

# 0' Big Data è meglio e' Pelè

## - BRIGHT 2015

Tra i ventiquattro Paesi europei e le venti città italiane impegnate nella notte dei ricercatori, c'era anche Pisa. Area della ricerca del CNR inclusa. E' qui che ieri, 25 settembre, ci siamo divertiti a provare "sul campo" le tecniche di acquisizione dati dai campi di calcio. Al Kddlabor dell'ISTI-CNR siamo abituati a lavorare sui dati dei calciatori professionisti, ma la rivoluzione dei dati si allargherà anche alle serie minori. Per questo, abbiamo invitato i visitatori di Bright a giocare sul nostro mini campo, allestito sul piazzale del CNR.



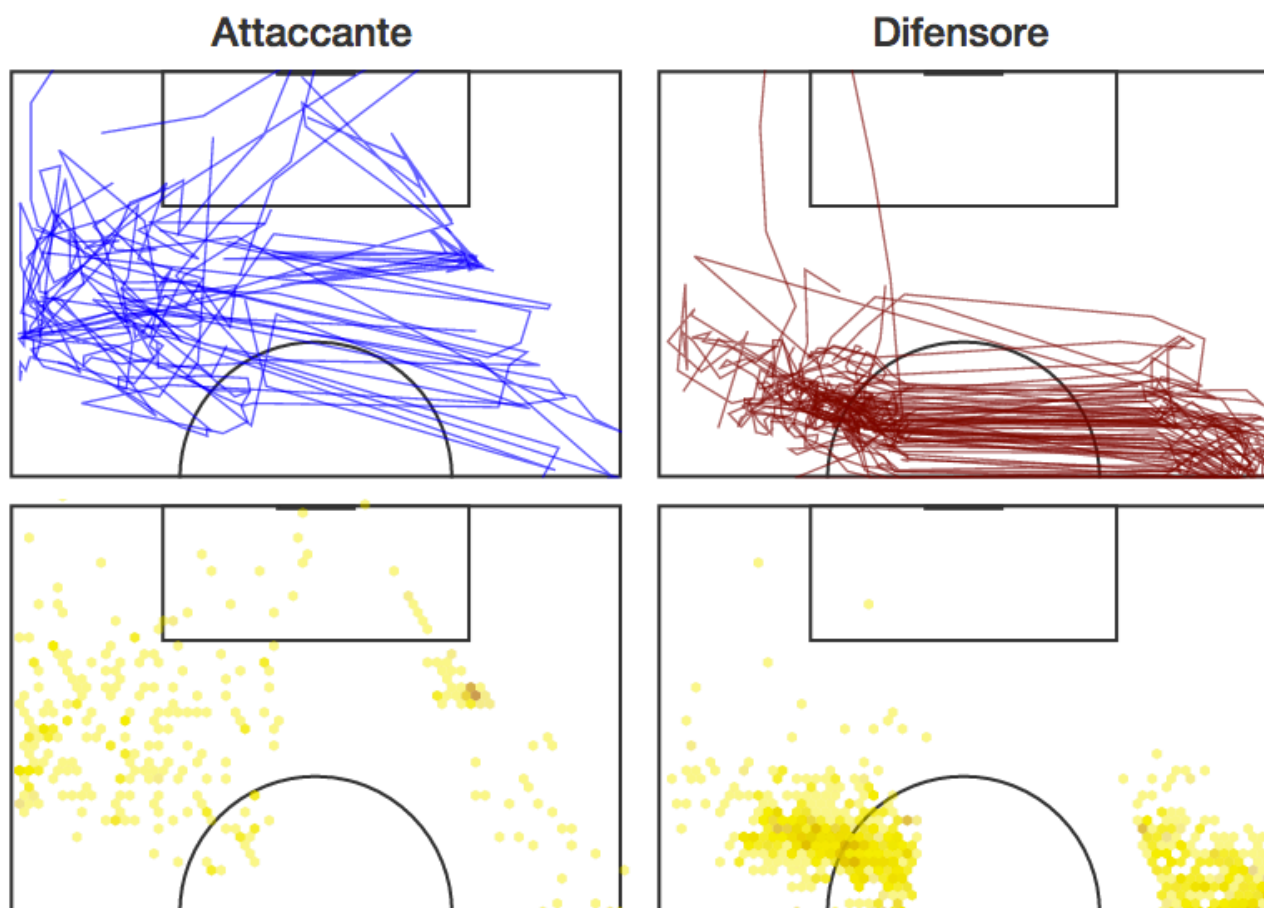
Bright 2015: tracking di giocatori in real time

Abbiamo proposto due sfide, a grandi e piccoli: segnare un gol affrontando il più arcigno dei difensori del CNR, al secolo Salvo Rinzi, e segnare da calcio piazzato, con tanto di barriera. La differenza, rispetto al calcio da strada con cui siamo cresciuti, è che le sfide uno contro uno sono state tracciate tramite videocamera e apposito software sviluppato da noi stessi, con le traiettorie e relative velocità mostrate in tempo reale. Nei calci piazzati, invece, un pallone con sensori e trasmettitore bluetooth ha registrato i dati sul colpo tirato dai partecipanti.

Dei tanti, bellissimo, stand presenti a Bright, siamo probabilmente gli unici a poter dire dove e come sono andati i nostri visitatori:

## O Big Data è meglio 'i Pelè

Bright 2015.



Bright: le traiettorie dei nostri visitatori

Dai due grafici in alto si vede la mobilità, limitata tra l'altro, del nostro Rinzivillo, comparata soprattutto con quella dei ben più giovani attaccanti che si sono succeduti nell'affrontarlo. Le velocità medie parlano chiaro: 11 km/h per gli attaccanti, 8 km/h per il difensore. Nei due plot in basso, invece, la densità di gioco in base alla zona del (mini) campo. La maggioranza di attaccanti di piede destro è evidente dal grafico relativo: la zona più densa è quella, appunto, da dove calcia in porta un destro. Il nostro

Rinzivillo, invece, ha difeso la (mini) porta con spostamenti prevalentemente laterali. Tanto per avere un confronto, un giocatore professionista arriva ad oltre 30 km/h, nei suoi scatti più intensi.

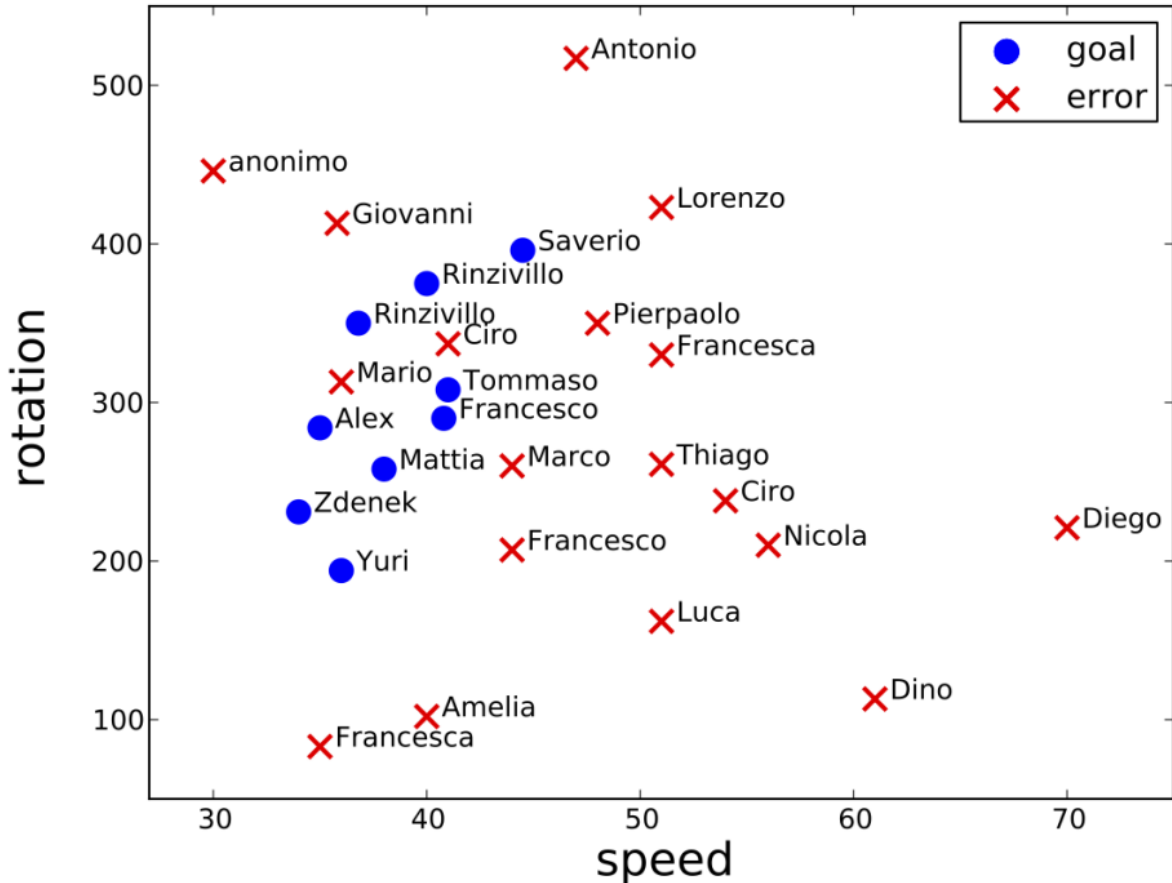


Sfida: battere la barriera dei brasiliani del CNR

La sfida sui calci piazzati è stata un po' più difficile. Se i bambini più piccoli hanno ancora poco controllo della palla, già dai 10 anni in su abbiamo potuto apprezzare dei piedini raffinati. La nostra smart ball ha registrato i dati di velocità della palla in km/h e di rotazione in giri al minuto.

La palla è stata sempre calciata dallo stesso punto, con tanto di barriera fissa. Il risultato più interessante è l'emergere di un pattern abbastanza preciso: per quanto la situazione fosse più un esempio giocattolo che un caso reale - distanza 5 metri, porta di 1,20x0.8 metri -, dal grafico finale si vede come la maggior parte dei gol siano stati segnati calciando ad una velocità tra i 37 e i 39 km/h. La velocità giusta per far abbassare la palla dietro la barriera, insomma. Altri numeri, per fortuna delle finestre del CNR, rispetto agli oltre 100 km/h di Pirlo. Dal grafico si nota anche la precisione del nostro Rinzivillo: da fine data scientist quale è, ha usato i giorni di preparazione a Bright per provare la palla bluetooth e sfruttare al meglio i dati raccolti. Altre due edizioni di Bright e sarà pronto per diventare il numero 10 del Pozzallo.

## Shot Performance



La bravura dei tiratori di Bright

Bright è stato un evento di successo, in tanti hanno visitato gli stand dei tanti laboratori presenti nell'area del CNR. Certo, l'organizzazione non è stata una passeggiata e al termine della giornata ci è sembrato di aver corso per 90 minuti di fila, ma d'altronde, come cantavano a Napoli:

*O' Big Data è meglio 'e Pelè, c'amm fatto 'o mazzo tanto pe l'avé. (\*)*

\*: Si ringrazia Luca Pappalardo per la consulenza sul lessico napoletano



In basso a destra, Diego e David (anpas Ponsacco) mentre attendono di osservare il risultato dei loro test

Post Scriptum: ai nostri esperimenti si sono prestati anche Diego e David, due calciatori dell'Anpas Ponsacco, venuti anche per ricordare il loro impegno nella partita del cuore. Partita che, l'indomani, hanno vinto. L'allenamento mirato a suon di dati è servito!

---

## Chi ha infranto la dura legge matematica del gol?

Il calcio porta con sé tanti misteri. Squadre che vincono senza quasi tirare in porta, amicizie sincere che si rovinano per un rigore negato rivisto alla moviola, città paralizzate dopo una partita vinta. Studiare il calcio, in fondo, significa cercare di svelare almeno una piccola parte di questi fenomeni al limite del paranormale. Messa così, è una sfida a cui un giallista non direbbe mai di no, soprattutto se il giallista è anche uno scienziato.

L'avvento dei Big Data del calcio è l'occasione giusta per poter spingere un po' più in là la nostra comprensione del

fenomeno calcistico. I dati sono gli indizi che servono alla nostra indagine. Fino a ieri, conoscevamo poche regole, perlopiù validate in modo empirico: il calcio è un gioco basato su episodi casuali che nel 99% dei casi sono favorevoli alla Juventus. Il mare di dati che oggi accompagna ogni partita, però, è in grado di farci capire qualcosa di più.

Prendiamo, ad esempio, la rete dei passaggi tra i giocatori di una squadra. Come per due amici in una rete sociale, due giocatori sono nodi di una rete che entrano in relazione tra di loro nel momento in cui uno passa la palla all'altro. In più, muovere la palla cercando di portarla dentro la porta avversaria è lo scopo primario del gioco del calcio. Il passaggio, inoltre, è l'evento più ricorrente durante una partita: mediamente una squadra ne effettua 400 ogni partita. Mentre il goal, al contrario, è l'evento più raro.

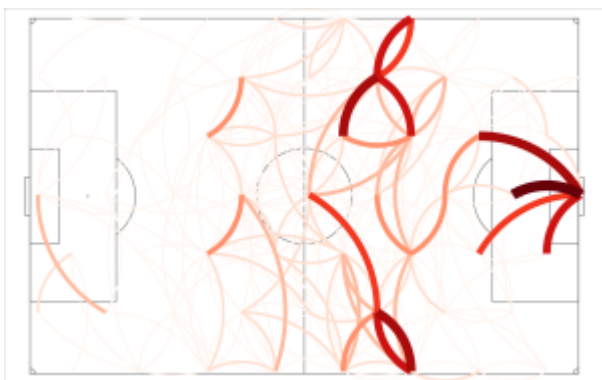


Juventus-Barcellona, giugno 2015. La rete della Juventus è incentrata su Pirlo.

Neanche Marco Malvaldi, lo scienziato giallista, ha saputo resistere quando gli abbiamo proposto di indagare il calcio. Niente morti ammazzati a Calambrone, né vecchietti ficcanaso: qui i delitti sono punti in classifica non meritati e i ficcanaso siamo noi. Gli indizi stavolta sono un po' di più della solita impronta digitale: 600mila passaggi, divisi in 148 partite di 4 campionati diversi. Per ogni partita possiamo costruire due reti, che esprimono *chi* effettua i passaggi

e dove i passaggi vengono effettuati.

Di queste due reti, analizziamo la distribuzione: la media di passaggi effettuata da ogni nodo e la relativa varianza. Possiamo esprimere due caratteristiche del gioco di una squadra: il volume di gioco e la sua imprevedibilità. Le caratteristiche spaziali sono invece analizzabili usando come nodi le zone del campo in cui avvengono i passaggi. Possono le caratteristiche di questa rete darci qualche informazione in più?



Rete dei passaggi del Barcellona durante l'ultima finale di Champions League. Il gioco si sviluppa su due zone precise, quelle di Messi e Neymar

L'ipotesi c'è, dunque. Ora servono le prove. Il fiuto di chi ha risolto decine di casi contorti sa dove andarle a cercare: analizziamo le reti di tutte le squadre dei campionati europei principali, durante una stagione intera. In pratica, simuliamo il campionato: per ogni partita calcoliamo volume e imprevedibilità delle reti di ogni squadra, li combiniamo in un'unica misura calcolandone la media armonica. Per comodità la chiamiamo  $H$ . Quindi, per ogni partita assegniamo i 3 punti alla squadra con l' $H$  più alto. Se la differenza di due squadre è ragionevolmente vicina, allora il risultato è un pareggio. Alla fine, la classifica finale secondo la misura  $H$  ha "confermato" l'ipotesi: le reti ci danno qualcosa in più. La correlazione tra la classifica simulata e quella reale è superiore all'80%. Arriva all'89% nel caso del campionato tedesco.



simulated ranking		real ranking	
Bayern	95	Bayern	90
Dortmund	75	Dortmund	71
Wolfsburg	62	Schalke	64
Leverkusen	59	Leverkusen	61
Augsburg	54	Wolfsburg	60
Hoffenheim	54	Mönchengladbach	55
Hannover	49	Mainz	53
Schalke	47	Augsburg	52
Hertha	43	Hoffenheim	44
Mönchengladbach	42	Hannover	42
Mainz	40	Hertha	41
Hamburg	40	Werder	39
Stuttgart	38	Freiburg	36
Frankfurt	34	Frankfurt	36
Nürnberg	29	Stuttgart	32
Braunschweig	26	Hamburg	27
Freiburg	24	Nürnberg	26
Werder	22	Braunschweig	25

Bundesliga 2013/2014, classifica reale e simulata secondo la misura H.

In generale, ciò che emerge è che l'analisi delle reti riesce a descrivere la performance di una squadra. Considerando che stiamo parlando solo di reti di passaggi, senza aver ancora incluso altri tipi di eventi (dribbling, tackle, etc.).

Tuttavia, c'è comunque un errore da migliorare. E per raggiungere la perfezione, vanno scomodati gli dèi. Il calcio l'hanno spiegato già in tanti prima di noi, e qualcuno ha anche provato a formularne le leggi che lo governano. Uno di questi è Max Pezzali. Già, la dura legge del gol.

Osservando le squadre per cui la nostra simulazione sbagliava di più, abbiamo notato che sono sia quelle che giocano in contropiede, che quelle che difendono male. Queste, però, sono osservazioni empiriche, la cui formalizzazione risulta difficile. Si possono invece analizzare gli eventi in campo ispirandosi a Pezzali. Ad esempio, una squadra dalla difesa debole subisce la dura legge del gol:

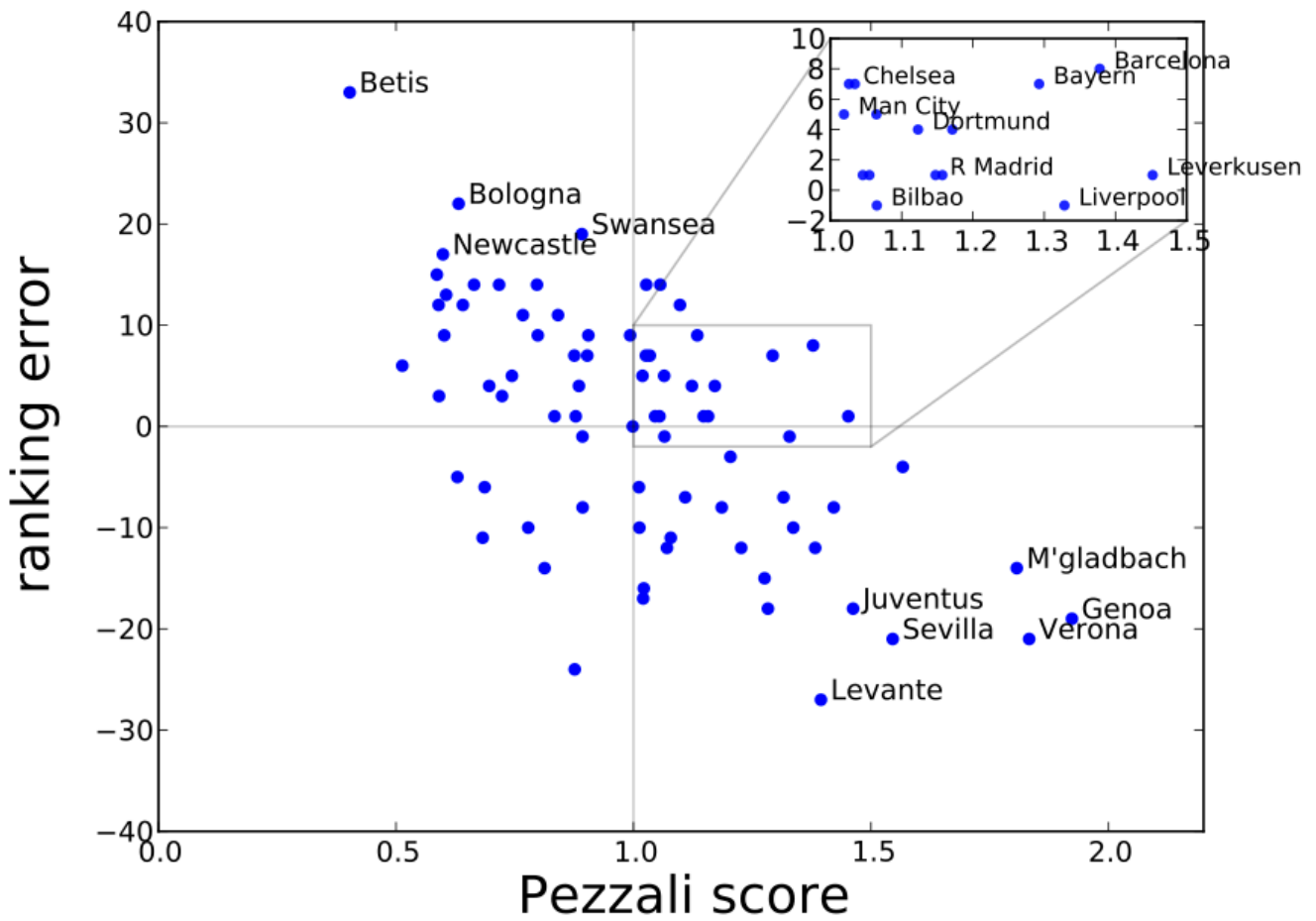


*E' la dura legge del gol, fai un gran bel gioco però, se non hai difesa gli altri segnano e poi vincono.*

Mentre ci sono squadre che questa legge la impongono:

*Loro stanno chiusi ma, alla prima opportunità, salgon subito e la buttan dentro a noi.*

A livello numerico, abbiamo definito il Pezzali score: il rapporto tra i tiri in porta e i gol fatti moltiplicato per il rapporto tra i gol subiti e i tiri fatti dall'avversario. Quando il Pezzali score è basso, la squadra è un po' come l'Inter, gioia e dolore dell'autore stesso: tanto gioco, ma agli avversari di turno basta un tiro per fare un gol. Quando, invece, il valore è alto, la squadra ottiene il massimo dal minimo sforzo offensivo. Per capire quanto la dura legge del gol sia decisiva nel dirci dove sbaglia la nostra simulazione, abbiamo confrontato il Pezzali score di ogni squadra con l'errore in termini di punti finali che la nostra simulazione assegna. Il risultato (vedi sotto) conferma quanto la visione di Pezzali, pur poetica e non scientifica, fosse azzeccata. Se con l'*H* siamo in grado di valutare quanto sia efficace il gioco di una squadra, la dura legge del gol ci riporta alla realtà: il calcio è un gioco dove esiste una componente casuale. Non basta il gran bel gioco per vincere, insomma.



La dura legge del gol: l'errore maggiore (Betis, Verona, etc) corrisponde a valori estremi del Pezzali score.

Il mistero non è ancora risolto, perché in fondo il giallo non è ancora finito. Con buona pace di Malvaldi e dei suoi vecchietti ficcanaso che risolvono casi di cronaca nera, c'ancora tanto da studiare, e da scrivere, prima di riuscire a spiegare il gioco più bello del mondo. E anche se la dura legge del gol continuerà a valere nei secoli dei secoli, proveremo fino in fondo ad infrangerla.

*I nostri risultati in tema di analisi dei dati calcistici li trovate qui:*

[The harsh rule of the goals: data-driven performance](#)

[indicators for football teams](#)

*P. Cintia, L. Pappalardo, D. Pedreschi, F. Giannotti, M. Malvaldi*

*2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA'15), Paris, France*

*Ne abbiamo parlato anche a [radio3 scienza](#) e a Zona Cesarini (Radio1):*

**Ringraziamo:** Marco Malvaldi, per averci prestato i vecchietti indagatori. Max Pezzali, perché siamo cresciuti negli anni '90. Mariano Tredicini e la Tim, per aver supportato la nostra ricerca.

---