

# This year the SERIE A champions will be...

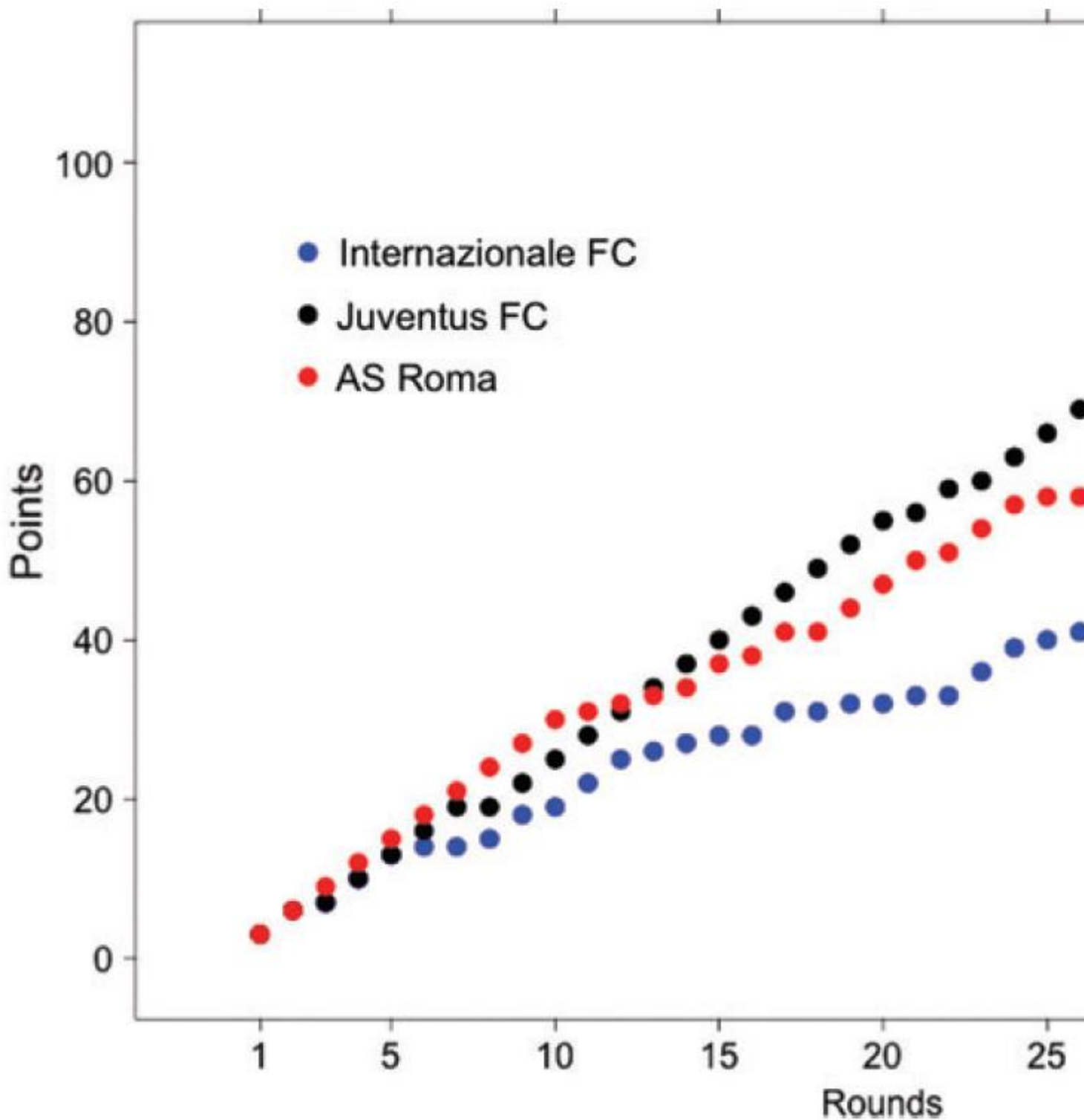
September 2, 2019

**Our scientist Giuseppe Jurman, who has always been a soccer fan, has applied predictive models to soccer and found that...**

Forecasting the results of sports matches and competitions is a growing field of research, which benefits from the growing amount of available data and the new data analysis techniques. Excellent forecasts can be achieved by advanced statistical and machine learning methods applied to detailed historical and socioeconomic data, particularly for the most popular sports such as soccer.

Giuseppe Jurman's study "[Seasonal Linear Predictivity in National Football Championships](#)" showed that, despite the large number of confounding factors, the results of a soccer team in longer competitions (such as a national championship) follow a substantially linear trend which can be also useful for predictive purposes.. In other words, **who said that the points made at the beginning of the championship are not enough to roughly guess how it will end?** To support this claim, the FBK researcher conducted a series of linear regression experiments compared to alternative approaches on a database collecting the annual results of **746 teams playing in 22 divisions up to five different levels from 11 countries, in 25 soccer seasons, for a total of 181,160 games grouped in 9,386 seasonal historical series.**

The study, conducted in his leisure time, showed results above expectations. The linear model adopted actually represents a coherent compromise between performance and simplicity. In other words, it achieves an excellent approximation with the smallest possible number of data and variables to be processed when compared with other more complex models. In particular, the model was trained in the initial part of the championship (recording the points made by the teams after the first 5-10 matches) to test the linear prediction of the number of points that would have been earned at the end of the championship. Well, the study has shown that even such a minimalist approach without using historical data can show good predictive results, reaching a margin of error of 2.5 points. We just have to wait for the end of the championship to see if there will be big surprises compared to what we can expect from the data.



The sporting vocation of Trento and the attention to scientific research applied to this sector is testified by the fact that the city hosted the first Hackathon of Italian soccer in October 2017 and that, since 2018, it has hosted by the Sports Festival, organized by the La Gazzetta dello Sport newspaper and by Trentino Marketing, with the cooperation of the Autonomous Province of Trento, the City of Trento and the Trento Agency for Tourism, and with the patronage of CONI, the Italian national olympics committee, and the Italian Paralympic Committee. The second edition of the Festival will take place in Trento October 10 through 13, 2019. The program includes the Soccer data challenge, a competition promoted by SoBigData and open to all data and soccer enthusiasts.

It will be an analytical marathon about soccer: the participating teams will have 30 hours to solve an analytical problem linked to soccer, using the largest dataset of game events ever released before. The competition will start on October 10, and will end the day after with the award ceremony. In 30 consecutive hours the teams will develop a solution for the analysis of soccer matches, using data from the 5 largest European championships of a soccer season and two international competitions.

**A good 1,941 games to analyze, 4,299 players to monitor and 3,251,294 the game events tracked.** Participants will present their work to a jury of soccer and Big Data experts. For the winners, the prize is € 5,000. Let the best win!



The poster features a central 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 line pointing to a 'goals' label. The background is a dark red with a pattern of soccer-related terms like 'interceptions', 'off target shot', 'blocked shot', 'saved shot', 'wood work', 'takeons', 'failed takeons', 'reds', and '2nd yellow'. The text 'SOCCER DATA CHALLENGE' is prominently displayed in white, outlined letters at the top. The iFoS logo (Festival dello Sport) is in the top right corner. The dates '10 - 11 ottobre 2019' are at the bottom.

**SOCCER DATA CHALLENGE**

iFoS  
FESTIVAL dello SPORT

10 - 11 ottobre 2019

Last but not least, Jurman is a member of the management committee of the Master's Degree in [Data Science](#) at the University of Trento and teaches Data Visualization. At the end of the

academic year, in the summer of 2020, the first students will graduate: it might be that some of them will end up working for some top-flight soccer team or in other sports like basketball or volleyball.

#### **PERMALINK**

<https://magazine.fbk.eu/en/news/this-year-the-serie-a-champions-will-be/>

#### **TAGS**

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

#### **RELATED MEDIA**

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

#### **AUTHORS**

- Giancarlo Sciascia