PAPER • OPEN ACCESS

Association Rule Implementation Using Algorithm Apriori To Analize Fishing Pattern In Indonesia

To cite this article: Titin Kristiana *et al* 2020 *J. Phys.: Conf. Ser.* **1641** 012072

View the article online for updates and enhancements.

Association Rule Implementation Using Algorithm Apriori To Analize Fishing Pattern In Indonesia

Titin Kristiana[1],Sukmawati Anggraeni Putri[2], Nurmalasari[3], Rani Irma Handayani[4],Nita Merlina[5], Norma Yunita[6]

Program Studi Sistem Informasi, Sekolah Tinggi Manajemen dan Informatika,Jl Damai No .8 warung Jati Barat (Margasatwa) Jakarta Selatan

E-mail: `titin.tka@nusamandiri.ac.id`

Abstract. There are more than 80 species of fish caught by fishermen in the sea of Indonesia. To find out what kinds of fish mostly caught, it is necessary to analyse the data pattern of fish being caught. The activities of searching and associating the data pattern are closely related to data mining technique that being used to discover the rules of association of items. In this associative rule method, there are two process can be used: the process of generating frequent itemset and finding associative rules. The Frequent Itemset Generation is a process to get the connection of the itemset and the value of the association based on the value of support and confidence. The algorithm used to generate the frequent itemset is Apriori Algorithm. The Apriori Algorithm has a weakness in the extraction of the appropriate feature of the used attributes. This condition causes the rules formed in large number. This research applies Apriori Algorithm based on principal component analysis to obtain more optimal rules. After the experiments using the apriori algorithm applied with the magnitude $\phi = 30$, minimum Support 80% and Confidence 80%, the result of the rule formed are totally 82 rules.

Keywords: Catching Fish, Apriori algorithm, Association rule,

1. Introduction

As an Archipelago, Indonesia has huge and diverse fishery potential. Widely known that Indonesia has 17,508 islands with 104,000 km long coastline and 70% (5.8 million km2) ocean. (PUSDATIN, 2011). Utilization of marine resources stored in the territorial waters of the archipelago to date need to get adequate attention considering the amount of potential stored in marine resources. When compared with the total area of the oceans owned by Indonesia then the government's efforts in exploiting the potential of marine still needs a lot of improvement. Marine fisheries resources are one of the important aspects of marine resources that need to be explored and developed. The potential of fisheries in Indonesian waters is known to be abundant, considering Indonesia's waters are tropical regions with rich plankton content, as a gathering place for fish.

Fishing is an activity that aims to obtain fish in waters that are not cultivated by any means or means, including activities that use ships to build, transport, store, cool, or preserve. (Yaya Mulyana, 2008). It has been very well known that there are various types of fish, more than 80,00 species, caught by fishermen in the sea of Indonesia. To find out the information of the specific types of fish being caught, it is necessary to analyse the data pattern of the fish.

This searching o data pattern or associative relationships of large-scale data is closely related to data mining. Association analysis or association rule mining is a data mining technique for finding associative rules between combinations of items. One of the stages of association analysis is to produce an efficient algorithm with the analysis of the highest frequency pattern (Kusrini, 2009). In the association rule method, there are two processes; they are the Frequent Itemset generation process and the excavation of association rules. Generation Frequent Itemset is a process to get related itemsset, and it has a value of association based on the support and confidence value. The Algorithm used to generate this frequent itemset is the Apriori Algorithm.

Association rule using Apriori Algorithm is used for analysing the data pattern of fish caught.. (Tyas, 2008), Feature Selection for Large Scale Rule Mining and Information Gain uses a Hybrid Approach (Balamurugan P. r., 2009)

2. Methods
Research is an activity aimed at making an original contribution to science (Dawson, 2009). In the context of a study, the method approaches used to solve problems include: collecting data, formulating hypotheses or propositions, testing hypotheses, interpretation results, and conclusions that can be evaluated independently by others (Berndtsson, Olsson, & Lundell, 2008). Meanwhile, according to (Dawson, 2009) there are four research methods commonly used, including: Action Research, Experiment, Case Study, and Survey.

In the experimental research method, there are several stages of research conducted as follows:
1. Data collection
This section describes how and where the data in this study obtained, there are two types of data collection, namely primary data collection and secondary data collection. Primary data is data collected first to see what really happened. Secondary data is data previously made by someone either published or not (Kothari, 2004). Primary data collection can be obtained from interview models of respondents, as well as by observation model of a body under study (Gray, 2004). At this stage determined the data to be processed. Search for available data, obtain additional data required, integrate all data into data sets, including variables required in the process.
2. Preliminary data processing
This section describes the early stages of data mining. Preliminary data processing involves the process of inputting data into required formats, grouping and determining data attributes, and split data for use in the learning and training process.
3. The proposed model
At this stage the data is analyzed, grouped which variables are related to each other. After the data is analyzed and applied the models according to the data type. Data sharing into training data and test data is also required for modeling.
3. Experiments and model testing
In this section described the experimental steps include ways selection of appropriate architecture of the proposed model or method to obtain results that can prove that the method used is appropriate.
4. Evaluation and validation of results
In this section conducted evaluation and validation of the results of application of the research model conducted to determine the level of accuracy of the model. The amount of initial data obtained from data collection that is as much as 1,200 data, To obtain quality data, the data cleaned and transformed the desired shape so that it can be done in preparing the model. For that required techniques to filter data training or known as preprocessing techniques (Witten, Frank, & Hall, 2011). Where the results of preprocessing techniques produce different attributes. The following table results of preprocessing techniques ((Han & Kamber, 2007) are:
a. Data cleansing is used to clear empty values, inconsistent or missing data (missing value and

noisy).

b. Data integration is used to unify different storage places into the same storage area. In this case data is taken from the data warehouse that is customer data and sales transaction data.

c. Data reduction, where the amount of tuple data used for training data is too large and only a few attributes are required, so unnecessary attributes will be removed. If there is the same tuple then use only one tuple only.

Modeling stage to complete the fish catch pattern analysis using two methods namely a priori algorithm and principal component analysis algorithm

1. Apriori algorithm is a method to find frequent itemset run by a set of data

The model proposed in the research on fish catch pattern analysis is by applying a priori algorithm using principal component analysis. Application of a priori algorithm determines the itemset with support and confidence first. Then get the value of the ratio of the itemset that formed. While the a priori algorithm based on principal component analysis determines the largest cumulative variance of the transformation results so as to form a more efficient rule for solving the problem

After going through the cleaning process there are 437 records of fish catch data, with 31 species of fish to be used in association rules. In the table below can be seen 31 types of fish that will be used in association rules.

Table 1. attribute names used

No	Types of fish	No	Types of fish
1	Beronang kuning	16	Kerapu bebek
2	Beronang lingkis	17	Kerapu karang
3	Biji nangka	18	Kerapu karet
4	Gabus Laut	19	Kuniran
5	Gulamah/Tigawaja	20	Kurau
6	ikan gaji	21	Kurisi
7	Ikan kakap merah/Bambangan	22	Kuro/senangin
8	ikan lidah	23	Layur
9	Ikan sebelah (Terompa)	24	Lencam
10	Ikan sebelah mata kiri	25	Manyung
11	Jenaha	26	Pari burung
12	Kakap Batu	27	Pari hidung sekop
13	Kakap Merah	28	Pari kekeh
14	Kakap putih	29	Pari kembang/Pari macan
15	Kapas-kapas	30	Remang
		31	Swanggi

From this training data will be used to analyze the data. The algorithm steps in the association rule are:

1. Determine the magnitude of ϕ, the minimum value of support and the minimum value of confidence.

2. Determine all frequency itemsset frequently.

3. For each frequency itemset do the following:

a. Take an element, name it s

b. For the rest call ss-s

c. Enter the attributed elements into the rule if (ss-s) then s.

From training data to be used, first determined:

1. Determine the magnitude:

a. magnitude ϕ, ie = 30

b. the value of minimun support = 80%

c. the minimum value of confidence = 80%

2. Determine all frequencies of frequent itemsets from training data.

In this step, a set of frequent itemsets is prepared, ie a temset having a minimum set of itemset frequencies of $\phi = 30$, as defined in the first step, which begins with the following 1-item set:

F_1= {{ Beloso/buntu kerno}, {Beronang kuning}, {Beronang lingkis}, {Biji nangka}, {Gabus Laut}, {Gulamah/Tigawaja}, {ikan gaji}, {Ikan kakap merah/Bambangan}, {ikan lidah}, {Ikan sebelah {Terompa}}, {Ikan sebelah mata kiri}, {Jenaha}, {Kakap Batu}, {Kakap Merah}, {Kakap putih}, {Kapas-kapas}, {Kerapu bebek}, {Kerapu karang}, {Kerapu kare}, {Kuniran}, {Kurau},

{Kurisi}, {Kuro/senangin}, {Layur}, {Lencam}, {Manyung}, {Pari burung}, {Pari hidung sekop}, {Pari kekeh}, {Pari kembang/Pari macan}, {Remang}, {Swanggi}}

   After that proceed with discussing each 2-itemset follows:

F_2 ={{Beloso/buntu kerno, Biji nangka}, {Beloso/buntu kerno, Kerapu karang}, {Beloso/buntu kerno, Layur}, {Beloso/buntu kerno, Manyung}, {Beloso/buntu kerno, Pari kekeh}, {Beloso/buntu kerno, Swanggi}, {Biji nangka, Layur}, {Biji nangka, Manyung}, {Biji nangka, Swanggi}, {Gulamah/Tigawaja, Ikan kakap}, {Gulamah/Tigawaja, Ikan sebelah mata kiri}, {Gulamah/Tigawaja, Kakap putih}, {Gulamah/Tigawaja, Kapas-kapas}, {Gulamah/Tigawaja, Kerapu karang}, {Gulamah/Tigawaja, Kuniran}, {Gulamah/Tigawaja, Kurisi}, {Gulamah/Tigawaja, Layur}, {Gulamah/Tigawaja, Manyung}, {Gulamah/Tigawaja, Pari burung}, {Ikan kakap merah/Bambangan, Kurisi}, {Ikan kakap merah/Bambangan, Layur}, {Ikan kakap merah/Bambangan, Manyung}, {Ikan sebelah mata kiri, Kakap putih}, {Ikan sebelah mata kiri, Kapas-kapas}, {Ikan sebelah mata kiri, Kerapu karang }, {Ikan sebelah mata kiri, Kuniran}, {Ikan sebelah mata kiri, Kurisi}, {Ikan sebelah mata kiri, Layur}, {Ikan sebelah mata kiri, Pari burung}, {Kakap Merah, Manyung}, {Kakap putih, Kapas-kapas}, {Kakap putih, Kerapu karang}, {Kakap putih, Kuniran}, {Kakap putih, Kurisi}, {Kakap putih, Layur}, {Kakap putih, Pari burung}, {Kapas-kapas, Kuniran}, {Kapas-kapas, Kurisi}, {Kerapu karang, Kuniran}, {Kerapu karang, Kurisi}, {Kerapu karang, Layur}, {Kerapu karang, ManyungKerapu karang, Manyung}, {Kerapu karang, Pari burung}, {Kerapu karang, Swanggi}, {Kuniran, Kurisi}, {Kuniran, Layur}, {Kuniran, Pari burung}, {Kurisi, Layur}, {Kurisi, Lencam}, {Kurisi, Manyung}, {Kurisi, Pari burung}, {Kurisi, Swanggi}, {Layur, Lencam}, {Layur, Manyung}, {Layur, Pari burung}, {Layur, Pari kekeh}, {Layur, Swanggi}, {Lencam, Manyung}, {Manyung, Pari kekeh}, {Manyung, Pari kembang/Pari macan}, {Manyung, Swanggi}, {Pari kekeh, Swanggi}}

   The rule used is if x then y, where x is antecendent and y is consequent. Based on the rule, it takes two items which one of them as antecendent and the rest as consequent. From the second step obtained One item Fk that is F2. F1 is not included because it only consists of 1 item only. For antecedent may be more than 1 element, while for consequent consists of 1 element

3. Result and Discussion

The result of model testing of the analysis of fish catching pattern with a priori algorithm and a priori algorithm of experimental results using a priori algorithm with the amount $\phi = 30$, min 80% Support and min Confidence 80% shows that the rules formed are amounted to 82 rules as seen in the picture below.

   After that, the experimental results with the application of feature extraction are applied using principal component analysis for rule efficiency. Results Maximizes Mutual Information
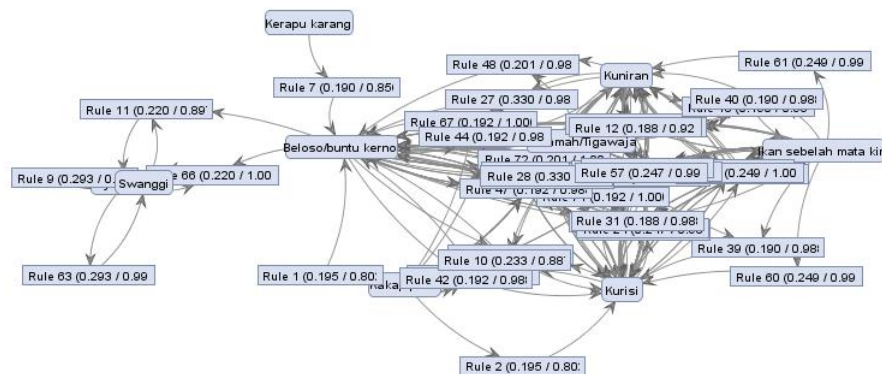
Figure 1. result rule that formed

on Gaussian Data, with entropy:

$$H(\mathbf{y}) = -\int p(\mathbf{y})\log_2 p(\mathbf{y})d\mathbf{y} = \frac{1}{2}\log_2(e(2\pi)^{d'}) + \frac{1}{2}\log_2\det(C_y)$$

Figure 2.

4. Conclusion

This study shows the results of the experiments conducted to analyse the fish catching patterns. It can be concluded that the results of the experiments using a priori algortima with the amount of $\phi = 30$, min 80% Support and min Confidence 80% shows the rules formed amounted to 82 rules. Thus, from the above model test results, it can be said that the a priori algorithm can solve the problem in the analysis of fish catching patterns. From the results of the tests that have been done and the results of conclusions given, it could be suggested that:

1. The data used should be the newest data because the results of the analysed data can be used for making and composing the strategies of fishing.
2. The researcher should pay more attention on the dimension or the number of items in the data set because it will require lots of space to store the calculation of support for each item.
3. The researcher could develop another association rules algorithm, such as quantitative association rule and generalized association rule algorithm.

References

[1] arge Scale Data by Combining Class Association Rule Mining and Information : a Hybrid Approach. Internetworking Indonesian Journal , 17.
[2] alamurugan, P. r. (2009). Feature Selection for Large Scale Rule Mining and Information Gain : a Hybrid Approach . Internetworking Indonesia Journal , 17-24.
[3] erndtsson, M., Olsson, J. H., & Lundell, B. (2008). Thesis Projects A Guide for Students in Computer Science and Information Systems. Verlag London: Springer.
[4] awson, C. W. (2009). Projects in Computing and Information System A Student's Guide. England: Addison-Wesley.
[5] ray, D. E. (2004). Doing Research in the Real World. New Delhi: SAGE.
[6] an, J., & Kamber, M. (2007). Data Mining Consepts and technique. San Fransisco: Morgan Kaufmann.

[7] anash, C. C. (2003). Mining gene expresion databases for association rule. Bioinformatics , 79-86.

[8] , G. S., & Deepa, D. S. (2011). Analysis of Computing Algorithm using Momentum in Neural Networks. Journal of computing, volume 3, issue 6 , 163-166.

[9] othari, C. R. (2004). Research Methology methodes and Technique. India: New Age Interntional.

[10] usrini, E. t. (2009). Algoritma Data Mining. Yogjakarta: Andi.

[11] arose, D. T. (2007). Data Mining Methods And Models. New Jersey: A John Wiley & Sons.

[12] aimon, O. (2010). Data Mining and Knowledge Discovery Handbook. New York Dordrecht Heldelberg London: Springer.

[13] USDATIN. (2011). Kelautan dan Perikanan dalam angka 2010. Jakarta: Kementerian kelautan dan Perikanan.

[14] antosa, B. (2007). Data Mining Teknik Pemanfaatan Data Untuk keperluan Bisnis. Yogjakarta: Graha Ilmu.

[15] iombo, M. r. (2010). Hukum perikanan nasional dan internasional. Jakarta: Gramedia Pustaka Utama.

[16] yas, E. W. (2008). Penerapan Association rule dengan menggunakan algoritma apriori untuk analisa pola data hasi tangkapan ikan. Konferensi dan temu Teknologi Informasi dan Komunikasi Untuk Infonesia .

[17] itten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining Practical Machine Learning Tools and Techniques Third Edition. USA: Morgan Kaufmann.

[18] aya Mulyana, A. D. (2008). Konservasi kawasan perairan indonesia bagi masa depan dunia. Jakarta: Direktorat Konservasi dan Taman Nasional laut, Kementerian Kelautan dan Perikanan.