



Optimization of The Best Line-up in Football using Binary Integer Programming Model

Mahrudinda^{1*}, Sudrajat Supian¹, Subiyanto², Diah Chaerani¹

¹*Departement of Mathematics,*

Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia

²*Departement of Marine Sciences,*

Faculty of Fishery and Marine Sciences, Universitas Padjadjaran, Indonesia

Corresponding Author(s) email: mahrudinda19001@mail.unpad.ac.id

Abstract

This paper aims to find the formation with the best line-up of the Liverpool FC football team in the English Premier League in the 2020/2021 season. Researchers used binary integer programming (BIP) modeling to determine optimum solutions. The data used for this optimization is the *rating* value of the players recorded in the performance data from the previous matches. The optimum result of this problem is the selection of variables that are valued at 1, namely $\{x_1, x_4, x_6, x_8, x_{11}, x_{18}, x_{21}, x_{28}, x_{34}, x_{37}, \text{ dan } x_{39}\}$ for formations 4-3-3 with a maximum value of 82.47, and variables $\{x_1, x_6, x_7, x_8, x_{11}, x_{14}, x_{16}, x_{29}, x_{31}, x_{32}, \text{ dan } x_{42}\}$ for 4-2-3-1 formations with a maximum value of 80.04. The 4-3-3 formation is more effective because it has a higher maximum *rating* than the 4-2-3-1 formation. 4-3-3 formation is an attacking formation with a higher intensity of attack and faster than 4-2-3-1 formation that tends to defend moderately.

Keywords: Sport, football, football Line-up, football formations, optimization, binary linear programming, binary integer programming

1. Introduction

Formation in football is a very important thing element in implementing a coach's strategy. In addition to the formations used, the coach is also very concerned about how the selection of the line-up of players will be used as starting eleven. The coach will prepare some line-up plans that will be used in a match to be played. The line-up used will see and depend on the pattern or formation that the opponent will takedown. Many previous studies have discussed formation in football. The paper (Bradley et al., 2011), describes the comparison of formations in football used in the English premier league. The study formulated the influence of playing formations on high-intensity runs and technical performance during elite football matches. The results state that 4-3-3 formations produce the highest intensity compared to other formations. Meanwhile, for the best ball possession obtained by a 4-2-2 formation. It is well known that the 4-3-3 formation does not touch the ball much 1-2, but instead relies on long passes. It was also mentioned in the research (Carling, 2011), that the 4-2-2 formation gives more role to the midfielder in

doing the bait spread. A formation can be said to be good if it gets an effective playing effect in the match. In addition to the winning score, there are other important elements, namely, the movement pattern determined by the Global Positioning System (GPS). On the 5 most common game formations (4-4-2; 4-3-3; 3-5-2; 3-4-3; 4-2-3-1) used in 11-on-11 football games in England, GPS includes, Total distance data (TD), high speed (HSR), high metabolic load distance (HMLD), high-speed acceleration (Acc), and deceleration (Dec) (Venter et al., 2011; Aquino et al., 2019; Aquino et al., 2019). The 5 formations can be applied capriciously depending on the situation of the game during the game. Teams choose between defensive and attacking formations and between hard playing styles and soft playing styles (Dobson and Goddard, 2010). In other conditions, the trainer will also apply win-stay loss-shift, in determining formation decisions. In other words, they tend to stick to the current formation after winning and switch to another formation after losing. This is practical, but not a good way to choose a formation (Tamura and Masuda, 2015).

This research discusses how a trainer can determine the best line-up that can be used on starting eleven formations. Performa players in previous matches are considered, then selected the most maximum players to occupy the position in the formation that the coach will determine. Unlike previous studies (He et al., 2018; Ge et al., 2020; Be et al., 2020), which used stage multi modeling in determining formations, this study used optimization methods in selecting the maximum performance of the players obtained from historical data. So the study selected a formation, then determined a suitable line-up in the formation. The optimization method used is the optimization of the binary integer programming (BIP) model. The previous most common and frequently used BIP was in scheduling issues distributing an object (Pan and Chen, 2005; Correa et al., 2015; Wong et al., 2006; Brey et al., 2012; Ziaee and Sadjadi, 2007; Balouchzahi et al., 2015; Gholamnejad and Osanloo, 2007). While scheduling on this research expanded its use into the formation of player line-ups. In this study also, we can see that the problem-solving BIP complex and many variables, will be able to be solved with mathematical computing.

2. Materials and Methods

2.1. Materials

This research seeks to determine the best line-up by choosing the optimal solution of player *rating* performance based on historical data, by identifying the position of each player that will be used as an obstacle. The research focuses on data on Liverpool FC players who competed in the Premier League in the 2020/2021 season until week 11. This optimization maximizes the *rating* of the players (see Table 1) already loaded on the web: <https://www.whoscored.com/>. The player's *rating* is accumulated based on several aspects of value, namely, minutes played, total assists, pass success percentage, total goals, shots per game, aerial duels won per game, and man of the match.

The position of the players is generalized into 10 important positions, namely, GK (goalkeeper), DC (defender center), DL (defender left), DR (defender right), DM (defensive midfielder), CM (central midfielder), AM (attack midfielder), FWL (forward left), FWR (forward right), and FW (forward). From those 10 positions, the next 11 players will be selected to fill the starting eleven, based on the team formation often used by Liverpool FC. These are 4-3-3 attack formations and medium defensive formations 4-2-3-1 (see Figure 1).

Table 1. Player rating based on frequently played positions in the 2020/2021 season

Player	Position									
	GK	DC	DL	DR	DM	CM	AM	FWL	FWR	FW
Alisson	6.81	-	-	-	-	-	-	-	-	-
Adrián	5.86	-	-	-	-	-	-	-	-	-
van Dijk	-	6.62	-	-	-	-	-	-	-	-
Gomez	-	6.91	-	-	-	-	-	-	-	-
Robertson	-	-	6.85	-	6.97	6.64	-	-	-	-
Matip	-	6.69	-	-	-	-	-	-	-	-
Phillips	-	7.24	-	-	-	-	-	-	-	-
Arnold	-	-	-	6.93	6.4	-	-	-	-	-
Williams	-	-	-	6.63	7.66	7.77	-	-	-	-
Fabinho	-	7.11	-	-	-	6.58	-	-	-	-
Wijnaldum	-	-	-	-	6.64	6.74	7.16	-	-	-
Thiago	-	-	-	-	-	7.38	-	-	-	-
Milner	-	-	-	8.15	-	6.97	-	-	-	-
Keïta	-	-	-	-	-	6.43	-	-	-	-
Henderson	-	-	-	-	6.8	6.7	-	-	-	-
Jones	-	-	-	-	-	7.13	-	-	-	-
Shaqiri	-	-	-	-	-	6.49	6.52	-	-	-
Firmino	-	-	-	-	-	-	6.8	-	-	6.99
Mané	-	-	-	-	-	-	7.24	7.56	-	-
Salah	-	-	-	-	-	-	-	-	7.42	6.8
Minamino	-	-	-	-	-	6.05	-	-	-	6.03
Jota	-	-	-	-	-	9.39	6.77	7.23	7.84	8.22
Origi	-	-	-	-	-	-	-	6.4	-	5.84



Figure 1. The line-up formations that will be used.

The completion of optimum solutions in this research will use computing with the help of *R Programming* software.

2.2. Methods

The data obtained, then done BIP modeling by maximizing an objective function. The modeling is generally written as follows (Hillier and Lieberman, 1995):

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (1)$$

subject to the restrictions,

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \quad (2)$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \quad (3)$$

\vdots

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m, \quad (4)$$

and,

$$x_1, x_2, \dots, x_n = 0 \text{ or } 1 \text{ (binary)} \quad (5)$$

Maximize of the Z function in (1), is the maximum value of the *rating* of Liverpool FC players to be selected into the Line-up. x_n is a variable that represents each cell that contains the *rating* of the players (see Table 1). x_1, x_2, \dots, x_{43} sorted from top to bottom from the first column to the last column in table 1. For example, x_1 is Alisson as GK position, x_2 is Adrian as GK position, so x_{43} is Origi as FW position. Note that equation (5) requires that each variable x_n to be specified is a binary value. The Binary criteria, meaning 0 is "NO", and 1 is "YES". "YES" means the absolute decision that the player in question is selected as one of the players for the starting eleven.

Note figure 1 that for 4-3-3 formations require DC as many as 2 players, CM as many as 3 players, and for GK, DL, DR, FWR, FW, and FWL will be selected as many as 1 player each. As for the 4-2-3-1 formation, 2 dc players will be selected, 2 players as DM, 3 players in AM position, and for players with GK, DL, DR, and FW positions, 1 player will be selected.

3. Results and Discussion

Based on the equation (1) - (4) it can be written a following objective function,

$$\text{Maximize } Z = 6.81x_1 + 5.86x_2 + 6.62x_3 + \dots + 6.03x_{41} + 8.22x_{42} + 5.84x_{43} \quad (6)$$

The constraint function is separated into 2 parts each to maximize the line-up of 4-3-3 formations and 4-2-3-1 formations.

Subject to the restrictions (for a formation 4-3-3),

$$x_1 + x_2 + x_3 + \dots + x_{41} + x_{42} + x_{43} = 11 \quad (7)$$

$$x_1 + x_2 = 1, \quad (8)$$

$$x_3 + x_4 + x_5 + x_6 + x_7 = 2, \quad (9)$$

$$x_8 = 1, \quad (10)$$

$$x_9 + x_{10} + x_{11} = 1, \quad (11)$$

$$x_{17} + x_{18} + x_{19} + x_{20} + x_{21} + x_{22} + x_{23} + x_{24} + x_{25} + x_{26} + x_{27} + x_{28} = 3, \quad (12)$$

$$x_{34} + x_{35} + x_{36} = 1, \quad (13)$$

$$x_{37} + x_{38} = 1, \quad (14)$$

$$x_{39} + x_{40} + x_{41} + x_{42} + x_{43} = 1, \quad (15)$$

$$x_8 + x_{17} \leq 1, \quad (16)$$

$$x_{10} + x_{18} \leq 1, \quad (17)$$

$$x_7 + x_{19} \leq 1, \quad (18)$$

$$x_{11} + x_{22} \leq 1, \quad (19)$$

$$x_{37} + x_{40} \leq 1, \quad (20)$$

$$x_{27} + x_{41} \leq 1, \quad (21)$$

$$x_1, x_2, x_3, \dots, x_{43} \text{ are binary} \quad (25)$$

$$x_1, x_2, x_3, \dots, x_{43} \text{ are binary} \quad (42)$$

[illegible]

From computing completion, optimum solutions are obtained to determine the best line-up in the 4-3-3 formation of $\{x_1, x_2, x_3, \dots, x_{43}\} = \{1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0\}$ with a maximum value of 82.47. While the optimum solution to determine the best line-up in the 4-2-3-1 formation is $\{x_1, x_2, x_3, \dots, x_{43}\} = \{1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0\}$ with a maximum *rating* of 80.04. So each variable is selected as the best line-up of the two successive formations, namely $\{x_1, x_4, x_6, x_8, x_{11}, x_{18}, x_{21}, x_{28}, x_{34}, x_{37}, x_{39}\}$ and $\{x_1, x_6, x_7, x_8, x_{11}, x_{14}, x_{16}, x_{29}, x_{31}, x_{32}, x_{42}\}$ (see Table 2).

4-3-3 Formation				4-2-3-1 Formation			
Binary variable has a value of 1	Player Name	Position	Rating	Binary variable has a value of 1	Player Name	Position	Rating
x_1	Alisson	GK	6.81	x_1	Alisson	GK	6.81
x_4	Gomez	DC	6.91	x_6	Philips	DC	7.24
x_6	Philips	DC	7.24	x_7	Fabinho	DC	7.11
x_8	Robertson	DL	6.85	x_8	Robertson	DL	6.85
x_{11}	Milner	DR	8.15	x_{11}	Milner	DR	8.15
x_{18}	Williams	MC	7.77	x_{14}	Williams	DM	7.66
x_{21}	Thiago	MC	7.38	x_{16}	Henderson	DM	6.80
x_{28}	Jota	MC	9.39	x_{29}	Wijnaldum	AM	7.16
x_{34}	Mane	FWL	7.56	x_{31}	Firminho	AM	6.80
x_{37}	Salah	FWR	7.42	x_{32}	Mane	AM	7.24
x_{39}	Firminho	FW	6.99	x_{42}	Jota	FW	8.22
Max Z			82.47	Max Z			80.04

The optimum solution, shown in Table 2, can be formed into the line-up formation in Figure 2.



Figure 2. The best line-up for the 4-3-3 and 4-2-3-1 formations

Note that the maximum *rating* in the 4-3-3 (82.47) formation is greater than the maximum *rating* in the 4-2-3-1 (80.04) formation. It can be said that a 4-3-3 formation is more effective in a match. The 4-3-3 formation has a very high intensity of attack and applies a faster pattern of play. The formation also provides many opportunities for forwards (FWL, FW, and FWR) to create assists and goals. In contrast to the 4-2-3-1 formation that applies a slightly defensive strategy with the two DM players he uses. In figure 2 also shown several players can be used in two different formations. The players are Alisson, Robertson, Philips, Milner, Williams, Jota, Mane and Firminho. Those players have excellent ratings in both line-ups.

4. Conclusion

BIP optimisation process in this study obtained optimum results for line-up formation 4-3-3 by producing selected variables worth 1 namely: $\{x_1, x_4, x_6, x_8, x_{11}, x_{18}, x_{21}, x_{28}, x_{34}, x_{37}, x_{39}\}$, with a maximum value of 82.47. As for the 4-2-3-1 formation, it produces selected variables that are valued at 1: $\{x_1, x_6, x_7, x_8, x_{11}, x_{14}, x_{16}, x_{29}, x_{31}, x_{32}, x_{42}\}$ with a maximum value of 80.04. These variables can be represented as player names and player positions that can be seen in Table 2 and Figure 2. The author may conclude also that 4-3-3 formations are more effectively applied than 4-2-3-1 formations. This can be seen from the comparison of the maximum *rating* values of the two formations. The 4-3-3 formation is more attacking (compared to the 4-2-3-1 formation) relying on the speed of the forwards with the support of more dominating midfielders.

References

- Barthels, H., Psarras, C., & Bientinesi, P. (2020, June). Automatic generation of efficient linear algebra programs. In *Proceedings of the Platform for Advanced Scientific Computing Conference* (pp. 1-11).
- Aquino, R., Vieira, L. H. P., Carling, C., Martins, G. H., Alves, I. S., & Puggina, E. F. (2017). Effects of competitive standard, team formation and playing position on match running performance of Brazilian professional soccer players. *International Journal of Performance Analysis in Sport*, 17(5), 695-705.
- Aquino, R., Machado, J. C., Manuel Clemente, F., Praça, G. M., Gonçalves, L. G. C., Melli-Neto, B., & Carling, C. (2019). Comparisons of ball possession, match running performance, player prominence and team network properties according to match outcome and playing formation during the 2018 FIFA World Cup. *International Journal of Performance Analysis in Sport*, 19(6), 1026-1037.
- Balouchzahi, N. M., Fathy, M., & Akbari, A. (2015). Optimal road side units placement model based on binary integer programming for efficient traffic information advertisement and discovery in vehicular environment. *IET Intelligent Transport Systems*, 9(9), 851-861.
- Bradley, P. S., Carling, C., Archer, D., Roberts, J., Dodds, A., Di Mascio, M., & Krstrup, P. (2011). The effect of playing formation on high-intensity running and technical profiles in English FA Premier League soccer matches. *Journal of sports sciences*, 29(8), 821-830.
- Brey, J. J., Carazo, A. F., & Brey, R. (2012). Using AHP and binary integer programming to optimize the initial distribution of hydrogen infrastructures in Andalusia. *International Journal of hydrogen energy*, 37(6), 5372-5384.
- Beal, R., Chalkiadakis, G., Norman, T. J., & Ramchurn, S. D. (2020). Optimising Game Tactics for Football. *arXiv preprint arXiv:2003.10294*.
- Carling, C. (2011). Influence of opposition team formation on physical and skill-related performance in a professional soccer team. *European Journal of Sport Science*, 11(3), 155-164.
- Corrêa, R., Cardoso Jr, G., de Araújo, O. C., & Mariotto, L. (2015). Online coordination of directional overcurrent relays using binary integer programming. *Electric Power Systems Research*, 127, 118-125.
- Dobson, S., & Goddard, J. (2010). Optimizing strategic behaviour in a dynamic setting in professional team sports. *European Journal of Operational Research*, 205(3), 661-669.
- Ge, T., An, Z., Cai, H., & Wang, Y. (2020, August). An Analysis on the Effectiveness of Cooperation in A Soccer Team. In *2020 15th International Conference on Computer Science & Education (ICCSE)* (pp. 787-794). IEEE.
- Gholamnejad, J., & Osanloo, M. (2007). Using chance constrained binary integer programming in optimising long term production scheduling for open pit mine design. *Mining Technology*, 116(2), 58-66.
- He, K., Liang, Z., Cui, T., Ke, Z., Liu, Z., Zhao, Q., & Fang, F. (2018, June). Formation optimization of RoboCup3D soccer robots using delaunay triangulation network. In *2018 Chinese Control And Decision Conference (CCDC)* (pp. 224-229). IEEE.
- Hillier, F. S., & Lieberman, G. J. (1995). Introduction to operations research. New York: McGraw-Hill Science, Engineering & Mathematics.
- Pan, J. C. H., & Chen, J. S. (2005). Mixed binary integer programming formulations for the reentrant job shop scheduling problem. *Computers & Operations Research*, 32(5), 1197-1212.

- Tamura, K., & Masuda, N. (2015). Win-stay lose-shift strategy in formation changes in football. *EPJ Data Science*, 4(1), 9-22.
- Venter, R. E., Opperman, E., & Opperman, S. (2011). The use of Global Positioning System (GPS) tracking devices to assess movement demands and impacts in Under-19 Rugby Union match play: sports technology. *African Journal for Physical Health Education, Recreation and Dance*, 17(1), 1-8.
- Wong, J. K. L., Mason, A. J., Neve, M. J., & Sowerby, K. W. (2006). Base station placement in indoor wireless systems using binary integer programming. *IEE Proceedings-Communications*, 153(5), 771-778.
- Ziaee, M., & Sadjadi, S. J. (2007). Mixed binary integer programming formulations for the flow shop scheduling problems. A case study: ISD projects scheduling. *Applied mathematics and computation*, 185(1), 218-228.