

# Instagram Fake Users Recognition with different Classification Techniques

Fernando De Nitto  
Francesco Fornaini

Data Mining  
Department of Information Engineering  
University of Pisa

26/02/2020

## **Abstract**

Qui ci scriviamo un summary di TUTTO breve

## **1 Project**

Scrivere cosa volevamo raggiungere con questo progetto del cazzo dio cane. Dire che il progetto innovativo visto che la maggior parte li fanno su twitter.

### **1.1 Instagram**

Scrivere cosa vogliamo fare

### **1.2 Why Fake Detection is important?**

Per giustificare che l'applicazione ha un'utilita'

### **1.3 Problems and Challenges**

In this section we want to discuss why spend efforts on Instagram Social Network and which challenges are linked with this kind of Social Network. All the considerations could be extend in every social network that follows the pattern of Instagram.

Fare elenco puntato delle problematiche tipo: API (chiuse e piccolo confronto con Twitter) , Costruzione del dataset (che non si trova in giro)

## **1.4 Steps**

Individuare gli step che portano dalla costruzione dei dati (scraping etc) fino all'applicazione finale.

## **1.5 Tools**

Scrivere cosa usiamo per portare avanti il progetto cio python e tutte le librerie (moduli).

# **2 Dataset**

Scrivere cosa vogliamo fare

## **2.1 Approach Utilizzato per costruire il dataset**

Reali seguono reali fake seguono fake

## **2.2 Attribute Selection**

Scrivere gli attributi giustificando che non ci sono correlazioni evidenti tra di loro e che abbiamo preso ordini statistici standard invece per gli attributi relativi ai contenuti.

## **2.3 Problems and Challenges**

Alcuni fake venivano cancellati Problema che utenti privati diventavano pubblici e viceversa e risoluzione delle incongruenze (questo ci ha dato spunto per non utilizzare le statistiche a volte)

# **3 Classification**

Scrivere in generale quelli che sono i classificatori utilizzati e che si sono fatte le 10 cross validation su tutti giustificando il motivo (vedere slides lezioni) e poi confrontare in fondo

## **3.1 SVM**

Spiegare per ogni metodo in generale come funziona ed i risultati ottenuti con relativi grafici e risultati.

## **3.2 KNN**

Scrivere cosa vogliamo fare

### **3.3 Multi-Layer Perceptron**

Scrivere cosa vogliamo fare

### **3.4 Decision Tree**

Scrivere cosa vogliamo fare

### **3.5 Random Forest**

Scrivere cosa vogliamo fare

### **3.6 Classifications results**

Scrivere quali classificatori hanno fornito i risultati migliori e giustificare il fatto che come ci aspettavamo l'albero decisionale dava un migliore risultato visto che gli oggetti non avevano dei forti legami statistici (non andava bene Bayes) e che il classificatore che pi si avvicina al comportamento umano di riconoscimento dei fake.

FARE TABELLA (tutti con 10-cross fold validation da separare tra quelle con Stats e quelle senza Stats)

## **4 Final Results**

Scrivere cosa vogliamo fare

### **4.1 Final Application**

Scrivere cosa vogliamo fare

## **5 Conclusion**

Scrivere cosa vogliamo fare

## **6 Further Applications**

Sentiment analysis etc. migliorarla con fake detection per analizzre il sentimento reale.