

National College of Ireland

Project Submission Sheet – 2015/2016

School of Computing

Student Name: Siddharth Thanga Mariappan

 15000401
Student ID:
 M.Sc. Data Analytics 2016
Programme: **Year:**
 Research in Computing
Module:
 Dr. Jason Roche
Lecturer:
Submission Due Date: 22-Aug-16

Project Title: Can Machine Learning bring glory to a football team?

 4400
Word Count:

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Siddharth Thanga Mariappan

 22-Aug-16
Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. **Please do not bind projects or place in covers unless specifically requested.**
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| Office Use Only | |
|----------------------------------|--|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

CAN MACHINE LEARNING BRING GLORY TO A FOOTBALL TEAM?

SIDDHARTH THANGA MARIAPPAN



National
College *of*
Ireland

SUBMITTED AS PART OF THE REQUIREMENTS FOR THE DEGREE
OF MSc IN DATA ANALYTICS
AT THE SCHOOL OF COMPUTING,
NATIONAL COLLEGE OF IRELAND
DUBLIN, IRELAND.

August 2016

Supervisor Dr. Jason Roche

Abstract

Sports analytics in the past decade has grown to an extent more than ever. Analytics have visibly reshaped how sports are being played compared to the past. The evolution of technologies and the data collection technique improvement has enabled the video analysts to catch each and every movement and occurrences of the game. The project has handled a set of Spanish league data for Real Madrid football club for the last 15 years and the tactical datasets of the Real Madrid season 10-11 and season 15-16. With many of the past researches not considering the impact that a manger brought into his team and the fatigue of the players, this project will resolve the previously aroused doubts about trusting the work done. The techniques used for analysis purposes were decision trees and Anova.

Acknowledgements

The dissertation work on Can Machine Learning bring glory to a football team? for Masters in Data Analytics is accomplished at National College of Ireland, Cloud Competency Center.

I would like to extend my gratitude and thankfulness to my supervisor Dr. Jason Roche. I would like to sincerely thank him for guidance, motivation and discussions during the course work. Every meeting helped me understand my goal and target to achieve the desired results in the research work.

I would also like to thank my family and friends who helped me directly and indirectly in order to complete the research work.

Declaration

I confirm that the work contained in this MSc project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specially acknowledged and all verbatim extracts are distinguished by quotation marks.

Place: Dublin

Signed(Siddharth Thanga Mariappan)

Date

Contents

| | |
|--|------------|
| Abstract | ii |
| Acknowledgements | iii |
| Declaration | iv |
| 1 Introduction | 1 |
| 1.1 Overview | 1 |
| 1.2 Manager's Work Rate | 1 |
| 1.3 Thesis Contribution | 2 |
| 1.4 Thesis Structure | 3 |
| 2 Literature Review | 4 |
| 2.1 Background | 4 |
| 2.2 Different Attributes | 5 |
| 2.3 Feature Engineering | 5 |
| 2.4 Machine Learning Technique - Decision Tree | 5 |
| 2.5 Statistical Analysis - Anova | 6 |
| 2.6 Summary | 6 |
| 3 Methodologies | 7 |
| 3.1 The Sample Dataset with Example Procedure | 7 |
| 3.2 Feature Engineering | 7 |
| 3.3 Statistical Testing | 8 |
| 3.3.1 ANOVA Test Case | 10 |
| 3.4 Machine Learning - Classification | 10 |
| 3.5 Summary | 10 |
| 4 Results and Findings | 11 |
| 4.1 Through Anova Hypothesis Testing | 11 |
| 4.2 Through Machine Learning Classification - Season 2010/2011 | 11 |

| | | |
|----------|--|-----------|
| 4.3 | Through Machine Learning Classification - Season 2015/2016 | 12 |
| 4.4 | Summary | 13 |
| 5 | Conclusion | 14 |
| | Bibliography | 15 |
| A | Using L^AT_EX | 17 |
| A.1 | Structure of this Template | 17 |
| A.2 | Using Figures | 18 |
| A.3 | Referencing | 19 |
| A.4 | Citing Bibliographic References | 20 |
| | A.4.1 Compiling a BibTeX Database | 20 |
| | A.4.2 The Vancouver Bibliography Style | 20 |
| A.5 | Working with Tables | 21 |
| A.6 | Inserting an Algorithm | 21 |
| A.7 | Inserting Program Code Samples | 21 |
| A.8 | Working with Maths | 22 |
| A.9 | Required Software | 23 |
| A.10 | Working with Quotes | 24 |
| A.11 | Further Information | 24 |
| A.12 | Conclusion | 24 |

List of Figures

| | | |
|-----|---|----|
| 3.1 | Feature | 8 |
| 3.2 | Data types | 9 |
| 3.3 | Classification | 10 |
| 4.1 | Test Results | 11 |
| 4.2 | ML Classification | 12 |
| 4.3 | 15-16 | 13 |
| A.1 | Caption for a Chart | 18 |
| A.2 | Mobile Phone Suppliers, Market Share 2006 (Millions of Units Shipped) | 19 |
| A.3 | Caption for Figure | 19 |
| A.4 | Caption for Bitmap Image Example | 19 |

List of Algorithms

| | | |
|---|--|----|
| 1 | A Sample Algorithm | 22 |
|---|--|----|

Listings

| | |
|--|----|
| A.1 The Caption for the Code Listing | 22 |
| A.2 Sample Program Code Listing | 22 |

Chapter 1

Introduction

1.1 Overview

According to Biggestglobalsports.com, Association Football is the most popular and biggest sport around the globe. The amount of money that is spent by each club on football is increasing every year. The tactics of a manager also holds the difference between being a winning team or a losing team. There is no such thing called as good or bad Manager. It is being right and what works for the club. Bringing the right manager for a club is essential to a team's on-field success right manager can keep the squad happy and bring success to a team [1]. Gary linker believes the right manager in the modern era would be able to keep a player happy even if he is not in the team. Different managers adopt to different playing styles and they call it as philosophy in their terms. Through Statistical Analysis and Machine learning this Manager's work rate on the team's success can be addressed through various metrics.

1.2 Manager's Work Rate

[13] Perry describes the manager position in a football club as organizationally vague. often the managers end up having no job description, or even clearly specified objectives or clear accountabilities [13]. This vagueness and variability in remit and expectations is also highlighted on further researches [17]. He believes that the manager has responsibility for strategy (e.g. playing style), operational tactics (e.g. game decisions), player development, opponent analysis. Therefore the work rate of a Manager is the result that the team has attained through the roles and the active changes that he has made for every matches. The important thing that the manager has to consider over

here in order to attain these goals without failing is fatigue of a player. Fatigue mainly occurs when a player is overused through the season without proper rest. The squad rotation comes into play here and implies the importance of squad Strength. As the former Manchester United Manager and Player Ryan Giggs admitted on BBC about the essentials of squad strength in attaining success. The Tactics of the Manager plays a key role in determining the result of the game. Different Managers follow different Tactics. It includes from providing the individual player roles to the formations and styles of the team when they are winning and losing. The Manager also needs to change tactics according to the opponent's playing style [20]. The Playing style is something that the manager has to adopt according to the resources he has in the team to be successful and stick with it for a longer time irrespective of the result. Generally a team manager adopts to either counter attacking football or possession football. Each has its own positives and negatives. This possession game is also called as Tiki-Taka which revolves the game around short passing till a hole is found on the opponents defense. The Counter Attacking football is the team defend very deep till they get the ball and when they have ball they look to hit on opponents immediately in numbers than normal. The managers generally choose one of them and convey them into their players. The manager also has to choose whether he wants to protect his lead or to score one more and kill the match. One way of predicting performance and team behavior is to examine the tactical strategies of individuals or teams with a view to identify any common patterns that exists. These tactical observation therefore serves two purposes. It not only provides a conceptual basis to coaching theory, but also provides a meaningful and useful practical tool for the analysts, coach and the performers[11]. The tactics of the manager when chosen correctly against an opposition can have a huge impact in the game result. These tactics are analyzed using a machine learning classification technique in-order to establish the relationship between the result and the tactics. The rest day that the team had for each match are further statistically analyzed using T-test to find out the impacts of fatigue and the importance of squad rotation in a game. The team that is winning at half time and failed to win the match at full time can also identified through this statistical analysis. So, the findings through tactical analysis, fatigue analysis and opting to choose between protecting the lead or to kill the match paves way to find out the strength of a manager.

1.3 Thesis Contribution

The motive of this research is to utilize machine learning and statistical analysis inside the football field. In particular the tactics of the manager has been classified using machine learning algorithm to find out the most important attribute that affects the

result. Also, the fatigue of a player that the manager needs to consider and its impacts has been founded out.

1.4 Thesis Structure

The research paper is followed with the structure as mentioned here. Chapter 2-provides details on background information about this research. Chapter 3-describes the various attributes and methodologies used on this research. Chapter 4-explains the results and its evaluation. Chapter 5- presents the conclusion and the future work of the research.

Chapter 2

Literature Review

2.1 Background

The success of a team cannot be measured through a single match but only through a period or a season. Match analysis is generally the recording of behavioral events on a single game or a competition and then examining through it for new findings and learnings [15]. It allows the manager to identify the weaknesses and areas of improvement inside his team and to exploit the weakness of the opposition. The attributes that are noted during matches analyzed uniquely or with the combination of other attributes to define some aspect of performance which in turn helps to achieve the success is called as Performance indicators [10]. The Performance indicators can further be used to build a statistical or machine learning model and predict the future result or behavior of a sport [19]. Most of the researches ([14], [21]) focus upon goal scoring and identifying patterns of build-up play which leading to shots. However, playing styles within the past researches have shown different findings. For example Hughes, ([9]) found that teams who utilised wings were failed to progress beyond the group stages of the world cup 1986. This study proves that teams plays different styles and which distinguishes them between winning team and losing team. However these findings may be less applicable to modern football due to the time period in which it was conducted. Low et al., 2002 ([22]) has reported that on 40 matches within the 2002 soccer World Cup which produced similar results to those of Hughes et al., 1988 ([9]). For example Griffiths ([6]) has reported that France was able to create more shots while also having the ability to retain possession for long periods.

2.2 Different Attributes

Each manager has its own approach and philosophy towards football. Gonzalez-Rodenas ([5]) has founded out the strikers in the team usually finds it difficult when playing under a new manager . Hughes et al ([10]) has said that limitations arises from the usage different variables in the analysis. But these variables are great in order to get the knowledge about the game. (Bruinshoofd and Ter Weel; 2003) also founded out the manager has an impact on his player other than the tactics that he puts to display on the pitch.

2.3 Feature Engineering

Feature engineering is nothing but finding and extracting the information that is hidden on the dataset. This feature engineering plays an important role in statistical analysis and machine learning ([3]). The feature engineering will also help to avoid cross validation of the datasets ([18]). Minimal optimal problem and all relevant problem are the two problems that are expected in feature engineering ([7]). Minimum optimal method relies on the small set of features whereas all relevant relies on large set of features.

2.4 Machine Learning Technique - Decision Tree

Predicting the score lines of football matches or any other sport poses an tough and interesting challenge due to the fact that the sport is so popular, widespread and ever changing. predicting the outcomes becomes a difficult problem because of the number of factors in the game which must be taken into account that cannot be quantitatively valued or modeled. Machine Learning is a technique that could very well address these problems and bring out a solution ([8]). Bivariate Poisson regression has been used to built the forecasting models for goals scored and conceded[m2]. Ordered probit regression is used to estimate the forecasting models for match results ([4]). Bayesian nets has been combined with the decision tree ([12]) which had a better output, but it was a complicated model that can make serious mistakes with the lesser number of input samples.

2.5 Statistical Analysis - Anova

Statistical Analysis is one of the best technique to see if there is a difference exists between two groups ([2]). The statistical techniques are Anova, Manova and T-test. Statistical Analysis Anova has been used on most of the analysis that has been done on football before. The reason being Anova is very handy when the independent variable is continuous and the dependent is binary ([16]).

2.6 Summary

The past researches has been studied carefully. The various attributes and their outcomes were identified from the past works. Also, the techniques that are mostly used in the football analytics are also reviewed.

Chapter 3

Methodologies

3.1 The Sample Dataset with Example Procedure

Real Madrid football club which plays in the Spanish League has been chosen for this study. The data about the match date and the result at half time and full time for all the matches of Spanish League has been collected from Sports data mart website. This data holds record for the last 15 Seasons. There were a total of 20 teams played against each other twice in each season. The match date and the result will help to evaluate the effects of fatigue in match result. To understand the importance of manager Real Madrid's 2010-2011 Season is chosen and the data has been extracted from whoscored.com website by using web scraping technique. Python-selenium has been used to do the web scraping effectively. Selenium is generally used to scrape web pages with dynamic contents .The tactics along with the result will help to find out the manager's work rate with the club. The experimental procedure for this research is as follows. The impacts of fatigue is analyzed prior to the tactics of a manager. The data holds records for all the 20 teams which is filtered out to feature only Real Madrid matches in excel and saved for analysis. In-order to analyze the fatigue the days between two consecutive matches are needed. The Feature Engineering has been chosen for this purpose.

3.2 Feature Engineering

The dataset for fatigue analysis holds attributes like match date, home team, away team and match result. Fatigue occurs mainly due to the presence of continuous fixtures without a proper break. This was founded out by finding out the difference between

| A | B | C | D | E | F | G | H | I | J |
|----------|-------------|-------------|-----------------|-----------------|----------|------------|---------|-----------------|----------------|
| Date | HomeTeam | AwayTeam | HomeTeamGoalsFT | AwayTeamGoalsFT | ResultFT | Season | Fatigue | binomialfatigue | binomialresult |
| 08-09-01 | Real Madrid | Malaga | 1 | 1 | D | season0102 | 14 | 1 | 0 |
| 29-09-01 | Real Madrid | Valladolid | 2 | 2 | D | season0102 | 7 | 0 | 0 |
| 21-10-01 | Real Madrid | Celta | 1 | 1 | D | season0102 | 8 | 1 | 0 |
| 27-10-01 | Villarreal | Real Madrid | 2 | 3 | A | season0102 | 6 | 0 | 1 |
| 17-11-01 | Real Madrid | Sevilla | 2 | 1 | H | season0102 | 6 | 0 | 1 |
| 25-11-01 | Vallecano | Real Madrid | 0 | 3 | A | season0102 | 8 | 1 | 1 |
| 08-12-01 | Tenerife | Real Madrid | 0 | 2 | A | season0102 | 7 | 0 | 1 |

Figure 3.1: Feature

two consecutive date at which the match occurred. A new column named fatigue was created on excel and a formula field was created by using the '-' operator. This is then converted into binomial field in excel with values less than 7 as '0' and more than 7 as '1'. The value 7 is the average difference between two matches. The results are generally draw, win and lose. In-order to treat this column as binomial the losing and draw were considered as '0' while the wining was considered as '1'. This job was done on excel by creating a new column named as binomial result. Then the difference in days between two matches are calculated by using formula field in excel.

The dataset for manager tactic analysis holds attributes like Real Madrid's playing styles , opponent playing styles and the points that Real Madrid has got from that match. The dataset for 2010-2011 season is chosen which is the year before they won the league at 2011-2012. The dataset for 2015-2016 season is then later chosen to identify the areas that needs to be improved in tactics wise to perform well in the upcoming season 2016-2017. It has got 25 independent variables and 1 dependent variable.

3.3 Statistical Testing

The analysis was to check whether there are any differences between the binomial result of the match and binomial fatigue. Hypothesis Testing is the best way to do this. It involves Null hypothesis and hypothesis testing. Null hypothesis implies there is no significant differences between the two groups. Alternative Hypothesis implies there is differences between the two groups. Anova has been preferred for this purpose. Since there are two variables and both are binomials. The data also meets all the requirements for anova testing.

| | A | B |
|----|---|------------|
| 1 | Attribute | Data Type |
| 2 | Home | Binomial |
| 3 | Real Madrid : Attacked through the middle | Binomial |
| 4 | Real Madrid : Attacked down the Right side | Binomial |
| 5 | Real Madrid : Had a high shot frequency when in possession | Binomial |
| 6 | Real Madrid : Had a large quantity of possession in their opponent's half | Binomial |
| 7 | Real Madrid : Favoured long shots | Binomial |
| 8 | Real Madrid : Favoured long balls | Binomial |
| 9 | Real Madrid : Favoured through balls | Binomial |
| 10 | Real Madrid : Attacked down the left side | Binomial |
| 11 | Real Madrid : Favoured short passing | Binomial |
| 12 | Real Madrid : Favoured crossing the ball | Binomial |
| 13 | Real Madrid : Dominated possession | Binomial |
| 14 | Real Madrid : Played with width | Binomial |
| 15 | Opponent : Played with width | Binomial |
| 16 | Opponent : Had a large quantity of possession in their opponent's half | Binomial |
| 17 | Opponent : Had a high shot frequency when in possession | Binomial |
| 18 | Opponent : Attacked through the middle | Binomial |
| 19 | Opponent : Attacked down the Right side | Binomial |
| 20 | Opponent : Attacked down the left side | Binomial |
| 21 | Opponent : Dominated possession | Binomial |
| 22 | Opponent : Favoured short passing | Binomial |
| 23 | Opponent : Favoured through balls | Binomial |
| 24 | Opponent : Favoured crossing the ball | Binomial |
| 25 | Opponent : Favoured long shots | Binomial |
| 26 | Opponent : Favoured long ball | Binomial |
| 27 | Real Madrid Points | Polynomial |

Figure 3.2: Data types



Figure 3.3: Classification

3.3.1 ANOVA Test Case

Dependent Variable : binomial fatigue (data type : binomial)

Independent Variable: binomial result (data type: binomial)

This anova hypothesis testing was performed in RStudio.

3.4 Machine Learning - Classification

In-order to identify the manager's work rate and the areas of improvement in tactics the Real Madrid 2010-2011 season and Real Madrid 2015-2016 dataset are both loaded into Rapid Miner database. The classification technique decision tree has been preferred for analysis. A new process was initiated in Rapid Miner and the dataset was given as input for the process with all the attributes except Real Madrid Points is termed as binomial. The Real Madrid Points attribute is termed as polynomial. Then the roles were set using the set role operator. Through the set role operator the dependent variable which is the Real Madrid Points is setted as label. Then the decision tree was connected to the set role operator and output. Then the process was made to run. This was repeated for both the seasons. once the process was completed, the results were evident on the decision tree graph.

3.5 Summary

The chapter has explained the datasets that were used in the research. This Chapter has also explained the statistical analysis - anova as well as the machine learning classification technique.

Chapter 4

Results and Findings

4.1 Through Anova Hypothesis Testing

The anova hypothesis testing was done between binomial fatigue attribute and the binomial result attribute. The results seen in R as below, P value is less than 0.5, that means our hypothesis is true. There exists a relationship between the result of the game and the fatigue that the team has generated because of consistent matches. This as a manager has to be careful about while choosing his squad for each match.

4.2 Through Machine Learning Classification - Season 2010/2011

Real Madrid always win when they dominate possession in the opponents half and attack through the right side. Their maximum loses came when they favored long shots when in possession. The long rangers as well as controlling the possession were the two tactics used by the Real Madrid manager impacts the result of the match

```
> data = read.xlsx("data1.xlsx", colNames = T)
> fit <- aov (binomialfatigue ~ binomialresult, data=data)
> summary(fit)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----------------|-----|--------|---------|---------|----------|
| binomialresult | 1 | 1.07 | 1.0708 | 4.843 | 0.0282 * |
| Residuals | 568 | 125.60 | 0.2211 | | |

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.1: Test Results

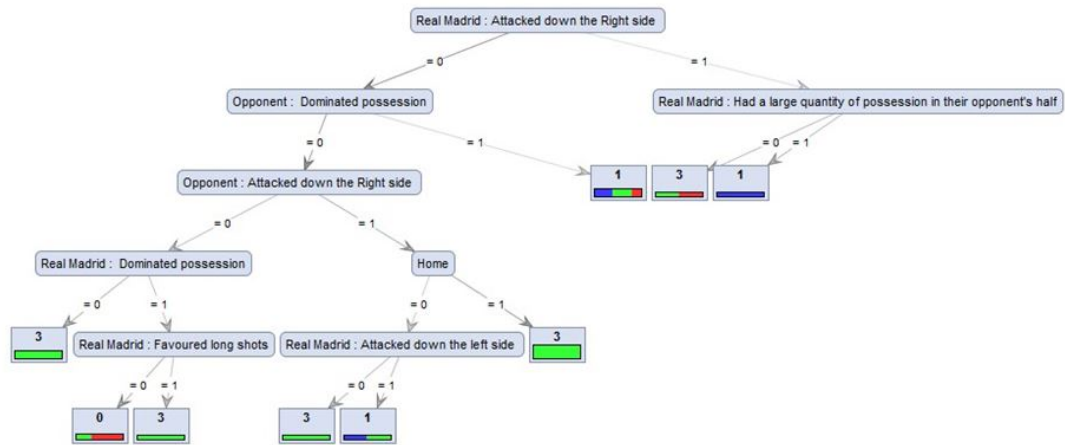


Figure 4.2: ML Classification

heavily. These are the two factors that are heavily linked with the mid fielders of a team. As the mid fielders are the one who is responsible for controlling the passage of play through this season and will be in a good position to take long range shots at goal. The season ended Real Madrid finishing 2nd in the table. One area which Real Madrid needed to improve if they had to improve their chances of winning the league was their midfield. The next season 2011/2012, they did the same and ended up being the Laliga Winners by securing a record of scoring 100 points. Xabi Alonso - Real Madrid mid fielder has been chosen as the player of the winning season despite their forward scoring the most number of goals in the league.

4.3 Through Machine Learning Classification - Season 2015/2016

The result of the decision tree that has been generated in rapid miner for the season 2015/2016 is as below, Basically every time that Real Madrid attacked through the right side, they haven't lost the match. They have managed a draw to the minimum. But unfortunately, Real Madrid are restricted by injuries to one of their key player Gareth Bale, who is a record signing usually plays on the right side of the pitch wasn't fully fit for the season. He had injuries regularly to an aggregate of 80 days in total, which made him to miss 17 of the 37 league games. There were also areas of concern when it comes to attacking through the middle. So the areas of improvement was to keep the player fit to most part of the next season and as well as to sign a centre attacking mid fielder. According to bleacher reports, Real Madrid had been linked with

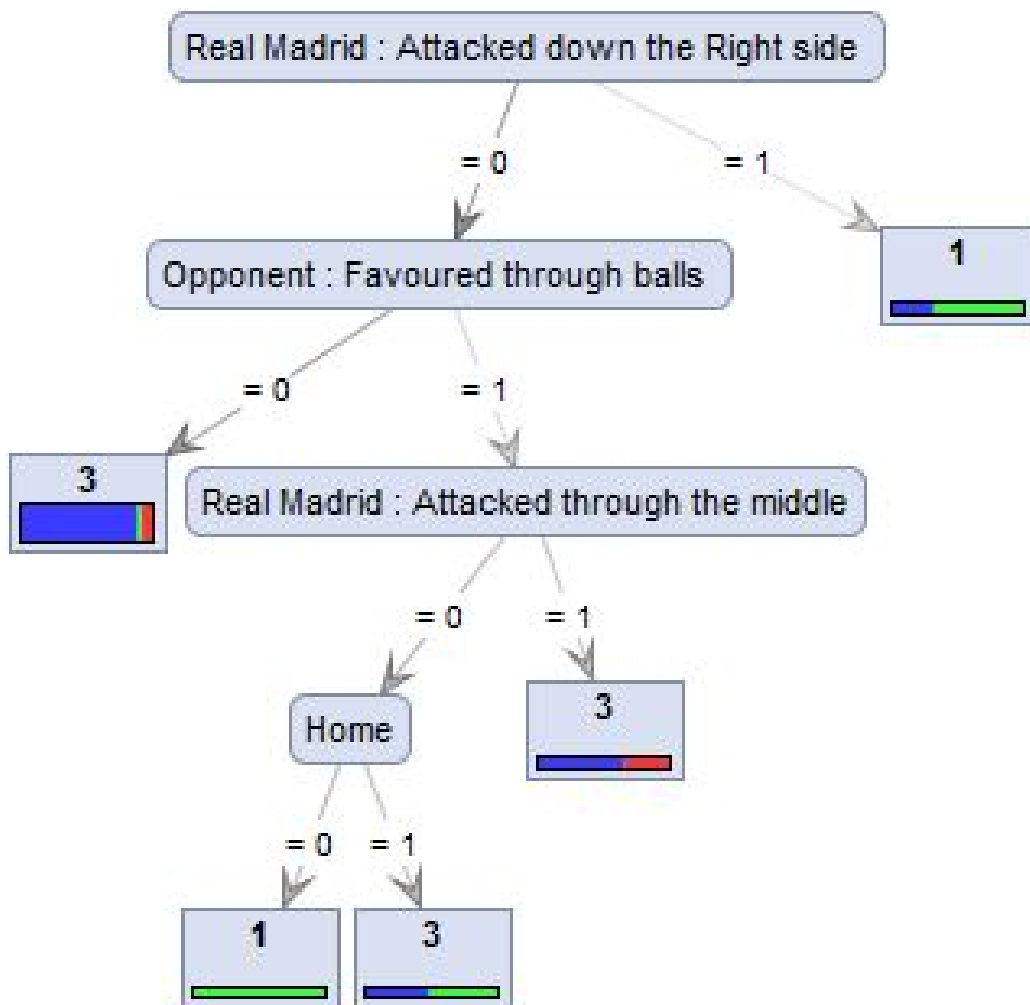


Figure 4.3: 15-16

Paul Pogba and Andre Gomez which eventually failed and the club brought back their young player Marco Asensio from loan.

4.4 Summary

The Statistical Analysis - Anova has concluded that there is a difference in result when team suffers from fatigue and not suffering from it. Decision Trees for season 2010/2011 has proved to be caution for the success of the next season 2011/2012. Decision Trees for season 2015/2016 has identified the areas that the team needs to improve to succeed.

Chapter 5

Conclusion

what makes a team to be on the winning side?. As Johan Cryuff () once said football is played with your head, and your legs are there only to support you . This research has supported the Johan Cyuff's lines and at the same time answers the question through machine learning - classification technique and statistical analysis - anova. Previous studies have made use of machine learning techniques and statistical analysis on the player work rates like the goals scored, assists, successful tackles etc., But this research was rather more focused on the playing style of a team in each match to identify the areas to improve as a seasonal objective and come back stronger next season by rebuilding the holes. Decision trees built on playing style of a team has identified the areas that the team needs to improve to be more successful. The statistical analysis - anova built on fatigue versus result of the match proved that there is a difference in result when the players suffer from fatigue. The dataset has failed to include the other matches that occurs in between the league games, but for now, anova was still was able to prove that fatigue affects the results of the game. This research was primarily focused on the playing style (Manager's Tactic) of a team and the future work is expected to be with the inclusion of roles of each individual players in the squad. The future work will also have the inclusion of predictions for next season if the identified holes through decision trees are filled.

Bibliography

- [1] Adrian Bell, Chris Brooks, and Tom Markham. The performance of football club managers: skill or luck? *Economics & Finance Research*, 1(1):19–30, 2013.
- [2] Juan de Dios Tena and David Forrest. Within-season dismissal of football coaches: Statistical analysis of causes and consequences. *European Journal of Operational Research*, 181(1):362–373, 2007.
- [3] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.
- [4] John Goddard. Regression models for forecasting goals and match results in association football. *International Journal of forecasting*, 21(2):331–340, 2005.
- [5] Joaquín González-Ródenas, Ignacio López Bondía, Ferran Calabuig Moreno, and Rafael Aranda Malavés. Indicadores tácticos asociados a la creación de ocasiones de gol en fútbol profesional.(tactical indicators associated with the creation of scoring opportunities in professional soccer). *CCD. Cultura.Ciencia.Deporte*. -- doi: 10.12800/ccd, 10(30):215–225, 2015.
- [6] Mark Griffiths. Gambling technologies: Prospects for problem gambling. *Journal of gambling studies*, 15(3):265–283, 1999.
- [7] Isabelle Guyon and André Elisseeff. An introduction to feature extraction. In *Feature extraction*, pages 1–25. Springer, 2006.
- [8] Josip Hucaljuk and Alen Rakipović. Predicting football scores using machine learning techniques. In *MIPRO, 2011 Proceedings of the 34th International Convention*, pages 1623–1627. IEEE, 2011.
- [9] M Hughes, K Robertson, and A Nicholson. Comparison of patterns of play of successful and unsuccessful teams in the 1986 world cup for soccer. *Science and football*, pages 363–367, 1988.
- [10] Mike D Hughes and Roger M Bartlett. The use of performance indicators in performance analysis. *Journal of sports sciences*, 20(10):739–754, 2002.
- [11] Nic James, Stephen D Mellalieu, and Chris Hollely. Analysis of strategies in soccer as a function of european and domestic competition. *International Journal of Performance Analysis in Sport*, 2(1):85–103, 2002.
- [12] A Joseph, Norman E Fenton, and Martin Neil. Predicting football results using bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 19(7):544–553, 2006.
- [13] Seamus Kelly. Understanding the role of the football manager in britain and ireland: A weberian approach. *European Sport Management Quarterly*, 8(4):399–419, 2008.
- [14] Xanthi Konstadinidou and Nikolaos Tsigilis. Offensive playing profiles of football teams from the 1999 women’s world cup finals. *International Journal of Performance Analysis in Sport*, 5(1):61–71, 2005.

- [15] Carlos Lago. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of sports sciences*, 27(13):1463–1469, 2009.
- [16] Patrick E McSharry. Effect of altitude on physiological performance: a statistical analysis using results of international football games. *Bmj*, 335(7633):1278–1281, 2007.
- [17] Stephen Morrow and Brian Howieson. The new business of football: A study of current and aspirant football club managers. *Journal of Sport Management*, 28(5):515–528, 2014.
- [18] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1):1, 2015.
- [19] Peter O’Donoghue and Shane King. Activity profile of men’s gaelic football. In *Science and Football V: The Proceedings of the Fifth World Congress on Sports Science and Football*, page 207. Routledge, 2005.
- [20] Luca Prestigiacomo. *Coaching soccer: match strategy and tactics*. Reedswain Inc., 2004.
- [21] A Scoulding, Nic James, and J Taylor. Passing in the soccer world cup 2002. *International Journal of Performance Analysis in Sport*, 4(2):36–41, 2004.
- [22] Markus Waldén, Martin Hägglund, and Jan Ekstrand. Uefa champions league study: a prospective study of injuries in professional football during the 2001–2002 season. *British journal of sports medicine*, 39(8):542–546, 2005.