Inventory model of goods availability with apriori algorithm

Radhiatul Husna*, Riri Lestari, Yomei Hendra
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Andalas University, Indonesia,

* radhiatulhusna@sci.unand.ac.id

Abstract. In this study, mathematical modeling will be discussed to determine the probability of goods purchase based on other goods. Research on this topic is considered to be important and interesting to study because it describes the implementation of applied mathematics by constructing a mathematical model and solving it using an apriori algorithm related to the application of the probability concept. In this case, the apriori algorithm is used to determine the pattern of dependency relationships between goods and other goods so that the probability of goods purchase based on other goods can be found. To get the output of this research which is in the form of the mathematical modeling for determining the probability of goods being purchased depended on the purchase of other goods, then selecting and classifying the types of goods based on the sales, using the apriori algorithm, constructing models, finding a solution of models and interpreting of the model are conducted.

1. Introduction
Along with the development of competition in the business world, especially in the sales industry, business people are required to find a strategy that can increase sales and marketing of sold products/goods. Before the strategy is designed, it is necessary to continuously analyze the sales results within a certain period. This is intended to ascertain how much an increase in sales occurs. Sales increase can be attempted as much as possible by looking at the relationship of dependence on the sale of goods with other goods. For example, if a consumer buys a toothbrush, then the consumer will also certainly buy toothpaste. By observing the trend pattern of purchasing an item depending on the sale of other items/goods, business people can make a strategy by providing both goods stocks in the same quantity for each interdependent item. In addition, in order to make the maximum sales and increase profits, business people can also put these items together in a container with promised price cuts so that all goods can be sold at the same time.

Such a problem above can be considered as a mathematical modeling. Mathematical modeling is one of the branches of applied mathematics commonly used by mathematicians in predicting a variable or condition in the future. Therefore, to predict how much the sales will increase in the future and the probability of goods purchase depends on the sale of other goods, it is necessary to construct a mathematical model, namely an inventory model of goods availability.

The solution of the mathematical model uses a priori algorithm to determine the relationship pattern of the dependence between an item and other items in the sales transaction data. A priori algorithm is too important to analyze the transaction data and help customers in buying the items/goods they want more easily such that the sales can be risen. This algorithm has been used in various problems, such as A. Nursikuwagus and T.Hartono [2] used the algorithm for analyzing sales with web based and Kuo
M.H., Kusniruk A.W., Borycki E.M., & Greig D. [6] utilized it for detection of adverse drug reaction. In addition, Not only did L. Hakim and A. Fauzy [8] apply the same algorithm for determining the pattern of traffic accidents, but also Pria Nita andRB Fajriya Hakim [10] applied it for analyzing the data pattern of aircraft accidents. In this paper the application of this algorithm is given for analyzing the sales transaction data to construct an inventory model of goods availability.

2. Literature reviews
In this section some reviews about set theories, probabilities, and apriori algorithms will be discussed.

2.1. Set Theories
The following is given some definitions related to the sets and several operations on the sets.

2.1.1. Definition [3] A set is a collection of objects which are unordered and no repetition determined by its elements/members and usually represented by a capital letter. Elements of a set are the objects in the collection.

Suppose that x is an element of a set A. Then, it can be written as x ∈ A. The number of elements in the set A is called a cardinality of A, denoted by #A.

2.1.2. Definition [3] Given A and B are two sets. A is a subset of B, written as A ⊆ B if and only if every element of A is an element of B. A is a proper subset of B if and only if A ⊂ B and A ≠ B. A is a subset of B, denoted by A ⊆ B if and only if A ⊂ B and A = B.

There are several operations on sets as follows:

2.1.3. Union
The union of two sets A and B is the set

\[ A \cup B = \{ x \mid x \in A \text{ or } x \in B \}\]

In a form of a Venn diagram, it can be illustrated as

![Figure 2.1 A union of two sets](image)

2.1.4. Intersection. The intersection of two sets A and B is the set

\[ A \cap B = \{ x \mid x \in A \text{ and } x \in B \}\]

In a form of a Venn diagram, it can be depicted as

![Figure 2.2 An intersection of two sets](image)
2.2. Probabilities

The set of all possible results from an experiment is called a sample space, denoted by $\Omega$, while an event is a subset of the sample space $\Omega$ [8].

2.2.1. Definition[9] Events $A_1, A_2, \ldots$ are called as disjoint events if $A_i \cap A_j = \emptyset, i \neq j$.

2.2.2. Definition[9] For an experiment, $\Omega$ expresses the sample space and $A_1, A_2, \ldots$ are possible events of the experiment. A set function that relates the real value $P(A)$ with each event is called a probability set function and $P(A)$ is called the probability of the event $A$, if the following properties are met:

1. $0 \leq P(A) \leq 1, \forall A$,
2. $P(\Omega) = 1$,
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ if $A_1, A_2, \ldots$ are disjoint.

2.2.3. Definition[9] Two events $A$ and $B$ are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B).$$

Otherwise, two sets $A$ and $B$ are called as dependent events.

2.2.4. Definition[9] A conditional probability of the event $B$ if the event $A$ has occurred, is as follows:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0.$$  

2.3. Apriori Algorithm

Apriori algorithm was first developed by Agrawal and Srikant in 1994. It is an innovative way to find association rules on large scale allowing implication outcomes that consist of more than one item. The following will be given some definitions and explanation related to the apriori algorithm.

**Definition 2.3.1**[1] An itemset is a set of items that occur together.

**Definition 2.3.2** [1] Association rule is a rule that shows a probability of particular items which are purchased together.

For an example, the rule {milk, tea} $\Rightarrow$ {sugar} found in the sales data of a supermarket would indicate that if a customer buys milk and tea together, they are likely to also buy sugar. Such information can be used as the basis for decisions about marketing activities such as e.g. promotional pricing or product placement.

**Definition 2.3.3** [1] Support of an item $X$, denoted by $\text{supp}(X)$ is the ratio of transactions in which an itemset appears to the total number of transactions.

Moreover, the support count of an item $X$ is the frequency or the total number of the item which appears in all transaction. Item sets that meet a minimum support threshold are referred to as frequent itemsets.

**Definition 2.3.4** [1] Confidence is the occurrence of goods; the chances of an item $X$ happening given an item $Y$ has already happened. On the other hand, confidence measure how often goods in $Y$ appear in transactions that contain $X$.

Confidence of rule $X \Rightarrow Y$ denoted by $\text{conf}(X \Rightarrow Y)$ is given as follows.

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}.$$  

(3)
Confidence can also be defined in terms of the conditional probability namely

$$\text{conf}(X \Rightarrow Y) = P(Y \mid X) = \frac{P(X \cap Y)}{P(X)}.$$  \hspace{1cm} (4)

The apriori principle states that any subset of a frequent itemset must be frequent. The following will be given the procedures of apriori algorithm:

1. Given data in the database in regard to sales transaction data.
2. Set the minimum support.
3. Calculate the support or frequency of all goods.
4. Discard the goods with the support which is less than minimum support.
5. Combine two items and count their support.
6. Go back to step 4.
7. Combine three goods and count their support.
8. Go back to step 4.
9. Stop if there is no possible combination.

Then, apriori algorithm can be illustrated as follows.

![Illustration of Apriori Algorithm](image)

**Figure 2.3 Illustration of Apriori Algorithm**

3. Research Methods

In this section, it will be given the steps that have been done during the research period as follows.

1. Preliminary research, includes literature study related to the research topic.
2. Taking data, which is sales transaction data from a minimarket; Minang Mart.
3. Choosing and classifying the types of goods based on the sales transaction data.
4. Using apriori algorithm, namely
   a. Set the minimum number of items/goods in all transactions (minimum support).
   b. Calculate the number of items from sales results (support count).
   c. Calculate the probability percentage of those appeared goods.
   d. Construct the group combination of the possible goods and calculate the support and its percentage.
5. Constructing an inventory model of goods availability.
6. Solving the model.
7. Interpreting the solution of the model.
4. Discussion And Results

Based on the sales transaction data, it can be seen the trend pattern of buying an item depending on the sale of other goods. The purchase can be used as a reference for making product analysis. The following will be explained an example that expresses the description how apriori algorithm works.

Given the sales transaction data \( D \) in the table below.

| TID | Items Bought          |
|-----|-----------------------|
| 1   | eggs, milk, bread     |
| 2   | milk, tea, sugar      |
| 3   | eggs, milk, tea, sugar|
| 4   | tea, sugar            |
| 5   | fruit                 |

In the table above, it can be see the goods bought for each 5-transaction. This data can be stored as a binary \( m \times n \) binary matrix as follows.

| TID | A | B | C | D | E | F |
|-----|---|---|---|---|---|---|
| 1   | 1 | 1 | 0 | 0 | 1 | 0 |
| 2   | 0 | 1 | 1 | 0 | 0 | 1 |
| 3   | 1 | 1 | 1 | 0 | 0 | 1 |
| 4   | 0 | 0 | 1 | 0 | 0 | 1 |
| 5   | 0 | 0 | 0 | 0 | 1 | 0 |

In table 4.2, TID refers to a transaction identity. So there are 5 transactions chosen. In addition, there are 6 goods bought, namely \( A \), \( B \), \( C \), \( D \), \( E \), and \( F \). Let \( A \) be eggs, \( B \) be milk, \( C \) be tea, \( D \) be bread, \( E \) be fruit, and \( F \) be sugar. The binary matrix above gives the value of 1 whenever the item is bought and otherwise 0. Then, suppose that the minimum support is 40% with the support count of 2. It means that every itemset to be considered as a valid candidate must appear at least in two out 5 of the five transactions. At each iteration, the support of candidate item sets is calculated eliminating those which support is under the threshold. So based on the data given in Table 4.2, the support for 1-itemsets is calculated in the following table.

| 1-itemsets | Support Count (SC) | Support |
|------------|--------------------|---------|
| \( \{A\} \) | 2                  | 0,4     |
| \( \{B\} \) | 3                  | 0,6     |
| \( \{C\} \) | 3                  | 0,6     |
| \( \{D\} \) | 1                  | 0,2     |
| \( \{E\} \) | 1                  | 0,2     |
| \( \{F\} \) | 3                  | 0,6     |

In table 4.3, the itemsets \( \{D\} \) and \( \{E\} \) are discarded since the support does not meet the minimum support threshold. So, there are 4 goods, namely \( A \), \( B \), \( C \), and \( F \) that will be considered for the next step. Then, the combination of 2-itemsets among these goods will yield 6 possible combinations which is \( C(4,2) = 4!/[2!(4-2)!] = 24/(2*2)=6 \).

Note that it can be obtained using the formula

\[
C(n, r) = \frac{n!}{r!(n-r)!}
\]  \hspace{1cm} (5)
which is the total number of all possible combinations for choosing \( r \) elements at a time from \( n \) distinct elements without considering the order of the elements. Among these combination, the support will be calculated as given in Table 4.4 below.

### Table 4.4 2-itemsets with their support

| 2-itemsets | SC | Support |
|------------|----|---------|
| \{A, B\}  | 2  | 0.4     |
| \{A, C\}  | 1  | 0.2     |
| \{A, F\}  | 1  | 0.2     |
| \{B, C\}  | 2  | 0.4     |
| \{B, F\}  | 2  | 0.4     |
| \{C, F\}  | 3  | 0.6     |

It can be seen from Table 4.4 that only four combinations have satisfied the minimum support since \{A, C\} and \{A, F\} do not meet the minimum support. Hence, these 4-combinations are selected for the next step. Moreover, the combination of 3-itemsets among 4 items will have 4 possible combinations and their support can be seen in the following table.

### Table 4.5 3-itemsets with their support

| 3-itemsets | SC | Support |
|------------|----|---------|
| \{A, B, C\} | 1  | 0.2     |
| \{A, B, F\} | 1  | 0.2     |
| \{A, C, F\} | 1  | 0.2     |
| \{B, C, F\} | 2  | 0.4     |

Table 4.5 leads to the only one combination left that meets the minimum support threshold summarized in the table below.

### Table 4.6 Final frequent itemsets

| 3-itemsets | SC | Support |
|------------|----|---------|
| \{B, C, F\} | 2  | 0.4     |

Using the result in table 4.6, the possible association rules with their support and confidence can be made as follows.

### Table 4.7 The possible association rules

| Association rules | Support | Confidence |
|-------------------|---------|------------|
| \{B\} \Rightarrow \{C, F\} | 40% | 67% |
| \{C\} \Rightarrow \{B, F\} | 40% | 67% |
| \{F\} \Rightarrow \{B, C\} | 40% | 67% |
| \{C, F\} \Rightarrow \{B\} | 40% | 67% |
| \{B, F\} \Rightarrow \{C\} | 40% | 100% |
| \{B, C\} \Rightarrow \{F\} | 40% | 100% |

From the table 4.7, it is chosen the best association rules, namely

1. \{B, F\} \Rightarrow \{C\}.
2. \{B, C\} \Rightarrow \{F\}.

The rule \{B, F\} \Rightarrow \{C\} will indicate that if people buy milk (\(B\)) and sugar (\(F\)), then they will buy tea (\(C\)) with the confidence of 100%. However, if they buy tea, then they will buy milk and sugar with the confidence of 67%. Furthermore, the rule will state that if people buy milk and tea, then they will buy sugar with the confidence of 100%. In contrast, if they buy sugar, then they will buy milk and tea with the confidence of 67%. In this case, the items which are milk, tea, and sugar should be put together so that their position is closed each other.
Based on the example that have explained above, the inventory model can be constructed as follows. Given the database $D$ which is sales transaction data consisting of events $T_1, T_2, \ldots, T_m$. So $D$ can be written as $D = \{T_1, T_2, \ldots, T_m\}$. Then, let there be the $n$-th itemset $X_n$ that is a subregion of the event $T_m$, that is $X_n \subseteq T_m$. The database $D$ can be written as a binary matrix with the size of $m \times n$ in which $X_{an} = \begin{cases} 1, \text{if the item is bought} \\ 0, \text{otherwise} \end{cases}$.

So the inventory model can be written as a binary $m \times n$ matrix as $D_{mn}$ with the confidence defined as

$$\text{conf}(X_a \Rightarrow X_b) = \frac{\text{supp}(X_a \cup X_b)}{\text{supp}(X_a)}$$  \hspace{1cm} (6)

or

$$\text{conf}(X_a \Rightarrow X_b) = P(X_b | X_a) = \frac{P(X_a \cap X_b)}{P(X_a)}.$$  \hspace{1cm} (7)

Next, from that model, the association rule $X_a \Rightarrow X_b$ would like to be found such that $X_a$ and $X_b$ are the subregions of event $T_m$, that is $X_a \subseteq T_m$ and $X_b \subseteq T_m$.

In which $X_a \cap X_b = \emptyset$.

Now, the model will be solved for the previous example. Suppose that $D = \{T_1, T_2, T_3, T_4, T_5\}$ in which there are 5 goods available, that is $X = \{X_1, X_2, X_3, X_4, X_5\}$. Let $T_1 = \{1,1,0,1,0,0\}, T_2 = \{0,1,1,0,1,1\}, T_3 = \{1,1,1,0,0,1\}, T_4 = \{0,0,1,0,0,1\}$, and $T_5 = \{0,0,0,0,1,0\}$. So the inventory model $D$ is a binary $5 \times 6$ matrix, namely

$$D_{5 \times 6} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Then, using apriori algorithm will yield the association rule as follows.

1. $\{X_2, X_6\} \Rightarrow \{X_3\}$.

2. $\{X_2, X_3\} \Rightarrow \{X_6\}$.

The first rule means that if people buy $X_2$ and $X_6$, then they will buy $X_3$ with the confidence of 100%. Furthermore, the second rule indicates that if people buy $X_2$ and $X_3$, then they will buy $X_6$ with the confidence of 100%.

Using the same way as the previous example, the inventory model is also obtained from a 22-sale transaction data chosen for the case study at Minang Mart. In this case, the model is represented as a binary $22 \times 51$ matrix or $D_{22 \times 51}$ in which there are 51 goods that are considered in those 22 transactions. Given the minimum support of 13%. Then, solving the model applying apriori algorithm will give the solution of association rules, namely

1. $\{\text{lifeboy}\} \Rightarrow \{\text{pepsodent}\}$ with confidence of 0.6.

2. $\{\text{pepsodent}\} \Rightarrow \{\text{pepsodent}\}$ with confidence of 0.428.

The first rule says that if people buy lifeboy (soap), then they will buy pepsodent (tooth paste) with the confidence of 60%. Otherwise, the second rule mentions that if people buy pepsodent, they will buy
lifeboy with the confidence of 42.8%. This result shows that the lifeboy and tooth paste should be placed close to each other. The purchase of lifeboy will increase the sales of pepsodent with the probability of 60%. Otherwise, the purchase of pepsodent will increase the sales of lifeboy with the probability of 42.8%. So the purchase of lifeboy is better than that of pepsodent in order to boost another item. Hence, lifeboy should be supplied more as a stock than pepsodent.

5. Conclusion
Nowadays, competition in business world, especially in sales industry leads the business people to find out a strategy to increase their sales. Based on the sales transaction data, the transactions can be examined to decide what items typically appear together, e.g., which items customers typically buy together in a database of supermarket transactions. This in turn gives insight into questions such as how to market these products more effectively, how to classify them in store layout or product packages, or which items to offer on sale to boost the sale of other goods. One of the methods that can be applied is apriori algorithm.

Apriori algorithm is an innovative method to determine association rules on large scale allowing implication outcomes that consist of more than one item. Finding that rules will require to find the confidence and the support of all goods with all possible combination itemsets. Based on 22 transactional data, the inventory model is represented as a binary matrix $D_{22 \times 51}$ in which there are 51 goods that are considered in those 22 transactions. Given the minimum support of 13%. The solution of the model will lead to the association rules that if people buy lifeboy (soap), then they will buy pepsodent (tooth paste) with the confidence of 60%. Otherwise, if people buy pepsodent, they will buy lifeboy with the confidence of 42.8%. In this case, the purchase of lifeboy is better than that of pepsodent in order to increase another item. Moreover, lifeboy should be supplied more as a stock than pepsodent.

References
[1] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), Advances in knowledge discovery and data mining (pp. 307–328). Menlo Park, CA: AVAAI Press.Springer.
[2] A. Nursikuwagus and T. Hartono. (2016). Implementation of Apriori Algorithm for Analyzing Sales with Web Based. Journal Simetris.
[3] Dave Witte Morris & Joy Morris. (2016). Proofs and Concepts the Fundamentals of Abstract Mathematics. University of Lethbridge.
[4] Frank R.G, William P.F, & Steven B.H. (2010). A First Course in Mathematical Modeling. USA: Brooks/Cole.
[5] Hermawati, Fajar Astitu. (2013). Data Mining. Yogyakarta : Andi Offset.
[6] Kuo M.H, Kusiruk A.W, Borycki E.M, & Greig D. (2009). Application of the Apriori Algorithms for Adverse Drug Reaction Detection. JournalNCBI.
[7] Kusrini dan Luthfi, E.T. (2009). Mining Data Algorithms. Yogyakarta: Andi Offset.
[8] L.Hakim & A.Fauzy. (2015). The Pattern Determination of Traffic Accident Relationship Using Association Rule Methods with Apriori Algorithms. Research Gate.
[9] L.J.Bain dan M.Engelhardt. (2000). Introduction to Probability & Mathematical Statistics. Cengage Learning, Inc.USA: Brooks/Cole.
[10] Pria Nita & RB Fajriya Hakim. (2015). Application of Association Rule Methods Using Apriori Algorithms to Analyze the Data Pattern of Aircraft Accidents from 1967-2014 in Indonesia. Proceeding of National Seminar on Mathematics and Educational Mathematics, UMS
[11] Rohayati, Sinambela Sarton. (2008). The Effect of Promotion Mix on Sales Increase at PT. Aquasolve Sanari. Ekonomika. Vol IV1-19.
[12] Salahuddin, Muhammad & Sukamto, Rosa Ariani. (2011). Software Engineering. Bandung : Modula.