

ATTACHMENT 2.6

to the Technical - Scientific Report of the Implementing Entity
Reporting period no. 5 (months 17-20)

PRIN 2022 PNRR

***Normative and Digital Solutions to Counter Threats
during National Election Campaigns
(RightNets)***

Project Code: P2022MCYCK | CUP: D53D23022340001

University of Macerata

PI - Prof. Giovanni Di Cosimo



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



unimc
UNIVERSITÀ DI MACERATA

Normative and Digital Solutions to Counter Threats during National Election Campaigns (RightNets)

Guidelines for pilot case replication (D5.2)

Project Title	Normative and Digital Solutions to Counter Threats during National Election Campaigns
Project Acronym	RightNets
Call Identifier	PRIN 2022 PNRR - Decreto Direttoriale n. 1409 del 14-9-2022
Grant. no	P2022MCYCK
CUP	D53D23022340001
Duration	24 months
Work Package	5
Abstract	<p>This deliverable outlines methodological guidelines for replicating the pilot study conducted within the RightNets project, which analyzed digital political engagement during the 2024 European Elections. The report details the operational steps for data collection, preprocessing, and model evaluation, emphasizing the importance of contextual calibration. In fact, the proposed approach integrates automated data collection across multiple social media platforms using scraping tools, guided by thematically selected hashtags and keywords. The guidelines suggest how to process and analyze through a computational pipeline leveraging supervised classification techniques to categorize engagement levels of political content.</p>
Authors	Paolo Sernani (UNIMC)
PI	Giovanni Di Cosimo (UNIMC)

The project 'Normative and Digital Solutions to Counter Threats during National Election Campaigns - RightNets', is funded by the European Union - Next Generation EU under the Prin 2022 PNRR call, Mission 4, Component 2, Investment 1.1, P2022MCYCK, CUP D53D23022340001.

Views and opinions expressed in this document are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



Guidelines for pilot case replication

1. Applicabilità del metodo a nuovi contesti elettorali

Sebbene gli esperimenti in RightNets siano stati incentrati sulle Elezioni Europee del 2024, l'approccio metodologico adottato – sia nella fase di raccolta dati che in quella di analisi sperimentale – si presta a essere applicato con opportune modifiche a una vasta gamma di casi di studio, sia a livello nazionale che locale, sia in contesti europei che extraeuropei. In particolare, la strategia di raccolta multi-piattaforma, fondata sull'uso di strumenti automatizzati di scraping (Apify) e sulla selezione tematica tramite hashtag e parole chiave, può essere facilmente adattata a campagne elettorali diverse, modificando i parametri di selezione in funzione della lingua, del contesto politico e delle specificità della consultazione. La natura flessibile delle API utilizzate consente infatti di aggiornare il processo di acquisizione per includere nuove parole chiave o finestre temporali differenti, rendendo possibile un monitoraggio sistematico anche durante campagne in tempo reale o retrospettive.

Dal punto di vista analitico, la pipeline sperimentale basata sull'estrazione automatica di features da contenuti visivi e la successiva classificazione tramite modelli supervisionati può essere replicata per valutare l'engagement politico anche in contesti non elettorali, come campagne referendarie, consultazioni interne ai partiti, o mobilitazioni civiche promosse attraverso i social media. La categorizzazione dell'engagement in classi discrete di popolarità, pur costruita su soglie empiriche definite nel caso delle elezioni europee, può essere ricalibrata in funzione della distribuzione delle metriche osservate nel nuovo dataset, mantenendo inalterato l'impianto computazionale. Allo stesso modo, l'infrastruttura per la classificazione può essere potenziata includendo ulteriori fonti di informazione – come i metadati testuali dei post o i dati demografici del pubblico interagente – con l'obiettivo di cogliere più a fondo le specificità comunicative di ciascun contesto.

L'adozione di modelli pre-addestrati, come ViT-B16 o VGG16, consente una rapida generalizzazione a nuovi insiemi di immagini politiche, riducendo la necessità di training ex novo e garantendo tempi contenuti di deployment anche in assenza di grandi quantità di dati etichettati. Ciò rende il metodo adatto a contesti in cui la rapidità di analisi è cruciale, come nei giorni immediatamente precedenti al voto o durante eventi politici imprevisti. Inoltre, l'utilizzo di metriche di valutazione consolidate (accuratezza, AUC, F1-score) consente un confronto trasparente dei risultati ottenuti in contesti diversi, contribuendo alla costruzione di una base empirica comparativa sulle forme di interazione politica online. In prospettiva, l'applicazione iterata di questa metodologia a campagne diverse potrebbe generare una serie storica di osservazioni digitali, utile per studiare in modo longitudinale l'evoluzione della comunicazione politica sui social media, le variazioni nei livelli di engagement e la ricorrenza di tecniche persuasive o narrative.

2. Linee guida per la replicazione della raccolta dati e degli esperimenti in altri contesti elettorali

Dal punto di vista operativo, il primo passo per replicare la raccolta dati (e successivi esperimenti) in altri contesti elettorali consiste nell'identificare il contesto politico di riferimento, come ad esempio elezioni nazionali, locali, europee o consultazioni referendarie. È fondamentale definire con precisione l'arco temporale in cui si intende raccogliere i dati, idealmente coprendo i mesi che precedono l'evento elettorale, in modo da cogliere l'intera evoluzione della campagna digitale. Contestualmente, va stabilita la lingua di riferimento e il paese in cui si svolge la competizione politica, elementi essenziali per guidare la successiva selezione delle parole chiave e dei contenuti da monitorare. Questa fase serve a contestualizzare l'intero processo analitico e ad assicurare che le tecniche adottate siano sensibili alle specificità sociopolitiche e linguistiche del caso di studio.

Definito il contesto, l'arco temporale e le parole chiave per guidare la ricerca, la raccolta dati procede nelle seguenti fasi e operazioni:

Preparazione per la raccolta dati

- Selezionare le piattaforme social da monitorare (ad es. X/Twitter, Facebook, Instagram)
- Individuare parole chiave e hashtag rilevanti per l'evento, in lingua locale (es. #elezioni2025, #nomeCandidato, #votoPaese).
- Opzionalmente, selezionare gli account ufficiali dei partiti e candidati principali.

Raccolta dati

- Utilizzare scraper automatizzati, come quelli disponibili su Apify e usati per le Elezioni Europee 2024 negli esperimenti in RightNets, come Facebook Hashtag Scraper¹, Tweet Scraper V2² e Instagram Scraper³;
- Configurare lo scraping: ad esempio limitando il numero di post massimo scaricabile e aggiungendo o rimuovendo vincoli temporali.
- Registrare le metriche di engagement (like, commenti, condivisioni, etc.) almeno 48 ore dopo la pubblicazione dei post per garantire la propagazione adeguata sulla piattaforma
- Organizzare i dati in file CSV organizzati in colonne con variabili quali id del post, timestamp, numero di "likes", numero di commenti, numero di condivisioni.

Successivamente, al fine di utilizzare i dati raccolti in esperimenti di classificazione e previsione di livelli di engagement come fatto in RightNets, le fasi e le operazioni sono le seguenti:

¹ <https://apify.com/apify/facebook-hashtag-scraper>

² <https://apify.com/apidojo/tweet-scraper>

³ <https://apify.com/apify/instagramscraper>

Preprocessing dei dati

- Pulire i dati eliminando duplicati e contenuti irrilevanti.
- Calcolare il punteggio totale di engagement (es. like + commenti + condivisioni).
- Ipotizzare delle classi di post per popolarità, ad esempio bassa (< 1000 interazioni), media: 1000–5000, alta: > 5000 (le soglie possono essere ricalibrate in base alla distribuzione del nuovo dataset).

Analisi computazionale

- Dividere il dataset in training e test (ad esempio 70% e 30%), mantenendo bilanciamento tra classi.
- Addestrare classificatori supervisionati sui vettori di features estratti da immagini e/o testo.
- Valutare le performance con metriche standard, quali accuratezza, AUC, F1-Score, Precisione, Recall e analizzare la matrice di confusione per individuare bias o errori sistematici (es. sovrapposizione tra classi).
- Valutare i pattern ricorrenti nei post più (o meno) popolari per ottenere insight qualitativi.

I punti descritti costituiscono una base operativa per la replicazione della raccolta dati e della metodologia sperimentale implementate in RightNets. Tuttavia, ogni applicazione richiede un adattamento puntuale ai vincoli tecnici, politici e linguistici del contesto specifico: i passi, le soglie e le metriche suggerite vanno intesi come parametri di riferimento iniziali, più che come standard generalizzabili.