# The Utilization of Ontology to Support The Results of Association Rule Apriori

1st Dewi Wardani
*Department of Informatics*
*Universitas Sebelas Maret*
Surakarta, Indonesia
dww_ok@uns.ac.id

2nd Achmad Khusyaini
*Department of Informatics*
*Universitas Sebelas Maret*
Surakarta, Indonesia
achmadk@students.uns.ac.id

***Abstract*—Association rule is one of the data mining techniques to find associative combinations of items. There are several algorithms including Apriori, FP - Growth, and CT-Pro. One of the advantages of the Apriori algorithm is that it produces many rules. To improve its result, one of the methods is by using the semantic web technology. This work proposes how the hierarchical type of ontology can be utilized by the Apriori algorithm to improve the results. The Apriori with ontology implements the Interestingness Rule (IR) which is a parameter to determine the degree of association between combinations of items in a dataset. The series of experiments show that the proposed idea can improve the results compare to the default Apriori algorithm.**

***Keywords*—*Association Rule, Apriori, Ontology, Interestingness***

## I. Introduction

Data mining (DM) is the process of finding patterns or interesting information in selected data using a particular technique or method. The association analysis or association rule mining is a data mining technique for finding associative rules between a combination of items. Association rule analysis is also known as one of the data mining techniques that became the basis of one of the other data mining techniques [1]. Apriori algorithm is an algorithm for finding frequent itemsets patterns on association rules. The main steps in Apriori algorithm are: first, look for frequent itemset (set of items that meet the minimum support) of the transaction database, second eliminate the itemset with low frequency based on the predetermined minimum level of support. Next, building the association rule of the itemset that meets the minimum confidence value in the database [2].

The main problem of applying conventional methods of association rule by using Apriori algorithm is [3]: 1) Produce many rules, so that the process of tracing information becomes difficult within the scope of the rule that becomes widespread. 2) Produce many unattractive rules in high numbers resulting in a meaningless rule. 3) Long-running time process. If the database is used in large numbers and varies, it will take a long time to run the algorithm.

Semantic Web is an approach developed specifically on the World Wide Web (WWW) technology that aims to enrich the information provided so that it becomes better in defining it. Semantic Web allows a web to become more intelligent because it has a knowledge base (knowledge base) in it in the form of ontology. In the semantic web technology, ontology serves as the core (core technology) so that it can be called as semantic web ontology [4]. Ontology is a set of hierarchically structured terms to describe a domain that can be used as the basic framework of a knowledge base [5].

The development of ontology aims to capture knowledge into a format that can be used in the system. Next, by populating the knowledge base to get the instances to fill the knowledge base. The benefit of ontology is to explain a domain of knowledge; provide a hierarchical structure of concepts to explain a domain and how they relate [3].

One of the technologies to improve the default Apriori is by using semantic web technology which applied knowledge base in the form of ontology. By using the domain ontology, it has a positive impact to support the association rule, which can do the pruning on the rule that is not interesting [6]. By using ontology can improve knowledge discovery process in association rule [7]. Ontology is used to support data integration process mining, that is combining knowledge domain into data mining process [8]. The ontology in the hierarchy approach can simplify all data components to a certain level of form so that they can be comprehensively understood. Then the data is classified according to its category to facilitate the searching of information in the dataset.

This research considers the lack of application of conventional association rule method using Apriori algorithm and excess ontology. Hence, required a merger between association rule and ontology. The use of ontology is expected to provide a solution to the problem of conventional association rule methods, namely: the problem of producing multiple rules and producing many rules that are not interesting in high numbers. The main novelty in this work is how to combine the hierarchical ontology into the Apriori algorithm, and compose the new IR formula as well.

## II. Related Work

A work that the using an ontology domain has a positive impact to support association rule. It can anticipate the existence of ambiguous data, reducing the results of the rule so that more accurate [6]. The other study found a diverse collection of ontology entries on the KDD process. Tests were conducted on two domains, namely medical and sociological fields [7].

The research by Mahmoodi et al. [3] provides information on the use of association rule model with ontology as the solution to solve the problem of tracing information on factors that cause cancers. Test obtained results with clinical databases show that the prediction model according to the desired goal is to produce the best rules. The application of association rule with ontology gives a good impact in the search results of the rule becomes more accurate, unlike the results using only the Apriori algorithm. The test results obtained show that the prediction model according to the desired goal can produce the best rules. Therefore, the method of combining the Apriori algorithm and the concept of ontology gives a good impact in the search results of the rule more accurate, in contrast to the results using the Apriori algorithm alone.

The other work suggested that the proposed method which combined the using ontology is more efficient than other association rule methods. The number of rules generated can be reduced even in a large number of datasets [9].

A similar work which used the association rule mining with ontology (ARMO) method gives more efficient results than other association rule methods. The number of rules generated can be reduced though on the number of large datasets. The results of this study suggest that the proposed method can reduce the rule results produced from the association rule mining process on a large number of datasets [10].

In the other previous study [8] that using the ontology domain has a positive impact to support association rule. It can do the pruning on the rule that is not interesting. It also can anticipate the existence of ambiguous data, and reduce the results of the rule so that more accurate. The model seeks to separate the independent sub-processes of the DM process and combine the mining results.

## III. Considered Problem And The Approach

The main proposed idea in this work is to implement the modified interestingness rule (IR) formula. Here, it is assumed that the ontology is consists of classes with no hierarchy among them. Each of class is a single stand class. All individuals of each class is a group hierarchy under the class. Therefore, the graph of ontology is a simple one.

The Apriori algorithm will be implemented with ontology and by considering the value of support and calculation of IR value. This IR value is used to improve the results of the rule search between items in the dataset. The formula calculates the following IR values which modified and inspired from [9], that the IR is customized for three itemsets A, B and C:

$$IR = \left[\log\left(2\,x\,Trans\left(A,B\right)\right)+\log\left(1\,x\,Trans\left(C\right)\right)\right]x\left[\frac{Trans(A,B,C)}{Total_{trans}}\right]$$

where Trans (A, B) is the number of stransactions from itemset A and B, Trans (C) is the number of transactions from itemset C, and TotalTrans is the total number of transactions in the database.

## IV. Experiment And Result

### A. Dataset

This experiment uses two real datasets: 1. Student data (actual data from students of Universitas Sebelas Maret), 2. Internet movie database data (IMDB). The student data provides information about the student's personal and academic data. Examples of student data include student name, the tuition class (UKT), scholarship id, study program, faculty, address, place of birth date (TTL), cumulative grade point (GPA), senior high school, study duration, entry point, parent income, etc.

The IMDB dataset provides information on films from around the world, including those involved in it from actors/actresses, directors and writers. In IMDB data there are some attributes such as director name, genre, movie title, country, number review, budget, etc.

Data with some attributes will be elected again at the preprocessing data stage. Preprocessing data is done by selecting the attributes that will be used. At this stage, it is also done refinement of data contents if the data is not complete. Data that is empty or less completse will be selected so that not included in the data processing. Completed data will be adjusted. Therefore, a program can understand the data.

### B. Ontology Development

At this stage will be designed ontology based on the used dataset. The design of ontology is based on hierarchy ontology approach. The ontology hierarchy approach aims to simplify all data components to form certain levels so that they can be comprehensively understood. Then, the data is classified according to the category. Examples of hierarchical forms of student data ontology can be seen in Figure 2 and Figure 3 below:

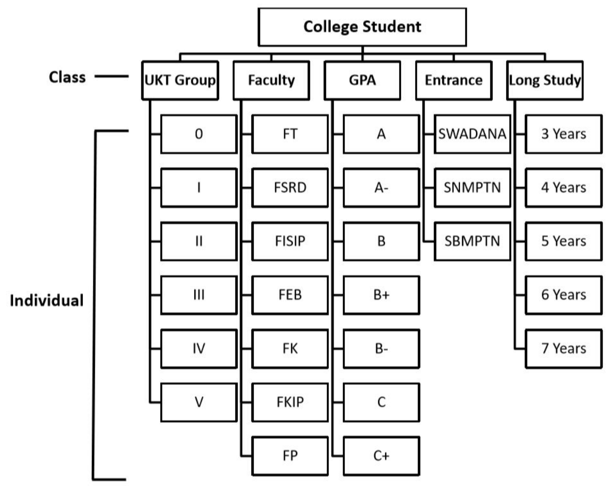


Fig. 1. Ontology of Student

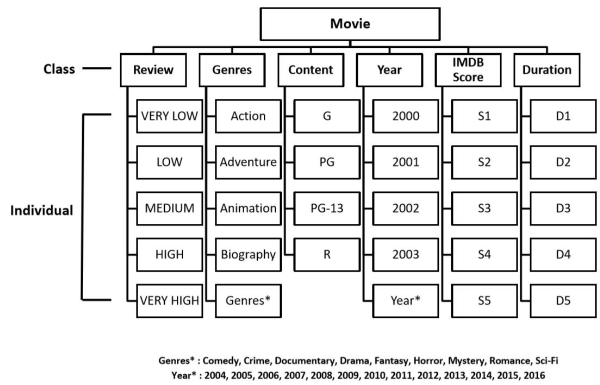


Fig. 2. Ontology of IMDB

### *C. Experimental setting*

A few attributes are written in the abbreviation for the shake of name simplification. Such as, A means GPA 3.69 – 4.00, FEB means Faculty of Economics and Business, etc. The scenario of the experimental is as below described in Table 1. Each experiment performers two type of experiment: Association Rule Apriori (ARA) and Association Rule Apriori with Ontology (ARAO).

TABLE I. THE SCENARIO OF EXPERIMENTAL

| Code | Dataset | Amount | Support | Confidence | IR |
|---|---|---|---|---|---|
| 1 | Student | 1000 | 0.2 | 0.5 | 0.1 |
| 2 | Student | 1000 | 0.2 | 0.3 | 0.05 |
| 3 | IMDB | 1000 | 0.2 | 0.5 | 0.1 |
| 4 | IMDB | 1000 | 0.2 | 0.3 | 0.05 |
| 5 | Student | 100 | 0.25 | 0.1 | 0.1 |
| 6 | Student | 100 | 0.2 | 0.1 | 0.1 |
| 7 | Student | 100 | 0.15 | 0.1 | 0.1 |
| 8 | IMDB | 2000 | 0.2 | 0.5 | 0.25 |

### *D. The results*

TABLE II. THE RESULT OF EXPERIMENT 1

| No | ARAO | No | ARAO |
|---|---|---|---|
| 1 | (B+), III→ SNMPTN : 1* | 1 | (FKIP) , III→SNMPTN :0.76 * |
| 2 | V→ SWADANA : 1 | 2 | 5 TH , III→SNMPTN :0.59 |
| 3 | (B), III→ SNMPTN : 0.99 | 3 | (B+) , SWADANA→V :0.58 |
| 4 | (B+), SNMPTN→ III : 0.99 | 4 | (B) , SWADANA→V :0.56 |
| 5 | (B), SNMPTN→ III : 0.99 | 5 | (FISIP) , III→SNMPTN :0.54 |
| 6 | 4 TH, SNMPTN→ III : 0.99 | | |
| 7 | SNMPTN→ III : 0.99 | | |
| 8 | III→ SNMPTN : 0.98 | | |
| 9 | SWADANA→ V : 0.98 | | |
| 10 | 4 TH,III→ SNMPTN : 0.97 | | |
| 11 | (B+),4 TH→ SNMPTN : 0.77 | | |
| 12 | 4 TH→ III : 0.75 | | |
| 13 | (B+)→III,SNMPTN : 0.74 | | |
| 14 | (B+)→SNMPTN : 0.74 | | |
| 15 | (B+)→III : 0.74 | | |
| 16 | 4 TH→III,SNMPTN : 0.73 | | |
| 17 | 4 TH→SNMPTN : 0.73 | | |
| 18 | (B)→SNMPTN : 0.64 | | |
| 19 | (B)→III : 0.64 | | |
| 20 | (B)→III,SNMPTN : 0.63 | | |
| 21 | III,SNMPTN→(B+) : 0.59 | | |
| 22 | SNMPTN→(B+),III : 0.59 | | |
| 23 | 4 TH,SNMPTN→(B+) : 0.59 | | |
| 24 | SNMPTN→(B+) : 0.59 | | |
| 25 | III→(B+),SNMPTN : 0.58 | | |
| 26 | III→(B+) : 0.58 | | |
| 27 | 4 TH→(B+) : 0.56 | | |
| 28 | III→4 TH : 0.5 | | |

The highest score of information can be obtained from table 2 as follow:

(i). ARAO produces fewer number of rules than ARA. The rules that ARAO generates are 5 rules while ARA has 28 rules.

(ii) In ARA, a combination of item "V→ SWADANA" is obvious information. Subjectively, this rue is not interesting. Meanwhile, in ARAO, one item appears "(B +)" as an additional item from the combination of items V and SWADANA so that it becomes more interesting.

TABLE III. THE RESULT OF EXPERIMENT 2

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | (B+),III→SNMPTN : 1* | 1 | (FKIP), |
| 2 | V→SWADANA : 1 | | III→SNMPTN :0.76 * |
| 3 | (B),III→SNMPTN : 0.99 | 2 | 5 TH , III→SNMPTN :0.59 |
| 4 | (B+),SNMPTN→III : 0.99 | 3 | (B+) , SWADANA→V |
| 5 | (B),SNMPTN→III : 0.99 | 4 | (0.58) |
| 6 | 4 TH,SNMPTN→III : 0.99 | | (B) , SWADANA→V :0.56 |
| 7 | SNMPTN→III : 0.99 | 5 | (FISIP) , |
| 8 | III→SNMPTN : 0.98 | | III→SNMPTN :0.54 |
| 9 | SWADANA→V : 0.98 | 6 | (FEB) , 3 |
| 10 | 4 TH,III→SNMPTN : 0.97 | | TH→SNMPTN :0.51 |
| 11 | (B+),4 TH→SNMPTN :<br>0.77 | 7 | (FSRD) ,<br>III→SNMPTN :0.47 |
| 12 | 4 TH→III : 0.75 | 8 | 4 TH , |
| 13 | (B+)→III,SNMPTN : 0.74 | | SWADANA→V :0.45 |
| 14 | (B+)→SNMPTN : 0.74 | 9 | (FKIP) , |
| 15 | (B+)→III : 0.74 | | SWADANA→V :0.45 |
| 16 | 4 TH→III,SNMPTN : 0.73 | 10 | (FKIP) , 4 TH→III :0.41 |
| 17 | 4 TH→SNMPTN : 0.73 | | (FKIP) , 4 |
| 18 | (B)→SNMPTN : 0.64 | 11 | TH→SNMPTN :0.41 |
| 19 | (B)→III : 0.64 | | (FEB) , |
| 20 | (B)→III,SNMPTN : 0.63 | 12 | SWADANA→V :0.41 |
| 21 | III,SNMPTN→(B+) : 0.59 | 13 | (FEB) , 4 |
| 22 | SNMPTN→(B+),III : 0.59 | | TH→SNMPTN :0.38 |
| 23 | 4 TH,SNMPTN→(B+) :<br>0.59 | 14 | (FK) , III→SNMPTN :0.35<br>(B) , |
| 24 | SNMPTN→(B+) : 0.59 | 15 | (FKIP)→SWADANA :0.3 |
| 25 | III→(B+),SNMPTN : 0.58 | | (B+) , (FEB)→V :0.29 |
| 26 | III→(B+) : 0.58 | 16 | (B) , (FKIP)→4 TH :0.29 |
| 27 | 4 TH→(B+) : 0.56 | 17 | (FISIP) , 4 TH→III :0.29 |
| 28 | SNMPTN→4 TH : 0.5 | | (B) , (FKIP)→V :0.29 |
| 29 | III,SNMPTN→4 TH : 0.5 | 18 | (FISIP) , 4 |
| 30 | (B+),SNMPTN→4 TH : 0.5 | 19 | TH→SNMPTN :0.29 |
| 31 | III→4 TH : 0.5 | 21 | (B+) , |
| 32 | SNMPTN→4 TH,III : 0.49 | | (FEB)→SWADANA :0.29 |
| 33 | III→4 TH,SNMPTN : 0.49 | 22 | 3 TH , |
| 34 | (B+)→4 TH : 0.48 | | SWADANA→V :0.28 |
| 35 | 4 TH→(B+),SNMPTN :<br>0.43 | 23 | (FP) , III→SNMPTN :0.27<br>5 TH , |
| 36 | (B+)→4 TH,SNMPTN :<br>0.37 | 24 | SWADANA→V :0.23 |
| 37 | III,SNMPTN→(B) : 0.32 | 25 | |
| 38 | SNMPTN→(B) : 0.32 | | |
| 39 | SNMPTN→(B),III : 0.31 | | |

| | | | |
|---|---|---|---|
| 40 | III→(B) : 0.31 | | |
| 41 | III→(B),SNMPTN : 0.31 | | |

The highest score of information can be obtained from table 3 as follow:

(i). ARAO produces fewer rules than ARA. The rules generated by ARAO are 25 rules while ARA has 41 rules.

(ii). Similar with the result of table 2. In ARA, a combination of item "V→ SWADANA" subjectively is not an interesting rule. In ARAO, one item appears "(FEB)" as an additional item from the combination of items "V and "SWADANA". Hence, it becomes more attractive.

TABLE IV. THE RESULT OF EXPERIMENT 3

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | LOW→S3 : 0.85* | 1 | Comedy , LOW→S3 :0.82 * |
| 2 | S3→LOW : 0.82 | 2 | (R) , LOW→S3 :0.75 |
| 3 | (PG-13),S3→LOW : 0.79 | 3 | (PG) , LOW→S3 :0.65 |
| 4 | (PG-13),LOW→S3 : 0.78 | | |
| 5 | Action,LOW→S3 : 0.78 | | |
| 6 | Action,S3→LOW : 0.73 | | |
| 7 | (PG-13)→S3 : 0.71 | | |
| 8 | (PG-13)→LOW : 0.71 | | |
| 9 | Action→S3 : 0.71 | | |
| 10 | Action→LOW : 0.67 | | |
| 11 | Action→(PG-13) : 0.63 | | |
| 12 | (PG-13)→LOW,S3 : 0.56 | | |
| 13 | LOW→(PG-13) : 0.56 | | |
| 14 | S3→(PG-13) : 0.53 | | |
| 15 | Action→LOW,S3 : 0.52 | | |
| 16 | LOW,S3→(PG-13) : 0.51 | | |
| 17 | (PG-13)→Action : 0.5 | | |

The highest score of information can be obtained from table 4 as follow:

(i). ARAO produces fewer rules than ARA. The rules that ARAO produces are 3 rules while ARA is 17 rules.

(ii). In ARA, the highest score combination of item is "LOW→S3". Meanwhile, a more information with the addition term Comedy is obtained from ARAO, "Comedy , LOW→S3". Hence, it is more interesting

TABLE V. THE RESULT OF EXPERIMENT 4

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | LOW→S3 : 0.85* | 1 | Comedy , LOW→S3 :0.82 * |
| 2 | S3→LOW : 0.82 | 2 | (R) , LOW→S3 :0.75 |
| 3 | (PG-13),S3→LOW : 0.79 | 3 | (PG) , LOW→S3 :0.65 |
| 4 | (PG-13),LOW→S3 : 0.78 | 4 | (R) , Action→LOW :0.37 |
| 5 | Action,LOW→S3 : 0.78 | 5 | (R) , Action→S3 :0.36 |
| 6 | Action,S3→LOW : 0.73 | 6 | Adventure , LOW→S3 :0.36 |
| 7 | (PG-13)→S3 : 0.71 | 7 | (PG),Adventure→LOW :0.35 |
| 8 | (PG-13)→LOW : 0.71 | 8 | Drama , LOW→S3 :0.3 |
| 9 | Action→S3 : 0.71 | 9 | (PG) , Adventure→S3 :0.28 |
| 10 | Action→LOW : 0.67 | | |
| 11 | Action→(PG-13) : 0.63 | | |
| 12 | (PG-13)→LOW,S3 : 0.56 | | |
| 13 | LOW→(PG-13) : 0.56 | | |
| 14 | S3→(PG-13) : 0.53 | | |
| 15 | Action→LOW,S3 : 0.52 | | |
| 16 | LOW,S3→(PG-13) : 0.51 | | |
| 17 | (PG-13)→Action : 0.5 | | |
| 18 | LOW→(PG-13),S3 : 0.44 | | |
| 19 | S3→Action : 0.43 | | |
| 20 | S3→(PG-13),LOW : 0.42 | | |
| 21 | LOW→Action : 0.42 | | |
| 22 | LOW,S3→Action : 0.39 | | |
| 23 | LOW→Action,S3 : 0.33 | | |
| 24 | S3→Action,LOW : 0.32 | | |

The highest score of information can be obtained from table 5 as follow:

(i). ARAO produces fewer rules than ARA. The rules that ARAO produces are 9 rules while ARA has 24 rules.

(ii). Similar with the result of table 4. In ARA, the highest score combination of item is "LOW→S3". Meanwhile, ARAO obtained "Comedy , LOW→S3". Hence, it is more interesting

TABLE VI. THE RESULT OF EXPERIMENT 5

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | (B+)→III : 0.79 | 1 | (FSRD) , 5 TH→III :0.4 |
| 2 | (FISIP)→III : 0.68 | 2 | (B) , (FISIP)→5 TH :0.37 |
| 3 | 4 TH→III : 0.64 | 3 | (B-) , (FSRD)→5 TH :0.34 |
| 4 | 5 TH→III : 0.63 | 4 | (FISIP) , 5 TH→III :0.34 |
| 5 | (B)→III : 0.62 | 5 | (B+) , 4 TH→III :0.33 |
| 6 | (FISIP)→(B) : 0.57 | 6 | (FISIP) , 4 TH→III :0.32 |
| 7 | 4 TH→(B) : 0.56 | 7 | (B+) , (FISIP)→III :0.32 |
| 8 | (B)→5 TH : 0.5 | | |
| 9 | 5 TH→(B) : 0.49 | | |
| 10 | III→5 TH : 0.48 | | |
| 11 | III→(B) : 0.47 | | |
| 12 | (B)→(FISIP) : 0.42 | | |
| 13 | (B)→4 TH : 0.4 | | |
| 14 | III→(FISIP) : 0.38 | | |
| 15 | III→4 TH : 0.35 | | |
| 16 | III→(B+) : 0.33 | | |

The highest score of information can be obtained from table 6 as follow:

(i) ARAO produces fewer rules than ARA. The rules generated by ARAO are 7 rules while ARA has 16 rules.

(ii). ARA obtained the rule "(B +) → III" which is also found in the 7th place of ARAO. Meanwhile, ARAO obtained a more interesting rule, "(FSRD) , 5 TH→III", because the other term "FSRD" which is absent in ARA, is obtained in ARAO.

TABLE VII. THE RESULT OF EXPERIMENT 6

| No | Apriori | No | ARAO |
|---|---|---|---|
| 1 | (FISIP)→III : 0.68 | 1 | (FSRD) , 5 TH→III :0.4 |
| 2 | 5 TH→III : 0.63 | 2 | (B) , (FISIP)→5 TH :0.37 |
| 3 | (B)→III : 0.62 | 3 | (B-) , (FSRD)→5 TH :0.34 |
| 4 | (B)→5 TH : 0.5 | 4 | (FISIP) , 5 TH→III :0.34 |
| 5 | 5 TH→(B) : 0.49 | 5 | (B+) , 4 TH→III :0.33 |
| 6 | III→5 TH : 0.48 | 6 | (FISIP) , 4 TH→III :0.32 |
| 7 | III→(B) : 0.47 | 7 | (B+) , (FISIP)→III :0.32 |
| 8 | III→(FISIP) : 0.38 | | |

The highest score of information can be obtained from table 7 as follow:

(i). ARAO produces fewer rules than ARA. The rules generated by ARAO are 7 rules while ARA has 8 rules.

(ii). Similar with the result of table 6. ARA obtained the rule "(FISIP)→III" which is also found in the 4th place of ARAO. Meanwhile, ARAO obtained a more interesting rule, "(FSRD) , 5 TH→III", because the other term "FSRD" is obtained in ARAO.

TABLE VIII. THE RESULT OF EXPERIMENT 7

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | (B-)→5 TH : 0.85 | 1 | (FSRD) , 5 TH→III :0.4 |
| 2 | (B+)→III : 0.79 | 2 | (B) , (FISIP)→5 TH :0.37 |
| 3 | (FSRD)→III : 0.77 | 3 | (B-) , (FSRD)→5 TH :0.34 |
| 4 | (FSRD)→5 TH : 0.77 | 4 | (FISIP) , 5 TH→III :0.34 |
| 5 | (FISIP)→III : 0.68 | 5 | (B+) , 4 TH→III :0.33 |
| 6 | 4 TH→III : 0.64 | 6 | (FISIP) , 4 TH→III :0.32 |
| 7 | (B),5 TH→III : 0.64 | 7 | (B+) , (FISIP)→III :0.32 |
| 8 | 5 TH→III : 0.63 | | |
| 9 | (B)→III : 0.62 | | |
| 10 | (FISIP)→(B) : 0.57 | | |
| 11 | V→(B) : 0.56 | | |
| 12 | 4 TH→(B) : 0.56 | | |
| 13 | V→5 TH : 0.56 | | |
| 14 | (B),III→5 TH : 0.52 | | |
| 15 | 5 TH,III→(B) : 0.5 | | |
| 16 | (B)→5 TH : 0.5 | | |
| 17 | 5 TH→(B) : 0.49 | | |
| 18 | (FISIP)→5 TH : 0.49 | | |
| 19 | III→5 TH : 0.48 | | |
| 20 | III→(B) : 0.47 | | |
| 21 | (B)→(FISIP) : 0.42 | | |
| 22 | (B)→4 TH : 0.4 | | |
| 23 | (B)→V : 0.38 | | |
| 24 | III→(FISIP) : 0.38 | | |
| 25 | 5 TH→V : 0.37 | | |
| 26 | III→4 TH : 0.35 | | |
| 27 | 5 TH→(FISIP) : 0.35 | | |
| 28 | III→(B+) : 0.33 | | |
| 29 | 5 TH→(FSRD) : 0.33 | | |
| 30 | 5 TH→(B-) : 0.33 | | |
| 31 | (B)→5 TH,III : 0.32 | | |
| 32 | 5 TH→(B),III : 0.31 | | |
| 33 | III→(FSRD) : 0.26 | | |
| 34 | III→(B),5 TH : 0.24 | | |

The highest score of information can be obtained from table 8 as follow:

(i). ARAO produces fewer rules than ARA. The rules generated by ARAO are 7 rules while ARA has 34 rules.

(ii). Similar with the result of table 7. ARA obtained the rule "(B-) → 5 TH" which is also found in the 3rd place of ARAO. Meanwhile, ARAO obtained a more interesting rule, "(FSRD) , 5 TH→III", because the other term "FSRD" is obtained in ARAO.

TABLE IX. THE RESULT OF EXPERIMENT 8

| No | ARA | No | ARAO |
|---|---|---|---|
| 1 | (LOW)→S3 : 0.7 | 1 | (VERY_LOW), Comedy→S3 :0.3 |
| 2 | Action→S3 : 0.69 | | |
| 3 | (PG-13)→S3 : 0.67 | 2 | (PG-13),(VERY_LOW)→ Comedy:0.24 |
| 4 | S3→(LOW) : 0.65 | | |
| 5 | (R)→(LOW) : 0.64 | 3 | (MEDIUM) , Action→S3 :0.23 |
| 6 | (R)→S3 : 0.63 | 4 | (PG) , Adventure→S3 :0.22 |
| 7 | (PG-13)→(LOW) : 0.6 | 5 | (MEDIUM),(PG-13)→Action :0.21 |
| | | 6 | (MEDIUM), (PG-13)→S3 :0.21 |

The highest score of information can be obtained from table 9 as follow:

(i). ARAO produces fewer rules than ARA. The rules that ARAO produces are 6 rules while ARA has 7 rules.

(ii). ARA obtained "(LOW) → S3". Whereas, ARAO obtained "(VERY_LOW), Comedy→S3"

(iii). The combination of items produced by ARAO appears several items that are not in ARA, namely: "MEDIUM", "VERY LOW", "Adventure", "Comedy".

TABLE X. THE COMPARISON OF THE NUMBER OF THE OBTAINED RULES

| Experiment | Apriori | Apriori with Ontology |
|---|---|---|
| 1 | 28 | 5 |
| 2 | 41 | 24 |
| 3 | 17 | 3 |
| 4 | 24 | 9 |
| 5 | 8 | 7 |
| 6 | 16 | 7 |
| 7 | 34 | 7 |
| 8 | 7 | 6 |

From experimental results 1 to 8 can be concluded:

- ARAO generates fewer rules than ARA as explained in Table 10.
- The use of minimum support, confidence, and IR parameters and the amount of data used can affect the number of generated rules.
- Rules that have high confidence or IR value will always appear even if the minimum parameter value is changed.
- The results of experiments 1 and 2 return pretty similar ruless as well as experiments 3 and 4. For the experimental 1 and 2, the most top rule is "III → SNMPTN". For the experiment 3 and 4, the highest rule is "LOW → S3" (rating at level 3). Until the experiment 4 can be concluded that Apriori with ontology reduces trivial rules (have a meager score). Therefore the number of obtained rules is diminished.
- The experimental 5,6 and 7 show that the result of combining items using Apriori with ontology is more consistent although the minimum support values are various: 0.25.0.20, and 0.15. Also, the testing of the default of the Apriori algorithm with a minimum value of support of 0.25 and 0.20 there are items still trimmed, in example the item "FSRD", and only appear when the minimum condition of support is 0.15. Whereas in Apriori with an ontology for "FSRD" item always seems at the minimum value of support equal to 0.25 and 0.20.
- Rules that have consequence non-item "III" in experimental 5,6 and 7 the result of the default Apriori algorithm shows that the rules which have 2 until 16 rules. Whereas, in Apriori with ontology has 2 rules with a combination of more complex items. It makes easier in the process of tracking information.
- From experiment 8, it is also can see, in case that to reduce the number of the obtained rule only by increasing the minimum support and confidence as has shown in the default Apriori algorithm, the

obtained rules are not interesting. Comparing to the obtained rules of Apriori with ontology.

## V. CONCLUSION AND FUTURE WORK.

This research has proposed the utilization of ontology to improve the result of association rule of Apriori algorithm. The ontology is a simply hierarchical ontology. The IR is calculated based on three itemsets. From the obtained results of this study, the utilization of ontology in Apriori can facilitate the tracking of information because it produces fewer amount rules compared to the default Apriori algorithm. The obtained rules of Apriori with ontology are generally also more interesting compare to the result of the default Apriori algorithm.

In the near future study is expected to use more complex ontology which perhaps influences the formula of IR. The use of more efficient IR formulas and measurements of the level of interrelationship of more item combinations, is also the task which should be solved soon.

## REFERENCES

[1] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.

[2] R. Agarwal, R. Srikant, and others, “Fast algorithms for mining association rules,” in Proc. of the 20th VLDB Conference, 1994, pp. 487–499.

[3] S. A. Mahmoodi, K. Mirzaie, and S. M. Mahmoudi, “A new algorithm to extract hidden rules of gastric cancer data based on ontology,” SpringerPlus, vol. 5, no. 1, p. 312, 2016.

[4] T. Berners-Lee, J. Hendler, O. Lassila, and others, “The semantic web,” Scientific american, vol. 284, no. 5, pp. 28–37, 2001.

[5] B. Swartout, R. Patil, K. Knight, and T. Russ, “Toward distributed use of large-scale ontologies,” in Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems, 1996, pp. 138–148.

[6] I. N. Ferraz and A. C. B. Garcia, “Ontology in association rules,” SpringerPlus, vol. 2, no. 1, p. 452, 2013.

[7] V. Svátek, J. Rauch, and M. Ralbovskỳ, “Ontology-enhanced association mining,” in Semantics, Web and mining, Springer, 2006, pp. 163–179.

[8] P. Ding, S. Jun-yi, and Z. Mu-xin, “Incorporating domain knowledge into data mining process: An ontology based framework,” Wuhan University Journal of Natural Sciences, vol. 11, no. 1, pp. 165–169, 2006.

[9] T. Bharathi and A. Nithya, “Enhanced Way of Association Rule Mining With Ontology,” International Journal Of Engineering And Computer Science, vol. 5, no. 10, 2016.

[10] B. S. Balasubramani, V. R. Shivaprabhu, S. Krishnamurthy, I. F. Cruz, and T. Malik, “Ontology-based urban data exploration,” in Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics, 2016, p. 10.