Polyclinic Visitor Pattern Discovery Using Apriori Algorithm

I Gede Teguh Satya Dharma[a1], I Ketut Gede Suhartana[a2]

[a]Informatics Department, Faculty of Math and Science, Udayana University
Bali, Indonesia
[1]teguhsatyadhr@gmail.com
[2] ikg.suhartana@unud.ac.id

Abstract

Living in an information era, the presence of data is super important. With the data exponentially grows from decades to decades, it reflect the old saying that people today are so rich with data yet so poor on information. To dissect the information that contained within the large amount of data, a method is introduced called "Data Mining." Data mining is a process of retriving a unique, unseen, and valuable information/insight from the data. Data mining comes with a lot of methods branches, one of those is pattern analysis, or also known as "Association Rules". With the help of Association Rules, people can discover the relational pattern within the data, so people can make the best decision on the period of time. In this study, the writer is implementing Apriori Algorithm (one of the Association Rules algorithm) to see the pattern of a polyclinic visitor.

*Keywords: Data, Data Mining, Association Rules, Apriori Algorithm, Polyclinic Visitor*

1. Introduction

Apart from education and infrastructure, health service is one of the most prominent aspect in human life. With the population increasing, the number of health service is also need to be increased in order to facilititate the people's needs. The presence of data also holds a significant roles in health service, with the medical data being recorded from each patients, this helps the development of the health service such as a useful material of analysis, research and evaluation of the quality of services provided to the patient and also crucial data for research and educational purposes or even to spot trends and anomalies on the patient record.

Data mining is the perfect solution to analyze this gigantic amount of data[1]. Data mining makes it easier for health services manager to take decision upon several cases happening in the company itself, such as provision of rooms, calculating the number of labors they should provide and more. There are several method of data mining, such as classification, clustering, and association. In this study, the writer use associaton approach to discover the pattern of a polyclinic visitation. The algorithm that match the association method is apriori algorithm. The data that used in this study were contained from a polyclinic located in Tabanan Regency, Bali.

2. Reseach Methods

This study implement one of the techniques of data mining, namely association rule. The algorithm that used in this study is Apriori Algorithm. Apriori Algorithm is the algorithm used to find association among the items which come together in a transaction. It takes transaction database as inputs and gives frequent itemset which occur together as an output. It takes the help of minimum support and minimum confidence to find the strong association rules[2]. This research method is divided to some sections to make it easier for the reader. The section can be seen by the figure below :
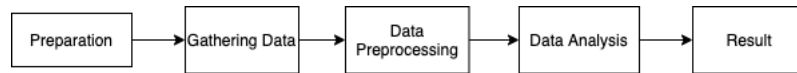
Figure 1. Research Section

The research method are consist of several section, namely the preparation sector, data-gathering section, data preprocessing section, data analysis section, and result section. The following is the explanation :

The first section is the preparation section. Just like the name, this section focused on what to prepare in order to execute the research. At this section also, the formulation of the problem is being constructed, so the study will have a clear purpose on what to solve.

The second is section is data-gathering section. At this section, data is being gathered from sources like database, secondary-data, data warehouse or anything basically that hold large amount of data that used in this particular research. For this study, the writer uses secondary data which come from a polyclinic.

The third ones, is the data preprocessing. Data quality is important in data mining. Low-quality data will lead to low-quality mining results. There are many factor comprising data quality, including accuracy, completeness,concistency, timeliness,believability, and interpretability[3]. Data preprocessing is used to maintain the data quality, it eleminates inconcistency, and incompleteness from the data source.

The fourth, one of the most important section is the analysis. At this section the data mining technique is being implemented. Apriori algorithm is going to process, train and evaluate the data that has been pre-processed before and will generate a result, which is the association rules between items.

The final section, the result. The association rule between items will be displayed at the final section, complete with the value of confidence, lift and support that comes with it.

3. Result and Discussion
3.1. Dataset

The dataset is acquired from a polyclinic located in Tabanan Regency, Bali. Total data in the dataset is 50 data(s). The items/properties that contained within the dataset can be seen in the tables below:

Table 1. Dataset

Gender	Age (Year)	Occupation	Diagnose
L	62	Swasta	Low Back Pain, Asthma
…	…	…	…
P	57	Petani	Senile Cataract

3.2. Pattern Discovery Using Apriori Algorithm

After the data is gathered, the next step is to analyze the data, discovering the unique, novel pattern that were hidden in the dataset. This process include reading the data file, executing apriori algorithm and displaying result. In this study, the writer uses python(3.8.3 version) programming languange, and a python library named "apyori". For the IDE, the writer uses Pycharm. In order to implement this to a device, python needs to be downloaded and installed first (with the IDE) then the library need to be imported.

3.3. Apriori Algorithm Flowchart

To understand how the apriori algorithm occur, it can be seen through the flowchart diagram written below :
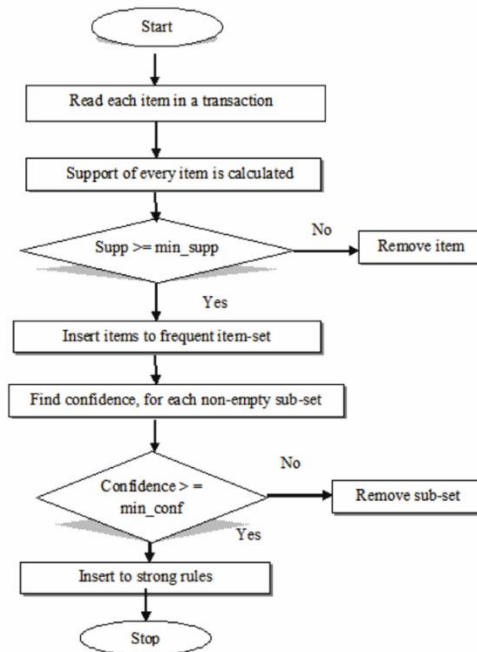

Figure 2. Apriori Algorithm Flowchart Diagram

3.4. Calculating Support Value, Confidence Value, and Lift Value

Item support value calculation in Apriori Algorithm can be seen on equation below[4]:

$$Support(A) = \frac{Amount\ of\ transaction\ that\ contains\ A}{Total\ Transaction} \qquad (1)$$

$$Support(A\ U\ B) = \frac{Amount\ of\ transaction\ that\ contains\ A\ and\ B}{Total\ Transaction} \qquad (2)$$

$$Confidence = P(B|A) = \frac{Amout\ transaction\ that\ contains\ A\ and\ B}{Amount\ transaction\ that\ contains\ A} \qquad (3)$$

$$Lift\ Ratio = \frac{Confidence(A,B)}{Benchmark\ Confidence\ (A,B)} \qquad (4)$$

With the apyori library in python, this repeated calculation can be done simply just by typing a simple syntax. The result can be displayed on the picture below :

```
association_rules = apyori.apriori(dataRecords, min_support=0.015, min_confidence=0.5, min_lift=3, min_length=2)
association_results = list(association_rules)
```

Figure 3. Apyori Library Syntax

3.5. The Result

After executing the calculation, the system will print the result. The result consist of the association rules that has been created, complete with the value of support and confidence each rules.

```
Run:   main
/Users/dedesmacbook/.conda/envs/paperCondaProject/bin/python /Users/dedesmacbook/PycharmProjects/paperCondaProject/main.py
    0  1        2                                               3
0  L  62       SWASTA                         Low back pain, asma
1  P  51       SWASTA              Abdominal pregnancy, Hipertensi
2  L  73       PETANI                  Hipertensi primer/essensial
3  P  45       SWASTA                          Abdominal pregnancy
4  L  12  BLM SEKOLAH  Superficial injury of unspecified body region
Aturan Asosiasi: PELAJAR => 10
Nilai Support: 0.02
Nilai Confidence: 1.0
Nilai Lift: 50.0
===================================
Aturan Asosiasi: Superficial injury of unspecified body region => 10
Nilai Support: 0.02
```

Figure 4. Result

```
Run:   main
Aturan Asosiasi: Superficial injury of unspecified body region => 10
Nilai Support: 0.02
Nilai Confidence: 1.0
Nilai Lift: 8.333333333333334
===================================
Aturan Asosiasi: 12 => BLM SEKOLAH
Nilai Support: 0.02
Nilai Confidence: 1.0
Nilai Lift: 50.0
===================================
Aturan Asosiasi: 12 => Superficial injury of unspecified body region
Nilai Support: 0.02
Nilai Confidence: 1.0
Nilai Lift: 8.333333333333334
```

Figure 5. Result (Continued)

4.   Conclusion

Data mining is a process of retrieving a unique, unseen, and valuable information/insight from large amount of data. Data mining is applicable on almost every aspect today's living, which one of them is health service. There are many technique available in data mining, including association analysis. On this study of polyclinic study, a pattern has been discovered such as 0.5% visitors aged 38 had other peripheral vertigo, abdominal pregnancy diagonosed, 1% of the visitors who aged 42 had chronic renal failure, and many more. With this analysis result obtained, can be a provision in making decisions to develop the health care services such as adding polyclinic rooms, increasing the workforce, and even increasing/reducing the polyclinic operating hours.

References

[1]   Pujari, A. K. (2001). *Data Mining Technique.* India: University Press.

[2]   More, N. (2014). Recommendation of Books Using Improved Apriori Algorithm.
        International Journal for Innovative Research in Science & Technology, 80.

[3]   Jiawei Han, Micheline Kamber, Jian Pei. (2012). *Data Mining Concept and Techniques.*
        USA: Morgan Kaufmann.

[4]   Mohamad Fauzy, Kemas Rahmat Saleh W, Ibnu Asror. (2015). Penerapan Metode
        Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah
        Kota Bandung. *e-Proceeding of Engineering*.

This page is intentionally left blank