Advancing Action-Level Soccer Analytics:
A Comparative Study of VAEP Model Enhancements Using Division 1 Women's Collegiate Soccer Event Data

Benjamin Thorpe
Duke University Honors Thesis in Statistical Science
April 2024

Abstract

This thesis seeks to enhance the analytical modeling of player actions in soccer utilizing Division 1 women's collegiate soccer event-level data. By focusing on the Valuing Actions by Estimating Probabilities (VAEP) framework, this research addresses the need for model updates in light of a new and more detailed data version available to D1 women's teams and dives deeper into understanding the components that render a player's actions valuable. Employing data accessed through an API from the Duke team's provider, Wyscout, this study intricately analyzes the transition from Version 2 (V2) to Version 3 (V3) of the Wyscout event data, adapting it to the Soccer Player Action Description Language (SPADL) for model compatibility and analysis. Additionally, through evaluating the performance of the VAEP model and completing a variable importance analysis on it, the research provides a detailed comparison between model versions and the significance of various actions within the game. Findings highlight the model's enhancements and the critical factors contributing to player performance, offering a useful tool for coaches to refine their in-game tactics and player recruitment plans. Moreover, this thesis outlines potential future directions, including utilizing this same structure to analyze players and leagues around the world, to augment the applicability of VAEP across broader soccer contexts. Overall, this study aims to contribute to soccer analytics by elucidating key aspects that influence positive game outcomes, thus offering a replicable foundation for further research in this field.

Introduction

In the evolving landscape of sports analytics, the quest for precision and actionable insights has led to the development of sophisticated models aimed at quantifying player performance in a myriad of contexts. This thesis is positioned at the intersection of this quest and the practical needs of the Duke University women's soccer team, aiming to both advance the understanding of player actions in soccer through the Valuing Actions by Estimating Probabilities (VAEP) model framework [1] and provide tangible benefits to the team's strategy and performance analysis. The impetus for this project stems from a desire to leverage the latest advancements in event-level data analytics to enhance the coaching staff's decision-making process and player evaluation methods, all while contributing to the broader field of soccer analytics.

The VAEP model gives a unique and thorough way to analyze the game of soccer. By quantifying individual player actions within the game, it offers a nuanced perspective on performance, beyond traditional statistics. Considering soccer is such a low-scoring game, having the ability to value the many actions on the pitch that are not goals or assists allows for a much more robust way to evaluate players who are not often directly involved in scoring opportunities. It is clear that scoring a difficult goal outside of the six yard box contributes positively to a team's success, which coaches and pundits will easily pick up on, but what about the incisive pass that leads to a great scoring opportunity that happens to be missed? The latter is an example of where utilizing VAEP can paint a more accurate picture of which actions occurring throughout the course of a match are most valuable. This approach enables us to calculate a quantitative measure of an action's contribution to an attack, irrespective of whether or not a goal was scored, and this independence from outcomes is a key principle in what makes VAEP so powerful. Additionally, this model delivers a way to assess how likely a team is to concede a goal after a certain action is performed. Effectively quantifying defense has been a difficult problem not only in soccer but throughout the world of sports analytics, which is why this side of the VAEP model output will be analyzed as well.

My research tackles the challenge of updating and refining the model to incorporate new, more detailed versions of event-level data, specifically transitioning from Version 2 (V2) to Version 3 (V3) of the Wyscout event data [2]. More detail on the differences between these feature sets and their effect on VAEP model performance come throughout the paper. This process not only necessitates adapting to a new data format but also critically analyzing the inputs to the model to determine if any specific action types stand out in creating opportunities to score or concede.

Interpretability in machine learning has emerged as a paramount concern in the field, especially in applications where understanding the “why” behind predictions can be as crucial as the predictions themselves. In the realm of sports analytics, and soccer in particular, the ability to explain the rationale behind model outcomes enables coaches, analysts, and stakeholders to make informed decisions based on more than just numerical outputs. It bridges the gap between complex statistical models and practical, tactical insights applicable in real-world scenarios. By focusing on the variable importance within the VAEP model and examining how specific player actions contribute to match outcomes, this research contributes to this dialogue on interpretability. It underscores the necessity of transparent, comprehensible models that not only predict but also illuminate the underlying dynamics of the game. Through this research the aim is to advance the field of soccer analytics by providing a framework that not only leverages the latest in data analytics but also emphasizes the importance of model interpretability for enhancing strategic decision-making in the sport.

The API access the Duke women’s team’s partnership with Wyscout provides is utilized to access the data for this study. In the data section of the thesis the transition between different versions of event data and the rationale behind the preprocessing that is done to have it in a form compatible with the VAEP model is explained. The methodology ensures that while the insights derived are mainly only relevant to ACC women’s soccer, the code is still replicable and applicable to other teams and leagues through the plethora of event stream data publicly available online.

The thesis unfolds in several key sections, beginning with a comprehensive literature review that sets the stage for understanding the VAEP framework and its implications for soccer analytics. This is followed by a detailed examination of the data collection process, the modeling methodology employed to update and evaluate the VAEP model, and an analysis of the results that underscores the improvements to the model and insights gained from the research. The discussion extends into potential future directions and acknowledges the limitations of the study, providing a balanced view of its contributions and constraints.

Finally, the thesis culminates by highlighting the significant findings in the context of the research objectives and outlines how these contributions can be leveraged by coaching staffs and analysts within the ACC collegiate women's soccer domain and beyond. The main goal with this project is to lay out a resource for decision makers in soccer to consult when evaluating player and team performance and to explain the importance of the VAEP framework to future efforts in this field, especially in collegiate women's soccer, where access to such rich data is likely to be underutilized. By offering a detailed account of the research process and its outcomes, this thesis hopes to enrich the field of soccer analytics through an examination of VAEP as a refined tool that can help shape the game going forward.

Literature Review

The advancement of analytics in soccer has transformed the way player actions are evaluated, moving to a more complete understanding of performance. As event-level data has become publicly available through providers such as Wyscout (the data provider in use here) and Statsbomb, it has enabled papers to be written about analyzing the game of soccer on a deeper level. An example of this is Bjertnes's master's thesis on applying a zero-sum two-agent Markov game model to soccer matches to better understand individual player contributions [3]. This paper brings up ideas which have become central to today's analysis of individual actions, including recognizing the game-state to supply more context to a predictive model. More recently, Lotte Bransen and Jan Van Haaren's work [4] demonstrates how action data can be split into possession sequences and then utilized to effectively quantify the value of passes in leading to goals. The VAEP model which is the basis for this thesis builds on this approach to value all actions from a match in a similar manner.

A result of this evolution is the work of Tom Decroos, whose thesis "Actions Speak Louder than Goals: Valuing Player Actions in Soccer" includes the creation of the Valuing Actions by Estimating Probabilities metric [1]. The paper, published in 2018, as a whole discusses a "new language for describing individual player actions" and develops a "framework for valuing any type of player action" while accounting for game context [1]. This framework is designed to quantify the value of individual player actions within the context of a game, providing a more objective and comprehensive measure of a player's contribution to the team's success. By analyzing event-level data, VAEP assigns a numerical value to each action based on its impact on the likelihood of scoring or conceding a goal, thereby offering a granular view of team and player performance. However, evaluating their model is a difficult proposition as there are "no objective ground truth action values or player ratings" [1]. Although the intuition behind their model makes sense in theory, the researchers performed different experiments and created visualizations to demonstrate the face-validity of VAEP. Examples of these tests include analyzing game sequences to determine if the importance VAEP assigns to specific actions is reasonable as well as identifying which players provide the most value per action and seeing whether it passes a sense check. Plots taken from the original VAEP paper [1] demonstrating both of the analyses described above, respectively, are shown below in Figure 1 and Figure 2.

	TIME	PLAYER	ACTION	P_{scores}	VALUE
1	92m4s	S. Busquets	pass	0.03	0.00
2	92m6s	L. Messi	pass	0.02	- 0.01
3	92m8s	S. Busquets	pass	0.03	+ 0.01
4	92m11s	L. Messi	take on	0.08	+ 0.05
5	92m12s	L. Messi	pass	0.17	+ 0.09
6	92m14s	A. Vidal	shot	1.00	+ 0.83

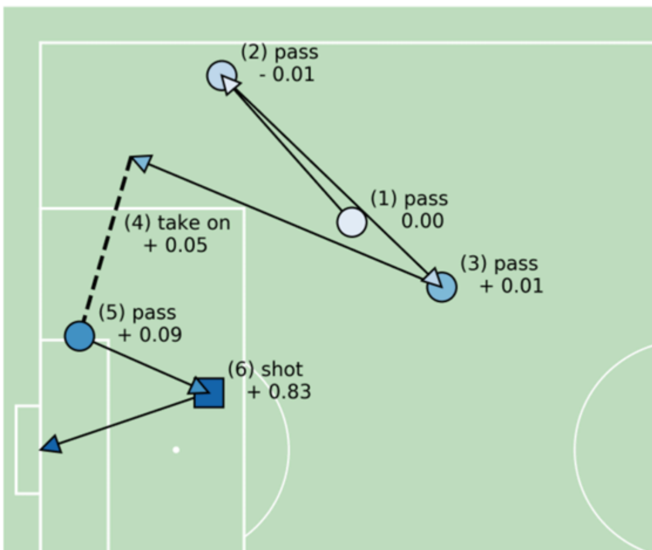
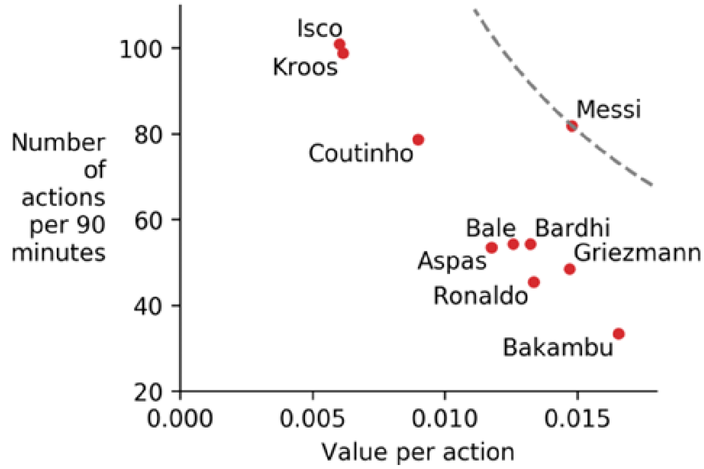


Figure 1: The attack leading up to Barcelona's final goal in their 3-0 win against Real Madrid on December 23, 2017.



(c) Top-10 players in the Spanish league

Figure 2

Additionally, the thesis explains the Soccer Player Action Description Language "as an attempt to unify the existing event stream formats into a common vocabulary" [1], which will be delved into more deeply in the data section. It also briefly describes how VAEP can be used to characterize an athlete's playing style, which can have extremely useful applications in player recruitment, and attempts to spot potential future stars of the sport with the metric. It claims that the uniqueness of VAEP comes from how it encompasses all actions that occur on the pitch rather than just the select few that are often the focus (shots, passes, etc.) and accounts for the context wherein the action occurs.

"Actions Speak Louder than Goals" gives an incredible platform for future research into the relatively new arena of event-level soccer data. As Decroos and his colleagues mention, there is no "ground truth" to what numerical value each action throughout the course of a match holds which allows for a broad range of directions to expand on this quantification process. For this project, the topics of how improved data affects the accuracy of the VAEP model and why certain actions end up with relatively high, or low, VAEP scores are examined. In regards to the latter, the Decroos thesis does not include any analysis on the makeup of the model itself outside of solely studying its output and its potential applications. This paves the way for this work, which attempts to deconstruct the model into the variables which make up an action and then determine whether there is insight to be gained through identifying any important variables at play.

Another source that plays an essential part in enabling this research was the SoccerAction python package documentation website [5]. While not a typical source for a paper, the clear manner in which the package is described, including great explanations of how the SPADL data setup and the VAEP model itself can be used through code, gives a strong platform on which to start this analysis from and thus should be highlighted. The documentation makes it possible to go through the source code and alter it to both fit the V3 data as well as figure out exactly where in the code the model building takes place so the process can be investigated for the variable importance analysis. Sports analytics often faces the problem of having its most cutting-edge analysis be private due to the competitive nature of the athletics landscape; however soccer analytics has grown on a very similar timeline with the desire for open-source research which has enabled tools such as SoccerAction to be created and shared.

Data

The bedrock of this thesis is built upon a dataset gathered through the Duke women's soccer team's access to the Wyscout API. This data acquisition process requires a two-step API call: an initial call to gather all available ACC matches information, followed by subsequent calls for each match to collect detailed event data. This approach ensures a complete dataset encompassing every individual action that occurred on the pitch throughout the ACC matches for the 2022-2023 and 2023-2024 seasons. Given the Duke student analyst team's past implementation of VAEP using V2 data, the necessary data cleaning functions which were already created could be adapted for the newer V3 data. This prior knowledge and code base is crucial in efficiently processing and preparing the ACC data for VAEP model training.

The decision to focus solely on ACC data from the two most recent seasons was strategic, aiming to maintain a consistent baseline for assessing player abilities within the context of their conference. This choice acknowledges the vast disparities in play styles and player skill levels across Division 1 women’s soccer, and is made upon a suggestion by a contact on Duke’s coaching staff. The introduction of V3 data in fall 2022 marked the beginning of this data accumulation phase, setting a new precedent for depth and detail in event-level collegiate women’s soccer data.

The Wyscout event data's format evolved significantly from V2 to V3. While V2 offers basic action-level data - such as broad play types, starting action positions, and subtype identification - V3 introduces a richer layer of detail. This newer version includes written action subtypes, angles, both starting and ending locations of passes, and duels between players, among other enhancements. These additions not only improve one’s ability to understand exactly what occurred on the pitch through the data but also provide a wealth of context that could potentially refine the accuracy of the VAEP model. Shown in Figure 3 and Figure 4, respectively, are one example of actions represented in the V2 format and then only a portion of the many columns that make up the V3 dataset. Notice how much more specific the V3 format gets, especially on specific actions like passes, shots, and duels.

Game ID	Action ID	Match Time (seconds)	Team ID	Player ID	Start X	Start Y	End X	End Y	Event ID	Bodypart ID	Action Type ID	Result ID
5517601	1	0.00269	61493	559277	51.45	34.68	39.90	26.52	1805426688	0	0	1
5517601	1	0.00485	61493	760252	39.90	26.52	71.40	0.00	1805425095	0	0	0
5517601	1	0.01237	61483	762917	75.60	0.00	79.80	10.88	1805424963	2	2	1
5517601	1	0.01479	61483	798102	79.80	10.88	72.45	7.48	1805424964	0	0	1
5517601	1	0.01850	61483	762917	72.45	7.48	91.35	12.24	1805424965	0	0	1

Figure 3: V2 Data

Action ID	Match Period	Minute	Second	Action Type	Action Location	Pass	Shot	Ground Duel	Aerieal Duel
1830086247	1H	6	51	{'primary': 'duel', 'secondary': ['aerial_duel', 'recovery', 'counterpressing_recovery']}	{'x': 61, 'y': 23}	NaN	NaN	NaN	{'opponent': {'id': 688145, 'name': 'H. Hershfelt', 'position': 'RDMF', 'height': 172}, 'firstTouch': True, 'height': 175, 'relatedDuelId': 1830086316}
1830086250	1H	6	54	{'primary': 'interception', 'secondary': ['progressive_run', 'carry']}	{'x': 77, 'y': 10}	NaN	NaN	NaN	NaN
1830086329	1H	6	54	{'primary': 'duel', 'secondary': ['defensive_duel', 'ground_duel']}	{'x': 18, 'y': 89}	NaN	NaN	{'opponent': {'id': 803081, 'name': 'J. Dudley', 'position': 'CF'}, 'duelType': 'defensive_duel', 'keptPossession': None, 'progressedWithBall': None, 'stoppedProgress': False, 'recoveredPossession': False, 'takeOn': False, 'side': 'right', 'relatedDuelId': 1830086255}	NaN
1830086255	1H	6	56	{'primary': 'duel', 'secondary': ['dribble', 'ground_duel', 'offensive_duel']}	{'x': 82, 'y': 11}	NaN	NaN	{'opponent': {'id': 759695, 'name': 'M. Duff', 'position': 'LCB'}, 'duelType': 'dribble', 'keptPossession': True, 'progressedWithBall': True, 'stoppedProgress': None, 'recoveredPossession': None, 'takeOn': False, 'side': 'left', 'relatedDuelId': 1830086329}	NaN
1830086256	1H	7	0	{'primary': 'pass', 'secondary': ['short_or_medium_pass', 'shot_assist']}	{'x': 84, 'y': 24}	{'accurate': True, 'angle': 110, 'height': None, 'length': 12, 'recipient': {'id': 689757, 'name': 'J. Echegini', 'position': 'LW'}, 'endLocation': {'x': 80, 'y': 40}}	NaN	NaN	NaN
1830086337	1H	7	0	{'primary': 'duel', 'secondary': ['defensive_duel', 'ground_duel']}	{'x': 20, 'y': 60}	NaN	NaN	{'opponent': {'id': 689757, 'name': 'J. Echegini', 'position': 'LW'}, 'duelType': 'defensive_duel', 'keptPossession': None, 'progressedWithBall': None, 'stoppedProgress': True, 'recoveredPossession': False, 'takeOn': False, 'side': 'right', 'relatedDuelId': 1830086261}	NaN
1830086261	1H	7	1	{'primary': 'duel', 'secondary': ['dribble', 'ground_duel', 'offensive_duel']}	{'x': 80, 'y': 40}	NaN	NaN	{'opponent': {'id': 688145, 'name': 'H. Hershfelt', 'position': 'RDMF'}, 'duelType': 'dribble', 'keptPossession': True, 'progressedWithBall': False, 'stoppedProgress': None, 'recoveredPossession': None, 'takeOn': False, 'side': 'left', 'relatedDuelId': 1830086337}	NaN
1830086262	1H	7	2	{'primary': 'shot', 'secondary': ['goal', 'opportunity']}	{'x': 79, 'y': 44}	NaN	{'bodyPart': 'right_foot', 'isGoal': True, 'onTarget': True, 'goalZone': 'gt', 'xg': 0.05795, 'postShotXg': 0.07749, 'goalkeeperActionId': 1830086339, 'goalkeeper': {'id': 688133, 'name': 'H. Mackiewicz'}}	NaN	NaN

Figure 4: V3 Data

Transitioning the V3 data to a format compatible with the VAEP model requires converting it to SPADL (Soccer Player Action Description Language) through the intermediary step of using the Kloppy Python package. Kloppy [6] serves as a great tool for working with varied soccer event data formats, offering serialization and deserialization capabilities, standardized data models, and data transformation functions. However, adapting Kloppy for the specific needs of this project entails diving into its source code to edit the Wyscout V3 deserializer to handle peculiarities like team formations post-red card incidents and ensuring player identifiers are correctly formatted. Once the source code is tweaked to successfully transform the Wyscout data forming the basis of this research into the Kloppy format, it is transformed again into SPADL as the input to the VAEP model.

SPADL conceptualizes a game as a series of on-ball actions, each defined by a tuple of twelve attributes, including action type, result, and body part used, among others. This structured approach allows for a granular analysis of game events, with actions categorized into types like passes, shots, or dribbles, and outcomes marked as successful or failed. The adaptation of V3 data into SPADL is a critical step in preparing the dataset for analysis with the VAEP model, enabling a detailed examination of player actions and their impact on the game. A brief SPADL data dictionary [5] is shown in Figure 5.

Attribute	Description
game_id	the ID of the game in which the action was performed
period_id	the ID of the game period in which the action was performed
seconds	the action's start time
player	the player who performed the action
team	the player's team
start_x	the x location where the action started
start_y	the y location where the action started
end_x	the x location where the action ended
end_y	the y location where the action ended
action_type	the type of the action (e.g., pass, shot, dribble)
result	the result of the action (e.g., success or fail)
bodypart	the player's body part used for the action

Figure 5: SPADL Data Dictionary

By meticulously gathering and processing this event-level data, this research lays the groundwork for a thorough analysis of soccer player performance. The dataset not only facilitates an updated application of the VAEP model but also ensures the scalability of this methodology to include broader datasets from various leagues and competitions in future work.

Methodology

This section delineates the methodological structure to evaluating and enhancing the VAEP model through the analysis of event-level data in ACC women's collegiate soccer. A rigorous approach is undertaken to ensure the reliability and validity of the findings, starting with the division of the dataset into training and testing subsets.

Data Partitioning

The dataset, comprised of ACC matches from the 2022 to 2023 seasons, is partitioned using an 80-20 train-test split. This allocation method designates the first 80% of games chronologically as the training data, which is used to train the VAEP model. The most recent 20% of games is reserved as the testing data, serving to evaluate the model's predictive accuracy and generalizability to unseen data. This split is chosen to mirror the temporal progression of the seasons, so that the model can be tested on the most current data available.

VAEP Model Overview

The VAEP model operates on a principle of quantifying the value of individual player actions within a soccer match. Building upon the foundational goals outlined previously, the model assesses each action's impact on the game's outcome through a dual probabilistic framework. This framework consists of two distinct models: one for scoring and another for conceding. The scoring component calculates the probability that a given action will lead to a goal for the team executing the action, while the conceding model evaluates the probability that an action could result in the team giving up a goal. Both models consider various factors, including the type of action, the action's location on the pitch, the game context, and other relevant parameters that might influence the likelihood of scoring or conceding.

Below is a mathematical notation of the value for the *i*th game state in a given soccer match:

$$V(S_i) = P_{score}(S_i, t) - P_{concede}(S_i, t)$$

where the probabilities shown are the chance team *t* which possesses the ball in game state S_i scores or concedes, respectively, in the next *k* actions. A game state is defined as a set number of actions, a parameter which we will call *j*. So when *j*=3, for example, game state 20 is defined as $S_{20} = \{a_{18}, a_{19}, a_{20}\}$ where a_{20} is the 20th action that occurs in a game and more generally a_i represents action *i* in a soccer match. For estimating these probabilities, each game state is assigned a value of 1 if the team with possession scores (or concedes, depending on the model) a goal in the next *k* actions and gets assigned a value of 0 otherwise. Continuing with the example of game state 20, if a goal is scored in any action between a_{20} and a_{30} then S_{20} would be assigned a positive label in the scoring model. The probabilities are then used to generate offensive and defensive values, respectively, for each action, by calculating the difference in the chance a team scores or concedes in consecutive game states:

$$\Delta P_{\text{score}}(a_i, t) = P^k_{\text{score}}(S_i, t) - P^k_{\text{score}}(S_{i-1}, t)$$

Finally, we repeat the same process for conceding goals and take the difference between the change in probability in scoring and that for conceding to obtain a total VAEP value for each action.

The model learning process is a “gradient boosted binary classifier trained on historical data to predict how a game state will turn out based on what happened in similar game states that arose in past games” [5]. Specifically, XGBoost [7] is selected as the underlying prediction method, which will be discussed in more detail below under model selection. Thus, each row in the data directly fed into the model consists of all game action information and its context, as well as the same for each of the previous *j* - 1 actions. The greater the *j* value, the farther back in the game the model will go. This allows the model to potentially pick up on patterns in play that lead to scoring or conceding goals. As *k* increases, reflecting a broader look-ahead window, the model's ability to incorporate longer-term outcomes of actions enhances, allowing for a more comprehensive assessment of each action's potential impact over an extended sequence of gameplay. A complete analysis on selecting optimal *j* and *k* values for the dataset, and a setup that can be used in other event-level datasets, is examined after the model overview.

The output of the VAEP model is a score assigned to each action, representing its overall value in terms of influencing the game's outcome. This score is derived by calculating the difference between the probabilities generated by the scoring and conceding models for each action. This score encapsulates the action's dual potential to contribute to scoring while mitigating the risk of conceding, or vice-versa, offering a nuanced measure of its value.

The VAEP score, thus, serves as a quantitative assessment of an action's effectiveness, providing coaches, analysts, and researchers with a powerful tool to evaluate player performance and make informed decisions. By applying this methodology to the analysis of event-level data from women's collegiate soccer, insights can be uncovered that enhance our understanding of what makes certain actions more valuable than others and how these actions contribute to a team's success on the field.

Comparing V2 vs. V3 in VAEP

A pivotal aspect of this research involves the comparative analysis of model outputs based on two versions of Wyscout event data: V2 and V3. This comparison aims to assess the impact of data version updates on the performance of the VAEP model, thereby providing insights into Wyscout’s evolution of data quality and its implications for soccer analytics.

To conduct a thorough comparison, the previously described training and testing sets are utilized, ensuring consistency in the evaluation process. The methodology entails training separate VAEP models for both the V2 and V3 datasets, applying a series of *j*-values—3, 6, and 9—and *k*-values—3, 6, 10, and 13—to explore the effect of different variable sizes on the model's performance. These values are chosen to span the reasonable range of *j*’s and *k*’s, respectively. The lowest value tested was 3 for each as the advantages of the VAEP model are realized when accounting for more than just the passing and shooting actions which are directly related to a goal. The highest value tested for *j* is at 9 since the runtime for generating the dataset and then training the model became unreasonable past this point, while for *k* 13 is the high point.[1]

Each model is then tested against the reserved testing set, with the outcomes evaluated based on two key metrics: Brier Score and Area Under the Receiver Operating Characteristic (AUROC). The Brier Score, a measure of the accuracy of probabilistic predictions, quantifies the mean squared difference between the predicted probability assigned to the possible outcomes and the actual outcome. AUROC, on the other hand, measures the ability of a model to distinguish between classes. An AUROC score closer to 1 indicates a high degree of separability achieved by the model. Since the AUROC is a measure of the true positive rate compared to the false positive rate for different thresholds, it is a strong metric to use in evaluating imbalanced datasets. This is relevant in this analysis since goals are such a rare event in a soccer match and they act as the response variable in the VAEP model. On the other hand, Brier Scores use the predicted probabilities from the model output rather than solely looking at the binary classification results, which places more importance on how well the model does from a probabilistic standpoint. However, this metric does not have an inherent mechanism for handling highly imbalanced datasets well.

The process of generating an XGBoost model includes a degree of randomness, primarily due to the stochastic nature of the algorithms it employs, such as random sampling of data points and features when building trees [7]. To account for this variability and ensure the robustness of the study's findings, each model configuration is trained and tested five times. The average Brier Score and AUROC are then calculated for each model and *k*-value, providing a stable basis for comparison.

The results of this comparative analysis highlight the differences in performance between the VAEP models trained on V2 and V3 data across the chosen *j*-values and *k*-values for both the scoring and conceding models. This comparison sheds light on the impact of the enhanced detail and complexity present in V3 data on the model's predictive accuracy and classification ability.

By examining the average Brier Scores and AUROC values, we gain valuable insights into how the updates from V2 to V3 data influence the effectiveness of the VAEP model. This analysis not only contributes to our understanding of the model's adaptability to evolving data formats but also underscores the importance of continuous data refinement and model recalibration in analyzing event data in soccer.

Model Selection

An important step in the methodology of this research is selecting the optimal machine learning model to underpin the VAEP framework. The original VAEP paper utilizes CatBoost as the

---

[1] *k*=13 was chosen arbitrarily based on knowledge of game context as a value to test higher than the *k*=10 used in Decroos's paper

foundational model, known for its high performance in handling categorical features and complex interactions within datasets. Given the close performance metrics between CatBoost and XGBoost documented in the initial VAEP study, a comparative analysis is conducted to determine the most suitable model for this research.

The comparison focused on evaluating both models' performance on the dataset derived from the Wyscout V3 event dataset. The criteria for comparison included not only the predictive accuracy of each model but also their efficiency in terms of computational resources and time required for training and testing. Upon conducting tests on the specified dataset, it is observed that CatBoost exhibits very similar performance in terms of AUROC compared to XGBoost. However, utilizing CatBoost as Decroos did comes at a significant cost in computational efficiency, especially as *j* increases as this enlarges the size of the training and testing datasets. Given the relatively equal test set accuracies between the two models and the considerable difference in execution time, a decision is made to use XGBoost for further analyses. This choice is driven by the need for efficiency considering the extensive *j* and *k*-value analysis and variable importance tests which make up the bulk of this study.

For selecting the *j* and *k*-values for the model to be analyzed for variable importance, values of 3 and 6 are picked, respectively. This choice is predicated on a combination of the strong testing performance it demonstrated as well as the rationale that a relatively low *k*-value (at least compared to the *k*=10 used in the VAEP paper) would enable a stronger variable importance analysis. The assumption was that by allowing the model to consider a smaller range of events preceding a goal, it would be possible to better capture the interactions and dependencies among the variables in a more direct leadup to the scoring event, thereby enhancing the analytical depth of the study.

In summary, the model selection process culminates in the choice of XGBoost, coupled with a *j*-value of 3 and a *k*-value of 6, for conducting the variable importance analysis. This decision balances the need for computational efficiency with the desire to extract meaningful insights from the data, setting the stage for a comprehensive exploration of factors influencing player action value in women's collegiate soccer.

<u>VAEP Variable Importance Analysis</u>

We have discussed how the VAEP model offers a nuanced approach to analyzing player actions within soccer matches, but understanding the relative importance of different variables within this model is crucial for deepening our insights into game dynamics. This section delves into the variable importance analysis conducted to identify which features most significantly influence the VAEP model's predictions.

To explore variable importance, five experimental model runs are executed, each using the *j*=3 and *k*=6 parameters to ensure a consistent context for comparison. These runs include a scoring model trained exclusively on passes, a scoring model focused solely on crosses, and both scoring and conceding models trained on random results. This diverse array of experiments is designed to uncover how specific actions or outcomes contribute to the model's predictive accuracy and to isolate the effect of individual variables on the overall VAEP score. The objective of the passing and crossing models are to see if any interesting correlations appear when taking out the final shooting action that occurs when a goal is scored, thus putting a greater importance on the actions leading up to that shot. The random results models are utilized to look at a more process oriented view of the VAEP model, as the unaltered models rely heavily on outcomes (i.e. whether a shot attempt was successful) to assign values. This makes sense since the purpose of the model is to accurately attribute team success to individual actions, but is not very useful in analyzing variables that correlate with scoring or conceding. By randomly assigning true and false values to these result columns they no longer correlate strongly with the model’s response variable, allowing for the action types and game context to become more prevalent in shaping the model.

The analysis of variable importance is grounded in the concept of Shapley values from cooperative game theory, which provides a principled approach to attributing the contribution of each variable to the prediction of a model [8]. Shapley values offer several advantages over the main alternative measures of feature importance used in conjunction with tree-based models: XGBoost’s built-in feature importance function and permutation feature importance. The XGBoost function calculates the aggregate decrease in model impurity from splits on each feature, averaged across all trees. While informative, this method can obscure the individual impact of features in complex models and does not specify how different values of the key model inputs might affect the response variable [9]. Permutation feature importance assesses importance by evaluating the change in model performance after permuting each feature, identifying those with the most significant impact on accuracy. However, this approach can miss nuances in how features interact within the model [9].

The SHAP python package gives an in-depth explanation of how the SHAP values used in this analysis are derived from the Shapley values described above. At a high level, these SHAP values are “Shapley values applied to a conditional expectation function of a machine learning model” [8]. This measure is unique from the prior two in that it directly estimates the contribution of each feature to the model's output for individual predictions, providing a clear, interpretable measure of variable importance that accounts for interaction effects and the non-linear nature of tree-based models. Unlike the other methods, SHAP values reveal both the magnitude and direction of a feature's impact, offering detailed insights into how specific variables influence model predictions. Figure 6 shows an example of how SHAP values for the top seven most influential variables for a single observation contribute to the overall model prediction for that action, where E[f(x)] is the expected log-odds of a goal being scored in the next *k* actions prior to looking at the specific data from that action and F(x) is the XGBoost model output (also log-odds of scoring in the next *k* actions).

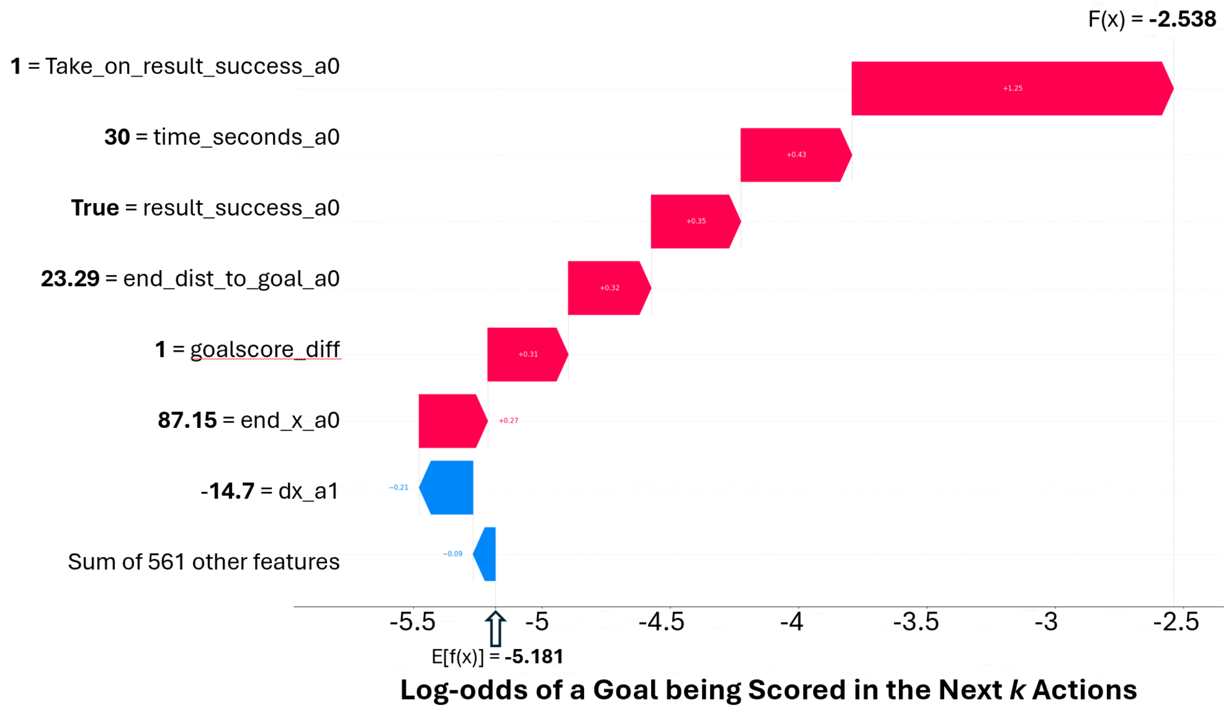


Figure 6: Waterfall Plot of SHAP Values for One Observation

The SHAP methodology identifies the successful take on which occurs in the above action as the most important variable here, with its SHAP value of 1.25 suggesting that it makes a goal being scored in the next *k* actions about 3.5 (or $e^{1.25}$) times more likely to occur.

For visualizing the results of this analysis, the SHAP package's beeswarm plot [10] is employed. This plot type effectively displays the distribution and impact of Shapley values for the top features, organizing them by mean absolute Shapley value to highlight those with the broadest average impact on the model's predictions. Each beeswarm plot is a detailed visual representation of the SHAP values for individual predictors across the dataset. In these plots, each dot represents a single observation for a given feature in the dataset, where the x-position of the dot is determined by the SHAP value of that feature for that particular instance. This positioning reflects how much each feature value contributes to pushing the model's output from the base value (the prediction that would be made if no features were known) to the actual model output. Dots "pile up" along each feature row to show the density of the SHAP values, indicating how frequently similar values occur, thus demonstrating the distribution of impact each feature has across different instances. This piling up or dispersion imparts insight into the consistency of a feature's influence on the model's predictions, highlighting which features generally have more uniform effects and which vary more significantly between observations.

The color of each dot enhances this visualization by corresponding to the original value of the feature, helping to correlate the feature's actual observed value with its impact on the output. For example, in the context of soccer analytics, where each dot is one action, a dot's color might show the angle of a pass, with its position on the pitch in relation to the goal indicating how this particular angle increases or decreases the probability of scoring in that given action. Beeswarm plots excel in presenting an information-dense summary that emphasizes both the average influence of features and how individual data points impact model output. The use of beeswarm plots in this analysis allows for a nuanced understanding of how different variables contribute to the scoring and conceding models, and these graphs will be displayed and examined in the following section. A data dictionary explaining the variable names will be included in the appendix as Figure 16.

The variable importance analysis, underpinned by SHAP values and visualized through beeswarm plots, provides a comprehensive overview of how different actions and outcomes are weighted within the VAEP model. By focusing on the j=3 and k=6 scoring and conceding models, the analysis highlights the critical role of context in determining the value of player actions in soccer matches. This approach not only confirms the significance of key variables but also illuminates the complex interplay between them, offering valuable insights for coaches, analysts, and researchers seeking to understand the intricacies of soccer performance.

Analysis and Discussion

Wyscout Data Improvement

The transition from Wyscout V2 to V3 event data represents a significant evolution in the granularity and quality of information available for Division 1 women's soccer teams. The comparison between models trained on V2 and V3 data utilized the AUROC as the primary evaluation criterion. As discussed above, the AUROC is chosen due to its effectiveness in handling unbalanced datasets, a common characteristic in event-level soccer data, where certain actions (e.g., goals) are much rarer than others (e.g., passes). This metric presents a robust measure of the model's ability to distinguish between classes and accurately predict outcomes. Figure 7 and Figure 8 show the average model performance over five model builds trained on the earliest 80% of games in the ACC women's dataset and tested on the most recent 20% when j is held constant at 3 and k is held constant at 10, respectively. The full output from every j and k combination tested is shown in the appendix.

k	Model Type	V2 AUROC	V3 AUROC	V3 AUROC Improvement
3	concedes	0.930162	0.951997	0.021836
3	scores	0.957101	0.968783	0.011682
6	concedes	0.837349	0.886612	0.049262
6	scores	0.853636	0.856935	0.003298
10	concedes	0.741794	0.793788	0.051994
10	scores	0.785398	0.777370	-0.008028
13	concedes	0.712236	0.745270	0.033034
13	scores	0.752618	0.744161	-0.008457

Figure 7: VAEP performance at *j*=3

j	Model Type	V2 AUROC	V3 AUROC	V3 AUROC Improvement
3	concedes	0.741794	0.793788	0.051994
3	scores	0.785398	0.777370	-0.008028
6	concedes	0.739896	0.793511	0.053615
6	scores	0.785585	0.779724	-0.005861
9	concedes	0.738152	0.793884	0.055731
9	scores	0.783248	0.776421	-0.006827

Figure 8: VAEP Performance at *k*=10

The analysis reveals that the V3-based model exhibited superior performance across most comparisons when *j* is held constant. The improvement is much greater for the conceding model compared to the scoring model. First notice that as the *k*-value increases in Figure 7, the model's test accuracy decreases. This makes sense given that it should be more difficult to predict that an event which occurs 12 actions prior to a goal will ultimately lead to one as opposed to looking at an event only three actions before a goal is scored. On the other hand, Figure 8 demonstrates that adding additional actions into a game state, and thus increasing the number of independent variables that are used in model training, does not have a discernible correlation with any changes in VAEP model accuracy. This is surprising given that expanding the context an action is placed in should potentially improve the amount of variation in goal scoring (or conceding) opportunities the model can account for. Rather, Figure 8 shows that only examining the last three actions provides a sufficient amount of context in defining a game state. Further evidence to this point is that *j*=3 has been empirically found to work well with other soccer datasets [1].

Interestingly, both the V2 and V3 versions of the model demonstrated slightly better performance at *j*=3 and *k*=10 (the parameters used in the original VAEP study's model) than the model trained on a significantly larger dataset encompassing various men's leagues, as detailed in Decroos's paper [1]. The V3 data AUROC for this model is around 0.777 for scoring and about 0.794 for conceding while these respective values are 0.7693 and 0.7313 [1], respectively, for the model trained on multiple years of European soccer data across multiple leagues. The results from that model test as a whole are shown in the appendix in Figure 17. This outcome underscores the VAEP model's applicability to the ACC women's soccer dataset, suggesting that the specific context and characteristics of this dataset may be particularly well-suited to the VAEP analysis framework, especially from a defensive standpoint given the relative improvement of the conceding model. While these findings are encouraging, it is important to note that there is likely much less variability between teams and players in ACC women's soccer compared to in the many different leagues analyzed in Decroos's study.

Additionally, it is possible that over the course of the six-year timeframe which makes up the original paper's dataset play styles changed, while this dataset only looks at a smaller two-year period over which team strategies are much more likely to remain similar. Both of these ideas might impact the amount of variability the VAEP model is able to account for.

The observed improvements in model performance from V2 to V3, and the comparison with previous studies, offer compelling evidence of the positive impact of enhanced data quality on soccer analytics. The more detailed and context-rich V3 data not only improves the model's predictive accuracy, specifically in the conceding model, but also suggests that as event-level data continues to evolve and improve, we can anticipate corresponding enhancements in the accuracy and utility of VAEP and similar analytical tools. This progression points to a promising future for soccer analytics, where improved data quality enables deeper insights into the game and more effective support for decision-making by coaches, analysts, and other team consultants.

<u>VAEP Variable Importance</u>

This section synthesizes takeaways from the VAEP variable analysis, emphasizing unexpected correlations and their potential applications for coaches and players. The analysis utilizes the SHAP values displayed in the beeswarm plots to elucidate the impact of individual variables on the likelihood of scoring or conceding goals, focusing on ones that reveal less intuitive aspects of gameplay and strategy. As a reminder, dots further away from zero on the x-axis of the beeswarm plots indicate a greater effect on model output for the observation that dot is a part of.

Initial observations from the base scoring and conceding models (shown in Figure 9 and Figure 10) highlight the pivotal role of outcome variables, such as "result_success_a0" or "result_fail_a0," which is a binary classifier of whether the action in question is succeeds or fails, respectively, in influencing goal-scoring probabilities. Intuitively, the likelihood of scoring increases with successful actions near the opponent's goal, while the probability of conceding is heightened by unsuccessful actions in defensive areas.

However, the analysis also demonstrates the influence of positional and contextual factors on game outcomes. Notably, the timing within a match and the current score have relatively important impacts on scoring probabilities, suggesting that game dynamics shift based on these factors. Figure 9 insinuates that goals are more likely to come later in games (as the time increases). This finding opens avenues for teams to try and exploit the game becoming more open as time goes on. An intriguing aspect revealed by the base conceding model is the correlation between a high level of movement between the third to last and last actions and giving up a goal. This implies that quick counterattacks where the ball advances up the pitch rapidly may be particularly effective in the ACC women's soccer context, underscoring an area for potential tactical emphasis in training and game planning upon further research. This is further supported by the negative correlation shown between "Team_2", which is a binary indicator of whether the team with possession had possession two actions prior, and a goal being scored in the next six actions, as goals occurring soon after turnovers are a sign of a counter attack goal.

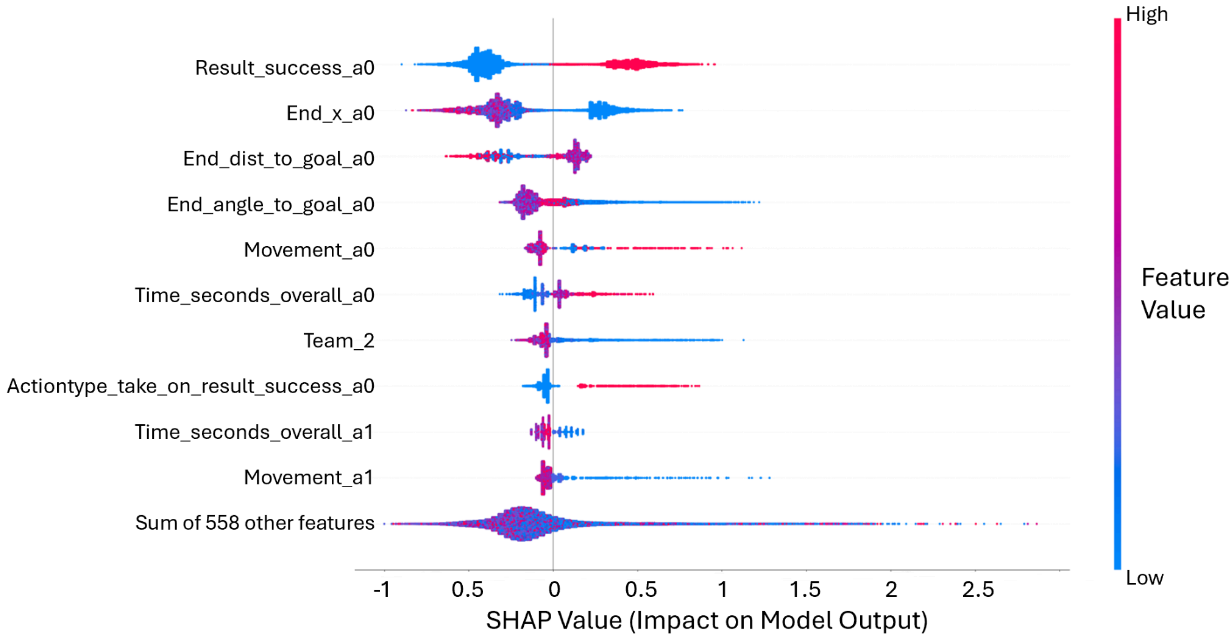


Figure 9: *j*=3 and *k*=6 Scoring Model Beeswarm Plot

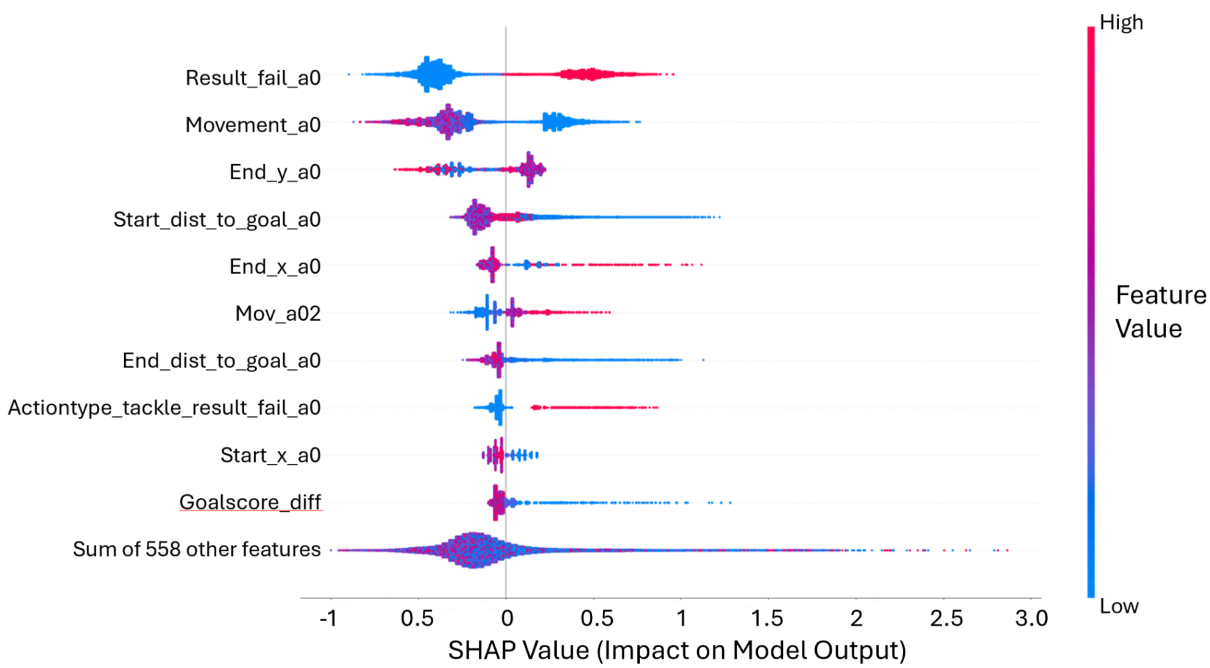


Figure 10: *j*=3 and *k*=6 Conceding Model Beeswarm Plot

The analysis of models focused on passing and crossing unearths correlations that reinforce the value of central play in creating scoring opportunities. A lower end angle to the goal's center, identified in both models, indicates the effectiveness of actions that penetrate through the middle of the pitch. Moreover, the relevance of angles and positioning at both the start and end of a cross suggests an importance in delivering the ball both from and to optimal areas on the pitch. Specifically, the plot highlights that to maximize scoring chances from crosses they should be played from wide areas to the center of the box. This finding aligns with conventional wisdom on how and when crosses should be created in a match. The passing model is similar to the base scoring model in that game context (time and score) has a relatively strong correlation with a goal being scored, and otherwise it does not display any other correlations of note.

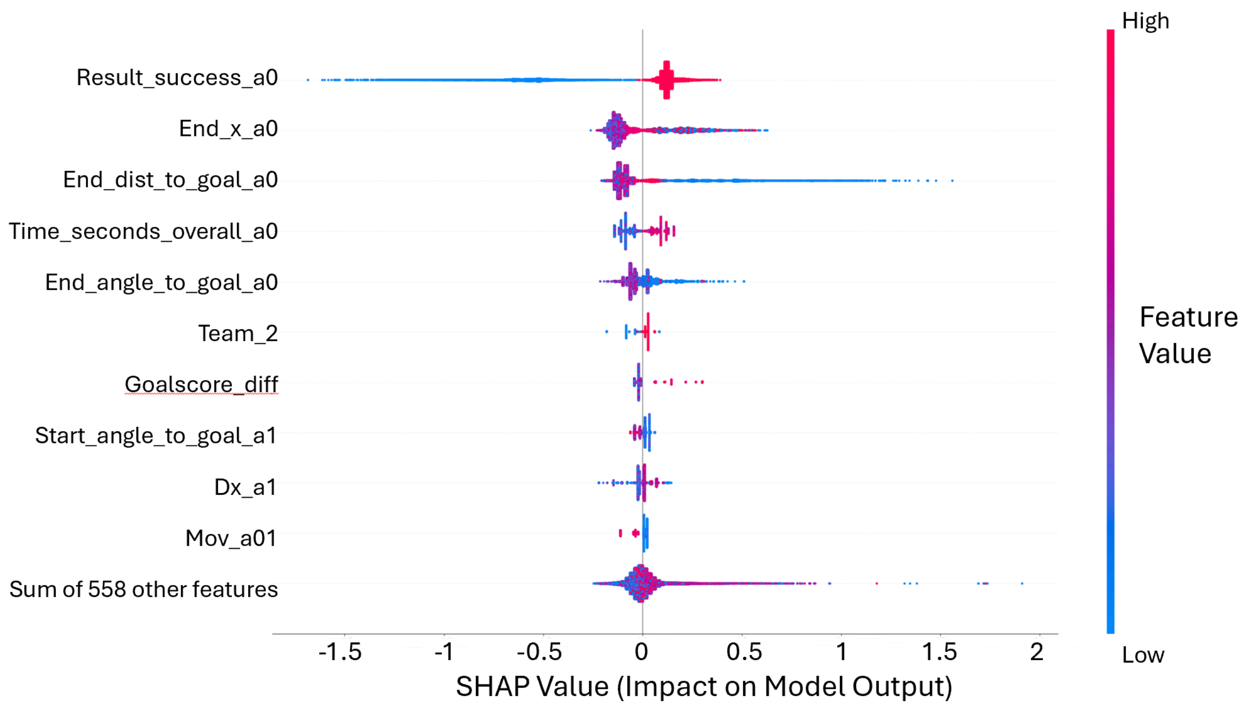


Figure 11: Passing Scoring Model

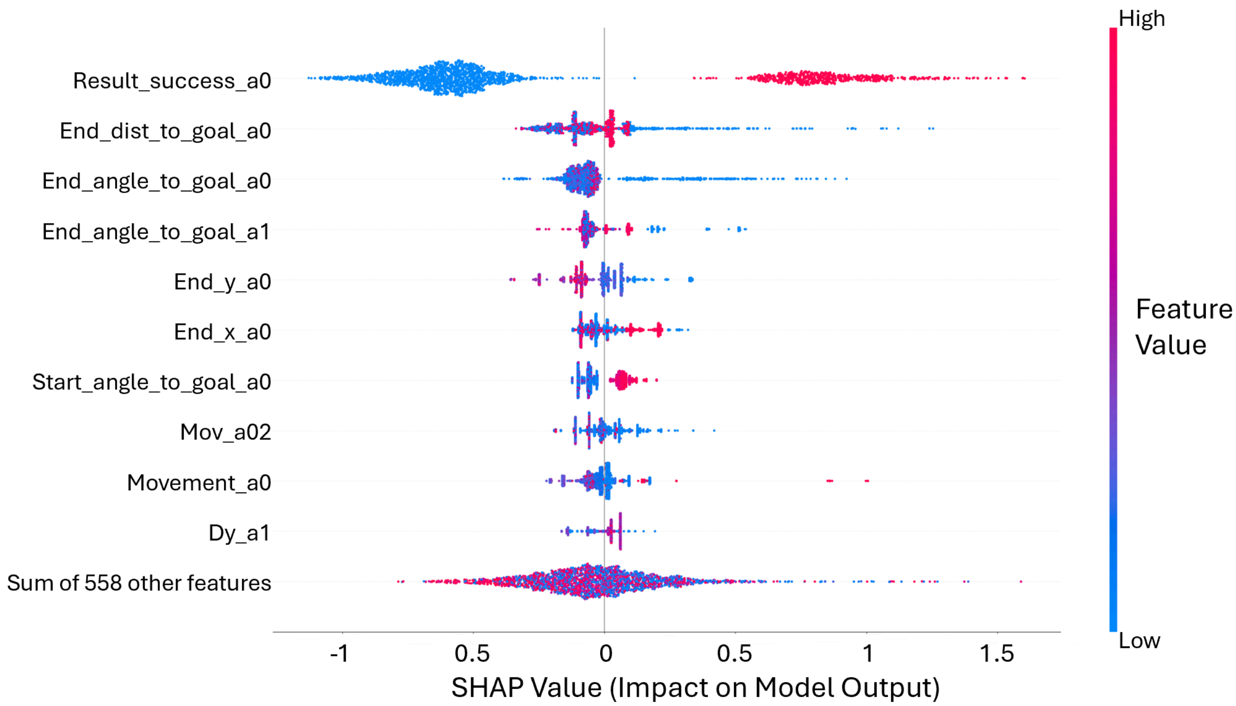


Figure 12: Crossing Scoring Model

Models trained on random results, devoid of outcome success indicators, demonstrate the importance of ball movement in both scoring and conceding goals. Notably, the scoring model exhibits a strong positive relationship with the movement in an action leading up to a goal, while the conceding model shows the reverse correlation. This finding gives credence to the idea that counterattacks which traverse the pitch seem to be a good tactic in ACC women's soccer. The lack of movement before a concession of a goal is most likely due to that often occurring when a team loses possession off a dribble, which is supported by the relevance of tackles shown in the plot even when the outcome of these are essentially unknown to the model. The fact that both of these variables show up in the conceding model could indicate that coaches should focus on mitigating dribbling risk close to their own goal, either by clearing the ball up field more often or focusing on quick passing when in possession on their own half of the field. These insights from the random results models suggest a more complete understanding of how both offensive and defensive strategies can be optimized.

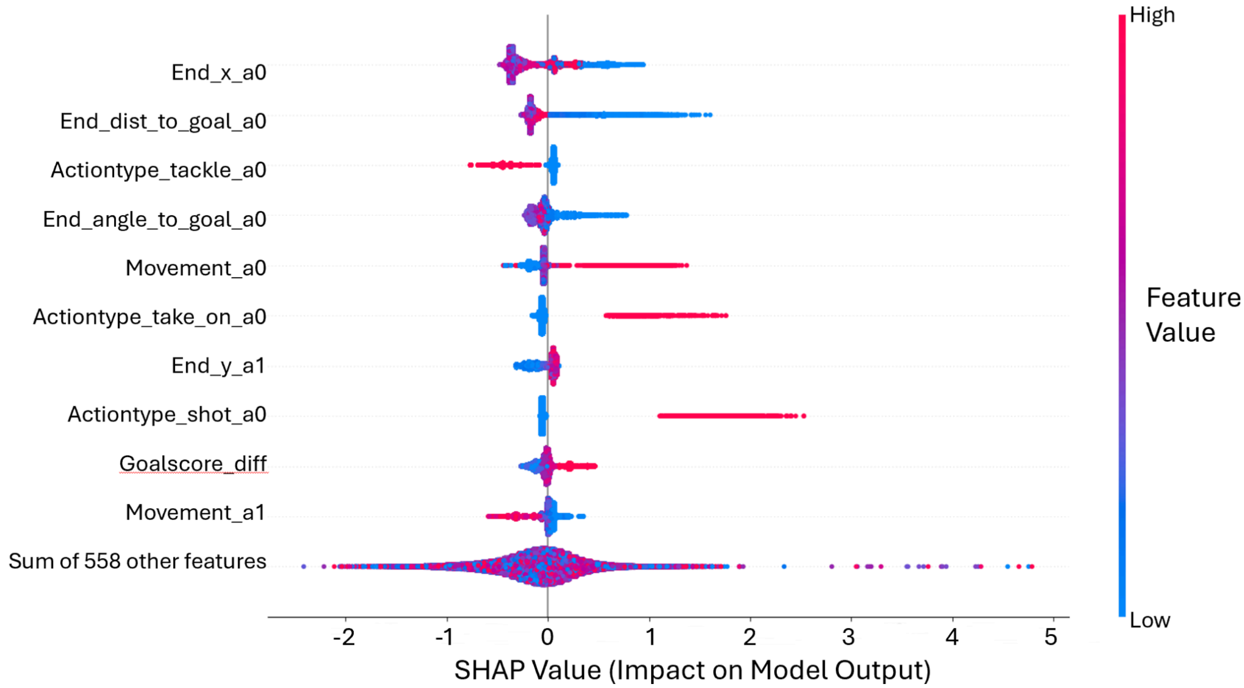


Figure 13: Random Results Scoring Model

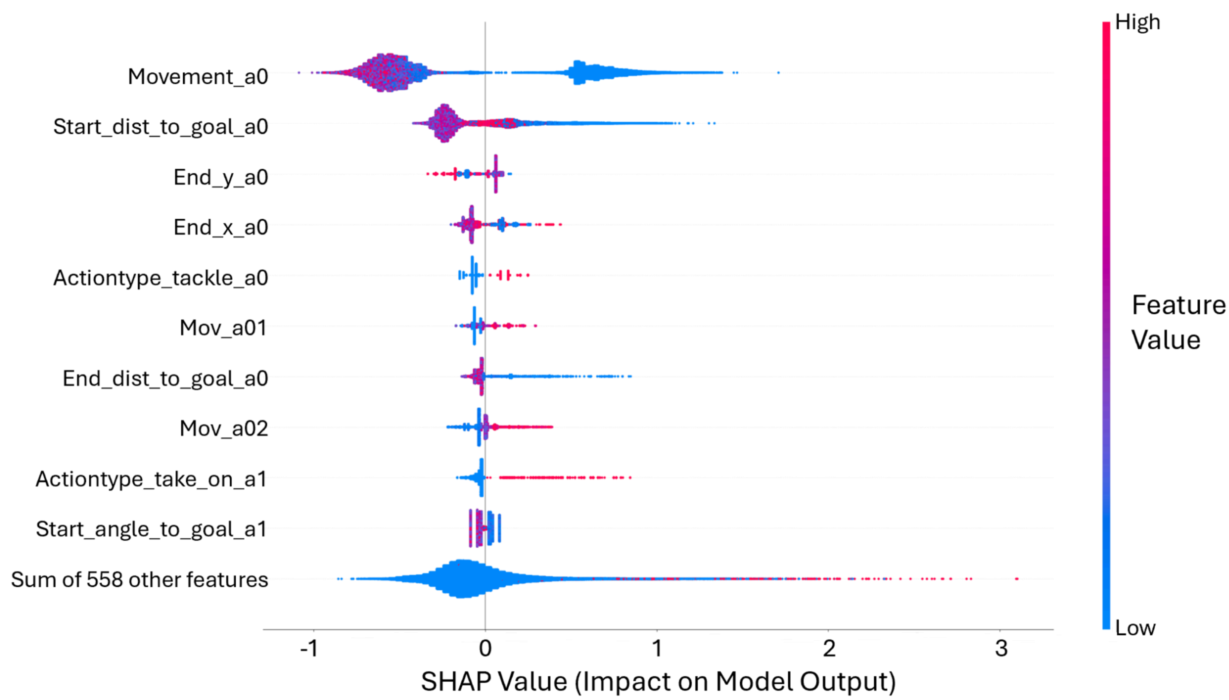


Figure 14: Random Results Conceding Model

The findings from this variable importance analysis illustrate the utility of SHAP values in conjunction with the VAEP model for generating actionable insights into soccer gameplay. While the specifics of this study provide targeted recommendations for enhancing performance in ACC women's soccer, the methodology exemplifies a broader approach to analyzing soccer that can be applied across different leagues and levels of play. The detailed examination of variable importance not only identifies areas for tactical innovation but also underscores the potential for data-driven strategies to refine coaching and training practices.

In summary, this examination of VAEP variable importance through SHAP values has unveiled nuanced understandings of the factors influencing game outcomes in ACC women's collegiate soccer. By focusing on both expected and unexpected determinants of scoring and conceding probabilities, this research offers a foundation for strategic enhancements that are deeply informed by data analytics. As the field of soccer analytics continues to evolve, the application of such sophisticated tools can become increasingly integral to the development of competitive strategies and the broader understanding of the game.

Limitations

This research, while offering insights into the application and enhancement of the VAEP model in analyzing women's collegiate soccer, is not without its limitations. These constraints are primarily related to the scope of the dataset, the generalizability of the findings, and the nature of the analysis conducted. Acknowledging these limitations is necessary for understanding the context in which the discoveries should be interpreted and for guiding future research in the field.

<u>Sample Size and Scope</u>

One of the primary limitations of this study is the relatively small sample size. The dataset encompasses only two full seasons of data from the Atlantic Coast Conference, totaling just 151 games. For the purposes of model training, only 80% of these games (120 games) are utilized. This sample size is modest, especially when compared to larger datasets used in similar studies, such as the dataset consisting of 11,565 games played in top divisions throughout Europe over a six-year period used in Tom Decroos' paper [1]. The English Premier League alone features 380 games in a single season, highlighting the constrained scope of this project's dataset. Given this context, the initial results, while encouraging, should be interpreted with caution. A larger sample size can provide a stronger foundation for confidence in the study's conclusions. The current dataset's limited scope, restricted to a single conference within Division 1 women's soccer over just the two seasons in which the data is available, narrows the applicability of the conclusions drawn from this research.

<u>Generalizability</u>

The study's takeaways are directly applicable only to ACC women's soccer, given the dataset's composition. The variability in the level of play across different conferences and divisions raises questions about the generalizability of the variable importance analysis conducted in this study to other teams and leagues in Division 1 women's soccer. It remains uncertain whether the insights gained can be applied to this broader context, and they are very likely not transferable to the men's game. However, it is noteworthy that the code framework developed in this research is adaptable and can be applied to larger and more diverse datasets, potentially broadening the study's relevance in the future.

Lack of a Causal Analysis
Another limitation is the correlational nature of this analysis. This study identifies interesting correlations that merit further exploration and may offer valuable insights for coaches in terms of game strategy and player traits. However, it does not establish causal relationships. The findings should thus be viewed as starting points for more in-depth investigations rather than conclusive evidence of specific dynamics within the game.

Future Work

The results and methodologies presented in this research open several avenues for future work, aiming to refine and expand the application of the VAEP model in soccer analytics. The following areas are identified as promising directions for further investigation, offering potential to enhance the model's accuracy, applicability, and utility across different soccer contexts.

Expanding Dataset Use
The capability to apply the developed code framework to broader datasets is a compelling direction for future research. Numerous publicly available datasets, such as those offered through Statsbomb Open Data github repository [11] or via the load_open_data() function in the Kloppy package [6], provide access to professional-level event data across various leagues. Leveraging these data sources would not only increase the sample size but also enable the application of VAEP analyses to other leagues, thereby enhancing the generalizability and relevance of the conclusions. The ease of acquiring publicly available data makes this the most direct and feasible next step for expanding the research scope.

Hyperparameter Tuning and Analysis
Currently, the model operates on the base configuration of XGBoost, a decision-tree-based ensemble machine learning algorithm that is efficient for predictive modeling. Incorporating a hyperparameter tuning framework represents a next step in optimizing the model's performance. Hyperparameter tuning can systematically explore a range of configurations to find the most effective model settings for a specific dataset, potentially improving accuracy and yielding richer insights from the variable analysis process. Focusing on the implementation of such a framework can refine the model's predictive capabilities and allows for an evaluation on how these enhancements affect the different variable importance experiments presented.

Simulations of Game Sequences
Another intriguing area for future exploration involves the simulation of game sequences based on VAEP model outputs. If game sequences can be accurately generated in a manner that reflects the contextual underpinnings of the VAEP model, this could significantly advance the understanding of the model's robustness and applicability. Such simulations would allow for the

hypothetical modeling of game outcomes based on different actions and strategies, providing valuable insights into the dynamics of soccer matches. This approach could be particularly beneficial for leagues and conferences, like the ACC women's conference from this study, where comprehensive event data has only been available for a limited duration. By simulating game sequences, researchers and analysts could explore hypothetical scenarios and strategies, enriching the strategic planning and evaluation processes within teams and leagues.

Through these avenues of future work, the research community can continue to build on the foundation this thesis lays down, driving forward the development of soccer analytics and enhancing the understanding of the game's complexities. The integration of hyperparameter tuning, expansion of dataset sources, and exploration of game sequence simulations present exciting opportunities to elevate the sophistication and impact of VAEP model applications in the field.

Conclusion

This research embarks on a journey to explore the depths of soccer analytics through the lens of the Valuing Actions by Estimating Probabilities model, with a specific focus on advancing the model's application and understanding within the context of ACC women’s soccer. By transitioning from the Wyscout V2 to the more detailed V3 data, this thesis has not only updated and refined the VAEP model but also provided a deeper analysis of the variables that significantly influence the game's outcomes, offering valuable insights into player and team performance.

The findings from this study highlight the importance of context and positional play in influencing the likelihood of scoring or conceding goals, with certain correlations offering unique perspectives on game strategy and player evaluation. The most notable conclusion which appears throughout the analysis of the beeswarm plots is the perceived effectiveness of quick counterattacks in ACC women’s soccer. This key insight can inform coaching strategies and player development within the conference. In addition, the superior performance of the V3-based VAEP model underscores the critical role of enhanced data quality in improving the accuracy and applicability of soccer analytics tools in general.

While these insights are mainly applicable to ACC women’s soccer, the project also offers broader implications for the field of soccer analytics, especially within the collegiate landscape where access to rich data sources like Wyscout V3 is becoming increasingly prevalent. The methodology and key takeaways this study presents provide a blueprint for leveraging event-level data to uncover nuanced understandings of player actions and their impact on game outcomes. Coaches, analysts, and other decision-makers in the sport can utilize these insights to refine their tactical approaches, player evaluations, and training methodologies, thereby enhancing the overall competitiveness and strategic sophistication of their teams.

Overall, this thesis contributes to the growing body of soccer analytics research by offering a strong mechanism for evaluating player performance and shaping game strategies through the

VAEP model. By demonstrating the potential of improved data quality and advanced analytical techniques to yield implementable conclusions, hopefully the groundwork has been laid for future research and development in the field. As soccer analytics continues to evolve, the approaches and findings from this study will serve as a valuable resource for enhancing our understanding of the beautiful game and the countless actions that define its outcomes.

Acknowledgements

First off, thank you to my thesis advisor, Dr. Jerry Reiter, whose dedication and expertise have been pivotal in the completion of this work. His willingness to meet with me weekly, keep me on track, and provide continuous guidance has profoundly shaped this research process. Thank you to Leo Biral for his instrumental role in helping me select a meaningful topic and providing strategic direction for my project. His insights and advice have been critical in giving my work a purposeful trajectory. Thank you to Coach Kieran Hall for the opportunity to assemble an analyst team for the Duke Women's Soccer team two years ago, whose work laid the groundwork for my thesis. Thank you to Dr. Joan Combs Durso for providing much of the organization around the thesis in the statistical science program. Lastly, thank you to my mom for encouraging me to pursue graduation with distinction in the first place, I couldn’t have done this without you.

References

[1] Tom Decroos, Jan Van Haaren, Lotte Branson, and Jesse Davis. 2018. Actions Speak Louder than Goals: Valuing Player Actions in Soccer.
[2] Wyscout Data. https://footballdata.wyscout.com/
[3] Vegard Rødseth Bjertnes, Olav Nørstebø, and Eirik Vabo. 2016. Valuing Individual Player Involvements in Norwegian Association Football. *Norwegian University of Science and Technology*.
[4] Lotte Bransen, Jan Van Haaren. 2018. Measuring football players' on-the-ball contributions from passes during games.
[5] SoccerAction. 2020. https://socceraction.readthedocs.io/en/latest/index.html. Last accessed April 1, 2024.
[6] kloppy 3.14.0. 2020. *PySport*. Last accessed April 1, 2024
[7] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *University of Washington*
[8] Scott Lundberg. 2018. An introduction to explainable AI with Shapley values. Last accessed April 1, 2024.
https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html
[9] Emily K Marsh. 2023. Calculating XGBoost Feature Importance.
https://medium.com/@emilykmarsh/xgboost-feature-importance-233ee27c33a4#:~:text=The%20SHAP%20method%20is%20more,category%20of%20the%20dependent%20variable.
[10] Scott Lundberg. 2018. Beeswarm plot. Last accessed April 1, 2024.
https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html.
[11] StatsBomb Open Data. 2023. Last accessed April 1, 2024.
https://github.com/statsbomb/open-data

Appendix

	Choice	P_{scores} Brier	P_{scores} AUC	$P_{concedes}$ Brier	$P_{concedes}$ AUC
Algorithm	CatBoost	**0.01376**	**0.7693**	**0.00547**	**0.7313**
	Logistic Regression	0.01601	0.7231	0.00562	0.6578
	Random Forest	0.01409	0.7050	0.00552	0.6457
	XGBoost	0.01390	0.7556	0.00550	0.7255

Figure 15: "Actions Speak Louder than Goals" VAEP Evaluation Results

Transformer	Feature	Description
`actiontype()`	actiontype(_onehot)_ai	The (one-hot encoding) of the action's type.
`result()`	result(_onehot)_ai	The (one-hot encoding) of the action's result.
`bodypart()`	actiontype(_onehot)_ai	The (one-hot encoding) of the bodypart used to perform the action.
`time()`	time_ai	Time in the match the action takes place, recorded to the second.
`startlocation()`	start_x_ai	The x pitch coordinate of the action's start location.
	start_y_ai	The y pitch coordinate of the action's start location.
`endlocation()`	end_x_ai	The x pitch coordinate of the action's end location.
	end_y_ai	The y pitch coordinate of the action's end location.
`startpolar()`	start_dist_to_goal_ai	The distance to the center of the goal from the action's start location.
	start_angle_to_goal_ai	The angle between the action's start location and center of the goal.
`endpolar()`	end_dist_to_goal_ai	The distance to the center of the goal from the action's end location.
	end_angle_to_goal_ai	The angle between the action's end location and center of the goal.
`movement()`	dx_ai	The distance covered by the action along the x-axis.
	dy_ai	The distance covered by the action along the y-axis.
	movement_ai	The total distance covered by the action.
`team()`	team_ai	Boolean indicating whether the team that had possesion in action a_{i-2} still has possession in the current action.
`time_delta()`	time_delta_i	Seconds elapsed between a_{i-2} and the current action.
`space_delta()`	dx_a0i	The distance covered by action a_{i-2} to a_i along the x-axis.
	dy_a0i	The distance covered by action a_{i-2} to a_i along the y-axis.
	mov_a0i	The total distance covered by action a_{i-2} to a_i.
`goalscore()`	goalscore_team	The number of goals scored by the team executing the action.
	goalscore_opponent	The number of goals scored by the other team.
	goalscore_diff	The goal difference between both teams.

Figure 16: VAEP Feature Descriptions (from the SoccerAction documentation website for a *j*=3 model)

j	k	Model Type	V2 AUROC	V3 AUROC	V2 Brier Score	V3 Brier Score
3	3	concedes	0.9302	0.9520	0.0010	0.0013
3	3	scores	0.9571	0.9688	0.0031	0.0021
3	6	concedes	0.8373	0.8866	0.0025	0.0025
3	6	scores	0.8536	0.8569	0.0077	0.0060
3	10	concedes	0.7418	0.7938	0.0058	0.0049
3	10	scores	0.7854	0.7774	0.0127	0.0105
3	13	concedes	0.7122	0.7453	0.0081	0.0069
3	13	scores	0.7526	0.7442	0.0165	0.0135
6	3	concedes	0.9210	0.9434	0.0010	0.0013
6	3	scores	0.9560	0.9681	0.0031	0.0021
6	6	concedes	0.8319	0.8819	0.0025	0.0025
6	6	scores	0.8519	0.8572	0.0077	0.0060
6	10	concedes	0.7399	0.7935	0.0058	0.0049
6	10	scores	0.7856	0.7797	0.0127	0.0104
6	13	concedes	0.6987	0.7456	0.0081	0.0069
6	13	scores	0.7495	0.7505	0.0165	0.0135
9	3	concedes	0.9341	0.9361	0.0010	0.0013
9	3	scores	0.9582	0.9670	0.0031	0.0021
9	6	concedes	0.8323	0.8826	0.0025	0.0025
9	6	scores	0.8485	0.8484	0.0077	0.0060
9	10	concedes	0.7382	0.7939	0.0058	0.0049
9	10	scores	0.7832	0.7764	0.0127	0.0104
9	13	concedes	0.6943	0.7375	0.0081	0.0069
9	13	scores	0.7496	0.7466	0.0165	0.0135

Figure 17: All Model Evaluation Scores

Table 3: Different design choices evaluated on both scoring and conceding probabilities using the Brier score and ROC AUC. For the Brier score lower values are better, whereas for ROC AUC higher values are better.

		P_{scores}		$P_{concedes}$	
Design	**Choice**	**Brier**	**AUC**	**Brier**	**AUC**
Feature set	All features	**0.01376**	**0.7693**	**0.00547**	**0.7313**
	No features	0.01632	-	0.00564	-
	Location	0.01562	0.7330	0.00560	0.6770
	Action type	0.01590	0.6405	0.00562	0.6348
	Loc + Action type	0.01549	0.7417	0.00550	0.6912
Algorithm	CatBoost	**0.01376**	**0.7693**	**0.00547**	**0.7313**
	Logistic Regression	0.01601	0.7231	0.00562	0.6578
	Random Forest	0.01409	0.7050	0.00552	0.6457
	XGBoost	0.01390	0.7556	0.00550	0.7255

Figure 18: Decroos Study VAEP Model Evaluation