

Mining football players' behavioral profile: identifying candidate proxy features from event data.

Soccer Track
844641

1. Introduction

Understanding individual athletic performance in soccer is not trivial. When we watch a player scoring a goal, we cannot quantify the precise amount of the underlying technical, tactical, physical, and psychological skills that are involved. Rather than directly influencing performance *per se*, those skills are heavily intertwined, so it is impossible to isolate the influence of each individual set of skills and quantify how much of performance is caused by one or another [1]. From the point of view of Sport Psychology, using huge event and tracking datasets can be fruitful and epitomize a substantial increase in observational power as long we keep in mind that the data contained in it belongs to the realm of behavior. In other words, the data is about performed behaviors, not the hidden skills that make it possible. Thus, those skills can be inferred but not measured as absolute entities.

Therefore, we can question the value of event/tracking data to answer questions like: How can we identify the best-performing players under competitive pressure? Can we track a player's ability to perform under competitive pressure across time? Or, to put it differently, how is my player evolving regarding his ability to cope with normative contextual stress?

Following the steps of Bransen et al. [2], we argue that, under the bold assumption that both competitive contextual pressure and the human ability to deal with it vary normatively, it is possible to assign a relative value to football players' ability to cope with normative contextual pressure. Moreover, we suggest that such a metric has practical worth for talent identification and development.

In that sense, this paper will present our approach to modeling normative contextual pressure from event data. In the sections below, we will report the data preparation jobs to generate the features required for our modeling, present the theoretical background inspiring it, and provide a use case in which we used unsupervised machine learning methods to identify natural groups of football players with different profiles regarding performing under normative contextual pressure.

1.1. – Theoretical Fundamentals Regarding Competitive Stress

Whether it is reasonable to presume, like in Bransen et al. [2], that contextual pressure (or 'mental pressure' as named by the authors in the original paper) is a dynamic construct influenced by pre-game and in-game factors, it is advised to understand why it is dynamic and also why, by only

using behavioral data (i.e., described in event streams), we can speak of contextual stress as a relative descriptor rather than as an absolute measure. Therefore, following the old adagio that “there is nothing more practical than a good theory” [3], it is substantial to establish a theoretical foundation to ground the modeling effort.

Since our attention is devoted to the performance under pressure topic, we root our analysis of football players’ performing profiles under varying levels of normative contextual pressure in the Theory of Stress and Coping [4], which was already explored mainly within the domain of Sport Psychology (e.g., [5, 6]).

The Theory of Stress and Coping primary postulation is that stress or anxiety is a relational process mediated by subjective cognitive appraisals rather than a direct response to external stimuli (see [4]). As depicted in Figure 1, primary and secondary appraisals mediate the coping response on a transactional basis, in which the individual judges or appraises situations (primary cognitive appraisals) and their ability to cope (secondary cognitive appraisals) with the demands imposed by them in an ongoing fashion.

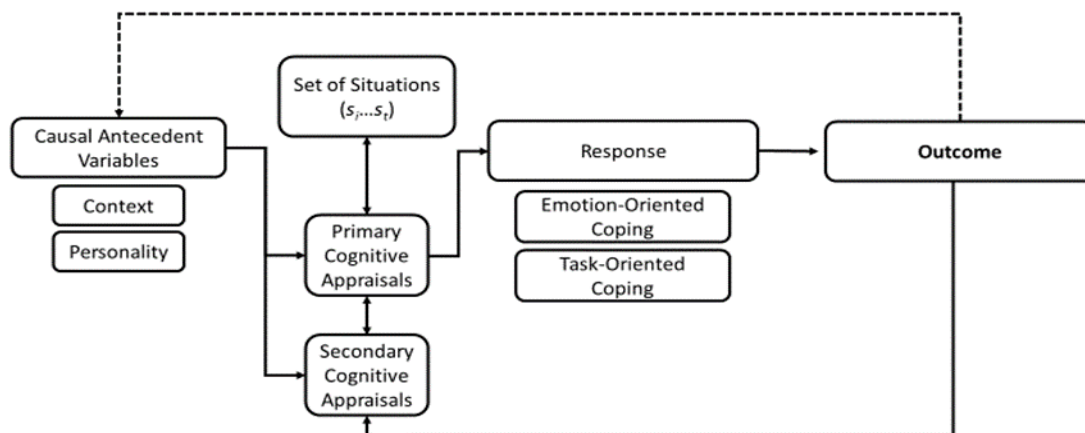


Figure 1 – An overview of Lazarus and Folkman’s (1984) Model of Stress and Coping.

Therefore, as stated by Cruz [7], “anxiety as an emotional response only happens when athletes perceive a threat or, in other words, when they appraise situations as meaningful and significant for their “ego” or subjective wellbeing while, at the same time, they judge their resources to cope or deal with such situations as insufficient” ([7], p.30).

It is essential to notice that, contrary to Lazarus and Folkman [4], rather than focusing on anxiety as a personal experience and according to our aim, we are herein looking at contextual pressure as an environmental variable. We understand that the experience of anxiety is highly idiosyncratic and cannot be grasped from event data only since it depends on each player's cognitive and affective

personal background, so what we are looking for is the effects of contextual pressure on individual performance that can be moderated/facilitated by anxiety and other emotions. In this sense and within this study, we recognize pressure as a dynamic situational property that may characterize football players' performance rather than an individual feature.

2. Data Preparation

Given our primary goal of creating a metric or a group of metrics that capture football players' relative ability to perform under normative contextual pressure, several data mining goals must be accomplished. Figure 2 summarizes all the preprocessing jobs we executed before modeling normative contextual pressure using the FiveThirtyEight and the Public WyScout data.

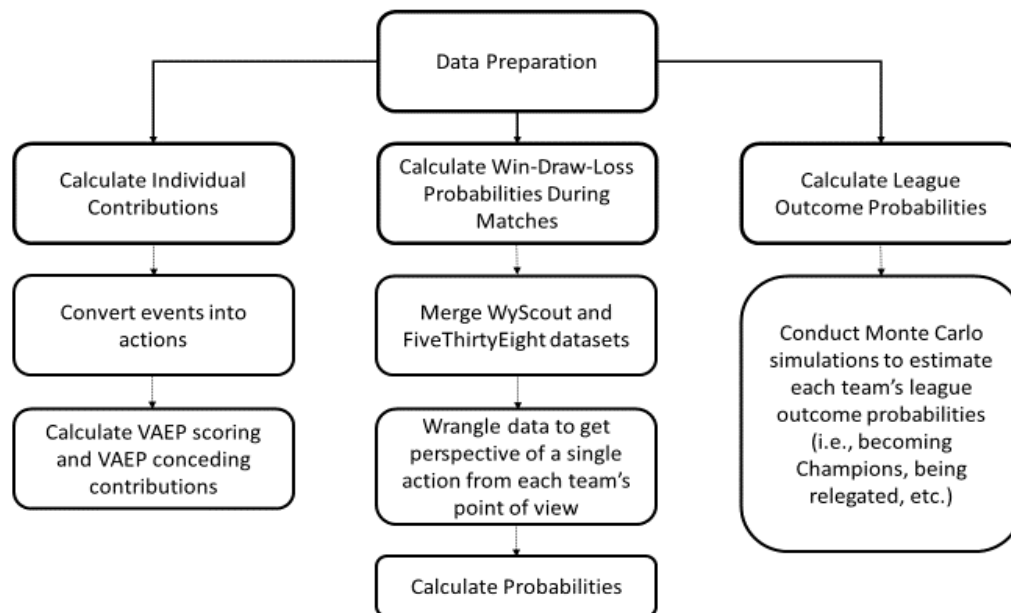


Figure 2 – Scheme of data preparation tasks

2.1. Calculate Individual Contributions

Since soccer is a game driven by two primary objectives (i.e., scoring and not conceding goals), examining the probabilities of scoring and not conceding goals associated with each performed action is a viable approach. Therefore, we used Valuing Actions by Estimating Probabilities (VAEP) [3] method to compute the scoring and conceding scores associated with each action. In sum, this

method extends the idea of expected goals to actions other than shots by calculating the variation in scoring and conceding probabilities between consecutive actions given three different types of features: simple event data descriptors (e.g., actions start and end locations, actions start and ending times, kind of actions), complex features derived from several attributes within and across subsequent actions (e.g., distance and angle to the goal, distance covered during the actions in both x and y directions), and contextual features containing information regarding the actual game state (e.g., the goal difference).

The assignment of a value (probability of scoring/not conceding) can be described formally by Equation 1:

Given: an on-the-ball action a_i

Do: learn a function that assigns a value $V(a_i)$ to the action. Since every on-the-ball action alters the game state (s) consecutively, then:

$$V(a_i) = V(s_i) - V(s_{i-1}) \quad (1)$$

If the action (the distance between s_{i-1} to s_i) positively affects the respective team's scoring odds, it shall be valued positively. Otherwise, it should be negatively valued. Then, as stated in Equation 2:

$$V(s_i) = P_{scores}(s_i) - P_{scores}(s_{i-1}) \quad (2)$$

A further parameter (k) can be added to this equation, explicitly stating the number of succeeding actions considered when calculating the goal-scoring/conceding probabilities for s_i , which adds further context to the computation of the associated probabilities. Indeed, since soccer is a territorial game in which space or pitch control makes a significant difference in team performance, harmonizing the computation by the background play within each action occurs shall more accurately assign a truthful value to each action. Therefore, we end with Equation 3:

$$V(s_i) = (\Delta P_{scores}(^k s_i)) + (- \Delta P_{concedes}(^k s_{i-1})) \quad (3)$$

Several steps were followed to calculate individual contributions. First, we loaded the game events from Public WyScout API and divided our data into training and testing conditions. We used data from the 2017-2018 season regarding the German Bundesliga, the English Premier League, France Ligue 1, and Spanish La Liga for the training condition. Data from the Italian Calcio A from the same season was chosen for the testing condition. Due to computational restrictions, we limited the

number of features used to calculate VAEP scores to the previous three actions (i.e., $k=3$), unlike the original implementation [2, 3].

We used the Random Forest Classifier (RFC) to estimate the associated probabilities of scoring and conceding goals and calculated two scores: an execution VAEP score and a decision VAEP score. These scores differ because the execution score considers the outcome of each action (i.e., successful or unsuccessful) as a training feature. In contrast, the decision score does not (i.e., is outcome unaware). Such a distinction regards our supposition that while the first is a better proxy for execution quality, the second would capture aspects more related to decision quality. This is a relevant distinction because, although connected, execution and decision are separate constructs. For instance, a player may make a decision that would contribute much to his team's probability of scoring but execute poorly. Conversely, he can perform a highly-skilled pass that contributes very little to his team's likelihood of scoring a goal shortly.

Regarding the parameterization of the RFC, we used 100 trees as the number of estimators' parameter and set the minimum number of samples required to split an internal node to 50. We do not carry any optimization procedure (e.g., grid search, Bayesian optimization) because we knew beforehand that those were the best parameters to tune the model [4]. The results of the classification model are presented in Table 1 in Appendix A.

2.2. – Calculate In-Game Win-Draw-Loss Probabilities

Given that contextual pressure is dynamic and influenced by both antecedent and current events, besides the tension that stems from the expectation of confronting an opponent team with an associated value of relative strength that we may describe as naturally intrinsic to each game, we also need to consider how pressure mounts or lowers as each game evolves (e.g., goals are scored, yellow and red cards are given, etc.).

We addressed this problem by estimating how win-draw-loss probabilities evolved as games progressed. As illustrated in Figure 2, we first merged data from the Public WyScout and the FiveThirtyEight datasets to join within the same framework the required features to calculate the abovementioned probabilities. In the last dataset, we had information about the game's importance and each team's strength indexes as calculated by FiveThirtyEight (these features are explained in Appendix B). At the same time, the first contained the event data streams describing all the actions. Since every action can change the game outcome, we assumed that every action could also alter the level of contextual pressure during the game.

Since football events have contrasting meanings for each competing team, we had to transform the data to obtain each team's perspective on the ongoing result to calculate the win-draw-loss probabilities for each team. For instance, a scored goal means getting closer to winning the game for the scoring team and the opposite for the team conceding the goal.

The probability of each team winning, drawing, or losing was estimated using a multiclass classification model with each team's strength index (i.e., FiveThirtyEight's Soccer Power Index), the current time, the current score, and visiting/visited conditions as features, and the result (win,

draw, and loss) as the labels. Working under Python's Sklearn standard parameterization and using a 10-fold cross-validation approach, we have observed that the RFC evidenced a better classification performance in terms of accuracy and a similar standard deviation compared to the other algorithms we tested (see Appendix C for more detailed results).

2.3. – League Outcome Probabilities

Finally, we conducted Monte Carlo simulations to calculate the league outcome probabilities for the target season (i.e., the likelihood of becoming Champions, qualifying for the Champions League, qualifying for the Europa League, or becoming relegated). Taking advantage of the Law of Great Numbers, we repeatedly generated simulated league tables for each round from the FiveThirtyEight winning, drawing, and losing odds to estimate the outcome, as mentioned earlier.

3. Modeling Normative Contextual Pressure

As in Bransen et al. [2], we modeled normative contextual pressure by assuming that it is a composed measure that can be operationalized as the intersection between the pre-game (or antecedent) and in-game (or current) contextual pressure. The former relates to the presumed factors likely to influence cognitive appraisal processes before a match begins. At the same time, the latter corresponds to the variation of pressure as events in the game succeed each other and keep updating a given team's odds of winning, drawing, or losing the game. Formally, from Equation 4, we have:

$$\text{Normative Contextual Pressure} = \text{Pre - Game Pressure} * \text{In - Game Pressur} \quad (4)$$

3.1. – Modeling Pre-Game Contextual Pressure

Significantly, unlike Bransen et al. [2] and based on Sport Psychology literature [4, 7], we considered that the level of confidence of a given player could influence the level of pressure a player can feel before a match begins. In short, the level of confidence, or the degree to which the player appraises himself as a more or less competent performance, largely influences stress appraising processes.

We modeled pre-game normative contextual pressure considering the game's importance for the statistical outcome of the two confronting teams, the closeness between each team's strength indexes, and the player's confidence levels according to Equation 4.

$$Pre - Game Pressure = Importance + Closeness + Confidence \quad (5)$$

For the closeness factor, following an Elo-based logic, we conjectured that games between teams with closer strength indexes would tend to be perceived by the players as more stressful than games for which that difference was higher given that the unpredictability about the outcome of the game is also higher. According to Lazarus and Folkman [4] and Cruz [7], unpredictability or uncertainty is a facilitative stress factor. In sum, we considered the following arguments when modeling pre-game normative contextual pressure:

Importance: the importance of the game for each team participating in the match for the season's outcome (see Appendix C).

Closeness (between team strength indexes): how apart are each team's strength indexes using the normalized difference between the teams' Elo ratings, obtained by Equation 4.

$$Closeness = 1 - norm(|SPI_{home} - SPI_{away}|) \quad (6)$$

Confidence: the normalized inverse of the (rolling) average VAEP scores for the past month obtained by Equation 5:

$$Confidence = 1 - norm(\text{mean}_{last\ 30\ days}(VAEP)) \quad (7)$$

3.2- Modeling In-Game Contextual Pressure

Herein, we have assumed that in highly disputed or tighter matches, the pressure tends to mount until the end of the game. The opposite occurs when the scoring difference is too wide for the losing team to recover and get back into the game. Therefore, we devised a measure of a game's tightness to model the in-game normative contextual pressure. Our approach took advantage of the win-draw-loss probability estimation and the normalized standard deviation between the three probabilities; basically, the smaller the standard deviation, the tighter the matches. After normalizing the values between 0 and 1, we can define in-game contextual pressure as defined in Equation 7:

$$In - Game Pressure = 1 - norm(std([P_{win}, P_{draw}, P_{lose}])) \quad (8)$$

The correlation matrix in Figure 3 shows the correlation between all the pressure metrics. The fact that the in-game pressure metric is more correlated to the normative contextual pressure metric makes sense since the former depends on the win-draw-loss probabilities during the game, which already considers each team's value and the current score. Intuitively, the pre-game pressure metric is a more "hypothetical" kind of pressure because it does not acknowledge the unfolding events of a game; likewise, its influence on normative contextual pressure is more limited.

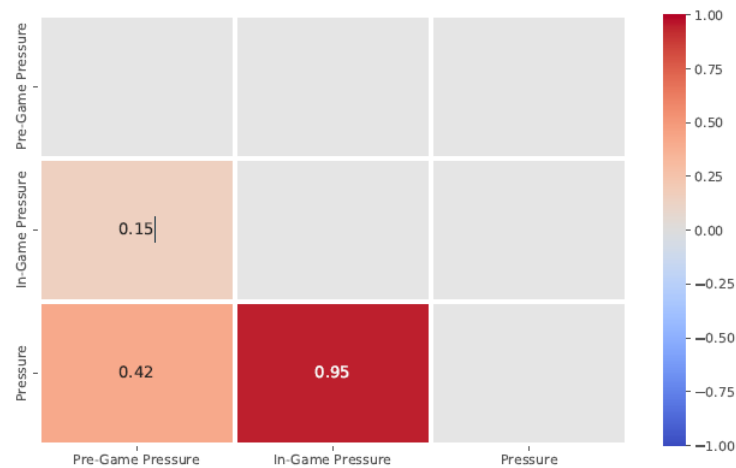


Figure 3 – Correlation Matrix for the several normative contextual pressure metrics

3.4- Discretizing Normative Contextual Pressure

Unlike Bransen et al. (2019), which opted for relativist criteria to define the intervals corresponding to different levels of pressure, we opted for a more business-oriented approach. Accordingly, we used more stringent standards instead of simply dividing the distribution so that the 20 percent lower pressure values correspond to the category low-pressure and the 20 percent higher to higher pressure.

High-pressure situations in a soccer match are rarer than 20%, and many actions captured in event streaming data do not involve high-risk situations. For instance, about half of the events recorded in this data set are passes (see Appendix D), and many of the passes performed in a game do not imply a high level of risk (e.g., those between two center backs when the team is in an offensive organization). Hence, by inspecting the distributions in a histogram (see Figure 5), we discretized pressure by setting the following categories:

Low Pressure: pressure values below percentile 65.

Medium Pressure: pressure values between percentile 65 and 91

High Pressure: pressure values above the percentile 92.

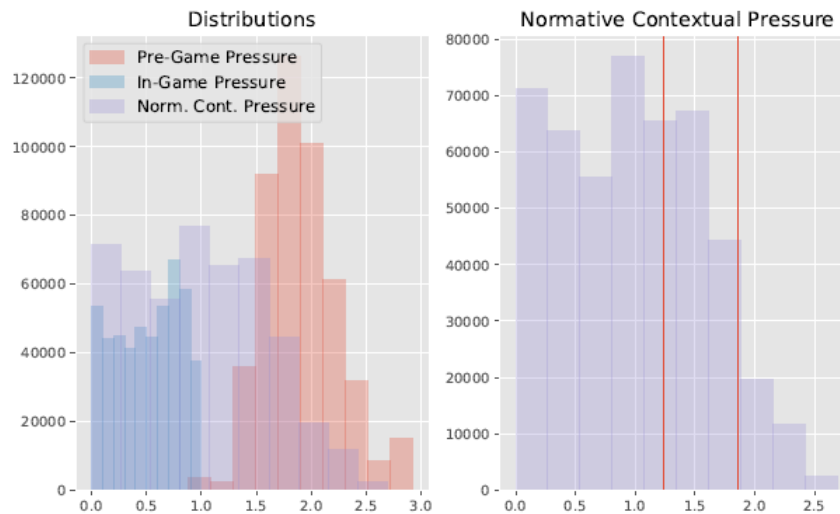


Figure 5: On the left are pre-game, in-game and normative contextual pressure histograms. In the right, normative contextual pressure histograms with vertical lines signaling the percentile values (65, 92) used for discretization in low, medium, and high-pressure situations

As reported in Tables 1 and 2, the average values for VAEP scoring execution and VAEP decision execution follows the same pattern. They are slightly higher for moderate-pressure situations than low-pressure situations and sharply decrease for high-pressure situations. This pattern narrowly mimicked the Yerkes and Dodson law (Yerkes et al., 1908) for the relationship between stress and performance and was expected. The Yerkes and Dodson law states that the stress and performance relationship follows an inverted U-shaped curve, according to which performance tends to be lower for low and high-stress situations than for moderate (or optimal) stress levels. Furthermore, the results indicate that execution and decision are impaired in high-pressure situations.

Table 1: VAEP scoring execution average values and considered several instances after defining intervals for low, medium, and high-pressure levels

	VAEP Scoring Execution		
	Low Pressure	Moderate Pressure	High Pressure
Avg Performance	0,00296	0,00299	0,00251
Number of Instances	311035	129199	38282

Table 2: VAEP scoring average decision values considering the number of instances after defining intervals for low, medium, and high-pressure levels.

VAEP Scoring Decision			
	Low Pressure	Moderate Pressure	High Pressure
Avg Performance	0,0041	0,00413	0,0038
Number of Instances	311035	129199	38282

Given that event stream data is clearly unbalanced towards attacking actions, we inspected how pressure influenced performance on two types of these actions: take-one and shots. Although, the two kinds of action suffer from increased contextual anxiety. As shown in Figure 6, pressure tends to be especially detrimental to shot performance. On average, the shooting performance is much lower for high-pressure situations than for lower and moderate-pressure conditions.

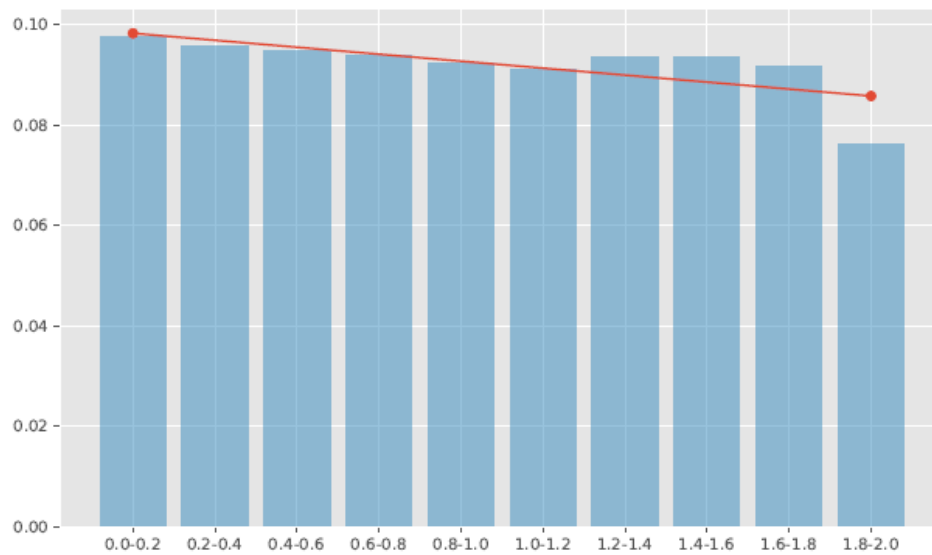


Figure 6: VAEP execution scores according to normative contextual pressure for shots

3.5- Exploring the Performance Under Pressure Profile of the Top Scorers and the Best-Rated Players for the Target Season

To examine the explainability power of our pressure metric, we explored how the top scorers and the best players from Italian Serie A for the 2017-2018 season performed in situations with low, moderate, and high-pressure levels. While always speculative, we can identify different performance profiles by looking at these Figures, particularly when attending to additional information regarding each player's career evolution or other relevant developmental and demographic data.

In the case of top scorers (see Figure 7), high pressure seems to be generally

detrimental to performance. However, there were two cases in which two players, Higuaín and Quagliarella, performed more valuable actions when the pressure was high. Curiously, these players were two veteran international-level players aged over 30 and 34 years old, respectively. It can be the case (as it is very likely) that players may learn how to cope better with high moments of pressure as their careers unfold. The issue with Giovanni Simeone is the opposite; he was a young talented player at 22 years old, and as pressure mounted, his performance declined. In the last few years, he did not confirm all the potential that he early has shown up. Regarding the best players (see Figure 8), we can see that several players have contributed the most to the probability of their team scoring during high-pressure situations. The case in which a decline in high-pressure situations was most evident was that of J. Ilicic - a player whose career in the last years has been reportedly affected by psychological issues.

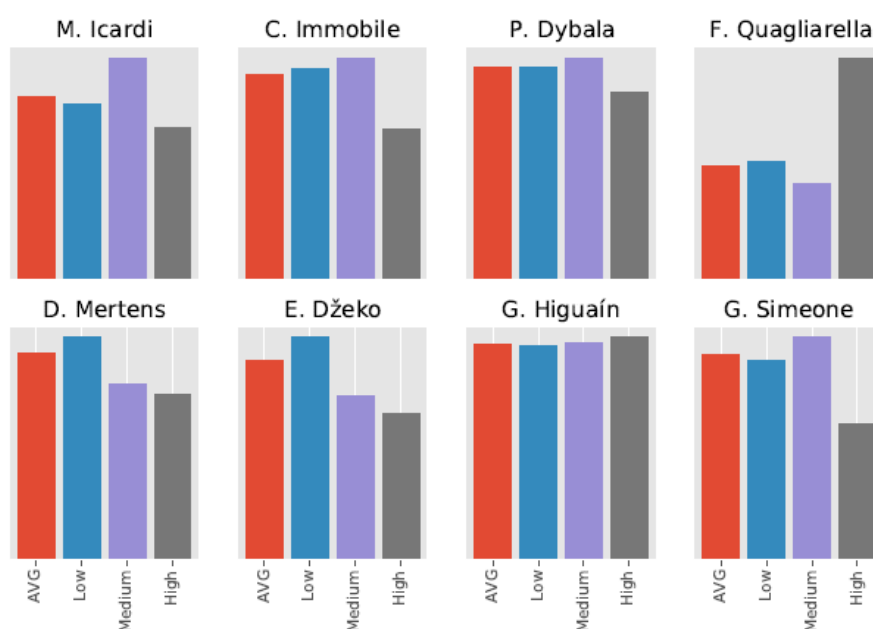


Figure 7 – VAEP execution scores for the top scorers in the target league for low, medium, and high-pressure situations.

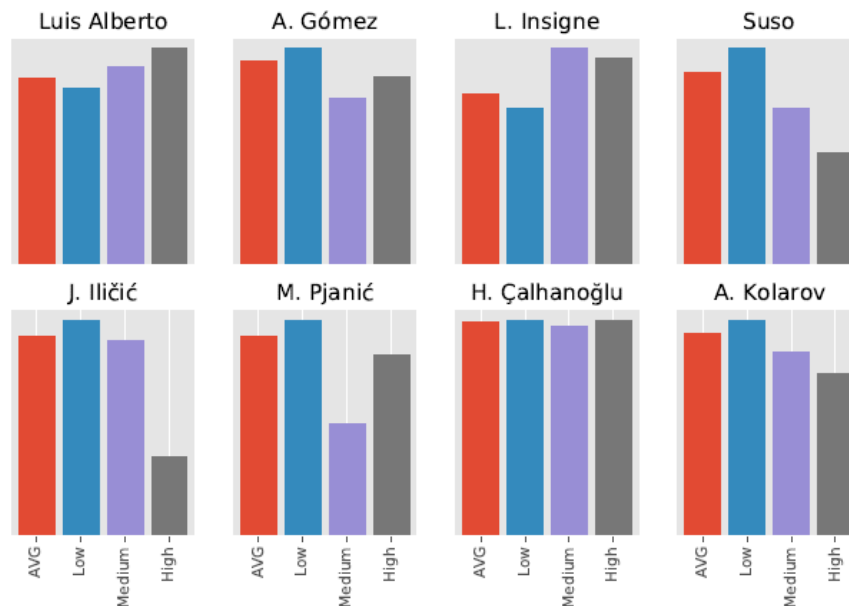


Figure 8 – VAEP execution scores for the best-rated players in the target league for low, medium, and high-pressure situations.

4. Use Case: Finding natural groups of players regarding their performance under pressure profiles

One of the fundamental processes within football clubs is talent identification and recruitment. However, talent identification is not a trivial task because the concept of talent is itself complex. Investigation sport sciences suggest that rather than the sum of technical, tactical, physical, and psychological skills, a talented football player exhibits the right balance of them [1]. Nevertheless, given all the attention devoted to football, a new player that does not fit the team rapidly becomes noticed and criticized, which inflicts tremendous pressure on all those involved in the talent identification and recruitment process.

Herein, we used our pressure metric to inform player recruitment about the ability of football players to cope with normative contextual stress. Regarding this use case, we have followed an unsupervised machine learning approach to identify groups of distinct players concerning their performance under low, moderate, and high-pressure levels.

Specifically, we have used the K-Means algorithm using the predefined 'sklearn' criteria to cluster football players regarding their performance at low, moderate, and high-pressure levels. We used

the elbow method to identify the most appropriate number of clusters (see Appendix E), with 5 being the number of clusters indicated by the method.

While Table 3 provides the average values for each cluster given the features of low, moderate, and high pressure, Figure 9 affords a more direct visualization of the identified profiles.

Cluster 1 comprises players that perform more valuable actions in low-pressure conditions, with similar VAEP scores for both moderate and high-pressure situations. Cluster 2 groups players that perform much better in high-pressure situations and are likely to be substitute players thrown into tight matches that eventually make an assist or score a goal. Cluster 3 contains players with very acceptable scores for both low and moderate levels categories but that seem to choke under pressure; as pressure continuously mounts, their performance level decreases. Cluster 4 contains the players whose performances follow the typical inverted U-shape with better scores in moderate-pressure situations. Finally, cluster 5 holds players whose performance increases as pressure mounts, with very acceptable values for all the varying levels of pressure.

Table 3 - Frequency and average values for the low, medium, and high-pressure

	Cluster 1 n = 59	Cluster 2 n=61	Cluster 3 n=54	Cluster 4 n=46	Cluster 5 n=56
Low	0,9990	0,3860	0,8716	0,3272	0,6650
Moderate	0,3443	0,3187	0,9314	1	0,8421
High	0,4671	1	0,3510	0,3352	0,9134

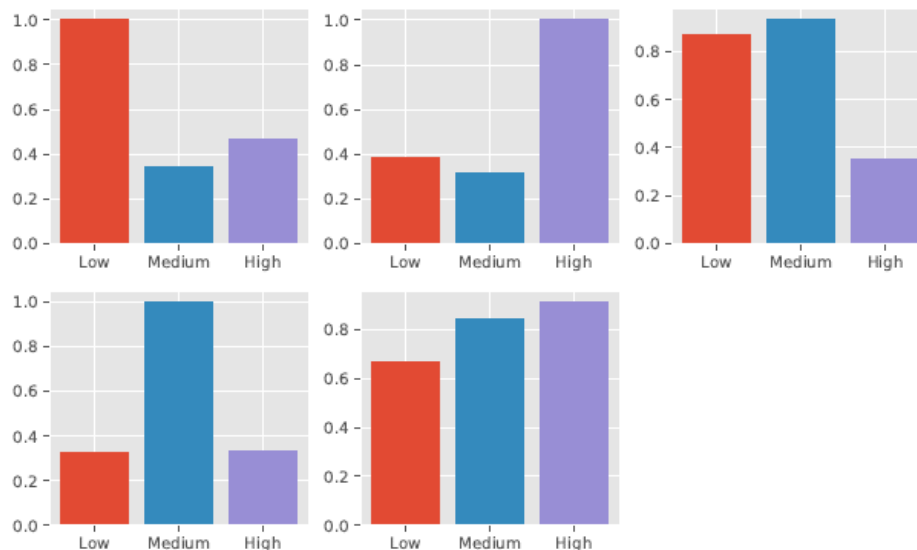


Figure 9 - Graphical representation of the clusters

Conclusions

In general terms, we think we provide a fair reflection regarding how Psychology sees football players' performance, particularly regarding the influence of normative contextual pressure and how the psychological processes embedded in stressful generations occur.

By the way, those processes were considered when modeling normative contextual pressure, the second aim. We made a considerable effort of data wrangling and feature engineering to calculate the features required to model normative contextual stress in such a way that it accorded one of the most prominent theories of stress - the Transactional Model of Stress by Lazarus and [4]. Although future improvements are undoubtedly welcome, our work represents a step ahead of the only known similar work [2].

Rather than presenting a solid and polished product, we wanted to demonstrate that we could use raw data (i.e., event streams) to generate valuable knowledge regarding something as abstract as football players' psychological profiles. Furthermore, we wanted to show that such knowledge could inform decision-making at a club level (e.g., talent identification and development). Indeed, our model can be improved, and Machine Learning methods other than unsupervised knowledge can and should be tested in the future.

Limitations and Future Work

Despite its merits, this work also has several significant limitations that shall be acknowledged. In the first place, the available data to train the models we used to generate features was limited. We only had data from one season for five different leagues. Thus, we needed data from competitions other than the target league to use machine learning methods and generate the features.

Ideally, we would rather have a historical data set in which we would use data from the past to train the models to be tested in current data. This would prevent several putative cultural biases regarding the natural differences between each league. Furthermore, it would have enabled us to use different data-mining approaches. For instance, we could have followed a time series paradigm and inspected how vulnerability to stress evolved over extended time frames. We could even have studied how coping ability changed throughout a player's career and confirm/infirm the hypothesis that more experienced football players are better than novices at dealing with normative competitive stress.

Due to the extensive time required to treat the raw data and make it manageable to model normative contextual pressure, we needed more time to optimize/tune each model we used during our feature engineering process. In the future, such an optimization/tuning shall be performed using grid search or Bayesian optimization methods.

Finally, it would be interesting to investigate how normative contextual pressure relates to subjective measures of stress, such as the threat/challenges perception scales [7]. More than providing us with a "ground truth" to evaluate the external validity of our metrics since they may

estimate different things, it could be interesting to study in which cases normative pressure correlates with perceived stress and the cases in which that doesn't happen. Do the players performing better in high-pressure situations perceive them as stressful? It is an open question.

References

- [1] Meylan, Cesar, et al. "Talent identification in soccer: The role of maturity status on physical, physiological and technical characteristics." *International Journal of Sports Science & Coaching* 5.4 (2010): 571-592.
- [2] Bransen, Lotte, et al. "Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure." *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. MIT SLOAN; <http://www.sloansportsconference.com/wp-content/uploads/2019/02/Choke-or-Shine-Quantifying-Soccer-Players-Abilities-to-Perform-Under-Mental-Pressure.pdf>, 2019.
- [3] Lewin, Kurt. "Field theory in social science: selected theoretical papers (Edited by Dorwin Cartwright.)." (1951).
- [4] Lazarus, Richard S., and Susan Folkman. *Stress, appraisal, and coping*. Springer publishing company, 1984.
- [5] Dias, Cláudia, José F. Cruz, and António Manuel Fonseca. "The relationship between multidimensional competitive anxiety, cognitive threat appraisal, and coping strategies: A multi-sport study." *International Journal of Sport and Exercise Psychology* 10.1 (2012): 52-65.
- [6] Hammermeister, Jon, and Damon Burton. "Stress, appraisal, and coping revisited: Examining the antecedents of competitive state anxiety with endurance athletes." *The Sport Psychologist* 15.1 (2001): 66-90.
- [7] Cruz, José Fernando A. "Stress e ansiedade na competição desportiva: natureza, efeitos e avaliação." (1996).

Appendix

A. VAEP scores estimation

Table 1. Classification Model Results for VAEP scores estimation

		Train			Test		
		R2	MAE	MedAE	R2	MAE	MedAE
execution	scoring	0.339	0.023	0.008	0.144	0.026	0.009
	conceding	0.224	0.008	0.002	0.030	0.009	0.002
decision	scoring	0.256	0.026	0.008	0.040	0.029	0.010
	conceding	0.214	0.009	0.002	0.021	0.009	0.002

We can observe that the Coefficient of Determination (R^2) score decreased from the training to the testing condition, which was expectable, and also that these values are low. In ordinary situations, such low scores would mean that the model cannot explain (or you only could explain very little) the relationship between the variance of the features and the target variables. However, the R^2 score is probably not the best metric to evaluate probability models regarding soccer data [4]. Concretely, since goals are sporadic events that significantly influence the calculation of the VAEP scores, the R^2 score is much more sensitive to the variance, and the Mean Absolute Error (MAE,) and the Median Absolute Error (MedAE) may be better estimators. These values are pretty low and vary very little from the training to the testing conditions.

B. Explanation of FiveThirtyEight's Soccer Power Index (SPI) and Importance Metrics

The SPI rating is a team's overall strength metric obtained through an attacking and a defensive rating. The offensive rating corresponds to the number of goals a team is expected to score against an average team in a neutral field. At the same time, the defensive consists of the number of goals a team is likely to concede within the same condition. Given the SPI rating of two confronting teams, the probable result of a match between them can be predicted according to an Elo-model fashioned way (i.e., by direct pairwise comparison).

Though, as the season unfolds, the ratings are adjusted after each match based on the team's performances, the SPI rating is calculated as defined in Equation 3 before the beginning of a season:

$$SPI_{Pre-season} = \frac{2}{3} * SPI_{Previous Season} + \frac{1}{3} * SPI_{Market Value} \quad (4)$$

In its turn, importance measures how much a particular match outcome will affect the team's statistical outlook on the season. FiveThirtyEight calculates the importance of a game by generating probabilities for each factor conditional on winning (or losing) the match and then finding the difference between those two possible numbers. The factor with the maximum range of the difference is taken, and the result is normalized between 0 and 100. Formally, the importance of a game for a team is given by Equation 5:

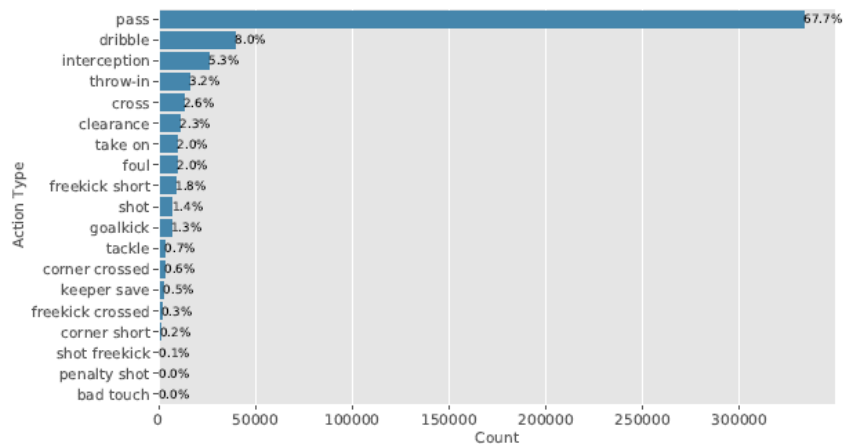
$$Importance = \max(P(win) - P(lose)), O \in [Champion, UCL, EL, Relegated] \quad (5)$$

C. In-Game Win-Draw-Loss Probabilities Estimation Models Performances

Table 2 - Classification model results for win-draw-loss probabilities estimation

	Training Results	
	Avg. Accuracy	Standard Deviation
Random Forest	0.865	0.0538
XG Boost	0.639	0.0585
Logistic Regression	0.594	0.0492
Kneighbors Classifier	0.493	0.0537

D. Frequency of actions in our dataset



E. The elbow curve for determining the optimal number of clusters

