

DATA MINING REKAM MEDIS UNTUK MENENTUKAN PENYAKIT TERBANYAK MENGGUNAKAN DECISION TREE C4.5

Yuli Mardi

Akademi Perkam dan Informasi Kesehatan, Jl. Gajah Mada No. 23 Padang
email: adimardi@gmail.com

Submitted: 26-01-2018, Reviewed: 03-04-2018, Accepted 20-04-2018
<http://doi.org/10.22216/jsi.v4i1.3077>

Abstract

Medical record data in hospitals is rarely used for research and increase knowledge. Medical record data stored electronically or stored in the form of archives, periodically will be removed by the hospital according to existing rules, because the data is considered waste that will burden the storage media only. The main purpose of this research is how to utilize the medical record data that is considered to be a waste in order to give positive contribution for all parties both for hospital in making policy, for health facility, and for government in handling health. From data mining obtained at Citra BMC Padang General Hospital in January 2013, data analysis, data classification and decision tree making using algortima c4.5 were used, so that from total of 21 patients who got treatment got total entrophy 2,5061441 with amount most cases were found in CHAPTER XVIII (R00-R99) as many as 8 patients from 21, with sex details (female 5 patients and 3 men), age (elderly 5 patients, young and adults 1 patient, infant and child 2 patient), address (Padang Timur 4 patients, North Padang 1 patient, Lubuk Begalung 2 patient and Padang Barat 1 patient).

Keywords : C4.5; Data Mining; Decission Tree; Knime; Medical Record

Abstrak

Data rekam medis yang terdapat di rumah sakit jarang sekali digunakan untuk penelitian dan menambah pengetahuan. Data rekam medis yang disimpan secara elektronik maupun yang disimpan berupa arsip, secara berkala akan dihapus oleh pihak rumah sakit menurut aturan yang ada, karena data tersebut dianggap sampah yang akan membebani media penyimpanan saja. Tujuan utama dari penelitian ini adalah bagaimana memanfaatkan data rekam medis yang dianggap sampah tersebut agar dapat memberikan kontribusi yang positif bagi semua pihak baik bagi rumah sakit dalam membuat kebijakan, bagi fasilitas kesehatan, maupun bagi pemerintah dalam penanganan kesehatan. Dari data mining yang diperoleh di Rumah Sakit Umum Citra BMC Padang yang berobat pada bulan Januari 2013, dilakukan analisis data, klasifikasi data dan pembuatan pohon keputusan menggunakan algortima C4.5, sehingga dari total 21 pasien yang berobat didapatkan total entrophy 2,5061441 dengan jumlah kasus terbanyak terdapat pada BAB XVIII (R00-R99) yaitu sebanyak 8 pasien dari 21, dengan rincian jenis kelamin (perempuan 5 pasien dan laki-laki 3 pasien), usia (tua 5 pasien, muda dan dewasa 1 pasien, bayi dan anak 2 pasien), alamat (Padang Timur 4 pasien, Padang Utara 1 pasien, Lubuk Begalung 2 pasien dan Padang Barat 1 pasien).

Kata Kunci : C4.5; Data Mining; Decision Tree; Knime; Rekam Medis



PENDAHULUAN

Rekam medis menurut Edna H Kuffman merupakan berkas yang menyatakan siapa, apa, mengapa, dimana, kapan dan bagaimana pelayanan yang diberikan kepada pasien selama masa perawatan yang memuat pengetahuan mengenai pasien dan pelayanan yang diperolehnya serta memuat informasi yang cukup untuk mengidentifikasi pasien, membenarkan diagnosis dan pengobatan serta merekam hasilnya (Mawarni & Wulandari, 2013).

Keterangan-keterangan yang terdapat dalam berkas rekam medis tersebut berdasarkan pemeriksaan, pengobatan, observasi dan wawancara. Keterangan atau informasi yang terdapat dalam rekam medis tersebut tidak boleh disebarluaskan kepada pihak-pihak yang tidak berwenang, karena berkas rekam medis tersebut merupakan dokumen yang bersifat rahasia. Isi dari dokumen rekam medis terdapat informasi tentang diagnosis akhir pasien yang digunakan dalam proses pengkodean. Pengkodean ini dilakukan dengan menggunakan standar klasifikasi penyakit yang sesuai dengan ICD-10 atau *International Statistical Classification of Diseases and Related Health Problem Tenth Revision* (Yuliani, 2010).

Adapun tujuan penggunaan ICD-10 tersebut adalah untuk menyeragamkan nama dan golongan penyakit serta faktor-faktor yang mempengaruhi kesehatan (Rusliyanti, Hidayat, & Seha, 2016).

Data-data rekam medis / medical record yang tersimpan di rumah sakit jarang sekali digunakan untuk kegiatan penelitian dan hanya sebatas kewajiban

yang harus dilaporkan kepada pimpinan ataupun kepada dinas kesehatan.

Seiring dengan perkembangan ilmu pengetahuan dan teknologi sekarang ini, berkembang pula konsep data mining untuk mengolah data dalam jumlah yang besar yang dulunya dianggap sampah dan diabaikan begitu saja menjadi sesuatu yang ternyata dapat memberikan manfaat yang begitu banyak, baik untuk ilmu pengetahuan, bisnis maupun untuk pengambilan keputusan.

Dari data yang sangat besar tersebut, kemudian dapat dirumuskan permasalahan nya sebagai berikut :

- a) Bagaimana memanfaatkan data mining dengan metode *Decision Tree C4.5* untuk mengelompokkan penyakit terbanyak berdasarkan kode *International Classification of Diseases (ICD)*
- b) Bagaimana menggunakan salah satu software *Data Mining KNIME* untuk menggali informasi dan melakukan klasifikasi penyakit dari data rekam medis yang ada di sebuah rumah sakit, sehingga didapatkan hasil yang optimal

Penelitian yang dilakukan dibatasi hanya untuk data rekam medis pasien rawat inap rumah sakit umum Citra BMC Padang yang berobat pada bulan Januari 2013 dan data tersebut hanya data pasien yang berasal dari Kota Padang saja, karena rumah sakit Citra BMC juga menerima pasien dari luar Kota Padang.

LANDASAN TEORI

Istilah *Data Mining* dan *Knowledge Discovery in Database (KDD)* seringkali

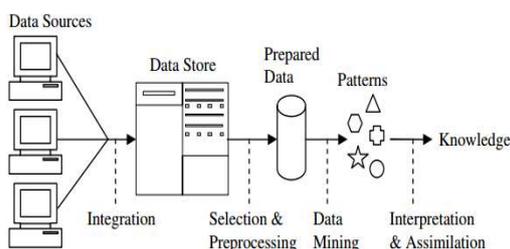
digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*.

Data mining sudah digunakan dalam beberapa penelitian, diantaranya untuk memperoleh model prakiraan yang sesuai agar memudahkan proses analisa dan prakiraan cuaca (Mujiasih, 2011), untuk memprediksi kriteria nasabah kredit, sehingga dapat diketahui apakah nasabah yang bersangkutan merupakan nasabah yang berpotensi menjadi nasabah kredit yang produktif atau tidak (Mabrur & Lubis, 2012), dan lain sebagainya

Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Kusrini. & Luthfi, 2009) :

- a) *Data Selection*
- b) *Pre-processing / Cleaning*
- c) *Transformation*
- d) *Data Mining*
- e) *Interpretation / Evaluation*

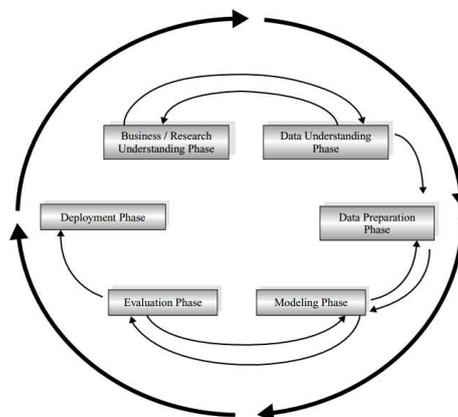
Sementara itu menurut (Bramer, 2007), proses Knowledge Discovery in Database (KDD) dapat digambarkan sebagai berikut :



Gambar 1. Proses Knowledge Discovery in Database

Menurut Gartner Group data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika.

The Cross-Industry Standard Process for Data Mining (CRISP DM), sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam tahapan (Larose, 2005).



Gambar 2. Proses Data Mining Menurut CRISP-DM

Dengan data mining, kita dapat mengelompokkan sesuatu menurut kelas, ciri-ciri persamaan ataupun perbedaan sebuah kebiasaan. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi (Widodo, Handayanto, & Herlawati., 2013).

Decision Tree merupakan metode yang paling efisien. Ibaratnya, kita



menyaring sesuatu lewat pohon keputusan, apakah suatu data lolos atau tidak terhadap saringan kita dengan proses yang cukup cepat.

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target (Achmad & Fauzi, 2012).

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Kusrini. & Luthfi, 2009). Algoritma C4.5 memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan) (Widodo et al., 2013).

Dalam pembuatan decision tree, perlu dicari terlebih dahulu nilai entropy dan gain, dengan rumus sebagai berikut :

a) menghitung nilai *entropy* digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

Di mana

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

b) menghitung nilai *gain* digunakan rumus :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

(S_i)

Di mana

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

S_i = proporsi S_i terhadap S

S = jumlah kasus dalam S

Langkah-langkah pembuatan pohon keputusan :

1. Menyiapkan *data training*.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*.
3. Kemudian hitung nilai *gain*.
4. Ulangi langkah ke-2 hingga semua *record* terpartisi
5. Proses partisi pohon keputusan akan berhenti pada saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi
 - c. Tidak ada *record* di dalam cabang yang kosong

HASIL DA PEMBAHASAN

Data bersih yang didapatkan dari rumah sakit umum Citra BMC Padang dapat dilihat di tabel 1 sebagai berikut :

Tabel 1. Data Bersih

NO	SEX	USIA	KECAMATAN	BAB ICD
1	L	Bayi dan anak-anak	Padang Timur	XVIII (R00-R99)
2	L	Bayi dan anak-anak	Padang Utara	I (A00-B99)
3	L	Tua	Padang Selatan	XIV (N00-N99)



4	L	Tua	Padang Utara	XVIII (R00-R99)
5	L	Muda dan dewasa	Lubeg	I (A00-B99)
6	P	Muda dan dewasa	Padang Timur	XIV (N00-N99)
7	P	Bayi dan anak-anak	Padang Selatan	I (A00-B99)
8	L	Bayi dan anak-anak	Padang Timur	I (A00-B99)
9	P	Bayi dan anak-anak	Padang Timur	XVIII (R00-R99)
10	P	Tua	Padang Timur	IV (E00-E90)
11	L	Bayi dan anak-anak	Padang Selatan	X(J00-J99)
12	P	Tua	Lubeg	XVIII (R00-R99)
13	P	Tua	Padang Barat	XVIII (R00-R99)
14	L	Tua	Padang Timur	XVIII (R00-R99)
15	P	Muda dan dewasa	Padang Timur	XVIII (R00-R99)
16	P	Bayi dan anak-anak	Padang Utara	III (D50-D89)
17	P	Tua	Padang Barat	IX (I00-I99)
18	P	Tua	Padang Barat	IX (I00-I99)
19	P	Tua	Lubeg	XVIII (R00-R99)
20	P	Bayi dan anak-anak	Padang Utara	I (A00-B99)
21	P	Muda dan dewasa	Lubeg	XV (O00-O99)

Dari data yang didapatkan, kemudian variabel-variabel tertentu dikelompokkan sebagai berikut :

a) Usia

Menurut WHO, pengelompokan usia untuk kesehatan mendasar yaitu dikelompokkan menjadi tiga kelompok jenjang usia. Atribut usia yang bernilai angka tersebut dibuat dalam bentuk kategori, yaitu usia < 15 tahun klasifikasinya bayi dan anak-anak, usia > 15 tahun dan < 50 tahun klasifikasinya muda dan dewasa, kemudian usia > 50 tahun klasifikasinya adalah tua. Klasifikasi tersebut bisa dilihat pada table 2 sebagai berikut :

Tabel 2. Klasifikasi Usia

Usia	Klasifikasi
< 15 tahun	Bayi dan anak-anak
> 15 tahun dan < 50 tahun	Muda dan dewasa
> 50 tahun	Tua

b) Pengelompokan Wilayah

Isi pada atribut alamat secara keseluruhan pada abstraksi rekam medis yang sesuai dengan UHDDS merupakan keterangan alamat secara lengkap, di mana alamat yang tercantum di dalamnya memuat jalan, kelurahan, kecamatan, kabupaten/kota, kode pos, kebangsaan dan seterusnya. Dalam hal ini atribut yang ada pada kolom alamat tidak akan dipilih semuanya, yang akan dijadikan atribut hanyalah atribut alamat kecamatan, karena alamat kecamatan bisa mewakili kota Padang secara keseluruhan, sedangkan atribut lainnya yang terdapat pada kolom alamat tidak akan digunakan.

c) Pengelompokan Kode ICD-10

Dalam asuhan kesehatan, pengkodean adalah pemberian angka untuk mewakili penyakit, prosedur dan bahan-bahan yang digunakan dalam pelaksanaan asuhan kesehatan. Pengkodean diagnosa berdasarkan ICD-10, bisa dikelompokkan berdasarkan bab sesuai dengan ketentuan ICD-10 Volume 3. Pengelompokan kode tersebut bisa dilihat pada tabel 3.

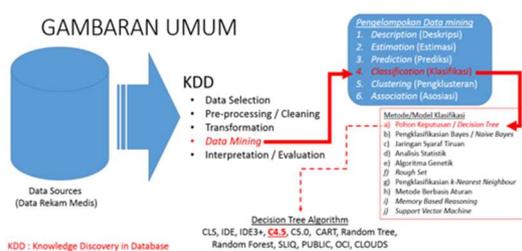
Tabel 3. Pengelompokan Kode Berdasarkan BAB pada ICD 10

Kode ICD	BAB ICD	Judul
A00-B99	I	Certain Infectious and parasitic diseases
C00-D48	II	Neoplasms
D50-D89	III	Diseases of the blood-forming organs and certain disorders involving the immune mechanism
E00-E90	IV	Endocrine, nutritional, and metabolic diseases
F00-F99	V	Mental and behavioural disorders
G00-G99	VI	Diseases of the nervous system
H00-H59	VII	Diseases of the eye and adnexa
H60-H95	VIII	Diseases of the ear and mastid process
I00-I99	IX	Diseases of the circulatory system
J00-J99	X	Diseases of the respiratory system
K00-K93	XI	Diseases of teh digestive system
L00-L99	XII	Diseases of the skin and subcutaneous tissue
M00-M99	XIII	Diseases of the musculoskeletal system and connective tissue
N00-N99	XIV	Diseases of the genitourinary system
O00-O99	XV	Pregnancy, childbirth and the puerperium
P00-P96	XVI	Certain condition originating in the perinatal period



Q00-Q99	XVII	Congenital malformation, deformations and chromosomal abnormalities
R00-R99	XVIII	Symptoms, sign and abnormal clinical and laboratory finding, not elsewhere classified
S00-T98	XIX	Injury, poisoning and certain other consequences of external causes
V01-Y98	XX	External causes of morbidity and mortality
Z00-Z99	XXI	Factors influencing health status and contact with health services

Gambaran umum penelitian dapat dilihat pada gambar 3.



Gambar 3. Gambaran Umum Penelitian

Data yang telah didapatkan kemudian dianalisis dan diolah, kemudian dilakukan proses Knowledge Discovery in Database (KDD) hingga mendapatkan data yang bersih dan siap diolah. Data tersebut kemudian diproses secara manual untuk mendapatkan decision tree dan menghasilkan pengetahuan berupa rule.

Setelah dilakukan analisis terhadap data mentah, kemudian didapatkan data bersih. Data bersih ini kemudian diolah menggunakan software Knime.

Knime (*Konstanz Information Miner*) merupakan software open source dan memiliki mode kerja grafis yang user friendly untuk akses data, transformasi data, penyelidikan awal, analisis prediktif, visualisasi dan pelaporan. Integrasi platform open source memberikan lebih dari 1000 modul atau node. KNIME juga menyediakan kemampuan untuk mengembangkan laporan berdasarkan

informasi data dan untuk mengotomatisasi penerapan wawasan baru kembali ke sistem produksi (Putra, 2015).

Pengolahan dan penggunaan *software* ini dilakukan pada sebuah laptop dengan spesifikasi sebagai berikut :

- Processor : *Intel Core i5 3317U*
- Memory : *DDR3 4 Gb*
- VGA : *Intel(R) HD Graphics 4000*
- Hardisk : *500 Gb*

Sedangkan perangkat lunak yang digunakan antara lain adalah sebagai berikut :

- Sistem Operasi : *Microsoft Windows 8 Pro 32 Bit*
- Aplikasi Office : *Microsoft Office 2013*
- Software Data Mining : *KNIME 2.8.2*

Sebelum dilakukan pengujian dengan menggunakan software Knime, dilakukan terlebih dahulu pengujian secara manual sesuai dengan teori dan rumus-rumusnya. Pengujian manual dilakukan terhadap data yang dijadikan sampel, yang kira-kira bisa mewakili data secara keseluruhan.

Sebelum dilakukan perhitungan nilai entropy untuk masing-masing atribut, terlebih dahulu dihitung nilai entropy total untuk semua atribut secara manual. Nilai entropy dapat dihitung dengan persamaan sebagai berikut : $Entropy (S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$

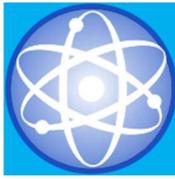
$$Entropy (S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

Entropy (total keputusan)

$$\left(-\frac{5}{21} \cdot \log_2 \left(\frac{5}{21}\right)\right) + \left(-\frac{1}{21} \cdot \log_2 \left(\frac{1}{21}\right)\right) + \left(-\frac{1}{21} \cdot \log_2 \left(\frac{1}{21}\right)\right) + \left(-\frac{2}{21} \cdot \log_2 \left(\frac{2}{21}\right)\right) + \left(-\frac{1}{21} \cdot \log_2 \left(\frac{1}{21}\right)\right) + \left(-\frac{2}{21} \cdot \log_2 \left(\frac{2}{21}\right)\right) + \left(-\frac{1}{21} \cdot \log_2 \left(\frac{1}{21}\right)\right) + \left(-\frac{8}{21} \cdot \log_2 \left(\frac{8}{21}\right)\right)$$

Entropy (total) = 2,5061441

Dari perhitungan entropy didapatkan nilai total entropy keseluruhan atribut



adalah 2, 5061441. Nilai inilah yang akan digunakan untuk menghitung nilai gain terhadap masing-masing atribut nantinya. Setelah didapatkan nilai entropy total untuk seluruh kasus, kemudian dihitung nilai entropy masing-masing atribut sebagai berikut :

Entropy (Sex)

Entropy (P) =

$$\left(-\frac{2}{13} \log_2 \left(\frac{2}{13}\right)\right) + \left(-\frac{1}{13} \log_2 \left(\frac{1}{13}\right)\right) + \left(-\frac{1}{13} \log_2 \left(\frac{1}{13}\right)\right) + \left(-\frac{2}{13} \log_2 \left(\frac{2}{13}\right)\right) + \left(-\frac{0}{13} \log_2 \left(\frac{0}{13}\right)\right) + \left(-\frac{1}{13} \log_2 \left(\frac{1}{13}\right)\right) + \left(-\frac{1}{13} \log_2 \left(\frac{1}{13}\right)\right) + \left(-\frac{5}{13} \log_2 \left(\frac{5}{13}\right)\right)$$

Entropy (P) = 1,668793615

Nilai entropy dari atribut jenis kelamin P adalah 1,668793615, kemudian dihitung nilai entropy atribut jenis kelamin L sebagai berikut :

Entropy (L) =

$$\left(-\frac{3}{8} \log_2 \left(\frac{3}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{1}{8} \log_2 \left(\frac{1}{8}\right)\right) + \left(-\frac{1}{8} \log_2 \left(\frac{1}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{3}{8} \log_2 \left(\frac{3}{8}\right)\right)$$

Entropy (L) = 1,811278124

Nilai entropy dari atribut jenis kelamin L adalah 1,811278124, kemudian dilanjutkan dengan menghitung nilai entropy dari atribut usia, diantaranya :

Entropy (Usia)

Entropy (Tua) =

$$\left(-\frac{0}{9} \log_2 \left(\frac{0}{9}\right)\right) + \left(-\frac{0}{9} \log_2 \left(\frac{0}{9}\right)\right) + \left(-\frac{1}{9} \log_2 \left(\frac{1}{9}\right)\right) + \left(-\frac{2}{9} \log_2 \left(\frac{2}{9}\right)\right) + \left(-\frac{0}{9} \log_2 \left(\frac{0}{9}\right)\right) + \left(-\frac{1}{9} \log_2 \left(\frac{1}{9}\right)\right) + \left(-\frac{0}{9} \log_2 \left(\frac{0}{9}\right)\right) + \left(-\frac{5}{9} \log_2 \left(\frac{5}{9}\right)\right)$$

Entropy (Tua) = 1,657742727

Nilai entropy dari atribut usia Tua adalah 1,657742727, kemudian dihitung

nilai entropy atribut usia Muda dan dewasa sebagai berikut :

Entropy (Muda dan dewasa) =

$$\left(-\frac{1}{4} \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{0}{4} \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{1}{4} \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{1}{4} \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{1}{4} \log_2 \left(\frac{1}{4}\right)\right)$$

Entropy (Muda dan dewasa) = 2

Nilai entropy dari atribut usia Muda dan dewasa adalah 2, kemudian dilanjutkan dengan menghitung nilai entropy atribut usia Bayi dan anak sebagai berikut :

Entropy (Bayi dan anak-anak) =

$$\left(-\frac{4}{8} \log_2 \left(\frac{4}{8}\right)\right) + \left(-\frac{1}{8} \log_2 \left(\frac{1}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{1}{8} \log_2 \left(\frac{1}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{0}{8} \log_2 \left(\frac{0}{8}\right)\right) + \left(-\frac{2}{8} \log_2 \left(\frac{2}{8}\right)\right)$$

Entropy (Bayi dan anak-anak) = 1,75

Nilai entropy dari atribut usia Muda dan anak adalah 1,75, kemudian dilanjutkan dengan menghitung nilai entropy dari atribut yang terakhir yaitu atribut kecamatan, diantaranya :

Entropy (Kecamatan)

Entropy (Padang timur) =

$$\left(-\frac{1}{7} \log_2 \left(\frac{1}{7}\right)\right) + \left(-\frac{0}{7} \log_2 \left(\frac{0}{7}\right)\right) + \left(-\frac{1}{7} \log_2 \left(\frac{1}{7}\right)\right) + \left(-\frac{0}{7} \log_2 \left(\frac{0}{7}\right)\right) + \left(-\frac{0}{7} \log_2 \left(\frac{0}{7}\right)\right) + \left(-\frac{1}{7} \log_2 \left(\frac{1}{7}\right)\right) + \left(-\frac{0}{7} \log_2 \left(\frac{0}{7}\right)\right) + \left(-\frac{4}{7} \log_2 \left(\frac{4}{7}\right)\right)$$

Entropy (Padang timur) = 1,664497779

Nilai entropy dari atribut kecamatan Padang Timur adalah 1,664497779, kemudian dilanjutkan dengan menghitung nilai entropy atribut kecamatan Utara sebagai berikut :

Entropy (Padang Utara) =



$$\left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) +$$

$$\left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right)$$

Entropy (Padang Utara) = 1,5

Nilai entropy dari atribut kecamatan Padang Utara adalah 1,5 kemudian dilanjutkan dengan menghitung nilai entropy atribut kecamatan Padang Selatan sebagai berikut :

Entropy (Padang Selatan) =

$$\left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) +$$

$$\left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right)$$

Entropy (Padang Selatan) = 1,584962501

Nilai entropy dari atribut kecamatan Padang Selatan adalah 1,584962501 kemudian dilanjutkan dengan menghitung nilai entropy atribut kecamatan Lubeg sebagai berikut :

Entropy (Lubeg) =

$$\left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) +$$

$$\left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4}\right)\right) + \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right)$$

Entropy (Lubeg) = 0,0000000

Nilai entropy dari atribut kecamatan Lubeg adalah 1,5 kemudian dilanjutkan dengan menghitung nilai entropy atribut kecamatan terakhir yaitu kecamatan Padang Barat sebagai berikut :

Entropy (Padang Barat) =

$$\left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{2}{3} * \log_2 \left(\frac{2}{3}\right)\right) +$$

$$\left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{0}{3} * \log_2 \left(\frac{0}{3}\right)\right) + \left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right)$$

Entropy (Padang Barat) = 0,918295834

Nilai entropy dari atribut kecamatan Padang Barat 0,918295834. Dengan demikian semua nilai atribut telah didapatkan, kemudian setelah semua nilai entropy masing-masing atribut didapatkan, dicari nilai gain untuk masing-masing atribut sebagai berikut :

Gain (total, sex) =

$$2,5061441 - \left(\frac{13}{21} * 1,668793615\right) + \left(\frac{8}{21} * 1,811278124\right)$$

Gain (total, sex) = 0,783070639

Gain (total, usia) =

$$2,5061441 - \left(\frac{9}{21} * 1,657742727\right) + \left(\frac{4}{21} * 2\right) + \left(\frac{8}{21} * 1,75\right)$$

Gain (total, usia) = 0,748063851

Gain (total, alamat) =

$$2,5061441 -$$

$$\left(\frac{7}{21} * 1,664497779\right) + \left(\frac{4}{21} * 1,5\right) + \left(\frac{3}{21} * 1,584962501\right) + \left(\frac{4}{21} * 1,5\right) + \left(\frac{3}{21} * 0,918295834\right)$$

Gain (total, alamat) = 2,5061441

Hasil perhitungan total jumlah kasus dan *entropy* nya, total nilai *entropy* untuk masing-masing atribut dan nilai *gain* untuk masing-masing atribut ditunjukkan pada tabel 4.

Tabel 4. Hasil Perhitungan Node 1

Node		JUMLAH KASUS								Entropy	Gain	
		I (S)	II (S)	III (S)	IV (S)	V (S)	VI (S)	VII (S)	XVIII (S)			
1	Total	21	5	1	1	2	1	2	1	8	2,506144067	
	Sex											0,783070639
	P	13	2	1	1	2	0	1	1	5	1,668793615	

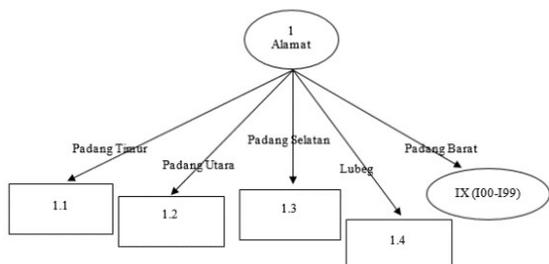


	L	8	3	0	0	0	1	1	0	3	1,811278124	
Usia												0,748063851
	Tua	9	0	0	1	2	0	1	0	5	1,657742727	
	Muda dan Dewasa	4	1	0	0	0	0	1	1	1	2	
	Bayi dan Anak	8	4	1	0	0	1	0	0	2	1,75	
Alamat												1,022274569
	Padang Timur	7	1	0	1	0	0	1	0	4	1,664497779	
	Padang Utara	4	2	1	0	0	0	0	0	1	1,5	
	Padang Selatan	3	1	0	0	0	1	1	0	0	1,584962501	
	Lubeg	4	1	0	0	0	0	0	1	2	1,5	
	Padang Barat	3	0	0	0	2	0	0	0	1	0,918295834	

Dari hasil tabel 4 dapat diketahui bahwa total nilai gain yang tertinggi adalah alamat, yaitu sebesar 1,022274569. Dengan demikian, alamat bisa dijadikan sebagai node akar. Ada 5 atribut dari dari alamat, yaitu Padang Timur, Padang Utara, Padang Selatan, Lubeg dan Padang Barat. Untuk Padang Barat sudah diklasifikasi keputusannya, yaitu IX (I00-I99). Adapun bentuk pohon keputusan sementara adalah seperti pada gambar 4 sebagai berikut :

perhitungan node 1.1.1, 1.3.1, dan 1.4.1 atau hasil perhitungan akhir dari node yang ada.

Setelah dilakukan proses perhitungan dan pembuatan *decision tree* secara manual, kemudian diuji dengan menggunakan salah satu *software data mining*, yaitu *KNIME (Konstanz Information Miner)*. Dari proses tersebut dihasilkan *decision tree* dan aturan yang sesuai dan sama dengan pengujian secara manual. Hasil akhir pohon keputusan yang terbentuk disajikan pada gambar 5 dan informasi ataupun rule yang dihasilkan dari proses pembuatan pohon keputusan ini dapat dilihat pada gambar 6.



Gambar 4. Pohon Keputusan Node 1

Adapun hasil pencarian entropy dan gain selanjutnya untuk masing-masing variable dan node berikut dapat disajikan pada tabel-tabel yang terdapat dibawah ini. Tabel 5 merupakan hasil perhitungan untuk node 1.1, 1.2, 1.3, dan 1.4, sedangkan tabel 6 merupakan hasil

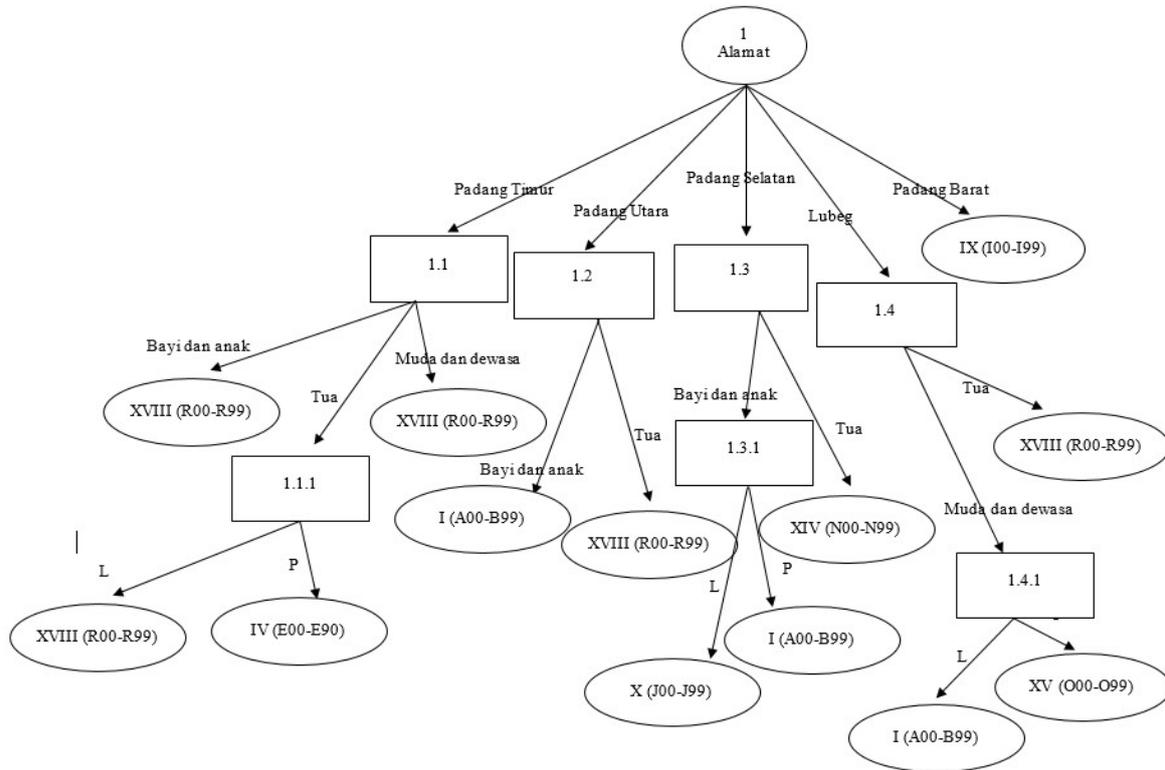
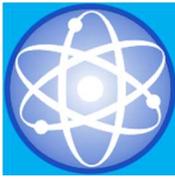


Tabel 5. Hasil Perhitungan Node 1.1, 1.2, 1.3, dan 1.4

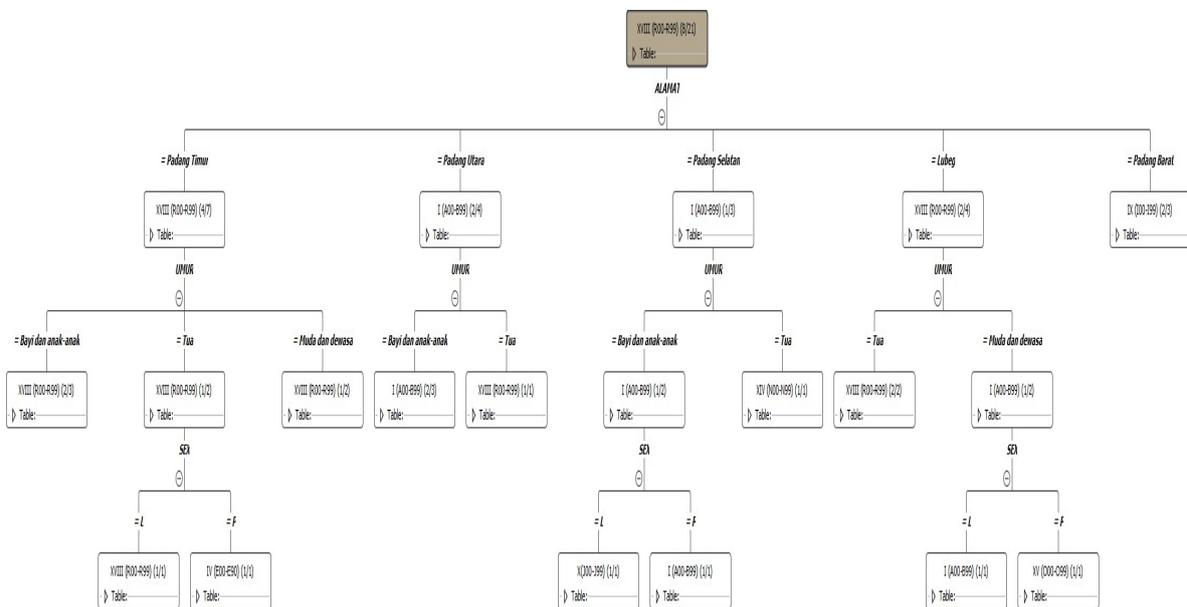
Node		Jumlah kasus	I (S ₁)	III (S ₂)	IV (S ₃)	IX (S ₄)	X (S ₅)	XIV (S ₆)	XV (S ₇)	XVIII (S ₈)	Entropy	Gian
1.1	Alamat-Padang Timur	7	1	0	1	0	0	1	0	4	1,664497779	0,69951385
	Usia											
	Bayi dan anak-anak	3	1	0	0	0	0	0	0	2	0,918295834	
	Tua	2	0	0	1	0	0	0	0	1	1	
	Muda dan dewasa	2	0	0	0	0	0	1	0	1	1	
	Sex											
	L	3	1	0	0	0	0	0	0	2	0,918295834	
	P	4	0	0	1	0	0	1	0	2	1,418295834	
1.2	Alamat-Padang Utara	4	2	1	0	0	0	0	0	1	1,5	1,040852083
	Usia											
	Bayi dan anak-anak	3	2	1	0	0	0	0	0	0	0,918295834	
	Tua	1	0	0	0	0	0	0	0	1	0	
	Muda dan dewasa	0	0	0	0	0	0	0	0	0	0	
	Sex											
	L	2	1	0	0	0	0	0	0	1	1	
	P	2	1	1	0	0	0	0	0	0	1	
1.3	Alamat-Padang Selatan	3	1	0	0	0	1	1	0	0	1,584962501	0,918295834
	Usia											
	Bayi dan anak-anak	2	1	0	0	0	1	0	0	0	1	
	Tua	1	0	0	0	0	0	1	0	0	0	
	Muda dan dewasa	0	0	0	0	0	0	0	0	0	0	
	Sex											
	L	2	0	0	0	0	1	1	0	0	1	
	P	1	1	0	0	0	0	0	0	0	0	
1.4	Alamat-Lubeg	4	1	0	0	0	0	0	1	2	1,5	1
	Usia											
	Bayi dan anak-anak	0	0	0	0	0	0	0	0	0	0	
	Tua	2	0	0	0	0	0	0	0	2	0	
	Muda dan dewasa	2	1	0	0	0	0	0	1	0	1	
	Sex											
	L	1	1	0	0	0	0	0	0	0	0	
	P	3	0	0	0	0	0	0	1	2	0,918295834	

Tabel 6. Hasil Perhitungan Node 1.1.1, 1.3.1, dan 1.4.1 (akhir)

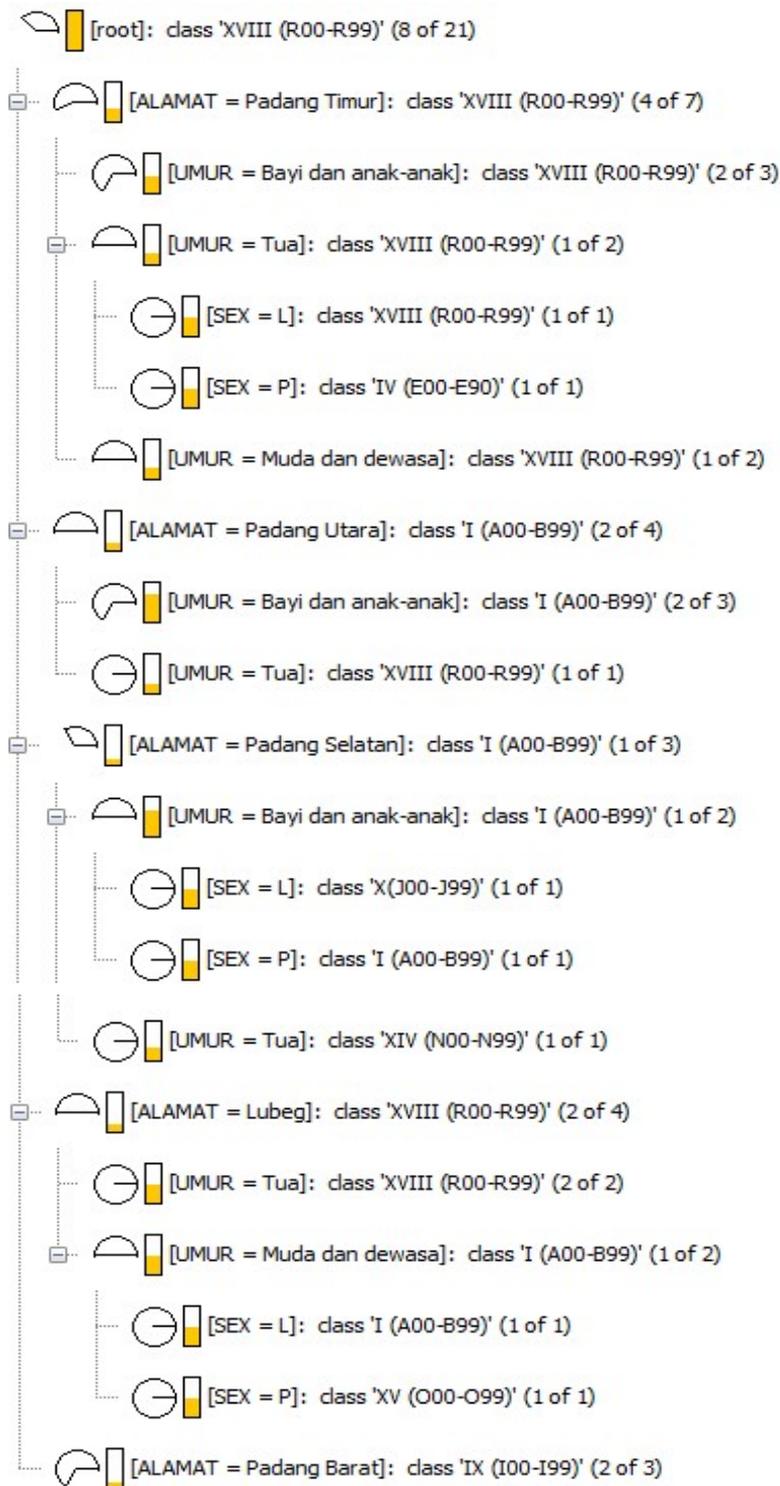
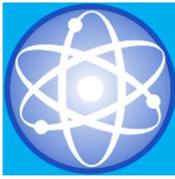
Node		Jumlah kasus	I (S ₁)	III (S ₂)	IV (S ₃)	IX (S ₄)	X (S ₅)	XIV (S ₆)	XV (S ₇)	XVIII (S ₈)	Entropy	Gian
1.1.1	Alamat-Padang Timur-Tua	2	0	0	1	0	0	0	0	1	1	1
	Sex											
	L	1	0	0	0	0	0	0	0	1	0	
	P	1	0	0	1	0	0	0	0	0	0	
1.3.1	Alamat-Padang Selatan-Bayi dan anak	2	1	0	0	0	1	0	0	0	1	1
	Sex											
	L	1	1	0	0	0	0	0	0	0	0	
	P	1	0	0	0	0	1	0	0	0	0	
1.4.1	Alamat-Lubeg-Muda dan dewasa	2	1	0	0	0	0	0	1	0	1	1
	Sex											
	L	1	1	0	0	0	0	0	0	0	0	
	P	1	0	0	0	0	0	0	1	0	0	



Gambar 5. Hasil Akhir Pohon Keputusan Secara Manual



Gambar 6. Hasil Akhir Pohon Keputusan Menggunakan KNIME



Gambar 7. Rule yang Dihasilkan



SIMPULAN

Dari penelitian dan penjelasan bab sebelumnya dapat ditarik beberapa kesimpulan, di antaranya :

1. Salah satu metode yang digunakan untuk klasifikasi adalah decision tree, metode ini sangat baik digunakan untuk menghitung sebaran penyakit terbanyak yang diderita pasien rawat inap di Rumah Sakit Umum Citra BMC Padang.
2. Keputusan *decision tree* tidak hanya untuk keputusan yang bersifat “Yes” atau “No” tapi juga bisa digunakan untuk banyak keputusan seperti menentukan penyakit apa saja yang banyak diderita di suatu wilayah seperti penelitian yang telah dilakukan, di mana keputusannya lebih dari 2 variabel.
3. Pohon keputusan dengan menggunakan algoritma C4.5 sangat mudah untuk digunakan, dianalisa dan diterjemahkan.
4. *KNIME* merupakan *software* yang sangat baik untuk digunakan dalam membuat *decision tree*. Selain tidak berbayar, *KNIME* juga mudah digunakan dan *output* yang dihasilkan juga mudah untuk direpresentasikan.

UCAPAN TERIMAKASIH

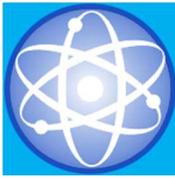
Ucapan terimakasih disampaikan kepada civitas akademika Apikes Iris Padang yang ikut membantu dan memberikan kontribusi dalam menyelesaikan penelitian ini. Kemudian juga kepada pihak RSU. Citra BMC Padang yang telah memberikan izin untuk dapat melakukan penelitian.

Terimakasih juga kepada semua pihak yang tidak dapat disebutkan satu persatu yang secara langsung atau tidak langsung

ikut membantu dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- Achmad, B. D. M., & Fauzi, S. (2012). Klasifikasi data karyawan untuk menentukan jadwal kerja menggunakan metode decision tree. *Jurnal IPTEK*, 16:1, 1–7. Diambil dari <http://jurnal.itats.ac.id/wp-content/uploads/2013/06/Klasifikasi-Data-Karyawan-.pdf>
- Bramer, M. (2007). *Principles of data mining*. London: Springer.
- Kusrini., & Luthfi, E. T. (2009). *Algoritma data mining*. Yogyakarta.: ANDI Yogyakarta.
- Larose, D. T. (2005). *Discovering Knowledge in Data An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Mabrur, A. G., & Lubis, R. (2012). Penerapan data mining untuk memprediksi kriteria nasabah kredit. *Jurnal Komputer dan Informatika (KOMPUTA)*, 1:1, 53–57. Diambil dari http://komputa.if.unikom.ac.id/_s/data/jurnal/volume-01/komputa-1-1-penerapan-data-mining-riani-8.pdf/pdf/komputa-1-1-penerapan-data-mining-riani-8.pdf
- Mawarni, D., & Wulandari, R. D. (2013). Identifikasi ketidaklengkapan rekam medis pasien rawat inap Rumah Sakit Muhammadiyah Lamongan. *Jurnal Administrasi Kesehatan Indonesia*, 1:2, 192–199. Diambil dari <http://download.portalgaruda.org/article.php?article=133649&val=1097>



- Mujiasih, S. (2011). Pemanfaatan data mining untuk prakiraan cuaca. *Jurnal Meteorologi dan Geofisika*, 12:2, 189–195. Diambil dari <http://puslitbang.bmkg.go.id/jmg/index.php/jmg/article/viewFile/100/94>
- Putra, E. (2015). Aplikasi software gratisan untuk keperluan predictive analytic. Diambil 26 April 2017, dari <http://www.bianalyst.net/2015/05/Aplikasi-Software-Gratisan-Predictive-Analytic.html>
- Rusliyanti, N. K. L., Hidayat, A. R., & Seha, H. N. (2016). Analisis ketepatan pengkodean diagnosis berdasarkan ICD-10 dengan penerapan karakter ke-5 pada pasien fraktur rawat jalan semester II di RSUD Mitra Paramedika Yogyakarta. *JURNAL PERMATA INDONESIA*, 7:1, 26–34. Diambil dari http://www.permataindonesia.ac.id/wp-content/uploads/2016/08/03.-Jurnal-PI_Lusi-Anas-Harinto.pdf
- Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). *Penerapan data mining dengan matlab*. Bandung: Rekayasa Sains, Bandung.
- Yuliani, N. (2010). Analisis keakuratan kode diagnosis penyakit commotio cerebri pasien rawat inap berdasarkan ICD-10 rekam medik di Rumah Sakit Islam Klaten. *ejournalinfokes*, 1:1, 17–31. Diambil dari http://www.apikescm.ac.id/ejournalinfokes/images/volume1/novita_voll.pdf