technische universität dortmund

fakultät statistik

2nd Workshop on *Sports Analytics in Soccer* at TU Dortmund University
25th of November 2025

Program

Tuesday, 25th November 2025

10:00: Welcome by orga team (Andreas Groll & Rouven Michels)

10:10: Talks

1. Gordana Marmulla (ETH Zurich)
Women's and men's soccer teams: Differences in their play beyond physiology and tournament winning probabilities

Association football (soccer) is played according to the same rules for both sexes, while at the same time both the FIFA tournament format and ranking algorithm differ between the sexes. In this talk, I present two separate but related projects. The first project addresses the question whether there are notable differences between men's and women's teams in how they play soccer, and, whether these differences go beyond what is to be expected due to physiological differences between the sexes. The second project exemplifies consequences of the different tournament formats and rankings on tournament winning probabilities, when the latter are used for individual match outcomes. We show for the two most recent world cups that exact computation of winning probabilities is possible and substantially faster than their approximation through simulating the tournaments tens of thousands of times, independent of the model used for individual match outcomes.

2. Rouven Michels (TU Dortmund Univ.)
Applying extensions of the Dixon and Coles model to women's soccer and the new Champions League format

The prevalent model by Dixon and Coles (1997) is an extension of the double Poisson model, in which the number of goals scored by each team is modeled by two independent Poisson distributions. This model involves moving probabilities between the scores 0–0, 0–1, 1–0 and 1–1. We demonstrate that this constitutes a particular instance of a multiplicative model known as the Sarmanov family. Based on this family, we create more suitable models by adjusting the probabilities and using other discrete distributions. We apply these new models to women's football scores, which exhibit characteristics that differ from those of men's football, and to the new Champions League format, which shows different patterns to previous years' formats.

11:10: Coffee break

11:30: Talks

3. Lander Rodríguez (BCAM Bilbao)
Research interests and future plans in sports analytics

For the moment, Lander is mainly interested in tracking data (football and basketball) and game outcome prediction, and in parallel he is working on a project related with synthetic data applied to sports. He will join the workshop there with more questions than answers, and will briefly present some of these ideas and hopes to discuss them.

technische universität dortmund

fakultät statistik

2nd Workshop on *Sports Analytics in Soccer* at TU Dortmund University
25th of November 2025

4. Alexander Gerharz (Borussia Dortmund)
Clustering of Training Drills

In professional football, there are a lot of different training drills used on the pitch to develop the players, teach tactics and level up and then keep up a high level of fitness. These drills can have quite some variations, but for load monitoring it would be nice to categorize them. In this presentation, I will show a clustering of training drills using the physical performance parameters derived from a gps tracking system. I will also show, how the training content varies and evolves over time.

12:30: Lunch break (Mensa)

13:30: Talks

5. Robert Bajons (WU Vienna)
Rethinking goals above expectation: Player evaluation in football using residualized scores

Expected Goals (xG) is one of the most popular metrics in modern football analytics. As a measure of a shot's value, it is commonly used to evaluate the shooting skills of players by considering goals over expectation (GAX), i.e., the difference for each shot between actual and expected goal. Recently, GAX have been criticized for being unstable over seasons, for not allowing for uncertainty quantification, and for being prone to biases in the data. In this work, we present residualized GAX (rGAX), a generalization of GAX. This generalization enables us to utilize black-box xG models while still obtaining valid uncertainty quantification. Thereby, rGAX addresses existing issues with GAX. Furthermore, rGAX can be related to player-specific effect estimates in an interpretable semiparametric model, providing a different view on and a deeper understanding of player evaluation via GAX. We present results for the 2015/16 season of the top 5 European leagues and discuss further extensions, as well as adaptations for women's football.

6. Andreas Groll (TU Dortmund Univ.)
Comparing Machine Learning and classic statistical approaches for injury prediction in young professional soccer players

Frequent injuries pose a problem in professional soccer that is being tackled with preventive measures. Consequently, injury prediction and prevention are also increasingly addressed from a statistical perspective. In a pilot study, several machine learning algorithms and conventional statistical approaches have been compared regarding their potential to predict time-loss non-contact lower-body injuries in professional youth soccer players, using data from a prospective cohort study with 56 players of which 22 were injured. The covariates considered here include basic soccer-related as well as neuro-muscular and biomechanical features derived from physical testing. Lasso regularized logistic regression, naive Bayes, linear discri-minant analysis, k-nearest neighbors, classification trees, random forests, XGBoost, and support vector machines are considered for binary classification and prediction of an injury occurrence. The prediction results from a cross-validated procedure are compared regarding multiple quality measures. In addition to the specific results on the available injury data set, the proposed comparison procedure of several models for binary prediction provides a generally applicable analysis guideline. This roadmap can also be applied in other contexts where similarly structured small but rich data sets are available.

technische universität dortmund

fakultät statistik

2nd Workshop on *Sports Analytics in Soccer* at TU Dortmund University
25th of November 2025

14:30: Coffee break

15:00 - 16:00: Discussion & Future joint research activities

18:00: Dinner (Restaurant *"meierabend."*, see Directions below)

21:00: UEFA CL (Borussia Dortmund vs. FC Villareal)

Directions

Arrival from airport Düsseldorf to Dortmund main station

Direction from Dortmund main station to TU Dortmund University
At the Dortmund main station take the S-Bahn: S1 (Platform 7)

EXIT: "Dortmund Universität". Leave the station and walk left passing the construction area of the library building. Cross "Vogelpothsweg" street and follow the crosswalk to the entrance of the Math-Tower. Walk up the stairs. Continue straight, at the end turn right and take the stairs or elevator up to the 2nd floor.

meierabend.

Direction from TU Dortmund to "meierabend." restaurant
with "S-Bahn S1" till (Hbf/main station):
At the main station take the following suburban train/S-Bahn:
U47->Aplerbeck, U41->Do-Hörde or U49->Hacheney:
EXIT "Stadtgarten"
Switch over: U42->Hombruch
EXIT: "Kreuzstraße".
Sidewalk: toward "Große Heimstraße", the restaurant will be after 250m on the left. For alternative routes check Google maps.

From city center or Dortmund Hbf, you can also simply walk!

Große Heimstraße 45
44137 Dortmund
Tel. 02 31 – 53 48 77 76

meierabend.

TU Dortmund University Campus Nord (Math building) Vogelpothsweg 87 2nd Floor/meeting room 217b 44221 Dortmund