**Bozhko Oleksandr**

postgraduate student

«Kharkiv National University of Radio Electronics», Ukraine

**Petrov Konstantin**

Dr. Sc., Prof., Head of Information Control Systems Department

«Kharkiv National University of Radio Electronics», Ukraine

**Using expected goals models for assessing the quality of soccer team play**

Expected Goals (xG) models are an interesting example of applying machine learning and statistical analysis techniques to predict soccer match results based on historical data. These models use detailed data about each kick, including its location, context, and outcome, to predict the probability of a goal. This provides a more accurate and informative assessment of game performance based on the probability of scoring goals under certain conditions than traditional metrics such as goal counts or shots on goal. The xG approach has proven to be particularly effective in predicting sports performance, providing a significant improvement in prediction accuracy over traditional methods.

Today, xG models are widely used not only by the coaching staffs of top soccer clubs, but also by sports analysts, bookmakers and even television commentators. Expected goals provide analysts with an advantage that can help in decision making at both the sporting and business levels of soccer. Not only can they help improve the on-field success of soccer clubs through tactical analysis of player and team performance, but they can also help in financial situations such as player acquisitions and contract negotiations [1].

This widespread use confirms the effectiveness and applicability of this approach.

The paper intends to analyze how the xG methodology can be adapted for application in different subject areas. The main objective is to investigate how the methods of data collection, feature selection and engineering, handling unbalanced data and using machine learning algorithms applied in xG models can be transferred to other subject areas requiring forecasting based on historical data.

Expected goals is a statistical measure that is used to assess the quality of shots in soccer. This measure was developed as an alternative to traditional statistical measures such as number of shots or possession, which may not always accurately reflect the performance of a team or player.

In game sports, especially soccer, where goals are scored less frequently and match scores can include a significant amount of randomness and may not always accurately reflect the game, it is important to have additional metrics to evaluate the performance of teams and individual players. Expected Goals allow further insight than just a scoreline [2].

xG models use detailed data about each kick, including its location on the field, context (e.g., whether it was a direct kick or after a pass), and other factors to predict the probability that the kick will become a goal. In this way, xG allows for a more sophisticated and in-depth analysis of the game.

Various approaches to xG modeling include logistic regression, gradient bousting, neural networks, support vector methods, and tree-based classification algorithms. Most of the features included in these models are generated from game event data, divided into two sections, events and positioning [3]. The choice of these features, as well as the size and values of the data used to train the model, can significantly affect its performance. For example, a model trained on event data from top-tier leagues may not perform as well when applied to data from lower-tier leagues, where the style of play and quality of players can vary significantly. Therefore, it is important to carefully select and tune model parameters to maximize model performance and applicability.

This study intends to use several open-source machine learning models that have been developed to calculate expected goals in soccer, as well as to develop our own models. This will allow for a thorough analysis and comparison of different approaches to building xG models.

In addition to using existing models, the paper considers the possibility of creating our own original open-source models based on them. This task involves enhancing existing models by adding new features, optimizing parameters, or applying new machine learning algorithms. The goal of these enhancements is to improve the predictive performance of the models and/or their interpretability, as well as to extend their applicability to other subject areas.

One of the key advantages of xG model research is the large amount of data available for analysis. Soccer, being one of the most popular sports in the world, generates a huge amount of data that is recorded and stored in every match. This data includes information about each kick, its location, context, and result, as well as many other parameters.

Most of this data is publicly available, making it easy to use for research purposes. This allows for thorough and detailed analysis using real match data and provides a rich basis for training and testing machine learning models.

Importantly, the availability of these data also contributes to the transparency and reproducibility of research findings.

Overall, the large amount of data available makes xG models an ideal subject to explore, allowing a deep dive into the analysis and modeling of this interesting and important aspect of soccer analytics.

The Expected Goals model, originally developed for analyzing soccer data, is an example of applying machine learning to predict outcomes based on historical data. This approach can be adapted for use in a variety of subject areas where prediction based on historical data is required. However, adapting the xG model to new subject areas requires careful selection and tuning of model features and parameters to maximize its performance and applicability.

An important first step is an in-depth study of the subject area in which the model is to be applied. This includes understanding the key variables and factors that may influence the predicted outcome, as well as the types of data that are available for use. Based on this understanding, the functions (or variables) that will be used to train the model should be selected. These functions should be relevant and informative to the predicted outcome.

The data used to train and test the model must be carefully prepared. This may include cleaning the data, transforming the data, and possibly creating new features from existing data. After that, you should select the machine learning model that best suits the task and tune its parameters to maximize its effectiveness.

Once the model has been trained, it is important to carefully evaluate its performance and perform validation using delayed or cross-validation datasets. This will help ensure that the model is generalizing to the data and not overtraining. Finally, it is important to interpret the results of the model and understand which features are most important to the prediction and how they affect the predicted outcome. This can help to draw informative conclusions regarding the importance of features and their impact on the outcome and understand how the model can be improved.

In general, adapting an xG model to new subject areas is a complex process that requires careful analysis and numerous experiments. However, due to the vast amount of available data and the transparency of open-source models [4, 5], this process provides a unique opportunity for deep theoretical research and new practical results.

Several key outcomes are expected to be achieved during the study. First, an in-depth study of the mechanisms of the xG model and an investigation of its application in the context of soccer is envisaged. This will include analyzing the various functions and parameters used in the model and their impact on the predicted outcome.

The second important outcome will be the adaptation of the xG model to the new subject areas. This will involve selecting and customizing model features and parameters for each new subject area and preparing and processing data for training and testing the model. It is expected that this process will lead to new, effective prediction models for different subject areas.

Another outcome will be to evaluate the performance of these new models. This will involve rigorously testing the models on delayed or cross-validation datasets and interpreting the results to understand the importance of different features and their impact on the predicted outcome.

However, the study may identify potential problems and limitations of this approach, as well as possible directions for further development of the model. Overall, the results of the study are expected to lead to the development of new, effective prediction models for various subject areas.

**List of References:**

1. James Mead , Anthony O’Hare , Paul McMenemy. Expected goals in football: Improving model performance and demonstrating value. April 5, 2023. <https://doi.org/10.1371/journal.pone.0282295>
2. James H. Hewitta, Oktay Karakus. A Machine Learning Approach for Player and Position Adjusted Expected Goals in Football (Soccer). arXiv:2301.13052v2 [cs.LG] 2 May 2023. <https://arxiv.org/pdf/2301.13052.pdf>
3. Andrew Jaegle, Felix Gimeno, Andrew Brock, Andrew Zisserman, Oriol Vinyals, Joao Carreira. Perceiver: General Perception with Iterative Attention. arXiv:2103.03206v2 [cs.CV] 23 Jun 2021. <https://arxiv.org/pdf/2103.03206.pdf>
4. Pieter Robberechts. A Python package for training and analyzing expected goals (xG) models in soccer. <https://github.com/ML-KULeuven/soccer_xg/blob/master/README.md>
5. David Sumpter. Soccer analytics. <https://github.com/CleKraus/soccer_analytics/blob/master/README.md>