IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

The Pattern Failure Analysis of Sulfuric Acid Production Process with the Association Rules Algorithm Apriori

To cite this article: W Septiani *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **528** 012069

The Pattern Failure Analysis of Sulfuric Acid Production Process with the Association Rules Algorithm Apriori

W Septiani[1], I A Marie[2], D Sugiarto[3] and L Hakim

Faculty of Industrial Technology, Kampus A Universitas Trisakti, Jakarta 11440, Indonesia
[1]winnie_septiani@trisakti.ac.id, [2]iveline.annemarie@trisakti.ac.id,
[3]dedy@trisakti.ac.id

Abstract. PT. Asam Sulfat is industrial Sulfuric Acid that was able to meet the needs of most battery manufacturers in Indonesia. Problems in production operations were mostly caused by production equipment, such as engine stop suddenly, length of set up and adjustment time and decrease of machine production speed. These results caused a decrease level in the efficiency and effectiveness of the engine. This study aims to identify and analyze the machine and components that dominate the sulfuric acid production process damage. The method of analysis used is the Association Rules Algorithm Apriori. Based on the results of initial identification obtained 23 machines and 9 components are often damaged. The variables analyzed include unit, machine type, component, time to repair and turbidity. Using a minimum value of support 0.1 and confidence 0.9 resulted in 17 rules dominated by low turbidity, especially in AT Pump engines and Gasket components. The High turbidity occurred in patterns with the Ash Pump component, The Time to Repair was high and it happened on the Sulfur Pump engine. By knowing the patterns that occurred absolutely the failure in the production process can be more quickly handled.
Keywords: association rules, machine, component, turbidity, sulphuric acid

1. Introduction

PT Asam Sulfat is the first Sulfuric Acid Industry in Indonesia. One of the products produced by PT Asam Sulfat is sulfuric acid supplied to the automotive battery industry that meets standards prescribed by the Indonesian National Standards (Indonesian: *Standar Nasional Indonesia*, abbreviated to SNI) as can be seen in Table 1. The capacity for sulfuric acid production is 82,500 ton per annum. The company has obtained the ISO 9001 certificate and meet standards prescribed by SNI for sulfuric acid.

The quality of sulfuric acid products produced by PT Asam Sulfat is able to meet the needs of most battery plants in Indonesia and the sulfuric acid offers good translucence and iron content that other sulfuric acid plants might find it hard to produce, giving it a big advantage over other plants. This is because of the highly corrosive nature of sulfuric acid itself, thus requiring that the operation of production machinery and supervision of the processes be arranged carefully. The production of sulfuric acid is carried out continuously. Cessation of a production process often results from problems with the production equipment, for example in the event of downtime, time-consuming set-up and adjustment, and decreased speed of production machinery. Those cause losses to the company as they reduce the efficiency and effectiveness of the machines and result in repair costs. In anticipation of those problems, an analysis needs to be performed using a method called association rules in order to reduce production failure.

Table 1. Specifications Prescribed by SNI 0030: 2011

Specification	Content	Specification	Content
Sulfuric Acid Content	Min 98.20%	Sulfuric Acid	Min 98.20%
Residue on Ignition	Max 0.02%	Residue on	Max 0.02%
Turbidity	Max 60 NTU	Turbidity	Max 60 NTU
Chloride (CI) Content	Max 5 ppm	Chloride (CI)	Max 5 ppm
Iron (Fe) Content	Max 40 ppm	Iron (Fe)	Max 40 ppm
Lead (Pb) Content	Max 9 ppm	Lead (Pb)	Max 9 ppm
Arsenic (As) Content	Max 0.2 ppm	Arsenic (As)	Max 0.2 ppm
Copper (Cu) Content	Max 1 ppm	Copper (Cu)	Max 1 ppm
Selenium (Se) Content	Max 0.5 ppm	Selenium (Se)	Max 0.5 ppm
Zinc (Zn) Content	Max 2.0 ppm	Zinc (Zn)	Max 2.0 ppm

Association rules are a useful data mining technique to find out a correlation or the most important/ interesting pattern of a large data set [1]. There are a great number of previous studies undertaken based on scientific works that use the method *association rules*, for example those conducted by: [2], which studied the implementation of data mining on the sale of plane tickets using the apriori algorithm; [3], which conducted research into mining and visualization of association rules over relational dbmss; [4], which studied data mining to analyze the rate of street crimes using the algorithm of the association rules and the apriori method; [5], which studied the application of association rules using the apriori algorithm to analyze fish catch data patterns; and [6], which applied association rules to determine traffic accident patterns.

Poor maintenance at PT Asam Sulfat greatly affect the resulting downtime level and frequent engine breakdowns as a result of the engine condition that is less suitable and the absence of production machine effectiveness calculation. This study aims to identify and analyze the machine and components that dominate the sulfuric acid production process damage using *association rules* with the apriori algorithm.

2. Literature Review

According to [7], association rules are defined as a data mining technique used to determine the rules of association between particular combinations of an item. Association rules are one of the methods that aim to find out patterns that often appear or occur between many transactions, where each transaction consists of several items, and, thus, this method will support the recommendation system through the discovery of patterns between items in the existing transactions [4].

Whether particular association is important or not can be determined using two benchmarks, namely: support and confidence. Support (the support value) refers to the percentage of combinations of the items in the database, while confidence (the value of certainty) is the extent to which the relationship between items in association rules is deemed strong [2]. Support in this research is defined as the probability of occurrence of several item sets in the variable *plane crash* resulting from the association rules of the overall variable *crash*.

3. Method

In this research, the data used were the ones on sulfuric acid production obtained from PT Asam Sulfat in 2017. The stage of research carried out as follows:

1. Preliminary data analysis

   The analysis used represents the overall description of each variable, especially the variables *machine type* and *component* using the Pareto chart. The next step was to take several data that run into problems most often.

2. Determine the analysis variable
   There were a total of five variables, namely: *unit*, *machine type*, *component*, *time to repair*, and *turbidity*. The categories for each variable are determined as follows:
   - Unit : Two, Four
   - Machine Type : AT pump, DT pump, Sulfur pum
   - Component : Ash pump, bearing, breaker, bushing pump, electro motor, gasket, impeller, pipa steam, hulley pump
   - Time to repair : high, low, medium
   - Turbidity : high, low
3. Compile association rules
   The basic methodology of the association analysis is divided into three stage :
   a. *High-frequency pattern analysis*
   This stage aims to discover the combination of items that meets the requirement of the minimum support value in the predetermined database. Support refers to a percentage of occurrence in the data wherein both Events A and B occur simultaneously. Thus, the support value of an item is obtained using the following formula [8]:

$$\text{Support (A} \rightarrow \text{B)} = (A \cup B) = \frac{n(A \cup B)}{n(S)} \tag{1}$$

$$\text{Support (A} \rightarrow \text{B)} = \frac{\text{number of transactions containing A and B}}{\text{total number of transactions}} \tag{2}$$

where:
$(A \cup B)$: Probability of simultaneous occurrence of Events A and B
$(A \cup B)$: Rate of simultaneous occurrence of Events A and B
(S) : Number of S members

   b. *Establishment of the association rules*
   After all high-frequency patterns have been discovered, the next step is to find out the association rules that meet the minimum confidence value requirement. Confidence refers to the percentage of occurrence where if the Event A occurs, then the Event B will occur as well. A combination of items is obtained using the following formula [8]:

$$\text{Confidence (A} \rightarrow \text{B)} = P(B|A) = \frac{P(A \cup B)}{P(A)} \tag{3}$$

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{number of transactions containing A and B}}{\text{number of transactions of A}} \tag{4}$$

where:
P(B|A) : Conditional probability of Event B if Event A has occurred
$P(A \cup B)$: Probability of simultaneous occurrence of Events A and B
P(A) : Probability of Event A

   In addition to the two parameters above, which are the support value and the confidence value, there is another parameter known as the lift ratio. The lift ratio indicates a particular level of power that rules have over the resulting association rules. [8] defines the lift ratio as the confidence of the association rules divided by the probability of Event A and the probability of Event B independent of each other, which is expressed in the following formula:

$$\text{Lift ratio (A} \rightarrow \text{B)} = \frac{P(A \cup B)}{P(A)P(B)} \tag{5}$$

$$\text{Lift ratio} = \frac{\text{confidence (A=>B)}}{P(B)} \tag{6}$$

$$= \frac{\text{confidence (A=>B)}}{\text{support B}} \tag{7}$$

where:
P(A∪B) : Probability of simultaneous occurrence of Events A and B
P(A) : Probability of Event A
P(B) : Probability of Event B

c. Apriori algorithm
The apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 to determine frequent item sets for association rules. The rules expressing the association between several attributes are often called *market basket analysis*.

4. The descriptive picture of the rules established using software *R*
Afterwards, association rules were established, followed by the descriptive picture of the rules established using software *R* with the assistance of several packages such as: *arules* and *arulesViz*.

4. Result and Discussion

4.1 Preliminary Data Analysis
Figure 1 below provides information about machine types that commonly run into problems.

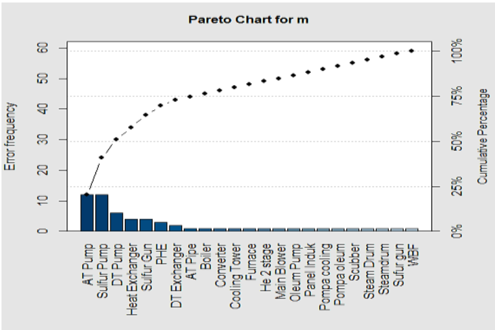

Figure 1. The type of machine with the highest level of problem

If looking closely at Figure 1, machine types *AT Pump, Sulfur Pump,* and *DT Pump* are the three machine types that run into problems most often compared to the other machine types. If there are no efforts made to solve this issue, it will result in unstable product quality and, in turn, will lead to customer dissatisfaction. In addition, components that often run into problems were also examined using the same diagram.

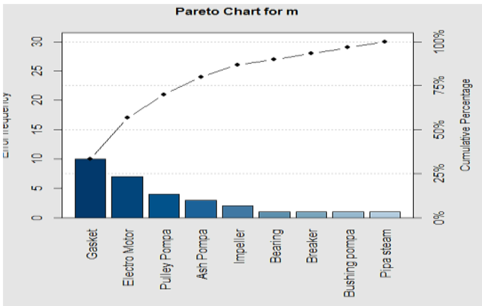

Figure 2. The type of machine with the highest level of problem

In Figure 2, it is revealed that *Gasket, Electro Motor,* and *Pulley Pump* are components that run into problems most often. It is not only damage to the engine that causes a high turbidity value, but damage to the components will also affect the value of turbidity.

After finding out the dominant damage taking place during the production process, the next step was to perform an analysis using the association rules of the apriori algorithm using only the machine types that often run into problems, namely AT Pump, Sulfur Pump, and DT Pump.

4.2 Determine the analysis of category variable
There were a total of five variables, namely: *unit, machine type, component, time to repair,* and *turbidity*. The categories for each variable are determined as follows:

- Unit : Two, Four
- Machine Type : AT pump, DT pump, Sulfur pum
- Component : Ash pump, bearing, breaker, bushing pump, electro motor, gasket, impeller, pipe steam, hulley pump
- Time to repair : high, low, medium
- Turbidity : high, low

4.3 Compile association rules

4.3.1 High-frequency pattern analysis. In this research, to determine the relation between the resulting patterns or rules of association, the researcher tried to set limits on the support value and the confidence value, which were equal to 0.03 and 0.9, respectively. Using those limits, there are a total of 104 rules established.

4.3.2 Establishment of the association rules. To illustrate how association rules work using the apriori algorithm using five sample, refer to the results below:

1. Data association rules

Table 2. Example data Association Rules of machine type, component and Turbidity

No	Machine Type	Component	Turbidity
1	AT Pump	Gasket	Low
2	Sulfur Pump	Impeller	High
3	AT Pump	Impeller	High
4	AT Pump	Gasket	Low
5	Sulfur Pump	Gasket	High

2. To convert data presented in the table above into a matrix of binary numbers, where 1 indicates that the process exists while 0 indicates that the process does not exist (Table 3).

Table 3. Convert association rules into a binary numbers

No	AT Pump	Sulfur Pump	Gasket	Impeller	High	Low
1	1	0	1	0	0	1
2	0	1	0	1	1	0
3	1	0	0	1	1	0
4	1	0	1	0	0	1
5	0	1	1	0	1	0
Total	3	2	3	2	3	2

3. To determine the total frequency of item sets and the number of item sets used (Table 4 – Table 11). The total frequency of item sets is equal to 2 with a total of 3 item sets.

IOP Conf. Series: Materials Science and Engineering **528** (2019) 012069 doi:10.1088/1757-899X/528/1/012069

Table 4. Combination 1 for Turbidity Low

AT Pump	Gasket	Low	x
1	1	1	1
0	0	0	0
1	0	0	0
1	1	1	1
0	1	0	0
Σ			2

Table 5. Combination 2 for Turbidity Low

AT Pump	Impeller	Low	x
1	0	1	0
0	1	0	0
1	1	0	0
1	0	1	0
0	0	0	0
Σ			0

Table 6. Combination 3 for Turbidity High

AT Pump	Gasket	High	x
1	1	0	0
0	0	1	0
1	0	1	0
1	1	0	0
0	1	1	0
Σ			0

Table 7. Combination 4 for Turbidity High

AT Pump	Impeller	High	x
1	0	0	0
0	1	1	0
1	1	1	1
1	0	0	0
0	0	1	0
Σ			1

Table 8. Combination 5 for Turbidity Low

Sulfur Pump	Gasket	Low	x
0	1	1	0
1	0	0	0
0	0	0	0
0	1	1	1
1	1	0	0
Σ			1

Table 9. Combination 6 for Turbidity Low

Sulfur Pump	Impeller	Low	x
0	0	1	0
1	1	0	0
0	1	0	0
0	0	1	0
1	0	0	0
Σ			0

Table 10. Combination 5 for Turbidity High

Sulfur Pump	Gasket	High	x
0	1	0	0
1	0	1	0
0	0	1	0
0	1	0	0
1	1	1	1
Σ	1		

Table 11. Combination 6 for Turbidity High

Sulfur Pump	Impeller	High	x
0	0	0	0
1	1	1	1
0	1	1	0
0	0	0	0
1	0	1	0
Σ	1		

In Column X, Number 1 indicates items that occur simultaneously, while Number 0 means that no items occur simultaneously or there are no transactions. The symbol Σ represents total frequency of item sets. The total frequency of item sets must be greater than or equal to the predetermined total frequency of set items. The predetermined frequency of item sets above is as many as 2 item sets, if there are less than 2 item sets, then the rule is not used. From the tables above, there is one combination with value of the frequency of item sets equal to the predetermined total frequency of item sets, which is Combination 1.

4. The next step is to take the combination that has met the predetermined requirement, i.e. the one with total frequency of item sets of 2 and a total of 3 item sets used. The combination that met the predetermined requirement is Combination 1 with total frequency of item sets of 2 (Table 12).
   Based on the table 12, it is revealed that the resulting rules are listed below: (AT Pump, Gasket, Low)
5. The rule used is “if x, then y”, where x is the antecedent while y is the consequent. Based on the rule, there must be 2 item sets, one of which serves as the antecedent and the other serves as the consequent (Tabel 13).

Table 12. Combination AT PUM and Gasket for Turbidity Low

AT Pump	Gasket	Low	x
1	1	1	1
0	0	0	0
1	0	0	0
1	1	1	1
0	1	0	0
Total	2		

Table 13. If Then Rule for Machine Type, Component and Turbidity

No	If antecedent, then consequent	Support	Confidence	Lift
1	If there is damage to the machine type *AT Pump* and to the component *Gasket*, then such a condition will result in *Low Turbidity*.	{Total simultaneous occurrence/ Total data} = {2/5=0.4}	{ Total simultaneous occurrence/ Antecedent (X)}= {2/2=1)	{Total data/ Consequent (Y)}={12/2=6}

Rule No. 1. If there is damage to the machine type *AT Pump* and to the component *Gasket*, then such a condition will result in *Low Turbidity* with support value (domination) of 0.4 or 40% and confidence value (level of confidence) of 1 or 100%.

4.3.3 Apriori algorithm. Based on the limits set, it is revealed that there are a total of 17 rules established. The following is information obtained from Table 14. In Rule No. 1. In the event of failure to Unit 4, especially to component Ash Pump, the resulting turbidity is usually low with support value (domination) of 10% and confidence value (the level of confidence) of 100%. In Rule No. 2. In the event of failure to Unit 2 and it occurs to machine type DT Pump, the resulting turbidity is usually low with support value (domination) of 10% and confidence value (the level of confidence) of 100%. In Rule No. 16. In the event of failure to Unit 4 and it occurs to machine type AT Pump, especially component Gasket, the resulting turbidity is usually low with support value (domination) of 16% and confidence value (the level of confidence) of 100%. In Rule No. 17. In the event of failure to Unit 4 and it occurs to machine type AT Pump, especially component Gasket, the resulting turbidity is high with support value (domination) of 10% and confidence value (the level of confidence) of 100%. Table 14 presents a summary of the 104 rules established.

Table 14. Association Rules

No	Sup	Conf	Lift	If (Antecedent)				Then (Consequent)
				1st Condition	2nd Condition	3rd Condition	4rd Condition	Response
1	0.10	1	2.73	Unit=Four	Component= Ash Pump			Turbidity=High
2	0.10	1	1.57	Unit=Two	Machine.Type= DT Pump			Turbidity=Low
--	----	---	---	------------------	---------------	------------		------------
--	----	---	---	------------------	---------------	------------		------------
16	0.16	1	1.57	Unit=Four	Machine.Type= AT Pump	Component= Gasket		Turbidity=Low
17	0.10	1	1.57	Unit=Four	Machine.Type= AT Pump	Component= Gasket	Time.to.Repair= Low	Turbidity=Low

4.4 Graph of items of the association Rules

Then, those association rules are visualized in the form of a Graph of Items of the Association Rules illustrated in Figure 3.

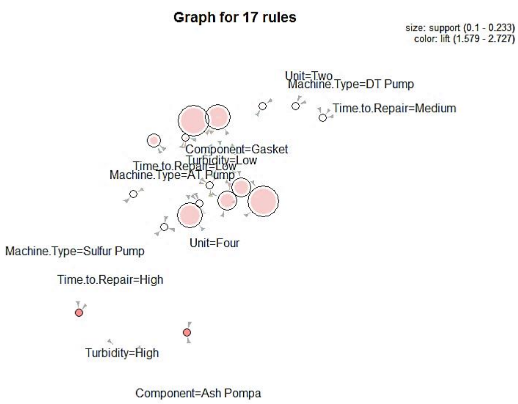

Figure 3. Visualization Graph

Based on Figure 3, it is revealed that the measures used in the graph of items were based on the support value and the lift ratio. Based on the figure above, it can be concluded that:

- A larger circle size indicates a higher support value (domination in the association rules). This can be seen from the association rules where machine type AT Pump, component Gasket, Unit 4 have a larger circle size than the other circles, thus it is revealed that the rule is the rule with the highest support value as can be seen in rules No. 4 and 14 with support value of 23%.
- A darker circle color indicates a higher lift value (the power of association to generate patterns). This can be seen from the association rules where machine type Sulfur Pump, High Time to Repair, High Turbidity, and component Ash Pump have a darker circle color than the other circles, thus it is revealed that the rule is the rule with the highest lift value as can be seen in rules No. 1 and 3. According to the rules, the existing patterns describe events that must happen or whose probability of occurrence is higher than that of other events.

5. Conclusions

Engine failure during the production process will certainly affect the sulfuric acid produced. Failure is more common in machine types AT Pump, Sulfur Pump, and DT Pump as well as components Gasket and Electro Motor. If there are no efforts made to solve this issue, it will affect the resulting turbidity, leading to the unstable production process. In addition, using a minimum support value of 0.1 and a minimum confidence value of 0.9, there are a total of 17 rules established, most of which generate low turbidity, especially in machine type AT Pump and component Gasket. On the other hand, high turbidity is generated by the pattern with the component Ash Pump, High Time to Repair, and machine type Sulfur Pump. Knowing the resulting patterns, failure during the production process can be solved. Thus, production outcomes will be stable and it in turn will affect customer satisfaction.

Acknowledgement

This paper is the result of a research grant funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia in accordance with the Research Contract: No. 011/KM/PNT/2018. The author is very grateful for all the data assistance and support from the leadership and staff of PT GS Battery and the Trisakti University Research Institute for the implementation of this research.

6. References

[1] Kuswardani D, Widyanto MR, Trihandini I. 2011 Metode association rules untuk analisis citra CT organ pasien kanker ovarium. *Kursor* **6**(2) pp 111-20

[2] Siregar S R 2014 Implementasi data mining pada penjualan tiket pesawat menggunakan algoritma apriori. *Pelita Informatika Budi Dharma* **7**(1) pp 152-6

[3] Zhang H 2000 *Mining and visualization of association rules over relational Dbmss*. (Florida USA: University Of Florida In Partial Fulfillment Of The Requirements For The Degree Of Master Of Science)
[4] Fadlina 2014 Data mining untuk analisa tingkat kejahatan jalanan dengan algoritma association rules metode apriori. *Informasi dan Teknologi Ilmiah (INTI)* **3**(1) pp 144-54
[5] Tyas E W 2008 Penerapan metode association rule menggunakan algoritma apriori untuk analisa pola data hasil tangkapan ikan. *e-Indonesia Initiative 2008* pp 1-4
[6] Hakim L and Fauzy A 2015 Penentuan pola hubungan kecelakaan lalu lintas menggunakan metode association rules dengan algoritma apriori. *The 1st University Research Colloquium (URECOL)*
[7] Kusrini and Luthfi E T 2009 *Algoritma Data Mining* (Yogyakarta: Andi Offset)
[8] Zhao Y and Cen Y 2013 *Data Mining Applications with R* (Waltham: Elsevier)