
Campionato mondiale di calcio 2018

-La Finale-

di Beatrice Fumagalli - matricola 784549 e Pierluigi Tagliabue - matricola 835211

Abstract

La finale del campionato mondiale di calcio 2018, in Russia, ha visto come protagoniste le due squadre avversarie di Francia e Croazia.

Nel presente elaborato si è cercato di analizzare il rapporto che intercorre tra la notorietà di un giocatore e l'effettiva bravura sul campo. In quanto alla prima sono stati raccolti i *tweets* italiani in cui erano presenti i profili ufficiali dei giocatori. Per lo *streaming* e *storing* dei *tweets* sono stati utilizzati rispettivamente Kafka, presente nella Virtual Machine Azure (Microsoft) e MongoDB¹, database NoSQL orientato ai documenti, basato sul formato JSON. Successivamente il database è stato elaborato in Python ed integrato con due ulteriori dataset in formato .csv, riguardanti le prestazioni di ogni singolo giocatore.

1 Introduzione

Il campionato mondiale di calcio² è il massimo torneo calcistico per squadre nazionali maschili. Nato nel 1930, si disputa ogni quattro anni e prevede una fase di qualificazione che si svolge nei tre anni precedenti per determinare le squadre che si qualificheranno per la fase finale. L'evento è a cura della FIFA, la federazione internazionale di calcio, che premia il vincitore con il famosissimo trofeo della Coppa del Mondo. Alla fase finale partecipano 32 squadre, incluse le nazioni ospitanti che si qualificano automaticamente.

La FIFA World Cup 2018 è stata la ventunesima edizione del campionato mondiale di calcio ed è stata la prima coppa del mondo che si è disputata nell'Europa orientale, infatti la fase finale si è svolta in Russia.

La finale, svoltasi il 15 luglio nello Stadio *Lužniki* di Mosca, ha visto come protagoniste le nazionali di Francia e Croazia e la sfida si è conclusa con la vittoria della Francia per 4-2 che ha così conquistato la coppa del mondo per la seconda volta nella sua storia. Il terzo posto è andato al Belgio.

La Coppa del Mondo è sicuramente il torneo di calcio più prestigioso in assoluto e, infatti, è anche l'evento sportivo più seguito sulle televisioni dell'intero pianeta. Per la finale Francia-Croazia, solamente Canale 5, in Italia, ha registrato 11 milioni 688mila spettatori e il 66,56% di share.³

Ma il grande successo non si ferma alla televisione, infatti prosegue sui Social Media, dove i due *hashtags* della finale, #FRACRO e #WorldCupFinal, hanno totalizzato 110 milioni di interazioni su Instagram e 356 mila di tweets su Twitter⁴.

2 Obiettivo

L'obiettivo della nostra ricerca è quello di mettere in evidenza la relazione che intercorre tra la notorietà di un giocatore e l'effettiva bravura sul campo, sia durante la partita finale dei Mondiali, che durante l'intero anno calcistico. Ci siamo infatti chiesti se i giocatori delle due squadre finaliste più twittati attraverso il loro profilo ufficiale Twitter fossero anche i giocatori più forti nelle statistiche FIFA e se fossero stati i migliori in campo durante la Finale, ossia se fossero i giocatori con il punteggio complessivo postpartita maggiore.

3 Twitter

Per svolgere la nostra analisi siamo ricorsi all'utilizzo del *social network* Twitter che mette a disposizione delle API per consentire agli utenti lo *streaming*.

Ogni giocatore possiede un proprio profilo ufficiale, fatta eccezione per i seguenti giocatori croati: Tin Jedvaj, Nikola Kalinić, Duje Ćaleta-Car e Lovre Kalinić. Grazie all'utilizzo di Kafka, abbiamo effettuato lo *streaming* in *real time* dei *tweets* prima, durante e dopo la finale dei Mondiali. Successivamente è stato eseguito lo *storing* di tutti i dati con MongoDB al fine di poter eseguire i passaggi successivi per l'analisi dei medesimi.

4 Architettura Kafka e MongoDB

Per lo *streaming* abbiamo utilizzato la piattaforma Apache Kafka, dell'ecosistema *Hadoop*, un sistema *open source* di messaggistica istantanea, che consente la gestione di un elevato numero di operazioni in tempo reale da migliaia di *client* ed eseguibile su un qualsiasi sistema dotato di JVM.

L'architettura di questa piattaforma si serve di due API principali:

- **Producer API**, che permette alle applicazioni di pubblicare flussi, nel nostro caso i *tweets*, in uno o più *topic* creati dall'utente.
- **Consumer API**, che consente l'accesso ai dati immagazzinati nel *topic*.

Per accedere alle API Twitter, è stata utilizzata *Tweepy*, una libreria Python che supporta l'autenticazione *OAuth* e ne consente il collegamento grazie a quattro *token* personali (*consumer key*, *consumer secret*, *access token* e *access secret*). Quest'ultimi sono messi a disposizione dell'utente direttamente da Twitter dopo la creazione di una nuova applicazione e

¹ <https://www.mongodb.com/>

² https://it.wikipedia.org/wiki/Campionato_mondiale_di_calcio

³ <http://www.today.it/media/tv/mondiali-2018-ascolti-tv-finale-francia-croazia.html>

⁴ <https://twitter.com/>

vengono utilizzati attraverso la creazione di un'istanza *OAuthHandler* nel *Producer*.

Successivamente abbiamo utilizzato MongoDB per immagazzinare i tweets, uno ad uno, in modo da avere pieno controllo sul salvataggio dello *streaming* durato cinque giorni.

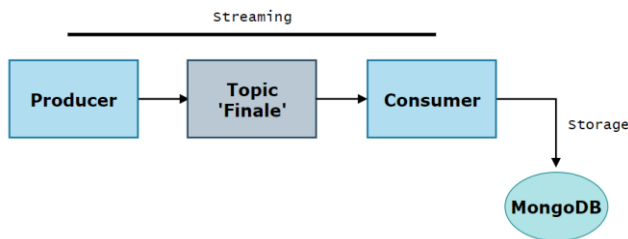


Fig. 1. Architettura implementata

5 Streaming

Abbiamo pensato ad un'architettura che ci permettesse di effettuare lo *streaming* dei tweets aventi come *keywords* i profili ufficiali dei giocatori delle squadre finaliste e che salvasse i tweets per le analisi future.

Per effettuare lo *streaming* dei tweets abbiamo deciso di ricorrere al metodo *'on status'*, offerto dalla libreria *Tweepy*, definito all'interno della classe *'listener'* ereditata da *StreamListener*.

Nel metodo scelto abbiamo poi selezionato gli attributi che abbiamo ritenuto essere più utili ai fini del progetto e li abbiamo inseriti in una lista. I tweets vengono scaricati in formato JSON e poi convertiti in stringa attraverso il comando *'json.dumps'* per poter essere inviati dal *Producer* al *topic 'Finale'*, da noi precedentemente creato con il seguente comando:

```
[root@sandbox bin]# ./kafka-topics.sh --create
--zookeeper sandbox.hortonworks.com:2181 --repl
lication-factor 1 --partitions 1 --topic Finale
```

Fig. 2. Creazione del *topic 'Finale'*

Infine, il *Consumer* ha il compito di ascoltare i messaggi in arrivo dal *topic* e, all'avvenuta ricezione di un messaggio, lo recupera e converte in formato JSON con il comando *'json.load'*. Questo nuovo messaggio viene poi aggiunto come nuovo *record* nella *collection 'Finale'* di MongoDB.

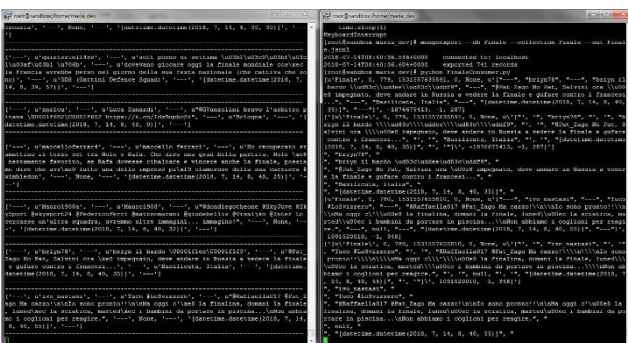


Fig. 3. *Streaming Producer (dx)* e *Consumer (sx)*

6 File JSON

Una volta ultimato il processo di *streaming* dei tweets abbiamo ottenuto un database MongoDB di 13015 righe. La struttura di ogni record è rappresentata in figura 4 e comprende:

- *"_id"* : identificativo univoco generato in maniera automatica da MongoDB
- *"username"* : nome dell'utente Twitter, autore del tweet

- *"author"* : *nickname* univoco dell'utente Twitter
- *"text"* : testo del tweet
- *"location"* : luogo indicato dall'utente come residenza
- *"data"* : data e orario di pubblicazione del tweet

```
{
  "_id": {"$oid": "5b4bf33ae5cde5fd6404bb78"},
  "username": "\", \"Mpadia Miafouna\", \"",
  "author": "\", \"Mpadia_miafouna\", \"",
  "text": "\", \"@Modric_FR @lukamodric10 Bravo maestro\", \"",
  "location": "\", \"Brazzaville, Congo\", \"",
  "data": "\", \"[datetime.datetime(2018, 7, 16, 1, 22, 1)]\", \""
}
```

Fig. 4. *Record* salvato in MongoDB

7 Manipolazione tweets ed integrazione dataset

Per analizzare i dati memorizzati nel database MongoDB abbiamo deciso di esportare la *collection* tramite il comando *'mongoexport'*:

```
[root@sandbox maria_dev]# mongoexport --db Finale
--collection Finale --out Finale14.json
```

Fig. 5. *Export* della *collection 'Finale'*

ottenendo così il file desiderato.

Una volta portato in locale, abbiamo manipolato il file in Python.

Innanzitutto, abbiamo creato un Data frame mantenendo tutti gli attributi precedentemente scelti e poi abbiamo apportato le operazioni necessarie di *data cleaning*. È importante specificare la modifica apportata all'orario di pubblicazione del tweet. Infatti, di *default* viene considerata l'ora solare con fuso orario inglese, pertanto abbiamo aggiunto due ore per uniformarlo con l'ora legale italiana. Inoltre, è corretto segnalare un nostro errore, poi prontamente corretto, riguardante l'inserimento delle *keywords* da ricercare all'interno dei vari tweets.

La parola *"Finale"* inserita per sbaglio nel filtro dello *streaming*, in concomitanza con la finale di tennis maschile di Wimbledon, ha portato allo scaricamento di tweets non voluti. L'errore è stato corretto dopo solo un'ora dall'inizio dello *streaming*, eliminando la *keyword* dalla lista delle parole ricercate. Pertanto, abbiamo in seguito provveduto all'eliminazione dei 124 tweets non inerenti alla partita tramite codice Python, ottenendo un dataset contenente effettivamente 12891 tweets. Successivamente abbiamo proseguito con l'integrazione di due dataset al fine di reperire le informazioni utili alla nostra analisi.

Il primo dataset *'players_convocati.csv'* è un dataset in formato *.csv* contenente tutti i giocatori di tutte le nazionali convocate ai Mondiali 2018. Il secondo dataset *'CompleteDataset.csv'* contiene invece tutte le statistiche calcistiche di tutti i giocatori per cui FIFA dispone della licenza.

Il primo dataset è stato utilizzato per creare una libreria con i nomi dei giocatori convocati di Francia e Croazia. Questa libreria ci è servita per creare un nuovo dataset che, in corrispondenza del nome di ciascun giocatore, contiene il corrispettivo *nickname* Twitter con il numero di volte in cui quest'ultimo è stato twittato (colonna *'Count_tweet'*). Al dataset è stato inoltre aggiunto, con un ciclo *for*, il punteggio post-partita assegnato a ciascun giocatore da Sport Mediaset⁵ (colonna *'Post_Match_Ratings'*).

Solamente a questo punto è stato possibile integrare le statistiche calcistiche relative ai giocatori contenute nel secondo dataset, utilizzando come chiave primaria il nome del giocatore. In particolare, abbiamo integrato il punteggio complessivo riguardo ciascun calciatore nella colonna *'Overall'*. Dal momento che il dataset di FIFA conteneva anche statistiche ritenute inutili per il nostro progetto, abbiamo deciso,

⁵ http://www.sportmediaset.mediaset.it/mondiali-russia-2018/mondiali-le-pagelle-della-finale-francia-croazia_1225651-201802a.shtml

Entrambe comprendono i medesimi attributi, ossia:

- **Name** : iniziale del nome del giocatore puntata e cognome
- **Nickname** : nickname Twitter del profilo ufficiale del giocatore
- **Club** : squadra di provenienza del giocatore
- **Preferred Position** : ruolo del giocatore
- **Overall** : punteggio FIFA totale del giocatore
- **Count_tweet** : numero di volte in cui è stato twittato il giocatore
- **Post_Match_Ratings**: pagella giocatore postpartita
- **Velocity**: velocità del giocatore
- **Attack** : capacità in fase offensiva del giocatore
- **Defense** : capacità del giocatore in fase difensiva
- **Passing** : precisione e tempestività nei passaggi
- **Movement** : agilità nei movimenti
- **Physicality** : prestanza fisica

8 Analisi preliminari

8.1 Flusso Streaming

Abbiamo deciso di scaricare i tweets per cinque giorni, al fine di fornire un'analisi preliminare circa la variazione dei profili dei giocatori più twittati nel prepartita, durante la partita e nel postpartita.

Il prepartita comprende i giorni 13 (a partire dalle ore 23), 14 e 15 fino le ore 17, orario di inizio della partita. La partita si è svolta il giorno 15, dalle ore 17 alle ore 18:32. Il postpartita comprende tutti i tweets scaricati dalle 18:32 del giorno 15 fino al giorno 17. Nonostante svariati problemi di connessione alla Virtual Machine e la velocità ridotta del *Consumer* abbiamo raggiunto il seguente numero di tweets scaricati:

- 1246 nel prepartita
- 3418 durante la partita
- 8227 nel postpartita

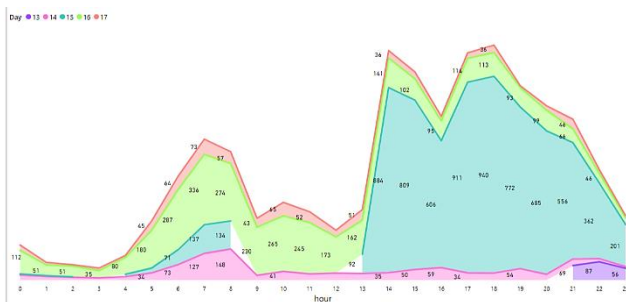


Fig. 6. Distribuzione del numero di tweets scaricati durante le cinque giornate di *streaming*

Come si evince dallo *Stacked area chart*⁷, se si prendono in considerazione i singoli giorni di *streaming* e le ore, il numero maggiore di tweets è stato raccolto durante la partita.

8.2 Word Cloud⁸

I giocatori più twittati durante il prepartita sono Modrić, Mandžukić e Perišić della Croazia e solamente Paul Pogba della Francia.



Fig. 7. Word Cloud Prepartita

Si raggiunge invece una situazione quasi di parità durante la partita, in cui i giocatori più twittati diventano Perišić e Mandžukić per la Croazia e Griezmann, Pogba e Mbappè per la Francia, ossia i cinque giocatori che hanno segnato durante la finale di Campionato.



Fig. 8. Word Cloud Partita

La partita si conclude con la vittoria della Francia per 4-2 ed i giocatori francesi sovrastano quelli croati, sia in campo, che nella Word Cloud postpartita.



Fig. 9. Word Cloud Postpartita

⁶ <https://sofifa.com/>

⁷ Stacked area chart, Power BI

⁸ Word Cloud, Power BI

9 Analisi e risultati

Lo scopo del nostro progetto è quello di evidenziare la relazione esistente tra la notorietà di un giocatore e l'effettiva bravura sul campo.

Prima di analizzare i nostri dati è importante fare un sunto di quello che è stato il mondiale per queste due nazioni.

La Croazia dopo aver largamente dominato il suo girone, ha poi progressivamente mostrato un leggero calo di prestazioni nelle sfide successive ricorrendo per i tre turni successivi ai supplementari o ai rigori. Questo non le ha comunque impedito di arrivare alla prima finale mondiale della sua storia, prima di questo mondiale il suo miglior piazzamento era stato un terzo posto conquistato nel 1998.

La Francia, a differenza della Croazia, ha avuto un cammino più regolare lungo tutta la competizione qualificandosi al primo posto nel girone e non dovendo mai ricorrere ai tempi supplementari nei turni successivi.

9.1 Croazia⁹

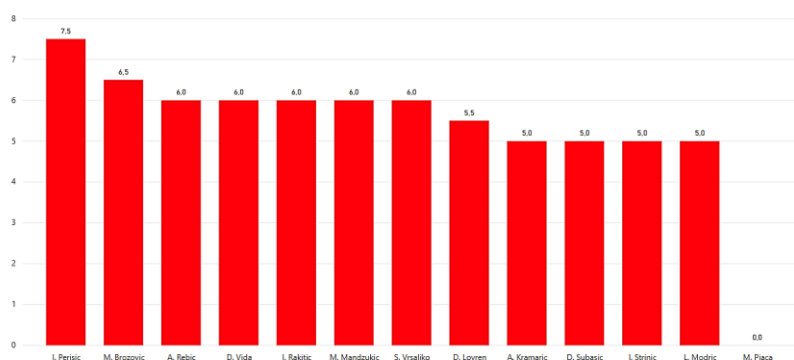


Fig.10. Punteggio postpartita

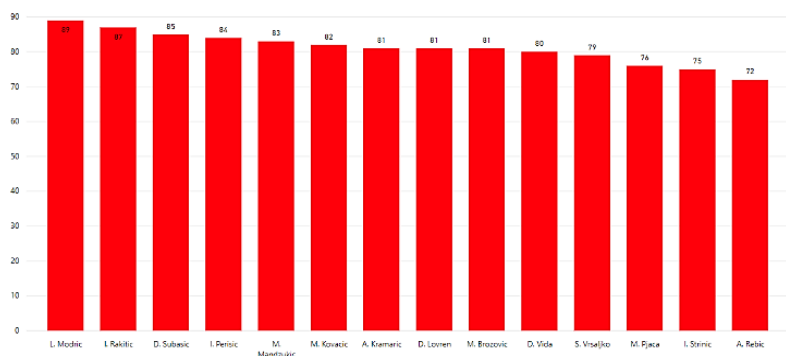


Fig. 11. Punteggio FIFA

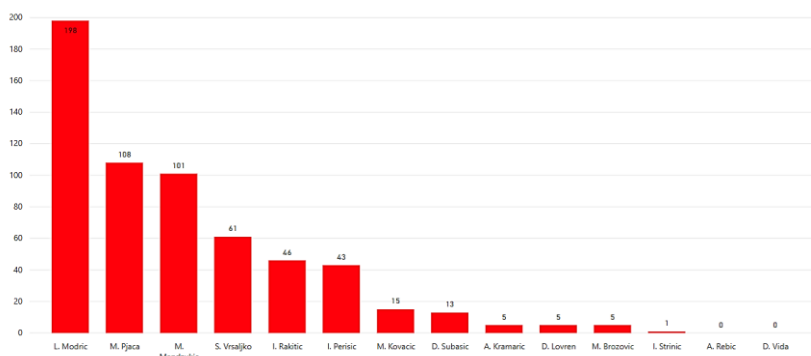


Fig. 12. Count tweets

⁹ Bar charts, Power BI

Per quanto riguarda la Croazia si può notare dai *Bar charts* che i giocatori più twittati, in figura 12, sono giocatori ritenuti estremamente forti, con un punteggio FIFA molto alto. Infatti, troviamo al primo posto sia come giocatore più twittato (198 tweets), che come giocatore più forte (punteggio FIFA: 89) il centrocampista Luka Modrić. Questa considerazione vale anche per Mario Mandžukić che, nonostante il goal, viene twittato meno rispetto a Modrić, ma si posiziona comunque terzo con 101 tweets. Mano a mano che i tweets decrescono, anche il punteggio FIFA circa le abilità del giocatore sul campo cala; ad esempio Rëbic, posizionato all'ultimo posto all'interno della Croazia, con un punteggio FIFA pari a 72, non è mai stato twittato.

A seguire, siamo andati ad analizzare le pagelle postpartita di diversi siti come il *Corriere*¹⁰ ed *Eurosport*¹¹, scegliendo poi di utilizzare i punteggi di *Sport Mediaset*¹², poiché, secondo noi, più equilibrati. Perišić viene valutato come miglior giocatore della Croazia con uno *score* di 7.5. Al contrario, la pagella postpartita del capitano della Croazia Luka **Modrić**, eletto miglior giocatore della competizione (vincitore del Pallone d'oro dei Mondiali), evidenzia come non sia riuscito a brillare abbastanza in finale, con un voto pari a 5.0. È da segnalare anche come il suo nome compaia sia nella Word Cloud inerente al prepartita (in figura 7), segno che era uno dei giocatori che aveva creato le aspettative più alte tra il pubblico italiano, ma che non compare nella Word Cloud del *match* (in figura 8). Ciò indica che sul campo non ha compiuto imprese degne di nota, rimanendo però un giocatore amato dal pubblico. Infatti, è l'unico giocatore croato nella Word Cloud postpartita (in figura 9). I tweets riguardanti Modrić risultano pertanto così distribuiti:

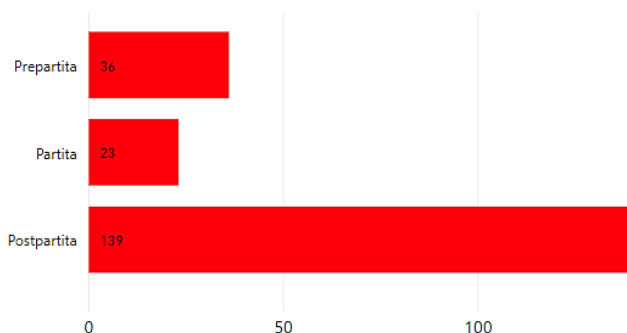


Fig. 13. Distribuzione tweets su Modrić

Seguendo l'ordine del *Bar Chart* raffigurante il conteggio dei tweets per ogni giocatore risulta che il secondo più twittato è Marko **Pjaca**, con 108 tweets, poco più della metà rispetto a quelli su Modrić. Il giovane talento Croato rappresenta una particolarità degna di un approfondimento, sia perché a livello di statistiche FIFA risulta uno dei meno prestanti con un punteggio pari a 76, sia perché è entrato in partita solo nei minuti finali, rendendo così impossibile la sua valutazione postpartita. Per capire il perché di questa specificità abbiamo esaminato i tweets legati al profilo ufficiale di Marko Pjaca. I risultati di questa analisi hanno evidenziato uno stretto collegamento tra il giocatore e la sua squadra di provenienza, ovvero la Juventus FC. Sempre collegati a quest'ultima, i testi dei tweets hanno rilevato una connessione con i compagni di squadra presenti nella finale, ovvero Mario Mandžukić e Blaise Matuidi. Infine, i restanti tweets riguardano il calcio mercato estivo che ha visto Marko Pjaca accostato alla ACF Fiorentina, da qui la stretta correlazione tra i due.

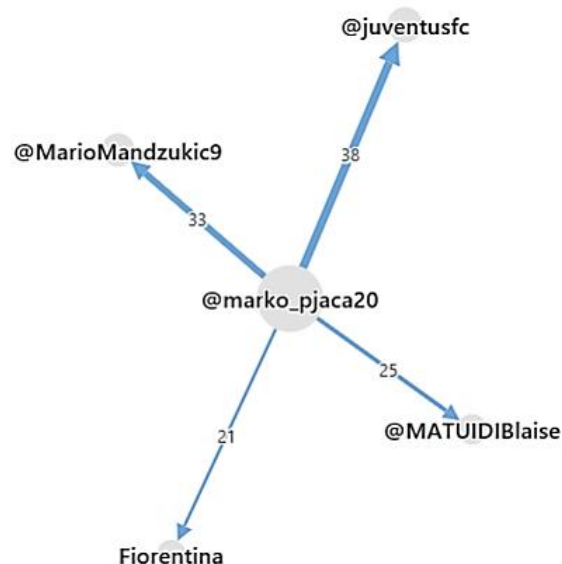


Fig. 14. GiGraph di Pjaca

Il sesto giocatore più twittato con 43 tweets è il centrocampista Ivan **Perišić** che, nonostante risulti il migliore della Croazia per l'ottima prestazione durante il *match* con un punteggio di 7.5, risulta avere un numero di tweets nella media, concentrati per lo più durante la partita, proprio per la bravura dimostratasi sul campo. Discorso analogo può essere applicato a **Brozović** che, nonostante sia stato il secondo miglior giocatore in campo con appena un punto in meno rispetto a Perišić, ha un numero estremamente basso di tweets (5). Questo potrebbe indicare la minor notorietà dei due all'interno della squadra.

Come per Pjaca anche Mateo **Kovačić** rappresenta un caso particolare, infatti pur non avendo giocato ha comunque avuto un buon riscontro per quanto riguarda il numero di tweets a lui dedicati (15).

Dato il suo punteggio FIFA mediamente alto (82) era plausibile aspettarsi il suo ingresso in campo durante la partita, questa ipotesi è sotto riscontrabile in un tweet preso ad esempio tra tutti quelli a lui collegati.

Avanzava dello spazio in panchina da scaldare? A questo punto venderanno @Mateo_Kova23

Fig. 15. Tweet su Kovačić

¹⁰ <https://www.corriere.it/sport/calcio/mondiali/francia/notizie/01-sport-pagellecorriere-web-sezioni-5de92efe-885a-11e8-923a-1c26282fdcfce.shtml>

¹¹ https://it.eurosport.com/calcio/mondiali/2018/le-pagelle-di-francia-croazia-4-2-mbappe-stellare-con-pogba-e-griezmann-modric-stecca-la-finale_sto6849572/story.shtml

¹² http://www.sportmediaset.mediaset.it/mondiali-russia-2018/mondiali-le-pagelle-della-finale-francia-croazia_1225651-201802a.shtml

9.2 Francia¹³

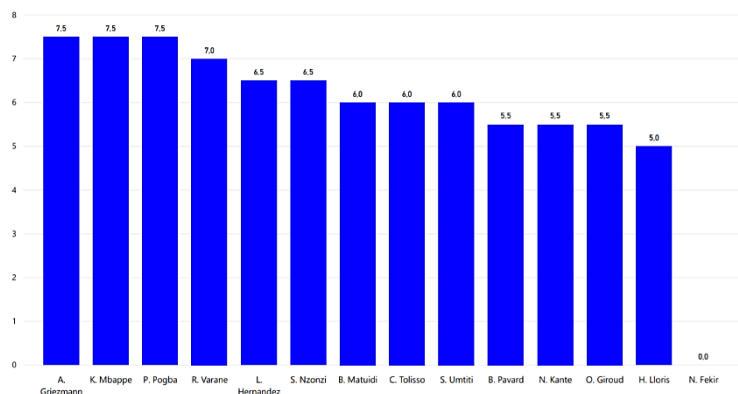


Fig.16. Punteggio postpartita

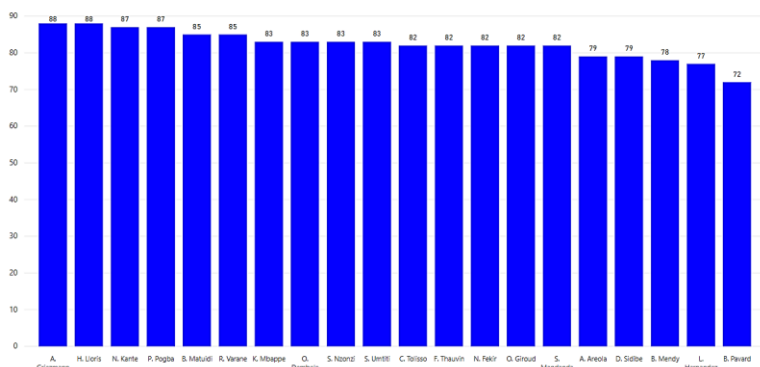


Fig. 17. Punteggio FIFA

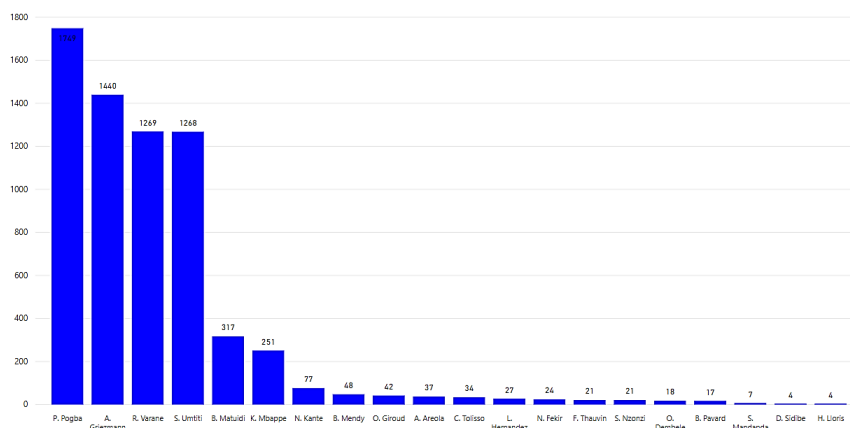


Fig. 18. Count tweets

Come ci si aspettava, tra i giocatori della Francia i più twittati (in figura 18) risultano essere **Pogba** e **Griezmann**, con rispettivamente 1749 e 1440 tweets. Si tratta infatti di giocatori molto forti, con un punteggio FIFA pari a 87 e 88 e che hanno giocato molto bene in partita, segnando una *goal* ciascuno e aggiudicandosi entrambi uno *score* del 7.5. Pogba spicca in prima posizione per numero di tweets anche perché gode di una notorietà affermata tra gli italiani grazie al suo precedente ruolo da centrocampista nella Juventus.

La prima anomalia si riscontra con **Varane** e **Umtiti**, che sono stati twittati rispettivamente 1269 e 1268 volte nonostante non abbiano né un punteggio FIFA elevato né tantomeno abbiano eseguito azioni degne di nota sul campo. Infatti, entrambi non compaiono nella Word Cloud della partita (in figura 8), ma appaiono magicamente in quella postpartita (in figura 9). Andando ad

analizzare i tweets ci siamo accorti della forte correlazione tra i due giocatori e il profilo ufficiale del Museo del *Louvre*. Infatti, il numero di tweets estremamente elevato deriva dal fatto che gli italiani, sdegnati per il post pubblicato dal Museo ritraente la Gioconda che indossa la maglia francese, hanno twittato la loro indignazione verso il post inserendo anche il tag di questi due giocatori francesi, oltre a quello di Pogba e Griezmann che però, a differenza dei primi, se si detraessero dal totale i tweets in questione, si riconfermerebbero tra i giocatori più twittati, assieme a **Matuidi** e **Mbappé** con rispettivamente 317 e 251 tweets.

Nel grafico si può notare il numero di tweets riferiti a questo episodio rispetto al *count* totale di ciascuno dei quattro calciatori.

¹³ Bar charts, Power BI

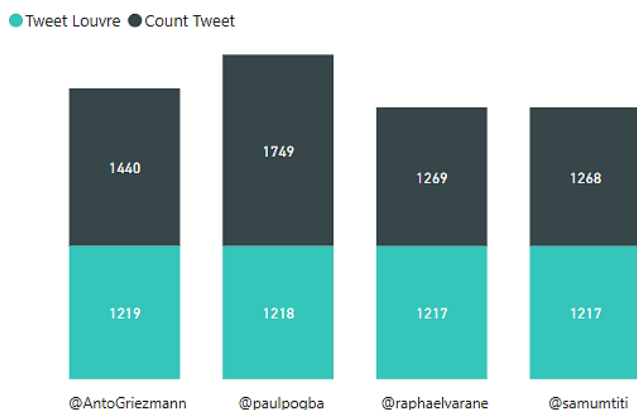


Fig. 19. Tweets relativi all'episodio del Museo del Louvre

Come per Pjaca e Kovačić della Croazia, anche per la Francia sono presenti dei giocatori che, nonostante non abbiano giocato, sono stati twittati. Questo è il caso di **Dembélé** (18 tweets), **Thauvin** (21), **Sidibé** (4), **Mendy** (48), **Mandanda** (7) e **Areola** (37).

Al contrario **Fekir**, essendo entrato in campo all'ottantunesimo minuto, non ha ricevuto il punteggio postpartita. Solamente Areola rappresenta un'anomalia giustificata, in quanto tutti i giocatori sopra citati sono stati in realtà inseriti in tweets di congratulazioni circa la vittoria della squadra francese e i tifosi ne hanno citato tutti i componenti. Analizzando i tweets di Areola tramite Word Cloud abbiamo constatato come questi ultimi si riferiscano in realtà alle ultime vicende di calciomercato riguardanti il giocatore, ovvero al suo possibile approdo alla squadra A.S. Roma.

10 Criticità e conclusioni

Le principali criticità incontrate durante l'elaborazione di questo progetto riguardano lo sviluppo del codice del Producer e del Consumer per effettuare lo *streaming* dei tweets inerenti alla finale dei Mondiali 2018. Dopo numerosi tentativi siamo però riusciti ad ottenere un codice in grado di scaricare i dati di nostro interesse e effettuare lo *storage* dei medesimi. Una ulteriore difficoltà incontrata e risolta in maniera soddisfacente riguarda la lunghezza dei tweets che dai 140 caratteri inizialmente concessi è passata a 280. Pur rappresentando un grande miglioramento per il *social network* Twitter, è importante sottolineare come le librerie e i metodi utili per effettuare lo *streaming* non siano ancora riusciti ad aggiornarsi in modo completo e uniforme a questa novità.

L'utilizzo del metodo "classico" per la cattura dei testi dei tweets portava alla creazione di stringhe interrotte qualora queste contenessero più di 140 caratteri. Per ovviare a questo problema abbiamo fatto ricorso all'utilizzo dell'attributo '*extended_tweet*' presente nel metodo "*on_status*" precedentemente descritto e da noi utilizzato. Ciò ha poi portato ad un'ulteriore complicazione, dovuta al fatto che il metodo *on_status*, non possiede l'attributo *extended_tweet* per i tweets con lunghezza inferiore ai 140 caratteri. Questo portava inevitabilmente ad errori ed interruzioni dello streaming ogni qualvolta il *Producer* si imbatteva in una di queste stringhe. Per gestire l'eccezione abbiamo utilizzato un blocco *try-except*. Ogni volta in cui il codice nel blocco del *try* genera un'eccezione dovuta alla lunghezza del *text* inferiore ai 140 caratteri, allora nel blocco dell'*except* viene eseguito il comando base '*status.text*' per gestirla.

Un'ulteriore questione affrontata riguarda le *keywords* inserite nel codice del Producer. Ci siamo domandati se fosse meglio scaricare i tweets solamente utilizzando i profili ufficiali o se fosse opportuno introdurre anche gli *hashtags*. Effettivamente come si può vedere nel caso del giocatore francese Pogba (in figura 20), utilizzare

solamente il profilo ufficiale risulta riduttivo, in quanto utilizzando anche l'*hashtag* si ha un numero maggiore di *records* scaricati. Infatti, abbiamo eseguito una prova inserendo l'*hashtag* del giocatore come *keyword* e solamente in dodici ore sono stati scaricati 90 tweets. Estendendo tale scelta a tutti i giocatori si sarebbe raggiunto un numero complessivo di tweets sicuramente più elevato rispetto a quello ottenuto, ma abbiamo deciso in linea definitiva di mantenere solamente i profili ufficiali, al fine di non falsare la veridicità dei dati, in quanto i *nicknames* di Twitter non solo rappresentano una chiave primaria, ma bensì e soprattutto una chiave univoca per l'utente e in questo caso specifico per i calciatori.

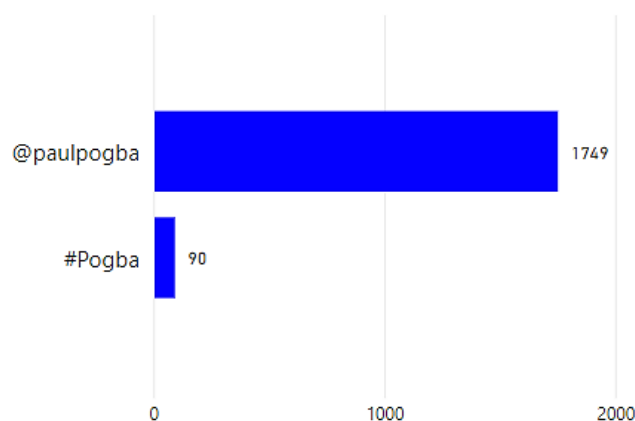


Fig. 20. Profilo ufficiale vs Hashtag

Dall'analisi effettuata emerge che i giocatori che godono di una maggiore notorietà e dunque i più twittati, sono i giocatori con un punteggio FIFA elevato e che hanno dimostrato la loro bravura anche sul campo della finale dei Mondiali, fatta eccezione per le anomalie riscontrate e precedentemente analizzate, legate per lo più ad eventi esterni al campionato e ad azioni svoltesi invece durante la partita. Una cosa certa è che il giocatore più twittato di ciascuna delle due squadre Francia e Croazia sono rispettivamente Paul Pogba e Luka Modrić, due calciatori ben noti tra tutti i tifosi italiani.

Come si può notare nel *Radar chart* sotto riportato, nonostante Modrić sembri essere poco più forte tra i due, Pogba, con i suoi 1749 tweets, risulta essere in assoluto il giocatore più twittato della finale dei Mondiali 2018.

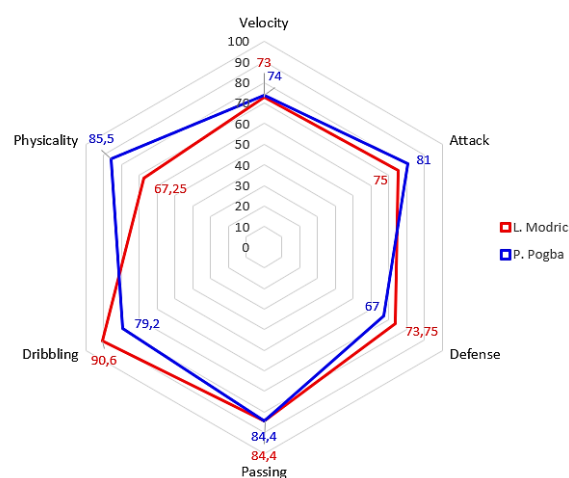


Fig. 21. Radar chart¹⁴ Pogba vs Modrić

¹⁴ Radar chart, Power BI