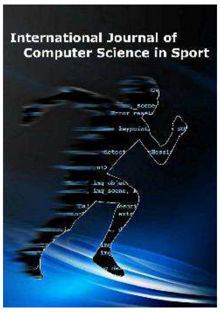
International Journal of Computer Science in Sport

Volume 22, Issue 2, 2023

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2023-0010



The Kos Angle, an optimizing parameter for football expected goals (xG) models

Hassani Karim[1], *Lotfi Marwane*[2]

[1]*Hassan II University, Machine Intelligence Department, Casablanca, Morocco;* [2]*Fath Union Sport Football Club, Rabat, Morocco*

Abstract

The utilization of metrics such as expected goals (xG) has the potential to provide teams with a competitive edge. By incorporating xG into their analysis and decision-making processes, teams can gain valuable insights. This study proposes a new approach to football xG modeling using Kos Angle which represents the shooting angle, from which we substract the angles occupied by players inside the shot angle. The objective of this study is to evaluate the impact of the Kos Angle feature on the performance of football xG models. After developing the mathematical formula of the Kos Angle, we selected additional features and built different xG models. Subsequently, the impact of the Kos Angle feature on the models' performances was evaluated, revealing an increase in Recall and Precision and a decrease in Brier score and RMSE. We also found that the Kos Angle accounted for a significant portion of the models' predictive power. By providing a more realistic representation of shot situations, the addition of the Kos Angle feature allows the improvement of xG models performances, which can give a more valuable insights to football professionals who rely on xG metrics and their variations.

KEYWORDS: FOOTBALL, EXPECTED GOALS, MACHINE LEARNING, FEATURE IMPORTANCE, CLOSED ANGLES

Introduction

There are many definitions and ways to explain expected goals, the most complete one for many experts is the one presented by (Tippett, 2019, p. 4): "*xG indicates how many goals a team could have expected to score based on the quantity and quality of chances that they created in a match*".

During the English Premier League match between Manchester City and Tottenham Hotspur on November 21, 2020, Manchester City had more shots on goal than Tottenham Hotspur, but the final score was 2-0 in favor of Tottenham Hotspur. When we look at the xG statistics for the match, we see that Manchester City had an xG of 3.07, which means that based on the quality of their shots at goal, they were expected to score around three goals. Tottenham Hotspur, on the other hand, had an xG of only 0.55, which means that they were expected to score less than 1 goal based on the quality of their scoring chances. This example illustrates how xG can provide a more accurate representation of a team's performance by taking into account the quality of their scoring chances rather than just the number of goals scored. In this case, Manchester City was unfortunate not to score more goals despite creating better chances, while Tottenham Hotspur was able to convert their limited opportunities to secure a surprising victory.

The expected goals models can differ by the parameters used to evaluate shot occasions. Usually, the features used to train xG models are derived from two categories of data: positional data also known as tracking data representing the coordinates of the ball and the players throughout play at regular intervals, and event data representing events on the field such as tackles, dribbles, shots, and pressures.

Various studies have been conducted on xG models, utilizing differing datasets in terms of size and information, to train various machine learning algorithms, resulting in varying outcomes. Lucey et al. (2015) used almost 10,000 shots and considered various features such as features derived from game phase information (e.g., corner, free-kick, open-play, counter-attack), the distance from the shooter to the defenders within the shot angle, a variable containing information about the interaction of surrounding players, speed of play, shot location x, and shot location y to build a Logistic Regression xG model with an average error of 0.1439. Bertin (2016) used shot distance and shot angle as features to train a Logistic Regression model using nearly 50,000 shots and found an R-Squared error of 0.121. Eggels et al. (2016) used a dataset of 128,667 shots to build different xG models using a Logistic Regression, a Decision Tree, a Random Forest, and a Decision Tree boosted with ADA Boost. The features they selected to build these models include features derived from game phase (e.g., corner, free-kick, open-play, counter-attack), features derived from body part used to shoot the ball, shot distance, shot angle, features derived from the type of action preceding the shot (pass, carry, etc.), the EA Sports FIFA video-game ratings of player quality, and the EA Sports FIFA video-game ratings of goalkeeper quality. Their most performant model was a Random Forest xG model with a Precision of 0.785 and a Recall of 0.822. Rathke (2017) divided the pitch into zones and used information about the ball's location at the moment of shooting and shot outcomes to calculate the frequency of goals in each zone. He used a dataset including 18,218 shots to build his xG model. Herbinet (2018) achieved 0.509 accuracy using a Random Forest Classifier built from 88,340 shots and taking shot distance, shot angle, features derived from the type of shot, and the number of goals of each team at the moment of shooting as features. Tiippana (2020) used an xG model built from over 100,000 shots that were processed in a Poisson Regression. He achieved a MAE of 6.5 using features such as shot distance, shot angle, features derived from body part used to shoot the ball, features derived from game phases (e.g. open play, penalty, free kick, corner), features derived from possible errors committed in the scoring situation, number of dribbled players before the shot, and features derived from the game state (i.e. whether the sgame is even or whether the home or away team is leading). Pardo (2020) constructed several xG models and attained a Log Loss performance of 0.2536, using an XGBoost algorithm on a dataset of 87,161 shots. Pardo´s xG models were created using the following features: shot location x, shot location y, shot distance, shot angle, goal angle (which is the angle between the line passing through the position of the shooter and the center of the goal line, and the goal line), features derived from contextual variables of actions preceding the shot (such as corner, direct freekick, indirect freekick, pass to penalty box), and the player performance ratings in the EA Sports FIFA video game. Bransen and Davis (2021) utilized Generalized Additive Models (GAMs) and XGBoost Classifier, which considered features such as shot location x, shot location y, shot distance, shot angle, features derived from body part used to shoot the ball, features derived from type of assist, scoreline at the moment of the shot, and the moment in the match when the shot was taken. Their models were fed with 200,000 shots, and the Brier score was used to evaluate the performance of their models. The XGBoost they used achieved the best performance of 0.0863 Brier score. Umami et al. (2021) developed an xG model using a Logistic Regression and a dataset containing 32,000 shot events. They considered features such as shot location x, shot location y, shot distance, shot angle, and features derived from body part used to shoot the ball. Their model achieved a Recall of 0.9671 and a Precision of 0.1277. Anzer and Bauer (2021) utilized a dataset of 105,000 shot events to develop their models. They used features including the shot location x, shot location y, player speed at the time of shooting, the number of defenders in the shot's path, features derived from goalkeeper’s location, features derived from the type of pressure applied on the shooter, features derived from the type of the shot, a variable indicating how the shooter gained control of the ball before/when taking the shot, a variable indicating whether the shot followed a freekick, and a variable indicating whether the shot is a direct freekick. Anzer and Bauer (2021) employed multiple machine learning algorithms, including Logistic Regression, ADA Boost, Random Forest, and Gradient Boosting. Their Gradient Boosting model provided the best results, achieving a Recall of 0.181 and a Precision of 0.646. Cavus and Biecek (2022) used a dataset of 315,430 shots to train Random Forest, CatBoost, XGBoost, and LightGBM models. They considered features such as the minute of the shots, a variable indicating if the shooter is from home or away team, features derived from the action preceding the shot (pass, carry, etc.), shot distance, and shot angle. Their best model in terms of precision was a Random Forest xG model with a Precision of 0.922, which they achieved after using a ROSE random over-sampling method.

Although defender‘s influence on a shot is critical, only few researchers have incorporated this aspect into their xG models. Anzer and Bauer (2021) and Lucey et al. (2015) attempted to account for the presence of defenders by using discrete measures such as the number of defenders in front of the shooter and the shooter's proximity to the defenders. However, this approach does not accurately reflect reality as the real positioning and orientation of defenders can have varying impacts on the shot.

This study presents an innovative approach for building xG models that takes into account the precise position of players at the moment of the shot and the orientation of players in front of the shooter, through a new feature called the Kos Angle.

Methods

Football Shot Modelling

In the rest of this study, we will be considering a football pitch having L_p length and W_p width. The first parameter considered in our approach is the distance between the position of the ball during the shot, and the center of the goal line modeled in Figure 1.

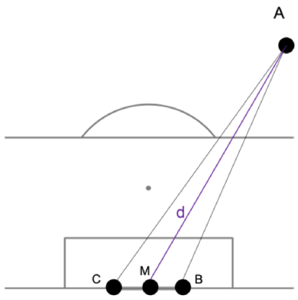


Figure 1: Shot distance, d

We will be using the following variables :

- A(x,y) to represent the ball location;
- B(L_p ,y_B) to represent the location of the right extremity of the goal box;
- C(L_p ,y_C) to represent the location of the left extremity of the goal box;
- M(L_p , $\frac{W_p}{2}$) to represent the location of the center of the goal box.

The distance parameter is calculated using the following formula (1):

$$d = \sqrt{\left(x - L_p\right)^2 + \left(y - \frac{W_p}{2}\right)^2} \tag{1}$$

The second parameter we will consider in our approach is the shot angle $\boldsymbol{\theta}$ between the position of the ball during the shot, and the extremities of the goal box (Figure 2**Error! Reference source not found.**).

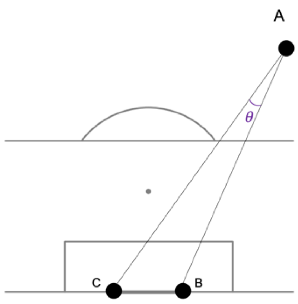


Figure 2: Shot angle, θ

Ghiyath al-Din Jamshid Mas'ud al-Kashi's formula (Pickover, 2009) is used to calculate the shot angle of a shot taken from a position A(x,y). We obtain the formula below (2):

$$\theta = Arccos\left(\frac{2 \times \left(x - L_p\right)^2 + (y - y_B)^2 + (y - y_C)^2 - (y_B - y_C)^2}{2 \times \sqrt{((x - L_p)^2 + (y - y_B)^2) \times (\left(x - L_p\right)^2 + (y - y_C)^2)}}\right) \tag{2}$$

The third parameter we will consider in our approach is the Kos Angle representing the shooting angle θ, from which we subtract the angles occupied by the players located inside θ.

For instance, in Figure 3, the Kos Angle can be observed as the blue area, which appears when two players are located inside the shot angle.

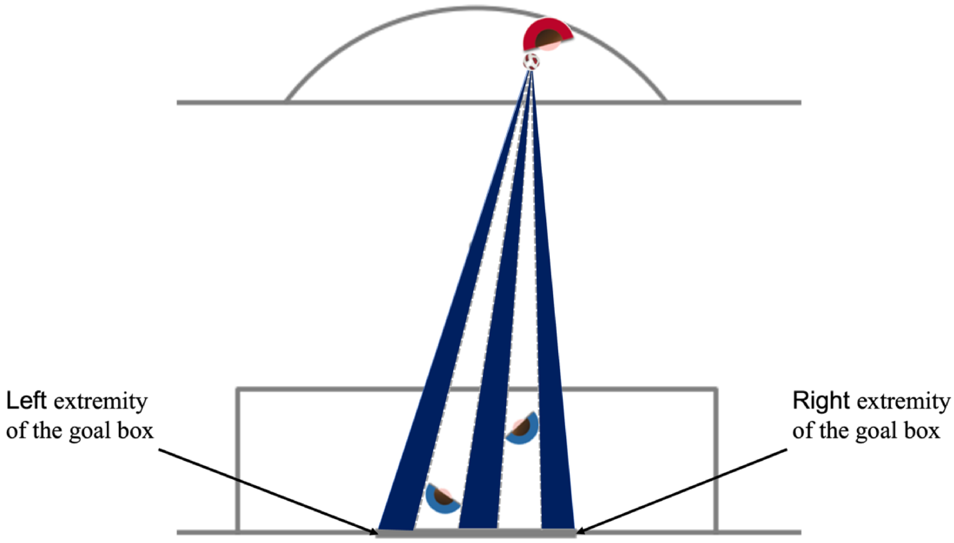


Figure 3: Kos Angle in the case of two players inside the shot angle

We consider n players P_i ($i \in [\![1, n]\!]$) inside the shot angle at the moment of the shot, (x_i, y_i) ($i \in [\![1, n]\!]$) the coordinates of the players P_i and α_i their orientation in relation to the line parallel to the goal line as shown in Figure 4.

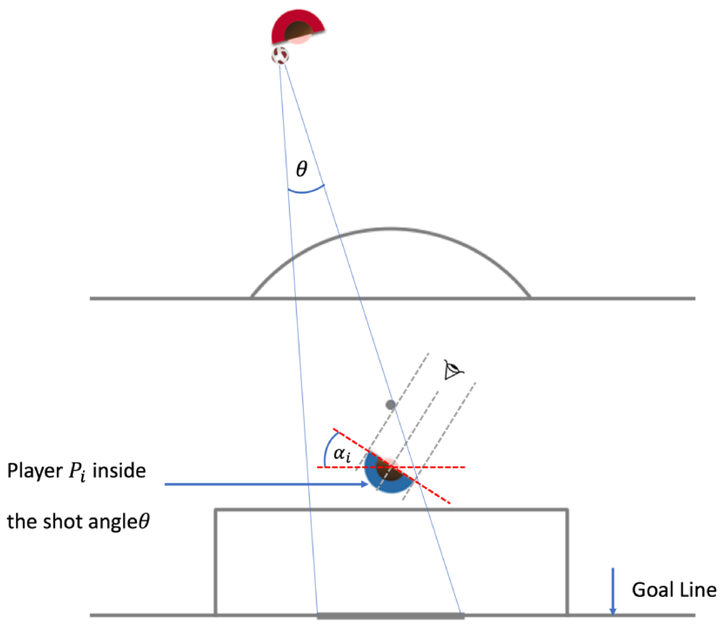


Figure 4: Orientation $\boldsymbol{\alpha_i}$ of the player $\boldsymbol{P_i}$ located inside the shot angle $\boldsymbol{\theta}$

Using the same formula of Ghiyath al-Din Jamshid Mas'ud al-Kashi (Pickover, 2009) for n players inside the shot angle, we obtain the following expression for the Kos Angle (3):

$$Kos\ Angle = Arccos\left(\frac{2\times\left(x-L_p\right)^2+(y-y_B)^2+(y-y_C)^2-(y_B-y_C)^2}{2\times\sqrt{((x-L_p)^2+(y-y_B)^2)\times(\left(x-L_p\right)^2+(y-y_C)^2)}}\right)$$
$$-\sum_{i=1}^{n} Arccos\left(\frac{\left[(x-x_i)^2+(y-y_i)^2-\left(\frac{l_i}{2}\right)^2\right]}{\sqrt{\left[\left(x-x_i-\sin(\alpha_i)\times\frac{l_i}{2}\right)^2+\left(y-y_i-\cos(\alpha_i)\times\frac{l_i}{2}\right)^2\right]\times\left[\left(x-x_i+\sin(\alpha_i)\times\frac{l_i}{2}\right)^2+\left(y-y_i+\cos(\alpha_i)\times\frac{l_i}{2}\right)^2\right]}}\right) \quad (3)$$

By utilizing the Surveyor’s formula (Braden, 1986), we developed a pre-processing algorithm that verifies whether a point resides within a triangle. This algorithm enabled us to ascertain whether each player's extremity lies within the shooting range. There are four potential outcomes: players outside the shooting range, players inside the shooting range, players with their right extremity inside the shooting range, and players with their left extremity inside the shooting range. According to each case, our algorithm calculates the Kos Angle.

We have considered fixed values of players widths: teammate’s width: $l_{\mathrm{TM}} = 0.4\mathrm{m}$, opponent goal keeper’s width: $l_{\mathrm{GK}} = 1.4\mathrm{m}$, opponent other player’s width: $l_{\mathrm{O}} = 0.5\mathrm{m}$. These values were chosen in collaboration with subject matter experts from FUS Football Club (*Moroccan First Division Club, Rabat*) according to multiple game observations and considering a two-dimensional modelling. Due to the absence of player’s orientation information in the shot dataset we utilized, we have made the assumption that all players within the shot angle were facing the ball and looking in its direction at the moment of the shot.

In addition to the three geometric features defined previously (distance, shot angle and Kos Angle), we have selected in collaboration with the technical staff of FUS Football Club and its Coach Mr Jamal Sellami (*season 2022-2023*), a list of other features that can influence the outcome of shots: location x of the shot, location y of the shot, under pressure (Boolean value indicating if the shot was made under opponent pressure), follows a dribble (Boolean value indicating if the shot followed a dribble or not), a bining of body part (the body part used to shoot the ball), a bining of shot techniques (the shot technique used when shooting the ball).

Data Preprocessing

We have formed a dataset composed of 68,382 shots using different leagues and competitions: African Cup of Nations, Botola Pro 1, Bundesliga, Ligue 1, NWSL, Premier League, UEFA Euro, World Cup Qualification (Asia, Europe and South America).

After binning body part and shot technique features, we calculated, for each shot, its distance to goal line, its shot angle θ, and its Kos Angle according to formulas (1), (2), and (3).

We divided finally the resulting dataset into training and test sets. The training data will be used to train the models, the test data will be used to evaluate the performances of our trained models. We have considered the following proportions: 80% of our data will be used as training data and 20% as testing data.

Expected Goals Calculation and Hypothesis

In the case of non-continuous random variable x, the expected value of x is:

$$E[x] = \sum_i x_i \times p_i \quad (4)$$

where x_i are the possible outcomes of the random variable x and p_i are their corresponding probabilities.

When applied on shots, the expected value of a shot is given by the formula (5):

$$E[x] = 1 \times p_{goal} + 0 \times \left(1 - p_{goal}\right) = p_{goal} \tag{5}$$

where p_{goal} is the goal probability of a shot.

Until now, we can mathematically consider that the expected value of a shot is exactly equal to its probability becoming a goal (this specificity is not present in other ball sports like basketball where different outcomes are possible for successful shots: 1 point, 2 points and 3 points). Then, to calculate goal probabilities, we have considered the hypothesis of Large Numbers (Sandon, 1946). This principle can be used to estimate goal probabilities from goal frequencies when having a large amount of data.

Choice of Machine Learning Models

After explaining the mathematical methodology for calculating xG values, we will now introduce the machine learning algorithms we have used to construct our xG models. Machine Learning Classifiers are commonly used in xG models due to the binary nature of shot outcomes: goal or non-goal.

Initially, we have opted to use only a Decision Tree Classifier for building our xG model due to its high interpretability. However, to achieve satisfactory performance with this type of classifiers, we often need to decrease the minimum number of samples in each leaf, which may violate the hypothesis of the Law of Large Numbers. Moreover, since a fixed xG value is assigned to each leaf, the model's outcomes are discrete and limited to the number of leaves in the trees. Finally, using Decision Tree Classifiers poses a high risk of overfitting the data. Therefore, we have decided to explore ensemble learning models in addition to Decision Tree Classifiers.

We have decided to utilize Random Forest and Gradient Boosting Classifiers as ensemble learning models in order to address three weaknesses of Decision Trees:

- The number of trees hyperparameter offered by these models enables us to expand the effective sample size by employing multiple trees, thereby fulfilling the requirement of the Law of Large Numbers. Consequently, our xG models becomes more reliable;
- These models have a stochastic nature that can generate a continuous model outcome, allowing for an almost infinite number of xG values;
- The likelihood of overfitting the data is reduced with these models.

Choice of Performance Measures

We have chosen various test measures, and each one serves a specific purpose:

- Precision: tests quality of positive predictions made by the models.
- Recall: tests the ability of the models' predicting goals.
- RMSE: tests the gap between the calculated xG values and real shot outcomes.
- Brier Score: evaluate the accuracy of our probabilistic predictions.

Results

We will consider the following notation:

- DTC: Decision Tree Classifier;
- RFC: Random Forest Classifier;
- GBC: Gradient Boost Classifier.

In the results section of this study, we will present the performance of our built models in two different scenarios. To begin with, we will present the performances obtained without incorporating the Kos Angle as a feature. Then, we will present the performances when incorporating the Kos Angle as a feature.

The primary objective of this study is to evaluate the impact of the use of Kos Angle as a feature on the performance of xG models. Hence, the progress of the models' performances from the scenario of not incorporating the Kos Angle to the one where it is utilized, will hold more significance than the models' overall performance.

Table 1 summarizes the final performance results of our trained models using the features described previously, excluding the Kos Angle.

Table 1: Performance of trained xG models, without using the Kos Angle as a feature

Model	Precision	Recall	RMSE	Brier Score
DTC	0.6104	0.0757	0.2840	0.0806
RFC	0.6991	0.0569	0.2834	0.0803
GBC	0.635	0.0916	0.2825	0.0798

NB: For Precision and Recall, values near 1 are the best, for RMSE and Brier Score values near 0 are the best

As a result, the Decision Tree Classifier showed a lower Precision compared to the other models, while the Random Forest Classifier had a higher Precision (69.91%) compared to the Gradient Boost Classifier (63.5%). However, the Gradient Boost Classifier demonstrated better performance in correctly predicting goals, with a higher Recall (9.16%) compared to the Random Forest Classifier (5.69%).

The observed low Recall values are attributed to data imbalance, as explained by Cavus and Biecek (2022). Specifically, only 9.97% of our training data consists of shots that result in goals, thereby limiting our model's exposure to this minority class during the training process.

After evaluating the performances of our models without considering the Kos Angle, we proceeded to include this angle as a feature and re-evaluate the same models with the same hyper-parameters. The outcomes of these models, incorporating the Kos Angle, are presented in Table 2.

Table 2: Performance of trained xG models, when using the Kos Angle as a feature

Model	Precision	Recall	RMSE	Brier Score
DTC	0.6766	0.0815	0.2824	0.0797
RFC	0.7608	0.0757	0.2806	0.0787
GBC	0.6812	0.1125	0.2806	0.0787

NB: For Precision and Recall, values near 1 are the best, for RMSE and Brier Score values near 0 are the best

A noticeable improvement in all performance metrics is observed, and Table 3 provides a summary of this enhancement (by the use of the Kos Angle).

Table 3: Impact of the Kos Angle on model's performances

Model	Precision	Recall	RMSE	Brier Score
DTC	+10.84%	+7.66%	-0.56%	-1.11%
RFC	+8.82%	+33.04%	-0.98%	-1.99%
GBC	+7.27%	+22.81%	-0.67%	-1.37%

NB: For Precision and Recall, positive percentages are increasing the performances, for RMSE and Brier Score negative percentages are increasing the performances

The impact of incorporating the Kos Angle varies across different models, and despite the general enhancement observed, particular emphasis should be placed on the substantial improvement in Recall, which has been identified as a weak component in existing football expected goals models Cavus and Biecek (2022). In other words, even with a limited number of goals in the training data, the inclusion of the Kos Angle improves the ability of our models to predict shots that result in goals. Consequently, this suggests that the xG values generated by our models will tend more to 1. To verify this trend, we will utilize SHAP values.

SHAP values estimate the impact of a feature on predictions. In our case, we will use SHAP values to observe the impact of Kos Angle on the two possible outcomes of shots: goals and misses.

In Figure 5, Figure 6 and Figure 7, for each feature, the number of points on the left indicates how well the feature aids in miss prediction by the model, the number of points on the right indicates how well the feature aids in goals prediction.

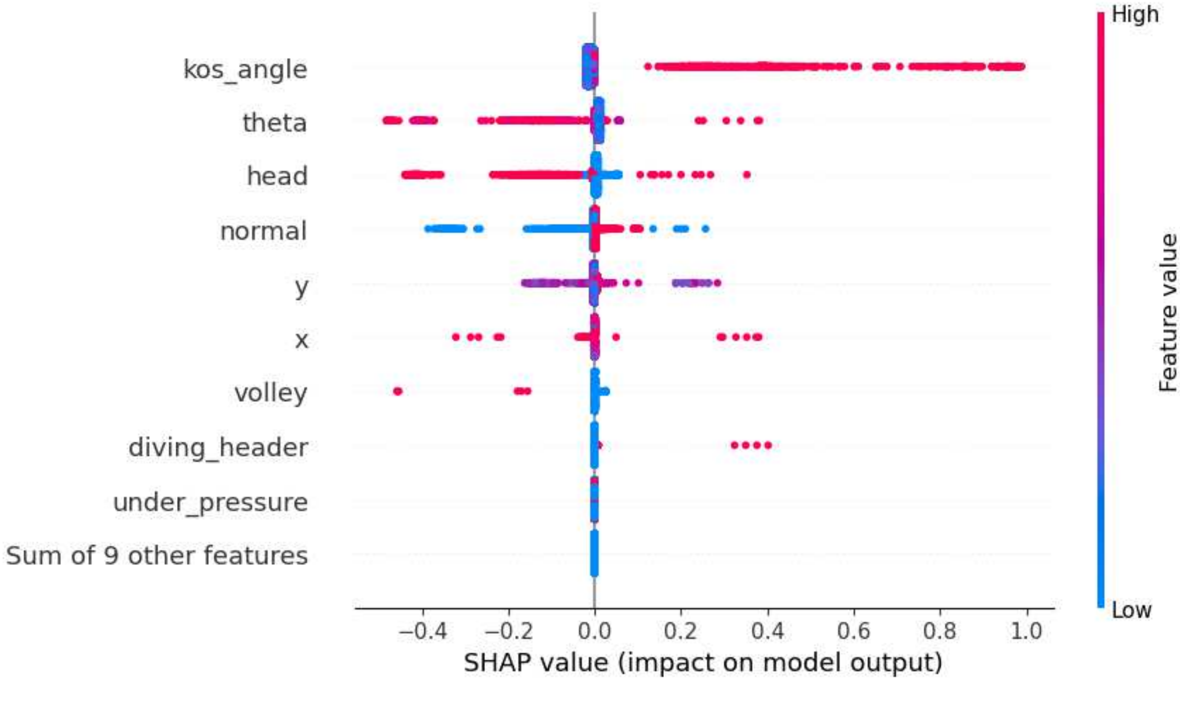


Figure 5: SHAP feature importance in the Decision Tree Classifier model (features are sorted from most important to less important)

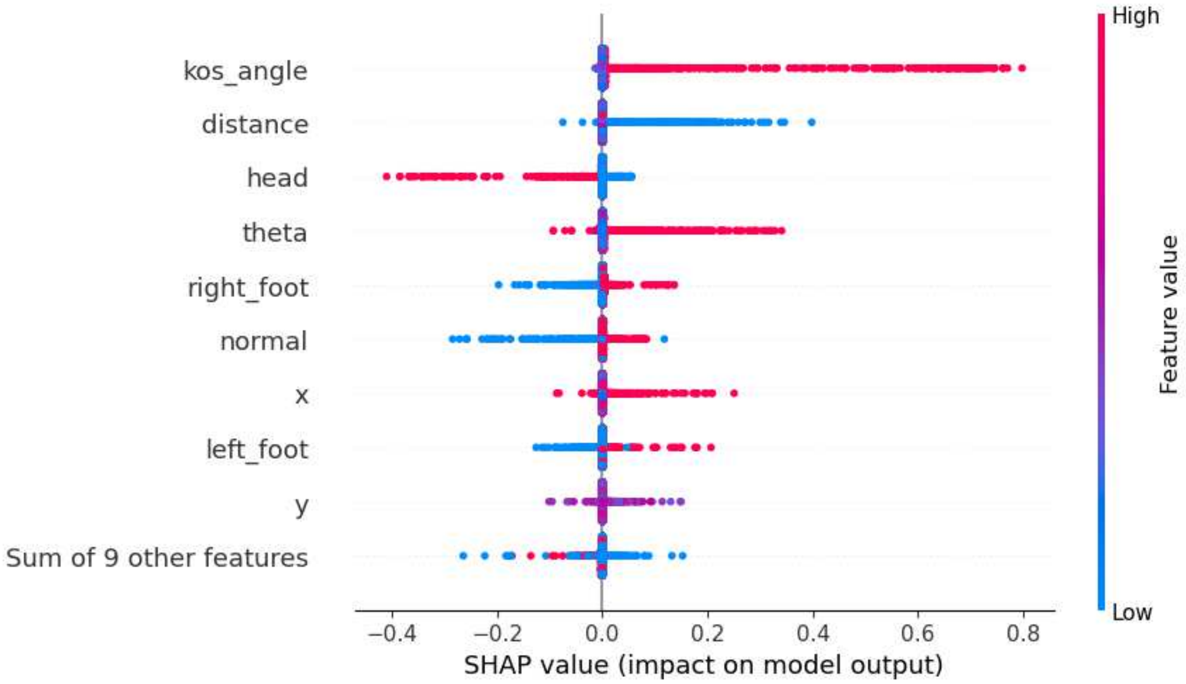


Figure 6: SHAP feature importance in the Random Forest Classifier model (features are sorted from most important to less important).

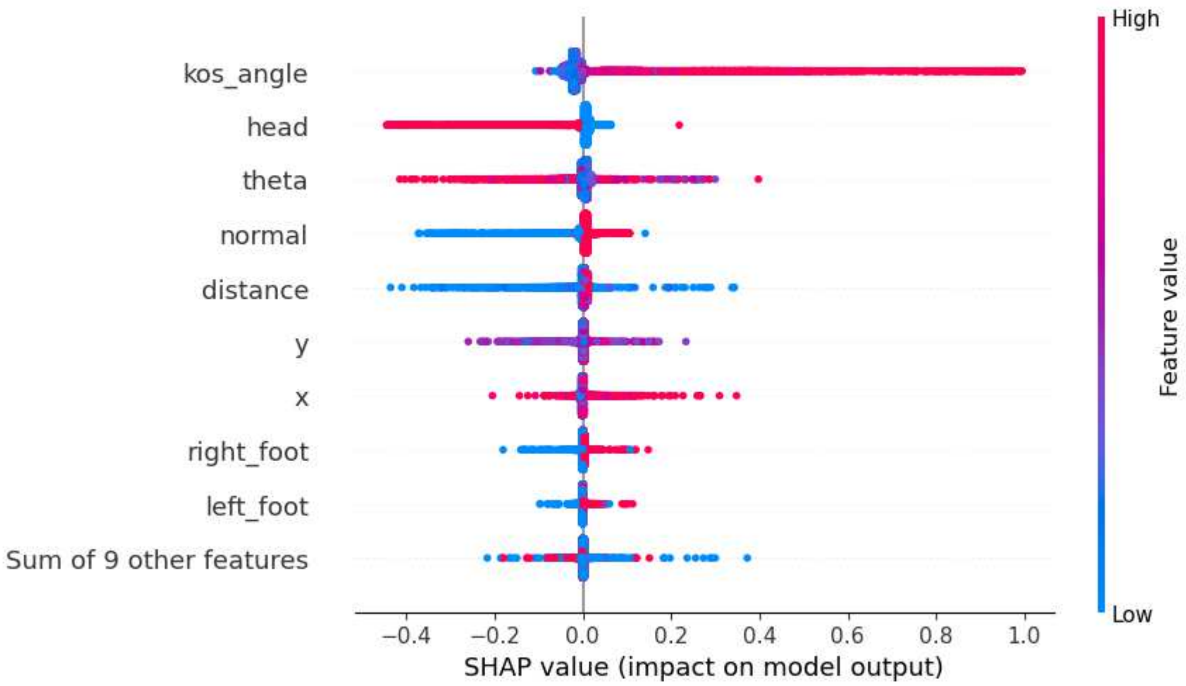


Figure 7: SHAP feature importance in the Gradient Boost Classifier model (features are sorted from most important to less important).

The impact of the Kos Angle on the model's forecast is greater than that of any other feature, as evidenced by the findings depicted in Figure 5, Figure 6 and Figure 7. Specifically, a low Kos Angle value (represented by blue dots) aids the model in predicting misses, while a high Kos Angle value (represented by pink dots) aids the model in predicting goals. Moreover, there exists a proportional relationship between the magnitude of the Kos Angle and the resulting xG values, such that larger Kos Angles lead to higher xG values, and vice versa for smaller Kos Angles.

The Kos Angle is usually the most crucial component in our models. This is also approved by Table 4:

Table 4: Feature importance in used models

Model	Kos Angle	Shot Angle	Distance	Other features
DTC	53.61%	18.43%	5.18%	22.78%
RFC	37.08%	18.35%	16.39%	28.18%
GBC	45.76%	17.75%	9.65%	26.84%

The Kos Angle accounts for 53.61% of our DTC model feature importance, for 37.08% of our RFC model feature importance, and for 45.76% of our GBC model feature importance. Comparatively, other features such as Shot Angle, Distance, and additional other features contribute to a lesser extent. These findings emphasize the significance of the Kos Angle in our models and its substantial influence on the forecasting outcomes.

Discussion

There are various possibilities for improving the accuracy of xG models, these include optimizing the hyperparameters of the machine learning algorithms used, carefully selecting the most relevant features for constructing these models, and utilizing larger shot databases. Since our dataset comprised only 68,382 shots, we have kept the focus on enhancing our models by introducing a new and improved method for shot modeling and fine-tuning the hyperparameters of the machine learning algorithms used.

Despite our access to a smaller dataset of shots, we believe that it would be valuable to compare the performance of our models to those of other studies that utilized larger and different datasets, such as Anzer and Bauer (2021) with 105,000 shots, Pardo (2020) with 120,000 shots, Eggels et al. (2016) with 126,000 shots, Bransen and Davis (2021) with 200,000 shots, and Cavus and Biecek (2022) with 315,000 shots.

We have selected the most performant models in the studies that have used at least one similar performance measure to the ones we have used, and found out that:

- Our GBC Model achieved a Precision of 68.12%, surpassing that of Anzer and Bauer (2021): 64.60%. Nonetheless, their Recall :18.1% exceeds ours :11.25%. Unlike Umami's model, which emphasizes a high Recall of 96.71% at the expense of Precision: only 12.77%, Anzer and Bauer's xG model maintains a balanced Precision/Recall ratio.
- With a Brier Score of 0.0787, our GBC model is more performant than Bransen and Davis (2021) according to this measure.
- Eggels et al. (2016) outperformed our model, achieving a Precision of 78.5% and a Recall of 82.2%, partly because of their use of a much larger dataset containing 128,667 shots, while our dataset consisted of only 68,382 shots. On the other hand, the model developed by Cavus and Biecek (2022) has by far the best performance, with a Precision of 92.1% and a Recall of 95.5%. These remarkable results are probably due to the balancing technique they utilized, and the 315,430 shots they have employed.

Our xG modeling strength does not rely only on the quantity of data used or balancing techniques employed. Rather, it primarily derives from incorporating the effective location and a hypothetical orientation of players. With the growing demand for football data, the integration of 3D tracking technologies is imminent. We are confident that incorporating 3D geospatial data to calculate the KOS Angle will introduce a new dimension and significantly enhance the performance of our model.

Conclusion

This research paper emphasizes the importance of integrating a realistic feature called the Kos Angle to optimize the performance measures of xG models.

Through this study we developed a general mathematical formula for the Kos Angle, and used it to observe the impact of this feature on different machine learning models we have built.

We observed a general enhancement in all the performance measures we have selected, especially in the Recall which is a weak component in football xG models. It also turned out that the Kos Angle is the most important feature in all our models, as can be noticed by the distribution of the SHAP values we get.

This is a good demonstration of how thorough mathematical modelling of real phenomena can positively impact the performance of machine learning models.

References

Tippett, J. (2019). The Expected Goals Philosophy: A Game-Changing Way of Analysing Football. Independently published.

Bertin, M. (2016). Why Soccer's Most Popular Advanced Stat Kind Of Sucks. Récupéré sur Deadspin: https://deadspin.com/why-soccers-most-popular-advanced-stat-kind-of-sucks-1685563075

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. (2015). "Quality vs Quantity": Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data. 9th Annual MIT Sloan Sports Analytics Conference.

Eggels, H., Van Elk, R., & Pechenizkiy, M. (2016). Explaining soccer match outcomes with goal scoring opportunities predictive analytics. Proceedings of the Workshop on Machine Learning and Data Mining for Sports Analytics 2016 co-located with the 2016 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Riva del Garda, Italy.

Herbinet, C. (2018). Predicting Football Results UsingMachine Learning Techniques. MEng, Imperial College Of Science, Technology and Medicine.

Tiippana, T. (2020). How accurately does the expected goals model reflect goalscoring and success in football? Bachelor’s Thesis, Aalto University.

Pardo, M. (2020). Creating a model for expected goals in football using qualitative player information. Master’s thesis, Universitat Politècnica de Catalunya.

Bransen, L., & Davis, J. (2021). Women’s football analyzed: interpretable expected goals models for women. AI for Sports Analytics (AISA) Workshop at IJCAI 2021.

Pickover, C. (2009). The Math Book: From Pythagoras to the 57th Dimension, 250 Milestones in the History of Mathematics. Sterling.

Cavus, Mustafa; Biecek, Przemyslaw. (2022). Explainable expected goal models for performance analysis in football analytics. 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA). Shenzhen, China. doi:10.1109/dsaa54385.2022.10032440

Anzer, G., & Bauer, P. (2021). A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). Frontiers in sports and active living, 3. doi:10.3389/fspor.2021.624475

Braden, B. (1986, September). The Surveyor's Area Formula. The College Mathematics Journal, 17(4), 326-337. doi:10.1080/07468342.1986.11972974

Rathke, A. A. (2017, August 29). An examination of expected goals and shot efficiency in soccer. 12(Proc2). doi:10.14198/jhse.2017.12.proc2.05

Sandon, F. (1946, July). An experimental introduction to the theory of probability. By J. E. Kerrich Pp. 98. Danish Cr. 8·50. 1946. (Munksgaard, Copenhagen). The Mathematical Gazette, 30(290), 178-179. doi:10.2307/3609116

Umami, I., Gautama, D. H., & Hatta, H. (2021). implementing the Expected Goal (xG) model to predict scores in soccer matches. IJIIS: International Journal of Informatics and Information Systems, 4(1), 38-54. doi:10.47738/ijiis.v4i1.76