

Progettare sistemi di AI per analizzare contenuti social riguardo al tema dei vaccini

Alessandro Renda

Dipartimento di Ingegneria dell'Informazione



UNIVERSITÀ DI PISA

Agenda

- **Social sensing**: i social network come reti di sensori
- Il caso di studio: la ***stance*** rispetto al tema dei vaccini
- L'approccio **text mining** all'analisi dei tweet
- L'emergenza coronavirus e l'opinione sul **green pass**

Social Sensing



<https://fredcavazza.net/2017/04/19/social-media-landscape-2017/>

- Un **sensore** è un dispositivo, in diretta interazione con il sistema misurato, che fornisce una lettura di una determinata grandezza di interesse
- **Social Sensors** sono utenti del social network che, tramite la condivisione di informazioni, forniscono indicazioni sulla popolazione
- L'esempio di **Twitter**
 - Condivisione di un breve messaggio per riferire su un evento pubblico o privato, o per esprimere la propria opinione su un topic
 - Svariate applicazioni di *social sensing*:
 - 🐦 Monitoraggio in tempo reale del traffico
 - 🐦 Analisi feedback sulla soddisfazione di un servizio
 - 🐦 Analisi e previsione delle tendenze della moda
 - 🐦 Previsione dei risultati elettorali

Un altro esempio...

... un po' datato: *MathCompEpi2018 – Erice – 28 Agosto 2018*



Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate

David A. Broniatowski PhD, Amelia M. Jamison MAA, MPH, SiHua Qi SM, Lulwah AlKulaib SM, Tao Chen PhD, Adrian Benton MS, Sandra C. Quinn PhD, and Mark Dredze PhD (show fewer authors)

[+] Author affiliations, information, and correspondence details

Accepted: May 22, 2018 Published Online: August 23, 2018

Il caso di studio: la *stance* rispetto al tema dei vaccini

- Dibattiti e gruppi di discussione online sui social sul tema vaccinazione
 - Ad esempio: presunta connessione tra **autismo** e **vaccino MPR**



- Le notizie e le opinioni diffuse nei social network **influenzano i comportamenti** e i sentimenti individuali

- I comportamenti individuali possono avere implicazioni:
 - Marzo 2017 - Il Ministero della Salute rileva un **calo generale della copertura vaccinale**
 - Rischio riemergenza malattie eradiccate o aumento incidenza
 - Evidenze più recenti con Covid-19

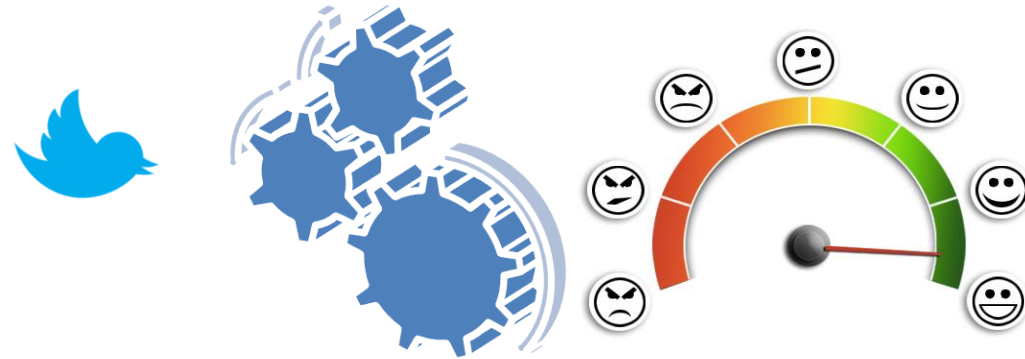


La IA per l'analisi dei contenuti social

- **Obiettivo:** monitoraggio automatico della *stance* espressa su Twitter sul tema dei vaccini
 - Misurare andamento e variazioni nella *stance* della popolazione
 - Strumento per sondare l'opinione pubblica e promuovere contromisure o campagne informative
- Sistema di **text mining** per la **classificazione della *stance***

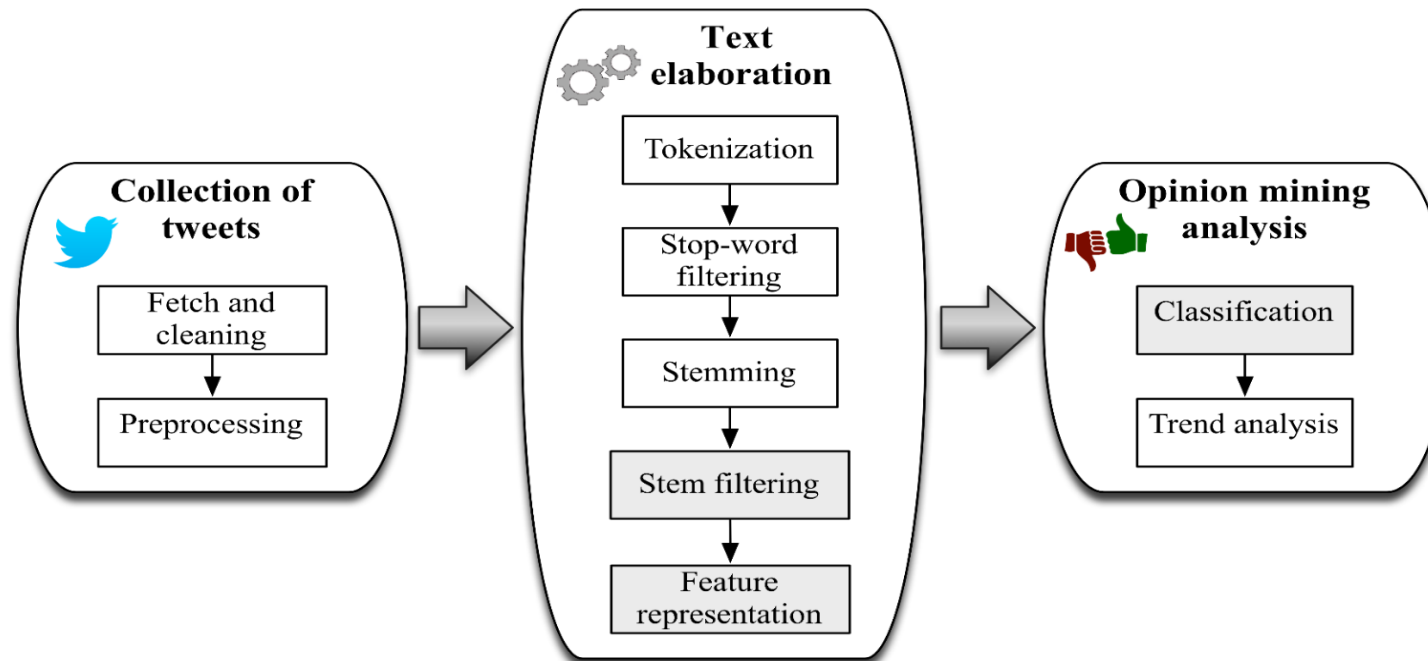
Per ogni tweet sul tema, determinare se esso è

- A favore della vaccinazione
- Contrario alla vaccinazione
- Neutrale



Sistema di classificazione della *stance*

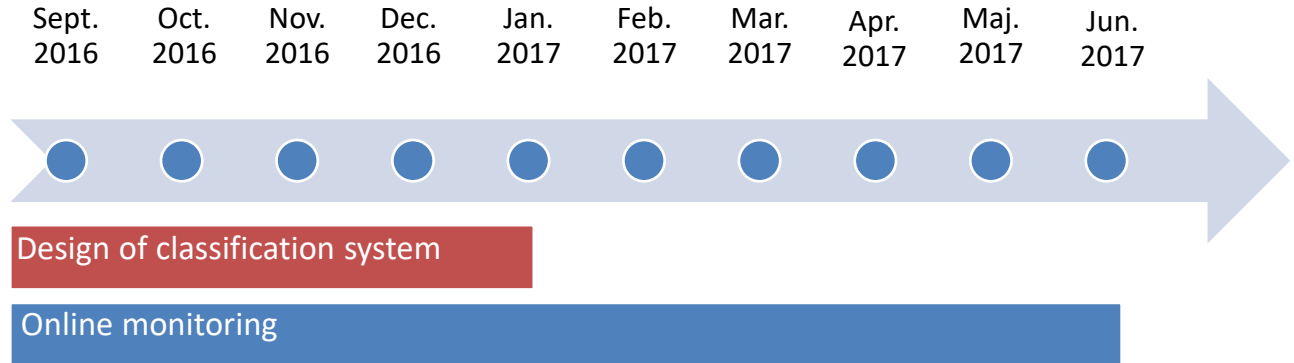
- **Text mining:** estrazione di informazioni significative dal testo in linguaggio naturale. Nel caso dei tweet:
 - Testo non strutturato e irregolare, di lunghezza limitata, con acronimi, hashtag, mention, URL
 - Linguaggio colloquiale, talvolta con errori di ortografia o grammaticali
 - Ambiguità di opinione, sarcasmo, ironia



Sistema di classificazione della *stance*

- **Setup sperimentale:**

- Lingua italiana
- Query basata su 37 keywords
- ~ 100.000 tweet tra 9/2016 – 6/2017



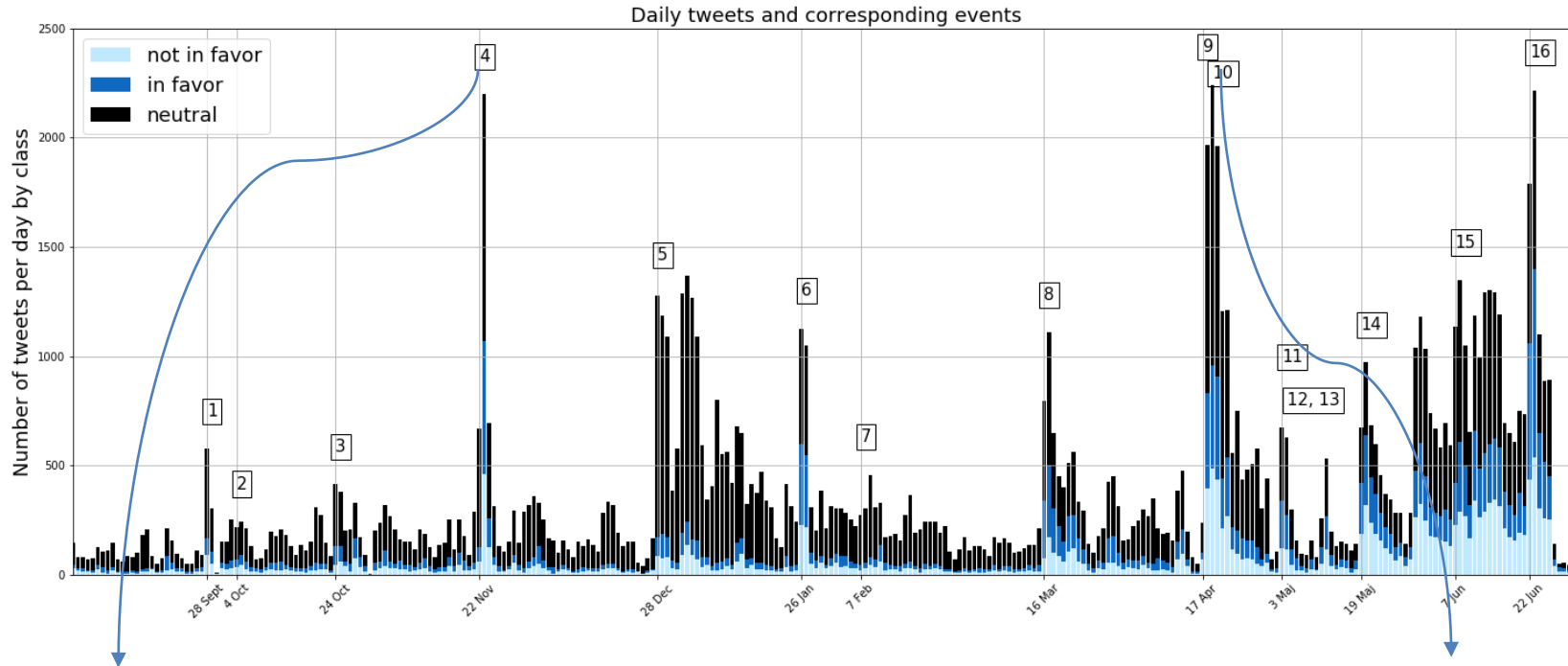
- **Progettazione del sistema di classificazione**

- ~ 700 tweet etichettati manualmente per utilizzo come training set
- Selezione modello più adeguato in 10-fold CV: approccio *Bag-of-Words + SVM* (Accuracy: 64.8%)

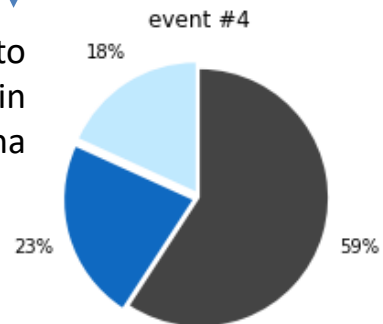
- **Online monitoring**

- Fase di inferenza per individuare le tendenze dell'opinione pubblica nel tempo

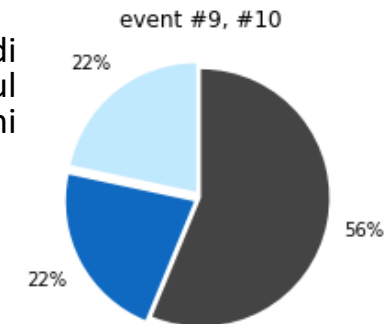
Sistema di classificazione della *stance*



Approvazione regolamento
obbligo vaccinale per bambini in
età scolare in Emilia Romagna



Puntata di
Report sul
tema vaccini

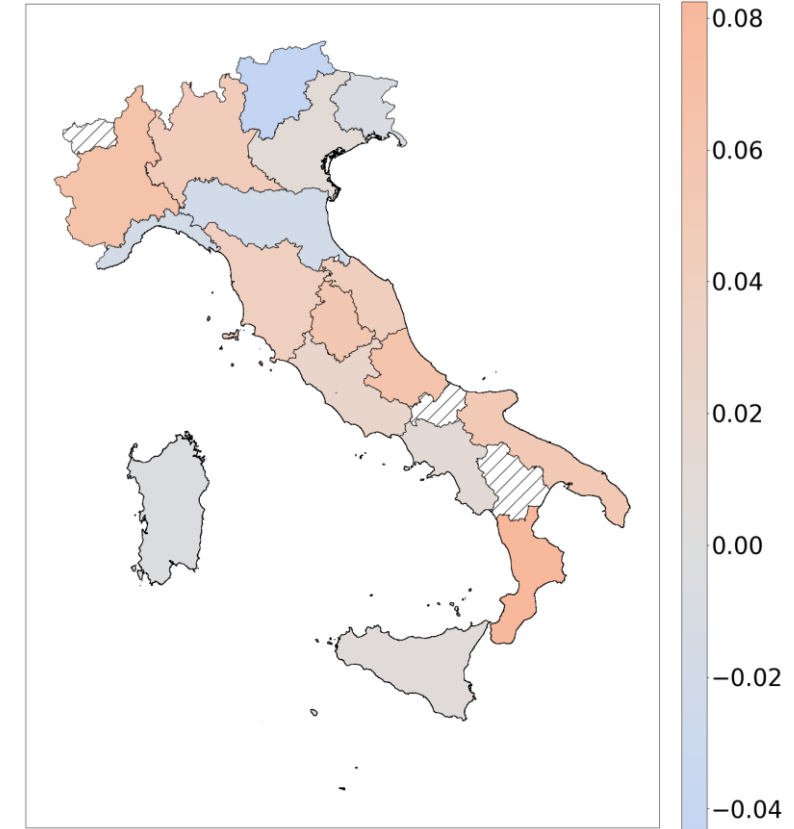
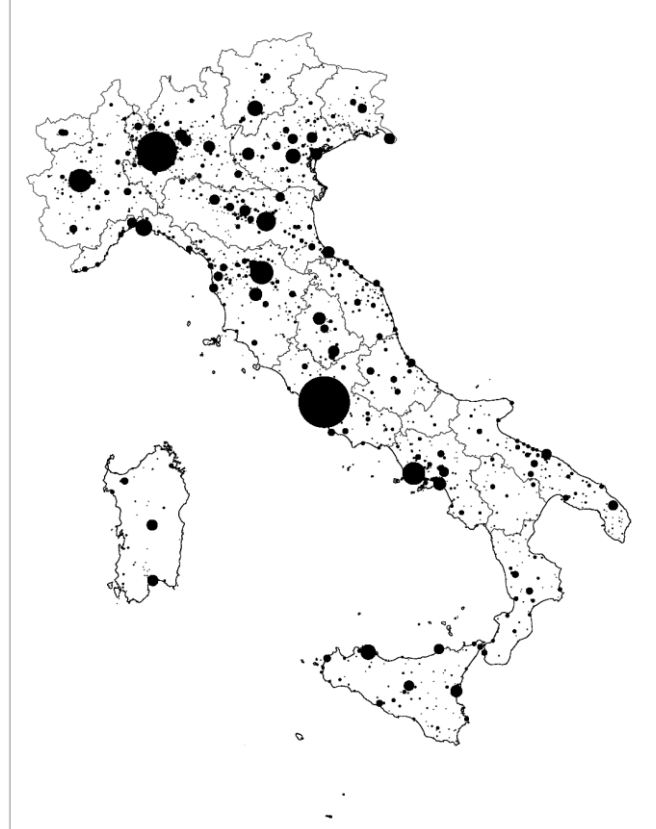


Sistema di classificazione della *stance*

I *metadati* consentono ulteriori analisi:

- Distinzione fra utenti **verificati** (agenzie, canali di informazione, ...) e **non verificati** (individui)
- Analisi su **scala regionale** dell'attività e della *stance*

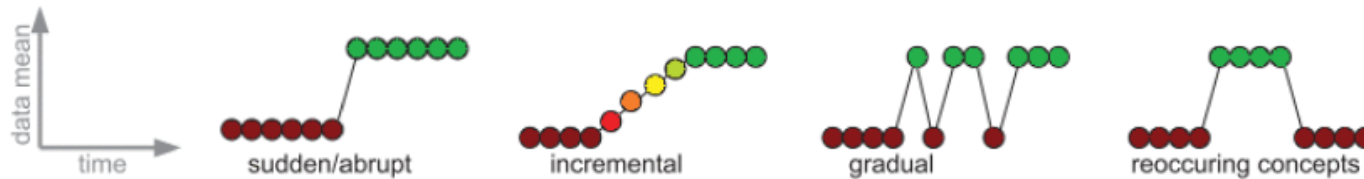
$$\text{RegionalScore}_r = \frac{\text{InFavor}_r - \text{NotInFavor}_r}{\text{Total}_r}$$



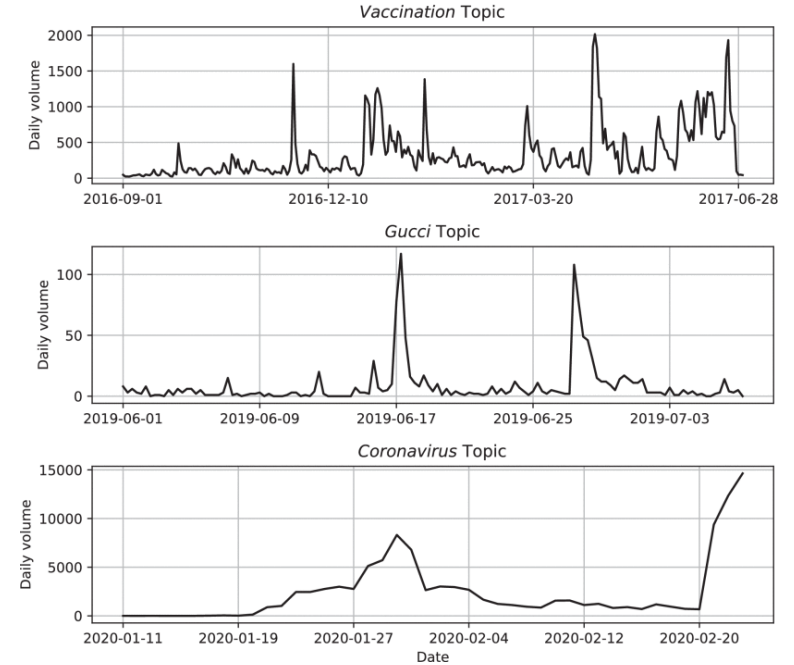
Il problema del *concept drift* nell'analisi dei tweet

- La raccolta di tweet può essere caratterizzata da *concept drift*, possibilmente determinato da avvenimenti di cronaca /sociali / politici
- Con p_{ti} distribuzione di probabilità congiunta all'istante t_i tra l'insieme delle variabili di input X e la variabile target y , il *concept drift* tra t_0 e t_1 è definito come:

$$\exists X: p_{t_0}(X, y) \neq p_{t_1}(X, y)$$



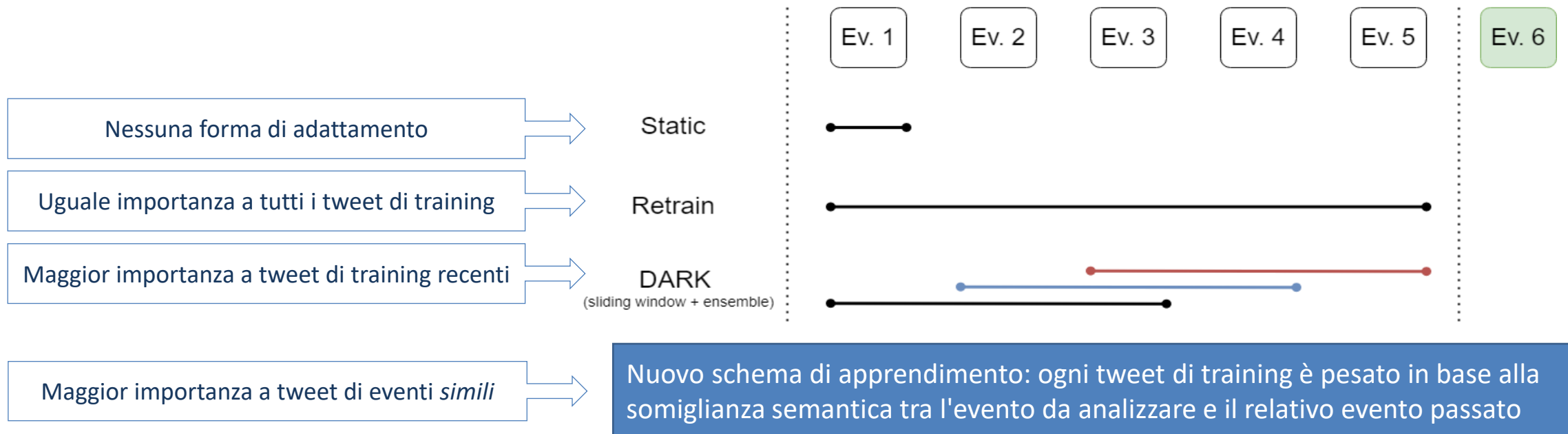
- Un modello di classificazione *iniziale* è adatto ad un monitoraggio a lungo termine?
- Come gestire il *concept drift* in un task di classificazione di tweet?



Il problema del *concept drift* nell'analisi dei tweet

- **Setup sperimentale:**

- Confronto di diversi schemi di apprendimento su una campagna di monitoraggio estesa (~ 3 anni)
- Etichettatura manuale di alcuni tweet ad ogni evento, abilitando **valutazione accuratezza nel tempo** e, possibilmente, **aggiornamento del modello**



Il problema del *concept drift* nell'analisi dei tweet

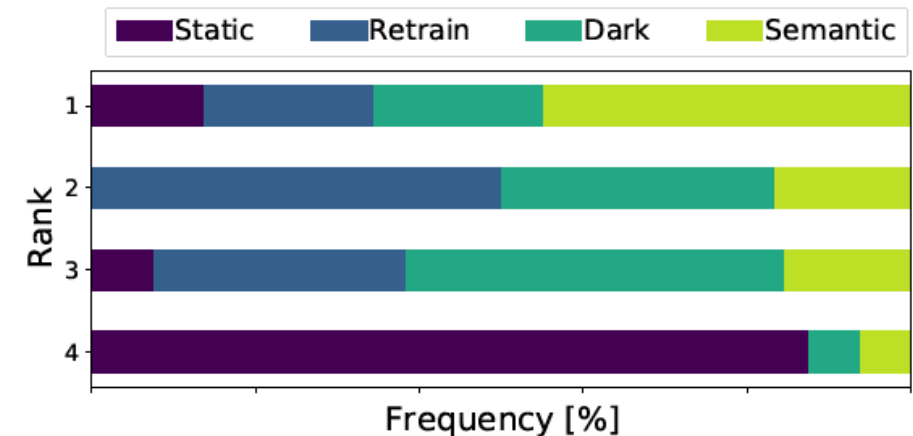
	Aumento (preoccupante) dei casi di morbillo in Italia. No, ma l'importante è che le merendine non contengano olio di palma
NN-1	+230% di casi di #morbillo in Italia in un solo anno. Ma mi raccomando, continuate a dire che i vaccini fanno male
NN-2	Preoccupante boom dei casi di #morbillo in #Italia a fronte del calo dei #vaccini. Chi fa propaganda anti-vaccinazione è un delinquente.
NN-3	@RaiRadio2 come dimostrereste che l'aumento dei casi di morbillo nel 2017 "é sicuramente dovuto" alla diminuzione delle vaccinazioni?

	Obbligo dei vaccini, 3 bimbi non possono entrare a scuola a Pagani
NN-1	Vaccini, comincia demolizione decreto Lorenzin: tutti i bambini potranno andare a scuola anche senza certificato Asl "basterà autocertificazione"
NN-2	Vaccini: nessun bambino sarà escluso da scuola a settembre
NN-3	Salvini, inutili 10 vaccini obbligatori , tutti i bimbi a scuola - Focus vaccini

- Valutazione qualitativa: esempi di tweet e rispettivi *nearest neighbors*



- Frequenze della "posizione in classifica" dei vari approcci, in base ai valori di accuratezza ottenuti



L'emergenza coronavirus e l'opinione sul green pass

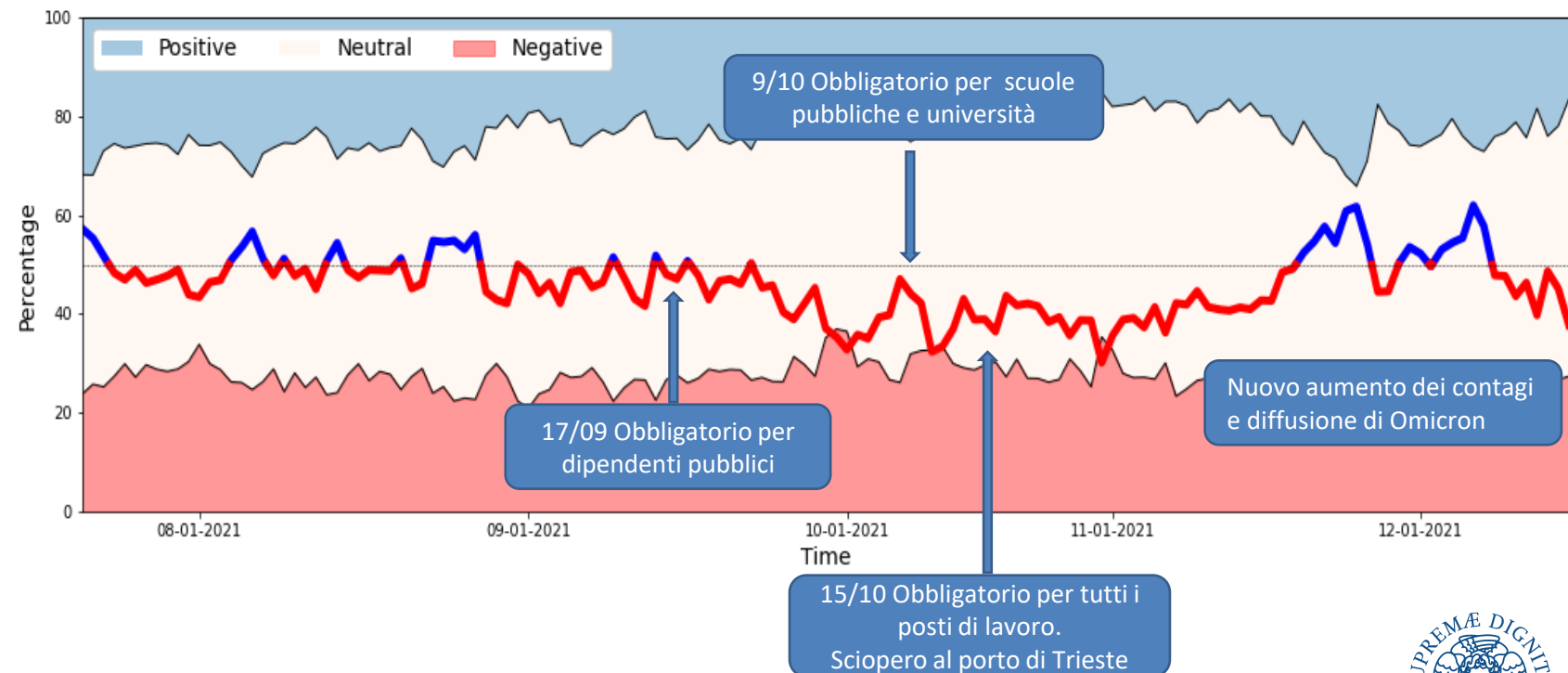
La piattaforma di monitoraggio sviluppata può essere agevolmente adattata a nuovi casi di studio

- La *stance* verso il **green-pass**: visualizzazione polarità come *stacked bar-plot* normalizzato
- Definizione di *overall stance* su base giornaliera (linea nella parte centrale del grafico)



$$Stance_i = \frac{Positive_i}{Positive_i + Negative_i}$$

- Blu: prevalenza *a favore*
- Rosso: prevalenza *contro*



Grazie per l'attenzione!



Riferimenti bibliografici:

E. D'Andrea, P. Ducange, A. Bechini, A. Renda & F. Marcelloni
Monitoring the Public Opinion about the Vaccination Topic from Tweets Analysis.
Expert Systems with Applications, 116:209–226, (2019)

A. Bechini, P. Ducange, F. Marcelloni & A. Renda
Stance Analysis of Twitter Users: the Case of the Vaccination Topic in Italy
IEEE Intelligent Systems, 36(5), 131-139. (2020)

A. Bechini, A., Bondielli, A., Ducange, P., Marcelloni, F., & A. Renda.
**Addressing event-driven concept drift in twitter stream:
 A stance detection application**
IEEE Access, 9, 77758-77770. (2021)

Bondielli, A., Tortora, G. C., Ducange, P., Macri, A., Marcelloni, F., & A. Renda,
Online Monitoring of Stance from Tweets: The case of Green Pass in Italy.
 In *2022 IEEE EAIS* (pp. 1-8). IEEE. (2022)



Contatto mail
alessandro.renda@unipi.it