Paper 3828-2019

SCOUTING IN SOCCER WITH APPLIED MACHINE LEARNING

Vishal Gaurav, Oklahoma State University; Goutam Chakraborty, Oklahoma State University

ABSTRACT

After netting 65 goals in 102 league appearances between 2007 & 2011, Fernando Torres became the most sought after player in the 2011 January transfer window. Chelsea spent a whopping £50m to procure the services of the most lethal striker of his era. Torres' move set a British record for most expensive transfer at that time. In the seasons to follow with the Blues, the Spaniard failed to live up to his high price tag. He scored only 19 league goals in 103 matches for Chelsea. Was it the injuries, or, the pressure of the huge expectations? Or, was he just not a right fit for the Blues?

Every year we see major soccer clubs spend hundreds of millions of dollars to sign contracts with the players of their interest. More often than not, these deals are based on a player's skills, potential for growth, performance in prior seasons, and brand value. While often these transfer deals pay off, there is an abundance of failed transfers every season. And in the end, it all invariably depends on the right fit between the playing style of the player and the team. The primary objective of this paper is to leverage analytics to identify players who are the best fit for the teams. The idea is to analyze, and decipher underlying quantitative patterns and make an association that could help identify-

1. Players with similar skill sets.
2. Potential market value of such players.
3. Positions within teams where these players would be most effective.

INTRODUCTION

Association football, aka soccer, is by far the most popular and globally recognized sport of our time. Over the last decade, its popularity has soared significantly. The recent conclusion of the FIFA 2018 world cup in Russia, displayed the immense popularity this game has garnered across the globe. With the ever-increasing media coverage, the game has become a multibillion-dollar business. Recent advancement in technology has enhanced our ability to capture statistical data on each game, which in turn has essentially transformed player selection for this beautiful game from subjective judgment by scouts to analytics-driven selection of players just like Moneyball transformed player selection in baseball.

The FIFA video game series from EA Sports is a commendable effort in this direction and is a reflection of the aforementioned phenomenon. FIFA 18 offers detailed quantitative and statistical information on individual players and teams. In modern day football specific positions represent a player's main role and their primary area of operation on the field. It is extremely complicated and equally important to characterize a player according to their position on the field, given most players these days are able to fit into various roles. However, most players still prefer to play in a limited range of positions, as each position requires a different combination of skills and physical attributes.

With the rapid increase in the volume of football data in digital form, the use of specific metrics for characterizing and ranking players in their specific positions according to their perceived abilities has attracted the attention of coaches and data scientists alike. Most

European clubs have enlisted the help of data scientists or quantitative analysts in one form or the other. The ability to quantitatively analyze the performance of football players based on available data offers an important competitive advantage to the teams who employed such data analysts. Given the advancement in modern computing and democratization of machine learning, state of the art machine learning methods are being successfully deployed in football. These analyses have been used to predict outcomes of a match, to analyze a team's performance and in a multitude of other scenarios. However, the problem of grouping and selecting players based on performance data using machine learning methods is an interesting field that has not seen much-published research.

DATA COLLECTION & PREPARATION

The data sets for this analysis were acquired from https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset. These three individual datasets were merged together and consolidated using player ID as the primary key.

ID	Acceleration	Aggression	Agility	Balance	Ball control	Composure	Crossing	Curve	Dribbling	Finishing	Free kick accuracy
20801	89	63	89	63	93	95	85	81	91	94	76
158023	92	48	90	95	95	96	77	89	97	95	90
190871	94	56	96	82	95	92	75	81	96	89	84
176580	88	78	86	60	91	83	77	86	86	94	84
167495	58	29	52	35	48	70	15	14	30	13	11
188545	79	80	78	80	89	87	62	77	85	91	84
193080	57	38	60	43	42	64	17	21	18	13	19
183277	93	54	93	91	92	87	80	82	93	83	79
182521	60	60	71	69	89	85	85	85	79	76	84
167664	78	50	75	69	85	86	68	74	84	91	62
155862	75	84	79	60	84	80	66	73	61	60	67
192985	76	68	80	75	87	84	90	83	85	83	83
192119	46	23	61	45	23	52	14	19	13	14	11
184941	88	80	90	87	87	86	80	78	90	85	78
177003	75	62	93	94	92	84	78	79	86	71	77
173731	93	65	77	65	87	85	87	86	89	87	85
153079	90	63	86	91	89	90	70	82	89	90	72
138956	68	92	59	64	57	82	58	60	58	33	31
1179	49	38	55	49	28	70	13	20	26	15	13
211110	88	48	91	85	93	84	80	88	92	85	84
200389	43	34	67	49	16	55	13	13	12	11	14
194765	87	69	90	80	86	86	82	84	87	88	75
189509	77	57	90	86	92	83	72	85	90	69	77

Table 1. Attribute Data

ID	Name	Age	Nationality	Overall	Potential	Value	Wage	Special	Club
20801	Cristiano Ronaldo	32	Portugal	94	94	95.5M	565K	2228	Real Madrid CF
158023	L. Messi	30	Argentina	93	93	105M	565K	2154	FC Barcelona
190871	Neymar	25	Brazil	92	94	123M	280K	2100	Paris Saint-Germain
176580	L. SuÃ¡rez	30	Uruguay	92	92	97M	510K	2291	FC Barcelona
167495	M. Neuer	31	Germany	92	92	61M	230K	1493	FC Bayern Munich
188545	R. Lewandowski	28	Poland	91	91	92M	355K	2143	FC Bayern Munich
193080	De Gea	26	Spain	90	92	64.5M	215K	1458	Manchester United
183277	E. Hazard	26	Belgium	90	91	90.5M	295K	2096	Chelsea
182521	T. Kroos	27	Germany	90	90	79M	340K	2165	Real Madrid CF
167664	G. HiguaÃn	29	Argentina	90	90	77M	275K	1961	Juventus
155862	Sergio Ramos	31	Spain	90	90	52M	310K	2153	Real Madrid CF
192985	K. De Bruyne	26	Belgium	89	92	83M	285K	2162	Manchester City
192119	T. Courtois	25	Belgium	89	92	59M	190K	1282	Chelsea
184941	A. SÃ¡nchez	28	Chile	89	89	67.5M	265K	2181	Arsenal
177003	L. ModriÄ‡	31	Croatia	89	89	57M	340K	2228	Real Madrid CF
173731	G. Bale	27	Wales	89	89	69.5M	370K	2263	Real Madrid CF
153079	S. AgÃ¼ero	29	Argentina	89	89	66.5M	325K	2074	Manchester City
138956	G. Chiellini	32	Italy	89	89	38M	225K	1867	Juventus
1179	G. Buffon	39	Italy	89	89	4.5M	110K	1335	Juventus
211110	P. Dybala	23	Argentina	88	93	79M	215K	2063	Juventus

Table 2. Personal Data:

ID	Preferred Positions	CAM	CB	CDM	CF	CM	LAM	LB	LCB	LCM	LDM	LF	LM	LS	LW	LWB	RAM	RB	RCB	RCM	RDM	RF	RM	RS	RW	RWB	ST
158023	RW	92	45	59	92	84	92	57	45	84	59	92	90	88	91	62	92	57	45	84	59	92	90	88	91	62	88
20801	ST LW	89	53	62	91	82	89	61	53	82	62	91	89	92	91	66	89	61	53	82	62	91	89	92	91	66	92
190871	LW	88	46	59	88	79	88	59	46	79	59	88	87	84	89	64	88	59	46	79	59	88	87	84	89	64	84
183277	LW	88	47	61	87	81	88	59	47	81	61	87	87	82	88	64	88	59	47	81	61	87	87	82	88	64	82
176580	ST	87	58	65	88	80	87	64	58	80	65	88	85	88	87	68	87	64	58	80	65	88	85	88	87	68	88
41	LM CM	87	63	74	83	84	87	69	63	84	74	83	84	76	84	72	87	69	63	84	74	83	84	76	84	72	76
211110	ST CAM	86	43	55	86	78	86	55	43	78	55	86	86	83	87	60	86	55	43	78	55	86	86	83	87	60	83
192985	RM CM CAM	86	57	70	85	84	86	66	57	84	70	85	85	81	85	71	86	66	57	84	70	85	85	81	85	71	81
189509	CDM CAM CM	86	66	76	83	85	86	72	66	85	76	83	83	77	83	75	86	72	66	85	76	83	83	77	83	75	77
188350	LW ST LM	86	52	64	86	80	86	63	52	80	64	86	84	82	86	67	86	63	52	80	64	86	84	82	86	67	82
177003	CDM CM	86	72	80	83	86	86	78	72	86	80	83	84	76	83	80	86	78	72	86	80	83	84	76	83	80	76
173731	RW	86	67	71	87	81	86	72	67	81	71	87	87	87	87	74	86	72	67	81	71	87	87	87	87	74	87
9014	RW RM	86	47	59	86	78	86	58	47	78	59	86	85	82	87	63	86	58	47	78	59	86	85	82	87	63	82
198219	LW	85	39	54	83	77	85	55	39	77	54	83	84	76	85	60	85	55	39	77	54	83	84	76	85	60	76
197781	LM RM CAM CM	85	61	71	84	84	85	67	61	84	71	84	82	79	83	70	85	67	61	84	71	84	82	79	83	70	79
190460	LM RM CAM	85	53	68	82	83	85	64	53	83	68	82	84	77	83	69	85	64	53	83	68	82	84	77	83	69	77
189242	CAM LW	85	49	61	83	79	85	60	49	79	61	83	83	77	84	64	85	60	49	79	61	83	83	77	84	64	77
184941	RM LW ST LM	85	56	64	85	79	85	62	56	79	64	85	85	83	86	66	85	62	56	79	64	85	85	83	86	66	83
176635	RW CAM	85	41	57	82	79	85	52	41	79	57	82	83	76	83	58	85	52	41	79	57	82	83	76	83	58	76
175943	ST LW CF	85	47	60	85	78	85	60	47	78	60	85	84	78	86	65	85	60	47	78	60	85	84	78	86	65	78

Table 3. Playing Position Data

Note: To understand abbreviations in this table such as RW meaning Right Wing requires knowledge of the game. These fields were not directly used in the analysis. Rather we used their preferred positions to explain some of the results. A detailed description of each of the position abbreviation is available in the appendix section of this paper.

The final consolidated data set consists of 15,178 observations and 44 columns. A complete description of each of the columns are available at *https://www.fifauteam.com/fifa-18-attributes-guide/*

ID	Name	Club	Nationality	Age	Overall	Potential	Value	Wage	Acceleration	Aggression	Agility	Balance
20801	Cristiano Ronaldo	Real Madrid CF	Portugal	32	94	94	95500000	565000	89	63	89	63
158023	L. Messi	FC Barcelona	Argentina	30	93	93	105000000	565000	92	48	90	95
190871	Neymar	Paris Saint-Germain	Brazil	25	92	94	123000000	280000	94	56	96	82
176580	L. SuÃ¡rez	FC Barcelona	Uruguay	30	92	92	97000000	510000	88	78	86	60
167495	M. Neuer	FC Bayern Munich	Germany	31	92	92	61000000	230000	58	29	52	35
188545	R. Lewandowski	FC Bayern Munich	Poland	28	91	91	92000000	355000	79	80	78	80
193080	De Gea	Manchester United	Spain	26	90	92	64500000	215000	57	38	60	43
183277	E. Hazard	Chelsea	Belgium	26	90	91	90500000	295000	93	54	93	91
182521	T. Kroos	Real Madrid CF	Germany	27	90	90	79000000	340000	60	60	71	69
167664	G. HiguaÃ-n	Juventus	Argentina	29	90	90	77000000	275000	78	50	75	69
155862	Sergio Ramos	Real Madrid CF	Spain	31	90	90	52000000	310000	75	84	79	60
192985	K. De Bruyne	Manchester City	Belgium	26	89	92	83000000	285000	76	68	80	75
192119	T. Courtois	Chelsea	Belgium	25	89	92	59000000	190000	46	23	61	45
184941	A. SÃ¡nchez	Arsenal	Chile	28	89	89	67500000	265000	88	80	90	87
177003	L. ModriÄ‡	Real Madrid CF	Croatia	31	89	89	57000000	340000	75	62	93	94
173731	G. Bale	Real Madrid CF	Wales	27	89	89	69500000	370000	93	65	77	65
153079	S. AgÃ¼ero	Manchester City	Argentina	29	89	89	66500000	325000	90	63	86	91
138956	G. Chiellini	Juventus	Italy	32	89	89	38000000	225000	68	92	59	64
1179	G. Buffon	Juventus	Italy	39	89	89	4500000	110000	49	38	55	49
211110	P. Dybala	Juventus	Argentina	23	88	93	79000000	215000	88	48	91	85

Table 4. Consolidated Data

METHODOLOGY

Clustering (or, unsupervised classification) is one of the most important and fundamental tasks in machine learning. It helps organize unlabeled data into distinct groups, such that the observations within each group are similar to each other, while observations in different groups are different from each other based on metrics such as distance between the data points. In this paper, we explore the possibility of a model-based clustering method to help the scouts and football clubs in their search for professional football players. We had 35 attributes such as acceleration, aggression and so on (each measured on a 100 points scale according to EA Sports) about each player and found that many of them are highly correlated with each other. Therefore, we decided to reduce the dimensionality and handle the problem of potential multicollinearity using Principal Component Analysis (PCA). PCA was used to replace a large number of possibly correlated variables with a much smaller set of uncorrelated variables while capturing as much information in the original variables as possible. Using Principal Component Analysis, we reduced the 35 attributes into 4 linear combinations of observed variables i.e., principal components. The first principal component accounted for the most variance in the original set of variables. The second principal component accounted for the most variance in the original variables while maintaining its orthogonality to the first principal component. Each subsequent component maximized the variance accounted for, while at the same time remaining uncorrelated with all previous principal components.

MODEL

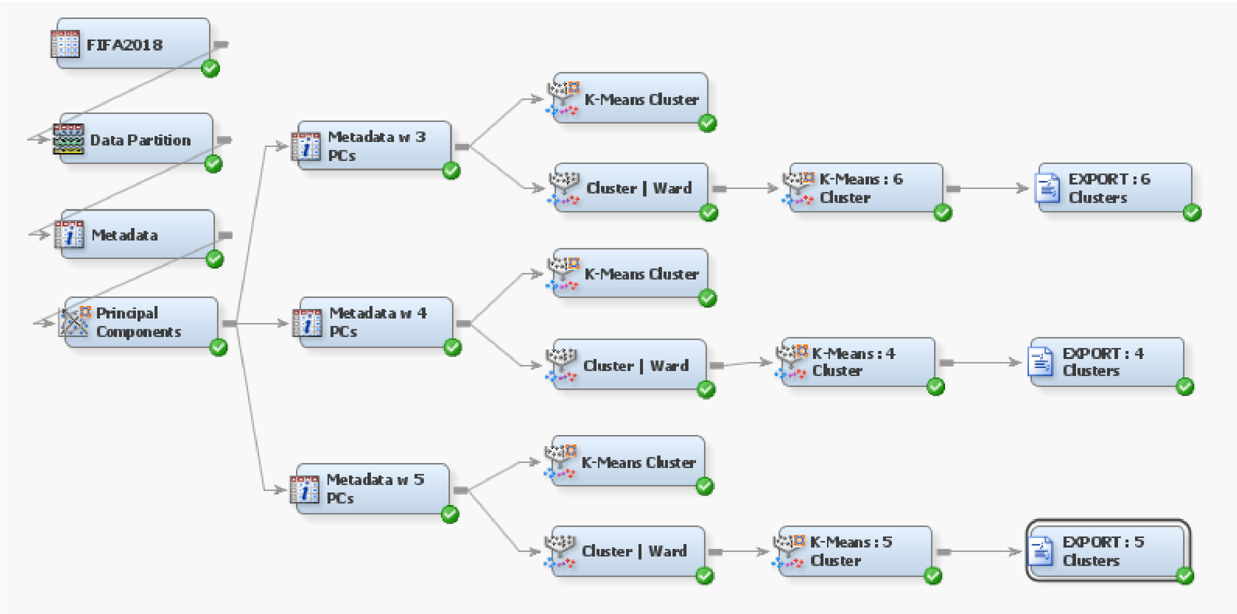

Figure 1. Model Diagram in SAS Enterprise Miner

RESULTS

PRINCIPAL COMPONENT ANALYSIS

As seen in Figure 2, the first four PCs accounted for nearly 77% of the variance. While we could have used perhaps 7 PCS to account for up to 86% of the variance, for the purpose of this research, we proceeded with just 4 principal components, for further analysis.

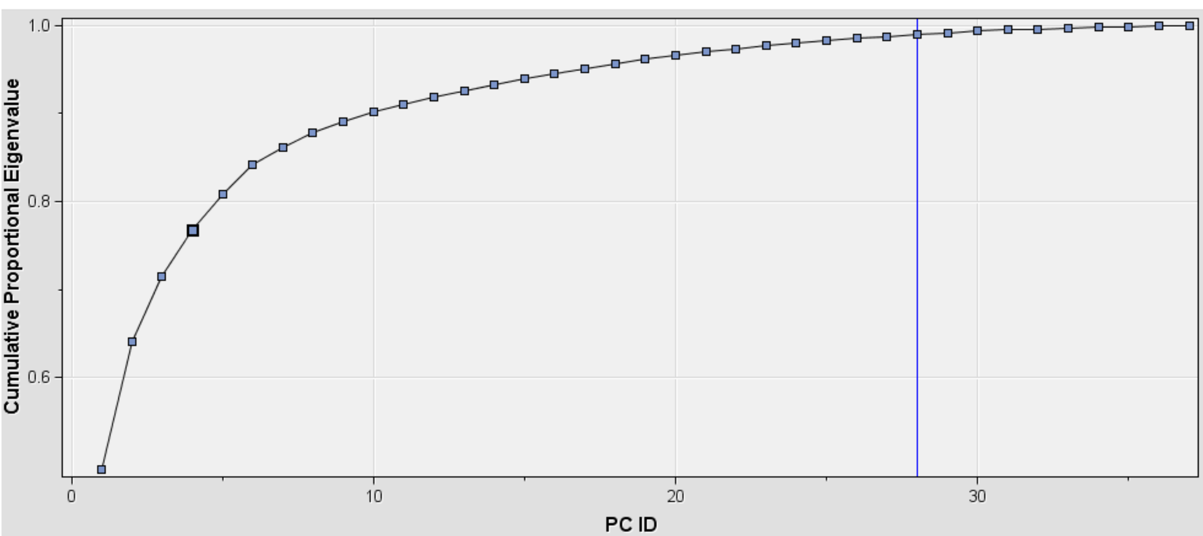

Figure 2. Selection of Number of PCs

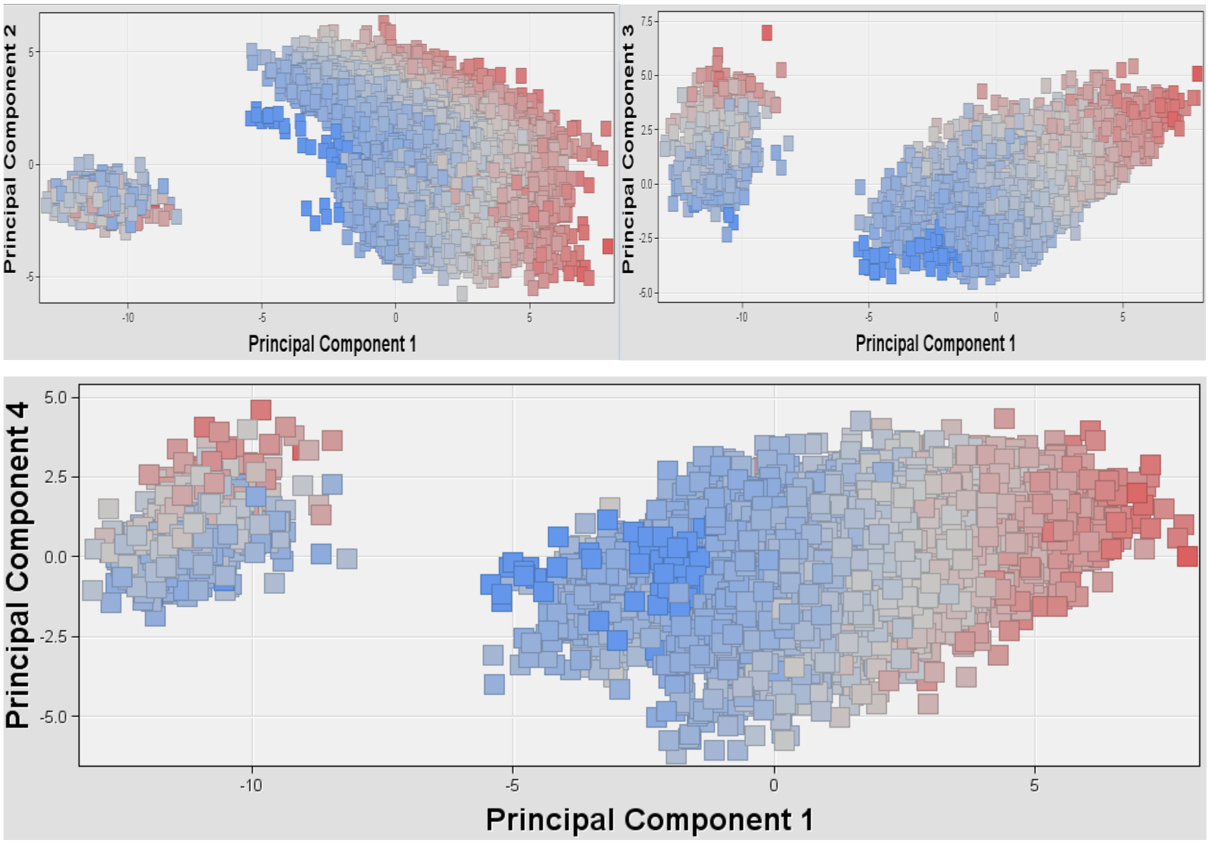

Figure 3. Plots of Data Points in the PC Space

The above graph (Figure 3) reveals the presence of two major groups in the space of PCs. Upon further probing (by looking at players and attributes in each of the groups and applying domain expertise) we realized that the smaller groups are a representation of attributes that characterize a goalkeeper, while the bigger group is made of players with no goalkeeping skills.

CLUSTERING

Once the number of PCs was decided, we used SAS EM to run cluster analysis using the PC values. We used Ward's method for pre-clustering with a CCC cut-off of 2 before running k-Means in SAS EM to estimate the number of clusters. The analysis resulted in four reasonably sized clusters as shown in Figure. 4. From the Cluster Proximities plot in Figure. 5, it seems the centers are also fairly well separated in the two-dimensional space suggesting a reasonable solution.

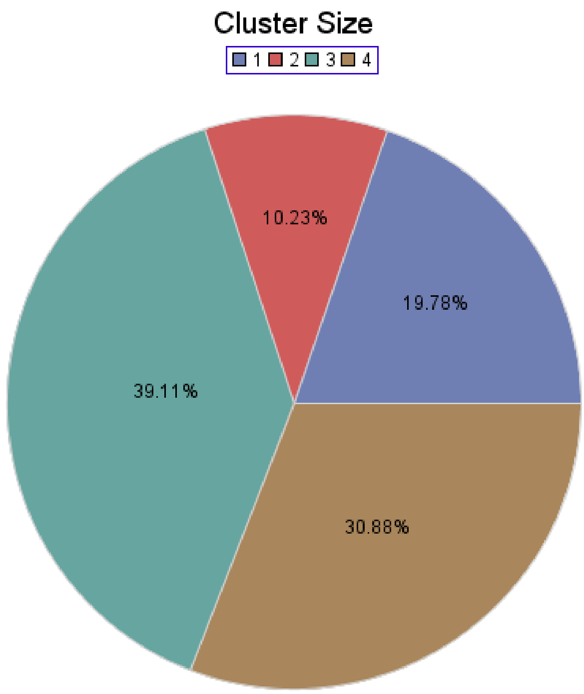

Figure 4. Cluster Size

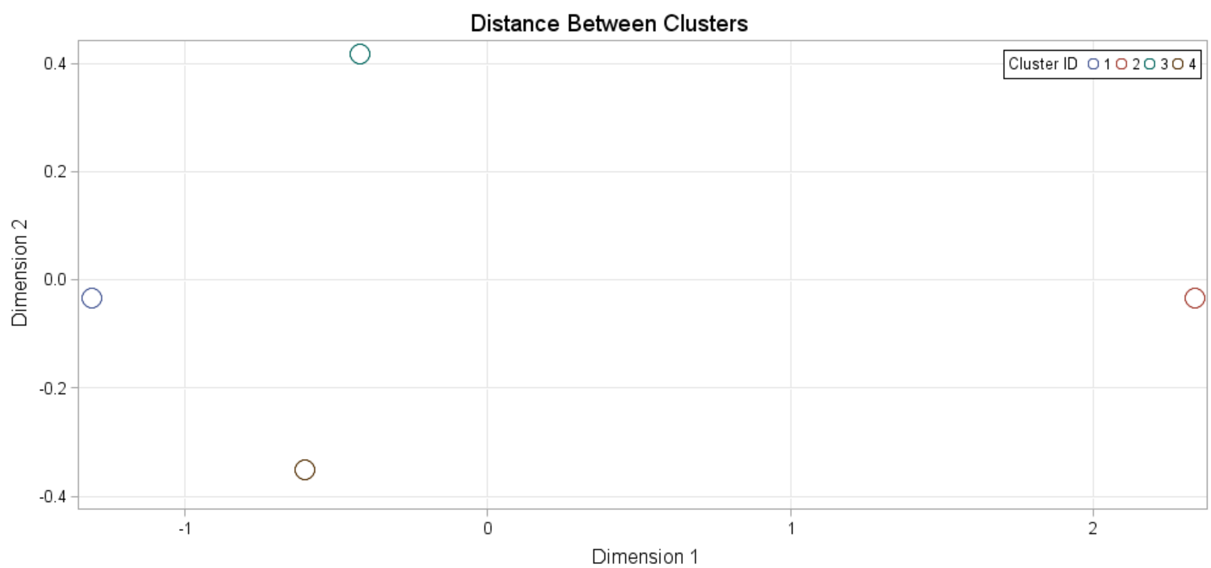

Figure 5. Separation of Cluster Centers

Once we obtained fours clusters, we first labeled them by looking at the top players in each cluster and applying domain knowledge as shown below.

Cluster 1 = Midfielder, Cluster 2 = Goalkeeper, Cluster 3 = Defender, Cluster 4 = Forward

It seems k-means has essentially split out the large group that we observed in the PC space into three smaller groups where top players' roles are not goalkeepers. Next, we profiled each of the four clusters using the raw attribute values to get a better sense of how these clusters differ among each other. These profile analyses are presented in a series of boxplots below. In these boxplots, PlayerRole in the X-Axis refers to the four clusters.

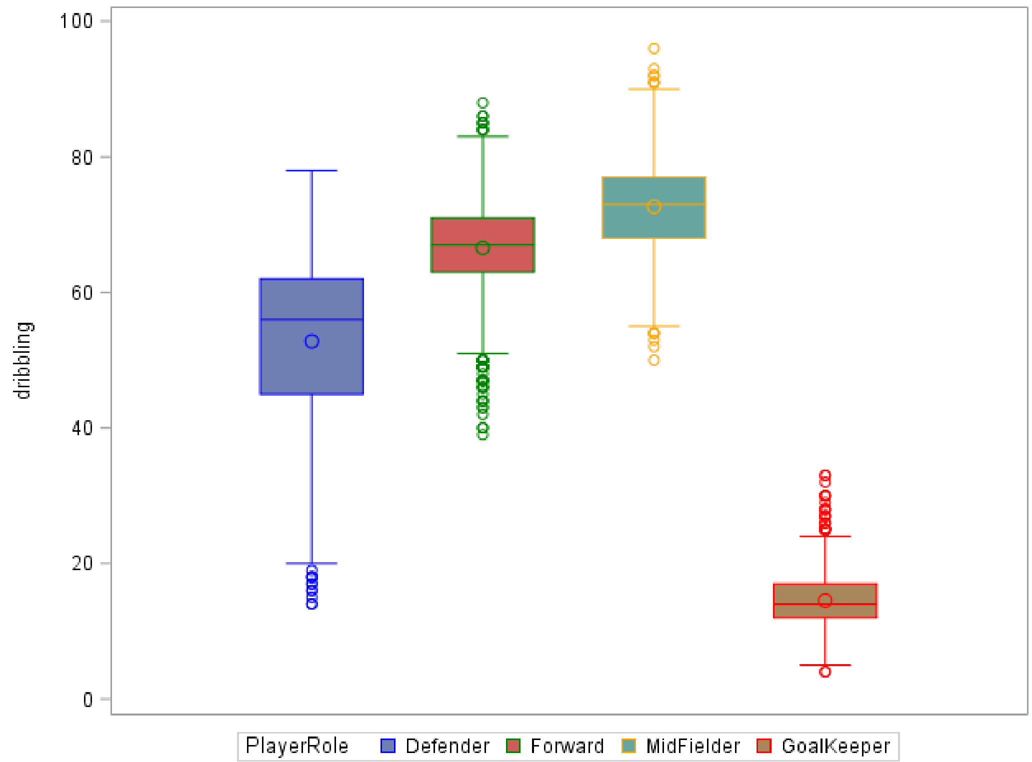

Figure 6. Variation in Dribbling Skills across Clusters (PlayerRole)

Figure. 6 shows high dribbling skills for clusters labeled as forward and midfielders, which makes sense given the fact that these are the players who hold the ball the most and create goal-scoring opportunities.

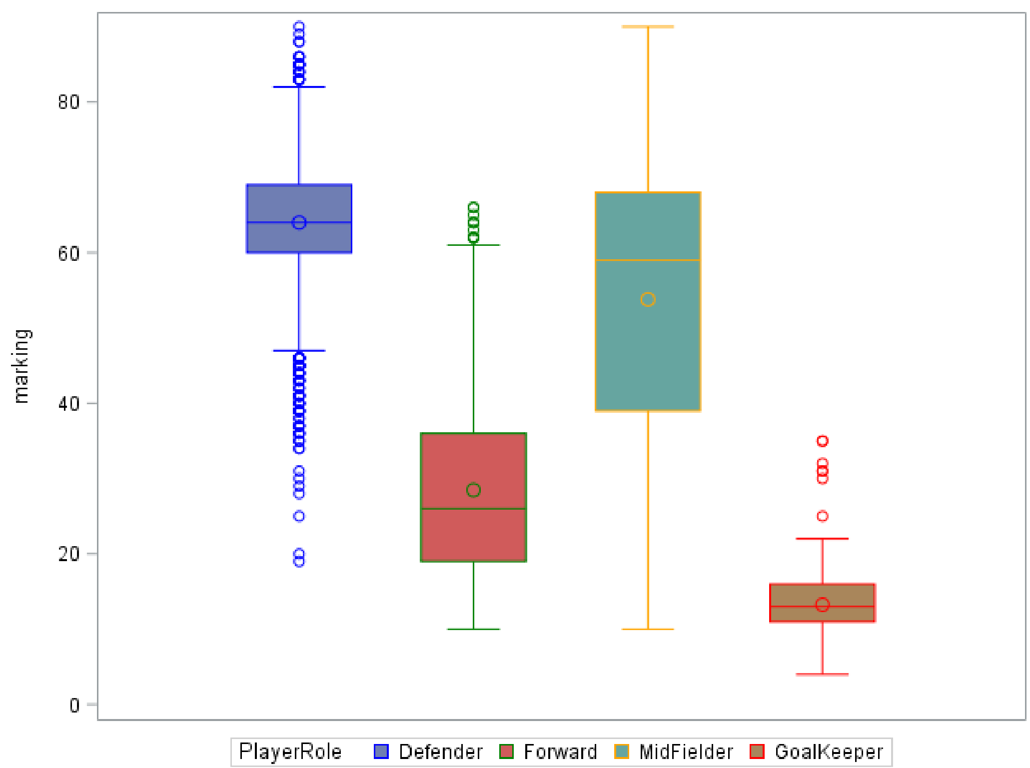

Figure 7. Variation in Marking Skills across Clusters (PlayerRole)

From Figure. 7, we see a significantly higher skill score for marking for the clusters labeled as defenders and midfielders. This is because it is the responsibility of these players to “mark” the opposing teams’ forwards and wingers so as to prevent them from creating any goal scoring opportunities.

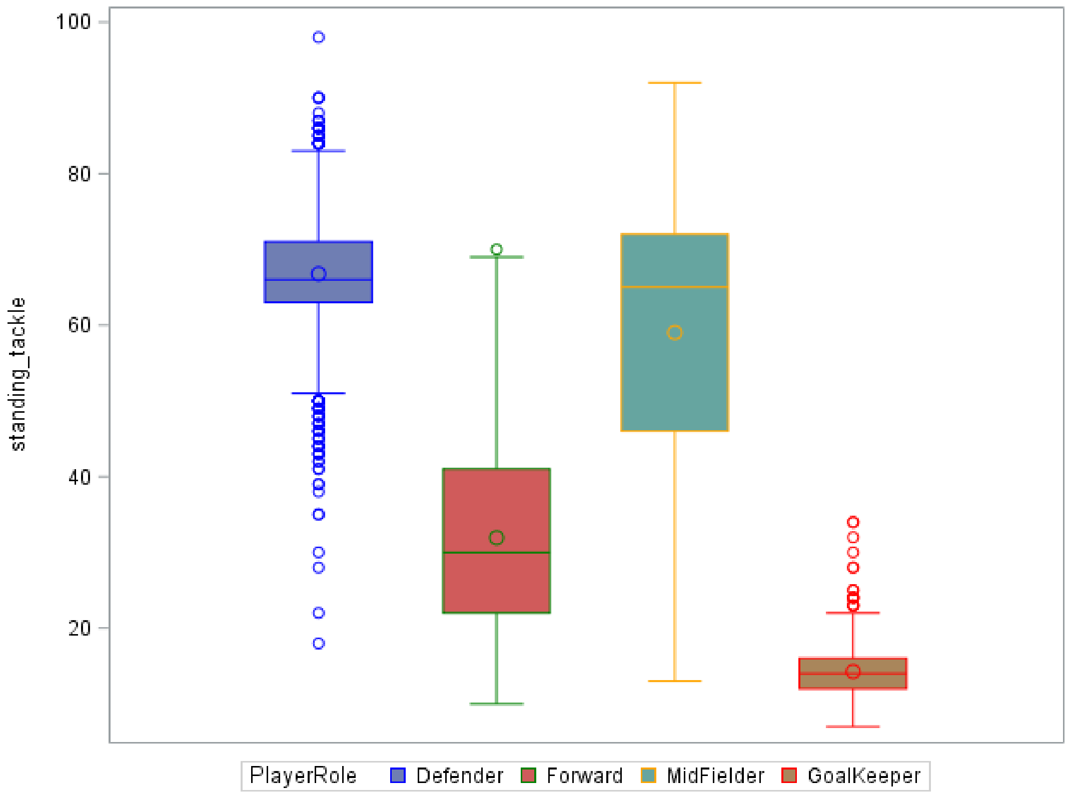

Figure 8. Variation in Standing Tackle Skills across Clusters (PlayerRole)

As expected, from Figure. 8, we notice that clusters labeled as defenders have the highest tackle score (standing and sliding) as tackling is the old school skill and the best way for a clean defense in the game of football. The Midfielder cluster has a higher rating on tackle scores as well, as this group includes several defensive midfielders.

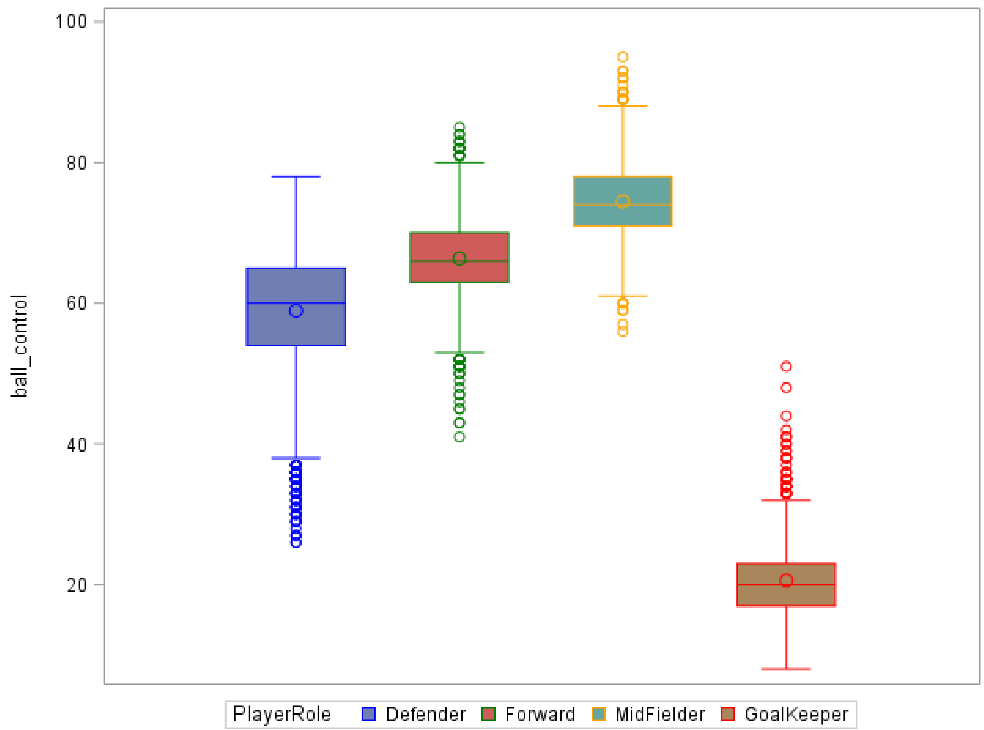

Figure 9. Variation in Ball Control Skills across Clusters (PlayerRole)

From Figure. 9, we see higher skill scores for ball control for the clusters labeled as midfielders and forwards. This is because it is the responsibility of forwards (wingers) and midfielders to create goal scoring opportunities.

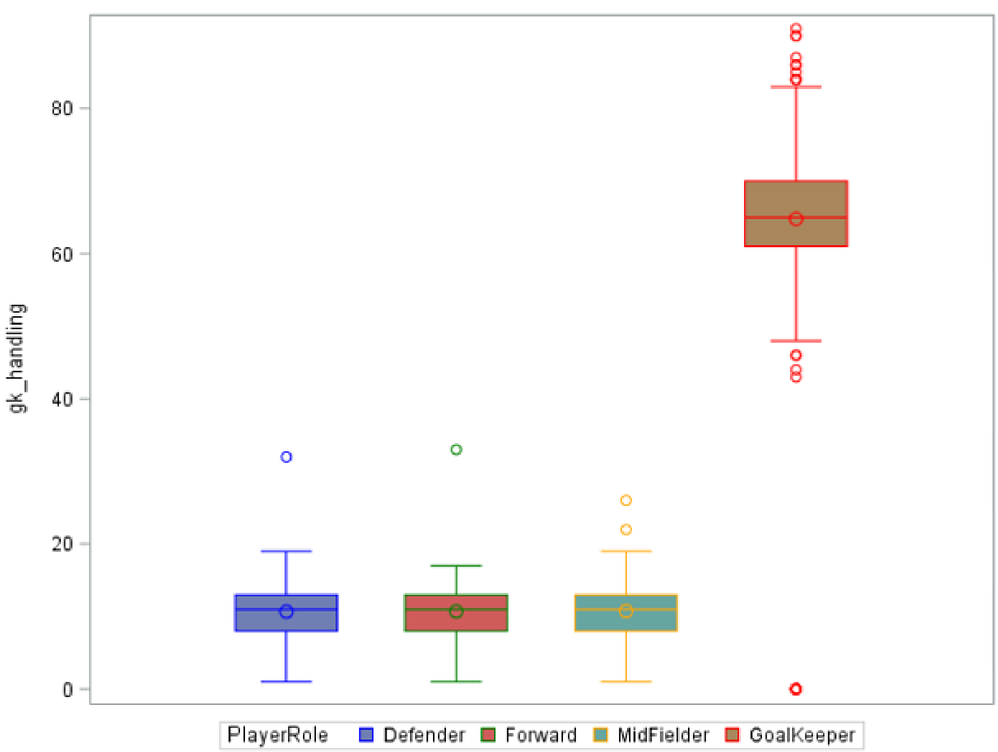

Figure 10. Variation in Goal Keeper Handling Skills across Clusters (PlayerRole)

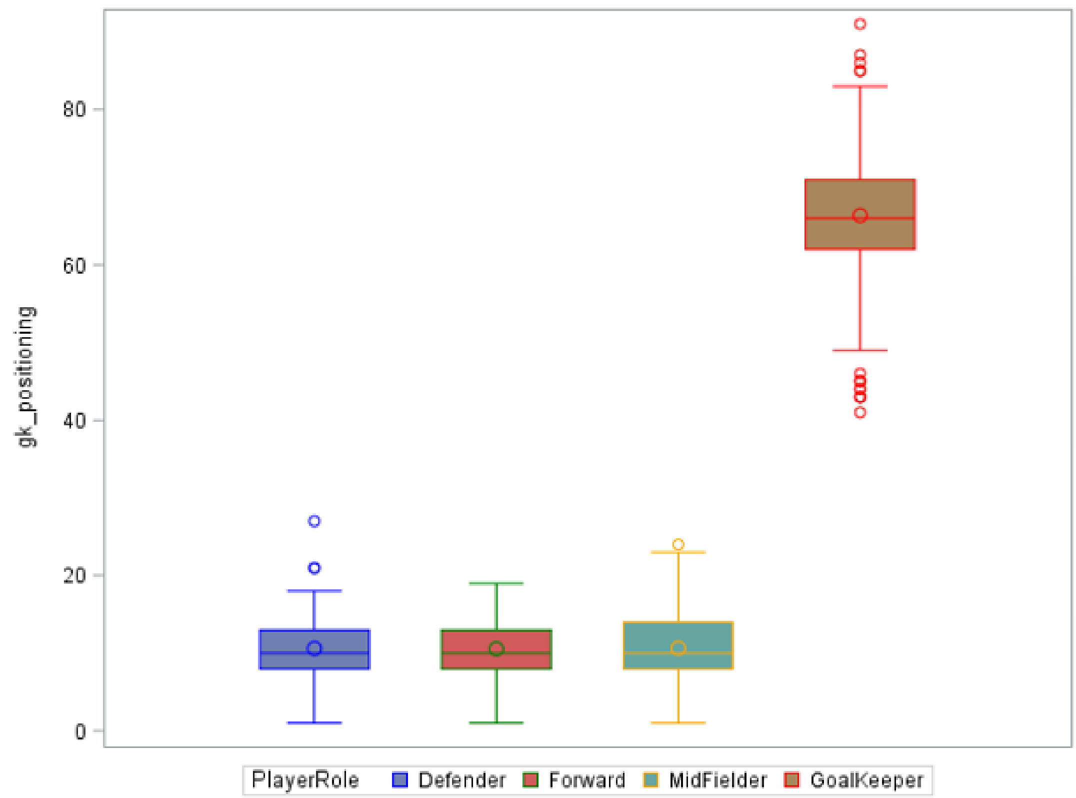

Figure 11. Variation in Goal Keeper Positioning Skills across Clusters (PlayerRole)

As expected, in Figure. 10 and 11, we notice that clusters labeled as Goalkeepers have significantly high rating on goalkeeping skills such as handling and positioning.

ANALYSIS

While the box plots provide a univariate visualization of each attribute across the clusters, they are not suitable for multivariate visual comparisons. To get a better understanding of the players' skills within each of the four clusters across all 32 attributes, we created a radar chart (Fig. 12). In Fig.12, the four clusters are represented by four colors: defenders (blue), midfielders (yellow) forwards (red) and goalkeepers (grey). The radar chart shows mean values by each attribute for each cluster in one comprehensive visual. When we look at the chart, it becomes evident that some variables related directly to goalkeepers (Diving, Handling, Kicking, Positioning and Reflexes) have lower values for the rest of the roles as compared to goalkeepers' profile. However, when we look at the attributes of other players we see a clear variation in mean value of attributes for different roles. For example, for forwards, we observe high value for skills such as finishing, sprint speed, and acceleration. For midfielders, we see high values on ball control, dribbling, passing etc. Attributes such as Overall Rating, Age, Wage, and Potential, did not show any significant differences in their mean values of performance among all players' roles. Therefore, these were not depicted in the radar chart.

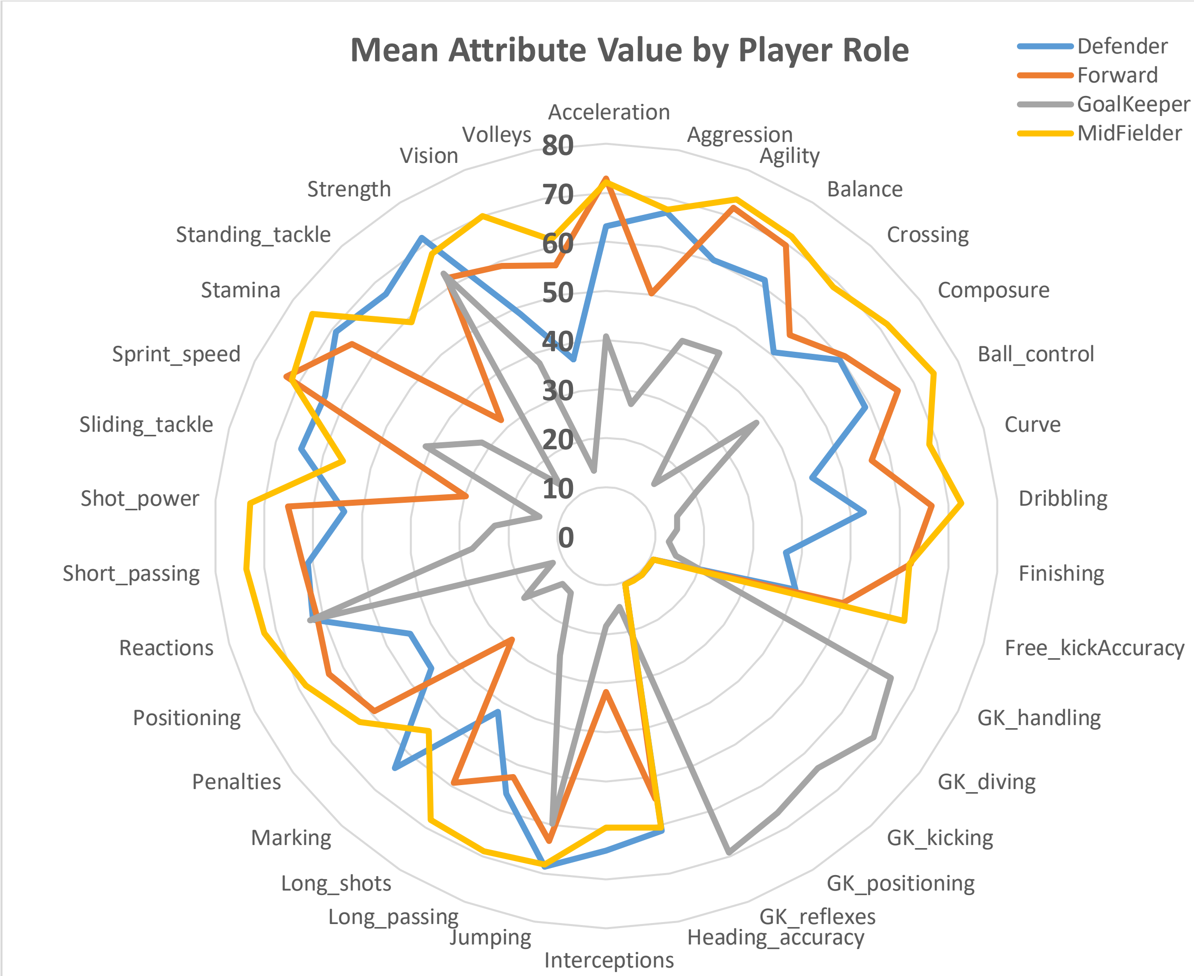

Figure 12. Radar Chart of Mean Attributes across Clusters (PlayerRole)

To further generate insights into our solution, we used a decision tree to identify the most important variables within each cluster.

Variable Importance			
	Training		
Variable	Relative	Importance	Count
GK_reflexes	1.0000	4.5375	4
GK_diving	0.9451	4.2885	2
GK_positioning	0.6321	2.8680	2
GK_handling	0.6114	2.7742	2
Short_passing	0.4968	2.2540	2

Figure 13. Variable Importance for Goalkeepers

From Figure. 13, we can see that the top skills that are desired in goalkeepers are gk_reflexes, gk_diving, gk_positioning, and gk_handling.

Variable Importance			
	Training		
Variable	Relative	Importance	Count
Ball_control	1.0000	6.3778	3
Standing_tackle	0.5188	3.3090	2
Sliding_tackle	0.4190	2.6726	2
Finishing	0.4008	2.5563	2
Dribbling	0.3976	2.5359	2
Reactions	0.3898	2.4859	1
Stamina	0.3280	2.0920	1

Figure 14. Variable Importance for Midfielders

From Figure. 14, we can see that the top skills that are desired in midfielders are ball control, tackling, dribbling, and finishing.

Variable Importance			
	Training		
Variable	Relative	Importance	Count
Ball_control	1.0000	4.3486	2
Finishing	0.8745	3.8030	4
Dribbling	0.8387	3.6472	3
Free_kickAccuracy	0.5165	2.2459	1
Heading_accuracy	0.4360	1.8958	1
Penalties	0.3847	1.6731	1
Strength	0.3283	1.4277	1
Acceleration	0.3282	1.4272	1

Figure 15. Variable Importance for Forwards

From Figure. 15, we can see that the top skills that are desired in forwards are ball control, finishing and dribbling. A forward (striker, winger) needs to have excellent ball control, finishing and dribbling skills in order to dribble through opposition’s defense and score or create scoring opportunities.

Variable Importance			
	Training		
Variable	Relative	Importance	Count
Standing_tackle	1.0000	7.8665	6
Heading_accuracy	0.3141	2.4706	2
Sprint_speed	0.3133	2.4649	2
Interceptions	0.2809	2.2093	1
Dribbling	0.2677	2.1060	1
Ball_control	0.2512	1.9758	1
Reactions	0.2228	1.7529	1
Sliding_tackle	0.1995	1.5697	1

Figure 16. Variable Importance for Defenders

From Figure. 16, we find that the top skills that are desired in defenders are standing tackle and heading accuracy. Defenders need to constantly put pressure and prevent the opposition from scoring. Their ability to tackle and win the ball back, as well as their ability to win headers play a vital role in diffusing the pressure when their team is under attack.

CONCLUSION

To generate further insights, we plotted the data points for players using two dimensions, Potential and Overall Rating (from EA Sports/FIFA18 dataset) in Figure 17. In addition, we have grouped the points based on their growth potential (defined as a difference of their Potential and Overall Rating). The growth potential may be used to narrow down the most promising players, i.e., perhaps players early in their career.

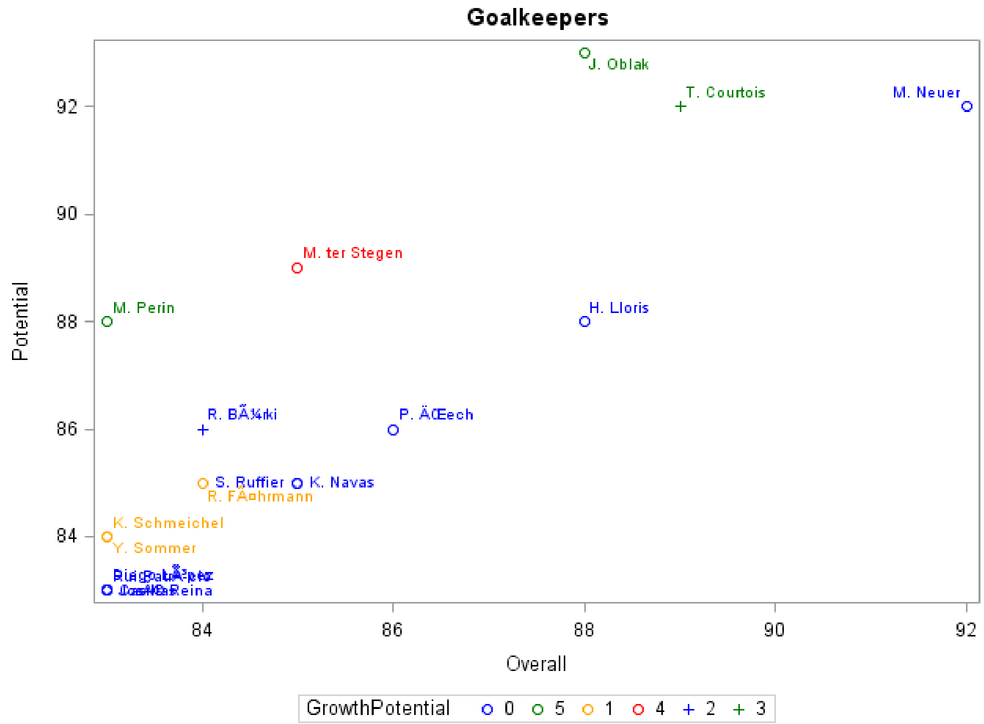

Figure 17. Scatterplot of Potential vs. Overall rating grouped by Growth Potential

Figure 17 shows that the clustering algorithm has done an excellent job at pooling in similar players together. But convincing an iconic star to move from an established team is often a challenging and expensive undertaking, however, identifying a talented player young player who is early in his career could be worth the investigation.

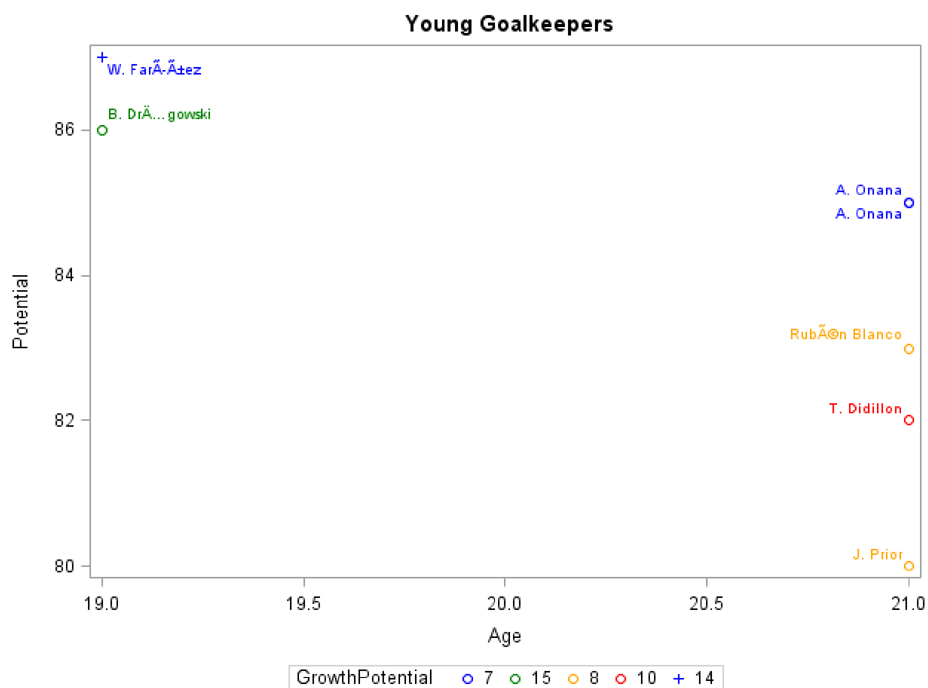

Figure 18. Scatterplot of Potential rating vs Age grouped by Growth Potential

We can see that B. Dragowski (Empoli, Goalkeeper), André Onana (Ajax, Goalkeeper), and Wuilker Faríñez Aray (Milonarios, Goalkeeper) show high potential to be an excellent goalkeeping prospect. It is a widely accepted notion among the soccer community, coaches and scouts that goalkeepers peak past 30 years of age. Given these players are in their sub and early twenties and have projected Potential in excess of 85, they are definitely worth tracking as they can easily play for another decade.

Similarly, let's look into the defender cluster, which is shown in the chart below, Figure 19.

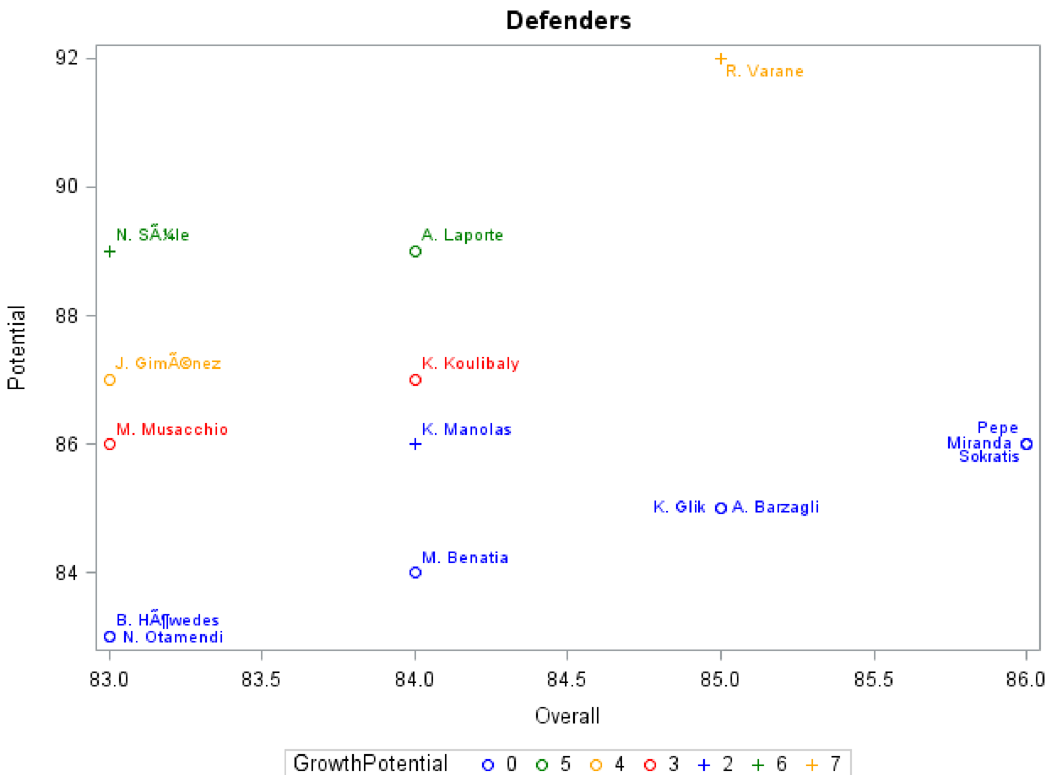

Figure 19. Scatterplot of Potential vs Overall rating grouped by Growth Potential

In the plot above, by looking at players such as Pepe, Miranda, Sokratis, Barzagli, and Varane we can confirm that the clustering algorithm has pooled together very similar players.

In order to further narrow down the most promising players, we need to look at players who are fairly young (Age < 20 years) with an Overall rating between 70 and 80 having high growth potential.

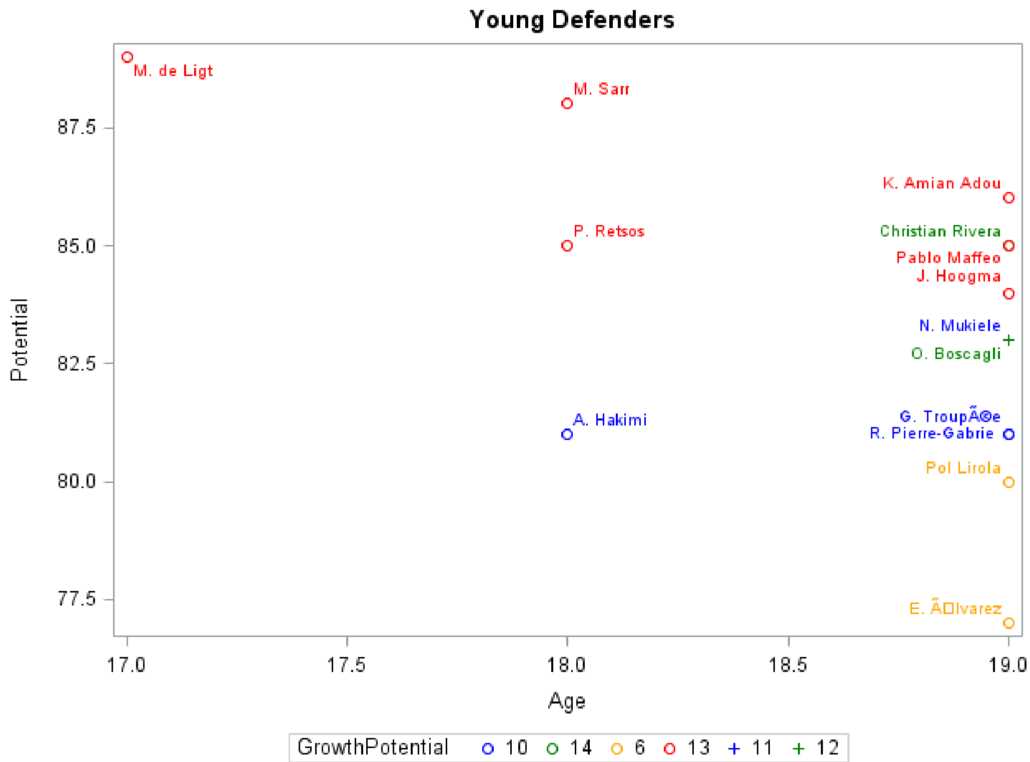

Figure 20. Scatterplot of Potential rating vs Age grouped by Growth Potential

From Figure. 20, we can quickly see that three players immediately stand out: M. de Ligt (Central Defender, Ajax), P. Retsos (Bayer 04 Leverkusen, Defender) and M. Sarr (Center Back, Nice) are extremely young and show immense potential for growth.

Similar approaches may be used to identify budding talents in the forward and midfield playing positions.

FUTURE WORK

In subsequent research, we plan to incorporate team level performance statistics and conduct a network analysis using in-game stats to decipher insights on a player's fit, given the playing style of a team. Such inferences can further empower scouts and coaches in their hunt for a suitable player based on the needs of their team's strengths and weaknesses. By bringing in historical data about the most successful players in their early and formative ages, we can significantly improve the predictive power of the model, thus enabling it to identify exceptional talents right at the beginning of their career. Based on these empirical values, we would also be able to develop a model to predict the potential market value of each player within the specified role.

REFERENCES

[1] Markovits, A. S., & Green, A. I. (2017). FIFA, the video game: a major vehicle for soccer's popularization in the United States. Sport in Society, 20(5-6). http://dx.doi.org/10.1080/17430437.2016.1158473

[2] Kabacoff, R. I. (2011). Principal components and factor analysis. In R in Action. Shelter Island, NY: Manning Publications Co.

[3] César Soto-Valero. A Gaussian mixture clustering model for characterizing football players using the EA Sports' FIFA video game system. https://www.researchgate.net/publication/315812505

[4] Di Salvo, V.; Baron, R.; Tschan, H.; Calderon Montero, F. J.; Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. International Journal of Sports Medicine, 28(3).
http://dx.doi.org/10.1055/s-2006-924294

[5] https:sofifa.com

[6] https://support.sas.com/rnd/app/stat/procedures/ClusterAnalysis.html

[7] http://logfact.com/football-soccer-field-player-positions-abbreviations/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Vishal Gaurav
Master's in Business Analytics
Oklahoma State University
+1 918-853-8112
Vishal.gaurav@okstate.edu

Goutam Chakraborty
SAS Professor of Marketing Analytics
Director of Masters in Business Analytics
Oklahoma State University
Goutam.chakraborty@okstate.edu

APPENDIX

1. GOAL KEEPER (GK)
For most of the time, goalkeeper remains in and around the penalty area. Their main work is to punch, kick, palm or catch the ball from crosses, kicks, headers, and shots from opponent players.

2. DEFENDERS (DF)
The defenders are positioned behind the midfielders with the sole aim of providing support and preventing the opponent players from scoring. They are mostly within the field half with their goal they are going to defend.

CENTRE BACKS (CB)
Also called Central Defender, their job is to stop the opposing strikers from scoring a goal and clear the ball from their penalty area.

SWEEPER (SW)
The sweeper comes into play when the opponent players breach the defensive line and try to score a goal. Sweeper is a kind of Center Back but their position is more versatile and they have more freedom compared to other defenders.

WINGBACKS (RWB) OR (LWB)

The wingbacks are a kind of defenders but they concentrate more on attacking. When the team is on the attack, they can be considered to be midfielders.

FULL BACKS OR RIGHT (RB) AND LEFT (LB)
The full-backs are those defenders that are positioned on either side of Center Backs. These are needed for protection when the wide players from opponent team attack.

3. MIDFIELDERS (MF)
The midfielders are positioned between the defenders and forwards. They are responsible for keeping the possession of the ball and taking the ball from a defender and passing it to the attackers.

DEFENSIVE MIDFIELDER (DM)
This player is positioned at the center and supports the defense by being in front of defensive players. When other midfielders support the attack, the defensive midfielder holds back.

WIDE OR LEFT AND RIGHT MIDFIELDERS (LM) AND (RM)
Also called side midfielder, they are on the left or right-hand side of central midfield. They can also be stationed at a wider position for providing more defensive protection.

CENTER MIDFIELDERS (CM)
The center midfielders have a number of duties to perform as they mainly link the attack and defense. These players initiate the attacks as well as provide defensive protection against the opponent's attack.

ATTACKING MIDFIELDERS (AM)
The attacking midfielder has a role on the offensive side and can be found between forwards and central midfield. Depending on the situation, the attacking midfield is classified into left, right and central midfield.

4. FORWARDS (FW)
Forwards are the players, who are nearest to the opponent's goal. Their job is to score a maximum number of goals or at least provide goal scoring opportunities for their team members.

CENTER, RIGHT AND LEFT FORWARDS (CF), (RWF), (LWF)
The center-forward is the main striker of the team and has the sole responsibility to score goals for his team. These players also contribute to their team's success in many other ways.

SUPPORT STRIKERS (ST)
These players are second strikers in the team and are often called support strikers. They have to create scoring opportunities for forwards, pick up loose balls, use the spaces that are created in the defense of opposition and in some cases, score themselves.