere supportati dal **framework Apache Sqoop**. Se vogliamo acquisire flussi continui di dati, rapidamente generati, tramite sistemi capaci di catturare eventi, elaborarli e salvarli su un Database in modo efficiente, possiamo utilizzare tecnologie come **Apache Flume, Apache Kafka, Apache Pulsar, o Microsoft StreamInsight**.

2. **Estrazione e Pulizia delle informazioni:** La maggior parte dei dati raccolti non sono disponibili nel formato richiesto per la successiva fase di elaborazione. Pertanto, bisogna creare un processo di estrazione che prelevi le informazioni richieste dalla grande fonte dei dati e le rappresenti in una forma standard e strutturata pronta per l'analisi.
3. **Storage dei Dati:** Vari tipi di database sono stati proposti negli anni nell'intento di memorizzare, gestire ed organizzare dataset caratterizzati da grandezze, strutture e provenienze diverse. A causa della staticità delle strutture tabellari, le soluzioni basate su RDBMS si sono rivelate inadatte ai requisiti di varietà e volume propri dei Big Data. Per tale motivo utilizziamo Database NoSQL, che abbandonano le restrizioni imposte dal modello relazionale e dal linguaggio SQL, e che invece possiedono caratteristiche fondamentali per i Big Data come l'essere "*schemaless*" e distribuite, avere una facile replicazione dei dati, eventual consistency e, non ultimo, il supporto per la memoria persistente di grandi moli di dati. Esempi sono **MongoDB, Apache Cassandra, Apache HBase, Apache Druid, e Neo4J**: quest'ultimo è uno dei DB NoSQL più diffusi ed apprezzati, basato su una strutturazione dei dati a grafo. Spesso vi è anche una fase di integrazione, che consiste in ulteriori elaborazioni e trasformazioni dei dati per prepararli alla successiva fase di analisi. Ciò permette di ottenere una visione unificata e normalizzata dei dati.
4. **Analisi ed Elaborazione del Dato:** l'obiettivo è estrarre valore in forma di conoscenza dai Big Data, esaminando gli enormi dataset a disposizione alla scoperta di correlazioni, trend, pattern ed ulteriori indici statistici nascosti nei dati. Possiamo fare l'analisi di testi con tecniche di NLP, Machine Learning ed Analisi Statistica, oppure l'analisi di dati multimediali con algoritmi di Machine Learning e librerie come **TensorFlow**. Per l'elaborazione e l'analisi dei dati è possibile usare **Apache Spark e MLib**, ma esistono tanti altri framework, come ad esempio **Apache Flink, Apache Storm, Airone di Apache**.

5. **Visualizzazione:** infine segue l'interpretazione dei parametri analizzati, in modo tale da produrre grafici o ipotesi empiriche, che possono essere utili per il supporto a decisioni di business. Esistono tantissimi framework aziendali che potremmo utilizzare, degli esempi sono **Tableau** e **GraphX**.

I software per l'estrazione, gestione e analisi dei dati disponibili sul mercato sono tanti e diversi, ciascuno con le sue caratteristiche e funzionalità peculiari. La selezione richiede consapevolezza e conoscenza delle tecnologie più adatte al proprio caso d'uso. Gli strumenti ETL (Extract Transform Load) per l'estrazione, gestione e analisi dei dati permettono di superare quegli ostacoli dapprima insormontabili per le piccole e medie imprese.

Oggi, infatti, grazie a piattaforme come data lakes e Hadoop per l'archiviazione dei dati, a tecnologie incaricate di estrarre, gestire e fare analisi in tempo reale – è finalmente possibile dominare questo immenso volume di dati che arrivano a noi in forma grezza, non leggibile, perché generati da ogni fonte (transazioni, Internet of Things, apparecchi industriali, e-mail, video, audio, social media, siti web ecc).

Nei contesti in cui si vogliano analizzare *near-realtime* grandi volumi di dati, emerge la necessità di adottare come modello di calcolo lo Stream Processing. In figura è schematizzato lo Stream Processing e si mette in evidenza come sia sua caratteristica l'elaborazione dei dati ancora in movimento, senza l'esigenza di memorizzarli e quindi senza dover introdurre ritardi "artificiali" nel calcolo.

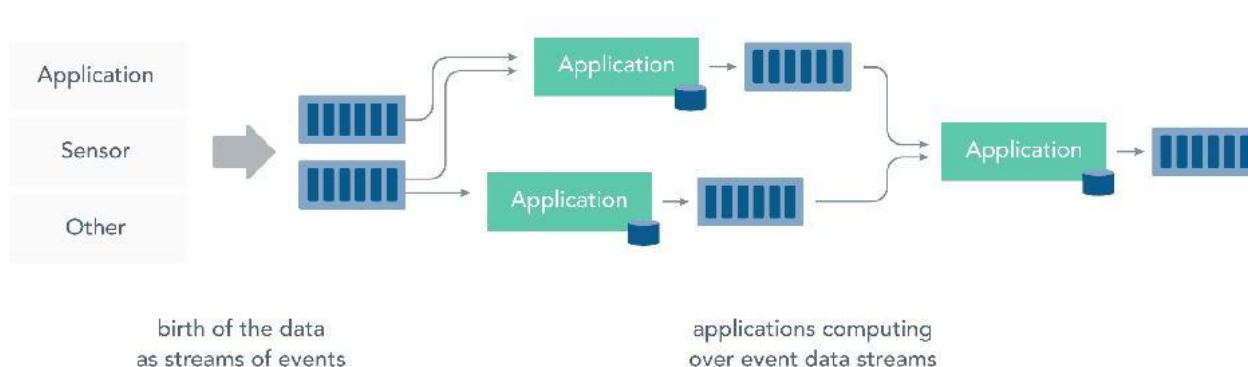


Figura 23. Stream Processing

Apache Flink è uno dei più attivi player nell'ambito dello stream processing. Esso è un framework open source per lo stream processing nato come fork del progetto di ricerca Stratosphere, finanziato dalla Comunità Europea e di cui fanno parte alcune università tedesche dell'area di Berlino (TU

Berlin, HU Berlin, HPI Potsdam). Flink è attualmente sviluppato dalla Apache Software Foundation. Il core di Apache Flink è un motore per l'elaborazione stateful e distribuita di stream di dati scritto in Java e Scala, che permette di lavorare sostanzialmente con la medesima API su dataset bounded e unbounded, adattandosi quindi sia come piattaforma per le esigenze di batch processing che per quelle di stream processing. A grandi linee, un'applicazione Flink può essere immaginata come un insieme di operatori (distributed, stateful) che rappresentano i core di calcolo, connessi tra di loro da vere e proprie pipeline, con Flink che si occupa sia della implementazione delle connessioni che dell'esecuzione degli operatori su thread separati eventualmente distribuiti su molteplici host. Apache Flink è stato costruito sin dal principio per supportare l'esecuzione su cluster di centinaia o addirittura migliaia di nodi (con supporto anche per YARN and Mesos). Flink si occupa di rendere relativamente facile lo sviluppo di applicazioni distribuite per l'analisi di dati ma non fornisce alcuna primitiva o componente per lo storage o il salvataggio dei risultati delle elaborazioni. Piuttosto, Flink mette invece a disposizione delle API ad alto livello per descrivere sorgenti e sink di dati, con le quali possono essere realizzati connettori con gli esistenti strumenti. Numerosi connettori sono forniti anche insieme al runtime di Flink, tra i quali è interessante ricordare **Apache Kafka**, utilizzabile sia come sorgente che come destinazione dei dati è un motore di Stream Processing distribuito per costruire pipeline di dati in real-time e applicazioni in streaming. Permette di ricevere dati da diversi tipi di sorgenti (detti producer), elaborandoli all'interno della sua architettura, e rendendoli disponibili ai riceventi (consumer). All'interno dell'architettura è possibile costruire applicazioni attraverso la libreria Kafka Streams per operazioni di filtraggio e arricchimento dati.

L' **API di streaming di Twitter** fornisce l'accesso al flusso di tweet messo a disposizione da Twitter. Flink Streaming viene fornito con una **TwitterSourceclass** incorporata per stabilire una connessione a questo flusso.

3.5.3 Tecniche di Big Data Analytics

La sfida più grande per la maggior parte delle aziende oggi è nella capacità di acquisire un vantaggio competitivo lavorando sui dati. Tutte le aziende e tutte le Pubbliche Amministrazioni stanno diventando grandi "fabbriche di dati". Noi stessi contribuiamo costantemente, consapevolmente e spesso anche inconsapevolmente, alla "produzione di dati". La definizione di big data analytics fa riferimento al processo che include la raccolta e l'analisi dei big data per ottenerne informazioni utili

al business. Le tecniche di big data analytics consentono infatti di fornire alle aziende intuizioni originali, per esempio, sulla situazione del mercato. D'altra parte offrono idee sul comportamento dei clienti su come raffinare le strategie di customer experience e così via. Per compiere le attività tese a fornire queste e tante altre informazioni preziose per migliorare l'attività dell'impresa è necessario quanto segue.

- Software (dai database e strumenti utili per acquisire ed elaborare informazioni agli applicativi dedicati per specifici processi aziendali).
- Servizi (per esempio, per customizzare la tecnologie e integrarle con successo nei sistemi pre esistenti).
- Risorse infrastrutturali (capacità di calcolo, storage eccetera).

I benefici che l'analisi big data può dare sono parecchi. Si ricorda che il significato di analisi big data è la capacità di analizzare, estrapolare e mettere poi in relazione una grande quantità di dati eterogenei strutturati o meno. Questo per scoprire legami e correlazione tra fenomeni e addirittura arrivare a prevederli.

L'analytics dei big data consiste nell'utilizzo di tecniche analitiche avanzate rispetto a data set molto grandi e diversificati, che includono dati strutturati, semi-strutturati e non strutturati, da fonti differenti e di diverse dimensioni, da terabyte a zettabyte.

Esistono una moltitudine di fasi e attività che insieme concorrono alla formazione del processo di analisi dei dati nella sua interezza. Alcuni di questi a volte vengono svolti solo in parte, altri invece sono maggiormente importanti e time-consuming in alcune tecniche e contesti piuttosto che in altri; di seguito vengono riportate le principali:

- **Requirement Gathering:** La fase primordiale del processo di analisi dei dati non può prescindere da un'attenta e chiara definizione di quello che è il problema, il bisogno, la necessità che l'analisi stessa ha come obiettivo. Identificare i desiderata e il valore che l'analisi deve portare al business, aiuta a guidare le fasi successive che stanno a valle; come scegliere quelle che sono le corrette sorgenti dati (o porzioni di esse) che dobbiamo considerare, cosa dobbiamo misurare dentro questo perimetro informativo e come i risultati

andranno poi opportunamente comunicati. Questo primo passo, aiuta anche a direzionare la scelta della migliore metodologia e degli opportuni strumenti da utilizzare.

- **Data Collection:** Basandosi su quello che è l'output di una prima fase di analisi del requisito, si passa alla raccolta dei dati necessari da poter soddisfare le esigenze finali, i comportamenti che si vogliono valutare e agli aspetti che si devono misurare. I dati vengono collezionati da una varietà di sorgenti (DB, ERP, sensori, website feed,...) contenenti informazioni sia strutturate che non. Spesso in questo contesto, risulta necessario intraprendere azioni tecniche/commerciali per recuperare determinati informazioni ancora non presenti nei sistemi di riferimento.

- **Data Processing:** Dopo aver collezionato i dati dalle sorgenti, questi devono essere processati e organizzati opportunamente per essere utilizzati in fase di analisi. In questo momento vengono applicate misure quali i controlli di integrità referenziale o la conversione dei dati in un formato utile alle lavorazioni successive.

- **Data Cleansing:** Una volta organizzati e processati, i dati possono risultare incompleti, contenere duplicati o errori. Per fare in modo che i risultati generati dall'analisi che si sta preparando siano coerenti e affidabili è importante prevedere iniziative di Data Cleansing che siano in grado di fornire un adeguato livello di Data Quality. Spesso questa fase, insieme alla precedente, è quella che risulta più time consuming vista la varietà e il volume di dati che sono coinvolti in processi di analisi.

- **Analysis/Communication:** Puliti e organizzati, i dati sono pronti per la vera e propria fase di analisi. A seconda di quelle che sono le tecniche scelte, questo step può essere approcciato in maniera profondamente differente. Quello che però accomuna questi diversi modi di affrontare il problema è la comunicazione verso gli stakeholder che sono interessati o hanno direttamente commissionato l'analisi dati in questione: le informazioni possono essere riportate in diversi formati per rispondere ai requisiti iniziali. Per fare questo, spesso vengono applicate diverse metodologie di Data Visualization in modo da guidare la comunicazione dei messaggi chiave contenuti nelle informazioni analizzate. Gli utenti finali, sulla base dell'intero processo, potranno decidere di prendere le dovute azioni e fornire feedback riguardo l'analisi generata che genererà analisi ulteriori, dando vita a un processo analitico iterativo.

I diversi metodi che possono essere utilizzati per analizzare i dati rientrano generalmente in due macrogruppi: le analisi **quantitative** e quelle **qualitative**. Le prime sono quelle dove l'informazione è espressa numericamente, può essere utilizzata di conseguenza in calcoli di diversa natura e può essere rappresentata in maniera visuale tramite tabelle o grafici. Forniscono quindi indicazioni come classificare le possibili cause di problemi, o quantificare il loro impatto, ma non ci dicono direttamente *come* e *quale* problema affrontare per primo. Il secondo macrogruppo di analisi risponde a domande del tipo “come, perché, cosa” in forma testuale e danno la possibilità di definire un problema e delle azioni per affrontare il problema stesso.

● 4 Energy Communities e Social Analysis

Le Comunità Energetiche (o Energy Community) rappresentano un paradigma emergente il cui obiettivo è quello di creare una comunità di stakeholders (cittadini, attività commerciali, attività private) che possano condividere energia rinnovabile e pulita, in uno scambio tra pari.

Numerose sono le definizioni fornite per caratterizzare la Comunità Energetica: il gruppo ACEA² ha fornito la seguente definizione: *“una comunità attiva in grado di produrre, scambiare e consumare energia in modo da essere autosufficiente è una comunità energetica rinnovabile, in cui l'energia viene condivisa e in cui i cittadini sono coinvolti nello sviluppo sostenibile della loro città”*.

Una visione più dettagliata su questa tipologia di realtà emergente è fornita anche dalla stessa ENEA³, che nelle più recenti pubblicazioni a riguardo, fornisce un quadro di riferimento completo e dettagliato delle Comunità Energetiche in Italia, fornendo linee guida per i cittadini che vogliano aderire a questa nuova opportunità.

In una comunità energetica, i clienti finali, consumatori di energia elettrica, possono associarsi per produrre localmente, tramite fonti rinnovabili, l'energia elettrica necessaria al soddisfacimento dei propri fabbisogni, attraverso la “condivisione”. Questa opportunità è stata supportata dalla entrata

² <https://www.acea.it/acea-energia-stories/comunita-energetiche>

³ https://www.enea.it/it/seguici/pubblicazioni/pdf-volumi/2020/guida_comunita-energetiche.pdf

in vigore del decreto-legge 162/19 (articolo 42bis) e dei relativi provvedimenti attuativi, quali la delibera 318/2020/R/eel dell'ARERA e il DM 16 settembre 2020 del MiSE.

L'energia elettrica "condivisa" (computata come il minimo, su base oraria, tra l'energia elettrica immessa in rete dagli impianti di produzione e l'energia elettrica prelevata dai consumatori che rilevano per la configurazione) può beneficiare di un contributo economico riconosciuto dal GSE a seguito dell'accesso al servizio di valorizzazione e incentivazione.

Le tipologie di configurazione ammesse al servizio sono due⁴:

- gruppi di autoconsumatori di energia rinnovabile che agiscono collettivamente;
- comunità di energia rinnovabile;

Per **autoconsumatore** di energia rinnovabile si intende un cliente finale che, operando in propri siti ubicati entro confini definiti, produce energia elettrica rinnovabile per il proprio consumo; può immagazzinare o vendere energia elettrica rinnovabile autoprodotta purché il beneficiario sia un autoconsumatore di energia rinnovabile diverso dai nuclei familiari, e che tali attività non costituiscano l'attività commerciale o professionale principale. Un **gruppo di autoconsumatori** è un insieme di almeno due autoconsumatori di energia rinnovabile, che agiscono collettivamente in virtù di un accordo privato e che si trovano nello stesso condominio o edificio.

Una **comunità di energia rinnovabile** è un soggetto giuridico con le seguenti caratteristiche:

1. è dotato di autonomia ed è effettivamente controllato da azionisti o membri che sono situati nelle vicinanze degli impianti di produzione detenuti dalla comunità di energia rinnovabile; è basato su partecipazione aperta e volontaria;
2. i suoi azionisti o membri sono persone fisiche, piccole e medie imprese (PMI), enti territoriali o autorità locali, comprese le amministrazioni comunali, sotto la condizione che, per le imprese private, la partecipazione alla comunità di energia rinnovabile non costituisca l'attività commerciale e/o industriale principale;
3. l'obiettivo principale è fornire benefici ambientali, economici o sociali a livello di comunità ai propri azionisti o membri o alle aree locali in cui opera, piuttosto che profitti finanziari.

⁴ <https://www.gse.it/servizi-per-te/autoconsumo/gruppi-di-autoconsumatori-e-comunita-di-energia-rinnovabile>

Le comunità energetiche rappresentano, quindi, un modello innovativo per la produzione, la distribuzione e il consumo di energia proveniente da fonti rinnovabili. Questo modello fonda i suoi valori sulla lotta allo spreco energetico e sulla condivisione di un bene fondamentale a un prezzo concorrenziale, grazie all'innovazione che sta rivoluzionando il mercato dell'energia.

Il concetto di una comunità energetica “smart” estende quello della Smart Grid, che fornisce lo specchio dei cambiamenti che la digitalizzazione ha introdotto anche nella rete elettrica. Il mondo digitale, che ha permesso la connessione tra nodi peer to peer, ha permesso il passaggio da una rete fisica centralizzata, con delle trasmissioni one-to-many (il gestore elettrico che fornisce energia alle case), a una rete digitale decentralizzata, con collegamenti one-to-one e many-to-many. Questa tipologia di rete è considerata smart dal momento che è in grado di incorporare, oltre ai necessari sensori di misura, anche algoritmi di intelligenza artificiale, che abilitano la partecipazione attiva anche dei singoli cittadini.

Grazie a questa caratteristica di essere “smart” oltre che energeticamente sostenibili, le EC possono beneficiare delle informazioni provenienti dalla elaborazione dei dati raccolti dai propri stakeholders, per calibrare e migliorare l'efficienza e l'efficacia delle stesse. L'analisi dei dati raccolti e l'utilizzo dei social network possono rivestire il ruolo di efficaci volani, per la diffusione di una migliore e più accurata informazione in merito a questo rivoluzionario paradigma; essi possono consentire, dunque, non solo la messa a conoscenza dei cittadini della esistenza di questo innovativo paradigma, ma anche la condivisione di esperienze da parte dei partecipanti alle EC, utili a vincere la diffidenza nei confronti di un sistema non ancora noto ai più e, pertanto, approcciato con eccesso di cautela.

● 5 Conclusioni

Nella nostra epoca ricca di dati, capire come analizzare ed estrarre il vero significato dalle intuizioni digitali della nostra azienda è uno dei principali fattori di successo. Nonostante l'enorme volume di dati che creiamo ogni giorno, solo lo 0,5% viene effettivamente analizzato e utilizzato per la scoperta, il miglioramento e l'intelligence dei dati. Anche se potrebbe non sembrare molto, considerando la quantità di informazioni digitali che abbiamo a portata di mano, il mezzo per cento rappresenta ancora una grande quantità di dati. L'analisi dei dati è il processo di raccolta, modellazione e analisi dei dati per estrarre informazioni a supporto del processo decisionale. Esistono diversi metodi e tecniche per eseguire l'analisi a seconda del settore e dello scopo dell'analisi. Tutti questi vari metodi per l'analisi dei dati si basano in gran parte su due aree centrali: *metodi quantitativi e metodi qualitativi nella ricerca*.

Utilizzando l'analisi dei dati per avere una visione a 360° di tutti gli aspetti relativi ai tuoi clienti, puoi capire quali canali usano per comunicare con te, i loro dati demografici, interessi, abitudini, comportamenti di acquisto e altro ancora.

Il lavoro portato avanti in questo report si è rivolto allo studio e alla comparazione di tecniche per l'analisi di contenuti testuali, tramite tecniche di Natural Language Processing, condivise sui Social Network. Avendo come obiettivo quello di rivolgersi a piattaforme di microblogging, lo studio dello stato dell'arte effettuato si è focalizzato in particolare sulla cosiddetta Sentiment Analysis, volta a riconoscere i trend espressi dagli utenti dei social e ad associarvi emozioni e sentimenti specifici. Il report ha sottolineato quali siano i vantaggi, per una compagnia o un'organizzazione con una certa attenzione alla Customer Care, apportati dalle tecniche di Sentiment Analysis all'andamento complessivo del Business. Inoltre, si è soffermato sull'analisi e la comparazione di diversi tool utilizzati in questo campo, identificandone differenze, difetti e punti di forza.

Requisito imprescindibile per una corretta Sentiment Analysis si è rivelata essere la gestione delle consistenti fonti di dati costituite dai Social Network, che prevedono l'applicazione di specifiche tecniche di Big Data analysis per una corretta interpretazione delle informazioni. Sottolineando l'importanza dell'utilizzo di una Big Data pipeline ben definita per immagazzinare, ripulire e processare i dati provenienti da fonti eterogenee e non strutturate, quali appunto i Social media, il report ha mostrato le principali tecniche e tecnologie di realizzazione di una piattaforma di Big Data.

Future attività di ricerca si muoveranno verso lo studio di tecniche di integrazione delle tecnologie e delle metodologie analizzate nel presente report, nonché sull'uso di specifiche tecnologie semantiche per l'arricchimento e l'annotazione dei dati ottenuti dai Social Network. Un altro importante aspetto da prendere in considerazione riguarda l'utilizzo di Servizi Cloud per la gestione dei dati e delle analitiche: il Cloud, con la sua scalabilità e flessibilità, si presta infatti a molteplici applicazioni pratiche di gestione ed analisi massiva dei dati.

Le tecniche di data analytics discusse in questo report possono essere applicate a supporto di ulteriori scenari emergenti, quali le community energetiche, che rappresentano una realtà dell'immediato futuro.

Una comunità energetica, ovvero una comunità locale che genera e auto-consuma energia green, sostenendo in questo modo la comunità locale e favorendo la sostenibilità, rappresenta la principale direzione in cui si stanno muovendo Energia e Comunità: l'energia può essere condivisa tra i cittadini di una comunità, abitanti, attività commerciali e imprese del territorio allo scopo di favorire la sostenibilità, ridurre la povertà energetica e generare un ciclo economico a basse emissioni di carbonio.

In questa ottica, la data science, nelle sue specifiche declinazioni della big data analytics e della sentiment analysis possono più che mai supportare il processo di scambio delle informazioni tra gli stakeholders della comunità energetica, permettendo di ottenere e monitorare in tempo quasi reale i fabbisogni, le esigenze e le emergenti criticità a cui questo nuovo paradigma della comunità energetica potrà essere soggetta, soprattutto nella fase precedente alla condizione di regime e consolidamento.

● Appendice

Beniamino Di Martino è Professore Ordinario presso l'Università della Campania "Luigi Vanvitelli" (già Seconda Università di Napoli), per il settore Scientifico Disciplinare INGINF/05 - Sistemi per l'Elaborazione dell' Informazione. E' Adjunct Professor presso la Asia University – Taiwan. E'è stato Delegato del Rettore per il Coordinamento e Potenziamento delle Reti di Ateneo, per l' Informatica, per il CINECA e per il GARR. E' stato vice-Direttore del Dipartimento di Ingegneria Industriale e dell' Informazione.

Dal 1995 al 1998 e' stato Ricercatore presso l' Institute for Software Technology and Parallel Systems dell'Università' di Vienna (Austria). Dal 1998 al 2002 è stato Ricercatore Universitario, e dal 2002 al 2005 Professore Associato, presso la Seconda Università di Napoli.

Ha pubblicato 14 testi scientifici a diffusione internazionale ed oltre 300 pubblicazioni scientifiche a diffusione internazionale, di cui più di 80 articoli su riviste internazionali.

Ha partecipato ed è stato responsabile scientifico a numerosi progetti di ricerca internazionali (EC Esprit, EC CEI-Pact, EC TMR, Austrian FWF), nazionali (PRIN, FAR) e regionali. E' stato Project Coordinator del Progetto Europeo (FP7-ICT) **mOSAIC** (su Cloud Computing), e Responsabile di Unità per i Progetti Europei FP7-SMARTCITIES **CoSSMIC** (su Smart Energy Grids) e JU-ARTEMIS **Crystal** (su sistemi software affidabili). E' stato responsabile scientifico di un progetto nazionale CNR (Agenzia 2000), di un progetto MUR FAR (Laboratori Pubblico-privati) e dell'unità di ricerca della Seconda Università di Napoli per i progetti europei e Nazionali: IST (FP5-ICT) Working Group APART, Thematic Network (FP5-ICT) OntoWEB, PRIN-MUR Cloud@Home. E' attualmente WorkPackage leader dei Progetti Europei H2020-ICT **Toreador**, su Big Data, ed H2020-MG **GreenCharge** su Smart, Green and Integrated Transport.

E' Editor o Associate Editor di 7 riviste scientifiche internazionali (tra cui IEEE Transactions on Cloud Computing – TCC - ed IEEE Transactions on Parallel and Distributed Systems - TPDS), membro in numerosi editorial board di riviste internazionali e guest editor per numerose riviste internazionali. E' stato general e program chair di numerosi congressi internazionali, steering committee member di congressi internazionali, membro di numerosi comitati di programma di congressi internazionali.

E' Vice Chair dell' IEEE Technical Committee on Scalable Computing (IEEE-TCSC); è stato membro dell' Executive Board of the IEEE CS Technical Committee on Supercomputing Applications (IEEE-TCSA).

E' membro dell' IEEE Standardization Working group su Cloud Interoperability, degli IEEE Technical Committees on Scalable Computing (TCSC), on Big Data (TCBD), on Data Engineering (TCDE), on Semantic Computing (TCSEM), on Services Computing (TCSVC), on Intelligent Informatics (TCII), on Pattern Analysis and Machine Intelligence (TCPAMI), on Software Engineering (TCSE), on Distributed Processing (TCDP), on Parallel Processing (TCPP), on Cloud Computing (TCCLD), del Cloud Standards Customer Council, dell' OMG – Cloud Working Group, del Future Cloud Experts' Group della Commissione Europea – Internet of Services, Software and Virtualization Unit, dell' Innovation Advisory Board (IAB) of VECMA H2020 EC Project, dell' Advisory Board of OntoChain H2020 EC Project, del Comitato di Indirizzo della Federazione IDEM (IDentity Management). E' Chair del Nomination Committee for the "IEEE TCSC Award of Excellence in Scalable Computing", membro del Nomination Committee for the "IEEE TCSC Award for Early and Medium Career Researchers", e membro dell' Award Committee dell' IEEE Technical Committee on Cloud Computing.

E' revisore di progetti per la Commissione Europea, per i programmi ICT, ICT-PSP ed eInfrastructures, e per l' European Research Council (ERC). E' revisore di progetti scientifici per i Ministeri della Ricerca del Belgio, del Lussemburgo e del Cile. E' revisore di progetti industriali in ambito ICT per i Ministeri dell' Università e della Ricerca e dello Sviluppo Economico, per la Regione Piemonte, per la Regione Campania, per la Regione Calabria, per la Regione Lazio, per la Regione Puglia, per la Regione Toscana, per la Regione Sardegna e per numerosi enti pubblici locali. E' membro del Comitato "Agenda Digitale" ed è stato membro del "Comitato Tecnico per la predisposizione ed implementazione del Piano Strategico della Società dell'Informazione" della Regione Campania. E' stato membro di Commissione per l'abilitazione a ruoli di Senior Researcher e Professore Associato per l' Università di Cork (IR) e di Lille (Fr). E' stato valutatore di esami finali per il PhD per le Università di Oxford, Cipro (Cy), Sidney (Au), Vienna (A), La Laguna (Sp), Genova, Calabria, Politecnico di Torino, Roma Tor Vergata, Roma Tre, Pavia, Pisa, Napoli2. E' membro del Consiglio di Dottorato del Gran Sasso Science Institute. E' membro del Consiglio Direttivo del Consorzio Interuniversitario Nazionale per l' Informatica – CINI, membro del Consiglio Consortile del Consorzio Interuniversitario CINECA, e membro del Consiglio Scientifico del Consorzio Interuniversitario RIMIC; è stato membro del CdA del Consorzio Interuniversitario per l' Università a distanza NETTUNO. E' Direttore dei Nodi locali dei Laboratori Nazionali CINI "Artificial Intelligence and Intelligent Systems", "Big Data", CyberSecurity" and "Smart Cities and Communities". E' Responsabile dell'Unità della Università della Campania "Luigi Vanvitelli" del Gruppo Nazionale di Ingegneria Informatica - G.I.I. E' membro del Comitato di Indirizzo dell' Iniziativa Nazionale IDEM – Identity Management, e Responsabile IDEM per l' Ateneo. E' stato vice-Direttore del Dipartimento di Ingegneria Industriale e dell'Informazione dell' Università della Campania. Ha svolto attività di ricerca sui seguenti temi: Modelli, paradigmi, linguaggi ed ambienti di programmazione ed esecuzione avanzati per sistemi informatici Distribuiti, Cloud, BigData, Edge ed IoT, Tecniche e strumenti di Software Engineering e Program Comprehension, Semantic Web e Semantic Web Services, Natural Language Processing, Big Data Analytics, Deep Learning, Ingegneria della Conoscenza, Intelligenza Artificiale e Business Intelligence.

Antonio Esposito è attualmente Ricercatore presso il Dipartimento di Ingegneria dell'Università della Campania "Luigi Vanvitelli". La sua tesi di dottorato si è concentrata sul riconoscimento e l'applicazione di Design e Cloud Patterns allo sviluppo di Software in ambiente Cloud, con il supporto di Tecnologie Semantiche. È stato coinvolto nel progetto FP7-ICT finanziato dall'UE mOSAIC e nel progetto Horizon 2020 Toreador, ed è attualmente coinvolto nel progetto Horizon 2020 GreenCharge e nel progetto di ricerca applicata "Big data Giustizia e Datawarehouse" promosso dal Ministero italiano di Giustizia nell'ambito del Consorzio Interuniversitario Nazionale per l'Informatica (CINI). È inoltre responsabile del progetto Tenacious, finanziato nell'ambito del progetto cascata Horizon 2020 Ontochain.

I suoi principali interessi sono l'ingegneria del software, il Cloud computing, i Design e i Cloud Pattern, e le tecnologie per il Web Semantico.

Fiammetta Marulli è attualmente Ricercatore presso il Dipartimento di Matematica e Fisica dell'Università della Campania "Luigi Vanvitelli". Ha conseguito il Dottorato di Ricerca presso il Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione (DIETI) della Università degli Studi di Napoli "Federico II", difendendo una tesi sulle tecniche di rappresentazione della conoscenza e la generazione automatica di contenuti testuali nel dominio del Cultural Heritage. Durante il Dottorato di ricerca si è specializzata nella elaborazione del linguaggio naturale attraverso tecniche automatizzate (NLP) e ha raffinato le competenze in tale area durante la Research Fellowship tenuta con il Consiglio Nazionale delle Ricerche (CNR), presso l'Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR). In particolare, durante questa fellowship ha affinato l'applicazione di tecniche di Deep Learning e l'impiego di Deep Neural Networks per risolvere problemi di ambito NLP. Nel 2017 Ha partecipato al progetto C.R.E.A. nell'ambito dei programmi Horizon 2020 finanziati per il settore dell'E-Justice, coordinata dal Prof. Francesco Romeo, presso la Facoltà di Giurisprudenza dell'Università degli Studi di Napoli Federico II, occupandosi della progettazione dello sviluppo del portale web e di un motore, basato su tecniche machine learning, per il supporto alle attività di conciliazione durante le dispute giudiziarie. Attualmente partecipa al progetto di ricerca applicata "Big data Giustizia e Datawarehouse", promosso dal Ministero Italiano di Giustizia nell'ambito del Consorzio Interuniversitario Nazionale per l'Informatica (CINI). Ha svolto servizio di didattica, come docente a contratto, presso il Dipartimento di Ingegneria dell'Università degli Studi della Campania "L. Vanvitelli" dal 2016 al 2021.

Attualmente gli interessi di ricerca e le attività didattiche sono incentrate sulle tecniche e metodi di Intelligenza Artificiale applicate nei settori della Data Analytics, Text Analytics e nell'ambito NLP.

Mariangela Graziano è attualmente assegnista di ricerca presso il Dipartimento di Ingegneria dell'Università della Campania "Luigi Vanvitelli". Ha conseguito la laurea magistrale in ingegneria informatica nel 2020 con una tesi nell'ambito di Knowledge Engineering and Big Data Intelligence incentrata sull'applicazione di tecniche di rappresentazione semantica e inferenza logica applicate al domino BIM - Building Information Modeling. E' stata coinvolta nel progetto Horizon 2020 GreenCharge su Smart, Green and Integrated Transport. Attualmente partecipa al progetto di ricerca applicata "Big data Giustizia e Datawarehouse", promosso dal Ministero Italiano di Giustizia nell'ambito del Consorzio Interuniversitario Nazionale per l'Informatica (CINI).