The Analysis and Implementation of Algorithm of Frequent Pattern – Growth to Support the Promotion Strategy in Victory University Sorong

Iriene Surya Rajagukguk[#1], Sri Yulianto Joko Prasetyo[*2], Irwan Sembiring[#3]
[#]*Faculty of Information System, Satya Wacana Christian University*
Diponegoro Street Number 52-60 Salatiga, Central Java, Indonesia

Abstract

Nowadays, the increasing number of campuses is growing, which makes want to get a better promotion strategy. As for how to find the right promotional strategy will be able to reduce the cost of promotion and achieve the right promotional goals. One way that can be done for the determination of a promotion strategy is to use data mining techniques. The technique used in this case is FP-Growth Algorithm is an alternative algorithm that can be used to determine the most frequent item set in a data set. Research is done by observing some research variables that are often considered by universities, especially the marketing section in determining the promotional target of the last education, the origin of the region, the choice of major and promotion strategy. The result of this research is a suitable promotion strategy to be applied at Victory University Sorong namely advertising, personal selling and publicity. And the most influential is advertising, this can be seen from the amount of the advertising value of 35 transaction from the total sample number of 43 transaction with a significant level 81.40 % of the other dimensions. Personal selling 6 transactions, the effect is 13.95 % a publicity as much as 2 transactions or 4.65 % of the other dimensions.

Keywords

Data mining, association rules, frequent item set, FP-growth algorithm.

I. INTRODUCTION

The competition in this world business is very strict, especially in promoting the university. Every campus tries to give the best facility. Therefore each campus tries to find out the exact strategy to promote its university. According to the developers to find a strategy that can determine marketing strategy in promoting university by maximizing services to the community.

One of technique that use in the implementation of Data Mining is in the promotion scale. If the promotion target is not determined well, in the sense of not seeking potential promotional target, then just spend a lot of time and cost that should be minimized through the selection of good promotional target.

As for in this research will be discussed how to apply one of algorithm in the data mining, that is Frequent Pattern-Growth (FP-Growth) algorithm. This algorithm is part of association technique on the data mining. As for FP-Growth is one of algorithm alternative that can be used to determine frequent item set in a set of data. The characteristic of FP-Growth algorithm is the data structure that used on the tree called FP-Tree. By using FP-Tree, FP-Growth algorithm can directly extract frequent item set from FP-Tree.

Victory University Sorong is private university that is located in Sorong West Papua. As time goes by, then began many emerging private universities and high schools in Sorong or Sorong Regency. Of course to face this competition required strategies to establish the existence of Victory University Sorong. The increasing number of private university and requires Victory University Sorong to conduct a diverse range of active promotions, in order to not less competitive with other universities.

This research will do the grouping based on Victory Sorong student data. The attributes to be used in determining the grouping of territories for promoting the Victory University is the latest education, origin, choice of major and promotion strategy. The result of this student data processing aims to assist the marketing of Victory University Sorong in conducting promotion and search for new students from various regions in Sorong.

The purpose of this research is to analyze and apply the promotion strategy in Victory University Sorong.

II. LITERATUR REVIEW

Research [1] [Improved Algorithm for Frequent Item sets Mining Based on A priori and FP-Tree] In this paper the researchers experiments based on two algorithms into one such APFT, which combines a priori algorithm and FP-tree structure of the FP-growth algorithm. The advantage of APFT does not

produce conditional & sub-conditional patterns of trees recursively and experimental results show that FP-tree works quickly than A priori and is almost as fast as FP-Growth. They have proposed to go further and modify APFT to include correlated items & uncorrelated item set. This additional feature optimizes the FP-Tree & removes the assembled items directly from the item set frequent. They chose to call this method as APCFT method with APFT correlation.

Research [2] Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data] Implementation of data mining on medical data to generate rules and patterns using FP-Growth algorithm. The research study in this paper presents how data mining can apply to medical data. Therefore, data and empirical results are presented in this paper to provide more guidance to the doctors as well as to better understand the relationship of the doctors and a patient. For that, the first discuss about the importance of data mining using medical data and then the discussion of general data mining techniques have been presented. Next, methodology describes conceptual model for extraction rules about the medical database finally the result can guide the relationship between the different attributes presented in the data.

Research [3] [FP Growth Algorithm Implementation] This paper discusses FP Tree concept and apply it uses Java for general social survey dataset. They use this approach to determine the association rules that happens in dataset. By this way, can set relevant rules and patterns in a set of records.

Research [4] [FP-Tree Based Algorithms Analysis: FP Growth] Mining frequent item sets from large transactional databases have important and very critical tasks. Many algorithms have been proposed from recent years. Yet FP-Tree-like algorithms are regarded as highly effective algorithms for efficient frequent item set circuits. This algorithm is considered efficient because of its compact structure and also a slightly item set generation compared to the A priori Algorithm. Therefore, the purpose of this paper is to provide a basic concept of some algorithms (FP-Growth, COFI-Tree, CT-PRO) based on FP-Tree such as the structure to mine frequent item sets equal to the ability and comparison.

Research [5] [FP Growth Tree Implementation In Bank Transaction Databases] In this paper describes the application of FP-Growth algorithm in the database of bank transactions to ensure the best performance. The purpose of this research to ponder and test existing association techniques for collecting frequent mining items and successful implementation of the FP-Growth algorithm in a database of banking transactions to improve the performance of frequent item sets in data mining.

Research [6] [Overview on Methods for Mining High Utility] In this paper, proposed two algorithms, namely the growth of utility pattern (UP-Growth) and UP-Growth +, to mine item set with high utility with a set of effective strategies for item set candidate. The information of high utility item set is maintained in Up-Tree, item set candidate can be generated efficiently with two database scans. The result of experiments shows that the proposed algorithm, especially UP-Growth+, not only reduces the number of candidates effectively but also outperforms other algorithms sub-stance in terms of run time, especially when the database contains many long transactions.

Research [7] [A Survey on Association Rule Mining in Market Basket Analysis] This method examine customer purchase patterns by identifying associations between different items where customers are in the shopping cart. Identification of such associations can help retailers broaden their marketing strategy by gaining insight into what customers often buy. This paper presents a survey of existing data mining algorithms for shopping cart analysis.

Research [8] [Market Basket Analysis for a Supermarket based on Frequent Item set Mining] In this research, shopping cart Analysis is am importance component in the system analysis in retail organization to determine placement of stuff, design the marketing promotion for various segment of customer to increase customer satisfaction and profit from the supermarket. These issues are aimed at leading supermarkets by using frequent items set mining. The item set is mined from the market basket database using the efficient K-A priori algorithm and the resulting association rules.

The difference between this research and the previous one is on this research has been found promotion strategy that is suitable to use at Victory University Sorong along with 3 promotion mixes are: advertisement dimension, personal selling dimension, and publicity dimension.

A. FP-Growth Algorithm

FP-Growth is one of algorithm alternative that can be use to determine the frequent item set in a data set. On the FP-Growth algorithm uses tree development concept, commonly called FP-Tree, in searching frequent item set is not using generate candidate as it is done on A priori algorithm. By using the concept, FP-Growth algorithm becomes faster than A priori algorithm. [9]

FP-Growth Method is divided into three main stages:

- Generation stage conditional pattern base,
- Generation stage conditional FP-Tree, and
- Searching stage frequent item set.

B. Association Rules

Association rule is one of method that aims to find the frequent pattern that often apper among many transactions, where each transaction consists of several items so that this method will support the recommendation system through the discovery of

patterns between items in transactions that happened. [10] The basic methodology of association analysis is divided into two stages:

- High frequency pattern analysis

This stage looks for a combination of items that fulfil the minimum requirements of the support value in the database. The value of an item's support is obtained by the following formula:

$$Support\ (A) \frac{\text{The number of transaction contains A}}{\text{Transaction Total}} \quad (1)$$

Explanation:

A: which school are you from

B: origin

On the formula 1 explains that support value (A) is obtained by finding the number of transactions containing A in divide by total transaction.

- Formation of Associative Rules

After all the frequency patterns are found, then find the associative rules that supply the minimum requirements for confidence by calculating the confidence of the associative rules A_B The confidence value of rule A_B is obtained from the following formula:

$$Confidence = \frac{\in \text{Transaction contains A and B}}{\in \text{Transaction contains A}} \quad (2)$$

Explanation:

A: which school are you from

B: origin

III.RESEARCH METODOLOGY

In the research methodology there is a sequence of frameworks to be followed, the order of this framework is a description of the steps that must be passed so that this research can work well. The framework to follow can be seen on figure 1.

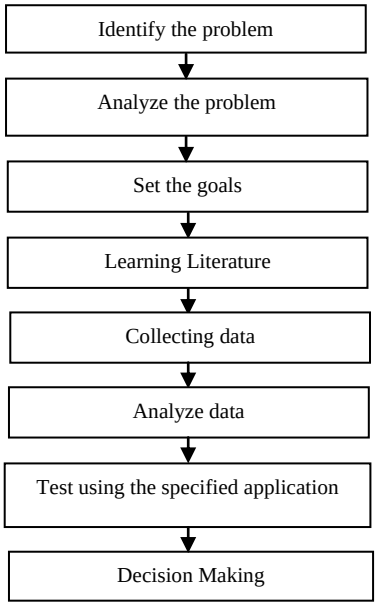

Fig 1: Research Framework [11]

Based on Figure 1 the work sequence steps are described as follows:

A. *Identifying Problem*

The problem identified in this research is to find the cause of the decreasing number of new student candidates amidst intense competition among universities.

B. *Analyzing the Problem*

Problem analysis in this research is done by two methods that are descriptive method and comparative method.

- Descriptive method

In this method, the data is submitted, arranged, classified and analyzed so that the clear description of the problem which will be discuss can be taken.

- Comparative method

In this method, analysis is done by comparing the theory and practice so that the description about the similarities and the differences between of them can be taken.

C. *Determining the Aim*

Based on the explanation above, the aim of this research is to know how far is the implementation of promotion strategy which has been done by Victory Sorong University.

D. *Studying Literature*

The literature which is used as reference in this research is from scientific journals, learning module and books about Data Mining. These literatures will be guidelines to do the research in order to make easy the process of the research.

E. *Collecting Data*

The method of collecting data is done by doing the observation to the College of Victory Sorong University. Besides of observation, the interview is also done toward the related parties of this research.

F. *Analyzing Data*

In this stage, the data which has been collected will be analyzed. Analysis using Association Rule is done by support and confidence counting mechanism from an item relation. An association rule can be said as important if the support value is bigger than minimum support and also the confidence value is bigger than minimum confidence. By using FP-Growth technic which result Frequent Itemset without doing candidates generation with the purpose to get new knowledge in form of information about the effective and efficient promotion strategy in getting the candidate of new student in the middle of the binding competition between Universities.

G. *Testing Use Certain Application*

In this stage, rules are tested or examined using data mining system. Tools which are used as system testing is Rapidminer Studio 6.4.for the first, the manual data "the data of the new student candidate

2008-2016" which is available is transformed into Microsoft Excel.

As the first stage in the testing process is importing the data of new student in the collage of Victory Sorong University which will be made as the table of item frequency transaction on tools Rapidminer Studio 6.4. FT-Growth analysis is definite as a process to find all of FT-Growth rule which fulfill the minimum support requirement and minimum requirement for confidence. From the process of Data Input and Affinity, it will be found a certain pattern or criteria on getting promotion strategy and will be processed back to result an association rule. From the association rule, the result will be seen. The result is a new knowledge and also as reference to analyze the strategy on promoting education.

H. Taking Decision

After being examined, the result of analysis will show the comparison between the manual manner and tool utilizing. The next stage is determining and decision taking toward new knowledge which is gotten in form of finding related with the strategy in promoting education.

IV. RESULT OF THE RESEARCH AND DISCUSSION

The characteristic of FP-Growth algorithm is data structure which using FP-Tree. By using FP-Tree, FP-Growth algorithm can directly extract frequent itemset from FP-Tree. Itemset excavation by using FP-Growth algorithm will be done by generating the structure of data tree or usually known as FP-Tree.

Input: FP-Tree Tree

Output: Rt complete group of frequent pattern

Method: FP-Growth (Tree, null)

Procedure: FP-Growth (Tree, Θ)

[

01: if Tree consist single path P;

02: then to every combination (detonated β)

From the nodes in path do

03: generate $_{\beta}$ ά pattern with the support from nodes in $_{\beta}$;

04: else for every a1 on the header of Tree do

05: generating pattern

06: form $_{\beta}$=a1 ά with the support = a1.

Support

07: if Tree $_{\beta}$ = ά

08: then summon FP-Growth (Tree, $_{\beta}$)

}

}

FP-Growth Algorithm [11]

Before doing the mining process, preprocessing is done first by changing the format of the data so that it can be used on the software. The appearance of the stages and the result is as the follows:

TABLE I
ITEMSET

TID	Itemset
1	{A3,B4,C6}
2	{A4,B4,C6}
3	{A3,B6,C6}
4	{A1,B4,C6}
5	{A3,C6}
6	{A3,B6,C6}
7	{A4,B4,C6}
8	{A1,B4,C6}
9	{A1,B1,C6}
10	{A1,B1,C6}
11	{A1,B4,C6}
12	{A1,B4,C6}
13	{A2,B1,C6}
14	{A4,B1,C6}
15	{A3,B4,C6}
16	{A2,B1,C6}
17	{A1,B4,C6}
18	{A1,C6}
19	{A1,C6}
20	{A1,C6}
21	{A1,C6}
22	{A1,C6}
23	{A2,B4,C6}
24	{A1,B4,C6}
25	{A2,B4,C6}
26	{A2,B4,C6}
27	{A3,B4,C6}
28	{A2,B4,C6}
29	{B4,C6}
30	{A3,B4,C6}

So, the next stage is forming the tree of FP-Tree by seeing the table 1 as a follow:

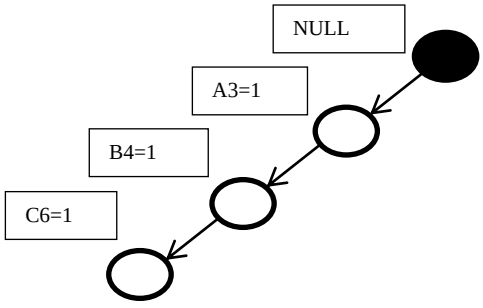

Fig 2: The result of fp-tree forming after reading tid 1

Explanation:

= Root

= Data Flows

= Branch

Figure 2: is the explanation of FP-Tree forming after reading TID 1 which contains: NULL - A3 (Public SMA) = 1 – B4 (Ayamaru) = 1 – C6 (Indonesian language education) = 1.

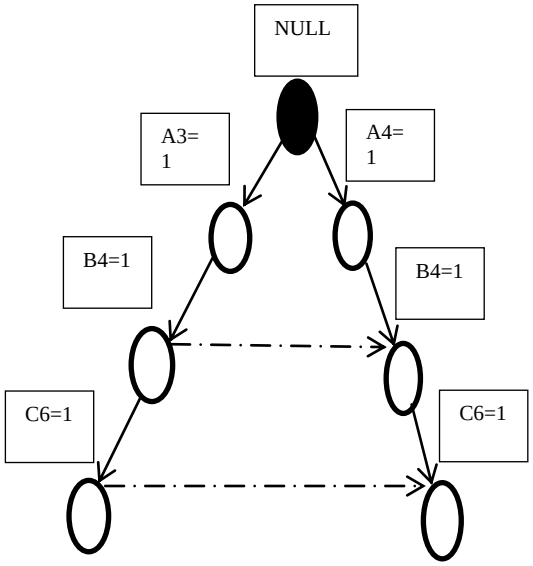

Fig 3: The result of fp-tree forming after reading tid 2

Explanation:

= Root
= Data Flows
= Branch
= same data flows

Figure 3: is gotten after doing TID 2 that is: NULL - A4 (Private SMA) = 1 - B4 (Ayamaru) = 1 – C6 (Indonesian language education) = 1.

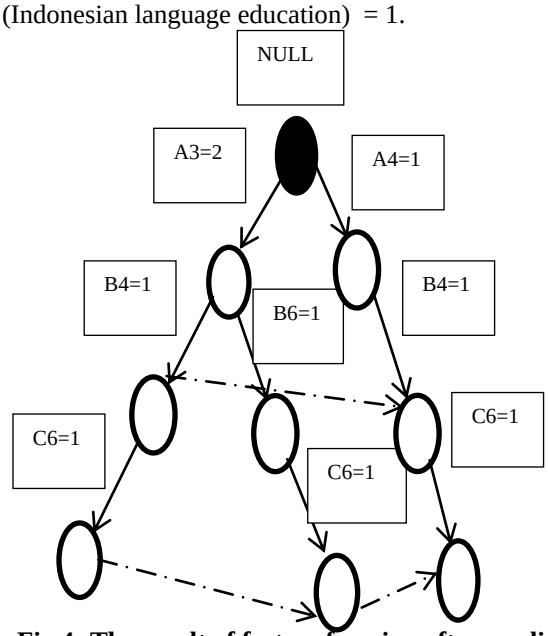

Fig 4: The result of fp-tree forming after reading tid 3

Explanation:

= Root
= Data Flows
= Branch
= Same data flows

Figure 4: is gotten after doing TID 3 that is: NULL – A3 (Public SMA) = 2 - B6 (Tamrauw) = 1 – C6 (Indonesian language education) = 1.

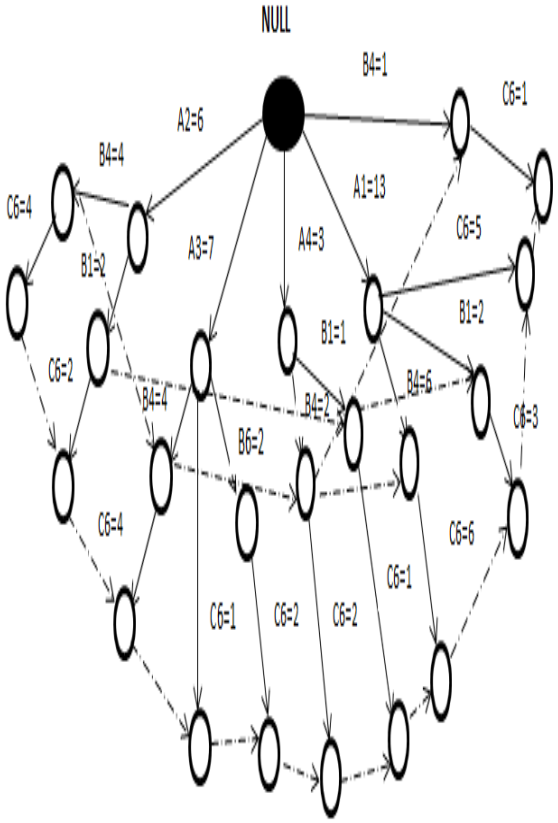

Fig 5: The result of fp-tree forming after reading tid 30

Explanation:

= Root
= Data Flows
= Branch
= same data flows

Figure 5: is gotten after doing TID 3 that is : NULL – A3 (Public SMA) = 7 – B4(Ayamaru) = 4 – C6 (Indonesian language education) = 4.

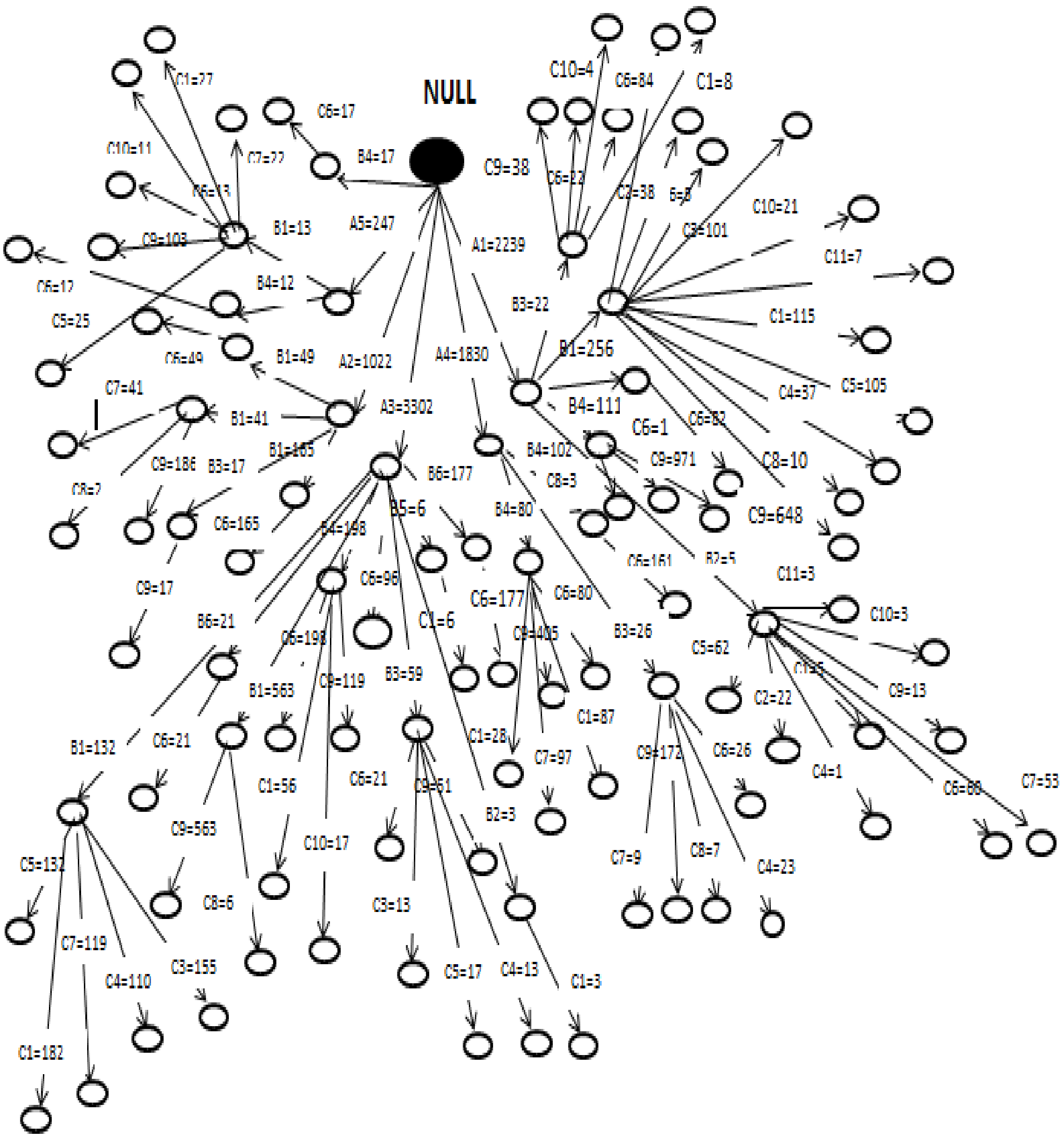

Fig 6: The result of fp-tree after reading tid 8729

Figure 6 is gotten after doing TID 8729. From frequent itemset which is gotten from FP-Tree forming and FP-Growth so the support and confidence value can be counted as the follows:

- Support (Public SMA, Sorong, Electro) = Count (Public SMA, Sorong, Electro)/ the sum of transaction = 1/8729.

While for the confidence our trusted value is as the follows:

- Confidence (Public SMA, Sorong, Electro) = Count (government supported senior high school, Sorong, Electro)/ the sum of transaction = 1/51.

After support and confidence value is gotten from the whole combination on the data with FP-Tree and FP-Growth counting so there is found the highest and most accurate support and confidence value that is the combination (A3, B1, C9) {Public SMA, Sorong, Information system} which has support value: 563/8729)=0,064 and confidence value : 563/4139=0,136.

A. Promotion Strategy

This research shows that in doing promotion the Victory Sorong University did 3 promotion strategies, that are advertisement, personal selling and publicity. Advertisement is the most affected and publicity is non-affected promotion in increasing the sum of student.

The dimension of advertisement is done by making billboard, banner, poster, brochure and tassle that they set on the strategic place. Such as in the edge of street, surrounding school environment like government

supported and private SMA/SMK and another busy place.

Advertisement can be done through media, such as:

- Electronic media (Television, radio, cinema, internet).
- Print media (newspaper, magazine, bulletin)
- Outdoor media (electronic board which is set in the edge of the street which moved electronically, billboard, banner, poster, ect).

The dimension of personal selling is done by visiting or socialization to some schools like goverment supported and private SMA/SMK. Usually, promotion team is consisted of 2 until 3 people who has a job to give information directly to the school and the third grade student who will join national final examination. The dimension of publicity is done by making a workshop or event in the collage and in the outside of collage then inviting reporter from print media to pervade the event. Besides of that the promotion team uses internet media through social network FB and website to share the picture or photos of the workshop and event which they have done to the entire user of another social network. This thing is expected to make good image of the collage in the society. The result of data analysis which is gotten from the research by using quantitative research method to answer the problem statement. First to see the effect, the whole result is as the follows:

TABLE II
THE RESULT OF PROMOTION USING RAPIDMINER

Index	Nilai Nominal	Frequent	Fraction
1	Advertisement	5778	0.662
2	Personal selling	1801	0.206
3	Publicity	1150	0.132

To get the whole information about the promotion strategy which is done by Victory Sorong University, choosing research information is done based on the research aim that is by using sample.

Data sample which is taken are 43 people on October 2017 as it seen in the table III.

TABLE III
THE RESULT OF PROMOTION STRATEGI IN THE FIELD

Index	Nilai Nominal	Frequent	Fraction
1	Advertisement	35	0.8140
2	Personal selling	6	0.1395
3	Publicity	2	0.0465

It shows that advertisement is the most affected promotion strategy toward the increasing sum of student in Victory Sorong University than using personal selling and publicity.

The effect of advertisement dimension is most significant. It can be seen that 35 transaction from the whole sample with significant 81.40% than another dimension. This research explains that advertisement is the most dominant variable in affecting the sum of student. In making advertisement, university uses the electronic media which consist of Television, radio, and internet, and also print media through banner, brochure and tassle.

The effect of Personal Selling dimension is quiet significant. It can be seen that the effect of 6 transaction is 13.95% than another dimension. The effect of publicity dimension is not significant. It can be seen that the value is 2 transactions with significant rate 4.65%.

The inhibiting factors of promotional activities are the funds promotional activities provided for operational activities felt not sufficient to carry out advertising promotion activities, personal selling, and publicity to the maximum in order to get results as expected.

V. SYSTEM IMPLEMENTATION

To prove the truth of the analysis results required a testing process to test the truth of the results of data processing done manually in chapter 3 that has been done before, for the testing process we can use one software application such as Rapidminer, with the steps as follows:

The dataset consists of 8729 data records and contains the variables attributes of public SMK, private SMK, public SMA, private high school, Madrasah Aliyah, Sorong, information systems, etc. are stored in Microsoft Excel with the new student candidate name 2008-2016. xls and which will be attempted using Rapidminer studio 6.4 software to see the same results as a frequent itemset search through:

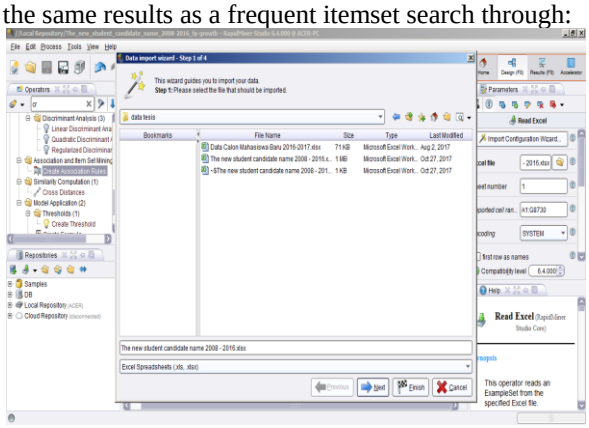

Fig 7: The appearance of excel sheet

Figure 7: is the process of data import of new student candidate.xls into rapidminer.

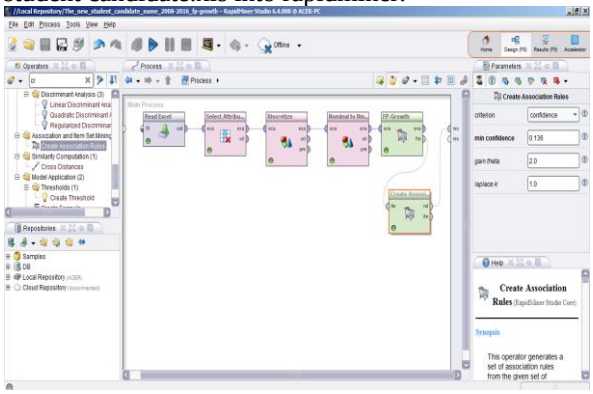

Fig 8: The process of connecting fp-growth to first race and create association rules into second races.

In the figure 8: it is explained about the process of connecting FP-Growth into first race and create association rules into second races and set into minimum support and minimum confidence.

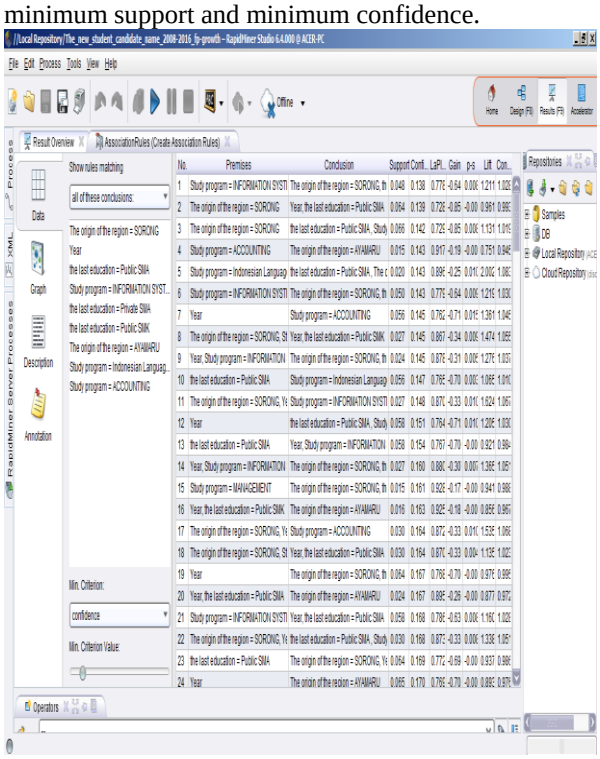

Fig 9: The appearance of the result of data processing use fp-growth

In figure 9: it is explained about the appearance of the result of data processing use FP-Growth and its relation to association rule.

Lift ratio is important parameter besides of support and confidence on association rule. Lift ration measures how important the rule that has been formed based on the value of support and confidence. Lift ratio is values that indicates the validity of the transaction process and provides information whether it is true that product A is purchased together with product B.

Improvement Ratio is obtained by the following formula:

$$\frac{Support\ (A \cap B)}{Support\ A \bullet Support\ B} = \qquad (3)$$

To find valid rule value is if have lift ration value >1 by way of support lift ration = support containing value A and value B divided support A * support B support value containing value A and value B is minimum result from minimum support divided with items (A3, B1, C9) {public SMA, Sorong, system information} At each occurrence transacted.

As for the results of the rule that most influence is: if the prospective student is from high school public school, Sorong address then he will choose the system of information systems with 100% confidence level and supported 47% of the overall data. The results of that rule will be targeted in promoting education.

VI. CONCLUSION AND SUGGESTION

A. *Conslusion*

After finishing the analysis and also the system implementation, it can be concluded as the follows:

1. As big as the sum of transaction which is processed so as the FP-Growth load process will be longer.
2. After doing research in the field, by using 43 transaction sample, it results: advertisement 81.40% or 35 transaction, personal selling 13.95% or 6 transaction and publicity 4.65% or 2 transaction. The suitable promotion strategy to being applied in Victory Sorong University is advertisement, Personal selling and publicity.
3. The obstacles which are found along promotion event is the minimum fund which is given to the whole promotion so that the promotion gives less maximal contribution toward the increasing sum of student.

B. *Suggestion*

Suggestions which can be noticed for the implementation of this technique are:

1. To get the good promotion result, it is needed more data sources and more complete data resource.
2. To implement rule which is resulted from FP-Growth algorithm with the management side can develop the strategy on education promotion so that the collage side will be more advance.

ACKNOWLEDGMENT

I hereby express my gratitude to the Victory Sorong University foundation that has provided funding and support during the course of this research.

REFERENCES

[1] Sujatha Dandu, B.L. Deekshatulu & Priti Chandra.“Improved Algorithm for Frequent Item sets Mining Based on Apriori and FP-Tree”. Global Journal of Computer Science and Technology Software & Data Engineering, Volume 13 Issue 2 Version 1.0 Year 2013, Type: Double Blind Peer Reviewed International Research Journal, Publisher: Global Journals Inc. (USA), Online ISSN: 0975-4172 & Print ISSN: 0975-4350.

[2] Abdullah Saad Almalaise Alghamdi. “Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data”. IJCSNS International Journal of Computer Science and Network Security, Vol.11 No.12, Dec 2011.

[3] Shivam Sidhu, Upendra Kumar Meena, Aditya Nawani, Himanshu Gupta, Narina Thakur. “FP Growth Algorithm Implementation”. International Journal of Computer Applications (0975 – 8887) Vol. 93 – No.8, May 2014.

[4] Bharat Gupta, Dr. Deepak Garg. “FP-Tree Based Algorithms Analysis: FPGrowth”. Bharat Gupta et al. / International Journal on Computer Science and Engineering (IJCSE). ISSN : 0975-3397 Vol. 3 No. 7 Jul 2011.

[5] Aroosa Hameed. “FP Growth Tree Implementation In Bank Transactional Databases”. ITEE Journal Information

Technology & Electrical Engineering, ISSN: - 2306-708X, Vol.5, Issue 3 Jun 2015.

[6] Smita R. Londhe, Rupali A. Mahajan, Bhagyashree J. Bhoyar. "Overview on Methods for Mining High Utility Itemset from Transactional Database". International Journal of Scientific Engineering and Research (IJSER),www.ijser.in, ISSN (Online): 2347-3878, Vol. 1 Issue 4, DEC 2013.

[7] Savi Gupta and Roopal Mamtora. "A Survey on Association Rule Mining in Market Basket Analysis". International Journal of Information and Computation Technology. ISSN 0974-2239 Vol. 4, No. 4 (2014), pp. 409-414 © International Research Publications House http://www. irphouse.com /ijict.htm.

[8] Loraine Charlet Annie M.C. and Ashok Kumar D. "Market Basket Analysis for a Supermarket based on Frequent Itemset Mining". IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, Sep 2012, ISSN (Online): 1694-0814. www.IJCSI.org.

[9] Ririanti. "Implementation of FP-GROWTH Algorithm on Motorcycle Prediction Application (Case Study of PT Pilar Deli Labumas)". Pelita Informatika Budi Darma, Vol : VI No: 1, Mar 2014. ISSN : 2301-9425.

[10] Ali Ikhwan, Dicky Nofriansyah, Sriani. "Implementation of Data Mining with Fp-Growth Algorithm to Support Education Promotion Strategy (Case Study Campus STMIK Triguna Dharma)". Jurnal Ilmiah SAINTIKOM, ISSN : 1978-6603, Vol.14, No.3, Sep 2015.

[11] Erwin. "Market Basket Analysis With Apriori Algorithm and FP-Growth". JURNAL GENERIC, ISSN 1907-4093, Vol. 4 No. 2, Jul 2009.