



EUtwiz

Analisi sui dati presenti all'interno del Social Network "Twitter" relativi alle elezioni per il rinnovo del Parlamento Europeo

Daniele Ciciani 846618, Davide Mancino 847700, Louis Fabrice Tshimanga 847529

Università degli Studi di Milano - Bicocca

Abstract

Nel presente progetto viene proposto un sistema di raccolta in tempo reale e analisi del flusso di tweet scritti da cittadini e leader politici europei nel periodo precedente e successivo alle elezioni per il rinnovo del Parlamento Europeo. Gli aspetti cardinali dei Big Data considerati in questo studio sono Volume e Velocity. L'architettura implementata segue il modello di una Lambda Architecture. Quanto ottenuto è stato poi visualizzato in dashboard, realizzate tramite l'ausilio del software Tableau, contenenti dati sull'andamento dei follower dei leader politici monitorati, wordcloud dei termini più ricorrenti, polarità dei discorsi pubblicati online e opinioni dei cittadini europei in riferimento ai temi elettorali. Il sistema è una base di analisi per tematica, finestra temporale e volume di dati, nonché per elaborazioni più specifiche come NLP, Sentiment Analysis e Data Semantics.

Contents

1 Introduzione.....	2	5 Risultati.....	4
2 Obiettivi.....	2	6 Conclusioni e prospettive future.....	6
3 Architettura.....	3	Bibliografia, link e riferimenti.....	6
4 Dashboard	3		

1 Introduzione

Sempre più politici utilizzano i social media per interagire con gli elettori e sempre più le piattaforme online diventano luoghi di discussione politica orizzontale e verticale. **Twitter**, definito come sito di microblogging, fornisce uno spazio uguale per tutti gli utenti, limite di testo per **tweet** fisso per tutti, più la possibilità di un'interazione pubblica e diretta tra profili verificati (ossia istituzionali o rilevanti per altre ragioni) quanto fra profili di cittadini privati. Inoltre, c'è la possibilità di sottolineare parole tematiche tramite hashtag, indicizzando gli argomenti di discussione e costituendo trend specifici (per l'appunto, trending topics).

Queste caratteristiche rendono Twitter una sorta di termometro sociale che viene spesso preso come riferimento o corollario per comprendere l'umore delle masse o di personaggi pubblici. La piattaforma offre anche delle API per interagire con grande disponibilità in maniera algoritmica, automatica, rapida e generalizzata. Tanto che i data journalists, quanto i data analysts di aziende e organizzazioni per cui l'opinione comune è un fattore economicamente o strategicamente importante, i ricercatori ed i comuni cittadini possono usufruire dei dati presenti all'interno della piattaforma.

Nel progetto qui presentato, la raccolta dati ha avuto due linee principali, e altrettante tipologie analitiche: una divisione del flusso di lavoro è quella tra il tracciamento di determinati leader politici coinvolti nelle Elezioni Europee tramite un "ascolto" in streaming dei relativi account, ed il tracciamento di determinate parole-chiave associate alle stesse elezioni o alle istituzioni europee in generale, prescindendo dalla fonte specifica per valutare il sentire comune; l'altra distinzione si può tracciare tra statistiche quantitative in termini di numero di follower, retweet, attività riguardante gli account dei leader e tra semantica e sentiment analysis (sia dei tweet da account verificati, sia dei tweet da account comuni), in breve l'analisi delle parole ricorrenti per i topic citati, la polarità dei tweet (positiva, neutra, negativa) e in futuro un approfondimento sulla semantica e la conformazione dello "spazio di parole" che emerge dal corpo testuale dei tweet pubblicati dai leader.

2 Obiettivi

Gli obiettivi originali dell'analisi ipotizzata consistevano nella raccolta ed analisi basilare dello stream di dati generato alcune settimane prima e dopo le Elezioni Europee, ed il confronto con una raccolta dei tweet delle Elezioni Europee del 2015. Il confronto è caratterizzato da un problema convertibile in oggetto di riflessione, seguito da un problema che invece rende impossibile il compito prefigurato, con i mezzi a disposizione. Infatti, la distanza di tempo è tale per cui si può parlare di epoche differenti per il Social Network in questione, ad esempio per la differenza di caratteri ammessi per tweet, numero di utenti e differente modalità d'uso da parte dei leader politici, questi fattori renderebbero più interessante un confronto adeguatamente contestualizzato. Indipendentemente, però, la piattaforma non permette agli sviluppatori di accedere a qualsiasi query storica dei tweet di intervallo superiore alla settimana, imponendo un approccio manuale che esula dagli scopi presenti.

A fronte di ciò gli obiettivi si sono focalizzati nella programmazione di listener sui leader europei per un'analisi real-time del numero di followers e retweet, listener focalizzati su topic identificati da parole-chiave, nell'archiviazione ordinata dei dati provenienti dai due flussi e nell'analisi batch dei vocabolari e della polarità associata al testo dei tweet estratti.

Su questi obiettivi, si è scelto quindi di focalizzare l'analisi sulla gestione di 2 delle 3V Big Data

- Volume
- Velocity

In particolare, è stato rilevante l'aspetto della Volume in quanto sono stati gestiti circa 8.5 GB di dati, mentre per la Velocity si è giunti ad elaborare dati in near real-time, in finestre di 10 sec.

LEADER	STATO
SILVIO BERLUSCONI	Italia
EMMA BONINO	Italia
LUIGI DI MAIO	Italia
GIORGIA MELONI	Italia
MATTEO SALVINI	Italia

NICOLA ZINGARETTI	Italia
YANIS VAROUFAKIS	Grecia
GUY VERHOFSTADT	Belgio
FRANCOIS-XAVIER BELLAMY	Francia
MARINE LE PEN	Francia
EMMANUEL MACRON	Francia
JEAN-LUC MELENCHON	Francia
ANNALENA BAERBOCK	Germania
ANDREA MARIA NAHLES	Germania
ALICE WEIDEL	Germania
JEREMY CORBYN	Regno Unito
NIGEL FARAGE	Regno Unito
THERESA MAY	Regno Unito
NICOLA STURGEON	Regno Unito
PABLO CASADO BLANCO	Spagna
PABLO IGLESIAS	Spagna
ALBERT RIVERA	Spagna
PEDRO SANCHEZ	Spagna

Tab. 1 Leader europei monitorati

TREND MONITORATI
EUROPEE2019
ELEZIONIEUROPEE
26MAGGIO
EUROPARLAMENTO2019
EUELECTIONS2019
EP2019

Tab. 2 Trend monitorati

3 Architettura

L'architettura implementata, segue il modello di una **Lambda Architecture**, in particolare un'architettura di elaborazione dati progettata per gestire grandi quantità di dati (big data) sfruttando sia metodi di elaborazione batch che real-time. A seguire una rappresentazione ad alto livello in cui sono state rappresentate le principali tecnologie utilizzate e il flusso dati end to end all'interno delle due pipeline di gestione dati:

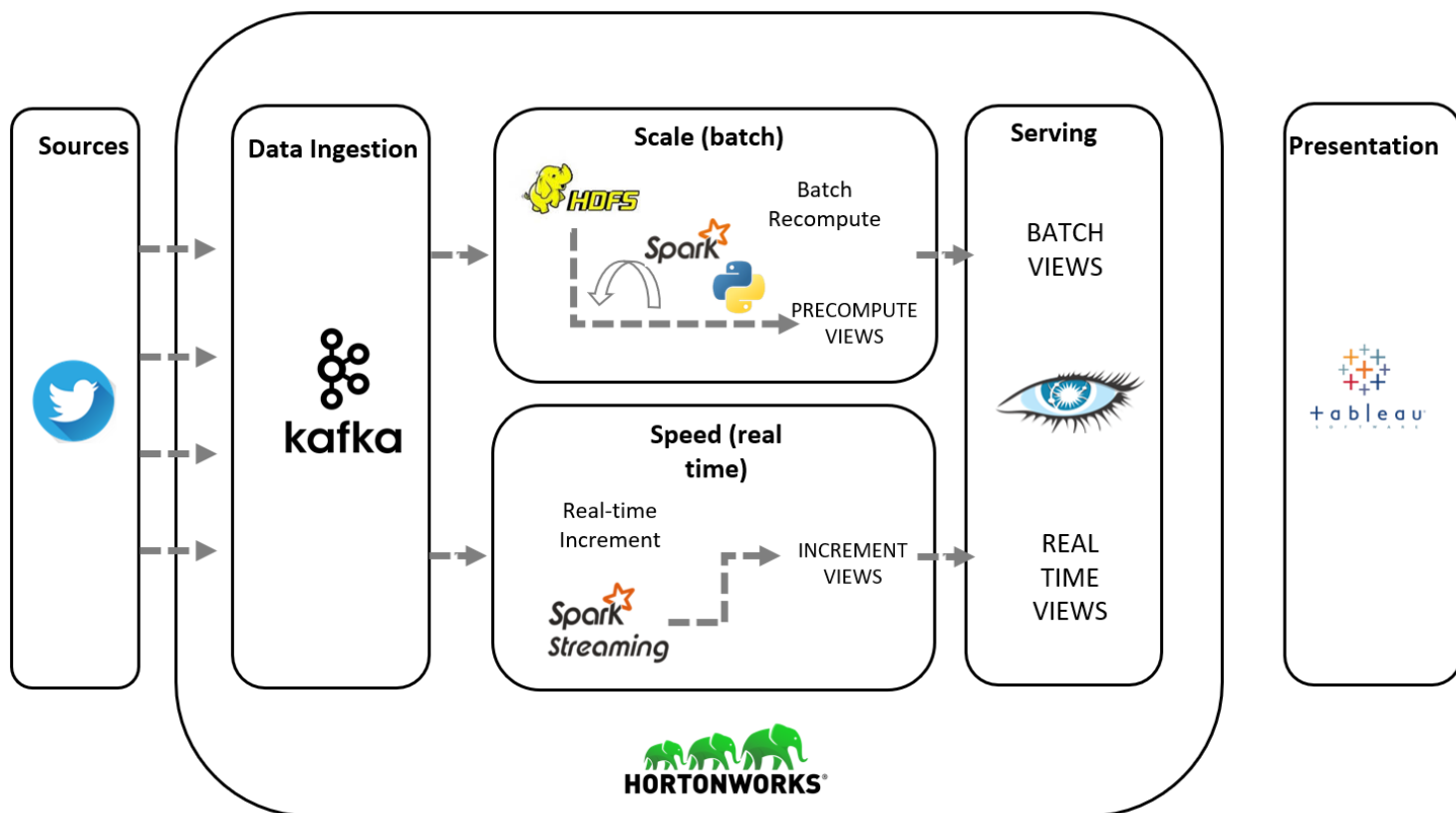


Immagine 1. Architettura

Partendo da sinistra, è indicata la sorgente dati, Twitter. In particolare è stata utilizzata tramite **Python**, la libreria **tweepy** in grado di interfacciarsi con i servizi del popolare Social Network e poterne richiamare dati relativi a tweet, commenti, follower dei vari account, ecc. sia in modalità batch che in streaming.

Proseguendo verso destra, è presente un macro blocco rappresentante la suite finalizzata alla gestione dei big data. In particolare è stata utilizzata la distribuzione **Hortonworks** in esecuzione, tramite container **Docker** (single node) all'interno di una macchina **IaaS Azure**.

Il primo componente della lambda architecture è il popolare message queue publish-subscribe **Kafka**. Tramite questo strumento, è stato possibile andare a definire un layer di **Data Ingestion**, altamente scalabile e ridondato se configurato a dovere, in grado di dividere logicamente i vari flussi dati provenienti dai client di acquisizione su vari topic e renderli disponibili ai vari consumer che in necessità o in real-time sono in grado di richiedere i dati. L'impiego di Kafka si è rivelato particolarmente vincente in quanto potendo trattenere il dato in memoria (su disco) per un certo periodo di tempo (di default 168 ore), è stato possibile continuare a raccogliere i dati twitter, anche in situazioni d'indisponibilità degli altri sistemi (**HDFS** in particolare), cosa che durante il progetto più volte è capitata per via delle limitate risorse hardware. A seguirei un focus sui componenti agenti come Consumer e Producer per Kafka:

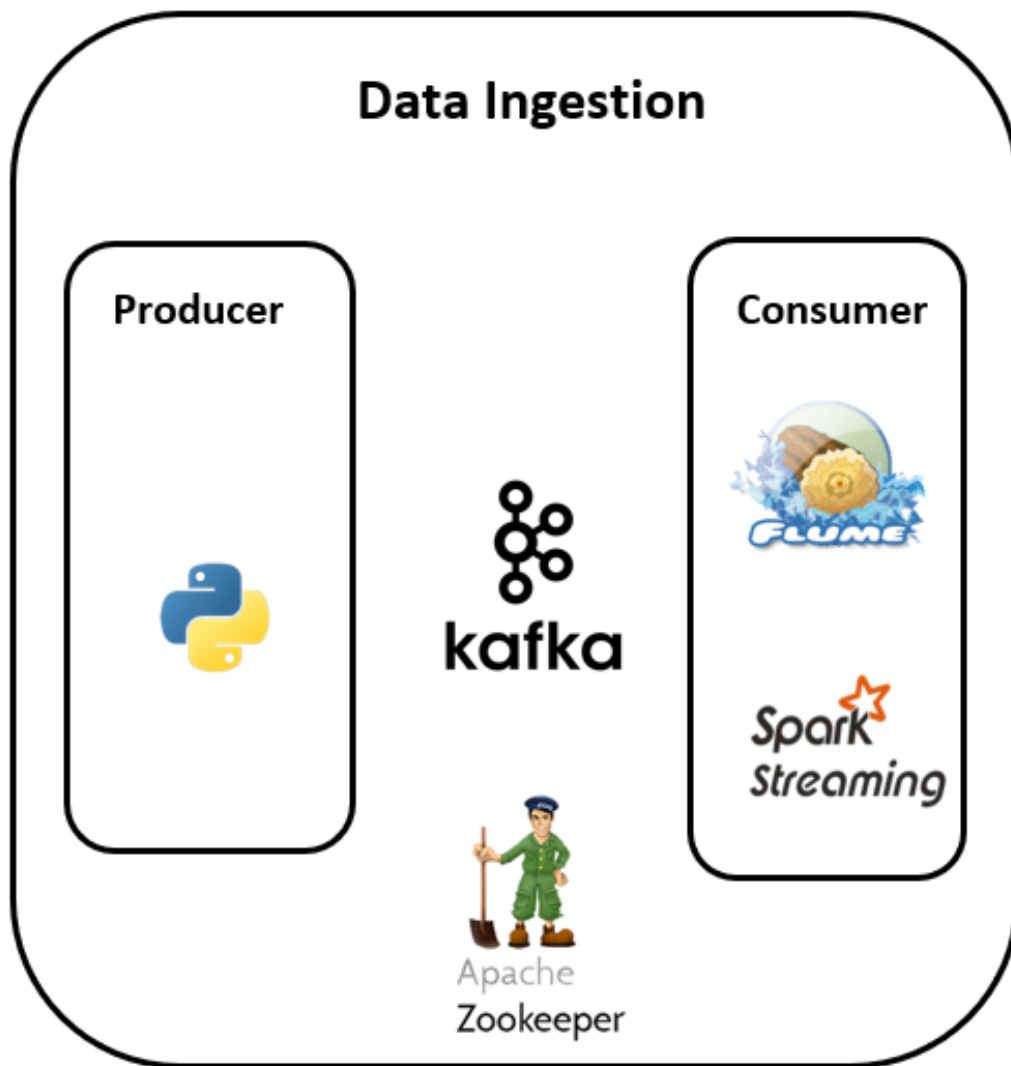


Immagine 2. Dettaglio producer-consumer kafka

In particolare, la logica del producer in scrittura sui topic è stata sviluppata in Python sullo stesso script client che richiedeva i dati da twitter. La parte relativa ai consumer è stata invece gestita sia con spark streaming per il calcolo di alcuni kpi in real-time e con Flume per la scrittura dello stesso dato su HDFS.

A seguire un dettaglio della configurazione di flume:



Immagine 3. Dettaglio source-sink flume

Proseguendo nell'architettura principale, per la parte relativa all'elaborazione batch (box in alto), sono stati utilizzati i seguenti tools: HDFS, come storage persistente distribuito di ogni tipologia di dato (sia streaming che batch), **Spark** (versione 2.3) per una prima fase di elaborazione massiva su tutti i dati comprendente logiche di pulizia, where condition, rimozione duplicati e infine **Python** per la parte relativa alla traduzione dei tweet in lingua inglese, **sentiment analysis** e **NLP**.

Nella parte invece dedicata all'analisi real-time, è stato utilizzato **Spark Streaming** per effettuare delle aggregazioni in finestre temporali (micro batch) di pochi secondi e ottenere delle informazioni in tempo reale quali gli hashtag più ricorrenti istante per istante.

Il livello successivo sulla destra, il **service layer**, rappresenta il punto in cui confluisce il dato elaborato e eventualmente arricchito realmente utile per le nostre analisi, da cui si aggancia il sistema di reportistica. In particolare è stato individuato in **Cassandra** installato su singolo nodo, un valido database NoSql in grado di interfacciarsi in maniera facile con Spark.

Lo strumento di reportistica utilizzato è stato **Tableau Public**, versione free comprensiva di tutte le funzionalità di editing ma limitata in feature di connessione con servizi esterni, che ha reso quindi necessario un'esportazione dei dati calcolati nelle varie pipeline in formato CSV e importati manualmente nello strumento.

Infine per la parte esplorativa dei dati è stato utilizzato il notebook **Apache Zeppelin**, in grado di, tramite diversi interpreti (.sh, spark, r) sviluppare rapide analisi esplorative sui dati.

3 Linguaggi e Librerie

Il linguaggio di programmazione principale utilizzato è stato python.

In particolare è stato utilizzato nelle seguenti fasi:

- **Acquisizione Dati:** utilizzato per configurare i client d'acquisizione da twitter ed eseguire query strutturata di parole-chiave, parole da escludere, lingue da considerare, device da valutare.
- **Streaming dati:** utilizzato per l'utilizzo del framework Spark che ha reso possibile l'analisi di dati in finestre temporali molto brevi.
- **Pre-processing:** utilizzato per la manipolazione e pulizia massiva dei dati tramite framework Spark.
- **Sentiment Analysis e NLP:** utilizzato principalmente per la traduzione dei tweet da italiano ad inglese, analisi del sentiment di ogni tweet e applicazione di tecniche di NLP (es: rimozioni stop words, ecc.)

A seguire si riportano le principali librerie per area e una breve descrizione:

Acquisizione Dati:

LIBRERIA	DESCRIZIONE
TWEEPY	Libreria Python per l'accesso all'API di Twitter.
KAFKA-PYTHON	Client Python per il sistema di elaborazione del flusso distribuito Apache Kafka.
JSON	Libreria per interagire con i JavaScript Object (JSON).

Tab. 3 Librerie utilizzate per acquisizione dati.

Pre-processing:

LIBRERIA	DESCRIZIONE
PYSPARK.SQL	Modulo Spark per l'elaborazione di dati strutturati.
RE	Libreria per utilizzo di espressioni regolari.

Tab. 4 Librerie utilizzate durante la fase di pre-processing.

Streaming dati:

LIBRERIA	DESCRIZIONE
PYSPARK.SQL	Modulo Spark per l'elaborazione di dati strutturati.
PYSPARK.STREAMING	Estensione dell' engine Spark per elaborazioni dati in streaming.
PYSPARK.STREAMING.KAFKA	Modulo di estensione di spark streaming per permettere l'interfacciamento con kafka

Tab. 5 Librerie utilizzate per lo streaming dei dati.

Sentiment Analysis e NLP:

LIBRERIA	DESCRIZIONE
VADERSENTIMENT	Libreria di analisi del sentimento basato su regole e lessico di sentimenti espressi nei social media.
GOOGLETRANS	Libreria che implementa le API di Google Traduttore.
RE	Libreria per utilizzo di espressioni regolari.
SEABORN	Libreria di visualizzazione dei dati Python basata su matplotlib.
WORDCLOUD	Libreria per creare wordcloud
PANDAS	Libreria atta alla manipolazione dati sotto forma di data frame
NUMPY	Pacchetto per il calcolo scientifico con Python
NLTK	Strumento per lavorare con i dati del linguaggio umano (NLP).

Tab. 6 Librerie utilizzate per effettuare NLP e Sentiment Analysis.

4 Dashboard

A valle del processo di analisi, sono state realizzate tre dashboard interattive in ambiente **Tableau**, per la visualizzazione dei risultati.

La prima dashboard [1] presenta alcuni tra i più importanti leader politici italiani del momento (Silvio Berlusconi, Emma Bonino, Luigi Di Maio, Giorgia Meloni, Matteo Salvini, Nicola Zingaretti) tracciati, su un asse **UE – Italia** riferito alle frequenze di citazione dell'una e dell'altra area semantica all'interno dei rispettivi tweet durante la campagna elettorale per le elezioni sul rinnovo del Parlamento europeo. Cliccando sul volto di ciascuno, si accede alla visualizzazione della **wordcloud** delle parole più utilizzate, della **sentiment analysis** dei tweet pubblicati, e infine del tweet più **ricondiviso**.

La seconda dashboard [2] presenta la mappa europea con alcuni Stati evidenziati e segnalati dalla relativa bandiera. Cliccando su di una delle bandiere presenti si accede ad una lista di leader monitorati per nazione, per ognuno è stato creato un grafico che indica l'incremento percentuale di follower giornaliero e un altro che indica l'incremento del numero assoluto di follower giornaliero. Si possono selezionare contemporaneamente tutti i leader di un gruppo nazionale per confrontarne andamenti relativi e assoluti in termini di follower.

La terza dashboard [3] rappresenta la visualizzazione delle opinioni che ha l'elettorato europeo riguardo i trend presenti all'interno della **Tab. 2** e i cinque leader politici italiani monitorati. Per i trend monitorati è stata effettuata una sentiment analysis con rispettiva **wordcloud** per rappresentare le parole ricorrenti in base ai tweet positivi e negativi delle parole monitorate. Sono stati raccolti tweet in lingua inglese ed italiana, quindi si è deciso di visualizzarne i risultati in grafici

separati.

Per quanto riguarda le opinioni che l'elettorato europeo ha sui leader italiani, dopo aver diviso i tweet tra quelli in lingua inglese e quelli in lingua italiana, si è deciso di utilizzare un **side-by-side bars chart** per visualizzare la percentuale di tweet positivi, neutrali e negativi per ogni leader.

5 Risultati

Il monitoraggio del coinvolgimento online da parte dei vari leader politici ha seguito alcune figure principali provenienti da Belgio, Francia, Germania, Grecia, Italia, Regno Unito e Spagna.

Belgio

Il belga **Guy Verhofstadt** si è distinto come presidente dell'ALDE-PDE come liberale e convinto europeista, ha visto una crescita moderata e costante dei propri follower, da 370k a 384k, con picchi percentuali maggiori per il 30 aprile (in seguito al confronto tra candidati alla Presidenza della Commissione Europea) ed il giorno successivo alle elezioni europee, il 27 maggio.

Francia

In Francia sono stati tracciati gli andamenti del Presidente **Emmanuel Macron**, la candidata del *Rassemblement National* **Marine Le Pen**, **Jean-Luc Mélenchon** come esempio politico di sinistra promettente in termini di consenso e infine **François-Xavier Bellamy** leader emergente del centrodestra francese. Macron guida con 3,8 – 4 milioni di follower la crescita maggiore in termini assoluti, notevole anche in termini relativi. Seguono tra i 2 e i 2.25 milioni Le Pen e Mélenchon, mentre si distingue per lo scarso séguito Bellamy tra 32k e 38k follower. Il picco più significativo si ravvisa per Macron, escludendo quello successivo agli exit poll per Le Pen, in data 16 aprile (incendio della cattedrale di Notre Dame). Si noti che le rilevazioni per gli altri candidati sono iniziate il 27 aprile e la serie temporale degli incrementi mediamente più intensi è quella relativa a Bellamy, per cui le oscillazioni sono probabilmente giustificate dalla bassa “inerzia” del bacino di follower.

Germania

La figura politica più importante del Paese, **Angela Merkel**, non possiede un account Twitter ufficiale, questo non ci ha permesso il suo monitoraggio. Sono state monitorate tre leader, la segretaria SPD **Andrea Maria Nahles**, la co-segretaria **Annalena Baerbock** dei Verdi in ascesa e **Alice Weidel** della nuova destra di AfD. A sorpresa, data la loro età e la crescita dei due movimenti, queste ultime si posizionano rispettivamente tra 33 e 38k, e tra 68k e 71k follower. Nahles guida con più di 2 milioni di follower, con una leggera crescita nel periodo complessivo dal 27 aprile al 4 giugno, caratterizzato però da una flessione in data 10 maggio e ben 6 giorni di perdita di follower nella settimana precedente il 26 maggio. In questo periodo si riscontra l'anniversario dell'alleanza di governo con il centro democristiano ed il rilancio di un programma socialista ritenuto insufficiente o sfumato da attivisti ed elettori giovani e tendenti all'ecologismo. Il 2 giugno Nahles si è dimessa, mantenendo un saldo positivo giornaliero quanto a follower del suo account.

Grecia

Con più di 950k follower, l'ex Ministro dell'Economia **Yanis Varoufakis** ha un account potenzialmente interessante, cui si aggiunge il progetto di un movimento politico transnazionale, DiEM25, che però non si è manifestato ad esempio in Italia, e ha eletto infine un solo membro del Parlamento. Varoufakis presenta una crescita di quasi 8k follower tra 11 aprile e 4 giugno, crescita però non monotona, positiva ma in flessione nel weekend delle elezioni ed infine una decrescita dal 30 maggio.

Regno Unito

Il Regno Unito ha partecipato alle ultime elezioni europee, se non a sorpresa, contro le intenzioni di una porzione consistente della popolazione avente diritto e della classe politica incaricata di portarle concretamente a termine (la permanenza del Regno Unito in UE è stata prorogata, ufficialmente l'11 aprile, dal 29 marzo al 31 ottobre 2019). I capi politici monitorati sono **Jeremy Corbyn**, leader dei Laburisti, considerato ambiguo rispetto all'uscita o permanenza del Regno Unito all'interno dell'UE, **Theresa May**, Prime Minister espressa dai Conservatori e incaricata di trattare i termini della “Brexit”, **Nicola Sturgeon**, PM della Scozia e tra le principali figure politiche europeiste in UK, e infine **Nigel Farage**, principale promotore della Brexit e candidato con un partito eponimico (rivelatosi il più votato) proprio al Parlamento Europeo.

Il bacino di follower per Corbyn è il maggiore, vicino ai 2 milioni, seguito da Farage a circa 1.3 milioni, Sturgeon attorno a 970k e May oltre 850k. Gli incrementi assoluti sono contenuti fra circa 8k (Sturgeon) e quasi 45k (Farage), con picchi percentuali molto simili per entità e sincronia, con due eccezioni particolari: l'aumento di 0.85% per May il 24 maggio (segue Farage con 0.21%) e l'aumento di 0.58% di Farage il 27 maggio (segue Sturgeon con 0.11%). In queste due date si registrano rispettivamente la vigilia del voto dell'House of Commons sull'ultima versione dell'accordo di uscita proposto da May, e successivamente la pubblicazione dei risultati generali del voto, come anticipato, favorevoli a Farage.

Spagna

In Spagna si sono svolte le elezioni generali il 28 aprile, e le elezioni europee il 26 maggio. I leader monitorati sono i capi politici dei primi quattro partiti per numero di voti, in entrambe le elezioni. In ordine di voti ottenuti: **Pedro Sánchez** (PSOE), **Pablo Casado Blanco** (PP), **Albert Rivera** (Ciudadanos) e **Pablo Iglesias** (Unidos Podemos). La situazione spagnola vede quindi due partiti di centrosinistra e centrodestra tradizionali in risalita, una formazione liberale ed una populista-socialista-ecologista più recenti. Il seguito su twitter è invece dominato da Iglesias con circa 2.3 milioni di follower, Rivera con 1.3 milioni, Sánchez appena oltre 1mln ed infine Casado Blanco quasi a 250k.

Tutti hanno vissuto il maggior incremento durante le elezioni generali (+1.1% Sánchez, quasi +0.6% in due giorni consecutivi per Casado), successivamente il 9 maggio più che il 26, probabilmente per le trattative e le alleanze di governo (nessuna lista ha ottenuto la maggioranza assoluta). Le elezioni europee hanno confermato le relazioni d'ordine, con un aumento di consensi percentuali dei due partiti tradizionali ed un leggero peggioramento per i due partiti più recenti, senza fluttuazioni notevoli nella piattaforma.

Italia: seguito e sentiment analysis

In Italia **Matteo Salvini** (Viceministro, Ministro dell'Interno, segretario della Lega) guida sia quanto a follower totali sia quanto ad incremento (da 1.023 a 1.066 milioni, una crescita di 43k unità), in linea con l'elevata percentuale di consensi rilevati alle urne. Di contro, a classificarsi come seconda per numero di follower, si trova **Giorgia Meloni** (Fratelli d'Italia), tra i 750k e i circa 770k follower nel periodo di rilevazione. Sopra i 500k si stabilisce **Luigi Di Maio** (Viceministro, Ministro per lo Sviluppo Economico, capo politico M5S), poi **Nicola Zingaretti** (neoeletto segretario del Partito Democratico) attorno a 420k follower. Infine, i più anziani, anagraficamente e politicamente, **Emma Bonino** (+Europa) e **Silvio Berlusconi** (Forza Italia), si attestano l'una attorno ai 190k e l'altro intorno ai 57k. Nessun leader ha visto una perdita netta di follower ed i picchi incrementali sono piuttosto sincronizzati. I maggiori si ravvisano il giorno successivo alle elezioni (Berlusconi e Salvini oltre +0.3%), e secondariamente il 20 maggio (giorno di scontri a Bologna tra manifestazioni neofasciste e contromanifestazioni, ricovero per Giachetti a causa dello sciopero per i fondi pubblici a Radio Popolare, nuova stesura del decreto "Sicurezza-bis").

Per quanto riguarda la sentiment analysis dei tweet pubblicati dagli stessi leader, si può riportare parzialmente l'umore citando il podio di chi ha la più alta percentuale di termini positivi: Berlusconi, Zingaretti e Salvini rispettivamente a 58%, 51% e 43%; il podio per termini negativi vede invece Meloni, Di Maio e Berlusconi con 35%, 32% e 28%. Salvini si segnala anche per la maggior percentuale di termini neutrali (37%), secondo lo standard in uso.

Quanto a massimo coinvolgimento degli utenti, Emma Bonino supera gli altri politici citati con 1904 retweet di record personale, equivalenti allo 0.98% dei suoi follower, seguita da Salvini con 1738 retweet, equivalenti invece allo 0.16% dei suoi follower (ricordando che i follower non sono gli unici autori dell'operazione di retweet).

Sentiment della comunità

Per quanto riguarda il "sentiment" degli utenti, si è raccolto il flusso di tweet in italiano ed inglese seguendo tanto le menzioni dei politici italiani considerati in precedenza, quanto le menzioni di alcune parole-chiave (**Tab. 2**) relative ad istituzioni e politiche europee.

I punteggi migliori si registrano dai tweet in lingua inglese per Emma Bonino (57%) e in lingua italiana per Giorgia Meloni (45,83%). Per Salvini, il punteggio è comunque migliore per i tweet in inglese.

Osservando invece i risultati dei tweet in inglese verso i topic europei, si nota l'uso di termini legati a Brexit, UK, così fra i flussi a polarità positiva, come fra quelli a polarità negativa. Si potrebbe desumere che, almeno su twitter, l'inglese non sia lingua franca europea per trattare temi comuni e comunitari, o che perlomeno questo ruolo sia comunque oscurato dalla quantità di commenti circa la situazione originale del quasi completamente ex-Paese membro.

6 Conclusioni e prospettive future

L'architettura ed il paradigma di analisi qui presentati si prestano ad un'estensione sia nei volumi di dati trattati, sia nel tempo di monitoraggio, e conseguentemente ad una flessibilità e varietà di topic osservati e una maggiore intensità e profondità delle analisi statistiche, testuali e di supporto giornalistico per commentari sociologici e politologici, fuori dallo scopo primario delle operazioni svolte.

I testi dei tweet raccolti potrebbero costituire un data set di valore per il **natural language processing** e le attività nel campo della **data semantics**, quali named entity recognition e sviluppo di word embeddings per valutare nei termini della semantica distribuzionale il linguaggio delle comunità online e dei loro leader.

Bibliografia, link e riferimenti

[1] link prima dashboard :

<https://public.tableau.com/profile/davide.mancino5338#!/vizhome/DatiLeaderItalia/Europaitalia>

[2] link seconda dashboard : <https://public.tableau.com/profile/davide.mancino5338#!/vizhome/Europee/Europe>

[3] link terza dashboard : <https://public.tableau.com/profile/davide.mancino5338#!/vizhome/elettoriEuropei/People>