PERBANDINGAN KLASIFIKASI PENYAKIT HIPERTENSI MENGGUNAKAN REGRESI LOGISTIK BINER DAN ALGORITMA C4.5

(Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul)

SKRIPSI

Disusun Oleh :

WELLA RUMAENDA

24010211130037

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2016

PERBANDINGAN KLASIFIKASI PENYAKIT HIPERTENSI MENGGUNAKAN REGRESI LOGISTIK BINER DAN ALGORITMA C4.5

(Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul)

Disusun Oleh :

WELLA RUMAENDA

24010211130037

Skripsi

Diajukan Sebagai Salah Satu Syarat untuk memperoleh Gelar

Sarjana Statistika pada Jurusan Statistika

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2016

LEMBAR PENGESAHAN I

Judul : Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5 (Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul)

Nama : Wella Rumaenda

NIM : 24010211130037

Telah diujikan pada sidang Tugas Akhir tanggal 21 Maret 2016 dan dinyatakan lulus pada tanggal 28 Maret 2016.

Semarang, 28 Maret 2016

Mengetahui,
Ketua Jurusan Statistika
FSM UNDIP

Dra. Dwi Ispriyanti, M.Si
NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir
Ketua,

Drs. Tarno, M.Si
NIP. 196307061991021001

LEMBAR PENGESAHAN II

Judul : Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5 (Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul)

Nama : Wella Rumaenda

NIM : 24010211130037

Telah diujikan pada sidang Tugas Akhir tanggal 21 Maret 2016.

Semarang, 28 Maret 2016

Pembimbing I	Pembimbing II
Yuciana Wilandari, S.Si, M.Si	Diah Safitri, S.Si, M.Si
NIP. 197005191998022001	NIP. 197510082003122001

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang diberi judul **"Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5 (Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul).** Tugas Akhir ini tidak akan terselesaikan dengan baik tanpa adanya dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada :

1. Ibu Dra. Dwi Ispriyanti, M.Si. selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Ibu Yuciana Wilandari, S.Si, M.Si dan Ibu Diah Safitri, S.Si, M.Si selaku dosen pembimbing I dan II.
3. Bapak/Ibu dosen jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
4. Kepala UPT Puskesmas Ponjong I Gunungkidul beserta jajarannya yang telah mengizinkan dan membantu penulis dalam pengambilan data penelitian.
5. Orang tua beserta semua pihak yang telah membantu kelancaran penyusunan Tugas Akhir ini.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran demi kesempurnaan penulisan selanjutnya.

Semarang, 28 Maret 2016

ABSTRAK

Hipertensi masih menjadi masalah utama dunia sampai saat ini. Di Indonesia prevelensi hipertensi terbilang masih cukup tinggi. Terdapat dua jenis hipertensi berdasarkan penyebabnya, yaitu hipertensi primer dan sekunder. Pada tugas akhir ini difokuskan pada pengklasifikasian jenis hipertensi berdasarkan penyebabnya menggunakan metode regresi logistik biner dan algoritma C4.5 dengan studi kasus pasien hipertensi di UPT Puskesmas Ponjong I, Gunungkidul bulan Oktober-November 2015. Regresi logistik biner adalah metode yang menjelaskan hubungan antara variabel respon dan beberapa variabel prediktor dengan variabelnya bernilai 1 untuk menyatakan keberadaan suatu karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan suatu karakteristik. Algoritma C4.5 merupakan salah satu metode klasifikasi *data mining* yang digunakan untuk membentuk pohon keputusan (*desicion tree*). Variabel prediktor yang digunakan pada tugas akhir ini adalah jenis kelamin, umur, tekanan darah sistolik, tekanan darah diastolik, riwayat berobat, serta penyakit lain. Untuk mengevaluasi hasil klasifikasi digunakan perhitungan APER (*Apparent Error Rate*). Berdasarkan analisis tersebut, klasifikasi penyakit hipertensi dengan metode regresi logistik biner diperoleh nilai APER=27,4648% dan ketepatan klasifikasi sebesar 72,5352%, sedangkan menggunakan algoritma C4.5 diperoleh nilai APER=35,9155% dan ketepatan klasifikasi sebesar 64,0845%. Pada uji beda dua proporsi didapatkan bahwa ada perbedaan signifikan dari kedua metode.

Kata kunci : Jenis Hipertensi, Klasifikasi, Regresi Logistik Biner, Algoritma C4.5, APER

ABSTRACT

Hypertension is a major problem in the world today. In Indonesia prevalence of hypertension is still high. There are two types of hypertension based on cause, primary and secondary hypertension. In this thesis focused on the classification of types of hypertension based on the cause using binary logistic regression and C4.5 algorithms with case studies in UPT Puskesmas Ponjong I, Gunungkidul of October-November 2015. Binary logistic regression is a method that describes the relationship between the response variable and several predictor variables with the variable equal to 1 to declare the existence of a characteristic and the value 0 to declare the absence of a characteristic. C4.5 algorithm is one method of classification of data mining is used to create a decision tree. The predictor variables were used in this thesis are gender, age, systolic blood pressure, diastolic blood pressure, treatment history, other diseases. To evaluate the result of classification use APER (Apparent Error Rate) calculation. Based on this analysis, classification of hypertension using binary logistic regression method obtained APER=27,4648% and 72,5352% of accuracy, while C4.5 algorithm obtained APER=35,9155% and 64,0845% of accuracy. In two different test proportion was found that there were significant differences of the two methods.

Keywords : Types of Hypertension, Classification, C4.5 Algorithm, Binary Logistic Regression, APER

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sampai saat ini, hipertensi atau penyakit darah tinggi masih menjadi masalah utama di dunia, baik di negara maju maupun di negara berkembang. Menurut WHO (2013), penyakit yang dijuluki *the silent killer of death* ini merupakan penyebab kematian nomor satu di dunia. Berdasarkan penelitian yang telah dilakukan, terdapat sekitar sembilan juta jiwa penduduk dunia meninggal setiap tahun karena hipertensi. Sedangkan di Indonesia sendiri, menurut Riset Kesehatan Dasar (2013) bahwa prevelensi hipertensi di Indonesia terbilang masih cukup tinggi yaitu sebesar 25,8 %.

WHO (2013) mendefinisikan hipertensi (tekanan darah tinggi) sebagai peningkatan tekanan darah sistolik lebih dari 140 mmHg dan atau tekanan darah diastolik lebih dari 90 mmHg pada dua kali pengukuran dengan selang waktu lima menit dalam keadaan cukup istirahat/tenang. Peningkatan tekanan darah yang berlangsung dalam jangka waktu lama dan tidak terkendali dapat menimbulkan kerusakan pada organ lain, seperti kerusakan pada ginjal, jantung ataupun otak (stroke).

Menurut Mansjoer (2001), hipertensi sendiri tidak menunjukkan gejala tertentu. Terdapat sekitar 95% kasus hipertensi yang tidak diketahui penyebabnya, sedangkan sisanya ditimbulkan akibat adanya penyakit lain seperti penyakit jantung koroner, gangguan fungsi ginjal, dan gangguan fungsi kognitif atau stroke. Tidak jarang hipertensi ditemukan secara tidak sengaja saat pemeriksaan

kesehatan rutin atau datang dengan keluhan lain. Bahkan terkadang gejala paling parah yang dirasakan adalah ketika tekanan darah sudah sangat tinggi.

Menurut Kementerian Kesehatan (2013), faktor risiko hipertensi antara lain umur, jenis kelamin, riwayat keluarga, genetik (faktor yang tidak dapat diubah), serta faktor yang dapat diubah seperti kebiasaan merokok, konsumsi garam, konsumsi lemak jenuh, penggunaan jelantah, obesitas, kebiasaan minum minuman beralkohol, stres, kurang aktivitas fisik, dan faktor lainnya. Berdasarkan penyebabnya, hipertensi dibagi menjadi dua golongan, yaitu hipertensi primer dan hipertensi sekunder. Hipertensi primer adalah suatu kondisi dimana terjadi tekanan darah tinggi yang tidak diketahui penyebabnya secara pasti. Sedangkan hipertensi sekunder merupakan suatu kondisi terjadinya tekanan darah tinggi yang penyebabnya secara spesifik diketahui seperti adanya penyakit lain. Faktor-faktor penyebab timbulnya kedua jenis hipertensi tersebut hampir tidak bisa dibedakan, karena faktor penyebab hipertensi primer bisa jadi terjadi pada hipertensi sekunder dan atau sebaliknya. Berdasarkan faktor-faktor di atas, akan diidentifikasi faktor-faktor apa saja yang memungkinkan menjadi penyebab terjadinya kedua jenis penyakit hipertensi tersebut.

Berdasarkan data UPT Puskesmas Ponjong I tahun 2015, penyakit hipertensi primer sepanjang tahun 2014, berada diurutan pertama pada sepuluh besar penyakit yang banyak diderita oleh masyarakat, yaitu sebanyak 3.112 jiwa. Sekitar 90 % dari jumlah tersebut yaitu sebanyak 2.799 jiwa, diderita oleh mereka yang berusia lebih dari 60 tahun. Berdasarkan jumlah tersebut, sebagian besar tidak diketahui penyebabnya secara pasti.

Statistika memiliki beberapa metode yang dapat digunakan untuk mengidentifikasi faktor-faktor apa saja yang menyebabkan timbulnya penyakit hipertensi, diantaranya metode regresi logistik dan algoritma C4.5. Hosmer dan Lemeshow (2000) mengatakan bahwa metode regresi logistik adalah suatu metode analisis statistika yang mendeskripsikan hubungan antara variabel respon yang memiliki dua kategori atau lebih dengan satu atau lebih variabel prediktor. Salah satu model regresi logistik adalah regresi logistik biner. Model regresi logistik biner merupakan metode regresi logistik yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik.

Algoritma C4.5 adalah salah satu metode klasifikasi dari *data mining* yang digunakan untuk mengkonstruksikan pohon keputusan (*decision tree*). Menurut Prasetyo (2014), pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Menurut Witten *et al*., (2011), Algoritma C4.5 merupakan pengembangan dari ID3 yang mampu mengatasi nilai yang hilang (*missing value*), mengatasi data bertipe kontinu, dan melakukan pemangkasan pohon (*prunning trees*). Selain itu, dengan menggunakan Algoritma C4.5 dapat diketahui pula nilai ketepatan klasifikasi.

Sehubungan dengan tugas akhir ini, kedua metode digunakan karena keduanya dapat mengatasi data yang bertipe kategorik. Selain itu regresi logistik biner digunakan untuk mengidentifikasi dua variabel respon bertipe kategorik

dalam tugas akhir ini variabel hipertensi primer dan sekunder. Sementara algoritma C4.5 memiliki perhitungan sederhana dalam mengklasifikasikan jenis pernyakit hipertensi dan dapat mengatasi data bertipe kontinu. Variabel yang diduga mempengaruhi jenis penyakit hipertensi primer dan hipertensi sekunder diantaranya variabel jenis kelamin, umur, tekanan darah sistolik, tekanan darah diastolik, riwayat berobat, serta penyakit lain. Variabel-variabel tersebut akan diuji dan dianalisis menggunakan metode regresi logistik biner dan metode algoritma C4.5, yang selanjutnya kedua metode tersebut akan dibandingkan ketepatan klasifikasinya berdasarkan nilai akurasi masing-masing metode.

1.2 Perumusan Masalah

Permasalahan yang akan dibahas dalam tugas akhir ini adalah menganalisis serta memodelkan faktor-faktor yang mempengaruhi jenis penyakit hipertensi menggunakan metode regresi logistik biner dan Algoritma C4.5 yang selanjutnya kedua metode dibandingkan ketepatan klasifikasinya.

1.3 Pembatasan Masalah

Permasalahan pada tugas akhir ini dibatasi untuk data pasien yang menderita penyakit hipertensi di UPT Puskesmas Ponjong I selama bulan Oktober-November 2015. Metode yang digunakan dalam tugas akhir ini adalah metode regresi logistik biner dan Algoritma C4.5

.

1.4 Tujuan

Tujuan dari penulisan tugas akhir ini adalah:

1. Menentukan faktor-faktor yang mempengaruhi terjadinya jenis penyakit hipertensi di UPT Puskesmas Ponjong I
2. Mendapatkan permodelan penyebab terjadinya jenis penyakit hipertensi menggunakan metode regresi logistik biner serta mengukur ketepatan klasifikasinya
3. Membentuk pohon klasifikasi menggunakan metode Algoritma C4.5 dan mengukur ketepatan klasifikasinya
4. Membandingkan ketepatan klasifikasi antara metode regresi logistik biner dengan metode Algoritma C4.5.