Workshop

HATE SPEECH AND SOCIAL MEDIA

Cristina Bosco, Viviana Patti,
Giancarlo Ruffo, Marco Stranisci

di.unito.it

Fondazione CRT

27 ottobre 2017, Torino

Hate speech: perchè?

- L’**istigazione pubblica all’odio** nei confronti di una minoranza (etnica, religiosa o di genere) è un fenomeno pervasivo e crescente in Italia e nel mondo.

- Si osserva una produzione di hate speech soprattutto nel web, in particolare nei **social media** che rappresentano oggi il “sentire comune” della popolazione

- I **fenomeni migratori** coinvolgono oggi un numero enorme di persone, ma suscitano atteggiamenti discriminatori e di intolleranza che hanno un forte impatto negativo sull’integrazione

Hate speech: dove?

La **scuola**:

- soprattutto secondaria, è uno dei luoghi in cui si manifestano atteggiamenti discriminatori e si produce hate speech
- Gli allievi **immigrati** sono oggi in Italia oltre il **12% della popolazione scolastica**
- Gli interventi contro l'hate speech **non** sono interventi di **censura**, ma di **sensibilizzazione ed educazione**

Hate speech: dove?

Il web e i **social media**:

- Per loro natura si prestano alla **diffusione** capillare rapida ed ubiquitaria dei contenuti che veicolano, quindi anche dell’haute speech

- Anche grazie alla sorta di **anonimato** che offrono agli utenti, sono il luogo ideale per la produzione di hate speech

Contrastare l’hate speech!

- Il progetto *Hate Speech and Social Media* ha lo scopo di **promuovere la consapevolezza** soprattutto **negli studenti** della scuola secondaria italiana dell’hate speech e delle sue dinamiche di produzione e diffusione

- Per raggiungere questo scopo stiamo applicando un **approccio basato su tecniche di intelligenza artificiale** che consentano agli utenti di entrare in contatto con le informazioni sull’hate speech

Contrastare l’hate speech!

- Il progetto *Hate Speech and Social Media* ha lo sco
  di **promuovere la conoc**

  ...ntano agli utenti di entrare in
  contatto con le informazioni sull’hate speech

Contrastare l’hate speech!

Obiettivi:

- **Raccogliere i dati** sull’hate speech in tempo reale e **trasformarli in conoscenza**
- **Rendere disponibile la conoscenza relativa all’hate speech** in una forma adatta agli utenti (insegnanti e studenti)
- Favorire di conseguenza l’**integrazione multi-etnica ed inter-culturale**

Dai dati alla conoscenza

La trasformazione dei dati in conoscenza comporta che i dati vengano

- **raccolti** dai social media
- sottoposti ad **analisi linguistiche automatiche** per riconoscere ed identificare quelli che portano hate speech
- **aggregati** statisticamente per leggere i fenomeni sottostanti
- **mostrati** tramite una interfaccia grafica

Raccolta dei dati

Raccolta dei dati

Raccolta dei dati

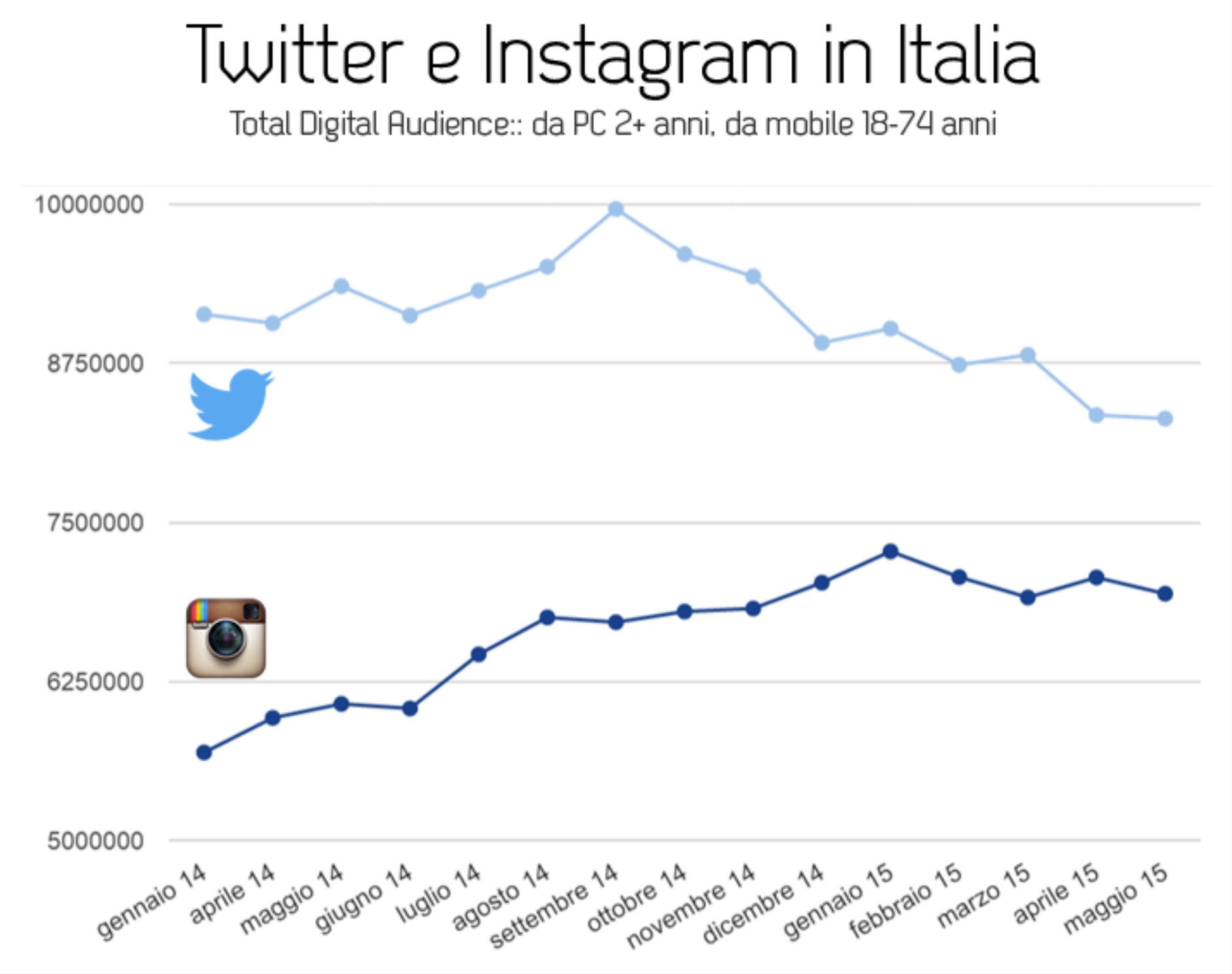
Raccolta dei dati

I dati prodotti nei social media fanno parte dei cosiddetti **bigdata** e occorrono tecniche adeguate per trattarli

Twitter in Italia: circa 9 milioni di utenti nel 2015 (http://www.internetlivestats.com/twitter-statistics/)

Milioni di **tweet italiani** ogni giorno: una vera **miniera per acquisire conoscenza sull'uso del linguaggio per i sistemi di analisi automatica**

Raccolta dei dati



Analisi dei dati

I dati estratti dai social media sono prevalentemente in forma testuale, quindi il progetto considera **centrale il linguaggio** che è:

- veicolo dell’odio, medium di comunicazione privilegiato nel web, strumento per denotare le persone come facenti parte di determinati gruppi sociali

Riconoscere l’hate speech

Per riconoscere l’hate speech applichiamo tecniche di analisi automatica, sentiment analysis e opinion mining

che ci consentono di isolare i testi che esprimono opinioni, giudizi, sentimenti, emozioni (stati privati) di utenti su argomenti, eventi, persone, prodotti di consumo, programmi tv, etc.

Ci interessano quelli che contengono hate speech, e per poterlo riconoscere dobbiamo conoscerlo bene

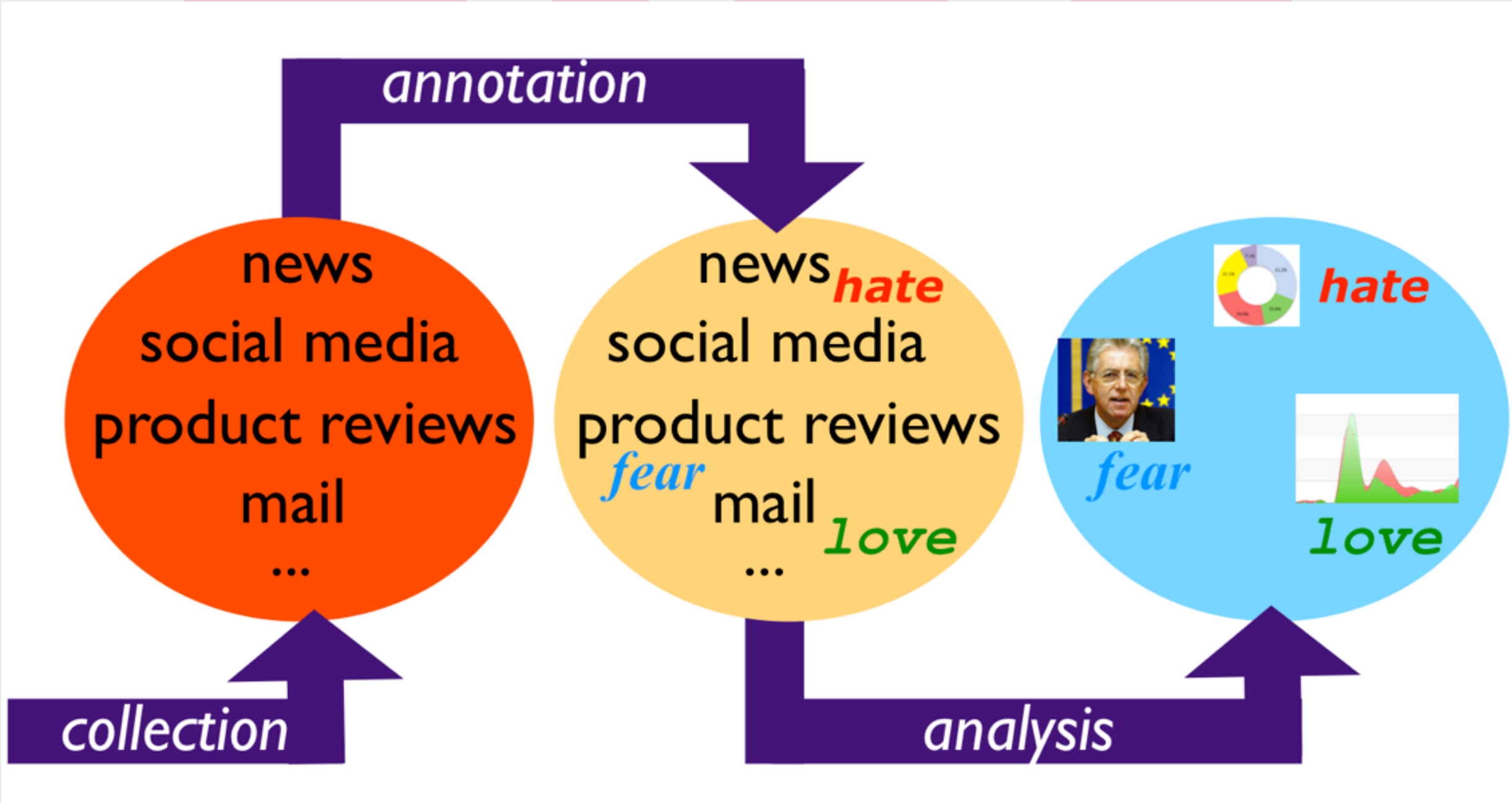
Riconoscere l’hate speech

Abbiamo analizzato molti testi e dati e abbiamo scoperto molti modi in cui l’odio si esprime nel social media (e non abbiamo finito!)

Abbiamo costruito insiemi di esempi dove le forme di hate speech sono annotate a mano da gruppi di annotatori esperti

I nostri sistemi vengono addestrati sulla base di questi esempi e sono pronti a riconoscere esempi simili dentro nuovi testi e ad annotarli nello stesso modo

Riconoscere l’hate speech



Riconoscere l'hate speech

Table 1: Combinations of values allowed by our annotation scheme

subj	pos	neg	iro	**description**
0	0	0	0	an objective tweet example: *l'articolo di Roberto Ciccarelli dal manifesto di oggi* `http://fb.me/1BQVy5WAk`
1	0	0	0	a subjective tweet with neutral polarity and no irony example: *Primo passaggio alla #strabrollo ma secondo me non era un iscritto*
1	1	0	0	a subjective tweet with positive polarity and no irony example: *splendida foto di Fabrizio, pluri cliccata nei siti internazionali di Photo Natura* `http://t.co/GWoZqbxAuS`
1	0	1	0	a subjective tweet with negative polarity and no irony example: *Monti, ripensaci: l'inutile Torino-Lione inguaia l'Italia: Tav, appello a Mario Monti da Mercalli, Cicconi, Pont...* `http://t.co/3CazKS7Y`
1	1	1	0	a subjective tweet with positive and negative polarity (mixed polarity) and no irony example: *Dati negativi da Confindustria che spera nel nuovo governo Monti. Castiglione: "Avanti con le riforme"* `http://t.co/kIKnbFY7`
1	1	0	1	a subjective tweet with positive polarity, and an ironic twist example: *Letta: sicuramente non farò parte del governo Monti . e siamo un passo avanti. #finecorsa*
1	0	1	1	a subjective tweet with negative polarity, and an ironic twist example: *Botta di ottimismo a #lInfedele: Governo Monti, o la va o la spacca.*

Riconoscere l’hate speech

testo
Nuova gaffe della Boldrini: «Mi sento più italiana quando sto con gli immigrati» https://t.co/sI77XBVHhP
ROMA – ORRORE ALLA STAZIONE: DUE CARABINIERI ACCERCHIATI DA 15 IMMIGRATI INFEROCITI - Notiziaro Online
Milano, soldati aggrediti da immigrati: un ferito e un arresto https://t.co/JbrJk0IbEN #thexeon #milanocronaca
Battipaglia, immigrato piscia in strada nei pressi di una scuola:... https://t.co/jVKXOVupfK
Un giudice blocca in tutti gli Stati Uniti lo stop di Trump agli immigrati islamici https://t.co/y5Mys5Djrl
Altro che pagarci le pensioni Immigrati sono solo un costo: Un luogo comune molto diffuso in Italia racconta ...
Italia: immigrati 92.300, emigrati 107.529. Chi è che invade?
@dawac72 immigrati a volontà e questi che ha fatto Renzi città invase e disagi ovunque
MILANO: RIVOLTA IMMIGRATI, ZONA DEVASTATA E CASSONETTI ROVESCIATI | VoxNews https://t.co/ndzFs1VHOC
@VincenzoDeLuca È giunto il momento di fare un po di pulizia seria, tra espellere immigrati irregolari e arrestare i nostri malviventi!!!
Che futuro possiamo garantire a tutti questi immigrati? Che futuro possiamo garantire a noi e ai nostri figli? SIETE DEI CRIMINALI...
vicino a #Spagna ci sono Marocco/Tunisia, #Italia vicino a #Libia dove parte flusso immigratorio. #Salvini la #geografia salla!
@Roccia9 Leggiti questo articolo per vedere quanto sono amati gli immigrati italiani. https://t.co/NYaJEVjTj9 @Italy_Mauri
Immigrato stupra la figlia di 7 anni | Tutti i Crimini degli Immigrati https://t.co/b6gIbN3FNc
Isis a Milano: è la prova delle frottole raccontate da Renzi sugli immigrati ED E SOLO L INIZIO VEDRAI CHE FUOCHI EHH
Le "social card" di #Renzi? Sette su dieci agli immigrati https://t.co/Lb9b08LHvY il 67% dei contributi per l'affitto va a non italiani
@ACeschia Assolutamente NO..Renzi e il PD stanno distruggendo l'italia..con gli immigrati clandestini.. Berlusconi ha fatto il suo tempo!
#Gentiloni ce la farà a bloccare il flusso immigrati in Libia?in Spagna Zapatero c è riuscito a bloccare tutto muri e guardie!#speriamo
Sotto la neve alla fermata del bus: studentessa salva un immigrato https://t.co/2Ud9ICD5TN
Capotreno a giudizio per aver fatto scendere immigrato senza biglietto https://t.co/Uar6aLRApX https://t.co/qae8uDxs61

Riconoscere l’hate speech

testo	stance	ironia	hate speech	offensivo
Nuova gaffe della Boldrini: «Mi sento più italiana quando sto con gli immigrati» https://t.co/sI77XBVHhP	NEUTRALE	IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
ROMA – ORRORE ALLA STAZIONE: DUE CARABINIERI ACCERCHIATI DA 15 IMMIGRATI INFEROCITI - Notiziaro Online	CONTRO	NON IRONICO	HATE SPEECH: SI	POCO OFFENSIVO
Milano, soldati aggrediti da immigrati: un ferito e un arresto https://t.co/JbrJk0IbEN #thexeon #milanocronaca	NEUTRALE	NON IRONICO	HATE SPEECH: SI	POCO OFFENSIVO
Battipaglia, immigrato piscia in strada nei pressi di una scuola:... https://t.co/jVKXOVupfK	CONTRO	NON IRONICO	HATE SPEECH: SI	MOLTO OFFENSIVO
Un giudice blocca in tutti gli Stati Uniti lo stop di Trump agli immigrati islamici https://t.co/y5Mys5Djrl	NEUTRALE	NON IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
Altro che pagarci le pensioni Immigrati sono solo un costo: Un luogo comune molto diffuso in Italia racconta ...	CONTRO	NON IRONICO	HATE SPEECH: SI	MOLTO OFFENSIVO
Italia: immigrati 92.300, emigrati 107.529. Chi è che invade?	CONTRO	IRONICO	HATE SPEECH: SI	ABBASTANZA OFFENSIVO
@dawac72 immigrati a volontà e questi che ha fatto Renzi città invase e disagi ovunque	CONTRO	NON IRONICO	HATE SPEECH: SI	ABBASTANZA OFFENSIVO
MILANO: RIVOLTA IMMIGRATI, ZONA DEVASTATA E CASSONETTI ROVESCIATI | VoxNews https://t.co/ndzFs1VHOC	CONTRO	NON IRONICO	HATE SPEECH: SI	PER NULLA OFFENSIVO
@VincenzoDeLuca È giunto il momento di fare un po di pulizia seria, tra espellere immigrati irregolari e arrestare i nostri malviventi!!!	NEUTRALE	NON IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
Che futuro possiamo garantire a tutti questi immigrati? Che futuro possiamo garantire a noi e ai nostri figli? SIETE DEI CRIMINALI...	NEUTRALE	NON IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
vicino a #Spagna ci sono Marocco/Tunisia, #Italia vicino a #Libia dove parte flusso immigratorio. #Salvini la #geografia salla!	PRO	IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
@Roccia9 Leggiti questo articolo per vedere quanto sono amati gli immigrati italiani. https://t.co/NYaJEVjTj9 @Italy_Mauri	NEUTRALE	NON IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
Immigrato stupra la figlia di 7 anni | Tutti i Crimini degli Immigrati https://t.co/b6glbN3FNc	CONTRO	NON IRONICO	HATE SPEECH: SI	MOLTO OFFENSIVO
Isis a Milano: è la prova delle frottole raccontate da Renzi sugli immigrati ED E SOLO L INIZIO VEDRAI CHE FUOCHI EHH	CONTRO	IRONICO	HATE SPEECH: SI	ABBASTANZA OFFENSIVO
Le "social card" di #Renzi? Sette su dieci agli immigrati https://t.co/Lb9b08LHvY il 67% dei contributi per l'affitto va a non italiani	CONTRO	NON IRONICO	HATE SPEECH: SI	PER NULLA OFFENSIVO
@ACeschia Assolutamente NO..Renzi e il PD stanno distruggendo l'italia..con gli immigrati clandestini.. Berlusconi ha fatto il suo tempo!	CONTRO	NON IRONICO	HATE SPEECH: SI	POCO OFFENSIVO
#Gentiloni ce la farà a bloccare il flusso immigrati in Libia?in Spagna Zapatero c è riuscito a bloccare tutto muri e guardie!#speriamo	CONTRO	IRONICO	HATE SPEECH: SI	ABBASTANZA OFFENSIVO
Sotto la neve alla fermata del bus: studentessa salva un immigrato https://t.co/2Ud9ICD5TN	PRO	NON IRONICO	HATE SPEECH: NO	PER NULLA OFFENSIVO
Capotreno a giudizio per aver fatto scendere immigrato senza biglietto https://t.co/Uar6aLRApX https://t.co/qae8uDxs61	CONTRO	NON IRONICO	HATE SPEECH: SI	ABBASTANZA OFFENSIVO

Aggregare i dati

Una volta estratti i contenuti di hate speech occorre:

- **classificarli** (target, modalità espressive, livello di aggressività, …)
- aggregarli in base a **luogo**, **tempo** e che cosa può interessare l’utente

Visualizzare l’hate speech

La visualizzazione consente:

- di **mettere in evidenza la conoscenza** che sta dietro al dato analizzato
- di **democratizzare la conoscenza**, nel senso di esprimerla in forme diverse e modulabili in base alle caratteristiche, capacità cognitive di chi deve fruirne
- di rendere **azionabile** la conoscenza

Visualizzare l’hate speech

La visualizzazione consiste nella **costruzione dinamica di mappe, istogrammi ed altre forme grafiche** che rappresentano vari aspetti, tra cui la diffusione nel tempo e nello spazio dell’hate speech

Educare all’hate speech

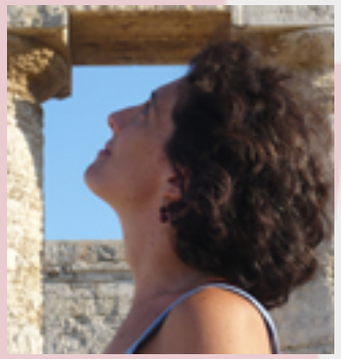
- Il cuore del progetto è il **percorso educativo** verso il quale tutte le attività sono orientate

- Gli studenti e i loro insegnanti di riferimento sono coinvolti nelle fasi di test dell’interfaccia grafica e diventeranno gli utenti finali di essa

Team multidisciplinare

di.unito.it

Cristina BOSCO



Viviana PATTI

Delia Irazu HERNANDEZ FARIAS

Manuela SANGUINETTI

LabNET

Rossano SCHIFANELLA

Marcello BOGETTI

Giancarlo RUFFO

Massimo FOLLIS

ACMOS associazione di promozione sociale

Marco STRANISCI

CPS CULTURE POLITICA SOCIETÀ

Michelangelo CONOSCENTI

New entry: Alessandra CIGNARELLA e Fabio POLETTO e ???

HATE SPEECH AND SOCIAL MEDIA

Fondazione CRT

I HATE PREJUDICE

Compagnia di San Paolo

GRAAAAAZIE
dell’ATTENZIONE!!!