Erasmus University Rotterdam

Master Thesis BA&QM: SciSports

Master Econometrics and Management Science

Erasmus School of Economics

Valuing Defensive Performances of Football Players

Author: J.P.T. Piersma 400162
Supervisor: dr. A.J. Koning
Second Assessor: dr. S. Sharif Azadeh
SciSports Supervisor: dr. ir. J. van Haaren
SciSports Supervisor: drs. L. Bransen

Abstract

Football clubs are increasingly using large volumes of data that are collected during football matches to discover suitable reinforcements in the transfer market. Although several data-driven performance metrics have been proposed in the past few years, most existing metrics focus on the offensive performances of football players alone and disregard their defensive performances. To bridge this gap, we introduce a reliable metric to quantify football players' defensive abilities. In particular, our metric measures each player's ability to intercept the passes performed by the opposing team. We investigate the difficulty to perform of an interception in combination with the support of all teammates and, most exciting, we consider the missed opportunities to perform an interception as well. By involving statistical machine learning techniques, we gain insight into the defensive performances of players. According to pass, player and match specific factor characteristics, we clarify these performances. Moreover, we introduce a model to predict the probability of an interception and subsequently predict the probability of each player to perform an interception in two independent stages and one correlated step. To optimize our models, we evaluate the quality of each model's predictions according to LogLoss and AUC-PR ranks the performances of the models. Our models outperform baseline models based on averages and domain knowledge. In line with our expectations, we reveal established names such as Van Dijk and Kanté to provide outstanding defensive performances according to evaluations out of season 2018/2019. Moreover, we provide insight into young talented players who at the moment, lack the experience of performing interceptions, but are mostly well positioned.

May 1, 2020

Acknowledgements

First, I would like to thank my university supervisor dr. (Alex) A.J. Koning for supervising my master thesis. Next, I want to extend my extraordinary thanks to Jan Van Haaren and Lotte Bransen, both working at SciSports, for their support with detailed feedback and advice, whom I learned a lot from. I also would like to thank my mother, Margriet Piersma, and my sparring partner Peter Boerema, for their guidance, trust, patience and financial support during my study. Finally, I want to dedicate this thesis to my father, Henk Piersma, who always remains the most critical motivator in my life.

Contents

1 Introduction **1**
1.1 Research Context . . . 1
1.2 Motivation and Relevance . . . 2
1.3 Research Questions . . . 3
1.4 Contributions . . . 3
1.5 Thesis Structure . . . 3

2 Data **4**
2.1 Data set . . . 4
2.2 Passing Outcomes . . . 6
2.3 Limitations and Noise in Data . . . 8

3 Methodology Background **9**
3.1 Machine Learning Algorithms . . . 9
3.1.1 Tree-Based Methods . . . 9
3.1.2 Generalized Additive Model . . . 11
3.2 Evaluation Metrics . . . 12
3.3 Football Metrics . . . 15
3.3.1 Defensive metrics . . . 15
3.3.2 SciSkill Index . . . 16

4 Football Analytics **18**
4.1 General Approach . . . 18
4.1.1 Feature Engineering . . . 18
4.1.2 Data Pre-Processing . . . 21
4.1.3 Model Optimization . . . 21
4.2 Expected Interceptions . . . 22
4.2.1 Learning Task . . . 22
4.2.2 Baseline Models . . . 23
4.2.3 Performance Analysis . . . 23
4.2.4 Model Inspection . . . 25
4.2.5 Discussion . . . 27
4.3 Defensive Positional Expectations . . . 27
4.3.1 Learning Task . . . 28
4.3.2 Baseline Models . . . 28
4.3.3 Performance Analysis . . . 29
4.3.4 Model Inspection . . . 32
4.3.5 Discussion . . . 34
4.4 Expected Positional Interceptions . . . 35
4.4.1 Learning Task . . . 35
4.4.2 Baseline Models . . . 35
4.4.3 Performance Analysis . . . 35
4.4.4 Model Inspection . . . 38
4.4.5 Discussion . . . 40
4.5 Defensive Ability Valuation . . . 40
4.5.1 Two-Step Approach Valuation . . . 41

4.5.2 One-Step Approach Valuation . . . 42
4.5.3 Normalize Valuation . . . 43
4.6 Discussion Football Analytics . . . 43

5 Results and Evaluation **44**
5.1 Structure of Results . . . 44
5.2 Results of Defensive Performances . . . 45
5.2.1 Results of Defenders . . . 47
5.2.2 Results of Midfielders . . . 49
5.2.3 Results of Goalkeepers and Attackers . . . 50
5.2.4 Results of Talents . . . 51
5.2.5 Results of Defensive Actions . . . 53
5.2.6 Missing Players . . . 54
5.3 Discussion Results . . . 55

6 Conclusion and Discussion **56**
6.1 Research Questions . . . 56
6.2 Further Research . . . 57

A Binary Classification Validation Set **59**

B Expected Defensive Actions **60**

C Positioning **61**

D Multiclass Classification Validation Set **63**

E Multitask Classification Validation Set **64**

F Expected Positional Defensive Actions **65**

G Results Interceptions **67**

H Results Defensive Actions **68**

1 Introduction

This section introduces the primary assignment of this research. To start, we introduce relevant information related to current defensive football metrics. Subsequently, we clarify our motivation to provide a new metric. Next, we present our research questions to this study. Furthermore, we provide our contributions to a new advanced metric, and lastly, we provide the structure of this thesis.

1.1 Research Context

One of the most popular sports all over the world is football which records 3.3 - 3.5 billion fans and is played by 250 million players in 200 countries (Das, 2019). Moreover, transfers of players in big leagues confirm staggering amounts of money where the average salary of athletes in the major European leagues' top teams is in millions of euros (Sportingintelligence, 2019). Deciding which players to trade and develop often depend on "gut" feeling or cling to old habits which suffers from confirmation bias. Perhaps, we should think differently.

In the late nineties, Billy Beane manifests with his companion Bill James into baseball. They applied an evidence-based decision-making process to reveal players who made it more likely to win games (Biermann, 2019). Inspired by their success, the philosophy to manage a sports club based on statistical analysis spread to football, which was significantly different because of the dynamic aspect of the game. Under the guidance of Matthew Benham, a former hedge fund manager and sports bettor, Denmark's FC Midtjylland signed players on the advice of statistical explanations which identify undervalued players. Evaluations of the matches were based on mathematical perspectives, and free kicks became a science project (de Hoog, 2015).

In recent years, the usage of data in football has increased enormously and improved even more. Nowadays, football has been considered as a research area in mathematics and computer science (Rahnamai Barghi, 2015). Some major clubs, such as BV Borussia 09 Dortmund and Arsenal FC, adjust their scouting system to the use of data, whereas Manchester City FC organises events to generate insights into the data (Biermann, 2019).

Obviously, data analyses require data input to provide insights. Two types of data are available regarding football analytics. Player tracking data records every movement in the match as snapshots of the whole pitch, which allows us to obtain all the players. The major downside is the sparseness of this type of data. It is only available in the biggest competitions. On the other hand, we could apply data which only records the events on the ball in the match, which is called on-the-ball event data. This data type is widely available. For instance, we can evaluate players in the second league of Poland; therefore, it is useful for recruitment. We prefer the latter type of data and continue to focus on now.

Most of the researches analyse the attacking intentions of players. This makes sense since football fans and clubs are mainly interested in goals and the decisive players. At the end of the match, it is all about the result whereas one of the fascinating statements of Johan Cruijff explains: 'you have to score more goals than your opponent'. Since we can easily connect the data to evaluate the contribution of passes and goals, the available data is convenient to implement for attacking play (Lucey et al. (2014), Cronin (2017), Kwiatkowski (2017), Decroos et al. (2019), Link et al. (2016), Biermann (2019)).

Let us put the statement that you have to score more goals than your opponent slightly differently, which we now interpret as conceding fewer goals than your opponent. From this point of view, we focus on defending. Defending is more than covering your direct opponent; it is about controlling all the spatial areas on the pitch and disruption of the opponent's possession (Payne, 2015). According to this way of thinking, we assign an objective which challenges a significant limitation of most existing football metrics; they largely ignore actions other than goals and shots (Routley, 2015).

However, to gain a first impression, many statistical approaches in football start by inspecting the number of occurrences of a specific event such as passes or goals, which we define as count-based statistics. But in so doing regarding interceptions, we do not incorporate the difficulty of the performed interception. Moreover, players of teams with relatively little possession control are more likely to score well, which makes it an unfair comparison (Thompson, 2020). Partly because of this, SciSports provides a metric which computes the impact of each defensive action performed. Before and after a specific defensive action, it evaluates the differences in probabilities of scoring and conceding a goal. Within this approach, a particular action only relates to the corresponding match, which we can not extrapolate in general (SciSports, 2020).

Analyzing defence in football is quite a challenge. A critical part of defence relies on the amount of pressure on the opponent. By inspecting the ratio between the opponent's number of passes and the number of defensive actions, we get insight into the amount of pressure performed by a team (Trainor, 2014). In this sense, a lower value corresponds to more pressure. This train of thought enables us to investigate the effectiveness of performed defensive pressure by considering the trade-off between recovering the ball versus the cost of leaving the team's defensive structure (Robberechts, 2019).

Next, the research of Bojinov and Bornn (2016) creates an overview of a specific team's abilities according to controlling phases when in possession and disrupting stages when not in possession. Their model depends on fitting the correlation between occurred events during matches which incorporates the behaviour of players in specific situations. Since we can not interpret the causes of these phases, we run into a significant shortcoming of this approach.

Since most of the interceptions in football are recorded after unsuccessful passes, and passes are the most frequently occurring actions on the pitch, we focus on interceptions corresponding to passes (Power et al., 2017). Passes are conducted under different conditions, which ensures one interception to add more value compared to another. Therefore, we evaluate the relevant information of each pass and describe it as explicitly as possible. For instance, a centre back gives a pass forward, which is intercepted by the opponent. We need to distinguish this event of a bad pass, a good pass missed by a teammate, or perhaps an excellent defending action of the opponent to prevent an opening in yards of space. Since we do not know the intention of the pass, especially the intended recipient of the pass, it is hard to determine which alternative occurred. Possibly because of this inconvenient decision, there is not yet much research done in discovering the defensive abilities of players in football.

1.2 Motivation and Relevance

Since there is currently no generalizable accurate metric which gives insight into the defensive performances of football players, the main objective of this research is to accurately quantify defensive abilities of players to intercept the opponent's passes. These findings could be used as support for scouting players according to their defensive abilities and tactical decisions in football.

Current metrics suffer from incorporating relevant information which provides insight into the difficulty of an interception. For instance, we should appreciate a relatively difficult interception to perform more than a relatively easy interception. To gain these insights, we investigate relevant information of the intercepted pass.

From our point of view, interceptions are mostly the merits of collective resistance, which insinuates not to reward an individual player for an interception. An idealized model should rate the influences of teammates as well.

Since Bojinov and Bornn (2016) lacks to reveal insight into significant factors related to interceptions, we desire to design an approach which enables us to obtain the sensitivity of the corresponding features. These interpretations clarify the defensive performance of players.

Finally, to process players' failures as well, we prefer to get insight into the missed opportunities of players to execute an interception. Since this component depends on the defensive expectations of players, it is the most challenging tool of our research. In this sense, we prevent to rely our analyses on performed interceptions only, which could provide a distorted picture of reality.

1.3 Research Questions

In this research, we aim to design an accurate metric to quantify players' defensive abilities to intercept the opponent's passes. By including relevant context, we improve the interpretation and reliability of our findings. To achieve this goal, we address the following main research question: "How to evaluate the defensive performances of football players?".

Since we intend to discover the difficulty of an interception under various conditions, we need to design a setup which involves this objective. Moreover, we state interceptions of the opponent's passes are the merits of multiple players. For this reason, we take into account the contributions of all players of a squad. Therefore, we address the first research sub-question: "How to estimate the likeliness of a pass being intercepted by a given player while accounting for the defensive support of the player's teammates?".

To generalize our models and allow them for all relevant information, we describe the context of each pass as explicit as possible. According to this context, we enable us to reveal the difficulty to perform an interception. This drives us to the next research sub-question: "What factors contribute to the likeliness of a pass being intercepted?".

Since we prefer to investigate the predictions and explanations of interceptions, we apply several statistical techniques. We test these methods compared to baseline models. In addition, we provide the last research sub-question: "What statistical method is best suited to predict the likeliness of a pass being intercepted?".

1.4 Contributions

We provide a reliable metric for defenders and midfielders to analyze their defensive abilities. Our models predict the difficulty of an interception according to specific conditions for each player of the defending team. Within this approach, we incorporate the contributions of all players.

Most exciting, we involve the missed opportunities of players to execute an interception in players' defensive performance evaluation as well.

By inspecting practical cases, we provide insights into the evaluations of our models and performance of players where we clarify the substantial factors involved.

Finally, we provide insight into players who manage to attend many duels but do not achieve to perform an interception. For instance, these players lack experience, which makes it an exciting insight into young players' defensive performance.

1.5 Thesis Structure

In Chapter 2, we describe the data used for this research. In Chapter 3, we investigate the methods and techniques applied to this research. In Chapter 4, we explain the implementation of the methods and techniques into our scope of practice. In Chapter 5, we clarify the exciting results of our findings. In Chapter 6, we discuss the conclusions of the study and future work.

2 Data

In this section, we explain the data set used in this research. First, we provide the information available in this set. Next, we clarify all possibilities of the passing outcomes. Finally, we pay attention to the limitations and noise of the data set.

2.1 Data set

During this research, we use a data set which contains information regarding specific seasons with corresponding teams and matches, the positions of the players in each match, and the on-the-ball events performed in each match. SciSports collected this data set which is initially provided by Wyscout. It contains information of 7,300 matches in the seasons 2016/2017, 2017/2018 and 2018/2019 in seven European competitions; the English Premier League, the Spanish LaLiga, the Italian Serie A, the French Ligue 1, the German Bundesliga, the Dutch Eredivisie, and the Belgian Jupiler Pro League.

We utilize match sheet data which states the home playing side, intermediate and final scores for each specific match. In this data, at least once per half, for both teams, the line-up and formation (e.g. 4-3-3) are available. The data presents the relative positions of all the players given a specific formation. Besides, it shows substitutes who play in a different formation caused by a tactical change as well.

The critical data contains event data which includes the on-the-ball events on the pitch (e.g. passes, shots, duels). Every match approximately 1,500 events are manually recorded. In total, over 10 million events are present. Each event includes 73 variables which represent details of a specific event. In Figure 1, we show a snapshot of the match event data from TSV Bayer 04 Leverkusen - 1. FSV Mainz 05 on the second recorded match of the season (Bundesliga 2017/2018) including the most useful variables.

	competition_id	season_id	match_id	event_id	period_id	period_milliseconds	team_id	player_id	type_id	subtype_id	location_start_x	location_start_y	location_end_x	location_end_y	accurate
1826	426	181137	2516916	224290604	1	3588	2446	86489	8	85.0	50	49	47.0	52.0	True
1827	426	181137	2516916	224290605	1	5226	2446	15188	8	85.0	47	52	29.0	68.0	True
1828	426	181137	2516916	224290606	1	6988	2446	228768	8	85.0	29	68	31.0	93.0	True
1829	426	181137	2516916	224290607	1	8540	2446	14781	8	85.0	31	93	43.0	95.0	False
1830	426	181137	2516916	224290689	1	8547	2460	15399	7	72.0	57	5	53.0	0.0	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3522	426	181137	2516916	224292172	2	2835547	2446	15231	1	12.0	15	7	9.0	9.0	True
3523	426	181137	2516916	224292303	2	2837755	2460	21469	1	13.0	91	91	97.0	88.0	True
3524	426	181137	2516916	224292173	2	2837917	2446	40657	1	13.0	9	9	3.0	12.0	False
3525	426	181137	2516916	224292304	2	2839652	2460	21469	8	80.0	97	88	100.0	100.0	False
3526	426	181137	2516916	224292175	2	2841619	2446	15231	7	72.0	2	14	0.0	0.0	False

1701 rows × 15 columns

Figure 1: Most relevant variables of the match event data subset. Note that *period_id* indicates the first or second half, *period_milliseconds* records the time of an event in milliseconds after the start of the half, and *accurate* presents the accuracy of the event (if a shot event is True, we obtain a goal).

The first point of interest concerns *type_id*, which provides the type of action executed in the event. We divide all the actions into subactions (*subtype_id*) which enables to clarify the type of action in detail. For instance, a pass divides into a cross, hand pass, head pass, high pass, launch, simple pass, and smart pass. There are ten different types of actions possible. In Table 1, we obtain all type of actions.

Action Type	**Description**
Duel (1)	Playing 1 vs 1 against an opposite player; ground or air.
Foul (2)	Foul on the pitch.
Free Kick (3)	Start of play after interruption of the game.
Goalkeeper leaving line (4)	Goalkeeper leaves its area to take or punch the ball.
Interruption (5)	Interruption of the game; ball out of the field or sign of referee.
Offside (6)	Player receives the ball in illegal possession.
Others on the ball (7)	Other actions on the ball.
Pass (8)	Pass from player A to player B.
Save attempt (9)	Goalkeeper performs an attempt to stop a shot.
Shot (10)	Shot in the direction of the target.

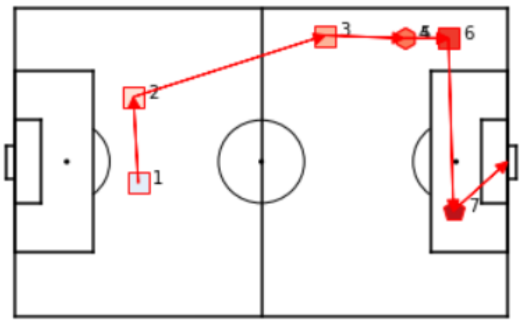
Table 1: Most critical type of actions (*type_id*). Free Kick could be a regular free-kick, penalty, goal kick or throw-in.

A type of action to clarify in more depth is *others on the ball* (7) which covers three possibilities. A player is using and changing his speed in a run with the ball for at least 10 meters which we note as an acceleration, and we could interpret as a dribble. When the player is just touching the ball but does not perform a pass or some other event, we note a touch. We note a clearance if the player decides to clear the ball forward or out of play, playing safe, which is used most of the times when he is under pressure.

Next, we obtain the location of the events. The data maps an x,y-graph of the pitch to provide a clear description of the event's location. These values are normalized from 0 to 100 for both axes. The opponent's goal is at the right side of the pitch (x = 100) for both teams; thus the play mirrors after a turnover. In some cases two sequential actions, for instance, two passes, do have space between the end and start location. The difference between these locations is most likely a short dribble which we define as a carry and is not indicated as an action. To provide a realistic view of the Euclidean length of the events, we normalize the size of the pitch. We use the widely applied standard size for the pitch which format is 68 meters wide and 105 meters long.

Finally, we link specific player data to the events executed on the pitch. Information about the player's player ID and name is most useful since it enables us to realize the connection. Moreover, we obtain if a player gets a yellow or red card, and even if a player receives a second yellow card. Other available information contains the player's birthday, nationality, gender, height, weight, preferred foot and line code (e.g. defender, midfielder).

Due to standardization of the pitch and all available information in our data set, it's possible to visualize events on the pitch. In Figure 2 we visualize, with the help of matplotsoccer library (Decroos, 2019), a counter-attack which results in the 3-0 from the match BV Borussia 09 Dortmund - FC Schalke 04 (Bundesliga 2017/2018).



	Time	Action Type	Player	Team
1	19:33	Pass	Kehrer	Schalke 04
2	19:36	Pass	Yarmolenko	Borussia Dortmund
3	19:42	Acceleration	Aubameyang	Borussia Dortmund
4	19:44	Duel	Stambouli	Schalke 04
5	19:44	Duel	Aubameyang	Borussia Dortmund
6	19:46	Pass	Aubameyang	Borussia Dortmund
7	19:47	Shot	Götze	Borussia Dortmund

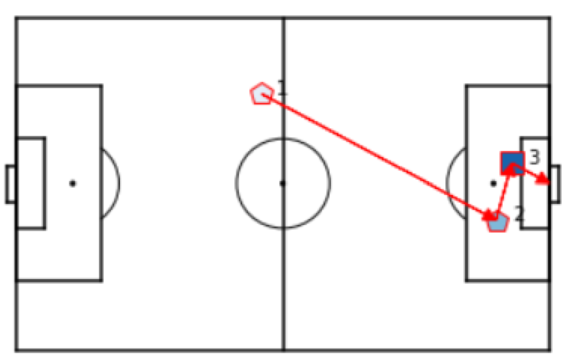
Figure 2: Counter visualization. Note that according to the variable *goal* of the shot event we notice a goal. Event 4 and 5 overlap caused by a duel of two players. Time measured in minutes and seconds.

(1) A bad pass of Thilo Kehrer results in a turnover. (2) Andriy Yarmolenko picks up the ball and passes to Pierre-Emerick Aubameyang. (3) Pierre-Emerick Aubameyang dribbles to his opponent. (4/5) Pierre-Emerick Aubameyang duels with Benjamin Stambouli. (6) Pierre-Emerick Aubameyang crosses the ball in front of the target. (7) Mario Götze heads the ball into the back of the net.

2.2 Passing Outcomes

Since this thesis mainly focuses on quantifying defending abilities corresponding to the opponent's passes, it is interesting to provide insight into the possible outcomes and statistics of the passes. Firstly, let us discuss the various options as a result of a specific pass and explain the situations that could occur. We illustrate every situation from the match Liverpool FC - Arsenal FC (Premier League 2018/2019), where Liverpool FC plays from left to right, and Arsenal FC vice versa.

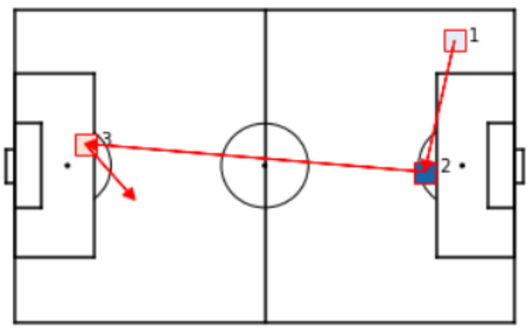
(1) Pass reaches teammate. A pass enables a teammate to generate an action which we interpret as a successful pass. If the following action is a shot which turns into a goal, we call the concerning pass an assist. In Figure 3, we obtain an attack from Liverpool, where Robertson passes the ball to Salah, who passes to Mané who performs a shot.



	Time	Action Type	Player	Team
1	31:25	Pass	Robertson	Liverpool
2	31:29	Pass	Salah	Liverpool
3	31:31	Shot	Mané	Liverpool

Figure 3: Pass reaches teammate. We note two passes which both reaches a teammate, followed by a shot.

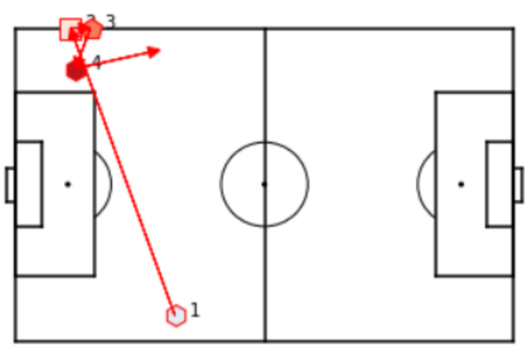
(2) Pass intercepted by opponent. A pass ends at the opponent, which generates an action continuously. This type of unsuccessful passes is the most valuable for this research because we need these interceptions to credit a player's defending abilities. The data allows obtaining a failed feint of a teammate, and the occurrence of an impossibility of a teammate to control the ball. Both cases result in an interception. In Figure 4, Papastathopoulos passes the ball to his goalkeeper Leno, who launches the ball forward. Liverpool's goalkeeper Alisson intercepts the pass and passes the ball to the right side of the pitch.



	Time	Action Type	Player	Team
1	30:18	Pass	Papastathopoulos	Arsenal
2	30:23	Pass	Leno	Arsenal
3	30:30	Pass	Alisson	Liverpool

Figure 4: Pass intercepted by opponent. We note a pass which reaches a teammate from Papastathopoulos to Leno, followed by an intercepted pass.

(3) Pass out of play. This situation concerns the alternative possibility of an unsuccessful pass. It could be a simple bad pass, or it could be necessary because of the opponent's pressure which we can not obtain from the data. The passes could end over the sideline or goal line, and result in a throw-in or goal kick respectively. In Figure 5, we obtain a cross of Maitland-Niles missed by his teammates. The ball goes out of play. Alexander-Arnold performs a throw-in towards Lovren, who passes the ball forward along the left side of the pitch.



	Time	Action Type	Player	Team
1	60:39	Pass	Maitland-Niles	Arsenal
2	60:48	Interruption	-	-
3	61:39	Free Kick	Alexander-Arnold	Liverpool
4	61:40	Pass	Lovren	Liverpool

Figure 5: Pass out of play. We note a pass sent from Maitland-Niles which ends out of play. Subsequently, Alexander-Arnold throws the ball to Lovren who sent a pass forward.

(4) Pass followed by referee's interruption. These situations are not valuable for this research since we have too few information to judge these results. Firstly, we can obtain a pass into the offside. For example, the concerning pass launched too late, the player in offside could make a bad run, or the defence made a right decision to move up the field and create the opponent's illegal possession. Next, the opponent could make a foul followed by a pass which could be a wrong defending action or an intelligent action to prevent a promising attack. Lastly, the referee could interrupt the game on the ground of a dubious situation on the pitch. For instance, we could think of a head injury. In Figure 6, Torreira passes to Aubameyang, which receives the ball in offside possession. The referee interrupts the game to provide a free kick in Liverpool's advance. Van Dijk performs the free-kick and passes the ball to Fabinho who passes the ball to the right side of the pitch.

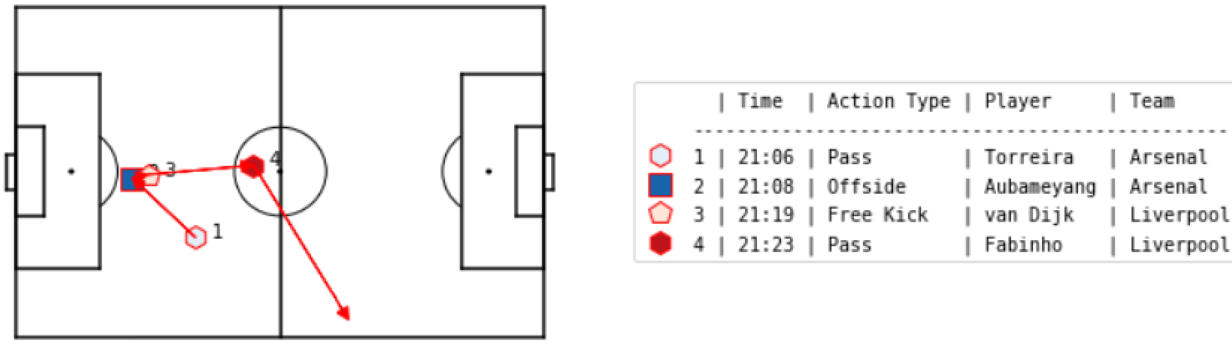


Figure 6: Pass followed by referee's interruption. We note a pass from Torreira sent into offside. Subsequently, Van Dijk passes a free-kick to Fabinho who sends a pass sideways.

To gain a better understanding of the passing outcomes in our data set, we provide the descriptive statistics in Table 2. We obtain that most of the on-the-ball events performed on the pitch are passes. Nearly one out of six passes are intercepted by the opponent, which results in a reasonably imbalanced set. Furthermore, we obtain that passes out of play are incredibly rare, and the number of passes followed by referee's interruption is negligible.

Type of event	**Number of occurences**	**Ratio**
Number of on-the-ball events	12,027,451	N/A
Number of passes	6,121,991	N/A
Number of passes reaches teammate	5,041,217	82.3%
Number of passes intercepted by opponent	987,377	16.1%
Number of passes out of play	98,788	1.6%
Number of passes followed by referee's interruption	3,609	0.0%

Table 2: Descriptive statistics of the passing outcomes. Ratio contains the type of event's rate to number of passes.

2.3 Limitations and Noise in Data

As discussed, we use match event data in this research, which provides only the on-the-ball events executed by players on the pitch. Consequently, we do not know where the other players are at the time of an event which makes it hard and sometimes not possible to be aware of the intention of the player. This could be the case in many situations such as unsuccessful passes where a player receives the ball in an offside position. Then we do not know the exact purposed path of action; we are only aware of the start and end location. Besides, as mentioned, we consider the Euclidean distance between two subsequent actions as a carry.

In some cases, the match sheet data subset does not provide the line up for both teams from the start of the match. To enable linkages between events and players, we assume the first seen line up in the match to be the starting line up.

Furthermore, manually entering the data could result in mistakes in location and time. Due to our relatively large data set (see Table 2), we assume these errors are minimized and negligible.

3 Methodology Background

In this section, we clarify the methods and techniques we apply during this research. To start, we give insight into the machine learning algorithms performed to most efficiently quantify players' defending abilities. Next, we on elaborate metrics to evaluate our designed models. Finally, we represent already existing metrics to evaluate players' defending abilities.

3.1 Machine Learning Algorithms

During this research, we explore methods and techniques to quantify players' defending abilities. In particular, we explore methods to quantify players' ability to disrupt the opponent's passes. To achieve this objective, we could choose parametric and non-parametric methods. Since we are quite uncertain about the distribution of the passes, we create a potential disadvantage with a distribution assumption required in the parametric approach. If the chosen statistical distribution is too far from the actual model, then the predictions will be weak. In contrast, we apply various machine learning techniques to provide predictive models. We base our explanation of methods mainly on James et al. (2013). Furthermore, we gather information from referenced papers.

3.1.1 Tree-Based Methods

To attain reliable evaluations, we use tree-based methods which are applicable in many circumstances and widely-used. These methods allow for relatively strong interpretability of complex relations between features which entails opportunities to improve the performance of the learning algorithms. Besides, in Kaggle competitions tree-based models are popular and do often achieve highly efficient scores. For instance, the winner of the Allstate Claims Severity Kaggle Competition, Alexey Noskov, attributes his success in the competition to variants of tree-based methods (Kaggle, 2017).

As we present in Figure 7, all the events of the data set to start at the top of the tree (root node), then splits into two branches further down the tree which we define as recursive binary splitting. Finally, each tree comes down to the leaf nodes. Within every tree, the number of leaf nodes (internal nodes) can differ, where more leaf nodes bring more complexity to the tree.

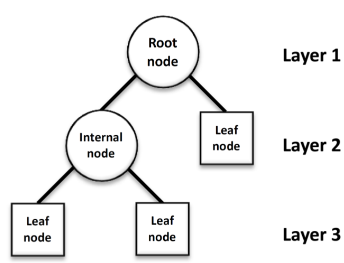


Figure 7: Decision Tree

Each new event starts at the root node of the decision tree. With the aid of binary splits, the event comes down to a leaf node which corresponds to a prediction. We apply decision trees to predict a particular class for each event which we define as classification problems.

To maximize the fraction of training events in that region which belongs to the most common prediction, we investigate how well each feature predicts an event. The more refined a feature turns out to be, resulting in more reliable forecasts, the higher we place them in the decision tree.

Several techniques are possible which utilize decision trees where we consider a trade-off between prediction accuracy and model interpretability. Because of our relatively imbalanced data set, we should be aware of high accuracy by default. For instance, if we classify none of the passes as intercepted, we still achieve an accuracy of about eighty per cent. We define this phenomenon as the accuracy paradox. For this reason, we start with a Random Forest model which allows for a relatively high level of interpretability compared to other tree-based methods, for example, due to the possibility to obtain the splits in the model.

Furthermore, we inspect to improve the performance of our predictions with Gradient Boosted Decision Trees. A fundamental property of Gradient Boosted Decision Trees is that trees grow sequentially with the notion to enhance the accuracy where it performs weakly. In contrast, trees in Random Forests grow in parallel.

3.1.1.1 Random Forest

Random Forests achieve a proper way to start since we take the majority vote of all predicted classes from multiple decision trees. In this sense, we avoid fitting the model too strictly to specific patterns in the training data defined as overfitting. Instead of high complexity, we prefer more robust models that mostly realize better predictions (Ho, 1995).

By learning decision trees for various parts of the training set, we could end up with entirely different prediction results for each event in each tree. Learning trees by taking repeated samples from the training data (bootstrapping) results in more representative decision trees of the training data. If we apply the test data to this model of trees and average the prediction results, we reduce the prediction variance. Building trees with the aid of bootstrapping, followed by averaging the predictions we define as Bagging. However, an issue occurs when we deal with a powerful feature hidden in the root node of each tree which results in quite similar decision trees.

Random Forests improve the idea of bagging by a small tweak of the process. When building decision trees, each time we consider a split, we choose a random subset of features as split candidates from the full set of features. This rule avoids that all of the trees use the strongest feature in the top split, and so other features have more chance to influence the prediction process. In this sense, we possibly create more variance between the prediction outcomes of the trees. However, the overall model will be less dependent on the mentioned strongest feature and provide more reliable results. Hence, we decorrelate the trees within the Random Forest.

The Random Forest's prediction depends on a majority vote of all trees for each classification problem.

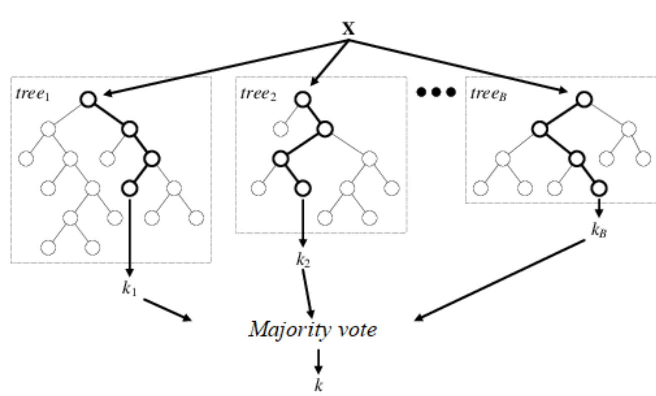


Figure 8: Random Forest

3.1.1.2 Gradient Boosted Decision Trees

We introduce a highly accurate and efficient procedure which optimizes its model performance over sequentially grown trees defined as Gradient Boosted Decision Trees (GBDT). Moreover, GBDTs contain properties that avoid potentially overfitting (Schapire, 1990).

The procedure in the standard GBDT starts by fitting a decision tree to the training data. After that, the trees grow sequentially; each next fitted tree uses information from previously grown trees. In this sense, each tree fits on a modified version of the original data set. Events that provide the most substantial deviations to the fitted model get a significant weight which increases the probability to involve the set of events in the next sequential tree. We track these deviations according to computing the gradients of the data instances where negative gradients can be seen as residuals. By fitting decision trees to considerable residuals, we slowly improve the prediction in areas where it does not perform well. We determine the pace of learning from previous errors by tuning the learning rate. So, with that, we enable to avoid overfitting.

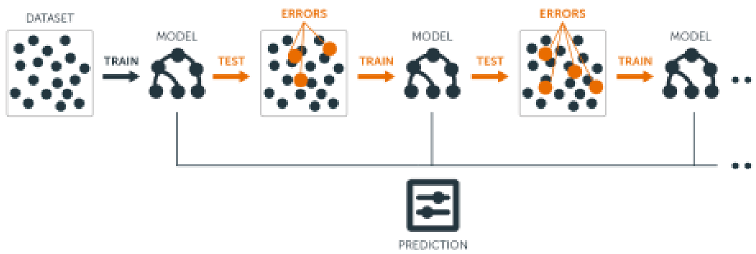


Figure 9: Gradient Boosted Decision Trees

In recent years, various designs to improve the standard GBDT are created. We apply three of them in our research which could improve our final results due to our relatively large data set and use of categorical features: XGBoost, LightGBM and CatBoost.

XGBoost. The most commonly used method which utilizes a quite similar technique applied in Random Forests, which involves feature subsampling to avoid overfitting (Chen and Guestrin, 2016).

LightGBM. A fast and highly scalable technique that only scans the most significant data instances to estimate the information gain, which speeds up the learning algorithm (Ke et al., 2017).

CatBoost. A method that incorporates similar effects for different categories within a feature. CatBoost provides efficient performance when dealing with a feature set which contains many high-ordered categorical features (Prokhorenkova et al., 2018).

3.1.2 Generalized Additive Model

Machine learning techniques often require a trade-off between efficiency and interpretability. Generalized additive models (GAM) provides cases that confirm both concepts (Caruana et al., 2015). GAMs provide an additive framework which extends a standard linear model by allowance for non-linear functions of each of the features (Hastie and Tibshirani, 1990).

GAM computes the expected value of the outcome of multiple feature functions (predictors) in the model. On account of additivity, we enable to interpret each predictor separately while holding other predictors fixed. Since we easily understand the contribution of each predictor to its expected value, we consider such models to be interpretable. Each response classification depends on the sum of predictors incorporated with a corresponding bias.

To improve efficiency, functions of pairwise interacted features adjust to standard GAMs, which we define as GA^{2}Ms. We can visualize the pairwise interactions in a heat map; hence the model remains to be interpretable.

3.2 Evaluation Metrics

By looking for methods to accomplish reliable results for the quantification of players' defending abilities. Since we are interested in predicting if a pass will be intercepted or not, and who performs the interception, we focus on classification trees. Therefore, we need metrics to evaluate the performance of the classifiers. As mentioned, because of our imbalanced passing data set, we have to deal with the accuracy paradox, so we investigate alternative evaluation metrics.

Since we design a system to assign a value to the players based on the probability of an interception and the probability of all players' positions to realize the interception, we are interested in measures which pay attention to the predicted probabilities. Two synthetic classifiers with predetermined performance difference designed in Witten and Frank (2002) are tested to the influence of probability noise. According to sensitivity analyses of Ferri et al. (2009), various measures calculate scores to determine the best classifier. We identify Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Mean Squared Error (Brier Score) and logarithmic loss function (LogLoss) as the most reliable metrics in case of our objective. However, since we deal with an imbalanced data set Area Under the Precision-Recall Curve (AUC-PR) gives the most accurate view of the algorithm performance (Davis and Goadrich, 2006). Since LogLoss and Brier Score provide a nearly similar arrangement of methods, we evaluate our methods according to the default metric in the machine learning community: LogLoss. Besides, we evaluate our models with the help of Calibration Curves and Learning Curves.

AUC-ROC. The metric evaluates the classifications of the classifier's predictions for every possible threshold between zero and one. Hence it measures how well the classifier distinguishes between classes based on the prediction probabilities. As we show in figure 10, the evaluation depends on the false positive rate and the true positive rate (recall) along the x and y axes, respectively. By testing all possible thresholds, we evaluate a specific classifier based on these measure objectives. We present a binary classifier (left) and a multiclass classifier (right). The area under the curve indicates the performance level of the classifier, where a larger area corresponds to better performance. Among an imaginary forty-five degrees line, we obtain a model which corresponds to randomly distributed classifications. Concavity of the graph means the classifier performs better than a random classifier. Ideally, the curve hits the upper-left corner; hence, a perfect classifier.

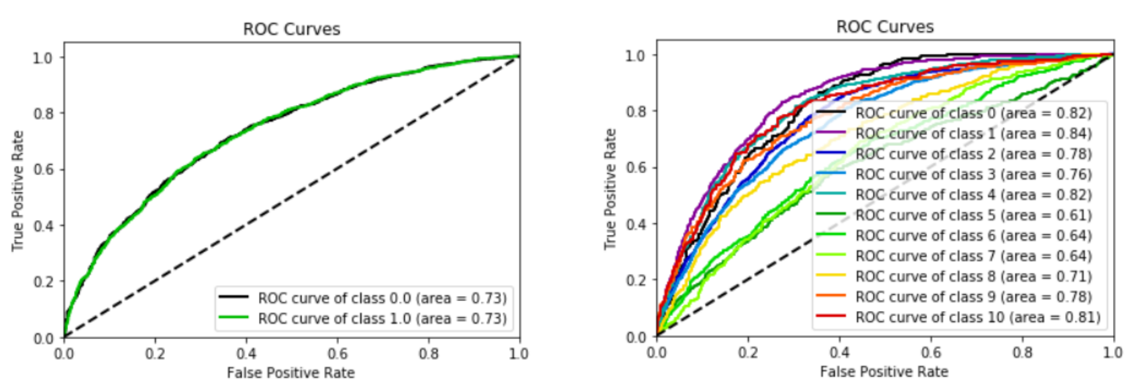


Figure 10: AUC-ROC. Note that every curve represents a class in the multiclass classification (right subfigure).

AUC-PR. An extension to the well-known F-score, where the ratio between recall and precision (proportion of true positives to all positives) are evaluated for all possible thresholds. In the case of our imbalanced data set, we should include the false positives in the evaluation. According to this line of reasoning, precision provides a more reliable performance indication compared to a false positive rate involved in AUC-ROC. In figure 11, we present an AUC-PR curve of a binary classifier (left) and a multiclass classifier (right). Again a more substantial area under the curve corresponds to better performance. An ideal graph hits the upper-right corner, which is getting only the true positives with no false positives and no false negatives; hence, a perfect classifier.

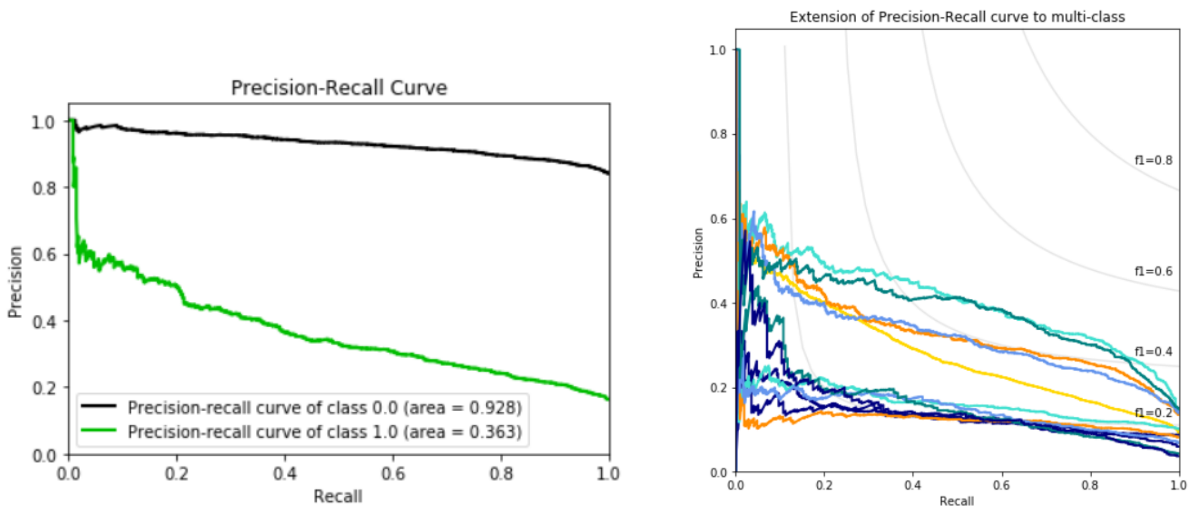


Figure 11: AUC-PR. Note that every curve represents a class in the multiclass classification. Moreover, the iso-f1 curves represent a particular f1-score in general.

LogLoss. A metric that measures the accuracy of a learning algorithm according to the calculated probabilities for every possible class. It works with a natural logarithmic design similar to cross-entropy. In this manner, the metric penalizes prediction failures. Also, an increasing number of failures are relatively harder punished. A lower score corresponds to better performance of the classifier.

Brier Score. Most-commonly used method of measurement in statistics which evaluates the probability forecasts to the actual response of the event, also known as Mean Squared Error. This metric entails an advantage due to its convenient use in classifying multiple outcomes. Again, a lower score corresponds to better performance of the classifier.

Calibration Curve. To measure how well-calibrated the binary classifier fits the data set, we introduce the Calibration Curve. As presented in figure 12, the curve depends on the fraction of positives for a given mean predicted value.

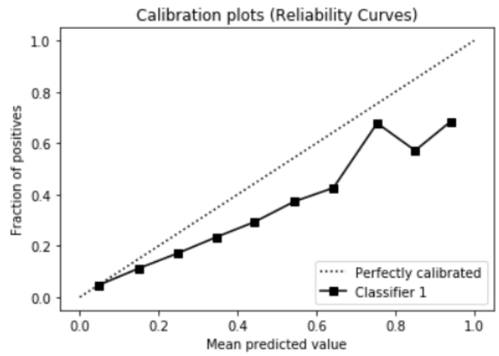


Figure 12: Calibration curve. The dotted forty-five degrees line corresponds to a perfectly calibrated model. The closer the model curve comes to this line, the better it calibrates to the data.

Ideally, we predict the fraction of positives similar to the mean predicted value. Since we have to deal with an imbalanced data set, the objective, predicting one, is hard to realize. For example, in our case, for about eighty per cent of the data consists of zeros (pass reaches teammate) and twenty per cent of ones (interceptions). So suppose the benchmark prediction is 0.2 for convenience. To predict an actual one, the benchmark prediction needs to be forced up by the algorithm with at least 0.3, because it then exceeds the classification threshold of 0.5, which could be difficult if the situation is unclear. In these conditions, the features barely influence the prediction resulting in a predicted zero.

Learning Curve. A metric which evaluates the variance-bias trade-off to provide accurate predictions. Learning Curves show the relationship between training set size and a chosen evaluation metric (e.g. LogLoss, Accuracy, etc.) on the training and validation set. If the curves converge, the model provides similar performance for both sets. As we present in Figure 13, the interaction point at approximately 250,000 events means the model needs 250,000 events for training to provide similar performance. The performance of the model does not improve anymore after this point of convergence. If the curves do not converge, the model probably suffers from a high-variance problem which suggests we fit the model too strict to the training data.

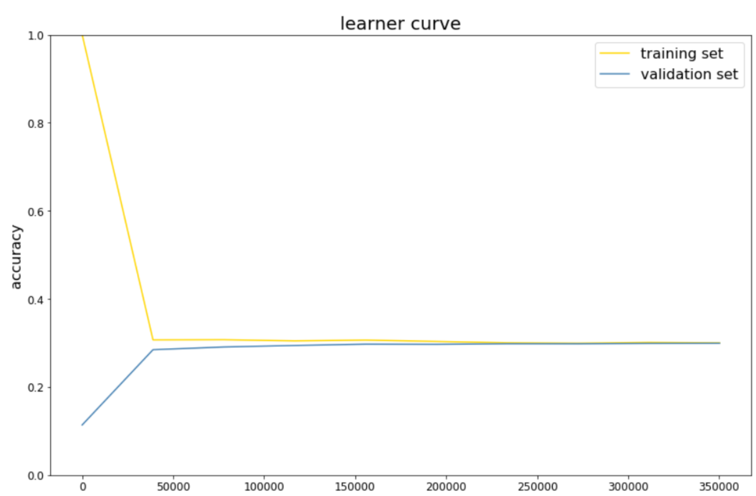


Figure 13: Learning Curve. If the curves converge, the model provides similar performance for both sets.

3.3 Football Metrics

Now we discuss the most relevant existing football metrics to this research. We first introduce three metrics that quantify the defensive abilities of players and then present a player index that estimates the overall level of players.

3.3.1 Defensive metrics

So far, not many football metrics are developed to explore the defensive side of football in contrast to attacking play, such as the well-known expected goals metric. One of the main reasons could be that it is hard to determine the value of an interception. For example, we could deal with a bad pass which is simply intercepted by the opponent, or a well-intended pass intercepted by a smart and intelligent heroic tackle. Either the difficulty of a pass to reach a teammate, as well as the challenge of an interception is hard to establish out of the data.

However, there are a few existing useful metrics which experts apply to measure the defensive abilities of players. To provide insights into these metrics, we discuss the three most commonly used metrics.

3.3.1.1 Interceptions P90

Most of the primary metrics in football provide superficial count-based statistics, such as known for goals and assists (i.e., a pass preceding a shot that results in a goal). Interceptions P90 tracks all the interceptions executed by the players per ninety minutes. We usually report the number of interceptions per 90 minutes of playing time, which is the default duration of a football match, to allow for a fair comparison between players who have spent different amounts of time on the pitch.

Within this focus solely on executed interceptions, some problems occur. For example, when team A is terrible at holding possession and probably weaker than the opponent, a player of team A has by nature more opportunities to make an interception than a player of the opponent, which makes it an unfair comparison.

Quite interconnected to the latter issue, a drawback of this approach is that we similarly treat every interception. Probably, it is more challenging to make an interception when playing against a relatively strong team compared to a relatively weak team. For instance, if Real Madrid FC is playing against FC Groningen, it is probably more comfortable for a defender of Real Madrid FC to make an interception compared to a defender of FC Groningen because of the difference in performance level between those teams.

3.3.1.2 Defensive Contribution Ratings

SciSports provides a metric which computes the impact of each action on the pitch to the outcome of the match defined as contribution ratings (Decroos et al., 2019). Before and after a specific action, the probability to score and concede a goal is calculated by evaluating all outcome results of previously obtained situations. The contribution depends on the difference in probabilities in both moments. The action receives a positive reward if the next game state increases the probability of scoring a goal or decreases the probability of conceding a goal, and a negative reward if the next game state decreases the probability of scoring a goal or increases the probability of conceding a goal.

Based on the same principle, according to SciSports (2020), the defensive contribution of each player is computed. For each defensive action, the impact on the outcome of the match is determined.

Besides, preventing the opponent from scoring is valued as well. In advance, a specific contribution is expected for each player in the match. Given a particular player, if its contribution is less than expected at the end of the match, the direct defensive opponent receives a positive credit.

A drawback of this approach is that we only judge obtained defensive actions executed, and do not have insight into the missed opportunities to make an interception. Besides, the impact of the defensive action relates only to the specific corresponding match, which we can not extrapolate in general. Moreover, it does not incorporate the difference in performance level and defensive tactics of both teams.

3.3.1.3 Passes Allowed Per Defensive Action

Trainor (2014) introduces an approach that evaluates the defensive intensity of each defending team by taking the passes allowed of the opponent per defensive action (PPDA) into account. Hence it indicates the pressing utilized by a team in their matches. To achieve this score of pressure, we provide a ratio between the number of passes by the attacking team and the number of defensive actions executed by the defending team. We define defensive actions as interceptions and duels followed by an opponent's pass. Let us consider an example where team A plays against team B. We compute the PPDA for team A as follows

$$\text{PPDA}_A = \frac{\text{Number of passes}_B}{\text{Number of defending actions}_A}, \tag{1}$$

note that a smaller value means a higher level of defensive intensity.

If we scale the defensive actions of a team over players who execute these defensive actions, we provide a PPDA-score for each player. A significant detail; we need to be aware of the exact time of the players on the pitch to link the number of passes to each player.

In advance, we expect every team to execute defensive actions when the opponent is near their goal. Therefore it makes no sense to track the number of defensive actions in your own box since every team would behave similarly. Along with this reasoning, we should make a feasible region to investigate where we enable us to distinguish between the defensive behaviour of teams. To return to the previous example; to provide the PPDA of team A, we only track team B's number of passes and team A's number of defending actions on the half of team B.

3.3.2 SciSkill Index

Competitions all over the world differ in performance level due to the contributing teams and players. Within a country, there are distinct levels of leagues which represent noticeable levels in descending order. Between different countries, it is harder to distinguish each performance level of the league.

In general, teams and players represent the level of the league; thus, these three aspects are related in their performance levels. SciSports' SciSkill Index assigns an index number to each player that reflects the player's contribution to their team's performance. Since the SciSkill Index algorithm accounts for the strength of the opposition, the resulting index numbers can be compared across leagues (SciSports, 2014).

The SciSkill index depends on five components which expresses a combination of individual player contributions and its team's match results. (1) Offensive skill computes the expected goals a player scores in a match. (2) Defensive skill calculates the expected goals a player concedes in a match. (3) resistance incorporates the performance level of a player. This level is related to its team and the league it participates. According to the player's position, we determine its (4) offensive and (5) defensive share, which together sums up to one.

Before a match, an expectation based on the current performance levels of both teams and players determine a predicted final score. According to the final result follows an update of the performance levels of all contributing players. Undervaluation leads to an upgrade, whereas overvaluation leads to a downgrade. Hence an iterative algorithm updates each player's index after each match according to match sheet data. As a reminder, match sheet data contains, among other things, the line up of both teams, substitutes, exact minutes played of every player and the final score of the match which applies for the algorithm.

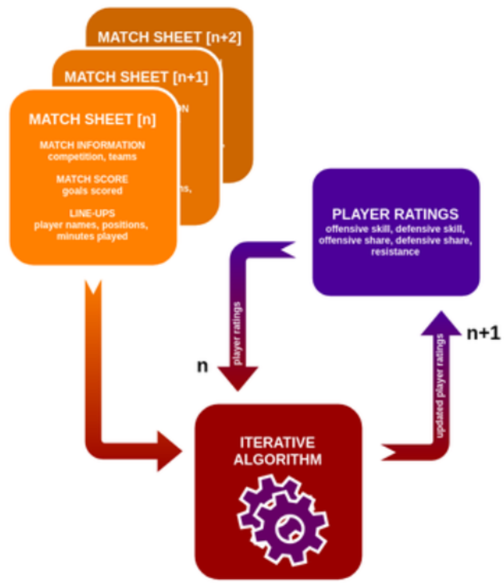


Figure 14: SciSkill index algorithm.

4 Football Analytics

In this section, we implement our methods and techniques to evaluate the defensive performances of players. Firstly, we clarify our general approach. Next, we elaborate and give insight into our classification models. Lastly, we explain the players' performance valuation regarding our approaches.

4.1 General Approach

Widely-used descriptive statistics, such as interceptions P90, only determine how many interceptions players provide. In this sense, more interceptions correspond to a better valuation of the player's defensive abilities. As a result, players of both relatively weak teams and team's with relatively little possession control probably score well because they have a lot of ground to cover (Thompson, 2020). Besides, we know nothing about the difficulty of each specific interception. Note that we could differ the weight of interceptions according to the opponent's team strength, but it would still give a superficial insight of interceptions. In this spirit, we only credit the interceptor by his defending contribution. In addition, a significant shortcoming is that we can not judge the situations where you expect an interception of a pass, but it does not occur. In this research, we tackle these problems.

To assess the defensive performances of players, we evaluate every situation on the pitch where a player sends a pass which we define as a game state. Out of every game state, we gain information, where some information is more relevant compared to other aspects. Therefore, we decompose the most significant factors of each game state and describe them as well as possible.

According to combinations of features and their corresponding outcomes, we train a model to recognize the most substantial combination structures of different game states. When we observe a new game state, the algorithm detects similarities to previous game states and knows how to act with a prediction outcome.

Using the machine learning models, we compute the added defensive value of an interception of a pass, where we take the collaboration of all defensive players into account. In this manner, we enable us to assess players pro-rata for their contribution. Most interestingly, we inspect the missed opportunities of defensive players to execute interceptions as well.

Next to interceptions, it creates another relevant source of information if we enable us to evaluate all types of defensive actions (e.g. touches, clearances, sliding tackles, duels) as well. Similarly, as we process interceptions, we investigate opportunities for defensive actions as well. Since our applied methods work equally for both interceptions and all defensive actions, only the target feature differs, we focus, for convenience, on an explanation related to interceptions.

4.1.1 Feature Engineering

One of our primary goals is to discover essential structure combinations of features which indicate the outcome of a pass. Therefore, we set properties which restrict us to describe features in a particular feasible region explicitly. Since we only obtain the start and end location of a pass and do not obtain the locations of other players on the pitch, it is hard to extract the primary intention of each pass out of the data. For instance, a pass could miss a teammate which results in an interception or intends to cover a relatively large distance, but an opponent intercepts the pass along the way. For this reason, we allow all information up till a fraction after the pass to include as context.

Inspired by Gurpinar-Morgan (2018), the most critical aspects of the pass are the location and the direction of the pass. So we start our set by explicitly describing these pass specific features.

For example, we could disaggregate direction into the angle of the pass, backwards or forwards, or to the side of the pitch. Hence we initialize our feature set by expressing the location and direction of the pass explicitly.

In some cases, interceptions relate to the difference in performance level between two teams. Probably, it is easier to execute an interception when playing against a relatively weak team compared to a relatively strong team. We interpret these differences in a broad sense to a multi-level analysis and apply this correction by incorporating the SciSkill index. During a season, each player's SciSkill index does not change a lot. For this reason, we use a fixed value for each player for each season in our process. On the other hand, some players operate on experience. As a result of many similar situations, they know how to handle a specific game state. Due to these reasons, we add player specific features.

Some players improve their performance when playing for a cheering home crowd. On the other hand, for some players, it works the other way around and get positive energy out of the hissing. Besides, the score of the match could influence the performance as well. Probably, most of the players are more concentrated when scores are equal in the dying minutes compared to a five-goal lead. For these reasons, we implement match specific features; we add the home or away side and update the intermediate score during the match.

To facilitate the algorithms in positional insight on the pitch, we introduce a few more match specific features. We involve both teams' formations and the corresponding relative position of the player who sends a pass.

Naturally, there are teams which play in different defensive styles after losing possession. For instance, we noticed a style of José Mourinho at several clubs such as Internazionale FC and Chelsea FC where the players were ordered to fall-back positions when the opponent is in possession. Parking the bus is a well-known term for this tactic. On the other hand, we obtain a tactic of Jürgen Klopp at Liverpool FC where the players immediately fight to return in possession after losing the ball called gegenpressing. Perhaps, predictions of positional interceptions are more reliable if we adjust a component that indicates the virtual central location of a team on the pitch.

By tracking the start location of the attacking passes of a team, we compute the average attacking location of a team. Equivalently, we track a team's locations of defensive actions to determine the average defensive location of a team. Throughout a match, in some cases, it is reasonable for a team to differ its tactic. For instance, a team could play higher up the pitch in terms of defensive pressure when it is one goal down in the dying minutes compared to the start of the match, which results in a shift of the average defensive location. To estimate both current locations of a team, we continuously update these locations during the match. From our point of view, recently occurred events are more of a determining factor for the current tactic compared to events relatively long ago. For this reason, we apply an exponential weighted moving average to estimate the current locations.

Related to these locations, we provide insights for the model at the time of a pass. For example, we indicate if the start location of the pass is behind or in front of the opponent's average defensive location, and we calculate the Euclidean distance to the opponent's average defensive location. We present our total feature set in Table 3.

Genre of features	Feature	Description
Pass specific *Direction*	Angle	Angle of passing direction.
	Categorical direction	A pass forward, wide or backward.
	Side	A pass to the right, left or straightforward.
	Pass to centre	A pass directed to the centre of the pitch.
Pass specific *Location*	Half	A pass started from own or opponent's half.
	Zone	A pass starting from the centre or flanks.
	Field line	Dividing the length pitch into three field lines.
	Distance goal	Distance to the centre of the opponent's goal.
	Distance centre spot	Distance to the centre spot of the pitch.
Player specific *Characteristic*	Weight	Weight of a player who sends a pass.
	Height	Height of a player who sends a pass.
	Age	Age of a player who sends a pass.
	SciSkill	Quality estimate of player who sends a pass.
Player specific	Characteristic average	Averages of opponent's player specific characteristics.
	Relative characteristics	Absolute difference between player specific characteristics of player who sends a pass and opponent's average team/defense/midfield.
	Sign of relative components	Positive or negative value of difference (point of view: passer).
Match specific	Home side	Indicator of home and away side.
	Score	Intermediate score of the match.
Match specific *Possession*	Formation passer	Team formation of passing player.
	Relative position passer	Relative position of passing player.
	Formation opponent	Team formation of opponent.
	Number of defenders	Boolean for 3/5 or 4 defenders opponent.
	Attacking passes location	Location of passes (for both teams).
	Defensive actions location	Location of defensive actions (for both teams).
	Distance to location	Euclidean distance to location (both locations, for both teams).
	Sign of location	In front or behind location (both locations, for both teams).

Table 3: Description of feature set.

Normalization. In some of the cases, various continuous features extremely differentiate in the variance of their values. For instance, the angle of a pass (in radians) varies between 0 and, roughly estimated, 1.57, in contrast, distance to the centre of the opponent's goal varies between 0 and 105. As a result, according to the difference of magnitude in terms of variance, the feature with more extensive range might dominate the optimization criterium. By rescaling the features, we prevent this issue. Note that most of the algorithms rescale the data set by default, but it does not harm to pre-process it in advance.

Several techniques to rescale the features are possible. Since we know there are no disproportionate outliers available in the feature set, and we are not sure about the distribution of the features itself (e.g. Gaussian distribution), we transform all features to a common scale (Lakshmanan, 2019). The transformation to a common scale makes all features' range lie between zero and one defined as normalizing. We apply Min-Max scaling which is done via the following equation

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{2}$$

where X_{norm}, X, X_{min}, X_{max} corresponds the normalized feature value, actual feature value, minimum feature value and maximum feature value, respectively.

4.1.2 Data Pre-Processing

As discussed in the data section, there are four possible outcomes for a pass; pass reaches a teammate, a pass is intercepted, a pass ends out of play, or the referee's interruption follows a pass. Let us briefly explain the added value of these circumstances.

In some situations, a defensive action after a pass causes a disruption of the game by the referee. For instance, a player makes a violation to prevent a counter-attack. The action provides information of the defending abilities of a player. Unfortunately, our on-the-ball event data does not allow us to reveal these situations. Therefore, we do not have enough information to substantiate this claim and come up with useful analyzes.

Furthermore, passes out of play do not contribute to our research as well according to the reasoning that some of these passes are not able to intercept or reach a teammate.

Our primary focus depends on an accurate prediction of an interception. Within this design, we exclude passes out of play and passes followed by referee's interruption since these situations do not have added value to this research. Contrarily, we include passes to our models that reach a teammate and intercepted passes.

Furthermore, some events do not match to player-specific data. For example, a relative position does not conform to an event, which causes that we are not able to reveal the interceptor of a pass. These passes become useless; therefore, we eliminate them as well. In the remainder of this study, we refer to usable passes.

4.1.3 Model Optimization

Repeatedly, we train models according to different settings and validate each model's performance in terms of LogLoss and AUC-PR. According to LogLoss, we evaluate the quality of each model's predictions and AUC-PR ranks the performances of the models. In this sense, we search for the best performing model.

As is common in machine learning, we divide the available data into three sets: a training set, a validation set, and a test set. We use the training set to train the model, we use the validation set to tune the hyperparameters of the machine learning algorithm that learns the model, and we use the test set to evaluate the performance of the best model. In our setup, the test set includes the matches that are used to assess the defensive qualities of the players.

For a specific model, we desire to provide similar performances for all sets which we assign as a reliable model. Therefore, we keep track of the evaluation metrics' differences concerning the training, validation and test set. If these differences are too substantial, we probably fit the model too strict to the training data, so in an ideal situation, the performances of all sets are nearly similar. In this sense, we prevent under- and overfitting. Finally, we present the performances of the evaluation metrics evaluated on the test set.

We apply this approach for our tree-based methods; Random Forest, XGBoost, LightGBM and CatBoost. Each tree-based method has various properties to set up its model defined as hyperparameters. For instance, a restriction to the number of trees in a Random Forest, or adjusting the minimum number of events per leaf in a Gradient Boosted Decision Tree. These hyperparameters enable us to restrict each model to certain conditions manually. As a result, we improve the performance of the model. On the other hand, we need to keep in mind to prevent the model for overfitting. Hence we optimize the hyperparameters according to various candidate models where we deal with a trade-off regarding performance and overfitting. In this research, we focus on learning a model as accurate as possible, which calibrates well.

4.2 Expected Interceptions

To provide more insight into the value of an interception, we incorporate the game situation in which a player performs an interception of a pass. For instance, an interception near the halfway line could have diverse added value for a team. When this interception occurs for a team which is three goals up, there is little importance. In contrast, when a team is one goal up, and the interception prevents a counter-attack, it should be highly valued. In this sense, we look at factors which influence the value of an interception that we define as the context of a pass.

By gaining relevant information of all occurred passes, we evaluate the difficulty of each pass to reach a teammate. We convert the difficulty of a pass into a probability which we define as an expected interception. To assess a player for an interception, it is reasonable to judge an interception of a pass according to the probability of the pass to reach a teammate. In advance, if we highly expect a pass to be intercepted, it makes little sense to reward high value to the interception because we expect many players to perform this interception. Contrarily to a pass which is hard to intercept, we provide a relatively high reward to the interception since it probably requires abilities to execute the interception. Following the same train of thought, we enable us to punish missed interceptions where a pass reaches a teammate.

4.2.1 Learning Task

To evaluate the difficulty of a pass to reach a teammate, we fit several learning algorithms to the training data to predict the probability of an interception for each pass. Hence, we train binary classification models with interceptions as target and pass reaches a teammate as a failure; Expected Interceptions (EI). Each classifier has benefits concerning prediction accuracy and interpretation. Moreover, by designing multiple different methods, we enable space to compare each classifier's prediction performance to the rest.

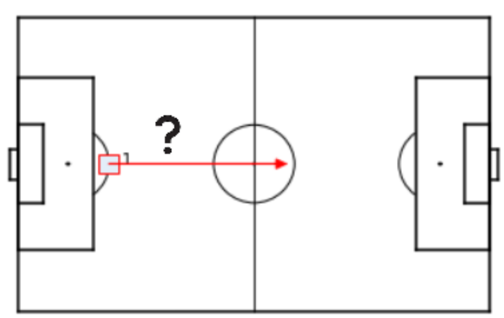


Figure 15: Learning Task Expected Interceptions.

However, to get insights into the strength of our classifiers, we firstly provide simple models to create a performance benchmark for comparison.

4.2.2 Baseline Models

We develop baseline models based on domain knowledge. The baseline models consider binary classifications with a target relating to an interception and failure determines a pass that reaches a teammate.

Baseline 1. Since a forward pass is in most cases more valuable to score compared to a pass sideways or backwards; a team possibly allows the opponent more frequently to play sideways or back, in contrast to a pass forwards. Instead, this argument goes, to provide the difficulty of a pass, it is essential to know if the pass is directed backwards or sideways, or is directed forwards. Therefore, we introduce a baseline model with the calculated averages of all obtained interceptions of passes in both categories, separately. Passes which we attempt to predict the outcome with our learning algorithms, we now attach the computed averages as replacement predictions.

Baseline 2. Since we believe it makes a difference if you give a pass in the back or near the opponent's goal, we design a model which depends on the same basics as the latter model but includes a division of the pitch in three parts as well. Hence we divide the pitch into three field lines and calculate the averages of all passes backwards or sideways, and passes forwards.

Baseline 3. Passes within a range of five meters often fail to reach a teammate because these plays suffer from opponent's pressure resulting in blocks or interceptions. Passes over twenty meters could be risky because of the relatively long time frame it takes in transit. For these reasons, we compute averages of obtained interceptions of passes within a range of five up to twenty meters, and the averages outside this range. Again, similar passes we intend to predict with the algorithms we now label these calculated averages for both categories, separately.

Now, we do have a starting point in terms of binary classification. Next, we develop using machine learning techniques several models. Eventually, we compare them to the baseline models to get a better insight into the strength of our trained models.

4.2.3 Performance Analysis

We analyze the performance levels of the binary classification models related to interceptions. To keep it clearly structured, we focus in our analysis on interceptions. In Appendix B, we present the performances of our models related to all defensive actions as well.

We develop the following classifiers: Random Forest, XGBoost, LightGBM and GA^{2}M. Especially, we focus on hyperparameters which prevent fitting too strict to the training set. We train each model on 1,231,853 passes which we subsequently validate on 350,000 passes. In Table 4, we show each method's optimized hyperparameters.

Type of hyperparameter	Random Forest	XGBoost	LightGBM
Number of decision trees	800	100*	100*
Maximum depth of tree	10	6	-1*
Learning rate	N/A	0.1	0.2
Minimum number of events per leaf	200	1*	30
Minimum number of samples per split	10	1*	1*
Subsample of features per tree	ln(2)	0.8	1*
Fraction of randomly events per tree	0*	0.8	0*
Minimum child weight	0*	5	1

Table 4: Hyperparameters of binary classification models. The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this method, and * corresponds to the method's default parameter value.

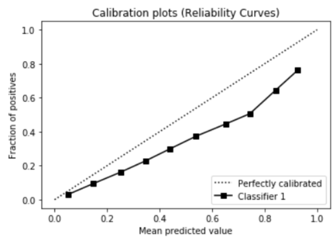
We evaluate the best performing models of each method on the test set, which contains 858,020 passes of the season 2018/2019. As we present in Table 5, we obtain that Random Forest and LightGBM provide most accurate and efficient predictions for an interception of a pass. However, as we show in Figure 16, a relatively weak calibration of Random Forest is probably caused by overfitting to the data. The dotted forty-five degrees line corresponds to a perfectly calibrated model. The closer the model curve comes to this line, the better it calibrates to the data.

When inspecting the training, validation, and test set, we notice relatively significant differences of the evaluation metrics compared to the Gradient Boosted Decision Trees. In particular, LightGBM fits exceptionally well to the data. As we see in optimizing hyperparameters, we barely need to tune them.

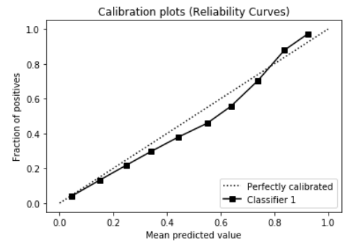
Apparently, there are a few features with relatively significant impact which influence predictions the most. When only focusing on these features, we provide the best results. Furthermore, we obtain that the baseline models are still lagging far behind compared to our trained models. Note that we do not evaluate the calibration curves for the baseline models since they are terrible calibrated by nature caused by the input of relatively imbalanced averages. In general, LightGBM, Random Forest and XGBoost provide reliable models to predict an interception of a pass, whereas GA^{2}M provides noticeable fragile performance.

	Evaluation Metric	
Method	AUC-PR	LogLoss
Random Forest	**0.950-0.330**	**0.344**
LightGBM	0.950-0.329	**0.344**
XGBoost	0.950-0.326	0.346
GA^{2}M	0.919-0.224	0.415
Baseline 1	0.888-0.164	0.388
Baseline 2	0.908-0.190	0.379
Baseline 3	0.883-0.160	0.392

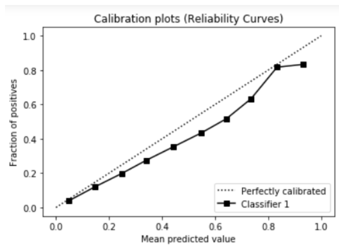
Table 5: Evaluation metrics of binary classification models and all baseline models, including only interceptions as target feature (EI). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR contains two values, where this metric evaluates: precision recall curve of pass reaches a teammate-precision recall curve of an interception.



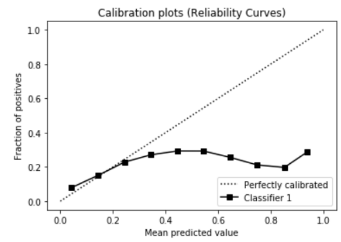
(a) Random Forest



(b) LightGBM



(c) XGBoost



(d) GA^{2}M

Figure 16: Calibration Curves, including only interceptions as target feature (EI). The dotted forty-five degrees line corresponds to a perfectly calibrated model. The closer the model curve comes to this line, the better it calibrates to the data. Note that these curves are evaluated on the test set.

Next to the models related to interceptions, we investigate all defensive actions as well. In Appendix B, we present the optimized hyperparameters and performance of binary classifications for all defensive actions (EDA). For this design, LightGBM and XGBoost provide the best performances, which suspects there are more features with relatively small impact on predictions important to consider compared to predicting an interception. Since the proportion of failure and target is more in balance, we immediately notice a performance downgrade of all models in terms of LogLoss. Note that again the performances of the models are way more accurate and efficient compared to the baseline models, except for GA^{2}M which even provides weaker performs weaker performances in terms of LogLoss compared to the baseline models. Lastly, all tree-based methods are relatively well-calibrated, compared to the models where we include only interceptions as a target feature. Overall, these models provide relatively good performances as well.

4.2.4 Model Inspection

To gain insight into the predictions of our models, we elaborate feature influences on predictions according to our Random Forest model. We provide examples in the validation set of both a correctly predicted interception and a correctly predicted pass which reaches a teammate.

Figure 17 shows a correctly predicted interception of a pass performed by player five. We note a few passes from player one to five on the pitch, where this team plays from left to right. In advance, we assess the pass of player five as a relatively risky pass directed to the box near the opponent's goal. In the waterfall chart, we obtain the most influencing features in terms of probabilities related to a prediction. Since we obtain the Random Forest classifier, feature influences are scaled over all decision trees, which is in our case eight hundred trees. The dotted line confirms the average interception rate of all passes, about sixteen per cent. The green blocks are the feature influences which force the prediction into an interception, where we notice the blue block as the total influence of all features to the prediction outcome. For instance, about ten per cent of the forecast is caused by the location's distance to the centre spot (*distance_centre_spot*), ceteris paribus. All feature influences are combined, so we can not interpret each feature separately. Note that *other* contains the rest of all features total impact to the prediction outcome.

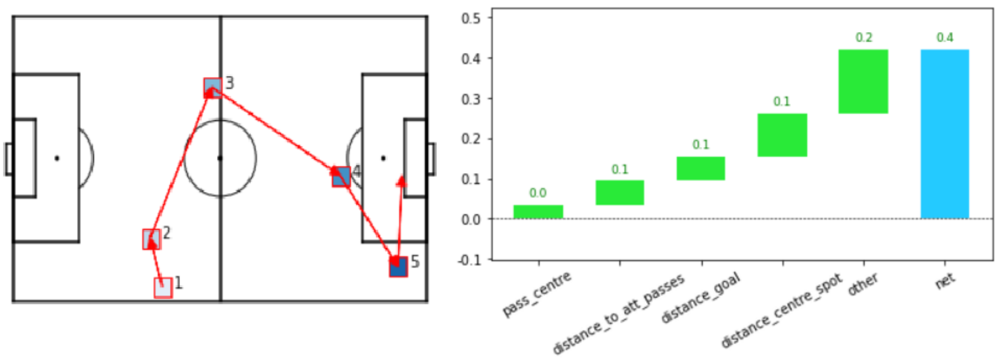


Figure 17: Event visualization of well-predicted interception. Random forest evaluates pass sent by player five. The combination of features forces the prediction to a risky pass. Note that *distance_to_att_passes* is the Euclidean between the start location of the pass and the average location of its team's attacking passes, and *other* relates to all remainder features.

Figure 18 shows a correctly predicted pass that reaches a teammate performed by player four. Again, We note a few passes on the pitch, where this team plays from left to right. On beforehand, we assess the pass of player four as a relatively risk-averse pass to the side of the pitch near the halfway line. In the waterfall chart, we obtain the red blocks as the feature influences which force the prediction into a pass that reaches a teammate. Since the pass is backwards, in this event, the categorical direction of the pass profoundly influences the prediction outcome (*cat_dir*), ceteris paribus.

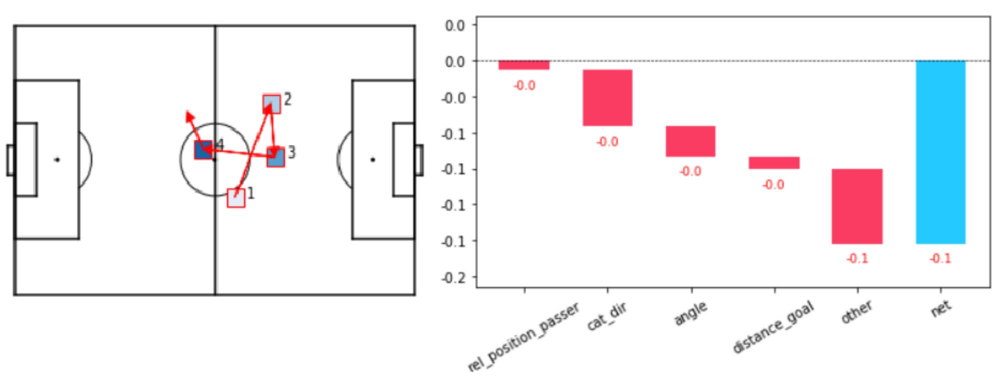


Figure 18: Event visualization of well-predicted pass which reaches a teammate. Random forest evaluates pass sent by player four, which the model assesses as a risk-averse pass according to the feature combination. Note that *rel_position_passer* corresponds to the relative position of the player who sends the pass, and *other* relates to all remainder features.

Moreover, we inspect the feature importance over all predictions as well. Now we evaluate the best performing and calibrated model according to the test set; LightGBM. One of the main properties of LightGBM entails the scan of only the most significant data instances. According to this way of thinking, we suggest a few features to be highly crucial to the prediction outcomes. In Figure 19, we present the most critical features overall predictions. If we select only the fifteen strongest features, we see nearly similar AUC-PR evaluations compared to the assessment, including all features in Table 5.

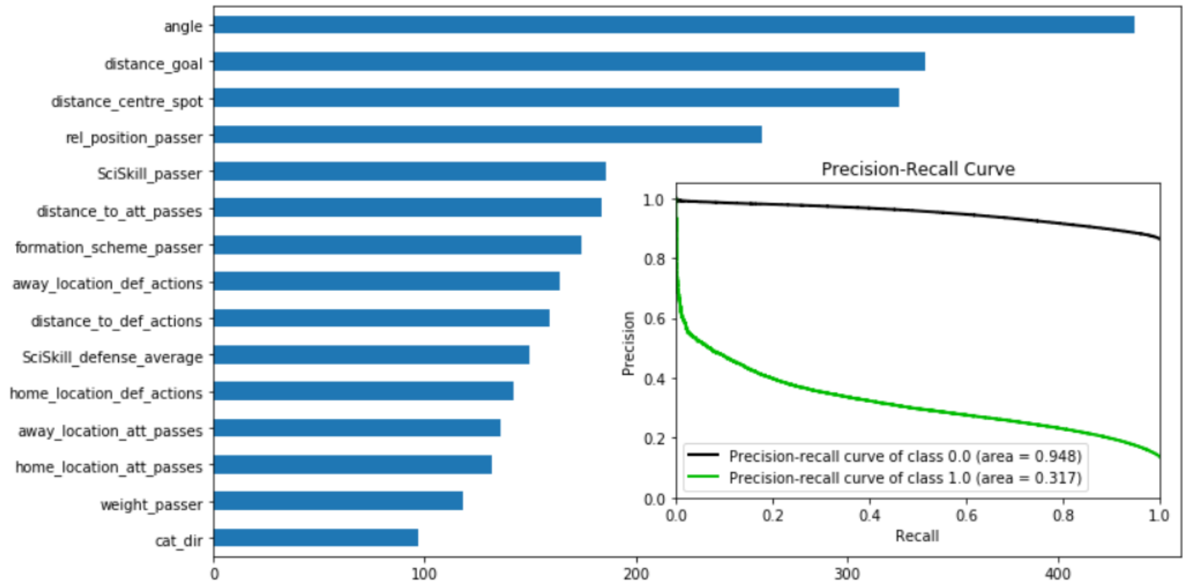


Figure 19: Feature importance LightGBM. AUC-PR curve evaluated on fifteen strongest features.

In the spirit of Gurpinar-Morgan (2018), we note the location and angle of the pass as the most crucial influences to predict the outcome of a pass. Moreover, it seems logical to incorporate the positional features and quality of the passing player, which correspond to $formation\_scheme\_passer$, $rel\_position\_passer$ and $SciSkill\_passer$.

However, many side specific features appear to be significant as well, which might be slightly unexpected. Apparently, the model gains relevant information out of the home and away playing performances to predict the outcome of a pass. Another remarkable strong feature corresponds to the weight of the passer ($weight\_passer$). Mostly playmakers do not belong to the heavyweight champions although it is an exciting finding.

4.2.5 Discussion

So far we assumed an interception is the merit of the player who performs the interception. However, defending is not always performed by a single player; we even dare to state defending is mainly achieved by multiple players. For instance, several players put pressure on the ball, which leads to an interception, or several players defend their direct opponent, which decreases the passing possibilities or perhaps teammates support each other for defensive back up. Thus we should invent ways to relate all players who are accountable for the interception and apportion responsibility where it should be.

4.3 Defensive Positional Expectations

Along with each specific pass's difficulty, we relate to the pass which players should possibly make an interception. For instance, if a central midfielder passes to a left-winger we probably expect the right back to perform an interception with a relatively high probability, perhaps the centre back and right half as well, but unlikely to expect from your central forward to conduct an interception.

This makes sense based on domain knowledge and possessional insight of the players on the pitch. Moreover, it confirms to appear in the data as well.

4.3.1 Learning Task

We provide predictions of the defending players to appear on the pitch. The first option could be to split the pitch into areas where each player defends a zone. In practice, we notice that successful defending performance often appears to be teamwork. In this sense, the defensive tasks of players are, in most cases, related and should be judged jointly.

An alternative, which incorporates the collaboration of defence, tracks all players per position on the pitch and comes up with areas where they mostly appear. A significant difference to the first option is the overlap where players possibly appear on the pitch. However, from our point of view, each significantly different game state entails another defensive possession, which we do not incorporate in this approach.

Our approach tends to this latter approach; instead, We introduce a model which evaluates the opportunity for each position to make an interception for each pass. We involve an overlap of players who can execute the interception, which can be seen as a correlation of defensive tasks. Note that these expectations depend on the assumption where, for both teams, eleven players are out on the pitch.

We fit various learning algorithms to the training data to predict the probability for each relative position to execute an interception. Hence we train a multiclass classification model with interceptions by relative positions as the target; Defensive Positional Expectations (DPE).

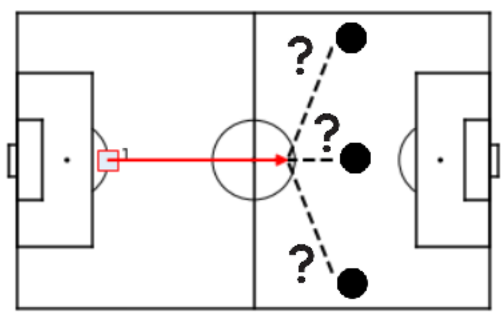


Figure 20: Learning Task Defensive Positional Expectations. For convenience we show only three defenders.

But let us first develop a defensive positional benchmark which we finally compare to our learning algorithms to gain insight into the strength of these models.

4.3.2 Baseline Models

We design baseline models where we involve the expectations of each position to make an interception. We achieve this task by incorporating the computed averages of all interceptions per position.

Baseline 1. From our point of view, it does make a considerable difference for the team possession on the pitch if it consists of either three or five defenders, compared to a team which is playing with four defenders. This significant difference relates to the defensive roles of the players in both formations. For example, we probably expect the backs of a formation either three or five defenders to play relatively higher up the pitch compared to a team which is playing with four defenders.

According to this way of thinking, we expect them to perform interceptions at different locations on the pitch. For this reason, we distinguish between these team formations in processing each position's average of interceptions.

Baseline 2. To improve the reliability of the baseline models, we divide the pitch into three parts (three field lines). We design a model which depends on the same basics as the latter model and calculate the averages of the positional interceptions based on the passes sent from each field line.

Baseline 3. Again, we start with the basics both of the two previous baseline models. To further extend, we process each position's interceptions if passes are directed backwards or sideways, or is directed forwards.

Since we provided a benchmark for the multiclass classification, we now focus on the learning classification models related to defensive positional expectations.

4.3.3 Performance Analysis

Now we analyze the performance levels of the multiclass classification models related to interceptions. In Appendix C, we present the performances of our models related to all defensive actions. We implement Gradient Boosted Decision Trees to test for accurate and reliable performance; CatBoost and XGBoost. We train each model on 159,494 passes which we subsequently validate on 50,000 passes. Since we train the model which player intercepts a pass, we only incorporate intercepted passes, whereas we incorporate passes related to all defensive actions in Positioning. Within that approach, we train each model on 350,301 passes which we subsequently validate on 50,000 passes. In Table 6, we present each method's optimized hyperparameters. By default, the learning rate of CatBoost is defined automatically close to the optimal one based on the data set properties and the number of iterations.

Type of hyperparameter	**CatBoost**	**XGBoost**
Number of decision trees	100	100*
Maximum depth of decision tree	6	6
Learning rate	0.15*	0.1
Objective	Multiclass	Multi:softprob
Amount of randomness for scoring splits	1	N/A
Coefficient L2 regularization cost function	1	N/A
Subsample of features per tree	1	0.8
Fraction of randomly events per tree	0*	0.8
Minimum child weight	1*	5

Table 6: Hyperparameters of multiclass classification models, including interceptions as target feature (DPE). The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this particular method, * corresponds to the method's default parameter value.

When inspecting the performance of each player on the pitch separately, we need to classify each player to a specific position on the pitch as clear as possible regarding its occurrence at spatial areas and defensive tasks. The process of assigning players efficiently is widely defined as the alignment challenge (Sha et al. (2017), Le et al. (2017)).

During this research, we classify each possible position into a relative position which corresponds to a specific formation (Wyscout, 2019). Position 0 always corresponds to the goalkeeper.

Since most teams play with four defenders, mainly position 1 to position 4 are defenders. Counting upwards, we define the midfielders and attackers. The relative positions are compiled in overlapping player roles as much as possible for the most common formations, although it is a hard task to provide these positions efficiently. For example, position 6 corresponds to a defensive midfielder in a 4-3-3 formation and a right central midfielder in a 4-4-2 formation. Another example is more varied. Position 4 is a left-back in a 4-3-3 formation; however, it is a right-wing-back in a 3-4-3 formation although both positions are located on the flanks. Moreover, players possibly swap over positions at off-the-ball events during the match which we can not incorporate in these relative positions.

As we show in Tables 7 and 8, we obtain a considerable advantage for XGBoost compared to CatBoost. In comparison to our baseline models, we note relatively good performance. Particularly, we notice efficient performance for defenders (position 1 through position 4) which confirm for about twice the AUC-PR performance of the baseline models. Since midfielders probably appear in more different situations compared to defenders, it is reasonably harder to provide efficient predictions for each game state.

Relative position	**AUC-PR**				
	CatBoost	XGBoost	Baseline 1	Baseline 2	Baseline 3
Position 0	0.19	0.22	0.07	0.10	0.10
Position 1	0.31	0.33	0.12	0.12	0.13
Position 2	0.28	0.29	0.14	0.15	0.17
Position 3	0.29	0.30	0.14	0.15	0.16
Position 4	0.32	0.33	0.12	0.12	0.13
Position 5	0.16	0.17	0.09	0.10	0.11
Position 6	0.17	0.17	0.11	0.12	0.12
Position 7	0.15	0.16	0.08	0.09	0.11
Position 8	0.16	0.16	0.06	0.07	0.10
Position 9	0.15	0.15	0.04	0.05	0.09
Position 10	0.15	0.15	0.03	0.04	0.09

Table 7: AUC-PR of multiclass classification models and all baseline models, including interceptions as target feature (DPE), for each relative position separately. The performances are evaluated according to the test set.

Method	**Evaluation Metric**	
	AUC-PR	LogLoss
CatBoost	0.240	2.006
XGBoost	**0.250**	**1.991**
Baseline 1	0.120	2.317
Baseline 2	0.130	2.304
Baseline 3	0.140	2.247

Table 8: Evaluation metrics of multiclass classification models and all baseline models, including interceptions as target feature (DPE). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR are weighted averages over all relative postitions.

However, when comparing these performances to the performances of the same models to our validation set in Tables 9 and 10, we observe significantly different performances for XGBoost which suggests overfitting. For this reason, we inspect the learning curves of both models. If the curves of the training and validation set converge, the model provides similar performance on both sets. Moreover, the point of convergence declares how many events we need in our training set to provide the corresponding performance.

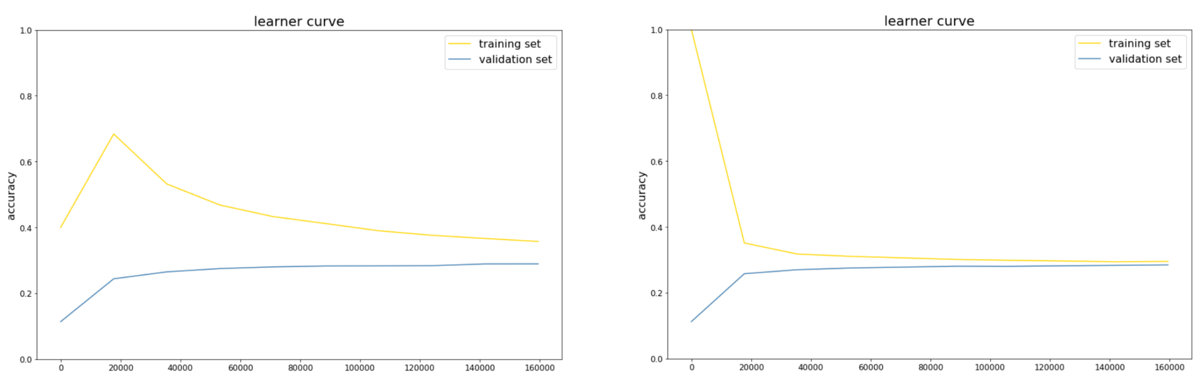
In Figure 21, we note XGBoost suffers from overfitting since the curves do not converge to a similar accuracy. For this reason, we consider CatBoost to provide more reliable performances.

Relative position	**AUC-PR** CatBoost	XGBoost
Position 0	0.21	0.30
Position 1	0.33	0.39
Position 2	0.33	0.38
Position 3	0.32	0.37
Position 4	0.34	0.39
Position 5	0.18	0.26
Position 6	0.19	0.28
Position 7	0.18	0.29
Position 8	0.19	0.26
Position 9	0.20	0.27
Position 10	0.17	0.27

Table 9: AUC-PR of multiclass classification models, including only interceptions as target feature (Defensive Positional Expectation), for each relative position separately. The performances are evaluated according to the validation set.

Method	**Evaluation Metric** AUC-PR	LogLoss
CatBoost	0.270	1.933
XGBoost	**0.320**	**1.854**

Table 10: Evaluation metrics of multiclass classification, including only interceptions as target feature (Defensive Positional Expectation). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR are weighted averages over all relative postitions.



(a) XGBoost (b) CatBoost

Figure 21: Learning curves, including only interceptions as target feature (DPE). If the curves of the training and validation set converge, the model provides similar performance on both sets.

Since we are also interested in players who are well-positioned but lack in feasible interceptions, we train a multiclass classification model which involves all occurred defensive actions as target feature as well (Positioning). For instance, these could be young players who perhaps miss stratagems or experience to execute the interception.

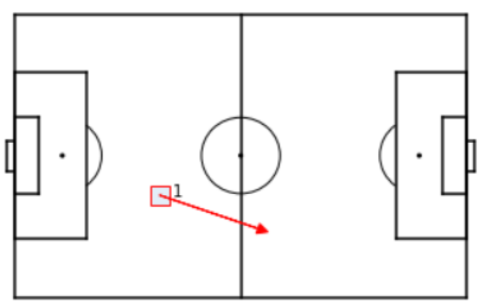
In Appendix C we show the multiclass classification performances and its corresponding hyperparameters that applies this setup. Since the baseline performances of including interceptions as a target are nearly similar to adding all defensive actions as the target, we enable us to compare the learning algorithms, including different targets. Most of the relative positions perform similar performances except for a few defensive positions; here, we notice a slight advantage for the learning algorithms, including all defensive actions as the target feature.

Overall, we see again that XGBoost provides the best performance. According to the same line of reasoning as we did for the models including only interceptions, we obtain overfitting when investigating the performance according to the validation set and considering the learning curve in Appendix D.

4.3.4 Model Inspection

To gain insight into the predictions of the models, we elaborate on the feature influences related to these predictions. We provide an example of our XGBoost model evaluated on the validation set. Since we are not able to use the waterfall chart as in binary classification, because the target feature has more than two classes, we apply a commonly used method which provides each feature's impact on the prediction of a relative position to perform an interception, while keeping the rest of the features fixed defined as SHAP value (Parsa et al., 2020).

As we present in Figure 22, we see a pass from the back near the halfway line directed to the right centre of the pitch. On beforehand, we suspect a midfielder or perhaps a left-back to make an interception. XGBoost anticipates by assigning the highest prediction to the central midfielders to intercept the pass (position 6 and 7). Furthermore, the centre backs in most formations get a relatively high probability as well (position 2 and 3), which is slightly unexpected. The actual outcome is an interception executed by position 7.



Relative position	Probability
Position 0	0.036
Position 1	0.047
Position 2	0.162
Position 3	0.126
Position 4	0.049
Position 5	0.050
Position 6	0.172
Position 7	0.261
Position 8	0.045
Position 9	0.032
Position 10	0.020

Figure 22: Defensive positional expectations to event. All probabilities of each relative position to execute an interception.

In Figure 23, we get more insight into the effects of the features concerning an interception performed by a midfielder, relative position 7. Each feature's range is scaled over high values in red and low values in blue. When we notice an observation (coloured dot) on the right side of the vertical central baseline (0.0), it means the impact of this feature value increases the probability of relative position 7 to execute an interception, ceteris paribus. In contrast, we note an effect which decreases the probability on the left side.

The importance of each feature on average to all predictions of relative position 7 to perform an interception is lined up in descending order. All impacts of the features to the forecasts together form a curve over the width. When we note a thick part of the feature's curve, it means this feature realizes many impacts of this specific magnitude to the predictions. For example, the effect which increases the probability of an interception of relative position seven according to *cat_dir* varies a lot, whereas the impact which decreases the probability is often similar.

Let us take a look at some features' impacts. For instance, the model tends to classify relative position 7 as an interceptor when we obtain a relatively low relative position of the passing player (*rel_position_passer*). This seems logical since these players mostly appear at the back or the centre of the pitch in the neighbourhood of the midfielders.

When the passing player gets closer to the opponent's goal, we notice a decreasing probability of relative position 7 to execute an interception (*distance_goal*), which seems to be quite interconnected to the latter case. Probably, the most likely players to make an interception shifts to the defenders.

Returning to Figure 22, we note a pass relatively far away from the opponent's goal, probably sent by a player with a low relative position. This line of reasoning confirms an argument for the model to predict relative position 7 to make an interception highly.

We note some remarkable effects as well. All averages note the indicated proportions of the defending team. We obtain a stimulating effect to the probability for relative position 7 to make an interception in case of a relatively low valued defence according to the SciSkill indices (*SciSkill_defence_average*). This means when playing with weaker teammates in defence, the probability to execute an interception increases. On beforehand, we expect a team to put relatively many pressure on the weak spots of the opponent. This tactic suspects to skip possession on the midfield which should decrease the probability of an interception performed by relative position 7. Perhaps, players at this position get a more defensively focused task on average when their team has a relatively weak defence.

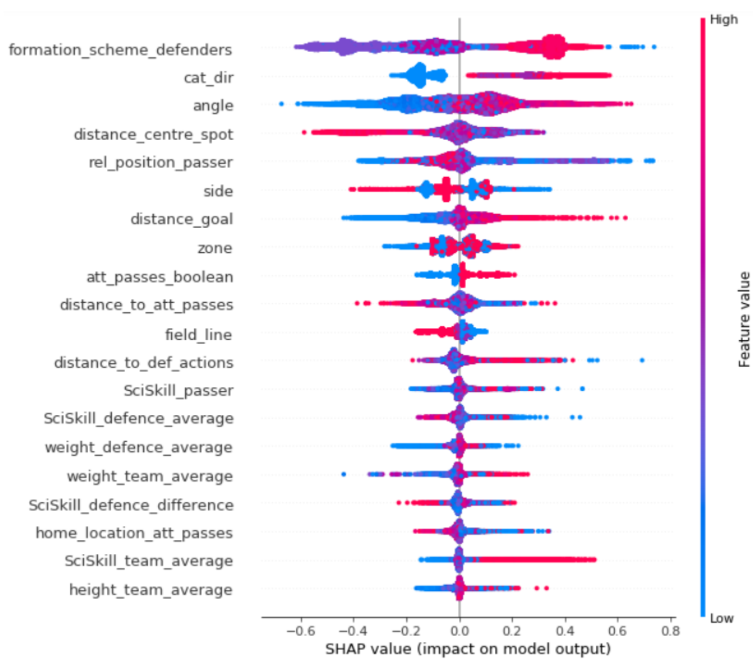


Figure 23: Feature importance of relative position 7 (midfielder). Indicated feature values on the right side increase the probability of an interception executed by relative position 7, and the left hand side vice versa.

In Figure 24, we note the overall feature importance related to all predictions of relative positions to perform interceptions in the validation set. We note as expected the pass and positional specific features to be highly crucial in predicting a relative position to execute an interception.

Each coloured area within an average feature impact describes the importance of this feature to the corresponding relative position to execute an interception on average. A larger area relates to a more useful feature. For instance, we obtain the most extensive area for *formation_scheme_defenders* related to relative position 7. This means the formation scheme of the defensive team is the most crucial feature to predict an interception performed by this relative position over all predictions, which is in line with our findings in Figure 23.

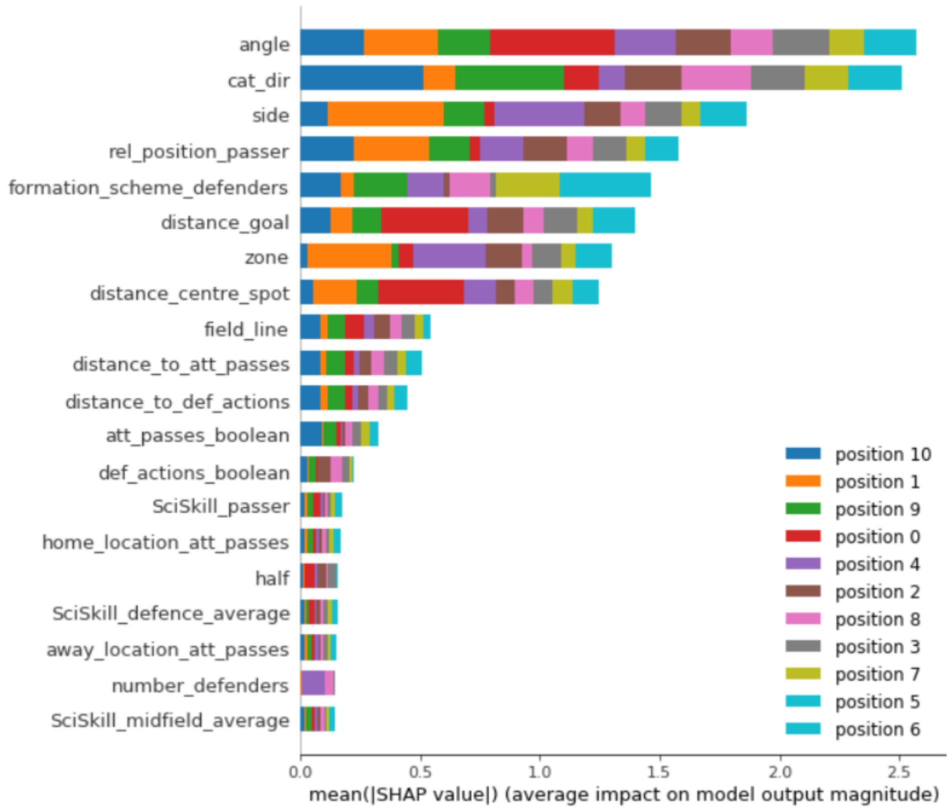


Figure 24: Feature importance of Defensive Positional Expectations. We obtain the discriminating power of each feature concerning all relative positions to execute an interception.

4.3.5 Discussion

We introduce a way of thinking to evaluate possible interceptions in two steps; "Will the pass be intercepted?" computed by the passing difficulty, followed by the assignment "Who intercepts the pass?". In this sense, we approach both tasks independent of each other. According to our performance evaluations, we provide reliable methods to quantify the defensive abilities of players.

However, it is conceivable that the assignments of an interception and the player who possibly performs the interception are related to each other, which triggers to design an experimental setting considering this approach.

4.4 Expected Positional Interceptions

Now, we establish a design which evaluates the passing difficulty and the defensive positional expectation of the players at once. From our point of view, a possible correlation between these two objectives could entail a decisive factor by determining who we expect to execute an interception. For instance, a simple pass in the back will probably not be intercepted. In our current approach, we predict the probability for every player of the opponent to intercept this pass. Perhaps, it is unrealistic to expect any player to make an interception since it is simply out of their reach. For this reason, this setup brings a potential advantage.

4.4.1 Learning Task

We fit two learning algorithms to the training data to predict the probability of an interception per position and the probability of no interception at once. Hence, we train multitask classification models with interceptions by relative positions and no interception as the target; Expected Positional Interceptions (EPI).

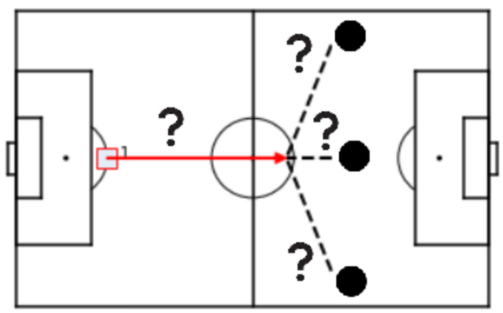


Figure 25: Learning Task Expected Positional Interceptions. For convenience we only show three defenders.

Again, let us first develop a benchmark for comparison based on this one-step approach.

4.4.2 Baseline Models

We develop baseline models in a similar way of thinking as we did for defensive positional baseline models. The main difference of these averages holds that the two-step approach only deals with intercepted passes, where the one-step approach contains passes which reach a teammate as well. Obviously, the proportions of the defensive positional baseline models are probably more substantial considered in absolute value compared to the baseline models we now design but are relatively the same.

Since we provide a benchmark for the one-step approach, we now focus on the learning classification models.

4.4.3 Performance Analysis

To evaluate the expected positional interceptions' approach, we train multitask classification models (EPI). In Appendix F, we show the performances of our models related to all defensive actions. We implement Gradient Boosted Decision Trees to test for the best performance; CatBoost and XGBoost. We train each model on 1,231,853 passes which we subsequently validate on 350,000 passes. In Table 11, we present each method's optimized hyperparameters. Note that the rest of the hyperparameters are set by default.

Type of hyperparameter	**CatBoost**	**XGBoost**
Number of decision trees	100	100*
Maximum depth of decision tree	6	6
Learning rate	0.15*	0.1
Objective	Multiclass	Multi:softprob
Amount of randomness for scoring splits	0.5	N/A
Coefficient L2 regularization cost function	4	N/A
Subsample of features per tree	0.8	0.8
Fraction of randomly events per tree	0*	0.8
Minimum child weight	1*	5

Table 11: Hyperparameters of multitask classification models, including only interceptions as target feature (EPI). The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this method, and * corresponds to the method's default parameter value.

In Table 12 and Table 13, we notice a slightly better performance for XGBoost compared to CatBoost. We obtain for both the multitask classification models and the baseline models a large difference between AUC-PR of all relative positions to intercept a pass and a pass which reaches a teammate (no interception). Apparently, the task to distinguish between a particular relative position to execute an interception and all other possibilities is relatively hard to accomplish for the model compared to the task to discriminate for no interception. On the other hand, the model provides relatively good performance for most of the relative positions compared to the baseline models.

Again, we note relatively high performance for defenders, even better than multiclass classification performance. Probably, the model enables to distinguish more successfully when to incorporate a potential interception of a defender for each pass. For example, when the opponent's right back passes to the centre back, the model predicts a negligible prediction for a defender to intercept this pass. Again, the midfielders' interceptions are harder to predict. Perhaps we need to explain each game state more explicitly in our feature set to provide better performance regarding these relative positions.

Moreover, it is imaginable that the model's low performance regarding midfielders has to do with the assignment of midfielders to the corresponding relative positions. Notably, at the midfield, these relative positions are difficult to accurately generalize due to their possible different movements and functions on the pitch. For instance, a relative position 6 of team A could play in a more defensive tactic compared to a relative position 6. However, both teams confirm similar formations, which makes it hard for the model to act appropriately at all time for this relative position.

Furthermore, at the midfield players probably swap more frequently over positions which we can not trace continuously during the match. This provides issues when, for example, a defender makes an interception, and we assign wrong rewards to the relative positions of midfielders.

For attackers (position 9 and 10), we notice weak performance. If we apply the averages computed by baseline 2 or 3 to predict the probability of an interception for each pass, we confirm similar performance. Probably, the model predicts a value near to zero for each pass to be intercepted by an attacker.

Avoid the use of overfitted models we inspect the performances according to the validation set and learning curves of our best performing XGBoost model. In Appendix E, we present reliable performances for this model.

The overall AUC-PR scores are relatively large compared to the multiclass classification performances, which results from a relatively large number of correct predictions where passes reach a teammate. Moreover, LogLoss of the multitask classification models are relatively small compared to the multiclass classification models.

Relative	**AUC-PR**				
position	CatBoost	XGBoost	Baseline 1	Baseline 2	Baseline 3
Position 0	0.06	0.08	0.01	0.02	0.02
Position 1	0.06	0.07	0.02	0.02	0.03
Position 2	0.09	0.09	0.02	0.03	0.03
Position 3	0.08	0.09	0.02	0.03	0.03
Position 4	0.06	0.07	0.02	0.02	0.03
Position 5	0.03	0.03	0.01	0.01	0.02
Position 6	0.03	0.04	0.01	0.02	0.02
Position 7	0.02	0.03	0.01	0.01	0.01
Position 8	0.02	0.02	0.01	0.01	0.01
Position 9	0.01	0.01	0.01	0.01	0.01
Position 10	0.01	0.01	0.00	0.01	0.01
No interception	0.95	0.95	0.87	0.88	0.91

Table 12: AUC-PR of multitask classification models and all baseline models, including interceptions as target feature (EPI), for each relative position separately. The performances are evaluated according to the test set.

	Evaluation Metric	
Method	AUC-PR	LogLoss
CatBoost	**0.880**	0.628
XGBoost	**0.880**	**0.618**
Baseline 1	0.780	0.712
Baseline 2	0.800	0.705
Baseline 3	0.830	0.684

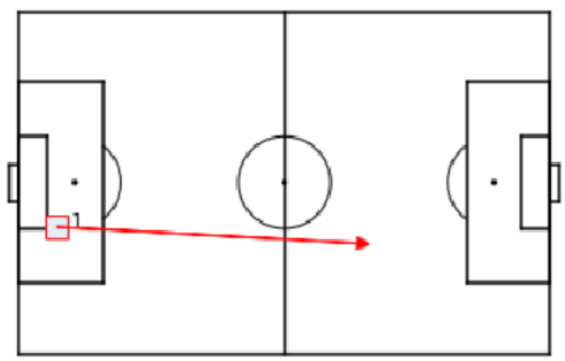
Table 13: Evaluation metrics of multitask classification models and all baseline models, including interceptions as target feature (EPI). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR are weighted averages over all relative postitions.

Next, we design a multitask classification model for defensive actions as well (EDA). As we present in Appendix F, especially the performance of XGBoost for predicting relative positions' interceptions is excellent compared to the baseline model. Moreover, if we consider the strength compared to each own baseline model, the performance of this set up is even better than including only interceptions as target feature for most of the relative positions. Although, the overall performance is considerably worse than the multitask classification, including interceptions as target feature which is caused by the large number of passes that reach a teammate and creates a distorted picture.

4.4.4 Model Inspection

To provide insight into the predictions of our models, we elaborate on the feature impacts to these predictions. We give an example in the validation set of our XGBoost model. Again, we apply SHAP values to explain the impact of the features to the predictions.

In Figure 26, we see a player, probably a goalkeeper, who directs the ball straight forward through the centre. In some cases, teams remain calm and try to solve it with their positional play, such as Manchester City, but most of these gameplays are launches for better safe than sorry. When obtaining such a game state, we expect an aerial duel most likely performed by centre backs or in some cases midfielders. As expected, these relative positions retain the highest probabilities to execute an interception. Note that we expect an interception for about eighty per cent and no interception for approximately twenty per cent.



Relative position	Probability
Position 0	0.007
Position 1	0.042
Position 2	0.298
Position 3	0.167
Position 4	0.011
Position 5	0.113
Position 6	0.121
Position 7	0.017
Position 8	0.008
Position 9	0.005
Position 10	0.004
No interception	0.207

Figure 26: Correctly predicted event by Expected Positional Interceptions. All probabilities of each relative position to execute an interception, and the probability of no interception.

As we present in Figure 27, we see the most significant effects related to an interception performed by relative position 2. This position refers to a centre back. Firstly, we notice two quite interconnected features in terms of the direction of a pass (*angle* and *cat_dir*). These features denote that when a pass has a low angle, which means in our feature description a pass forwards, the probability for relative position 2 to make an interception increases. On the other hand, when sending a pass backwards or sideways decreases this probability. This makes sense in terms of positional overview of the pitch since the centre backs are the last resort of a team.

The distance to the centre spot can be interpreted in various settings (*distance_centre_spot*). On the one hand, these could apply for passes near the opponent's goal. On the other hand, they could relate to launches sent from the back of the pitch. Moreover, we notice when the pass is sent near the halfway line, the model expects a low probability of the centre back to make an interception.

So when coming back to 26, we note a launch from the box, straightforward to the other side of the pitch. The combination of these features increases the probability for a centre back to make an interception.

However, we see some notable odd feature effects as well. For example, when inspecting the relative position of the passer (*rel_position_passer*). According to the latter argumentation, when players in the back or front send a pass, we would suspect the centre back's probability of an interception to increase. This feature tends to give a distorted picture by indicating relatively low relative positions of passers to both a decrease and an increase of the probability for a centre back to execute an interception. Perhaps this example emphasises the combination of these direction features.

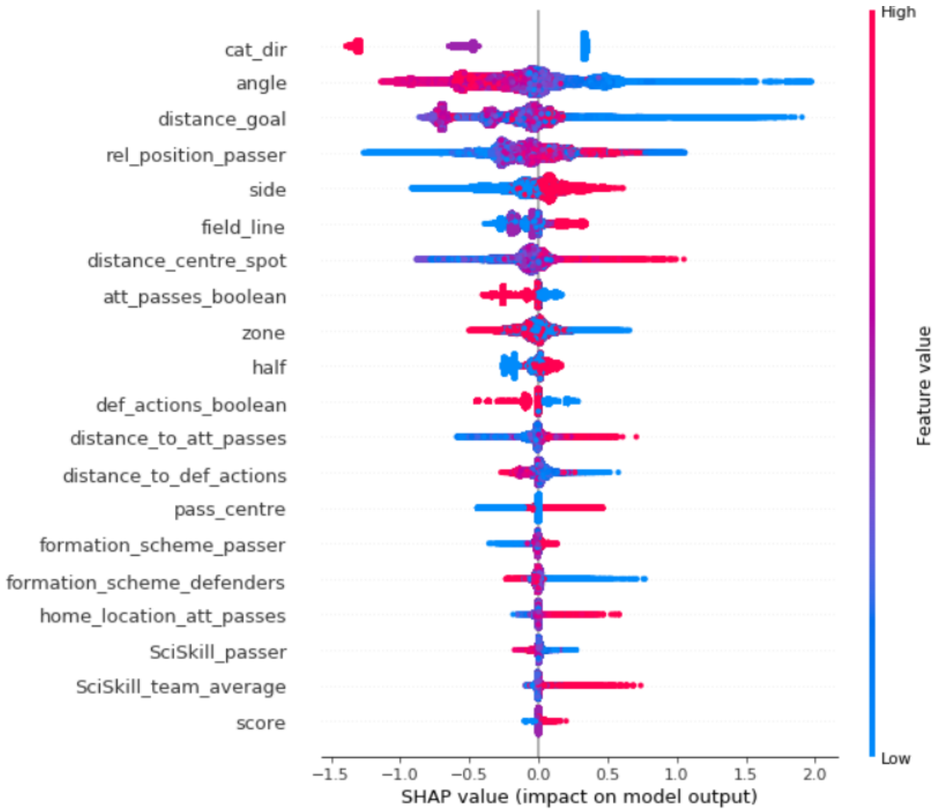


Figure 27: Feature importance of relative position 2. Indicated feature values on the right side increase the probability of an interception executed by relative position 2, and the left hand side vice versa.

Figure 28 presents the average features' impact to predict interceptions for all relative positions and no interception. We note quite similar significant features as we do for the Defensive Positional Expectations approach.

Obviously, a significant difference is the impact of the features for the model to predict no interception. The model gains most of the power from the categorical direction of the pass (*cat_dir*).

We expected the location of the attacking team's passes and primarily the defending team's defensive actions to be more substantial. Employing these features, we can reveal most of the positions of the opponent to make an interception. In particular, we could explain there is no opportunity to make an interception.

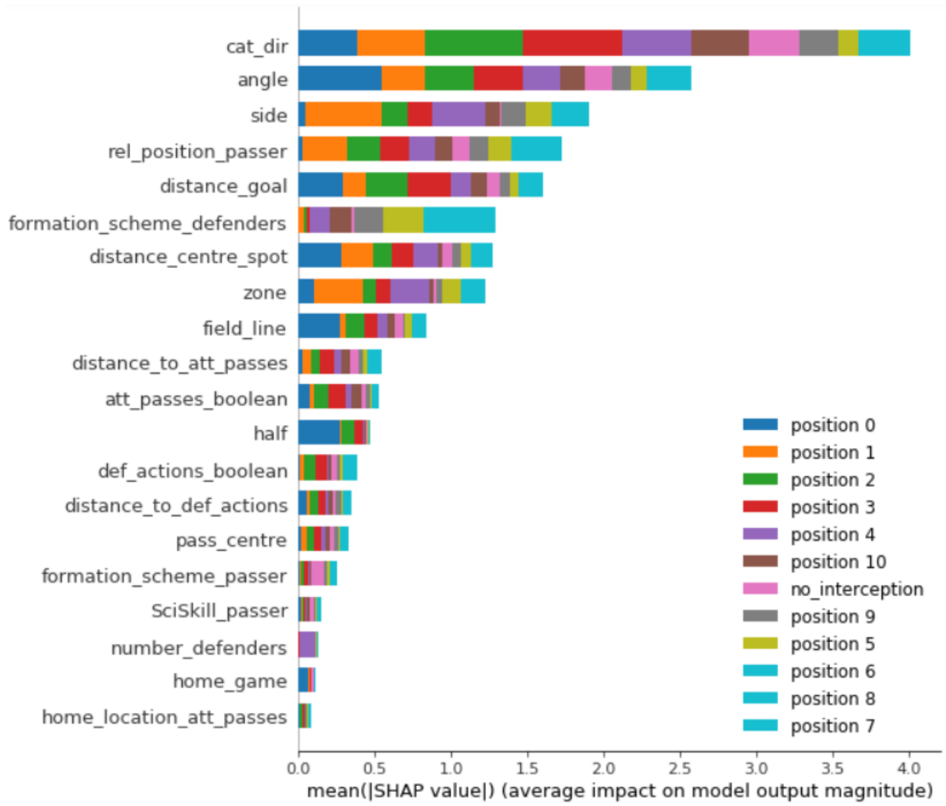


Figure 28: Feature importance of Expected Postional Interceptions. We obtain the discriminating power of each feature concerning all relative positions to execute an interception and no interception.

4.4.5 Discussion

Next to the approach to evaluate the defensive performances of players in two stages, we now provide a method in one go. In advance, we foresee a potential advantage to incorporate the correlation if a pass will be intercepted, and which relative position probably executes an interception. Eventually, it seems to be hard for the model to discriminate between an interception of a particular relative position and the rest of the outcomes. However, especially regarding defenders, we provide reliable performances compared to the baseline models.

The last step considers the transformation of these approaches to the real value of each player.

4.5 Defensive Ability Valuation

We achieve methods to evaluate the passing difficulty and the defensive positional expectations of all players for each specific pass. We realize these objectives in two independent steps and, on the other hand, in one correlated step. To quantify players' defending abilities, we need to convert our approaches to a real value.

First, let us provide a quick recap with the most relevant information to accomplish the valuation process. Based on different game states, which we implement as context, we train our models to recognize various game states on the pitch. This approach enables us, in contrast to existing metrics, to judge off-the-ball events as well.

Besides, we improve the reliability of this research by depending our metrics on the collaborations of a team instead of focussing on individual player performances. As a result, when a team performs well, probably multiple players of the team are highly evaluated, and vice versa.

Since our valuation systems work similarly for either interception as defensive actions, only the target feature differs, we focus, for convenience, on an explanation related to interceptions.

4.5.1 Two-Step Approach Valuation

Within this approach, we judge players according to the pass difficulty and the corresponding defensive positional expectation of the players on the pitch for each game state. When focusing purely on interceptions, we evaluate Expected Interceptions (EI) and Defensive Positional Expectations (DPE). By concentrating on all defensive actions of players, we value Expected Defensive Actions (EDA) and positioning. As a result of each game state, several possibilities could occur. Now let us elaborate on the different situations and explain our validation mechanisms.

(1) Pass not intercepted. We obtain a pass that reaches a teammate. We penalize all players at the level we expect them to make an interception of each specific pass j. For instance, if a convenient opportunity occurs to make an interception, we predict a significant interception probability. If no one grasps this opportunity, we penalize relatively hard. On the other hand, when it is hard to realize an interception, we assign a relatively small penalty. Note that a penalty corresponds to a negative reward. Besides, we scale the penalty by the rate we expect each player i to make the interception according to its position

$$Penalty_{ij} = Defensive\ Positional\ Expectation_{ij} \times \Pr[Interception_j]. \tag{3}$$

Note that the probability of an interception scales the penalties which correct for idle passes of a team. For example, when a team is passing a relatively high frequency on their own half, this probably not results in high penalties courtesy of the relatively low interception probabilities. Contrarily, an occurrence which we can not compensate is when two players (or even more) swap over positions. We can not obtain these movements out the data, so we always punish the players according to their original position, which is a research deficiency.

(2) Pass intercepted by most likely player. We obtain an interception of a pass by player i we most likely expect to make the interception. We reward the players pro-rata by their contribution to the interception provided by the positional expectation. If a pass is likely to be intercepted, the reward should be relatively low and vice versa. Hence now we scale the rate by the probability of a pass reaching a teammate for each specific pass j

$$Reward_{ij} = Defensive\ Positional\ Expectation_{ij} \times (1 - \Pr[Interception_j]). \tag{4}$$

In this sense, we reward teammates of the player who performs the interception as well. This logic relies on controlling the pitch as a team.

(3) Pass intercepted by unexpected player. We obtain an interception of a pass by player i we do not most likely expect to make the interception. So we obtain an interception by player B where we most likely expected player A to make the interception. Most conceivable situations which could occur in this case is that player A and B swap over positions, or player A neglects its defending task which allows player B to correct. In the long run, we think it is more likely to assume that players swap over positions because when a team is defensively outnumbered, which occurs when player A neglects its defensive task, it will mostly be played out by the attacking team. In case they not succeed to play out this situation, it is more attacking inability than extra defending ability, so we do not grant an additional reward for the interception. The rest of the players, we reward pro-rata according to their contribution to the interception provided by defensive positional expectations. Hence we switch the rates of the players' A and B and the rest of the players i ((A,B) $\notin i$) get a reward according to their own defensive positional expectation

$$Reward_{Aj} = Defensive\ positional\ expectation_{Bj} \times (1 - \Pr[Interception_j]), \tag{5}$$

$$Reward_{Bj} = Defensive\ Positional\ Expectation_{Aj} \times (1 - \Pr[Interception_j]), \tag{6}$$

$$Reward_{ij} = Defensive\ Positional\ Expectation_{ij} \times (1 - \Pr[Interception_j]). \tag{7}$$

In this sense, we still mostly reward the player that performs the interception, which is reasonable thanks to the execution of the interception.

4.5.2 One-Step Approach Valuation

This approach entails the assignment of the pass difficulty and the defensive positional expectation of all players for each pass in one correlated design. The model estimates the probability of a pass reaching a teammate, and it computes the probability of each defending player to execute an interception. When we focus strictly on interceptions, we evaluate Expected Positional Interceptions (EPI). If we focus on all defensive actions, we evaluate Expected Positional Defensive Actions (EPDA). For each possible situation, we convert the mentioned probabilities to make an interception of each player into a valuation. The logic behind the mechanisms correspond to the two-step approach valuation.

(1) Pass not intercepted. We obtain a pass that reaches a teammate. As discussed, the opportunity for each player to make an interception connects to a corresponding probability for each relative position. We penalize player i with this corresponding value of pass j

$$Penalty_{ij} = Expected\ Positional\ Interception_{ij}. \tag{8}$$

(2) Pass intercepted by most likely player. We obtain an interception of pass j by player i we most likely expect to make the interception. We apply the same logic of valuation as done with the penalty. However, we now reward the players for making an interception of pass j

$$Reward_{ij} = Expected\ Positional\ Interception_{ij}. \tag{9}$$

(3) Pass intercepted by unexpected player. We obtain an interception of pass j by a player we do not most likely expect to make the interception. So for convenience, we use the same example as we did in the two-step approach; we obtain an interception by player B where we most likely expected player A to make the interception. Due to a similar way of thinking, we switch the expected positional interceptions of player A and B, and we keep the rates of all other i players fixed ((A,B) $\notin i$)

$$Reward_{Aj} = Expected\ Positional\ Interception_{Bj}, \tag{10}$$

$$Reward_{Bj} = Expected\ Positional\ Interception_{Aj}, \tag{11}$$

$$Reward_{ij} = Expected\ Positional\ Interception_{ij}. \tag{12}$$

4.5.3 Normalize Valuation

According to the valuation systems, we provide a value for the defensive abilities of players, where we assess all situations related to opportunities for players to execute an interception. Logically, when players gain relatively many opportunities to execute an interception, they acquire a more significant probability to get a larger valuation compared to players who attend relatively few opportunities. This line of thought brings more opportunities for players of teams which are mainly under defensive pressure. In general, these are the relatively weaker teams.

Therefore, we count the defensive positional expectations and the expected positional interceptions for the two-step and one-step approach, respectively, which we define as interception opportunities. By scaling the valuation of each player with its interception opportunities' sum, we even the playing field. So in a broad sense, we provide a normalization of the defensive valuation to prevent an unfair comparison of players.

4.6 Discussion Football Analytics

In this section, we develop a metric using machine learning techniques to measure the defensive performance of football players. By training models including various game states, we estimate the difficulty of an interception and predict which players to execute an interception. We provide these objectives as two independent stages in our two-step model and provide these tasks in once by incorporating the correlation of them in our one-step model. Both approaches confirm their pros and cons for predicting the interception of a pass. In general, we notice better performance related to defenders compared to the rest of the players.

Within our approach, we enable us to involve the collaboration of players in terms of defensive performance. Above all, we allow us to assess missed opportunities of players to make an interception.

In the process to generate our models, we continuously consider the trade-off between the performance of the model and preventing the model for overfitting. By avoiding the models for overfitting, we allow to provide the optimal models within our range.

Finally, we rate the players according to their performances with the aid of our valuation systems.

5 Results and Evaluation

In this section, we convert the performances of players into valuations. To start, we clarify the structure of ranking lists regarding players' valuations. Next, we present and interpret the final results.

5.1 Structure of Results

Our main learning tasks require to estimate the difficulty of an interception and predict which players to execute an interception. Therefore, we provide these objectives as two independent stages in our two-step model and provide these tasks in once in our one-step model. We apply our valuation system to our models which confirm the best and most reliable performances according to our evaluation metrics. We assess the performances of the players on 858,020 passes out of 2018/2019 season. These passes contain interceptions made by the opponent and passes which reaches a teammate.

Both approaches enable us to focus on strictly interceptions or all defensive actions. To focus purely on interceptions, we apply Expected Interception in combination with Defensive Positional Interceptions (two-step) or Expected Positional Interceptions (one-step). When concentrating on all defensive actions, we use Expected Defensive Actions and Positioning (two-step) or Expected Positional Defensive Actions (one-step).

The performance gap between interceptions and all defensive actions indicates relevant information which relates to experience, stratagems or even ingenuity of players to execute an interception. Since the procedures to provide ranking tables of players within each focus works similarly, for convenience, we relate our explanation to interceptions.

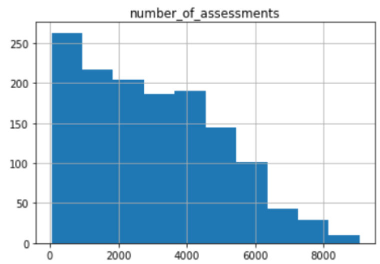
Our analysis depends on the defensive performances of all players on the pitch. We provide ranking tables with the best performing players for goalkeepers, defenders, midfielders and attackers in terms of their defensive performances according to our approaches. Since the most relevant categories for this research depend on evaluating defenders and midfielders, we focus on them. For scouting purposes, we track young players of both categories up to and including twenty-four years old, which we define as talents.

Ranking Table Settings. Firstly, we restrict players to gain a minimum number of assessments to provide reliable results. For instance, a player plays well for five minutes in a whole season and offers a few interception opportunities; this player probably confirms a relatively high score because of a low scaling of its valuation. However, from our point of view, this score is not reliable since it possibly entails a significant amount of coincidence. Therefore, we explore methods to look for convergence of a player's valuation.

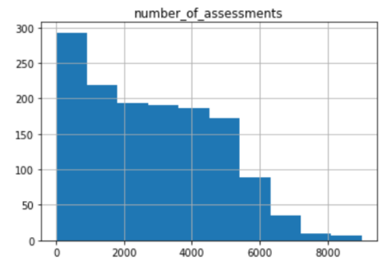
We use histograms related to the number of players' assessments to create multiple boundary restrictions. We search for a boundary which filters out as many noise of not converged player valuations by trial and error in combination with domain knowledge. Finally, for each list, we normalize the values of all players in the feasible region to create a clear ranking.

Since talents possibly get fewer chances to play in matches compared to older and more experienced players, they perhaps face less interception opportunities. To prevent these defensive talents from falling off our watch list, we conduct an exception minimum boundary for these players.

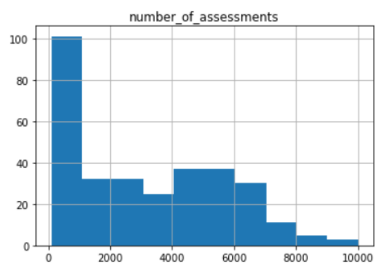
Figure 29 presents distributions for all type of players, whereas we notice the number of assessments on the x-axis and the number of players on the y-axis. We note nearly similar distributions for both defenders and midfielders. For this reason, we select the same boundaries for both categories. By inspection of these histograms, it seems reasonable to provide a first limit which catches many players at a minimum of 3,500 assessments. Next, we choose 4,500 and 5,000 assessments as well, which reveals a relatively more substantial proportion of established names on top. Note that these limits are suitable for goalkeepers and attackers as well. For talents, we provide two boundaries of a minimum number of assessments at 2,500 and 3,000 assessments.



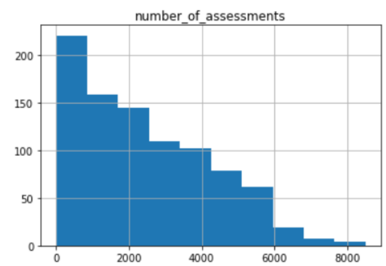
(a) Defenders



(b) Midfielders



(c) Goalkeepers



(d) Attackers

Figure 29: Histograms related to number of assessment. It presents the number of assessments on the x-axis and corresponding number of players on the y-axis. It uses ten different intervals to divide the number of assessments.

Although both the two-step and one-step approach provide in all models relatively good performance compared to the baseline models, they probably deal with advantages and disadvantages due to their independence and correlation when providing the defensive performance of players. For this reason, we design an ensemble method where we average the results of both methods. Hence we average the listings for defenders and midfielders separately over all alternatives. For both categories, we have three records according to the two-step approach, three lists according to the one-step approach and nine lists by averaging both approaches.

Since there is no widely accepted truth in ranking the players according to their defensive performances, we involve market values of the players. We compute the Pearson correlation of the rankings and the market values of all players to get an indication of the most accurate ranking. Within our reach, we suggest this alternative to be the most reliable indication, although we keep in mind that the market values of players depend on more aspects than defensive performance separately.

We provide market values gathered by Transfermarkt.de, which contain the market values of players in the season 2018/2019 established on January 1st (Tranfermarkt.de, 2019). For these correlations, a value closer to minus one provides more reliable results, since an increase of the market value correlates to a higher place on the ranking.

5.2 Results of Defensive Performances

To start, we provide a ranking with our focus strictly on interceptions. We select our ranking according to the most considerable correlation between the ranking and the market values of players.

Table 14 provides the one-step model with a minimum boundary of 4,500 assessments to be most reliable ranking for defenders. In case of the midfielders, we note the ensemble method with a minimum limit of 5,000 assessments according to the two-step model and a minimum boundary of 3,500 assessments according to the one-step model to confirm most reliable ranking for midfielders. Furthermore, the one-step model with a minimum boundary of 3,500 assessments to be most reliable ranking for goalkeepers.

Remarkably, we note the one-step model with a minimum boundary of 5,000 assessments to be most reliable ranking for attackers. However, considering attackers, the one-step approach provides similar performance compared to our baseline models in terms of the evaluation metrics. Therefore, we do not offer a reliable ranking regarding attackers.

		Corr. Defenders		**Corr. Midfielders**		**Corr. Keepers**		**Corr. Attackers**	
Approach	**Boundary**	**MV**	**# I**	**MV**	**# I**	**MV**	**# I**	**MV**	**# I**
Two-Step	3500	0.03	-0.20	0.02	-0.32	-0.07	-0.39	0.01	-0.43
	4500	0.01	-0.27	0.05	-0.32	-0.06	-0.53	-0.03	-0.40
	5000	-0.01	-0.25	0.02	-0.28	-0.02	-0.50	-0.09	-0.50
One-Step	3500	-0.19	-0.44	-0.08	-0.63	**-0.13**	-0.32	-0.07	-0.57
	4500	**-0.24**	-0.50	-0.03	-0.64	-0.12	-0.46	-0.12	-0.57
	5000	-0.23	-0.50	-0.05	-0.65	-0.08	-0.44	**-0.13**	-0.61
Ensemble Method	3500-3500	-0.07	-0.36	-0.04	-0.53	-0.10	-0.36	-0.03	-0.55
	3500-4500	-0.06	-0.64	-0.04	-0.59	-0.12	-0.64	0.02	-0.55
	3500-5000	-0.02	-0.64	-0.08	-0.54	-0.06	-0.66	0.08	-0.57
	4500-3500	-0.09	-0.68	-0.05	-0.63	-0.11	-0.62	0.00	-0.59
	4500-4500	-0.09	-0.43	0.02	-0.54	-0.09	-0.51	-0.07	-0.54
	4500-5000	-0.04	-0.50	-0.10	-0.49	-0.02	-0.58	-0.03	-0.54
	5000-3500	-0.11	-0.70	**-0.12**	-0.62	-0.08	-0.65	-0.01	-0.63
	5000-4500	-0.10	-0.54	-0.09	-0.50	-0.06	-0.58	-0.04	-0.57
	5000-5000	-0.11	-0.42	0.00	-0.54	-0.04	-0.49	-0.10	-0.59

Table 14: Correlations related to ranking lists regarding interceptions as target feature. The bold values confirm the most reliable rankings. Note that boundary incorporates the minimum number of assessments, whereas the ensemble method firstly considers the two-step approach, followed by the one step approach boundary. The correlations of the market values and interceptions are related to the ranking of each corresponding list. Moreover, Corr., MV and # I corresponds to correlations, market value and number of interceptions, respectively.

According to our evaluation metrics regarding defenders, our one-step approach models provide relatively better performance compared to the two-step approach models. Therefore, we expected the one-step approach to provide a more reliable ranking.

Contrarily to defenders, we do not obtain an outstanding approach for midfielders. Since the performances of both approaches are nearly similar for midfielders, it seems to be logical that both approaches confirm almost identical correlations. Probably, both of them has their pros and cons. In this sense, it is more likely to reveal an ensemble method to provide the most reliable ranking.

Furthermore, we compute for each ranking the correlation between the ranking order of players and the number of interceptions executed by the players. We obtain a relatively high negative correlation for most of the models. This seems to be logical since we always provide the highest reward for the player who performs an interception. The negative sign relies on the opposite direction of interceptions and the ranking; more performed interceptions of a players causes a higher rank, ceteris paribus. However, these correlations are still far away from perfect correlation which provides a ranking according to the number of interceptions. We could relate this to the concept of InterceptionsP90 since we utilize minimum boundary restrictions, so most of the players are evaluated on a similar period. However, this could still differ a lot between various players.

5.2.1 Results of Defenders

Table 15 presents the best defenders according to the one-step approach with a minimum restriction of 4,500 assessments. The players are listed in descending order according to their valuation. The valuations of each ranking are normalized by scaling the ranking between zero and one. We note some established names as Rüdiger (Chelsea FC), Piqué Bernabéu (FC Barcelona) and Van Dijk (Liverpool FC). Although, we notice some relatively unknown players as well, such as Høegh (SC Heerenveen), Deschacht (KSC Lokeren Oost-Vlaanderen) and Valdés Díaz (SD Eibar).

Overall, we note that many couples in terms of teammates score nearly similar according to the valuation in the list. For instance, Orban, Konaté and Halstenberg (Rasen Ballsport Leipzig), Høegh and Pierie (SC Heerenveen) and Rüdiger, Moreira Marinho and Alonso Mendoza (Chelsea FC). Since we incorporate the collaboration of a team by providing rewards and penalties, this seems to be logical. When teams perform well, many players acquire rewards and vice versa. Moreover, this uncovers a drawback of our model as well. Apparently, the model does not reveal one specific relative position to execute an interception. Instead, it enables the distribution among several relative positions which is less accurate.

We present opportunities as the sum of all predictions according to Expected Positional Interceptions to execute an interception provided for all player assessments which we rounded to give a clear representation. Furthermore, we note the number of interceptions performed and number of assessments for each player. We obtain relatively small values for opportunities of all players, for about two per cent on average. This means the model assigns on low average predictions to defenders to execute an interception. Since the evaluation metrics provide the best performance for defenders, the model probably predicts that most of the passes will not be intercepted. This suggests another indication we would do well to ensure a more explicit description of game states to improve the efficiency of our model since we might miss some essential features.

First Name	Last Name	Valuation	Opportunities	# Interceptions	# Assessments	Age
Daniel	Høegh	1.00	95	125	4709	29
Willi	Orban	0.99	106	152	5211	27
Antonio	Rüdiger	0.87	124	121	5919	27
Olivier	Deschacht	0.86	100	106	5088	39
Kik	Pierie	0.84	96	142	4816	19
Bart	Straalman	0.84	95	116	4756	23
José Ángel	Valdés Díaz	0.83	86	138	5132	30
Gerard	Piqué Bernabéu	0.81	108	141	4785	33
Virgil	van Dijk	0.81	129	170	5948	28
Jakov	Filipović	0.80	108	111	5423	27
Ibrahima	Konaté	0.80	117	121	6163	20
Toby	Alderweireld	0.80	99	125	4601	31
Marcos	Alonso Mendoza	0.79	96	123	5579	29
Diego Carlos	Santos Silva	0.78	106	125	5425	27
David Luiz	Moreira Marinho	0.78	142	127	6459	33
Aymeric	Laporte	0.77	101	109	4615	25
John Anthony	Brooks	0.77	109	97	4941	27
Obite Evan	N'Dicka	0.77	89	100	4636	20
Ron	Vlaar	0.76	104	111	5147	35
Kurt Happy	Zouma	0.76	102	99	5135	25
Mike	te Wierik	0.75	123	136	6439	27
Maximilian	Rossmann	0.75	96	123	4869	24
Marcel	Halstenberg	0.75	88	109	5865	28
Danny	Vieira da Costa	0.75	73	77	5691	26
Dario	Van den Buijs	0.74	105	106	5203	24

Table 15: Defenders ranking. Best defenders according to the one-step approach with a minimum restriction of 4,500 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Interceptions.

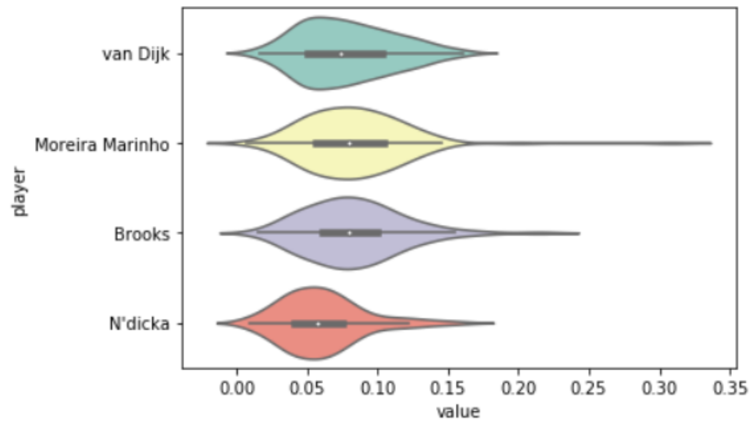
The valuation of players depends on three aspects: (1) interceptions performed, (2) proportion of missed opportunities to execute an interception and (3) support of interceptions executed by teammates. Naturally, we take the difficulty of these alternatives into account in the valuation process. For instance, when a pass is hard to intercept, we highly reward the responsible players for a successful interception. On the other hand, if this opportunity is missed, we penalize them relatively gentle.

In Figure 30, we present the rewards gathered from executed interceptions by the corresponding players and their teammates in sub-figures a and c, respectively. Furthermore, we note the penalties of the corresponding players in sub-figure b. We evaluate players which correspond to relative position 3. Concerning the penalties for these players, we note nearly similar results for all players. However, we observe a difference in the rewarded values.

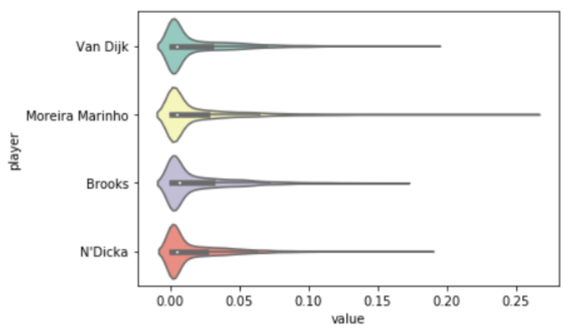
Firstly, we note a slightly better performance of Moreira Marinho and Brooks in terms of values rewarded by interceptions compared to Van Dijk, which we interpret as relatively more valued interceptions of Moreira Marinho and Brooks (VfL Wolfsburg) on average compared to Van Dijk. Although this concerns a minor advantage, we state with more certainty that N'Dicka (Eintracht Frankfurt) executes significantly less valued interceptions on average compared to the other players.

Secondly, we notice a slightly better performance of Moreira Marinho and Brooks when investigating the rewards gained from interceptions by teammates compared to the other players. We interpret this as better support of Moreira Marinho and Brooks in the defensive performances of their teammates.

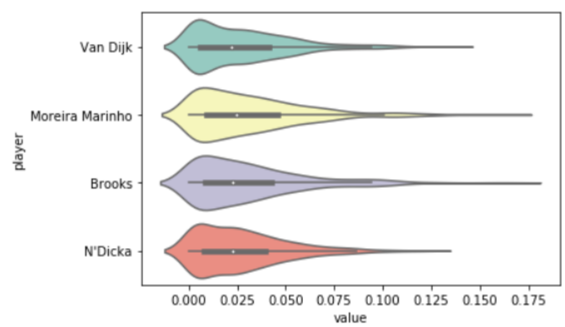
For these reasons, we notice more opportunities in Table 15 for Moreira Marinho and Brooks compared to their interceptions, and we state that their added value in defensive performances are more valuable per interception compared to Van Dijk and N'Dicka. This presents a quite exciting result since Van Dijk nearly won the Ballon d'Or last December for best player in the world. However, as we show in Table 16, Van Dijk executes significantly more interceptions, which leads us to a quality versus quantity trade-off to decide who performs best. Note that the number of missed opportunities to perform an interception and the number of interceptions of teammates where these players support are relatively similar.



(a) Interceptions



(b) Missed Interceptions



(c) Interceptions by Team mates

Figure 30: Probability plots related to interceptions. Evaluated on defenders playing on relative position 3.

Player	**# Interceptions**	**# Missed Interceptions**	**# Interceptions by Team mates**	**# Total Assessments**
van Dijk	170	4941	837	5948
Moreira Marinho	127	5415	917	6459
Brooks	97	4121	723	4941
N'Dicka	100	3868	668	4636

Table 16: Absolute values relates to interceptions. Evaluated on defenders playing on relative position 3.

5.2.2 Results of Midfielders

Now, we look at the defensive abilities of midfielders. Table 17 presents the best midfielders according to the ensemble method with a minimum boundary of 5,000 assessments for the two-step model and a minimum limit of 3,500 assessments for the one-step model. Again we obtain some established names in this list as Wijnaldum (Liverpool FC), Gueye (Everton FC) and Kanté (Chelsea FC). However, we note relatively unknown players as well, such as Midstjø (Alkmaar Zaanstreek), De Sart (KV Kortrijk) and Reis (FC Groningen) although Reis made a transfer to FC Barcelona possibly caused by outstanding performances.

Similarly, as we obtain for defenders, we note relatively many couples of teammates. For example, Midstjø and Koopmeiners (Alkmaar Zaanstreek), Moreira Marinho, Frello Filho and Kanté (Chelsea FC). Moreover, we note couples between highly valued defenders and midfielders as well. For instance, te Wierik, Reis and Doan (FC Groningen), Orban, Konaté and Kampl (Rasen Ballsport Leipzig) and Van Dijk and Wijnaldum (Liverpool FC). When a team provides relatively good defensive performance, many players will be rewarded due to the collaborative aspect.

A crucial difference compared to Table 15 is the valuation. Since we average the valuation of both models, we do not necessarily obtain a player at the maximum valuation of one; we even observe that the best player is far from this maximum score. On the other hand, for both tables, we note nearly similar variance among the rest of the players. These observations indicate that the two-step and one-step approach provide substantially different results.

Another significant difference is that we note relatively more interception opportunities since the model scales the probabilities over all relative positions without taking into account no interception. It provides, on average, more significant probabilities since the multitask model often predicts large probabilities for no interception. The opportunities we obtain for midfielders depend on an equally weighted average of both concerning models.

First Name	Last Name	Valuation	Opportunities	# Interceptions	# Assessments	Age
Fredrik	Midtsjø	0.83	380	109	5659	26
Hannes	Van der Bruggen	0.83	417	131	5876	27
Georginio	Wijnaldum	0.83	322	82	5132	29
Idrissa Gana	Gueye	0.81	372	108	5482	30
Julien	De Sart	0.79	381	103	5836	25
Ludovit	Reis	0.79	326	88	5085	19
Maximilian	Arnold	0.79	322	83	5515	25
Jorge Luiz	Frello Filho	0.78	495	161	6630	28
N'Golo	Kanté	0.77	366	99	6258	29
Killian	Overmeire	0.75	398	115	5754	34
Teun	Koopmeiners	0.74	334	123	5693	22
Kevin	Kampl	0.74	334	87	5178	29
Kai	Havertz	0.73	284	63	5418	20
Pierre-Emile	Højbjerg	0.72	343	104	5258	24
Diego	Demme	0.70	332	82	5085	28
Alexander	Merkel	0.70	385	94	5683	28
Ritsu	Doan	0.70	275	50	5721	21
Filip	Kostić	0.69	322	77	5490	27
Davy	Klaassen	0.69	295	72	5588	27
Leandro	Trossard	0.67	206	63	5192	25
Mohammed	Osman	0.67	332	63	5277	26
Hans	Vanaken	0.67	284	62	5137	27
Nemanja	Maksimović	0.66	400	76	5935	25
Igor	Zubeldía Elorza	0.66	359	91	5336	23
Joris	van Overeem	0.65	382	116	5460	25

Table 17: Midfielders ranking. Best midfielders according to the ensemble method with a minimum boundary of 5,000 assessments for the two-step model and a minimum boundary of 3,500 assessments for the one-step model. Note that opportunities confirm the average of Defensive Positional Expectations and Expected Positional Interceptions according to the sum of all probabilities.

5.2.3 Results of Goalkeepers and Attackers

From our point of view, when scouting a goalkeeper or an attacker, a nearly negligible part of interest depends on their abilities to intercept the opponent's passes. For this reason, these type of players are not relevant to investigate regarding their defensive abilities regarding our approach. However, we still provide their ranking in Appendix G. Note that, in the case of the attackers, we do not provide reliable performance according to the evaluation metrics.

5.2.4 Results of Talents

We apply the same process for detecting the talents of defenders and midfielders according to their defensive performances. Table 18 provides the one-step model with a minimum boundary of 2,500 assessments to be most reliable rankings for defenders. In the case of the midfielders, we note the ensemble method with a minimum limit of 2,500 assessments for both the two-step model and one-step model to confirm the most reliable rankings for midfielders.

		Correlation Defenders		Correlation Midfielders	
Approach	**Boundary**	**Market Values**	**# Interceptions**	**Market Values**	**# Interceptions**
Two-Step	2500	0.08	-0.24	-0.09	-0.30
	3000	0.04	-0.24	-0.07	-0.41
One-Step	2500	**-0.23**	-0.39	-0.07	-0.56
	3000	-0.22	-0.44	-0.07	-0.67
Ensemble	2500-2500	-0.08	-0.32	**-0.10**	-0.46
Method	2500-3000	-0.07	-0.48	-0.02	-0.62
	3000-2500	-0.11	-0.49	-0.00	-0.64
	3000-3000	-0.10	-0.36	-0.08	-0.58

Table 18: Correlations related to ranking lists of talents regarding interceptions as target feature. The bold values confirm the most reliable rankings. Note that boundary incorporates the minimum number of assessments, whereas the ensemble method firstly considers the two-step approach, followed by the one step approach boundary. The correlations of the transfer prices and interceptions are related to the ranking of each corresponding list.

By inspecting the talents in Table 19 and Table 20, we obtain fewer options of couples we make up by teammates, which is a logical result since not many talents often get a chance to play in official matches. Talents in the Dutch Eredivisie get relatively many opportunities to play compared to other leagues. For this reason, we notice a lot of players from this league, such as de Ligt, de Jong, Mazrraoui and van de Beek (AFC Ajax), Chabot, Matusiwa, Reis and Doan (FC Groningen) and Dumfries and Rosario (PSV Eindhoven).

However, when players in the major league such as the Premier League get opportunities to play, and their team performs well, they are highly valued according to our models. For instance, we obtain couples between of all players and talents, such as Alderweireld and Sánchez Mina (Tottenham Hotspur FC), Moreira Marinho, Frello Filho, Kanté and Nunes do Nascimento (Chelsea FC).

First Name	Last Name	Valuation	Opportunities	# Interceptions	# Assessments	Age
Niklas	Süle	1.00	79	87	3326	24
Matthijs	de Ligt	0.93	91	112	3862	20
Noussair	Mazraoui	0.91	57	70	3169	22
Kik	Pierie	0.87	96	142	4816	19
Bart	Straalman	0.87	95	116	4756	23
Ibrahima	Konaté	0.83	117	121	6163	20
Alessandro	Ciranni	0.83	50	68	2908	23
Jeremiah	St. Juste	0.81	53	65	2792	23
Enock	Kwateng	0.81	51	93	3248	23
Hervé	Matthys	0.80	67	66	3638	24
Obite Evan	N'Dicka	0.79	89	100	4636	20
Jake	Clarke-Salter	0.79	81	75	3879	22
Denzel	Dumfries	0.79	78	89	4479	24
Maximilian	Rossmann	0.78	96	123	4869	24
Dario	Van den Buijs	0.77	105	106	5203	24
Milos	Veljković	0.77	67	67	3375	24
José Ángel	Esmoris Tasende	0.76	83	95	4479	23
Nordi	Mukiele Mulere	0.75	48	62	2786	22
Dan-Axel	Zagadou	0.74	65	67	3134	20
Timo	Baumgartl	0.73	63	63	3146	24
Kingsley	Ehizibue	0.73	66	87	4319	24
Joseph	Aidoo	0.71	99	98	4276	24
Julian	Chabot	0.71	93	79	4632	22
Jonathan	Tah	0.71	117	135	5458	24
Davinson	Sánchez Mina	0.70	55	52	2501	23

Table 19: Defenders Talents. Best talents of defenders according to the one-step approach with a minimum restriction of 2,500 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Interceptions.

First Name	Last Name	Valuation	Opportunities	# Interceptions	# Assessments	Age
Azor	Matusiwa	1.00	243	114	3413	22
Selim	Amallah	0.89	136	37	2636	23
Frenkie	de Jong	0.86	240	93	3395	22
Noussair	Mazraoui	0.84	180	70	3169	22
Pablo Paulino	Rosario	0.81	278	97	4117	23
Manuel	Benson Hedilazio	0.77	123	37	2756	23
Konrad	Laimer	0.75	257	74	4518	22
Benjamin	Van Durmen	0.72	246	57	3736	23
Ludovit	Reis	0.71	326	88	5085	19
Teun	Koopmeiners	0.69	334	123	5693	22
Robert Kenedy	Nunes do Nascimento	0.69	178	47	3607	24
Gabriel	Boschilia	0.68	115	21	2546	24
Frank	Boya	0.68	282	97	4300	23
Florian	Grillitsch	0.68	298	92	4459	24
Martin	Ødegaard	0.67	233	73	4675	21
Leon	Bailey Butler	0.66	155	32	3459	22
Kai	Havertz	0.66	284	63	5418	20
Pierre-Emile	Højbjerg	0.65	343	104	5258	24
Donny	van de Beek	0.65	161	32	3111	23
Arne	Maier	0.63	266	66	3970	21
Mikel	Merino Zazón	0.62	184	51	3185	23
Gustavo	Hamer	0.62	185	52	2810	22
Julian	Brandt	0.62	242	56	4993	23
Lucas	Torreira Di Pascua	0.62	195	61	3032	24
Ritsu	Doan	0.61	275	50	5721	21

Table 20: Midfielders Talents. Best midfielders according to the ensemble method with a minimum boundary of 2,500 assessments for the two-step model and a minimum boundary of 2,500 assessments for the one-step model. Note that opportunities confirm the average of Defensive Positional Expectations and Expected Positional Interceptions according to the sum of all probabilities.

5.2.5 Results of Defensive Actions

Next, we focus on the performances of the players according to all defensive actions. The most relevant information we gather from comparing this different target to interceptions provides players who often appear in duels but probably in many cases do not achieve to execute an interception. Many reasons can be applied to explain these occurrences, such as experience or timing to enter into a duel.

Appendix H presents the most reliable ranking list for midfielders according to the ensemble method with a minimum boundary of 5,000 assessments for the two-step model and a minimum limit of 3,500 assessments for the one-step model. We offer the list in Table 21.

Since we set similar boundary limits as we do for Table 17, we can compare players among these lists. A big name which misses in the list for best-performing midfielders according to all defensive actions is Kanté. Moreover, we note a significant performance downgrade of Wijnaldum. These differences could provide an indication for their experience how to handle a duel efficiently.

This interpretation could mainly be valuable when inspecting the talents list (Table 22). We could notice young players who suffer from these stratagems at the moment but could improve these abilities along the way. For instance, this could be the case for Clarke-Salter (SBV Vitesse), Posch (TSG 1899 Hoffenheim) and Upamecano (Rasen Ballsport Leipzig).

First Name	Last Name	Valuation	Opportunities	# Defensive Actions	# Assessments	Age
Ludovit	Reis	0.87	374	218	5085	19
Idrissa Gana	Gueye	0.81	424	220	5482	30
Diego	Demme	0.79	373	152	5085	28
Jorge Luiz	Frello Filho	0.78	444	266	6630	28
Rani	Khedira	0.77	440	276	6464	26
Teun	Koopmeiners	0.77	383	213	5693	22
Hannes	Van der Bruggen	0.77	456	218	5876	27
Fredrik	Midtsjø	0.76	440	198	5659	26
Julien	De Sart	0.74	436	199	5836	25
Igor	Zubeldía Elorza	0.73	391	179	5336	23
Killian	Overmeire	0.73	428	206	5754	34
Hans	Vanaken	0.72	342	120	5137	27
Maximilian	Arnold	0.71	380	160	5515	25
Kevin	Kampl	0.71	373	136	5178	29
Otávio Henrique	Passos Santos	0.70	404	192	5358	25
Ritsu	Doan	0.69	329	101	5721	21
Alexander	Merkel	0.69	412	176	5683	28
Leandro	Trossard	0.69	226	111	5192	25
Pol	Llonch Puyalto	0.69	411	224	5759	27
Nemanja	Maksimović	0.68	445	138	5935	25
Mohammed	Osman	0.68	384	121	5277	26
Youssef	El Jebli	0.66	349	109	5686	27
Georginio	Wijnaldum	0.66	381	120	5132	29
Kai	Havertz	0.66	333	93	5418	20
Daniel	Baier	0.65	515	236	6974	35

Table 21: Midfielders. Best midfielders according to the ensemble method with a minimum boundary of 5,000 assessments for the two-step model and a minimum boundary of 3,500 assessments for the one-step model. Note that opportunities confirm the average of Positions and Expected Positional Defensive Actions according to the sum of all probabilities.

First Name	Last Name	Valuation	Opportunities	# Defensive Actions	# Assessments	Age
Bart	Straalman	1.00	179	205	4756	23
Kik	Pierie	1.00	184	267	4816	19
Niklas	Süle	0.97	145	156	3326	24
Jake	Clarke-Salter	0.92	147	138	3879	22
Ibrahima	Konaté	0.91	213	238	6163	20
Milos	Veljković	0.88	129	155	3375	24
Matthijs	de Ligt	0.87	171	175	3862	20
Stefan	Posch	0.87	109	148	2774	22
Dario	Van den Buijs	0.86	187	192	5203	24
Maximilian	Rossmann	0.86	173	191	4869	24
Dayot	Upamecano	0.85	97	131	2610	21
Julian	Chabot	0.84	167	161	4632	22
Davinson	Sánchez Mina	0.84	101	99	2501	23
Joseph	Aidoo	0.82	184	202	4276	24
Dan-Axel	Zagadou	0.82	116	117	3134	20
Frank	Boya	0.80	170	209	4300	23
Enock	Kwateng	0.79	106	160	3248	23
Zinho	Vanheusden	0.79	135	161	3485	20
Jeremiah	St. Juste	0.78	104	131	2792	23
Clément	Lenglet	0.76	109	108	2790	24
Noussair	Mazraoui	0.75	114	134	3169	22
Obite Evan	N"Dicka	0.75	172	195	4636	20
Joachim	Andersen	0.74	209	193	5614	23
Jonathan	Tah	0.74	218	220	5458	24
Nordi	Mukiele Mulere	0.73	90	107	2786	22

Table 22: Defenders Talents. Best talents of defenders according to the one-step approach with a minimum restriction of 2,500 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Defensive Actions.

5.2.6 Missing Players

Obviously, there are a lot of players missing in our most reliable rankings. Since we inspected other rankings, which seems to be less reliable, as well, we find it worthwhile to report the most common players who appear on those rankings. Table 23 presents these players.

Azpilicueta, Andersen and Koundé appear on many rankings but just drop out the ranking because of their valuation corresponding to the most reliable rankings. Hummels, Blind and Issa Diop drop outside the ranking because they just lack the minimum number of assessments, but provide promising results.

Since we provide an ensemble method for midfielders, some players lack the number of assessment of one of those rankings. Fernandinho and Aránguiz lack the number of assessments for the ranking with a minimum boundary of 5,000 assessments; however, both players provide good results when we incorporate a minimum limit of 3,500 assessments. Furthermore, Winks confirms too few assessments for both boundaries. Lastly, Witsel, Ndidi and Guendouzi appear on many rankings but just drop out of the table because of their valuation corresponding to the most reliable ranking.

Genre	Defenders	Midfielders
All players	Azpilicueta* (Chelsea FC)	Fernandinho** (Manchester City FC)
	Hummels** (FC Bayern München)	Witsel* (BV Borussia 09 Dortmund)
	Blind** (AFC Ajax)	Aránguiz** (TSV Bayer 04 Leverkusen)
Talents	Andersen* (UC Sampdoria)	Ndidi* (Leicester City FC)
	Koundé* (FC Girondins de Bordeaux)	Winks** (Tottenham Hotspur FC)
	Issa Diop** (West Ham United FC)	Guendouzi* (Arsenal FC)

Table 23: Missing players. Players who often appear on various lists, but not our most reliable lists. Note that * corresponds to players who drop just outside the most reliable lists because of their valuation, ** corresponds to players who do not confirm the minimum boundary.

Next, some midfielders from SD Eibar provide extremely promising valuations; Diop, Jordán Moreno and Orellana Valenzuela. When incorporating the minimum boundary of 3,500 assessments, they appear as top 10 players for all rankings, but since we apply the average approach, they drop outside the most reliable ranking. This founding would suggest that this team plays according to a tactic which pays close attention to the defensive part of the game. Although, we do not obtain corresponding remarkable results in their final league standing at twelfth place with an about average number of goals conceded.

5.3 Discussion Results

In this section, we provide results of our defensive ability metric. Since both the two-step and one-step approach entail their pros and cons, we offer an ensemble method of them as well. Notably, in the case of the midfielders, this process pays off.

After gaining various rankings, we get an indication of the most reliable ranking by inspecting the correlation to the market values of players. However, a perfect negative correlation between the ranking list and market values is probably not truthful since in general players are evaluated on more than just defensive performances.

Moreover, we provide a new tool to provide insight into the added value of performed interceptions and the support players offer to their teammates. Furthermore, it gives insight into the missed opportunities of players to execute an interception.

Since our models are not able to point out specific players to execute an interception, we observe that players of a team are highly correlated according to their valuations. This founding infers we need to describe each game state more explicitly.

Lastly, we provide an opportunity to reveal defensive talents by inspecting the different results according to all defensive actions and interceptions.

6 Conclusion and Discussion

In this section, we clarify how to provide a reliable metric to quantify players' defensive abilities. To start, we answer our research questions which explains our design. Next, we elaborate on our shortcomings and declare our suggestions to improve this metric.

6.1 Research Questions

The main research question concerns how to evaluate the defensive performances of football players. We allow to generalize our approaches by describing the most relevant information regarding each specific pass. As a result, we enable to focus on three crucial components which reveals the defensive performance of football players.

Firstly, we investigate the difficulty to execute an interception which we interpret as the quality of an interception. From our point of view, when an interception is hard to perform, a player needs strong defensive abilities to realize the interception. Therefore, the more difficult an interception seems to be, the more we valuate a successful interception.

Secondly, we inspect the support of players within a team according to their defensive performances. We grant players pro-rata for the interception of a player of a team. In this sense, we take the supportive contribution of players into account.

Lastly, we provide insight into the missed opportunities of players to execute an interception. This component enables us to penalize players whenever we expect them to perform an interception, but they fail. For instance, if we obtain a relatively easy opportunity to perform an interception, where we highly expect players to execute an interception, but they fail, we penalize them relatively hard and vice versa.

Regarding our first research sub-question, we clarify how to estimate the likeliness of a pass being intercepted by a given player while accounting for the defensive support of the player's teammates. We investigate the possibilities if a specific pass is intercepted and who will intercept the pass as two independent stages. Therefore, we present a classification model to predict the difficulty of an interception and a classification model to predict which player to perform an interception. Since we suspect these stages are related to each other, we investigate a classification model which provides these two stages as one correlated step as well. Since this latter approach offers the most reliable results regarding defenders' defensive performance, it pays, in particular, off for this type of players. Concerning midfielders, we notice that both models provide their pros and cons. Moreover, we provide the most reliable results according to a combination of both approaches.

The second research sub-question relates to what factors contribute to the likeliness of a pass being intercepted. In all our models, the pass specific factors related to the location and direction of a pass seem to be fairly most substantial to the predictions. To be clear, these factors entail the angle of a pass, the Euclidean distance of a pass to the centre of the opponent's goal and if the pass is sent forward, sideways or backward. Since we assigned a position to every player on the pitch, we gain an insight into the possession of both teams. We note that the position of the player who sends the pass appears to be crucial. Lastly, in predicting which player to intercept a pass, the side direction of the pass is substantial as well.

Our last research sub-question concerns what statistical method is best suited to predict the likeliness of a pass being intercepted. To provide our objectives regarding the interception of a pass, we implemented several statistical methods. Since a few factors are most significant in predicting a pass to be intercepted or not, LightGBM delivers the best and most reliable performance concerning this classification model. Since XGBoost provides many variances by evaluating different data sets for predicting which player to intercept a pass, we perceive CatBoost to provide the most reliable performance. However, for the prediction if a pass will be intercepted and which player to intercept the pass in once, we notice XGBoost to provide the best and most reliable performance.

6.2 Further Research

In this research, we provide a reliable metric to analyze the defensive abilities of defenders and midfielders. However, we run into some issues which restrict the performance level. In this section, we propose several potential avenues for future research.

Alignment Challenge. When inspecting the performance of each player on the pitch separately, we classified each player to a specific position on the pitch as evident as possible regarding its occurrence at spatial areas and defensive tasks. The process of assigning players efficiently is widely defined as the alignment challenge.

During this research, we classified each possible position into a relative position which corresponds to a specific formation. Since there are already many possible combinations of relative positions and formations, the classification model probably merges combinations with nearly similar output. Since these merged relative positions can vary a lot in their functions and appearances on the pitch, it is hard for the model to act appropriate continuously. Therefore, we notice a performance downgrade of the classification model.

In our approach, we consider each relative position in conjunction with a formation to entail similar behaviour, which is possibly not true. For this reason, we should implement the player role of each relative position as well, which makes it even less organized for the classification model. As a result, this adjustment increases the obtained issues of the latter problem even more.

The classification models depend its expectations on the occurred events on the ball of a relative position. When two players swap over positions, this will not be recognized by the classification model. Subsequently, we evaluate these relative positions in a wrong way.

Perhaps, we should propose the players differently to the classification model. A first suggestion would be to solely assign a player role to each player on the pitch. However, this probably entails too few training examples to investigate the behaviour of players, and we maintain an issue with players who swap over positions.

To avoid the problem of players who swap over positions, we could divide most related defensive players in groups and assign defensive tasks to each group. An issue which occurs is that we can not assess players separately in this manner, and still we are not able to guarantee players between groups to swap over positions.

Anomaly Detection. To improve the performance of our models, we should describe the relevant information concerning each pass more explicitly. We only provide factors which correspond to the exact time of a pass. It could entail relevant information if we implement a time span in front of the pass to get a better insight into each specific pass. These factors would be pass specific as well, where we could think of the level of overcrowding on the pitch or variation of play of a team in possession. We tested some factors related to the occurrence of a duel, the length of previously obtained passes and the number of passes on the same side of the pitch. However, we did not provide a performance upgrade, so these adjustments only made the model more complicated.

Extend Target Information. Our focus relies on interceptions and defensive actions. Perhaps we should investigate other specific defensive actions in more depth, such as tackles and clearances. This would indicate the specific defensive abilities of players as well.

Moreover, it could be interesting to inspect other types of defensive actions which are currently out of our scope, such as players who are purposely sidelined or fouls to prevent a counter-attack.

A final exciting investigation could be related to second balls, which are, by nature, unpredictable. However, players could prepare to be ready for different eventualities. These types of interceptions are precious in football.

Information-Richer Data. We only observe players' locations on the pitch when they provide an action on the ball. In some cases, this entails an obstacle, for instance, in predicting which player to intercept a pass. If we gain more information, we probably provide more accurate models. In this sense, we consider data which offers snapshots of the pitch to get an overview of all players on the pitch. As a result, it could be applicable for match analysis. However, This type of data is only available in the major leagues, which makes it unusable for player recruitment.

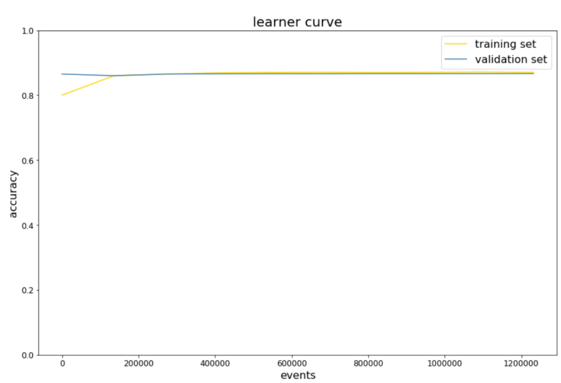
A Binary Classification Validation Set

Method	Evaluation Metric	
	AUC-PR	LogLoss
Random Forest	0.950-0.326	0.343
LightGBM	0.952-0.329	0.341
XGBoost	**0.953-0.340**	**0.339**
GA^{2}M	0.941-0.269	0.359

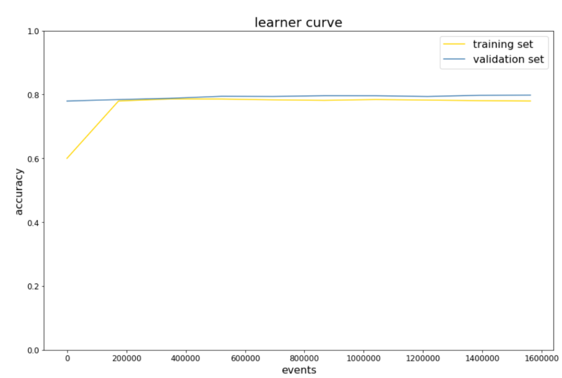
Table 24: Evaluation metrics of binary classification models and all baseline models, including only interceptions as target feature (EI). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR contains two values, where this metric evaluates: precision recall curve of pass reaches a teammate-precision recall curve of an interception.

Method	Evaluation Metric	
	AUC-PR	LogLoss
Random Forest	0.907-0.546	0.471
LightGBM	**0.935-0.520**	**0.421**
XGBoost	**0.915-0.564**	**0.461**
GA^{2}M	0.894-0.489	0.491

Table 25: Evaluation metrics of binary classification models and all baseline models, including all defensive actions as target feature (EDA). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR contains two values, where this metric evaluates: precision recall curve of pass reaches a teammate-precision recall curve of an interception.



(a) LightGBM Expected Interceptions



(b) LightGBM Expected Defensive Actions

Figure 31: Learning curves of binary classifiers. If the curves of the training and validation set converge, the model provides similar performance on both sets.

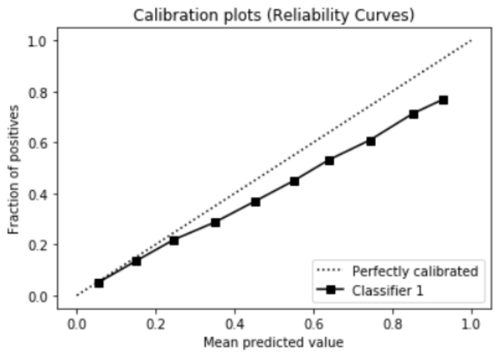
B Expected Defensive Actions

Type of hyperparameter	**Random Forest**	**XGBoost**	**LightGBM**
Number of decision trees	800	100*	100*
Maximum depth of tree	10	6	-1*
Learning rate	N/A	0.1	0.2
Minimum number of events per leaf	200	1*	30
Minimum number of samples per split	10	1*	1*
Subsample of features per tree	ln(2)	0.8	1*
Fraction of randomly events per tree	0*	0.8	0*
Minimum child weight	0*	1	1

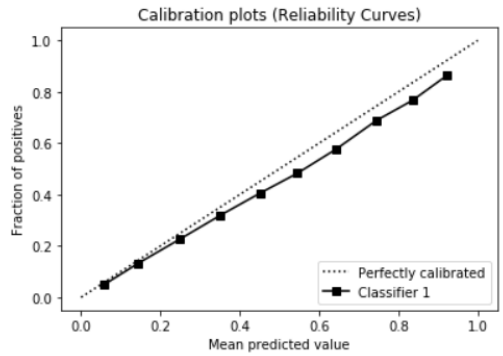
Table 26: Hyperparameters of binary classifiers, including all defensive actions as target feature (EDA). The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this method, and * corresponds to the method's default parameter value.

	Evaluation Metric	
Method	AUC-PR	LogLoss
Random Forest	0.910-0.530	0.466
LightGBM	**0.912-0.536**	**0.463**
XGBoost	**0.913-0.534**	0.464
GA^{2}M	0.848-0.397	0.601
Baseline 1	0.801-0.315	0.540
Baseline 2	0.838-0.355	0.525
Baseline 3	0.774-0.286	0.559

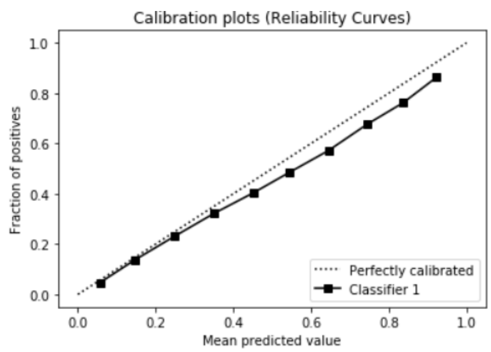
Table 27: Evaluation metrics of binary classification models and all baseline models, including all defensive actions as target feature (EDA). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR contains two values, where this metric evaluates: precision recall curve of pass reaches a teammate-precision recall curve of all defensive actions.



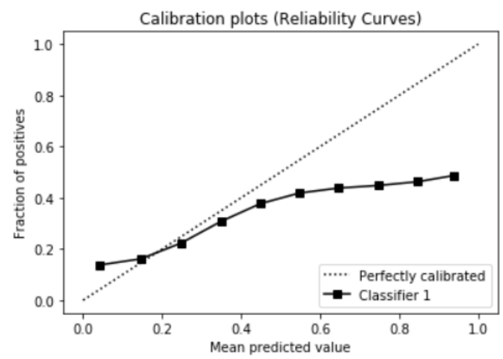
(a) Random Forest



(b) LightGBM



(c) XGBoost



(d) GA^{2}M

Figure 32: Calibration curves, including defensive actions as target (EDA). The dotted forty-five degrees line corresponds to a perfectly calibrated model. The closer the model curve comes to this line, the better it calibrates to the data. Note that these curves are evaluated on the test set.

C Positioning

Type of hyperparameter	**CatBoost**	**XGBoost**
Number of decision trees	100	100*
Maximum depth of decision tree	6	6
Learning rate	0.1*	0.1
Objective	Multiclass	Multi:softprob
Amount of randomness for scoring splits	0.5	N/A
Coefficient L2 regularization cost function	4	N/A
Subsample of features per tree	0.8	0.8
Fraction of randomly events per tree	0*	0.8
Minimum child weight	1*	1

Table 28: Hyperparameters of multiclass classification models, including all defensive actions as target feature (Positioning). The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this method, and * corresponds to the method's default parameter value.

Relative position	AUC-PR				
	CatBoost	XGBoost	Baseline 1	Baseline 2	Baseline 3
Position 0	0.12	0.14	0.04	0.06	0.06
Position 1	0.33	0.37	0.13	0.13	0.15
Position 2	0.30	0.31	0.14	0.14	0.16
Position 3	0.31	0.32	0.14	0.14	0.16
Position 4	0.33	0.35	0.12	0.13	0.14
Position 5	0.18	0.18	0.10	0.10	0.11
Position 6	0.16	0.17	0.11	0.11	0.12
Position 7	0.14	0.17	0.09	0.09	0.11
Position 8	0.16	0.17	0.07	0.07	0.10
Position 9	0.14	0.15	0.05	0.05	0.09
Position 10	0.14	0.16	0.03	0.04	0.10

Table 29: AUC-PR of multiclass classification models and all baseline models, including all defensive actions as target feature (Positioning), for each relative position separately. The performances are evaluated according to the test set.

Method	Evaluation Metric	
	AUC-PR	LogLoss
CatBoost	0.250	2.005
XGBoost	**0.270**	**1.963**
Baseline 1	0.130	2.305
Baseline 2	0.130	2.296
Baseline 3	0.140	2.243

Table 30: Evaluation metrics of multiclass classification models and all baseline models, including all defensive actions as target feature (Positioning). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR are weighted averages over all relative postitions.

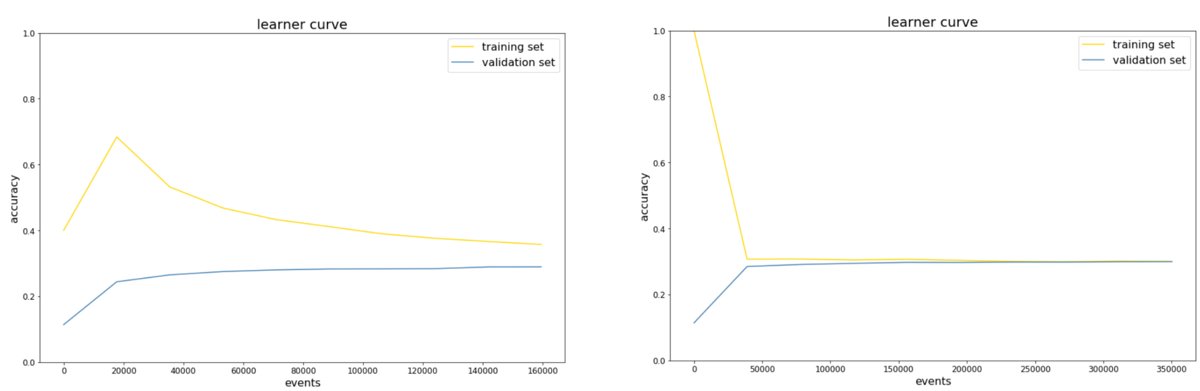
D Multiclass Classification Validation Set

Relative position	**AUC-PR** CatBoost	XGBoost
Position 0	0.13	0.21
Position 1	0.37	0.40
Position 2	0.32	0.35
Position 3	0.33	0.37
Position 4	0.36	0.39
Position 5	0.19	0.23
Position 6	0.18	0.23
Position 7	0.19	0.24
Position 8	0.20	0.23
Position 9	0.19	0.23
Position 10	0.16	0.22

Table 31: AUC-PR of multiclass classification models, including all defensive actions as target feature (Positioning), for each relative position separately. The performances are evaluated according to the validation set.

Method	**Evaluation Metric** AUC-PR	LogLoss
CatBoost	0.280	1.929
XGBoost	**0.310**	**1.865**

Table 32: Evaluation metrics of multiclass classification, including all defensive actions as target feature (Positioning). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR are weighted averages over all relative positions.



(a) XGBoost (b) CatBoost

Figure 33: Learning curves, including all defensive actions as target feature (Positioning). If the curves of the training and validation set converge, the model provides similar performance on both sets.

E Multitask Classification Validation Set

Relative position	AUC-PR	
	CatBoost	XGBoost
Position 0	0.07	0.11
Position 1	0.07	0.09
Position 2	0.10	0.11
Position 3	0.09	0.11
Position 4	0.07	0.09
Position 5	0.03	0.05
Position 6	0.04	0.05
Position 7	0.03	0.04
Position 8	0.02	0.04
Position 9	0.01	0.03
Position 10	0.01	0.03
No interception	0.95	0.96

Table 33: AUC-PR of multitask classification models, including only interceptions as target feature (Expected Positional Interception), for each relative position separately. The bold values confirm the best performances. The performances are evaluated according to the validation set.

Method	Evaluation Metric	
	AUC-PR	LogLoss
CatBoost	0.880	0.613
XGBoost	**0.890**	**0.585**

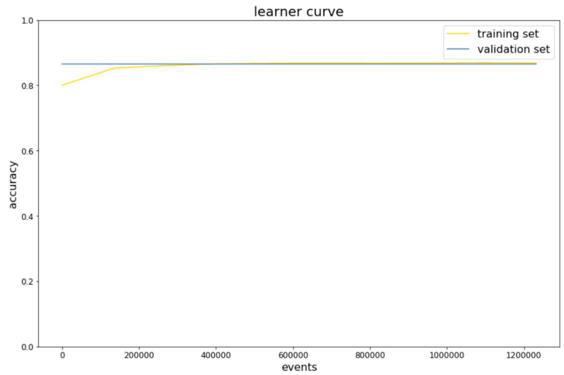
Table 34: Evaluation metrics of multitask classification, including only interceptions as target feature (Expected Positional Interception). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR are weighted averages over all relative positions.

Relative position	AUC-PR	
	CatBoost	XGBoost
Position 0	0.07	0.12
Position 1	0.15	0.17
Position 2	0.19	0.20
Position 3	0.20	0.22
Position 4	0.15	0.17
Position 5	0.06	0.08
Position 6	0.07	0.09
Position 7	0.05	0.07
Position 8	0.03	0.05
Position 9	0.03	0.05
Position 10	0.02	0.04
No defensive action	0.91	0.91

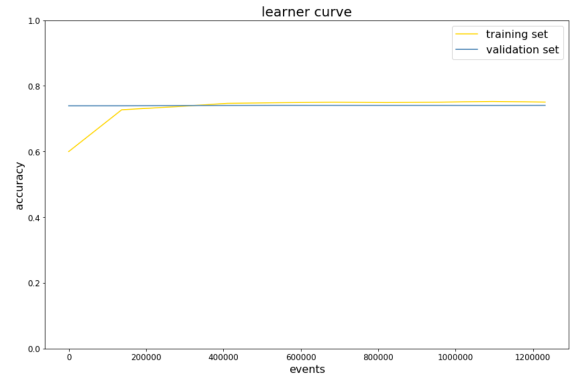
Table 35: AUC-PR of multitask classification models, including all defensive actions as target feature (Expected Positional Defensive Action), for each relative position separately. The performances are evaluated according to the validation set.

Method	Evaluation Metric	
	AUC-PR	LogLoss
CatBoost	0.780	0.982
XGBoost	**0.790**	**0.957**

Table 36: Evaluation metrics of multitask classification, including all defensive actions as target feature (Expected Positional Defensive Action). The bold values confirm the best performances. The performances are evaluated according to the validation set. Note that AUC-PR are weighted averages over all relative positions.



(a) XGBoost Expected Positional Interceptions



(b) XGBoost Expected Positional Defensive Actions

Figure 34: Learning curves of multitask classifiers. If the curves of the training and validation set converge, the model provides similar performance on both sets.

F Expected Positional Defensive Actions

Type of hyperparameter	**CatBoost**	**XGBoost**
Number of decision trees	100	100*
Maximum depth of decision tree	6	6
Learning rate	0.15*	0.1
Objective	Multiclass	Multi:softprob
Amount of randomness for scoring splits	0.5	N/A
Coefficient L2 regularization cost function	4	N/A
Subsample of features per tree	0.8	0.8
Fraction of randomly events per tree	0*	0.8
Minimum child weight	1*	3

Table 37: Hyperparameters multitask classifiers, including all defensive actions as target feature. The models are optimized according to the validation set. Note that N/A means it is not available to tune this hyperparameter for this method, and * corresponds to the method's default parameter value.

Relative position	**AUC-PR**				
	CatBoost	XGBoost	Baseline 1	Baseline 2	Baseline 3
Position 0	0.06	0.09	0.01	0.02	0.02
Position 1	0.13	0.15	0.03	0.04	0.05
Position 2	0.16	0.18	0.04	0.04	0.06
Position 3	0.17	0.18	0.04	0.04	0.06
Position 4	0.12	0.14	0.03	0.04	0.05
Position 5	0.05	0.06	0.02	0.03	0.03
Position 6	0.06	0.07	0.03	0.03	0.04
Position 7	0.04	0.05	0.02	0.02	0.03
Position 8	0.03	0.04	0.02	0.02	0.02
Position 9	0.02	0.02	0.01	0.01	0.01
Position 10	0.02	0.02	0.01	0.01	0.01
No defensive action	0.91	0.91	0.75	0.78	0.84

Table 38: AUC-PR of multitask classification models and all baseline models, including all defensive actions as target feature (EPDA), for each relative position separately. The performances are evaluated according to the test set.

Method	**Evaluation Metric**	
	AUC-PR	LogLoss
CatBoost	**0.780**	0.984
XGBoost	**0.780**	**0.967**
Baseline 1	0.610	1.151
Baseline 2	0.650	1.140
Baseline 3	0.700	1.094

Table 39: Evaluation metrics of multitask classification models and all baseline models, including all defensive actions as target feature (EPDA). The bold values confirm the best performances. The performances are evaluated according to the test set. Note that AUC-PR are weighted averages over all relative postitions.

G Results Interceptions

First Name	Last Name	Valuation	Opportunity	# Interceptions	# Assessments	Age
Péter	Gulácsi	1.00	74	90	6943	29
Marco	Bizot	0.99	66	94	5973	29
Iago	Herrerín Buisán	0.96	54	58	4628	32
Warner	Hahn	0.96	65	63	5523	27
Kepa	Arrizabalaga Revuelta	0.96	66	73	6253	25
Kevin	Trapp	0.91	59	72	5506	29
Stéphane	Ruffier	0.88	53	69	4938	33
Marc-André	ter Stegen	0.87	52	49	4612	28
Alexei	Coşelev	0.85	74	76	5853	26
Sergio	Padt	0.85	73	103	6594	29
Jordan	Pickford	0.82	66	68	6555	26
Lukáš	Hrádecký	0.82	58	56	5130	30
Jean	Butez	0.81	57	73	4542	24
Mike	Maignan	0.80	56	55	5007	24
Alisson Ramsés	Becker	0.78	64	73	5948	27
Ederson	Santana de Moraes	0.75	52	53	5057	26
Emil	Audero Mulyadi	0.74	67	65	6056	23
Yann	Sommer	0.67	74	78	7466	31
Hugo	Lloris	0.67	51	52	4745	33
Nicolas	Penneteau	0.67	70	81	5628	39
Janis	Blaswich	0.65	67	79	6116	28
Roman	Bürki	0.62	65	67	6252	29
Timon	Wellenreuther	0.62	57	77	5114	24
Jiří	Pavlenka	0.60	59	67	5945	28
Ron-Robert	Zieler	0.58	77	79	6813	31

Table 40: Goalkeepers ranking. Best goalkeepers according to the one-step approach with a minimum restriction of 3,500 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Interceptions.

First Name	Last Name	Valuation	Opportunity	# Interceptions	# Assessments	Age
Jordan Rolly	Botaka	1.00	70	72	5357	26
Teddy	Chevalier	0.79	44	42	5565	32
Oussama	Idrissi	0.76	39	48	5323	24
Wout	Weghorst	0.74	28	34	5464	27
Kevin	Volland	0.74	30	29	5240	27
Sadio	Mané	0.69	36	42	5558	28
Kristoffer	Peterson	0.68	39	46	5447	25
Brandley	Kuwas	0.68	42	29	5798	27
Jean-Paul	Boëtius	0.67	48	62	5896	26
Renaud	Ripart	0.65	47	50	5410	27
Mohamed	Salah Ghaly	0.64	38	33	5812	27
Nicolas	Pépé	0.64	45	28	5007	24
Sheraldo	Becker	0.61	49	27	5780	25
Iñaki	Williams Arthuer	0.60	29	27	5461	25
Richarlison	de Andrade	0.60	36	29	5175	22
Gyrano	Kerk	0.59	53	46	6685	24
Eden	Hazard	0.58	34	33	5720	29
Max	Kruse	0.57	31	33	5642	32
Thorgan	Hazard	0.56	55	54	7260	27
Ismaïla	Sarr	0.56	43	32	5423	22
Sam	Lammers	0.56	23	22	5180	23
Timo	Werner	0.54	35	27	6063	24
Mikhail	Rosheuvel	0.53	49	30	5635	29
Yussuf	Yurary Poulsen	0.52	34	17	6016	25
Felipe	Anderson Pereira Gomes	0.51	55	72	8035	27

Table 41: Attackers ranking list. Best attackers according to the one-step approach with a minimum restriction of 5,000 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Interceptions.

H Results Defensive Actions

		Correlation Defenders		Correlation Midfielders	
Approach	**Boundary**	**Transfer Prices**	**Defensive Actions**	**Transfer Prices**	**Defensive Actions**
Two-Step	3500	0.10	-0.22	0.06	-0.22
	4500	0.04	-0.27	0.08	-0.24
	5000	0.01	-0.26	0.08	-0.22
One-step	3500	-0.19	-0.43	-0.04	-0.59
	4500	-0.23	-0.41	0.02	-0.62
	5000	**-0.24**	-0.44	0.00	-0.61
Average	3500-3500	-0.07	-0.37	0.00	-0.47
	3500-4500	-0.04	-0.63	-0.01	-0.57
	3500-5000	-0.00	-0.65	-0.05	-0.52
	4500-3500	-0.10	-0.66	-0.03	-0.62
	4500-4500	-0.10	-0.38	0.06	-0.50
	4500-5000	-0.06	-0.51	-0.04	-0.48
	5000-3500	-0.12	-0.69	**-0.10**	-0.63
	5000-4500	-0.12	-0.52	-0.06	-0.51
	5000-5000	-0.13	-0.41	0.04	-0.48

Table 42: Correlations related to ranking lists regarding all defensive actions as target feature. Note that boundary incorporates the minimum number of assessments, whereas the average approach firstly considers the two-step approach, followed by the one step approach boundary. The correlations of the market values and defensive actions are related to the ranking of each corresponding ranking. The bold values confirm the most reliable ranking.

Approach	Boundary	Correlation Defenders: Transfer Prices	Correlation Defenders: Defensive Actions	Correlation Midfielders: Transfer Prices	Correlation Midfielders: Defensive Actions
Two-Step	2500	0.08	-0.14	**-0.07**	-0.19
	3000	0.09	-0.18	0.02	-0.30
One-step	2500	**-0.26**	-0.35	-0.01	-0.53
	3000	-0.24	-0.41	-0.02	-0.65
Average	2500-2500	-0.12	-0.26	**-0.07**	-0.41
	2500-3000	-0.10	-0.45	0.01	-0.57
	3000-2500	-0.14	-0.44	0.04	-0.60
	3000-3000	-0.10	-0.33	-0.02	-0.53

Table 43: Correlations related to ranking lists of talents regarding all defensive actions as target feature. Note that boundary incorporates the minimum number of assessments, whereas the average approach firstly considers the two-step approach, followed by the one step approach boundary. The correlations of the transfer prices and defensive actions are related to the ranking of each corresponding list

First Name	Last Name	Valuation	Opportunities	# Defensive Actions	# Assessments	Age
Willi	Orban	1.00	197	285	5211	27
Diego Carlos	Santos Silva	0.83	191	218	5425	27
Virgil	van Dijk	0.83	238	292	5948	28
Ibrahima	Konaté	0.81	213	238	6163	20
Mike	te Wierik	0.79	228	246	6439	27
Antonio	Rüdiger	0.77	229	215	5919	27
Michael	Keane	0.77	205	243	5727	27
Dario	Van den Buijs	0.76	187	192	5203	24
Samir	Memišević	0.75	223	275	6594	26
Ron	Vlaar	0.75	190	170	5147	35
David Luiz	Moreira Marinho	0.74	256	234	6459	33
Kurt	Zouma	0.73	182	170	5135	25
Serge Wilfried	Kanon	0.73	187	201	5151	26
Timo	Letschert	0.72	190	186	5182	26
Olivier	Deschacht	0.71	179	167	5088	39
José Ángel	Valdés Díaz	0.69	168	237	5132	30
Niklas	Moisander	0.68	187	178	5335	34
Djené	Dakonam Ortega	0.68	183	235	5510	28
Robin	Knoche	0.67	203	176	5088	27
Sébastien	Dewaest	0.67	232	210	5610	28
Marcel	Halstenberg	0.67	170	202	5865	28
Konstantinos	Laifis	0.67	198	221	5326	26
Germán Alejandro	Pezzella	0.67	214	216	5495	28
Joachim	Andersen	0.66	209	193	5614	23
Jonathan	Tah	0.66	218	220	5458	24

Table 44: Defenders. Best defenders according to the one-step approach with a minimum restriction of 5,000 assessments. Note that opportunities confirm the sum of all probabilities according to Expected Positional Defensive Actions.

First Name	Last Name	Valuation	Opportunities	# Defensive Actions	# Assessments	Age
Azor	Matusiwa	0.94	264	198	3413	22
Selim	Amallah	0.91	156	71	2636	23
Benjamin	Van Durmen	0.85	274	123	3736	23
Frank	Boya	0.82	323	209	4300	23
Manuel	Benson Hedilazio	0.80	142	52	2756	23
Ludovit	Reis	0.76	374	218	5085	19
Pablo	Rosario	0.72	340	170	4117	23
Teun	Koopmeiners	0.71	383	213	5693	22
Frenkie	de Jong	0.70	273	135	3395	22
Florian	Grillitsch	0.69	326	191	4459	24
Mikel	Merino Zazón	0.68	225	104	3185	23
Gabriel	Boschilia	0.67	138	44	2546	24
Gustavo	Hamer	0.66	207	108	2810	22
Konrad	Laimer	0.66	293	145	4518	22
Igor	Zubeldía Elorza	0.66	391	179	5336	23
Noussair	Mazraoui	0.65	218	134	3169	22
Bryan	Heynen	0.63	396	211	4990	23
Moussa	Djenepo	0.63	134	55	2944	21
Lucas	Torreira Di Pascua	0.61	234	104	3032	24
Samuel	Moutoussamy	0.60	183	66	2972	23
Răzvan Gabriel	Marin	0.60	327	108	4358	23
Matúš	Bero	0.59	264	100	4075	24
Robert Kenedy	Nunes do Nascimento	0.58	209	81	3607	24
Ritsu	Doan	0.58	329	101	5721	21
Martin	Ødegaard	0.58	282	136	4675	21

Table 45: Midfielders Talents. Best talents of midfielders according to the average approach with a minimum boundary of 2,500 assessments for the two-step model and a minimum boundary of 2,500 assessments for the one-step model. Note that opportunities confirm the average of Positions and Expected Positional Defensive Actions according to the sum of all probabilities.

References

Biermann, C. (2019). *Football Hackers*. Blink Publishing.

Bojinov, I. and Bornn, L. (2016). The pressing game: Optimal defensive disruption in soccer. *Proceedings of MIT Sloan Sports Analytics*.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., and Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1721–1730. ACM.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM.

Cronin, B. (2017). An analysis of different expected goals models. `https://www.pinnacle.com/en/betting-articles/Soccer/expected-goals-model-analysis/MEP2N9VMG5CTW99D`.

Das, S. (2019). Top 10 most popular sports in the world. `https://sportsshow.net/top-10-most-popular-sports-in-the-world/`.

Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240.

de Hoog, M. (2015). How data, not people, call the shots in denmark. `https://thecorrespondent.com/2607/how-data-not-people-call-the-shots-in-denmark/230219386155-d2948861`.

Decroos, T. (2019). matplotsoccer. `https://github.com/TomDecroos/matplotsoccer`.

Decroos, T., Bransen, L., Van Haaren, J., and Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1851–1861. ACM.

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38.

Gurpinar-Morgan, W. (2018). On quantifying passing skill. `https://2plus2equals11.com/2018/08/13/on-quantifying-passing-skill/`.

Hastie, T. and Tibshirani, R. (1990). Generalized additive models. chapman hall & crc. *Monographs on Statistics & Applied Probability. Chapman and Hall/CRC*, 1.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Kaggle, T. (2017). Allstate claims severity competition, 2nd place winner's interview: Alexey noskov. `https://medium.com/kaggle-blog/allstate-claims-severity-competition-2nd-place-winners-interview-alexey-noskov-f4e4ce18fcfc`.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pages 3146–3154.

Kwiatkowski, M. (2017). Quantifying finishing skill. `https://statsbomb.com/2017/07/quantifying-finishing-skill/`.

Lakshmanan, S. (2019). How, when and why should you normalize / standardize / rescale your data? `https://medium.com/@swethalakshmanan14/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff`.

Le, H. M., Carr, P., Yue, Y., and Lucey, P. (2017). Data-driven ghosting using deep imitation learning.

Link, D., Lang, S., and Seidenschwarz, P. (2016). Real time quantification of dangerousity in football using spatiotemporal tracking data. *PloS one*, 11(12).

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., and Matthews, I. (2014). quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th annual mit sloan sports analytics conference*, pages 1–9.

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., and Mohammadian, A. K. (2020). Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136:105405.

Payne, T. (2015). Tactical theory: Compactness. `https://spielverlagerung.com/2015/05/08/tactical-theory-compactness/`.

Power, P., Ruiz, H., Wei, X., and Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1605–1613. ACM.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). Catboost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, pages 6638–6648.

Rahnamai Barghi, A. (2015). *Analyzing dynamic football passing network*. PhD thesis, Université d'Ottawa/University of Ottawa.

Robberechts, P. (2019). Valuing the art of pressing. *StatsBomb Innovation in Football Conference*.

Routley, K. D. (2015). *A markov game model for valuing player actions in ice hockey*. PhD thesis, Applied Sciences:.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.

SciSports (2014). Sciskill index functional description.

SciSports (2020). Defensive contribution ratings in scisports platform enable in-depth data-driven scouting of defenders. `https://www.scisports.com/defensive-contribution-ratings-in-scisports-platform-enable-in-depth-data-driven-scouting-of-defenders/`.

Sha, L., Lucey, P., Zheng, S., Kim, T., Yue, Y., and Sridharan, S. (2017). Fine-grained retrieval of sports plays using tree-based alignment of trajectories. *arXiv preprint arXiv:1710.02255*.

Sportingintelligence (2019). Global sports salaries survey 2019. `https://globalsportssalaries.com/GSSS\%202019.pdf`.

Thompson, M. (2020). Possession adjusting or not possession adjusting. that is the question. `http://geommark.space/possession-adjusting-or-not-possession-adjusting-that-is-the-question/`.

Trainor, C. (2014). Defensive metrics: Measuring the intensity of a high press. `https://statsbomb.com/2014/07/defensive-metrics-measuring-the-intensity-of-a-high-press/`.

Tranfermarkt.de (2019). Market values. `https://www.transfermarkt.de`.

Witten, I. H. and Frank, E. (2002). Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, 31(1):76–77.

Wyscout (2019). Wyscout api, product documentation. `https://apidocs.wyscout.com/matches-wyid-events`.