# Discovering Drugs Combination Pattern Using FP-Growth Algorithm

Rini Anggrainingsih
Informatics Dept
Mathematics and Natural Science, UNS
Surakarta, Indonesia
rini.anggrainingsih@staff.uns.ac.id

Nach Rowi Khoirudin
Informatics Dept
Mathematics and Natural Science, UNS
Surakarta, Indonesia
Nach.rowi@student.uns.ac.id

Haryono Setiadi
Informatics Dept
Mathematics and Natural Science, UNS
Surakarta, Indonesia
haryonosetiadi@staff.uns.ac.id

***Abstract*— A drug can be used to deal more than one diseases and to deal an illness often need a combination of more than one drugs. This paper present how to discover a pattern of a combination of medicines related to a diagnosis of diseases using FP-Growth one of frequent pattern mining algorithm. We use FP-Growth because it has better performance than Apriori and Eclat. Data is collected from outpatients pharmacy of Sukoharjo state hospital, Central Java, Indonesia during January 2015 to June 2016 and obtain 526,195 records of prescription data and use a diagnosis of diseases base on the ICD-10 standard. This studies just apply on the top ten of the most frequently occurred illness in the outpatient's services of Sukoharjo state hospital. Then the pattern of association between diseases and combination of drugs was reviewed by pharmacist committed to being validated. These studies result in some combination of medicines for to top ten of the most frequent diseases. We also found 21 similar combinations of drugs for various diseases. In the future, this finding can be used to provide suggestions to physicians to select an appropriate mix of the drug to deal some diseases.**

***Keywords*— Combination of drug, Diseases, FP-Growth, Pattern recognising***

## I. INTRODUCTION

A drug can be used to deal more than one illness on the one hand and on the other hand to deal a disease we oft need a combination of more than one drugs. The question is how the association pattern between combinations of drug and diseases. The model of a combination can be discovered using frequent pattern mining algorithms such as *Apriori, Eclat,* and *FP-Growth[1]*.

The apriori algorithm finds itemset using *join and prune[2]*. It has a problem with candidate generation and scanning database process[3]. To deal this issue, researchers proposed a frequent pattern growth (FP-Growth) algorithm which uses FP-tree and also divide-and-conquer to reduce database[4], so it has a better performance result than Apriori algorithm[5]. On the other hand, Eclat[6] algorithm discovers itemset by applying intersection between ID transaction so it is only need once database scanning, but it consumes lots of memory when using on large transaction data. Furthermore, Eclat also needs a long time to compute intersection process[3]. So, FP-Growth has the best performance of others algorithm[7].

Previous studies have employed FP-Growth to finding a pattern of drugs which purchased by the customer in the drugstore[8] and the other study has used FP-Growth to find an association for identifying customers buying habits[9].

This paper present how to discover a pattern of a combination of drugs related to a diagnosis of diseases using FP-Growth. The data was collected from outpatients of Sukoharjo hospital pharmacy, a state hospital in Central Java, Indonesia during January 2015 to June 2016 and obtain 526,195 records of prescription data. This study used the diagnosis of diseases which included on the *International Classification of Diseases-10$^{th}$* (ICD-10)[10] standard. This study focus on the top ten frequent occurred conditions in the outpatient's services of Sukoharjo state hospital. Then the pattern of association between diseases and combination of drugs was reviewed by a committee of pharmacy to be validated.

This paper is organised as follows. In section 2, the frequent pattern mining is described. In section 3, the FP-Growth method is presented. The experimental results are shown in section 4. Finally, our work of this paper is summarised in the last section.

## II. FREQUENT PATTERN MINING

### A. Frequent Pattern Mining

A frequent pattern is a pattern (itemset, subsequence, or substructure) which frequently appears in the dataset[4]. Frequent pattern mining is the most important stage in the association rules technique.

The primary job of frequent pattern mining is how to find all itemset in the dataset that fulfils a minimum support threshold. Support from the itemset (I) is defined as part of the transaction in the database $T = \{T_1 \ldots T_n\}$ which contains I as a subset, notated as *sup(I)*[11]. When the support of *I* is equal or more than the minimum support, then *I* is called a frequent pattern.

978-1-5386-0549-3/17/$31.00 ©2017 IEEE

$$sup(I) = \frac{Total\ transaction\ containing\ subset\ I}{Total\ transaction} \quad (1)$$

The total number of common pattern depend on the level of minimum support. The less of minimum support lead to a total of familiar pattern higher. On the other hand, the higher minimum support result less frequent pattern.

### B. FP-Growth Algorithm to Find Pattern

The mining process using FP-Growth does not need candidate generation[3]. FP-Growth adopts the divide-and-conquer strategy. FP-Growth carries out a development of Frequent Pattern Tree (FP-Tree) to produce a common pattern and employs twice database scanning only. Firstly to find common item and secondly to develop FP-Tree. The steps to find patterns are as follow:

1) Scanning the transaction data, then accumulating frequency of each item. Then an item which does not meet the minimum support threshold should be eliminated.
2) Sorting items on each transaction from the highest frequency.
3) Building FP-Tree started with the root and reading all item in each transaction. When a transaction has the same prefixes with the previous transaction, then a line can be added to the same node, and it will add the number of support count on its nod. After that, if there does not have similar prefixes, there will be made a new line which has a value one for the support count on each node.
4) The next process finding pattern using the FP-Growth algorithm based on the FP-Tree which has developed in the previous phase as shown in Figure 2. There are three steps, *conditional pattern base generating, conditional FP-Tree generating*, and the *frequent itemset forming.*

```
Input   : a FP-tree, D – transaction database, s – minimum
          support threshold.
Output  : The complete set of frequent patterns.
Method :
call FP-growth(FP-tree, null).
Procedure FP-growth (Tree, A)
{
  if Tree contains a single path P
    then for each combination (denoted as B) of the nodes in
    the path P do
    generate pattern B ∪ A with support = minimum support of
    nodes in B
  else for each ai in the header of the Tree do
  {
    generate pattern B = ai ∪ A with support = ai.support;
    construct B's conditional pattern base and B's conditional
    FP-tree
    TreeB;
    if TreeB ≠ Ø
    then call FP-growth (TreeB, B)
  }
}
```

Fig.1. FP-Growth Algorithm[5]

## III. METHODOLOGY

This study collected data from outpatients pharmacy of Sukoharjo state hospital, Central Java, Indonesia during January 2015 to June 2016 and obtained 526,195 records of prescription data and used a diagnosis of diseases term base on the *International Classification of Diseases-10th* (ICD-10) standard. Then we conduct several steps are pre-processing data, applying the FP-growth algorithm to discover the pattern and evaluating the validity of discovered pattern.

Firstly, we do pre-process which consist of data selection, data cleaning, dan data transformation. We make data selection to select the attributes of the required data. The necessary attribute data are; the transaction date, the number of services, code of primary diagnosis, and the name of the drugs. Then, cleaning the data by removing transaction which containing incomplete data, i.e., invalid data or redundant data. And modifying the format of transaction date and unify the data of drugs based on the service number, then save these into the database. We obtain complete 90.477 data as result of preprocessing data. The sample of dataset shown in Table 1.

TABEL 1. SAMPLE OF *DATASET* AFTER *PREPROCESSING*

| Date of Transaction | Number of Services | Code of Diagnose | Drugs |
|---|---|---|---|
| 2015-01-05 | 00027004601003 | E149 | ALLOPURINOL 100 MG; MELOXICAM 7,5 MG; AMLODIPIN 10 MG BESILAT |

This study just applies on the top ten of the most frequently occurred diseases in the outpatient's services of Sukoharjo state hospital. These diseases and the code of diagnoses are *essential hypertension* (I10), *unspecified diabetes mellitus without complications* (E149), *asthma* (J459), *cataract* (H269), H524 (*presbyopia*), *acute upper respiratory infection* (J069), T*b lung confirm sputum microscopy with or without culture* (A150), *dyspepsia* (K30), *acute bronchitis* (J209), and *chronic obstructive pulmonary disease* (J449). Table 2 describes the top ten most occurred conditions

TABEL 2. THE TOP TEN OCCURRED CONDITIONS AT SUKOHARJO HOSPITAL

| No | Code of Diagnose | Description | Number of Occurrences |
|---|---|---|---|
| 1 | I10 | *Essential (primary) hypertension* | 1496 |
| 2 | E149 | *Unspecified diabetes mellitus without complications* | 950 |
| 3 | J459 | *Asthma* | 898 |
| 4 | H269 | *Cataract* | 837 |
| 5 | H524 | *Presbyopia* | 783 |
| 6 | J069 | *Acute upper respiratory infection* | 763 |
| 7 | A150 | *Tb lung confirm sputum microscopy with or without culture* | 680 |
| 8 | K30 | *Dyspepsia* | 586 |
| 9 | J209 | *Acute bronchitis* | 532 |
| 10 | J449 | *Chronic obstructive pulmonary disease* | 518 |

We developed an application to employ frequent pattern mining of combination using FP-Growth. We can find the number of combinations by choosing The code of diagnosis and adjust the value of minimum support. The interface of Application shown in Figure 1.

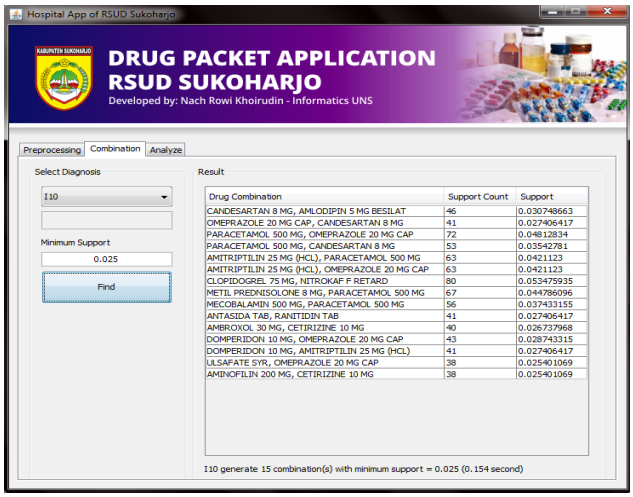


Fig.1. Interface of Application

## IV. Result

A sample result of drug combination pattern for code diagnose I10 and E149 with minimum support 0.025 are shown in Table 3 and Table 4. Table 3 illustrates that E149 has 16 combinations of drug patterns and Table 4 indicates that I10 has 25 combinations of drug patterns.

Table 3. Drug Combination Pattern for I10 Diagnosis.

| No | Combination Pattern | Support (%) |
|---|---|---|
| 1 | CLOPIDOGREL 75 MG, NITROKAF F RETARD | 0.05 |
| 2 | PARACETAMOL 500 MG, OMEPRAZOLE 20 MG CAP | 0.05 |
| 3 | METIL PREDNISOLONE 8 MG, PARACETAMOL 500 MG | 0.04 |
| 4 | AMITRIPTILIN 25 MG (HCL), PARACETAMOL 500 MG | 0.04 |
| 5 | AMITRIPTILIN 25 MG (HCL), OMEPRAZOLE 20 MG CAP | 0.04 |
| ⋮ | ⋮ | ⋮ |
| 16 | OMEPRAZOLE 20 MG CAP, AMITRIPTILIN 25 MG (HCL), PARACETAMOL 500 MG | 0.025 |

Table 4. Drug Combination Pattern for E149 Diagnosis.

| No | Combination Pattern | Support (%) |
|---|---|---|
| 1 | ACARBOSE 50 MG, METFORMIN 500 | 0.1 |
| 2 | GLIMEPIRIDE 2 MG, ACARBOSE 50 MG | 0.07 |
| 3 | GLIMEPIRIDE 2 MG, METFORMIN 500 | 0.07 |
| 4 | NITROKAF F RETARD, CLOPIDOGREL 75 MG | 0.05 |
| 5 | GLIBENCLAMIDE 5 MG, METFORMIN 500 | 0.05 |
| ⋮ | ⋮ | ⋮ |
| 25 | VIT B1 TAB 100MG, VITAMIN B6 TAB 10MG | 0.025 |

We used minimum support 0.025 and obtained the number of drugs combination patterns for the top ten frequent diseases as shown in Table 5.

Table 5. Some Drug Combination Pattern.

| Code of Diagnose | Number of Patterns |
|---|---|
| I10 | 16 |
| E149 | 25 |
| J459 | 1,671 |
| H269 | 18 |
| H524 | 15 |
| J069 | 156 |
| A150 | 387 |
| K30 | 50 |
| J209 | 355 |
| J449 | 3,566 |

According to Table 5, we get information that the most number of drug combination is J449 (*chronic obstructive pulmonary disease*) with 3,566 patterns followed by J459 (*asthma*) which has 1,671 patterns. Then the least combination of drugs is H524 (*presbyopia*) which has 15 combinations.

Then validity of these patterns were reviewed by a committee of pharmacy of Sukoharjo state hospital. There are so many various combinations, to make it simple, we only select ten combinations from the highest level of minimum support among other to be checked by a committee of pharmacy whether valid or not. The result of validity testing are shown in Table 6.

Table3. Validity Test of Drug Combination Pattern.

| Code of Diagnoses | Valid Pattern | % |
|---|---|---|
| I10 | 6 | 60 |
| E149 | 7 | 70 |
| J459 | 10 | 100 |
| H269 | 8 | 80 |
| H524 | 8 | 80 |
| J069 | 10 | 100 |
| A150 | 10 | 100 |
| K30 | 10 | 100 |
| J209 | 10 | 100 |
| J449 | 10 | 100 |
| AVG. | 8,9 ≈ 9 | 89 |

Based on Table 6 we get information that the average validity of pattern is 89%. The other interesting thing is we also found 21 similar combinations of the drug to deal several diagnoses of diseases as shown in Table 7.

TABLE 7. SIMILAR COMBINATION OF DRUG TO DEAL MULTI DIAGNOSES OF DISEASES

| No | Combination | Code of Diagnoses |
|---|---|---|
| 1 | CETIRIZINE 10 MG, SALBUTAMOL 4 MG | J459, J069, A150, J209, J449 |
| 2 | AMBROXOL 30 MG, CETIRIZINE 10 MG | J459, J209, J449 |
| 3 | CENDO CATARLENT 5 ML, VITAMIN C TAB 50 MG | H269, H524 |
| ⋮ | ⋮ | ⋮ |
| 21 | CETIRIZINE 10 MG, GLISERIL GUAIAKOLAT | J459, J449 |

## V. CONCLUSION

FP-Growth algorithm can be used to find the pattern of drugs based on the disease diagnosed. The result of this study is patterns of drug combinations for ten diagnose of diseases. These patterns can be used to suggest physician make a medical prescription. Also, this study also found 21 similar patterns of drug combination to deal several diseases.

## *References*


[1] J. Han, H. Cheng, D. Xin, X. Yan. Frequent pattern mining: Current status and future directions, Data Min Knowl Disc 15 (2007), no. 1, 55-86.

[2] R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large databases, VLDB Conference (1994), 487–499.

[3] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Elsevier, United States of America (2006).

[4] J. Han, J. Pei, Y. Yin and R. Mao, Mining frequent patterns without candidate generation: A frequent-pattern tree approach, Data Min Knowl Disc 8 (2004), no. 1, 53-87.

[5] K. Vanitha and R. Santhi, Evaluating the performance of association rule mining algorithms, Journal of Global Research in Computer Science 2 (2011), no. 6, 101-103.

[6] M. J. Zaki. Scalable algorithms for association mining, IEEE Transactions on Knowledge and Data Engineering 12 (2000), no. 3, 372-390.

[7] K. Garg and D. Kumar. Comparing the performance of frequent pattern mining algorithms, IJCA 69 (2013), no. 25, 29-32.

[8] Noma, N.G. and Ghani, M.K.A. Discovering Pattern in Medical Audiology Data with FP-Growth Algorithm, 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences (2012).

[9] A. Mehay, K. Singh and N. Sharma, Analyze market basket data using fp-growth and apriori algorithm, International Journal on Recent and Innovation Trends in Computing and Communication 1 (2013), no. 9, 693 – 696.

[10] WHO, International statistical classification of diseases and related health problems, vol. 2, WHO (2011).

[11] C. C. Aggarwal, Data Mining: The textbook, Springer, Swiss, 2015.