PENERAPAN *DATA MINING* UNTUK MENGKLASIFIKASIKAN PENYAKIT BERDASARKAN WILAYAH DENGAN METODE KLASIFIKASI DAN ALGORITMA C4.5

Danang Aristianto[1], Alif Finandhita[2]

[1,2] Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : danang.aristianto@gmail.com[1], finandhita@gmail.com[2]

ABSTRAK

Puskesmas Purwakarta merupakan salah satu pusat pelayanan kesehatan masyarakat yang ada di daerah Purwakarta. Setiap harinya, Puskesmas Purwakarta melayani banyak pasien dari berbagai wilayah yang ada di daerah Purwakarta, jumlah pasien bisa mencapai ratusan orang untuk sekali pelayanan dalam satu hari dan menghasilkan data laporan harian kunjungan pasien yang sangat banyak. Setiap bulannya pihak Puskesmas Purwakarta melakukan sebuah penyuluhan penyakit kepada masyarakat di suatu wilayah. Namun penyuluhan yang dilakukan oleh pihak Puskesmas Purwakarta tersebut memiliki suatu masalah yaitu tidak tepat sasarannya penyakit yang disampaikan kepada masyarakat. Masalah tersebut terjadi karena penyuluhan yang dilakukan tidak berdasarkan data laporan harian kunjungan pasien yang ada.

Data mining merupakan suatu teknik untuk menggali suatu informasi yang tersembunyi dari suatu kumpulan data. Metode yang digunakan untuk penelitian ini yaitu klasifikasi. Metode kalsifikasi ini bertujuan untuk mengklasifikasikan penyakit berdasarkan wilayah yang diperoleh dari data laporan harian kunjungan pasien. Algoritma yang digunakan adalah algoritma C4.5. Dengan metode dan algoritma tersebut nantinya akan menghasilkan suatu bentuk pohon keputusan yang bisa membantu pihak Puskesmas Purwakarta mengambil keputusan yang tepat sasaran dalam kegiatan penyuluhan penyakit di masyarakat.

Pengujian yang dilakukan dengan membandingkan hasil pohon keputusan yang dihasilkan secara manual dan aplikasi. Berdasarkan hasil pengujian, aplikasi yang dibangun dapat membantu pihak Puskesmas Purwakarta untuk kegiatan penyuluhan penyakit kepada masyarakat menjadi tepat sasaran berdasarkan data laporan harian kunjungan pasien yang telah diolah menggunakan aplikasi.

Kata Kunci : Algoritma C4.5, *Data mining*, Klasifikasi

1. PENDAHULUAN

1.1 Latar Belakang

Puskesmas Purwakarta merupakan salah satu pusat pelayanan kesehatan masyarakat yang ada di daerah Purwakarta, bertempat di Jalan Siliwangi No 4 Purwakarta. Seperti pusat pelayanan kesehatan pada umumnya, Puskesmas Purwakarta memberikan pelayanan kesehatan bagi masyarakat setiap harinya. Di Puskesmas Purwakarta sendiri mempunyai tiga macam pelayanan kepada para pasien, yaitu pelayanan Umum, Jampis (Jaminan Purwakarta Istimewa), dan BPJS (Badan Penyelenggara Jaminan Sosial). Yang membedakan ketiga jenis pelayanan tersebut yaitu di dalam hal administrasinya, jika pada pelayanan Umum, para pasien membayar biaya administrasi kepada Puskesmas Purwakarta, dan untuk pelayanan Jampis dan BPJS tidak dipungut biaya (gratis). Setiap harinya, Puskesmas Purwakarta melayani banyak pasien dari berbagai wilayah yang ada di daerah Purwakarta, jumlah pasien bahkan bisa mencapai ratusan orang untuk sekali pelayanan dalam satu hari.

Jumlah pasien yang banyak tersebut akan menghasilkan data laporan harian kunjungan pasien yang banyak pula. Dari data tersebut, Puskesmas Purwakarta membutuhkan pengolahan data penyakit lebih lanjut untuk menemukan informasi baru yang berguna sebagai pembantu mengambil keputusan, dalam hal ini yaitu kegiatan penyuluhan suatu penyakit ke suatu wilayah tertentu dengan cara melihat data klasifikasi yang telah terbentuk. Kegiatan penyuluhan penyakit yang selama ini dilakukan oleh pihak Puskesmas Purwakarta hanya berdasarkan penyakit di dalam program kerja yang sudah dibuat jadwalnya untuk kegiatan penyuluhan selama satu tahun. Penyakit yang terdapat di dalam program kerja tersebut memiliki beberapa macam penyakit diantaranya, Diabetes, Hipertensi, Pneumonia, Diare, dan TBC (*Tuberculosis*). Berdasarkan fakta yang sudah diungkapkan sebelumnya, maka kegiatan penyuluhan penyakit yang dilakukan oleh Puskesmas Purwakarta tersebut menimbulkan suatu masalah yaitu tidak tepat sasarannya informasi penyakit yang disampaikan kepada masyarakat. Hal ini dikarenakan penyakit yang ada di dalam program kerja tersebut bukan

berasal dari data laporan harian kunjungan pasien Puskesmas Purwakarta setiap harinya.

Berdasarkan hal tersebut, Puskesmas Purwakarta membutuhkan suatu alat bantu dalam pengolahan data laporan harian kunjungan pasien yang bisa memberikan informasi yang berguna. Dikarenakan juga Puskesmas Purwakarta menangani pasien dari berbagai wilayah, jadi perlu adanya pengklasifikasian penyakit yang digunakan untuk menentukan wilayah mana yang akan dijadikan tempat kegiatan penyuluhan penyakit. Teknik pengolahan data yang dibutuhkan Puskesmas Purwakarta untuk mengklasifikasikan penyakit berdasarkan wilayah yaitu dengan *Data Mining* dan metode Klasifikasi [1]. Dalam hal ini metode klasifikasi digunakan untuk mengklasifikasikan penyakit berdasarkan wilayah yang diperoleh dari data laporan harian kunjungan pasien. Dengan begitu informasi penyakit yang dipakai untuk kegiatan penyuluhan tersebut bisa tepat sasaran terhadap penyakit yang ada di masyarakat karena sesuai dengan data laporan harian kunjungan pasien yang ada dan bukan hanya berdasarkan program kerja. Algoritma yang digunakan adalah algoritma C4.5 karena algoritma tersebut cocok untuk metode klasifikasi dan pembentukan *decison tree* atau pohon keputusan [2] [3].

Dengan memanfaatkan teknologi *data mining* dengan metode klasifikasi dan algoritma C4.5 ini diharapkan dapat digunakan oleh pihak Puskesmas Purwakarta sebagai alat bantu untuk mengambil keputusan di dalam kegiatan penyuluhannya dan kegiatan tersebut menjadi tepat sasaran terhadap penyakit yang ada di masyarakat.

2. ISI MAKALAH

2.2 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari pembuatan aplikasi ini adalah menerapkan teknologi *data mining* untuk mengklasifikasikan penyakit berdasarkan wilayah dengan metode klasifikasi dan algoritma C4.5.

Sedangkan tujuan yang ingin dicapai dalam pembangunan aplikasi ini adalah :

Agar informasi penyakit yang disampaikan oleh pihak Puskesmas Purwakarta kepada masyarakat melalui kegiatan penyuluhan menjadi tepat sasaran dan membantu Puskesmas Purwakarta dalam mengambil keputusan wilayah mana yang akan menjadi tempat kegiatan penyuluhan berdasarkan informasi klasifikasi penyakit yang didapat dari hasil *data mining*.

2.3 *Data Mining*

Menurut Daniel T Larose di dalam bukunya yang berjudul *Discovering Knowledge in Data : An Introduction to Data Mining* menjelaskan, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1].

Beberapa definisi awal dari *data mining* menyertakan fokus pada proses otomatisasi. Berry dan Linoff dalam buku *Data mining Technique for Marketing, Sales, and Customers Support* mendefinisikan *data mining* sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semiotomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti [1].

2.4 Klasifikasi

Menurut Daniel T Larose di dalam bukunya yang berjudul *Discovering Knowledge in Data : An Introduction to Data Mining*, di dalam klasifikasi, data yang digunakan terdapat target variabel kategori atau target kelas [1]. Contoh untuk klasifikasi adalah penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Model *data mining* memeriksa sejumlah data yang jumlahnya besar, setiap data berisi informasi variabel target atau target kelas. Contohnya, seorang peneliti ingin mendapatkan pengklasifikasian golongan pendapatan dari seseorang yang tidak terdapat di dalam *database*, berdasarkan karakteristik yang berkaitan dengan orang tersebut, seperti usia, jenis kelamin, dan pekerjaan. Tugas ini sangat cocok dikerjakan dengan *data mining* dan metode klasifikasi. Proses algoritmanya kurang lebih seperti ini. Pertama, memeriksa data yang mengandung atribut prediktor (yang belum terklasifikasi) dan yang sudah terklasifikasi golongan pendapatannya. Dengan cara ini, algoritma (*software*) akan "belajar" mengenai kombinasi atribut yang berhubungan dengan golongan pendapatan. Sebagai contoh, wanita yang lebih tua mungkin berhubungan dengan golongan pendapatan yang tinggi. Data ini disebut *training set*. Lalu algoritma akan mencari data baru yang belum mempunyai informasi atau kelas golongan pendapatan. Berdasarkan klasifikasi pada di *training set*, algoritma akan menempatkan hasil klasifikasi ke data baru [1].

2.5 Algoritma C4.5

Dalam buku yang berjudul *Discovering Knowledge in Data : An Introduction to Data Mining* yang ditulis oleh Daniel T Larose menyebutkan bahwa algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an yang bertujuan untuk membuat suatu pohon keputusan. Algoritma C4.5 menggunakan konsep nilai *Entropy* dan *Gain* untuk membentuk pohon keputusan yang optimal [1].

Algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai akar
- Buat cabang untuk tiap – tiap nilai
- Bagi kasus dalam cabang
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk menghitung nilai *entropy* dapat dilihat pada persamaan berikut :

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

Keterangan : *S* : himpunan kasus
n : jumlah partisi *S*
pi : proporsi dari S_i ke *S*

Untuk menghitung *gain* digunakan rumus dari persamaan berikut :

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

Keterangan : *S* : himpunan kasus
A : atribut
n : jumlah partisi atribut A
S_i : jumlah kasus pada partisi ke – i
S : jumlah kasus dalam S

2.6 *Decision Tree*

Decision Tree / pohon keputusan merupakan pengubahan data menjadi pohon keputusan dan aturan – aturan keputusan. Pembentukan dimulai dari simpul akar, yang ditempatkan di paling atas sebuah pohon keputusan. Lalu setelah dari akar akan membentuk suatu cabang. Setiap cabang berisi nilai atribut, dan akan mengarah ke simpul lain atau langsung mengarah ke simpul daun atau kelas [1]. Data yang akan digunakan untuk membentuk pohon keputusan berbentuk tabel dengan atribut dan *record*. Satu atribut di dalam tabel menyatakan sebuah kelas keputusannya.

2.7 Tahapan *Data Mining*

Tahapan *data mining* yang digunakan pada penelitian ini adalah KDD (*Knowledge Discovery in Database*). KDD merupakan suatu area yang mengintegrasikan berbagai metode, yang meliputi statistik, basis data, kecerdasan buatan (*Artificial Intelegent*), *machine learning*, pengenalan pola (*patern recognition*), pemodelan yang menangani ketidakpastian, visualisasi data, optimasi, sistem informasi manajemen, dan sistem berbasis pengetahuan (*knowledge based system*) [5].

2.7.1 *Data Selection*

Yaitu pemilihan data dari sekumpulan data operasional yang sesuai untuk kebutuhan *data mining*. Di dalam penelitian ini data yang dipilih adalah data laporan harian kunjungan pasien di Puskesmas Purwakarta. Dikarenakan data laporan harian kunjungan pasien ini sudah mewakili hampir semua data yang ada di Puskesmas Purwakarta. Data laporan harian kunjungan pasien yang digunakan untuk penelitian ini adalah data bulan Juli 2015 dan yang menjadi sampel data adalah data tangal 31 Juli 2015.

2.7.2 *Preprocessing*

Proses ini bertujuan untuk memilih atribut dari suatu data dan membuang duplikasi data, memperbaiki kesalahan pada data seperti satu atribut tidak terisi atau kesalahan dalam menulis isi atribut. Berikut merupakan penjelasan pemilihan atribut dari data laporan harian kunjungan pasien Puskesmas Purwakarta.

Tabel 2.1. Tabel Pemilihan Atribut

Nama Atribut	Keterangan
Nama Penyakit	Atribut Nama Penyakit menjelaskan tentang nama penyakit yang diderita oleh masyarakat
Umur	Atribut ini dipilih karena bisa menjelaskan bagaimana membedakan penanganan terhadap pasien, seperti pemberian obat terhadap pasien anak – anak dan dewasa yang mempunyai dosis yang berbeda
Wilayah	Atribut Wilayah merupakan atribut penentu keputusan kegiatan sosialisasi Puskesmas Purwakarta yang menunjukan penyakit apa yang terdapat di Wilayah kerja Puskesmas Purwakarta yang terdiri dari Nagri Kidul, Nagri Tengah, Sindangkasih, dan Luar Wilayah

2.7.3 *Transformation*

Yaitu proses mengganti suatu nilai di sebuah atribut dengan nilai baru yang sesuai untuk proses *data mining*. Dalam data yang ada di Puskesmas Purwakarta atribut yang ada nilainya sudah sesuai untuk proses *data mining*. Contohnya seperti atribut Umur tidak terjadi perubahan dikarenakan nilai dari atribut umur sudah masuk ke dalam sebuah interval umur, jadi sudah sesuai untuk proses *data mining*. Begitu juga dengan atribut Wilayah yang sudah mengelompokan alamat dari setiap pasien ke dalam wilayah – wilayah tertentu sesuai dengan wilayah kerja Puskesmas Purwakarta.

Setelah proses *data selection*, *data prepocessing* hingga *transformation* selesai maka dihasilkan tabel yang siap untuk klasifikasi yang berbentuk pohon

keputusan menggunakan algoritma C4.5 seperti berikut.

Tabel 2.2. Tabel Kasus

Nama Penyakit	Umur	Wilayah
Diare	45 - 55 tahun	Sindangkasih
Diare	29 hari - 1 tahun	Nagri Tengah
Tuberkolosis	5 - 9 tahun	Luar Wilayah
Tuberkolosis	5 - 9 tahun	Luar Wilayah
Pneumonia	29 hari - 1 tahun	Sindangkasih
Diare	1 - 4 tahun	Luar Wilayah
Diare	5 - 9 tahun	Luar Wilayah
ISPA	1 - 4 tahun	Sindangkasih
ISPA	1 - 4 tahun	Sindangkasih
ISPA	1 - 4 tahun	Luar Wilayah
ISPA	1 - 4 tahun	Luar Wilayah
Dispepsia	20 - 44 tahun	Luar Wilayah
Dispepsia	45 - 55 tahun	Luar Wilayah
ISPA	5 - 9 tahun	Nagri Kidul
ISPA	5 - 9 tahun	Sindangkasih
ISPA	10 - 14 tahun	Sindangkasih
Gangguan Saraf	10 - 14 tahun	Nagri Kidul
ISPA	10 - 14 tahun	Nagri Kidul
Tukak Lambung	20 - 44 tahun	Luar Wilayah
Demam Tifoid	1 - 4 tahun	Sindangkasih
Demam Tifoid	5 - 9 tahun	Sindangkasih
ISPA	15 - 19 tahun	Luar Wilayah
Diabetes Melitus	56 - 59 tahun	Nagri Kidul
Diabetes Melitus	60 - 69 tahun	Luar Wilayah
Hipertensi Primer	20 - 44 tahun	Luar Wilayah
Hipertensi Primer	20 - 44 tahun	Luar Wilayah
Hipertensi Primer	45 - 55 tahun	Sindangkasih
Gangguan Kulit	20 - 44 tahun	Luar Wilayah
Cacar Air	20 - 44 tahun	Nagri Kidul
Common Cold	29 hari - 1 tahun	Sindangkasih

2.7.4 *Data Mining*

Yaitu proses mencari informasi berupa klasifikasi penyakit berdasarkan wilayah dengan bentuk *decision tree* atau pohon keputusan.

Sebelum membentuk pohon keputusan, data dari tabel di atas diolah terlebih dahulu menggunakan algoritma C4.5 dengan menghitung *entropy* dan memilih sebuah atribut sebagai akar didasarkan pada nilai *gain* tertinggi.

Berikut merupakan proses perhitungan untuk nilai *entropy*.

Entropy total = $((-\frac{5}{30} * \log_2(\frac{5}{30})) + ((-\frac{1}{30} * \log_2(\frac{1}{30})) + ((-\frac{10}{30} * \log_2(\frac{10}{30})) + ((-\frac{14}{30} * \log_2(\frac{14}{30})) = 1{,}6358$

1. Atribut Nama Penyakit

- *Entropy* Cacar Air = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* Common Cold = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* Demam Tifoid = $((-\frac{2}{2} * \log_2(\frac{2}{2})) = 0$
- *Entropy* Diabetes Melitus = $((-\frac{1}{2} * \log_2(\frac{1}{2})) + ((-\frac{1}{2} * \log_2(\frac{1}{2})) = 1$
- *Entropy* Diare = $((-\frac{1}{4} * \log_2(\frac{1}{4})) + ((-\frac{1}{4} * \log_2(\frac{1}{4})) + ((-\frac{2}{4} * \log_2(\frac{2}{4})) = 1{,}5$
- *Entropy* Dispepsia = $((-\frac{2}{2} * \log_2(\frac{2}{2})) = 0$
- *Entropy* Gangguan Kulit = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* Gangguan Saraf = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* Hipertensi Primer = $((-\frac{1}{3} * \log_2(\frac{1}{3})) + ((-\frac{2}{3} * \log_2(\frac{2}{3})) = 0{,}9183$
- *Entropy* ISPA = $((-\frac{2}{9} * \log_2(\frac{2}{9})) + ((-\frac{4}{9} * \log_2(\frac{4}{9})) + ((-\frac{3}{9} * \log_2(\frac{3}{9})) = 1{,}5305$
- *Entropy* Pneumonia = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* Tuberkolosis = $((-\frac{2}{2} * \log_2(\frac{2}{2})) = 0$
- *Entropy* Tukak Lambung = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$

2. Atribut Umur

- *Entropy* 29 hari – 1 tahun = $((-\frac{1}{3} * \log_2(\frac{1}{3})) + ((-\frac{2}{3} * \log_2(\frac{2}{3})) = 0{,}9183$
- *Entropy* 1 – 4 tahun = $((-\frac{3}{6} * \log_2(\frac{3}{6})) + ((-\frac{3}{6} * \log_2(\frac{3}{6})) = 1$
- *Entropy* 5 – 9 tahun = $((-\frac{1}{6} * \log_2(\frac{1}{6})) + ((-\frac{2}{6} * \log_2(\frac{2}{6})) + ((-\frac{3}{6} * \log_2(\frac{3}{6})) = 1{,}4591$
- *Entropy* 10 – 14 tahun = $((-\frac{2}{3} * \log_2(\frac{2}{3})) + ((-\frac{1}{3} * \log_2(\frac{1}{3})) = 0{,}9183$
- *Entropy* 15 – 19 tahun = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$
- *Entropy* 20 – 44 tahun = $((-\frac{1}{6} * \log_2(\frac{1}{6})) + ((-\frac{5}{6} * \log_2(\frac{5}{6})) = 0{,}6500$
- *Entropy* 45 – 55 tahun = $((-\frac{2}{3} * \log_2(\frac{2}{3})) + ((-\frac{1}{3} * \log_2(\frac{1}{3})) = 0{,}9183$
- *Entropy* 56 – 59 tahun = $((-\frac{1}{1} * \log_2(\frac{1}{1})) = 0$

- *Entropy* 60 – 69 tahun = $((-\frac{1}{1} * \log_2(\frac{1}{1}))$ = 0

Setelah perhitungan nilai *entropy*, selanjutnya adalah perhitungan nilai *gain* Untuk atribut Nama Penyakit dan Umur sebagai berikut.

1. *Gain* Nama Penyakit

$1{,}6358 - ((\frac{1}{30} * 0) + (\frac{1}{30} * 0) + (\frac{2}{30} * 0) + (\frac{2}{30} * 1) + (\frac{4}{30} * 1{,}5) + (\frac{2}{30} * 0) + (\frac{1}{30} * 0) + (\frac{1}{30} * 0) + (\frac{3}{30} * 0{,}9183) + (\frac{19}{30} * 1{,}5305) + (\frac{1}{30} * 0) + (\frac{2}{30} * 0) + (\frac{1}{30} * 0) = 0{,}8182$

2. *Gain* Umur

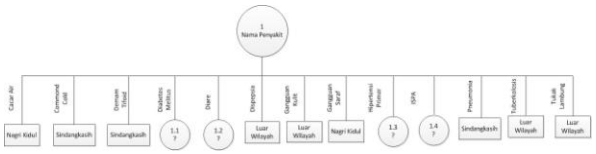
$1{,}6358 - ((\frac{3}{30} * 0{,}9183) + (\frac{6}{30} * 1) + (\frac{6}{30} * 1{,}4591) + (\frac{3}{30} * 0{,}9183) + (\frac{1}{30} * 0) + (\frac{6}{30} * 0{,}6500) + (\frac{3}{30} * 0{,}9183) + (\frac{1}{30} * 0) + (\frac{1}{30} * 0) = 0{,}7385$

Berikut merupakan hasil perhitungan untuk nilai *entropy* dan *gain* yang disajikan dalam bentuk tabel.

Nama Atribut	Nilai Atribut	Jml Kasus (total)	Nagri Kidul	Nagri Tengah	Sindangkasih	Luar Wilayah	Entropy	Gain
Total		30	5	1	10	14	1,6358	
Nama Penyakit								0,8182
	Cacar Air	1	1	0	0	0	0	
	Common Cold	1	0	0	1	0	0	
	Demam Tifoid	2	0	0	2	0	0	
	Diabetes Melitus	2	1	0	0	1	1	
	Diare	4	0	1	1	2	1,5	
	Dispepsia	2	0	0	0	2	0	
	Gangguan Kulit	1	0	0	0	1	0	
	Gangguan Saraf	1	1	0	0	0	0	
	Hipertensi Primer	3	0	0	1	2	0,9183	
	ISPA	9	2	0	4	3	1,5305	
	Pneumonia	1	0	0	1	0	0	
	Tuberkolosis	2	0	0	0	2	0	
	Tukak Lambung	1	0	0	0	1	0	
Umur								0,7385
	29 hari - 1 tahun	3	0	1	2	0	0,9183	
	1 - 4 tahun	6	0	0	3	3	1	
	5 - 9 tahun	6	1	0	2	3	1,4591	
	10 - 14 tahun	3	2	0	1	0	0,9183	
	15 - 19 tahun	1	0	0	0	1	0	
	20 - 44 tahun	6	1	0	0	5	0,6500	
	45 - 55 tahun	3	0	0	2	1	0,9183	
	56 - 59 tahun	1	1	0	0	0	0	
	60 - 69 tahun	1	0	0	0	1	0	

Gambar 2.1. Gambar Tabel Perhitungan *Entropy* dan *Gain*

Dari tabel di atas dapat diketahui bahwa nilai atribut *gain* tertinggi adalah Nama Penyakit yaitu sebesar 0,8182. Dengan demikian Nama Penyakit dapat menjadi *node* pertama dan juga menjadi *node* akar. Ada tiga belas nilai atribut dari Nama Penyakit. Dari ketiga belas nilai atribut tersebut masih perlu dilakukan perhitungan lagi untuk nilai atribut Diabetes Melitus, Diare, Hipertensi Primer, dan ISPA karena nilai dari atribut tersebut belum masuk pada sebuah kelas, dan hasil perhitungan dari tabel di atas dapat digambarkan pada pohon keputusan seperti gambar berikut.



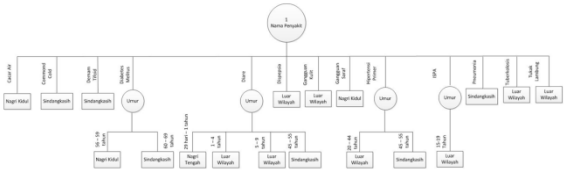
Gambar. 2.2. Bentuk Pohon Keputusan *Node* Pertama

Hasil perhitungan akhir *entropy* dan *gain* dan pohon keputusan dapat dilihat pada tabel dan gambar dibawah ini.

Nama Atribut	Nilai Atribut	Jml Kasus (total)	Nagri Kidul	Nagri Tengah	Sindangkasih	Luar Wilayah	Entropy	Gain
Nama Penyakit - ISPA		9	2	0	4	3	1,5305	
Umur								0,6416
	29 hari - 1 tahun	0	0	0	0	0	0	
	1 - 4 tahun	4	0	0	2	2	1	
	5 - 9 tahun	2	1	0	1	0	1	
	10 - 14 tahun	2	1	0	1	0	1	
	15 - 19 tahun	1	0	0	0	1	0	
	20 - 44 tahun	0	0	0	0	0	0	
	45 - 55 tahun	0	0	0	0	0	0	
	56 - 59 tahun	0	0	0	0	0	0	
	60 - 69 tahun	0	0	0	0	0	0	

Gambar 2.3. Gambar Tabel Perhitungan Akhir *Entropy* dan *Gain*

Dari tabel gambar tabel, masih ada nilai atribut yang belum memasuki kelas. Tetapi karena hanya atribut umur yang digunakan maka tidak ada perhitungan lagi untuk nilai atribut yang belum memasuki kelas. Maka perhitungan selesai sampai di sini.



Gambar 2.4. Bentuk Akhir Pohon Keputusan

2.7.5 *Interpretation / Evaluation*

Berdasarkan pohon keputusan yang sudah terbentuk pada gambar diatas dapat diketahui bahwa semua kasus sudah masuk ke dalam kelas. Maka dari itu pohon keputusan pada gambar diatas merupakan pohon keputusan yang bisa menjadi bahan bagi Puskesmas Purwakarta untuk menentukan kegiatan penyuluhan penyakit ke wilayah tertentu. Untuk hasil pohon keputusan tersebut akan disajikan ke dalam bentuk tabel sebagai berikut.

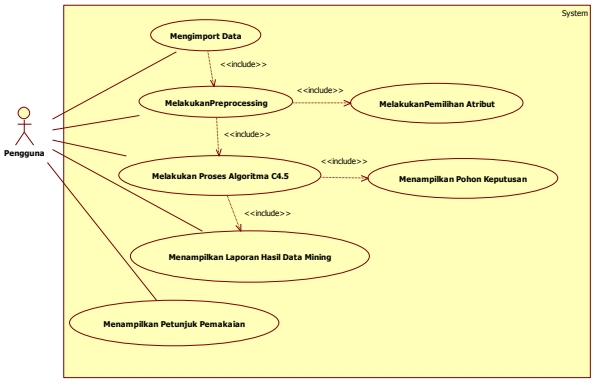
Tabel. 2.3. Tabel Hasil Pohon Keputusan

Nama Penyakit	Keterangan dan Wilayah
Cacar Air	Berada di Nagri Kidul
Commond Cold	Berada di Sindangkasih
Demam Tifoid	Berada di Sindangkasih
Diabetes Melitus	Umur 56 – 59 tahun berada di Nagri Kidul Umur 60 – 69 tahun berada di Sindangkasih
Diare	Umur 29 hari – 1 tahun berada di Nagri Tengah Umur 1 – 4 tahun bearda di Luar Wilayah Umur 5 – 9 tahun bearda di Luar Wilayah Umur 55 – 55 tahun bearda di Sindangkasih
Dispepsia	Berada di Luar Wilayah
Ganggua Kulit	Berada di Luar Wilayah
Gangguan Saraf	Berada di Nagri Kidul

Hipertensi Primer	Umur 20 – 44 tahun bearda di Luar Wilayah Umur 45 – 55 tahun bearda di Sindangkasih
ISPA	Umur 15 – 19 tahun berada di Luar Wilayah
Pneumonia	Berada di Sindangkasih
Tuberkulosis	Berada di Luar Wilayah
Tukak Lambung	Berada di Luar Wilayah

2.8 Analisis Kebutuhan Fungsional

2.8.1 Diagram *Use Case*



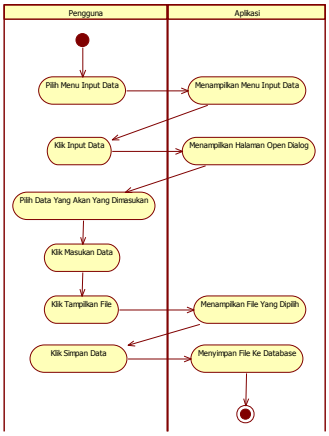
Gambar 2.5. Diagram *Use Case* Aplikasi *Data Mining*

2.8.2 Skenario *Use Case Input Data*

Use Case Name	**Input Data**	
Goal In Context	Masuk ke dalam aplikasi dan data bisa dimasukan ke dalam aplikasi dan *database*	
Precondition	-	
Succesful End Condition	Mengakses menu input data, menampilkan data ke dalam aplikasi dan berhasil memasukan data ke *database*	
Fail End Condition	Tidak bisa mengakses menu input data, tidak bisa menampilkan data ke dalam aplikasi dan tidak bisa memasukan data ke *database*	
Primary Actor	*User*	
Secondary Actor	-	
Trigger	Mengklik tombol simpan data	
Main Flow	Step	Action
	1.	Masuk ke dalam aplikasi
	2.	Memilih menu Input Data
	3.	Aplikasi akan menampilkan menu input data
	4.	*User* akan memilih data yang berformat Microsoft Excel
	5.	Setelah data terpilih, *user* akan memasukan data ke dalam aplikasi dan akan ditampilkan ke dalam aplikasi
	6.	Data yang sudah terpilih tersebut akan disimpan ke dalam *database* dengan mengklik tombol simpan data
Extension	-	-

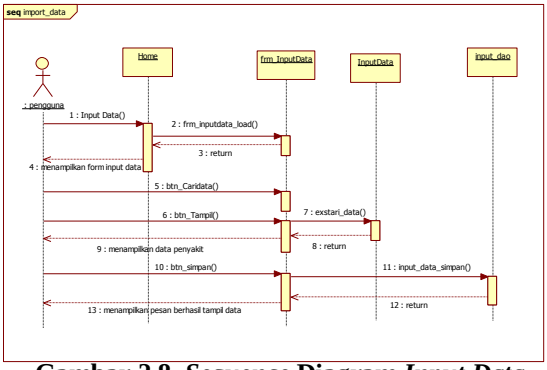
Gambar 2.6. Gambar Tabel Skenario *Use Case Input data*

2.8.3 Diagram *Activity Input Data*



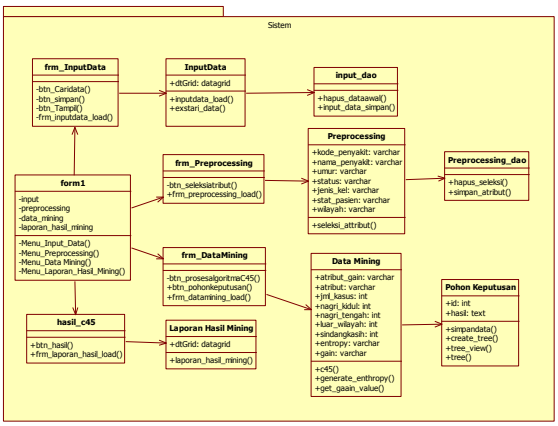
Gambar 2.7. Diagram *Activity Input Data*

2.8.4 *Sequence* Diagram *Input Data*



Gambar 2.8. *Sequence* Diagram *Input Data*

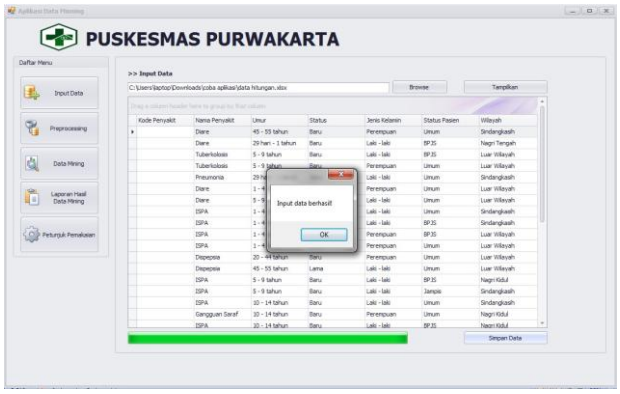
2.8.5 *Class* Diagram



Gambar 2.9. *Class* Diagram

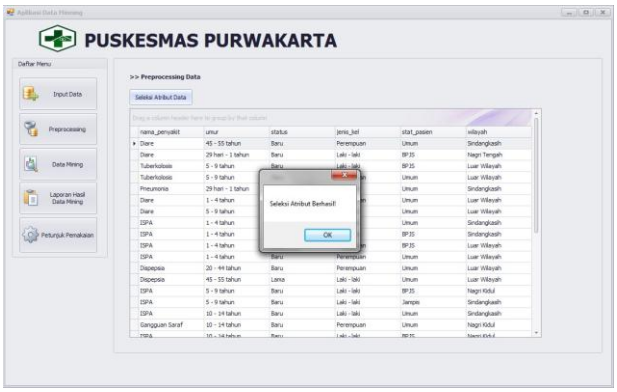
2.9 Implementasi Antar Muka

a. Tampilan Menu *Input Data*



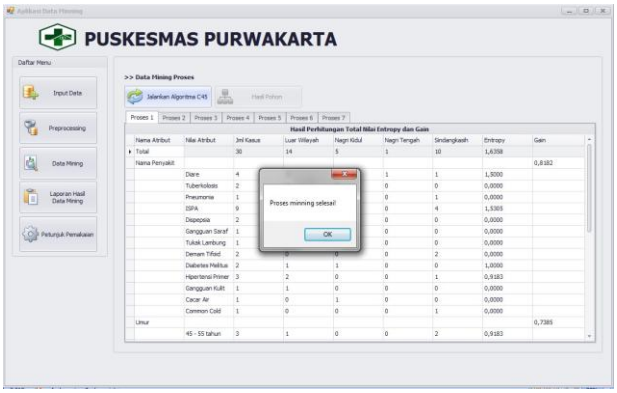
Gambar 2.10. Tampilan Menu *Input Data*

b. Tampilan Menu *Preprocessing*



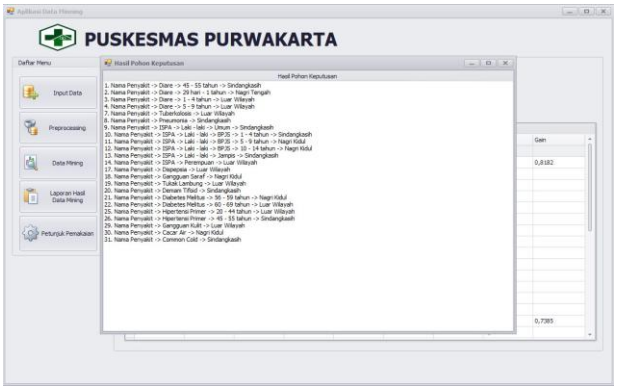
Gambar 2.11. Tampilan Menu *Preprocessing*

c. Tampilan Menu *Data Mining*



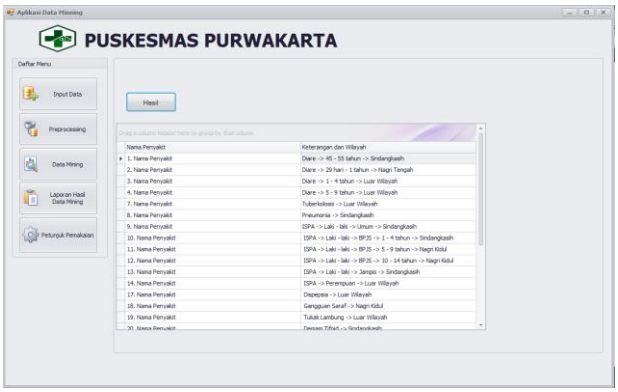
Gambar 2.12. Tampilan Menu *Data Mining*

d. Tampilan Pohon Keputusan



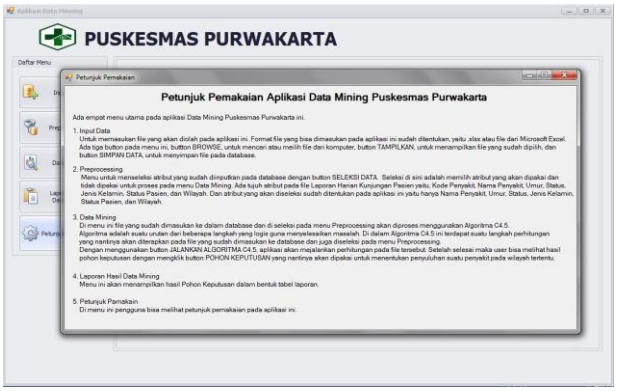
Gambar 2.13. Tampilan Pohon Keputusan

e. Laporan Hasil *Data Mining*



Gambar 2.14. Laporan Hasil *Data Mining*

f. Petunjuk Pemakaian



Gambar 2.15. Petunjuk Pemakaian

3. PENUTUP

Berdasarkan hasil analisis dan pengujian yang telah dilakukan pada sistem penerapan *Data Mining* dengan metode Klasifikasi dan Algoritma C4.5 ini maka dapat diambil kesimpulan sebagai berikut:

Aplikasi yang dibangun dapat membantu pihak Puskesmas Purwakarta untuk kegiatan penyuluhan penyakit kepada masyarakat menjadi tepat sasaran berdasarkan data laporan harian kunjungan pasien yang telah diolah menggunakan aplikasi. Informasi yang dihasilkan oleh aplikasi berupa pohon keputusan dan dapat menjadi bahan pengambilan

keputusan pihak Puskesmas Purwakarta wilayah mana yang akan menjadi tempat kegiatan penyuluhan penyakit.

UCAPAN TERIMA KASIH

Penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, namun berkat bantuan, bimbingan, kerja sama dari berbagai pihak dan berkah dari Allah SWT sehingga kendala-kendala yang dihadapi dapat diatasi. Untuk itu penulis menyampaikan rasa hormat dan terima kasih sebesar-besarnya kepada :

1. Allah SWT karena memberikan rahmat dan karunia-Nya.
2. Baginda Nabi Muhammad SAW karena kecintaan terhadap umatnya.
3. Kepada Siti Nurjanah dan Alm. Suratno selaku orang tua yang telah tulus mendoakan, memberikan dorongan moril dan materil, masukan, perhatian yang tidak ternilai.
4. Bapak Alif Finandhita, S.Kom., M.T. selaku dosen pembimbing yang telah banyak memberikan dorongan dan arahan selama proses penyusunan tugas akhir ini.
5. Ibu Dian Dharmayanti, S.T., M.Kom. selaku dosen reviewer yang telah memberikan saran serta kritiknya dalam penyusunan tugas akhir ini.
6. Ibu Gentisya Tri Mardiani, S.Kom., M.Kom. selaku dosen wali dan dosen penguji sidang skripsi yang telah memberikan saran serta kritiknya dalam penyusunan tugas akhir ini.
7. Ibu Anne Hediana Koesmah, Hj., dr., MM selaku kepala dinas kesehatan Kabupaten Purwakarta yang telah memberikan izin agar bisa melakukan penelitian di ruang lingkup Dinas Kesehatan Kabupaten Purwakarta.
8. Ibu Hj. Atik Hayati, AmKeb, MMKes selaku kepala Puskesmas Purwakarta yang telah memberikan izin agar bisa melakukan penelitian di Puskesmas Purwakarta.
9. Kepada teman – teman sesama mahasiswa Unikom terutama kelas IF-14 angkatan 2011 yang selalu mendukung, bersedia bertukar pikiran dan memberikan motivasi selama proses penyusunan tugas akhir ini.
10. Dan semua pihak yang tidak dapat saya sebutkan satu persatu saya ucapkan terimakasih atas segalanya.

DAFTAR PUSTAKA

[1] Larose, Daniel T. 2005. *Discovering Knowledge in Data : An Introduction to Data Mining.* John Willey & Sons, Inc.

[2] Kusrini dan Luthfi, Emha Taufiq. 2009. *Algoritma Data Mining*. Penerbit Andi. Yogyakarta.

[3] Adi Swara, Jaka. *Aplikasi Data Mining Menggunakan Metode Decison Tree Untuk Pemantauan Distribusi Penjualan Sepeda Motor Di PD. Wijaya Abadi Bandung.*

[4] Sommerville, Ian. 2011. *Software Engineering : 9th Edition.* Addison-Wesley. Boston Massachusetts.

[5] Sumanthi, S. Sivanandam, S.N. 2009. *Introduction to Data Mining and its Applications*. Spinger. Verlag Berlin Heidelberg.

[6] Kadir, Abdul. 2003. *Pengenalan Sistem Informasi.* Penerbit Andi Yogyakarta. Yogyakarta.

[7] Sutabri, Tata. 2005. *Sistem Informasi Manajemen.* Penerbit Andi Yogyakarta. Yogyakarta.

[8] Fathansyah. 2004. *Buku Teks Komputer Basis Data*. Informatika Bandung.

[9] Hariyanto, Bambang. 2004. *Sistem Manajemen Basis Data : Pemodelan, Perancangan, dan Terapannya.* Informatika Bandung.

[10] Connolly, T. Begg, C. 2010. *Database Systems : A Practical Approach To Design, Implementation, and Management. 5th Edition*. America : Pearson Education.

[11] Utdirartatmo, Firar. 2002. *Mengelola Database Server MySQL Di Linux Dan Windows*. Penerbit Andi Yogyakarta.

[12] Weisfeld, Matt. 2009. *Object-Oriented Thought Process, Third Edition*. Addison-Wesley

[13] Fowler, Martin. 2005. *UML Distilled. Panduan Singkat Bahasa Permodelan Objek Standar Edisi 3*. Penerbit Andi Yogyakarta. Yogyakarta.

[14] Presman, Roger S. 1997. *Software Engineering : A Practioner's Approach 4th Edition*. Mcgraw Hill.