Utilization of the FP-Growth Algorithm on MSME Transaction Data: Recommendations for Small Gifts from The Padang Region

Firman Noor Hasan* and Riyan Ariyansah1

1,2Informatics Engineering, Faculty of Industrial Technology and Information Technology, University of Muhammadiyah Prof. DR. HAMKA
1,2Jl. Tanah Merdeka No.6, subdistrict. Pasar Rebo, East Jakarta, 13830, Indonesia

ABSTRACT

The existence of adequate transaction data turns out to have a similar sales transaction pattern for MSMEs, so it would be a shame if it were left like that. Moreover, this data can be used to increase efficiency in MSMEs in the culinary sector, one of which is as a recommendation for small gifts. The study uses the Association Rules technique, whereas fp-growth is used to obtain a combination of elements. The goal is to facilitate MSMEs’ ability to suggest small gifts to clients. The fp-growth algorithm calculation was implemented to process 2043 data originating from transaction data in MSMEs, with the specified minimum support value being 15%, while the minimum confidence value determined was 55%. The results of the trial obtained the two best rules, namely, "If a customer buys a list of small gifts from Balado Sanjai Chips, then the customer will buy Jangek Crackers" and "If a customer buys Jangek Crackers, then the customer will buy Sanjai Balado Chips".

Keywords: FP-growth; MSME; small gifts; transaction data;
1. INTRODUCTION

Recent technological advancements have significantly impacted human lives [1]. One aspect is how people utilize the information [2]. Information is vital in everything. [3], including ways to boost sales outcomes, customer buying habits and preferences, and the ability to detect present market developments [4]. Saving information in a soft copy or cloud simplifies electronic data processing [5]. Big data is the accumulation of data that is kept on servers and in the cloud, with it expanding in size and volume over time [6]. With the advancement of technology nowadays, huge amounts of data even play a major role in producing outcomes [7], due to its relationship with social networks [8].

The number of tourists visiting Padang City in 2023 will reach 3,660,947 visits [9]. This figure has increased compared to 2022. The increasing number of tourist visits cannot be separated from the many significant events held in Padang City [10]. Culinary income is the most important tax contributor from the tourism sector in Padang, amounting to IDR 70.6 billion, or reaching 100.87 percent [11]. This kind of food business is the same as a tiny restaurant, which serves a selection of snacks and beverages and markets itself as a stylish spot to unwind and have casual conversations about personal and public issues [12]. In addition, this type of culinary business only provides a variety of food dishes and beverages, from coffee and non-coffee drinks to dine-in or take-out options, which are supplied in accordance with the theme concept of the particular type of business. Several types of businesses provide takeaway and delivery services [13]. Typical Padang small gifts are a type of business quite popular in the area. Initially, Small Gifts Khas Padang was a small business with a few employees, but Khas Padang already has approximately 4 employees. Typical Padang small gifts are a type of business quite popular in the area [14]. Consumers who come to Typical Padang Small gifts have problems in determining small gifts recommendations. Specifically, no ideal presentation idea has been developed to determine modest gift recommendations for clients. As a result, it is critical to select the best small gift recommendations so that clients do not waste too much time ordering a list of accessible small gift goods [15].

At Small Gifts Padang Region, a customer approaches the cashier, who then records the items the buyer wishes to purchase [16]. The current data is just kept or used as transaction history since the recorded information won't be fully utilized; instead, it will be retained as a sales transaction note. Naturally, the abundance of data on typical Padang little presents has sales transaction patterns that are similar to one other [17]. These patterns can be used and considered by applying data mining science knowledge. Data mining can be used with the aim of analysis to explore data where there has never been a previous idea about the concept of results that are considered interesting [18]. In order to discern the notion of useful outcomes and non-trivial information in enormous datasets and make inferences about the produced data, data mining is necessary for new searches, which is accomplished by balancing human knowledge with the ability of computers to visually represent specific problems and objectives [19].

Association rules are needed in the data mining technique utilized in the context of this research design to get research findings [20]. The directional association of each item in the dataset can be ascertained using rules in data mining techniques, which may aid in describing the relationship or correlation between various objects and other items. In other words, association rules will combine the appropriateness of the attributes created for every object in the collection [21]. Fp-growth is one of many association rules algorithms that, in this case, will be used to process the dataset and produce conclusions pertaining to the adequacy of the characteristics produced for each item. This study aims to process data in sales transactions using the fp-growth method in a way that will facilitate MSMEs' recommendation-making process for modest gifts to clients [22].

Previous research conducted by Sigit related to the analysis of the fp-growth algorithm for product recommendations in retail data on sales of cosmetic products concluded that there were significant obstacles in terms of promotion on packages of cosmetic products being sold, which were caused by insufficient effectiveness in organizing products that had been sold. Interconnected products between each other [23]. Meanwhile, research conducted by Alfanisa related to the application of data mining to analyze consumer purchasing...
patterns with the FP-Growth algorithm on motorbike spare part sales transaction data concluded that analysis of consumer purchasing patterns in motorbike spare part sales transactions resulted in the dominant spare part item being sold a lot in the form of oil, seal, axle, cylinder gasket, screw valve adjusting, and cable clutch [24]. Another research conducted by Abu related to searching for association patterns for arranging goods using a comparison of the apriori and FP-Growth algorithms (a case study of the EPO Store Pemalang distro), concluded by comparing two association search methods, namely apriori and FP-Growth, which resulted in the FP-Growth Algorithm producing that the level of accuracy is greater or significant than the apriori algorithm in sales transactions [25]. As far as researchers have observed from previous studies, no one has conducted research related to sales transaction data to produce product recommendations, especially culinary products.

2. METHODS

After learning the various scientific steps in the research stage, a framework of thought is needed as a reference in determining the design of the research structure flow from start to finish to obtain significant results for the continuity of the current research. Figure 1 shows the stages carried out in this research. We will start by analyzing the problem that will be used as a topic, analyzing needs, analyzing the system that is currently running, and analyzing the system that will be proposed. Next, the researcher began collecting data through interviews, field observations, and literature studies to make comparisons with previous research that other researchers had conducted. Then, the researchers searched for patterns using the association rules and fp-growth methods. After obtaining the pattern, the researcher calculated the level of accuracy using the lift ratio. Next, the researcher designed an interface application that users would use using the state transition diagram method with the Harel model. Then, the design results will be implemented and tested. If the testing succeeds, it will continue making a research report. The general research techniques used are as follows:

2.1. STD Design

Applications in this research are designed using State Transition Diagrams using the Harel model. The harel model offers more expressive power than simpler state diagrams [26]. In order to visually represent the current processing in a system whose states are interconnected with the existing states, State Transition Diagrams were invented [27].

2.2. Implementation

At this stage, the focus is on describing the results obtained from the design implementation and significant application testing. The design uses the Perl Hypertext Processor programming language at the implementation stage. Perl was chosen because it is feature-rich, the source code is simple, and it can be shortened [28].
2.3. Testing

Getting test results will provide a detailed explanation of the data outcomes of the system, which is finished [29]. The results of data calculations are the test results that form sales using the fp-growth method or algorithm transaction patterns at Padang Typical Small Gifts MSMEs and searching for recommendations items at Padang Typical Small Gifts MSMEs. Formulas Known More Commonly as fp-growth, frequent pattern growth is a component of the apriori algorithm, which was created to locate and identify a number of dominant data in the domain of itemsets (frequent itemsets) [20]. Candidate-generating activities are unnecessary because the fp-growth algorithm favors a strategy based on the notion of tree building (fp-tree) when acquiring dominating data in the itemset domain. Consequently, this approach outperforms the apriori technique in terms of collecting dominating information in the realm of itemsets (frequent itemsets) [30].

3. RESULTS AND DISCUSSION

The stages in application design begin with designing a State Transition Diagram, then proceed to the database design stage with the final output structure in the form of an introduction to the user interface or application interface display.

3.1. STD Design

The STD design of the FP-Growth page describes the computation of a dataset used with the FP-Growth algorithm technique. Figure 1 illustrates where a calculation sheet will be entered on the FP-Growth page when choosing current transaction dates and entering the minimum support and confidence values.

The Calculation Results page's State Transition Diagram includes a tool for creating association rules based on the FP-Growth algorithm technique. As demonstrated in Figure 2 below, if each rule is created, it can serve as a solution for suggestions for little gifts of food and drink.

As seen in Figure 3 below, the Recommendation Result Page of STD provides recommendations for little presents of food and drink and explains the results.

3.2. Fp-growth

The next step is forming association rules in the fp-growth algorithm, which generates frequent itemset steps that search for items in data transactions that have met the minimum support value obtained.

3.2.1. High-frequency pattern analysis

This stage seeks to maximize the conditions that must be met for the support value with the assumption that the goal is an effort to find a combination of items, with a significant formulation of obtaining the support value of an item as follows:

\[
\text{Support} \ (A, B) = \frac{\text{transactions contain } A \text{ or } B}{\text{transactions summary}} \quad (1)
\]

Meanwhile, the significant formulation for obtaining support values from two items can be known as follows:

\[
\text{Support} \ (A \cap B) = \frac{\text{transactions contain } A \text{ and } B}{\text{transactions summary}} \quad (2)
\]

3.2.2. Formation of Association Rules

After knowing the assumptions in identifying high-frequency patterns, as explained in the previous paragraph, the next step focuses on the mechanism for searching for associative rules, with the assumption that the
minimum requirement for confidence is determined by calculating the confidence of the associative rule A$\rightarrow$B. The confidence value of the rule A$\rightarrow$B is obtained from the following formula:

$$P(B \mid A) = \frac{\text{transaction summary contain } A \text{ and } B}{\text{transaction summary contain } A} \quad (3)$$

It can be seen the results of the generated frequent itemset step in small gifts of Padang region in obtaining the minimum support value as explained in Table 1 below:

<table>
<thead>
<tr>
<th>No</th>
<th>Item</th>
<th>Support</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Keripik Balado Sanjai</td>
<td>58.15%</td>
<td>1188</td>
</tr>
<tr>
<td>2</td>
<td>Kerupuk Jangek</td>
<td>52.11%</td>
<td>1065</td>
</tr>
<tr>
<td>3</td>
<td>Roti Randang</td>
<td>20.02%</td>
<td>411</td>
</tr>
<tr>
<td>4</td>
<td>Galamai</td>
<td>14.53%</td>
<td>299</td>
</tr>
<tr>
<td>5</td>
<td>Dendeng Balado Instan</td>
<td>13.20%</td>
<td>272</td>
</tr>
<tr>
<td>6</td>
<td>Arai Pinang</td>
<td>13.05%</td>
<td>269</td>
</tr>
<tr>
<td>7</td>
<td>Bolu Batik</td>
<td>11.83%</td>
<td>245</td>
</tr>
</tbody>
</table>

This aims to visualize the acquisition of specific data transactions on each domination item, which has similarities between the items that will be applied to an FP-Tree trajectory, so that at the ongoing stage of visualizing the acquisition of data transactions, the FP-Tree trajectory allows for the formation of overlapping each other, the more high dominance items that have similarities between items will increase the effectiveness of FP-Tree trajectory data processing.

### 3.3. User Interface Implementation

On the Add Data feature page, there is a form that stores specific additional data for Padang Typical Small gifts. In this form, you are required to input the Transaction ID, goods, or items, as well as the time and date specified in the columns provided.

The data import feature web page has a choose file button to use the application to input data, which will be processed using the fp-growth algorithm. It should be noted in the data input process that the Excel file format is required to be in .csv format and have a writing format in the form of data_id, transaction_id, item, and date fields. If the Excel file is not in .csv format, the system will indirectly reject the file that will import the data into the database, so it will not appear on the application page.

### 3.4. Accuracy Test

The primary objective of the test outcomes is to ascertain the precision of the outcomes for every data transaction that arises from association rules. The lift ratio is used as a reference assumption to assess the validity of the transaction patterns observed in this investigation. The lift ratio has a decision-making premise attached to it. In particular, when the lift ratio is smaller than one (<1), it can be presumed that the item's connection with item B is negatively correlated, meaning there is no meaningful linkage between the item and
other items. Conversely, the link between item A and item B is probably considered to be favorably associated. If the lift ratio result is larger than or equal to one (≥1). The assumption used to make decisions is that items A and B are categorized as independent if the final value is equal to one (=1).

The system testing technique uses one of the parameters in the Association Rule, namely the Lift Ratio, which is used in the current research process. The formula that can be used to calculate the Lift Ratio is as follows:

\[
Lift(A, B) = \frac{P(A \cup B)}{P(A)P(B)}
\]

To obtain confidence values from itemset product A with product B, the number of transactions containing product A is required, and the number of transactions containing product B is also necessary. The Lift Ratio itself is a mapping measuring tool of how important or dominant the rules that have been created or produced are based on the value of the support and confidence results. In this context, the Lift Ratio is part of the acquisition of value which can indicate the truth or authenticity of the ongoing process of a transaction. It contributes to providing significant information or information from purchasing product A with product B simultaneously [3].

The assumption that can be classified as genuine or positively associated because the lift ratio value is greater than 1, it can be categorized as invalid or adversely associated. Meanwhile, the next rule is "If a customer buys small gifts of Balado Sanjai Chips, then the customer will buy Jangek Crackers." A rule can be classified as genuine or positively correlated if it has a value of 1.15 for the lift ratio, 34.61% for support, 59.51% for confidence. Next, the last rule is "If a customer buys Jangek Crackers as a small gifts, the customer will buy Balado Sanjai Chips". This rule can be classified as genuine or positively associated because the lift ratio value is greater than 1, with a support value of 34.61%, a confidence value of 66.38%, and a lift ratio value of 1.15. From the results of the rules that have been created, as can be seen, there are just two guidelines that are considered to be reliable, and that can be followed when suggesting little gifts to clients.

3.5. Functionality Test

This testing aims to find system errors, especially for Small Gifts of the Padang Region recommendation items. This testing is carried out as an analysis of the objectives of the vision and mission of the system according to the criteria that must be met in the application design objectives. The test will be carried out directly by the owner who uses food and drinks in the recommendation feature. Overall test results are presented in Table 3 below.

<table>
<thead>
<tr>
<th>No</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If customers buy Roti Randang and Galamai small gifts, then customers will buy Balado Sanjai Chips</td>
<td>220/2043 = 10.77%</td>
<td>53.58%</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>If a customer buys small gifts of Balado Sanjai Chips, then the customer will buy Jangek Crackers</td>
<td>707/2043 = 34.61%</td>
<td>59.51%</td>
<td>1.15</td>
</tr>
<tr>
<td>3</td>
<td>If a customer buys Jangek Crackers as a small gifts, the customer will buy Balado Sanjai Chips</td>
<td>707/2043 = 34.61%</td>
<td>66.38%</td>
<td>1.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No</th>
<th>Scenario</th>
<th>Expected Result</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Login</td>
<td>Admin can enter the application</td>
<td>Valid</td>
</tr>
<tr>
<td>2</td>
<td>Dashboard</td>
<td>Admin can see information about food and drink menus</td>
<td>Valid</td>
</tr>
<tr>
<td>3</td>
<td>Viewing Transaction Data</td>
<td>Admin can view existing transaction data</td>
<td>Valid</td>
</tr>
<tr>
<td>4</td>
<td>Changing Transaction Data</td>
<td>Admin can change transaction data if necessary</td>
<td>Valid</td>
</tr>
<tr>
<td>5</td>
<td>Importing Transaction Data</td>
<td>Admin can import transaction data that has been entered into Microsoft Excel</td>
<td>Valid</td>
</tr>
<tr>
<td>6</td>
<td>Deleting Transaction Data</td>
<td>Admin can delete transaction data if necessary</td>
<td>Valid</td>
</tr>
</tbody>
</table>
Table 3 continued...

<table>
<thead>
<tr>
<th>No</th>
<th>Scenario</th>
<th>Expected Result</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Transaction Data Calculation Using the FP-Growth Method</td>
<td>Admin can determine the start date, end date, support, and confidence, which will determine food and drink menu recommendations</td>
<td>Valid</td>
</tr>
<tr>
<td>8</td>
<td>FP-Growth Method Calculation Results</td>
<td>Admins can view data and also metadata results from calculations that have been carried out on the FP-Growth page</td>
<td>Valid</td>
</tr>
<tr>
<td>9</td>
<td>Results of Food and Drink Menu Recommendations</td>
<td>Admins can see what menus are purchased the most and also see menu recommendations suggested by the system</td>
<td>Valid</td>
</tr>
<tr>
<td>10</td>
<td>Change Password</td>
<td>Admin can change the admin account password if necessary</td>
<td>Valid</td>
</tr>
</tbody>
</table>

3.6. Effectiveness Test

Effectiveness testing is carried out to see how useful or practical the system that has been created is, and the testing is carried out in a way that is directly related to the system and uses the system directly. The effectiveness of the recommendation items was tested using the fp-growth method. The effectiveness test was carried out face-to-face. Examiners make conclusions using a Likert scale with selected aspects. In general, there are 3 most important test questions related to the effectiveness test for this system, namely: is this system what the user expects? Can this system help users determine recommendations for typical Padang souvenirs? Does the system interface appear? Can make it easier for users to operate this system.

CONCLUSION

The fp-growth algorithm approach can be implemented to get a considerable output in understanding the transaction patterns that arise from buying little gifts or food and drink products. This implementation can also precisely determine recommendations for small gifts or items from food and drinks based on sustainable patterns of purchasing transactions by customers in MSMEs in the culinary sector. The results obtained from the customer's sustainable purchasing transaction patterns or expressed as rules for typical Padang small gifts MSMEs, apparently produce two transaction patterns which are stated as unique rules, namely: "If a customer buys Balado Sanjai Chips, then the customer will buy Jangek Crackers ", then “If a customer buys Jangek Crackers as a small gifts, the customer will buy Balado Sanjai Chips". Only two rules can be deemed legitimate and utilized as a guide when making suggestions for food and drink based on the outcomes of the rules that were created for small gifts at Padang Typical Small gifts MSMEs. So data that can support customer decisions are obtained from two patterns, which are based on the assumption that the greater the support value used, the impact it will have on the confidence value and also the lift ratio value for the rule formed, the better the rule will be. Suggestions for further research include using other methods such as apriori or market basket analysis.

The display on the small gifts recommendation results web page shows recommendations for small gifts for food and drink and descriptions of recommended recommendations from Padang Typical Small gifts in graphical form, which is formed using calculations from the fp-growth algorithm. This will make it easier for Padang Typical Small Gifts MSME managers to determine the list of small gifts recommended to customers. So the results of this research can be useful for applying to MSMEs, especially in terms of small gift recommendations.

Based on the results of effectiveness tests that have been carried out, this system has an effectiveness level of 87% in recommending typical Padang small gift products. So it can be concluded that this system helps users in determining recommendations for typical Padang small gifts products.

REFERENCES


[20] A. Sani, Samuel, D. Suryadi, F. N. Hasan, A. D. Wiranata, and S. Aisyah, “Predicting the Success of Garment...


