

Stock management using K-means method and Time Series method as Stock Order

Komang Sri Utami^{1*}, I Gede Wira Darma², and Ni Wayan Sri Aryani³

^{1,2}Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University,

³Department of Electrical and Computer Engineering, Udayana University

*Email: sriutami821@gmail.com

Abstract Good stock management is one of the keys to success for sales businesses. A stable stock flow will affect the cost of purchasing goods and income. This condition can be achieved when the prediction of the required stock is right, so there is no accumulation of stock or empty stock. The case to be taken is for drug management of a pharmacy. This study uses the K-means method and time series method. The K-means method is a grouping method that is very easy to use and implement. Drug groupings will be made into 3 types, namely the best-selling, selling, and less-selling groups. While the regression time series method is used to predict the stock to be purchased that will be used in two weeks so that there is no stock buildup. Both of these methods are used to provide a grouping of drugs and the right amount of medicine to buy so that the management of drug stocks can be done well. The results of the tests carried out using 1000 test data, in which the K-means grouping test was $C1 = 13$, $C2 = 29$, $C3 = 958$ which was obtained from 11 iterations that had been done. In addition, each drug item has been predicted for the number of drugs to be purchased according to the sales performance of the last 3 months. From both of these results, it can be a reference in making order decisions to better manage stocks.

Keywords — Stock Manajement, K-means, Regression Time Series, Stoct Order

I. INTRODUCTION

Some data mining systems for trend prediction, to find patterns that are not known before, or for decision making have been investigated to date. Most of the research studied is related to the sale of goods and services both online and standalone in a computerized manner. Previous research, there are several approaches based on data mining [2,3,5] which focus on market segmentation and consumer interest, [1] which focuses on performance quality evaluation, APBD clustering [6] using k-means.

All of the research focuses on getting results that can later be used as decision making. In this study, a grouping of drugs will be conducted using the K-means method and the Regression Time Series method. K-means is known as a very simple media to implement and use. This grouping is made to facilitate the taking of priority levels of drugs that will be ordered so that there is no accumulation of stock or lack of stock in the future. While the Regression Time Series method is used for forecasting the required stock. Of course, this can help management to do better stock management.

The contribution of this paper is to introduce the application of the K-means method as a drug clustering to the sale of pharmacies and Regression Time Series as a determinant of order quantity. K-means clustering is an

unsupervised learning method which is used on data that does not have a label. The purpose of the unsupervised learning method is to do clustering, which is to group data with similar characters. This method is very suitable for data that has a lot of capacity. Which will be used in this study. The Regression Time Series method is also a simple method that can be used to predict the amount of stock that must be purchased by taking into account the history of transactions that have been carried out.

This paper is organized as follows. In the next section, the research objectives are mentioned. In the third part, several theories relating to this research will be mentioned. In the fourth part, the methodology used will be explained. In the fifth section, the results of the tests will be explained. And the sixth part contains conclusions from this study.

II. PURPOSE OF PAPER

The purpose of this study is to group data based on existing data, namely sales, number of transactions, and average sales, and predict the amount of stock needed as consideration of order decisions.

III. LITERATUR REVIEW

A. Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. The term data mining has the essence as a discipline whose sole purpose is to discover, dig, or mine the knowledge of the data or information we have. Data mining, also known as Knowledge Discovery in Database (KDD). KDD is an activity that includes the collection, use of data, historically to find regularity, patterns or relationships in large data sets. [1]

(1) Training Methods

Broadly speaking the training methods used in data mining techniques are divided into two approaches, namely

- a) Unsupervised learning, this method is applied without any training and without any teacher. The teacher here is the label of the data
- b) Supervised learning, that is learning method with training and trainer. In this approach, to find the decision function, separator function or regression function, multiple examples of data that have output or labels are used during the training process.

(2) Grouping Data Mining

There are several techniques that data mining has on the task that can be done, namely:

- a) Description
Researchers usually try to find ways to describe hidden patterns and trends in the data.
- b) Estimates
Estimates are similar to classifications, except destination variables that are more numerical in direction than on the category
- c) Prediction
Predictions bear a resemblance to estimation and classification. Only, the prediction of the results shows something that has not happened (may happen in the future).
- d) Classification
In the classification of variables, objectives are categorical. For example, we will classify income in three classes, which are high income, medium income, and low income.
- e) Clustering
Clustering is more toward grouping records, observations, or cases in a class that have similarities.
- f) Association
Identify the relationship between the events that occur at one time.

B. K-means Method

One of the clustering algorithm in data mining is the K-Means Clustering algorithm to be able to produce groups that share similar attributes. K-Means is the most popular and widely used clustering method in many fields because it is

simple, easy to implement, has the ability to cluster large data, able to handle data outliers. K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters. This method partitions the data into clusters / groups so that data that have the same characteristics are grouped into the same clusters and data that have different characteristics are grouped into other groups. The K-Means algorithm is a relatively simple algorithm for classifying or grouping large numbers of objects with certain attributes into groups as much as K. K-Means one of the non-hierarchical clustered data methods that attempts to partition the existing data into in the form of one or more clusters or groups. The K-Means grouping algorithm will result in a group of notes as much as k pieces. The K-Means algorithm was first conceived by J. Mac Queen. The steps of doing Clustering with the K-Means method are as follows [1].

- Select the number of clusters k.
- The initialization of this cluster center k can be done in various ways. But the most frequently done is by random. Cluster centers are preliminarily scored with random numbers
- Allocate all data / objects to the nearest cluster. The proximity of two objects is determined by the distance of the two objects. Likewise the proximity of a data to a particular cluster is determined the distance between the data with the cluster center. In this stage we need to calculate the distance of each data to each cluster center. The most distance between one data and one particular cluster will determine which data to enter in which cluster. To distance all data to each cluster center point can use Euclidean distance theory formulated in formula 1:
- Recalculate cluster center with current cluster membership. The cluster center is the average of all data / objects in a particular cluster. If desired it can also use the median of the cluster. So the mean (mean) is not the only size that can be used

C. Regression Time Series

Time series data is a sequence of data collected over a period of time. Time series data for stock market predictions can be collected daily, weekly, monthly or yearly[4]. Time series data analysis extracts statistical information that is useful for understanding data characteristics. Time series forecasting techniques involve the use of models to predict future values based on past information.

Often time series economics are based on monthly or quarterly data that follows a seasonal pattern (regular up and down movements). Therefore it is necessary to eliminate seasonal factors from the time series data so that it can focus on conditions such as trend. The process of eliminating seasonal factors is called deseasonalization or seasonal adjustment. One of the methods of deseasonalization is the dummy variable method.

The following is an example of a model using the dummy technique with a quarterly seasonal pattern:

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t$$

Figure 1 Regression Time Series Model

IV. RESEARCH METHOD

This research was conducted at ABC Pharmacy which focuses on the main objectives of drug clustering based on the transactions owned and also the predictions of stocks needed for better stock management. This research consists of several stages that can be used as clear and easy solutions to problem solving.

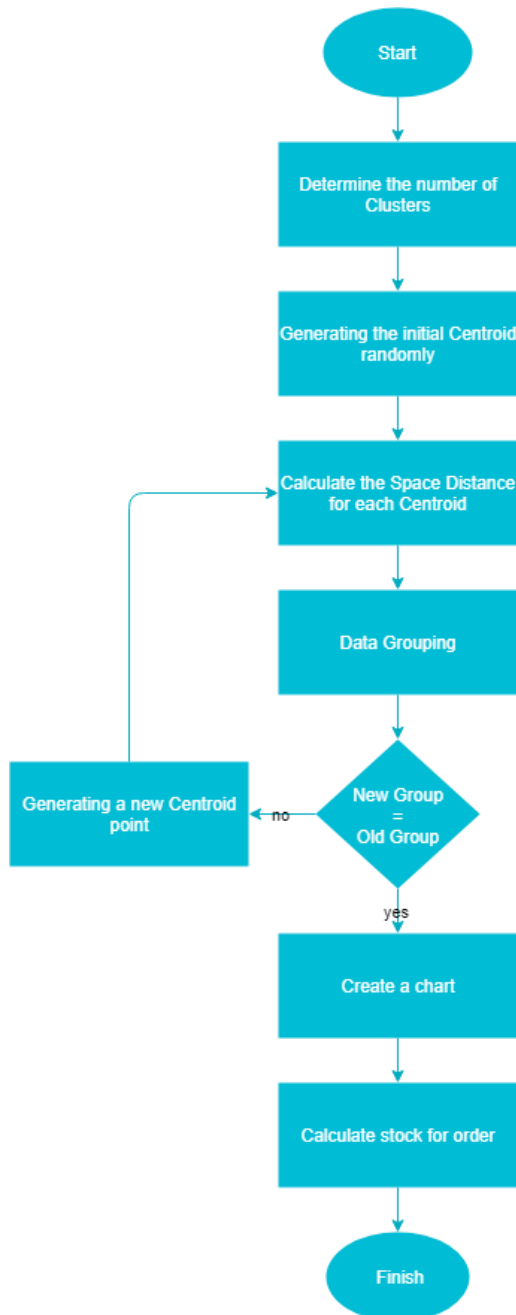


Figure 2. Research Methods

In the Figure 2, show step by step in this research, which describe the steps being taken:

- 1) Determine the number of clusters.
- 2) Produce a centroid that is determined from the maximum, medium and minimum values of the sales transaction.
- 3) Calculate the distance data to each centroid.
- 4) Data grouping. If the new group value is formed = the old group value, then make a graph, if not then make new cluster and return to process 3. As such then until the new group and the old group get the same value.

V. ANALYSIS AND RESULT

In In this section, the researcher will discuss each process described in the methodology section and will discuss the results of the tests that have been carried out. Performed three (3) times of testing, namely by using 30 data, 100 data, and 1000 data.

A. Number of Cluster

In this study used 3 (three) groupings, namely the best-selling (C1), standard-selling (C2), and less-selling (C3) groups.

B. Centroid Value

In the first test, with 30 test sample data, the centroid used was determined from the maximum, medium and minimum values, namely: 30th data, 15th data, and 22nd data.

C. Distance Space Data

After the centroid is determined, it will calculate the distance space data to each centroid. Following is the formula used to calculate distance space:

$$d_{euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Figure 3 Distance space formula

The results of calculations that have been done are presented in table 1, and then make marked with a value of 1 if the value of the centroid group is the value of the distance space obtained.

Table 1. Result of distance space 1

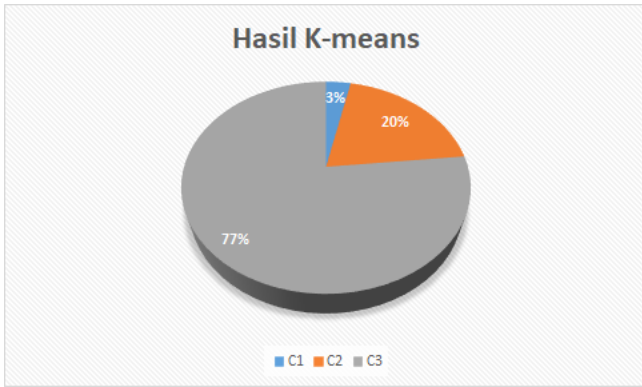


Figure 5 Chart of cluster

After the first test, two more tests were conducted, using 100 data and 1000 data, the results are as follows:

Testing 2: 100 data with 6 iterations

C1 = 2 item, C2 = 10 item, C3 = 88 item

Testing 3: 1000 data with 11 iterations

C1 = 13 item, C2 = 29 item, C3 = 958 item

E. Calculate stock for order

The following is an example of calculating a stock of drugs that must be ordered from Acyclovir 400 mg tab.

Year	Time Period (X)	X ²	Total Item Sold (Y)	X.Y
2019				
January	1	1	40	40
February	2	4	65	130
March	3	9	60	180
April	4	16	40	160

$$\sum X = 10$$

$$\sum X^2 = 30$$

$$X_{average} = \frac{10}{4} = 2.5$$

$$\sum Y = 200$$

$$Y_{average} = \frac{200}{4} = 50$$

$$\sum XY = 510$$

$$b = \frac{\sum XY - nX_{average} nY_{average}}{\sum X^2 - nX^2} = \frac{510 - 4(2.5)(50)}{30 - 4(2.5)^2} = \frac{10}{5} = 2$$

$$a = Y_{average} - b X_{average} = 50 - 2(2.5) = 45$$

Note :

a = stock for order

b = buffer stock

From these results, the number of people who must be ordered for Acyclovir 400 mg tab is as many as 45 items.

VI. CONCLUSION

There are several that can be concluded from this study, that is:

- 1) Centroid values are divided into 3 clusters where C1 is the highest transaction level, C2 is the average transaction level and C3 is the lowest transaction level.
- 2) The more data sets tested, the more iterations are carried out.
- 3) Obtained amounts must be ordered so there is no accumulation of drug items
- 4) Based on the testing data above, it is known that only a few items of drugs which are in demand and the amount of stock that must be ordered, as a material consideration for regulating the purchase of stock items for drugs.

REFERENCES

- [1] Dwi Ardiada, Pasek Agus Ariawan, and Made Sudarma. 2018. Evaluation of Supporting Work Quality Using K-Means Algorithm. 3, 1 (2018), 4.
- [2] Jejen Arisandi. 2016. Prediksi Kebutuhan Buah dengan Segmentasi Pasar Menggunakan K-Means. 2, 1 (2016), 3.
- [3] Penda Sudarto Hasugian. 2018. PENERAPAN DATA MINING UNTUK KLASIFIKASI PRODUK MENGGUNAKAN ALGORITMA K-MEANS. 2, 2 (2018), 8.
- [4] A Subashini and Dr M Karthikeyan. Forecasting on Stock Market Time Series Data Using Data Mining Techniques. *Internet Of Things*, 8.
- [5] Teguh Nurhadi Suharsono and Maulana Akbar. 2017. PREDIKSI MINAT KONSUMEN SESUAI MUSIM MENGGUNAKAN ALGORITMA K-MEANS PADA PERANGKAT LUNAK PEMESANAN TEMPAT DAN MAKANAN ONLINE. 2, 2 (2017), 6.
- [6] I Putu Ari Wijaya and Ida Bagus Alit Swamardika. 2017. Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm. 2, 1 (2017), 7.