COMMUNICATIONS IN
SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

Clustering seasonal performances of soccer teams based on situational score line

Canggih Puspo Wibowo*

Sagasitas Research Center, Jalan Cendana 9, Yogyakarta, 55223, Indonesia

Article history:
Received: 7 May 2016 / Received in revised form: 13 May 2016 / Accepted: 14 May 2016

Abstract

In this research, the basic pattern of seasonal performances of soccer teams is investigated. We propose a clustering method to reveal the seasonal performance. In the proposed method, a new performance indicator called situational score line is used as a feature describing the seasonal performance. It consists of score line, opponent rating, and away rating. Using *k*-means, the features are clustered into four clusters. Cluster 1, which has a pattern of decreasing performance, is the basic pattern of Italian Serie A and German Bundesliga. Cluster 2 has a stable performance, which is mostly shown in English Premier League, Italian Serie A, and Spanish La Liga. Cluster 3 has the highest competitiveness and is one of the most common patterns in French Ligue 1 and Spanish La Liga. Finally, Cluster 4, which has a rising performance, is the basic pattern of the English Premier League.

Keywords: clustering; seasonal performance; situational score line

1. Introduction

In soccer games, there are many factors that make the performance of a certain team during a season stable. The variation in performance is due to players' conditions, such as fatigue during some matches [1], endurance [2,3], or injuries [4]. Numerous teams' performance dropped because their star players could not play the games. This is a common issue especially for teams which depend too much on particular players. Factors affecting those variations are considered seasonal. Thus, there is a pattern of team performances which might be repeated in other seasons.

Researchers have suggested some performance indicators of the soccer teams. A performance indicator is a selection, or combination, of action variables that aim to define some or all aspects of a performance [5]. Some of action variables are total shots and shots on targets [6], passing patterns [7], ball possession [8,9], and ball recovery patterns [10]. Apart from that, score line, whether a team is winning, draw, or loses, is seen as the ultimate performance indicators [11]. However, it is known that the same score line could mean differently in other matches, depending on the situation. For example, winning a match against a tough team is valued higher than against a team which is in the last place. Therefore, the quality of opponent is regarded as a crucial variable determining the performance indicator [12]. Furthermore, home-field advantage is also reckoned as one of situational variables [13,14]. It is well-known that, when playing against the same opponent, winning in the away match reflects a better result than in the home match. Hence, a new performance indicator is proposed by incorporating quality of opponent and home-field advantage into the score line.

In this work, patterns are extracted from the seasonal variation of team performance by clustering the proposed indicator. The aims are to discover and analyze the basic pattern of seasonal team performances.

2. Materials and Methods

Soccer match results are collected from five biggest leagues in Europe, namely La Liga (Spain), Premier League (England), Bundesliga (Germany), Serie A (Italy), and Ligue 1 (France), during 1993-2014 [15]. For each league, only the winning team in a season is selected as a representative. Totally, there were 110 seasonal team performances. The winning teams here are selected based on their performance on the fields. Hence, even though Serie A winner in 2004/2005 was awarded to Inter and in 2005/2006 was awarded to none (due to Calciopoli Scandal [16]), Juventus is still picked in this work because they earned the biggest points at the end of the season.

This section includes two parts. In the first part, the calculation method for the seasonal team performance indicator calculated will be explained. Using the indicator, the feature vectors can be defined. In the second part, the seasonal team performances are clustered to reveal the basic pattern.

2.1. Situational Score Line

Here, a new seasonal performance indicator is proposed.

* Corresponding author. Tel.: +62-89610868608
Email: canggih.p.w@ieee.org

This indicator consists of three measurements, namely score line, away rating, and opponent rating.

2.1.1. Score line

Score line is the outcome of a match played by a particular team, which is calculated using the same formula as the standard soccer rule. The score line ($p(t_n)$) at a matchday t_n is defined as

$$p(t_n) = \begin{cases} 3, \text{if the team won at matchday } t_n \\ 1, \text{if the team drew at matchday } t_n \\ 0, \text{if the team lost at matchday } t_n \end{cases} \quad (1)$$

2.1.2. Away Rating

Home-field advantage gives a huge benefit for a team. Therefore, in order to measure the team's performance, we should take a look at performances in the away matches. The away rating measures the performance of a team in the away matches. It is defined as the total point earned in the away matches divided by the total maximum away points. Thus, the away rating $l(t_n)$ is calculated as

$$l(t_n) = \frac{2}{3N}\sum_{n=0}^{N-1} p(t_n)h(t_n), \quad (2)$$

where,

$$h(t_n) = \begin{cases} 1, \text{if away match at matchday } t_n, \\ 0, \text{if home match at matchday } t_n, \end{cases}$$

and N is the number of matches played by the team in a season.

2.1.3. Opponent Rating

Opponent rating measures the quality of the opponent team in a particular match. It is calculated by dividing the total point earned by the opponent team at the end of the season by the total maximum point that can be earned. The opponent rating at a matchday t_n is defined as follows

$$o(t_n) = \frac{1}{3N}\sum_{n=0}^{N-1} p(t_n), \quad (3)$$

The three measurements are then combined into a seasonal performance called situational score line, $s(t_n)$, given by

$$s(t_n) = p(t_n)[l(t_n) + o(t_n)], \quad (4)$$

The comparison of performance indicator based on the score line ($p(t_n)$) and situational score line ($s(t_n)$) is shown in Fig. 1. It can be seen that $s(t_n)$ has more number of peaks and more distinct values of performance than $p(t_n)$. It means that more changes in the performance can be observed by using $s(t_n)$.

It is known that each league has different performance scale (UEFA country coefficient). Thus, in order to make it on the same scale, normalization with respect to the scale was performed. The method was discussed in [17]. Furthermore, the numbers of matches in a season are not all entirely the same for different leagues, e.g., every Bundesliga team had to play 42 matches during 1995-1996 while every Serie A team only had to play 34 matches for each season within 1993-2003. Thus, the number of matches needs to be normalized. In this case, length normalization with the same method as [18] was performed. It was done by connecting two adjacent points with a straight line.

Let $\hat{s}(t)$ be the result of scale and length normalization of $s(t_n)$. Then, in order to reveal the pattern of team performance as well as to reduce the fluctuation, it is assumed that $\hat{s}(t)$ is an even periodic function of time. Afterwards, $\hat{s}(t)$ is expanded into a Fourier series as

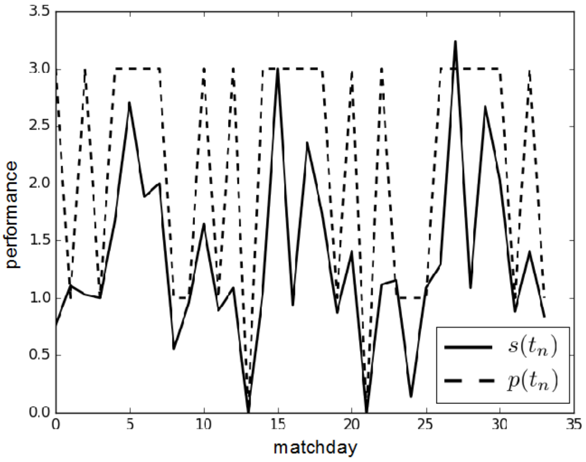


Fig. 1 Comparison between situational score line $s(t_n)$ and score line $p(t_n)$.

$$\hat{s}(t) = \frac{a_0}{2} + \sum_{k=1}^{K-1} a_k \cos\left(\frac{k\pi t}{T}\right), \quad (5)$$

where,

$$a_k = \frac{2}{T}\int_0^T \hat{s}(t)\cos\left(\frac{k\pi t}{T}\right)dt, \quad (6)$$
$$(k = 0,1,2,\ldots,K-1),$$

and K and P are the number of Fourier coefficients and period of function of time, respectively. The example of performance indicator and its Fourier approximation in a season is shown in Fig. 2. Using the Fourier approximation, the performance curve can be made smoother and the fluctuation is reduced.

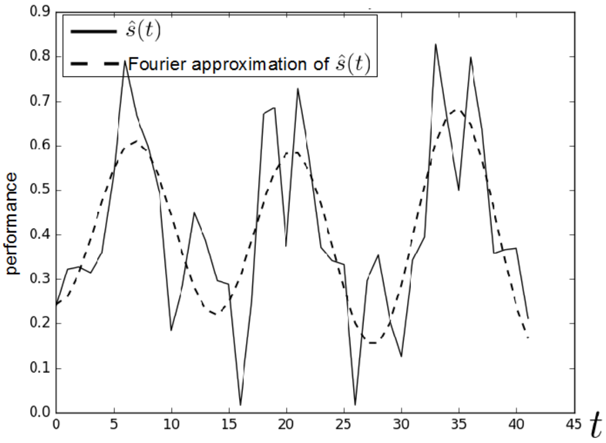


Fig. 2 Fourier approximation of situational score line $\hat{\mathbf{s}}$(t).

2.2. Clustering Seasonal Teams' Performance

The basic pattern of the seasonal performance is revealed by clustering the features of performance. Let X be a set of feature vectors $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_M\}$, where M is the total number of data points. The feature vectors are defined as

$$\mathbf{x}_p = [a_0, a_1, a_2, \ldots, a_{K-1}]', (p = 1,2,3,\ldots,M) \quad (7)$$

where $\mathbf{x}_p{}'$ means the transposition of $\mathbf{x}_p$. Here, Fourier coefficients a_k were considered as features reflecting the seasonal team performance in a season. Then, one of the most used clustering method, k-means, is employed here. The goal of k-means is to minimize the within-cluster distance (J_w) defined as

$$J_w = \sum_{i=1}^{C} \sum_{\mathbf{x}_p \in S_i} d(\mathbf{x}_p, \mu_i)^2, \qquad (8)$$

where μ_i, C, and S_i are the cluster center, the number of clusters, and the set of data points that belong to cluster i, respectively. Whilst, $d(\mathbf{x}_p, \mu_i)$ is the Euclidean distance between $\mathbf{x}_p$ and μ_i. To solve the clustering problem in (8), Lloyd's algorithm is used as follows.

1. Initialize the cluster center μ_i and decide the number of cluster$C$.
2. Assign data points to nearest cluster
   $$S_i = \{\mathbf{x}_p : \|\mathbf{x}_p - \mu_i\|^2 \leq \|\mathbf{x}_i - \mu_j\|^2 \; \forall j, 1 \leq j \leq C.$$
3. Recalculate cluster center
   $$\mu_i = \frac{1}{|S_i|} \sum_{\mathbf{x}_p \in S_i} \mathbf{x}_p.$$
   where $|S_i|$ is the number of elements in S_i.
4. Repeat step 2 and 3 until convergence is obtained.

In Lloyd's algorithm, there are two parameters that have to be determined in advance. They are the initial cluster center and the number of clusters. It is known that both of them affect the results of clustering. To set the initial cluster centers, instead of using random values, a method called k-means++ can be employed [19]. The algorithm is as follows.

1. Choose a cluster center μ_1 at random from the set of features X.
2. Compute $D(\mathbf{x}_p)^2$, the distance between data point $\mathbf{x}_p$ to the nearest cluster center.
3. Choose a new cluster center μ_i from data point which has a probability $D(\mathbf{x}_p)^2 / \sum_{\mathbf{x}_p \in \mathrm{X}} D(\mathbf{x}_p)^2$.
4. Repeat step 2 and 3 until C cluster centers are selected
5. Proceed with Lloyd's algorithm.

Furthermore, in order to determine the number of clusters (C), a method involving cluster validity is used [20]. Cluster validity is described as a ratio between average of within-cluster distance J_w (8) and the minimum distance between cluster centers J_b.

$$\text{validity} = \frac{J_w}{MJ_b} \qquad (9)$$

where,

$$J_b = \min\left(\|\mu_i - \mu_j\|^2\right),$$
$$(i = 1,2,\ldots,C-1), (j = i+1,\ldots,C)$$

The k-means clustering was computed repeatedly with different number of clusters. The appropriate number of clusters (C) is determined using the Elbow method [21], employing cluster validity as the cost function.

3. Results and Discussion

The clustering process results in four clusters (the complete results are shown in Table 2). In this work, the cluster centers are considered as the basic pattern of seasonal performance. Furthermore, cluster results for each league are presented to show the difference of pattern in each league.

3.1. Seasonal Performance Patterns

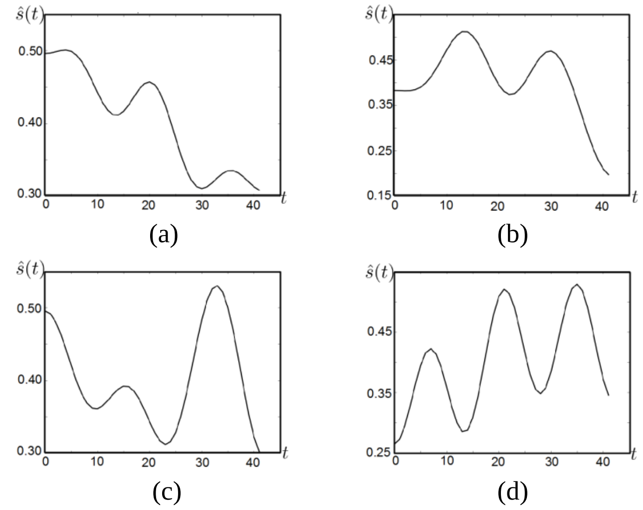
Fig. 3 shows the seasonal performance of each cluster, represented by its centers. It can be observed that all patterns have up- and down-periods. No teams could maintain top performance throughout the season. Moreover, all basic patterns even dropped at the end of the season. This is common to happen since most of the teams, 82 out of 110, were also competing in European competition until the second half of the season. Dealing with more matches against many tough opponents from different countries made a team struggling harder in the later half.

The basic pattern of cluster 1 is shown in Fig. 3(a). It is shown that the team performance in this cluster was dropping down until the end of the season. The teams only had a slight improvement in the middle of the season. Then, after mid-season break, they managed to bring a different performance. As a result, in the beginning of the second half, the performance was raised. However, in the end, it still dropped. Price *et al.* mentioned that rates of injury are increasing after the mid-season break [22]. Obviously, the team's performance will be influenced by the player's injury. Apart from that, even though the teams had a significant decrease in performance, at the end they still managed to win the league. Therefore, this cluster shows the dominance of the winners.

Cluster 2, which is shown in Fig. 3(b), tends to have a stable performance. From the beginning until about 1.5 months before the season ended, there was no significant decrease in performance. It means that teams started falling down when they only had approximately 5 matches to play. Such performance is expected because 36% teams in this cluster were qualified at least until semifinals of European competition (26%, 18%, and 30% in cluster 1, 3, and 4 respectively). Moreover, if we look at the average point margin to the second places in the corresponding league, it shows dominance. The margin in this cluster is 9.36 points (6.6, 4.32, and 5.26 in clusters 1, 3, and 4, respectively). Therefore, teams in this cluster are considered dominant in their league.

In cluster 3, shown in Fig. 3(c), for the first half of the season, the performance dropped significantly. However, after the mid-season break, it raised high until the last quarter of the season, and finally dropped again. Similar to the trend of cluster 1, the turning point was in the middle of the season. Furthermore, among all four clusters, teams in this cluster had the smallest winning margin, that is 4.32 points. It can be concluded that this cluster has the highest level of competitiveness.

The basic pattern of cluster 4 is shown in Fig. 3(d). Generally, this cluster has a slight rising performance. It can be seen from the end of the season which has a higher performance than the beginning. In this case, it is distinct from the other three. This cluster also has a repeatable pattern, repeated three times. We can figure out that the first turning point, happened around September, is the start of European competition group stage, in which teams were taking part. The second one is around February, when the knockout phase is starting. Hence, the teams in this cluster were heavily affected by the European competition.



(a) (b)

(c) (d)

Fig. 3 Basic pattern of each cluster: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, and (d) Cluster 4.

3.2. Cluster Results in Each League

Clustering result displaying the corresponding cluster number in each league is shown in Table 1. It is known that Serie A of Italy was once recognized as the best football leagues in the world. However, the quality started dropping since 2000 due to financial problems [23]. Teams could not afford to pay their star players. Thus many star players left Italy and moved to clubs in another country. As a result, it was tough for Italian teams to compete consistently in European competitions. Furthermore, it also affected the performance of the teams. In Table 1 column Serie A, it is shown that after 2000, there were changes in the performance pattern. The winners tent to have performance similar to that in clusters 1 and 2. The winners that belong to cluster 1 became the UEFA Champions League (UCL) finalists three times (Juventus 2002/2003, Inter 2009/2010, and Juventus 2014/2015). This means that for these teams the decreasing performance in the league did not affect the performance in the UCL. On the other hand, only Milan in 2003/2004, teams who belong in cluster 2, who qualified until quarter finals of UCL. Thus, the teams in cluster 2 were only dominating the league but not European competition.

In the English Premier League, most of the team performances are grouped in cluster 2 and 4. However, after 2003/2004, pattern of cluster 2 appeared more frequently. Even though cluster 2 was the major pattern, the teams' most successful period was belonging to cluster 4, when Manchester United played in UCL finals two times in a row (2007/2008 and 2008/2009).

In the Spanish La Liga, Barcelona and Real Madrid are the two most successful teams. Since 1993, for the last 22 seasons, Barcelona has won the league 10 times, while Real Madrid 7 times. Moreover, with exceptions in 1995/1996 and 2001/2002, at least one of them ended the league in the top two. It can be implied that both are taking rule the league. The results in Table 1 align with the facts. Most performance patterns belong to cluster 2, as we expect from a league with such dominating teams. The second most common pattern is cluster 3. In this cluster, other teams also had quality to oppose those two big teams, such as Deportivo La Coruna, Athletic Bilbao, Valencia, and Atletico Madrid. Thus, the league was more competitive when other teams become real contenders to Barcelona and Real Madrid.

In German Bundesliga, we can see that before 2007/2008 the performance trends used to belong to cluster 1, which are decreasing performance throughout the season. However, after that, the performance patterns became stable, with four seasons are classified in cluster 2. In those four seasons, the winners were leading with large margins, 15.25 points difference to the second place. For the last three years, Bayern Munich has been winning by an average 18 points margin. With these trends, we can expect that in the following seasons, the winners of Bundesliga still will follow the performance trends of cluster 2.

In the French Ligue 1, performance patterns in cluster 3 were mostly shown in the winner's team. It means that actually the competition here was tight. In the duration of 2001/2002 until 2007/2008, Lyon created history by winning the league seven times consecutively. In that period, the competition becomes quite strict in 2001/2002, 2002/2003, and 2007/2008. While in the other four seasons, Lyon dominated the league. After that, a team dominated the Ligue 1 again starting 2012/2013 when PSG won the league 3 times in a row.

Table 1. Cluster results in each league: SA = Serie A; PL = Premier League; LL = La Liga; BL = Bundesliga; L1 = Ligue 1

Season	SA	PL	LL	BL	L1
1993/1994	2	1	3	2	2
1994/1995	2	2	2	1	1
1995/1996	3	3	1	1	4
1996/1997	3	4	2	3	4
1997/1998	4	3	3	1	4
1998/1999	4	4	4	1	3
1999/2000	3	3	2	4	2
2000/2001	1	2	2	4	3
2001/2002	4	4	4	1	3
2002/2003	1	4	2	2	3
2003/2004	2	1	3	2	2
2004/2005	3	2	2	4	1
2005/2006	1	1	2	1	2
2006/2007	2	2	3	1	2
2007/2008	1	4	1	1	3
2008/2009	2	4	2	3	4
2009/2010	1	3	3	4	3
2010/2011	2	2	2	2	3
2011/2012	3	1	4	4	3
2012/2013	3	2	1	2	4
2013/2014	2	4	3	2	2
2014/2015	1	2	3	2	3

4. Conclusion

Four categories of soccer performance were derived from clustering the seasonal team’s performance in Europe's five biggest leagues. Cluster 1, which has a pattern of decreasing performance, is the basic pattern of the Italian Serie A and German Bundesliga. Cluster 2 has a stable performance. It is mostly shown in the English Premier League, Italian Serie A, and Spanish La Liga. Cluster 3 has a highest competitiveness and is one of the most common patterns in French Ligue 1 and Spanish La Liga. Finally, Cluster 4, which has a rising performance, is the basic pattern of the English Premier League.

Table 2. Clustering results

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Man United(1993/1994)	Milan(1993/1994)	Barcelona(1993/1994)	Auxerre(1995/1996)
Dortmund(1994/1995)	Bayern Munich(1993/1994)	Man United(1995/1996)	Man United(1996/1997)
Nantes(1994/1995)	Paris SG(1993/1994)	Milan(1995/1996)	Monaco(1996/1997)
Atletico Madrid(1995/1996)	Blackburn(1994/1995)	Juventus(1996/1997)	Juventus(1997/1998)
Dortmund(1995/1996)	Juventus(1994/1995)	Bayern Munich(1996/1997)	Lens(1997/1998)
Kaiserslautern(1997/1998)	Real Madrid(1994/1995)	Arsenal(1997/1998)	Man United(1998/1999)
Bayern Munich(1998/1999)	Real Madrid(1996/1997)	Barcelona(1997/1998)	Milan(1998/1999)
Roma(2000/2001)	La Coruna(1999/2000)	Bordeaux(1998/1999)	Barcelona(1998/1999)
Dortmund(2001/2002)	Monaco(1999/2000)	Man United(1999/2000)	Bayern Munich(1999/2000)
Juventus(2002/2003)	Man United(2000/2001)	Lazio(1999/2000)	Bayern Munich(2000/2001)
Arsenal(2003/2004)	Real Madrid(2000/2001)	Nantes(2000/2001)	Arsenal(2001/2002)
Lyon(2004/2005)	Real Madrid(2002/2003)	Lyon(2001/2002)	Juventus(2001/2002)
Chelsea(2005/2006)	Bayern Munich(2002/2003)	Lyon(2002/2003)	Valencia(2001/2002)
Juventus(2005/2006)	Milan(2003/2004)	Valencia(2003/2004)	Man United(2002/2003)
Bayern Munich(2005/2006)	Werder Bremen(2003/2004)	Juventus(2004/2005)	Bayern Munich(2004/2005)
Stuttgart(2006/2007)	Lyon(2003/2004)	Real Madrid(2006/2007)	Man United(2007/2008)
Inter(2007/2008)	Chelsea(2004/2005)	Lyon(2007/2008)	Man United(2008/2009)
Real Madrid(2007/2008)	Barcelona(2004/2005)	Wolfsburg(2008/2009)	Bordeaux(2008/2009)
Bayern Munich(2007/2008)	Barcelona(2005/2006)	Chelsea(2009/2010)	Bayern Munich(2009/2010)
Inter(2009/2010)	Lyon(2005/2006)	Barcelona(2009/2010)	Real Madrid(2011/2012)
Man City(2011/2012)	Man United(2006/2007)	Marseille(2009/2010)	Dortmund(2011/2012)
Barcelona(2012/2013)	Inter(2006/2007)	Lille(2010/2011)	Paris SG(2012/2013)
Juventus(2014/2015)	Lyon(2006/2007)	Juventus(2011/2012)	Man City(2013/2014)
	Inter(2008/2009)	Montpellier(2011/2012)	
	Barcelona(2008/2009)	Juventus(2012/2013)	
	Man United(2010/2011)	Atletico Madrid(2013/2014)	
	Milan(2010/2011)	Barcelona(2014/2015)	
	Barcelona(2010/2011)	Paris SG(2014/2015)	
	Dortmund(2010/2011)		
	Man United(2012/2013)		
	Bayern Munich(2012/2013)		
	Juventus(2013/2014)		
	Bayern Munich(2013/2014)		
	Paris SG(2013/2014)		
	Chelsea(2014/2015)		
	Bayern Munich(2014/2015)		

References

1. M. Mohr, P. Krustrup, and J. Bangsbo, *Match performance of high-standard soccer players with special reference to development of fatigue*, J. Sports Sci. 21 (2003) 519–528.

2. C. Castagna, F. Impellizzeri, E. Cecchini, E. Rampinini, and J. C. B. Alvarez, *Effects of intermittent-endurance fitness on match*

performance in young male soccer players, J. Strength Cond. Res. 23 (2009) 1954–1959.

3. P. S. Bradley, C. Carling, A. G. Diaz, P. Hood, C. Barnes, J. Ade, M. Boddy, P. Krustrup, and M. Mohr, *Match performance and physical capacity of players in the top three competitive standards of English professional soccer*, Hum. Mov. Sci. 32 (2013) 808–821.
4. Arnason, S. B. Sigurdsson, A. Gudmundsson, I. Holme, L. Engebretsen, and R. Bahr, *Physical fitness, injuries, and team performance in soccer*, Med. Sci. Sports Exerc. 36 (2004) 278–285.
5. M. D. Hughes and R. M. Bartlett, *The use of performance indicators in performance analysis*, J. Sports Sci. 20 (2002) 739–754.
6. J. Castellano, D. Casamichana, and C. Lago, *The use of match statistics that discriminate between successful and unsuccessful soccer teams*, J. Hum. Kinet. 31 (2012) 139–147.
7. Redwood-Brown, *Passing patterns before and after goal scoring in FA Premier League Soccer*, Int. J. Perform. Anal. Sport 8 (2008) 172–182.
8. Lago and R. Martín, *Determinants of possession of the ball in soccer*, J. Sports Sci. 25 (2007) 969–974.
9. P. D. Jones, N. James, and S. D. Mellalieu, *Possession as a performance indicator in soccer*, Int. J. Perform. Anal. Sport 4 (2004) 98–102.
10. D. Barreira, J. Garganta, P. Guimaraes, J. Machado, and M. T. Anguera, *Ball recovery patterns as a performance indicator in elite soccer*, J. Sport. Eng. Technol. 228 (2014) 61–72.
11. F. Clemente, M. Couceiro, F. M. L. Martins, and R. Mendes, *Team's performance on FIFA U17 World Cup 2011: Study based on notational analysis*, J. Phys. Educ. Sport 12 (2012) 13–17.
12. H. Folgado, R. Duarte, O. Fernandes, and J. Sampaio, *Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches*, Public Libr. Sci. ONE 9 (2014) 1–9.
13. C. H. Almeida, A. P. Ferreira, and A. Volossovitch, *Effects of match location, match status and quality of opposition on regaining possession in UEFA Champions League*, J. Hum. Kinet. 41 (2014) 203–214.
14. F. Carmichael and D. Thomas, *Home-field effect and team performance:Evidence from English Premiership football*, J. Sports Econom. 6 (2005) 264–281.
15. J. Buchdahl, *Historical football results*. Available: http://football-data.co.uk
16. FIGC, *Testo Della Decisione Relativa al Comm. Uff. N. 1/C – Riunione del 29 Giugno / 3 - 4 - 5 - 6 - 7 Luglio 2006 (Italian)*, 2006.
17. C. P. Wibowo, P. Thumwarin, and T. Matsuura, *On-line signature verification based on angles of pen-motion*, Simulation Technology, 32nd Int. Conf., Tokyo, Japan, 2013, pp 1-2.
18. C. P. Wibowo, P. Thumwarin, and T. Matsuura, *On-line signature verification based on forward and backward variances of signature*, Information and Communication Technology, Electronic, and Electrical Engineering, 4th Joint Int. Conf., Chiang Rai, Thailand, 2014, pp. 1–5.
19. D. Arthur and S. Vassilvitskii, *k-means++: The advantages of careful seeding*, Discrete algorithms, 18th Annual ACM-SIAM Sympos., Philadelphia, PA, USA, 2007, pp. 1027–1035.
20. S. Ray and R. H. Turi, *Determination of number of clusters in K-Means clustering and application in colour image segmentation*, Advances in Pattern Recognition and Digital Techniques, 4th Int. Conf., Calcutta, India, 1999, pp. 137-143.
21. T. M. Kodinariya and P. R. Makwana, *Review on determining number of cluster in K-Means clustering*, Int. J. Adv. Res. Comput. Sci. Manag. Stud. 1 (2013) 90–95.
22. R. J. Price, R. D. Hawkins, M. A. Hulse, and A. Hodson, *The football association medical research programme: an audit of injuries in academy youth football*, Br. J. Sports Med. 38 (2004) 466–471.
23. J. Goddard and P. Sloane, *Handbook on the economics of professional football*, Cheltenham, UK: Edward Elgar Pub, 2014.