Jurnal Mandiri IT
Vol. 12 No. 2, October (2023), pp. 123-134

ISSN 2301-8984 (Print)

Data mining for Indonesian National Football Team

Luki Ardiantoro[1], **Ronny M. Akbar**[2], **Yanuarini Nur S** [3]
[1,2,3] Department of Informatics Engineering, Majapahit Islamic University, Mojokerto, Indonesia

A R T I C L E I N F O

Article history:

Received Oct 6, 2023
Revised Oct 12, 2023
Accepted Oct 13, 2023

Keywords:

Data Mining;
Football;
FP-Growth;
Playing Pattern.

ABSTRACT

The Indonesian national football team is controlled by the Indonesian Football Association, which is a member of FIFA and also a member of the AFC. The current coach of the Indonesian national team is Shin Tae-yong (STY), a former South Korean soccer player. In this study, the researcher intends to examine, study and obtain an overview of the game patterns, tactics and game strategies carried out by STY against the Indonesian national team. The goal to be achieved is gained of tactics and strategies for local coaches in the future. FP Growth algorithm was used for this research. Method to analyze match videos and record them in the form of statistical data in a set of data tupples (rows/records) for a particular match session was used. Then normalize the data, by forming a series of pattern numbers as a representation of the direction of attack in certain situation. The data set is formed with one set per round. As conclusion, data mining can be used to provide an overview of the national team's playing pattern. Thus, it makes it easier for coaches to determine the right strategy to beat opponents.

Corresponding Author:

Luki Ardiantoro,
Department Informatics Engineering,
Majapahit Islamic University,
Jl. Raya Jabon No.KM.0,7, Tambak Rejo, Gayaman, Mojokerto, Jawa Timur, 61364, Indonesia.
Email: ipan.ardianto@gmail.com

1. INTRODUCTION

In data generalization, there are two major topics that are concentrated on this subject : Data Mining and Statistics. This approach enables companies and organizations to predict and prepare reasonable future plans. One particular applications of prediction is in sport such as football (Che Mohd Rosli et al., 2018). Data mining is the process of analyzing data using various methods to produce useful information. Changes in player performance, age, and squad size can be analyzed using statistical methods. These digital data will provide information about the club's market value (Andry et al., 2023). Most importantly, though, analytics in sports, and football in particular, will keep developing further in the upcoming years. We will witness increasingly sophisticated technology, especially in real-time applications (Kröckel, 2019).

Tactical information in football can be obtained by applying analytics techniques from the fields of network science, machine learning and process mining. The main motivation behind this research is, on the one hand, the lack of studies that use professional tracking data in football, and on the other, the lack of studies investigating real-time decision support in football based on data analytics. As a theoretical basis for the analytics concept, the dynamic system theory is used. According to this theory football teams are dynamic systems, composed of elements (the players) who interact constantly with each other and their environment and who, by their dynamic interactions, form behavioral patterns over time. These patterns are due to the self-organization ability of the players, and thanks to this, they are able to reorganize themselves and regain a state of balance

following a perturbation occurrence (e.g., a counter attack). By following the principles of this theory, analytics techniques such as social network analysis, self-organizing maps, and process mining are applied on football event data (Kröckel, 2019). Analyzing the tactical behavior in team sports is of paramount importance in sports performance analysis. The individual actions performed are of interest when analyzing the team's tactics. For quite some time, action frequencies of teams, and players have been the only way to gain insight into this performance aspect. However, this is not enough to get a complete picture of the performance, and especially the tactical behavior. Therefore, action/event sequences have been suggested for deeper insight into the game (Kröckel & Bodendorf, 2020).

By using computer technology to find information as well as knowledge in ever-increasing amounts of data has become feasible, inexpensive, and crucial. These factors necessitate the development of innovative approaches to convert large quantities of target data into meaningful information as well as knowledge in a sufficient length of time (Khajuria et al., 2023), It is nowadays relatively easy to calculate such indicators not only pre- and post-match, but during live matches as well. Not only coaches and their staff, but also football fans have access to basic statistics about their favorite team and player as the game develops. However, the field of performance analysis in football keeps developing, and latest research has called for more indepth analysis of the game that goes beyond simple statistics and indicators like shots on goal, tackles, distance run etc. New methods have been suggested in the literature that look not only at the team's and player's actions but also their interactions that are bounded by environment factors like the pitch dimensions, location (home or away), opponent's strength, position on the field, and current score. These interactions change and adapt constantly depending on the mentioned factors. The theoretical basis for this new direction in performance analysis research is inspired by the dy- namic system theory By using this option, it is possible to not only visualize the sequences leading to shot on goal for certain (Indonesia's) team but also to find out which players mostly started or ended a sequence. These analyses can be very useful in assessing the value of a player in a game (Kröckel & Bodendorf, 2020). Data mining is the process of looking for patterns or interesting information based on the results of analysis of a selected data population using certain algorithmic techniques or methods (Bompa & Haff, 2019).The use of data mining has developed in various fields of life, especially in business to assist in solving decision-making problems.

On 28 December 2019, the Indonesian Football Association (PSSI) confirmed the appointment of Shin as Indonesia's new coach, replacing Simon McMenemy. STY was born on May 11, 1970, in Yeongdeok, Gyeongbuk, South Korea. STY is a former South Korean football player who played in the midfield position. STY is contracted by PSSI for 4 years until 31 December 2023. STY is known as a coach who is detailed, careful in data, thorough in preparation, and has very varied tactics and game strategies. This research aims to study and obtain an overview of the playing patterns, tactics and game strategies used by STY for the Indonesian football national team. STY is also known as a coach who is difficult for opponents to predict. So under STY, the Indonesian national team has performed very well. Proven capable of appearing in the U-20 Asian Cup (after being absent for 15 years) as well as appearing in the senior level Asian Cup. Apart from that, the U-20 National Team will also appear in the U-20 World Cup. FIFA rankings, the Indonesian national team also soared from a country ranked 175th to 150th in the world, as well.So far, football development in Indonesia has not been sustainable and well planned. Every time a national team manager changes, a coach is replaced, and the national team's playing pattern also changes. So it is not surprising that the Indonesian national team has not achieved any major achievements. Athletics, in particular football, put significant demands on the athlete due to the large number of competitions and excessive physical and psycho-emotional stress. They demonstrate reduced motivation and, accordingly, the lack of interest and motivation provokes a feeling of fatigue. It has been empirically found that the respondent football players with a pronounced feeling of fatigue demonstrate a high level of desire to stop their training and competition activities (Shcherbak et al., 2023). If indeed physicality is a determining factor in achieving achievement in national football team, then STY's role as a coach in this case can be said to be successful. Therefore, more in-depth research is needed regarding the game patterns, tactics and game strategies carried out by STY

against the Indonesian national football team. It is hoped that with this research, Indonesian football will not lose momentum to achieve better achievements in the future, after STY era.

The objective of the research is to model the playing activities of the national team on the football field, the data is presented in quantitative form, then a simulation is carried out on the data using a data mining algorithm, to identify the dominant patterns of tactics during the matches carried out by the national team. It is hoped, Indonesian football will not lose momentum to achieve better achievements in the future. So far, football development in Indonesia has not been sustainable and well planned. Every time the manager changes, the coach changes, and the playing pattern also changes. So it is not surprising that the Indonesian national team has not achieved any achievements. Apart from that, the knowledge gained can be useful for adding to the repertoire of tactics and strategies for local coaches.

In previous research, the author conducted research on the application of the FP Growth algorithm to observe the playing patterns of Indonesian badminton players, Jonathan Christe (Ardiantoro et al., 2019). In this research FP-Growth algorithm was used also, but the research object was badminton players in individual events. Whereas in this research, the object of research is football, with observations of team play patterns. Research by (Che Mohd Rosli et al., 2018), focusing to find the most accurate data mining technique that fits the nature of football data. The techniques tested are Decision Trees, Neural Networks, Bayesian Network, and k -Nearest Neighbors. Other research has been carried out by Mochamad Ali Akbar, et.al (2019) who conducted research on the position of football players, using the k-means clustering algorithm data mining using the WEKA simulator (Akbar et al., 2019). The difference between this research and this research is the algorithm and research objectives. The clustering algorithm aims to group players according to the same characteristics. Meanwhile, in this research, the author aims to find the dominant pattern of a team's game, by modeling data that is suitable for pattern analysis needs, the FP Growth algorithm. The simulation tool also different, it was Tanagra compared to WEKA (which was used in previous research).

Research by Ladha & Deepa, 2011 shows general scheme of feature selection. Feature selection is one of the key issues in pattern recognition. A feature selection algorithm (FSA) is a computational solution that is motivated by a certain definition of relevance. The purpose of a FSA is to identify relevant features according to a definition of relevance(Ladha & Deepa, 2011).The quality of the feature selection has a direct impact on the classification accuracy and generalization performance of the classifier. In order to reduce the size of the feature subset and improve the efficiency of the algorithm without reducing the accuracy. On that paper they proposes a feature selection algorithm based on association rules, known as ARFS (association rule feature selection). The algorithm uses association rules to mine the frequent 2-items set of the feature attributes and category in the dataset. Then the algorithm sorts the features according to the confidence of the frequent 2-items set, and then combines the sequential forward selection method, and uses the classification performance of the decision tree classifier as the evaluation criteria of the feature subsets. This algorithm is simillar to the one discussed in this research, in this paper it applies to the field of soccer from Indonesian nasional team.

The next research, by Yonata Laia, et.al (2019) involved applying data mining to predict Champions League winners using the C4.5 algorithm (Tandian et al., 2019).Meanwhile, in this research, data mining analysis was carried out by modeling the national team game to be simulated using a data mining algorithm. This analysis is to identify frequently used game patterns based on ball distribution during the match. The data obtained was then simulated using the FP-Tree and FP-Growth algorithms to determine the dominant pattern of a team's ball distribution using TANAGRA. In this way, the national team's weaknesses and strengths can be identified when anticipating the situation on the field.

2. RESEARCH METHOD

2.1. Algorithm

Data mining is the process of looking for patterns or interesting information based on the results of analysis of a selected data population using certain algorithmic techniques or methods (Bompa & Haff, 2019).Data Mining is a term used to describe the discovery of knowledge in a

database, with the algorithms being: prediction, estimation, clustering, classification and association algorithms (Nurasiah, 2021). Data Mining in the process uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various large databases (Tandian et al., 2019).The Apriori algorithm was introduced by Agrawal and Srikant in 1994. Until now, this algorithm is an association algorithm that is widely used and developed by researchers(Tahir & Sitompul, 2021). The Apriori association rule algorithm is a data mining technique for finding association rules between a combination of items. These hidden relationships can be represented in the form of association rules for a set of items that frequently appear (frequent itemset) as shown in Fig 1.(Nurasiah, 2021)

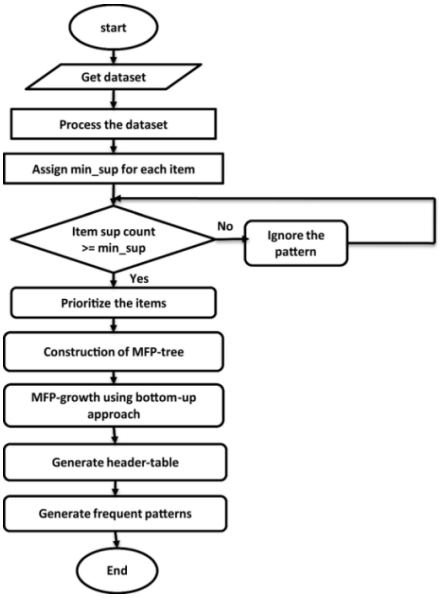

Figure 1. Flow-chart to Generate FP-Tree & FP-Growth Algorithms

Due to its practical importance, since theoritically become popular in (J. Han et al., 2011) frequent itemset mining has received extensive attentions and many algorithms are proposed. The existing frequent itemset mining algorithms can be classified into two groups: candidate-generation-based algorithms and pattern-growth-based algorithms (X. Han et al., 2019). The FP-Growth algorithm has stages that must be passed in order to provide maximum results, these stages are: 1. Conditional pattern base generation stage. 2. ConditionalFP-Tree generation stage. 3. Frequent itemset search stage. In the a priori algorithm there are two parameters used to measure associative rules, namely, Support (supporting value), namely the percentage of combinations of items in the database and Confidence (certainty value), namely the strength of the relationship between items in the associative rule (Larose & Larose, 2014)as in formula (i) below(Yudha et al., 2021)(Kamber & Han, 2018):

$$Support(A) = \frac{Number\ of\ Transactions\ contains\ A}{Total\ Transactions} \quad (1)$$

Confidence (c) is a measure of how often item A appears in transactions containing item A'. Confidence is calculated after support is determined. The Confidence value of the A-> B rule is obtained from the following formula (ii)(J. Han et al., 2011):

$$Confidence\ P(B|A) = \frac{Number\ of\ Transactions\ contains\ A \text{ and } B}{Number\ of\ Transactions\ \ contains\ A} \quad (2)$$

The a priori algorithm for handling association problems is to reduce the number of itemsets considered. Users set minimum support. So the value (A C) ≤ minimum support for each association rule A → C or (A → C) ≤ min- support(Syaifuddin et al., 2020). The first a priori generates all sets of

items that satisfy support. This set of items is often called a set of items (frequent item sets)(Ardiantoro & Sunarmi, 2020). Frequent item set generation is presented in Figure 2. The next stage is to collect all itemsets that meet the support ≥ minsupport requirements. These itemsets are called frequent itemsets. The next stage is rule generation, which aims to form rules with high confidence values from the frequent itemsets that have been obtained previously. These rules are called strongrules (Ardiantoro et al., 2019). Frequent itemset mining is an important operation to return all itemsets in the transaction table, which occur as a subset of at least a specified fraction of the transactions (X. Han et al., 2019).

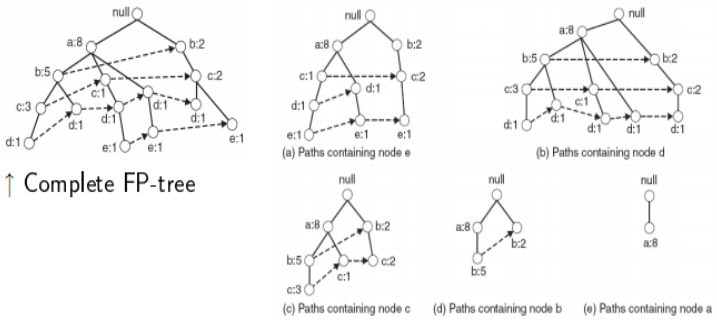

Figure 2. Frequent Item Set Generation Model

2.2. Data Model

In a soccer game, a team consists of 11 players, consisting of: 1 goalkeeper, 2-5 defenders, 2-5 midfielders, and 1-3 attackers. The goalkeeper (goalkeeper) is the only player on a team who can use his hands to block the ball from attacking the opposing team towards the goal. Generally, goalkeepers wear clothes (jersey) that are different from other players. The defender (back) has the main task of blocking and closing the opposing team's movements. Midfielders (midfielders/playmakers) usually consist of attacking midfielders who play close to the attackers and defensive midfielders who play close to the defenders. The main task of the midfielder is to set the tempo of the game and support the role of defenders and strikers in the team. There are also wingers who work on the right or left side of the field. The attacker (striker) has the main task of scoring goals against the opponent.In order to make it easier to compute, the data is modeled by dividing the football field playing areas into those in Figure 3.

Next, each area will be given a number notation according to the sequence presented in Figure 4. This number sequence is in accordance with that used by the famous Spanish coach, Pep Guardiola (Akbar et al., 2019). Based on the division of the field area, a data set of national football team games during defense and attack was then developed, using the number of the field area used. In the rightmost column, the notation called Item is used, this is intended to be in accordance with terms known to FP-Growth theory in general.

Next, using a data mining algorithm (FP-Growth) analysis was carried out on the Indonesian national football team game data set. The innovation carried out is the use of Data Mining for the scouting process for the coaching team to analyze the overall team performance. Data analysis is carried out according to the association rule algorithm which includes.

Figure 3. Division of the Football Field

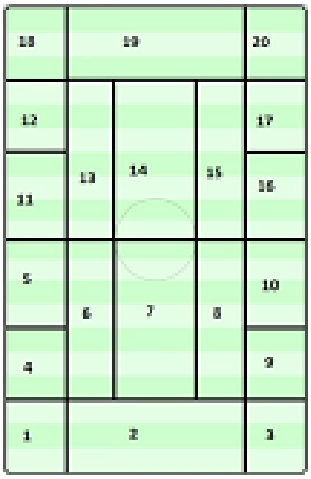

Figure 4. Numbering of Football Field Fields

2.3. Data Simulation

The observation object used is by observing video files of Indonesian national team matches in the STY coach era, in the 2020 and 2022 AFF cup tournaments, 2022 U-20 Asian Cup Qualification, 2022 Asian Cup Qualification, 2022 U-20 Asian Cup. Based on observations In this video, match data and statistics will be collected, especially in important phases of the match. The match phases include the defensive phase, transition phase and attack phase. Observations in each phase were made of the movement patterns and positions of the players, the role of each player, and the ball distribution patterns. This data is used as an illustration of the tactical patterns used by STY coaches, according to the opponents they face. This is because the tactics and strategy patterns applied are very dynamic. Each pattern will then be recorded in dBase, and then analyzed using Data Mining (Kamber & Han, 2018).

Table 1. Indonesian National Team Game Data Table (15 min) Round I

No	Menit	Area	Status
1	0,51	8, 7, 6, 11	Counter Fail
2	1,17	5, 7, 8, 7	Recovery Fail
3	1,19	8, 7, 10, 8, 15	Counter Fail
4	1,51	6, 4	Recovery Fail
5	1,59	2, 9, 3, 8, 7, 4, 6, 5, 6, 7, 10	Build up
6	2,38	7, 6, 5, 17, 16	Counter Fail
7	3,05	3, 9, 8, 7	Build up
8	3,21	2, 1	Build up
9	4,09	2, 6	Build up
10	4,52	7, 9, 7, 4, 5	Recovery Fail
11	5,27	7, 8, 9	Recovery Fail
12	6,06	8, 10	Recovery Fail
13	6,28	2, 1, 6, 7, 11	Build up
	...		
	dst		

The algorithm used to solve the problem is a pattern recognition algorithm. In this case, suitable data analysis approaches are the frequent item set algorithm and FP-Growth. Considering that there is no need for data analysis in real time, while the match is taking place, this research did not use cloud computing. This also takes into account security and the size of the data base which continues to grow. The results of the analysis are used for feedback, as well as technical discussions by the coaching team after the match, as well as analysis for the next match.

3. RESULTS AND DISCUSSIONS

In the first 15 minutes, it was clear that areas 2, 7 and 8 were the dominant areas of the national team's play. The data from Table 1 is then normalized, so that data is obtained as in Table 2. Normalization is carried out, so that the data makes it easier to analyze using Tanagra.Based on Table 2, it can be seen that the activities that often occur are the buildup attack and recovery failure phases, so based on this data it can be seen that the national team had great difficulty breaking through the opponent's playing area in the first 15 minutes.

Table 2. Table of Normalization of National Team Game Data

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0
4	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
7	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
13	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0
14	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
16	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
18	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
19	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
20	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0
21	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
23	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	1	1	0	1	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0
Jmlh	5	10	7	7	6	9	14	14	8	6	2	1	2	3	3	2	1	0	0

From table 2, dominant playing area are in national team own gaming area, and does not even produce a shot towards goal. This result relate with simulation with Tanagra, obtained 5 dominant rules consisting of: 9∩7 as much as 25%, 3∩8 as much as 25%, 6∩7 as much as 29.2%, 2∩8 as much as 25%, and 8∩7 as much as 45, 8%, so there is no attacks leading towards the opponent's goal.

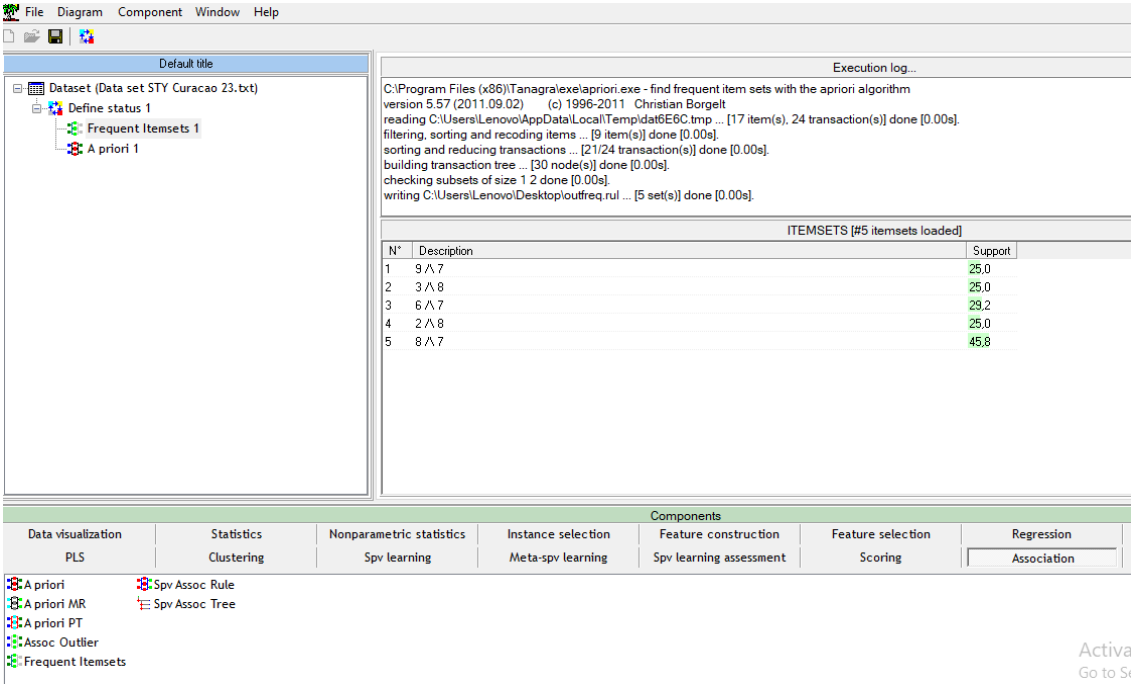

Figure 5. Frequent Itemset & Rules Result Using TANAGRA Simulation

Refer to Figure 5, after simulating with Tanagra, we obtained 5 dominant rules consisting of: 9∩7 as much as 25%, 3∩8 as much as 25%, 6∩7 as much as 29.2%, 2∩8 as much as 25% %, and 8∩7 as much as 45.8%. This means that mostly ball playing are directed towards the midfield (areas 7 and 8), but there is a deadlock in ball distribution during the game. It can even be seen that areas 8 and 7 have a very high correlation (ie 45.8%), if the ball is in area 8 then 45.8% is directed to area 7. Likewise what happens if the ball comes from 2 (goalkeeper area) and 3 (right back), 25% directed to area 8 (midfield/midfield), so it can be seen here that during the 15 minutes of play the national team very often used short passing patterns for ball distribution, between the lines. However, the distribution of the ball cannot flow to the front of the opponent's goal (area 19). During the first 15 minutes, the national team failed to attack the opponent's area, instead blocking the opponent's attacks in the middle area. On the other hand, this pattern of play always fails to enter the opponent's. Figure 5 show that that the dominant areas that appear most often are 9∩7, 3∩8, 6∩7, 2∩8, and 8∩7. So the dominant area during the first 15 minutes was the build up attack pattern from the center of defense, through the right wing back, flowing into the midfield. However, he then experienced difficulties in forward passing progression, due to position, one-on-one pressure from the opponent, and poor pass accuracy.

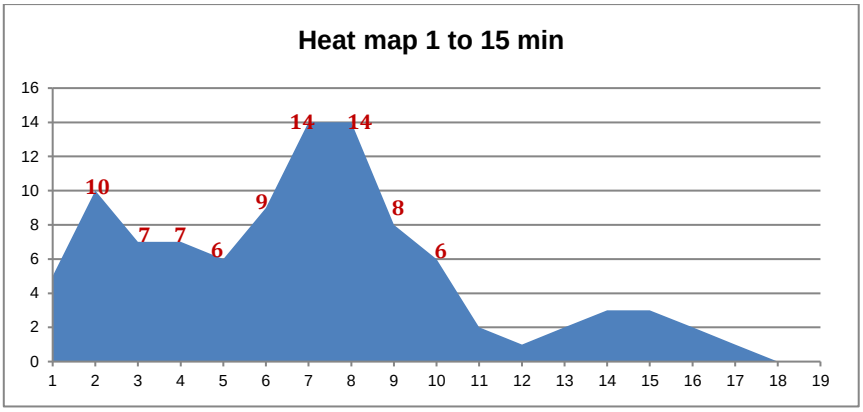

Figure 6. Ball Distribution (1 - 15 min)

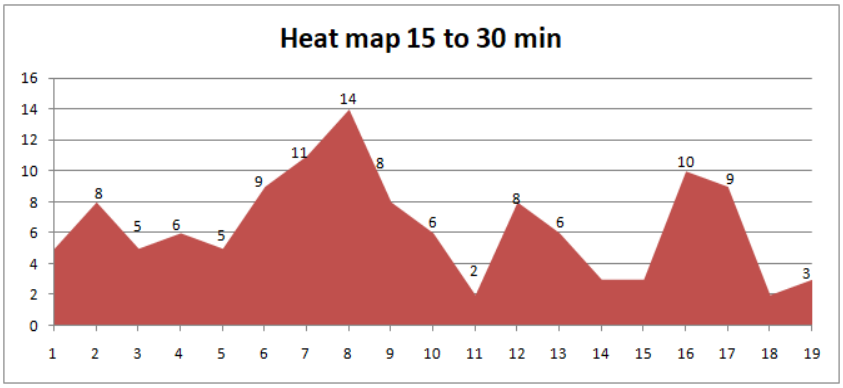

Figure 7. Ball Distribution (15 - 30 min)

Based on the results of data processing presented in Figures 6 and 7, it was found that there was a change in the ball distribution approach carried out by the STY coach. In the first 15 minutes, STY ordered to close the midfield area, building attacks from the left wing, but with very minimal attack results. Furthermore, in the second 15 minute phase, STY made changes by continuing to close the central area, but passing the ball to the right wing and area number 13. This proved capable of breaking into the opponent's defense area more effectively.

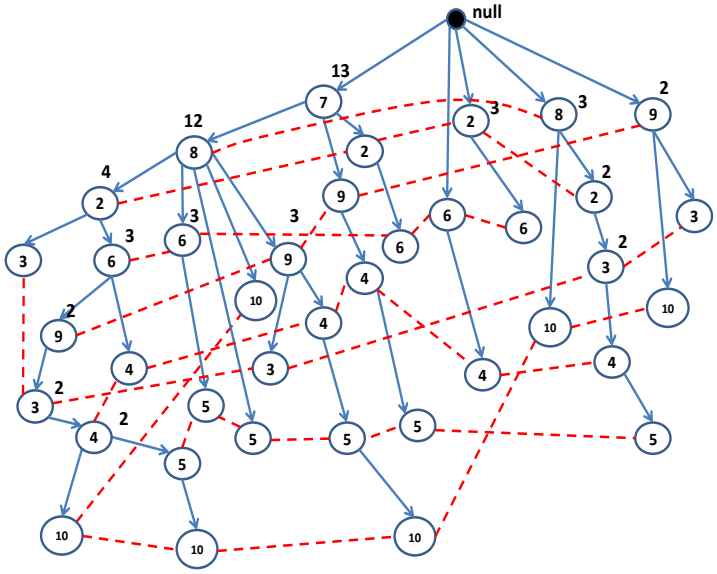

Figure 8.FP-Tree National Team Game Pattern (1-15 min)

Apart from that, based on the same method, the results obtained were that STY made (changed) data for the second 15 minutes (minutes 16-30) to see how dynamic a STY was as a coach ordering his team to change game tactics during the match. By adjusting to the opposing team's performance, a coach plays a role in anticipating, so that the national team's playing pattern is not easily read by the opposing team. As well as responding to the opponent's playing patterns, in order to win the match (the final result of the match was 1-0 for the national team to win). This is reflected in Figure 6 and Figure 7. Furthermore, by using the data pattern from the simulation results, processed to be presented in the FP-Tree and FP-Growth that occurred, results were obtained as in Figure 7. Based on the data in Figure 7, it can be seen that the area Important for the Indonesian national team game are areas number 7 and 8. Area 7 is the center of the game with 13 events, followed by area 8 with 12 events. Therefore, if a team (opponent) will play against Indonesia's national team under the coach STY, one of the efforts made is to focus the game on areas number 7 and 8 of the Indonesian national team. Apart from the formation used by the national team, as well as the rotation of players during the match, this is where the strengths and weaknesses of the Indonesian men's national football team play. This happened in the first 15 minutes. This playing pattern still has to be compared with the national team's playing pattern when facing different opponents.

The research model produced in this research requires large amounts of computing and data input, so analysis can only be carried out after the match takes place. In addition, considering the very large amount of data produced, manual processes should be replaced with electronic devices (or movement sensors) that support data collection during the match. If this can be done, more useful results will be obtained as evaluation material before the match ends. Apart from that, it is necessary to develop a more integrated research model involving non-technical factors of players. This is because the evaluation of overall tactics in the game of football is very complex, so it is necessary to take into account the factors of stamina, player rotation, individual skills, competitive mentality, individual intelligence, player instincts, etc.

4. CONCLUSION

Based on the results of data analysis and discussion in the match above, it can be concluded that the association algorithm (FP-Tree, FP-Growth) is able to provide an overview of the playing pattern of a football team. As a national team coach, STY always changes the way of playing according to game conditions and situations. Future research development suggestions focus on developing cloud-based and real-time applications for more mobile and responsive soccer match analysis. By

utilizing association algorithms such as FP-Tree and FP-Growth, this research can provide football team coaches with more accurate and quick insights into their team's playing patterns. It requires real-time data integration, fast analysis, customizability, and efficient team collaboration, as well as special attention to data security. This research will pave the way for more advanced use of technology in the world of soccer and help national teams make smarter decisions in matches.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Research, Technology and Higher Education, because this research was carried out in the framework of the Beginner Lecturer Research Grant (PDP) for the 2023 fiscal year.

REFERENCES

Akbar, M. A., Fatimah, F., & Jaenudin, J. (2019). Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering. *Semnati 2019*, *x*, *No.x*(x), 1278–1282.

Andry, J. F., Riama, S., & Yefta, V. N. (2023). Analysis of Big Data Football Club Market Value Using K-Means and Linear Regression Mining Methods. *Journal of Computer Science*, *19*(2), 286–294. https://doi.org/10.3844/JCSSP.2023.286.294

Ardiantoro, L., & Sunarmi, N. (2020). Badminton player scouting analysis using Frequent Pattern growth (FP-growth) algorithm. *Journal of Physics: Conference Series*, *1456*(1). https://doi.org/10.1088/1742-6596/1456/1/012023

Ardiantoro, L., Sunarmi, N., & Z, S. (2019). Pemanfaatan Knowledge Data Discovery(KDD) Pada Pola Permainan Atlet Bulutangkis. *Explore IT : Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, *11*(1), 1–6. https://doi.org/10.35891/explorit.v11i1.1467

Bompa, T. O., & Haff, G. G. (2019). Periodization: Theory and Methodology of Training, 6th Edition. In *Medicine & Science in Sports & Exercise* (Vol. 51, Issue 4). https://doi.org/10.1249/01.mss.0000554581.71065.23

Che Mohd Rosli, C. M. F., Saringat, M. Z., Razali, N., & Mustapha, A. (2018). A Comparative Study of Data Mining Techniques on Football Match Prediction. *Journal of Physics: Conference Series*, *1020*(1). https://doi.org/10.1088/1742-6596/1020/1/012003

Han, J., Kamber, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

Han, X., Liu, X., Chen, J., Lai, G., Gao, H., & Li, J. (2019). Efficiently Mining Frequent Itemsets on Massive Data. *IEEE Access*, *7*(c), 31409–31421. https://doi.org/10.1109/ACCESS.2019.2902602

Kamber, M., & Han, J. (2018). Data Mining: Concepts and Techniques : Concepts and Techniques. In *The Fundamentals of Political Science Research*.

Khajuria, R., Sharma, A., Sharma, S., Sharma, A., Baliya, J. N., & Singh, P. (2023). Performance analysis of frequent pattern mining algorithm on different real-life dataset. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(3), 1355–1363. https://doi.org/10.11591/ijeecs.v29.i3.pp1355-1363

Kröckel, P. (2019). *Big Data Event Analytics in Football for Tactical Decision Support* [der Friedrich-Alexander-Universität]. https://opus4.kobv.de/opus4-fau/frontdoor/deliver/index/docId/12365/file/Kröckel+Diss.pdf

Kröckel, P., & Bodendorf, F. (2020). Process Mining of Football Event Data: A Novel Approach for Tactical Insights Into the Game. *Frontiers in Artificial Intelligence*, *3*(July), 1–16. https://doi.org/10.3389/frai.2020.00047

Ladha, L., & Deepa, T. (2011). Feature Selection Methods and Algorithms. *International Journal on Computer Science and Engineering (IJCSE)*, *3*(5), 1787–1797. https://doi.org/10.1016/s0007-1226(78)90078-4

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data. In *Discovering Knowledge in Data*. https://doi.org/10.1002/9781118874059

Nurasiah. (2021). Implementasi Algoritma FP-Growth Pada Pengenalan Pola Penjualan. *TIN : Terapan Informatika Nusantara*, *1*(9), 438–444.

Shcherbak, T., Popovych, I., Kariyev, A., Duisenbayeva, A., Huzar, V., Hoian, I., & Kyrychenko, K. (2023). Psychological causes of fatigue in football players. *Journal of Physical Education and Sport*, *23*(8), 2193–2202. https://doi.org/10.7752/jpes.2023.08251

Syaifuddin, A., Harianto, R. A., & Santoso, J. (2020). Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA. *INSYST: Journal of Intelligent System and Computation*, *2*(1), 12–19. https://doi.org/10.52985/INSYST.V2I1.150

Tahir, M., & Sitompul, N. (2021). Penerapan Algoritma Fp-Growth Dalam Menentukan Kecenderungan Mahasiswa Mengambil Mata Kuliah Pilihan. *Network Engineering Research Operation*, *6*(1), 59. https://doi.org/10.21107/nero.v6i1.216

Tandian, C., Laia, Y., & Saputra, A. (2019). Penerapan Data Mining Dalam Memprediksi Pemenang Klub Sepak Bola Pada Ajang Liga Champion Dengan Algoritma C.45. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *2*(2), 91–98. https://doi.org/10.34012/jusikom.v2i2.397

Yudha, Y. N., Winarso, D., & syahril. (2021). Analisis Sentimen Masyarakat Pada Twiter Terhadap Isu Covid-19 Menggunakan Metode Lexicon Based. *JURNAL FASILKOM (Teknologi InFormASi Dan ILmu KOMputer)*, *11*(2), 97–103. https://doi.org/10.37859/JF.V11I2.2772

Yudho Ardianto, M., & Adinugroho, S. (2021). Penentuan Tata Letak Produk menggunakan Algoritma FP-Growth pada Toko ATK. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(9), 3826–3832. http://j-ptiik.ub.ac.id