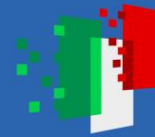




Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Big Data and AI - Social and Evaluation Research

Vincenzo Miracula
University of Catania
Workshop session



FOSSR

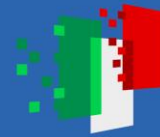
Fostering Open Science in Social Science Research
Innovative tools and services to investigate economic and societal change



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Main contents



Lectures on the basic principles of
evaluation research



Discussion Class on AI-BD and
Evaluation



Hands-on exercises on real cases



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Aims



Digital platforms for data collection - where and how?



AI tools for data analysis – sentiment and emotion detection



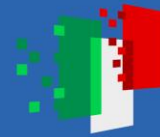
Group work on real cases



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Main aims

Four meetings left



Digital platforms & data collection exercise



AI tools & data analysis exercise



Presentation of selected topics and group work



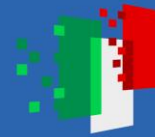
Group essay presentation (*see essay structure document*)



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Big Data and Evaluation



In today's hyper-digitized reality, individuals and groups constantly leave behind digital traces of their behaviors. Big Data, collected via interconnected platforms, reveals hidden patterns and trends. AI systems, particularly ML algorithms, harness this data to generate valuable insights for understanding complex social systems.



Today's evaluations must account for digital contexts alongside physical environments

- Social and economic interactions now occur through global, interconnected digital platforms
- Social Media (Facebook, Twitter, YouTube..)
- Digital Communities (Reddit, TikTok, Twitch...)
- Online Behavioral Data (likes, comments, search activity...)



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

The main aim is to highlight the value AI technologies can add to analyzing and evaluating complex social phenomena offering a venue for methodological innovations and the development of tools geared toward understanding and assessing large-scale complex social events

This is a necessary step to intercept the mechanisms resulting from participating in debates within the *digital agora*.





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Industry Giants (Most Popular and Globally Recognized Social Networks)

Platform	User
Facebook	3.065 billion monthly active users
YouTube	2.504 billion monthly active users.
Instagram	2 billion monthly active users.
WhatsApp	2 billion monthly active users.
WeChat	1.343 billion monthly active users.
Twitter (X)	Over 400 million users.
LinkedIn	More than 900 million active users.
Snapchat	800 million monthly active users
TikTok	1.582 billion monthly active users
Pinterest	450 million monthly active users
Reddit	430 million monthly active users



Emerging Social Platforms (Growing and Expanding Followings)

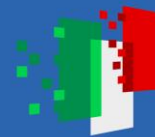
Platform	User
Threads	275 million monthly active users
Mastodon	13 million monthly active users.
Bluesky	15 million monthly active users.
Telegram	900 million monthly active users.
Discord	700 million monthly active users.
Twitch	140 million users.
Vimeo	170 million active users.
Kuaishou	500 million monthly active users
Douyin	750 million monthly active users
Strava	100 million monthly active users
Goodreads	90 million monthly active users



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



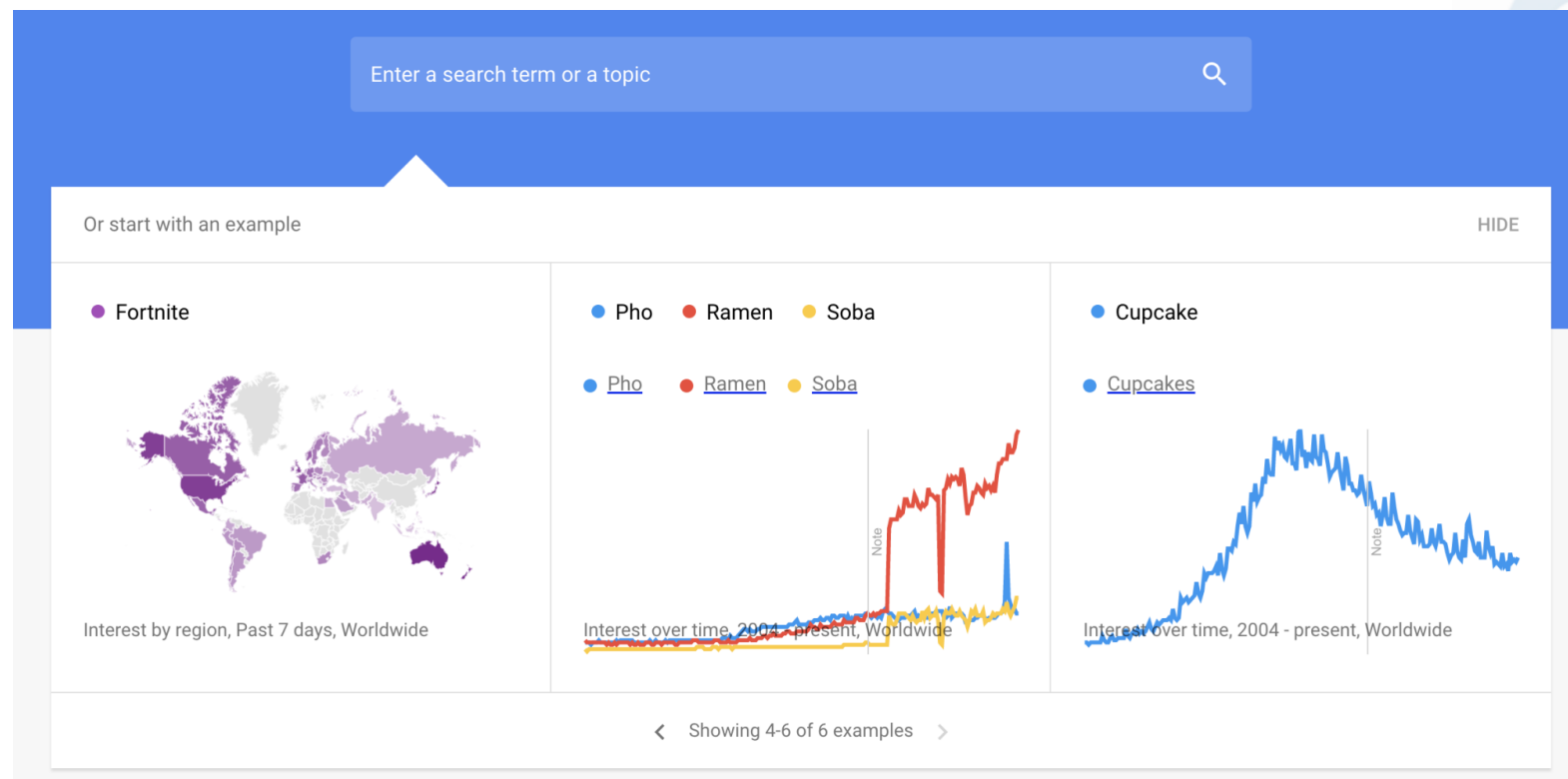
Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Google Trends

Google Trends allows users to see the relative popularity of search terms over time, providing insight into the public's interest in a particular topic





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Google Trends provides a time series index of the volume of queries users enter into Google. The maximum share of queries in a given time period is normalised to 100.

Queries such as "nuclear threat" are counted in the calculation of the query index for "nuclear".

Note that the Google Trends data are calculated using a sampling method, so the results vary by a few percentage points from day to day.

The Role of Twitter and Google Trends in Identifying the Perception of Russia-Ukraine Wars

<https://riunet.upv.es/handle/10251/201800>

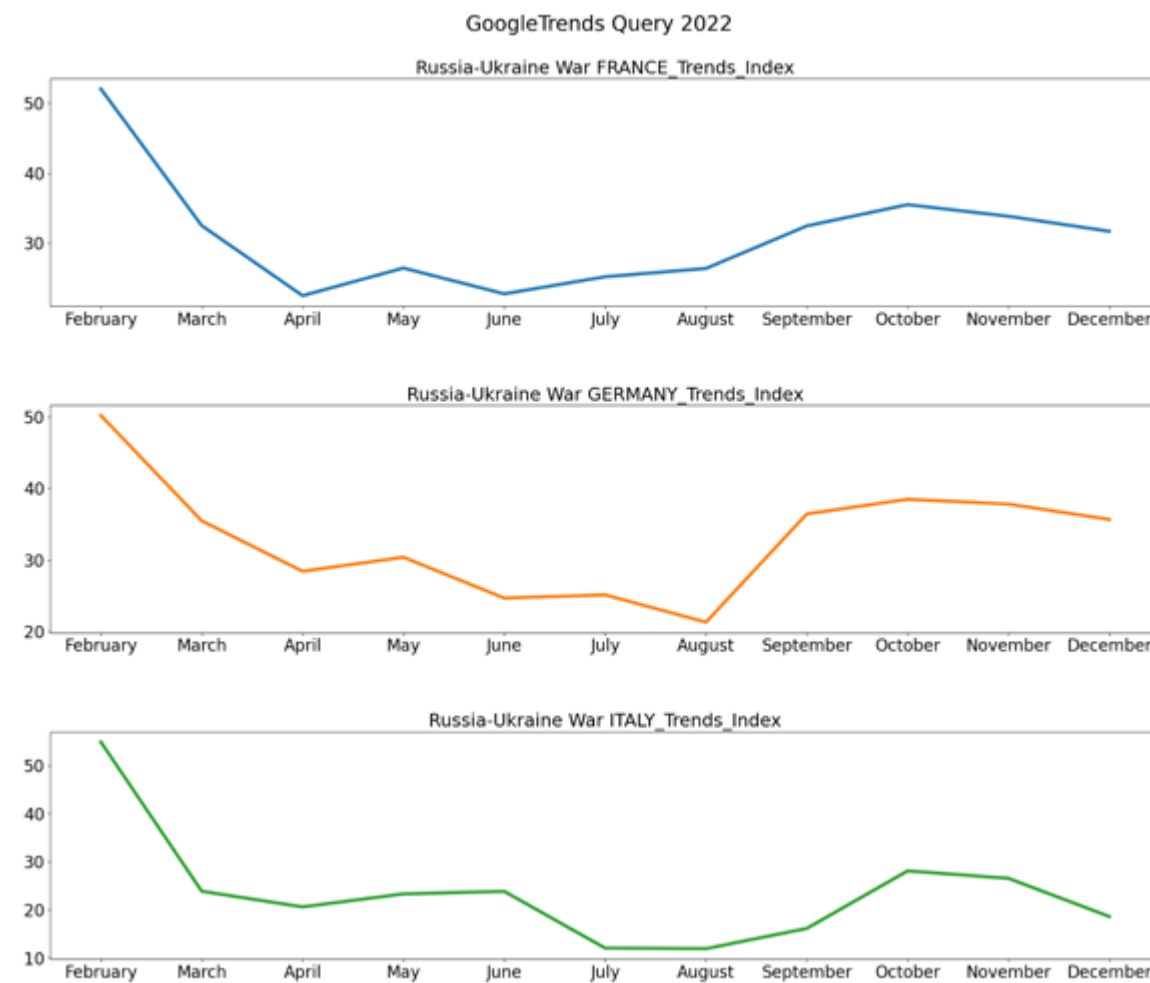


Figure 1. Google Trends Index for countries



Finanziato
dall'Unione europea
NextGenerationEU



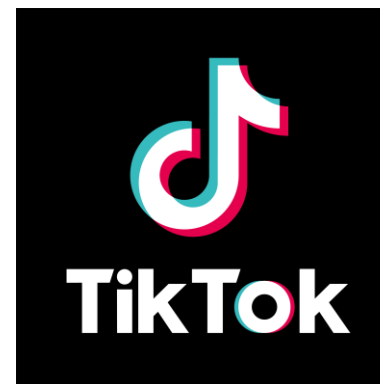
Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



Telegram





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



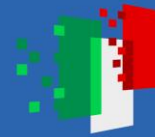
Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Choose one of the following themes

- Percezione dei brand sui social media** (es. monitoraggio delle opinioni su prodotti o servizi).
- Analisi delle reazioni a eventi politici o sociali** (es. elezioni, crisi globali, populismo, conflitti).
- Opinioni su tematiche ambientali e sostenibilità** (es. Fridays for Future, greenwashing).
- Polarizzazione delle opinioni su temi controversi** (immigrazione, eutanasia, aborto).
- Diffusione e impatto delle campagne di sensibilizzazione** (es. #BlackLivesMatter, #MeToo).



Dictionary-based approach

A dictionary-based approach for sentiment and emotion analysis uses predefined sets of words associated with specific sentiments (positive, negative, neutral) or emotions (joy, sadness, fear, anger, etc.). These dictionaries, such as SentiWordNet, AFINN, or NRC Emotion Lexicon, assign scores or categories to each term.

For example:

- In the context of sentiment, a word like "fantastic" may have a positive score (+3), while "terrible" might be strongly negative (-3).
- In the context of emotions, "happiness" can be associated with joy, while "panic" is linked to fear.

The analysis relies on identifying and counting the words in a text, comparing them with the dictionary to calculate aggregated metrics, such as sentiment scores or emotion distribution.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



VADER-Sentiment-Analysis

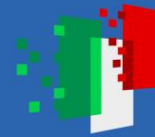
VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is *specifically attuned to sentiments expressed in social media*. It is fully open-sourced under the [\[MIT License\]](#) (we sincerely appreciate all attributions and readily accept most contributions, but please don't hold us liable).



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



Transformers

build passing license Apache-2.0 website online release v2.11.0

State-of-the-art Natural Language Processing for PyTorch and TensorFlow

Key Component: Self-Attention

The central part of the Transformer architecture is the **self-attention** mechanism. Here's how it works:

1.Weight of words: Each word in a sentence is represented as a numerical vector (a set of numbers that describe the word). The model assigns a "weight" to each word, indicating how relevant a word is to the others.

2.For example, in the sentence "The cat eats the mouse," the model might learn that the word "cat" is important for understanding "eats," and that "mouse" is related to "eats."

3.Updated representations: Instead of treating each word in isolation, the model updates the representation of each word based on its "weight" relative to the others. In other words, the words are interconnected. So, "cat" and "eats" influence the meaning of a word like "mouse."

Block Structure: Encoder and Decoder

Transformers are mainly composed of two parts:

- Encoder:** The encoder's job is to take the input (e.g., a sentence) and represent it in a way that the model can understand the relationships between the words. It uses self-attention to analyze each word in relation to the others.
- Decoder:** The decoder, typically used for tasks like translation, takes the processed information from the encoder and generates the final output (e.g., the translation of the sentence).

Parallelization

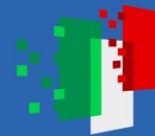
Unlike previous models (like RNNs or LSTMs), which process words one by one, Transformers can look at all the words simultaneously. This makes them much faster, especially when dealing with long sentences or large amounts of data.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



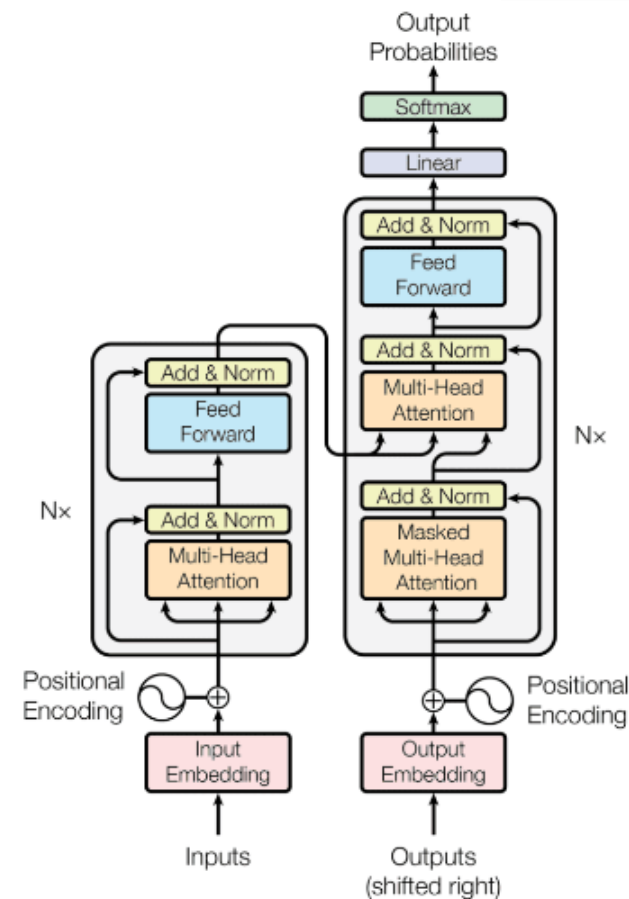
Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



Hugging Face

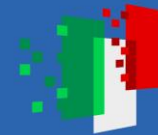




Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



Power BI

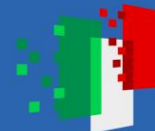
Power BI is a business intelligence suite developed by Microsoft that allows users to collect, analyze, visualize, and share data in a simple and effective manner. Power BI is widely used for creating interactive reports and dashboards that help businesses make data-driven decisions.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Key Features of Power BI:

1. Data Modeling:

Power BI allows importing data from various sources (databases, Excel files, cloud services, APIs, etc.) and modeling this data by creating relationships between tables. The ability to create complex data models and transform data with Power Query is one of its most powerful features.

2. Visualizations:

Power BI offers a wide range of visualizations, such as bar charts, line graphs, geographical maps, pivot tables, and scatter plots, to visually represent data and trends. These visualizations can be customized and combined into interactive reports.

3. Interactivity:

Reports created in Power BI are highly interactive. Users can filter, zoom, and explore data dynamically, enhancing the user experience and making data analysis easier.

4. Sharing and Collaboration:

Power BI offers cloud-based sharing capabilities, allowing users to publish and share reports and dashboards with colleagues and stakeholders. Real-time collaboration is also possible through Power BI Service.

5. Integration with Other Microsoft Applications:

Power BI integrates seamlessly with other Microsoft tools such as Excel, SharePoint, Azure, Dynamics 365, and Teams, enabling a more efficient workflow for users already within the Microsoft ecosystem.

6. Advanced Features:

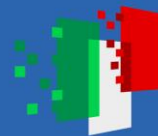
Power BI supports the use of programming languages like DAX (Data Analysis Expressions) for creating complex measures and calculations, and M for data transformation. It also allows integration with Python and R for advanced analytics and machine learning.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche