Market Basket Analysis with Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) on Outdoor Product Sales Data

Wiwit Pura Nurmayanti[1,*], Hanipar Mahyulis Sastriana[2], Abdul Rahim[3]
Muhammad Gazali[4], Ristu Haiban Hirzi[5], Zuhut Ramdani[6] Muhammad Malthuf[7]

[1,2,4,5,6] Department of statistics, Faculty of Mathematics and Natural Sciences
Universitas Hamzanwadi, Lommbok Timur, Nusa Tenggara Barat 83611, Indonesia.
[3]Department of Pharmacy, Faculty of Health
Universitas Hamzanwadi, Lommbok Timur, Nusa Tenggara Barat 83611, Indonesia.
Indonesia.
[7]Social Studies Education, Faculty of Teacher Training and Education
Universitas Islam Negeri Mataram, Mataram, Nusa Tenggara Barat 83125, Indonesia

[*] Corresponding author:
Email: wiwit.adiwinata3@gmail.com

***Abstract**.*

Indonesia is an equatorial country that has abundant natural wealth from the seabed to the top of the mountains, the beauty of the country of Indonesia also lies in the mountains that it has in various provinces, for example in the province of West Nusa Tenggara known for its beautiful mountain, namely Rinjani. The increase in outdoor activities has attracted many people to open outdoor shops in the West Nusa Tenggara region. Sales transaction data in outdoor stores can be processed into information that can be profitable for the store itself. Using a market basket analysis method to see the association (rules) between a number of sales attributes. The purpose of this study is to determine the pattern of relationships in the transactions that occur. The data used is the transaction data of outdoor goods. The analysis used is the Association Rules with the Apriori algorithm and the frequent pattern growth (FP-growth) algorithm. The results of this study are formed 10 rules in the Apriori algorithm and 4 rules in the FP-Growth algorithm. The relationship pattern or association rule that is formed is in the item "if a consumer buys a portable stove, it is possible that portable gas will also be purchased" at the strength level of the rules with a minimum support of 0.296 and confidence 0.774 at Apriori and 0.296 and 0.750 at FP-Growth.

***Keywords**: Outdoor goods, MBA (Market Basket Analysis), Association Rule, Apriori Algorithm, FP-Growth Algorithm.*

1. INTRODUCTION

Indonesia is one of the country's destinations that is rich in natural tourism, this will certainly attract tourists to explore tourist destinations in Indonesia. The beauty of the country of Indonesia also lies in the mountains that it has in various provinces, one of which is in the province of West Nusa Tenggara, namely Mount Rinjani. Currently, the term climbing mountains and hills around it is a new trend for young people in particular, climbing activities with the aim of exploring natural beauty and preserving

nature. With difficult terrain and unpredictable weather, of course outdoor activities have a high risk, for that we need good planning, including in terms of equipment and supporting tools. Outdoor equipment meets all the physical needs of consumers while in the wild,. The increase in outdoor activities in the province of West Nusa Tenggara has attracted a lot of interest from someone to open an outdoor shop in the West Nusa Tenggara region, especially the area close to Mount Rinjani. Currently, outdoor retail is growing and developing rapidly along with the increasing number of young people who choose outdoor activities, one of which is in the Selong sub-district. Outdoor footprints provide for climbers' needs. Every day there are sales transactions of goods resulting in transaction data. Sales transaction data can be processed into information that can be profitable for the store itself. In this study, the writer will try to conduct an experiment on the transaction data of outdoor sales of goods by using a market basket analysis method to see the association (rules) relationship between a number of sales attributes. The algorithms that will be used are the apriori algorithm and the frequent pattern growth (FP-growth) algorithm. By carrying out this research, it is hoped that it can provide results in the form of useful information for related parties in carrying out the managerial decision-making process, especially those related to the formulation or creation of marketing and sales strategies, especially for goods on Tapak tilas outdoor.

Literature Review
Research related to Data Mining Market Basket Analysis was conducted by Goldie and Dana Indra Sensuse (2012) with the title Application of the Data Mining Market Basket Analysis Method to Book Product Sales Data Using the Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) by taking a case study of printing. PT. Gramedia [4]. The purpose of his research is to see the association (correlation) between a number of sales attributes. The formation of rules uses these two algorithms, and the conclusion of the research is that there are 10 rules that are formed. The strength level of the association rules shows that the association rules generated by the Apriori algorithm have a higher level of strength than those generated by the FP-growth algorithm.

Further research related to the Fp-Growth Algorithm has been carried out by Arnandia and Arief (2018) with the title Searching for Consumer Purchasing Patterns Using the Fp-Growth Algorithm [2]. Arnandia and Arief discussed how to implement one of the algorithms in data mining, namely the FP-Growth algorithm. This algorithm is an alternative that can be used to determine data sets that often appear in data sets. Produces items that appear frequently, simultaneously, with 5 formed rules. If someone purchases 3kg LPG Gas, that person also purchases Djarum Super 12 with a support value of 0.02 and a confidence value of 1.00.

Another study was conducted by Maulidiya H and Arief Jananto (2020) with the title Data Mining Association Using Apriori Algorithms and Fp-growth as a Basic Consideration for Determining Food Packages [7]. Maulidiya H and Arief Jananto

implemented and then compared two association rule algorithms, namely the a priori algorithm and FP-Growth to produce the highest frequent itemsets. So that the results of this study will later find out what algorithm is best in forming frequent itemset. The Apriori Algorithm rule with a minimum support of 0.06 and confidence 0.01 obtained 8 rules. Then the FP-Growth Algorithm also succeeded in forming 14 rules.

The difference with the research that the author does is that the author raises transaction data on the sale of outdoor goods, in this case, which is the center of today's community activities. In the above studies, it is explained how the use of the Association Rules method in various different case studies, while in the research the author explains how the form of the implementation of the MBA (Market Basket Analysis) uses the Apriori algorithm and the FP-Growth algorithm using 2 software, namely Rstudio and Rapid Miner. And this study complements the research of the MBA (Market Basket Analysis) method with Apriori algorithm and FP-Growth algorithm.

II. METHODS

3.1. Data Mining

Data miningis the process of extracting valid, unknown, understandable and actionable information from large databases used for critical decision-making by Connolly and Begg [3]. Data mining focuses on data analysis and information retrieval in the form of hidden patterns and relationships from data using certain software techniques. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases by Kusrini and Emha [6].

3.2. MBA (Market Basket Analysis)

Han and Kamber [5] presented Market basket analysis is a process that analyzes the purchasing habits of customers by finding associations between different items in the customer's shopping cart. This association is needed to find out what items the customer may buy at the same time. This analysis is very helpful for business owners in improving their marketing strategy. Market basket analysis can be analyzed using the association rule. The purpose of market basket analysis is to find out which products may be purchased simultaneously.

3.3. Association Rules

Association analysis or association rule mining is a data mining technique to find association rules between a combination of items. The interestingness measure that can be used in data mining is Support, which is a measure that shows how much domination an item or itemset is from the entire transaction. Confidence, is a measure that shows the relationship between two items conditionally (based on certain conditions) by Kusrini and Emha [6].

- **Apriori Algorithm**

Apriori algorithm including the types of association rules in data mining. One of the stages of association analysis that has attracted the attention of many researchers to produce efficient algorithms is frequent pattern mining, the importance of whether an association can be identified by two benchmarks, namely: support and confidence. Support (value of support) is the percentage of the combination of these items in the data base, while confidence (value of certainty) is the strength of the relationship between items in the association rule. The initial stage in the a priori algorithm is the analysis of high frequency patterns, namely by finding combinations of items that meet the minimum requirements of the support value in the database. The support value of an item is obtained by the following formula by Agung and Nurhadiyono [1] :

$$support = \frac{number\ of\ transactions}{total\ transactions} \times 100\%$$

The item set frequency shows the itemset that has an appearance frequency of more than the specified minimum value. The next step is the formation of association rules, that is, after all high frequency patterns are found, then you can search for association rules that meet the minimum confidence requirements, by calculating the value A => B. The formula is as follows :

$$confidence = \frac{number\ of\ transaction\ containing\ A\ and\ B}{number\ of\ transaction\ containing\ A} \times 100\%$$

- **FP-Growth Algorithm**

    FP-growth is an alternative algorithm that can be used to determine the most frequent itemset in a data set. FP-growth uses a different approach from the paradigm used in the Apriori algorithm.

3.4. **Analysis step**

In this research, the application of the MBA algorithm is carried out by collecting data on the sale of outdoor goods. The steps in the analysis are as follows:
1. Recap sales data
2. Descriptive analysis
3. Determine the value of minimum support and confidence
4. The formation of associative rules with the Apriori algorithm and FP-Growth is as follows:
    a) Find association rules
    b) Make combinations of 2 dataset items
    c) Make combinations of 3 dataset items
    d) Select a frequency greater than or equal to the minimum limit that has been determined
    e) Create visuals with Graph Rules

III. RESULT AND DISCUSSION
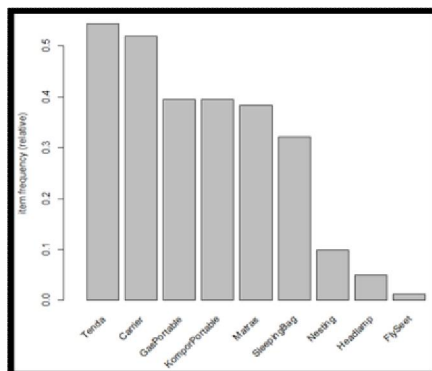
3.1 Descriptive analysis

Fig 1. Frequency plot of goods

Interpretation:

Figure 1 shows the frequency of outdoor goods most often purchased by consumers. It can be seen that the Tent is the number one item that is often purchased, then the mountain bag (Carrier) is the second item that is often purchased, and for the item with the least transactions among other items, namely Flyseet.

3.2. Implementation of the Apriori Algorithm

The following are the results of a priori algorithm analysis on RStudio, testing data using the a priori algorithm produces 10 rules, as shown in the picture :

```
> inspect(rules)
     lhs                               rhs                 support   confidence coverage  lift     count
[1]  {Matras}                       => {Carrier}           0.2962963 0.7741935  0.3827160 1.493088 24
[2]  {GasPortable}                  => {KomporPortable}    0.2962963 0.7500000  0.3950617 1.898438 24
[3]  {KomporPortable}               => {GasPortable}       0.2962963 0.7500000  0.3950617 1.898438 24
[4]  {GasPortable,SleepingBag}      => {KomporPortable}    0.1481481 0.8571429  0.1728395 2.169643 12
[5]  {KomporPortable,SleepingBag}   => {GasPortable}       0.1481481 0.9230769  0.1604938 2.336538 12
[6]  {KomporPortable,Matras}        => {Carrier}           0.1234568 0.8333333  0.1481481 1.607143 10
[7]  {Carrier,KomporPortable}       => {Matras}            0.1234568 0.7142857  0.1728395 1.866359 10
[8]  {Carrier,GasPortable}          => {KomporPortable}    0.1111111 0.7500000  0.1481481 1.898438  9
[9]  {Carrier,GasPortable}          => {Tenda}             0.1111111 0.7500000  0.1481481 1.380682  9
[10] {Carrier,KomporPortable}       => {Tenda}             0.1234568 0.7142857  0.1728395 1.314935 10
```

Fig 2. *Apriori algorithm rules*

Interpretation:

The first rule, has a support value of 0.297, meaning that 29.7% of the total transactions consist of "Matras" and "Carriers" of the entire transaction. A confidence level of 0.774 indicates that of all transactions containing "Matras", 77.4% also contain "Carriers" in their basket. Meanwhile, the value of the lift is 1.49, indicating that "Carriers" are 1.49 times more likely to be bought by customers who buy "mattresses". The second rule, if you buy a Portable Stove, is to buy Protebel Gas with 29.6% support and 75% confidence. And so on until rules 10.
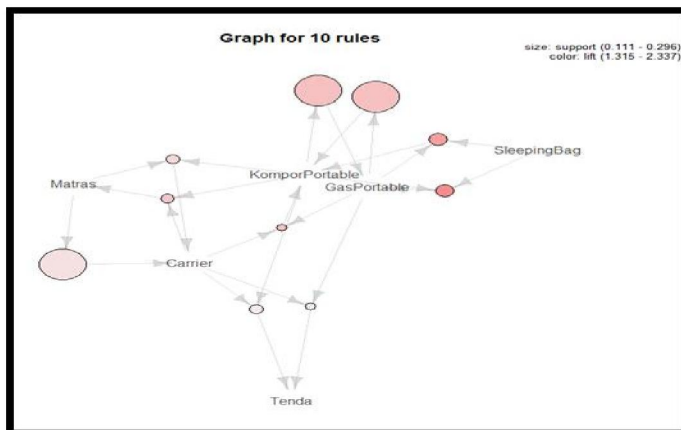
Fig 3. Graph of Apriori algorithm rules

Interpretation:

Mattresses with Carriers, portable stoves with portable gas have a higher support value compared to other items, which are characterized by a larger circle size than other items. and in items of portable stoves, portable gas, with a sleeping bag has a higher lift ratio compared to other items, which is interpreted by a darker color with other items.

3.3. Implementation of the FP-Growth Algorithm

The following are the results of the analysis of the FP-Growth algorithm in the RapidMiner software, testing using the FP-Growth algorithm produces 4 rules, as shown in the picture :
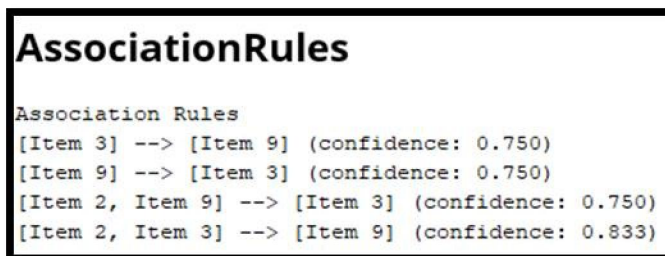
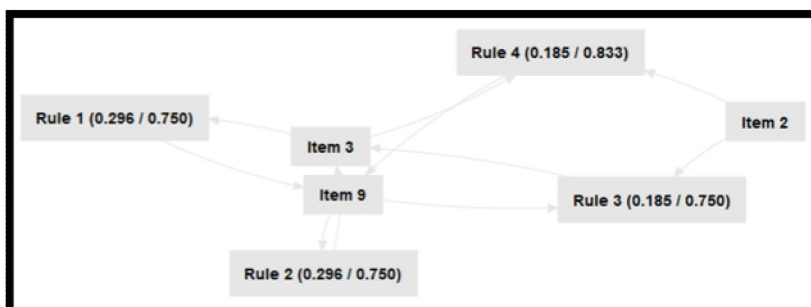

Fig 4. *FP-Growth algorithm rules*

Fig 5. *Graph of FP-Growth algorithm rules*

Table 1. Summary of the FP-Growth Rules

Rules	Premises	Conclusion	Support	Confidence	Lift
Rules 1	Portable stove	Portable gas	0.296	0750	1.90
Rules 2	Portable gas	Portable stove	0.296	0750	1.90
Rules 3	Carrier, portable gas	Portable stove	0.185	0750	1.90
Rules 4	Carrier, portable stove	Portable gas	0.185	0833	2.11

Interpretation:

The first *Rules* has a support value of 0.296, meaning that there are 29.6% of the total transactions consisting of portable stoves and portable gas from all transactions, a confidence level of 0.750 indicates that of all transactions containing portable stoves, 75% also contain portable gas in their shopping cart. Whereas the Lift value of 1.90 indicates that portable gas is 1.90 times more likely to be purchased by customers who buy portable stoves, and so on.

IV. CONCLUSION

1. Transaction patterns in the sale of outdoor goods that are formed with the Apriori algorithm are as many as 10 rules: the strength level of the rules at minimum support is 0.296, the confidence is 0.774 and the lift value is 1.49 indicating the pattern of consumers who buy Portuguese stove items also have the possibility of buying portable gas items.
2. Transaction patterns in the sale of outdoor goods that are formed with the FP-Growth algorithm are as many as 4 rules: the strength level of the rules at minimum support minimum support is 0.296, the confidence is 0.750 and the lift value is 1.90 shows that the pattern of consumers who buy portable stove items also have the possibility of buying portable gas items.

Recommendation for Further Researc

1. Further research examines more theoretical sources and explores the Market Basket Analysis method (Assosiacion Rules with Algorithms).
2. Subsequent research uses a different case study that has complete data and uses other algorithms that exist in the Association Rules.

REFERENCES

[1] Agung M.T,& Nurhadiyono B.( ) . *Penerapan Data Mining Pada Data Transaksi Penjualan Untuk Mengatur Penempatan Barang Menggunakan Algoritma Apriori*. Jurusan Teknik Informatika Fakultas Ilmu Komputer. Universitas Dian Nuswantoro. Semarang.

[2] Arnandia R.W, and Arief Jananto. 2018. *Mencari Pola Pembelian Konsumen Menggunakan Algoritma Fp-Growth.* Program Studi Sistem Informasi. Situbank University.

[3] Connolly T, Begg C. 2005. Database System: A Practical Approach to Design,Implementation, and Management. Fourth Edition, Addison Wesley, England.

[4] Goldie, dan Dana Indra Sensuse. 2012. Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori

dan Frequent Pattern Growth (FP-Grwoth) : Studi Kasus Percetakan PT.Gramedia. Jurusan Ilmu Komputer. Universitas Indonesia.

[5]   Han dan Kamber. 2006. Data Mining : Concepts and Techniques, 2nd ed.

[6]   Kusrini dan Emha Taufiq Luthfi. 2009. Algoritma Data Minin. Penerbit Andi.

[7]   Mauliiya H and Arief Jananto. 2020. *Asosiasi Data Mining Menggunakan Algoritma Apriori Dan Fpgrowth Sebagai Dasar Pertimbangan Penentuan Paket Sembako*. Program Studi Sistem Informasi. Situbank University.