PRESSING PHASE TRIGGERS & SUCCESS FACTORS IN PROFESSIONAL FOOTBALL

MSc Thesis in Data Science
Department of Mathematics and Computer Science
University of Southern Denmark

JUNE 2026

Student: Jonas Alexander Aalbæk Jensen
jonje24@student.sdu.dk
Data Science, IMADA, SDU Odense.

Supervisor: Panagiotis Tampakis
ptampakis@imada.sdu.dk
Data Science, IMADA, SDU Odense.

Table of Contents

List of Abbreviations

For the following a sufficient football understanding is recommended for the most optimal reading. However, it should be possible to obtain reasonable insights either way.

Passes Allowed Per Defensive Action written as PDDA

$$\rightarrow PDDA = \frac{Number\ of\ Passes\ made\ by\ Attacking\ Team}{Number\ of\ Defensive\ Actions}$$

Opta F24 = Opta event-data file containing match events & **Opta F73** = Opta possession-data

Shapley Additive Explanations can be written as SHAP = Model interpretation method

Receiver Operating Characteristic - Area Under the Curve written as ROC-AUC

Precision-Recall Area Under the Curve written as PR_AUC

Light Gradient Boosting Machine written as LightGBM

Footedness = Dominance of either left foot or right foot

Recall = Ability to classify number of positive cases

$$\rightarrow Recall = \frac{\text{True Positives}}{\text{True Positive + False Negative}}$$

Precision = Reliability to predict the positive classes

$$\rightarrow Precision = \frac{\text{True Positives}}{\text{True Positive + False Positive}}$$

Odense Boldklub = Danish Football Team (Men)

Extreme Gradient Boosting written as XGBoost

English FA = English Football Association

F1-Score = Machine Learning Metric

$$\rightarrow F1\ Score = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision+Recall}}$$

Divisionsforeningen can be written as Danish League

→ Represents 48 Danish football clubs in Denmark's top four tiers.

Second Spectrum = Tracking Data

False Positive can be written as FP

True Positive can be written as TP

False Negative can be written as FN

ABSTRACT

Pressing has become a central tactical component in modern football, yet it remains challenging to quantify. The most widely known pressing metric is PPDA, which measures the number of passes allowed per defensive action. While other recent research has applied tracking and event data to detect counterpressing, less attention has been given to modelling the success of high pressing. This thesis investigates which spatial, temporal, and tactical factors are associated with successful high-pressing episodes. The study uses full-season data from the Danish Superliga 2024/2025, provided by The Danish League, with tracking/event data from Second Spectrum and Opta for 192 matches, aligned and processed. With the aim of detecting pressing episodes using a detector inspired by football analytics literature. The fundamental press detection was further refined through video validation of 400 sampled episodes and domain-expert feedback. A successful press was likewise defined in process with domain experts and evaluated using three modelling approaches: a baseline feature model, an expanded model, and a simplified one-feature intensity model. The results showed that successful pressing is not explained solely by intensity. Features such as passing lane occlusion, trap intensity, and spatial pressure likewise contributed to the model's prediction of a successful press. This was further supported by the press trigger analysis and a heatmap, which indicated that successful presses occurred more often in wide areas near the sideline. The best-performing model was XGBoost with the expanded feature set, achieving an F1-score of 0.6538 and a recall of 0.9040 on the held-out test set, suggesting predictive signal despite the complexity in football analytics. The model can support football analysts by identifying pressing situations, evaluating and highlighting conditions which yields towards more effective high pressing. The findings suggest that successful high pressing is not determined by press intensity alone, but by a coalition of spatial, temporal and tactical factors.

1. Introduction

Pressing is a fundamental tactical component in modern professional football, where the English FA describes *"… pressing is when pressure is applied on the player or the team that's in possession. It's a skill used in all areas of the pitch – to win the ball back, dictate play, or delay the opposition."*. (England Football Learning, 2022) While pressing is widely discussed in the coaching literature and match analysis, quantifying it remains challenging due to its reliance on off-ball movements, spatial coordination, and contextual factors. The most widely used and recognised pressing metric in today's football is opponent passes allowed per defensive action, also known as PPDA. (HUDL, 2014) However, this pressure measurement is somewhat one-dimensional, as described by StatsPerform. Previous studies have used machine learning in pressing, with the most relevant work being *"Data-driven detection of counterpressing in professional football"*.(Bauer, P., & Anzer, G. 2021)

However, as the title suggests, this study uses machine learning to capture the factors in counterpressing, which is defined as *"pressure after losing the ball"*. (Fernández Navarro, J. 2018) This pressing form is tactically and temporally different from organised pressing. This thesis looks at pressing from a new perspective, which aims to capture the complexity in organised pressing, which will be done using a machine model approach and furthermore compare this performance against a single feature defined as pressing intensity. (Bekkers, J. 2025)

This will help assessing whether the machine model with multiple feature sets captures more predictive signal than a pressing-intensity model. Pressing, as described earlier, is a fundamental tactical component in modern professional football and covers a wide net of different pressing methods. (Chambers, T. 2022) However, this thesis has been provided in collaboration with the Danish Football Club Odense Boldklub and the Danish League, with a descriptive definition and video material for identifying high pressing. In this perspective, the thesis has condensed the many facets of pressing to only focusing on high pressing situations, which is defined as pressing episodes on the opponent's third of the pitch.

Additionally, this thesis aims to analyse pressing success factors by using match and event data and to analyse phase triggers. The thesis focuses on identifying which contextual situations and which spatial and temporal features contribute to a successful pressing outcome. This thesis seeks to provide a data-driven and reproducible framework for analysing pressing behaviour in professional football.

This thesis therefor tries to answer the following question:

1) Which tracking- and event features are associated with successful high pressing in the Danish Superliga?

2) Which pressing triggers can be derived from the thesis analysis of high pressing situations?

3) How does a multi feature model compare against a simpler interpretable pressing intensity model?

2. Related Work

A commonly used event-based measure of pressing in today's modern football is Passes Allowed per Defensive Action (PPDA), introduced by Colin Trainor in 2014. (HUDL, 2014) PPDA is calculated as the number of passes made by the team in possession divided by the number of defensive actions made by the out-of-possession team. It takes defensive actions using Opta definitions as tackles, interceptions, challenges/failed tackles, and fouls. The PPDA implementation measured high pressing using a defined pitch zone, with the boundary set at x = 40 in Opta's 0–100 coordinate system. This area includes the attacking half and a small part of the defending team's own half, allowing defensive actions and pressures to be counted. (HUDL, 2014) This logic is used as an important conceptual reference for defining pressing occurring high up the pitch. The definition of high pressing is conceptually equivalent to a higher pitch-location threshold, with x = 60.

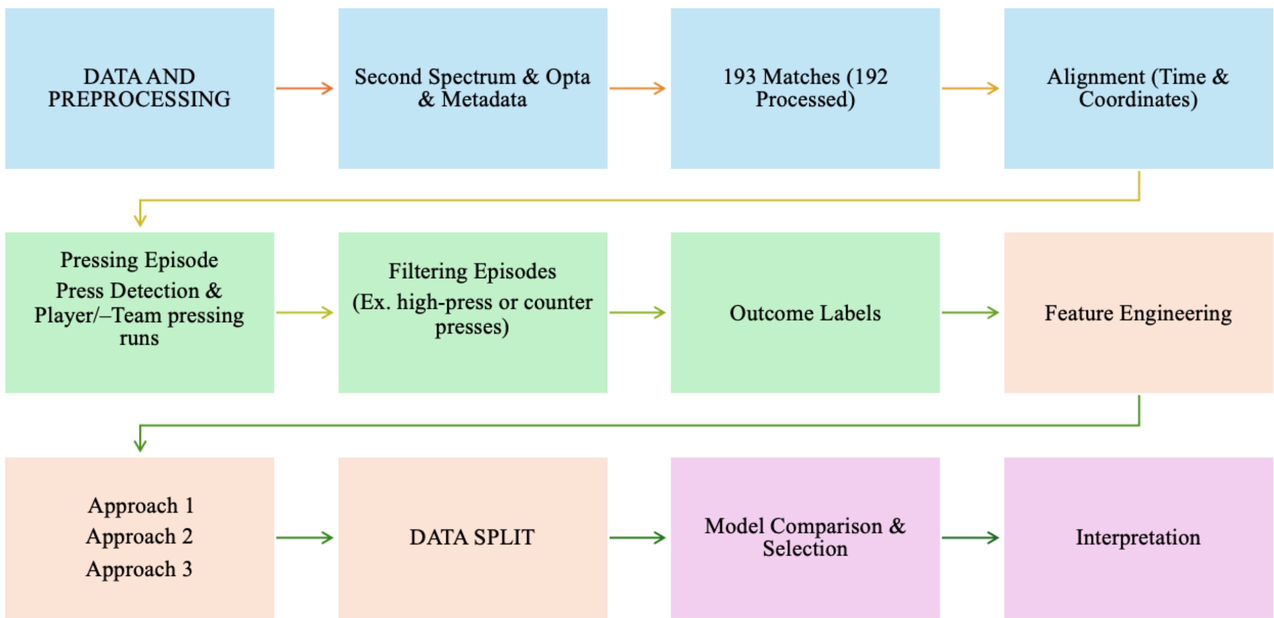
The present thesis also closely relates to the supervised machine-learning approach used in data-driven counterpressing-detection research. That paper similarly combines tracking and event data and formulates pressing-related behaviour as a classification task. Furthermore, it also shows the value of 134 expert-based features and the use of machine-learning models. This thesis follows a similar general structure by combining tracking and event data, constructing episode-level features, and evaluating tree-based models such as Random Forest and XGBoost. However, the present study differs in both tactical focus and practical annotation design, as this thesis focuses specifically on predicting the success of high-pressing. Also, the labelling procedure differs, where the counterpressing study used multiple human experts in the annotation process. This thesis was carried out by a single internal researcher, with video validation on a limited subset because of labour constraints. The counterpressing study has therefore been seen as a methodological inspiration rather than a directly replicated framework. (Bauer, P., & Anzer, G. 2021)

3. Method

This thesis investigates whether the success of high pressing in professional football can be predicted from event and tracking data. The methodological design is built around a single shared analytical pipeline that will be consistent across the thesis: match data are collected and synchronised, pressing episodes are detected from tracking data, tactical outcomes are labelled from event and possession data, features are extracted at the episode level, and models are evaluated using leakage-safe grouped validation. This shared framework is the methodological foundation for all three approaches.

- Approach 1 established the initial implementation with a working pressing detector, labelling logic, and baseline feature space.
- Approach 2 refined the tactical scope by isolating settled high pressing and expanding the feature set with temporal windows.
- Approach 3 introduced a press-intensity threshold model to test whether a simple tactical rule could explain pressing success without relying on a wider feature space.

Figure 1: Methodological Design



The illustrated pipeline in Figure 1 was first tested on a small sample of 10 matches before being scaled to the full season data. Furthermore, the following describes the shared components first, while the approach-specific differences are described afterwards.

3.1 Dataset & Data Sources

The empirical basis of the thesis is a full-season archive from the Danish Superliga 2024-2025 season. The raw archive contained 193 match folders with event and tracking feeds. Before any modelling was conducted, the match folders were audited for the files required by the pipeline: Second Spectrum tracking data, Second Spectrum metadata, Opta F24 event details, Opta F73 possession data, and match-results data. (Genius Sports, n.d. & Stats Perform, n.d.) This audit identified 192 tracking-ready matches. One match, 2024-10-06 SønderjyskE - FC Nordsjælland (2442608), was excluded because Second Spectrum tracking data were missing. The exclusion was made solely because of data availability, as no matches were removed based on tactical style, team identity, model performance, scoreline, or pressing outcome. (1)

The final sample therefore represents the complete set of matches provided from the available Danish League's Superliga archive. The data sources serve complementary purposes. Second Spectrum tracking data provides the high-frequency spatial positions of players and the ball. These data enabled the identification of pressure as movement in time and space. Opta F24 event data describes discrete on-ball actions, while Opta F73 possession data describes possession sequences and transitions. Match results data provide metadata such as teams, periods, lineups, and match identifiers. Which connects the detected pressing episode to the related footballing outcome.

All data were transformed into a common internal representation before feature engineering. Tracking data were loaded with assistance from the package "Kloppy" where possible, while the Opta XML files were parsed separately. (PySport, n.d.) Opta minute and second values were converted into 'gameClock' seconds within each period, and event coordinates were converted into metric pitch coordinates. Attack direction was inferred for each team and period, allowing spatial variables to be expressed relative to the opponent's goal. (2;3) Lastly, the package "MPL Soccer" was also used during the thesis, as a generating tool for football pitches. **(**Durgapal, A., & Rowlinson, A. 2025)

3.2 Pressing Episodes Detecting

In this thesis, the press detection formula is the primary core of the study, enabling the identification of relevant press episodes to evaluate. Meaning that the press detection in this study is based on a relevant combination of football analytical research papers and articles. Additionally, when the press detection was designed, the detection parameters were further fine-tuned through internal video validation, which involved a total of 400 pressing episodes. This fine-tuning process was performed four times, with each containing 100 pressing episodes. (4) A minimum running speed threshold of 5.0 m/s was selected, which is equivalent to 18 km/h. This threshold was chosen based on a research paper that compared high-speed running thresholds in elite football. (Silva et al., 2024)

The study also includes a constraint on the player's movement when detecting a pressing episode. This was defined as a maximum of 60 degrees angle towards the ball. This angle was chosen with regard to the research paper *"Analysis of player speed and angle toward the ball in soccer"*, which states, *"running directly toward the ball (angle close to 0 degrees) or parallel to it (angle close to 90 degrees)." & "... angle to the ball is approximately 60°... joining the players on the front line"* (Novillo et al., 2024). This threshold was used to ensure that detected runs represented purposeful movement toward the ball carrier rather than lateral repositioning or recovery movement. Furthermore, to distinguish pressing from general defensive running, the pressing detector was an added requirement that players during the detected action should close at least four meters towards the ball. This aligns with a research paper regarding pressing intensity, where effective pressing is described as a rapid reduction of space around the ball carrier. (Bekkers, J. 2025)

Similarly, a minimum run distance threshold of five meters was introduced to exclude small positional adjustments and movement that do not represent meaningful pressing behaviour.
In alignment with previous studies in football analytics, which used three passing actions and a five-second sequence window, this study has chosen to apply this method in the following way, with the model having the requirement of three consecutive pressing frames and merging nearby pressing runs occurring within the same interval into a single pressing episode. This allows the study of both pressing and coordinated pressing by multiple players.

Additionally, in the same alignment, a five-second outcome window was used to determine whether the pressing episode resulted in a binary outcome: successful or failed. (Forcher et al., 2022) Furthermore, an aerial ball height exclusion of one meter was applied, meaning that episodes where the ball was above one meter were excluded. This was done to detect and analyse pressing behaviour under controlled possession, without the chaotic patterns that may occur in aerial phases. (2;3) The applied method for press detection was additionally validated and discussed with Sønderjyske Fodbold's head of data. (C. E. Lopdrup, personal communication, 28 April 2026)

3.3 Successful Press - Labelling

The main target is a binary label indicating whether a detected pressing episode resulted in a tactically successful outcome. The label was designed to connect spatial pressure with the practical objectives of high pressing. A press was not considered successful simply because it looked intense in the tracking data. It had to produce a favourable outcome for the pressing team. The definition of success was informed by feedback from Odense Boldklub. (A.V. Nielsen, personal communication, 25 March, 2026) The desired outcomes were: winning the ball and creating a chance immediately after the regain, winning the ball and retaining possession, or forcing the opponent to play long, clear the ball, or send the ball out of play.

These priorities were converted into event-based labelling rules. A pressing episode was labelled as successful if subsequent Opta event or possession data indicated a regain, sustained possession, chance creation, forced clearance, long-ball escape, or an out-of-play outcome caused by the pressure. Episodes where the opponent retained controlled possession or progressed out of the pressured area were labelled unsuccessful. This is done through an implementation that links each detected pressing episode to later F73 possession information and F24 events. (2;3)

3.4 Feature Engineering

The approaches 1 & 2 use the same general feature-engineering principle. The features are computed from the tracking state at or around press initiation, while labels are computed from later events. The shared and total feature set are shown in Table 1 below.

TABLE 1: SHARED FEATURE SET APPROACH 1 & 2.

Feature	Approach	Tactical meaning	Definition
press_centroid_x	1 + 2	Describes the horizontal pitch location of the pressing unit.	Mean x-coordinate of pressing-team players.
press_centroid_y	1 + 2	Describes whether the pressing unit is central or wide.	Mean y-coordinate of pressing-team players.
opp_centroid_x	1 + 2	Describes where the team in possession is positioned.	Mean x-coordinate of opponent-team players.
opp_centroid_y	1 + 2	Describes whether the opponent structure is central or wide.	Mean y-coordinate of opponent-team players.
team_length	1 + 2	Measures vertical spread of the pressing team.	Difference between maximum and minimum x-coordinates of pressing-team players.
team_width	1 + 2	Measures horizontal spread of the pressing team.	Difference between maximum and minimum y-coordinates of pressing-team players.
shape_aspect_ratio	1 + 2	Describes whether the pressing shape is elongated or wide.	team_length / team_width, with protection against division by very small width.
def_line_height	1	Captures how high the defensive structure is positioned.	Defensive-line height derived from the pressing team's player positions.
opp_team_length	1 + 2	Measures vertical spread of the team in possession.	Difference between maximum and minimum x-coordinates of opponent players.
opp_team_width	1 + 2	Measures horizontal spread of the team in possession.	Difference between maximum and minimum y-coordinates of opponent players.
opp_shape_aspect_ratio	1 + 2	Describes the opponent's structural shape.	opp_team_length / opp_team_width, with protection against division by very small width.
dist_nearest_1	1 + 2	Measures immediate pressure on the ball carrier.	Distance from the ball to the nearest pressing player.
dist_nearest_2	1 + 2	Captures supporting pressure around the ball.	Distance from the ball to the second-nearest pressing player.
dist_nearest_3	1 + 2	Captures collective pressure density.	Distance from the ball to the third-nearest pressing player.
n_pressers_5m	1 + 2	Measures close-range pressure.	Number of pressing players within 5 metres of the ball.
n_pressers_10m	1 + 2	Measures local support around the press.	Number of pressing players within 10 metres of the ball.
press_surface_area	1 + 2	Larger values indicate a more spread pressing structure.	Convex-hull area covered by pressing-team players.
press_stretch_index	1 + 2	Measures how stretched or compact the pressing team is.	Average spatial spread of pressing-team players around their centroid.
opp_stretch_index	1 + 2	Measures whether the opponent structure is compact or stretched.	Average spatial spread of opponent players around their centroid.

ball_speed	1 + 2	Captures whether the ball is moving quickly during the press.	Speed of the ball in the current frame.
nearest_presser_speed	1 + 2	Measures the physical intensity of the closest presser.	Speed of the nearest pressing player to the ball.
avg_pressing_team_speed	1 + 2	Captures collective movement intensity.	Mean speed of pressing-team players.
n_sprinting_pressers	1 + 2	Measures how many players are pressing at high speed.	Number of pressing players above the sprint-speed threshold.
closing_speed_nearest	1 + 2	Measures whether the nearest presser is actively closing down the ball.	Change in distance between the nearest presser and the ball across frames.
pressing_intensity	1 + 2	Measures overall pressure applied to the ball.	Sum or aggregation of pressure contributions based on distance, speed, and movement toward the ball.

Furthermore, defensive line height is used only in approach 1 because it is measured differently in approach 2, as shown in the entire feature space in Appendix 1.

The third approach uses the same labelling; however, the feature engineering for the comparison is based on one feature design, "Press Intensity," where pressure is represented by player movement toward the ball and estimated time-to-intercept. Which is combined based upon the research paper *"Pressing Intensity / An Intuitive Measure for Pressing in Soccer".* This approach is used to evaluate whether intensity alone can measure successful pressing. *(*Bekkers, J. 2025; 5)

The distribution across approaches is shown in Table 2.

Table 2: The distribution of the features across all three approaches:

Feature set	Count	Explanation
Approach 1 base features	24 +1	Spatial, structural, proximity, and physical pressing features.
Additional Approach 2 base features	15	Tactical and contextual features added in the advanced settled high-press representation.
Total base feature concepts before temporal expansion	39	Approach 1 base features plus additional Approach 2 base features.
Approach 3 Single Feature	1	Sum or aggregation of pressure contributions based on distance, speed, and movement toward the ball.

Source: Self-made from appendix 1.

3.5 Model Training

The main predictive models were tree-based machine learning algorithms: Random Forest, XGBoost, and LightGBM. (LightGBM developers, 2025; scikit-learn developers, n.d.; XGBoost developers, 2026) These models were selected because the data are tabular and interaction-heavy. (Cisco ThousandEyes Engineering, 2025; Grinsztajn et al., 2022)Pressing success may depend on combinations of variables rather than isolated linear effects. For example, a short distance to the ball may be valuable only if the pressing team is compact, the opponent has limited passing lanes, and the ball is near a trapping zone. Tree-based models can capture such threshold effects and interactions without requiring the same assumptions as linear models. To avoid leakage between training and evaluation data, all splits were grouped by 'match_name'. This means that episodes from the same match were not allowed to appear in both training and validation or test data. Grouped splitting is necessary because pressing episodes within a match are not independent: they share teams, tactical plans, score context, pitch conditions, and temporal continuity. Random row-level splitting would risk overestimating model performance. Hyperparameter optimisation was performed with Weights & Biases, which tracked 59 hyperparameter searches, totalling approximately 4.800 runs, as seen in Appendix 2. The primary tuning procedure used grouped five-fold cross-validation, with Precision-Recall AUC (PR-AUC) as the main optimisation metric. PR-AUC was prioritised because the focus was on identifying successful presses. (Roberts, A., 2022) Furthermore, the F1-score was included in the evaluation of the best training model as it describes the balance between precision and recall, also, a threshold-finetuning of the F1-score was conducted across all three approaches. (Brownlee, 2020; Friedman, 2024)

3.6 Approach 1 // Machine Learning Model

Approach 1 was the first complete implementation of the pressing-analysis pipeline. Its main purpose was to establish whether raw tracking and event data could be transformed into a usable supervised learning dataset. It therefore included the first working versions of the data-loading, conversion tracking, press detection, episode grouping, outcome labelling, and baseline feature extraction steps. Approach 1 was also the stage where imbalance-aware modelling choices were most relevant. The notebook used or proposed settings such as 'class_weight="balanced" ' for Logistic Regression, Random Forest, and LightGBM, and 'scale_pos_weight' for XGBoost. (6) These settings were only included in the first broad modelling setup, where the aim was to make the baseline robust to what became an uneven target distribution based on the earlier mentioned OB definition. Because this was the first approach, it created the methodological baseline. And the focus was on defining pressure from first principles using tracking-derived movement and spatial criteria shown earlier in Table 1.

The contribution of Approach 1 showed that the thesis could move from raw match feeds to episode-level observations with meaningful labels. However, it also exposed a limitation. The detected pressing contained highly imbalanced data, including different tactical situations, such as counterpresses. Situations which share movement features but differ in tactical meaning. This motivated a second approach.

3.7 Approach 2 // Machine Learning Model

Approach 2 differed from Approach 1 by refining the tactical scope and expanding the feature representation. Instead of modelling all detected pressing actions together, the analysis was restricted to settled high pressing. This was done to exclude counterpress candidates by removing all detected pressing episodes within five seconds of ball possession, removing all episodes involving corner situations, and using field-position logic such as 'pressing_progress_to_opponent_goal >= 0.60'. (2) This change was made because Approach 1 was deemed too broad in its search. Approach 2 also introduced temporal windows. Instead of representing a press only through a single snapshot, features were computed across four temporal windows, being the first two seconds prior to the press: 'pre_2s', 'pre_1s', and the first second of the press 'start', and 'early_1s'. Frame-level features were summarised using statistics such as mean, maximum, and standard deviation. This allowed the model to compare the information available at press start with the information contained in the first second of pressure. The 'start_only' and 'early_1s' feature views made it possible to test whether early pressing dynamics improve prediction beyond static press-start structure. The approach 2 feature set included tactical concepts such as passing-lane occlusion, trap intensity, escape difficulty, compact pressure, and opponent shape, which should contribute to a higher tactical context. Approach 2, therefore, differed from Approach 1 in both population definition and feature design, as it narrowed the population to settled high-pressing and described each press with more predictors. (2)

3.8 Approach 3 // Press Intensity Model

Approach 3 used the same pressing-episode definition and target-labelling logic as the previous approaches. However, instead of using the broader spatial, contextual, and temporal feature sets from Approaches 1 and 2. Approach 3 represented pressure through one research-derived feature design: press intensity. This feature is based on player movement toward the ball and estimated time-to-intercept, following the logic of the research-paper concept "Pressing Intensity: An Intuitive Measure for Pressing in Soccer". (Bekkers, J. 2025)

This formular in figure 2 illustrates the pressure on object (P_J) equal one minus the probability that non active pressing players (i) can get to object (j), in approach 3 the calculated pressure is the detected frame-level intensity in the pressing episode. The first step in Approach 3 was to compute research-oriented press-intensity variables for each episode.

Figure 2: Press Intensity, Formular.

$$\mathrm{P_J} = 1 - \prod_{\mathrm{i}} (1 - \mathrm{P_{i,j}})$$

These probabilities were aggregated into episode-level measures, yielding a compact representation of pressure intensity while retaining the original episode identifiers, match names, team information, split assignments, and success labels. (UnravelSports, n.d; 5)

4. Results

This chapter presents primarily findings from the full-season modelling pipeline for predicting successful settled high pressing in the Danish Superliga 2024-2025 season. The purpose of the chapter is threefold: first, to compare the candidate machine learning models, second, to report the predictive performance of the selected final model, and third, to translate the model outputs into football-relevant interpretations of successful high pressing. Only full-season results are reported in this chapter.

4.1 Model Comparaison

Three tree-based classifiers were compared as candidate models: Random Forest, XGBoost, and LightGBM. Each approach and model was evaluated using the same train, validation, and test structure seen in **Table** 3. Furthermore, a validation phase was used to select classification thresholds through F1-based threshold tuning, while the held-out test set was used for final performance comparison. However, as mentioned earlier, approach 1 included several other forms of pressing episodes, which resulted in higher episode and split counts in the train, validation, and test sets: 5.172 (train), 3.789 (validation), and 4.717 (test), respectively. (6) The training and validation set for approach 2 respectively contained 4.348 and 1.071 pressing episodes. The test set contained 1.432 episodes. (7)

Table 3: Comparison across approaches.

Approch	Model (Best)	PR-AUC	ROC-AUC	F1-score	Precision	Recall
1	Random Forest	0.2469	0.6427	0.3305	0.2469	0.6496
2	**XGBoost**	0.6448	0.6503	0.6538	0.5121	0.9040
3	Single-Feature	0.5092	0.5242	0.6381	0.4779	0.9601

Source: (5; 6; 7)

The following will be in perspective of the selected approach 2, which performed best with the model XGBoost as illustrated in Table 3.

Table 4: Comparison across models, Selected Approach 2.

Model	PR-AUC	ROC-AUC	F1-score	Precision	Recall
Random Forest	0.6207	0.6321	0.6490	0.4957	0.9394
XGBoost	0.6448	0.6503	0.6538	0.5121	0.9040
LightGBM	0.6311	0.6377	0.6512	0.4946	0.9527

Source: (7)

Comparing the models internally within the selected approach 2, as shown in Table 4. It's seen that the difference between Random Forest, XGBoost and LightGBM was not one of complete dominance, however, XGBoost offered the strongest ranking performance.

The held-out test set provided clearer separation between the models. XGBoost achieved the highest test PR-AUC (0.6448), the highest test ROC-AUC (0.6503), and the highest test F1-score (0.6538). Random Forest achieved a test PR-AUC of 0.6207, ROC-AUC of 0.6321, and F1-score of 0.6490. LightGBM achieved a test PR-AUC of 0.6311, ROC-AUC of 0.6377, and F1-score of 0.6512. The test results therefore support using XGBoost as the final model because it produced the strongest ranking and threshold-based classification performance on the untouched test set. (7)

PR-AUC is especially relevant in this setting, as the model must identify successful pressing episodes among many structurally similar unsuccessful ones. ROC-AUC describes the model's overall ranking ability across both classes, whereas PR-AUC is more sensitive to the model's ability to separate successful episodes. For this reason, the test PR-AUC of 0.6448 is treated as the primary performance result for the selected XGBoost model. (Brownlee, 2020; Friedman, 2024; Roberts, 2022)

4.1.1 Model Selection

The final selected model is the XGBoost classifier trained on the 180-feature reduced numerical feature set. The selected validation threshold for XGBoost was 0.2799, determined via F1-score threshold analysis on the validation set. At this threshold, XGBoost achieved a validation F1-score of 0.6379, precision of 0.5072, and recall of 0.8595. This threshold was then applied to test-set evaluation, where the final model achieved a test PR-AUC of 0.6448 and a test ROC-AUC of 0.6503. (7)

The choice of XGBoost was based on its strongest test PR-AUC among the three models. Additionally, it produced the strongest test ROC-AUC and F1-score, suggesting that its advantage was not limited to a single metric. Furthermore, XGBoost supports SHAP-based interpretation, which is valuable because this thesis aims not only to classify pressing episodes but also to interpret which tactical conditions are associated with success. Also, the model's feature importance allowed predictive performance to be discussed in terms of pressing intensity, spatial control, trap dynamics, opponent escape difficulty, and the contextual contribution to pressing in football. The model selection was based on a combination of predictive performance, test-set ranking, etc. The choice was XGBoost from approach 2 as the final model, because it performed best on the held-out test set for PR-AUC, ROC-AUC, and F1-score, as seen in Table 4.

4.1.2 Error Analysis

On the held-out test set of 1,432 pressing episodes, the XGBoost model achieved a PR-AUC of 0.6448, an ROC-AUC of 0.6503, an F1-score of 0.6538, a precision of 0.5121, and a recall of 0.9040. The confusion matrix contained 612 true positives, 583 false positives, 172 true negatives, and 65 false negatives. In practical terms, this means the model captured most successful pressing episodes but also classified many unsuccessful episodes as successful. (7) The high recall of 0.9040 indicates that the model was sensitive to successful pressing situations: only 65 successful pressing episodes were missed on the test set. However, the precision of 0.5121 shows that the model's positive predictions were not always correct. Of all episodes predicted as successful, just over half were successful. This reflects the difficulty of the task, football outcomes are noisy, and a pressing situation with many favourable structural characteristics may still fail because of technical execution, opponent skill, ball bounce, referee decisions, or context not fully captured by the features. This is evident in the feature comparison, which is shown in Table 5 below.

Table 5: Feature comparison, TP & FP.

prediction_category	False Positive	True Positive	FP_vs_TP_Diff
pre_2s_press_surface_area_std	35.20	38.62	-3.42
pre_1s_press_surface_area_std	18.90	21.39	2.49
early_1s_opp_surface_area_std	16.73	18.58	-1.85
early_1s_press_surface_area_std	20.53	22.30	-1.77
previous_event_seconds_before_press	3.96	2.36	1.60
early_1s_dist_to_central_axis_max	17.56	18.99	-1.44
pre_2s_opp_surface_area_std	33.70	32.28	1.42
early_1s_team_length_mean	60.65	61.92	-1.27
last_pressing_team_event_seconds	22.47	21.28	1.20

Source: Self-made. (7)

This shows how false positives and true positives are very similar in feature values, even though they are two different outcomes. A false positive is an episode in which the model predicted pressing success, but the opponent escaped, retained control, or avoided the target outcome. These cases may appear structurally favourable to the pressing team, but were effectively resolved by the team in possession, as illustrated in Table 5. Meaning that the model is struggling to truly distinguish successful and failed pressing episodes. However, as mentioned earlier, the model has a high recall, which can indicate that the model can be used for detecting pressing episodes useful for a club football analyst.

4.2 Overview of Results

The overall results indicate that successfully settled high pressing contains a measurable signal in the tracking-derived feature space. The selected XGBoost model performed better than a naive or purely random expectation, with test PR-AUC of 0.6448, ROC-AUC of 0.6503, and F1-score of 0.6538. These values do not imply perfect prediction, but they do show that pressing success is not random with respect to the engineered tactical features. The model used spatial and temporal information to rank successful pressing episodes above unsuccessful ones, with meaningful differences. The comparison across candidate models suggests that the predictive task is learnable but challenging, as seen in Table 4. All three tree-based models produced broadly similar F1-scores on the test set, ranging from 0.6490 to 0.6538. This narrow range indicates that the result is not simply a quirk of one algorithm. At the same time, XGBoost's higher PR-AUC and ROC-AUC suggest that its probability ranking was more informative than the alternatives in the final test. The similarity between the models also implies that the feature set itself carries the main tactical information, while the algorithm provides incremental performance differences. The results also show that the model benefits from a broad range of contextual features. The most important XGBoost features were not limited to one type of variable. They included pressure intensity, passing-lane occlusion, area ratio, trap intensity, pressing stretch, number of pressers, escape difficulty, opponent width, and nearest-presser distance. This pattern suggests that high-pressing success is associated with the interaction of intensity and spatial context rather than with a single isolated feature family. (7)

Table 6: Chosen hyperparameters for best model.

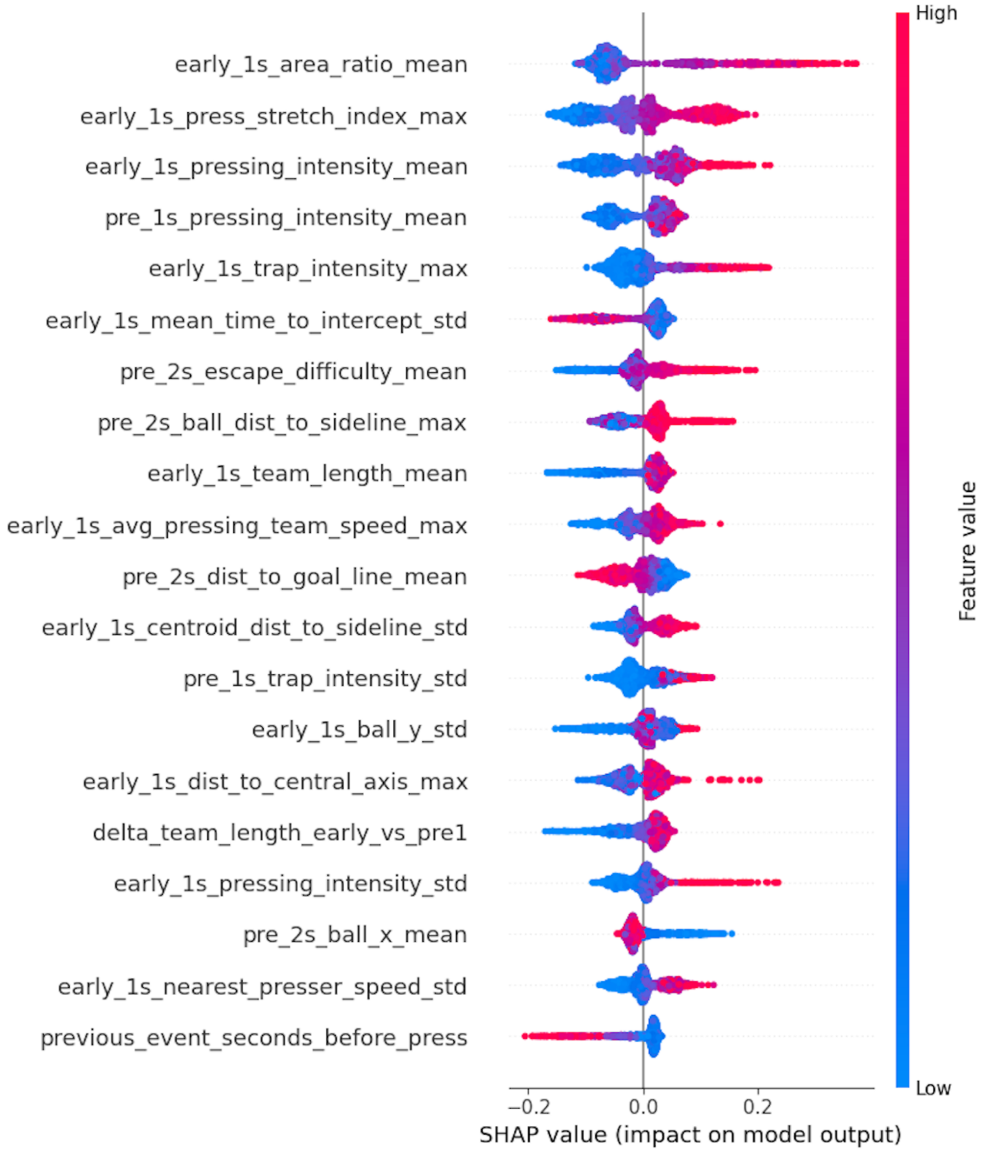
Hyperparameter	Describtion	Range	Model (Best) - Chosen
n_estimators	Number of boosting trees built by the model.	100 to 500	344
max_depth	Maximum depth of each tree.	3 to 50	5
learning_rate	Size for each update.	0.01 to 0.5	0.015
subsample	Fraction of training rows used for each tree.	0.5 to 1.0	0.665
colsample_bytree	Fraction of features used for each tree.	0.5 to 1.0	0.573
min_child_samples	Minimum number of samples required in a child node.	1 to 20	15

Source: (7; 8)

4.3 Which tracking- and event features are associated with successful high pressing in the Danish Superliga?

The football interpretation of the results is that successful high pressing appears to depend on both the amount of pressure applied and the structure around that pressure. The selected XGBoost model assigned high importance to pressing intensity before and during the first second of the press, but it also relied on variables describing passing-lane occlusion, trap intensity, area ratios, pressing-team stretch, opponent width, escape difficulty, and nearest-presser distance. In tactical terms, this suggests that pressure is most meaningful when it is embedded in a structure that limits the opponent's next action.

Figure 3: SHAP-Values, XGBoost.

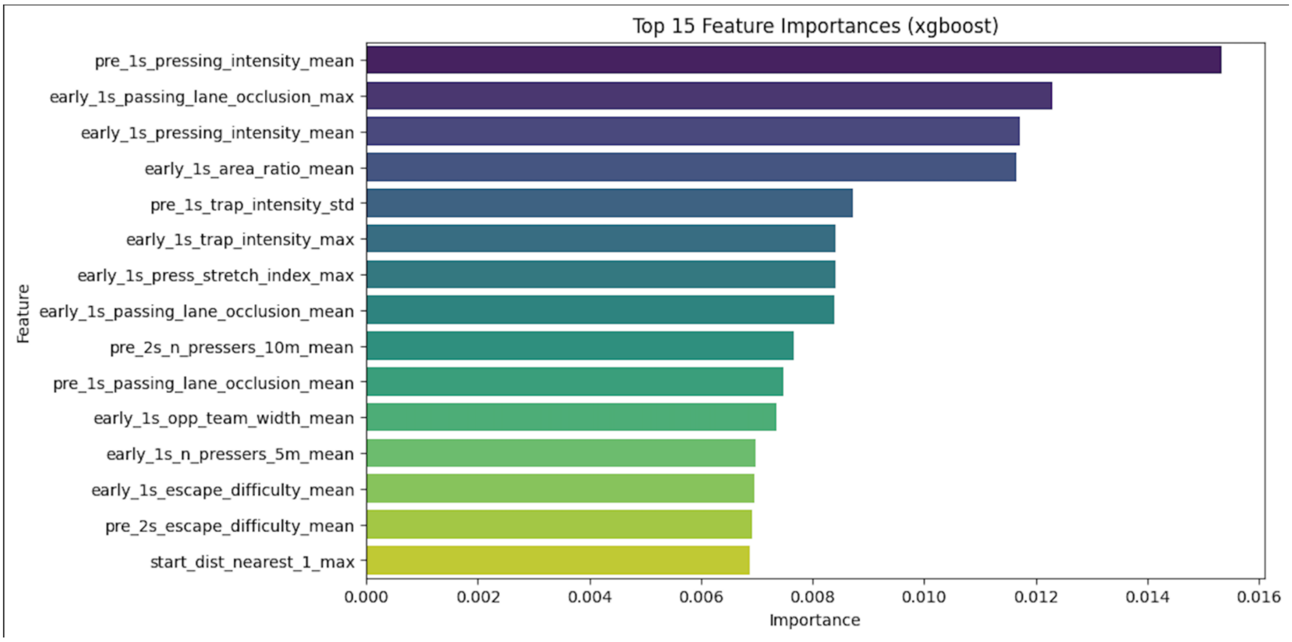


SOURCE: (7).

Figure 3 supports this idea that the first second after pressure begins contains important information about whether the press is likely to succeed. The model does not rely only on the initial snapshot, it also responds to how the pressing situation develops immediately after initiation.

At the same time, pre-press context also matters, as many pre-features are among the top SHAP values. This indicates that successful pressing may be partly prepared before the formal start of the detected episode. The model appears to value whether the pressing team was already close, compact, and positioned to limit options before pressure fully materialised.

Figure 4: XGBoost feature-importance.



SOURCE: (7)

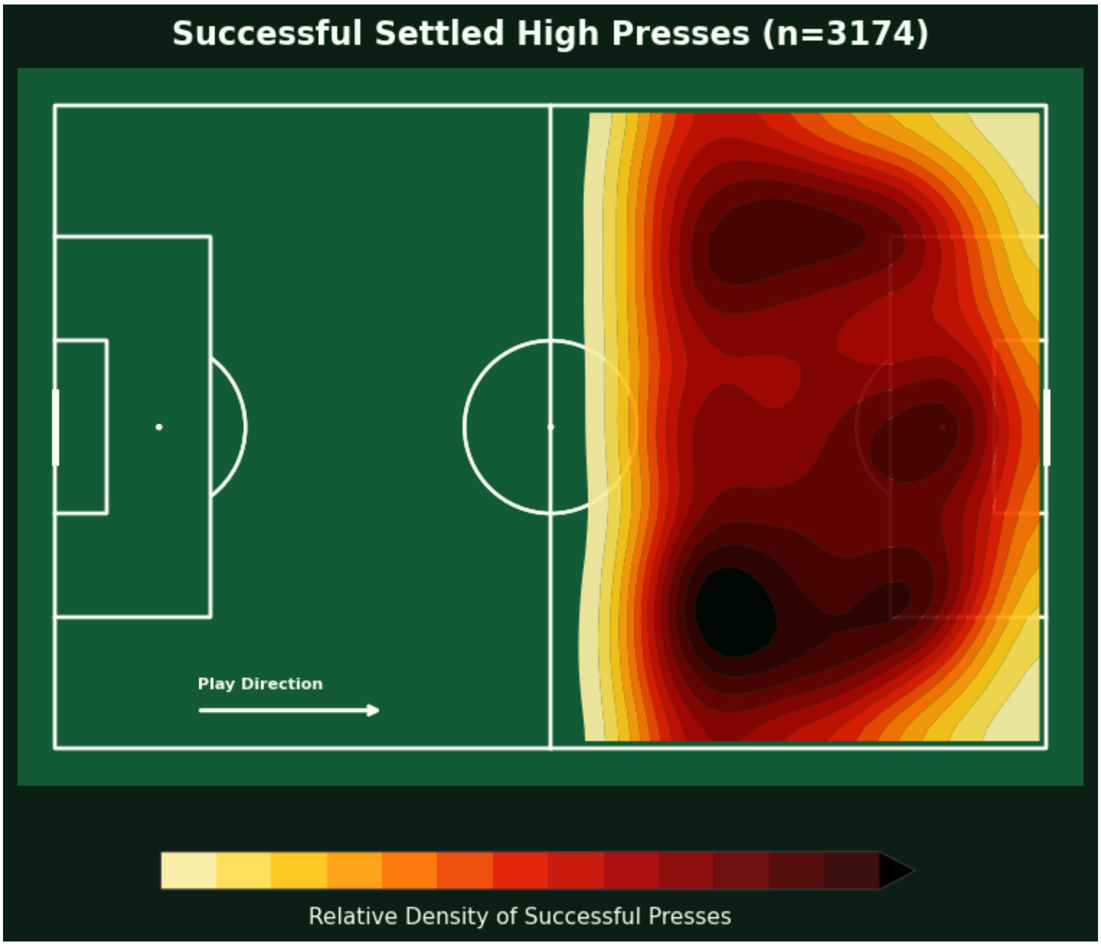
These features can be grouped into several football concepts. The first is pressure intensity, represented by pre-press and early-press intensity variables. The second is spatial restriction, represented by passing-lane occlusion, trap intensity, and escape difficulty. The third is collective structure, represented by area ratio, pressing stretch index, and opponent team width. This grouping can be more meaningful than listing column names alone because it shows a quick tactical aspect of pressing outcomes. It is important to distinguish the two interpretation methods. Built-in feature importance, as seen in Figure 4, indicates which variables the model used strongly in its tree structure. SHAP values in Figure 3 describe how feature values contributed to individual predictions and show whether high or low values tended to push predictions toward success. (7)

4.4 Which pressing triggers can be derived from the thesis analysis of high pressing situations?

The model-derived findings suggest several pressing triggers that are associated with successful high pressing. The first trigger is early pressure intensity. Both ‘pre_1s_pressing_intensity_mean’ and ‘early_1s_pressing_intensity_mean’ were highly ranked in both Figures 3 and 4, indicating that successful presses are associated with strong pressure before and immediately after the press begins. This does not mean that intensity alone guarantees success, but it suggests that the speed and immediacy of the pressure are central to a successful press.

The second trigger is the restriction of the opponent's passing options. Passing-lane occlusion and escape-difficulty variables were among the most important features. This indicates that the model associated successful pressing with situations where the opponent had fewer options away from pressure. In football terms, the press is more likely to succeed when the first defender's pressure is supported by the surrounding players, who block or disrupt the next pass.

Figure 5: All successful high settled presses, 3.174 episodes.



Source: (2)

The third trigger is spatial trapping, trap intensity. As seen in Figure 5, pressing near constrained areas, particularly where the ball is closer to the sideline or where the opponent's available space is reduced, appears to be associated with successful outcomes. This aligns with the generally known concept of footedness in football, which states that most players are right-footed. This can further explain why more successful presses happen on the left side, attacking a defender who is right-footed and therefore has more difficulty passing, as the sideline acts as an extra defender. This generalisation is backed by this paper on footedness in a previous World Cup tournament, which showed that 79 per cent of all players were right-footed. (Carey et al., 2001; 7)

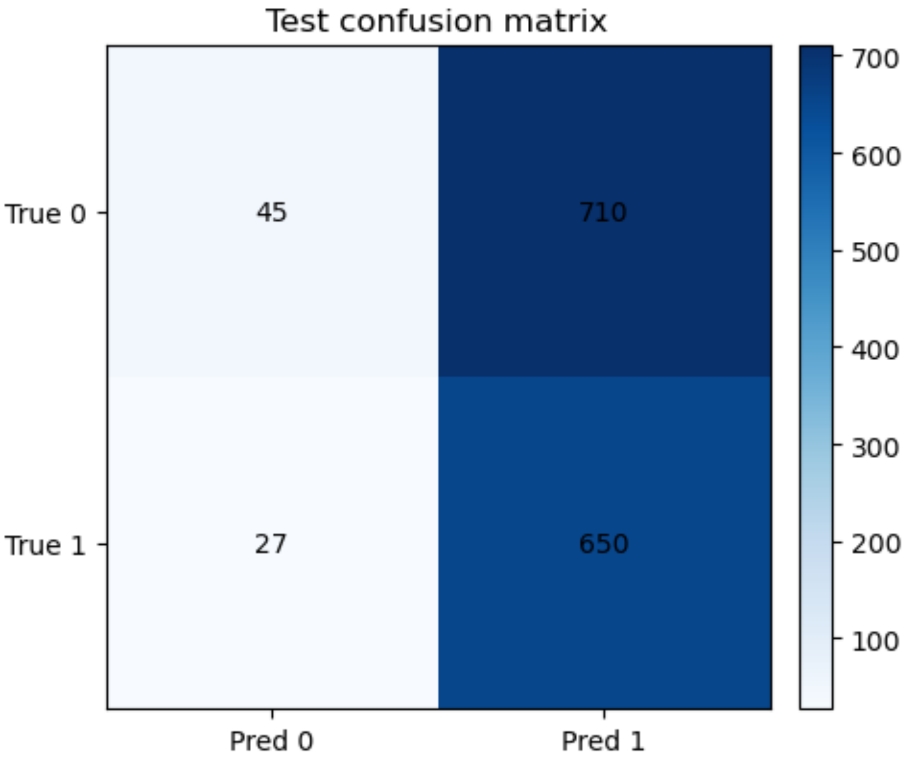
The fourth trigger is collective compactness and shape. Features such as area ratio, pressing stretch index, team length, and opponent width indicate that the model used information about team structure, not only individual actions. Successful pressing appears more likely when the pressing team can combine pressure on the ball with a compact supporting shape and when the opponent's structure is vulnerable to being constrained. The model-derived triggers therefore suggest that successful high pressing is not only a matter of individual pressure, but also of whether the surrounding spatial context limits the opponent's next action. (5)

4.5 How does a multi feature model compare against a simpler interpretable pressing intensity model?

The selected multi-feature XGBoost model can be compared with the one-feature press-intensity model. The model selected a fixed threshold of 0.10. On an untouched test set of 1432 episodes, which achieved a precision of 0.4779, recall of 0.9601, F1-score of 0.6381, ROC-AUC of 0.5242, and PR-AUC of 0.5092, as seen in Table 3. (5)

This comparison suggests that press intensity alone contains some useful signal, but it is not sufficient to match the broader multi-feature model. The intensity-only model achieved a PR-AUC of 0.5092, while the selected XGBoost model achieved a test PR-AUC of 0.6448 on the final multi-feature evaluation. The model also had extremely high recall with 0.96 but a below 50 per cent precision, which was expected, seeing 45 true negatives and 710 false positives demonstrated in Figure 6 below. This indicates that the low threshold captured nearly all successful presses but did little to separate successful and unsuccessful episodes.

Figure 6: Confusion Matrix, Single Feature Model (TEST-set).



Source: (5)

The practical interpretation is that intensity is useful but incomplete. A simple threshold on pressure intensity is easy to communicate and may serve as a transparent tactical baseline. However, the multi-feature model adds information about passing-lane occlusion, escape difficulty, trap intensity, compactness, opponent width, nearest-presser distance, and early spatial dynamics. The stronger performance of the multi-feature model therefore supports the argument that successful high pressing depends on more than applying pressure. It also depends on the press's geometry, the opponent's available options, the ball's location, and the timing of collective movement.

5. Discussion

The central result is that successful settled high pressing in the Danish Superliga can be predicted with meaningful, although imperfect, accuracy from tracking-derived tactical features. The selected XGBoost model outperformed the simpler press-intensity model, especially in PR-AUC and ROC-AUC, suggesting that pressing success is not explained by physical pressure alone. Instead, the results indicate that intensity becomes most valuable when combined with spatial restriction, timing, compact support, and opponent constraint.

5.1 Main Predictive Results

The final XGBoost model achieved the strongest overall performance among the evaluated tree-based models, with a test PR-AUC of 0.6448, ROC-AUC of 0.6503, and F1-score of 0.6538. These values indicate that the engineered feature space contains meaningful information about whether a settled high press will succeed. The result is important because it suggests that successful pressing is not random with respect to tracking-derived spatial and dynamic conditions. Instead, the model learned patterns that better distinguish successful from unsuccessful pressing episodes than a simple baseline or a single-variable intensity rule. At the same time, the model performance should be interpreted as insightful into successful pressing, although football is a highly dynamic and adversarial environment, and pressing success depends not only on the pressing team's structure but also on the opponent's technical quality, decision-making, body orientation, passing options, match context, and individual execution. The XGBoost model captures many features related to space and movement, but it cannot fully observe every relevant tactical or technical detail. Therefore, the model's imperfect performance is not simply a weakness of the algorithm, it reflects the complexity of the phenomenon being modelled.

The classification profile provides further insight. The model's recall of 0.9040 shows that it identified most successful pressing episodes in the test set. In practical terms, this means that the model is sensitive to the conditions under which successful high pressing tends to occur. However,, as previously discussed, the model still maintains a high recall and an adjustable pressing detection threshold, which may indicate it could be used in practice as a press detection tool. This could help support a club analyst by tagging pressing episodes for further video and contextual analysis.

5.2 Why Pressing is predictable, but not perfect

The results show that pressing success is predictable to a meaningful degree because pressing is partly structured. Teams do not press randomly. Successful high pressing often involves coordinated pressure on the ball, compact support behind the first presser, restricted passing lanes, and positioning that reduces the opponent's options. These aspects are visible in tracking data and can therefore be represented through engineered features such as pressing intensity, passing-lane occlusion, escape difficulty, trap intensity, compactness, and nearest-presser distance.

However, the outcome of a press remains imperfectly predictable because the outcome of a press depends on more than the spatial structure at the moment of pressure. A well-organised press can fail if the opponent executes a technically difficult pass, uses a press-resistant player, wins a duel, or benefits from a small timing advantage. Similarly, a less structurally optimal press can succeed if the opponent mishandles the ball, selects a poor pass, or is forced into an error by an individual defensive action. This was illustrated in Table 5 earlier.

The imperfection of the model means the thesis cannot claim that pressing success has a fixed outcome based on the model's prediction. Rather, it shows that tracking and event data contain a measurable tactical signal that can support the analysis of high pressing. The model should be understood as a description of pressing conditions, not as an explanation of every individual pressing outcome in professional football.

5.3 Intensity is not enough

The comparison with the press-intensity model clarifies the role of intensity in pressing success. The model showed that the research-based 'press_intensity' contains a predictive signal. However, the selected fixed threshold was low at 0.10, which caused the model to classify most test episodes as successful. This produced a high recall of 0.9601 and an F1-score of 0.6381. In other words, the threshold rule captured almost all successful presses, but it failed to identify differences from unsuccessful ones. The ROC-AUC of 0.5242 and PR-AUC of 0.5092 provide a more cautious interpretation, which states that intensity alone has the capability to rank pressing episodes, however, not as well as the multi-feature XGBoost model. The stronger performance of XGBoost in PR-AUC and ROC-AUC indicates that additional contextual features add value. Passing-lane occlusion, trap intensity, escape difficulty, area ratio, opponent width, nearest-presser distance, and early-window dynamics all contribute to a clearer description of the pressing situation. The central tactical interpretation is therefore that successful high pressing is not simply applying pressure, but it is a matter of applying pressure in a spatial and temporal context that restricts the opponent's next action.

5.4 Limitations

Firstly, the target variable depends on an operational definition of pressing success. Although the label was designed to reflect football-relevant outcomes such as regains, sustained possession, chance creation, forced clearances, and out-of-play events, any rule-based label simplifies the complexity of football decision-making. Some presses may create tactical value without meeting the formal success definition, while some labelled successes may be influenced by opponent mistakes rather than pressing quality alone.

Secondly, the data capture important but incomplete aspects of football behaviour. Tracking data provide player and ball positions, movement, and derived spatial relations, but they do not directly measure player intention, communication, body orientation, scanning, tactical instructions, or psychological pressure. Event data provide outcome context, but they may not capture every micro-action that shapes whether a press succeeds. As a result, the model can describe many structural conditions of pressing but cannot fully observe the tactical and technical mechanisms behind each outcome.

6. Future Works

The findings of this thesis have provided the foundations for future work on data-driven football pressing analysis and detection. While the results show pressing can be modelled and interpreted from tracking and event-derived data, future work described below could strengthen the generalisation.

6.1 Generalizing across multiple seasons

Future research could extend the thesis in several directions. First, the pipeline should be tested across multiple seasons and leagues to examine whether the identified pressing patterns generalise beyond the Danish Superliga 2024-2025 season. A broader dataset would make it possible to study whether the same features predict pressing success across different tactical cultures and competition levels.

6.2 Team Specific Pressing Style

Second, future work could model team-specific pressing styles. Teams may press with different triggers, levels of compactness, and risk profiles. A global model can identify general patterns, but team-level models may reveal that successful pressing depends on different feature combinations across tactical systems.

6.3 Improving feature representation

The results showed that successful and unsuccessful presses can sometimes be difficult to separate from immediate tracking features alone. This may be because the outcome of a press depends not only on its structure, but also on its timing within the match, pitch location, and whether the opponent has already experienced similar pressing situations. Future models could therefore include explicit time- and zone-based features. A press early in a match may affect the opponent differently than a similar press late in the game, and a sideline press may create different escape options than a central press. Opponents may also adapt after facing the same pressing pattern earlier in a match or season. Including match-time, repeated-exposure, and pitch-zone variables could help explain why similar pressing structures sometimes produce different outcomes.

6.4 Beyond Binary Pressing Outcomes

Fourth, future research could move beyond episode-level classification and model the sequence after the press. Instead of only predicting binary success, a model could estimate how pressure changes the probability of different possession outcomes: regain, forced long ball, clearance, progression, chance creation, or escape. This would create a more nuanced valuation of pressing actions.

7. Conclusion

This thesis investigated whether successful high pressing in professional football can be identified, modelled, and interpreted using synchronised tracking and event data. Motivated by the limitations of traditional aggregate measures, the thesis focused on pressing as a dynamic tactical behaviour involving player movement, compactness, and pitch location. To answer the central research questions, a reproducible pipeline was developed to transform raw data into pressing episodes, engineer tactical features, and evaluate predictive models using data from the Danish Superliga 2024-2025 season. The main finding is that successful high pressing can be predicted from meaningful performance metrics derived from tracking data. The selected XGBoost model achieved a test PR-AUC of 0.6448, ROC-AUC of 0.6503, and F1-score of 0.6538. The model's high recall successfully identified most pressing successes, but its moderate precision indicates that predicting outcomes remains difficult. The results also show that successful high pressing relies on a combination of pressure intensity and spatial context. Key feature families included early pressing intensity, nearest-presser distance, pressers near the ball, passing-lane occlusion, trap intensity, escape difficulty, and compactness. In tactical terms, pressure is most valuable when supported by compact positioning, blocked passing routes, favourable pitch locations, and limited escape possibilities. From this, several model-derived pressing triggers were identified: applying pressure early, positioning multiple defenders close to the ball, and creating spatial constraints. Comparing the multi-feature model with a press-intensity benchmark further clarified the role of intensity. While the benchmark showed that intensity alone contains useful information, the multi-feature model's stronger performance confirms that intensity alone is not a sufficient explanation. Successful pressing depends on a wider tactical context, including geometry, opponent shape, and collective timing. Methodologically, the thesis demonstrates the value of combining tracking data with machine learning to analyse off-ball defensive behaviour. While grouped evaluation and feature cleaning strengthened the model's validity, limitations remain, such as rule-based success labels and the inability of tracking data to capture player intention, communication, or body orientation. Ultimately, the thesis contributes a data-driven framework for detecting, modelling, and interpreting high pressing, showing how machine learning can generate football-relevant tactical insight and potentially be used as a practical pressing tool by club analysts. The main conclusion is therefore that successful high pressing in the Danish Superliga is associated with a combination of intensity, compact support, spatial restriction, timing, and opponent constraint. In other words, applying pressure in a spatial and temporal context that restricts the opponent's next action.

Acknowledgments

The completion of this master's thesis would not have been possible without the support, encouragement, and contributions of several people.

First and foremost, I would like to thank my supervisor, Panagiotis, for his guidance, valuable feedback, and continuous support throughout this project. His insights, encouragement, and willingness to discuss ideas have been greatly appreciated and have played an important role in shaping this thesis.

I would also like to thank my study group, **Data Crewet**, for making the master's journey both academically rewarding and enjoyable. The many discussions, shared frustrations, laughs, and mutual support over the years have meant a great deal and have contributed far beyond this thesis alone.

Furthermore, I am grateful to Asser and Mounir from the Danish League for their openness, engagement, and willingness to share knowledge and data, which have strengthened the practical relevance of this work.

I would also like to extend my thanks to Carl-Emil from Sønderjyske Fodbold for valuable discussions and for sharing his expertise within football analytics. Also, a sincere thank you to Odense Boldklub for taking the time to help clarify the definition of pressing, which provided important domain knowledge for this thesis.

Finally, I would like to thank everyone who has contributed to this project in one way or another. Through academic discussions, professional insights, encouragement, or simply showing interest in the work, the support has been greatly appreciated and has helped make this thesis possible.

8. References

Bauer, P., & Anzer, G. (2021). *Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction*. Data Mining and Knowledge Discovery. https://doi.org/10.1007/s10618-021-00763-7

Bekkers, J. (2025). *Pressing intensity: An intuitive measure for pressing in soccer* (arXiv:2501.04712v2). arXiv. https://doi.org/10.48550/arXiv.2501.04712

Brownlee, J. (2020, September 16). *ROC curves and precision-recall curves for imbalanced classification*. Machine Learning Mastery. https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/

Carey, D. P., Smith, G., Smith, D. T., Shepherd, J. W., Skriver, J., Ord, L., & Rutland, A. (2001). *Footedness in world soccer: An analysis of France '98*. Journal of Sports Sciences, *19*(11), 855–864. https://doi.org/10.1080/026404101753113804

Chambers, T. (2022, November 6). *Different types of presses in football: 4 variations* (Updated December 28, 2023). Soccer Source Coaching. https://soccersourcecoaching.com/different-types-of-presses-in-football/

Cisco ThousandEyes Engineering. (2025, June 25). *Machine learning-based network event detection*. Medium. https://medium.com/thousandeyes-engineering/machine-learning-based-network-event-detection-0a6560a19c09

Durgapal, A., & Rowlinson, A. (2025). *Pitch basics* (mplsoccer 1.6.1 documentation). Read the Docs. https://mplsoccer.readthedocs.io/en/latest/gallery/pitch_setup/plot_pitches.html#juego-de-posicion

England Football Learning. (01. Feb 2022). *What is pressing in football*. The Football Association. https://learn.englandfootball.com/articles-and-resources/coaching/resources/2022/what-is-pressing-in-football

Fernández Navarro, J. (2018). *Analysis of styles of play in soccer and their effectiveness* [Doctoral thesis, Universidad de Granada]. http://hdl.handle.net/10481/54554

Forcher, L., Forcher, L., Altmann, S., Jekauc, D., & Kempe, M. (2022). The keys of pressing to gain the ball – Characteristics of defensive pressure in elite soccer using tracking data. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2022.2158213

Friedman, Y. (2024, June 13). *Understanding F1 score, accuracy, ROC-AUC, and PR-AUC metrics for models*. Deepchecks. https://deepchecks.com/f1-score-accuracy-roc-auc-and-pr-auc-metrics-for-models/

Genius Sports. (n.d.). *Game-changing technology for sports leagues*. https://www.geniussports.com/technology-for-sports-leagues/

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* (arXiv:2207.08815). arXiv. https://doi.org/10.48550/arXiv.2207.08815

Hudl. (30. Jul 2014). *Defensive metrics: Measuring the intensity of a high press*. https://www.hudl.com/blog/defensive-metrics-measuring-the-intensity-of-a-high-press

Microsoft Corporation. (2025). *Welcome to LightGBM's documentation!* (Version 4.6.0). Read the Docs. https://lightgbm.readthedocs.io/en/stable/

Novillo, Á., Cordón-Carmona, A., García-Aliaga, A., Refoyo Roman, I., López del Campo, R., Resta, R., & Buldú, J. M. (2024). Analysis of player speed and angle toward the ball in soccer. *Scientific Reports*, *14*, Article 11780. https://doi.org/10.1038/s41598-024-62480-7

PySport. (n.d.). *Kloppy: Soccer data processing, simplified* (Version 3.18.0). https://kloppy.pysport.org/

Roberts, A. (2022, September 30). *What is PR AUC?* Arize AI. https://arize.com/blog/what-is-pr-auc/

Scikit-learn developers. (n.d.). *RandomForestClassifier* (scikit-learn 1.8.0 documentation). scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Silva, H., Nakamura, F. Y., Loturco, I., Ribeiro, J., & Marcelino, R. (2024). *Analyzing soccer match sprint distances: A comparison of GPS-based absolute and relative thresholds*. *Biology of Sport*, *41*(3), 223–230. https://pmc.ncbi.nlm.nih.gov/articles/PMC11167479/pdf/JBS-41-52032.pdf

Stats Perform. (n.d.). *Opta data*. https://www.statsperform.com/opta/

UnravelSports. (n.d.). *pressing_intensity.py* [Source code]. GitHub. https://github.com/UnravelSports/unravelsports/blob/main/unravel/soccer/models/pressing_intensity.py

XGBoost developers. (2026, February 9). *XGBoost documentation* (Version 3.2.0). Read the Docs. https://xgboost.readthedocs.io/en/release_3.2.0/

GitHub Public LINK: https://gitlab.com/jonje24-group/Jonje24-project/-/tree/main/MASTER_2026?ref_type=heads

GitHub Reference:
Find the corresponding number, which will provide the path in /Master_2026 for citied notebook OR look in README.md.

1) README.md
2) Approach 2 - Second ML Method/Data_Processing/Master_Full_Season_New_Features.ipynb
3) Approach 1 - First ML Method/Data_Processing/Numerical_Upload_Data.ipynb
4) Video_Validation
5) Approach 3 - Intensity Model/Model/Threshold_Approch_DIFFER_MASTER_PRESS_INTENSITY_MODEL.ipynb
6) Approach 1 - First ML Method/Model/EQUAL_BALANCED_BINARY_MODEL.ipynb
7) Approach 2 - Second ML Method/Model/Less_Features_New_Hyper.ipynb
8) W&B/Chosen_Config

9. Bibliography

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T., & Weber, H. (2017). *Visual analysis of pressure in football. Data Mining and Knowledge Discovery, 31*(6), 1793–1839. http://geoanalytics.net/and/papers/dmkd17.pdf

Bhalodia, S. (2025, Mar 28). *Mastering imbalanced datasets in machine learning: Effective techniques and solutions*. Medium. https://medium.com/@shradhdha.bhalodia/mastering-imbalanced-datasets-in-machine-learning-effective-techniques-and-solutions-c51c6e32a942

Bunker, R. P. (2025). *Machine learning and data mining for sports performance analytics: Frameworks & applications for match outcome prediction and pattern identification.* (Doctoral thesis, Flinders University). https://matthewstephenson.info/papers/students/Machine%20Learning%20and%20Data%20Mining%20for%20Sports%20Performance%20Analytics.pdf

Carling, C., Williams, A. M., & Reilly, T. (2005). *Handbook of soccer match analysis: A systematic approach to improving performance*. Routledge. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=http://ndl.ethernet.edu.et/bitstream/123456789/75970/1/29.pdf&ved=2ahUKEwiG1IKyxeSUAxWVzgIHHeYuBr0QFnoECBcQAQ&usg=AOvVaw26rzfOvaZmlOhrgRouGBfD

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://dl.acm.org/doi/epdf/10.1145/2939672.2939785

Chiarandini, M. (n.d.). *Football analytics: Pointers on football analytics*. Marco Chiarandini — Pointers and Other Stuff. https://imada.sdu.dk/u/march/Blog/football/analytics/2025/12/15/football.html

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). *Actions Speak Louder than Goals.* https://www.researchgate.net/publication/334719246_Actions_Speak_Louder_than_Goals_Valuing_Player_Actions_in_Soccer

Duarte, R., Araújo, D., Correia, V., & Davids, K. (2012). *Sports teams as superorganisms: Implications of sociobiological models of behaviour for research and practice in team sports performance analysis*. Sports Medicine, *42*(8), 633–642. https://doi.org/10.1007/BF03262285

Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). *Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone.* Human Movement Science, *31*(6), 1639–1651. https://doi.org/10.1016/j.humov.2012.03.001

Fernández, J., & Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *Proceedings of the 12th Annual MIT Sloan Sports Analytics Conference*. MIT Sloan Sports Analytics Conference. https://www.researchgate.net/publication/324942294_Wide_Open_Spaces_A_statistical_technique_for_measuring_space_creation_in_professional_soccer

Frencken, W., Lemmink, K., Delleman, N., & Visscher, C. (2011). *Oscillations of centroid position and surface area of soccer teams in small-sided games*. European Journal of Sport Science, *11*(4), 215-223.https://www.researchgate.net/publication/233024646_Oscillations_of_centroid_position_and_surface_area_of_soccer_teams_in_small-sided_games

GeeksForGeeks. (2025, Jul 23). *Cross-validation in machine learning*. GeeksForGeeks. https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/

GeeksForGeeks. (n.d.). *GradientBoosting vs. AdaBoost vs. XGBoost vs. CatBoost vs. LightGBM*. GeeksForGeeks. https://www.geeksforgeeks.org/machine-learning/gradientboosting-vs-adaboost-vs-xgboost-vs-catboost-vs-lightgbm/

Gudmundsson, J., & Horton, M. (2016, Feb 22). *Spatio-Temporal Analysis of Team Sports - A Survey.* https://www.researchgate.net/publication/301844832_Spatio-Temporal_Analysis_of_Team_Sports_--_A_Survey

Graser, A. (2019). *MovingPandas: Efficient structures for movement data in Python.* GI_Forum - Journal of Geographic Information Science, *2019*(1), 54–68. https://doi.org/10.1553/giscience2019_01_s54

hyunsungkim-ds. (2025). *elastic* [GitHub repository]. GitHub. https://github.com/hyunsungkim-ds/elastic

Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Garcia Badaracco, A., Farmer, C., Hjelle, G. A., & Leblanc, F. (2020, Jul 15). *geopandas/geopandas: v0.8.1* [Software]. Zenodo. https://doi.org/10.5281/zenodo.3946761

Jain, M. (2025, Dec 13). *A deep dive into tree-based ML models*. Medium. https://medium.com/mlworks/a-deep-dive-into-tree-based-ml-models-6ed32aee5c88

Memmert, D. (Ed.). (2022). *Match analysis: How to use data in professional sport*. Routledge. https://api.pageplace.de/preview/DT0400.9781000463767_A41909350/preview-9781000463767_A41909350.pdf

Olthof, S. B. H., Frencken, W. G. P., & Lemmink, K. A. P. M. (2015). *The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games*. Human Movement Science, 4*1*, 92–102. https://doi.org/10.1016/j.humov.2015.02.004

Peixoto, G. H. C., Bredt, S. T., Chagas, M. H., Teoldo, I., Greco, P. J., Praça, G. M., Pérez Morales, J. C., & Moreira, P. E. D. (2017). *Tactical behavior in soccer small-sided games: Influence of team composition criteria.* Revista Brasileira de Cineantropometria & Desempenho Humano, *19*(4), 481–493. https://www.researchgate.net/publication/318755881_Tactical_behavior_in_soccer_small-sided_games_Influence_of_team_composition_criteria

Softwarechasers. (2024, Dec 28). *Understanding normalization: Why it matters for regression and classification in machine learning*. Medium. https://medium.com/@softwarechasers/understanding-normalization-why-it-matters-for-regression-and-classification-in-machine-learning-5855962ef44d

Spearman, W. (2018). *Beyond expected goals*. MIT Sloan Sports Analytics Conference Research Papers Competition. https://www.researchgate.net/publication/327139841_Beyond_Expected_Goals

Stats Perform. (n.d.). *How we measure pressure*. Stats Perform. https://www.statsperform.com/resource/how-we-measure-pressure/

Taki, T., Hasegawa, J.et al. (2000). *Visualization of dominant region in team games and its application to teamwork analysis.* https://www.researchgate.net/publication/3852471_Visualization_of_dominant_region_in_team_games_and_its_application_to_teamwork_analysis

ThousandEyes Engineering. (2025, Jun 25). *Machine learning-based network event detection*. Medium. https://medium.com/thousandeyes-engineering/machine-learning-based-network-event-detection-0a6560a19c09

Mihindukulasuriya, H. (2026, Feb 10). Different *Types of decision tree algorithms and their uses*. Creately. https://creately.com/guides/types-of-decision-trees/

Vizuara. (2024) *ML Teach by Doing Lecture 13: Magic of features* [Video]. YouTube. https://www.youtube.com/watch?v=22EC0kdaR8w&list=PLPTV0NXA_ZSi-nLQ4XV2Mds8Z7bihK68L&index=14

Weight&Biases, (n.d.). *Overview of Project Folder: MASTER_SDU.* Personal Account. https://wandb.ai/jonje24-university-of-southern-denmark/projects

xgboost developers (2022). *XGBoost Python model* (Version 2.0.0) [Software documentation]. Read the Docs. https://xgboost.readthedocs.io/en/release_2.0.0/python/model.html

10. Appendix

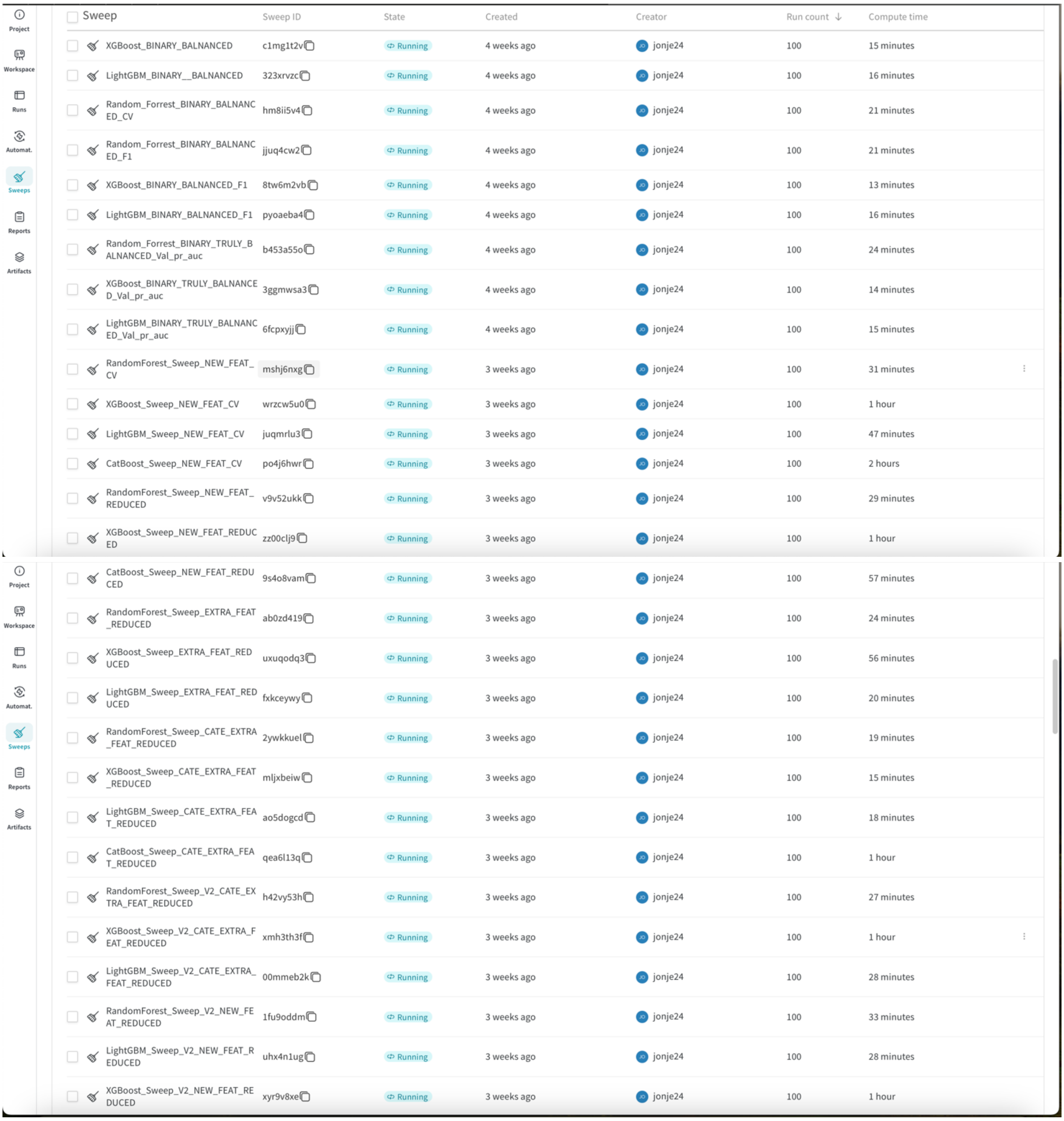
Appendix 1: Feature Categories, Approach 1 & Approach 2.

Feature	Approach	Tactical meaning	Definition	Category
press_centroid_x	1 + 2	Describes the horizontal pitch location of the pressing unit.	Mean x-coordinate of pressing-team players.	Pressing-team location
press_centroid_y	1 + 2	Describes whether the pressing unit is central or wide.	Mean y-coordinate of pressing-team players.	Pressing-team location
opp_centroid_x	1 + 2	Describes where the team in possession is positioned.	Mean x-coordinate of opponent-team players.	Opponent-team location
opp_centroid_y	1 + 2	Describes whether the opponent structure is central or wide.	Mean y-coordinate of opponent-team players.	Opponent-team location
team_length	1 + 2	Measures vertical spread of the pressing team.	Difference between maximum and minimum x-coordinates of pressing-team players.	Pressing-team shape
team_width	1 + 2	Measures horizontal spread of the pressing team.	Difference between maximum and minimum y-coordinates of pressing-team players.	Pressing-team shape
shape_aspect_ratio	1 + 2	Describes whether the pressing shape is elongated or wide.	team_length / team_width, with protection against division by very small width.	Pressing-team shape
def_line_height	1	Captures how high the defensive structure is positioned.	Defensive-line height derived from the pressing team's player positions.	Pressing-team shape
opp_team_length	1 + 2	Measures vertical spread of the team in possession.	Difference between maximum and minimum x-coordinates of opponent players.	Opponent-team shape
opp_team_width	1 + 2	Measures horizontal spread of the team in possession.	Difference between maximum and minimum y-coordinates of opponent players.	Opponent-team shape
opp_shape_aspect_ratio	1 + 2	Describes the opponent's structural shape.	opp_team_length / opp_team_width, with protection against division by very small width.	Opponent-team shape
dist_nearest_1	1 + 2	Measures immediate pressure on the ball carrier.	Distance from the ball to the nearest pressing player.	Ball-pressure proximity
dist_nearest_2	1 + 2	Captures supporting pressure around the ball.	Distance from the ball to the second-nearest pressing player.	Ball-pressure proximity
dist_nearest_3	1 + 2	Captures collective pressure density.	Distance from the ball to the third-nearest pressing player.	Ball-pressure proximity
n_pressers_5m	1 + 2	Measures close-range pressure.	Number of pressing players within 5 metres of the ball.	Ball-pressure proximity
n_pressers_10m	1 + 2	Measures local support around the press.	Number of pressing players within 10 metres of the ball.	Ball-pressure proximity
press_surface_area	1 + 2	Larger values indicate a more spread pressing structure.	Convex-hull area covered by pressing-team players.	Pressing-team compactness
press_stretch_index	1 + 2	Measures how stretched or compact the pressing team is.	Average spatial spread of pressing-team players around their centroid.	Pressing-team compactness
opp_stretch_index	1 + 2	Measures whether the opponent structure is compact or stretched.	Average spatial spread of opponent players around their centroid.	Opponent-team compactness
ball_speed	1 + 2	Captures whether the ball is moving quickly during the press.	Speed of the ball in the current frame.	Physical movement
nearest_presser_speed	1 + 2	Measures the physical intensity of the closest presser.	Speed of the nearest pressing player to the ball.	Physical movement
avg_pressing_team_speed	1 + 2	Captures collective movement intensity.	Mean speed of pressing-team players.	Physical movement
n_sprinting_pressers	1 + 2	Measures how many players are pressing at high speed.	Number of pressing players above the sprint-speed threshold.	Physical movement
closing_speed_nearest	1 + 2	Measures whether the nearest presser is actively closing down the ball.	Change in distance between the nearest presser and the ball across frames.	Physical movement
pressing_intensity	1 + 2	Measures overall pressure applied to the ball.	Sum or aggregation of pressure contributions based on distance, speed, and movement toward the ball.	Physical pressure

ball_x	2	Captures where the press occurs lengthwise on the pitch.	x-coordinate of the ball.	Ball location
ball_y	2	Captures whether the press occurs centrally or wide.	y-coordinate of the ball.	Ball location
ball_dist_to_sideline	2	Smaller values indicate stronger sideline-trap potential.	Distance from the ball to the nearest sideline.	Pitch-zone context
dist_to_goal_line	2	Captures how close the pressing situation is to goal.	Distance from the ball to the nearest goal line.	Pitch-zone context
dist_to_central_axis	2	Measures how central or wide the possession is.	Absolute distance of the ball from the pitch centre line.	Pitch-zone context
centroid_dist_to_sideline	2	Indicates whether the pressing unit itself is shifted wide.	Distance from the pressing-team centroid to the nearest sideline.	Pressing-team pitch context
opp_surface_area	2	Measures how much space the team in possession occupies.	Convex-hull area covered by opponent players.	Opponent-team structure
area_ratio	2	Compares pressing-team spread to opponent-team spread.	press_surface_area / opp_surface_area, with protection against division by very small opponent area.	Relative team structure
vertical_compactness	2	Measures how vertically compact the pressing team is relative to its size.	team_length / number of pressing players.	Pressing-team compactness
compact_pressure	2	Captures whether pressure is concentrated in a compact structure.	pressing_intensity / press_surface_area, with protection against very small area.	Tactical pressure
local_overload_10m	2	Measures whether the pressing team has local numerical superiority.	Number of pressing players near the ball relative to opponent players within a 10-metre radius.	Local overload
passing_lane_occlusion	2	Captures pressure as restriction of passing options.	Estimates how many nearby passing options are blocked by pressing players positioned along passing lanes.	Passing-lane control
escape_difficulty	2	Estimates how difficult it is for the opponent to play out of pressure.	Combination of local_overload_10m and passing_lane_occlusion.	Opponent escape constraint
trap_intensity	2	Captures intense pressure in constrained wide areas.	pressing_intensity / ball_dist_to_sideline, with protection against very small sideline distance.	Sideline trap pressure
speed_differential	2	Measures whether the pressing team is moving faster than the opponent.	Mean pressing-team speed minus mean opponent-team speed.	Relative movement

Source: (Andrienko et al., 2017; Bauer & Anzer, 2021; Bekkers, 2025; Bunker, 2025; Carling et al., 2005; Decroos et al., 2019; Fernández & Bornn, 2018; Frencken et al., 2011; Gudmundsson & Horton, 2016; Memmert, 2022; Peixoto et al., 2017; Spearman, 2018; Taki et al., 2000)

Appendix 2: Weight & Biases, Screenshots.



Sweep	Sweep ID	State	Created	Creator	Run count	Compute time
XGBoost_BINARY_BALNANCED	c1mg1t2v	Running	4 weeks ago	jonje24	100	15 minutes
LightGBM_BINARY__BALNANCED	323xrvzc	Running	4 weeks ago	jonje24	100	16 minutes
Random_Forrest_BINARY_BALNANCED_CV	hm8ii5v4	Running	4 weeks ago	jonje24	100	21 minutes
Random_Forrest_BINARY_BALNANCED_F1	jjuq4cw2	Running	4 weeks ago	jonje24	100	21 minutes
XGBoost_BINARY_BALNANCED_F1	8tw6m2vb	Running	4 weeks ago	jonje24	100	13 minutes
LightGBM_BINARY_BALNANCED_F1	pyoaeba4	Running	4 weeks ago	jonje24	100	16 minutes
Random_Forrest_BINARY_TRULY_BALNANCED_Val_pr_auc	b453a55o	Running	4 weeks ago	jonje24	100	24 minutes
XGBoost_BINARY_TRULY_BALNANCED_Val_pr_auc	3ggmwsa3	Running	4 weeks ago	jonje24	100	14 minutes
LightGBM_BINARY_TRULY_BALNANCED_Val_pr_auc	6fcpxyjj	Running	4 weeks ago	jonje24	100	15 minutes
RandomForest_Sweep_NEW_FEAT_CV	mshj6nxg	Running	3 weeks ago	jonje24	100	31 minutes
XGBoost_Sweep_NEW_FEAT_CV	wrzcw5u0	Running	3 weeks ago	jonje24	100	1 hour
LightGBM_Sweep_NEW_FEAT_CV	juqmrlu3	Running	3 weeks ago	jonje24	100	47 minutes
CatBoost_Sweep_NEW_FEAT_CV	po4j6hwr	Running	3 weeks ago	jonje24	100	2 hours
RandomForest_Sweep_NEW_FEAT_REDUCED	v9v52ukk	Running	3 weeks ago	jonje24	100	29 minutes
XGBoost_Sweep_NEW_FEAT_REDUCED	zz00clj9	Running	3 weeks ago	jonje24	100	1 hour
CatBoost_Sweep_NEW_FEAT_REDUCED	9s4o8vam	Running	3 weeks ago	jonje24	100	57 minutes
RandomForest_Sweep_EXTRA_FEAT_REDUCED	ab0zd419	Running	3 weeks ago	jonje24	100	24 minutes
XGBoost_Sweep_EXTRA_FEAT_REDUCED	uxuqodq3	Running	3 weeks ago	jonje24	100	56 minutes
LightGBM_Sweep_EXTRA_FEAT_REDUCED	fxkceywy	Running	3 weeks ago	jonje24	100	20 minutes
RandomForest_Sweep_CATE_EXTRA_FEAT_REDUCED	2ywkkuel	Running	3 weeks ago	jonje24	100	19 minutes
XGBoost_Sweep_CATE_EXTRA_FEAT_REDUCED	mljxbeiw	Running	3 weeks ago	jonje24	100	15 minutes
LightGBM_Sweep_CATE_EXTRA_FEAT_REDUCED	ao5dogcd	Running	3 weeks ago	jonje24	100	18 minutes
CatBoost_Sweep_CATE_EXTRA_FEAT_REDUCED	qea6l13q	Running	3 weeks ago	jonje24	100	1 hour
RandomForest_Sweep_V2_CATE_EXTRA_FEAT_REDUCED	h42vy53h	Running	3 weeks ago	jonje24	100	27 minutes
XGBoost_Sweep_V2_CATE_EXTRA_FEAT_REDUCED	xmh3th3f	Running	3 weeks ago	jonje24	100	1 hour
LightGBM_Sweep_V2_CATE_EXTRA_FEAT_REDUCED	00mmeb2k	Running	3 weeks ago	jonje24	100	28 minutes
RandomForest_Sweep_V2_NEW_FEAT_REDUCED	1fu9oddm	Running	3 weeks ago	jonje24	100	33 minutes
LightGBM_Sweep_V2_NEW_FEAT_REDUCED	uhx4n1ug	Running	3 weeks ago	jonje24	100	28 minutes
XGBoost_Sweep_V2_NEW_FEAT_REDUCED	xyr9v8xe	Running	3 weeks ago	jonje24	100	1 hour

Random_Forrest_TOP15_BINARY_TRULY_BALNANCED_Val_pr_auc	g14rytl1	Running	3 weeks ago	jonje24	100	18 minutes
XGBoost_BINARY_TOP15_TRULY_BALNANCED_Val_pr_auc	r25vvj66	Running	3 weeks ago	jonje24	100	9 minutes
LightGBM_TOP15_BINARY_TRULY_BALNANCED_Val_pr_auc	pntoyovw	Running	3 weeks ago	jonje24	100	9 minutes
RandomForest_Sweep_V2_EXTRA_FEAT_REDUCED	s32j94pw	Running	3 weeks ago	jonje24	100	35 minutes
XGBoost_Sweep_V2_EXTRA_FEAT_REDUCED	jqklkxtd	Running	3 weeks ago	jonje24	100	3 hours
LightGBM_Sweep_V2_EXTRA_FEAT_REDUCED	akegma2k	Running	2 weeks ago	jonje24	100	45 minutes
Random_Forrest-Sweep	4y7nb4yn	Running	4 weeks ago	jonje24	50	42 minutes
XGBoost-Sweep	zwa5707b	Running	4 weeks ago	jonje24	50	14 minutes
LightGBM-Sweep	sndcoq11	Running	4 weeks ago	jonje24	50	18 minutes
Random_Forrest-Sweep_PR	2uhb8jv2	Running	4 weeks ago	jonje24	50	47 minutes
XGBoost-Sweep_PR	p5e7n9ia	Running	4 weeks ago	jonje24	50	18 minutes
LightGBM-Sweep_PR	vzp7vtu3	Running	4 weeks ago	jonje24	50	15 minutes
Random_Forrest-Sweep_PR_2.0	bxhmdhb1	Running	4 weeks ago	jonje24	50	38 minutes
XGBoost-Sweep_PR_2.0	omy0gwo4	Running	4 weeks ago	jonje24	50	19 minutes
LightGBM-Sweep_PR_2.0	ecnrumhu	Running	4 weeks ago	jonje24	50	14 minutes
XGBoost_Multiclass_SWEEP_F1	uhcu6cm4	Running	4 weeks ago	jonje24	50	22 minutes
Random_Forrest-Sweep_3.0	7nswclbd	Running	4 weeks ago	jonje24	50	42 minutes

Random_Forrest-Sweep_PR_2.0	bxhmdhb1	Running	4 weeks ago	jonje24	50	38 minutes
XGBoost-Sweep_PR_2.0	omy0gwo4	Running	4 weeks ago	jonje24	50	19 minutes
LightGBM-Sweep_PR_2.0	ecnrumhu	Running	4 weeks ago	jonje24	50	14 minutes
XGBoost_Multiclass_SWEEP_F1	uhcu6cm4	Running	4 weeks ago	jonje24	50	22 minutes
Random_Forrest-Sweep_3.0	7nswclbd	Running	4 weeks ago	jonje24	50	42 minutes
Random_Forrest-Sweep_4.0	ghq4wn68	Running	4 weeks ago	jonje24	50	34 minutes
LightGBM-Sweep_4.0	2hgf3hlf	Running	4 weeks ago	jonje24	50	19 minutes
XGBoost-Sweep_4.0	hktk3g6o	Running	4 weeks ago	jonje24	50	16 minutes
Random_Forrest-Sweep_NEW_FEAT	4b2vbh2g	Running	4 weeks ago	jonje24	50	44 minutes
XGBoost-Sweep_NEW_FEAT	b8vrfqdg	Running	4 weeks ago	jonje24	50	27 minutes
LightGBM-Sweep_NEW_FEAT	vbo2yacf	Running	4 weeks ago	jonje24	50	14 minutes
Random_Forrest_BINARY_1.1	quy8a802	Running	4 weeks ago	jonje24	50	38 minutes
XGBoost_BINARY_1.1	82uxvyly	Running	4 weeks ago	jonje24	50	29 minutes
LightGBM_BINARY_1.1	4ewgytnv	Running	4 weeks ago	jonje24	50	7 minutes
Random_Forrest_BINARY_BALNANCED	0mxdinyu	Running	4 weeks ago	jonje24	50	11 minutes
XGBoost_BINARY_BALNANCED_CV	4fvmlhvi	Running	4 weeks ago	jonje24	50	7 minutes
LightGBM_BINARY_BALNANCED_CV	11fjdsqk	Running	4 weeks ago	jonje24	50	6 minutes
XGBoost_BINARY_BALNANCED_F1	k6hrei2b	Running	4 weeks ago	jonje24	50	8 minutes

Source: (Weight&Biases, n.d.)