

Tactical Analysis of Indonesian National Football Team using FP-Tree Algorithm

Luki Ardiantoro ¹, Ronny M. Akbar ², Yanuarini Nur S. ³

^{1,2,3} Department of Informatics, Majapahit Islamic University, Mojokerto, East Java

*Correspondence Email: ipan.ardiantogmail.com

ABSTRACT

Football is the most popular sport in the world, including Indonesia. 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 study and obtain an overview of the game patterns, tactics and game strategies carried out by STY for the Indonesian national team. The goal to be achieved is gained of tactics and strategies for local coaches in the future. Frequent Pattern Tree (FP-Tree) 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 modelling (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. As conclusion, the national team's playing pattern, including team formation, how to defend, how to attack. Thus, it makes it easier for coaches to determine the right strategy to beat opponents, when dealing in official matches.

KEYWORDS

Tactical, FP Growth, playing pattern, STY



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

Football (soccer) is one of the most popular sports on earth today. In a soccer match, to achieve victory, various tactics, strategies and formations of good players are developed in a duration of 2x45 minutes (normal). Therefore, insight and tactics are needed from a soccer coach to apply tactics that are suitable for the opponents they face. The Indonesian national football team is the national team that represents Indonesia in men's senior international football. The 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). STY was born on May 11, 1970, in Yeongdeok, Gyeongbuk, South Korea. STY is a former South Korean soccer player who played as a midfielder.

The use of technology in sports (sport science) has grown rapidly. The process of scouting, talent scouting and analysis of match processes against opposing players is

carried out using innovative methods, in order to boost the athlete's achievements. The scouting process itself can be done before, during and after the game. A series of statistical analyzes were carried out, thus providing an accurate approach in understanding the strengths and weaknesses of the opponent. The development of each player is monitored with accurate data, and formulated in effective training methods to boost performance.

Data mining is the process of looking for patterns or interesting information based on the results of analysis in a selected data population using certain algorithmic techniques or methods [1]. Data Mining is a term used to describe the discovery of knowledge in databases [2]. Data Mining in its process uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases [3]. 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 [4].

On December 28, 2019, the Indonesian Football Association (PSSI) confirmed Shin's appointment as Indonesia's new coach, replacing Simon McMenemy. STY is contracted by PSSI for 4 years until December 31, 2023. STY is known as a coach who is detailed, meticulous in preparation, and has very varied game tactics and strategies. 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 problem formulation in the research are: a. How to develop an analytical model for soccer players based on data collected from observing video files, by utilizing information technology ? b. How to use the FP-Growth algorithm to map the playing pattern of a soccer athlete?

The aims of this research are: a). Providing innovation to the coaching model that is applied to soccer athletes based on their needs, based on empirical data. b). Developing an analytical model for soccer players by utilizing information technology as part of development in implementing sport science in Indonesia. c). Improving the achievements of the Indonesian national football team, with more varied tactics and strategies, attacking, and keeping up with the times.

The benefits to be achieved are: a). It is hoped that this research will be developed as a role model for the scouting process for football athletes in the area. b). Increase the competitiveness of the athletes, so they can reach a higher level of play. So if this is applied to national athletes, it is hoped that it can play a role in increasing Indonesia's competitiveness in world football competition. c). become input and tools to develop talent as well as training programs for athletes that are more efficient.

RESEARCH METHOD

In this study, the authors developed a model for data analysis based on match video files. By utilizing one of the data mining algorithms (FP Growth), an analysis was carried out on PSSI national team, under coach STY. In the first stage, data collection was carried out by analyzing match videos and recording them in the form of statistical data in a set of data tuples (rows/records) for a particular match session. Then normalize the data, by forming a series of pattern numbers as a representation of the direction of attack from the team as a whole. The data set is formed with one set per round. This data is then collected in one round, to obtain a data set in one round. The second stage is processing the main data using a data mining algorithm, namely: FP-Growth. This is for main data analysis according to the needs of the algorithm and association rule analysis which includes searching for patterns from data that often appears (frequent itemset mining), as well as determining the best rules (rules) using the calculation of the final value. The results of the

analysis of the data obtained will provide an overview of the data patterns formed and expressed as sequences and probabilities. The third stage is interviews, feedback and improvement of the analysis process, for further adjustments if necessary. The results of the analysis are carried out by drawing conclusions and recommendations.

FP tree is one of the algorithms used in data mining. Data mining is the process of looking for patterns or interesting information based on the results of analysis in a selected data population using certain algorithmic techniques or methods [1]. The Apriori association rule algorithm or association analysis is a data mining technique for finding association rules between a combination of items. This hidden relation can be represented in the form of association rules for frequent itemsets. There are two parameters used to measure associative rules, namely support (supporting value), namely the percentage of combinations of these items in the database and confidence (certainty value), namely the strength of the relationship between items in associative rules [8].

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

Confidance (c) is a measure of how often item A appears in transactions that contain item A'. Confidance is calculated after support is determined. The confidence value of the A-> B rule is obtained from the following formula:

$$\text{Confidence} = P(B | A) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}}$$

The Apriori algorithm in dealing with association problems is to reduce the number of itemsets considered. The user sets the minimum support. So the value of (AC) ≤ minimum support for each association rule A → C or (A → C) ≤ min-support. The first a priori returns all item sets that satisfy support. This set of items is often called a set of items (frequent item sets) [8][9]. Frequent item set generation is presented in Figure 1. The first stage is to collect all itemsets that meet the support ≥ minsup requirements. These itemsets are called frequent itemsets. The second 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 strong rules[8][9].

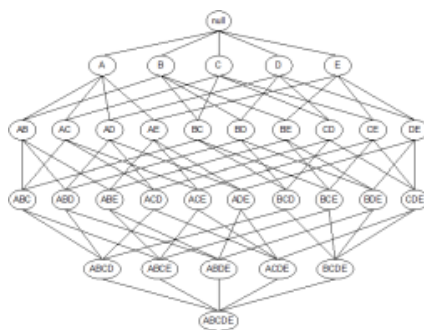


Figure 1: *Frequent Item Set*

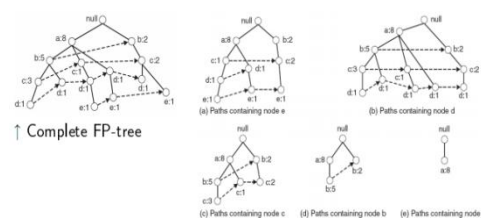


Figure 2 : *Frequent Item Set Generation*

The Frequent Pattern Growth (FP-Growth) algorithm is a development of the Apriori algorithm. So that the shortcomings of the Apriori algorithm are corrected by the FP Growth algorithm [3]. FP-Growth is an alternative algorithm that can be used to determine the most frequently occurring data set (frequent itemset) in a data set [10].

The data modeling used is to divide the game space into areas numbered as follows (Figure 3).



Figure.3 : Data Modelling Playing Pattern

RESULT AND DISCUSSION

Data observation in this research, was football match between Indonesia vs Curacao, the first leg FIFA Matchday held on 27 Sept 2022, in Senayan Jakarta. This data captured during first 15 minutes is presented in Table 1.

Table 1: Data-set Indonesia vs Curacao

No	Menit	Posisi	Item	Status	Situasi
1	0,51	Menyerang	8, 7, 6, 11	Counter Fail	merebut bola, masuk ke area lawan gagal
2	1,17	Menyerang	5, 7, 8, 7	Recovery Fail	merebut bola, masih di area sendiri gagal
3	1,19	Menyerang	8, 7, 10, 8, 15	Counter Fail	merebut bola, masuk ke area lawan gagal
4	1,51	Menyerang	6, 4	Recovery Fail	merebut bola, masih di area sendiri gagal
5	1,59	Menyerang	2, 9, 3, 8, 7, 4, 6, 5, 6, 7, 10	Build up	membangun serangan dari bawah
6	2,38	Menyerang	7, 6, 5, 17, 16	Counter Fail	merebut bola, masuk ke area lawan gagal
7	3,05	Menyerang	3, 9, 8, 7	Build up	membangun serangan dari bawah
8	3,21	Menyerang	2, 1	Build up	membangun serangan dari bawah
9	4,09	Menyerang	2, 6	Build up	membangun serangan dari bawah
10	4,52	Menyerang	7, 9, 7, 4, 5	Recovery Fail	merebut bola, masih di area sendiri gagal
11	5,27	Menyerang	7, 8, 9	Recovery Fail	merebut bola, masih di area sendiri gagal
12	6,06	Menyerang	8, 10	Recovery Fail	merebut bola, masih di area sendiri gagal
13	6,38	Menyerang	2, 1, 6, 7, 11	Build up	membangun serangan dari bawah
14	7,3	Menyerang	2, 1, 2, 3, 8	Build up	membangun serangan dari bawah
15	8,01	Menyerang	16, 15, 13, 14	set piece	lemparan ke dalam
16	8,41	Menyerang	5, 4, 7, 9, 8, 7, 9, 10	set piece	lemparan ke dalam
17	9,19	Menyerang	10, 9	set piece	lemparan ke dalam
18	9,57	Menyerang	2, 3, 2, 8, 4, 6, 7, 9, 10	Build up	membangun serangan dari bawah
19	10,58	Menyerang	3, 9	Recovery Fail	merebut bola, masih di area sendiri gagal
20	11,39	Menyerang	2, 3, 8, 2, 4, 5, 13, 14	Build up	membangun serangan dari bawah
21	13,08	Menyerang	2, 3, 8, 7	Build up	membangun serangan dari bawah
22	13,25	Menyerang	8, 7, 6	Recovery Fail	merebut bola, masih di area sendiri gagal
23	13,48	Menyerang	2, 1	Build up	membangun serangan dari bawah
24	14,07	Menyerang	1, 2, 6, 4, 7, 8, 15, 7, 14	Build up	membangun serangan dari bawah

Table 2: Normalized Data-set

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

The data is then normalized, to make it easier to process using Tanagra (Table 2). Based on the normalized data, Fig 5 shows a heat map of the national team's game pattern based on the division of the field area (see Fig 3). Then, by using the support parameter of 6, the results are obtained as shown in Figures 5 and 6.

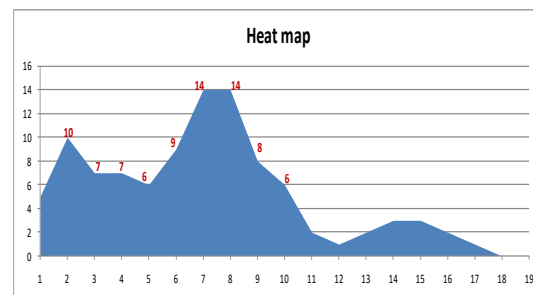


Figure.5 : Data set Heat Map Playing Pattern

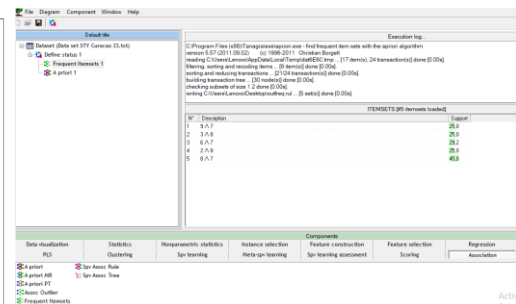


Figure.6 : Frequent Itemset simulation by Tanagra

After obtaining the data modeling as mentioned above, further analysis is carried out with the FP-Growth algorithm, which can be divided into 3 main stages as follows [8][10]:

1. The conditional pattern base generation stage,
2. FP-Tree conditional generation stage,

Fig 7 shows a complete picture of the national team's playing pattern for the first 15 minutes, it can be seen that the dominant sectors are areas 7 and 8 (midfield area). However, when building up an attack, the national team tends to start from the right-back sector, aiming for the left half space. Obviously, at those 15 minutes, the national team still had difficulty penetrating the opposing team's defense. This data gives an illustration of the tactic pattern used by the STY coach, according to the opponent he is facing. This is considering that the patterns of tactics and strategies that are applied are very dynamic, at least in first 15 minutes of the game against Curacao.

Table 3:Rearranged Itemset after simulation

No	Menit	Item	F. Itemset	Status
1	0,51	8, 7, 6, 11	7,8,6	Counter Fail
2	1,17	5, 7, 8, 7	7,8,5	Recovery Fail
3	1,19	8, 7, 10, 8, 15	7,8,10	Counter Fail
4	1,51	6, 4	6,4	Recovery Fail
5	1,59	2, 9, 3, 8, 7, 4, 6, 5, 6, 7, 10	7,8,2,6,9,3,4,5,10	Build up
6	2,38	7, 6, 5, 17, 16	7,6,5	Counter Fail
7	3,05	3, 9, 8, 7	7,8,9,3	Build up
8	3,21	2, 1	2	Build up
9	4,09	2, 6	2,6	Build up
10	4,52	7, 9, 7, 4, 5	7,9,4,5	Recovery Fail
11	5,27	7, 8, 9	7,8,9	Recovery Fail
12	6,06	8, 10	8,10	Recovery Fail
13	6,28	2, 1, 6, 7, 11	7,2,6	Build up
14	7,3	2, 1, 2, 3, 8	8,2,3	Build up
15	8,01	16, 15, 12, 13, 14		Set piece
16	8,41	5, 4, 7, 9, 8, 7, 9, 10	7,8,9,4,5,10	Set piece
17	9,19	10, 9	9,10	Set piece
18	9,57	2, 3, 2, 8, 4, 6, 7, 9, 10	7,8,2,6,9,3,4,10	Build up
19	10,58	3, 9	9,3	Recovery Fail
20	11,39	2, 3, 8, 2, 4, 5, 13, 14	8,2,3,4,5	Build up
21	13,08	2, 3, 8, 7	7,8,2,3	Build up
22	13,25	8, 7, 6	7,8,6	Recovery Fail
23	13,48	2, 1	2	Build up
24	14,07	1, 2, 6, 4, 7, 8, 15, 7, 14	7,8,2,6,4	Build up

Build up	11
Counter Fail	3
Recovery Fail	7
Set piece	3
Shoot (on/off) target	0
Total	24

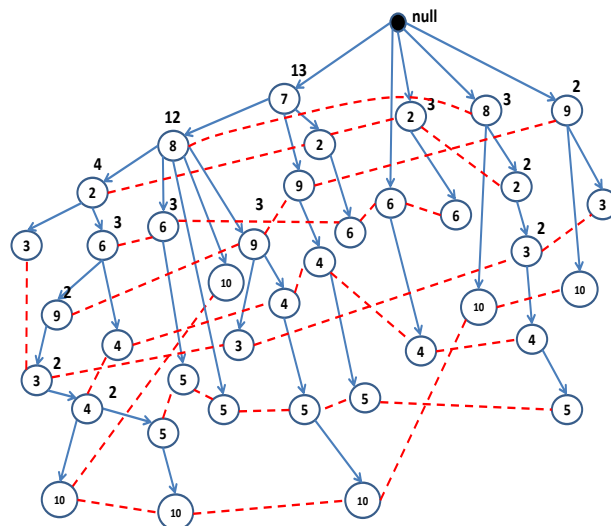


Figure 7. Complete FP Tree of Playing Pattern

CONCLUSION

FP Tree algorithm can be used to determine the playing pattern of a football team, by dividing the team's playing patterns, especially when attacking. This is done by modeling actual data into numerical data which is then normalized. The next process is carried out by forming a conditional pattern base, generating conditional FP-Tree, searching for frequent Itemsets and continuing with forming FP-Growth. Based on this pattern, if carried out in combination with technical and non-technical observations (stamina, physique, formation, individual player abilities, etc.) an analysis can be obtained along with an overview of the characteristics (strengths and weaknesses) of a team in a certain time duration. For example, when the national team game was known to use the midfield as the main playing area, (during 15 minutes of the first half) against Curacao. However, in the following minutes, he used the center area which leaned to the right wing to break into the opponent's playing area. Occasionally breaking through on the left wing, by closing the midfield as the center of defense and attack.

REFERENCES

- Goldie Gunadi., dan Dana Indra Sensuse (2012). "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) : Studi Kasus Percetakan PT. Gramedia." Jurnal TELEMATIKA MKOM. 1 (1). 118 132.
- Turban, Aronson, Liang., (2005). "Decision Support System and Intelligent System." Seventh Edition, Pearson/ Prentice Hall, New Jersey, USA.
- Kusrini, Luthfi, E.T., (2009). "Algoritma Data Mining. ", Andi Offset. Surabaya.
- Larose, Daniel T, (2006). "Data Mining Methods and Models." Hoboken New Jersey : Jhon Wiley& Sons, Inc.
- <https://www.guidetofm.com/squad/scout-reports>, diakses tanggal 20 September 2022.
- Bompa. (1990). "Theory and Methodology of Training.", Dubugus, Iowa: Kendall. Hunt Publishing Company,
- Setiawan Dwi, (2010). "Identifikasi Bakat Olahraga Siswa Putra Kelas 1 Di Sekolah Menengah Pertama Negeri Se-Kecamatan Jumantono Kabupaten Karanganyar Tahun 2008 / 2009. " Skripsi.Fakultas Keguruan Dan Ilmu Pendidikan Universitas Sebelas Maret : Surakarta.
- Affriantari Rochmah., (2010). "Perbandingan algoritma apriori dan Algoritma FP Growth dalam Optimasi Penentuan Frequent Itemset". FT UNS, Solo.
- Santosa, B. Data Mining: (2017). "Teknik Pemanfaatan Data untuk Keperluan Bisnis." Graha Ilmu. Yogyakarta,
- David Samuel, (2008). "Penerapan Stuktur FP-Tree dan Algoritma FP Growth dalam Optimasi Penentuan Frequent Itemset. " Bandung: Institut Teknologi Bandung,

<https://sport.detik.com/aboutthegame//pandit/d-3054769/menilik-cara-kerja-scout>, diakses tanggal 19 Agustus 2022. (Handbook style) *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44–60.2022.

L Ardiantoro, S Zahara, N Sunarmi, (2019). “Pemanfaatan Knowledge Data Discovery (KDD) Pada Pola Permainan Atlet Bulutangkis”, - *Explore IT!:* Jurnal, <https://jurnal.yudharta.ac.id/v2/index.php/EXPLORE-IT/article/view/1467>, Universitas Yudharta, Pasuruan. <https://doi.org/10.35891/explorit.v1i1i.1467>, Indonesia.