

18-090 vom 12.06.2018

Statistik der TU Dortmund sieht Spanien als wahrscheinlichen Fußball-Weltmeister Deutschland folgt – vor Brasilien, Frankreich und Belgien

Fußball-Weltmeister beim Turnier in Russland wird Spanien, Deutschland hat die zweitbeste Chance auf den Titelgewinn – das jedenfalls leiten Statistikerinnen und Statistiker aus einem Modell ab, das sie für diese WM entwickelt haben. Sie beziehen dabei Modelle ein, die sie für die letzten großen Turniere – die Europameisterschaften (EM) 2012 und 2016 sowie die Weltmeisterschaft (WM) 2014 – entwickelt hatten. Die Ergebnisse, die maßgeblich von der Statistik der TU Dortmund erarbeitet wurden, liegen jetzt als Working Paper vor.

In dieses Paper wurden 32 Wahrscheinlichkeiten für den Turniersieg aller Mannschaften, aber auch „Überlebenswahrscheinlichkeiten“ für die verschiedenen Turnierrunden sowie der wahrscheinlichste Turnierverlauf eingearbeitet. Gemäß dem Modell wird es diesmal besonders spannend, Spanien liegt nur ganz knapp vor Deutschland als Favorit auf den Turniersieg.

Spanien wird mit einer Wahrscheinlichkeit von 17,8 Prozent Weltmeister, gefolgt von Deutschland mit 17,1 Prozent. Dahinter kommen Brasilien (12,3 Prozent), Frankreich (11,2 Prozent) und Belgien (10,4 Prozent). Interessanterweise ist aber eigentlich Deutschland laut Modell das stärkste Team und würde auch im direkten Vergleich mit Spanien oder Brasilien mit der höheren Wahrscheinlichkeit als Sieger hervorgehen. Deutschland hat aber den schwereren „Turnier-Ast“ als Spanien. Dies wird vor allem am Achtelfinale deutlich, wo Spanien auf ein Team der Gruppe A (Uruguay, Russland, Saudi-Arabien, Ägypten) trifft und Deutschland auf ein Team der Gruppe E (Brasilien, Schweiz, Costa Rica, Serbien).

Während Spanien und Deutschland mit ähnlichen Wahrscheinlichkeiten mindestens das Achtelfinale erreichen (88,4 Prozent bzw. 86,5 Prozent), unterscheiden sich dementsprechend die Wahrscheinlichkeiten, mindestens das Viertelfinale zu erreichen, recht deutlich: Da liegt Brasilien mit 73,1 Prozent deutlich vor Deutschland mit 58 Prozent. Würde Deutschland aber das Viertelfinale erreichen, wäre es ab diesem Zeitpunkt laut Modell der Favorit auf den Titel.

Wie sind die Forscherinnen und Forscher vorgegangen? Zunächst haben sie drei verschiedene Modellierungsansätze betrachtet:

- o einen Regressionsansatz, den sie vom Grundprinzip her bereits in den vergangenen Jahren benutzt haben (EM 2012, WM 2014, EM 2016)
- o einen sogenannten Random Forest, eine moderne Technik aus dem Machine Learning
- o einen Rankingansatz, der ähnlich zum FIFA Rang – allerdings basierend auf einem statistischen Modell – allen Mannschaften eine aktuelle Stärke zuordnet

Die ersten beiden Ansätze basieren dabei auf Variablen der Nationalmannschaften: beispielweise deren Wettquote auf den WM-Sieg, deren aktuellem FIFA Rang, dem Durchschnittsalter der Spieler oder dem Erfolg der Spieler in der vorangegangenen Champions-League (CL)-Saison. Dies wurde gemessen an der Anzahl der Spieler im Kader einer Mannschaft, die mit ihrem nationalen Club mindestens das CL-Halbfinale erreicht haben.

Um die Vorhersagegüte zu testen, haben die Forscherinnen und Forscher auf den Daten der vier vergangenen FIFA WMs 2002-2014 die Modelle jeweils auf den Spielen dreier Turniere kalibriert und dann die Prognosegüte auf den Spielen der weggelassenen WM berechnet. Dabei schnitten der Random Forest und der Rankingansatz sehr gut ab. Sie kamen bereits sehr nah an die Prognosegüte der Wettanbieter zu diesen Spielen.

Um die Prognose weiter zu verbessern, kamen die Forscherinnen und Forscher auf die Idee, die durch den Rankingansatz geschätzten Teamfähigkeiten dem Random Forest als weitere Variable zu geben; dies steigerte die Prädiktionsgüte noch einmal erheblich und die Statistikerinnen und Statistiker waren nun sogar in der Lage, bei den Spielen der vergangenen vier WMs die Wettanbieter zu schlagen. Damit, so JProf. Andreas Groll, „haben wir das beste Modell in unserer bisherigen Modellierungserfahrung. Mit diesem Modell haben wir die WM 2018 100.000-mal simuliert und bekommen so Wahrscheinlichkeiten für alle 32 Teilnehmer für den WM-Sieg, aber auch Wahrscheinlichkeiten für das Weiterkommen in den einzelnen Turnierrunden für alle Teams.“

Um die Vorhersagegüte des Modells weiter zu validieren, werden die Wissenschaftlerinnen und Wissenschaftler auf alle 64 Spiele der kommenden WM 2018 gemäß einer Erwartungswert-Maximierungs-Strategie einen kleinen Betrag wetten; ein Wettgewinn wäre ein weiteres Indiz für eine gute Passung der Wahrscheinlichkeiten, die sie durch das Modell erhalten.

Weitere Informationen

<https://arxiv.org/pdf/1806.03208.pdf>

Bildinformation:

Vertraut der eigenen Prognose und setzt auf Sieg: JProf. Andreas Groll Foto:
Roland Baeye/TU Dortmund

Ansprechpartner für Rückfragen:

JProf. Andreas Groll

Fakultät Statistik

Telefon: 0231-755 4229

E-Mail: andreas.groll@tu-dortmund.de

Die Technische Universität Dortmund hat seit ihrer Gründung vor 50 Jahren ein besonderes Profil gewonnen, mit 16 Fakultäten in Natur- und Ingenieurwissenschaften, Gesellschafts- und Kulturwissenschaften. Die Universität zählt rund 34.600 Studierende und 6.200 Mitarbeiterinnen und Mitarbeiter, darunter etwa 300 Professorinnen und Professoren. Das Lehrangebot umfasst rund 80 Studiengänge. In der Forschung ist die TU Dortmund in vier Profildbereichen besonders stark aufgestellt: (1) Material, Produktionstechnologie und Logistik, (2) Chemische Biologie, Wirkstoffe und Verfahrenstechnik, (3) Datenanalyse, Modellbildung und Simulation sowie (4) Bildung, Schule und Inklusion. Beim QS-Ranking „Top 50 under 50“ belegt die TU Dortmund Rang drei der bundesdeutschen Neugründungen.