Measurement of health service performance through machine learning using clustering techniques

R D Dana*, A R Dikananda, D Sudrajat, A Wanto and F Fasya
STMIK IKMI Cirebon, Indonesia

*radith_danar@yahoo.com

Abstract. The health center is the first level public health service center that is run by the Indonesian government. Organizing quality health center services of health center is the hope and satisfaction of every patient. The dimensions of patient satisfaction vary greatly and the scope is very wide. Patient satisfaction can be used as an indicator of quality and service performance of health center. The purpose of this study was to measure the performance of health center services in order to determine the level of patient satisfaction and grouping the Patient Satisfaction Index. With the knowledge of the dimensions of performance and quality of health services provided by the health center, it will facilitate the government in carrying out the function of guidance and control of the health center. This study uses a Machine Learning technology approach with clustering techniques, by grouping the Patient Satisfaction Index with K-Means (Hard Clustering) and Fuzzy C-Mean (Soft Clustering) methods. Based on the subsets produced, clustering techniques can be divided into 2 methods, namely hard clustering techniques and soft clustering techniques. The K-Mean method is widely used in clustering techniques and has advantages in computational speed and relatively easy process stages. While the Fuzzy C-Means method has advantages in terms of flexibility in determining clusters so that there is little possibility of converging failure. The experimental results of the Patient Satisfaction Index show that the K-Means method provides better performance with a value of 96% compared to the Fuzzy C-Means method with a value of 76%.

1. Introduction
Indonesian Health Ministry Decree No. 857/Menkes/SK/ IX/2009 states that primary health care is the first level of health care facilities. As a first-level health care centers, health centers are obligated to conduct health care quality, fair and equitable. The level of patient satisfaction is a key indicator of quality health services. Carrying out quality health services and in accordance with patient expectations is not easy, this is because the dimensions of patient satisfaction vary widely and the scope is very wide [1-4]. Health service data that are usually presented in very large quantities requires a technique or approach that can facilitate the data analysis process. Machine Learning can be used to process health service data and be able to find hidden patterns from data. By using the patient satisfaction data clustering method will be grouped into a group or group based on patterns of data similarity [5]. Measuring the level of patient satisfaction is done by conducting a survey using a questionnaire to get the patient's perception and experiences of the parameters or elements of health services [6-8].
2. Research method

The sampling process is performed using the Simple Random Sampling method. The amount of data sample size is determined by using the Slovin formula with a significance level of $\alpha$ of 10%. The data used in this study was obtained from the Cirebon City Health Office in the form of a patient satisfaction survey on the services of 21 health centers.

Questionnaires were developed based on 14 indicators of public services that have been determined by ministerial regulations namely service procedures, service requirements, clarity of service personnel, discipline of service officers, responsibility of service personnel, service personnel ability, speed of service, fairness to get service, courtesy and friendliness of officers, fairness service fees, certainty of service costs, certainty of service schedule, certainty of service schedule, environmental comfort and service security [9]. Total respondents in this survey are as many as 3,000 patients who will judge the 14 elements of health center services, namely:

- **Service procedures**: namely the ease of service stages provided to the public in terms of the simplicity of the service flow
- **Service Requirements**: namely the technical and administrative requirements needed to obtain services in accordance with the type of service.
- **Clarity of Service Officers**: namely the presence and certainty of officers who provide services (name, position and authority and responsibility).
- **Discipline Service Officers**: namely the seriousness of officers in providing services mainly to the consistency of working time according to regulations.
- **Responsibilities of Service Officers**: namely clarity of authority and responsibility of officers in the administration and completion of services.
- **Ability of Service Officers**: namely the level of expertise and skills that officers have in providing / completing services to the community.
- **Service Speed**: namely the target service time can be completed within the time specified by the service provider unit.
- **Justice Gets Service**: namely the implementation of services by not distinguishing the class / status of the community served.
- **Courtesy and hospitality Officer**: the attitude and behavior of officers in providing services to the community as polite and friendly as well as mutual respect.
- **Fairness of Service Costs**: namely the affordability of the community to the amount of fees set by the service unit.
- **Certainty of Service Fees**: that is the suitability between the fees paid and the fees that have been set.
- **Assurance Services Schedule**: namely the implementation of service time, in accordance with the conditions set.
- **Environmental comfort**: namely the condition of service facilities and infrastructure that are clean, neat, and orderly so as to provide a sense of comfort to service recipients.
- **Service Security**: namely the guaranteed level of environmental security of the service provider unit or the facilities used.

The answer form of each statement contained in the questionnaire in general will reflect the level of service quality, that is from the very good to the bad. For the bad category given the perception value 1. Less good given the value of perception 2, both given a value of 3, very well given the value of perception 4. Data obtained through surveys will be processed and grouped based on data that has similar patterns. This grouping effort can be done using the Clustering method [10].

In this study will be grouped Patient Satisfaction Index using clustering techniques. Clusters are generally a form of subsets of a set of data. Based on the subsets produced, clustering techniques can be divided into 2 methods, namely hard clustering techniques and soft clustering techniques. To provide an overview of the performance generated from each clustering technique, the K-Means (Hard Clustering) and Fuzzy C-Mean (Soft Clustering) methods will be used. The K-Mean method is widely used in the field of research...
used in clustering techniques, K-Mean has advantages in computational speed and relatively easy process stages. Fuzzy C-Means method has advantages in terms of flexibility in determining clusters, so there is little possibility of converge failure [11-12].

2.1. Determine the number of cluster
The process of assessing quality of health services at the health center was begun with determining the number of clusters to be formed [13]. The number of clusters is determined based on the grouping of patient perceptions which can be described in table 1.

| Perception Value | Interval | Value of Conversion Interval Value | Service Quality | Information |
|------------------|----------|------------------------------------|-----------------|-------------|
| Patient Satisfaction Index | 1,76 – 2.50 | 43,76 – 62.50 | C | Poor |
| 2 | 2,51 – 3.25 | 62,51 – 81.25 | B | Good |
| 3 | 3,26 – 4.00 | 281,26 – 100.00 | A | Very Good |

2.2. Experiments and testing algorithms
K-Means is an algorithm used to cluster or to group data or data sets based on the attributes that the data has to a number of K groups or clusters. K is a positive integer. Clustering process is done by finding the closest square of the object with the centroid cluster. For example, there are 4 data samples that will be carried out in the training process, each data has 2 attributes, each attribute represents the coordinates of the object [14].

The following are the steps in Clustering using the K-Means algorithm
- Determine the number of clusters based on the possibility of patient perceptions
- Select the object that will be used as the centroid point of each cluster
- Group object objections to the nearest centroid point based on Euclidean Distance
- Recalculate all centroid point values again
- Repeat step 2 until the centroid point value does not change

The algorithm Fuzzy C-Means Clustering (FCM) is a clustering algorithm of data that each data to be a member of a cluster with a degree defined by membership level based on the minimization of the objective function is formulated as follows
- Input data that will be in the cluster, in the form of a matrix size x
- Determine number of custer, maximum repetition and smallest desired error
- Calculates the cluster center point
- Calculate matrix changes
- Check the stop computing algorithm condition. The stop condition will occur if the minimum error value expected has been reached or the maximum number of repetitions of the calculation process has been exceeded
- If the stop condition is not fulfilled then return to step 3

3. Results and discussion
Measurement of Patient Satisfaction Index as an indicator of the quality of health center services is calculated by using the "value - average weighted". From the calculation process of the entire population of 3,000 data, a Patient Satisfaction Index was obtained with a range of values from 49.7 to 99.4. If referring to the table 1 (Value of Patient Perception), it can be seen that there is no Patient Satisfaction Index which is a member of the category of "Not Good", so in this study the cluster "Not good" will not be used as an object of research or in other words omitted. By eliminating the "Bad" cluster, there are only 3 clusters to be studied, namely "Poor", "Good" and "Very Good"
The process of calculating "weighted average values" results in the following patient satisfaction index values:

- 10% of the respondents rated the quality of Puskesmas services with the perception of "Poor" with an average score of 56,978
- 65% of the number of respondents rated the quality of Puskesmas services with the perception of "Good" with an average value of 74,687
- 25% of the respondents rated the quality of Puskesmas services with the perception of "Very Good" with an average value of 89,602

The process of calculating the K-Means method is done by finding the minimum distance between objects with each centroid point of each cluster. The process of calculating the K-Means method produces clusters as follows:

- 14% of the total respondents rated the quality of Puskesmas services with a perspective that was "Poor" with the average distance of objects with cluster 1 was 3.5663
- 61% of the respondents rated the quality of Puskesmas services with a perspective of "Good" with an average distance of object points with cluster 2 is 2.6724
- 25% of the total respondents rated the quality of Puskesmas services with a perspective of "Very Good" with an average distance of object points with cluster 3 is 2.9312

![Figure 1. Results the clustering process uses K-means.](image)

The process of calculating the Fuzzy C-Means method is done by finding the largest membership degree value of each object against each cluster. The process of calculating the Fuzzy C-Means method produces clusters as follows:

- 34% of the respondents rated the quality of Puskesmas services with the perception of "Poor" with an average grade of 0.525
- 41% of the respondents rated the quality of Puskesmas services with the perception of "Good" with an average membership value of 0.583
- 25% of the number of respondents rated the quality of Puskesmas services with the perception of "Very Good" with an average grade of 0.589
Performance measurement will be carried out by analyzing the results of clustering from the K-Means and Fuzzy C-Means methods. The cluster members of each method will be compared with the target cluster that has been obtained by calculating the method of "weighted average".

### Table 2. K-means and fuzzy C-means method performance.

|           | K-Means | Fuzzy C-Means |
|-----------|---------|---------------|
| Accurate  | 96      | 76            |
| Not Accurate | 4     | 24            |
| Performance | 96%  | 76%           |

The table above shows the level of performance of the K-Means method in the case of grouping of Patient Satisfaction Index with 100 data samples showing 96%, when compared with the level of accuracy based on calculations using the Fuzzy C-Means method with cases and the same number of data samples showing 76%, there is an increase in the accuracy rate of 20%.

From the results of experiments conducted, it can be seen that the performance of hard clustering method with K-Means offers better results compared to soft clustering method with Fuzzy C-Means. In this study, the method used in the cluster analysis of the health center patient satisfaction perception value is K-Means and Fuzzy C-Means. From the results of experiments that have been carried out in the form of testing and performance measurement of these methods, it was found that the hard clustering method with K-Means is more optimal than the soft clustering method with Fuzzy C-Means. In the case of cluster analysis of patient satisfaction perception value health centers, with 90 of sample data showed the K-Means method accuracy by 96% and and Fuzzy C-Means 76%.

### 4. Conclusion

So, in the case of cluster analysis of patient satisfaction perception value health centers, the use of the hard method with K-Means clustering offers better performance than the soft clustering method with Fuzzy C-Means. This is because K-Means Clustering is more suitable for grouping data where the number of clusters can be determined at the beginning of computing, namely health services with the perception of "Less Good", "Good" and "Very Good", where data objects will be grouped exclusively as on the issue of measuring patient satisfaction perceptions where a patient at the same time is unlikely to have a different perception of satisfaction. On the other hand, Fuzzy C-Means Clustering will provide
better performance if it is applied to grouping complex data containing high uncertainties where data is possible to become part of more than one cluster [15-17].

Acknowledgement
Thanks to the STT Multimedia Education Foundation Cirebon Indonesia who have contributed in this research.

References
[1] Graham B 2016 Defining and Measuring Patient Satisfaction J. Hand Surg. Am. 41 9 pp 929-931
[2] Kementrian Kesehatan Republik Indonesia 2009 Keputusan menteri kesehatan republik Indonesia
[3] Sadegh S S, Saadat P K, Sepehri M M and Assadi V 2018 International Journal of Medical Informatics A framework for m-health service development and success evaluation Int. J. Med. Inform. 112 pp 123–130
[4] Ahenkan A and Aduo-Adjei K 2017 Predictors of Patient Satisfaction With Quality of Healthcare in University Hospitals in Ghana Hosp. Pract. Res. 2 1 pp 9–14
[5] Ogbuabor G and Ugwoke F N 2018 Clustering Algorithm For A Healthcare Dataset Using Silhouette Score Value Int. J. Comput. Sci. Inf. Technol. 10 2 pp 27–37
[6] Khoie M R, Tabrizi T S and Khorasani E S 2017 A Hospital Recommendation System Based on Patient Satisfaction Survey Appl. Sci.
[7] Yadav P and Mangwana S 2014 Analysis of patient satisfaction survey in multi-specialty accredited tertiary care hospital in northern India International Journal of Basic and Applied Medical Science 4 3 pp 33–39
[8] Wolf J A, Victoria N and Dianne M 2014 Defining Patient Experience Patient Exp. J. 1 1
[9] Kementerian Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi Republik Indonesia 2017 Peraturan Menteri Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi Republik Indonesia Nomor 14 Tahun 2017
[10] Schüller D and Pekárek J 2018 Customer Satisfaction Measurement – Clustering Approach 66 2 pp 561–569
[11] Madhukumar S and Santhiyakumari N 2015 Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain Egypt. J. Radiol. Nucl. Med. 46 2 pp 475–479
[12] Velmurugan T 2014 Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data Appl. Soft Comput. J. 19 pp 134–146
[13] Elbattah M and Molloy O 2016 Clustering-Aided Approach for Predicting Patient Outcomes with Application to Elderly Healthcare in Ireland Fractures in Ireland AAAI-17-Jt. Work. onHealth Intell. pp 533–541
[14] Cebeci Z and Yildiz F 2015 Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures J. Agric. Informatics 6 3 pp 13–23
[15] Dibya J B and Gupta D A K 2014 A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm Int. J. Comput. Trends Technol. 10 2 pp 108–113
[16] Ghosh S and Dubey S K 2013 Comparative Analysis of K-Means and Fuzzy C- Means Algorithms Int. J. Adv. Comput. Sci. Appl. 4 4 pp 35–39
[17] Bücker T 2016 Customer Clustering in the Insurance Sector by Means of Unsupervised Machine Learning