Identification of Adverse Event Patterns in Loperamide Using FP-Growth Algorithm

Alwis Nazir¹, Herlina¹, Yusra¹, Siska Kurnia Gusti¹, Roza Linda²

¹ Informatics Engineering Department, Faculty of Science and Technology
Universitas Islam Negeri Sultan Syarif Kasim Riau
² Chemistry Education Department, Faculty of Teacher Training and Education,
Universitas Riau
alwis.nazir@uin-susk.ac.id

Abstract. Loperamide is a drug to treat diarrhea, and should not be used for self-medication. There are cases of cardiac disorders due to the use of this drug, this is called an adverse event. Based on this problem, this research will identify the adverse event pattern in Loperamide using the FP-Growth algorithm. The data used is FAERS data from 2015-2017 with the attributes is the age group, sex, drug name, and adverse event. The amount of data after the preprocessing process is 2,840 records. The results is we found the pattern with the highest support values is a combination of adult age groups, male sex, the drug name Loperamide and adverse event cardiac disorders that have a support value of 3.8%, confidence 28.35% and lift ratio 1.483. The implemented using Matlab software and tested using SPMF tools.

1. Introduction
One drug commonly used for the treatment of diarrhea is Loperamide. Loperamide is a drug that is commonly consumed by most households that are sold freely in the market and to get it can be without using a doctor's prescription [1]. Loperamide is used to treat diarrhea symptoms. In addition, Loperamide is a hard drug that should not be a drug that can be recommended to clients who self-medication [2]. The Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) is a database containing reports of adverse events, medication errors, and product quality complaints that are submitted to the FDA in America. The database is designed to support the FDA’s post-marketing safety supervision program for drugs and therapeutic biology products. Adverse events and medication errors are coded using terms in the Medical Dictionary for Regulatory Activities (MedDRA) term [3].

Adverse event means any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related. [4]

In the 39 years from when loperamide was first approved in 1976 through 2015, FDA received reports of 48 cases of serious heart problems associated with use of loperamide. This number includes only reports submitted to FDA, so there are likely additional cases about which we are unaware. Thirty-one of these cases resulted in hospitalizations, and 10 patients died. More than half of the 48 cases were reported after 2010 [5].
Based on this, the research objective is the identification of adverse events to find out what adverse events can occur in the use of this drug so that the public knows the impact of using this drug if consume it freely.

The identification process that will be carried out will apply one method to data mining, namely the association method with the Frequent Pattern Growth (FP-Growth) algorithm. Data Mining is one of the stages of Knowledge Discovery in Database (KDD). KDD is the process of finding knowledge stored in a large database, data warehouse, web, or other large information repository. KDD is often called the same as data mining. Frequent Growth Pattern (FP-Growth) is one of the algorithms in the data mining association for finding frequent itemsets. FP-Growth is built by creating FP-Tree to extract transactions in the database [6]. There are 4 attributes that will be used in this research, namely: drug name, sex, age group, and adverse events.

2. Research Metodology

Workflows of this research can be seen in the following picture:

![Workflows](image)

**Figure 1. Workflows**

2.1 Data Collection

The data used in this study is data from the FAERS (Food and Drug Administration Adverse Event Reporting System) from 2015-2017. Data is taken from the site [www.fda.gov](http://www.fda.gov). The amount of data after the preprocessing stage is 2,840 records.

2.2 Analysis

2.2.1 Data Requirement Analysis. FAERS data is divided into 7 files. Each file stores different data relating cases of adverse events. This research just will use 3 FAERS files, namely Demo, Drug and Reac files. Demo file contains patient demographic and administrative information. Drug file contains drug/biologic information for as many medications were reported for the event. Reac File contains all "Medical Dictionary for Regulatory Activities" (MedDRA) terms coded for the adverse event (1 or more). The research attributes needed for this can be seen in Table 1.
Table 1. Attributes

| No | Attribute    | Explanation                                                                 |
|----|--------------|-----------------------------------------------------------------------------|
| 1  | Primaryid    | Unique number to identify FAERS report                                      |
| 2  | I_F_Cod     | Code for the status report that occurs. Initial (Initial) or follow up       |
|    |              | (Follow Up)                                                                 |
| 3  | Age          | The numerical value of the patient's age                                    |
| 4  | Age_cod      | Unit stands for age of the patient                                          |
| 5  | Sex          | Sex                                                                         |
| 6  | Occr_country | The country where the event occurred                                        |
| 7  | Drug_seq     | Unique number to identify the order of drugs used                           |
| 8  | Role_cod     | Codes for reported drug roles                                               |
|    |              | - The first suspect (Primary Suspect)                                       |
|    |              | - Second Suspect (Secondary Suspect)                                        |
|    |              | - Concomitant                                                               |
|    |              | - Interacting                                                               |
| 9  | Drugname     | Name of medicinal product                                                   |
| 10 | Prod_ai      | Drug Content                                                                |
| 11 | Pt           | Adverse Event                                                               |

2.2.2 Knowledge Discovery in Database Analysis. Selection, this stage will selection of several attribute categories was conducted so that the data used in the research was more focused. The attribute categories to be selected can be seen in Table 2.

Table 2. Attribute Requirement

| No | Attribute Requirement | Kategori                             |
|----|------------------------|--------------------------------------|
| 1  | Primaryid              | -                                    |
| 2  | I_F_Cod                | I                                    |
| 3  | Age                    | -                                    |
| 4  | Age_cod                | -                                    |
| 5  | Sex                    | -                                    |
| 6  | Occr_country           | US (United States)                   |
| 7  | Drug_seq               | I                                    |
| 8  | Role_cod               | PS (Primary Suspect)                 |
| 9  | Drugname               | -                                    |
| 10 | Prod_ai                | Loperamide dan                       |
|    |                        | Loperamide Hydrochloride             |
| 11 | Pt                     | -                                    |

The amount of data after the selection process from 2015-2017 was 6,366 records. The selection process is then carried out again to select adverse events in accordance with the issues raised. Adverse events that are used are adverse events related to disorders of the body. The selection process is carried out on the "pt" attribute. The amount of data after the final selection process was 4,079 records.

Preprocessing, Preprocessing stage or also called data cleaning will be performed cleaning data from missing values, inconsistent data, redundant data, and outliers.

- Check The Missing Value
  There were 1,239 data records were missing values. The missing value is found in the attributes "age" and "sex". Handling is done by removing data. The total data after the missing value handling stage is 2,840 records.
- Check The Inconsistent Data
There are 30 inconsistent data records on the "age_cod" attribute, the inconsistent data is data with the age code "DEC" or a decade, "DY" or day, and "MON" or month. The data to be used is the unit "YR" or year. The handling process is done by changing all units to "YR".

- Check The Redundant Data
  There is no redundant data, so the amount of data remains 2,840 records.
- Check The Outliers data
  There is no outlier data, so the amount of data remains 2,840 records.

Transformation, At this stage, the process of transforming data into a form or group is needed according to the research. In this research, there are 2 attributes that will be transformed, namely "age" into the age group and "pt" into the SOC (System Organ Class) group.

- Transformation age to age group
  The age group in this research uses age groups based of the Indonesian Department of Health. The age groups to be used can be seen in Table 3.

| No | Age Number | Age Group   |
|----|------------|-------------|
| 1  | 0 – 5      | Toddler     |
| 2  | 6 – 11     | Child       |
| 3  | 12 – 25    | Teenager    |
| 4  | 26 – 45    | Adult       |
| 5  | 46 – 65    | Elderly     |
| 6  | >66        | Older Person|

- Transformation attribute “pt” to “SOC”

The attribute "pt" will be transformed into group Organ System Class (SOC) using the Medical Dictionary for Regulatory Activities (MedDRA) conducted online through the site https://bioportal.bioontology.org/ontologies/MEDDRA. The SOC groups are in the 27 classes listed in Table 4.

| No | System Organ Class                                                                 |
|----|--------------------------------------------------------------------------------------|
| 1  | Blood and lymphatic system disorders                                                 |
| 2  | Cardiac Disorders                                                                   |
| 3  | Congenital, familial and genetic disorders                                          |
| 4  | Ear and labyrinth disorders                                                         |
| 5  | Endocrine disorders                                                                 |
| 6  | Eye disorders                                                                       |
| 7  | Gastrointestinal Disorders                                                          |
| 8  | General disorders and administration site conditions                                |
| 9  | Hepatobiliary disorders                                                             |
| 10 | Immune system disorders                                                             |
| 11 | Infections and infestations                                                         |
| 12 | Injury, poisoning and procedural complications                                       |
| 13 | Investigations                                                                     |
| 14 | Metabolism and nutrition disorders                                                  |
| 15 | Musculoskeletal and connective tissue disorders                                     |
| 16 | Neoplasms benign, malignant and unspecified (incl cysts and polyps)                 |
| 17 | Nervous System Disorders                                                            |
| 18 | Pregnancy, puerperium and perinatal conditions                                      |
Total data that will be used for the next stage amounted to 2,840 records and 4 attributes namely age, sex, drug name, and adverse events that have been transformed into System Organ Class levels (SOC).

2.3 Mining process with FP-Growth Algorithm

The FP-Growth algorithm work process will be explained. This table is 10 sample data used in this research. The minimum support to be used is 20%.

| Table 5. Sample Data |
|----------------------|
| No | Age_group | Sex | Drugname   | Adverse Event         |
|----|-----------|-----|------------|-----------------------|
| 1  | Adult     | M   | Loperamide | Cardiac Disorders     |
| 2  | Adult     | M   | Hydrochloride | Cardiac Disorders   |
| 3  | Adult     | M   | Loperamide | Psychiatric Disorders |
| 4  | Elderly   | F   | Imodium A-D | Gastrointestinal Disorders |
| 5  | Adult     | F   | Loperamide  | Psychiatric Disorders |
| 6  | Manula    | F   | Loperamide | Psychiatric Disorders |
| 7  | Adult     | M   | Loperamide | Cardiac Disorders     |
| 8  | Elderly   | F   | Loperamide | Gastrointestinal Disorders |
| 9  | Adult     | M   | Hydrochloride | Cardiac Disorders   |
| 10 | Elderly   | F   | Imodium A-D | Gastrointestinal Disorders |

The stages to be carried out are:

2.3.1 Count the support item. Count the support of each item. The formula of support is

\[
\text{Support} (A \rightarrow B) = \frac{\text{jumlah anggota (A$\cup$B)}}{\text{Total Transaksi}} \times 100\%
\]  

The value of support can be seen in Table 6.

| Table 6. Support item |
|-----------------------|
| No | Item          | Support | Support (%) |
|----|---------------|---------|-------------|
| 1  | Adult         | 6/10 = 0.6 | 60          |
| 2  | M             | 5/10 = 0.5 | 50          |
| 3  | F             | 5/10 = 0.5 | 50          |
| 4  | Loperamide    | 5/10 = 0.5 | 50          |
Items that have value below the minimum support of 20% will be eliminated. In the table above the item ‘Older Person’ item has 10% support, then the ‘Older Person’ item will be eliminated.

2.3.2 Sort items based on the highest support value. The sorted data can be seen in Table 7.

| No | Data |
|----|------|
| 1  | Adult, M, Loperamide, Cardiac Disorders |
| 2  | Adult, M, Cardiac Disorders, Loperamide Hydrochloride |
| 3  | Adult, M, Loperamide, Psychiatric Disorders |
| 4  | F, Elderly, Gastrointestinal Disorders, Imodium A-D |
| 5  | Adult, F, Loperamide Hydrochloride, Psychiatric Disorders |
| 6  | F, Loperamide, Psychiatric Disorders |
| 7  | Adult, M, Loperamide, Cardiac Disorders |
| 8  | F, Loperamide, Elderly, Gastrointestinal Disorders |
| 9  | Adult, M, Cardiac Disorders, Loperamide Hydrochloride |
| 10 | F, Elderly, Gastrointestinal Disorders, Imodium A-D |

2.3.3 FP-Tree. Based on Table 7 that has been formed, then made FP-Tree. FP-Tree that has been formed can be seen in Figure 2.
FP-Tree formed for research data totaling 2,840 records can be seen in Figure 3.

2.3.4 Conditional Pattern Base. Conditional pattern base is done by determining the tree branch with a path that ends with the smallest support value. Based on the minimum support value, the item that has smallest support value starts from the Imodium A-D, Gastrointestinal Disorders, Psychiatric Disorders, Loperamide Hydrochloride, Elderly, Cardiac Disorders, Loperamide, F, M, and Adults. This is the conditional pattern base of the Imodium A-D pattern.
This table is the formation of the Conditional Pattern Base Imodium A-D.

**Table 8. Conditional Pattern Base Imodium A-D**

| Item             | Conditional Pattern Base                      |
|------------------|-----------------------------------------------|
| Imodium A-D      | {F, Elderly, Gastrointestinal Disorders:2}    |

2.3.5 *Conditional FP-Tree.* FP-Tree Conditional is formed by counting the number of Support Count for each item that comes from the Conditional pattern base. Items that have a value greater than or equal to the minimum support will be raised while those that are not eliminated. Furthermore, any items that do not have a path that ends in Imodium A-D are discarded. Figure 5 is the form of the FP-Tree Conditional on Imodium A-D items.
This table is the formation of the Conditional FP-Tree Imodium A-D

| Item            | Conditional FP-Tree                      |
|-----------------|-----------------------------------------|
| Imodium A-D     | {F:2, Elderly:2, Gastrointestinal Disorders:2} |

2.3.6 Frequent Itemset. Next, find the frequent items set from forming a Conditional FP-Tree by combining items. This table is the result of frequent itemset of Imodium A-D that were formed:

| Item            | Frequent Itemset                                      |
|-----------------|------------------------------------------------------|
| Imodium A-D     | {F, Imodium A-D:2}, {Elderly, Imodium A-D:2}, {Gastrointestinal Disorders, Imodium A-D:2}, {F, Elderly, Imodium A-D:2}, {F, Gastrointestinal Disorders, Imodium A-D:2}, {F, Elderly, Gastrointestinal Disorders, Imodium A-D:2}, {F, Elderly, Gastrointestinal Disorders, Imodium A-D:2} |

2.3.7 Combination of patterns. The next stage is the formation patterns of frequent itemset that have been discovered. Based on the researcher’s goal, which is to find adverse event patterns based on age group, sex, and drug name, the frequent itemset that will be used is 4 items. In Table 10 the frequent itemset that has 4 items are {F, Elderly, Gastrointestinal Disorders, Imodium A-D} then the combination of patterns formed is as follows:

| Frequent Itemset | Pattern Combination                                      |
|------------------|---------------------------------------------------------|
| F, Elderly, Gastrointestinal Disorders, Imodium A-D | F, Elderly, Gastrointestinal Disorders → Imodium A-D |
| Imodium A-D      | Imodium A-D, Elderly, Gastrointestinal Disorders → F    |
|                  | Imodium A-D, F, Gastrointestinal Disorders → Elderly    |
|                  | Imodium A-D, F, Elderly → Gastrointestinal Disorders    |

The combination of patterns used is only a combination of patterns with consequent adverse events. Then the support, confidence, and lift ratio values are calculated from the pattern obtained. The formula of confidence is

\[
\text{Confidence}(A \rightarrow B) = \frac{\text{Jumlah anggota}(A \cup B)}{\text{Jumlah anggota}(A)} \times 100\%
\]
Lift Ratio is a measure to evaluate whether or not an association rule is strong. Lift Ratio compares the confidence of a rule with the value of benchmark confidence. Benchmark Confidence is a comparison between the number of all. If the value of the lift ratio is greater than 1, it indicates that the rule or rules have benefits and can be used. The higher the lift ratio, the greater the association strength [5]. Formula of lift ratio is

\[
\text{Lift Ratio} = \frac{\text{Confidence (A, B)}}{\text{Benchmark Confidence (A, B)}}
\]

(3)

\[
\text{Benchmark Confidence} = \frac{\text{Nc}}{\text{N}}
\]

(4)

Note:

Nc = Number of transactions with items that are consequent
N = Number of database transactions

| Rule | Confidence | Support | Nc | Be | Lift Ratio |
|------|------------|---------|----|----|------------|
| Imodium A-D, F, Elderly → Gastrointestinal Disorders | $\frac{2}{2} \times 100\% = 100\%$ (1) | $\frac{2}{10} \times 100\% = 20\%$ | 3 | 3/10 = 0.3 | 1/0.3 = 3.33 |

2.4 Implementation and Testing

2.4.1 Implementation with Matlab. The process of finding adverse event patterns in Loperamide using the Frequent Pattern Growth (FP-Growth) algorithm is implemented using matlab software. There are 5 fields to be filled in, namely adverse event data in .xls format, minimum support value, minimum confidence value, and selecting a pattern to be generated and file names that will be stored in the .xls format. The application will display the pattern information and the support, confidence, and lift ratio values for each pattern. Using 2,840 adverse event data with a minimum support value of 1% and minimum confidence of 10%, found 791 overall patterns and 23 patterns in accordance with the research objectives, namely 4 itemset patterns with consequent adverse events. This figure is from the Application view implemented.
2.4.2 Testing with SPMF. This test is carried out to determine the accuracy of the number of patterns generated, the value of support, confidence, and lift ratio between applications built with SPMF tools. Based on the testing that has been done, it is known that the number of association patterns result between applications built with SPMF tools is equal to 791 for all patterns and 23 for 4 itemset patterns with consequent adverse events, and the results of the value of support, confidence, and lift ratio also has the same between the application and SPMF. Based on this, it can be concluded that the results of the application built with the results of the tools have the same value. This figure is a test with SPMF that has been done.
3. Result and Discussion
The pattern used in accordance with the research objectives is a 4 itemset pattern with consequent adverse events. There are 23 matching patterns. The pattern that can be used as useful information is a pattern that has a lift ratio value ≥ 1. The patterns that meet the lift ratio are 20 patterns. The pattern can be seen in Table 13.

| No | If | Then | Support (%) | Confidence (%) | Lift Ratio |
|----|----|------|-------------|----------------|-----------|
| 1  | Adult, M, Loperamide | Cardiac Disorders | 3.803 | 28.346 | 1.483 |
| 2  | Adult, M, Loperamide | Psychiatric Disorders | 2.817 | 20.997 | 1.302 |
| 3  | Remaja, M, Loperamide | Psychiatric Disorders | 2.711 | 26.014 | 1.613 |
| 4  | Remaja, M, Loperamide | Cardiac Disorders | 2.606 | 25 | 1.308 |
| 5  | Manula, F, Imodium A-D | General Disorders And Administrations Site Conditions Gastrointestinal Disorders | 2.535 | 36.923 | 1.945 |
| 6  | Manula, F, Imodium A-D | General Disorders And Administrations Site Conditions | 2.5 | 36.41 | 2.929 |
| 7  | Manula, F, Loperamide Hydrochloride | General Disorders And Administrations Site Conditions | 2.43 | 53.488 | 2.818 |
| 8  | Adult, M, Loperamide Hydrochloride | Cardiac Disorders | 2.007 | 29.843 | 1.561 |
| 9  | Adult, F, Loperamide | Psychiatric Disorders | 2.007 | 26.636 | 1.652 |
| 10 | Remaja, F, Loperamide | Psychiatric Disorders | 2.007 | 23.171 | 1.437 |
| 11 | Adult, F, Loperamide | Cardiac Disorders | 1.866 | 24.766 | 1.295 |
| 12 | Remaja, F, Loperamide | Cardiac Disorders | 1.831 | 21.138 | 1.106 |
| 13 | Adult, M, Loperamide | Investigations | 1.444 | 10.761 | 1.634 |
| 14 | Adult, M, Loperamide | Nervous System Disorders | 1.373 | 10.236 | 1.14 |
| 15 | Manula, M, Imodium A-D | Gastrointestinal Disorders | 1.338 | 42.697 | 3.435 |
| 16 | Remaja, M, Loperamide Hydrochloride | Cardiac Disorders | 1.162 | 25.984 | 1.359 |
| 17 | Adult, M, Loperamide Hydrochloride | Psychiatric Disorders | 1.127 | 16.754 | 1.039 |
| 18 | Manula, F, Loperamide Hydrochloride | Gastrointestinal Disorders | 1.092 | 24.031 | 1.933 |
| 19 | Remaja, M, Loperamide | Nervous System Disorders | 1.092 | 10.473 | 1.166 |
| 20 | Manula, M, Imodium A-D | General Disorders And Administrations Site Conditions | 1.056 | 33.708 | 1.776 |

Based on the table, the research objective to find the adverse event patterns in Loperamide was successfully, and also in the pattern results table above it is known that the pattern with the highest support value is a combination of adult age groups, male sex, drug name Loperamide and adverse event cardiac disorders has a support value of 3.8%, confidence 28.35% and lift ratio 1.483. From this
5. Conclusion
Based on the results of the research, the following conclusions are obtained:
1. Successfully built an application to identify adverse event patterns in Loperamide using the Frequent Pattern Growth (FP-Growth) algorithm. 2. Using 2,840 adverse event data with a minimum support value of 1% and minimum confidence of 10%, found 791 overall patterns and 23 patterns in accordance with the research objectives, namely 4 itemset patterns with consequent adverse events. 3. Patterns with the highest support values are a combination of adult age groups, male sex, drug names Loperamide and adverse event cardiac disorders which have a support value of 3.8%, confidence 28.35% and lift ratio 1.483.

Reference
[1] Raini M, Gitawati R and Rooslamiati I 2015 Kerasionalan Penggunaan Obat Diare yang Disimpan di Rumah Tangga di Indonesia. Jurnal Kefarmasian Indonesia, 5(1), 49-56. doi:10.22435/jki.v5i1.4089.49-56
[2] Hasanah F, Puspitasari H P and Sukorini A I 2013 Profil Penggalian Informasi dan Rekomendasi Pelayanan Swamedikasi Oleh Staf Apotek Terhadap Kasus Diare Anak di Apotek Wilayah Surabaya. Farmasains, 2(1).
[3] FAERS, Questions and Answers on FDA’s Adverse Event Reporting System (FAERS). April. 2018 [Online]. Available: https://www.fda.gov/drugs/surveillance/fda-adverse-event-reporting-system-faers (Accessed: December. 2018)
[4] FAERS, CFR- Code of Federal Regulation Title 21. April 2018[Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=312.32 (Accessed: December. 2018)
[5] FAERS, FDA Drug Safety Communication: FDA warns about serious heart problems with high doses of the antidiarrheal medicine loperamide (Imodium), including from abuse and misuse. Januari.2018.[Online]. Available: https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-about-serious-heart-problems-high-doses-antidiarrheal (Accessed: December.2018)
[6] Han J, Kamber M and Pei 2012 Data Mining Concepts and Techniques, Third. Morgan Kaufmann
[7] Santoso B 2007 Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu