University of Amsterdam
Faculty of Science

Analysing the Serie A 2019/20: Finding the underlying patterns of winning in Italian football

by

Rabie Afqir
11606843

Spring 2021

A Dissertation submitted in part fulfilment of the Degree of Bachelor of Science Information Sciences

Faculty of Science
University of Amsterdam

Supervised by Paul Groth
Co-supervised by Robin Langerak

Abstract

In the current climate of football, the performance of football teams on the pitch is essential to generate revenue. An improvement in performance generates many new ways of income such as prize money or new sponsorships. It is thus of importance for a football club to perform as best as they can. A combination of machine learning and feature selection techniques were used to determine which practices lead to a top 7 finish. The results of this research hint at a slight correlation between a top 7 finish in the Serie A in 2019/20 and territorial dominance. Football coaches and other administrators in football and other sports should thus look further into the role of territorial dominance and a higher spot in the standings.

Keywords: Classification, Football Data, Machine Learning, Predictive analysis, Linear SVC, Random Forest, Gaussian Process, Sequential Feature Selector

Word count: 14 528

Contents

1 Introduction

The Italian Serie A is the third highest level football league in the world (UEFA, 2021). With some of the most historic clubs in the world and some of the greatest players in the world that have ever played the game, such as Cristiano Ronaldo, Zlatan Ibrahimovic, Zinedine Zidane, Ronaldinho and Gianluigi Buffon, the Italian Serie A is an amazing league to watch as a sports fan. The Serie A is not only interesting for its history and legacies but also for the revenue it generates with that history and those legacies. An example of this is how having a long history gives the clubs a better chance to get an individual TV deal (Brandes & Franck, 2007). As the history and the status of the club grows, it often gets more marketable since the sporting performance and history makes the club more attractive to the general public (Brandes & Franck, 2007). Also, qualifying for European football is financially beneficial to Italian football clubs. UEFA distributed 2,6 billion euros across the contenders for the European trophies in the season 2019/20 (UEFA, 2019a, 2019b).

It is thus important to make sure that sporting performance is excellent. One way to improve that performance is to analyse the events that happen during a game. During the competition there are a great number of elements that determine the result of a match, including, but not limited to, the amount of crosses of the left and the right side of the pitch, the amount of ball possession and the way the goals were conceded as seen in the research of Zuccolotto, Carpita, Sandri, and Simonetto (2015) and Constantinou and Fenton (2017). These elements are recorded by companies like OptaSports, SciSports and WyScout. These companies record the events that happen in games from leagues around the world. The events are taken from a video feed of the games and recorded by tagging them from the video feed (Wyscout, 2021; Opta Sports, 2021; SciSports, 2021). These companies produce these datasets not just for clubs and companies. Researchers use these datasets as well, for various purposes (Constantinou & Fenton, 2017). One of these purposes is to seek the identifying features of teams (Constantinou & Fenton, 2017). These defining features give insight into the underlying patterns of team performance (Constantinou & Fenton, 2017). The defining features also matter because they give players, coaches and other administrators in the Serie A insight into either, why they are a winning team, or what they are lacking (Zuccolotto et al., 2015). This paper will seek to understand what the defining features of a winning team in the Serie A are.

1.1 Approach

To scope our research, we focus on Italian Serie A football and pose the following main research question:

> *Which features correlate most strongly with finishing in the top 7 in the Italian Serie A in the season 2019/20?*

To be able to answer this main question, a two sub-questions must be answered first. These sub-questions are:

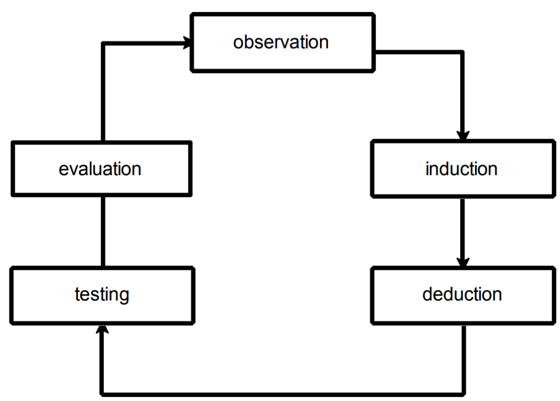
1. Which algorithm – out of random guessing, a possession-based baseline, Linear Support Vector Machine, Random Forest Classifier and Gaussian Process Classifier – is most effective to predict whether a team in the Italian Serie A finishes in the top 7?

2. Which subset of features is the most effective in predicting whether a team in the Italian Serie A finishes in the top 7?

1.2 Objective of research

Both these sub-questions will be answered through quantitative, inductive, deductive and empirical research. This research is inductive and deductive because it fits the attributes of both inductive and deductive as described by Thiel (2010). Inductive research describes a problem and creates a theory based on that while deductive research tests a theory (Thiel, 2010).

Figure 1: The empirical cycle as described by Dirkje Magrieta Mietus in 1962



This research does that by looking at which features correlate the most with a spot at the end of the season in the top 7. Deductive research tests a theory/knowledge and checks the validity of that theory/knowledge (Thiel, 2010) and that is exactly what happens when the performance of the models are checked on the real results. This research will follow the steps of the "empirical cycle" described by Mietus (1994).

1.3 Hypotheses

Following these research questions, the hypotheses are posed. The hypotheses are built upon assumptions of a correlation between an attacking style of football and increased performance. By researching these hypotheses, the objective is to affirm or reject these assumptions and thereafter being able make well-found assertions about the correlation of an attacking style of football and the performance. Hereby the sub-questions and main research questions will be answered. A few examples are:

- The Master thesis of Andrea Pirlo, Juventus, AC Milan, Internazionale and Italian national team legend, to complete his UEFA Pro license, the highest coaching badge in world football. The UEFA Pro license is needed to coach teams on the highest level in the world. In Pirlo's thesis, it was described that a proactive, attacking and quality football can give great benefits (Pirlo, 2020, p. 29).

- The Master thesis of Hernan Crespo to complete his UEFA Pro license. The Argentine is mostly known for his time as a player for Chelsea FC, Internazionale and AC Milan. Crespo (2013) underlines the importance of the attacking patterns as the basis of the tactics of his teams. The overlap of a fullback or exploiting the individual talent of a player are essential to the game according to Crespo (2013, p. 25).

- Zauli (2002) is a professional football coach who produced a comprehensive look at Italian football through the eyes of ten high level Serie A coaches such as Carlo Ancelotti, Arrigo Sacchi and Marcello Lippi. Zauli (2002) explains that "it is always better to arrive as fast as possible toward the opponents' goal" (p. 43). Zauli speaks about a direct play style when intercepting the ball to take advantage of the weakness of the opposition.

All these examples boil down to the opinion that having the opponent on the back foot can hinder their tactical plan. Based on this background, one would suspect the teams that attacking teams win more often since taking the initiative in a football match is deemed very important by Pirlo, Crespo and Zauli. Examples of this stance is that "a proactive, attacking and quality football can give great benefits" (Pirlo, 2020, p. 29) and that "it is always better to arrive as fast as possible toward the opponents' goal" (Zauli, 2002, p. 43).
We will thus select features that will indicate an offensive and a defensive playing style. The features that indicate an attacking playing style are:

- Fouls suffered in attack

- Passes received in attack

The features that will indicate a defensive playing style are:

- Shots suffered

- Ball possession

- Total goals conceded from set pieces

- Substitutes in

- Interceptions

These features are selected based on what Pirlo (2020), Crespo (2013) and Zauli (2002) define as an offensive and defensive playing style. These three authors are experts on the Serie A and have played and sometimes coached in it for years. Their work and opinions gives us an indication on what the styles in the Serie A are. We thus want to look at the difference between the performance of a model trained on the attacking features and the performance of a model trained on the defending features. Having this in mind, the hypotheses are as follows:

- **H0:** $\text{PoM}_{\text{attack}} = \text{PoM}_{\text{defence}}$
- **H1:** $\text{PoM}_{\text{attack}} \neq \text{PoM}_{\text{defence}}$

where PoM represents the Performance of the Model as a percentage of the correct predictions. Note that this hypothesis is not what this research is about but aims to aid in answering the aforementioned sub-questions by testing the truth of the background found in the work of Pirlo, Crespo and Zauli.

1.4 Relevance

To improve the performance of a football club in the Serie A is of immense importance for the team for various reasons. The clubs are not the only ones who can benefit from this study. The methods used in this study can begin to give new insight into the inner workings of the sports industry and the application of machine learning and big data therein.

1.4.1 Social relevance

The societal relevance is mostly found in the ways this research can increase the financial gain of clubs by improving the performance of football clubs. An example of that phenomenon is the fact that the distribution of the prize money the club receives after a season is partially dependent on the sporting performance of the club (Galardini, 2019).

Also, the bigger the clubs are, the more sponsorship money they get. The biggest and best clubs in the world get the most money in the world (Sartori, 2020). The clubs are also allowed to agree their own individual deal in which they sell their own broadcasting rights (Brandes & Franck, 2007). This obviously favours the biggest clubs in the league, like Juventus, Internazionale and AC Milan because the bigger and better a club is, the higher the value of the broadcasting rights is (Brandes & Franck, 2007). The performance also positively influences the stock price (Demir & Danis, 2011).

1.4.2 Scientific relevance

This research explores the methods used in earlier research to expand on what has already been researched. Firstly, Hubáček, Šourek, and Železný (2019) state:

> *Lastly, from the perspective of soccer outcome prediction modelling, it would be interesting to conduct a more thorough analysis of the individual constructed features towards the prediction performance.*

This research will look further into the causes of the prediction and which features influence the prediction the most, which is a gap in the research of Hubáček et al.. This will be done with a new part of the sklearn-learn package called the "Sequential Feature Selection" which was only released in December of 2020 which is a month ago at the time of writing. This research also explores the application of this new method in a practical way.

Secondly, this research aims to expand on the research of Zuccolotto et al. (2015) by using similar classification algorithms but also having different ones. This research has the usage of the Random Forest Classifier in common with the research of Zuccolotto et al. while all other algorithms are different. The reason this research chose methods that are different is to keep exploring the ever changing world of the application of data science on sports. This research also expands on the fact that the data is of the Italian Serie A from the same data provider. However, this research does in fact look at a single more recent season (2019/20) unlike the research of Zuccolotto et al. which looked at the seasons from 2008/09 until 2011/12. This research aims to improve the knowledge science has on the possibilities of predicting the driving factors behind winning a football match.

1.5 Thesis organisation

Section 2 will present the theoretical context of the topic this research explores by clarifying the setting this research finds itself in at the moment. Section 3 will give insight in the algorithms, the dataset and the overall steps used to come to the conclusion this research came to while this research was conducted. Section 4 will consist of the performance of the algorithms used and studied. This study will end with Section 5 that will discussing the interpretation of the reported performances of Section 4 and Section 6 containing the theoretical implications of the results and the future work that could follow after this study.

2 Theoretical Framework

This section will concern the concepts, definitions and the context wherein they lie. This section will clarify the way this study talks about this case study, the football vocabulary and football in general.

2.1 Conceptualisation

This section will clarify the theoretical setting wherein this research finds itself. This means that clarifying key terms is necessary to ensure it is clear what this research is talking about and any misunderstanding is avoided by setting a single definition of the terms. Only the terms that might need a clarification will further be explained. All terms that are not explained will be assumed to be commonly known among people with basic knowledge of football and data science.

2.1.1 Key term 1: Feature

In subquestion 1 as well as in the rest of the research, the term "feature" will be used. The term "feature" is a column of properties that describe the data points in machine learning (Müller & Guido, 2016). Although the term is quite basic one, it still warrants a clarification as to what a feature in this research is since the datasets of every company that collects sporting data has a different data structure and different features.

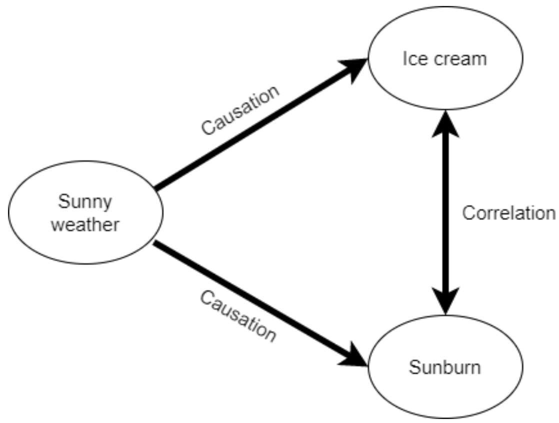
The dataset Panini Digital provided was an Excel file containing two sheets, one containing the data on the teams and one containing the data on the players of these teams. The data sheet of the players in the season 2019/20 has 573 rows and 857 columns. Meaning each of the 573 players that played a game in the season 2019/20 season of the Italian Serie A had 857 data points recorded before the data was cleaned. The data sheet of the teams in the season 2019/20, on the other hand, has 20 rows and 1352 columns. Meaning that for each team 1352 data points were recorded throughout the season. The best way to give an impression is to show the dataset.

The full list of all columns used and the first 17 columns of the two sheets of the dataset used for this research will be listed in Appendices as Appendix A, Appendix B and Appendix C, respectively.

2.1.2 Key term 2: Correlation

The main question uses the word "*correlate*". Burns and Burns (2012) define correlation as "a measure of the degree of correspondence between variables". As is commonly known and this definition also reinforces, the fact that there might be a correlation between facts, does not prove there is a form of causation (Figueroa, 2019). The fact that two events happen at the same time does not mean that either of them cause the other.

Figure 2: The example given by Figueroa (2019) which visualises how correlation does not mean causation



In Figure 2, a visualisation of Figueroa (2019) is shown on the difference between correlation and causation. An increase in the consumption of ice cream during sunny weather does correlate but obviously does not cause sunburn. This means that although this research will look at the correlation between features of teams in the Italian Serie A and the outcome of their matches, a causation between the features and the outcomes in any form whatsoever can not be disproved just yet.

2.1.3 Key term 3: Top 7 Finish

The relevance of a top 7 finish may seem clear to those who are known with the way football competitions in Europe work. The top teams in the European leagues qualify for the UEFA Champions League and the UEFA Europa League. The top 7 teams qualify for these leagues as of the season 2019/20 (UEFA, 2020). The reason the top 7 was chosen is because of the fact that being able to qualify for European football indicates the absence or presence of quality in a team. Only the best teams in a league are able to qualify. As such, taking the fact whether a team does or does not qualify for Europe is a reasonable metric to look at to identify a top team from a lesser team. There has not been a research paper or book which has done a research similar to this one. Further explanation on the gap this research fills is found in Section 2.3.

2.2 Scientific context

Sports obviously has a great application for predictive models, for example, in the form of predictive models to generate betting odds (Baboota & Kaur, 2018), increase the performance of teams (Zuccolotto et al., 2015) or help scout the correct players (Fried & Mumcu, 2016). Science therefore can get a decent indication of what influences the chance of a team to win a game. Fortunately, football is not predictable. Developing a model which predicts the outcome of a football match or a season with 100% accuracy will probably never be possible. We do have a fair indication of what works for teams in a certain league (Baboota & Kaur, 2018). Several of these indications are given in the research of Hucaljuk and Rakipović (2011). They attempted to predict the outcome of 96 Champions League games based on the form of a team, the outcome of the previous meeting between the two teams, the current position in the rankings, amount of injuries and the goal difference. They used several machine learning algorithms on this relatively small data set but their results still showed a satisfactory capability of prediction. The application of an Artificial Neural Network proved to be best with an accuracy of 68,8%.

These algorithms are, however, often applied to datasets of one league or competition. There is research where the data comprised of the data points of multiple seasons such as the work of Zuccolotto et al. (2015). Often, only one league is considered to account for the cultural playing styles of each country (Min, Kim, Choe, Eom, & McKay, 2008). Every country has a different playing style which is important to keep in mind. The Spanish possession based passing game is very different from the physical and defensive style of Italy, for example. It is therefore important to separate different leagues to account for these differences.

2.3 Gap to be filled

Looking at the same league at different moments, can be rather informative. That is exactly what this research will do. This research uses a different dataset from Zuccolotto et al. (2015) from the same data provider to explore the Italian Serie A after the work of Zuccolotto et al.. Zuccolotto et al. used the seasons 2008/09 until 2011/2012. This research will use the dataset of the season 2019/20, however. Gathering more recent data is important to gain a further knowledge on the developments of the relation between machine learning techniques and their application on the Italian Serie A. The size of the dataset also adds to the quality of this research by giving more insight into what happens during a football game. The 1228 data points for each team gives a more detailed impression into the style of play and possible success of the team.
The dataset is not the only thing that is different. The new feature selection method used will also be of interest. As already stated in Section 1.1, the new transformer of the sklearn-learn package called the "Sequential Feature Selection" was only recently released at the time of writing.

Also, at the time of writing, there was no other research to be found on the topic of classifications on a seasonal basis. Almost all research has been done on a match-to-match basis (such as (Zuccolotto et al., 2015) or looked at other features of a match, like the goal-scoring patterns, instead of the match result (such as (Garganta, Maia, & Basto, 1997)). As such, this research will present a new way to classify as: a top team or a team of lesser quality. Since these new methods in a new domain are used, this research takes a top 7 finish as an indication of being top. The substantiation for this decision is given in Section 2.1.3.

However, we also employ standard machine learning algorithms such as Support Vector Machines (SVM), Gaussian Process Classifier and the Random Forest Classifiers. It is however interesting to see these algorithms applied to a case where an entire season has to be predicted instead of just a single match.

3 Methodology

This section will discuss the methods used. The data file and the way it was treated will be highlighted. Afterwards the prediction algorithms and all other helper functions will be explained.

3.1 Dataset

The Italian Serie A has 20 teams. They all play each other twice, once at their own stadium and own once at the stadium of the opponent. This comes down to 38 games. These 38 games and their statistics were recorded. Panini Digital recorded some features on a team basis. A second dataset contained features of the individual players.
This research analyses the data of the season 2019/2020. The data file was provided by Panini Digital, one of the biggest players in the market of sports data. Panini Digital is the same data provider for the research of Zuccolotto et al. (2015). As such, the features recorded is the same type of information in the paper of Zuccolotto et al. such as free kicks and shots, action type, fouls, crosses, recovered balls, goal assists, average time of ball possession, saves, goals on free kicks (Zuccolotto et al., 2015).
The dataset contains 573 players from 58 different countries. Each team had an average of 26,7 players (±1,98) with an average age of 27,3 (±4,89).

3.2 Data cleaning

42 empty columns were deleted from the dataset because these empty values raised issues with some of the algorithms used in this research. Some empty columns were filled with hyphens to indicate the absence of a value. These hyphens were removed so they could be properly interpreted as actually empty.

3.3 Data transformation

The dataset contained characters that needed to be corrected for to be imported into scikit-learn. The algorithms only take a dataset with numerical values or booleans. As such, the columns with strings needed to be changed. Although all columns contained numerical values, multiple reasons made some columns be interpreted as columns containing strings. Three examples of these reasons are:

- A suffix for meters when a value concerns a distance. This suffix was removed to be left with just the distance in meters. "76 (mt)" meaning 76 meters would thus become "76".
- A space between numbers when they are bigger than 999. In Italian, numbers are separated by a space when there are 3 next to each other.

"42 500" would thus become "42500" and "1 050 600" would thus become "1050600".

- Change the comma to a point in decimal numbers. In Italian, decimal numbers are denoted with a comma instead of a point. "85,6" would thus become "85.6".

Also, the dataset contained a column called "Birth date" containing the dates of birth of the players. This column was transformed to a column which contains their age at the time of writing.
The data file also contained two datasets: one with all statistics with the data on a player basis and one with all the statistics on a team basis. These two datasets were merged into one wherein all features were recorded on a team basis. This was done by taking all the average of a team and adding those features in a new column. These new columns contained the prefix 'Average' and the suffix 'of all players' to denote the fact that these columns are generated.

3.4 Data selection

In this section the data split will be explained. The X-variable is the data which predict the Y-variable. These variables are explained to clarify which data is predicting what.

3.4.1 The X-variable

The dataset used for this research is a subset of all the features collected by Panini Digital. Various features were excluded from the dataset. These features were:

- Team
- Points
- Won
- Drawn
- Lost
- Goal favour
- Goal against
- Assists
- M Assists
- M Before assists
- Before goal assists from a play initiated

The column "Team" was not included since it is a string column and does not add any insight into the features of a top 7 team. All others were not included because these features will not give us any insight either. "Winning many games and scoring a lot makes a team a top team" is rather obvious and does not need scientific research to come to that conclusion. Instead this research aims to understand the underlying patterns of winning and performing well in a competition like the Serie A. These extremely obvious reasons for winning were thus excluded from the dataset.
All the above mentioned points means that the dataset contains 1960 columns and 20 rows.

3.4.2 The Y-variable

The Y-variable is derived from the standings at the end of the Serie A season of 2019/20. The Y-variable is all the teams that competed in the Serie A season of 2019/20 and whether they did or did not qualify for European football. The full presentation of the Y-variable is found as Appendix D.

3.5 Baselines

Two baselines were implemented to give a precedent for the performances of the algorithms that were used.

3.5.1 Algorithm 1: Random baseline

The random baseline randomly assigns a team as a top 7 team or not. It works as follows: A list of 20 booleans (7 times a "True" denoting a top 7 team and 13 times a "False" to denote a non-top 7 team) is generated and randomly shuffled to emulate a random assignment of the top teams. This list was then compared to the real assignment of top teams and lesser teams. This test was ran 1,000,000 because of how undemanding the code was. The performance of this baseline will function as the bare minimum an algorithm has to get to be considered as significant because if the algorithm performs that badly that randomly guessing works better than the algorithm, there will not be much meaning behind the most significant features it selects.

3.5.2 Algorithm 2: Possession-based baseline

The second baseline looked at the average ball possession teams had during each game. The prediction was that the higher a team's average ball possession was, the higher a team's chance to win was. As such, the seven teams with the most possession were identified as the top team. This prediction was then compared to the actual assignment of top teams and lesser teams.

3.6 Cross validation

The Repeated KFold was used to implement cross validation in this research. The Repeated KFold is a part of the scikit-learn module to help iterate over a dataset. This module provides different permutations of the rows in the dataset. These different permutations are used to determine a different X_train, X_test, Y_train and Y_test to have a different dataset to train and test on. These different permutations are used to determine different splits in the data to create different train and test data. By doing so, the statistical power increases. This creates a scenario where different splits are used instead of just a single split. By doing so, the chance that every row gets used as train and test data increases. It is thus essential to use in this research. This is the reason why the Repeated KFold is applied to every from algorithm 3 to 5.
This means that the train-test-split for every algorithm was done by the Repeated KFold. The rule of thumb in machine learning is a ratio of 80% train data and 20% test data. The Repeated KFold was configured for a split of 75% train data and 25% test data to account for a small dataset. This means that the train data consisted of the data of 15 teams and the test data consisted of 5 teams. These splits were done ten times. The choice for 10 iterations was made after considering having enough iterations to make the outcome as statistically as strong as possible and keeping the run time of the algorithms bearable. For each split, the respective algorithm was ran and its results collected.

3.7 Algorithms

This research used the Linear Support Vector Machine, Random Forest Classifier and the Gaussian Process Classifier to obtain its results. To get the same results every single time the algorithm was run, consistency in the algorithms had to be implemented. This was done by utilising the parameter "random_state" that the aforementioned algorithms and the Repeated KFold have. This means that whenever these algorithms were used the same random state was used and the same situation was used for every single time the algorithm is run on this data set. The parameter "random_state was assigned the number 1960 in reference to the 1960 features used in the data set of this research.

3.7.1 Algorithm 3: Linear Support Vector Machine

The Linear Support Vector Machine trains itself and creates an eclipse in a graph through the data points to classify them if the dataset contains two features. A two dimensional space is needed to visualise the Linear Support Vector Machine at work (Müller & Guido, 2016) if the data points have two features. A three dimensional space is needed to visualise data points with three features. This would mean the decision boundary would be a pane through a three dimensional space. This research, however, concerns a scenario with 1960 features. This is obviously impossible to visualise. Thankfully the Linear Support Vector Machine can still do its job. It is fitted on the X_train containing 1960 features. Afterwards the features are selected by the Sequential Feature Selector and its coefficients are shown. The parameters used for this algorithm are the default parameters apart from the parameter "random_state". This parameter is assigned to 1960 as has been explained earlier.

3.7.2 Algorithm 4: Random Forest Classifier

The Random Forest Classifier is characterised by the branches the model creates in the decision trees. These trees are each completely independently formed from each other. The trees are all a *bootstrap sample* of the X-data (Müller & Guido, 2016). This algorithm does however differentiate from the Linear Support Vector Machine and the Gaussian Process Classifier. *"Instead of looking for the best test for each node, in each node the algorithm randomly selects a subset of the features, and it looks for the best possible test involving one of these features"* (Müller & Guido, 2016). This means that instead of using one dataset, different sub sets of the X-data is used for the different branches in the classifier. Each branch with its own algorithm then provides a probability for a top 7 finish. The predictions are made based on the average of these probabilities. Afterwards the features are selected by the Sequential Feature Selector and its coefficients are shown. The parameters used for this algorithm are the default parameters apart from the parameter "random_state". This parameter is assigned to 1960 as has been explained earlier.

3.7.3 Algorithm 5: Gaussian Process Classifier

The Gaussian Process Classifier supports multi-class classification by performing one-versus-one or one-versus-rest based training and prediction. Because this research only has two classes, it uses the one-vs-one classification. The classifier is somewhat more advanced than the Linear Support Vector Machine. Gaussian processes are a type of kernel method, like SVM's, although they are able to predict highly calibrated probabilities, unlike SVMs.
This research trains itself on the X_train data. Afterwards the features are selected by the Sequential Feature Selector and its coefficients are shown. The parameters used for this algorithm are the default parameters apart from the parameter "random_state" and the 'multi_class' parameter which was changed to one-vs-one.

3.8 Sequential Feature Selector

After the aforementioned algorithms were trained, the features with the highest correlations were selected using the Sequential Feature Selector of the scikit-learn module. This transformer is a new addition to the module in version 0.24.0 released in December of 2020. This transformer looks at each feature and its correlation to the Y-variable. The feature with the lowest correlation get taken out of the equation. This process keeps happening until the amount of wanted features is reached. The amount of wanted features in this research was 5. The reason behind that amount is because it keeps the selected features comprehensible while also giving a proper insight into the results of the Sequential Feature Selector of the scikit-learn module. The selected features are then presented with their corresponding coefficients.

3.9 Semi-subjective predictions

All the previous algorithms make predictions without any help from a human. These objective predictions are made solely on data. Subjective predictions are however effective as well (Leung & Joseph, 2014). Most simple statistical algorithms consider the strength and weaknesses of teams to come to a prediction. It is however hard to do so if teams have not yet competed before. It is then useful to make predictions based on the expertise and knowledge of a professional in the field. This professional can provide insights the algorithm might not see or understand. This comparison is not new whatsoever. The expertise of an expert has been compared with data models as early as 2003. Boulier and Stekler (2003) looked at the performance of expertise versus the performance of data.

It is however interesting to look at the combination of expertise and algorithms. It is interesting to see if the Linear Support Vector Machine, Random Forest Classifier and Gaussian Process Classifier perform better with a helping hand in the form of expertise since football is not an exact science. As such, the aforementioned methods are replicated with a different dataset. This new dataset is a subset of the original dataset. We created a hand annotated dataset to incorporate expert football knowledge since it is rather difficult to meet the true experts of this domain due to COVID-19 restrictions at the time of writing. The attacking and defensive feature of a team tell us about their prowess in attack and defence. Attacking teams are often seen as superior to defensive teams. Defensive teams are often seen as defensive because they do not have the quality to be offensive. As such, it is interesting to see if the attack or the defence is indicative of the success of a team. Is a team that is good at attacking better than a team that is good at defending? Does a team need to be good at either attacking or defending to be able to be a top team? Do they need to be able to do both at the same time or do neither really matter at the end of the day?
To find that out, two subsets were created on which the three algorithms would be trained. The attacking subset of the dataset contained the following features:

- Fouls suffered in attack
- Passes received in attack
- Ball possession

The defensive subset of the dataset contained the following features:

- Shots suffered

- Ball possession
- Total goals conceded on set pieces
- Subs in
- Interceptions

The hypothesis concerns these attacking and defensive mentality. To be able to accept or reject this hypothesis, we need to see if the difference is significant. To detect a significant difference between the two mentalities, a paired samples t-test was used. The paired samples t-test is used to test for a significant difference between two samples that are related (Burns & Burns, 2012).

3.10 Related work

Although this research tries new techniques, there are resemblances with earlier research. Firstly, the usage of data provided by Panini Digital and the usage of the Random Forest Classifier are matching traits of the research of Zuccolotto et al. (2015) and this research.
Secondly, this research compares methods and algorithms with each other. The research of Khan and Kirubanand (2019) is an example of this comparison between algorithms as well. They compared the performance of an SVM model and an XGBoost model predicting the outcome of a football match. They compared them by looking at the F1 score and the accuracy. In this research, we will however look at the percentage of correct predictions and compare those among the algorithms.
Thirdly, the usage of subjective input to predict the outcome in a sports context is a common feature with the work of Leung and Joseph (2014) and Boulier and Stekler (2003). These authors created work where the expertise of experts in the sports domain where investigated and measured using various machine learning algorithms such as the Linear SVM, neural networks and decision trees. This research does the same but puts its own spin on it.
There are however plenty differences with earlier research. One of the most interesting ones is the usage of the Sequential Feature Selector. This research shows this extremely new feature selector in a practical way. This research uses this feature selector to see how this new feature of the scikit-learn module functions.

4 Results

In this section, the percentage of correct guesses of all baselines and algorithms are reported along with the features the Sequential Feature Selector found to be the highest correlated features. Then the performance of the algorithms trained on these highest correlated features will be presented.

4.1 Performance of the baselines

The baseline based on the average amount of possession a team had throughout the season was able to predict the classification of the correctly 90% of the time. However, randomly assigning whether a team was going to finish top 7 or not got a performance of 54,5%. This performance of predicting the classification 54,5% correctly will also show us how the algorithms performed compared to absolutely random guesswork.

4.2 Performance of the algorithms

The Linear SVC had an accuracy of 49% with the features with the highest correlated features together with the fact a team is or is not a top 7 team being:

- Territorial supremacy in attack (avversari) (-0.00008)
- Territorial supremacy (0.00008)
- Territorial supremacy in attack (0.00003)
- Balls played in attack area (avversari) (-0.00003)
- Balls won back (-0.00003)

After training the algorithm on only the aforementioned features, the Linear SVC had an accuracy of 58,5%.

The Random Forest Classifier had an accuracy of 54,5% with the features with the highest correlated features together with the fact a team is or is not a top 7 team being:

- Accelerations (0.28877)
- Acrobatic shots (0.24893)
- Anticipations (0.17744)
- Acrobatic shots from a set play (0.14854)
- Acrobatic shots from a play (0.13628)

After training the algorithm on only the aforementioned features, the Random Forest Classifier had an accuracy of 56%.

The Gaussian Process Classifier had an accuracy of 65% with the features with the highest correlated features together with the fact a team is or is not a top 7 team being:

- Attempted passes (0.31333)
- Attempted passes in opponent midfield (0.31333)
- Ball possession (0.31333)
- Balls played (0.31333)
- Balls played in attack area (avversari) (0.31333)

After training the algorithm on only the aforementioned features, the Gaussian Process Classifier had an accuracy of 51,5%. This 65% accuracy is however done by predicting that not a single team is going to finish top 7. As such 7 out of the 20 teams are misclassified.

4.3 Performance of the hypothesis based algorithms

The same algorithms were used but with two small subsets of the data set. The Linear SVC, Random Forest Classifier and the Gaussian Process Classifier were trained on attacking and defensive features as explained in Section 3.9. The performance of these algorithms are as reported in Table 1.

When looking at Table 1, a slight difference between the attacking and defensive-minded algorithms can be seen. The attacking Linear SVC performed 4% better than the defensive Linear SVC. However, the defensive Random Forest Classifier makes up for half of the difference. The defensive Random Forest Classifier performed 2% better than the attacking Random Forest Classifier. There is however no difference in performance between the attacking and defensive Gaussian Process Classifier. As such, the difference between the attacking and defensive algorithms is 2%. The difference between the attacking-minded algorithms (M=57,667,SD=6,526) and the defensive-minded algorithms (M=56,833,SD=7,286) was not significant however; t(2)=0,4336, p=0,7069. As such, the null hypothesis is accepted.

All the aforementioned performances amount to Table 2.

Algorithm	Linear SVC		Random Forest Classifier		Gaussian Process Classifier	
Mindset	Attacking	Defensive	Attacking	Defensive	Attacking	Defensive
% Correctly predicted	55,5	51	52,5	54,5	65	65

Table 1: Performance of the attacking and defensive-minded algorithms

	Algorithm	% correctly guessed	% incorrectly guessed
Baselines	Random guesswork	54,5	45,5
	Possesion based performance	90	10
Algorithms	Linear SVC	49	51
	Random Forest	54,5	45,5
	Gaussian Process	65	35
Feature Fitted Algorithms	Linear SVC	58,5	41,5
	Random Forest	56	44
	Gaussian Process	51,5	48,5
Attack-minded Algorithms	Linear SVC	55,5	44,5
	Random Forest	52.5	47.5
	Gaussian Process	65	35
Defence-minded Algorithms	Linear SVC	51	49
	Random Forest	54,5	45,5
	Gaussian Process	65	35

Table 2: Performance of all algorithms used in this research

5 Conclusion

The reason this research was conducted is to give coaches, players and other administrators insight into the driving forces behind finishing in the top 7 of the Italian football league and thus qualifying for European football. After looking at earlier research an approach was formulated to find these driving forces. Three algorithms were used in four different ways along with two baselines.

5.1 Interpretation of the results

The "Random guesswork" baseline reports an accuracy of 54,5%. As said in Section 3.5.1, this baseline is seen as the bare minimum to be considered as significant. This means that all algorithms are deemed insignificant excluding:

- The Gaussian Process Classifier using the entire data set;
- The Gaussian Process Classifier using the attacking features;
- The Gaussian Process Classifier using the defensive features;
- The Linear SVC using the features selected by the Sequential Feature Selector;
- The Linear SVC using the attacking features.

The reason all algorithms that perform under the baseline are deemed insignificant is the low quality of the feature selection caused by the low performance of the model. The quality of the prediction that the correlation of a set of features is high is not reliable if randomly guessing performs better than an actual model. As such, whenever a model performed worse than the "Random guesswork" baseline, its predictions were deemed unreliable and insignificant. Despite the 90% accuracy of the possession based baseline, it is in on of itself not a predictor of what makes a team a winning team. It goes without saying that merely having possession in football is not going to win any team any games. It is thus important to look at what needs to happen when a team does or does not have the ball.

The best performing algorithm was the Gaussian Process Classifier using the entire data set, the attacking features or the defensive features with an accuracy of 65%. However, it predicts the absence of any team finishing in the top 7 of the league to achieve this 65% accuracy. Predicting the absence of a top 7 is obviously impossible. The Gaussian Process Classifier using the entire data set, the attacking features or the defensive features is thus not suitable to predict whether a team is going to finish in the top 7 in this use case.

With the Gaussian Process Classifier out of the equation, only the Linear SVC performs better than the "Random guesswork" baseline. The Linear SVC using the attacking features performed 1% better than the "Random guesswork"

baseline while the Linear SVC using the features selected by the Sequential Feature Selector performed 4% better than the "Random guesswork" baseline. The features the Linear SVC together with the Sequential Feature Selector are in line with the finding.
Firstly, because the Sequential Feature Selector found only correlations which are linked to dominance which can be explained. These correlations are:

- A negative correlation with the territorial supremacy of the opponent in attack. This means that the opponent is unable to dominate when attacking;
- A positive correlation with territorial supremacy;
- A positive correlation with territorial supremacy in attack;
- A negative correlation with the amount of balls that were won back meaning that the more balls are able to be won back, the more chances the opponent has;
- A negative correlation with the amount of balls played in the attacking phase by the opponent.

Secondly, the correlation of the aforementioned features are low. This explains the slightly better performs of the Linear SVC using the features selected by the Sequential Feature Selector. The importance of dominance is stressed by the possession based baseline as well. The significance of keeping the ball and being able to dictate the game is underlined by the 90% accuracy of the possession based baseline.
Also, as the null hypothesis has been accepted in Section 4.3, there does not seem to be a difference in performance between the selected attacking and defensive features.

5.2 Answers on the research questions

The first sub-question handles the performance of the algorithms used in this research. As has been discussed, the best performing algorithm is the Linear SVC using the features selected by the Sequential Feature Selector with an accuracy of 58,5%. The second sub-question concerns the features that help getting this accuracy. The features the Sequential Feature Selector selected for the Linear SVC are

- The territorial supremacy of the opponent in attack;
- The territorial supremacy;
- The territorial supremacy in attack;
- The amount of balls that were won back;
- The amount of balls played in the attacking phase by the opponent.

.
The research question concerns what correlates most strongly with finishing in the top 7. These features are the features that answer the second sub-question together with the feature "Possession". The teams with the most possession qualify in the top 7 bar two exceptions. As such, these six features correlate most strongly with finishing in the top 7 in the Italian Serie A in the season 2019/20.

5.3 Significance of findings

The accuracy of 58,5% is lower than the accuracy of 70% of the models from Baboota and Kaur (2018). This research does however give a season wide view of performance instead of a match-to-match view Baboota and Kaur created. This research also gives insight into the reason why the better teams perform better and the worse teams perform worse.
This also means that players, coaches and administrators should look at the role of dominance in their game. Even though the findings are not conclusive, it hints at a significance of dominance on the pitch during the season 2019/20. This research suggests that a persisted dominance on the pitch during this season correlates with finishing in the top 7 of the Serie A. This correlation is thus interesting for those clubs to investigate.

6 Discussion

After assessing and interpreting the performances of the algorithms used, it is important to clarify what these interpretations mean. First, the quality of this research will be evaluated by looking at the reproducibility, generalisability and the transferability. Afterwards the conclusions will be put into its scientific and societal context.

6.1 Reproducibility

The steps of data transformation and selection have been discussed by not only telling what the steps are but also by substantiating the choices that have been made. Consequently, future researches are able to recreate the work and build on the work that has been done in this research. To be able to replicate the work of the algorithms, the parameter ”random_state” was used. As such, future researchers can replicate the procedures done in this thesis to verify and/or use the results and conclusions found in this research. To aid later researchers replicate this research, the code has been shared. The Jupyter Notebook with the code can be found here.

6.2 Generalisability and transferability

The main concern around transferability is the extent to which the findings of a research can be transferred from one case to another (Flick, 2018). This research will therefor also discuss the possibility of the findings to be transferable to another football competition or even another sport.

There are however different types of generalisability (Flick, 2018). The default understanding of generalisability is statistical generalisation – generalising from the sample to the population – but this research will use the so-called ”moderatum generalisations” proposed by Williams (2000).

6.2.1 Generalisability

This research concluded that during the season 2019/20 the dominance on the pitch correlated most strongly with a top 7 finish. If we were to apply the notion of Williams (2000) to this research, we would have to ask ourselves whether there is a feature that sets the teams in the top 7 apart rather than asking what that feature is. It is debatable whether the presence of such a feature actually existed for the season this research concerned itself with since the correlations the algorithms yielded were not decisive. However this effect could very well be generalisable to earlier and future seasons of the Italian Serie A. The presence or absence of this feature that set the top 7 teams apart from the rest of the league may take any form – whether as distance travelled, aggression or intensity – but cannot be determined and needs further research in the future.

6.2.2 Transferability

Although there are similarities in and between sports, many differences exist. The findings this research found may be applicable to other football competitions around the world. Football around the globe has found the tendency to be based around attacking power when in possession (Pirlo, 2020; Crespo, 2013). This generalisation might indicate a transferability between the identifying features of a top 7 teams of the Serie A and those of top teams in other competitions around the world.

Whether the findings of this research are transferable to similar ball sports is not unlikely. Sports such as field hockey and rugby are similar in pitch sizes, way of scoring and the amount of members per team and may be two examples of sports where the findings could be transferable. This suggestion of transferability would however have to be researched further before such transferability could be confirmed.

6.3 Strengths and weaknesses of this research

As was mentioned in Section 2.2, the research of Baboota and Kaur (2018) yielded a model with an accuracy of 70%. This research on the other hand was able to get an accuracy of 58,5%. This difference of 11,5% in accuracy does put a sense of uncertainty on the findings of this research.
Baboota and Kaur did however have other types of data. While this research used data of the entire season, Baboota and Kaur used data of the season separated game by game. By doing that, Baboota and Kaur was able to use data about the game that preceded the game that was to be predicted. By having that knowledge, factors like form or schedule congestion could be taken into consideration. Having the same data but on a match to match data could thus have been beneficial to this research.

This research concerned the application of machine learning on a real life use case in the field of football tactics. As such, it is important to combine the knowledge and views of the two field. While writing this research, very few researches have been found where the data mining was not left "untethered" but rather guided with the tactical background of the game. This research has implemented the knowledge and views of the experts who are still active in the field of football coaching by looking at the work of Pirlo, Crespo and Zauli. Also, this research applied its algorithms in three different ways: fully "untethered", "guided" by an algorithm and guided by interpreting the work of the experts out of the field of research. By doing so, three different approaches on applying machine learning on football tactics were tested. This research also explored a part of science on which very little research could be found. The application of machine learning algorithms on an entire football season instead of just a game is a part of science which is left greatly unexplored. It is thus important to add quality work to explore the application of machine learning algorithms on football tactics.

6.4 Implications for later research

Since the part of science where machine learning algorithms is applied on football tactics is still largely unexplored, it was difficult to compare this research with other relevant studies. It is thus interesting to see the application of more algorithms based on the data of a season rather than a match. During the writing of this research the lack of integration of the work of football coaches was noticed. The opinions and trends in the world of football tactics were rarely considered. As such it is interesting to see more work in collaboration with the world of people most knowledgeable on football tactics.

References

Baboota, R., & Kaur, H. (2018, 03). Predictive analysis and modelling football results using machine learning approach for english premier league. *International Journal of Forecasting*, *35*. doi: 10.1016/j.ijforecast.2018.01.003

Boulier, B. L., & Stekler, H. (2003). Predicting the outcomes of national football league games. *International Journal of Forecasting*, *19*(2), 257-270. Retrieved from https://www.sciencedirect.com/science/article/pii/S0169207001001443 doi: https://doi.org/10.1016/S0169-2070(01)00144-3

Brandes, L., & Franck, E. (2007, 02). Who made who – an empirical analysis of competitive balance in european soccer leagues. *Eastern Economic Journal*, *33*, 379-403. doi: 10.1057/eej.2007.32

Burns, R. B., & Burns, R. A. (2012). *Business research methods and statistics using spss.* SAGE.

Constantinou, A., & Fenton, N. (2017). Towards smart-data: Improving predictive accuracy in long-term football team performance. *Knowledge-Based Systems*, *124*, 93–104.

Crespo, H. (2013). *'il calciatore moderno: identita', nazionalismo e internazionalizzazione.*

Demir, E., & Danis, H. (2011). The effect of performance of soccer clubs on their stock prices: Evidence from turkey. *Emerging Markets Finance and Trade*, *47*(sup4), 58-70. Retrieved from https://doi.org/10.2753/REE1540-496X4705S404 doi: 10.2753/REE1540-496X4705S404

Figueroa, A. (2019, Aug). *Correlation is not causation.* Towards Data Science. Retrieved from https://towardsdatascience.com/correlation-is-not-causation-ae05d03c1f53

Flick, U. (2018). *The sage handbook of qualitative data collection.* SAGE Publications Ltd.

Fried, G., & Mumcu, C. (2016). *Sport analytics: A data-driven approach to sport business and management.* Taylor & Francis.

Galardini, G. (2019, Dec). *Mediapro offered $1.45 billion-a-year to make italian football great again.* Forbes Magazine. Retrieved from https://www.forbes.com/sites/giacomogalardini/2019/12/02/mediacom-deal-italian-football-serie-a

Garganta, J., Maia, J., & Basto, F. (1997). Analysis of goal-scoring patterns in european top level soccer teams. *Science and football III*, 246–250.

Hubáček, O., Šourek, G., & Železnỳ, F. (2019). Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning*, *108*(1), 29–47.

Hucaljuk, J., & Rakipović, A. (2011). Predicting football scores using machine learning techniques. In *2011 proceedings of the 34th international convention mipro* (pp. 1623–1627).

Khan, S., & Kirubanand, V. (2019). Comparing machine learning and ensemble learning in the field of football. *International Journal of Electrical and Computer Engineering*, *9*(5), 4321.

Leung, C. K., & Joseph, K. W. (2014). Sports data mining: Predicting results for the college football games. *Procedia Computer Science*, *35*, 710-719. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050914011181 (Knowledge-Based and Intelligent Information & Engineering Systems 18th Annual Conference, KES-2014 Gdynia, Poland, September 2014 Proceedings) doi: https://doi.org/10.1016/j.procs.2014.08.153

Mietus, D. M. (1994). Understanding planning for effective decision support. *Unpublished PhD thesis, University of Groningen, The Netherlands*.

Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. B. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, *21*(7), 551–562.

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with python: a guide for data scientists.* " O'Reilly Media, Inc.".

Opta Sports. (2021). *The Opta difference.* Retrieved from https://www.optasports.com/about/the-opta-difference/

Pirlo, A. (2020). *Il calcio che vorrei.*

Sartori, A. (2020, Jul). *The changing face of football sponsorship – key trends.* Retrieved from https://www.footballbenchmark.com/library/the_changing_face_of_football_sponsorship_key_trends

SciSports. (2021, 04). *Data Delivery.* Retrieved from https://www.scisports.com/services/data-delivery/

Thiel, S. v. (2010). *Bestuurskundig onderzoek: een methodologische inleiding.* Bussum: Coutinho.

UEFA. (2019a, 07). *How clubs' 2019/20 UEFA Champions League revenue will be shared.* Retrieved from https://www.uefa.com/uefachampionsleague/news/0253-0e99cd398188-f80e968d0ab3-1000--how-clubs-2019-20-uefa-champions-league-revenue-will-be-shared/

UEFA. (2019b, 07). *How clubs' 2019/20 UEFA Europa League revenue is shared.* Retrieved from https://www.uefa.com/uefaeuropaleague/news/0253-0e99cd69d503-aeecbfaee2a3-1000--how-clubs-2019-20-uefa-europa-league-revenue-is-shared/

UEFA. (2020, Feb). *Country coefficients: Uefa coefficients.* Retrieved from https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2020

UEFA. (2021, Jan). *Country coefficients: Uefa coefficients.* UEFA.com. Retrieved from https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2021

Williams, M. (2000). Interpretivism and generalisation. *Sociology*, *34*(2), 209–224. doi: 10.1017/s0038038500000146

Wyscout. (2021). *Homepage Wyscout.* Retrieved from https://footballdata.wyscout.com/

Zauli, A. (2002). *Soccer: modern tactics: Italys top coaches analyze game formations through 180 situations.* Reedswain.

Zuccolotto, P., Carpita, M., Sandri, M., & Simonetto, A. (2015, 01). Discovering the drivers of football match outcomes with data mining. *Quality Technology and Quantitative Management*, *12*, 537-553. doi: 10.1080/16843703.2015.11673436

Appendices

A Full list of column names of the used data set

Accelerations	M Left attack side throw-ins with shot
Acrobatic shots	M Left attack side throw-ins with shot (avversari)
Acrobatic shots from a play	M Left attack side throw-ins without shot
Acrobatic shots from a set play	M Left attack side throw-ins without shot (avversari)
Anticipations	M Left defensive side throw-ins
Anticipations by head	M Left foot goals
Attack area throw-ins	M Left midfield side throw-ins
Attack area throw-ins (avversari)	M Left side corners
Attempted passes	M Left side corners (avversari)
Attempted passes in opponent midfield	M Left side corners with shot
Ball possession	M Left side corners with shot (avversari)
Ball possession (avversari)	M Left side corners without shot
Balls overpassing the midfield area (avversari)	M Left side corners without shot (avversari)
Balls played	M Left side throw-ins
Balls played (avversari)	M Lob passes
Balls played in attack	M Lob passes (avversari)
Balls played in attack area (avversari)	M Lob passes in the opponents' half field
Balls played in central forward area	M Lob passes received
Balls played in defence	M Long passes
Balls played in defensive area (avversari)	M Long passes (avversari)
Balls played in forward area	M Long passes received
Balls played in midfield	M Long passes without cross field pass
Balls played in midfield area (avversari)	M Long punts
Balls played in opposition half	M Lost and played ball poss. ratio
Balls played in penalty area	M Lost ball poss.
Balls played in the left attack side	M Lost ball poss. in attack area
Balls played in the left defensive side	M Lost ball poss. in defensive area
Balls played in the left midfield side	M Lost ball poss. in midfield area
Balls played in the middle channel	M Lost ball reaction
Balls played in the middle channel (avversari)	M Lost ball reaction in opponent half-field
Balls played in the middle channel in attack	M Lost ball reaction in own half-field
Balls played in the middle channel in defence	M Lost dribbles
Balls played in the middle channel in midfield	M Lost headed wall passes
Balls played on the left side	M Low catches
Balls played on the left side (avversari)	M Low catches (avversari)
Balls played on the right attack side	M Low catches from a cross play
Balls played on the right defensive side	M Low catches from cross
Balls played on the right midfield side	M Low catches from dribbling
Balls played on the right side	M Low catches from goal line cross
Balls played on the right side (avversari)	M Low catches from lob pass
Balls played on the sides in forward area	M Low catches from long pass
Balls won back	M Low catches from set play
Before assists	M Low catches from through pass
Before assists from a play initiated	M Low passes in opponent half field
Before assists from a set play initiated	M Low passes received
Before goal assists	M Midfield area throw-ins
Before goal assists from a play initiated	M Number of passes in shooting
Before goal assists from a set play initiated	M Number of passes in shooting (avversari)
Bicycle actions	M Nutmeg dribbles
Bicycle goals	M Occasioni dirette da angolo (avversari)
Bicycle goals (avversari)	M Occasioni dirette da punizione in attacco (avversari)
Bicycle shots	M Occasioni dirette su calcio piazzato (avversari)
Bicycle shots from a play	M Occasioni indirette da angolo (avversari)
Bicycle shots from a set play	M Occasioni indirette da laterale in attacco (avversari)
Catches	M Occasioni indirette da punizione in attacco (avversari)
Catches from a cross play	M Occasioni indirette su calcio piazzato (avversari)
Catches from a play	M Occasioni su azione (avversari)
Catches from set plays	M Occasioni su calcio piazzato (avversari)
Clearances	M Off-sides
Corners	M Offensive ability

Continued on next page

– *Continued from previous page*

Corners (avversari)	M Offensive headers
Counterattack	M Opponent's balls played in penalty area (avversari)
Cross field passes	M Outswinging crosses
Cross field passes (avversari)	M Outswinging crosses from a set play
Cross on a play from goal line	M Outswinging crosses on a play
Cross on a right wing move	M Outswinging crosses on a play from a goal line
Cross on a rleft wing move	M Overlaps
Cross su azione da destra (avversari)	M Own-goals
Cross su azione da sinistra (avversari)	M Own-goals against
Crosses	M Palle giocate in area avv./Rigori a favore
Crosses (avversari)	M Palle giocate in area dagli avv./Rigori contro
Crosses from a goal line play (avversari)	M Penalty kicks
Crosses from a set play	M Penalty kicks (avversari)
Crosses from a set play (avversari)	M Penalty kicks against
Crosses on a play	M Penalty kicks missed
Crosses on a play (avversari)	M Penalty kicks missed (avversari)
Crosses on a play between 35 and 18 mts	M Penalty kicks saves (avversari)
Defensive area throw-ins	M Penalty kicks scored
Defensive headers	M Play speed
Defensive headers won	M Play speed (avversari)
Defensive off-sides	M Played minutes
Direct goals from corner (avversari)	M Plays in central forward area
Direct goals from free kick (avversari)	M Plays in forward area
Direct red cards	M Plays initiated in defensive area (avversari)
Direct shots from corner (avversari)	M Plays on the sides of forward area
Direct shots from free kick (avversari)	M Poss. palla avv./Falli contro
Direct shots from free kicks	M Possesso palla/Falli a favore
Direct shots from throw-ins (avversari)	M Pressing (mt)
Direct shots off target	M Punizioni alte laterali
Double marking	M Punizioni alte laterali (avversari)
Doubling-up	M Punizioni alte laterali con tiro
Dribbles	M Punizioni alte laterali con tiro (avversari)
Effect. rec. balls in aerial duels	M Punizioni alte laterali destre
Effect. rec. balls with anticipation	M Punizioni alte laterali destre (avversari)
Effect. rec. balls with tackle	M Punizioni alte laterali destre con tiro
Effective interceptions	M Punizioni alte laterali destre con tiro (avversari)
Effective lost ball poss.	M Punizioni alte laterali destre senza tiro
Effective lost ball poss. in attack area	M Punizioni alte laterali destre senza tiro (avversari)
Effective lost ball poss. in defensive area	M Punizioni alte laterali senza tiro
Effective lost ball poss. in midfield area	M Punizioni alte laterali senza tiro (avversari)
Effective recov. ball poss. in attack area	M Punizioni alte laterali sinistre
Effective recovered balls	M Punizioni alte laterali sinistre (avversari)
Effective time	M Punizioni alte laterali sinistre con tiro
External diagonal run (avversari)	M Punizioni alte laterali sinistre con tiro (avversari)
External diagonal runs	M Punizioni alte laterali sinistre senza tiro
Fake falls	M Punizioni alte laterali sinistre senza tiro (avversari)
Fast breaks	M Punizioni basse laterali
Fast breaks (avversari)	M Punizioni basse laterali (avversari)
Flick headers	M Punizioni basse laterali con tiro
Flicked passes	M Punizioni basse laterali con tiro (avversari)
Fouls committed	M Punizioni basse laterali destre
Fouls committed in attack area	M Punizioni basse laterali destre (avversari)
Fouls committed in defensive area	M Punizioni basse laterali destre con tiro
Fouls committed in midfield area	M Punizioni basse laterali destre con tiro (avversari)
Fouls committed next to opponent penalty area	M Punizioni basse laterali destre senza tiro
Fouls committed next to own penalty area	M Punizioni basse laterali destre senza tiro (avversari)
Fouls suffered	M Punizioni basse laterali senza tiro
Fouls suffered in attack area	M Punizioni basse laterali senza tiro (avversari)
Fouls suffered in defensive area	M Punizioni basse laterali sinistre
Fouls suffered in midfield area	M Punizioni basse laterali sinistre (avversari)
Free kicks	M Punizioni basse laterali sinistre con tiro
Free kicks from attack area	M Punizioni basse laterali sinistre con tiro (avversari)
Free kicks from attack area (avversari)	M Punizioni basse laterali sinistre senza tiro
Free kicks from central attack area	M Punizioni basse laterali sinistre senza tiro (avversari)
Free kicks from central attack area (avversari)	M Punizioni centrali
Free kicks from central attack area with shot	M Punizioni centrali (avversari)
Free kicks from central attack area with shot (avversari)	M Punizioni centrali con tiro

Continued on next page

– Continued from previous page

Free kicks from central attack area without shot	M Punizioni centrali con tiro (avversari)
Free kicks from central attack area without shot (avversari)	M Punizioni centrali dal limite
Free kicks from central defensive area	M Punizioni centrali dal limite (avversari)
Free kicks from central midfield area	M Punizioni centrali dal limite con tiro
Free kicks from defensive area	M Punizioni centrali dal limite con tiro (avversari)
Free kicks from left attack side	M Punizioni centrali dal limite senza tiro
Free kicks from left attack side (avversari)	M Punizioni centrali dal limite senza tiro (avversari)
Free kicks from left attack side with shot	M Punizioni centrali senza tiro
Free kicks from left attack side with shot (avversari)	M Punizioni centrali senza tiro (avversari)
Free kicks from left attack side without shot	M Punts, throws and clearances
Free kicks from left attack side without shot (avversari)	M Quick passes
Free kicks from left defensive side	M Quick passes (avversari)
Free kicks from left midfield side	M Quickness in shooting
Free kicks from midfield area	M Quickness in shooting (avversari)
Free kicks from right attack side	M Received passes
Free kicks from right attack side (avversari)	M Received passes from cross on action
Free kicks from right attack side with shot	M Received passes from cross on set play
Free kicks from right attack side with shot (avversari)	M Received passes in attack
Free kicks from right attack side without shot	M Received passes in defence
Free kicks from right attack side without shot (avversari)	M Received passes in midfield
Free kicks from right defensive side	M Recov. ball poss. and lost afterwards
Free kicks from right midfield side	M Recov. ball poss. and lost afterwards in attack
Give-and-go passes	M Recov. ball poss. and lost afterwards.1
Give-and-go passes (avversari)	M Recov. ball poss. in attack area
Goal assists	M Recov. ball poss. in attack area (avversari)
Goal assists (avversari)	M Recov. ball poss. in central forward area
Goal assists from a play initiated	M Recov. ball poss. in defensive area
Goal assists from a set play initiated	M Recov. ball poss. in defensive area.1
Goal chances	M Recov. ball poss. in forward area
Goal chances (avversari)	M Recov. ball poss. in penalty area
Goal chances direct free kicks	M Recov. ball poss. on the sides in forward area
Goal chances from direct attack free kicks	M Recovered ball
Goal chances from direct corner	M Red cards
Goal chances from field goal	M Red cards due to double yellow cards
Goal chances from indirect attack throw-in	M Right attack side throw-ins
Goal chances from indirect corner	M Right attack side throw-ins with shot
Goal chances from indirect free kicks	M Right attack side throw-ins with shot (avversari)
Goal chances from indirect set play	M Right attack side throw-ins without shot
Goal chances from set play	M Right attack side throw-ins without shot (avversari)
Goal kicks	M Right defensive side throw-ins
Goal posts hit	M Right foot goals
Goalkeeper’s punts	M Right foot goals (avversari)
Goalkeeper’s punts and throws	M Right foot shots
Goalkeeper’s throw passes	M Right foot shots (avversari)
Goalkeeper’s useful punts and throws	M Right foot shots on target
Goals	M Right midfield side throw-ins
Goals against	M Right side corners
Goals against from 8 to 16 mt	M Right side corners (avversari)
Goals against from a direct set play	M Right side corners with shot
Goals against from an indirect set play	M Right side corners with shot (avversari)
Goals against from less than 8 mt	M Right side corners without shot
Goals against from more than 16 mt	M Right side corners without shot (avversari)
Goals against from penalty kicks	M Right side throw-ins
Goals against on action	M Saves
Goals from 8 to 16 mt	M Saves (avversari)
Goals from direct attack free kicks	M Saves from chance
Goals from direct corner	M Saves from penalty kicks
Goals from direct free kicks	M Saves from shots from inside penalty area
Goals from indirect attack throw-in	M Saves from shots from outside of penalty area
Goals from indirect corner	M Set plays in attack
Goals from indirect free kicks	M Short passes received
Goals from indirect set play	M Shots
Goals from inside goal area	M Shots (avversari)
Goals from inside goal area (avversari)	M Shots against
Goals from less than 8 mt	M Shots against distance (mt)
Goals from more than 16 mt	M Shots against on target
Goals from out of pen. area	M Shots distance (mt)

Continued on next page

– *Continued from previous page*

Goals from out of pen. area (avversari)	M Shots form set plays in attack
Hand crosses from a throw-in	M Shots from a play initiated
Head goals	M Shots from a play initiated (avversari)
Head goals (avversari)	M Shots from a play with ball on the ground
Head passes	M Shots from a set play with ball on the ground
Head passes received	M Shots from attack throw-in
Head shots	M Shots from central free kick (avversari)
Head shots (avversari)	M Shots from central free kicks
Head shots from a play	M Shots from direct set play
Head shots from a set play	M Shots from direct set play (avversari)
Head shots on target	M Shots from free kicks
Headed lay-off	M Shots from goal area (avversari)
Headed wall passes	M Shots from indirect set play
Headers	M Shots from indirect set play (avversari)
Headers in opposition area	M Shots from inside goal area
Headers lost	M Shots from inside penalty area
Headers won	M Shots from inside penalty area (avversari)
Headers won in area	M Shots from left side corner
Headers.1	M Shots from left side corner (avversari)
Heel passes	M Shots from left side free kick (avversari)
High catches	M Shots from left side free kicks
High catches (avversari)	M Shots from left side throw-in
High catches from a cross play	M Shots from left side throw-in (avversari)
High catches from cross	M Shots from out of penalty area
High catches from goal line cross	M Shots from out of penalty area (avversari)
High catches from lob pass	M Shots from penalty kicks
High catches from long pass	M Shots from right side corner
High catches from set play	M Shots from right side corner (avversari)
High passes received	M Shots from right side corner.1
Indirect goals from corner (avversari)	M Shots from right side corner.2
Indirect goals from free kick (avversari)	M Shots from right side free kick (avversari)
Indirect goals from throw-ins (avversari)	M Shots from right side free kicks
Indirect shots from corner (avversari)	M Shots from right side throw-in
Indirect shots from free kick (avversari)	M Shots from right side throw-in (avversari)
Inside shots from a play initiated	M Shots off target
Inswinging crosses	M Shots off target (avversari)
Inswinging crosses from a set play	M Shots off target (avversari).1
Inswinging crosses on a play	M Shots off target distance (mt)
Inswinging crosses on a play from a goal line	M Shots on target
Intentional hand balls	M Shots on target (avversari)
Interceptions	M Shots on target against distance (mt)
Internal diagonal run (avversari)	M Shots on target against from 8 to 16 mt
Internal diagonal runs	M Shots on target against from less than 8 mt
Lay-off	M Shots on target against from more than 16 mt
Left attack side throw-ins	M Shots on target distance (mt)
Left attack side throw-ins with shot	M Shots on target from 8 to 16 mt
Left attack side throw-ins with shot (avversari)	M Shots on target from less than 8 mt
Left attack side throw-ins without shot	M Shots on target from more than 16 mt
Left attack side throw-ins without shot (avversari)	M Shots rejected or deflected
Left defensive side throw-ins	M Shots rejected or deflected (avversari)
Left foot goals	M Shots rejected or deflected.1
Left foot goals (avversari)	M Shots with ball on the ground
Left foot shots	M Sideways runs
Left foot shots (avversari)	M Sideways runs (avversari)
Left foot shots on target	M Substitutes in
Left midfield side throw-ins	M Substitutes out
Left side corners	M Successful lay-offs
Left side corners (avversari)	M Successful passes
Left side corners with shot	M Successful passes (avversari)
Left side corners with shot (avversari)	M Successful passes in attack
Left side corners without shot	M Successful passes in defence
Left side corners without shot (avversari)	M Successful passes in midfield
Left side throw-ins	M Successful passes in opponent midfield
Lob passes	M Supremazia territoriale (avversari)
Lob passes (avversari)	M Tackles
Lob passes in the opponents' half field	M Team length
Lob passes received	M Team length (avversari)

Continued on next page

– Continued from previous page

Long passes	M Temporary recov. ball poss. in attack area
Long passes (avversari)	M Territorial supremacy
Long passes received	M Territorial supremacy in attack
Long passes without cross field pass	M Territorial supremacy in attack (avversari)
Long punts	M Through passes received
Lost ball poss.	M Throw passes
Lost ball poss. in attack area	M Throw-ins
Lost ball poss. in defensive area	M Time to recover ball
Lost ball poss. in midfield area	M Time to recover ball in opponent half-field
Lost dribbles	M Time to recover ball in own half-field
Lost headed wall passes	M Total balls played (avversari)
Low catches	M Total goals against from set play
Low catches (avversari)	M Total goals from actions with assist
Low catches from a cross play	M Total goals from actions without assist
Low catches from cross	M Total goals from field goal
Low catches from dribbling	M Total goals from set play
Low catches from goal line cross	M Total inside shots from set play
Low catches from lob pass	M Total shots from set play
Low catches from long pass	M Touchline passes
Low catches from set play	M Touchline passes (avversari)
Low catches from through pass	M Touchline passes received
Low passes in opponent half field	M Useful cross on a play from goal line
Low passes received	M Useful crosses
M % Effective lost ball poss.	M Useful crosses (avversari)
M % attack to goal	M Useful crosses and total made ratio (avversari)
M % average ball possession	M Useful crosses on a play
M % balls overpassing the midfield area	M Useful dribbles
M % balls overpassing the midfield area (avversari)	M Useful dribbles (avversari)
M % balls played in attack	M Useful flick headers
M % balls played in attack area (avversari)	M Useful footed wall passes
M % balls played in defence	M Useful head passes
M % balls played in defensive area (avversari)	M Useful headed wall passes
M % balls played in midfield	M Useful high passes
M % balls played in midfield area (avversari)	M Useful lob passes
M % balls played in opposition half	M Useful lob passes (avversari)
M % balls played in the middle channel	M Useful lob passes and total made ratio (avversari)
M % balls played in the middle channel (avversari)	M Useful lob passes in the opponents' half field
M % balls played on the left side	M Useful long passes
M % balls played on the left side (avversari)	M Useful long passes (avversari)
M % balls played on the right side	M Useful long passes and total made ratio (avversari)
M % balls played on the right side (avversari)	M Useful long passes without cross field pass
M % cross su azione da destra (avversari)	M Useful long punts
M % cross su azione da sinistra (avversari)	M Useful low passes
M % crosses on a play from the left side	M Useful low passes in opponent half field
M % crosses on a play from the right side	M Useful plays
M % effective recov. ball poss.	M Useful plays (avversari)
M % effective recov. ball poss. on totals	M Useful plays in attack
M % effective recov./Opponent's PB	M Useful plays in defence
M % effectiveness of forward set plays	M Useful plays in forward area
M % from balls played in forward area	M Useful plays in midfield
M % goals against	M Useful plays in opposition half
M % goals from actions with assist	M Useful plays in own half
M % goals from actions without assist	M Useful punts, throws and clearances
M % goals from chances	M Useful punts, throws and clearances ratio
M % goals from chances (avversari)	M Useful short passes
M % goals from chances on action	M Useful short passes received
M % goals from chances on action (avversari)	M Useful through passes
M % goals from shots	M Useful through passes (avversari)
M % goals from shots on target (avversari)	M Useful through passes and total made ratio (avversari)
M % headers lost	M Useful through passes in the opponents' half field
M % headers won	M Useful touchline passes
M % of aerial attacks on goal	M Useful touchline passes (avversari)
M % offensive ball possession	M Volley goals
M % plays developed from backward area	M Volley goals (avversari)
M % plays in central forward area	M Volley shots
M % plays initiated in defensive area (avversari)	M Volley shots from a play
M % plays on the sides of forward area	M Volley shots from a set play

Continued on next page

– Continued from previous page

M % protection of area	M Wall passes
M % protection of own penalty area	M Wall passes (avversari)
M % recov. ball poss. because of opponents' ended play	M Wall passes.1
M % shots on target	M Yellow cards
M % successful backward passes	M clearance received
M % successful forward passes	M dribbles (avversari)
M % successful passes	M dribbles with sharp turn
M % successful passes in opponent midfield	M through passes
M % successful side passes	M through passes (avversari)
M % temporary recov. ball poss.	M through passes in the opponents' half field
M % useful crosses	Midfield area throw-ins
M % useful dribbles	Number of passes in shooting
M % useful lob passes	Nutmeg dribbles
M % useful long passes	Occasioni dirette da angolo (avversari)
M % useful plays	Occasioni dirette da punizione in attacco (avversari)
M % useful plays from balls played (avversari)	Occasioni dirette su calcio piazzato (avversari)
M % useful plays in opposition half	Occasioni indirette da angolo (avversari)
M % useful through passes	Occasioni indirette da laterale in attacco (avversari)
M % useful touchline passes	Occasioni indirette da punizione in attacco (avversari)
M Accelerations	Occasioni indirette su calcio piazzato (avversari)
M Acrobatic shots	Occasioni su azione (avversari)
M Acrobatic shots from a play	Occasioni su calcio piazzato (avversari)
M Acrobatic shots from a set play	Off-sides
M Anticipations	Offensive headers
M Anticipations by head	Opponent's balls played in penalty area (avversari)
M Assists	Outswinging crosses
M Assists (avversari)	Outswinging crosses from a set play
M Assists from a play initiated	Outswinging crosses on a play
M Assists from a set play initiated	Outswinging crosses on a play from a goal line
M Attack TEI	Overlaps
M Attack TEI.1	Own-goals
M Attack ability	Own-goals against
M Attack area throw-ins	Penalty kicks
M Attack area throw-ins (avversari)	Penalty kicks (avversari)
M Attempted passes	Penalty kicks against
M Attempted passes in opponent midfield	Penalty kicks missed
M Ball possession	Penalty kicks missed (avversari)
M Ball possession (avversari)	Penalty kicks saves (avversari)
M Balls overpassing the midfield area (avversari)	Penalty kicks scored
M Balls played	Played minutes
M Balls played (avversari)	Plays in central forward area
M Balls played in attack	Plays in forward area
M Balls played in attack area (avversari)	Plays initiated in defensive area (avversari)
M Balls played in central forward area	Plays on the sides of forward area
M Balls played in defence	Punizioni alte laterali
M Balls played in defensive area (avversari)	Punizioni alte laterali (avversari)
M Balls played in forward area	Punizioni alte laterali con tiro
M Balls played in midfield	Punizioni alte laterali con tiro (avversari)
M Balls played in midfield area (avversari)	Punizioni alte laterali destre
M Balls played in opposition half	Punizioni alte laterali destre (avversari)
M Balls played in penalty area	Punizioni alte laterali destre con tiro
M Balls played in the left attack side	Punizioni alte laterali destre con tiro (avversari)
M Balls played in the left defensive side	Punizioni alte laterali destre senza tiro
M Balls played in the left midfield side	Punizioni alte laterali destre senza tiro (avversari)
M Balls played in the middle channel	Punizioni alte laterali senza tiro
M Balls played in the middle channel (avversari)	Punizioni alte laterali senza tiro (avversari)
M Balls played in the middle channel in attack	Punizioni alte laterali sinistre
M Balls played in the middle channel in defence	Punizioni alte laterali sinistre (avversari)
M Balls played in the middle channel in midfield	Punizioni alte laterali sinistre con tiro
M Balls played on the left side	Punizioni alte laterali sinistre con tiro (avversari)
M Balls played on the left side (avversari)	Punizioni alte laterali sinistre senza tiro
M Balls played on the right attack side	Punizioni alte laterali sinistre senza tiro (avversari)
M Balls played on the right defensive side	Punizioni basse laterali
M Balls played on the right midfield side	Punizioni basse laterali (avversari)
M Balls played on the right side	Punizioni basse laterali con tiro
M Balls played on the right side (avversari)	Punizioni basse laterali con tiro (avversari)
M Balls played on the sides in forward area	Punizioni basse laterali destre

Continued on next page

– Continued from previous page

M Balls won back	Punizioni basse laterali destre (avversari)
M Barycentre (avversari)	Punizioni basse laterali destre con tiro
M Barycentre (mt)	Punizioni basse laterali destre con tiro (avversari)
M Before assists	Punizioni basse laterali destre senza tiro
M Before assists from a play initiated	Punizioni basse laterali destre senza tiro (avversari)
M Before assists from a set play initiated	Punizioni basse laterali senza tiro
M Before goal assists	Punizioni basse laterali senza tiro (avversari)
M Before goal assists from a play initiated	Punizioni basse laterali sinistre
M Before goal assists from a set play initiated	Punizioni basse laterali sinistre (avversari)
M Bicycle actions	Punizioni basse laterali sinistre con tiro
M Bicycle goals	Punizioni basse laterali sinistre con tiro (avversari)
M Bicycle goals (avversari)	Punizioni basse laterali sinistre senza tiro
M Bicycle shots	Punizioni basse laterali sinistre senza tiro (avversari)
M Bicycle shots from a play	Punizioni centrali
M Bicycle shots from a set play	Punizioni centrali (avversari)
M Catches	Punizioni centrali con tiro
M Catches from a cross play	Punizioni centrali con tiro (avversari)
M Catches from a play	Punizioni centrali dal limite
M Catches from set plays	Punizioni centrali dal limite (avversari)
M Clearances	Punizioni centrali dal limite con tiro
M Corners	Punizioni centrali dal limite con tiro (avversari)
M Corners (avversari)	Punizioni centrali dal limite senza tiro
M Counterattack	Punizioni centrali dal limite senza tiro (avversari)
M Cross field passes	Punizioni centrali senza tiro
M Cross field passes (avversari)	Punizioni centrali senza tiro (avversari)
M Cross on a play from goal line	Punts, throws and clearances
M Cross on a right wing move	Quick passes
M Cross on a rleft wing move	Quick passes (avversari)
M Cross su azione da destra (avversari)	Quickness in shooting
M Cross su azione da sinistra (avversari)	Received passes
M Crosses	Received passes from cross on action
M Crosses (avversari)	Received passes from cross on set play
M Crosses from a goal line play (avversari)	Received passes in attack
M Crosses from a set play	Received passes in defence
M Crosses from a set play (avversari)	Received passes in midfield
M Crosses on a play	Recov. ball poss. and lost afterwards
M Crosses on a play (avversari)	Recov. ball poss. and lost afterwards in attack
M Crosses on a play between 35 and 18 mts	Recov. ball poss. and lost afterwards.1
M Defence TEI	Recov. ball poss. in attack area
M Defense ability	Recov. ball poss. in attack area (avversari)
M Defensive ability	Recov. ball poss. in central forward area
M Defensive area throw-ins	Recov. ball poss. in defensive area
M Defensive headers	Recov. ball poss. in defensive area.1
M Defensive headers won	Recov. ball poss. in forward area
M Defensive off-sides	Recov. ball poss. in penalty area
M Direct goals from corner (avversari)	Recov. ball poss. on the sides in forward area
M Direct goals from free kick (avversari)	Recovered ball
M Direct red cards	Red cards
M Direct shots from corner (avversari)	Red cards due to double yellow cards
M Direct shots from free kick (avversari)	Right attack side throw-ins
M Direct shots from free kicks	Right attack side throw-ins with shot
M Direct shots from throw-ins (avversari)	Right attack side throw-ins with shot (avversari)
M Direct shots off target	Right attack side throw-ins without shot
M Double marking	Right attack side throw-ins without shot (avversari)
M Doubling-up	Right defensive side throw-ins
M Dribbles	Right foot goals
M Effect. rec. balls in aerial duels	Right foot goals (avversari)
M Effect. rec. balls with anticipation	Right foot shots
M Effect. rec. balls with tackle	Right foot shots (avversari)
M Effective interceptions	Right foot shots on target
M Effective lost ball poss.	Right midfield side throw-ins
M Effective lost ball poss. in attack area	Right side corners
M Effective lost ball poss. in defensive area	Right side corners (avversari)
M Effective lost ball poss. in midfield area	Right side corners with shot
M Effective recov. ball poss. in attack area	Right side corners with shot (avversari)
M Effective recovered balls	Right side corners without shot
M Effective time	Right side corners without shot (avversari)

Continued on next page

– *Continued from previous page*

M External diagonal run (avversari)	Right side throw-ins
M External diagonal runs	Saves
M Fake falls	Saves (avversari)
M Falli contro/Ammonizioni	Saves from chance
M Falli contro/Espulsioni	Saves from penalty kicks
M Fast breaks	Saves from shots from inside penalty area
M Fast breaks (avversari)	Saves from shots from outside of penalty area
M Flick headers	Set plays in attack
M Flicked passes	Short passes received
M Fouls committed	Shots
M Fouls committed in attack area	Shots (avversari)
M Fouls committed in defensive area	Shots against
M Fouls committed in midfield area	Shots against on target
M Fouls committed next to opponent penalty area	Shots form set plays in attack
M Fouls committed next to own penalty area	Shots from a play initiated
M Fouls suffered	Shots from a play initiated (avversari)
M Fouls suffered in attack area	Shots from a play with ball on the ground
M Fouls suffered in defensive area	Shots from a set play with ball on the ground
M Fouls suffered in midfield area	Shots from attack throw-in
M Free kicks	Shots from central free kick (avversari)
M Free kicks from attack area	Shots from central free kicks
M Free kicks from attack area (avversari)	Shots from direct set play
M Free kicks from central attack area	Shots from direct set play (avversari)
M Free kicks from central attack area (avversari)	Shots from free kicks
M Free kicks from central attack area with shot	Shots from goal area (avversari)
M Free kicks from central attack area with shot (avversari)	Shots from indirect set play
M Free kicks from central attack area without shot	Shots from indirect set play (avversari)
M Free kicks from central attack area without shot (avversari)	Shots from inside goal area
M Free kicks from central defensive area	Shots from inside penalty area
M Free kicks from central midfield area	Shots from inside penalty area (avversari)
M Free kicks from defensive area	Shots from left side corner
M Free kicks from left attack side	Shots from left side corner (avversari)
M Free kicks from left attack side (avversari)	Shots from left side free kick (avversari)
M Free kicks from left attack side with shot	Shots from left side free kicks
M Free kicks from left attack side with shot (avversari)	Shots from left side throw-in
M Free kicks from left attack side without shot	Shots from left side throw-in (avversari)
M Free kicks from left attack side without shot (avversari)	Shots from out of penalty area
M Free kicks from left defensive side	Shots from out of penalty area (avversari)
M Free kicks from left midfield side	Shots from penalty kicks
M Free kicks from midfield area	Shots from right side corner
M Free kicks from right attack side	Shots from right side corner (avversari)
M Free kicks from right attack side (avversari)	Shots from right side corner.1
M Free kicks from right attack side with shot	Shots from right side corner.2
M Free kicks from right attack side with shot (avversari)	Shots from right side free kick (avversari)
M Free kicks from right attack side without shot	Shots from right side free kicks
M Free kicks from right attack side without shot (avversari)	Shots from right side throw-in
M Free kicks from right defensive side	Shots from right side throw-in (avversari)
M Free kicks from right midfield side	Shots off target
M Frequency goals chances	Shots off target (avversari)
M Frequency goals chances against	Shots off target (avversari).1
M GEI	Shots on target
M General TEI	Shots on target (avversari)
M Give-and-go passes	Shots on target against from 8 to 16 mt
M Give-and-go passes (avversari)	Shots on target against from less than 8 mt
M Goal assists	Shots on target against from more than 16 mt
M Goal assists (avversari)	Shots on target from 8 to 16 mt
M Goal assists from a play initiated	Shots on target from less than 8 mt
M Goal assists from a set play initiated	Shots on target from more than 16 mt
M Goal chances	Shots rejected or deflected
M Goal chances (avversari)	Shots rejected or deflected (avversari)
M Goal chances direct free kicks	Shots rejected or deflected.1
M Goal chances from direct attack free kicks	Shots with ball on the ground
M Goal chances from direct corner	Sideways runs
M Goal chances from field goal	Sideways runs (avversari)
M Goal chances from indirect attack throw-in	Substitutes in
M Goal chances from indirect corner	Substitutes out
M Goal chances from indirect free kicks	Successful lay-offs

Continued on next page

– Continued from previous page

M Goal chances from indirect set play	Successful passes
M Goal chances from set play	Successful passes (avversari)
M Goal kicks	Successful passes in attack
M Goal posts hit	Successful passes in defence
M Goalkeeper's punts	Successful passes in midfield
M Goalkeeper's punts and throws	Successful passes in opponent midfield
M Goalkeeper's throw passes	Supremazia territoriale (avversari)
M Goalkeeper's useful punts and throws	Tackles
M Goals	Temporary recov. ball poss. in attack area
M Goals against	Territorial supremacy
M Goals against and shots on target ratio (avversari)	Territorial supremacy in attack
M Goals against distance (mt)	Territorial supremacy in attack (avversari)
M Goals against from 8 to 16 mt	Through passes received
M Goals against from a direct set play	Throw passes
M Goals against from an indirect set play	Throw-ins
M Goals against from less than 8 mt	Total balls played (avversari)
M Goals against from more than 16 mt	Total goals against from set play
M Goals against from penalty kicks	Total goals from actions with assist
M Goals against on action	Total goals from actions without assist
M Goals distance (mt)	Total goals from field goal
M Goals from 8 to 16 mt	Total goals from set play
M Goals from direct attack free kicks	Total inside shots from set play
M Goals from direct corner	Total shots from set play
M Goals from direct free kicks	Touchline passes
M Goals from indirect attack throw-in	Touchline passes (avversari)
M Goals from indirect corner	Touchline passes received
M Goals from indirect free kicks	Useful cross on a play from goal line
M Goals from indirect set play	Useful crosses
M Goals from inside goal area	Useful crosses (avversari)
M Goals from inside goal area (avversari)	Useful crosses on a play
M Goals from less than 8 mt	Useful dribbles
M Goals from more than 16 mt	Useful dribbles (avversari)
M Goals from out of pen. area	Useful flick headers
M Goals from out of pen. area (avversari)	Useful footed wall passes
M Hand crosses from a throw-in	Useful head passes
M Head goals	Useful headed wall passes
M Head goals (avversari)	Useful high passes
M Head passes	Useful lob passes
M Head passes received	Useful lob passes (avversari)
M Head shots	Useful lob passes in the opponents' half field
M Head shots (avversari)	Useful long passes
M Head shots from a play	Useful long passes (avversari)
M Head shots from a set play	Useful long passes without cross field pass
M Head shots on target	Useful long punts
M Headed lay-off	Useful low passes
M Headed wall passes	Useful low passes in opponent half field
M Headers	Useful plays
M Headers in opposition area	Useful plays (avversari)
M Headers lost	Useful plays in attack
M Headers won	Useful plays in defence
M Headers won in area	Useful plays in forward area
M Headers.1	Useful plays in midfield
M Heel passes	Useful plays in opposition half
M High catches	Useful plays in own half
M High catches (avversari)	Useful punts, throws and clearances
M High catches from a cross play	Useful short passes
M High catches from cross	Useful short passes received
M High catches from goal line cross	Useful through passes
M High catches from lob pass	Useful through passes (avversari)
M High catches from long pass	Useful through passes in the opponents' half field
M High catches from set play	Useful touchline passes
M High passes received	Useful touchline passes (avversari)
M Indirect goals from corner (avversari)	Volley goals
M Indirect goals from free kick (avversari)	Volley goals (avversari)
M Indirect goals from throw-ins (avversari)	Volley shots
M Indirect shots from corner (avversari)	Volley shots from a play
M Indirect shots from free kick (avversari)	Volley shots from a set play

Continued on next page

– *Continued from previous page*

M Inside shots from a play initiated	Wall passes
M Inswinging crosses	Wall passes (avversari)
M Inswinging crosses from a set play	Wall passes.1
M Inswinging crosses on a play	Yellow cards
M Inswinging crosses on a play from a goal line	clearance received
M Intentional hand balls	dribbles (avversari)
M Interceptions	dribbles with sharp turn
M Internal diagonal run (avversari)	through passes
M Internal diagonal runs	through passes (avversari)
M Lay-off	through passes in the opponents' half field

B The first 17 columns of the data sheet on the teams in the Italian Serie A in the season 2019/20 used in this research

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Team	Points	Played	Won	Drawn	Lost	Goal favour	Goal against	GENERAL DATA	M GENERAL DATA	Played minutes	M Played minutes	Effective time	M Effective time	Balls played	M Balls played	Balls played in opposition half
2	Squadra	Punti	Giocate	Vinte	Pareggiate	Perse	Reti fatte	Reti subite	DATI GENERALI	M DATI GENERALI	Minuti di gioco	M Minuti di gioco	Tempo effettivo	M Tempo effettivo	Palle giocate	M Palle giocate	Palle giocate nella metà campo avversaria
3	~~~ MEDIA ~~~	53	38	15	9	15	58	58			3 695'	97,2'	1 939':17"	51':02"	21 370	562,4	9 825
4	Atalanta	78	38	23	9	6	98	48			3 647'	96'	1 960':06"	51':35"	24 594	647,2	12 545
5	Bologna	47	38	12	11	15	52	65			3 700'	97,4'	1 853':11"	48':46"	21 367	562,3	10 152
6	Brescia	25	38	6	7	25	35	79			3 691'	97,1'	1 883':14"	49':34"	17 607	463,3	7 399
7	Cagliari	45	38	11	12	15	52	56			3 724'	98'	1 930':44"	50':49"	19 750	519,7	8 463
8	Fiorentina	49	38	12	13	13	51	48			3 722'	97,9'	1 916':55"	50':27"	19 944	524,8	9 107
9	Genoa	39	38	10	9	19	47	73			3 687'	97'	1 933':06"	50':52"	20 713	545,1	8 499
10	Hellas Verona	49	38	12	13	13	47	51			3 701'	97,4'	1 884':31"	49':36"	20 232	532,4	9 211
11	Inter	82	38	24	10	4	81	36			3 667'	96,5'	1 973':44"	51':56"	23 211	610,8	10 691
12	Juventus	83	38	26	5	7	76	43			3 706'	97,5'	1 971':03"	51':52"	25 781	678,4	13 834
13	Lazio	78	38	24	6	8	79	42			3 677'	96,8'	2 032':03"	53':29"	22 046	580,2	10 474
14	Lecce	35	38	9	8	21	52	85			3 707'	97,6'	1 990':01"	52':22"	19 662	517,4	8 043
15	Milan	66	38	19	9	10	63	46			3 690'	97,1'	1 943':42"	51':09"	23 288	612,8	11 411
16	Napoli	62	38	18	8	12	61	50			3 714'	97,7'	2 028':38"	53':23"	26 532	698,2	13 772
17	Parma	49	38	14	7	17	56	57			3 705'	97,5'	1 941':38"	51':06"	18 917	497,8	7 683
18	Roma	70	38	21	7	10	77	51			3 702'	97,4'	1 959':55"	51':35"	22 056	580,4	10 473
19	Sampdoria	42	38	12	6	20	48	65			3 698'	97,3'	1 891':20"	49':46"	18 617	489,9	8 486
20	Sassuolo	51	38	14	9	15	69	63			3 691'	97,1'	1 940':59"	51':05"	24 328	640,2	11 377
21	Spal	20	38	5	5	28	27	77			3 703'	97,4'	1 895':29"	49':53"	19 383	510,1	8 201
22	Torino	40	38	11	7	20	46	68			3 680'	96,8'	1 834':54"	48':17"	19 373	509,8	8 220
23	Udinese	45	38	12	9	17	37	51			3 694'	97,2'	2 020':22"	53':10"	20 006	526,5	8 464

C The first 17 columns of the data sheet on the players in the Italian Serie A in the season 2019/20 used in this research

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Player	Specific position	Nation	Birth date	Club	Team	Match played	Minutes played		CONTRIBUTION TO TEAM	M CONTRIBUTION TO TEAM	Plus/Minus	M Plus/Minus
2	Giocatore	Ruolo specifico	Nazione	Data di nascita	Club	Squadra	Presenze	Minuti giocati		CONTRIBUTO ALLA SQUADRA	M CONTRIBUTO ALLA SQUADRA	Plus/Minus	M Plus/Minus
3	Acerbi Francesco	Difensore centrale	Italia	10-8-1988	Lazio	Lazio	36	3396	94			-	2,9
4	Adjapong Claud	Centrocampista esterno destro	Italia	6-5-1998	Lecce	Hellas Verona	5	266	53			-	-7,9
5	Adopo Michel Ndary	Centrocampista offensivo	Francia	19-7-2000	Torino	Torino	2	10	5			-	-3,4
6	Adrien Tamèze	Centrocampista interno destro	Francia	4-2-1994	Hellas Verona	Atalanta	7	233	33			-	6,2
7	Agoume Lucien	Centrocampista interno sinistro	Francia	9-2-2002	Spezia	Inter	3	76	25			-	-6,5
8	Alan Empereur	Difensore centrale sinistro	Brasile	10-3-1994	Palmeiras (BRA)	Hellas Verona	13	946	73			-	-8,2
9	Alejandro Gomez	Centrocampista offensivo	Argentina	15-2-1988	Atalanta	Atalanta	36	2984	83			-	4
10	Alex Berenguer	Centrocampista offensivo	Spagna	4-7-1995	Athlétic Bilbao	Torino	29	1956	67			-	4,4
11	Alex Sandro	Difensore esterno sinistro	Brasile	26-1-1991	Juventus	Juventus	29	2585	89			-	-3,8
12	Alexis Sanchez	Attaccante seconda punta	Cile	19-12-1988	Inter	Inter	22	1013	46			-	5,6
13	Alfonso Enrico	Portiere	Italia	4-5-1988	Cremonese	Brescia	4	298	75			-	15
14	Allan	Centrocampista interno destro	Brasile	8-1-1991	Everton	Napoli	23	1362	59			-	-6,5
15	Amrabat Sofyan	Centrocampista interno destro	Marocco	21-8-1996	Fiorentina	Hellas Verona	34	3120	92			-	1,2
16	Anderson Djavan Lorenzo	Centrocampista esterno sinistro	Olanda	21-4-1995	Lazio	Lazio	5	188	38			-	-16,9
17	André Anderson	Centrocampista interno sinistro	Italia	23-9-1999	Salernitana	Lazio	7	94	13			-	3

D The Y variable: the top teams and the lesser teams

Club		Points	Top 7
1	Juventus	83	True
2	Inter Milan	82	True
3	Atalanta	78	True
4	Lazio	78	True
5	AS Roma	70	True
6	AC Milan	66	True
7	Napoli	62	True
8	Sassuolo	51	False
9	Verona	49	False
10	Fiorentina	49	False
11	Parma	49	False
12	Bologna	47	False
13	Udinese	45	False
14	Cagliari	45	False
15	Sampdoria	42	False
16	Torino	40	False
17	Genoa	39	False
18	Lecce	35	False
19	Brescia	25	False
20	SPAL	20	False