

Family Economic Correlation To Students Learning Achievement Using Apriori Method

Nurhayati^{1*}, Juliana Naftali Sitompul², Tri Kartika Sari³

^{*1,2,3}Program of Technical Information, STMIK Kaputama

*Corresponding author E-mail: nurhayatiazura059@gmail.com

Abstract

The education system in Indonesia as mandated in the GBHN aims to educate the nation while at the same time responding to new challenges to create a decent and prosperous life. Understanding, appreciation, and experience of cultural and religious values in the right and true form will be increasingly needed. The economic status of the family is one of the factors that is sufficient to support the level of continuing education, especially for teenagers who are still student in school. Apriori method is used to obtain association rules that describe the relationship between item in the transactional database. There are two databases used, each of which has a different number of transactions. This study aims to apply the apriori algorithm, as an analytical technique. The data taken as a case example is family economic data. This association search uses WEKA which will later find the rules and MySQL as the placeholder for the Database. From the results of the analysis using apriori, the highest confidence value was obtained at 0.9 with support 0.1 resulting in a students rule whose economics supported the learning achievement was very supportive, and the lowest confidence value of 0.2 with support 0.1 resulted in a students rule who had sufficient economics, so their learning achievement was also quite increased.

Keywords: Apriori Algorithm, Association Rule, Correlation, Minsupport, Min Confidence.

1. Introduction

The education system in Indonesia as mandated in the GBHN aims to educate the nation while at the same time responding to new challenges to create a decent and prosperous life. Understanding, appreciation and experience of cultural and religious values in the right and correct form will be increasingly felt needed[1].

Starting from the situation above, education is one of the most important tools to achieve this goal, especially for students as a milestone in the relay of the nations next generation. The economic status of the family is one of the factors that is quite decisive for the continuation of the education and behavior of children, especially teenagers who are still student in school [2].

The life conditions of students at the age of puberty are colored with various kinds of problems, because at this age they will bring negative traits. In this period of adolescence many needs that they want to get, so here the role of parents is very important, in other words their needs must be met. If their families economic needs do not support it, two possibilities will occur, namely from the positive and negative side, for example, if their families economy is supportive or sufficient, student learning achievement will increase and vice versa if the family economy is not supportive or sufficient, student learning achievement will automatically decrease.

Based on the research journal [3] said that the apriori can be further developed to further improve performance and performance in obtaining association rule.

2. Material

2.1 Correlation

Correlation (relationship between variables) of interest. Here two aspects of correlation analysis will be highlighted, namely whether the available sample data provide sufficient evidence that there is a relationship between the variables in the samples original population[4]. And second, if there is a relationship, how strong is the relationship between the variables. The closeness of the relationship is expressed by the name of the correlation coefficient (or can be called just correlation).

Correlation research is one part of ex-postfacto research. One part of the researcher does not manipulate the state of the existing variables and directly looks for the existence of the relationship and the level of the variables defined in the correlation coefficient.

Correlation into descriptive research because the research is an attempt to describe the correlation that has occurred in this study. The researcher tries to describe the current condition in the quantitative context defined in the variable.

2.2 Algoritma A Priori

The apriori algorithm is a type of association rule in data mining. In addition to apriori, what is included in this group are generalized rule induction and hash based methods. Rules that state association between multiple attributes are often affinity or market basket analysis[5]. Association analysis became famous for its application to analyze the contents of shopping carts in supermarkets. Association analysis is also known as one of the data mining techniques which is the basis of various other data mining techniques. In

particular, one of the stages of association analysis called high frequency pattern analysis (frequent pattern mining) has attracted the attention of many researchers to produce efficient algorithms. The importance of an associative rule can be determined by two parameters, support (supporting value) which is the percentage of the combination of these items in the database and confidence (certainty value) which is the percentage of the combination of these items in the database and confidence (certainty value) which is the strength of the relationship between items in the associative rule[6]. The associative rule is usually expressed in the form: {bread, butter} → {milk} (support=40%, confidence=50%). The rule means: "50% of transaction in the database containing bread and butter items also contain dairy items. Meanwhile, 40% of all transaction in the database contain those three items. "it can also be interpreted: "A consumer who buys bread and butter has a 50% chance of also buying milk. This rule is quite significant because it represents 40% of the transaction records so far." Association analysis is defined as a process to find all associative rules that meet the minimum requirements for support (minimum support) and minimum requirements for confidence (minimum confidence)[7].

This stage is looking for a combination of items that meet the minimum requirements of the support value in the database[1]. The support value of an item is obtained by the following formula :

$$\text{support} = \frac{\text{Jumlah Transaksi mengandung } A}{\text{Total Transaksi}} \quad (1)$$

While the support value of the 2 items is obtained from the following formula :

$$\text{Support}(A,B) = P(A \cap B) \quad (2)$$

$$\text{Support}(A,B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \quad (3)$$

3. Methodology

3.1. Research Stages

Systematic steps are organized as illustrated in Apriori to make it easier in performing research.

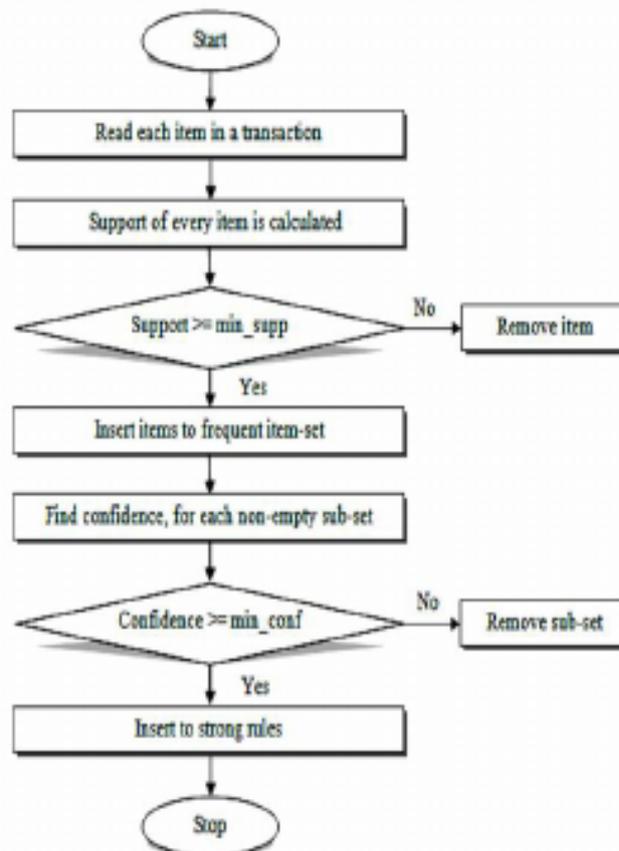


Fig. 1: K-Medoids algorithm

Data mining is the technique to find hidden patterns from a very large volume of historical data. Association rule is a type of data mining that correlates one set of items or events with another set of items or events. Another data mining strategy is clustering technique. This technique is used to create partitions so that all members of each set.

3.2. Training Data

accurate clustering of transactional data has many potential applications in retail industry, ecommerce intelligence, etc. Here, the term “large items” refers to the items contained in some minimum fraction of transactions in a cluster and is used as similarity measure of a cluster of transactions. The support of an item in cluster C_i is the number of transactions in C_i . Thus, for a minimum support s , an item is large in cluster C_i if its support is at least equal to $s \times C_i$, otherwise item is small. Thus, large items measure similarity in a cluster while small items measure dissimilarity. Two components of cost C are to be minimized consists of: the intra-cluster cost and the inter-cluster cost. The intra-cluster cost consists of the total number of small items and the inter-cluster cost measures the duplication of large items in different clusters. This clustering algorithm helps to minimize large items and small items cost.

Table 1: Achievement value data

siswa	nilai ekonomi
1	8, Mendukung
2	8, Mendukung
3	6, Cukup
4	8, Mendukung
5	8, Mendukung
6	9, Mendukung
7	7, Cukup
8	9, Mendukung
9	8, Mendukung
10	6, Cukup

Table 2. Transaction Data Tabular Format

No	6	7	8	9	Cukup	Mendukung
1	0	0	1	0	0	1
2	0	0	1	0	0	1
3	1	0	0	0	1	0
4	0	0	1	0	0	1
5	0	0	1	0	0	1
6	0	0	0	1	0	1
7	0	1	0	0	1	0
8	0	0	0	1	0	1
9	0	0	1	0	0	1
10	1	0	0	0	1	0

4. Results

4.1. Association Analysis

we can the frequency of itemset from the table above is known total Φ fro loan transactions $k = 1$, everything is more than Φ so:

$F1 = \{\{6\}, \{7\}, \{8\}, \{9\}\}$ to $k = 2$ (2 element), given for each pair of items. The possible sets are : $\{6,7\}, \{6,8\}, \{6,9\}, \{7,8\}, \{7,9\}, \{8,8\}, \{8,9\}$. The tables for the candidate 2 item set :

Table 3 : candidate 2-itemset

No	6	7	F	No	6	8	F	No	6	9	F
1	0	0	f	1	0	1	t	1	0	0	F
2	0	0	f	2	0	1	t	2	0	0	F
3	1	0	t	3	1	0	t	3	1	0	T
4	0	0	f	4	0	1	t	4	0	0	F
5	0	0	f	5	0	1	t	5	0	0	F
6	0	0	f	6	0	0	f	6	0	1	T
7	0	1	t	7	0	0	f	7	0	0	F
8	0	0	f	8	0	0	f	8	0	1	T
9	0	0	f	9	0	1	t	9	0	0	F
10	1	0	t	10	1	0	t	10	1	0	T
		Σ	7			Σ	3			Σ	6

No	7	8	F	No	7	9	F	No	8	9	F
1	0	1	t	1	0	0	f	1	1	0	T
2	0	1	t	2	0	0	f	2	1	0	T
3	0	0	f	3	0	0	f	3	0	0	F
4	0	1	t	4	0	0	f	4	1	0	T
5	0	1	t	5	0	0	f	5	1	0	T

6	0	0	f	6	0	1	t	6	0	1	T
7	1	0	t	7	1	0	t	7	0	0	F
8	0	0	f	8	0	1	t	8	0	1	T
9	0	1	t	9	0	0	f	9	1	0	T
10	0	0	f	10	0	0	f	10	0	0	F
		Σ	4			Σ	7			Σ	3

From the tables of the 2 element above, T means items that are used together, while F means that no items are used together or there is no transaction. Σ represents the number of item set frequencies. The number of item set frequencies must be greater then or equal to the number of item set frequencies ($\Sigma \geq \Phi$). From the table above, it is obtained : $F2 = \{\{6,7\}, \{6,9\}, \{7,9\}, \{6,8\}, \{7,8\}, \{8,9\}\}$. And the rule that will be used is if x then y, where x is antecedent and y is consequent. Based on this rule, it takes 2 items, one of which is as antecedent and the rest as consequent. From the steps above, we get 1 Fk, namely F2. F1 not included because it only consists of 1 item. The antecedent may have more than 1 element, while the consequent may consist of 1 element. Determine (ss-S) as the antecedent and s as the consequent of Fk that has been obtained based on the rule in the step above. In the step above. In F2, we get the set $F2 = \{\{6,7\}, \{6,9\}, \{7,9\}, \{6,8\}, \{7,8\}, \{8,9\}\}$.

Then it can be arranged :

to {6,7}

- a. If (ss-s) = 6 If s = 7 then enough
- b. If (ss-s) = 7 If s = 6 then enough

to {6,9}

- If (ss-s) = 6 If s = 9 then enough
- If (ss-s) = 9 If s = 6 then enough

to {7,9}

- If (ss-s) = 7 If s = 9 then enough
- If (ss-s) = 9 If s = 7 then enough

to {6,8}

- If (ss-s) = 6 If s = 8 then enough
- If (ss-s) = 8 If s = 6 then enough

to {7,8}

- If (ss-s) = 7 If s = 8 then enough
- If (ss-s) = 8 If s = 7 then enough

to {8,9}

- If (ss-s) = 8 If s = 9 then support
- If (ss-s) = 9 If s = 8 then support

From the steps above, gets 12 rule that can be used, namely:

- If buy 6 then buy 7
- If buy 7 then buy 6
- If buy 6 then buy 9
- If buy 9 then buy 6
- If buy 7 then buy 9
- If buy 9 then buy 7
- If buy 6 then buy 8
- If buy 8 then buy 6
- If buy 7 then buy 8
- If buy 8 then buy 7

count support and confidence

$$SUPPORT = \frac{\Sigma \text{ Transaksi yang digunakan}}{\Sigma \text{ Jumlah Seluruh Transaksi}} \times 100\% \tag{4}$$

$$CONFIDANCE = \frac{\Sigma \text{ Transaksi yang digunakan}}{\Sigma \text{ Jumlah Transaksi pada bagian Antecedent}} \times 100\% \tag{5}$$

For items that are borrowed at once in if buy A then buy B, there are 6 transactions, the total number of transaction is 20 transactions, so the support is :

$$SUPPORT = \frac{6}{20} \times 100\% = 0.3\% \tag{6}$$

For items used in if buy 6 then buy 7, there are 5 while the number of transactions using 6 is 3 transactions, so the confidence is :

$$CONFIDANCE = \frac{5}{6} \times 100\% = 83.4\% \tag{7}$$

Table 4. Association rules

If antecedence and consequent	Support	Confidence
If buy 6 then buy 7	(6/20)x 100% = 0.3%	(5/6) x 100% = 83.4%
If buy 7 then buy 6	(6/20)x 100% = 0.3%	(5/5) x 100% = 100%
If buy 6 then buy 9	(6/20)x 100% = 0.3%	(5/4) x 100% = 1.25%

If buy 9 then buy 6	$(6/20) \times 100\% = 0.3\%$	$(5/5) \times 100\% = 100\%$
If buy 7 then buy 9	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 9 then buy 7	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 6 then buy 8	$(6/20) \times 100\% = 0.3\%$	$(5/8) \times 100\% = 62.5\%$
If buy 8 then buy 6	$(6/20) \times 100\% = 0.3\%$	$(5/5) \times 100\% = 100\%$
If buy 7 then buy 8	$(6/20) \times 100\% = 0.3\%$	$(5/8) \times 100\% = 62.5\%$
If buy 8 then buy 7	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$

4.2 Association Rules

After getting support and confidence for each candidate, do the multiplication between support and confidence, so you get the following table :

Table 5. Association rules final

If antecedence and consequent	Support	Confidence
If buy 6 then buy 7	$(6/20) \times 100\% = 0.3\%$	$(5/6) \times 100\% = 83.4\%$
If buy 7 then buy 6	$(6/20) \times 100\% = 0.3\%$	$(5/5) \times 100\% = 100\%$
If buy 6 then buy 9	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 9 then buy 6	$(6/20) \times 100\% = 0.3\%$	$(5/5) \times 100\% = 100\%$
If buy 7 then buy 9	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 9 then buy 7	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 6 then buy 8	$(6/20) \times 100\% = 0.3\%$	$(5/8) \times 100\% = 62.5\%$
If buy 8 then buy 6	$(6/20) \times 100\% = 0.3\%$	$(5/5) \times 100\% = 100\%$
If buy 7 then buy 8	$(6/20) \times 100\% = 0.3\%$	$(5/8) \times 100\% = 62.5\%$
If buy 8 then buy 7	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 8 then buy 9	$(6/20) \times 100\% = 0.3\%$	$(5/4) \times 100\% = 1.25\%$
If buy 9 then buy 8	$(6/20) \times 100\% = 0.3\%$	$(5/8) \times 100\% = 62.5\%$

Then it can be read as follows :

if the family economy is supportive, learning achievement will increase, if the family economy is sufficient, the learning achievement is quite supportive.

This is done to find rules that can later be used for a recommendation system for student interest areas. The trial was carried out with a confidence range of 0.9 and 0.8 confidence. Then from the observations it will be found that the smaller the confidence, the more rules will appear. This happens because confidence is a condition for choosing or rejecting a rule, if the confidence value is lowered, it will provide more opportunities for it to be found.

Minimum support = 0.1 and confidence = 0.9, found a rule with 3 item set=2, and 2 item set=4 a total of 6 rules, namely :

achievement=75 caption=Enough 8 ==> economy=7 8 conf:(1)

This means that based on student data in the database that 100% of students who are economically sufficient, their learning achievement is also quite increased, this is supported by a minimum of 10% of transaction data in the database containing sufficient achievement, sufficient supporting information, and sufficient economy.

Economy = 7 30 ==> description=enough 30 [conf:\(1\)](#). This means that based on student data in the database, 100% of students who have sufficient economics with sufficient information support learning, this is supported by a minimum of 10% of transaction data in the database containing sufficient economics with sufficient supporting information.

Minimum support=0.1 and confidence=0.8, found the same rule with confidence=0.9.

Minimum support=0.1 and confidence=0.7, found a rule with 3 item set-1, and 2 item set=2 a total of 3 rules, namely :

Economy=7.30==> description= enough 30 [conf:\(1\)](#). This means that based on student data in the database, 100% of student who have sufficient economics with sufficient support learning, this is supported by a minimum of 10% of transaction data in database containing sufficient economics with sufficient supporting information.

Economy=8 21 ==> description = support 21 [conf:\(1\)](#). This means that based on student data in the database that 100% of students who are economically supportive with information supporting learning, this is supported by a minimum of 10% of transaction data in the database containing economic support with information support learning.

Minimum support=0.1 and confidence=0.6, found a rule with 3 item set=1, and 2 item set=2 total of 3 rules, namely :

Economy=7 30 ==> description= enough 30 [conf:\(1\)](#). This means that based on student data in the database, 100% of students who have sufficient economics with sufficient information support by a minimum of 10% of transaction data in the database containing sufficient economics with sufficient supporting information.

Economy=8 21 ==> description= supports 21 [conf:\(1\)](#)

This means that based on student data in the database that 100% of students who are economically supportive with information supporting learning, this is supported by a minimum of 10% of transaction data in the database containing economic support with information support learning.

Minimum support=0.1 and confidence=0.5, found the same rule with confidence=0.6.

Minimum support=0.1 and confidence=0.4, found the same rule with confidence=0.5.

Minimum support=0.1 and confidence=0.3, found a rule with 2 item set =4 rules, namely :

Economy=7 30 ==> description= enough 30 [conf:\(1\)](#)

This means that based on student data in the database, 100% of students who have sufficient economics with sufficient information support learning, this is supported by a minimum of 10% of transaction data in the database containing sufficient economics with sufficient supporting information.

Economy=9 13 ==> description= support 13 [conf:\(1\)](#)

This means that based on student data in the database that 100% of students who are economically supportive with information supporting learning, this is supported by a minimum of 10% of transaction data in the database containing economic support with information supporting learning.

This means that based on student data in the database that 100% of students who are economically supportive with information supporting learning, this is supported by a minimum of 10% of transaction data in the database containing economic support with information supporting learning.

Minimum support=0.1 and confidence=0.2, found the same rule with confidence=0.4.

The patterns found in this technique can be presented in the form of association rules. Rule and confidence are important in the process of obtaining a rule, in general the conditions are met. If the minimum support and minimum confidence are met. Support count is the number of the same item set appearing simultaneously in a transaction data in the shopping cart. Confidence is a measure that shows the relationship between two transactions. Based on the support value with the support value varying between 0.9% and 0.2%, it can be concluded that the smaller the support value, the better learning achievement, the greater the number of item sets produced, and the more association rules formed.

5. Conclusion

From the results of observations during the trial and evaluation process that has been carried out, several conclusions can be drawn as follows the highest confidence value was found at a value of 90%, so the family economy strongly supports student learning achievement, With this apriori method, it can help make it easier for schools to input family economic data.

From the results of this study, it can be seen that the family economy is related to student achievement. This is evidenced by the results of research where the family economy on student achievement greatly increases learning achievement.

Acknowledgement

We would like to thank STMIK Kaputama for supporting the completion and publication of this research.

References

- [1] R. Ruswati, A. I. Gufroni, and R. Rianto, "Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm," *Sci. J. Informatics*, vol. 5, no. 2, pp. 91–104, 2018, doi: 10.15294/sji.v5i2.16199.
- [2] T. Marnoto, "Drying of Rosella (Hibiscus sabdariffa) Flower Petals using Solar Dryer with Double Glass Cover Collector," *Int. J. Sci. Eng.*, vol. 7, no. 2, pp. 155–160, 2014, doi: 10.12777/ijse.7.2.150-154.
- [3] R. Rusdiansyah, N. Suharyanti, T. Triningsih, and M. Darussalam, "Application Of Pizza Sales Data Mining Using Apriori Method," *Sinkron*, vol. 4, no. 2, p. 1, 2020, doi: 10.33395/sinkron.v4i2.10500.
- [4] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.
- [5] A. Azwar, "Algorithm Apriori Use for a Consumer Behavior in," *Sains dan Inform.*, vol. 1, pp. 45–59, 2015.
- [6] R. Mustofa and Irhamah, "Topic Discovery pada Jurnal-jurnal di IEEE Explore menggunakan Association Rule Mining dengan Pendekatan Closed Frequent Itemset," *Ejurnal.Its.Ac.Id.*, vol. 8, no. 2, 2019, [Online]. Available: http://www.ejurnal.its.ac.id/index.php/sains_seni/article/view/43653.
- [7] W. W. Ariestya, W. Supriyatin, and I. Astuti, "Marketing Strategy for the Determination of Staple Consumer Products Using Fp-Growth and Apriori Algorithm," *J. Ilm. Ekon. Bisnis*, vol. 24, no. 3, pp. 225–235, 2019, doi: 10.35760/eb.2019.v24i3.2229.