Prediction of Disease Case Severity Level To Determine INA CBGs Rate

Sukma Puspitorini¹, Sri Kusumadewi², Linda Rosita³
¹Informatics Engineering, STMIK Nurdin Hamzah, Kota Jambi, Indonesia
²Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia
³Faculty of Medicine, Universitas Islam Indonesia, Yogyakarta, Indonesia

sukm4pit@gmail.com, cicie@uii.ac.id

Abstract. Indonesian Case-Based Groups (INA CBGs) is case-mix payment system using software grouper application. INA CBGs consisting of four digits code where the last digits indicating the severity level of disease cases. Severity level influence by secondary diagnosis (complications and co-morbidity) related to resource intensity level. It is medical resources used to treat a hospitalized patient. Objectives of this research is developing decision support system to predict severity level of disease cases and illustrate INA CBGs rate by using data mining decision tree classification model. Primary diagnosis (DU), first secondary diagnosis (DS 1), and second secondary diagnosis (DS 2) are attributes that used as input of severity level. The training process using C4.5 algorithm and the rules will represent in the IF-THEN form. Credibility of the system analyzed through testing process and confusion matrix present the results. Outcome of this research shows that first secondary diagnosis influence significant to form severity level predicting rules from new disease cases and INA CBGs rate illustration.

1. Background and Objective
Case - mix is a disease classification system that combines similar or same clinical characteristic disease, with the cost of medical treatment in a hospital that generally required same resources and treatment costs. Diagnosis grouping determined based on two principles, clinical homogeneity (patients who have clinical similarity) and resource homogeneity (patients using same sources intensity for similar therapy/resource consumption). INA CBGs (Indonesian Case-Based Groups) is case-mix payment system using software grouper application to determine the rates of health care facilities in accordance PMK No.69 Year 2013 about the Standard Rates of Health Services [1]. INA CBGs rates determined by an INA CBGs code consisting of four digits financing components write in alphanumeric code. Each digit explain case-mix main group (CMG), case-type group, case-based group (CBG), and severity level. CMG related with organ system that classify into 31 group, case-type group indicate 9 type of treatment and procedure type, and CBG refer to specific case-based group denoted with numeric from 01 to 99 [2]. Severity level (SL) as the last code that indicates disease cases severity associated with a secondary diagnosis that can prolong hospitalized patient’s length of stay (LOS) and caused rates varies on each case-mix main group (CMG).

The level of severity affect by complications and co-morbidity. Hospitalize patient grouped into 3 categories level of severity. Level I (mild) is diagnosis without complications and co-morbidity. Level II (medium) is diagnosis with mild complications and co-morbidity, and level III (severe) is diagnosis...
with major complications and co-morbidity [3]. If severity level could predict first during the treatment process, INA CBGs operators could anticipate increasing severity level during the treatment period and give illustration of INA CBGs rates before grouping process using INA CBGs software grouper.

In this study, severity levels determination relied on data grouping report by INA CBGs application. Result of the report is a large amount of data bases containing information and medical data or non-medical patients which analyzed to determine the pattern of prediction rules determining the severity level. Based on these, building decision support system (DSS) may help INA CBGs operators predicting severity level from disease diagnosis by knowing the patterns of patient data and classified using data mining techniques. Data mining is one of captivating soft computing techniques that widely use to analyze the hidden patterns in large amounts of information [3]. The rules of severity level determine by the classification model using C4.5 decision tree algorithm. Purpose of this research is implementing data mining decision tree classification model to develop severity level of disease case prediction and overview the rates. The C4.5 algorithm uses as decision tree classification. Attributes processed are age, primary diagnosis, and secondary diagnoses in hospitalized patients with hypertension, stroke, and diabetes mellitus type 2.

2. Methods

The following research analyzing 1000 patients from two government hospitals (RSUD Raden Mataher in Jambi and RSUD Dr. R. Soeprapto in Cepu) who had primary diagnosis of hypertension, stroke, and diabetes mellitus type 2 (non-insulin dependent diabetes mellitus) as seen in Table 1. This study focused on patient that had primary diagnosis with two secondary diagnosis to reduce the complexity in decision tree forming. First secondary diagnosis consist of 160 disease code while second secondary diagnosis consist of 109 disease code. Row data cannot directly use and needs transformation to categorize attribute. As the result, DS 1 categorized into 77 blocks of diseases and DS 2 categorized into 58 blocks of diseases. Secondary diagnosis categorized based on diseases block refer to ICD-10 chapter.

| Block         | Code | Disease Name                                      |
|---------------|------|---------------------------------------------------|
| Hypertension  | I10  | Essential (Primary) Hypertension                  |
|               | I119 | Hypertensive Heart Disease Without (Congestive) Heart Failure |
| Diabetes Mellitus Type 2 | E110 | Non-Insulin-Dependent Diabetes Mellitus With Coma |
|               | E111 | Non-Insulin-Dependent Diabetes Mellitus With Ketoacidosis |
|               | E117 | Non-Insulin-Dependent Diabetes Mellitus With Multiple Complications |
|               | E118 | Non-Insulin-Dependent Diabetes Mellitus With Unspecified Complications |
|               | E119 | Non-Insulin-Dependent Diabetes Mellitus Without Complications |
| Stroke        | I64  | Stroke, Not Specified As Haemorrhage Or Infarction |
The severity level prediction rules represented in the form of "IF-THEN" rules. This method does not require prior knowledge management and can solve the simple case which has dimensions of big data [4]. However, the constraint in building a decision tree is an attribute or a variable which will be at the root of the tree. One of the methods to determining the root of the decision tree is C4.5 algorithm. The roots of trees calculated using the concept of entropy-gain [5]. Attribute with the highest gain value will be elected as the root of the tree. Entropy compute using equation (1) below.

\[
Entropy(S) = \sum_{t=1}^{n} -pi \times \log_2(pi)
\]  

While the gain is calculated by equation (2) below

\[
Gain(S, A) = Entropy(S) - \sum_{t=1}^{n} \frac{|S_t|}{|S|} \times Entropy(S_t)
\]

Where
- $S$ = the reference node
- $A$ = attribute
- $n$ = number of attributes partitions $A$
- $|S_i|$ = Number of cases in the partition $i$ and $|S|$ = Number of cases in $S$
- $pi$ = proportion of $Si$ to $S$ and $\log_2 pi = \log/log_2$.

In addition, to find out INA CBGs rates, patient historical data use to determine CMG and CBG code. Data analyze using data mining tools WEKA 3.7.4 with J48 decision tree algorithm classification model. Testing process explain that decision tree generate tree rules of CMG and CBG code. Case-type in INA CBGs code describe treatment in hospitalize or outpatients, with or without medical procedure. There are 9 case types of cases and this study analyzing hospitalize patients without procedure with default code ‘4’.

3. Study Design

Classification worked to assess specific object from a number of available classes [6]. In data mining classification models, historical data could used as training data, to gain knowledge (such as a description or prediction). The data set divided into training set and testing set. Training set uses to build classification model while testing set uses to figure model accuracy. Decision tree classification model is supervises learning that proper to make predictions on this research [4]. Stages of this study are : 1) Data collection. This stage done through field studies from hospital and interviewing medical records staff to get information related INA CBGs code and severity level. 2) Data selection. Research focused on patient that had hypertension, diabetes mellitus type 2, and stroke primary diagnosis with two second diagnosis to reduce the complexity in decision tree forming. 3) Data analyzing. Result of this process setting primary diagnosis (DU), age, first secondary diagnosis (DS 1), and second secondary diagnosis (DS 2) as an input attribute. 4) Analysis data mining models using C4.5 decision tree classification algorithm to built prototype of decision making model. It will classify training set and predicted severity level from new cases. 5) Model implementation by building web-based decision support system applications to train data and performed testing. 6) Testing process to check credibility and validity of the system.

4. Decision Modelling

Decision modelling necessary to ease severity level determination. Model is predictive using decision tree classification model. There are several process to model the decision that is 1) Data input process. Data input is in csv format and user could export it from excel file or manually input it from application. 2) Set training data. in this study, all input data will set as training data. 3) Building decision tree using C4.5 algorithm. Decision tree will generate rule to classify severity level based on primary diagnosis, age, and disease block of secondary diagnosis. 4) Data testing. This process
conducted to determine validity from generated rule comparing with severity level result from INA CBGs grouping. 5) Credibility testing. System performance assess using confusion matrix.

5. Result
This application only serves as decision support software. Patient data that will predict has through the process of disease coding by medical record staff. User need entering medical record data like age, disease code of primary diagnosis (DU) and secondary diagnosis (DS 1 and DS 2) to determine severity level and rate prediction. Suppose user entering new input case. Patient age is 61 years with primary diagnosis code is I64, secondary diagnosis code are G21 and N309. Figure 1 shows the prediction result and information.

![Figure 1. Prediction Information Result](image1)

Display information explain that patient with primary diagnosis stroke not specified as hemorrhage or infarction with complication/co-morbidity secondary parkinsonism and cystitis unspecified has severity level II (SL 2) or medium severity level of disease case. Prediction of rate determine by clicking KLAIM button at the bottom and user will be redirected to page. Pres enter to assign CMG-CBG code. System classify new case into CMG-E and CBG-10. Then, select hospital type to determine regional and the type of the hospital. Suppose, user choose treatment in class 2. The result could see in Figure 2 below.

![Figure 2. INA CBGs Rates Prediction Result](image2)

System generate INA CBGs code E-4-10-I which mean that patient disease case classify as mild diabetes and metabolic nutritional disorder case-mix main group with specific CBG number 10. Patient hospitalized with no medical procedure treatment. Hospital region and type also influence INA CBGs rate.
Application testing performed to evaluate whether the severity level generated by the application fit with INA CBGs grouping result. The testing results from new cases in Table 2 below shows that from 15 cases, 13 cases predicted correctly while case 1, case 4, and case 5 has difference severity level result. This deviation is not surprising because there are partitions on the attributes that cannot determine the severity level class due to lack of base case. As the result, system classified severity level based on modus.

Table 2. Severity level prediction result using new data

| Case | Age | DU  | DS 1 | DS 2 | DSS | SL | INA CBGs SL | INA CBG CODE | RATE REG 1 TYPE B (IDR) | RATE REG 3 TYPE C (IDR) |
|------|-----|-----|------|------|-----|----|-------------|----------------|------------------------|------------------------|
| 1    | 40  | E110| A419 | 2    | 3   |    | E-4-10-II  | 4.235.205      | 2.798.255              |
| 2    | 47  | E117| N189 | C539 | 1   | 1  | E-4-10-I   | 3.059.460      | 1.820.720              |
| 3    | 50  | E118| I500 | 2    | 2   |    | E-4-10-II  | 4.235.205      | 2.798.255              |
| 4    | 52  | E119| A162 | 1    | 2   |    | E-4-10-I   | 3.059.460      | 1.820.720              |
| 5    | 67  | E119| L309 | 2    | 1   |    | E-4-10-II  | 4.235.205      | 2.798.255              |
| 6    | 38  | I10 | Z038 | 1    | 1   |    | I-4-17-I   | 3.502.018      | 2.928.961              |
| 7    | 78  | I10 | R51  | 1    | 1   |    | I-4-17-I   | 3.502.018      | 2.928.961              |
| 8    | 46  | I10 | G819 | 2    | 2   |    | I-4-17-II  | 4.747.470      | 3.970.611              |
| 9    | 57  | I10 | R42  | 1    | 1   |    | I-4-17-I   | 3.502.018      | 2.928.961              |
| 10   | 52  | I10 | J459 | N309 | 2    | 2   | I-4-17-II  | 4.747.470      | 3.970.611              |
| 11   | 79  | I10 | I500 | E119 | 2    | 2   | I-4-17-II  | 4.747.470      | 3.970.611              |
| 12   | 50  | I10 | G459 | 2    | 2   |    | I-4-17-II  | 4.747.470      | 3.970.611              |
| 13   | 44  | I119| K297 | 1    | 1   |    | I-4-17-I   | 3.502.018      | 2.928.961              |
| 14   | 71  | I64 | I10  | 1    | 1   |    | G-4-15-I   | 2.872.210      | 1.967.992              |
| 15   | 57  | I64 | E148 | 1    | 1   |    | G-4-15-I   | 2.872.210      | 1.967.992              |

INA CBGs rate is different in every regional. It depend on hospital type and class of treatment. Table 2 illustrate prediction if disease case occur on healthcare facilities type B at regional 1 and type C at regional 3 with the same class treatment. Although the INA CBGs code is same but the rate has differences.

6. Conclusion

Conclusion from this study is that data mining models classification with decision tree could use to compose severity level of disease cases rule. The first secondary diagnosis is attributes that significant determining severity level of disease cases. It is precondition that checked first before another attributes. Differences in severity level at any cases of disease caused differences in INA CBGs rates. Several disease case could have same INA CBGs code but the rate is different. It happen because hospitals regional differences and different class of treatment.

7. Reference

[1] Departemen Kesehatan Republik Indonesia 2013 Peraturan Menteri Kesehatan Republik Indonesia Nomor 69 Tahun 2013 Tentang Standar Tarif Pelayanan Kesehatan Pada Fasilitas Kesehatan Tingkat Pertama Dan Fasilitas Kesehatan Tingkat Lanjutan Dalam Penyelenggaraan Program Jaminan Kesehatan http://www.depkes.go.id/resources/download/general/PMK No. 69 Th 2013 ttg Tarif Pelayanan Kesehatan Program JKN.pdf

[2] Kementrian Kesehatan Republik Indonesia 2014 Peraturan Menteri Kesehatan Republik Indonesia No.27 Tahun 2014 Tentang Petunjuk Teknis Sistem Indonesian Case Base Groups
(INA-CBGs) http://www.jkn.kemkes.go.id/attachment/unduhan/PMK No. 20 ttg Juknis Sistem INA CBGs.pdf

[3] Tim Coding National Casemix Center (NCC) Departemen Kesehatan RI 2012 *Koding INA CBGs* http://bppsdmk.depkes.go.id/ckfinder/userfiles/files/KODING INA CBG.pdf

[4] Widodo P P, *et al* 2013 *Penerapan Data Mining Dengan Matlab* (Bandung: Penerbit Rekayasa Sains).

[5] Kusumadewi S, *et al* 2009 *Informatika Kesehatan* (Yogyakarta: Graha Ilmu).

[6] Prasetyo E 2012 *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab* (Yogyakarta: Andi Offset).