# APPLICATION OF DATA MINING TO DETERMINE THE LEVEL OF FISH SALES IN PT. TRANS RETAIL WITH FP-GROWTH METHOD

**Saut Parsaoran Tamba[1], Mario Sitanggang[2], Bimo Christhoper Situmorang[3], Gracia Laura Panjaitan[4], Marlince Nababan[5]**

[1,2,3,4]Information Systems Study Program, Faculty of Technology and Computer Science, North Sumatra, Prima Indonesia University

E-mail:
sautparsaorantamba@unprimdn@ac.id, mariostanggang02@gmail.com, bimositumorang24@gmail.com, gracia970@gmail.com, marlince@unprimdn.ac.id

## Abstract

| Article Info | |
|---|---|
| Received, 01 Juni 2022<br>Revised 20 Juni 2022<br>Accepted 22 Juni 2022 | PT. Trans Retail Indonesia is one of the shopping places in the city of Medan. Trans Retail Indonesia is engaged in providing raw materials such as selling fish. However, at Trans Retail Indonesia, sales data collection still uses a manual system, so this store has not been able to determine which type of fish has the highest level of sales. So this affects the availability of goods and results in a lack of stock in this store. So we need a data mining system with the FP-Grwoth method to analyze fish sales data so that the results of the analysis become a reference for stores in determining the supply of fish stocks. The results of the analysis carried out by researchers from the data obtained are the fish that is most in demand is the Jengka Split with a value of 90%, and if you take a split Jengka fish, you will take anchovy buntiaw with a value of 54%. If you take an anchovy, it will take a large anchovy and will take a white Peda fish with a value of 100%. |

Keywords: data mining, fp-growth, rapidminer

## 1. INTRODUCTION

In the development of information technology in the current era which has provided developments that are very helpful in growing the amount of data that can be collected and stored in a very large database and in collecting very large data, methods are needed that can help. a technology that can be used is data mining. Data mining is the process of looking for interesting patterns or information in selected data using techniques or methods.[1-3]. Data mining can be concluded as a process of processing data both small and large by extracting information from the database by extracting to get new knowledge that was previously unknown and the results can be useful for the future.[4-5], where in the application of data mining, one of the methods used is FP-Growth. FP-Growth is a frequent itemset search algorithm obtained from FP-Tree by exploring the tree from bottom to top [6-7]. The FP-Growth algorithm is a development of the Apriori algorithm. So that the shortcomings of the Apriori algorithm are corrected in the FP-Growth algorithm. The FP-Growth algorithm uses the concept of tree construction in the construction of itemset, not using generate candidates like the Apriori algorithm [8-10].

Company PT. Trans Retail Indonesia is one of the shopping places in the city of Medan. PT Trans Retail Indonesia is engaged in the provision of raw materials, one example of which is in the sales of fish. As time goes by, an increase in consumers occurs in the store, so that the increasing number of consumers affects the amount of stock available, especially in the availability of fish stocks.

With the increasing number of consumers, PT. Trans Retail Indonesia is still experiencing kThe lack of fish stocks is due to fish demand that is too high, due to an unsupportive system. A method is needed to find out what fish stocks must be prepared to anticipate a sudden shortage of goods. To overcome this problem, it is necessary to analyze data from fish sales for stock supplies by applying the FP-Growth method.

In previous studies similar but with different methods such as "*Implementation of Data Mining sales of cosmetic products at PT. Natural Nusantara uses the Apriori (2020) algorithm"*[11].In this

study, candidate generation is used through a set of items that are often generated. And in the research being conducted now is to use the FP-Growth method using Prefix-Tree(FP-Tree) to store data in compressed form where the FP-Growth method uses divide and conquer strategies to determine the set of items that often appear. In research "*Application of Data Mining K-Means Clustering Method for Sales Analysis at Banten Hijab Fashion Store*[12]. The data used in this study was only 3 months, while in this study, transaction data was used for 2 years.

**2.Method**

In this research, in determining the highest level of sales using the FP-Growth method, it requires sales data consisting of 1000 sales data for 2 years. The stages of research carried out in the application of Data Mining are shown in table 1.

Table 1. Research Stage
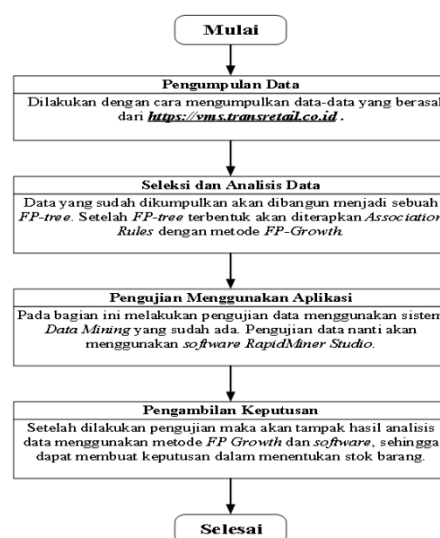


**3.RESULTS AND DISCUSSION**

**3.1Fish Sales Data**

Fish sales data that will be used in manual calculations are shown in Table 2.

Table 2.Fish Sales Data

| No | Date | Fish Name | Total Sales |
|----|------|-----------|-------------|
| 1 | Feb 05,2018 | Teri JENGKI SHOP | 3 |
| 2 | Feb 05,2018 | Teri Padang Besar | 0 |
| 3 | Feb 05,2018 | DRY SHRIMP | 0 |
| 4 | Feb 05,2018 | TERI BUNTIAW | 3 |
| 5 | Feb 05,2018 | OFFICIAL TERI JENGKI | 3 |
| 6 | Feb 05,2018 | LIMBAT SALE FISH | 10 |
| 7 | Feb 05,2018 | LARGE SALT REBON SHRIMP | 2 |
| 8 | Feb 05,2018 | TERI MEDAN | 0 |
| 9 | Feb 05,2018 | WHITE PEDA | 0 |
| 10 | Feb 21,2018 | Teri JENGKI SHOP | 2 |
| 11 | Feb 21,2018 | Teri Padang Besar | 0 |
| 12 | Feb 21,2018 | DRY SHRIMP | 0 |
| 13 | Feb 21,2018 | TERI BUNTIAW | 2 |
| 14 | Feb 21,2018 | OFFICIAL TERI JENGKI | 2 |
| 15 | Feb 21,2018 | LIMBAT SALE FISH | 5 |

| No | Date | Fish Name | Total Sales |
|----|------|-----------|-------------|
| 16 | Feb 21,2018 | LARGE SALT REBON SHRIMP | 1.9 |
| 17 | Feb 21,2018 | TERI MEDAN | 4.9 |
| 18 | Feb 21,2018 | WHITE PEDA | 0 |
| .. | ……………… | ………………… | ………… |
| 100 | Dec 05,2020 | TERI MEDAN | 0 |

To simplify the calculation process, the following is the name of the fish and the coding of each fish.

Table 3. Fish coding

| Data | Code |
|------|------|
| Teri JENGKI SHOP | 1 |
| Teri Padang Besar | 2 |
| DRY SHRIMP | 3 |
| TERI BUNTIAW | 4 |
| OFFICIAL TERI JENGKI | 5 |
| LIMBAT SALE FISH | 6 |
| LARGE SALT REBON SHRIMP | 7 |
| TERI MEDAN | 8 |
| WHITE PEDA | 9 |

**3.2 Fish Sales Transaction Data**

Then the transaction or purchase process is coded, if the fish is purchased, the code is (1), if the fish is not purchased, the code is (0) and the transaction is as follows:

Table 4. Sales Transactions

| Sunday | code | | | | | | | | |
|--------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 5 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 6 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 7 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| 8 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 9 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| 10 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| … | | | | | | | | | |
| 100 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Total | 90 | 50 | 17 | 49 | 47 | 49 | 49 | 61 | 33 |

**3.3 Calculation Stage**

The process of forming itemset with a minimum amount of support 20%, The formula for using the formula is as follows:

$$\text{Support (A)} = \frac{\sum \text{Transaksi yang mengandung A}}{\sum \text{Transaksi}} \times 100\%$$

1.Calculation for 1 itemset

Calculation      of      one      item      set      using      the      following      formula:

$$\frac{Ikan\ 1}{Jumlah\ data}x100\% = \frac{90}{100}x100\% = 90\%$$

With the following results:

Table 5. Results 1 Itemset

| Item Set | Presentation |
|----------|--------------|
| 1 | 90% |
| 2 | 50% |
| 4 | 49% |
| 5 | 47% |
| 6 | 49% |
| 7 | 49% |
| 8 | 61% |
| 9 | 33% |

2. Calculation For 2 Itemset

The process of forming the minimum number of itemset*support*20%, The formula for using the formula is as follows:

$$\frac{Ikan\ 1\ dan\ 2}{Jumlah\ data}x100\% = \frac{47}{100}x100\% = 47\%$$

Table 6. Results of 2 Itemset

| Itemset | Presentation | Itemset | Presentation | Itemset | Presentation |
|---------|--------------|---------|--------------|---------|--------------|
| 1.2 | 47% | 4.1 | 47% | 6.8 | 28% |
| 1.4 | 47% | 4.2 | 27% | 7.1 | 45% |
| 1.5 | 43% | 4.5 | 26% | 7.4 | 29% |
| 1.6 | 41% | 4.7 | 29% | 7.5 | 43% |
| 1.7 | 45% | 4.8 | 33% | 7.8 | 32% |
| 1.8 | 54% | 5.1 | 43% | 8.1 | 54% |
| 1.9 | 33% | 5.2 | 22% | 8.2 | 30% |
| 2.1 | 47% | 5.4 | 26% | 8.4 | 33% |
| 2.4 | 27% | 5.7 | 43% | 8.5 | 30% |
| 2.5 | 22% | 5.8 | 30% | 8.6 | 28% |
| 2.7 | 23% | 6.1 | 41% | 8.7 | 32% |
| 2.8 | 30% | 6.7 | 22% | 9.1 | 33% |

3. Calculation For 3Itemset

The calculation of the third combination is the same as the calculation of combinations 1 and 2. Then the result of combination 3 with the minimum number*support*20%, as follows.

$$\frac{Ikan\ 1, 2, dan\ 9}{Jumlah\ data}x100\% = \frac{47}{100}x100\% = 47\%$$

Table 7. Results of 3 Itemset

| Itemset | Presentation |
|---------|--------------|
| 1,2,9 | 33% |
| 1,6,9 | 33% |
| 1,9,2 | 33% |
| 1,9,6 | 33% |
| 2.1.9 | 33% |
| 2.9.1 | 33% |
| 6.1.9 | 33% |
| 2,4,1 | 27% |
| 1.5,2 | 22% |
| 1,5,4 | 25% |

From the combination above, the 10 itemset with the highest percentage are taken, the results are as follows.

Table 8. Highest Support Results

| Itemset | Support |
|---------|---------|
| 1.2 | 47% |
| 1.4 | 47% |
| 1.8 | 54% |
| 1,2,9 | 33% |
| 1,6,9 | 33% |
| 1,9,2 | 33% |
| 1,9,6 | 33% |
| 2.1.9 | 33% |
| 2.9.1 | 33% |
| 6.1.9 | 33% |

In the last stage of the calculation to find the frequency between items with a support value > 20% and confidence > 70%. This stage is done by finding the confidence value with the formula below:

$$Confidence(A|B) = \frac{Transaksi\ yang\ mengandung\ A\ dan\ B}{Transaksi\ yang\ mengandung\ A}X100$$

Table 9. Confidence Results

| Itemset | Confidence |
|---------|-----------|
| 1,2->2 | 47/50=94% |
| 1,4->4 | 47/49=95.91% |
| 1.8->8 | 54/61=88.5% |
| 1,2,9->9 | 33/33=100% |
| 1,6,9->9 | 33/33=100% |
| 1,9,2->9 | 33/33=100% |
| 1,9,6->9 | 33/33=100% |
| 2,1,9->9 | 33/33=100% |
| 2,9,1->9 | 33/33=100% |
| 6,1.9->9 | 33/33=100% |

**3.4 Development Stage of the Frequen Pattern Tree**

From*confident*Above is a decision tree. Where the first itemset of the twig is made to the right, the picture is as follows:
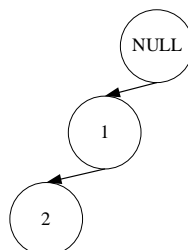


Figure 1. Formation of FP-Tree

Then the second itemset has a branch, while the branch is located at number 1 so that the branch is pointing to the left
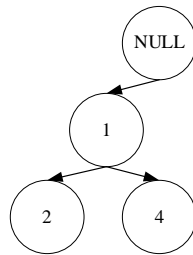
Figure 2. Formation of FP-Tree
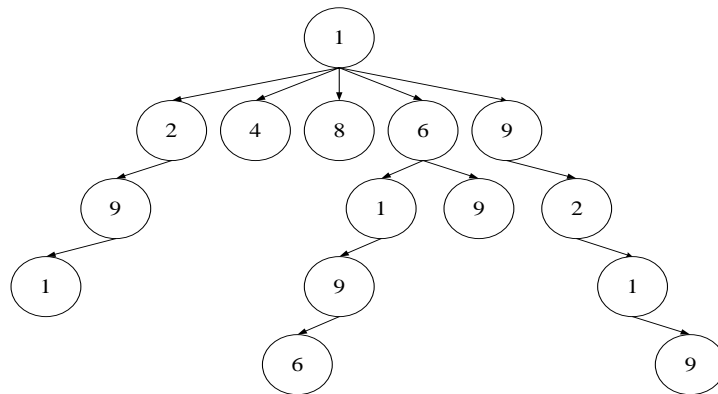
And so on so that the result is as follows:

Figure 3. Results of the Formation of FP-Tree

Information
1. If you buy item 1 and buy item 2, it has a support value of 47% and has a confidence value of 94%
2. If you buy item 1 and buy item 4, it has a support value of 47% and has a 95% confidence value
3. If you buy item 1 and buy item 8, it has a support value of 54% and has a confidence value of 88.5%
4. If you buy item 1, buy item 2 and buy item 9 then it has a support value of 33% and has a 100% confidence value
5. If you buy item 1, buy item 6 and buy item 9 then it has a support value of 33% and has a 100% confidence value
6. If you buy item 1, buy item 9 and buy item 2 then it has a support value of 33% and has a 100% confidence value
7. If you buy item 1, buy item 9 and buy item 6 then it has a support value of 33% and has a 100% confidence value
8. If you buy item 2, buy item 1 and buy item 9 then it has a support value of 33% and has a 100% confidence value
9. If you buy item 2, buy item 9 and buy item 1, it has a support value of 33% and has a 100% confidence value
10. If you buy item 6, buy item 1 and buy item 9 then it has a support value of 33% and has a 100% confidence value

**3.5Application Testing**

In the Testing Phase of this application, we use the Rapidminer Studio application to get the above calculations. The following are the application testing steps:

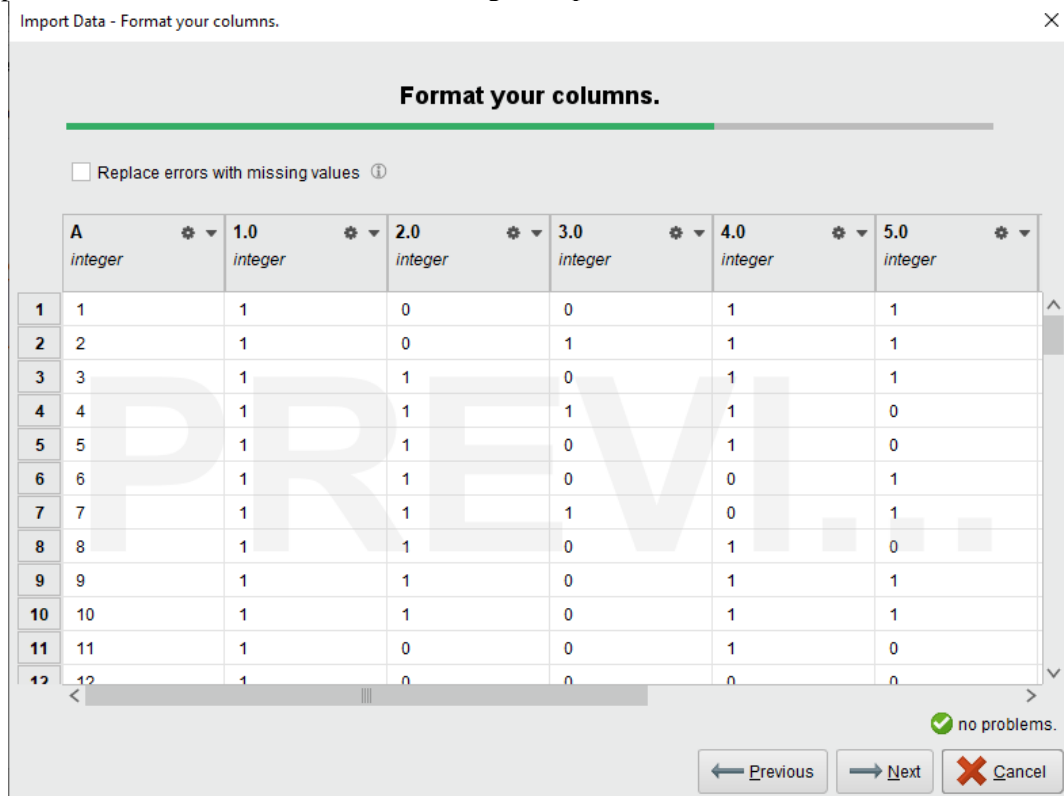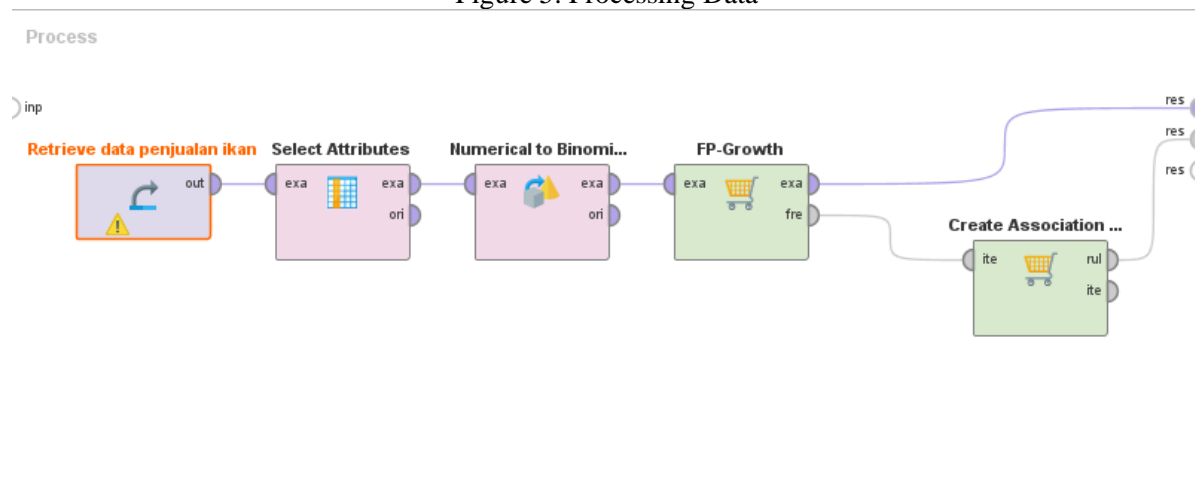1. Import data that will be used for Data Mining in Rapidminer



Figure 4. Import Data

2. Set the minimum support, Confidence and connect each operator.

Figure 5. Processing Data



3. Generated rules for the FP-Growth algoritma algorithm

**AssociationRules**

```
Association Rules
[ter mdn, rebon asin bsar, tri jngki tawar] --> [teri buntiaw] (confidence: 0.812)
[ter mdn, rebon asin bsar, tri jngki tawar] --> [teri_jgka_blah, teri buntiaw] (confidence: 0.812)
[teri_jgka_blah, ter mdn, rebon asin bsar, tri jngki tawar] --> [teri buntiaw] (confidence: 0.812)
[rebon asin bsar, udang krng] --> [teri buntiaw] (confidence: 0.846)
[rebon asin bsar, udang krng] --> [teri_jgka_blah, teri buntiaw] (confidence: 0.846)
[teri_jgka_blah, rebon asin bsar, udang krng] --> [teri buntiaw] (confidence: 0.846)
[ter mdn, rebon asin bsar, ikn sale limbat] --> [teri_jgka_blah] (confidence: 0.875)
[ter mdn] --> [teri_jgka_blah] (confidence: 0.885)
[tri jngki tawar, peda putih] --> [teri_jgka_blah] (confidence: 0.889)
[ter mdn, ikn sale limbat] --> [teri_jgka_blah] (confidence: 0.893)
[ter mdn, tri jngki tawar] --> [teri_jgka_blah] (confidence: 0.900)
[ter mdn, rebon asin bsar] --> [teri_jgka_blah] (confidence: 0.906)
[rebon asin bsar, ikn sale limbat] --> [teri_jgka_blah] (confidence: 0.909)
[ter mdn, ikn sale limbat, peda putih] --> [teri_jgka_blah] (confidence: 0.909)
[tri jngki tawar] --> [teri_jgka_blah] (confidence: 0.915)
[teri buntiaw, udang krng] --> [rebon asin bsar] (confidence: 0.917)
[teri buntiaw, udang krng] --> [teri_jgka_blah, rebon asin bsar] (confidence: 0.917)
[teri_jgka_blah, teri buntiaw, udang krng] --> [rebon asin bsar] (confidence: 0.917)
[rebon asin bsar] --> [teri_jgka_blah] (confidence: 0.918)
[rebon asin bsar, teri buntiaw] --> [teri_jgka_blah] (confidence: 0.931)
[ikn sale limbat] --> [teri_jgka_blah] (confidence: 0.932)
[teri buntiaw, peda putih] --> [teri_jgka_blah] (confidence: 0.933)
[ikn sale limbat, peda putih] --> [teri_jgka_blah] (confidence: 0.933)
[peda putih] --> [teri_jgka_blah] (confidence: 0.939)
[tri pdg bear] --> [teri jgka blah] (confidence: 0.940)
```

Figure 6. RapidMiner Results

## 4. Conclusion

From the results of analysis and application testing that has been carried out from transaction data of 1000 fish sales data, it can be concluded that: The best-selling fish is anchovy with a value of 90%, and if consumers take anchovies, they will take anchovy with a value of 54% If you take an anchovy, it will take a large anchovy and will take a white Peda fish with a value of 100%.

Reference
[1]     Anandi Ayu, Agus Perdana Windarto, Dedi Suhendro. 2021." Data Mining Implementation With the Fp-Growth Method of Goods Sales Data as a Sales Strategy     On     CV. A&A Copier". Pematangsiantar: STIKOM Tunas Bangsa. RESOLUTION : Engineering Informatics and Information Engineering Vol 2, No 2
[2]     Damanik Adela Ainun, Azhar Zulfi, Sapta Andy. " Application of Data Mining to Classify Customer Patterns Using the C4.5 Algorithm at TanjungBalai Pawnshops". Range: STMIK Royal. JUTSI (Journal of Information Technology and Systems) ISSN 2774-9029 (online) Vol. 2,No. 1.
[3]     Wulandari Rani Yuliani, 2018. "Data Mining Analysis with Classification Methods for Defective Products     at PT. Shuangying International Indonesia". Usedi: School Tall     Technologylamp of the nation,
[4]     Sari Rapita, 2018. "Aprio Algorithm Implementationri on Data Mining for Book Borrowing Patterns at the UIN Raden Fatah Library". Palembang: University     Islam     RadenFatah.
[5]     Maulana Yusuf, Winanjaya Riki, Rizki Fitri. "Application of Data Mining with the C4.5 Algorithm in Predicting Tempe Sales". Pematangsiantar : STIKOM Tunas Bangsa. BULLETIN OF COMPUTER SCIENCE RESEARCH Vol 2, No 2.
[6]     Batlayeri Yohanes Adi, 2019. "Shopping Basket Analysis Using the FP- Growth     On Happy Radiant Wholesale Store".Yogyakarta: Sanata Dharma University.
[7]     Rizky Muhamad, Umar Rusydi, and Yudhana Anton. 2020. "Application of FP-Algorithm Growthfor     Determination of Purchase Patterns for Sales Transactions in Stores KGS     Rizky Motor".Yogyakarta: Ahmad Dahlan University. Dynamic Journal Vol 25, No.1.
[8]     Setyo Wahyu Nur and Wardhana Sukma.2019. "ImplementaData Mining on Product Sales at CV Cahaya Setya Using the FP-Growth Algorithm", Jakarta:Universitas Mercu Buana.

[9]     Lestari Yuyun Dwi.2017."The Implementation of Meng Data Mininguse the FP-Tree and FP-Growth Algorithms on Drug Sales Transaction Data". Medan: High School        Technique Hope. JournalLightning.Vol.12 No.1.

[10]    Fajrin Alfannisa Annurullah and Maulana Algifanri. 2018."Application of Data Mining for Analysis of Consumer Purchase Patterns with FP-Growth Algorithm on Data        Sales TransactionMotorcycle Spare Parts". Batam: Batam Putera University. Journal Collection, Computer Science Vol 05, No.01.

[11]    Fajar Adhinda, Kusuma Wardani, and Titin Kristiana 2020. "Implementation of Product Sales Data Mining at PT.Natural Nusantara Using the Apriori Algorithm". STMIK Nusa Mandiri. Paradigm Journal of informatics and computers Vol 22, No.1.

[12]    Nurmah, Siti Nurajizah, and Arinda Salbinda. 2021. "Application of Data Mining K-Means Clustering Method for Sales Analysis at Banten Hijab Fashion Store". University of Nusa Mandiri, Journal of Computer Engineering AMIK BSI. Vol 7. No.2.

[13]    Tamba, SP, 2022. Application of Data Mining Apriori Algorithm in Determining Raw Material Stocks in Fisherman's Restaurants Using the Association Rule Method. Journal of Information Systems and Computer Science Prima (JUSIKOM PRIMA), 5(2), pp.97-102.