CLASSIFICATION OF HOSPITAL SERVICE PERFORMANCE QUARTERLY IN CENTRAL JAVA IN 2020 USING THE KNN METHOD

N G Ramadhanty [1], I A Risdiari [2], M I Ramdhani [3], E S Nugraha [4]
School of Business, Study Program of Actuarial Science, President University, Cikarang, Bekasi 17550, Indonesia
[1]*novia.ramadhanty@student.president*
[2]*inas.risdiari@student.president.ac.id*
[3]*muhammad.ramdhani@student.president.ac.id*
[4]*edwin.nugraha@president.ac.id*

ABSTRACT:
One of the most important health service institutions for citizen in every country is hospital. In order to evaluate the performance of hospital services, an indicator is needed to identify whether a hospital provides services well or not. The Ministry of Health determined standard for reviewing hospital performance as an indicator so health services evaluate themselves and can be improved in the future. There are several types of indicators given by Ministry of Health. Written in Pemenkes, there are 21 types of performance and 107 indicators. For this research, this research will use hospital inpatient indicator with 5 indicators. To calculate performance indicators, it is necessary to collect data and monitor in a hospital. By using the K-Nearest-Neighbors (k-NN) method, we can identify how good the services provided by a certain hospital. This method is a method for classification with dependence on the choice of "good value" for k. The output of this data mining can be used to identify whether the hospital fulfill the service standards set by the Ministry of Health in Indonesia or not. Based on the calculation in R, this research shows the output is the same as the original data. It means this method can be applied effectively in determining the quality of hospital services.

Keywords: *hospital services, K-Nearest-Neighbours, classification*

1. Introduction

In an effort to develop facilities and performance, hospitals need to continue to improve themselves and provide better services. By doing so, an indicator of hospital services needed to correct the performance of services provided by a hospital in order to survive in the development of health services (Heryant, 2013). By determining the indicators, we can see hospital's utilization, quality, and efficiency of public health services. The following indicators are sourced from the daily inpatient census (Heryant, 2013). The indicators used in this paper are BOR, LOS, TOI, NDR, and GDR.

BOR (Bed Occupancy Ratio) is an indicator that provides an overview of the utilization rate of hospital beds with a standard value of 60-85%. The percentage of bed usage is determined by Ministry of Health 2005. LOS (Length of Stay) is an indicator that provide an overview of service quality, when applied to certain diagnoses it can be used as a matter that needs further observation (www.hakayuci.com, n.d.). The ideal LOS according to Ministry of Health is between 6-9 days. TOI (Turn Over Interval) is an indicator which provides an efficient use of the bed level. According to the standard Pemenkes, empty beds are not filled in the range of 1-3 days. According to the Ministry of Health in 2011, NDR (Net Death Rate) is an indicator provides an overview of the mortality 48 hours after being treated for every 1000 patients out. The NDR value which is considered to be tolerable is less than 25 per 1000 (Ministry of Health 2011). GDR (Gross Death Rate) is the general death rate for every 1000 patients who are out from hospital. According to the Ministry of Health in 2011, The normal GDR (Gross Death Rate) standard is not more than 45 per 1000 patients who discharged.

K-Nearest Neighbor (K-NN) method that performs classification based on the proximity of the location (distance) of a data to other data (Fitri, 2016). This algorithm is an algorithm that relies on testing data and training data that tests the shortest distance between the two data or calculates the smallest distance value. This method aims to classify new objects based on the sample data used in the previous classification. In this method, the training data is the data that becomes the standard or the basis for grouping new data (test data). The advantage of the KNN algorithm is that it can classify data accurately and effectively for large

data. Disadvantages of the KNN algorithm require the value of the number of nearest neighbors (k) to test the data, while for a good k, certain calculations must be carried out.
Various studies already prove that KNN is a helpful method to make human work easier, especially on analyzing data. Fitri (2016) show that K-NN can classify diabetes mellitus type from patient data in Puri Husada Tembilahan hospital. The result has a significant output as it has 88% accuracy. The K-NN method also can be used in the agriculture field. K-NN can classify guava fruit quality based on colour and texture. The conclusion of this study is that the proposed KNN classification method can classify guava with an accuracy of 91.25% at the best neighbor value K = 3 with the extracted features for guava classification are r, g, b, defect area, energy, homogeneity, and contrast (Prahudaya & Harjoko, 2017). In the education field, K-NN can help to determine students who get scholarships in STMIK Sinar Nusantara. The result also shows a quite good output with accuracy 90,90% which sample data of this study can be used as a reference for determining scholarship acceptance, so by providing lots of good data, it can improve system performance with this method (Risman, Nugroho, & WU, 2015).
In this paper, we will apply an example of classifying the K-NN method using data on the performance of hospital services in Central Java. Hospital is a health care institution that organizes personal health services in plenary which organizes inpatient, outpatient, and emergency services. In providing health services, it is expected that hospitals can provide quality services (Hosizah & Maryati, 2018). The classification can help as material for evaluating the hospital's performance in providing services so that it can be better.

2. Literature Reviews

2.1 Classification System and K-NN

Data processing is one of the important things in information technology. Various kinds of data that are processed with effective and efficient methods will produce accurate information. In line with technological developments, various data processing methods have also been developed (Wibawa, 2018). This is to meet the need for the method used in accordance with the type of data to be processed. One method in data processing is classification. Classification is one way of grouping data according to the characteristics or characteristics of the data. Various kinds of common classification methods include: Naive Bayes, Support Vector Machine, KNN, Decision Tree, and Fuzzy.
KNN is a type of classification where the function is only approximated locally and all computation is deferred until evaluation of the function (Supplement to Loganathan & Mahindrakar, 2021). Because the algorithm is dependent on the distance to the classification, if the feature represents a different physical unit or has a very different scale, the data normalization training can dramatically improve its accuracy. The best value of k for this algorithm depends on the data. To determine the value of k, we can use parameter optimization such as cross validation and normalizing data (www.informatikalogi.com, 2017). The accuracy of the algorithm is highly influenced by the existence of features. If the feature irrelevant, it will affect the final result of classification. This algorithm has a strong consistency.
The classification process is carried out by finding the nearest point from the new nearest neighbours. One of the techniques to find the nearest neighbour is using the Euclidean distance formula. Euclidean Distance is often used to calculate distance which to test the approximately nearest distance between two objects. If the result of the value of the formula below is large, the further the level of similarity between the two objects will be. In the other hand, if the result of the value is smaller, the closer the level of similarity between the objects will be. The objects in question are training data and testing data. The formula above is a common way to calculate Euclidean Distance which is the formula is given by

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

Where, x_i is sample data, y_i is testing data, i is variable data, $d(x,y)$ is dissimilarity/distance, and n is data dimension.

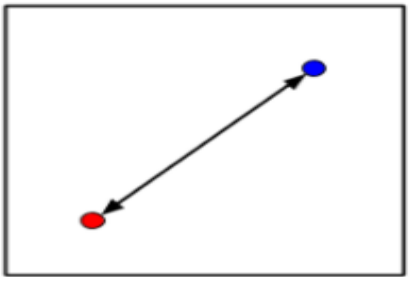

Figure 1 Euclidean distance illustration

From figure 1, The illustration shows two objects and the line between can be defined as the smallest distance between the two objects. This method of calculation is one of the easiest ways to define the smallest distance of two objects. The calculation of Euclidean distance use Pythagorean Theorem. This formula is known to compute different types of objects such us point to line or point to point.

3. Method

The work scheme is the stages in carrying out activities to produce a conclusion. In analyzing hospital performance data in Central Java Province using the work scheme below:

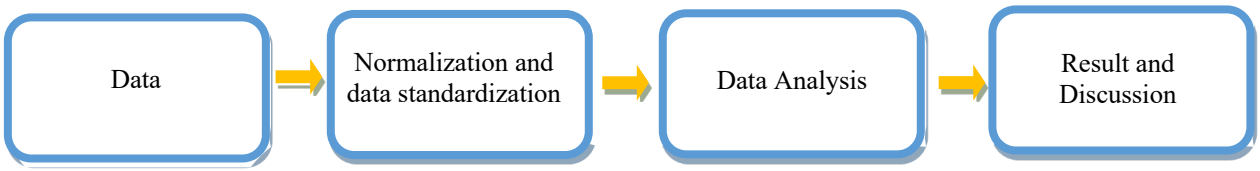

Figure 2 KNN work scheme

KNN work scheme on figure 2 will be our groundwork in this research. First, we have to search the data related to research theme. The data about hospital performance is taken from official government dataset website in Central Java. After we find the data, there will be a different range of value in the dataset. Therefore, we will do a normalization and standardization data to scale the value of each variable. After the value change into same range, we analyse the data using KNN method with Euclidean distance formula. After processing the data, we discuss about the result and how effective KNN is.

For using KNN method, we need KNN algorithm to map the steps that we will be doing in this research. Based on the figure 3, we need to input data as a start. Then we determine the training data and testing data. Because we use KNN, we need to decide the number of k (neighbour). After deciding the k, we calculate the distance between training data and test data that we determine before. Next, we select the nearest neighbour sample (k) and we classify the test data with the training data. Last, we will analyse the test data such as the accuracy with the training data, rate of classification, etc. Finally, we can classify the test data in the group based on the train data.

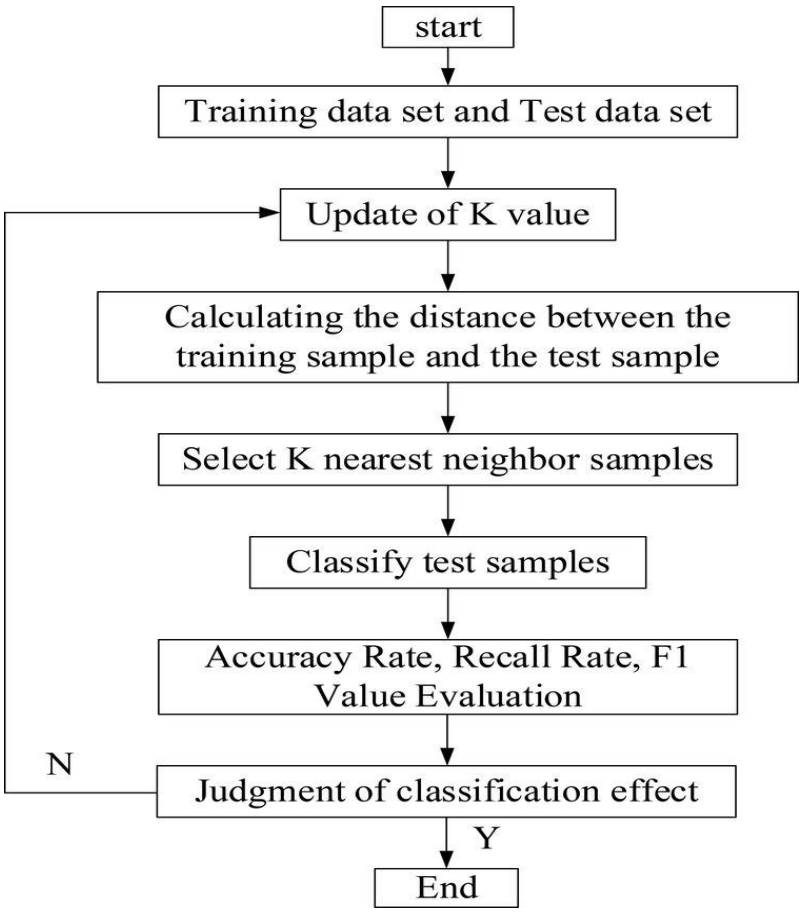

Figure 3 KNN Algorithm

3.1 Data

Data is the most important part to analyse the required information. The data we use is data on the performance of hospital services per quarter in the Central Java area. Data obtained from dataset Central Java government website (Indikator Kinerja Pelayanan Rumah Sakit Per Triwulan 3 2020, 2021) accessed on June 15, 2021. The data that we're about to use consist of two kinds of major variables, input variable and target variable. Each of them has their own variables within. Input variable constructed of 5 more variables which is BOR, LOS, TOI, NDR, and GDR. Target variable also known as the output or target unit to be produced from the process of analysing the whole data. The variable is inclusive to Variable 1 another term for good and Variable 2, the opposite of the other variable, not good.

3.2 Normalizing Data

Normalization is a logical design technique in a database that groups the attributes of various entities in a relationship so as to form a good relationship structure. The goal is to eliminate and reduce data redundancy and the second goal is to ensure data dependencies. We will normalize the data in Rstudio. We create a function name in R called normalize using normalization equation in the K-NN process which is used to change the variables into the same scale. In this case, we first separate the data into testing data and training data randomly with the "CaTools" package and use sample.split code with a ratio of 20% of training data including TRUE arguments and 80% testing data including FALSE arguments.

Table 1 Dataset of Hospital Performance Per Quarter in Central Java 2020

Hospital	Variable					Performance
	BOR	LOS	TOI	NDR	GDR	
RSUD Tugurejo Semarang	54.23	4.49	3.83	27.76	46.74	2
RSUD Prof. Dr.Margono Soekarjo Purwokerto	94.36	5.13	0.37	39.65	50.54	2
RSUD Dr.Moewardi Surakarta	61.19	3.86	3.56	77.44	96.47	2
RSUD Kelet Jepara	58.3	4	3	14	28	1
RSUP Dr. Kariadi Semarang	84.84	7.14	1.24	61.7	80.9	1
RSUP Prof.Dr.Soeradji T. Klaten	65.66	4.62	2.3	28.4	57.47	2
RSJD Dr. Amino Gondo Hutomo Semarang	62.97	11.99	5.15	2.28	6.64	1
RSJD Soedjarwadi Klaten	60.32	12.33	6.75	12.91	30.09	1
RSJD Surakarta	69.28	28	11	3.73	3.73	1
RSJP Prof. Dr.Soerojo Magelang	49.8	9.7	9.4	25.8	12.2	1
RSO Prof.Dr.Soeharso Sukoharjo	55.05	4.62	3.89	6.02	5.33	1
RSTP Dr.Ario Wirawan Salatiga	74.51	5	2	55	34	1
RSUD RAA. Soewondo Pati	56.79	3.33	2.55	26.35	61.4	2
RSUD Kraton Pekalongan	56.49	3.85	2.95	30.05	62.6	2
RSUD Kardinah Kota Tegal	59	3.94	3.4	28.01	51.61	2
RSUD Tidar Magelang	76.49	3.86	1.16	7.52	18.59	1
RSUD DR. Moewardi Surakarta	61.19	3.86	3.56	77.44	96.47	2
RSUD DR. Margono Soekarjo Purwokerto	94.36	5.13	0.37	39.65	50.54	2
RSUD Tugerejo Semarang	54.23	4.49	3.83	27.76	46.74	2

Formula of smallest distance is given by:

$$\sqrt{(a_k - a)^2 + (b_k - b)^2} \quad (2)$$

Where, a_k is training data, a is testing data, b_k is 2nd variable data, and b is 2nd variable testing data.

In the Table 2, Data training and data testing is chosen randomly using Rstudio. The result after running the program using sample split code with ratio 2 : 8 are in table 2. Hospitals that came out into data training are Tugurejo Semarang, Dr. Kariadi Semarang, Surakarta, and Prof. Dr. Soerojo Magelang. The rest of hospitals are becoming a data testing.

Table 2 Normalization and standardization in R

Hospital	Variable				
	BOR	**LOS**	**TOI**	**NDR**	**GDR**
Data *Testing*					
RSUD Prof. Dr.Margono Soekarjo Purwokerto	2.15112	-0.28880	-1.19726	0.37228	0.22232
RSUD Dr.Moewardi Surakarta	-0.34198	-0.50747	-0.05051	2.02354	1.83642
RSUD Kelet Jepara	-0.55920	-0.48337	-0.25182	-0.74850	-0.56979
RSUP Prof.Dr.Soeradji T. Klaten	-0.00601	-0.37661	-0.50346	-0.11928	0.46586
RSJD Dr. Amino Gondo Hutomo Semarang	-0.20819	0.89234	0.52105	-1.26062	-1.32044
RSJD Soedjarwadi Klaten	-0.40737	0.95088	1.09622	-0.79613	-0.49634
RSO Prof.Dr.Soeharso Sukoharjo	-0.80347	-0.37661	0.06811	-1.09719	-1.36647
RSTP Dr.Ario Wirawan Salatiga	0.65916	-0.31119	-0.61130	1.04301	-0.35893
RSUD RAA. Soewondo Pati	-0.67269	-0.59873	-0.41359	-0.20886	0.60397
RSUD Kraton Pekalongan	-0.69524	-0.50919	-0.26980	-0.04719	0.64614
RSUD Kardinah Kota Tegal	-0.50658	-0.49370	-0.10803	-0.13633	0.25992
RSUD Tidar Magelang	0.80798	-0.50747	-0.91327	-1.03165	-0.90048
RSUD DR. Moewardi Surakarta	-0.34198	-0.50747	-0.05051	2.02354	1.83642
RSUD DR. Margono Soekarjo Purwokerto	2.15112	-0.28880	-1.19726	0.37228	0.22232
RSUD Tugerejo Semarang	-0.86511	-0.39900	0.04654	-0.14725	0.08878
Data *Training*					
RSUD Tugurejo Semarang	-0.86511	-0.39900	-0.04674	-0.14725	-0.08878
RSUP Dr. Kariadi Semarang	1.43558	0.05727	-0.88451	1.33577	1.28925
RSJD Surakarta	0.26607	3.64893	2.62402	-1.19726	-1.12504
RSJP Prof. Dr.Soerojo Magelang	-1.19807	0.49805	2.04885	-0.23289	-1.12504

3.3 Data Analysis

By using the "CLASS" package and with the KNN code with the normalized training data added with a drop of the FALSE argument which is the opposite of the previous data argument which is a vector instead of a data frame, as well as normalized test data, then the class uses a training set before column normalization performance. By using k=1. To ensure suitability, you can use the table code, if it appears as shown below, it means that it is in accordance with the predictions of class 1 and 2. From Table 4, we know that the prediction and the original data are well execute by Rstudio. It means this method effectively work.

Table 3 Prediction Result from Rstudio

Hospital	Class Type
RSUD Tugurejo Semarang	2
RSUD Prof. Dr.Margono Soekarjo Purwokerto	2
RSUD Dr.Moewardi Surakarta	2
RSUD Kelet Jepara	1
RSUP Dr. Kariadi Semarang	1
RSUP Prof.Dr.Soeradji T. Klaten	2
RSJD Dr. Amino Gondo Hutomo Semarang	1
RSJD Soedjarwadi Klaten	1
RSJD Surakarta	1
RSJP Prof. Dr.Soerojo Magelang	1
RSO Prof.Dr.Soeharso Sukoharjo	1
RSTP Dr.Ario Wirawan Salatiga	1
RSUD RAA. Soewondo Pati	2
RSUD Kraton Pekalongan	2
RSUD Kardinah Kota Tegal	2
RSUD Tidar Magelang	1
RSUD DR. Moewardi Surakarta	2
RSUD DR. Margono Soekarjo Purwokerto	2
RSUD Tugerejo Semarang	2

Table 4 K-NN Prediction

Output	Comparison	
	R Prediction from K-NN	Original Data
Gol 1	9	9
Gol 2	10	10

3.4 Result and Discussion

After calculating the KNN predictions in excel and R studio, the results and comparisons are shown in Table 5. This classification validation using R studio with weighting K = 1 for the case of hospital performance can be concluded from the similarity of the K-NN calculation results with the original data. In this case, validation of the K-NN classification with R studio obtained data that has the same results from the K-NN calculation with the original data on hospital performance in Central Java Province. From the results obtained, the percentage of validation in this case data with a weight of K = 1 is 99% because the predictions we make may be different when done with other K-NN coding so it may not be 100% correct.

K-Nearest Neighbor (K-NN) includes a supervised learning algorithm which is a supervised learning where if the expected output is known beforehand. Although there are differences in the data collection of patients/patients between the predicted results of R studio and the original data, the K-NN method using R studio in the case of re-evaluating hospital performance data in Central Java Province has a considerable validation of 99 %.

By using CaTools and sample.split in R, They give a good recommendation to choose training data to be compared with the testing data as part of the development of technology in coding and data science world.

Also, ratio is playing a good influence in determining the amount of data testing and data training. Combining both codes, it will make a good result in KNN prediction.

Table 5 Result from R and Original Data

Hospital	Hospital Performance	Prediction R
RSUD Tugurejo Semarang	2	2
RSUD Prof. Dr.Margono Soekarjo Purwokerto	2	2
RSUD Dr.Moewardi Surakarta	2	2
RSUD Kelet Jepara	1	1
RSUP Dr. Kariadi Semarang	1	1
RSUP Prof.Dr.Soeradji T. Klaten	2	2
RSJD Dr. Amino Gondo Hutomo Semarang	1	1
RSJD Soedjarwadi Klaten	1	1
RSJD Surakarta	1	1
RSJP Prof. Dr.Soerojo Magelang	1	1
RSO Prof.Dr.Soeharso Sukoharjo	1	1
RSTP Dr.Ario Wirawan Salatiga	1	1
RSUD RAA. Soewondo Pati	2	2
RSUD Kraton Pekalongan	2	2
RSUD Kardinah Kota Tegal	2	2
RSUD Tidar Magelang	1	1
RSUD DR. Moewardi Surakarta	2	2
RSUD DR. Margono Soekarjo Purwokerto	2	2
RSUD Tugerejo Semarang	2	2

From table 5 above, The comparison of hospital services type is accurate. The Hospital performance in both original and predicted data shows 10 hospitals were giving a less services (not good) and the opposite 9 hospitals give good services. Since the prediction give a good result and specific output. KNN method is one of a good method to calculate and predict a big data of hospital services.

4. Conclusion

In K-NN there is a distance classification phase. This is generally done by looking for the nearest c point from the new c (nearest neighbour). The closest neighbour search technique is commonly done using the Euclidean distance formula because the calculation using the Euclidean formula on hospital performance indicators in Central Java Province, there are two classes which are types of hospital performance so that the system can recognize the class/type of hospital performance whether it is good (1) or not good (2) using the K-Nearest Neighbour (K-NN) method. However, in this case we have different results between the K-NN classifier using excel and R studio. Although there are differences in house data collection between the predicted results of K-NN both in excel and in R studio, the K-NN method using R studio in the case of re-evaluating hospital performance indicator data has a fairly large validation of.

K-NN itself there are also important parts that we need to know to facilitate the classification of a data, namely training data or can be interpreted as the main data for classification. The source of this training data is also usually not arbitrary. So, we really need to know where this training data comes from, such as those from hospitals, posyandu, puskesmas or so on. The next important part is data testing or commonly called test data. This data serves as data that is tested on K-NN so that the system analysis the label/category

of the data 99% because the predictions we make may be different when done with other K-NN coding so it is impossible to be 100%.

It is hoped that this paper will raise more information about the classification of K-NN clearly and with an easy-to-understand writing presentation. In addition, this paper is expected to provide information about the classification of hospital services that can be useful in the future. We realize that this paper may still find shortcomings and there is a need for developments regarding explanations related to classification information using the K-Nearest Neighbor (K-NN) method.

References

Informatikologi (2017). Retrieved from www.informatikalogi.com: https://informatikalogi.com/algoritma-k-nn-k-nearest-neighbor/

Fitri, Y. (2016). Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (KNN). Jurnal BAPPEDA*, Vol. 2 No. 1*, 8.

Heryant. (2013, Agustus 2). *heryant.web.ugm.ac.id*. Retrieved from https://heryant.web.ugm.ac.id/rekam-medis/indikatorindikator-pelayanan-rumah-sakit-bor-avlos-toi-bto-gdr-ndr/

Hosizah, & Maryati, Y. (2018). Statistik Pelayanan Kesehatan. In Sistem Informasi Kesehatan II (p. 153). Kementrian Kesehatan Republik Indonesia.

Indikator Kinerja Pelayanan Rumah Sakit Per Triwulan 3 2020. (2021, January 13). Retrieved from http://data.jatengprov.go.id/dataset/indikator-kinerja-pelayanan-rumah-sakit-per-triwulan-3-2020

Prahudaya, T. Y., & Harjoko, A. (2017). Metode Klasifikasi Mutu Jambu Biji Menggunakan KNN Berdasarkan Fitur Warna Dan Tekstur. 11.

Risman, H., Nugroho, D., & WU, Y. R. (2015). Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di Stmik Sinar Nusantara Surakarta. 7.

www.hakayuci.com. (n.d.). Retrieved from https://www.hakayuci.com/2015/10/rumus-bor-avlos-toi-bto-gdr-ndr-indikator-pelayanan-rawat-inap.html