

Medical Record to Simplify the Hospital Decision Making System

Edy Kurniawan

Muhammadiyah University of Ponorogo, 63471, Indonesia

*Corresponding author. Email: edy@umpo.ac.id

ABSTRACT

Piles of medical record data in hospitals or health clinics are personal information of a patient. The data is a patient's medical history that is recorded and stored in the hospital. Hospitals can use this data if at any time the patient returns to the hospital or is used for referrals to other hospitals. Medical record data in recording uses standard writing, and for most hospitals or health clinics in Indonesia using the ICD-10 standard. From the medical record data, information about health conditions or health conditions in community groups and age groups can be obtained. Health can take advantage of the information contained in the pile of medical record data to carry out counseling or early action to the community in certain groups. This study used a classification method with a total of 1601 clean medical records. The results of this study were to find patterns in analyzing disease trends in a group of people. As a determining attribute that has the highest gain in medical record data is the ICD grouping. From the results of data processing using 5 fold cross validation, the most optimal results are obtained with an accuracy of 70%.

Keywords: Medical Record, Data Mining, Hospital.

1. INTRODUCTION

The existing data both in the health department and in the hospital are currently only limited to providing graphs or statistics on the number of patients seeking treatment for the disease along with reports of the patient's return. Reports from this data are currently used by the health office to carry out policies when it is to provide education to the public. Regarding the pattern of disease tendencies suffered by a group of people, it has not been explored to be used as a reference when conducting counseling or disease prevention [1].

The purpose and use of medical records is to support the achievement of an orderly administration in an effort to improve health services. Without the support of a good and correct medical record management system, orderly administration will not succeed. The uses of medical records include [2]: Administrative aspects, medical records have administrative value, because their content involves actions based on authority and responsibility as medical personnel and nurses in achieving health service goals.

Medical Aspects, these records are used as a basis for planning treatment / virginity that must be given to patients. For example, the patient's identity, patient's age,

patient's address, status, and some other information regarding patient identity.

Legal Aspects, concerning the problem of the existence of legal certainty guarantees on the basis of justice, in the context of efforts to defend the law and provide evidence material to uphold justice.

Research aspects, medical record files have research value, because they contain data / information that can be used as an aspect of research [3]. The information in the medical record contains at least 23 variables that explain the identity of the patient. Some of the important attributes include:

- a) Gender
- b) Address
- c) Age
- d) Diagnosis
- e) Disease Information (ICD)

ICD (International Classification Of Diseases), a standard grouping of diseases carried out by the WHO (World Health Organization). This ICD code continues to evolve, every time there is a fundamental change in the development of new diseases in the world. Until now, the

ICD coding technique reaches ICD-10. ICD-10 is a reference for all health service providers (hospitals, medical centers, puskesmas) to be used as a guide in conducting archives.

In the ICD-10 guideline, disease groups are classified into 20 disease groups. In ICD-10, this grouping is based on the character and type of the disease. This function is to facilitate the administration of medical recorders because detailed information from a patient will be difficult in itself in filing by health service providers if not grouped [4].

2. METHOD

Data mining is a series of processes to explore the added value of a data set in the form of knowledge that has not been known manually. Therefore, data mining actually has long roots in scientific fields such as artificial intelligence, machine learning, statistics and databases. Some of the techniques that are often mentioned in the data mining literature include association rule mining, clustering, classification, neural networks, and others [5].

The data used for training in the classification method using the J48 decision tree algorithm is 1601 data obtained in extracting medical record data at Aisyiyah Hospital, Ponorogo. The raw data obtained by researchers is 1601, which is the first quarterly report data in 2018. To find raw data that is clean from missing value (cannot be processed because it has incomplete information), preprocessing is needed, namely cleaning data, and data selection.

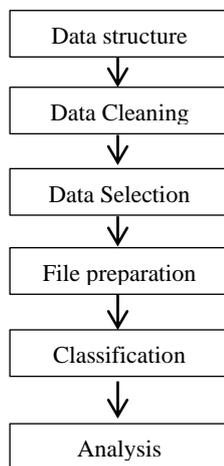


Figure 1 Research Systematics [6]

2.1. Cleaning Data

The initial stage of preprocessing data mining is data cleaning, this is done to remove data that has incomplete information. It is possible that a lot of data has incomplete information, such as the absence of age information, area information, gender information, disease grouping information.

2.2. Data Selection

Data that already has complete information in each attribute will then be selected. This selection is done to group the attributes according to the information needed. The attributes that were selected were the address attribute, age attribute and ICD-10 attribute. While the gender attribute is the target attribute.

In this study, data selection was carried out on data that had information on areas outside Ponorogo district. Because the hospital in question also accepts patients from outside Ponorogo, data containing information on areas outside Ponorogo regency will be discarded.

2.3. Age Groupings

The age grouping process has two methods (1) using describe (2) age grouping based on WHO rules. Discrete age grouping is not relevant when applied in this study, because the determination of the discrete in the preprocessing in the Weka 3.8.4 application uses the interval system. So the age grouping is not fundamental at levels [7].

The WHO methodology for age grouping for basic health is that age in years is grouped into three groups of age levels (infants and toddlers, young and adult groups, old groups) [8].

Table 1. Age grouping according to levels according to WHO

Age	Level
15 ≤	Babies and Children
15 to 50	Young and Adult
≥ 50	Old

2.4. Territorial grouping

The contents of the address attribute on the medical record are complete address information, where the address listed contains information on RT / RW, village / sub-district, road and number, District. In this regional grouping process the author uses MySQL and php to do the grouping automatically. This grouping technique is based on the address information that contains the word kecamatan in Ponorogo.

There are two stages of the regional grouping process, (1) converting addresses into sub-district names, (2) grouping sub-districts into regions as in the WHO division (urban areas, rural areas, coastal / mountainous areas) [4].

3. RESULT AND DISCUSSION

The preprocessing stage in this study had processed 1601 raw data. Disposing of missing value data results in

net data of 1421 medical record data. The next stage is to select data that are not included in the category for analysis. As an attribute that is carried out this selection is the attribute of the regional group. After the 1421 clean data were selected to find the actual data to be used for the analysis of this study, the researchers obtained 1400 clean data that had been selected.

3.1. The results of the data test used cross validation

To test the data with the cross-validation method, the researcher used 2 tests, where testing (1). using cross validation with 5-fold (2). testing using cross validation

Table 3. Results of the experiment using 10-fold

Test Data	Fold										Average
	1	2	3	4	5	6	7	8	9	10	
	53.2	33	67.9	52.3	51.9	49.1	76.9	51.9	44.4	57.4	53.8%

Table 4. Comparison of the results of cross validation between 5 folds and 10 folds

Test Data	Fold	
	5	10
	70%	53.8%

with 10-fold. From the test results, the following results are obtained:

3.1.1. Cross validation with 5-fold

Table 2. Results of the experiment using 5-fold

Test Data	Fold					Average
	1	2	3	4	5	
	73.7	78.8	77.4	71	49.1	70%

3.1.2. Cross validation with 10-fold

With Seeing from the results of the experiment using the cross validation method, it is known that the highest Correctly Classified Instances are in the experiment using 5-fold compared to using 10-fold.

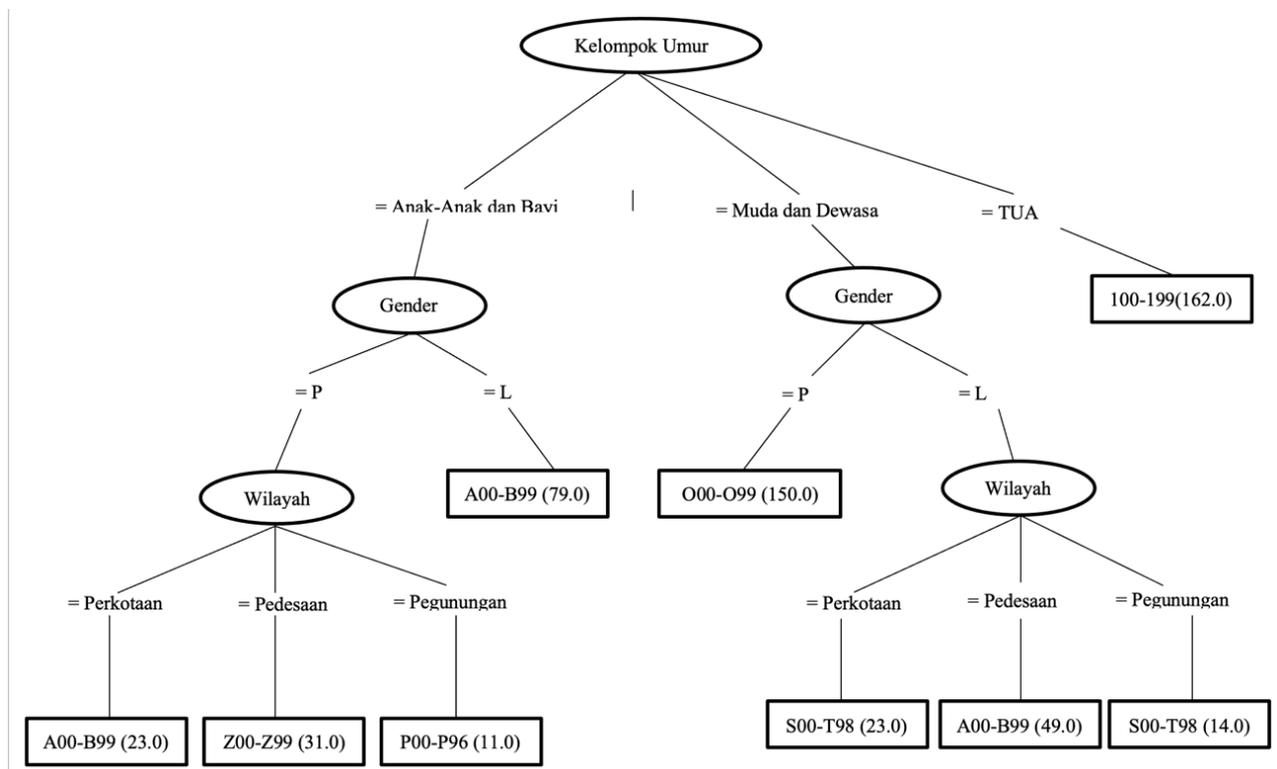


Figure 2 Decision tree of medical record results

Figure 2. provides information that in the "OLD" age group regardless of sex, the region for the most dominant disease group category is 100-I99 (Diseases of the

circulatory system or diseases of the circulatory system). Then information from young and adult age groups for women regardless of their location of residence, groups

of types of disease that have the potential are O00-O99 (Pregnancy, childbirth and the puerperium or related to the womb). Meanwhile, the male gender has two potentials, namely the S00-T99 (*Symptoms, signs and abnormal clinical and laboratory findings*, not elsewhere classified or groups of types of diseases that do not normally occur in other areas) and A00-B99 (Certain infectious and parasitic diseases). In the group of patients with children and infants, it is divided into three groups of diseases, the most dominant is the disease group A00-B99 (Certain infectious and parasitic diseases), then continued with the Z00-Z99 disease group (Factors influencing health status and contact with health services or factors that influence and direct contact with health services) while the minority of the disease group in these children is P00-P86 (Certain conditions originating in the perinatal period) which tends to be suffered by groups of children in the mountains.

4. CONCLUSION

Medical records have attributes that contain discrete values, so a decision tree is very suitable for use in medical record data. The medical record data has more than 15 attributes, but in this study, to determine the level of disease tendency suffered by a group of people in a certain area, only 4 attributes are needed. Attributes that affect the level of disease tendency in a group of regions are (1) gender, (2) region, (3) age, (3) ICD group. This study uses the J48 decision tree method.

From 1601 gross data, after cleaning the incomplete data and going through the selection process as required in the medical record analysis, 1400 clean data were obtained. Tests were carried out using the cross-validation method with fold 5 which has the most optimal results. By doing random cross validation of the four attributes tested, the highest gain information was obtained in the ICD group.

In this medical record data, in tests using pruning and unpruning, the results are the same. From the test results, the entire data using cross validation 5 fold can get the testing accuracy of 70%. The results of the decision tree provide information that the old age group is dominated by disease groups 100-199 (or groups of diseases related to blood vessels, while the age group of children and young and adults has a distribution of disease groups.

REFERENCES

- [1] E. Kurniawan, "Analysis of Medical Records to Determine the Pattern of Disease Groups Using Classification with the J48 Decision Tree," *Pros. Semin. Nas. Manaj. Technol.* XV, no. MMT-ITS, pp. 1–8, 2012.
- [2] L. Goodwin, M. Vandyne, S. Lin, and S. Talbert, "Data mining issues and opportunities for building nursing knowledge," *J. Biomed. Inform.*, vol. 36, no. 4–5, pp. 379–388, 2003, doi: 10.1016 / j.jbi.2003.09.020.
- [3] Y. Sebastian and PHH Then, "Domain-driven KDD for mining functionally novel rules and linking disjoint medical hypotheses," *Knowledge-Based Syst.*, vol. 24, no. 5, pp. 609–620, 2011, doi: 10.1016 / j.knosys.2011.01.008.
- [4] A. Mykowiecka, M. Marciniak, and A. Kupść, "Rule-based information extraction from patients' clinical data," *J. Biomed. Inform.*, vol. 42, no. 5, pp. 923–936, 2009, doi: 10.1016 / j.jbi.2009.07.007.
- [5] A. Ouali, A. Ramdane Cherif, and MO Krebs, "Data mining based Bayesian networks for best classification," *Comput. Stat. Anal Data.*, vol. 51, no. 2, pp. 1278–1292, 2006, doi: 10.1016 / j.csda.2005.09.012.
- [6] M. Delgado, D. Sánchez, MJ Martín-Bautista, and MA Vila, "Mining association rules with improved semantics in medical databases," *Artificial. Intell. Med.*, vol. 21, no. 1–3, pp. 241–245, 2001, doi: 10.1016 / S0933-3657 (00) 00092-0.
- [7] AK Sigurdardottir, H. Jonsdottir, and R. Benediktsson, "Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis," *Patient Educ. Couns.*, vol. 67, no. 1–2, pp. 21–31, 2007, doi: 10.1016 / j.pec.2007.03.007.
- [8] R. Pino-Mejías, MD Cubiles-de-la-Vega, M. Anaya-Romero, A. Pascual-Acosta, A. Jordán-López, and N. Bellinfante-Crocci, "Predicting the potential habitat of oaks with data mining models and the R system," *Environ. Model. Softw.*, vol. 25, no. 7, pp. 826–836, 2010, doi: 10.1016 / j.envsoft.2010.01.004.