Application of K-Mean Algorithm for Medicine Data Clustering in Puskesmas Rumbai

Taslim¹, Fajrizal², and Dafwen Toresa³
¹E-mail: taslim.malano@gmail.com, University of Lancang Kuning, Pekanbaru, Riau, Indonesia
²E-mail: Fajrizal@unilak.ac.id, University of Lancang Kuning, Pekanbaru, Riau, Indonesia
³E-mail: dafwentoresa@gmail.com, University of Lancang Kuning, Pekanbaru, Riau, Indonesia

Abstract.
Through the government’s health insurance program, efforts are made to ensure the health of the community through Puskesmas or community clinics. One of the most important components in health is the availability of medicines. The availability of medicines should be well managed to ensure that the medicines needed by the community are always available in sufficient quantities. Clustering on Data mining can be used to analyze the use of medicines during this time at a Puskesmas to be used as one of considerations for the Puskesmas to submit the demand of medicines in the period to come. The results of this study are expected to classify the level of medicines used in the pharmacy of Puskesmas in Rumbai Bukit Pekanbaru.

1. Introduction
Medicines is one of the important components in terms of good health to either prevent, reduce, eliminate or cure a disease or disease symptoms. That is why medicines need to be managed properly, effectively and efficiently. Planning on the needs of medicines is important to ensure the availability and distribution of medicines with the types and quantities sufficient so that medicines can be obtained quickly at the appropriate place and time at agencies related to health services, be it hospital, health centre, and so forth. The planning on the need of medicines will affect the procurement, distribution and usage of medicines in health care settings. Puskesmas Rumbai Bukit Pekanbaru is one of the public health service centres located in Pekanbaru which is government institution related to public health service in which one of its function isto provide medicine service for institution of health care settings in Pekanbaru. Effective and efficient analysis of the need of medicines is needed to ensure the availability of medicines in the pharmacy in Pekanbaru.

Clustering on the need of medicines is expected to be one of the considerations to ensure the availability of medicines in the health service of Pekanbaru. Data clustering is one method in data mining that can be used to get data mapping to classify into smaller groups based on the similarity of characteristics they have (Perim, Wandekokem, & Varejão, 2008). With these clustering results the distribution of medicines in health care agencies can be grouped according to the need based on the medicine distribution data in the previous year and it can be used as a reference for drug planning for the next year. It is hoped that the availability of medicines for the next year can be more secure and able to meet demand for medicines from health agencies.
One of the most well-known clustering methods among clustering algorithms is K-means (Patel & Mehta, 2011). The simplicity of this method makes the K-means algorithm applicable to various fields (K.Arai and A.R. Barakbah, 2007).

2. Phase stages in data mining
KDD is a nontrivial process of identifying the validity of data, potential, use, and ultimately yielding understandable data patterns. The stages in the stages of data mining are:

![Figure 1: Stages of data mining](image)

In this study, the data used are those taken from the Usage Report and Medicine Demand Sheet (LPLPO) Puskesmas Rumbai Bukit 2015. Then, the data are clustered to obtain patterns of medicine needs for the community of Rumbai in Pekanbaru.

2.1 Clustering
Clustering refers to grouping notes, observations, or cases into similar classes. A cluster is a collection of notes that resembles each other and differs from records in other clusters. Clustering is different from the classification which has no target variable for clustering. Instead, the algorithm clustering looks for the entire set of data segments into a relatively homogeneous subgroup or group, in which the similarity of records in the cluster is maximized, and the similarity of records out of this cluster is minimized.

Examples of grouping tasks in business and research include:
1) Determining product marketing targets
2) For accounting audit purposes, suspicious financial behavior is examined
3) It is used as a dimension reduction tool in which the data set has hundreds of Attributes

For grouping of gen expression, those in large number may exhibit the same behavior. In cluster analyzing, some factors need to note, for example such as determining
1) How to measure similarities
2) How to recode category variables
3) How to standardize or normalize numerical variable
4) How many clusters will be made
5) Interval-variable scale
Variable-scale interval is a continuous measurement which includes a linear scale. For instance the weight and height, latitude and longitude and weather temperature. The use of measurement units may affect cluster analysis [8]. For example, a changing measurement from meter to inch, or from kilogram to pound, may result in a very different cluster structure. For that reason, it requires normalization of data, so that all data have the same weight. Steps for normalizing the data are:

- Calculating the mean value
  \[ s = \frac{1}{n} (|x_1 - m| + |x_2 - m| + \cdots + |x_n - m|) \]

- Calculating the z-score
  \[ z = \frac{x - m}{s} \]

K-Means Algorithm
Clustering k-means algorithm [1] is a simple and effective algorithm for finding clusters in data with the following algorithms:
1) Determine the number of cluster
2) Determine the value of the location of the initial cluster.
3) Calculate the closest cluster center for each record
4) For each cluster k, calculate the centroid cluster and update Location of each cluster center
5) Repeat steps 3 through 5 until convergence or termination.

The k-means algorithm is known and widely used for the partitional method, which is to divide the set of data objects into a subset of non-overlapping clusters, so that each data object is exactly in one cluster.

3. Results and Discussion
The source of data used in this study is data from LPLPO Puskesmas Rumbai Bukit 2014. The data used can be seen in table 1.

Table 1. Recapitulation of prescription data 2014

| NO | NAMA OBAT | SERTUEN | Jan | Feb | Mar | April | Mei | Juni | Juli | Agust | Sept | Okt | Nov | Des |
|----|-----------|---------|-----|-----|-----|-------|-----|------|------|-------|------|-----|-----|-----|
| 1  | Acetamin 100 mg | tablet | 4   | 5   | 4   | 21    | 15  | 19   | 19   | 19    | 24   | 24  | 91  | 25  |
| 2  | Acetamin 200 mg | tablet | 23  | 23  | 23  | 100   | 100 | 100  | 100  | 100   | 100  | 100 | 100 | 100 |
| 3  | Acetamin 400 mg | tablet | 18  | 18  | 18  | 18    | 18  | 18   | 18   | 18    | 18   | 18  | 18  | 18  |
| 4  | Aspirin 100 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 5  | Aspirin 500 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 6  | Albendazole 400 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 7  | Albendazole 400 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 8  | Amoxicillin 250 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 9  | Amoxicillin 500 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |
| 10 | Amoxicillin 500 mg | tablet | 1   | 1   | 1   | 1     | 1   | 1    | 1    | 1     | 1    | 1   | 1   | 1   |

There are 133 types of medicines at pharmacies in Puskesmas of Rumbai Bukit. Some medicines and medicines spending each month can be seen in the following table.
Table 2. Monthly Recapitulation of Medicines

| NO | Names of Medicines | units | Jan | Feb | March | April | May | June | July | August | Sept | Oct | Nov | Dec |
|----|--------------------|-------|-----|-----|-------|-------|-----|------|------|--------|------|-----|-----|-----|
| 1  | Acyclovir krim     | tube  | 4   | 5   | 4     | 21    | 23  | 15   | 19   | 19     | 24   | 24  | 91  | 75  |
| 2  | Acyclovir 200 mg   | tablet| 20  | 110 | 20    | 95    | 110 | 132  | 23   | 0      | 0    | 0   | 0   | 0   |
| 3  | Acyclovir 400 mg   | tablet| 10  | 0   | 0     | 488   | 0   | 0    | 210  | 100    | 200  | 195 | 220 | 115 |
| 4  | Alatsuntik 1 ml    | set   | 0   | 0   | 0     | 0     | 0   | 0    | 0    | 100    | 0    | 0   | 0   | 0   |
| 5  | Alatsuntik 5 ml    | set   | 0   | 0   | 0     | 0     | 0   | 0    | 0    | 0      | 100  | 0   | 0   | 0   |
| 6  | Albendazol 400 mg  | tab   | 23  | 54  | 13    | 6     | 32  | 87   | 15   | 15     | 16   | 6   | 66  | 69  |
| 7  | Albendazol 100 mg  | tablet| 20  | 170 | 150   | 60    | 70  | 270  | 50   | 493    | 0    | 0   | 0   | 0   |
| 8  | Ambroxol 30 mg     | tablet| 191 | 200 | 3028  | 225   | 127 | 72   | 120  | 20     | 149  | 340 | 1602| 1035|
| 9  | Ambroxolinsirup    | botol | 72  | 56  | 57    | 45    | 28  | 22   | 50   | 30     | 193  | 0   | 0   | 0   |
| 10 | Aminofilin tab 200 mg | tablet | 10 | 0   | 152   | 0     | 72  | 10   | 0    | 0      | 40   | 0   | 45  | 33  |
| 11 | Amoksisilin tab 250 mg | kapsul | 1020| 924 | 844   | 669   | 750 | 1456 | 1344 | 780    | 1142 | 1308| 2002| 1474|
| 12 | Amoksisilin cap 500 mg | kapsul | 2245| 2310| 2370  | 1780  | 1503| 1765 | 1795 | 1210   | 4340 | 2110| 2754| 2042|
| 13 | Amoksisilinsyrkering 125 mg | botol | 110 | 110 | 77    | 83    | 96  | 166  | 162  | 91     | 145  | 161 | 193 | 150 |
| 14 | Ambprinbesilat 10 mg | tablet| 150 | 130 | 180   | 550   | 795 | 200  | 20   | 40     | 371  | 90  | 400 | 110 |
| 15 | Antasida tab doen  | tablet| 912 | 2474| 1023  | 1090  | 1098| 1145 | 960  | 268    | 529  | 693 | 2030| 1724|
| 16 | Antasidadoensuspensi | botol | 37  | 29  | 22    | 37    | 18  | 13   | 49   | 44     | 100  | 62  | 60  | 45  |
| 17 | Anti bakteridoensalap | tube | 14  | 21  | 20    | 15    | 10  | 19   | 24   | 7      | 12   | 17 | 35  | 25  |
|    |                    |       |     |     |       |       |     |      |      |        |      |    |    |    |
|    |                    |       |     |     |       |       |     |      |      |        |      |    |    |    |
| 133| Loperamid          |       | 94  | 122 | 128   | 28    | 0   | 0    | 0    | 0      | 0    | 0  | 0   | 0   |

From the above data (table 2), it leads to the process of data normalization so that the data has a smaller range of values that can accelerate the process of calculate of the normalization. This study uses z-score normalization. The next step is determine the number of clusters, that is 3 clusters in order to be able to determine whether a demand for the medicines are high", "high", or "low". For initial centroid values, it is randomly selected. The value of the centroid is taken from rows 50.60 and 70. In this step, updating of group cluster will be done. Next is clustering process d0 using Euclidean Distance, to get distance of matrix C1, C2 and C3. From clustering result after 33 times iteration, it is found out that data medicine is grouped into 3 groups, they are: very high, high, and low demands of medicines. The number of clustering results can be seen in the following table.

Table 3. Clustering 1

![Figure 1](image-url)
Some details of table on clustering medicine data can be seen in the following table.

**Table 4. Clustering Results of medicine demand**

|   | Medicine          | Type  | 4 | 5 | 4 | 21 | 23 | 15 | 19 | 19 | 24 | 24 | 91 | 75 |
|---|-------------------|-------|---|---|---|----|----|----|----|----|----|----|----|----|
| 1 | Acyclovir krim    | tube  | 4 | 5 | 4 | 21 | 23 | 15 | 19 | 19 | 24 | 24 | 91 | 75 |
| 2 | Acyclovir 200 mg  | tablet| 20 | 110 | 20 | 95 | 110 | 132 | 23 | 0 | 0 | 0 | 0 | 0 |
| 3 | Acyclovir 400 mg  | tablet| 10 | 0 | 0 | 0 | 0 | 210 | 0 | 100 | 200 | 195 | 220 | 115 |
| 4 | Alatantik 1 ml    | set   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| 5 | Alatantik 5 ml    | set   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| 6 | Albendazol 400 mg | tablet| 23 | 54 | 13 | 6 | 32 | 87 | 15 | 15 | 16 | 6 | 66 | 69 |
| 7 | Aluporinil tab 100 mg | tablet| 20 | 170 | 150 | 60 | 70 | 270 | 50 | 493 | 0 | 0 | 0 | 0 |
| 8 | Ambroxol 30 mg    | tablet| 191 | 200 | 3028 | 225 | 127 | 72 | 120 | 20 | 149 | 340 | 1602 | 1035 |
| 9 | Ambroxolkrup      | botol | 72 | 56 | 57 | 45 | 28 | 22 | 50 | 30 | 193 | 0 | 0 | 0 |
| 10| Amifostin 200 mg  | tablet| 10 | 0 | 0 | 0 | 0 | 72 | 10 | 0 | 0 | 40 | 0 | 45 | 33 |
| 11| Amoxicillin tab 250 mg | kapsul | 1020 | 924 | 844 | 660 | 750 | 1456 | 1344 | 780 | 1142 | 1308 | 2002 | 1474 |
| 12| Amoxicillin cap 500 mg | kapsul | 2245 | 2310 | 2370 | 1780 | 1503 | 1765 | 1795 | 1210 | 4340 | 2110 | 2754 | 2042 |
| 13| Ambosilin syruping 125 mg | botol | 110 | 110 | 77 | 83 | 96 | 166 | 162 | 91 | 145 | 161 | 193 | 150 |
| 14| Amflopsinbesilat 30 mg | tablet | 150 | 130 | 180 | 550 | 755 | 200 | 20 | 40 | 371 | 90 | 400 | 110 |
| 15| Antasil tab doen   | tablet | 912 | 2147 | 1028 | 1090 | 1098 | 1145 | 990 | 268 | 529 | 633 | 2030 | 1724 |
| 16| Antassidoenserupti | botol | 37 | 29 | 22 | 37 | 18 | 13 | 49 | 44 | 100 | 62 | 60 | 45 |
| 17| Anti bakteridensalap | tube | 14 | 21 | 29 | 15 | 10 | 19 | 24 | 7 | 12 | 17 | 35 | 25 |
| 18| Anti fungidensalap | pot   | 15 | 44 | 14 | 14 | 21 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19| Anti hemoroidsuppkomb | supp | 0 | 0 | 0 | 0 | 0 | 13 | 16 | 0 | 3 | 0 | 8 | 0 |
| 20| Asamaskorbat tab 50 mg | tablet | 1419 | 3812 | 3781 | 3005 | 2784 | 3392 | 3890 | 2703 | 3955 | 3567 | 7090 | 6112 |
| 21| Asammefenamat tab 100 mg | kapsel | 800 | 940 | 1040 | 1032 | 930 | 1000 | 570 | 420 | 820 | 715 | 1798 | 1510 |

4. Conclusion

From the results of data cluster of medicines, it can be concluded that medicines are grouped into low. The low demand of the medicines is caused by no demand at all for few months. For high demand, the average demand is 300 medicines. For the very high demand, the average demand is above 2000 medicines every month. This result of cluster analysis above still needs to be elaborated in order cluster data of medicines can be done more validly by setting the best centroid value.

5. References

[1] Arai. K and A.R. Barakbah “Hierarchical K-means : and algorithm for centroids initialization for K-means”.(2007)
[2] Hans-Peter. K, Karsten, Peer.K, Alexey.P, Matthias.S, Arthur.Z, “Future trends in data mining” (2007)
[3] Ranjan, J, “Application of data mining techniques in pharmaceutical industry” (2007)
[4] Tipawan.S and Kulthida. T “Data mining and its application for knowledge management: a literature review from 2007 to 2012” (2012)
[5] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth “From Data Mining to Knowledge Discovery in Databases” (1996).