

Analysis of Investment in Government Sector Shares in the Election of the Governor of DKI Jakarta Using the Mining Rule Method Association

Deny Pratama¹ and Yuli Karyanti²

^{1,2}Business Information System, Gunadarma University, Depok, West Java, Indonesia

¹denyuno@gmail.com, ²yuli@staff.gunadarma.ac.id

ABSTRACT

Data mining is extracting knowledge from very large data so that it produces useful information (Vaidya 2004). Currently data mining is not only for information gathering, but is used to recognize patterns and recognize specific characters (Neel 2011; Zaki and Meira 2013). The computational technique used is the Rule Mining Association (ARM) method, ARM can be used to determine the relationship between stocks (Agrawal et al. 1993; Lu et al. 2000). Apriori algorithm is a method to find the relationship pattern between one or more items in a dataset. This study aims to determine the minimum support value and confidence value in the ten banks listed on the IDX (Indonesia Stock Exchange) and able to select any stocks that are considered profitable. So as to make it easier for investors to decide what stocks are worth buying.

Keywords: *Investment, Stock, a Priori Algorithm and Association Rule Mining.*

1. INTRODUCTION

Many issuers offer shares in various attractive ways and methods, which can make potential investors give confidence that the shares offered will grow positively. In investing in shares, there are two types of long and short term investments, which have different returns and risks. Risks and returns in investing are closely related which is often known as high risk - high return, low risk - low return (Worthington & Higgs 2004).

Long-term investment has a smaller risk and a smaller profit than short-term investment, because short-term investments are volatile stock price movements and far more difficult to determine the pattern of stock price movements (Makaryanawati & Ulum 2009). Investor rationality can be measured from the extent to which they succeed in choosing stocks that give maximum results at certain risks, also influenced by investor preferences for different returns and risks.

The computational technique used is data mining with the ARM method. A priori algorithm is a method to find the

relationship pattern between one or more items in a dataset. A priori algorithm is to produce a pattern that was first introduced by Agrawal et al. (1993). This research aims to determine the minimum support value and confidence value in ten banks and is able to select what stocks are considered profitable. It is very important for investors to get relevant financial information. This research provides convenience in investing in the banking sector so that it can decide what stocks are worth buying and obtaining relevant financial information, thus helping investors to make decisions in investing.

2. RESEARCH METHODS

Association Rules Mining (ARM) is a data mining technique to find your association rules for item combinations. ARM aims to extract correlations and relationships between items contained in large data sets. ARM has two phases: frequent itemsets extraction and the formation of an associative rule from a collection of items, with the limitation of minimum support (minsup) values and minimum confidence (minconf) (Zang 2010; Lin 2000), from these two phases making ARM there are two problems. The first problem is having to find a collection of items that exceed the specified minutes and the second problem is to produce association rules from a large-scale collection of items with predefined minimum limits (Dudgikar 2000; Kotsiantis and Kanellopoulos 2006). An example of an analyst association rule in a supermarket is to know how likely a customer is to buy bread along with milk, so knowing the relationship between these items, the supermarket owner can arrange the placement of goods or design a marketing campaign using a combination of discount coupons for certain items in order to increase sales, examples from market basket analysis in Table 1. (Tan et al. 2006).



Table 1. Examples of market basketball transactions (Tan et al. 2006)

TID	ITEMS
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

ARM has the main function of expressing the relationship between items in the transaction.

Apriori algorithm is to produce a pattern that was first introduced by Agrawal et al. (1993). A priori algorithm uses knowledge about frequent itemsets that have been known beforehand, and processed the information further (Erwin 2009). In data mining there are problems that one of them is the amount of data, so that there is the possibility of data duplication (Tan et al. 2006). One solution is to use anti-monotonic, in Figure 1.

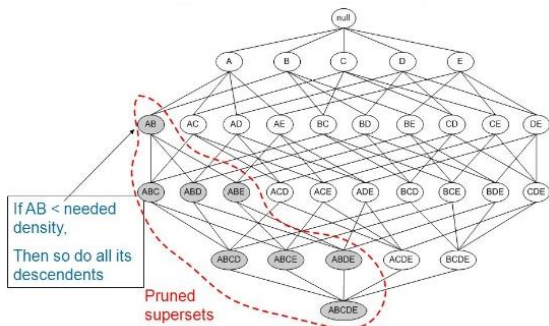


Fig. 1. Anti-monotonic inheritance of frequent itemset generation (Tan et al. 2006)

Anti-monotonic has the aim of reducing computational complexity by reducing the number of candidates using pruning techniques, reducing the number of comparisons, anti-monotonic is used in the apriori principle, with support values that never exceed the predetermined minimum support, so it is assumed that the items in each transaction stored and sorted in lexicographics, this is used to adapt algorithms if there are cases where database D is stored normally (Gurel 2008). There are many different types of transactions, but in doing lexographics each item must be made in a table first as in table 1. The use of normalized storage is illustrated in table 2.

Table 2: Normalized data storage

TID	ITEMS
1	Bread
1	Milk
2	Bread
2	Diaper
2	Beer
2	Eggs
.....
5	Coke

A priori principle illustration with minimum support is three, which corresponds to table 1 as an example of transaction data used in Figure 2.

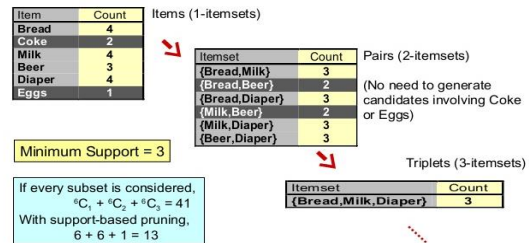


Fig. 2. A priori principle with a minimum support = 3 limit (Tan et al, 2006)

The process of a priori is finding the final pattern that often appears in the transaction, the illustration can be seen in Figure 2. Every repetition requires the original database. Many variants to increase a priori and have been proposed by reducing the number of candidates further, the number of transactions scanned or scanned from a very large amount of data so that this process is unfortunate because it scans the database repeatedly and checks one by one by matching patterns (Kaur 2013) The research flow is shown in Figure 3.

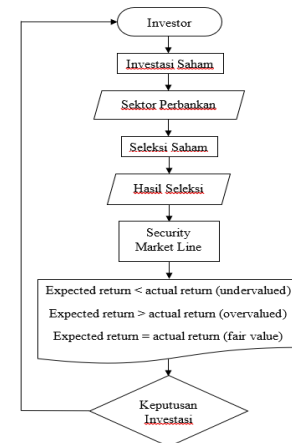


Fig. 3. Thought Framework

3. DISCUSSION

This research uses a quantitative approach using secondary data. In general, this research is a study conducted to determine whether data mining methods can select stocks to determine the size of a stock return. The object of research uses the banking sector because the banking sector is one of the vital sectors and is the heart of the economy in Indonesia and banking management is supervised and regulated by the government. The banking selected in this study uses the best 10 banking assets in quarter 1 of 2017 which refers to the Financial Services Authority (OJK) and the Indonesia Stock Exchange (IDX). The number of banking companies listed on the JSE until 2017 was 81 banking companies, but only 10 banking sectors were taken including 4 state-owned banks and 6 private banks.

Table 3. List of issuers' tables

Code Emiten	Nama Emiten
BBRI	PT. Bank BRI (Persero) Tbk.
BMRI	PT. Bank Mandiri (Persero) Tbk.
BBNI	PT. Bank BNI (Persero) Tbk.
BBTN	PT. Bank BTN (Persero) Tbk.
BBCA	PT. Bank Central Asia Tbk.
BNGA	PT. Bank CIMB Niaga Tbk
PNBN	PT. Bank Pan Indonesia Tbk
BNLI	PT. Bank Permata Tbk
BNII	PT. Bank Maybank Indonesia Tbk
BDMN	PT. Bank Danamon Indonesia Tbk

In this study the data taken are secondary data in the form of daily time series data taken from 3 months before and 3 months after the election of the Governor of DKI Jakarta and taken 3 election periods for DKI Jakarta Governor namely 2007, 2012 and 2017. Data used in the study uses data the historical price of each issuer listed on the Jakarta Stock Exchange (JKSE) and also registered at www.yahooofinance.com in the banking sector. Stock data used in the study are banking companies listed on the IDX.

The Role of Information Systems is by using data mining with a priori algorithms. In the process there is a data collection that is data obtained from www.yahooofinance.com in the form of tables.

In this study, which is used to select stocks is to use the difference between the close value, thus forming the actual return value. The actual return of the stock is the actual return received by the shareholders, the calculation using the formula.

$$R_{i,t} = \frac{(P_{i,t} - P_{i,t-1})}{P_{i,t-1}}$$

Information :

R_{i, t}: The actual return of the stock-I price index during the day

P_{i, t}: Stock-i price index during the t-day

P_{i, t-1}: Stock-I price index during the day to t-1

The main purpose of data preprocessing is to convert the untreated stock data obtained from YahooFinance into a friendly data association rule, so that the data can be extracted with the Mining Rule Association. In preprocessing data there are several steps, which are combining shares in one table. The second step changes the return value with 1 if the return value is positive and 0 if the return is negative or not generating. From table 4 produces return values such as table 5 for the period of 2017.

Table 5: Association of data set friendly rules for 2017

Date	BBRI	BMRI	BBCA	BBNI	BNGA	PNBN	BNLI	BNII	BDMN	BBTN
17/11/16	0	1	1	0	0	0	0	1	1	0
18/11/16	0	1	0	0	0	0	0	0	1	0
21/11/16	0	0	0	0	1	0	1	0	1	0
22/11/16	1	1	0	1	0	1	0	1	0	1
23/11/16	1	0	1	0	0	0	0	1	1	1
.....
Keangif	77	76	82	77	72	72	51	50	81	72
Hari	167	167	167	167	167	167	167	167	167	167
Jumlah%	46,11	45,51	49,10	46,11	43,11	43,11	30,54	29,94	48,50	43,11

The process of lifting data uses the IF and ELSE functions to find out whether on the same date these stocks have a positive rate of return. The a priori algorithm function can be formulated as follows:

Itemset-1:

Looping (n = 0; n <= number of shares; n ++)

IF (Support <minsup) then the stock is trimmed
 ELSE shares are not cut

Looping (m = 2; m <= count (stock not trimmed); m ++)

IF (count (stock not trimmed)> 1)

Itemset-m:

Looping (n = 0; n <= number of shares not cut; n ++)

Looping (i = 0; I <= number of days of transactions; i ++)

IF (Sahamn, 1 + Sahamn + 1, i
 (increases accordingly lots of m) = then
 the value of 1

ELSE value is 0

IF (Support <minsup) then the stock is trimmed

ELSE shares are not cut

Count (Stock is not trimmed);

ELSE stock selection end.

Data trimming aims to reduce computational complexity problems by being limited by minimum support. In Figure 4 is the result of 3-itemset value in 2017, for 2012 and 2007 using the same method. Where in this period, the 3-itemset value determines which



shares are suitable for use. The following is the connection between the shares that are suitable for use in the 3-itemset in Figure 4 in 2017.

NO	RULE	SUPPORT
1	BBRI + BMRI + BBKA	18,56
2	BBRI + BMRI + BBNI	20,96
3	BBRI + BMRI + BNGA	16,77
4	BBRI + BMRI + BDMN	20,36
5	BBRI + BMRI + BBTN	19,16
6	BBRI + BBKA + BBNI	19,76
7	BBRI + BBKA + BNGA	16,77
8	BBRI + BBKA + BDMN	22,75
9	BBRI + BBKA + BBTN	16,77
10	BBRI + BBNI + BNGA	18,56
11	BBRI + BBNI + BDMN	20,36
12	BBRI + BBNI + BBTN	18,56
13	BBRI + BNGA + BDMN	16,77
14	BBRI + BNGA + BBTN	14,97
15	BBRI + BDMN + BBTN	16,77

NO	RULE	SUPPORT
1	BBRI + BMRI + BBNI	20,96
2	BBRI + BMRI + BDMN	20,36
3	BBRI + BBKA + BDMN	22,75
4	BBRI + BBNI + BDMN	20,36

Fig. 4. Trimming 3-itemset data with a minimum support of 20%

The stock selection calculation for the 2012 and 2017 periods is the same as the minimum 20% support calculation in 2007. The difference is the result of the selection process which stops at the number of itemset and the issuer is above the minimum support value. Stock selection results in 2012 were carried out until 3-itemset, but 2 items were used because there was no support value that exceeded the minimum support. The results of the selection process get two issuers, namely BMRI and BBKA. The results of the 2017 stock selection are carried out until 4-itemset, but the items used are 3-itemset because the absence of a support value exceeds the minimum support.

The results of the selection process obtained five issuers, namely BBRI, BMRI, BBKA, BBNI and BDMN. From the results of the selection process that exceeds the minimum support value of 20% in the period 2007, 2012 and 2017 can be seen in table 6.

Tabel 6. Nilai minimum support sebesar 20% 2007, 2012 dan 2017

No	Year	Issuer
1.	2007	BMRI, BBKA, BBRI
2.	2012	BMRI, BBKA, BBNI, BBTN
3.	2017	BMRI, BBKA, BBRI, BBNI dan BDMN

From table 6. it can be determined to find the final pattern that often appears in the period 2007, 2012 and 2017 namely the issuer BMRI and BBKA. Then the selection is done again by looking at the results of the confidence value. The value of confidence has a role to know how sure the rules are formed. The greater the value of confidence, the better the rule. The limit of the minimum confidence (mincof) value used in this study is 50%, so the rules that have confidence values below the selected mincof value in table 7.

Table 7: Results of calculation of confidence values for 2007, 2012 and 2017

YEAR	RULES	CONFIDENCE
2007	BBKA -> BMRI	0,71
2007	BMRI -> BBKA	0,72
2012	BBKA -> BMRI	0,58
2012	BMRI -> BBKA	0,54
2017	BBKA -> BMRI	0,54
2017	BMRI -> BBKA	0,58

Results of table 7, it is known that the rules that are selected by BBKA -> BMRI and BMRI -> BBKA have a large confidence value above 50%, so that the rule can move positive values. These results can be used as a concept to buy a stock.

4. CONCLUSION

Determining the minimum support value is determined directly by the user, therefore to get the appropriate minimum support it needs trial and error. By using a minimum 20% support to be able to select any stocks that are considered profitable, it is evident that the selection process in the 2-itemset stage selects stocks that have a Beta value below 1, where the beta value below 1 means that the stock has low volatility . Calculations using ARM are also able to know the relationship between stocks, but the relationship cannot be used in the long term because the closing price of the stock changes every day.

The final result of the stock selection process there are 2 shares used, namely BMRI (Bank Mandiri) and BBKA (Bank Central Asia). Not all shares are suitable to be invested in pairs because each rule has different support and confidence values. Therefore, the greatest confidence value is used, because the confidence value symbolizes the level of confidence in the rule and that is worth investing is BMRI with BBKA with confidence values above 50% in three consecutive periods.

REFERENCES

- [1] Abid Shaikh, Ta-ware Pritam, et al 2014. "Stock Exchange Market Prediction". International Journal of Advance in Computer Science and Technology. India.
- [2] Agustria, Kharisma dan Khuzaini. 2016. "Pengaruh Kinerja Keuangan Terhadap Harga Saham Perusahaan yang Terdaftar dalam Indeks LQ 45". Jurnal Ilmu dan Riset Manajemen Volume 5 Nomor 3 Maret 2016 ISSN 2461- 0593.



- [3] Annie, Kumar. 2012. "Market Basket Analysis for a Supermarket based on Frequent". India: Government Arts College Trichy.
- [4] Arfan Rifqiawan. 2008. "Analisis Perbedaan Volume Perdagangan Saham – Saham yang Optimal pada Jakarta Islamic Index (JII) di Bursa Efek Indonesia (BEI)". Sains akuntansi. Universitas Diponegoro Semarang.
- [5] Arief Jananto. 2012. "Penggunaan Market Basket Analysis untuk Menentukan Pola Kompetensi Mahasiswa". Jurnal Teknologi Informasi. Universitas Stiku Bank.
- [6] Astuti, Pudji. (2002). Analisis Variabel-Variabel Yang Mempengaruhi Harga Pasar Saham Perusahaan Perbankan di BEJ. Kompak No.6 Hal 301-307.
- [7] Carol Anne Hargreaves, Pratek Dixit and Ankit Solanki. 2013. "Stock Portofolio using Data Mining Approach". Institute of Systems Sciences National University of Singapore.
- [8] Cipta Riang Sari. 2016. "Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan". Indonesian Journal on Networking and Security. Universitas Al Asyariah Mandar.
- [9] Conti Dante, Martinez and Pernia Alpha 2008. "Finding temporal associative rules in financial time series: A case of study in Madrid Stock Exchange". Venezuela. Unversidad de Los Andes.
- [10] Jiawei han, Micheline Kamber and jian Pei. 2006. "Data Mining Concepts and Techniques". Third Edition. Mourgan Kaupmann.
- [11] Jogyanto, Hartono. (2003). Manajemen Investasi. Jakarta. Salemba Empat.
- [12] Makar, Ulum. 2009. "Pengaruh tingkat Suku Bunga dan Tingkat Likuiditas Perusahaan terhadap Risiko Investasi Saham yang terdaftar pada Jakarta Islamic Index". Malang: Universitas Negeri Malang.
- [13] Mangasa Simatupang.(2010). Pengetahuan Praktis Investasi Saham dan Dana Reksa. Jakarta: Mitra Wacana Media.
- [14] Pratama, Aditya dan Teguh Ernawati. 2014. "Pengaruh Current Ratio, Debt to Equity Ratio, Return On Equity, Net Profit Margin dan Earning Per Share Terhadap Harga Saham (Studi Kasus pada Perusahaan Manufaktur yang Terdaftar di Bursa Efek Indonesia Periode 2008-2011)". Jurnal Akuntansi Vol. 2, No. 1, Juni 2014.
- [15] Ratih S. (2011). Pengaruh Risiko sistematis dan Risiko Tidak Sistematis terhadap Expected Return Portofolio Saham Perusahaan Manufaktur yang terdaftar di BEI. Fakultas Ekonomi Universitas Sebelas Maret : Surakarta.
- [16] Suhaiza Zailani. 2012. "A New Scheme for Extracting Association Rules : Market Basker Analysis Case Study". Int. J. Bussinerr Innovation and Research. University of Malaya.
- [17] Tan P, Kuma V, Srivastava J 2004. "Selecting the right objective measure for association analysis". Inf Syst 29(4):293-313.
- [18] Tanja Magoc, Xiaojing Wang. 2011. "Aplication of Fuzzy Measures and Intervals in Finance". Computer Science. University of Maryland.