UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Scienze Fisiche, Informatiche e Matematiche

CORSO DI LAUREA IN INFORMATICA

Costruzione di una dashboard intelligente per la composizione di Tweet in ambito Cultural Heritage per dispositivi Android

Relatore

Dott. Riccardo Martoglia

Laureanda

Chiara Boni

ANNO ACCADEMICO 2018/2019

INTRODUZIONE

La seguente ricerca, promossa dall'Università degli studi di Modena e Reggio Emilia, nasce per le istituzioni in **ambito culturale** che hanno la necessità di potenziare la loro presenza sui *social network*, così da portare la loro visibilità su una piattaforma globale.

Tramite questi strumenti possono **pubblicizzare** le proprie iniziative e coinvolgere il proprio pubblico.

È necessario essere **notati** e scrivere messaggi di successo, per contrastare la sempre più crescente concorrenza.

OBIETTIVI

Implementare un **modello** a classificatore che fosse in grado di categorizzare tweet, in base al loro effetto sul pubblico.

Creare una **dashboard** intelligente per dispositivi mobili, tale che fosse semplice ed intuitiva per ogni tipo di utente.
Essa deve aiutare i manager dei musei a comporre i tweet.

Collegare l'applicazione a **Twitter** per pubblicare, su un profilo esistente, i messaggi.

Struttura progetto

1. Tecnologie utilizzate
2. Il caso di studio
3. Costruzione classificatore
4. Applicazione Android
5. Prove sperimentali e risultati ottenuti

TECNOLOGIE UTILIZZATE

Lato Server

Flask Framework

PyCharm

PANDAS

Scikit-learn

Postman

JSON Objects

Seaborn

Lato Client

Android Studio

Volley library

Struttura progetto

1 Tecnologie utilizzate

2 Il caso di studio

Lato Server

3 Costruzione classificatore

Lato Client

4 Applicazione Android

5 Prove sperimentali e risultati ottenuti

1. Divisione musei

In questo studio sono presenti **25 musei**, i cui tweet sono stati divisi in sei gruppi distinti in base alla popolarità del mittente.
Sono stati esaminati più di 19.500 messaggi.

Table 1: Museum groups

Group 1 (G1)	# Followers	Group 2 (G2)	# Followers	Group 3 (G3)	# Followers
@MuseumModernArt	5,120,000	@britishmuseum	1,560,000	@CentrePompidou	970,000
@Tate	4,500,000	@vangoghmuseum	1,330,000	@NationalGallery	887,000
@metmuseum	3,680,000	@MuseeLouvre	1,250,000	@museofrodakahlo	847,000
@Guggenheim	3,350,000	@GettyMuseum	1,250,000	@MuseeOrsay	610,000
@saatchi_gallery	2,800,000	@museodelprado	1,180,000		
Group 4 (G4)	**# Followers**	**Group 5 (G5)**	**# Followers**	**Group 6 (G6)**	**# Followers**
@mfaboston	332,000	@visitmuve_it	88,000	@MuseoEgizio	21,600
@museiincomune	264,000	@museupicasso	65,000	@Uffizi	19,900
@philamuseum	247,000	@mart_museum	64,000	@MUSE_Trento	12,600
@maspmuseum	245,000				
@ngadc	216,000				
@Museo_MAXXI	190,000				

2. **Gestione dei dati**

Ogni data set è composto da i tweet salvati in formato tabellare, come mostrato in figura. Ogni riga corrisponde ad un **thread**, quindi ad un messaggio con tutte le sue componenti divise per colonne.

NRETWEET	NLIKE	NANS	MUSEUM	NURLS	NIMG	NHASH	NMENTION	LENGTH	DENSE	SENT	ENG	LONG	FULL_TEXT
5	46	1	Guggenheim	1	1	1	0	281	0	1	0	1	Happy birthday, Hilla Rebay! Enjoy 20% off the "Visionaries: Creating a Modern Guggenheim" catalogue, today and online only and learn more about Rebay's art-collecting practices and her role as the first director of the #Guggenheim.

Si deve scegliere un attributo sul quale calcolare il successo o meno del tweet.
Tra NRETWEET, NLIKE e NANS, si sceglie come **target** il NLIKE.

In base alla variabile target si distinguono **due classi**: 'GOOD' oppure 'BAD', che rappresentano i due possibili esiti del messaggio.

Struttura progetto

1 Tecnologie utilizzate

2 Il caso di studio

Lato Server

3 Costruzione classificatore

Lato Client

4 Applicazione Android

5 Prove sperimentali e risultati ottenuti

COSTRUZIONE CLASSIFICATORE

1. Scelta caratteristiche

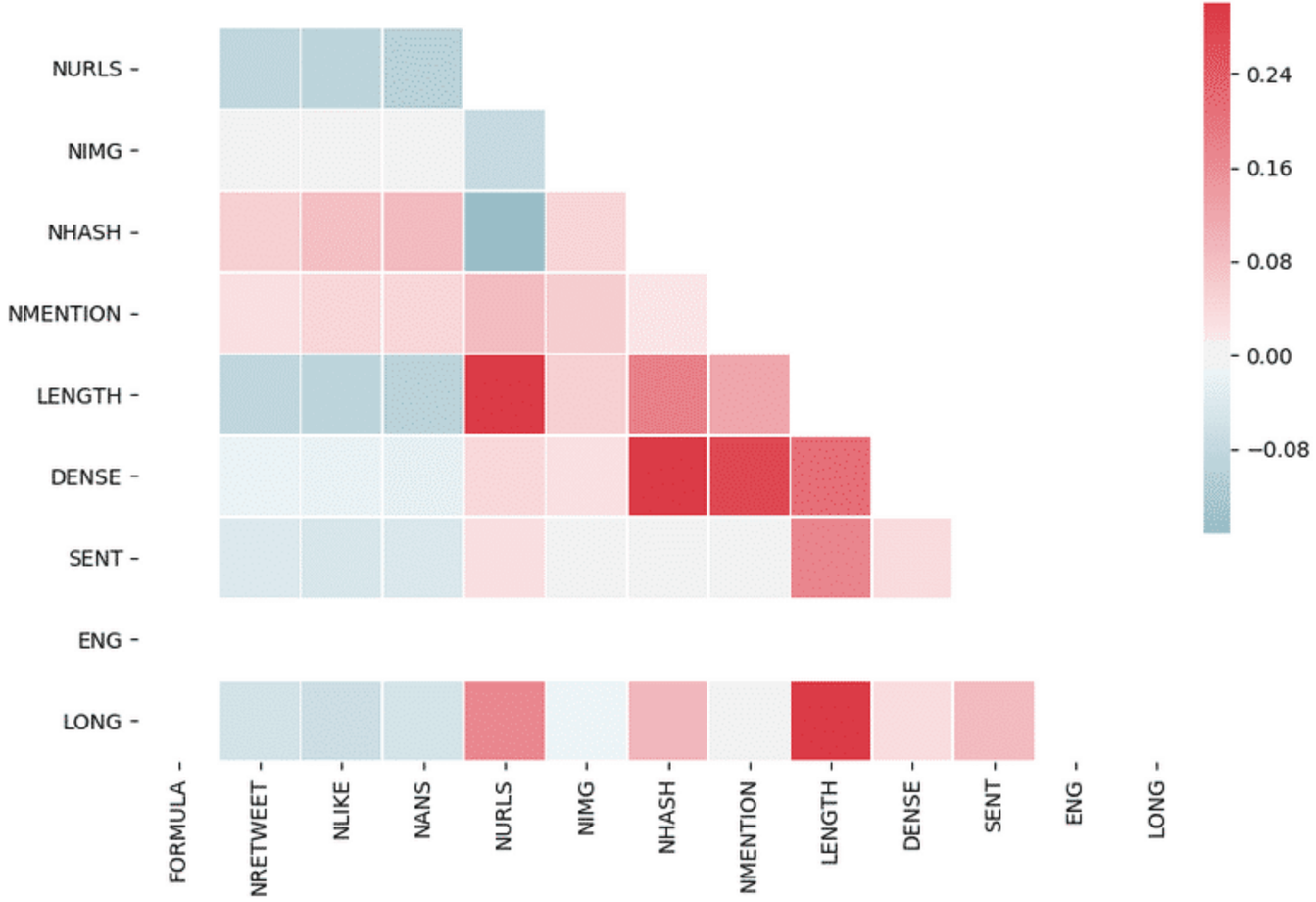
È necessario definire quali sono le proprietà del messaggio che sono in **relazione** con la variabile target, appena stabilita.
Per fare questo si studiano le relazioni tra le caratteristiche, mediante la matrice di correlazione.

Vengono scelte:

Num. Immagini

Num. Hashtag

Num. Menzioni



Matrice di correlazione su gruppo 1

COSTRUZIONE CLASSIFICATORE

1. Scelta caratteristiche

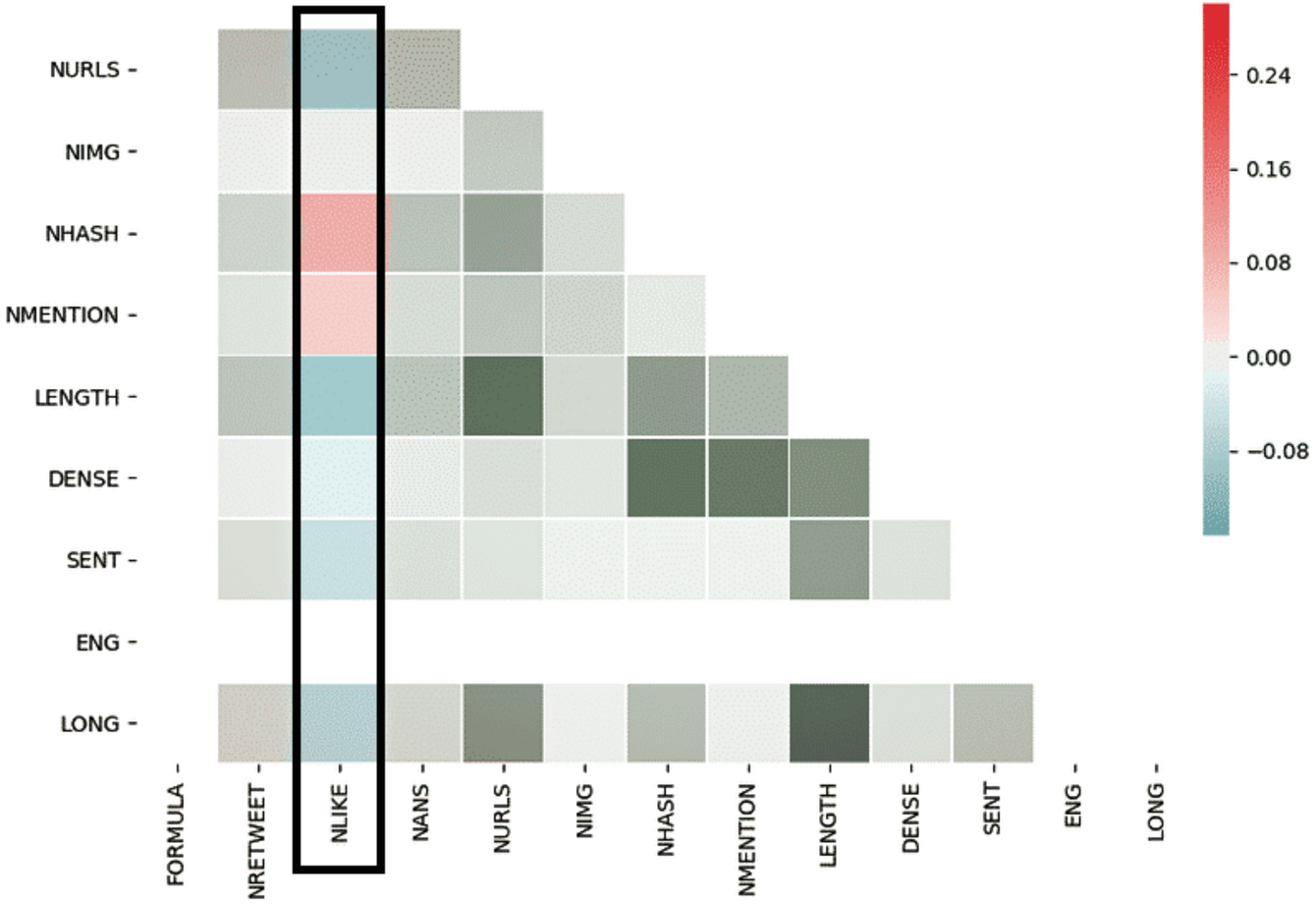
È necessario definire quali sono le proprietà del messaggio che sono in **relazione** con la variabile target, appena stabilita.
Per fare questo si studiano le relazioni tra le caratteristiche, mediante la matrice di correlazione.

Vengono scelte:

Num. Immagini

Num. Hashtag

Num. Menzioni



Matrice di correlazione su gruppo 1

2. Algoritmo K Nearest Neighbors

Una volta che i data set sono stati divisi in due sotto-insiemi (**training set** e il **testing set**), il classificatore deve essere istruito:

- Ogni elemento viene categorizzato in base alla classe dei **K** elementi a lui più vicino.
- Si confronta l'etichetta restituita con quella definita in precedenza, per calcolare l'**accuratezza** del modello.

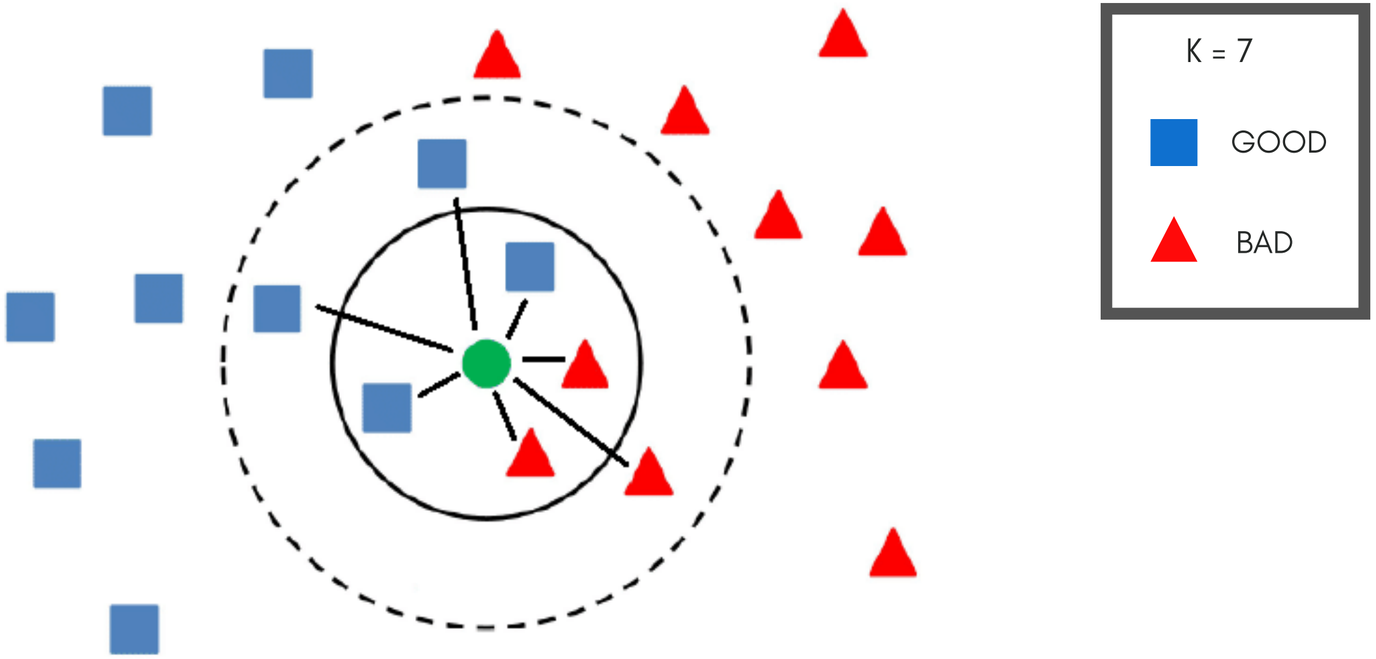
Con questo tipo di algoritmi, il concetto di distanza viene inglobato in quello di **similarità**: due elementi sono tanto più vicini, in uno spazio vettoriale, quanto simili.

3. Calcolo suggerimenti

Utilizzando nuovamente l'**algoritmo KNN**, vengono studiate le caratteristiche dei messaggi vicini a quello iniziale, per poter elaborare possibili suggerimenti.

VICINI

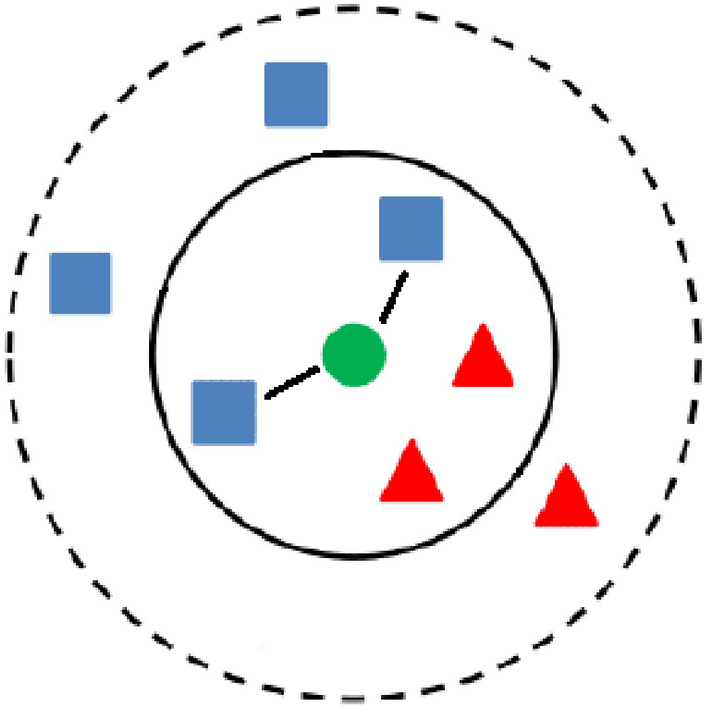
Il nuovo messaggio viene categorizzato dal modello, che ritorna i suoi vicini

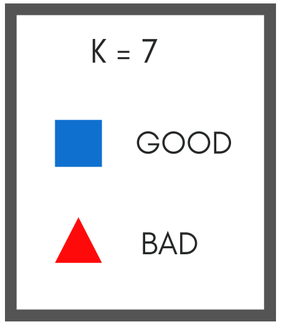


3. Calcolo suggerimenti

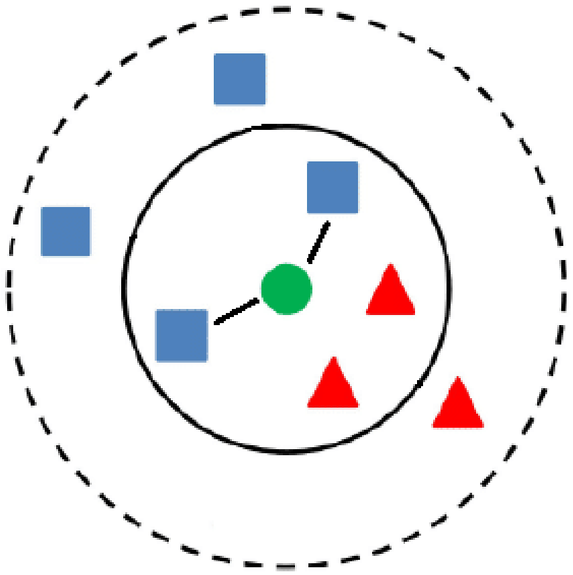
FILTRI

Vengono eliminati i vicini negativi e quelli che superano una soglia di distanza





3. Calcolo suggerimenti



MEDIA

Dei vicini rimasti, calcola la quantità media per ogni caratteristica

SUGGERIMENTO

Si confronta la media con le proprietà del messaggio iniziale, per formulare i suggerimenti

Media delle caratteristiche dei suoi vicini

NIMG	NHASH	NMENTION
2	3	0

Caratteristiche nuovo messaggio

NIMG	NHASH	NMENTION
3	1	0

Struttura progetto

1 Tecnologie utilizzate

2 Il caso di studio

Lato Server

3 Costruzione classificatore

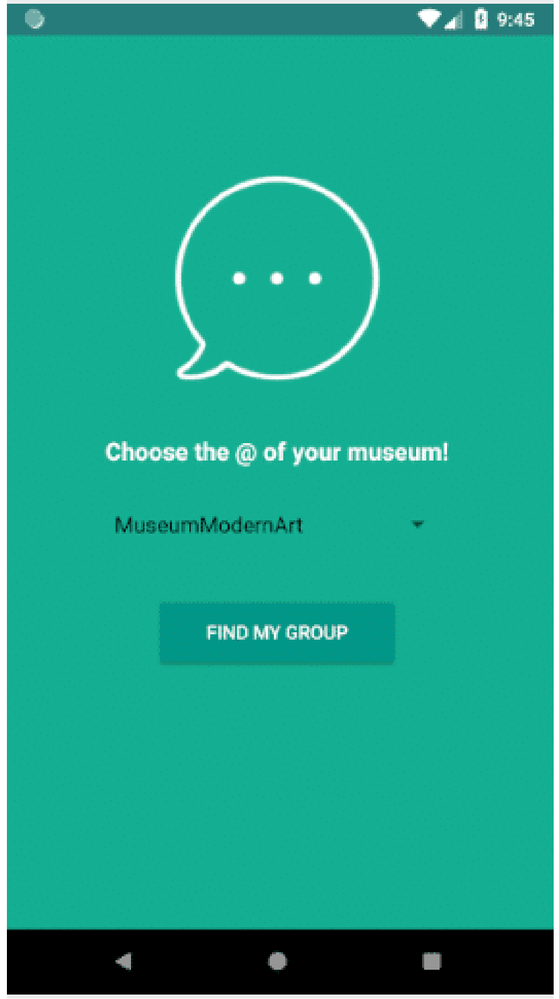
Lato Client

4 Applicazione Android

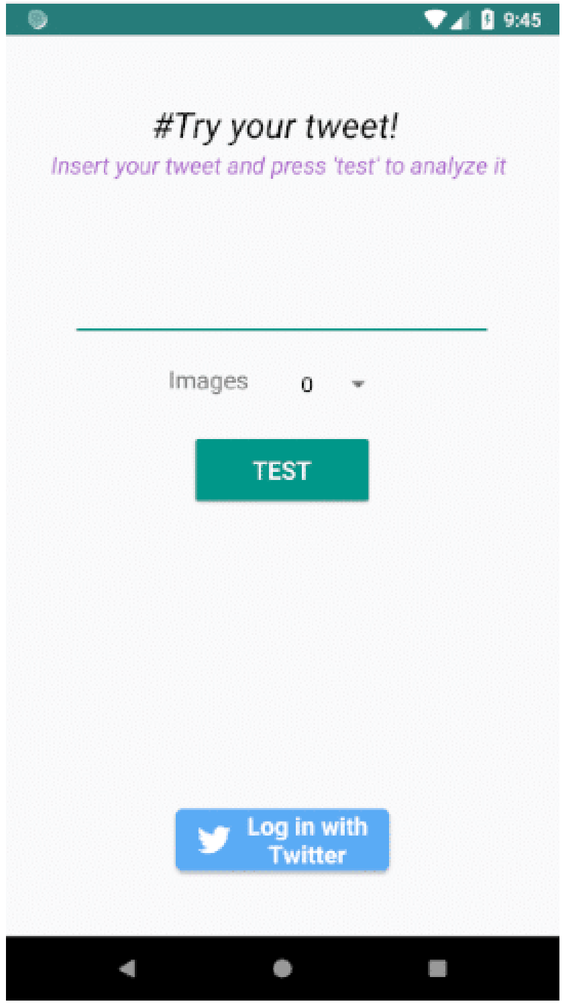
5 Prove sperimentali e risultati ottenuti

APPLICAZIONE ANDROID

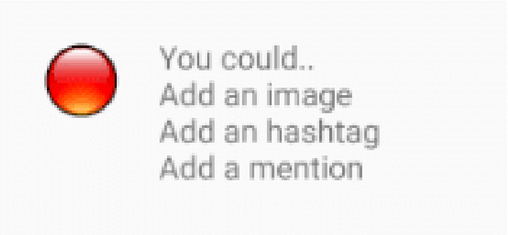
1



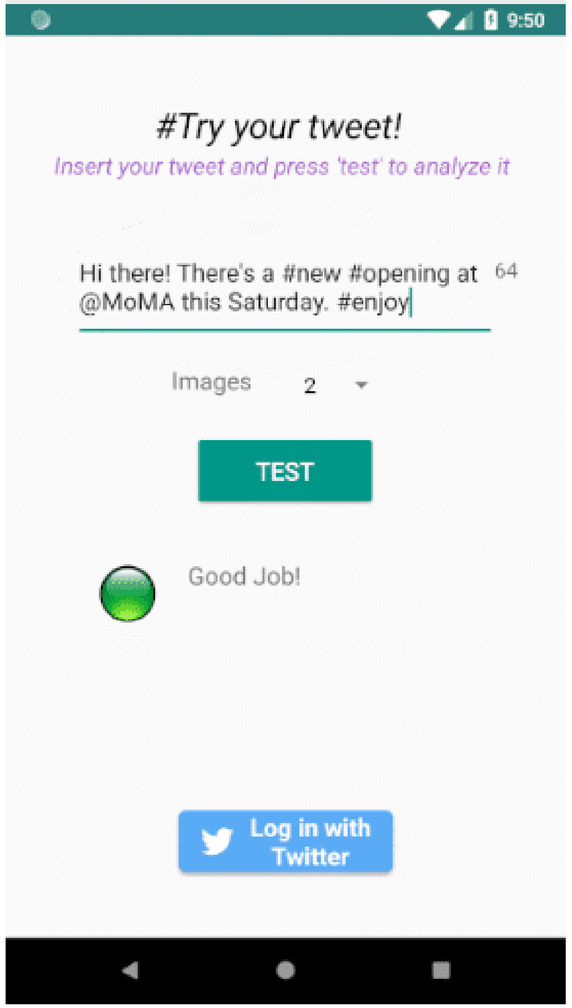
2



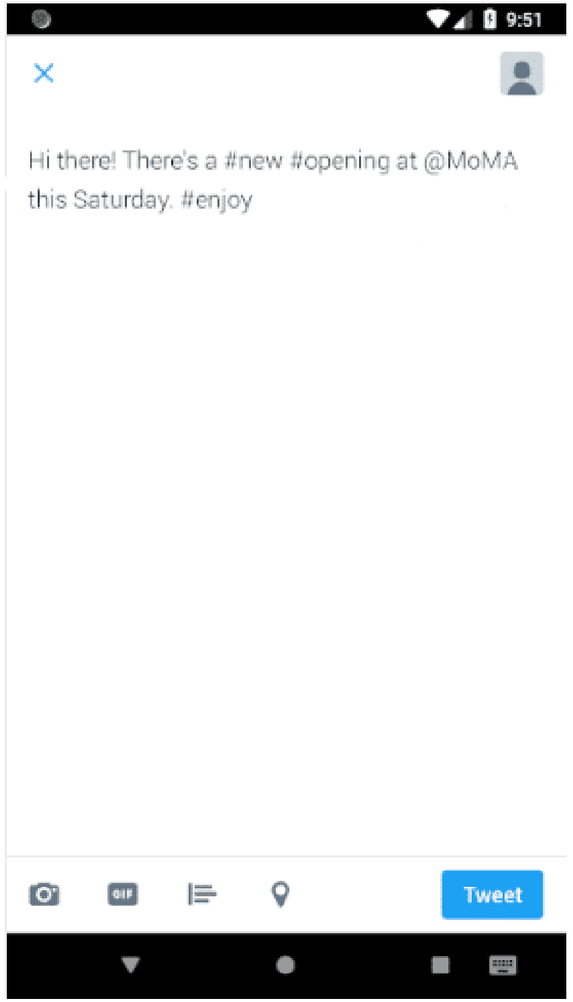
3


You could..
Add an image
Add an hashtag
Add a mention

4


9:50
#Try your tweet!
Insert your tweet and press 'test' to analyze it
Hi there! There's a #new #opening at @MoMA this Saturday. #enjoy
64
Images
2
TEST
Good Job!
Log in with Twitter

5


9:51
Hi there! There's a #new #opening at @MoMA this Saturday. #enjoy
Tweet

Struttura progetto

1 Tecnologie utilizzate

2 Il caso di studio

Lato Server

3 Costruzione classificatore

Lato Client

4 Applicazione Android

5 Prove sperimentali e risultati ottenuti

Prove sperimentali e risultati ottenuti

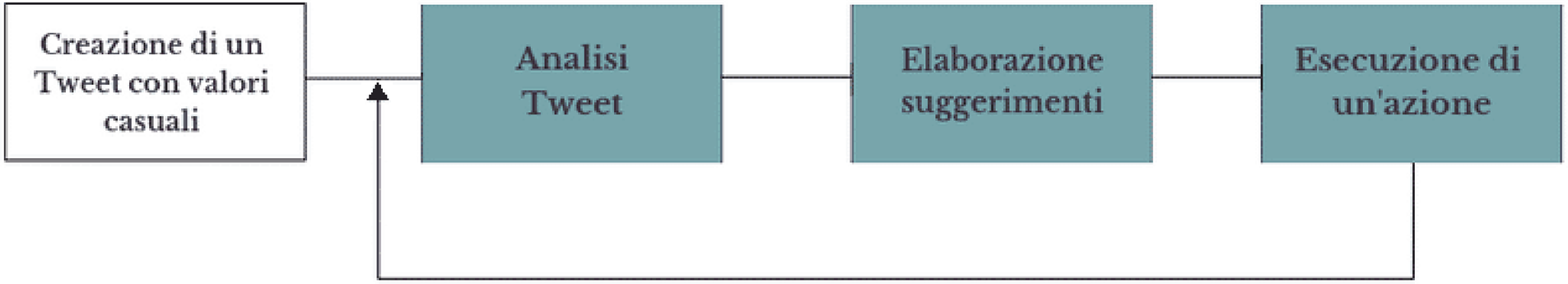
1. Accuratezza modello

Descrizione	Caratteristiche	Accuratezza % (dopo 10 run)					
		GRUPPO 1	GRUPPO 2	GRUPPO 3	GRUPPO 4	GRUPPO 5	GRUPPO 6
Caratteristiche scelte	**NIMG - NHASH - NMENTION**	**0,77**	**0,78**	**0,77**	**0,79**	**0,78**	**0,80**

Viene mostrata l'**accuratezza** del modello, in base alla scelta delle **caratteristiche**, la quale viene calcolata come media dopo 10 esecuzioni del programma e poi riportata in percentuale.

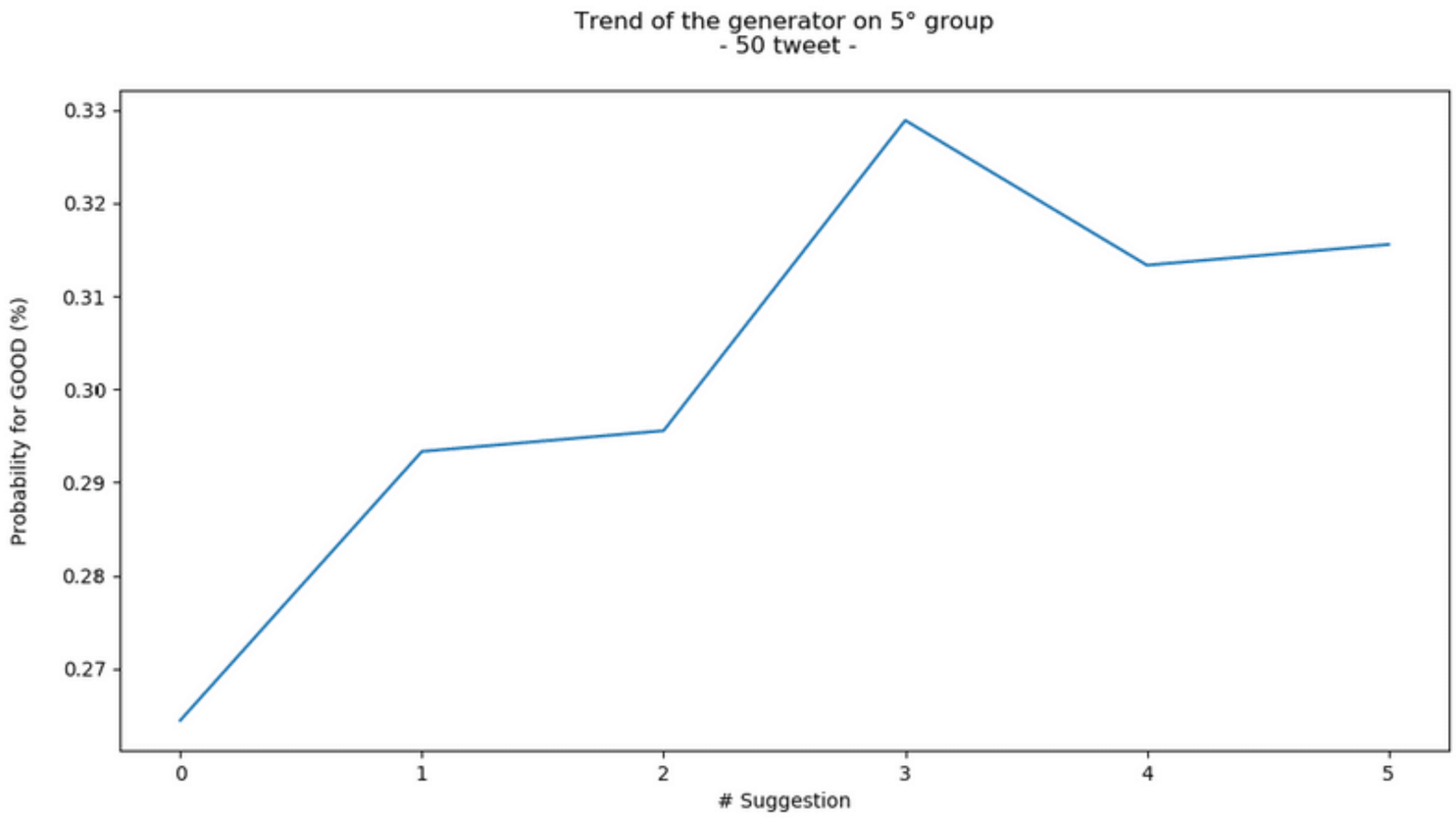
2. Accuratezza suggerimenti - generatore tweet

Per verificare l'accuratezza dell'algoritmo per i suggerimenti è stato ideato un **generatore randomico** di tweet che, ciclicamente, sceglie ed esegue un suggerimento restituito, tenendo traccia dell'esito del messaggio.



Il grafico a fianco mostra come l'**andamento** dell'esito medio di 50 tweet, generati in modo casuale, possa variare man mano che vengono seguiti i **suggerimenti**.

Si studia la **probabilità** dei messaggi di ricadere nella categoria "GOOD": ci si aspetta una tendenza crescente.



CONCLUSIONI & SVILUPPI FUTURI

Con il progetto presentato è stato possibile definire un classificatore che **categorizza** i tweet in due classi, restituendo suggerimenti per migliorare il loro effetto sul pubblico.

Come possibili **sviluppi futuri**:

Aumentare la capacità del classificatore aggiungendo **più etichette** disponibili per la categorizzazione.

Inserire una sezione per analizzare anche le **immagini** nei messaggi, tali che possano attirare meglio l'attenzione del pubblico.

Prendere informazioni da Twitter istruendo il classificatore sugli **hashtag in tendenza** e favorire quelli per i suggerimenti.

Ringrazio per l'attenzione.