

IZI SpA

**Metodi, analisi e
valutazioni
economiche**

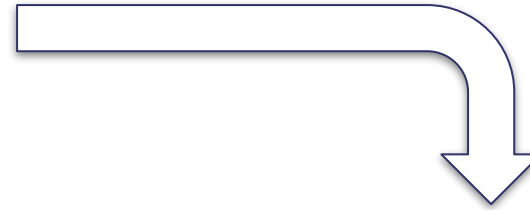
Chi siamo

IZI Spa

1984, Fondata a Roma, si occupa da quasi quarant'anni di supportare la Pubblica Amministrazione:

- ❖ nelle attività di **programmazione, gestione ed attuazione di politiche e programmi cofinanziati dai Fondi Strutturali**;
- ❖ nella realizzazione di **ricerche socio-economiche**, nella realizzazione delle **indagini statistiche**, nelle **ricerche di mercato e dei sondaggi di opinione**, con particolare riferimento al **monitoraggio della qualità dei servizi**.

L'integrazione tra **metodi di ricerca consolidati e tool tecnologici di ultima generazione** ci consente di garantire **un supporto qualificato e al passo con i tempi**.



IZILab

2020, nasce come **divisione digitale** di IZI Spa con l'intento di implementare le metodologie tradizionali di analisi e ricerca con **strumenti innovativi e nuove tecnologie, quali deep learning e AI** al fine di raccogliere, analizzare e processare **grandi set di dati, contenuti testuali e multimediali**.

I NUOVI DRIVER DELLA RICERCA

**BIG DATA
OPEN DATA
MACHINE LEARNING
ALGORITMO
DATA VISUALIZATION**

Cosa facciamo per Isfort

Dal 2018 collaboriamo nella realizzazione dell'indagine nazionale più importante sulla mobilità che viene svolta sulla domanda di mobilità e trasporti nel Paese.

Audimob "Osservatorio su stili e comportamenti di mobilità degli italiani"

L'indagine interessa un campione statisticamente significativo della popolazione italiana compresa fra 14 e 84 anni, stratificato per regione, sesso, ampiezza demografica del comune di residenza e per classi di età dell'intervistato. Le interviste vengono effettuate con un'equa distribuzione in tutti i giorni della settimana (campione feriale e festivo).



...qualche numero

- ★ realizziamo più di 17.000 interviste CATI/CAWI all'anno;
- ★ puliamo ed elaboriamo i dati fornendo un dataset finale contenente oltre 17.000 record;
- ★ elaboriamo informazioni in merito al tasso di mobilità, alla numerosità e lunghezza degli spostamenti ai mezzi di trasporto utilizzati e alle motivazioni degli spostamenti;
- ★ gestiamo un database costituito da 15,5 milioni di dati strutturati (oltre 17K record per quasi 1.000 campi).





Tecnologie e Metodologia della Ricerca



Progetto *Il Sentiment della Mobilità*

OGGETTO DI STUDIO

Indagine pilota sul tema della mobilità, online e sui social. **Chi** e **come** parla di **mobilità** sul web e su Twitter.

PERIODO DI RIFERIMENTO

Primo trimestre 2022.

PROGETTO

- **Ricerca nazionale**
Ricerca su tutto il territorio **nazionale**, tramite parole chiave legate alla mobilità e al trasporto pubblico.
- **Ricerca città campione**
Casi studio su **tre città** italiane: Milano, Bologna, Napoli.

1.451.740



8.780



175.273



35.354

47.054

Oggetto di studio

PAURA DEL CONTAGIO

QUALITÀ DEL SERVIZIO

**TRASPORTO
PUBBLICO**

TRAFFICO

DISPOSITIVI DI
INFOMOBILITÀ

INQUINAMENTO

MOBILITÀ URBANA

MOBILITÀ PEDONALE E
CICLISTICA

USO DEI MONOPATTINI

Fasi della ricerca

Definizione della domanda di ricerca

Scelta della metodologia
Scelta delle parole chiave per la query

Raccolta dati

Tweets tramite API ufficiali di Twitter
Articoli web tramite web scraping

Analisi dati & Deep Learning

KPIs
Top Words Correlation
Topic Analysis
Sentiment Analysis
Emotion Analysis
Trend Analysis
Tweet Geolocalization

Data visualization

Tool con dashboard interattiva

Definizione della domanda di ricerca

Definizione della domanda di ricerca

Scelta della metodologia
Scelta delle parole chiave per la query

Raccolta dati

Tweets tramite API ufficiali di Twitter
Articoli web tramite web scraping

Analisi dati & Deep Learning

KPIs
Top Words Correlation
Topic Analysis
Sentiment Analysis
Emotion Analysis
Trend Analysis
Tweet Geolocalization

Data visualization

Tool con dashboard interattiva

Scelta della metodologia

INDAGINE NAZIONALE

QUERY TRAMITE PAROLA CHIAVE

Per l'indagine nazionale abbiamo deciso di scaricare tutti i tweet e gli articoli web che contenevano alcune parole d'interesse legate alla mobilità.
Questo ha permesso di studiare il flusso di conversazioni in merito alla mobilità su scala nazionale.

INDAGINE sulle CITTÀ

QUERY ZONALE

Lo studio in merito alle città è stato invece effettuato raccogliendo tutti i contenuti pubblicati nelle tre città di interesse e successivamente isolando quelli che parlavano di mobilità (c.d. ricerca non condizionata).
Questo ci ha permesso di avere un'indicazione, a differenza dell'indagine nazionale, della mole di contenuti riguardanti la mobilità rispetto al totale dei contenuti della città indagata.
(Questo tipo di indagine non è stata possibile per gli articoli web, in quanto non contengono l'informazione geospaziale necessaria.)

Scelta delle parole chiave

Le **keywords** sono state scelte dopo un'accurata analisi di tutti gli ambiti potenzialmente rilevanti per il committente e delle parole interessanti ai fini della ricerca.

Twitter

a piedi, piedibus, pedibus, pedoni, bici, bicicletta, bicibus, monopattino, monopattini, auto, automobile, automobili, macchina, macchine, moto, motorino, motorini, scooter, taxi, parcheggio, parcheggi, traffico, ingorgo, ztl, autostrada, autostrade, tangenziale, area di sosta, aree di sosta, area di servizio, casello, caselli, telepass, code a tratti, trasporto pubblico, trasporti pubblici, mezzo di trasporto, mezzi di trasporto, trasportopubblico, trasportipubblici, mezzi pubblici, autobus, filobus, bus, pullman, tram, metro, metropolitana, treno, treni, trenino, stazione, stazioni, traghetto, traghetti, green mobility, car sharing, car pooling, colonnina di ricarica, colonnine di ricarica, mobilità sostenibile, mobilità dolce, bike sharing, pendolare, pendolari, pendolarismo, greenmobility, carsharing, carpooling, ncc

Web articles

trasporto trasporti trasportopubblico
treno treni autobus filobus bus pullman
tram metropolitana stazione traghetto
traghetti monopattino monopattini bici
bicicletta pedoni mobilityaservice
automobile automobili taxi moto
motorino scooter carsharing carpooling
greenmobility

Raccolta dati

Definizione della domanda di ricerca

Scelta della metodologia
Scelta delle parole chiave per la query

Raccolta dati

Tweets tramite API ufficiali di Twitter
Articoli web tramite web scraping

Analisi dati & Deep Learning

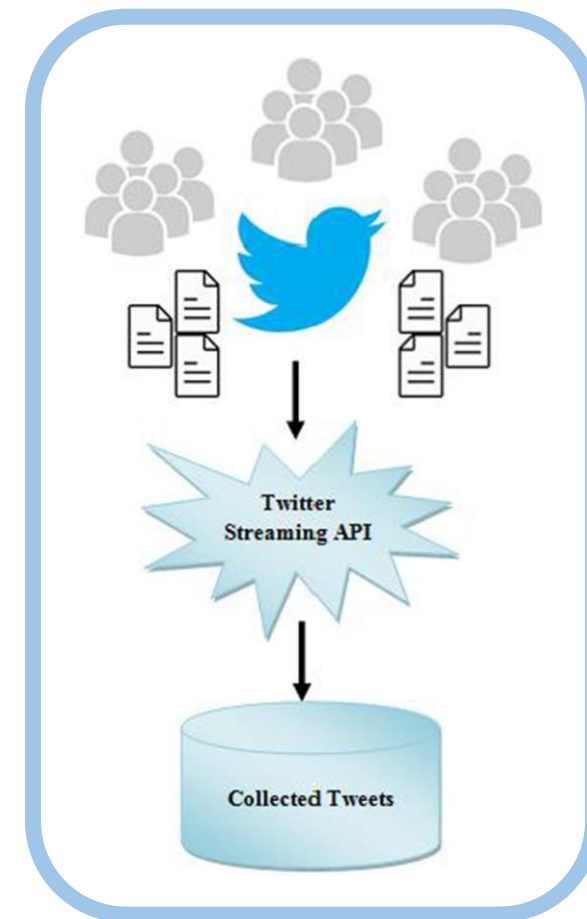
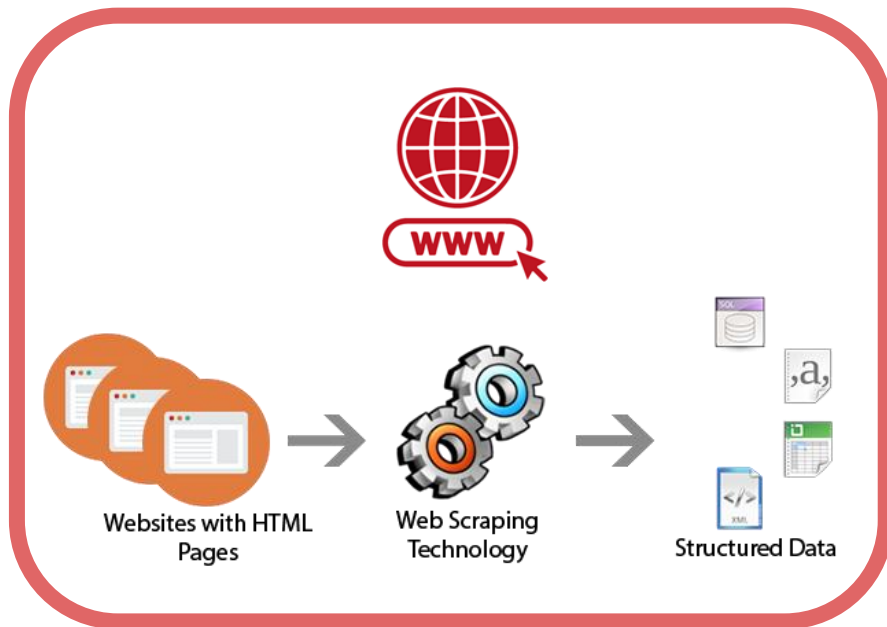
KPIs
Top Words Correlation
Topic Analysis
Sentiment Analysis
Emotion Analysis
Trend Analysis
Tweet Geolocalization

Data visualization

Tool con dashboard interattiva

Data scraping

Per scaricare i tweet abbiamo utilizzato le API ufficiali di Twitter. Esse permettono di recuperare il testo del tweet partendo dalle parole chiave o dalla geolocalizzazione ma forniscono molte altre informazioni quali la data di pubblicazione, il numero di like e di share ricevuti e molto altro.



Per scaricare gli articoli web abbiamo utilizzato una tecnica di web scraping che consiste nell'analizzare la struttura di una pagina web, studiarne il contenuto, e scaricare la parte testuale (quando siano presenti le parole chiave prestabilite).

Analisi dati & Deep Learning

Definizione della domanda di ricerca

Scelta della metodologia
Scelta delle parole chiave per la query

Raccolta dati

Tweets tramite API ufficiali di Twitter
Articoli web tramite web scraping

Analisi dati & Deep Learning

KPIs
Top Words Correlation
Topic Analysis
Sentiment Analysis
Emotion Analysis
Trend Analysis
Tweet Geolocalization

Data visualization

Tool con dashboard interattiva

Tipi di dati da analizzare

DATI TESTUALI:
contenuto dei tweet e degli
articoli web



NLP TECHNIQUES



Clustering

**Interpretazione dei
cluster**

Topic Analysis

Train del modello

Classificazione

**Sentiment &
Emotion Analysis**

**Varie informazioni scaricate
da Twitter**



Dato temporale



Trend Analysis

Dato spaziale



Geolocalized tweets on map

Like, n° di commenti, share etc.



Metriche e KPIs

Dati testuali: encoding

Il testo viene codificato in base alla sua semantica grazie all'utilizzo dei **Transformers**, reti neurali profonde che rappresentano lo stato dell'arte del Natural Language Processing.

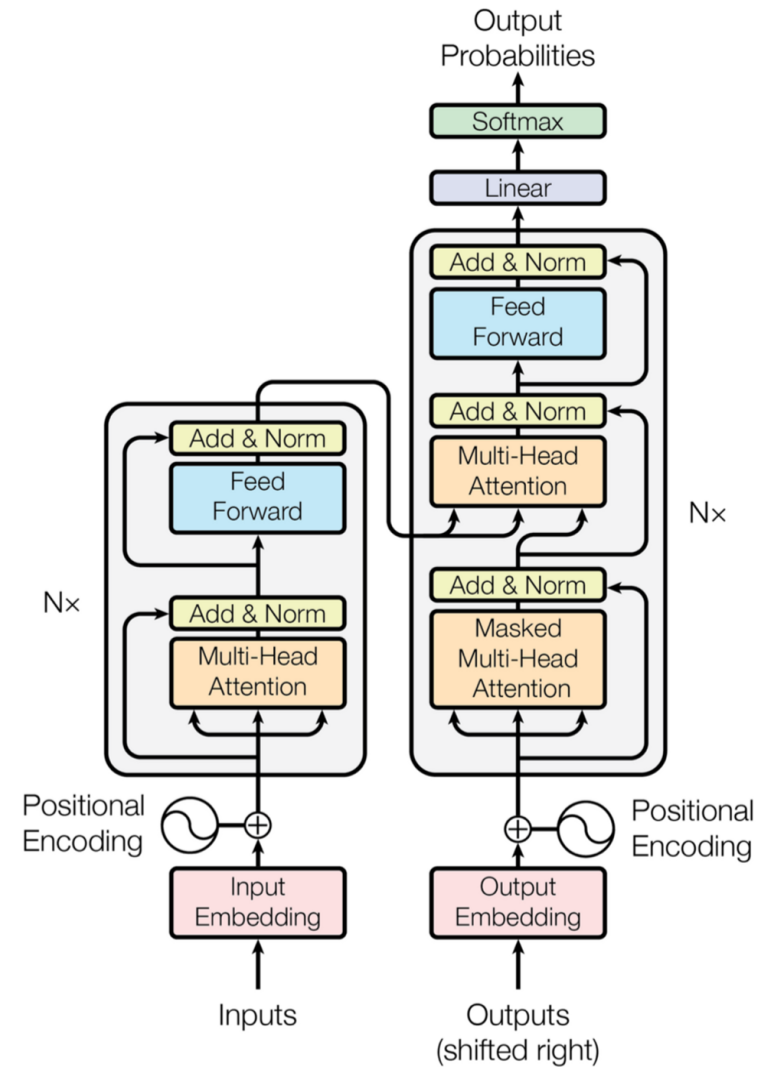
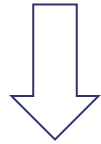


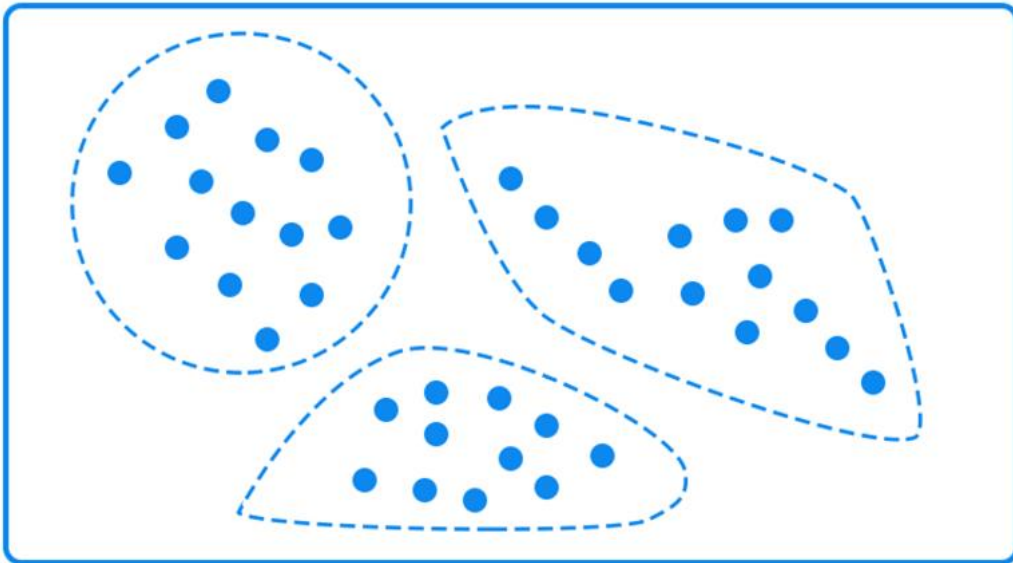
Figure 1: The Transformer - model architecture.

Dati testuali: analisi e algoritmi

Topic Modeling



Clustering



Per l'individuazione dei temi, **topic**, di cui si parla abbiamo utilizzato una **tecnica di machine learning non supervisionato** che permette di evidenziare le naturali aggregazioni dei dati, ovvero temi sorti spontaneamente nelle conversazioni.

Dopo aver creato (grazie ai Transformers) per ogni testo un vettore numerico che ne codifica il contenuto semantico, **i testi sono stati clusterizzati per similarità**.

In seguito abbiamo studiato le aggregazioni emerse per **interpretare quale fosse la semantica comune ai testi** di ogni cluster.

Abbiamo così individuato e riaggregato i **cluster relativi alla mobilità** e al trasporto pubblico e ne abbiamo analizzato il contenuto.

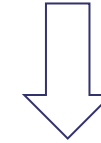
Dati testuali: analisi e algoritmi

Per l'analisi del sentiment e delle emozioni, invece, abbiamo utilizzato una tecnica di **machine learning supervisionato**: abbiamo utilizzato algoritmi di classificazione pre-allenati su grandi moli di dati etichettati, in base al **sentiment positivo o negativo e alle 4 emozioni di base** (rabbia, paura, gioia e tristezza).

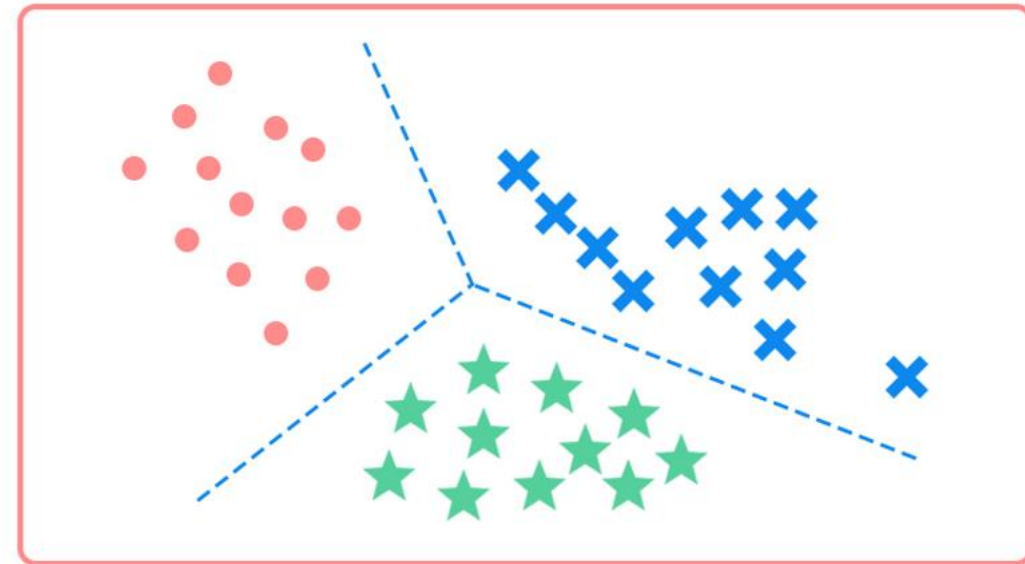
Gli algoritmi sono dunque in grado di classificare nuovi testi secondo le categorie imparate in fase di train.

Tali algoritmi sono stati allenati specificamente sulla lingua italiana e su post di Twitter.

**Sentiment &
Emotion Analysis**



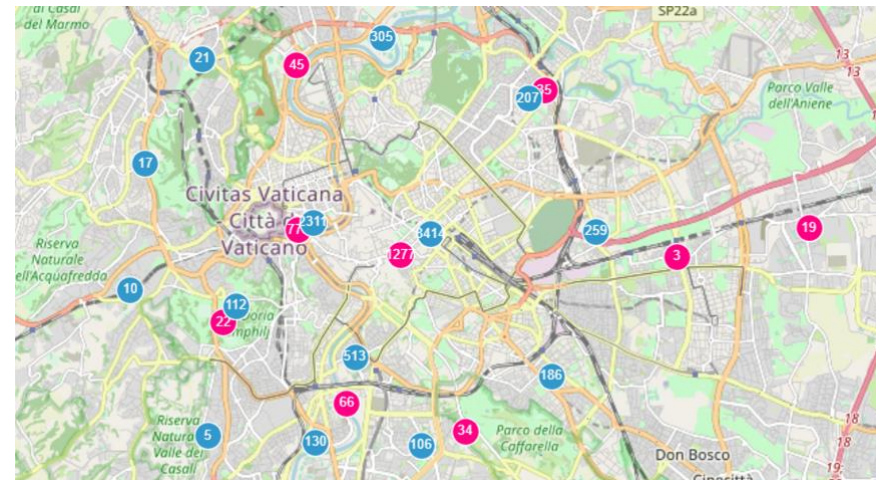
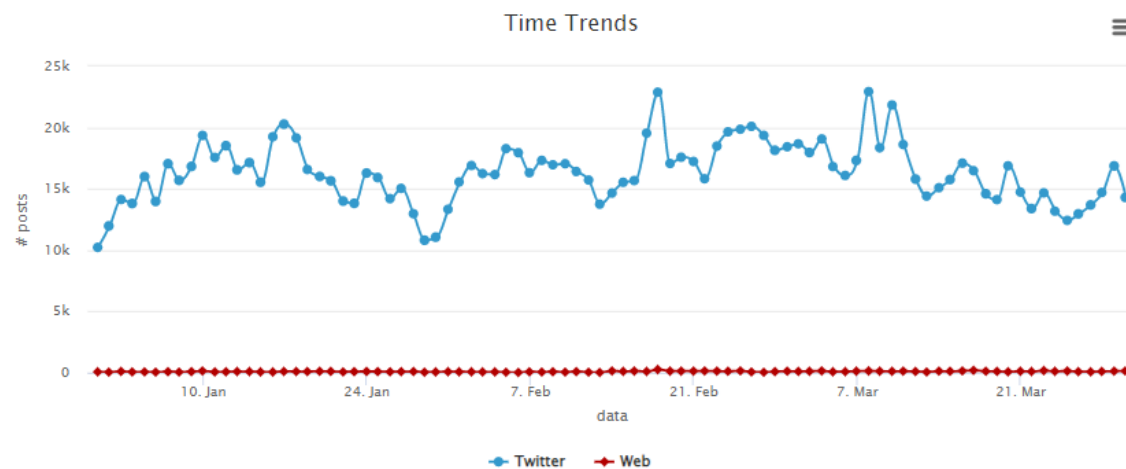
Classification



Altre analisi: spazio-temporale

Finora abbiamo visto le tecniche di analisi del contenuto testuale (applicabili sia ai testi dei tweet che degli articoli web scaricati).

Abbiamo però altre informazioni interessanti da poter analizzare con il supporto di grafici: l'**informazione temporale**, ovvero la data di pubblicazione del contenuto (disponibile sia per i tweet che per gli articoli web), l'**informazione spaziale** (disponibile solo per i tweet pubblicati con geolocalizzazione attiva).



Altre analisi: KPIs


In più, Twitter fornisce dati utili come il **numero di like**, il **numero di condivisioni** e il **numero di commenti**, che possono essere presi in considerazione per calcolare l'**engagement** prodotto da certi contenuti, ma è anche possibile risalire ad altre informazioni come il **numero di utenti unici** che hanno parlato di un certo argomento.



1.451.740

Tweets

Metric: Utenti

 252.799

Metric: Likes

 4.346.265

Sentiment: Posts


28,4%


71,6%

Engagement

Likes+Comment+Share

 195.437.385

Data visualization

Definizione della domanda di ricerca

Scelta della metodologia
Scelta delle parole chiave per la query

Raccolta dati

Tweets tramite API ufficiali di Twitter
Articoli web tramite web scraping

Analisi dati & Deep Learning

KPIs
Top Words Correlation
Topic Analysis
Sentiment Analysis
Emotion Analysis
Trend Analysis
Tweet Geolocalization

Data visualization

Tool con dashboard interattiva


Dashboard interattiva

Tutte le analisi finora illustrate, sono visualizzabili tramite il **tool proprietario sviluppato da IZILab**, che permette di consultare i dati a disposizione in maniera rapida, intuitiva ed interattiva, selezionando sotto-porzioni di dati di interesse e guadagnando nuovi insights sui temi studiati!

 **1.451.740**
Tweets

 **8.780**
Articoli

Metric: Utenti

 252.799

Metric: Likes

 4.346.265

Sentiment: Posts

 28,4%

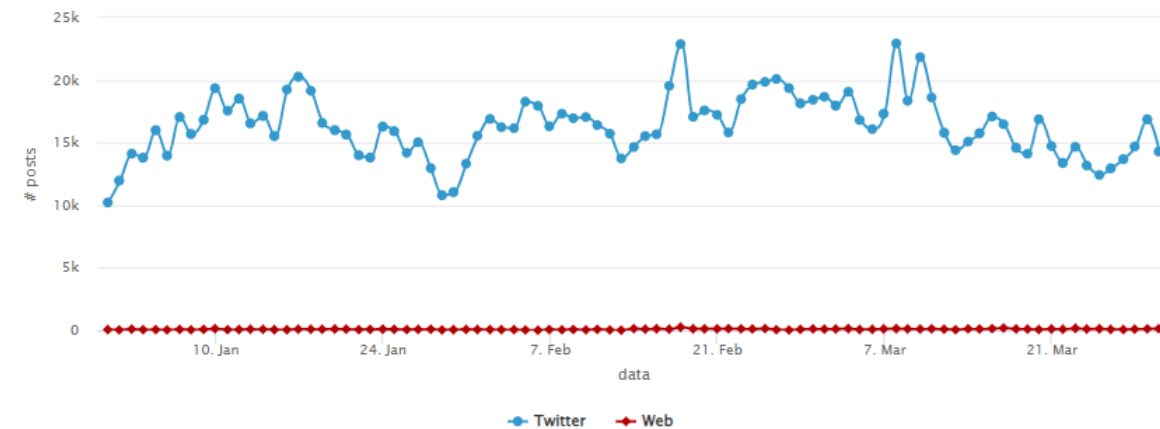
 71,6%

Engagment

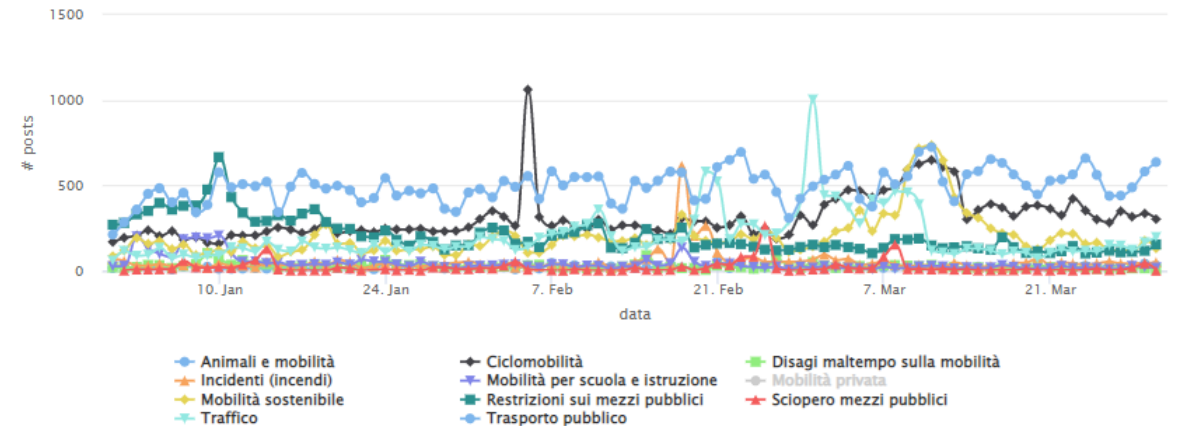
Likes+Comment+Share

 195.437.385

Time Trends



Twitter Topic Trends



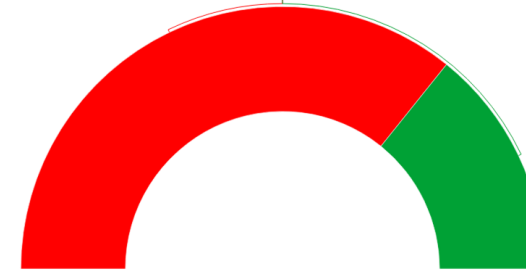
Dashboard interattiva

Topic emergenti

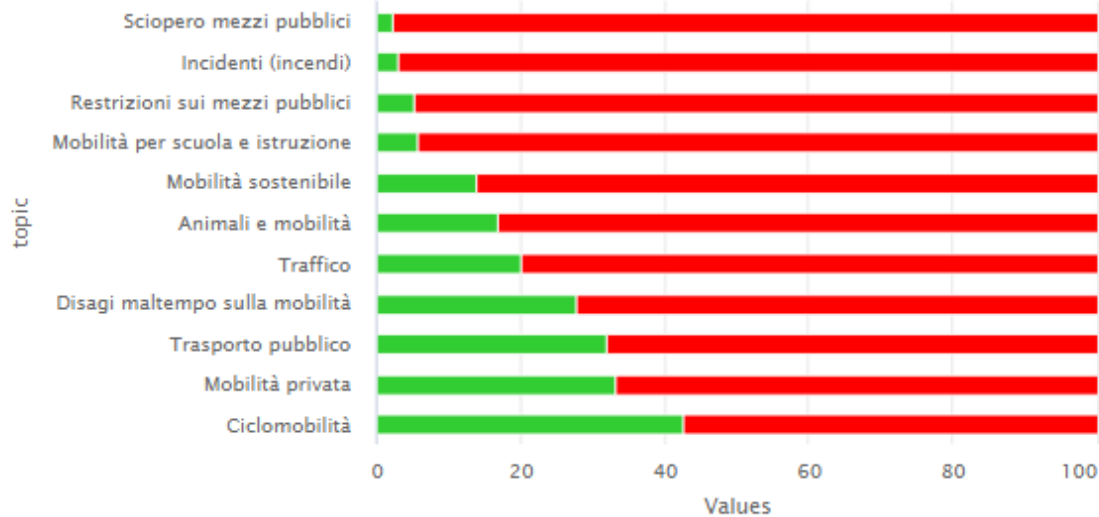


Sentiment

Negative 71.57% Positive 28.43%

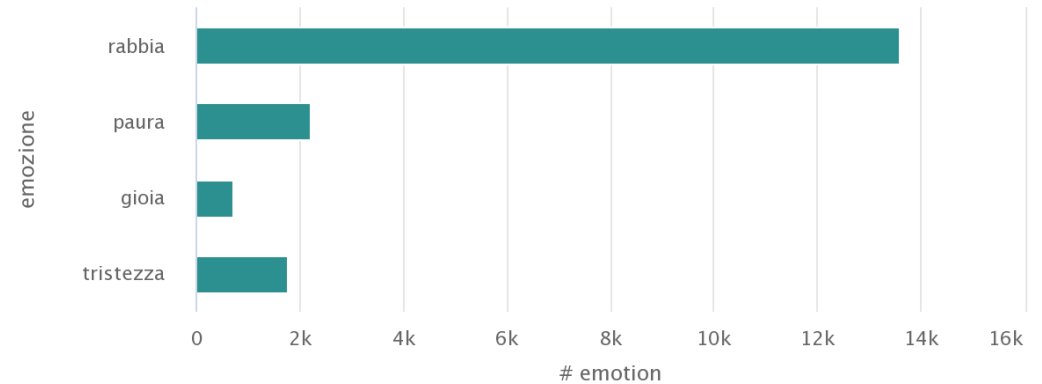


Twitter Sentiment per Topic



● negativo ● positivo

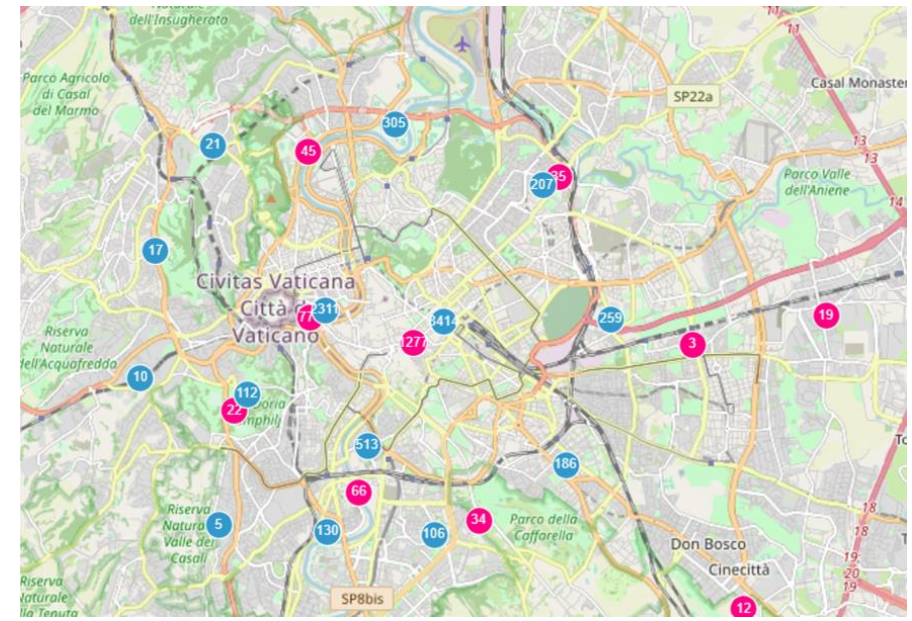
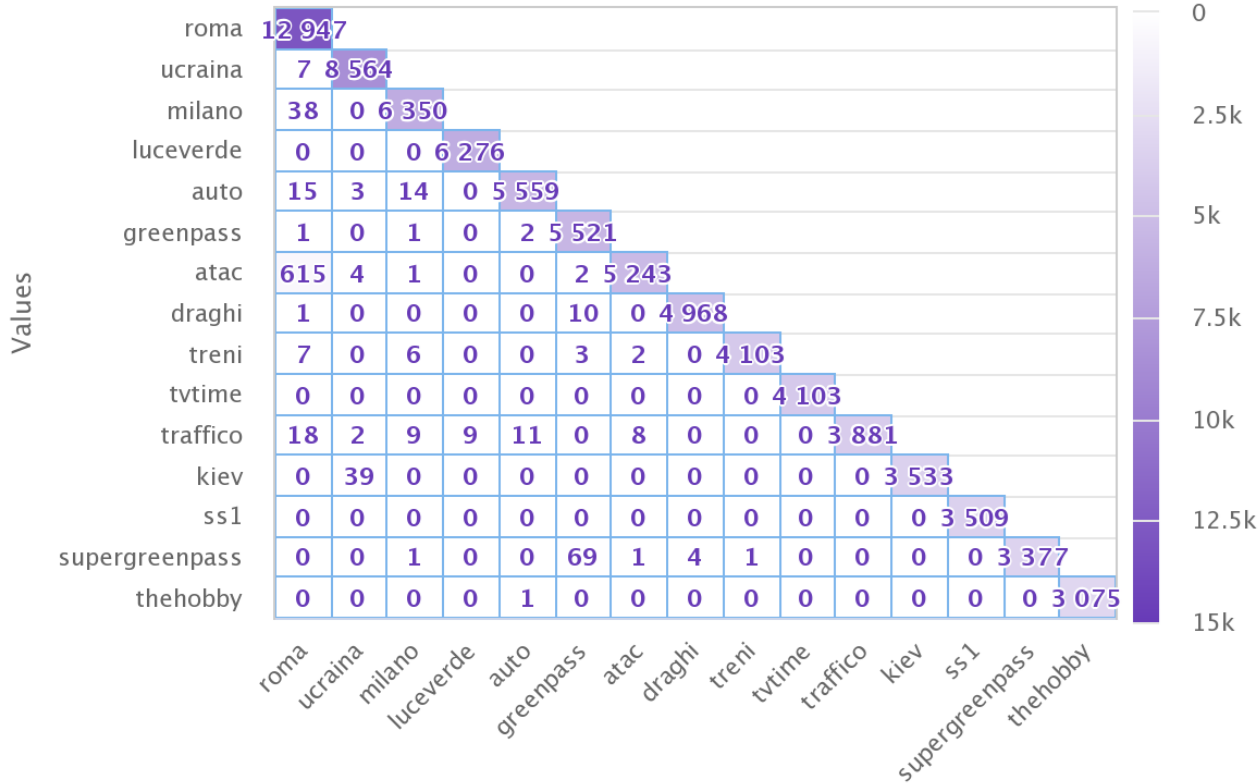
Twitter Topic Emofy



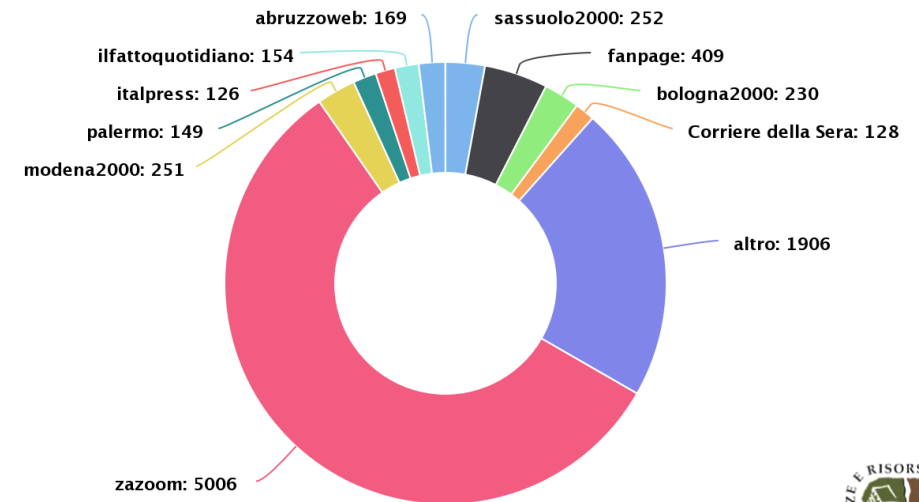
- Animali e mobilità
- Disagi maltempo sulla mobilità
- Mobilità per scuola e istruzione
- Mobilità sostenibile
- Sciopero mezzi pubblici
- Trasporto pubblico
- Ciclomobilità
- Incidenti (incendi)
- Mobilità privata
- **Restrizioni sui mezzi pubblici**
- Traffico

Dashboard interattiva

Matrice di Correlazione fra le parole chiave



Web sources





Grazie dell'attenzione