

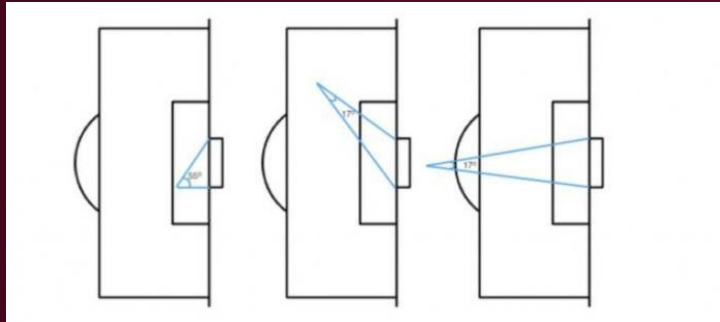


Analyzing Soccer Data: Can Statistics Accurately Predict Future Results?

Muhammed Hanan

Math of Soccer

- “A soccer game really is a field day for mathematicians.” (Linder, 4)
- From the first second of when a match starts to the last, almost everything that is going on in the field is being measured studied and analyzed in the forms of probability, statistics, graphs, patterns, geometry, etc, and converted into data.
- Distance and Geometry, more specifically angles play a big part on the likelihood of scoring a goal and making a successful pass. The larger the angle is to goal and the shorter the distance, the more likely a goal be scored.



History of Soccer Analytics - Charles Reep

- Charles Reep is widely considered to be the first soccer analyst and the godfather of soccer analysis.
- Reep was an accountant and a Royal Air Force Pilot.
- He developed a method of analyzing matches using a tally system, in which he recorded events such as shots, passes, and goals.
- He claimed that his analysis revealed that the majority of goals were scored from plays that were three passes or fewer, and that this information could be used to develop more efficient playing strategies.
- His early and amateur analysis lead to him working with a few different soccer clubs and their managers in the mid 1900s to whom he brought success.



Charles Reep Continued

- Reep's theories are widely criticized and thought of as flawed and controversial.
- Reep's tally system is considered unreliable.
- Reep's ideas of soccer are considered over-simplified and advocate "anti-football" and "route 1" types of play.
- His style of play make the sport boring and justify the lack of creativity

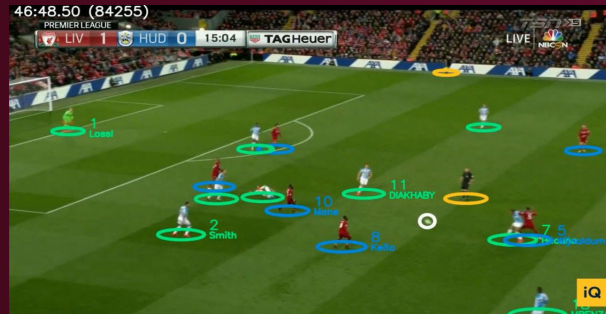
(17) Bolton Wanderers		Blackpool		(18) -1
Second		FA CUP, FINAL		1st 0 0 0 0
at Wembley		May 2nd 50		2 2 2 1
Half		10 15 NFA		1st 0
Conditions		Went down		- 1 0
Law		Against		- 0 -
B.O. BOLTON		K.O. BLACKPOOL		- 0 -
- 1 0		1 1 2		- 0 -
2 0		1		- 0 -
1 1 1 1 1 1 1 1 1 1		1 1 1 1 1 1 1 1 1 1		1st 0 0
1 1 1 1 1 1 1 1 1 1		1 1 1 1 1 1 1 1 1 1		2 0 -

History of Soccer Analytics Continued

- The Premier League was formed in the 90s, causing an evolution in ideas of how the game should be played and an evolution in technicalities and analytics.
- Opta a British sports analysis company with coverage in more than 30 sports in 70 countries began collecting Premier League game data in 1996, providing clubs with basic stats like distance run, tackles made, and passes made.
- Slowly statistics began a big a part of the game and managers such as Arsene Wenger have gained great success from the evolution of data as well as teams such as Liverpool.

How is Soccer Data Collected?

- Opta have professional analysts who monitor, record and annotate every touch of the ball.
- Automatic video analysis and fully automated systems
- ChyronHego's TRACAB Gen5 incorporates distributed camera architecture.
- The system can record 25 times in one second, giving about 135,000 records of data.
- Ball and player position data is automatically tracked by cameras or sensors. The position is usually measured by an x,y, and sometimes z coordinate system related to the pitch.
- Advanced image processing provide tracking data on each player, referee, and ball movement.
- Motion sensor in Adidas 2022 World Cup that sends out information 500 times a second.
- Wearable devices for consumers.



Soccer Data Explained

- Soccer data can be broken down into three data types, event, tracking, and physical data.
- Event data is essentially just what happens on or with the ball such as passes, shots, tackles, etc.
- Tracking data is the positions of each player on and off the ball on the pitch.
- Physical data measure different physical aspects of a player such as height, weight, speed, etc.



How is Soccer Data Applied?

- Scouting and player recruitment
- Player fitness and performance
- Tactics and training

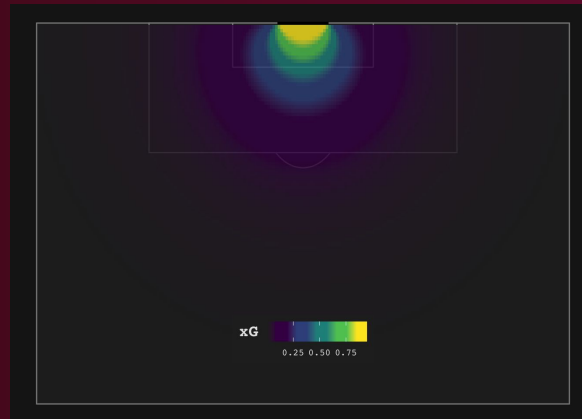


xG

- Expected Goals
- xG is a statistical metric that assesses the quality of scoring chances in soccer matches.
- xG calculates the likelihood of a shot resulting in a goal based on various factors like shot location, angle, type of play, and other relevant variables.
- It provides a more accurate representation of a team's or player's attacking effectiveness than simply counting goals scored or shots taken
- xG can identify players who are performing well but not scoring due to bad luck and players who are scoring but from low-quality opportunities.
- xG is a valuable tool in soccer analysis that provides a more nuanced understanding of a team's or player's attacking effectiveness.

xG Models

- xG models use historical information from thousands of shots with similar characteristics to estimate the likelihood of a goal on a scale between 0 and 1.
- xG models use logistic regression.
- Some important variables xG models consider are distance to the goal, angle to the goal, quality of chance, body part, assist, play pattern, and more.
- Free kicks and penalties are considered constants.



xG Related Metrics

- xGp/90 measures a player's expected goals per 90 minutes of play.
- xA measures the likelihood that a pass will lead to a goal.
- xGA estimates the probability of a team conceding a goal based on the quality of the opposition's scoring opportunities.
- PSxG estimates the probability of a shot resulting in a goal after it has been taken and is useful for evaluating goalkeepers.
- Expected statistics provide a more accurate and nuanced understanding of the game than simply looking at the number of goals scored or shots taken.
- More expected statistics are likely to be developed in the future to provide an even deeper understanding of the game.

Predictive Power of xG

- xG is not a perfect predictor of future performance in soccer.
- Soccer is a complex and dynamic game with many variables that can influence outcomes and xG cannot possibly cover all those variables.
- xG does not consider factors such as player mistakes.
- xG can provide valuable insights into a team's or player's performance and help identify areas for improvement.
- xG can be used in conjunction with other statistical and analytical methods to assess the effectiveness of tactics and strategies.
- As soccer analytics evolves, xG will remain an important statistic for evaluating the game and forecasting future outcomes.

Home/Away Model and xG Model

- Predicts matches for the 22/23 English Premier League season.
- Uses Poisson Distribution
- Poisson Distribution gives the probability of an event happening a certain number of times within a given interval of time or space.
- In this case, the event is scoring a goal and how many times it happens.
- Model uses a team's past GF and GA to for home and away to make predictions
- By comparing a team's GF and GA against the league as a whole, I'm able to get strength of a team's attack and defense.
- By using poisson distribution, I'm able to get the result of match and also get the probability of different amounts of goals being scored/conceded.
- The xG model follows the same principals however I replaced GF and GA with xG and xGA.

ML Model - ProphitBet

- Ability to train ML models with different algorithms including Neural Networks, Random Forests & Ensemble models
- Scrapes current data from the web
- A lot of different variables, parameters and tuning to use on ML model
- A lot of soccer data and statistics available
- Different analysis and evaluations methods

Results and Observations

- Unexpectedly, the models were not too far off in terms of accuracy.
- The ML model was the most accurate, followed by the xG model and lastly the home/away model.
- I noticed that the models had were able to predict well when the winner was obvious.
- When accounting for unexpected results, the models lacked in that.
- The ML models had a tough time dealing with matches ending in draws
- xG might be a slightly better predictor than just GF and GA

Ideal Model

- Machine learning model
- RandomForest
- xG statistics as parameters
- Trained on at least 3-5 seasons worth of matches
- Scrapes new soccer data

Take Away

- More of an appreciation of math and statistics
- Better understanding of mathematical and statistical concepts
- Better understanding of soccer
- A more mathematical perspective on soccer

Future

- Soccer analytics is rapidly growing and changing day by day
- New data collecting and analytical technologies and maybe the rise technologies similar to VAR
- The role of AI and machine learning in soccer analysis
- Overall just how analytics changes the sport as a whole
- I want to continue learning more about data science and analytics to see how it changes sports.
- Hopefully, be able to make that ideal model and make some safer bets.

Conclusion

- To answer the research question, soccer statistics can kind of predict results accurately, but I am leaning more towards no at the moment.
- I'd say with more resources, time, more processing power and LARGER datasets, yes.
- I think however, with technology and statistical analysis always evolving, in the future, it should definitely be more accurate.
- But soccer is an unexpected game with a lot possibilities and sometimes it's difficult to quantify the unexpected, which is I feel sports analysis lacks.