ATATÜRK ÜNİVERSİTESİ YAYINLARI
ATATURK UNIVERSITY PUBLICATIONS

Research Article *Araştırma Makalesi* DOI: 10.62425/rses.1943506

Predicting Football Shot Outcomes in the Turkish Super League: Support Vector Machine and Naive Bayes Models with Feature Importance Analysis

Türkiye Süper Ligi'nde Futbol Şut Sonuçlarının Tahmini: Destek Vektör Makineleri ve Naive Bayes Modelleri ile Özellik Önem Analizi

Muhammed Nasuhi GENÇ[1]
[1]İstanbul University, School of Business, Department of Quantitative Methods, İstanbul, Türkiye

Mustafa CAN[1]
[1]İstanbul University, School of Business, Department of Quantitative Methods, İstanbul, Türkiye

Neslihan FİDAN KEÇECİ[1]
[1]İstanbul University, School of Business, Department of Quantitative Methods, İstanbul, Türkiye

ABSTRACT

This study presents a modeling framework that formulates goal prediction as a binary classification problem (1 for a goal, 0 for no goal), based on 46,238 shot records from the Turkish Super League between the 2020-2025 seasons. In this study, Naive Bayes (NB) and Support Vector Machine (SVM), which are supervised machine learning algorithms, were used for prediction models. Considering the substantial class imbalance between goal and non-goal outcomes in the original dataset, a balanced sample was preferred to ensure a more balanced representation of the classes in the modeling process. The results of machine learning models created for a balanced dataset of 10,081 shots were compared, and the findings showed that SVM models (AUC 0.830 for RBF SVM and 0.824 for Linear SVM) discriminated between goals and non-goals more consistently than the NB models (Gaussian NB 0.801 and Complement NB 0.757). Feature importance analyses based on Permutation Importance and SHAP identified distance to goal as the dominant predictor, followed by the length of the preceding pass, while categorical shot descriptors carried substantially less weight. By combining a comparison of two algorithm families with model-agnostic explainability methods, this study provides league-specific evidence on which shot characteristics distinguish goals from non-goals and offers a transparent framework that analysts working in this league can apply directly.

Keywords: Shot outcome prediction, Supervised machine learning, Binary classification, Class imbalance, Sports analytics, Explainable AI (XAI), SHAP

ÖZ

Bu çalışma, 2020-2025 sezonları arasında Türkiye Süper Ligi'nde gerçekleşen 46.238 şut kaydına dayalı, gol tahminini ikili sınıflandırma problemi (gol için 1, gol olmaması durumu için 0) olarak formüle eden bir modelleme çerçevesi sunmaktadır. Çalışmada, tahmin modelleri için Naive Bayes ve Destek Vektör Makinesi, denetimli makine öğrenmesi algoritmaları, kullanılmıştır. Orijinal veri setindeki gol ve gol dışı durumlar arasındaki belirgin sınıf dengesizliği dikkate alınarak, modelleme sürecinde sınıfların daha dengeli temsil edilmesini sağlamak amacıyla dengeli bir örneklem tercih edilmiştir. 10.081 şutluk dengeli veri seti üzerinde oluşturulan makine öğrenmesi modellerinin sonuçları karşılaştırılmış ve bulgular, SVM modellerinin (RBF SVM için AUC 0.830; Linear SVM için 0.824) gol ile gol olmayan şutları NB modellerine (Gaussian NB 0.801; Complement NB 0.757) kıyasla daha tutarlı biçimde ayırt ettiğini göstermiştir. Permutation Importance ve SHAP ile yapılan değişken önem analizleri, kaleye olan mesafenin en baskın belirleyici, şuttan önceki pasın uzunluğunun ise ikinci sıradaki değişken olduğunu; kategorik şut değişkenlerinin belirgin biçimde daha az ağırlık taşıdığını göstermiştir. İki algoritma ailesinin karşılaştırılmasını model bağımsız açıklanabilirlik yöntemleriyle birleştiren bu çalışma, gol ile gol olmayan şutları hangi özelliklerin ayırdığına dair lige özgü kanıt sunmakta ve bu ligde çalışan analistlerin doğrudan uygulayabileceği şeffaf bir çerçeve önermektedir.

Anahtar Kelimeler: Şut sonucu tahmini, Denetimli makine öğrenmesi, İkili sınıflandırma, Sınıf dengesizliği, Spor analitiği, Açıklanabilir yapay zekâ (XAI), SHAP

Geliş Tarihi/Received 03.05.2026
Kabul Tarihi/Accepted 31.08.2026
Yayın Tarihi/Publication Date 30.09.2026

Sorumlu Yazar/Corresponding author: Neslihan FİDAN KEÇECİ
E-mail: neslihan@istanbul.edu.tr
Cite this article: Genç, M. N., Can, M., & Fidan Keçeci, N. (2026). Predicting football shot outcomes in the Turkish Super League: Support vector machine and naive Bayes models with feature importance analysis. Research in Sport Education and Sciences, 28(3), 176–193.

Content of this journal is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License.

Introduction

Machine learning (ML) has gained significant importance in sports analytics in recent years, both for performance evaluation and for developing prediction models. ML algorithms can generally be divided into two categories: supervised and unsupervised. Prediction models are also usually created using supervised ML algorithms. Recent developments in football analytics have increasingly focused on predicting shot outcomes using supervised machine learning techniques. In most studies, the prediction problem is formulated as a binary classification task in which each shot is labeled as either a goal (1) or a non-goal (0). Spatial variables such as shooting angle and distance to goal are generally considered the most influential predictors, while contextual variables including body part, assist type, set-piece situations, and match conditions are also frequently incorporated into predictive models.

One of the influential studies in this area was conducted by Lucey et al. (2015), who examined shot quality using spatiotemporal features derived from football tracking data. Using player tracking data, they showed that defender proximity, defensive structure and features describing the build-up all contribute to the likelihood of a shot resulting in a goal, beyond shot location alone. Similarly, Eggels (2016) approached shot prediction as a classification problem and evaluated the contribution of contextual variables such as body part, game situation, and set-piece conditions. His findings indicated that combining spatial and contextual features improves classification performance compared to models relying solely on shot distance. Anzer and Bauer (2021) further extended this line of research by integrating synchronized positional and event data into a goal-scoring probability framework. Their study employed supervised learning approaches to classify shot outcomes and highlighted the importance of variables such as shooting angle, distance to goal, crosses, and player positioning. Comparing their model against one trained on event data alone, the authors found that adding positional information improved discrimination between goals and non-goals, indicating that goalkeeper and defender positioning carry information event data cannot supply.

Machine learning algorithms have also been widely applied in football prediction studies beyond shot analysis. Tax and Joustra (2015) investigated several classification algorithms, including Support Vector Machines (SVM) and Naive Bayes (NB), in football-related prediction tasks using publicly available match data. Similarly, Hubáček et al. (2019) employed Gradient Boosting trees to predict football outcomes using relational match data and emphasized the importance of feature engineering in improving predictive performance. Many studies in the sports analytics literature address classification or regression problems using supervised ML algorithms. Pietraszewski et al. (2025) and Li and Hao (2025) demonstrate that classification, regression, and deep learning methods processing sports data are widely used in various decision support processes. These methods have made more objective performance assessment attainable even in tactically complex sports such as football. Scholtes and Karakuş (2024) show that positional effects diminish once richer shot descriptors are included, while player effects persist using a Bayes expected goal model. Fairchild et al. (2018) reveal that shot location is the primary determinant in goal prediction through fractal dimension-based spatial analysis using Major League Soccer (MLS) data. Mead et al. (2023) demonstrate that an expanded feature set and improved logistic regression approach using data from the five major European leagues provide higher prediction accuracy compared to traditional expected goal models.

Hewitt and Karakuş (2023) reveal that variable effects differ by position using SHAP in ML models. Iapteff et al. (2025) used a Bayesian generalized linear mixed-effects model in which team-level random effects capture differences between clubs in how shot angle and defensive pressure translate into scoring probability and interpreted individual predictions with SHAP. Davis and Robberechts (2024) emphasize the importance of fairness-based corrections by demonstrating that calibration and group-based biases in expected goal models can distort player performance evaluations. These studies demonstrate that supervised classification methods can effectively model football-related event outcomes when supported by appropriate feature engineering.

Moreover, the effectiveness and predictive accuracy of models depend heavily on how well they manage the scarcity of the target variable. Football shot datasets are naturally imbalanced because only a small proportion of shots result in goals. It becomes evident that imbalanced class distributions such as goal/non-goal are common in sports data. Natural class imbalance between goal and non-goal situations is a significant statistical problem. When events such as goals represent a small proportion of total observations, ML models tend to bias toward the majority class (non-goal shots) and fail to adequately learn the minority class. This situation can lead to misclassification of critical events and misinterpretation of model performance, despite seemingly high overall accuracy. Beyond model interpretability and variable effects, the class imbalance in sports analytics is a fundamental challenge to address to achieve high-performance modeling, particularly in modeling of rare events (goals, red cards, etc.).

Anzer and Bauer (2021) emphasized that goal instances represent a minority class in football shot prediction data and therefore evaluation based solely on accuracy may produce misleading results. Similarly, Senn (2024) highlighted that imbalanced shot datasets require the use of performance metrics such as Precision, Recall, F1-score, and ROC-AUC to evaluate classification performance more reliably. In addition, studies focusing on football shot classification have applied imbalance-handling approaches including class weighting and oversampling techniques such as SMOTE to improve minority-class prediction performance. These findings suggest that class imbalance should be considered when developing supervised machine learning models for football shot outcome prediction.

Despite this growing body of work, shot outcome modeling has been developed almost exclusively on data from the top European leagues, and evidence from the Turkish Super League remains scarce. This matters, since defensive organization and playing style differ across leagues, and shot patterns may not transfer. A second gap is methodological. Studies that address class imbalance often do so without examining how the resulting models arrive at their predictions, while studies that emphasize interpretability rarely make the imbalance problem explicit. The present study addresses both issues using five seasons of shot records from the Turkish Super League. We construct a balanced dataset, compare a geometric classifier (SVM) with a probabilistic one (NB), and interpret the resulting models through Permutation Importance and SHAP. The result is a set of models whose behaviour can be inspected variable by variable, rather than a black-box score, developed on data from a league that has so far received little attention.

In our study, creating a balanced sample is a crucial step in comparing two algorithm families under a common class distribution. Accordingly, we will perform our analyses using supervised learning algorithms, specifically SVM and NB, on the balanced dataset we have created. While gradient boosting algorithms and deep learning models stand out in the current sports analytics literature with their high prediction accuracy, our preference for SVM and NB algorithms in this study is based on specific theoretical and practical reasons. SVM constructs a decision boundary in a high-dimensional space and can model complex relationships through kernel functions, while Naive Bayes, despite its assumption of feature independence, offers a probabilistic framework with low computational cost. The comparison of these two algorithms was conceived to observe the characteristics of both geometric (SVM) and probabilistic (NB) approaches in analyzing the dynamic structure of football.

The following section provides a review of the literature regarding the predictive capabilities and application areas of these algorithms in football analytics.

Support Vector Machine (SVM) and Naive Bayes (NB) Algorithms for Sports Analytics Data

SVM is a ML method that offers strong classification performance in high-dimensional datasets and can model non-linear relationships through different kernel functions. SVMs are suitable for football shot classification because they are effective in handling high-dimensional feature spaces and nonlinear decision boundaries. Shot outcome prediction involves complex interactions among spatial and contextual variables such as shooting angle, distance, body part, and match situation. SVM models can capture these relationships efficiently, particularly when kernel functions are applied. In sports analytics

literature, SVM has been primarily applied in areas such as match outcome prediction and event detection through sensor data. In football, there are studies using SVM as a method for both predicting match results and detecting sensor-based injuries and head trauma.

One of the early studies on SVM in football was conducted by Igiri (2015), in which a Gaussian kernel-based SVM model was trained on a limited set of match data, reaching approximately 53.3% accuracy. The study demonstrates that despite SVM's theoretical strength, small sample sizes and limited feature sets significantly constrain performance. Ding (2020) compares the performance of different kernel functions on World Cup matches, highlighting the risk of overfitting in tournament-based small datasets and emphasizing that broader league-level datasets are more suitable for SVM. In sensor data-based applications, Wu et al. (2018) employ SVM to detect head impacts in American football using biomechanical features such as acceleration and angular velocity, demonstrating with high accuracy values that the method is a strong choice for high-dimensional data structures requiring non-linear separation. SVM has also been tested in various prediction problems in basketball and other sports. Kahraman (2022) compares ten algorithms for match outcome prediction in the NBA and the Turkish Basketball League across three seasons, reporting that accuracy varies considerably between seasons and that logistic regression and Gaussian NB outperformed more complex models such as Random Forest and gradient boosting, which the author attributes to the limited size of the training data. Guan (2025) proposes a hybrid model combining XGBoost and improved SVM for image processing-based features in basketball; finding that this structure, where SVM is used as the final classifier, provides higher accuracy compared to SVM alone. Studies in the literature show that SVM is effective in high-dimensional and non-linear datasets but exhibits sensitivity to overfitting and data quality in small datasets.

NB is another appropriate method for shot outcome prediction due to its computational simplicity and strong performance in probabilistic classification tasks. NB is a frequently used classification method across various domains due to its probabilistic foundation and computational efficiency. Although the algorithm assumes conditional independence among predictors, NB often performs well in practice even when this assumption is partially violated. Football event data typically include a mixture of categorical and numerical variables, making NB a practical baseline classifier. Furthermore, NB models are computationally efficient and easy to interpret, which makes them useful for comparing predictive performance against more complex machine learning algorithms. In football literature NB is primarily utilized as a supplementary model for match outcome prediction. Existing studies predominantly focus on macro-level problems such as match outcome prediction; the use of NB in shot-based models is quite rare.

Among the studies using Gaussian NB algorithm in football, Venkatachalam et al. (2023) are prominent. The researchers developed a model predicting match outcomes (win/draw/loss) using historical match statistics and league table data, reporting approximately 85.4% accuracy. The study reveals that despite the model's simple structure, it can demonstrate competitive performance, particularly showing that low computational cost methods can be effective in certain football prediction problems. Seyrek et al. (2021) built a Weka-based decision support system that runs eight classifiers, including Naive Bayes and SVM, on the same match-statistics dataset and selects the best performer automatically. Rudrapal et al. (2020) report a more direct comparison on 11,400 English Premier League matches: a multilayer perceptron performed best (73.57% accuracy), but Gaussian NB (65.84%) outperformed SVM (58.77%) by a clear margin. The relative ranking of the two families therefore appears to depend on the feature set rather than being fixed. Sharma et al. (2022) achieved higher accuracy and better ROC performance compared to classical NB by incorporating correlation-based weighting in their NB approach when predicting outcomes of various sports competitions; this result indicates that NB can be advantageous in datasets containing few but informative variables. Burhaein et al. (2024) employ NB to classify physical fitness levels in child and adolescent athletes, finding that the method offers both high accuracy and low computational cost.

Using both SVM and NB allows the study to compare two fundamentally different classification approaches. While SVM focuses on maximizing class separation through optimal decision boundaries, Naive Bayes estimates class probabilities based

on feature distributions. Comparing these models may provide additional insight into how different machine learning approaches handle football shot outcome prediction. After a model is built using these two supervised algorithms, the most important variables influencing shot outcome prediction can be examined based on the model outputs. Beyond prediction, model outputs can be used to rank the variables that explain goal and non-goal outcomes, which is the aim of the Explainable Artificial Intelligence approach.

In this study, after creating the supervised ML models, we will evaluate the importance of features and the key variables driving predictions. The following section examines how the current sports analytics literature addresses this interpretability.

Feature Importance Analysis and Explainability: Permutation Importance, SHAP, and Interpretability in Sports

As ML models spread through sports analysis, knowing why a model predicts a goal has become as important as the prediction itself; this is why model-agnostic methods such as Permutation Importance and SHAP have been widely adopted. Hewitt and Karakuş (2023) reveal the marginal effects of variables such as shot distance, angle, and shot type using SHAP and partial dependence plots in player and position-adjusted expected goal models; noting that these visualizations serve as position-based decision support tools for coaches. Mead et al. (2023) rank the contribution of additional variables (e.g., psychological proxies, team and player quality) in an extended expected goals model using odds ratios and SHAP, finding that their influence varies substantially: player value ranked among the top features in most leagues, while team form contributed little. Scholtes and Karakuş (2024) interpret player and position random effects in Bayesian hierarchical expected goal models through posterior distributions of group-level adjustments, providing both shot-level and player-level explainability. Similarly, Iapteff et al. (2025) apply SHAP to individual shots in a Bayesian mixed-effects model, using waterfall plots to show how each feature contributed to a given prediction and to diagnose poorly predicted shots. Cavus et al. (2026) propose glocal explanations, an intermediate level between local and global, by aggregating SHAP values and ceteris-paribus profiles across the shots of a given player or team rather than explaining single shots or the full dataset. SHAP and Permutation Importance are also extensively used in other sports. Ou-Yang et al. (2025) decompose the contribution of team and player performance to match outcomes in basketball using SHAP, showing coaches which statistics have the strongest effect on winning probability. Kim and Park (2025), in their study analyzing shooting performance in modern pentathlon, compare Mean Decrease Impurity, Permutation Importance, and SHAP for XGBoost; noting that Permutation Importance reduces certain biases while SHAP provides more detailed interpretation at global and local levels. Moustakidis et al. (2023) analyze team performance in football using SHAP and establish a model-based hierarchy of the statistics contributing most to team success. These studies demonstrate that explainability methods are what turn model outputs into information a technical staff can act on.

Methods

Data

The dataset used in this study consists of all shot records from matches in the Turkish Super League between the 2020-2025 seasons. The dataset was obtained for research purposes from the global sports data provider Opta (Stats Perform), facilitated by its Turkey office. The original dataset contains 46,238 shots, of which 5,081 resulted in goals and 41,157 were non-goals.

The target variable in this study is goal prediction. Therefore, the target variable is expressed as a categorical variable of 1 for situations resulting in goals and 0 for situations not resulting in goals. Given that the significant class imbalance could introduce performance bias, particularly in algorithms with high class separation sensitivity, a balanced sample was preferred for the analyses. To achieve this, all goal shots were retained, while 5,000 randomly selected observations from the non-goal class were included, resulting in a balanced dataset of 10,081 samples. To our knowledge, no previous shot-level study of this league has used a dataset of this size. Goal shots demonstrate different characteristics compared to non-goal shots. The

descriptive statistics of the continuous variables included in the analysis are given in Table 1.

Table 1.
The descriptive statistics of the continuous variables.

	Total Sample				Goal		Non-Goal	
Feature	**Min**	**Max**	**Average**	**Standard Deviation**	**Average**	**Standard Deviation**	**Average**	**Standard Deviation**
goal_angle	1.0	86.1	60.2	14.7	55.5	15.1	65.0	12.6
distance	0.0	89.4	16.3	8.8	12.6	6.8	20.0	9.0
match_minute	1.0	90.0	51.2	26.4	51.3	26.3	51.0	26.4
pass_distance	0.03	99.56	17.61	17.24	11.95	8.50	23.37	19.81

Combining these variables may improve the predictive capability of machine learning classification models in distinguishing goals from non-goals. The features used in the dataset are categorized as follows, where the first group comprises the continuous variables summarized in Table 1 and the remaining groups the categorical variables reported in Table 2:

(i) Geometric features and time: goal angle, distance to goal, pass distance, match minute;
(ii) Shot type: left foot, right foot, header, body;
(iii) Shot technique: regular, volley;
(iv) Position: penalty box status (in_box, out_box, penalty_spot);
(v) Set-piece situations: open play, corner, penalty (also counted under position), free kick, direct shot, throw-in;
(vi) Tactic: counter-attack, shot following a cross (assisted_by_cross).

Table 2.
The frequency distributions and goal success rates of the categorical variables.

	Feature	N	%	Goal Rate (%)
Shot Type	**Right foot**	5,289	52.5	**52.4**
	Left foot	2,934	29.1	48.0
	Header	1,820	18.1	48.2
	Body	38	0.4	**60.5**
Shot Technique	**Regular**	6,555	65.0	**47.5**
	Volley	1,041	10.3	**46.1**
Position	**In box**	6,729	66.7	**56.6**
	Out box	2,737	27.2	**24.8**
	Penalty spot	615	6.1	**96.3**
Set Piece situation	**Open play**	7,135	70.8	**49.0**
	Corner	1,345	13.3	**41.6**
	Penalty	615	6.1	**96.3**
	Free kick	585	5.8	**48.7**
	Direct shot	318	3.2	**34.0**
	Throw-in	83	0.8	**44.6**
Tactic **Counter attack**	**Yes**	580	5.8	**66.0**

	No	9,501	94.2	**49.4**
Assisted by Cross	**Yes**	2,048	20.3	**48.5**
	No	8,033	79.7	**50.9**

Note. Shot Technique: Regular and Volley are non-exclusive binary indicators; 661 shots (6.6%) carry both flags. Percentages are calculated relative to the total sample (N = 10,081) and therefore do not sum to 100. Technique information was not recorded in the source data for the remaining 3,146 shots (31.2%), which take a value of 0 on both indicators.

Table 1 and Table 2 tables provide an initial overview of the dataset structure and variable characteristics before the implementation of the machine learning classification models. Categorical variables were transformed using binary indicator encoding, resulting in a total of 21 binary and continuous variables. All calculations were performed using the Python (version 3.12) programming language.

Several variables were included in the model based on their theoretical and practical relevance in football shot outcome prediction. Spatial variables such as goal angle and distance to goal are considered fundamental determinants of scoring probability, as they directly influence the shooter's visual access to the goal and the available shooting space. Contextual variables were also incorporated to capture the conditions under which the shot occurred. Pass distance may reflect the speed and structure of attacking play, while penalty box status (in-box/out-box) represents the shooting location in relation to dangerous scoring areas. Match minute was included to account for potential temporal effects such as fatigue, tactical changes, or variations in defensive organization during different phases of the match. Body-part-related variables, including left foot, right foot, header, and body shots, were considered because shooting efficiency may vary depending on the execution technique and player positioning. Similarly, shot technique variables such as regular shots and volleys may influence shot control and accuracy. Set-piece and game-context variables including penalties, corners, free kicks, throw-ins, counter-attacks, and shots assisted by crosses were also included, as these situations represent structurally different attacking scenarios that may affect the likelihood of scoring.

Data Preprocessing

The significant imbalance between goal shots (11%) and non-goal shots (89%) in the dataset increases the tendency of algorithms sensitive to decision boundaries to learn the majority class, leading to misrepresentation of the minority class. Therefore, various balancing strategies such as oversampling (e.g., SMOTE; Chawla et al., 2002), undersampling, and class weighting are recommended in the literature. In this study, oversampling methods were not preferred due to the risk of synthetic sample generation distorting the spatial shot distribution and potentially increasing computational cost; instead, random undersampling was applied. In this context, all goal shots (5,081 observations) were retained, and combined with 5,000 randomly selected non-goal shots, an approximately balanced dataset of 10,081 total observations was obtained. The balanced dataset was first partitioned into training (80%, 8,064 observations) and test (20%, 2,017 observations) subsets by stratified sampling, so that the goal / non-goal ratio was preserved in both subsets. All preprocessing was then carried out inside a scikit-learn Pipeline together with the classifier.

Because the continuous variables are recorded on different scales, feature scaling was included as the first step of this pipeline. StandardScaler, which transforms variables to zero mean and unit variance, was used for the Gaussian NB, Linear SVM and RBF SVM models. MinMaxScaler was used for Complement NB and scaled all features to the [0,1] range. Categorical variables were already in binary indicator form and required no further transformation.

Hyperparameter optimization was performed on the training subset using Stratified 5-Fold cross-validation. Because scaling was embedded in the pipeline, the scaler was refitted on the training folds of each split and applied to the corresponding validation fold; no scaling statistic was estimated from a validation fold or from the held-out test set. The pipeline carrying the selected hyperparameters was then refitted on the full training subset and applied unchanged to the

test set, which was used once only, for the final performance evaluation. The partition was performed at the shot level, using stratified random sampling on the outcome variable, rather than at the match, team or player level. Shots from the same match may therefore fall on either side of the split.

To establish whether the findings depend on the balancing strategy, a separate robustness analysis was conducted on the original class distribution. The full dataset (N = 46,238) was split once into a training pool (N = 36,990) and a held-out test set (N = 9,248), with the natural prevalence of 10.99% goals preserved in both partitions. Two training sets were then derived from the same pool: a balanced set obtained by random undersampling (N = 8,130) and the full pool used with class weighting (N = 36,990). Class weighting was implemented as class_weight = 'balanced' for the SVM models and as balanced sample weights for the NB models. Each of the four models was tuned separately on each training set, following the procedure described above, and evaluated on the same fixed test set.

Supervised ML Algorithms: NB, SVM and Hyperparameter Optimization

Our aim was to contrast two classification paradigms with distinct decision logics (a geometric approach that separates classes through an optimal hyperplane (SVM) and a probabilistic approach that estimates class membership from feature distributions (NB)) rather than to identify the single best-performing algorithm. Both families are also computationally inexpensive and require few hyperparameters to be tuned, which makes the framework straightforward to replicate with limited computational resources. In this context, four models from the NB and SVM algorithm families were used to predict the probability of football shots resulting in goals: Gaussian NB, Complement NB, Linear SVM, and SVM using RBF (Radial Basis Function) kernel. Models were implemented using the Python scikit-learn (v1.3.0) library (Pedregosa et al., 2011).

Gaussian NB assumes that continuous features are normally distributed and is one of the most computationally efficient models. The primary hyperparameter optimized in this model is var_smoothing, which enhances numerical stability by adding a small value to variance estimates. Complement NB is a variant that particularly improves performance on imbalanced datasets and can provide more balanced predictions for small-sample positive classes as it calculates probabilities through the complement class. The hyperparameter optimized for this model was alpha (Laplace smoothing).

SVM algorithms perform classification by determining the optimal hyperplane that separates samples in a high-dimensional space. Linear SVM is effective in cases where linear separation is possible, while RBF SVM uses a Gaussian kernel to model non-linear relationships. For Linear SVM, C (regularization coefficient) was optimized, while for RBF SVM, both C and gamma, which determines kernel width, were optimized.

Hyperparameter optimization was performed using GridSearchCV. The parameter ranges searched were as follows:

- Gaussian NB: var_smoothing = [1e-09, 1e-08, 1e-07, 1e-06, 1e-05, 1e-04, 0.001, 0.01, 0.1, 1.0]
- Complement NB: alpha = [0.1, 0.5, 1.0, 2.0, 5.0]
- Linear SVM: C = [0.01, 0.1, 1, 10, 100], max_iter = 10,000
- RBF SVM: C = [0.1, 1, 10, 100, 1000], gamma = ['scale', 'auto', 0.001, 0.01]

The optimization process was conducted using Stratified 5-Fold on the training data, ensuring the preservation of goal/non-goal ratios in each fold. ROC-AUC was chosen as the performance metric. ROC-AUC is a robust evaluation criterion for problems where the positive class is rarer, such as goal/non-goal classification, as it is both insensitive to class imbalance and measures the model's discriminative power independent of the decision threshold. GridSearchCV calculated the mean ROC-AUC score from five-fold validation for each hyperparameter combination and selected the best parameter set. In the final stage, models trained with the selected optimal hyperparameters were run on the held-out test data to perform the final performance evaluation.

Evaluation Metrics

Multiple performance metrics were used to comprehensively evaluate model performance. The ROC-AUC (Receiver Operating Characteristic – Area Under Curve) metric measures the model's capacity to distinguish between positive and negative classes and is frequently preferred in classification problems due to its relative robustness to class imbalance. An ROC-AUC value of 0.5 indicates random prediction, while values approaching 1.0 indicate high discriminative power.

Precision shows what proportion of samples predicted as positive by the model are actually positive, while Recall (sensitivity) measures what proportion of actual positive samples are correctly detected. Since these two metrics often move in opposite directions in imbalanced data structures, the importance of F1-Score (harmonic mean of Precision and Recall) increases as a balanced performance indicator.

Feature Importance Analysis

Beyond evaluating model performance, revealing which variables are more determinative in explaining shot success represents an important analytical dimension in football analytics. For this purpose, two model-agnostic explainability techniques were used: Permutation Importance and SHAP (SHapley Additive exPlanations; Lundberg & Lee, 2017). Since both methods allow evaluation of feature effects regardless of model architecture, they are suitable for comparing results from different algorithms within a common framework. The Permutation Importance approach measures the decrease in model performance (ROC-AUC) by randomly shuffling a feature's values and uses this decrease as an indicator of feature importance. It was preferred due to producing more balanced results compared to impurity-based methods, which tend to overstate the importance of high-cardinality variables.

The analysis was conducted on test data using scikit-learn's permutation_importance function; mean importance scores were calculated over 10 permutations for each variable. Since analyzing the entire test set would not be practical, SHAP calculations were performed on 500 observations randomly selected from the test set. SHAP values were calculated using the shap library, and variable effects were evaluated through mean absolute SHAP values.

Results

Hyperparameter Optimization Results

Hyperparameter search was performed with Stratified 5-Fold cross-validation. The low standard deviations of performance scores obtained during cross-validation ($\sigma < 0.02$ for all models) indicate that hyperparameters yielded stable results across different data subsets and the risk of overfitting was limited. When evaluated in terms of computational time, hyperparameter optimization for Linear SVM was measured at approximately 1.4 minutes, while RBF SVM took approximately 4.9 minutes. Optimization time for NB models was completed in less than one minute. Following hyperparameter optimization performed with GridSearchCV, the optimal hyperparameter combinations for the four models were determined as shown in Table 3.

Table 3.
The optimal hyperparameter combinations.

Model	Hyperparameter	Optimal Value
Gaussian NB	var_smoothing	0.1
Complement NB	alpha	5.0
Linear SVM	C	1
RBF SVM	C	10
	gamma	0.01

Both Naive Bayes models initially selected hyperparameter values at the upper boundary of their search grids (Gaussian NB: var_smoothing=1e-04; Complement NB: alpha=5.0), prompting an extended search. For Gaussian NB, the expanded grid (up to 1.0) identified a true interior optimum (var_smoothing=0.1), with negligible change in AUC. For Complement NB, the cross-validated AUC continued to increase marginally across an extended range (up to α = 1000) without reaching a clear plateau; as this did not affect the model's relative ranking (Complement NB remained the weakest model throughout) or the study's principal conclusions, α=5.0 is reported as a representative value.

Model Performance

Table 4 presents the performance results of the four algorithms trained on the balanced dataset. Both Linear and RBF SVM outperform the NB in discriminative performance. RBF SVM achieves the highest AUC (0.830), indicating the strongest discriminative power between goals and non-goals. RBF SVM also provides the most balanced performance with the highest F1-Score of 0.762. Gaussian NB exhibits the highest Precision (0.786) but suffers from a lower Recall (0.532), meaning it missed close to half of the goals. In contrast, Complement NB improves Recall to 0.767 but at the cost of lower Precision (0.668).

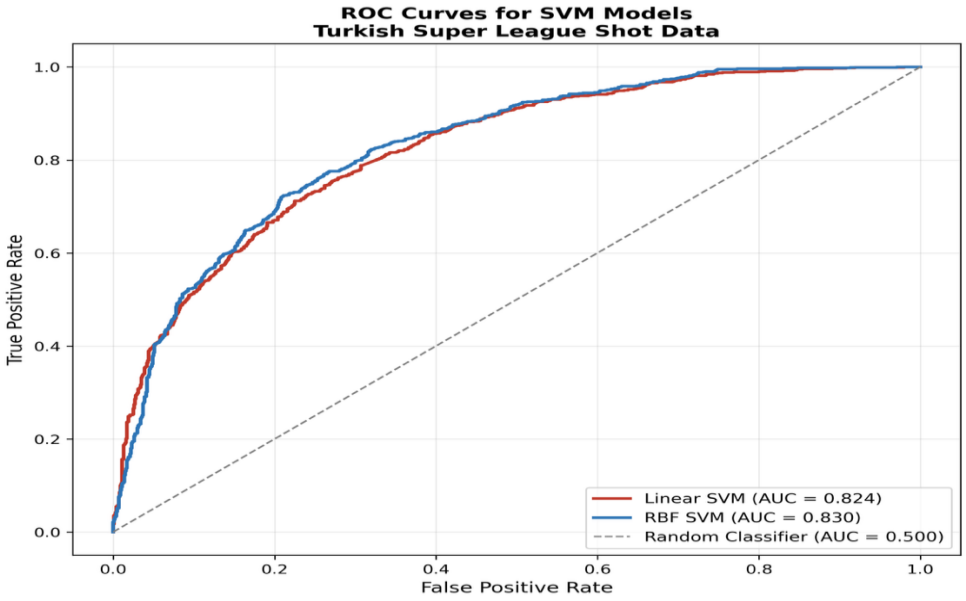
Table 4.
The test data performance results of the algorithms.

Model	AUC	Precision	Recall	F1-Score
Linear SVM	0.824	0.712	0.802	0.755
RBF SVM	0.830	0.726	0.802	0.762
Gaussian NB	0.801	0.786	0.532	0.635
Complement NB	0.757	0.668	0.767	0.714

The results show that SVM family is more suitable than NB for modeling shot outcomes. After model training was completed, predictions were obtained on the test dataset and all metrics were calculated as presented in Table 4. Based on the performance results observed in Table 4, the next step in the analysis focuses only on the two best-performing models in the SVM family: Linear SVM and RBF SVM. Following the evaluation of model discriminative power via ROC curves, confusion matrices are presented to provide a detailed breakdown of the classification errors.

The ROC curves were plotted for the SVM models and the variation in sensitivity-specificity balance at different threshold values was analyzed. These curves enabled comparative assessment of the models' sensitivity to decision thresholds and their overall discriminative capacity. Both models clearly outperformed random classification (AUC=0.500), indicating good predictive capability for goal scoring probability. Figure 1 shows the ROC curves for both models; Linear and RBF SVM. The curves are indistinguishable within the threshold range used in practice; the RBF SVM shows only a marginal advantage in the region of high false positive rates.

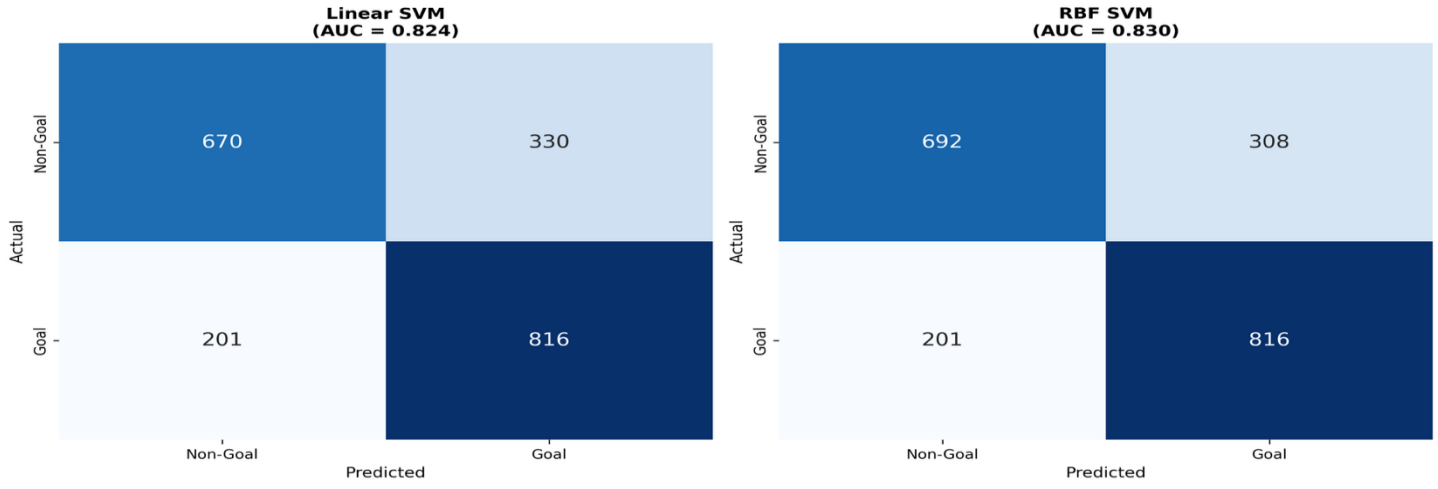
Figure 1.
ROC curves for the Linear and RBF SVM models (test set N = 2,017)



Additionally, confusion matrices were created to examine in detail the error distribution of models across positive and negative classes. The confusion matrix provides the opportunity to visualize which class models are more successful or weaker in by reporting the numbers of true positives, true negatives, false positives, and false negatives. The Confusion matrices for Linear and RBF SVM are given by Figure 2. The confusion matrices in Figure 2 further illuminate the performance differences between the Linear and RBF kernels.

Both models demonstrate strong sensitivity in true positive goal detection. Both the Linear and RBF SVM correctly identify 816 goals. The RBF SVM produces only 308 false positives compared to 330 for the Linear SVM. It appears that the linear model made goal decisions more easily compared to the RBF model. The RBF SVM achieves higher Precision and F1-Score, with identical Recall.

Figure 2.
Confusion matrices for the Linear and RBF SVM models (test set N = 2,017)



According to the True Negative Performance, the RBF SVM is more effective at identifying "Non-Goals," correctly classifying 692 instances compared to 670 by the Linear model. The False Negative Rate was the same; the models failed to correctly predict 201 goals. These findings form the foundation for the feature importance analysis conducted to understand the decision processes of models and interpret variable effects.

Feature Importance Analysis Results

Feature importance analyses were conducted using both Permutation Importance and SHAP. Analyses were restricted to the two best-performing models (Linear and RBF SVM). Tables 5 and 6 report the two measures for each model. SHAP values were computed on 500 observations drawn at random from the 2,017-shot test set.

Table 5.
Feature importance for Linear SVM

Order	Feature	Permutation Importance	Permutation Rank	SHAP	SHAP Rank
1	distance	0.223	1	0.194	1
2	pass_distance	0.059	2	0.081	2
3	counter_attack	0.004	3	0.006	9
4	assisted_by_cross	0.004	4	0.021	3
5	penalty_spot	0.004	5	0.004	10
6	set_piece_penalty	0.004	5	0.004	11
7	set_piece_corner	0.003	7	0.020	4
8	set_piece_direct_shot	0.002	8	0.004	12
9	shot_type(header)	0.001	9	0.012	5
10	in_box	0.001	10	0.003	13
Spearman rank correlation (Permutation vs. SHAP): $\rho = .685$, $p < .001$					

Across all 21 features the Spearman rank correlation between the two measures was $\rho = .685$ ($p < .001$) for the Linear SVM and $\rho = 0.885$ ($p < .001$) for the RBF SVM.

Table 6.
Feature importance for RBF SVM

Order	Feature	Permutation Importance	Permutation Rank	SHAP	SHAP Rank
1	distance	0.167	1	0.145	1
2	pass_distance	0.067	2	0.065	2
3	out_box	0.017	3	0.031	3
4	shot_type(header)	0.013	4	0.022	4
5	penalty_spot	0.012	5	0.010	10
6	set_piece_penalty	0.012	5	0.010	12
7	goal_angle	0.011	7	0.012	8
8	set_piece_corner	0.010	8	0.019	7
9	in_box	0.010	9	0.021	6
10	shot_type(right)	0.006	10	0.022	5
Spearman rank correlation (Permutation vs. SHAP): $\rho = .885$, $p < .001$					

Both models rank the same two features first and second on both measures: distance and pass_distance. In the Linear SVM these two account for 92% of the total permutation importance of the ten features listed (0.282 of 0.305); in the RBF SVM they account for 72% (0.234 of 0.325). The drop from the second to the third feature is from 0.059 to 0.004 in the Linear model and from 0.067 to 0.017 in the RBF model.

The two measures produce similar rankings but different magnitudes. In the Linear model the SHAP value exceeds the permutation importance for every feature except distance, where it is lower (0.194 against 0.223), and penalty_spot and set_piece_penalty, where the two are equal (0.004). In the RBF model the SHAP value is lower for distance (0.145 against 0.167), pass_distance (0.065 against 0.067), penalty_spot and set_piece_penalty (0.010 against 0.012), and higher for the remaining six features. The rank correlation between the two measures is $\rho = 0.685$ in the Linear model and $\rho = 0.885$ in the RBF model ($p < 0.001$). The rank differences occur below the second position: in the Linear model set_piece_corner moves from permutation rank 7 to SHAP rank 4, shot_type(header) from 9 to 5, and counter_attack from 3 to 9; in the RBF model shot_type(right) moves from 10 to 5 and in_box from 9 to 6. Several features share the same permutation importance at three decimal places. In the Linear model counter_attack, assisted_by_cross, penalty_spot and set_piece_penalty all have 0.004, and shot_type(header) and in_box both have 0.001. In the RBF model penalty_spot and set_piece_penalty both have 0.012, and set_piece_corner and in_box both have 0.010. The ordering within these groups is not resolved at the reported precision, while the SHAP values separate them: within the Linear 0.004 group, assisted_by_cross has a SHAP value of 0.021 (rank 3) and counter_attack 0.006 (rank 9). Penalties are encoded twice, as penalty_spot under position and as set_piece_penalty under set-piece situation. The two indicators receive identical values on both measures in both models: 0.004 permutation importance and 0.004 SHAP in the Linear model (SHAP ranks 10 and 11), and 0.012 permutation importance and 0.010 SHAP in the RBF model (SHAP ranks 10 and 12). Both share permutation rank 5 in each model. Each indicator carries the information the other loses when shuffled, which is a consequence of the encoding.

The two lists differ below the second position. goal_angle appears in the RBF top ten (permutation importance 0.011, rank 7; SHAP 0.012, rank 8) but not in the Linear top ten. out_box and shot_type(right) also appear only in the RBF table, and assisted_by_cross, set_piece_direct_shot and counter_attack only in the Linear table. shot_type(header), penalty_spot, set_piece_penalty, set_piece_corner and in_box appear in both. No shot technique indicator (regular or volley) appears in either table; the only body-part indicators listed are shot_type(header), in both tables, and shot_type(right), in the RBF table only.

Robustness to the Class Balancing Strategy

Table 7 reports the performance of the four models on a test set preserving the original goal prevalence, under the two training strategies described in the Methods.

Table 7.
Model performance under balanced and class-weighted training, evaluated on a common test set preserving the original class distribution.

Model	Training	AUC	Precision	Recall	F1-Score
RBF SVM	Balanced	0.8103	0.239	0.774	0.365
	Class-weighted	0.8142	0.241	0.781	0.369
Linear SVM	Balanced	0.8122	0.228	0.800	0.355
	Class-weighted	0.8119	0.233	0.780	0.359
Gaussian NB	Balanced	0.7913	0.338	0.395	0.364
	Class-weighted	0.7911	0.352	0.342	0.347
Complement NB	Balanced	0.7506	0.209	0.735	0.325
	Class-weighted	0.7532	0.205	0.750	0.321

Note. Fixed test set: n = 9,248, 10.99% goals, identical across both training strategies. Precision, Recall and F1-Score are computed at each model's default decision rule; AUC is threshold-independent.

The two Naive Bayes models keep the same positions under both strategies, and both SVM models rank above both NB

models in every case. The Linear and RBF SVM exchange places between the two strategies: the Linear SVM is higher under balanced training (0.8122 against 0.8103) and the RBF SVM under class weighting (0.8142 against 0.8119). The AUC difference between the two strategies is 0.004 or smaller for all four models.

Precision and F1-Score are considerably lower here than the corresponding values in Table 4, while Recall is comparable for the SVM models and lower for Gaussian NB. This reflects the composition of the test set rather than a deterioration in the models. These three metrics are computed at each model's default decision rule, and precision at a fixed threshold is bounded by the positive class rate, which is 10.99% in this test set against 50.4% in the balanced test set used for the primary analysis. AUC, being threshold-independent, is the quantity that can be compared directly across the two settings, and it is stable.

Discussion

In this study, shot outcomes in the Turkish Super League were modelled as a binary classification problem, two supervised algorithm families were compared, and the resulting models were interpreted with two model-agnostic explainability methods. Two findings stand out. SVM models were able to distinguish goals from non-goals more consistently than NB models. In both SVM models, the strongest predictor was goal distance, followed by pass distance.

Performance differences between the SVM and NB families

The two SVM variants produced comparable discriminatory performance in distinguishing between goal-scoring and non-goal shots, with AUC values (0.824 for Linear SVM and 0.830 for RBF SVM). Both SVM models outperformed Gaussian NB (AUC = 0.801) and Complement NB (AUC = 0.757). A likely explanation lies in the structure of the feature set. After binary indicator encoding, most of the 21 predictors are mutually exclusive binary indicators: a shot cannot simultaneously be in the box, outside the box and from the penalty spot, and penalties appear both as a position category and as a set-piece category. The conditional independence that NB assumes is therefore violated in a systematic rather than an incidental way. Tax and Joustra (2015) reported the opposite pattern at match level, where NB was among their best-performing classifiers once PCA had been applied, and they attributed its strong performance precisely to that decorrelation step. The two results suggest that the viability of NB depends less on the sport or the prediction target than on whether the input space has been made approximately independent beforehand. This also helps explain why NB performs competitively in match-level studies, such as the 85.43% accuracy reported by Venkatachalam et al. (2023) for a Gaussian NB model predicting tournament outcomes from squad- and team-level attributes, while remaining rare in shot-level work. Rudrapal et al. (2020) make the same point from the opposite direction: on 11,400 English Premier League matches, Gaussian NB reached 65.84% accuracy against 58.77% for SVM, the reverse of the ordering obtained here. Their predictors are aggregated team, player and head-to-head summaries rather than mutually exclusive event categories, which is the condition under which the independence assumption is least damaging. Kahraman (2022) reports the same pattern in basketball, where Gaussian NB outperformed Random Forest and gradient boosting on datasets of a few hundred games. The ordering between the two families therefore appears to follow the structure of the feature space rather than the sport or the inherent capacity of the algorithms.

For the comparison between the two kernels, the RBF SVM achieved the higher F1-Score (0.762 against 0.755), and produced fewer false positives (308 against 330). Ding (2020) found a clearer advantage for the Gaussian kernel over the linear kernel (F1 0.71 against 0.66; ROC 0.68 against 0.63) in a World Cup dataset, but also noted that the dataset was small enough for complex models to overfit and that league-level data would provide a more appropriate test. Our results are consistent with that expectation in the sense that the two kernels converge once the sample is large: on a balanced set of 10,081 shots, the additional flexibility of the RBF kernel altered the error profile of the model rather than its overall ranking ability. In practical terms, the choice between them is a choice about which type of error is more costly, not a choice about which model is stronger.

Gaussian NB returned the highest Precision of the four models (0.786) together with the lowest Recall (0.532): when it predicted a goal it was usually correct, but it missed close to half of the goals. A comparable divergence appears in Senn (2024) and in Eggels (2016). In all three cases the pattern reflects where the decision threshold falls relative to a distribution of predicted values concentrated at the lower end, rather than a defect in the classifier itself. This is also why the values reported here are best treated as internal comparisons rather than as figures that can be placed alongside accuracy rates from other studies. The meta-analysis by Pietraszewski et al. (2025) is instructive on this point: across sports-related artificial intelligence studies they observed substantial heterogeneity and attributed most of the variance to methodological and contextual differences rather than to sampling error.

Feature importance and its correspondence with the literature

Distance to goal dominated the importance ranking in both SVM models. This is one of the more stable results in the shot-modelling literature, and our data reproduce it in a league that has not previously been examined at this scale. Anzer and Bauer (2021) reported shot location as the most crucial factor in their SHAP analysis and described an almost linear relationship between goal distance and the predicted value. Fairchild et al. (2018) found location to be the significant determinant in their Major League Soccer model, while shot type, assist type and play type were not significant independent variables. Hewitt and Karakuş (2023) likewise identified distance from the centre of the goal among the highest-impact predictors. The descriptive statistics point in the same direction: goals were taken on average from 12.6 m and non-goals from 20.0 m.

Pass distance ranked second in both models, which is a less commonly reported result. Lucey et al. (2015) provide the closest reference point: adding attacking features, including the nature and origin of the incoming pass or cross, reduced their average prediction error from 0.1545 to 0.1439, a larger improvement than they obtained by adding defensive features (0.1554 to 0.1545). Senn (2024) similarly recommends capturing the origin and speed of the preceding pass as a route to more informative contextual models. In our data, goals were preceded by substantially shorter passes on average (11.95 m) than non-goals (23.37 m). This finding indicates an association between pass distance and shot outcome, but it does not imply that shorter passes directly increase the probability of scoring. Therefore, pass distance should be interpreted as a contextual characteristic of the attacking sequence rather than as a direct causal determinant of shot outcome.

The low ranking of goal angle (seventh on permutation importance and eighth on SHAP in the RBF SVM, and outside the top ten in the Linear SVM) is the point at which our results diverge most from the literature, where angle is reported among the two strongest predictors (Anzer & Bauer, 2021; Cavus et al., 2026; Hewitt & Karakuş, 2023). Two explanations are possible. First, angle and distance are geometrically related rather than independent quantities, and permutation-based measures distribute importance across correlated features: when one of two related predictors already carries the signal, shuffling the other produces only a small drop in performance. Kim and Park (2025) reported exactly this behavior when comparing importance measures on correlated sequential variables and concluded that no single measure should be relied upon. Second, the definition of the angle variable may differ from the shooter's-view angle constructed from goalpost coordinates in studies such as Senn (2024) and Eggels (2016); in our sample the mean angle value was lower for goals (55.5) than for non-goals (65.0). Mead et al. (2023) likewise attribute the unexpectedly low ranking of angle in their models to how the variable was constructed, noting that a bell-shaped distribution of angle values can distort its estimated effect. Clarifying this definition before comparing angle effects across data sources would be a useful step for subsequent work on this league. Kim and Park (2025) obtained different rankings from Mean Decrease Impurity, Permutation Importance and SHAP on the same data, and argued for using several methods in combination rather than one. Reporting the two measures separately, as here, makes the points of agreement and disagreement visible.

Limitations

Firstly, the dataset consists of event data and therefore contains no information on defensive pressure or goalkeeper positioning. Lucey et al. (2015) show that the number of defenders between the shooter and the goal carries independent information, and Anzer and Bauer (2021) report that goalkeeper positioning substantially alters the predicted value. In addition, to avoid information loss, instead of simplifying this distinction to goal or no goal as the target variable, a more detailed alternative could be suggested: the expected probability of a shot hitting the target (Yeung & Fujii, 2024). Another limitation arose from data sampling. Random undersampling was preferred over synthetic oversampling. Increasing the data by interpolating between two shot locations, as SMOTE does, generates positions from which no shot was actually taken and can distort the spatial distribution that the model needs to learn from. At the same time, retaining 5,000 of 41,157 non-goal shots at random means that some variation among unsuccessful attempts is not represented in the training data. The robustness analysis in Table 7 addresses this concern: training on the full distribution with class weighting changed the AUC of every model by 0.004 or less.

A further limitation concerns the validation design. The train–test partition was performed at the shot level, so shots from the same match can appear in both partitions. Shots within a match share the opposition, the tactical configuration and the match state in which they occurred, and a shot-level split may therefore yield a mildly optimistic estimate of performance on matches the model has not seen. The effect is likely to be limited here, since the feature set consists of shot-specific geometric and contextual variables rather than match- or team-level aggregates, none of which is constant within a match. A grouped partition, in which all shots from a given match are assigned to the same subset, or a temporal partition that trains on earlier seasons and tests on later ones, would provide a stricter test and is a natural next step.

Finally, the analysis covers a single league across five seasons. The feature dynamics reported here may reflect the tactical and defensive characteristics of the Turkish Super League during this period and should be tested against other leagues before being treated as general. Pietraszewski et al. (2025) note that sample composition is a major source of variability in the sports analytics literature, and Yeung and Fujii (2024) attribute limits in the generalization of their own models to the range of match scenarios represented in the training data.

Conclusion and Recommendation

This study examined whether shots in the Turkish Super League result in goals, using five seasons of event data and a balanced sample of 10,081 shots. Two findings define its contribution. First, the SVM family separated goals from non-goals more consistently than the NB family: RBF SVM produced both the highest discriminative power (AUC 0.830) and the highest F1-Score (0.762), while Gaussian NB combined the highest Precision (0.786) with the lowest Recall (0.532). This result is consistent with the expectation that NB's conditional independence assumption is not met. Second, Permutation Importance and SHAP both showed that distance to goal and the length of the preceding pass carry most of the discriminative weight, while body part, shot technique and the majority of set-piece categories contribute comparatively little. The prominence of distance replicates a well-established result in a league for which shot-level evidence has been scarce; the prominence of pass distance is a less familiar finding, and one that this dataset can identify but not explain.

Because the analysed sample was balanced by design, the model outputs are classification decisions rather than calibrated scoring probabilities: a model trained on a 50/50 sample produces values on a different scale from the true goal rate, so they cannot be aggregated as an expected goals total would be. A robustness analysis on the original class distribution confirmed that this design choice does not drive the reported performance, but it does not address the scale of the outputs. Probability calibration on the full distribution, a grouped or temporal validation design, and the addition of tracking-based features on defensive pressure and goalkeeper positioning are the most direct extensions of this work.

The framework is most useful as a structured input to post-match review rather than as a predictive instrument. Its importance ranking indicates that the location and build-up of a chance explain more about its outcome than the technique used to finish it, a claim that can be checked against video and used to focus the review of attacking play. Because the models produce a decision boundary rather than a set of rules, their outputs are best read alongside the judgement of the technical staff rather than as standalone recommendations.

Etik Komite Onayı: Veriler profesyonel maç kayıtlarından oluştuğu ve herhangi bir insan katılımcı veya hassas kişisel bilgi içermediği için bu araştırma için etik kurul onayı gerekmemiştir.
Katılımcı Onamı: Bu çalışmaya katılımcı dahil edilmemiştir.
Hakem Değerlendirmesi: Dış bağımsız.
Yazar Katkıları: Fikir- M.N.G., M.C., N.F.K.; Tasarım- M.N.G., M.C., N.F.K.; Denetleme- M.C., N.F.K.; Kaynaklar- M.N.G., M.C., N.F.K.; Veri Toplanması ve/veya İşlemesi M.N.G.; Analiz ve/ veya Yorum- M.N.G., M.C., N.F.K.; Literatür Taraması- M.N.G., M.C.; Yazıyı Yazan- M.N.G., N.F.K.; Eleştirel İnceleme- M.C., N.F.K.

Çıkar Çatışması: Yazarlar, çıkar çatışması olmadığını beyan etmiştir.
Finansal Destek: Yazarlar, bu çalışma için finansal destek almadığını beyan etmiştir.
Yapay Zeka Kullanımı: Bu makalenin hazırlanması sırasında yapay zeka araçları yalnızca dil düzenleme desteği amacıyla kullanılmıştır. Tüm bilimsel içerik, veri analizi ve entelektüel katkılar tamamen yazarların sorumluluğundadır.

Ethics Committee Approval: Since the data consists of professional match records and does not involve any human participants or sensitive personal information, ethics committee approval was not required for this research.
Informed Consent: This study does not involve participants.
Peer-review: Externally peer-reviewed.
Author Contributions: Concept - M.N.G., M.C., N.F.K.; Design- M.N.G., M.C., N.F.K.; Supervision- M.C., N.F.K.; Resources- M.N.G., M.C., N.F.K.; Data Collection and/or Processing- M.N.G.; Analysis and/or Interpretation- M.N.G., M.C., N.F.K.; Literature Search- M.N.G., M.C.; Writing Manuscript- M.N.G., N.F.K.; Critical Review- M.C., N.F.K.
Conflict of Interest: The authors have no conflicts of interest to declare.
Financial Disclosure: The authors declared that this study has received no financial support.
Use of Artificial Intelligence: Artificial intelligence tools were used solely for language editing support during the preparation of this manuscript. All scientific content, data analysis, and intellectual contributions are the sole responsibility of the author.

References

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*, 624475. https://doi.org/10.3389/fspor.2021.624475

Burhaein, E., Fadjeri, A., & Widiyono, I. P. (2024). Application of Naive Bayes Algorithm for Physical Fitness Level Classification. *International Journal of Disabilities Sports and Health Sciences*, *7*(1), 178-187. https://doi.org/10.33438/ijdshs.1330745

Cavus, M., Stańdo, A., & Biecek, P. (2026). Glocal Explanations of Expected Goal Models in Football. In: Guidotti, R., Schmid, U., Longo, L. (eds) Explainable Artificial Intelligence. xAI 2025. Communications in Computer and Information Science, vol 2580. Springer, Cham. https://doi.org/10.1007/978-3-032-08333-3_1

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Davis, J., & Robberechts, P. (2024). Biases in expected goals models confound finishing ability. *arXiv preprint* arXiv:2401.09940. https://arxiv.org/abs/2401.09940

Ding, S. (2020). Game Predication of FIFA Football World Cup Based on Support Vector Machine. *Academic Journal of Computing & Information Science, 3*(5), 17–22.

Eggels, H. (2016, October). Expected goals in soccer: Explaining match results using predictive analytics. In *The machine learning and data mining for sports analytics workshop* (Vol. 16, p. 131). Technische Universiteit Eindhoven.

Fairchild, A., Pelechrinis, K., & Kokkodis, M. (2018). Spatial analysis of shots in MLS: A model for expected goals and fractal dimensionality. *Journal of Sports Analytics, 4*(3), 165–174. https://doi.org/10.3233/JSA-170207

Guan, Z. (2025). Scalable Framework for Basketball Game Prediction Combining Image Processing XGBoost and Enhanced Support Vector Machine. *Scalable Computing: Practice and Experience*, *26*(4), 1727-1739. https://doi.org/10.12694/scpe.v26i4.4684

Hewitt, M., & Karakuş, O. (2023). A Machine Learning Approach for Player and Position Adjusted Expected Goals in Football (Soccer) *Franklin Open,* 4, 100034. https://doi.org/10.1016/j.fraope.2023.100034

Hubáček, O., Šourek, G., & Železný, F. (2019). Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning, 108*(1), 29–47. https://doi.org/10.1007/s10994-018-5704-6

Iapteff, L., Brito, L., & Zanella, A. (2025). Toward interpretable expected goals modelling using Bayesian mixed models. *Frontiers in Sports and Active Living, 7*, 1504362. https://doi.org/10.3389/fspor.2025.1504362

Igiri, C. P. (2015). Support Vector Machine–Based Prediction System for a Football Match Result. *IOSR Journals (IOSR Journal of Computer Engineering)*, 17(3), 21-26. https://www.europub.co.uk/articles/111323

Kahraman, H. T. (2022). *Prediction of basketball match results using machine learning algorithms* [Unpublished master's thesis]. Middle East Technical University.

Kim, J., & Park, J. (2025). Applying machine learning for analyzing shooting importance in modern pentathlon. Applied Sciences, 15(17), 9378. https://doi.org/10.3390/app15179378

Li, L., & Hao, A. (2025). Research on the sports training effect based on GABP neural network and artificial intelligence. *Scientific Reports, 15*(1), 39214.

https://doi.org/10.1038/s41598-025-20426-7
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. *Proceedings of the 9th Annual MIT Sloan Sports Analytics Conference*, 1–9.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.
Mead, J., O'Hare, A., & McMenemy, P. (2023). Expected goals in football: Improving model performance and demonstrating value. *Plos one*, *18*(4), e0282295. https://doi.org/10.1371/journal.pone.0282295
Moustakidis, S., Plakias, S., Kokkotis, C., Tsatalas, T., & Tsaopoulos, D. (2023). Predicting football team performance with explainable AI: Leveraging SHAP to identify key team-level performance metrics. Future Internet, 15(5), 174. https://doi.org/10.3390/fi15050174
Ou-Yang, Y., Hong, W., Peng, L., Mao, C. X., Zhou, W. J., Zheng, W. T., ... & Wang, Y. F. (2025). Explaining basketball game performance with SHAP: insights from Chinese Basketball Association. *Scientific Reports*, *15*(1), 13793. https://doi.org/10.1038/s41598-025-97817-3
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Pietraszewski, P., Terbalyan, A., Roczniok, R., Maszczyk, A., Ornowski, K., Manilewska, D., Kuliś, S., Zając, A., & Gołaś, A. (2025). The role of artificial intelligence in sports analytics: A systematic review and meta-analysis of performance trends. *Applied Sciences, 15*(13), 7254. https://doi.org/10.3390/app15137254
Rudrapal, D., Boro, S., Srivastava, J., & Singh, S. (2020). A Deep Learning Approach to Predict Football Match Result. In: Behera, H., Nayak, J., Naik, B., Pelusi, D. (eds) Computational Intelligence in Data Mining. Advances in Intelligent Systems and Computing, vol 990. Springer, Singapore. https://doi.org/10.1007/978-981-13-8676-3_9
Scholtes, A., & Karakuş, O. (2024). Bayes-xG: player and position correction on expected goals (xG) using Bayesian hierarchical approach. Front. Sports Act. Living 6:1348983. https://doi.org/10.3389/fspor.2024.1348983
Senn, W. (2024). *Addressing evaluation challenges on the expected goals (xG) metric in football analysis* [Unpublished bachelor's thesis]. University of Applied Sciences and Arts Western Switzerland. https://sonar.ch/global/documents/331251
Seyrek, M., Kiriktir, E., Doğan, Y., & Yapıcıoğlu, B. (2021). Football Match Result Prediction Decision Support System. Journal of Current Research on Engineering, Science and Technology, 7 (2), 173-202. https://doi.org/10.26579/jocrest.85
Sharma, M., Monika, Kumar, N., & Kumar, P. (2022). Naive bayes-correlation based feature weighting technique for sports match result prediction. *Evolutionary Intelligence*, *15*(3), 2171-2186. https://doi.org/10.1007/s12065-021-00629-3.
Tax, N., & Joustra, Y. (2015). Predicting the Dutch football competition using public data: A machine learning approach. https://doi.org/10.13140/RG.2.1.1383.4729
Venkatachalam, A., Rajeswari, D., & SS, S. N. (2023, March). Football prediction system using gaussian naïve bayes algorithm. In *2023 second international conference on electronics and renewable systems (ICEARS)* (pp. 1640-1643). IEEE. DOI:10.1109/ICEARS56392.2023.10085510
Wu, L. C., Kuo, C., Loza, J., Kurt, M., Laksari, K., Yanez, L. Z., Senif, D., Anderson, S. C., Miller, L. E., Urban, J. E., Stitzel, J. D., & Camarillo, D. B. (2018). Detection of American football head impacts using biomechanical features and support vector machine classification. *Scientific Reports, 8*, 855. https://doi.org/10.1038/s41598-017-17864-3
Yeung, C., & Fujii, K. (2024). A strategic framework for optimal decisions in football 1-vs-1 shot-taking situations: An integrated approach of machine learning, theory-based modeling, and game theory. *Complex & Intelligent Systems, 10*(5), 5989–6008. https://doi.org/10.1007/s40747-024-01466-4