Determine Types of Indonesian Hospital by Criteria-based Proses Model, K-means Cluster, and Hierarchical Average Linkage

Kristiana,
Department Business Statistics,
Institut Teknologi Sepuluh Nopember
email: anakristi18@gmail.com

Kelly Rossa Sungkono
Department of Informatics,
Institut Teknologi Sepuluh Nopember
e-mail: kelly@its.ac.id

Riyanarto Sarno
Department of Informatics,
Institut Teknologi Sepuluh Nopember
email: riyanarto@if.its.ac.id

***Abstract*—Hospital is an institution that provides health services to all communities. Hospitals in Indonesia have class type divisions, namely class A, class B, class C, and class D. Several criteria determine each classification. These criteria are human resources, such as the availability of doctors and pharmacists and time criteria. However, the values of those criteria are usually obtained by observation directly. The aim of this study is defining hospital values for each criterion through model process hospital directly. Firstly, this experiment forms a log of processes in the hospital instantly by using a workflow system. Then, a log is processes based on the hospital criteria. The final values of those criteria are processed by K-means Cluster and Hierarchical Average Linkage to obtain the hospital class. The final results of the experiment show that the K-means Cluster method is better with the Hierachical Average Linkage method in terms of accuracy, precision, and recall.**

***Keywords*—classification, Hierarchical Average Linkage, Indonesian hospital, K-means Cluster.**

I. INTRODUCTION

Hospital is a comprehensive part of a social and medical organization, which has the function of providing health services to the community both curative and preventive, where the services provided are inpatient, outpatient, and emergency units[1]. Hospitals in Indonesia have class type divisions, namely class A, class B, class C, class D and the difference in the four classes lies in facilities and medical support. Each hospital class has its own criteria. The used criteria in this study are the criteria of human resources and time criteria. The used human resources include doctors and pharmacists in outpatient care, inpatient care, and emergency units, while the used time criteria include waiting time in outpatient and waiting time in a prescription drug. The criteria values are obtained based on the hospital process model. To determine the hospital class, researchers use clustering methods, which are K-means and Hierarchical Average Linkage. Researchers choose the best result based on those methods.

So far, the determination of criteria is carried out by direct observation by the auditor. This way takes time, cost, and energy from the auditor. Therefore researchers want to streamline the method of determining the type of hospital by analyzing directly from the process model. Process models have been widely utilized for measuring the performance of processes in the company and pointing issues related to those processes[2]. In the process,the model containsan event that aims to help users in analyzing business processes[3]**.** Business processes divided into declarative models and imperative models**.** The declarative model has rules to describe the relation of flexible processes. The imperative model uses control-flow patterns to depict the relation of processes in detail**.** In this research, the business process uses the imperative model. The imperative model uses control-flow patterns uses to depict the detail of the activities flows. The imperative model is used in many ways, such as business, fraud, medical, and advertisement[4]. From the process model, it will be analyzed using the clustering method.

Cluster analysis is grouping objects or cases into smaller groups where each group contains objects that are similar to each other[5]. In the cluster, the measurement called a measure of distance or similarity is the principal part of the grouping. The used standard of distance is Euclidean. The Euclidean is the distance from the square root of the sum of the origin and destination values for all variables[6]. This paper uses two methods, namely the Hierarchy Method and the Non-Hierarchy Method. The Hierarchy Method is used as a weighting counter for each criterion[7]. As the basis for clustering, Hierarchy Method made clusters in terms of measures of spatial distance among data vectors in the space [8]. Non-Hierarchy Method begins by selecting a number of initial cluster values that correspond to the desired number and then the objects are combined into these clusters. The hierarchy method is Average Linkage Hierarchy while the Non-Hierarchy is K-means. Average Linkage is a clustering method that calculates the distance between two clusters by minimizing the average distance between pairs of clusters combined[9]. The K-means algorithm takes the parameters, k, and partitions a set of n objects into k clusters. The K-means algorithm results are randomly selected from the objects which originally were cluster mean or center. For each remaining objects, an object is assigned to the most similar cluster, based on the distance between the object and the cluster mean. After that, it computes the new mean for each cluster[10].

The data used is simulation data which amounts forty data. The data are analyzed to find accuracy, precision, and recall. Accuracy gives equal weight to correctly classified instances regardless of their class. Accuracy is not a sensible evaluation measure in such situations, as it over-values the always-negative classifier[11]. Finally, the best result between Hierarchy Average Linkage and K-means Cluster methods will be used to classify existing classes in the hospital.

II. LITERATURE REVIEW

The following are the methods used in analyzing the type of hospital in Indonesia.

A. Hospital Classification

A uniform hospital classification system is needed to

978-1-7281-3832-9/19/$31.00 ©2019 IEEE

make it easy to know the identity, organization, type of service provided, owner, and capacity of the bed. The information can be evaluated fo a particular group of hospitals.

The classification of hospitals according to various criteria such as ownership, type of service, length of stay, bed capacity, institutional affiliation, and accreditation status. But in this study the classification of hospitals based on accreditation status[12], which are:

a) Class A is general hospitals that have extensive facilities and broad medical and subspecialists medical service capabilities.
b) Class B are public hospitals that have medical service facilities and skills of at least 11 limited specialists and subspecialists.
c) Class C are public hospitals that have the facilities and capabilities of essential specialist medical services.
d) Class D are general hospitals that have the necessary medical facilities and skills.

B. *Hierarchical Average Linkage*

The hierarchical method starts with grouping two or more objects that have the closest similarity. Then the process is forwarded to another object that has a second closeness. And so on until the cluster will form a kind of "tree" hierarchy or clear the level objects from the most similar to the least similar. The hierarchy method isSingle Linkage, Complete Linkage, Average Linkage, and Ward's Linkage. However, this study only using the method of Hierarchical Average Linkage. The following is an explanation of the average linkage method.

Average Linkage calculates the distance between two clusters as the average distance between pairs of clusters combined. The step used in the average linkage is to calculate the distance between objects with the Euclidian distance formula and get the distance of object u with object v denoted by $D = d\{u, v\}$. This method can be formulated in Equation (1).

$$d_{(u,v)w} = \frac{\sum_u \sum_w d_{uw}}{N_{(uv)} N_w} \quad (1)$$

Where:

$d_{(u,v)w}$: distance of object u and object v with object w

C. *K-means Cluster*

Non-hierarchy methods (K-means cluster) begin by selecting a number of initial cluster values that correspond to the desired number and then the objects are combined into these clusters. Grouping using the K-means method is based on its membership function. The membership function is based on the minimum distance between objects and the cluster center (centroid).

The K-means algorithm aims to minimize objective functions which are error squared functions. For example, there are n objects and p variables. The distance between i object and the i-group is calculated using the Euclidean squared distance that can be seen in Equation (2).

$$d_{(i,k)} = \sum_{k=1}^{n} [x_{(i,k)} - \bar{x}_{(l,k)}]^2 \quad (2)$$

with:

i: 1,2, ..., n
k: 1,2, ..., m

D. *Process Modelling*

A business process is activities or related tasks that have a starting and an ending can be defined as inputs and outputs. The business process records what activities are carried out when activities occur, and what conditions, initial or final conditions, during the execution. The business process can be transformed into an image, which is called business process model[13]. The business process model can illustrate the process clearly. Process models have four main purposes:

a. Estimation

Estimation is used to estimating certain characteristics of the population or estimating population values (parameters) using sample values (statistics).

b. Prediction

The goal of prediction in process models is to help take the right business path, outcomes, and make business process risk aware.

c. Calibration

Calibration is used to ensure the measurement results from the process modeling carried are accurate and consistent.

d. Optimization

Optimization is used to achieve ideal results or to make or design a process model to the fullest. The goals are to maximize a process, to minimize the processing timeto make the product.

Business process models can be divided into Petri nets, Business Process Model Notation (BPMN), Event-driven Process Chain (EPC), and Yet Another Workflow Language (YAWL). In this research, the business processes were modeled with BPMN.

The goal of BPMN is to provide a notation which is easily understood by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes[14].

BPMN allows the creation of "end-to-end" business processes, is designed to cover many types of modeling. BPMN can communicate a wide variety of information to a wide variety of audiences. An "end-to-end" contains three basic types of sub-models: process, choreography, and collaboration. By combining the three basic types of sub-models, a detailed representation of business processes can be obtained but recommended for the designer to focus on a certain aspect of processes analysis to avoid creating too complex diagrams, which are difficult to understand.

III. METHOD

The grouping analysis process diagram uses the K-meanscluster method, the complete link hierarchy and the single link hierarchy can be seen in Figure 1 below.

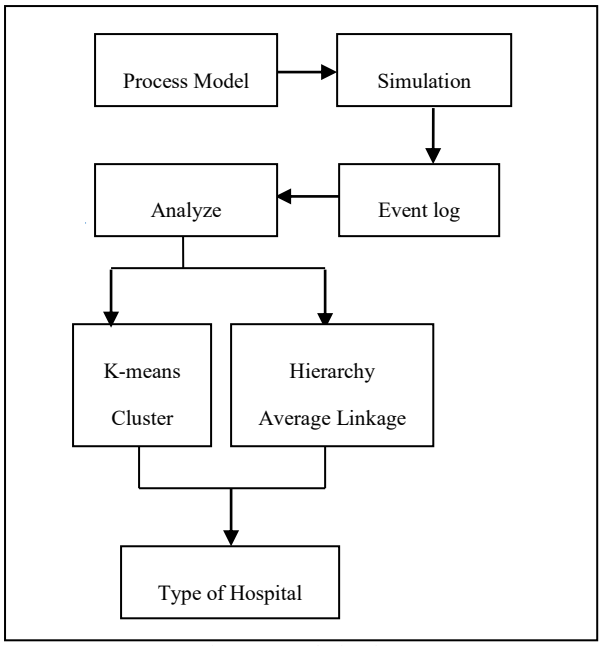

Figure 1. Analysis Diagram

First of all, model process is created. The business process is a set of things that work under coordination in an organizational and technical environment. The business process model in this study uses an application process maker that forms a model of data flow either in or out of patients in the hospital. The model reviews all processes in the hospital system which are then transformed into information. To make the process model into information, simulation is carried out on the process model that has been made. Simulations carried out in 40 hospitals with each of 10 patients both on the road, inpatient care, emergency installation, etc. The results of the simulation are eventlogs which are published into data. An event log contains a set of information conrcerning the processes, such as the performer of tasks, the name of the tasks. The data is taken based on hospital criteria that have been determined and then analyzed using k-mean cluster analysis andhierarchical average linkage. The results of the analysis will be compared the accuracy, precision, and recall of each analysis according to the hospital classification.Methods that have better accuracy, precisionand recall will be used to classify hospital classes.

IV. RESULT AND ANALYSIS

The following are data from 40 hospitals as a result of the model simulations process and criteria that have been set in Table I.

TABLE I. HOSPITAL DATA

N0.	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7
1	A	23	7	6	1	71	12
2	A	22	7	7	1	85	18
3	B	12	4	4	1	74	43
4	C	9	2	2	1	99	24
5	B	15	4	4	1	79	38
6	D	5	1	1	0	88	18
7	B	13	4	4	1	60	43
8	D	7	1	1	0	88	58
9	A	25	9	7	1	90	33
10	A	22	9	10	1	74	20
11	B	13	4	4	1	100	15
12	A	22	10	10	1	63	50
13	B	13	4	4	1	70	25
14	A	22	7	8	1	106	51
15	A	18	7	7	1	71	14
16	B	16	4	4	1	109	18
17	C	10	2	2	1	70	38
18	B	12	4	4	1	119	39
19	A	18	9	6	1	62	22
20	D	7	1	1	0	112	44
21	A	24	5	6	1	91	30
22	B	15	4	4	1	71	53
23	B	17	4	4	1	103	19
24	B	15	4	4	1	107	36
25	A	22	7	8	1	66	46
26	C	10	3	2	1	85	15
27	A	24	10	9	1	108	13
28	C	11	3	2	1	110	17
29	B	12	4	4	1	95	47
30	D	8	1	1	0	66	24
31	C	11	2	2	1	101	24
32	B	17	4	4	1	100	48
33	C	11	2	2	1	94	12
34	B	14	4	4	1	60	42
35	A	21	5	9	1	113	12
36	A	25	9	9	1	94	14
37	D	5	1	1	0	93	16
38	C	9	3	3	1	116	56
39	D	8	1	1	0	89	55
40	B	13	4	4	1	111	34

where:
Z_1 = Hospital Classification
Z_2 = Number of Doctors
Z_3 = Number of Pharmacists in the Outpatient Room
Z_4 = Number of Pharmacists in the Inpatient Room
Z_5 = Number of Pharmacists in the Emergency Room
Z_6 = Waiting Time in the Outpatient Room
Z_7 = Time to Wait for Taking Medication

Table I explains the simulation data of 40 hospitals divided into 4 hospital classifications based on Law No.44 of 2019 and using 7 criteria consisting of Z_1-Z_7, where the criteria have been mentioned in the description table I.

Then, hospital data in Table I will be analyzed using the Hierachical Average Linkage method and K-means Cluster so that later it can be chosen which method gives the right accuracy, prediction and recall.

A. Results of Hierarchical Average Linkage Method and K-means Cluster

Average Linkage Hierarchy is used for two groups or more objects that have the closest similarity minimizing the average distance between pairs of clusters combined. The following is an analysis of the average linkage hierarchy on hospital criteria data.

- Dendrogram

Dendrogram is used to see visually the number of clusters formed. The following is a dendrogram graph from hospital criteria data.

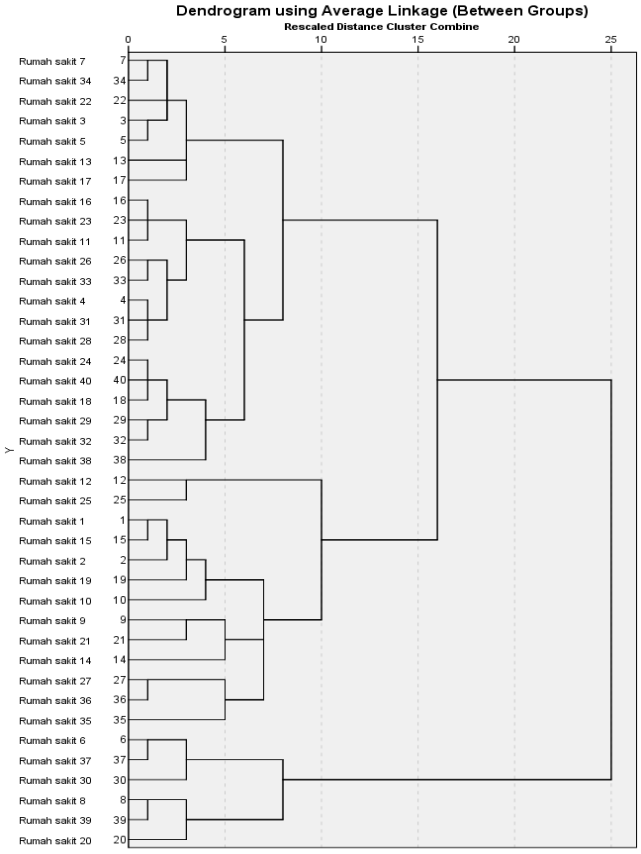

Figure 2. Dendrogram

Based on figure 2, it is known that the number of clusters formed in the hospital criteria data is four clusters. Members of each cluster formed can be seen in Table II.

TABLE II. *Cluster Membership* Metode Hierarchy

Cluster	Hospital List
1	H1, H2, H9,H10, H14, H15, H19, H21, H27, H35, H36
2	H3, H4, H5, H7, H11, H13, H16, H17, H18, H22, H23, H24, H26, H28, H29, H31, H32, H33, H34, H38, H40
3	H6, H8, H20, H30, H37, H39
4	H12 and H25

Based on the table that explained in Table II, it is known that cluster 1 which is categorized as class A type hospital consists of 11 hospitals, cluster 2 with class B type consists of 21 hospitals, cluster 3 class C types consist of 6 hospitals, and cluster 4 type D class consists of 2 hospitals where each member of the cluster has been mentioned in the table II.

K-means Cluster is a grouping method based on the value of the membership function. Next is the K-means Cluster analysis on hospital criteria data that explained in Table III.

TABLE III. *Cluster Membership* Metode *K-means*

Cluster	Hospital List
1	H1, H2, H6, H9, H10, H13, H15, H19, H21, H26, H30
2	H8, H14, H18, H20, H24, H29, H32, H38, H39, H40
3	H3, H5, H7, H12, H17, H22, H25, H34
4	H4, H11, H16, H23, H27, H28, H31, H33, H35, H36, H37

Based on Table III, it can be seen that cluster 1 which is categorized as class A type hospital consists of 11 hospitals, cluster 2 with type B class consists of 10 hospitals, cluster 3 type C class consists of 8 hospitals, and cluster 4 type D class consists of 11 hospitals where each member of the cluster is mentioned in Table III.

B) Acuration, Recall, and Precision

Here is a comparison of the accuracy from the average linkage hierarchy method and K-means Cluster method based on hospital criteria data.

Table IV. Prediction and Realties of Data Using the Hierarchical Average Method

		Reality				Total
		A	B	C	D	
Prediction	A	11	0	0	0	11
	B	0	14	7	0	21
	C	0	0	0	6	6
	D	2	0	0	0	2
Total		13	14	7	6	40

As can be seen in Table IV, true positive from hospital criteria data is 25, a true negative is 15, false positive is 15, and true negative is 105. Therefore, the accuracy of hospital criteria data can be calculated using the formula in Equation 3.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = 0.267 \qquad (3)$$

Based on the calculation in Equation (3), it is known that the accuracy of hospital criteria using the Hierarchical Average Linkage method is 26.7%.

Table V. Recall by using the Hierarchical Average Linkage Method

Recall	Formula	Result
A	TP(A)/total	0.175
B	TP(B)/total	0.125
C	TP(C)/total	0.025
D	TP(D)/total	0.025
Average		0.0875

The table V shows that the recall value of the hospital criteria using the Hierarchical Average Linkage is 8.75%, which means that the criteria are only able to explain the type of hospital in Indonesia at 8.75%.

Table VI. Precision by using the Average Linkage Hierarchy Method

Precision	Formula	Result
A	TP(A)/Tot.predict(A)	0.538
B	TP(B)/Tot.predict(B)	0.357
C	TP(C)/Tot.predict(C)	0.142
D	TP(D)/Tot.predict(D)	0.167
Average		0.301

Table VI shows that the precision value of hospital criteria using the Hierarchical Average Linkage is 30.13%, which means the accuracy of criteria in classifying types of hospitals in Indonesia is 30.13%.

Table VII. Prediction and Reality Using K-means Cluster Method

		Reality				Total
		A	B	C	D	
Prediction	A	7	1	2	3	13
	B	1	5	5	3	14
	C	1	1	1	4	7
	D	2	3	0	1	6
Total		11	10	8	11	40

Table VII shows that true positive from hospital criteria data is 14, the true negative is 26, false positive is 26, and true negative is 94. Therefore, the accuracy of hospital criteria data can be calculated using in Equation 4.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = 0.675 \qquad (4)$$

Based on the calculation in Equation 4, it is known that the accuracy of hospital criteria using the K-means Cluster method is 67.5%.

Table VIII. Recall by using K-means Cluster Method

Recall	Formula	Result
A	TP(A)/total	0.275
B	TP(B)/total	0.35
C	TP(C)/total	0
D	TP(D)/total	0
Average		0.15625

The table VIII shows that the recall value of the hospital criteria using the K-means Cluster is 15.62%, which means the criteria are able to explain the type of hospital in Indonesia by 15.62%.

Table IX. Precision by using K-means Method

Precision	Formula	Result
A	TP(A)/Tot.predict(A)	1
B	TP(B)/Tot.predict(B)	0.6666667
C	TP(C)/Tot.predict(C)	0
D	TP(D)/Tot.predict(D)	0
Average		41.67%

The table IX shows that the precision value of hospital criteria using the K-means Cluster is 41.67%, which means the accuracy of criteria in classifying types of hospitals in Indonesia is 41.67%.

V. CONCLUSION

Hospitals in Indonesia are classified into 4 classes, namely class A, class B, class C, and class D. Each classification of hospital classes has criteria. The criteria used in this study are human resources and time criteria. The hospital criteria were obtained from the process model. After that, the analysis was carried out using two methods, namely Average Linkage Hierarchy and K-means Cluster. From the two methods will be calculated accuracy, precision, and recall. After calculating accuracy, precision, and recall on two methods, it was concluded that the K-means Cluster method provides greater accuracy, precision, and recall.

REFERENCES

[1] A. Cahyapratama and R.Sarno,"Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) MethodsIn Singer Selection Process", 2018.

[2] B. Moseley and Joshua R. Wang,"Approximation Bounds for Hierarchical Clustering:Average Linkage, Bisecting K-means, and Local Search", 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. pp 1-2.

[3] D. Nughraheny, "Distance Value Method for Similarity or Characteristics of an Image (Cumulonimbus Cloud Detection Case Using Principal Component Analysis)", 2015.

[4] Endra Budi Setyawan. Febri, and Supriyanto, Stefanus, 2019. Hospital Management. Sidoarjo: Zifatama Jawara.

[5] Hermawan and R. Sarno, "A More Efficient Deterministic Algorithm In Process Model Discovery", Vol. 14, No. 3, pp 971-995, 2017.

[6] H. H. Ali and L. E. Kadhum, "Licensed Under Creative Commons Attribution CC BY K- Means Clustering Algorithm Applications in Data Mining and Pattern Recognition", Vol. 6, 2017.

[7] https://pdfs.semanticscholar.org/4f4a/857784bfd41c287813264c0099db79e4396a.pdf?_ga=2.116206918.1780103430.1557989915-1307034755.1556773688 accessed on Thursday at 14.30 WIB on May 16, 2109.

[8] K. S. Sungkono and R. Sarno,"Constructing Control-Flow Patterns Containing Invisible Task and Non-Free Choice Based On Declarative Model", International Journal of Innovative Computing, Information, and Control, Vol. 14, pp. 1285-1299, 2018.

[9] P. A. Flach dan M. Kull, "Precision-Recall-Gain Curves: PR Analysis Done Right.

[10] R. Aljumily, "Hierarchical and Non-Hierarchical Linear and Non-Linear Clustering Methods to "Shakespeare Authorship Question", 2015.

[11] R. Johanes, R. Sarno, and K. R. Sungkono,"Graph-Based Algorithms for Discoveing a Process Model Containing Invisible Tasks", 2018.

[12] R. Sarno and K. R. Sungkono,"Hidden Markov Model for Process Mining of Parallel Business Processes",International Review on Computers and Software, Vol.11, No. 4, pp. 290-300, 2016.

[13] Rencher, A.C., 2002. *Methods of Multivariate Analysis. Second Edition,* John Wiley and Sons. Inc., New York.

[14] Siregar, Charles J.P. 2004. Hospital Pharmacy(Theory and Application). Jakarta: Publishing the EGC Medical Book, pp 13-15.