

Quest'anno il campionato lo vincerà...

2 Settembre 2019

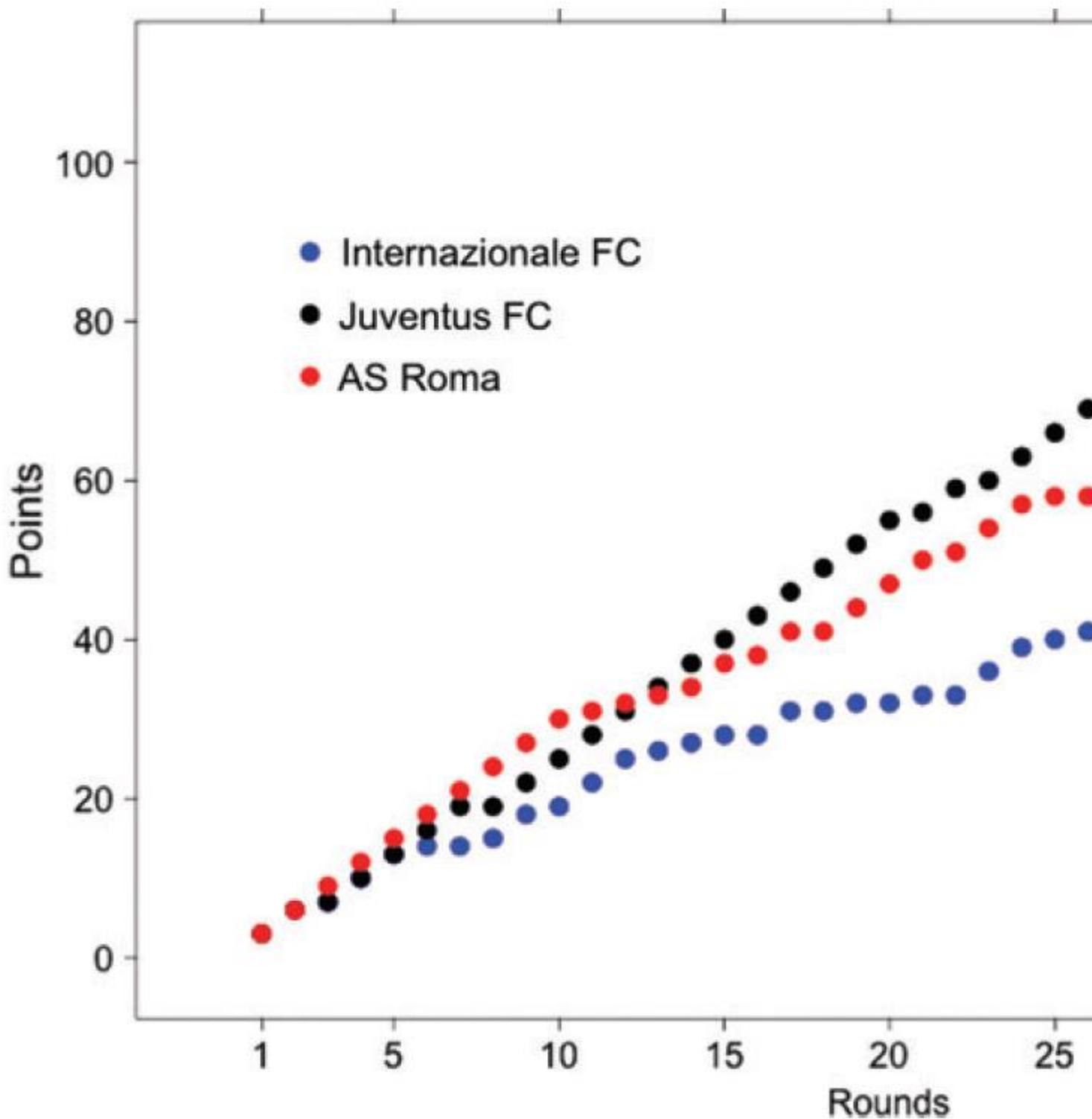
Il nostro data scientist Giuseppe Jurman, appassionato di calcio da sempre, ha applicato i modelli predittivi al calcio e ha scoperto che...

La previsione dei risultati di partite e competizioni sportive è un campo di ricerca in crescita, che beneficia dell'aumento della quantità di dati disponibili e nuove tecniche di analisi dei dati.

Previsioni eccellenti possono essere raggiunte da metodi statistici avanzati e di *machine learning* applicati a dati storici ed economici dettagliati, in particolare per gli sport più popolari come il calcio.

Lo studio di Giuseppe Jurman "[Seasonal Linear Predictivity in National Football Championships](#)", ha mostrato che, nonostante il gran numero di fattori concomitanti, i risultati di una squadra di calcio nelle competizioni più lunghe (ad esempio un campionato nazionale) seguono una tendenza sostanzialmente lineare che si può usare a fini predittivi. In altre parole, **chi l'ha detto che non bastino i punti fatti a inizio campionato per capire grosso modo come andrà a finire?**

A sostegno di questa affermazione, il ricercatore FBK ha condotto una serie di esperimenti di regressione lineare rispetto ad approcci alternativi su un database che raccoglie i risultati annuali di **746 squadre che giocano in 22 divisioni fino a cinque diversi livelli da 11 paesi, in 25 stagioni calcistiche, per un totale di 181.160 partite raggruppate in 9386 serie storiche stagionali**. Lo studio, condotto per passione nel tempo libero, ha mostrato dei risultati sopra le aspettative. Il modello lineare adottato rappresenta in effetti un compromesso coerente tra prestazione e semplicità. In altri termini: un'ottima approssimazione col minor numero possibile di dati e variabili da processare a confronto con altri modelli più complessi. In particolare, il modello è stato addestrato nella parte iniziale del campionato (registrando i punti raccolti dalle squadre dopo le prime 5-10 partite) per testare la previsione lineare del numero di punti che si sarebbero totalizzati a fine campionato. Ebbene, lo studio ha dimostrato che anche un approccio così minimalista e senza usare i dati storici può mostrare buoni risultati predittivi, raggiungendo un margine di errore di 2,5 punti (sulla classifica finale). Non ci resta che aspettare la fine del campionato per verificare se ci saranno grosse sorprese rispetto a quanto possiamo prevedere a partire dai dati.



La vocazione sportiva di Trento e l'attenzione alla ricerca scientifica applicata nel settore è testimoniata dal fatto che la città abbia ospitato il primo Hackathon del calcio italiano nell'ottobre 2017 e che dal 2018 ospiti il Festival dello Sport, organizzato da La Gazzetta dello Sport e da Trentino Marketing, con la collaborazione di Provincia Autonoma di Trento, Comune e Apt di Trento e col patrocinio di Coni e Comitato Italiano Paralimpico. Con il titolo "IL FENOMENO, I FENOMENI" si svolgerà a Trento dal 10 al 13 ottobre 2019 la seconda edizione del Festival. Fa parte del programma la Soccer data challenge, una competizione promossa da SoBigData e aperta a tutti gli appassionati di dati e calcio. Sarà una maratona analitica sul calcio: le squadre

partecipanti avranno 30 ore di tempo per risolvere un problema analitico legato al calcio, utilizzando il più grande dataset di eventi di gioco mai rilasciato prima. La competizione inizierà il 10 ottobre e si concluderà il giorno seguente con la cerimonia di premiazione. In 30 ore consecutive le squadre svilupperanno una soluzione per l'analisi di partite di calcio, utilizzando i dati dei 5 maggiori campionati europei di una stagione calcistica e di due competizioni internazionali.

Ben 1.941 le partite da analizzare, 4.299 i calciatori da monitorare e 3.251.294 gli eventi di gioco tracciati. I partecipanti presenteranno i loro lavori a una giuria di professionisti del mondo del calcio e dei BigData. Per i vincitori il premio è di 5000€. Vinca il migliore!



The poster features a central image of a soccer ball with a data visualization overlay. The visualization consists of a starburst pattern with numerical values (0.2, 0.4, 0.6, 0.8) and a red box labeled 'goals'. The background is a dark purple with various soccer-related terms like 'interceptions', 'off target shot', 'blocked shot', 'saved shot', 'wood work', 'takeons', 'failed takeons', 'reds', and '2nd yellow'. The top of the poster has a red banner with the text 'SOCCER DATA CHALLENGE' in large white letters. To the right of the banner is the logo for 'iFoS' (International Festival of Sports) with the text 'FESTIVAL dello SPORT' below it. At the bottom of the poster, a red banner contains the dates '10 - 11 ottobre 2019' in white text.

Infine, ma non per importanza, Jurman è membro del comitato di gestione della laurea magistrale in [data science](#) presso l'Università di Trento e insegna il corso di data visualization. A fine anno

accademico, nell'estate 2020, avremo i primi laureati: chissà che qualcuno di essi non vada a lavorare per qualche squadra di massima serie nel calcio o in altri sport come basket o pallavolo.

LINK

<https://magazine.fbk.eu/it/news/questanno-il-campionato-lo-vincerà/>

TAG

- #calcio
- #data science
- #festival
- #sport

MEDIA COLLEGATI

- Seasonal Linear Predictivity in National Football Championships:
<https://magazine.fbk.eu/wp-content/uploads/2019/09/jurman-soccer-predictivity.pdf>

AUTORI

- Giancarlo Sciascia