|  |  |
| --- | --- |
| Comprensione del comportamento scorretto nella discussione di COVID19 sui social network e sui media online | |
|  | |
| Studente/i | Relatore |
| Nogara Gianluca | Giordano Cremonese Silvia |
| Correlatore |
| Luceri Luca |
| Committente |
| - |
| Corso di laurea | Modulo |
| Ingegneria Informatica TP | C10344 Progetto di diploma |
| Anno |  |
| 2020/2021 |  |

|  |  |
| --- | --- |
| Data |  |
| - |  |

**Indice**

[Abstract 7](#_Toc81838690)

[Progetto assegnato 9](#_Toc81838691)

[1 Analisi 10](#_Toc81838692)

[1.1 Contesto 10](#_Toc81838693)

[1.2 Fruitori 10](#_Toc81838694)

[1.3 Requisiti 10](#_Toc81838695)

[1.3.1 Contenuti 10](#_Toc81838696)

[1.3.2 Rispetto della privacy 11](#_Toc81838697)

[1.4 Obiettivi 12](#_Toc81838698)

[2 Stato dell’arte 13](#_Toc81838699)

[2.1 Analisi dei dati 13](#_Toc81838700)

[2.2 Linguaggi di programmazione e ambiente di sviluppo 14](#_Toc81838701)

[2.3 Librerie 14](#_Toc81838702)

[2.3.1 Pandas 15](#_Toc81838703)

[2.3.2 Matplotlib e Plotly 15](#_Toc81838704)

[2.3.3 NetworkX 16](#_Toc81838705)

[2.4 Strumenti utilizzati 17](#_Toc81838706)

[2.4.1 Botometer 17](#_Toc81838707)

[2.4.2 Media Bias/Fact Check 18](#_Toc81838708)

[3. Implementazione 20](#_Toc81838709)

[3.1 Familiarizzare con i dati 20](#_Toc81838710)

[3.2 Parallelizzazione 21](#_Toc81838711)

[3.3 Score con Botometer 23](#_Toc81838712)

[3.3 Disinformation Dozen 25](#_Toc81838713)

[5 Piani di lavoro 27](#_Toc81838714)

[5.1 GitHub 27](#_Toc81838715)

[6 Conclusioni 27](#_Toc81838716)

[6.1 Problemi riscontrati 27](#_Toc81838717)

[6.2 Implementazioni future 27](#_Toc81838718)

[7 Fonti 28](#_Toc81838719)

**Indice delle figure**

[Figura 1: Esempio della struttura di tweets filtrata su poche colonne 11](#_Toc81838723)

[Figura 2: Esempio della struttura di profili filtrata su poche colonne 11](#_Toc81838724)

[Figura 3: Esempio di plot con matplotlib (10 utenti che effettuano più tweet) 15](#_Toc81838725)

[Figura 4: Esempio di mappa interattiva generata con Plotly 16](#_Toc81838726)

[Figura 5: Piani offerti da Botometer Pro su RapidAPI 17](#_Toc81838727)

[Figura 6: Esempio di classificazione della nota rivista statunitense "Forbes" 19](#_Toc81838728)

[Figura 7: Esempio di analisi sui dati condotta sul sample (Domini più condivisi) 20](#_Toc81838729)

[Figura 8: Differenza tra esecuzione in parallelo (a sinistra) e seriale (a destra) 22](#_Toc81838730)

[Figura 9: Risposta di Botometer API alla richiesta di score per un utente 23](#_Toc81838731)

[Figura 10: Distribuzione del CAP score con indicazione del primo dell'ultimo decile 24](#_Toc81838732)

[Figura 11: Tipologie di Bot ottenute 25](#_Toc81838733)

Abstract

I social network e i media online (OSNEM) sono tecnologie basate su computer che consentono agli utenti di creare contenuti e intrattenere relazioni sociali.

Le OSNEM sono rapidamente cresciute dall’essere semplicemente un canale di aggregazione all’essere un fenomeno globale che ha suscitato un cambio di paradigma della nostra società, trasformando il modo in cui le persone accedono alle notizie, condividono opinioni, fanno affari e fanno politica.

In uno scenario del genere, l'accuratezza, la veridicità e l'autenticità del contenuto condiviso sono ingredienti necessari per mantenere una sana discussione online.

Tuttavia, negli ultimi tempi, le OSNEM hanno dovuto affrontare una notevole crescita di account e attività dannose, che minano intenzionalmente l'integrità delle conversazioni online condividendo, nelle piattaforme OSNEM, informazioni false e provocatorie per influenzare l'opinione pubblica e creare conflitti su questioni sociali o politiche.

Ciò è stato, e continua ad essere, drammaticamente vero per la discussione sul COVID19, che è stata minata da campagne di manipolazione e disinformazione a livello globale. È quindi necessario studiare i comportamenti peculiari delle entità dannose nel dibattito COVID19 per fornire un cambiamento sociale verso una migliore comprensione (alfabetizzazione) delle reti sociali.

Questo include il rilevamento di malintenzionati, l'identificazione di informazioni di scarsa credibilità e notizie false, la consapevolezza degli utenti e il controllo dei comportamenti scorretti.

Social networks and online media (OSNEMs) are computer-based technologies that allow users to create content and engage in social relationships.

OSNEMs have rapidly grown from being simply a channel of aggregation to being a global phenomenon that has sparked a paradigm shift in our society, transforming the way people access news, share opinions, do business, and make policy.

In such a scenario, accuracy, truthfulness and authenticity of shared content are necessary ingredients to maintain a healthy online discussion.

However, in recent times, OSNEMs have faced a significant growth of malicious accounts and activities, which intentionally undermine the integrity of online conversations by sharing, in OSNEM platforms, false and provocative information to influence public opinion and create conflict on social or political issues.

This has been, and continues to be, dramatically true of the COVID19 discussion, which has been undermined by global campaigns of manipulation and misinformation. Therefore, it is necessary to study the peculiar behaviors of malicious entities in the COVID19 debate to provide social change towards a better understanding (literacy) of social networks.

This includes detection of malicious actors, identification of information of low credibility and fake news, user awareness, and control of misbehavior.

Progetto assegnato

Lo scopo di questo progetto è quello di analizzare entità malevoli presenti in un dataset pubblico su COVID19. L’ OSNEM di riferimento è Twitter e l’attività di analisi e classificazione delle entità malevoli (come bot e trolls) viene fatta tramite diversi algoritmi esistenti.

Il progetto in questione consente quindi di analizzare il comportamento all'interno della discussione e l’impatto che hanno tali entità malevole nella discussione globale.

**COMPITI**

• Identificazione degli strumenti usati per la propagazione di notizie

• Identificazione delle entità nella discussione

• Analisi comportamentale delle entità malevoli e confronto con entità competenti

**OBIETTIVI**

L'obiettivo principale di questa tesi è quello di arrivare a distinguere tra entità/azioni corrette e quelle malevoli nel dibattito COVID19.

Identificare quindi tramite quali mezzi e comportamenti le varie entità malevoli influsicono nelle discussioni, analizzando hashtags, link condivisi e attività, facendo una comparazione tra chi ha l’obiettivo di informare correttamente

**TECNOLOGIE**

• Python

• Complex networks

# 1 Analisi

## 1.1 Contesto

Il progetto si inserisce nel contesto del Dipartimento Tecnologie Innovative della SUPSI e in particolare nel corso di Ingegneria Informatica. Si tratta di un progetto interno alla SUPSI, anche se di grande importanza sociale dal momento che le informazioni raccolte possono essere utilizzate per limitare la propagazione di entità malevoli sui social network e sensibilizare su un tema tanto delicato e attuale.

## 1.2 Fruitori

Il progetto in questione è destinanto ad un pubblico intenzionato a informarsi sulle OSNEM nel dibattito da COVID19, e di come le varie entità interagiscono sul tema.

La finalità scientifica del progetto lo rende pubblicabile ed espandibile per nuovi studi di vario genere e operazioni di aggiornamento sui dati.

## 1.3 Requisiti

### 1.3.1 Contenuti

Il principale requisito a livello contenutisco è quello di avere dei costanti feedback visivi dell’analisi effettuata sui dati, in modo da effettuare operazioni di ricerca sempre più dettagliata in funzione di risultati.

Una volta identificate le strategie di condivisione delle informazioni è quindi possibile andare a stabilire di quali tipologie di utenti si tratta e definire quindi nuove metriche di valutazione per la credibilità di tali utenti.

Identificando utenti malevoli è inoltre possibile effettuare studi sulle loro interazioni attive e passive, definendo delle community potenzialmente utili per ricavere ulteriori inormazioni

### 1.3.2 Rispetto della privacy

Per lo svolgimento del progetto sono state utilizzate grosse quantità di dati pubblici ottenuti tramite un repository che contiene l’id dei tweets coinvolti nella discussione sul COVID19 (visitabile al link <https://github.com/echen102/COVID-19-TweetIDs>) divisi per mese.

I dati pubblicati nel repository in questione sono solo id (per questioni di policy di Twitter), pertanto non contengono informazioni ulteriori sugli utenti e/o sui tweets.

Dal momento che i dati così non avrebbero alcuna informazione, per avere accesso a informazioni complete è stato eseguito un processo di idratazione, che consiste nel richiedere a Twitter, via API, di avere i tweets completi a partire da una serie di tweets id.

Dal momento che si tratta di dati pubblici in una piattaforma aperta che possono essere visionati anche senza registrazione su Twitter non ci sono problemi legati alla privacy nella condivisione di tali.

Immagine che contiene testo

Descrizione generata automaticamente

Figura 1: Esempio della struttura del DataFrame dei tweets filtrata su poche colonne

Immagine che contiene testo

Descrizione generata automaticamente

Figura 2: Esempio della struttura del DataFrame dei profili filtrata su poche colonne

## 1.4 Obiettivi

Gli obiettivi di questo progetto sono i seguenti:

* Investigazione sulle strategie di condivisione (tweet, retweet, risposte)
* Classificazione delle enità malevoli
* Investigazioni sulle attività delle attività malevoli
* Studio degli utenti che interagiscono con tali entità
* Studio della credibilità dei domini
* Comparazione delle attività tra entità malevoli e attendibili

# 2 Stato dell’arte

## 2.1 Analisi dei dati

Come situazione di partenza è stato messo a disposizione un piccolo dataset di 50,000 tweets con i relativi utenti per prendere confidenza con dati eterogenei.

I dataset in questione riguardano i profili utente e i singoli tweet, la struttura dei due dataset è completamente differente, in particolare quello dedicato agli utenti è molto ricco di colonne (ben 46), mentre quello dedicato agli utenti ne ha poco meno della metà (18).

Questa differenza è data dal fatto che, per quanto riguarda i tweet, abbiamo a disposizione molte più informazioni rispetto a dei semplici profili utente.

Di seguito sono mostratie le colonne più utilizzate dei due dataset:

**Tweets**

|  |  |  |
| --- | --- | --- |
| Campo | Tipologia | Descrizione |
| id | int64 | Id univoco del tweet |
| created\_at | object | Data di creazione del tweet |
| user\_id | int64 | Id univoco del profilo |
| user\_screen\_name | object | Nome utente del profilo |
| text | object | Contenuto del tweet |
| retweet\_count | int64 | Numero di retweet del tweet |
| in\_reply\_to\_user\_id | float64 | Id dell’utente che riceve la risposta |
| in\_reply\_to\_screen\_name | object | Username dell’utente che riceve la risposta |
| rt\_user\_id | float64 | Id dell’utente che riceve il retweet |
| rt\_user\_screen\_name | object | Username dell’utente che riceve il retweet |
| hashtags | object | Hashtag usati nel tweet |
| urls | object | Link condivisi nel tweet |

**Profili**

|  |  |  |
| --- | --- | --- |
| Campo | Tipologia | Descrizione |
| id | int64 | Id univoco dell’utente |
| Screen\_name | object | Username dell’utente |
| location | object | Località dell’utente |
| follower\_count | int64 | Numero di follower dell’utente |
| verified | bool | Booleano che indica se il profilo è verificato |
| geo\_enabled | bool | Booleano che indica se il profilo ha una location |

I dati da analizzare risultano eterogenei, sono presenti infatti valori booleani, valori interi, oggetti e valori in virgola mobile, è quindi importante effettuare un preprocessamento per un corretto studio.

## 2.2 Linguaggi di programmazione e ambiente di sviluppo

Dal momento che i dati su cui lavorare sono file con formato .csv di grossa dimensione la scelta del linguaggio è ricaduta su Python, questo anche perché fornisce una grande varietà di librerie e tools per data analysi e complex networks.

Anaconda ha svolto un compito molto importante per quanto riguarda la programmazione in Python e la realizzazione di feedback visivi, infatti dal momento che per un’analisi corretta si ha bisogno di generare report grafici, Anconda contiene il software Jupyter Notebook.

Quest’ultimo è un'applicazione Web open source che rappresenta lo strumento perfetto per effettuare operazioni di data analysis grazie alle celle che consentono di salvare rappresentazioni grafiche e testo senza dover rieseguire necessariamente tutto il codice.

## 2.3 Librerie

Durante lo sviluppo del progetto si è fatto uso di una serie di librerie che hanno semplificato lo svolgimento del progetto, le più importanti hanno dato un contributo significativo e sono state utilizzate con una certa frequenza

### 2.3.1 Pandas

Pandas è tra le librerie più importanti in Python dal momento che è stato sviluppato per la manipolazione e l'analisi dei dati.

In particolare, offre strutture dati e operazioni per manipolare tabelle numeriche e serie temporali.

L’utilizzo di pandas ha consentito lo sfruttamento dei DataFrame per la lettura dei file .csv e la gestione dei relativi dati contenuti.

I DataFrame hanno rappresentato il cuore pulsante del lavoro svolto dal momento che sono altamente gestibili e semplici nel filtraggio.

Lo sfruttamento della libreria ha permesso inoltre l’utilizzo delle Series, ndarray monodimensionali che espongono un’infità di metodi per la manipolazione dei dati. Questa tipologia di strutture sono state largamente utilizzate per effettuare classificazioni basate sul numero di occorrenze.

### 2.3.2 Matplotlib e Plotly

Matplotlib è una libreria per la creazione di grafici ed è stata largmente utilizzata per la rappresentazione grafica di diverse tipologie di classificazioni, quali istogrammi, barplot e distribuzioni varie.

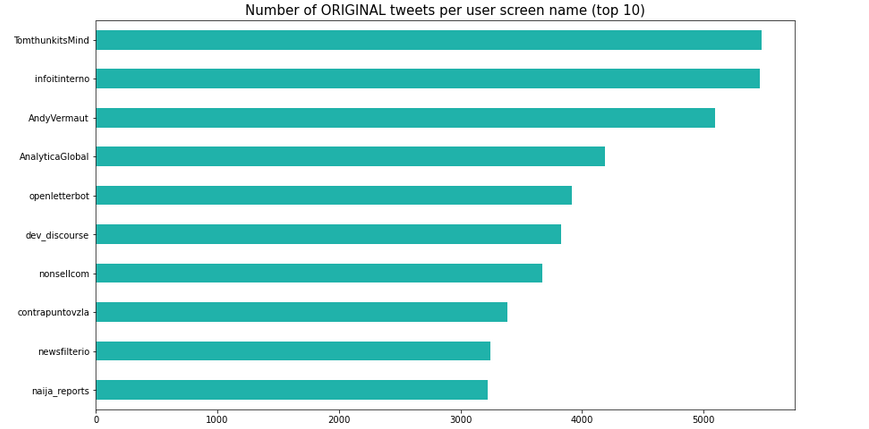
****

Figura 3: Esempio di plot con matplotlib (10 utenti che effettuano più tweet)

Dal momento che i plot generati sono statici essi risultano molto veloci da generare e leggeri da salvare, anche per quantità molto grosse di dati.

In contrapposizione vi è Plotly, libreria di grafici che viene creare grafici interattivi browser-based, che consente quindi di generare report dinamici ma che, in presenza di una quantità di dati significativa, risulta poco efficiente e pesante.

Immagine che contiene mappa

Descrizione generata automaticamente

Figura 4: Esempio di mappa interattiva generata con Plotly

### 2.3.3 NetworkX

Si tratta di una libreria realizzata per semplificare la generazione e lo studio di grafi e reti, è stata largamente utilizzata per classificare gli utenti in community e capire quindi i legami che hanno l’un l’altro.

NetworkX è una libreria adatta per il funzionamento su dati molto consistenti: ad esempio, grafici con più di 10 milioni di nodi e 100 milioni di archi, dal momento che è realizzata in puro Python, infatti è ragionevolmente efficiente, molto scalabile e altamente portabile per l'analisi di reti e social network.

## 2.4 Strumenti utilizzati

Durante lo svolgimento del progetto sono vi è stata la necessità di utilizzare strumenti realizzati da terzi per poter dare una validità ai dati generati o ottenere determinate informazioni.

### 2.4.1 Botometer

Si tratta di un progetto del gruppo OsoMe (Observatory on Social Media) sviluppato all'Indiana University.

OSoMe è una collaborazione tra il Network Science Institute (IUNI), il Center for Complex Networks and Systems Research (CNetS) e la Media School dell'Indiana University.

Botometer è facilmente fruibile tramite l’apposito sito web, raggiungibile al link <https://botometer.osome.iu.edu/>, ma espone anche un’API pubblica.

Tutte le informazioni per l’integrazione di Botometer sono disponibili sulla pagina GitHub del progetto (al link <https://github.com/IUNetSci/botometer-python>).

Per poter utilizzare l’API è necessario passare per RapidAPI e da Twitter.

RapidAPI è l'hub di API più grande al mondo, registrandosi gratuitamente è possibile accedere a Botometer Pro e sottoscrivere il piano adatto per il lavoro.

Immagine che contiene testo, monitor, screenshot, schermo

Descrizione generata automaticamente

Figura 5: Piani offerti da Botometer Pro su RapidAPI

Una volta sottoscritto il piano si ottengono due key, rispettivamente una key\_rapidapi e una consumer\_key, per poter accedere appunto all’API.

Una volta ottenuti i dati da RapidAPI è necessario accedere alla Twitter API e ottenere i permessi da sviluppatore e realizzare quindi una Twitter app.

Anche in questo caso si otterranno delle keys: consumer\_secret, bearer\_token, access\_token e access\_token\_secret.

Tutte queste keys sono essenziali dal momento che vengono utilizzate da Botometer per poter effettuare lo score sugli utenti passati.

È possibile utilizzare Botometer tramite un’apposita libreria:



Per quanto riguarda lo svolgimento del progetto si è utilizzato uno script già esistente con una serie di credenziali, il quale è stato realizzato in modo tale da effettuare un cambio di credenziali nel momento in cui il numero di richieste disponibili termini, in modo da effettuare quante più richieste possibili giornalmente.

Per l’esecuzione, lo script, richiede che vengs passato un file .csv formato da campi id\_utente,nome\_utente ritornando un file .json contente gli score di Botometer.

### 2.4.2 Media Bias/Fact Check

Media Bias / Fact Check è un sito Web che valuta qualitativamente i media (online, cartacei e radiofonici) in termini di verifica dei fatti e delle fonti.

Il sito è fruibile gratuitamente in lingua inglese al link <https://mediabiasfactcheck.com/>.

Si tratta di un sito utilizzato per verificare la credibilità di un dominio e non solo si occupa infatti di fornire infomazioni anche su orientamento politico-religioso ed un'eventuale manipolazione informativa, essenziale per lo studio di fattibilità delle fonti che vengono condivise sui social.

Nonostante il metodo utilizzato dal sito Web non sia affatto scientifico, la sua attendibilità è inopinabile. Il sito è una fonte largamente citata per notizie e studi sulla sifinromazione, è stato infatti utilizzato dai ricercatori dell’Università del Michigan per tracciare la diffusione delle “fake news” e fonti opinabili sui social media.

Immagine che contiene testo

Descrizione generata automaticamente

Figura 6: Esempio di classificazione della nota rivista statunitense "Forbes"

Dal momento che il sito non espone API per l’integrazione in Python, sono stati classificati manualmente oltre 800 domini, assegnando ad ogni nome uno score di credibilità (“high” o “low”) e una spiegazione della classificazione.

# 3. Implementazione

## 3.1 Familiarizzare con i dati

Per poter condurre correttamente uno studio efficace su dati nuovi è necessario prendere confidenza con tali.

Il primo task svolto per la realizzazione del progetto è stato un breve studio generico sui dati che ha riguardato sostanzialmente tutte le colonne di entrambi i DataFrame.

Le prime analisi condotte, seppur molto superfciali, hanno avuto un certo riscontro con analisi precedentemente condotte da studentesse del Master, dal momento che il sample è un subset di un DataFrame molto più grande in utilizzo.

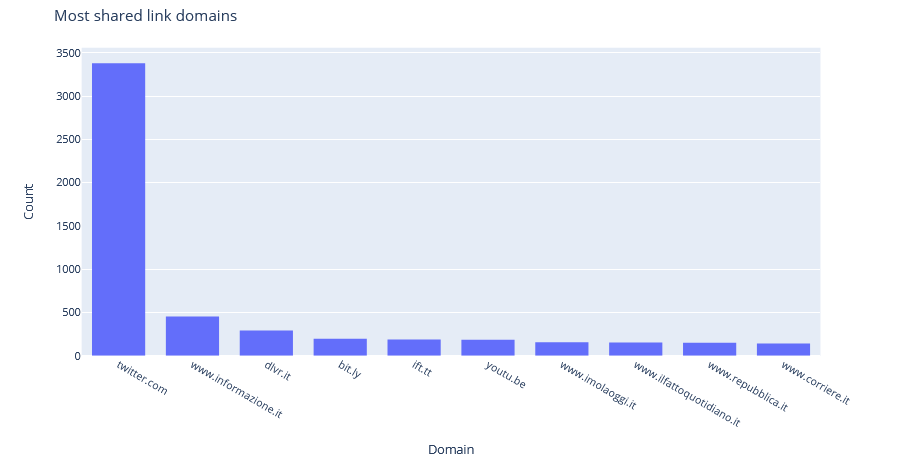


Figura 7: Esempio di analisi sui dati condotta sul sample (Domini più condivisi)

Le operazioni svolte sul set dei tweets sono state le seguenti:

* Divisione del DataFrame in SubDataFrame per tipologia di tweet (retweet, risposte e post originali)
* Plot degli utenti più attivi per tipologia di tweet
* Plot timeline dei tweets
* Plot domini più condivisi
* Ricerca di tweet contenenti una serie di hashtags sospetti (#iononmivaccino, #bigpharma, #nessunacorrelazione, #TuttiComplici e #iononmivaccinerò)

Le operazioni svolte sul set deglu utenti sono invece le seguenti:

* Plot timeline della creazione dei profili
* Plot piechart degli utenti verificati e non
* Plot piechart degli utenti con una località impostata e non
* Ricerca di potenziali utenti “to be scored” con poche informazioni
* Ricerca delle persone con la parola “no vax” in descrizione
* Plot di una mappa interattiva sulla località degli utenti

## 3.2 Parallelizzazione

Una volta effettuata una presa di confidenza con i dati di sample sono stati ottenuti i file veri e propri su cui è stato condotto lo studio.

Si tratta di un insieme di 10 file di tweets e profili, la cui origine e struttura è stata precedentemente spiegata.

La quantità dei dati è rilevante, in particolare vi sono:

* 66.412.411 di attività
* 13.999.715 di tweets
* 48.312.504 di retweets
* 4.100.192 di risposte

Che possiamo identificare meglio graficamente:

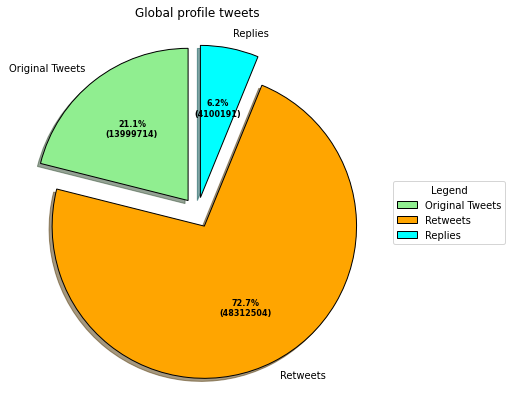


Figura 8: Statistiche in funzione della tipologia di tweet sull'intero DataFrame

Una volta presa visione dei dati è stato necessario effettuare un’integrazione con gli studi iniziali per generare un punto di partenza per le successive analisi.

Il primo approccio è stato aprire ogni mese come un semplice DataFrame, operazione che ha però portato alla saturazione del disco e crash.

Per ovviare a questo problema si è pensato di suddividere i grossi DataFrame in SubDataFrame più piccoli e processarli in parallelamente.

Si è usato l’attributo chunksize della funzione read\_csv() del modulo Pandas, in grado di prendere un numero di righe pari alla dimensione del parametro passato.

Si è poi processato in modo asincrono e prallelo i dati sfruttando il modulo concurrent.futures. Il modulo consente di lanciare task in parallelo in modo asincrono (usando ThreadPoolExecutor o ProcessPoolExecutor).

Per scelta implementativa si è utilizata una ProcessPoolExecutor, con numero di processi massimi uguali al numero di processori logici disponibili (nel mio caso 8).

La funzione process\_all\_data è la responsabile di quanto spiegato:



In questo modo si va a generare un corretto load balancing sulla CPU migliorando le prestazioni in modo non lineare, per la lettura dei files contenenti le informazioni degli utenti si passa da circa 1000 secondi a 390, e senza rischiare la saturazione del disco possia.

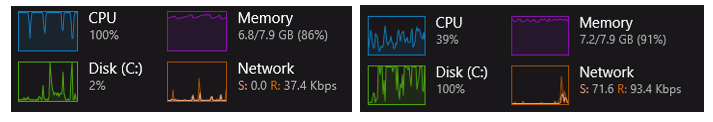


Figura 9: Differenza tra esecuzione in parallelo (a sinistra) e seriale (a destra)

## 3.3 Score con Botometer

Una volta integrato i dati nuovi con gli studi iniziali è iniziato il processo di collezionamento degli score degli utenti di Botometer.

Il criterio di scelta degli utenti “to be scored” è stato datto in funzione dell’attiività degli utenti e della tipologia di attività compiuta da tali, in particolari è andato a dividere il DataFrame in cinque diverse tipologie di attività:

* Tweet originali
* Retweet effettuati
* Retweet ricevuti
* Risposte effettuate
* Risposte ricevute

Una volta ottenute queste informazioni sono stati filtrati i 20.000 utenti più attivi per tipologia di attività, raggiungendo un numero di utenti “to be scored” di 100.000 unità.

Dopo aver definito il criterio di scelta di score è stato lanciato lo script responsabile della raccolta di tali informazioni.

La risposta di Botometer indica, secondo una serie di campi, delle informazioni sia sulla possibilità che l’utente sia o meno un bot, sia sulla tipologia di bot che rappresenterebbe.

Questo dato è piuttosto lungo e ricco di informazioni, per capirne meglio la struttura è necessario visionare un esempio:

Immagine che contiene testo

Descrizione generata automaticamente

Figura 10: Risposta di Botometer API alla richiesta di score per un utente

Questo oggetto viene trattato come dizionario in Python e contenente categorie e sottocategorie con i realivi score, suddivisi come segue:

**Categorie**:

* cap: probabilità condizionale che un utente con un punteggio maggiore o uguale a questo sia automatizzato (sia nella lingua dell'utente che in inglese)
* display scores: come raw score ma con un range che va da [0,5]
* raw scores: bot score con range che va da [0,1], sia in inglese che nella lingua dell'utente, contiene sottocategorie
* user: Utente Twitter con informazioni sull'id, nome e lingua dedotta dalla maggioranza dei tweet

**Sottocategorie**:

* fake\_follower: bot acquistati per aumentare il numero di follower
* self\_declared: bot da botwiki.org
* astroturf: bot politici etichettati manualmente e account coinvolti in follow trains che cancellano sistematicamente i propri contenuti
* spammer: account etichettati come spambot
* financial: bot che usando i cashtag
* other: vari altri bot ottenuti da annotazioni manuali, feedback degli utenti, ecc.

Per la categorizzazione bot/human, come illustrato nel seguente paper: <https://journals.uic.edu/ojs/index.php/fm/article/download/11431/9993> nella sezione “Data analysis”, nel capitolo “Automation”, si è andato ad individuare come umani gli utenti con CAP score inferiore al primo decile, mentre come bot quelli con CAP score superiore all’ultimo decile.

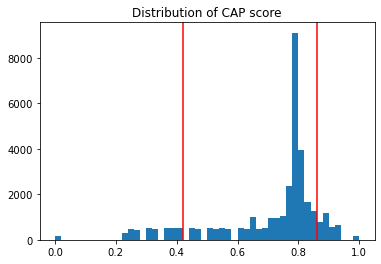


Figura 11: Distribuzione del CAP score con indicazione del primo dell'ultimo decile

Una volta effettuata la classificazione bot/human si è suddiviso per tipologia di bot in modo tale da verificare se i risultati avessero un riscontro con la tipologia di entità prevista.

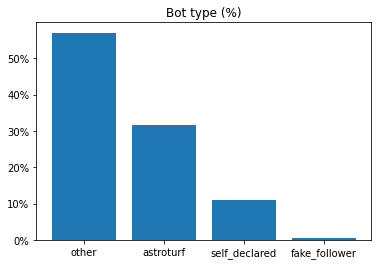


Figura 12: Tipologie di Bot ottenute

Come volevasi dimostrare non vi sono bot di tipo “financial” e “spammer”, tipologie di bot non inerenti all’argomento trattato.

Una volta effettuato questo genere di studi gli utenti con i relativi botscore sono stati riportati su un file per eventuali confronti incrociati nel corso del progetto.

## 3.3 Disinformation Dozen

In attesa dello score dei dati tramite Botometer si è passati ad un’altra atività di ricerca, dettata da un articolo molto interessante di Repubblica (fruibile al link <https://www.repubblica.it/esteri/2021/07/18/news/coronavirus_quella_sporca_dozzina_d_influencer_dietro_a_oltre_meta_delle_fake_news_sui_vaccini-310811514/>) che, riprendendo un altro articolo, introduceva i cosidetti “Disinformation Dozen”.

L’articolo in questione è un report del “Center for Countering Digital Hate”, organizzazione non-profit con sede a Londra che ha come scopo “distruggere l'architettura di odio e disinformazione online”.

L’organizzazione ha pubblicato il proprio il 24 marzo 2021 (report fruibile sul proprio sito al link <https://www.counterhate.com/disinformationdozen>) afferma che solo dodici novax sono responsabili di quasi due terzi dei contenuti anti-vaccini che circolano sulle piattaforme di social media.

Questa nuova analisi porta alla luce come come un piccolo gruppo di determinati novax sia responsabile di un'ondata di disinformazionee come gli OSNEM siano i responsabili di ciò.

I personaggi in questione sono riportati nella seguente tabella:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome | Mansione | Stato | Followers |
| Joseph Mercola | medico osteopata | Attivo | 314.302 |
| Robert Kennedy, Jr. | pseduo sostenitore ambientalista | Attivo | 311.481 |
| Ty and Charlene Bollinger | attivista della medicina alternativa | Rimosso | - |
| Sherri Tenpenny | medico osteopata | Sospeso | - |
| Rizza Islam | attivista e social media influencer novax | Rimosso | - |
| Rashid Buttar | medico osteopata | Attivo | 86.368 |
| Erin Elizabeth | attivista della medicina alternativa | Attivo | 39.384 |
| Sayer Ji | attivista della medicina alternativa | Attivo | 11.285 |
| Kelly Brogan | attivista della medicina alternativa | Attivo | 18.618 |
| Christiane Northrup | ostetrica e ginecologia | Attivo | 115.215 |
| Ben Tapper | chiropratico | Sospeso | - |
| Kevin Jenkins | CEO di un gruppo novax | Attivo | 897 |

(Aggiornato a settembre 2021, stato e followers sono in riferimento a Twitter)

Salta subito all’occhio come nessuno degli utenti in questione abbia una competenza scientifica in merito all’argomento trattato.

La prima attività svolta con queste preziose informazioni è stata una ricerca sull’effettiva presenza di questi “Disinformation Dozen” all’interno del nostro DataFrame di lavoro.

L’esito positivo di questa operazione ha rivelato poche centinaia di tweets

Si è quindi passati tramite la Twitter API per poter raccogliere tutti i tweets relativi a questi personaggi, nello specifico le attività in ingresso e in uscita da questi profili (quindi anche i retweet ricevuti, quando sono citati e le risposte).

Una volta ottenute le informazionni, un primo processamento sui dati ha mostrato subito dati interessanti da analizzare, si nota come le strategie di attività dei “Disinformation Dozen” siano differenti con l’overall delle attività ottenute nella rete globale.

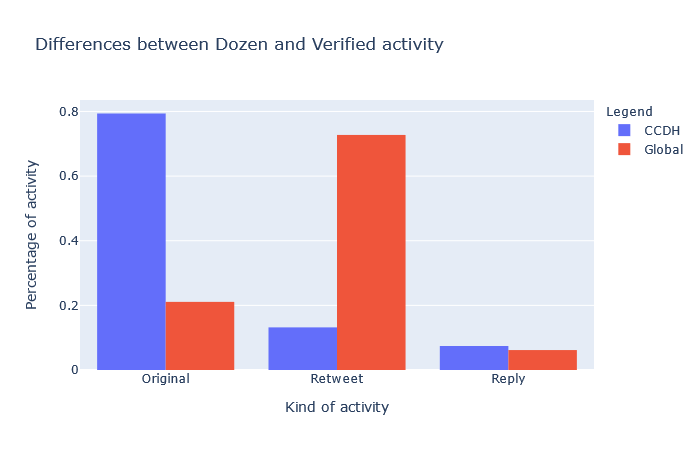


Figura 13: Confronto tra le attività degli utenti e i "Disinformation Dozen"

Una prima idea ipotizzata di questa differenza nelle attività è che i “Disinformation Dozen” hanno come finalità la produzione di fake news o notizie di dubbia veridicità, fungendo come fonte di informazione per chi interagisce con loro, in contrapposizione molti utenti tendono a retwettare un contenuto piuttosto che produrne uno.

Per capire la portata della disinformazione creata da questi 12 utenti basti pensare che coinvolgono olte 150.000 utenti diversi in termini di retweet e risposte, ricevendo oltre 2.000.000 di retweet nel periodo che va da gennaio 2020 a luglio 2021.

Il più popolare in assoluto è Robert Francis Kennedy Junior, nonché nipote di John Fitzgerald Kennedy, presidente americano ucciso nel 1963.

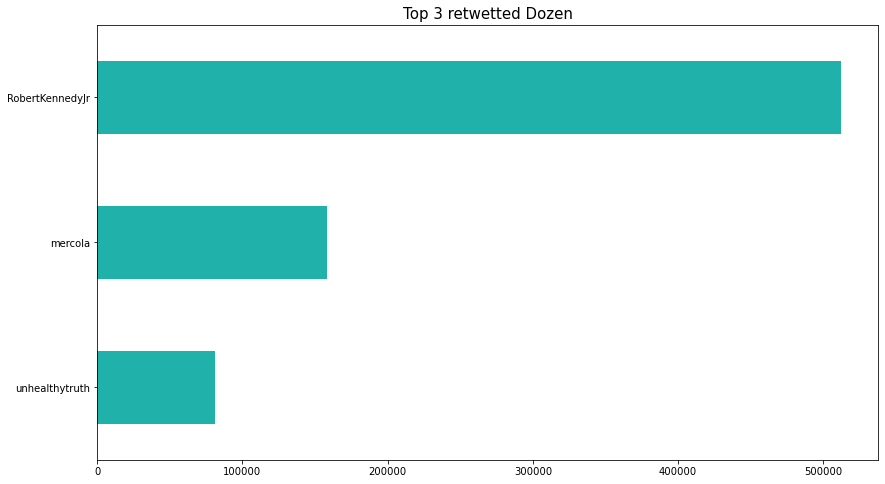


Figura 14: Dozen più retweettati nel DataFrame in utilizzo

Come è possibile notare in figura ottiene da solo oltre 500.000 di retweet, il suo legame con l’ex presidente, la sua attività politica e l’attivismo in campo ambientale e umanitario (presidente e fondatore di un’organizzazione non governativa, la “childrenshealthdefense”)

lo rendono agli occhi di molti una fonte attendibile da cui ottenere informazioni e giustificano la sua popolarità nel contesto attuale.

## 3.4 Disinformation Dozen e utenti verificati

Dal momento che il confronto dei “Disinformation Dozen” con dei semplici utenti qualsiasi, che potrebbero essere giornali, come medici o addirittura i Dozen stessi non rappresenta un metro di paragone valido si è scelto di effettuare una scrematura degli utenti con cui effettuare un paragone.

La prima scelta è stata effettuare un’operazione di filtering sul DataFrame generale estraendo tutti i tweets degli utenti verificati che rappresentano entità “autentiche, notorie e attive” (informazioni sul sito: <https://help.twitter.com/it/managing-your-account/about-twitter-verified-accounts>), che spiegato meglio significa:

* **Autentico**: verificato tramite documento, sito ufficiale o indirizzo email ufficiale
* **Notorio**: l’account deve rappresentare un individuo o un brand illustre oppure esservi in qualche modo associato
* **Attivo**: l’account deve essere attivo e vantare una storia di rispetto delle Regole di Twitter

Per ottenere i tweets degli utenti attivi si è filtrato il DataFrame degli utenti secondo la colonna “verified” e, una volta ottenuti gli id e il nome utente di questi, sono stati ottenuti i tweets di questi utenti dal DataFrame dei tweets.

A questo punto è stata effettuata un’attività di calcolo della credibilità delle fonti condivise dai singoli utenti di entrambi i DataFrame per farne un paragone.

La funzione che si occupa di questa classificazione è la seguente:



Il risultato è una lista di tuple che contengono il nome dell’utente e un valore di “high credibility shared” che va da 0 a 1, questo rappresenta per il singolo utente la percentuale di credibilità delle fonti condivise.

Per capire meglio questo valore, lo 0 significa che tutte le fonti condivise hanno credibilità bassa, 0.5 significa che l’utente condivide una metà di fonti con alta credibilità e l’altra metà di bassa credibilità (quindi al 50% sono “high credibility”) e 1 significa che le fonti condivise sono tutte fonti credibili (quindi al 100% sono “high credibility”).

Il risultato ottenuto è stato molto interessante, risultato riportato nel seguente grafico:

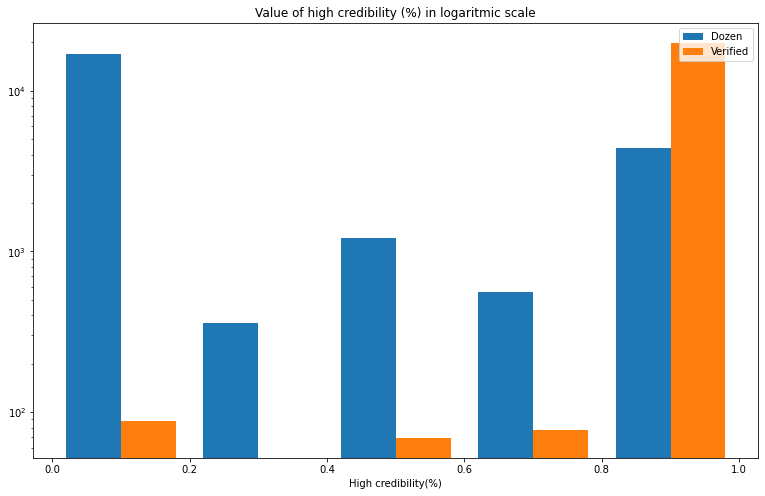


Figura 15: Confronto di credibility score su scala logaritmica tra Dozen e Verified

L’analisi condotta ha dato buoni risultati, una volta però effettuato uno studio sugli utenti più attivi che componevano questo DataFrame di utenti “Verified” è emerso come la maggior parte di questi utenti fossero teste giornalistiche o fonti di informazione, quindi non per forza utenti che portano buona informazione.

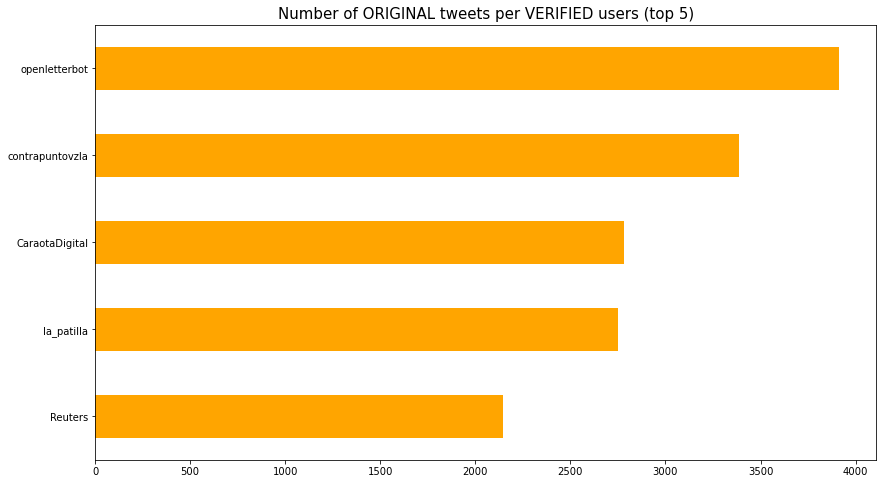


Figura 16: 10 utenti verificati che effettuano più tweets

Come anticipato salta immediatamente all’occhio come gli utenti verificati non siano un buon punto di partenza per effettuare un confronto sensato ma, anzi, hanno proprio come utente più attivo un bot.

Questi risultati, affiancati al fatto che anche i Dozen facciano parte degli utenti verificati, hanno fatto cambiare l’idea della ricerca, spostando l’attenzione su un'altra tipologia di paragone su una dozzina di utenti in contrapposizione con i Dozen.

## 3.4 Good e Bad Dozen

L’esigenza è stata quindi quella di trovare una dozzina di utenti responsabili di “good information” da contrapporre ai Dozen già trovati, per portare avanti questa ricerca si è iniziato a fare una ricerca manuale sugli utenti verificati più attivi, andando a classificare manualmente ogni utente per capire il tipo entità che rappresentasse.

Dopo aver trovato pochi utenti che potessero corrispondere alla figura di interesse e aver appurato che la ricerca fosse troppo onerosa in termini di tempo, si è scelto di verificare se ci fossero studi già esistenti per trovare questi utenti.

Fortunatamente è stata trovata una pubblicazione di Jonathan Oppenheim (Professore di Teoria Quantistica e ricercatore della Royal Society,University College London) e Kelly Truelove (ricercatore della University of California) riguardante i “Twitter accounts to follow on Covid-19”, i migliori account da seguire per ottenere informazioni scientificamente attendibili sul Covid-19.

La pubblicazione, facilmente raggiungibile al link <https://www.ucl.ac.uk/oppenheim/Covid-19_tweeps.shtml>, raccoglie ben 300 differenti account, scelti in base alla loro competenza e popolarita.

Prima di prendere i primi 12 utenti di questa lista composta da 300 nomi è stato prima verificato manualmente che ogni utente fosse effettivamente attivo e presente nel DataFrame in utilizzo in fino a questo momento.

Una volta verificato che i nuovi utenti fossero presenti si passati alla definizione di due nuove categorie: si è deciso di definire come “Bad Dozen” gli utenti precedentemente individuati, responsabili del 65% delle fake news, e “Good Dozen” i primi 12 utenti ottenuti dallo sudio condotto.

Per analizzare se gli utenti fossero attivi almeno quanto i “Bad Dozen” sono stati calcolati il numero di post e i retweet generati dai “Good Dozen”, ottenendo come risultato 14.348.506 di retweets generati, dimostrando la loro popolarità.

La tabella che fa riferimento ai “Good Dozen” è la seguente:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome | Mansione | Stato | Followers |
| Dena Grayson | medico, ricercatore ed esperta di Ebola | Attivo | 326.532 |
| Ian M. Mackay | virologo | Attivo | 120.086 |
| Eric Feigl-Ding | epidemiologo ed economista sanitario | Attivo | 588.577 |
| Ilona Kickbusch | fondatrice del Global Health Centre | Attivo | 30.435 |
| Ashish Jha | medico e ricercatore | Attivo | 264.490 |
| Helen Branswell | reporter di malattie infettive e salute globale | Attivo | 216.153 |
| Marc Lipsitch | epidemiologo e microbiologo di malattie infettive | Attivo | 234.412 |
| Trevor Bedford | ricercatore in virus, evoluzione e immunità | Attivo | 328.185 |
| Kai Kupferschmidt | giornalista scientifico e biologo molecolare | Attivo | 134.596 |
| Ed Yong | scrittore scientifico per The Atlantic | Attivo | 354.278 |
| Rochelle Walensky | direttrice del CDC del governo degli USA | Attivo | 332.234 |
| Tedros Ghebreyesus | direttore generale del WHO | Attivo | 1.530.289 |

(Aggiornato a settembre 2021, stato e followers sono in riferimento a Twitter)

Andando a confrontare le menzioni tra “Good” e “Bad” Dozen appare subito lampante la differenza di competenze abissale tra gli attori dell’informazione in questione.

# 4 Piani di lavoro

## 4.1 GitHub

# 5 Conclusioni

## 5.1 Problemi riscontrati

## 5.2 Implementazioni future

# 6 Fonti

<https://botometer.osome.iu.edu/api>

<https://pypi.org/project/botometer/#description>

https://github.com/IUNetSci/botometer-python

<https://cnets.indiana.edu/blog/2020/09/01/botometer-v4/>

<https://blog.quantinsti.com/detecting-bots-twitter-botometer/>

<https://rapidapi.com/developer/dashboard>

<https://networkx.org/documentation/stable/index.html>

<https://help.twitter.com/it/safety-and-security/public-and-protected-tweets>

<https://link.springer.com/article/10.1007/s42001-021-00139-3>

<https://www.counterhate.com/disinformationdozen>

<https://www.repubblica.it/esteri/2021/07/18/news/coronavirus_quella_sporca_dozzina_d_influencer_dietro_a_oltre_meta_delle_fake_news_sui_vaccini-310811514/>

<https://github.com/echen102/COVID-19-TweetIDs>

https://help.twitter.com/it/managing-your-account/about-twitter-verified-accounts