



Model *Data Mining* sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik *Decision Tree*

Ari Muzakir¹, Rika Anisa Wulandari²

^{1,2}Jurusan Informatika, F. ILKOM, Universitas Bina Darma Palembang

Email: ¹arimuzakir@binadarma.ac.id, ²ranisa.wulandari@yahoo.com

Abstrak

Prevalensi hipertensi pada wanita hamil terjadi sebanyak 1.062 kasus (12,7%). Dari 1062 kasus ibu hamil dengan hipertensi, ditemukan 125 kasus (11,8%) yang telah didiagnosis dengan hipertensi oleh tenaga kesehatan. RSIA YK Madira Palembang sebagai pusat kesehatan harus mengembangkan metode yang dapat memprediksi risiko tinggi ibu hamil dengan hipertensi dari data hasil pemeriksaan kehamilan. Dengan memanfaatkan sumber data yang terdiri dari data perawatan antenatal, diterapkan teknik *data mining* dengan algoritma *decision tree* C4.5, berdasarkan *Knowledge Discovery in Database* (KDD). Sehingga akan ditemukan pengetahuan, informasi, dan pola tersembunyi dari data pelayanan antenatal, yang merupakan prediksi hipertensi pada kehamilan. Metode yang digunakan yaitu Algoritma C4.5. Setelah mendapatkan *decision tree* dan *rules* yang dapat memprediksi penyakit hipertensi dalam kehamilan, dilakukan evaluasi dengan *supplied test set* menggunakan WEKA dihasilkan kesalahan (*error*) 7.3427% dan tingkat akurasi 92.6573%. *Data training* yang berjumlah 286 *instances*, hal ini menunjukkan bahwa terdapat 265 *instances* yang akurat dan 21 *instances* yang *error* atau prediksinya salah.

Kata Kunci: *Data mining, Decision tree, C4.5 algorithms, Prediction, Pregnancy*

1. PENDAHULUAN

Perkembangan teknologi informasi saat ini sangat berguna bagi semua kalangan masyarakat. Sekarang ini teknologi informasi telah menjadi salah satu kebutuhan yang sangat penting dalam kehidupan sehari-hari. Teknologi informasi dapat mempermudah manusia dalam mengerjakan pekerjaannya. Sehingga, teknologi informasi dapat diimplementasikan dalam berbagai bidang, termasuk dunia kedokteran dan kesehatan. Kesehatan merupakan aspek terpenting didalam setiap kehidupan, khususnya bagi ibu hamil. Hipertensi merupakan salah satu masalah kesehatan yang sering muncul selama kehamilan dan dapat menimbulkan komplikasi pada 2–3% kehamilan. Hipertensi pada kehamilan juga dapat berlanjut menjadi *preeklamsia* dan *eklamsia* yang dapat menyebabkan kematian pada ibu maupun janin [4].

Pada penelitian yang dilakukan oleh Riset Kesehatan Dasar tentang Prevalensi Hipertensi dalam Kehamilan di Indonesia, mencatat bahwa ditemukannya sebanyak 8.341 kasus (1,51%) ibu hamil dari semua sampel perempuan yang berusia 15–54 tahun. Didapatkan prevalensi hipertensi pada ibu hamil sebesar 1.062 kasus (12,7%). Dari 1062 kasus ibu hamil dengan hipertensi, ditemukan 125 kasus (11,8%) yang pernah didiagnosis menderita hipertensi oleh petugas kesehatan [3].

Hal ini menunjukkan bahwa penyakit hipertensi merupakan salah satu masalah kesehatan yang sering muncul selama kehamilan dan dapat menimbulkan komplikasi pada 2–3% kehamilan.

Salah satu metode yang dapat digunakan untuk memprediksi penyakit hipertensi adalah dengan menggunakan *data mining*. *Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [2]. Metode untuk menganalisis data dalam penerapan *data mining* ini adalah *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahapan, yaitu *Cleaning and Integration*, *Selection and Transformation*, *Data mining*, dan *Evaluation and Interpretation* [1].

Adapun teknik yang digunakan dalam *data mining* adalah *decision tree* (pohon keputusan) dengan menggunakan algoritma C4.5. Dengan menggunakan teknik *decision tree* menggunakan algoritma C4.5, akan menghasilkan pola atau *pattern*, yaitu berupa *rules* yang dapat memprediksi penyakit hipertensi dalam kehamilan.

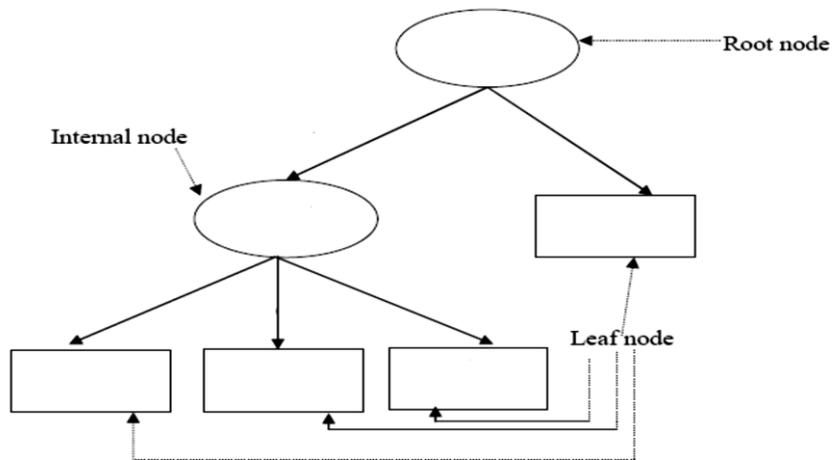
2. METODE

2.3 *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD). Adapun teknik yang digunakan pada proses *data mining* adalah dengan teknik *decision tree* dengan menggunakan algoritma C4.5.

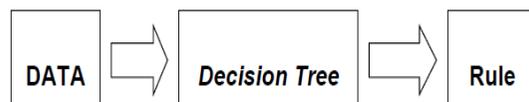
2.3.1 *Decision Tree*

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk basis data seperti *Structure Query Language* (SQL) untuk mencari *record* pada data tertentu. Sebuah *decision tree* adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pada *decision tree* setiap simpul daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian *record* yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat. Berikut adalah struktur *decision tree* seperti yang ditunjukkan pada Gambar 4.



Gambar 4. Struktur *decision tree*

Berikut adalah konsep *decision tree* seperti yang ditunjukkan pada Gambar 5.



Gambar 5. Konsep *decision tree*

Ada beberapa konsep dalam *decision tree*, antara lain:

- 1) Data dinyatakan dalam bentuk Tabel dengan *atribut* dan *record*.
- 2) *Atribut* menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi *per-item* data yang disebut dengan target atribut.
- 3) Atribut memiliki nilai-nilai yang dinamakan dengan *instance*.

2.3.2 Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang digunakan dalam *decision tree*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang dikembangkan oleh J.Ross Quinlan.

Secara umum, langkah-langkah algoritma C4.5 dalam membangun *decision tree* adalah sebagai berikut.

- 1) Pilih atribut sebagai akar.
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang.
- 4) Ulangi proses untuk setiap proses cabang sampai semua kasus pada cabang memiliki kelas yang sama.

C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area *machine learning* yang memiliki beberapa perbaikan dari algoritma sebelumnya, ID3, yaitu dalam hal metode pemangkasan (*pruning*). Adapun perbaikannya adalah sebagai berikut:

- 1) Algoritma C4.5 menghitung *gain ratio* untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul. Penggunaan *gain ratio* ini memperbaiki kelemahan dari ID3 yang menggunakan *information gain*.
- 2) Pemangkasan dapat dilakukan pada saat pembangunan pohon (*tree*) ataupun pada saat proses pembangunan pohon selesai.
- 3) Mampu menangani *continues attribute*.
- 4) Mampu menangani *missing data*.
- 5) Mampu membangkitkan *rule* dari sebuah pohon.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti tertera dalam persamaan 1 berikut.

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

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi atribut A

|S_i| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut ini.

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad 2)$$

Keterangan:

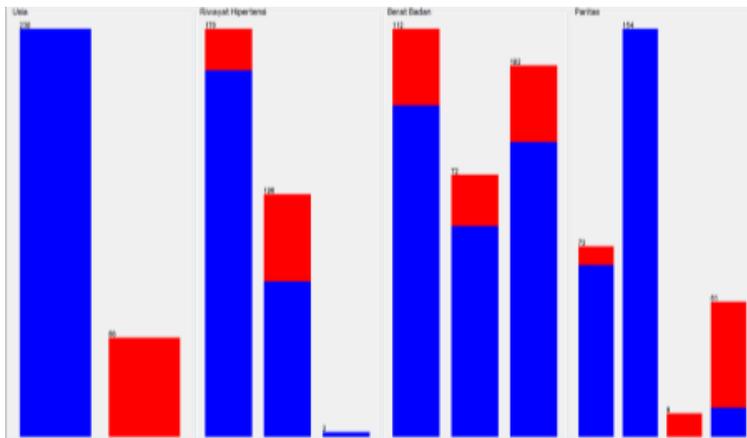
S = Himpunan Kasus

A = Fitur

n = Jumlah partisi S

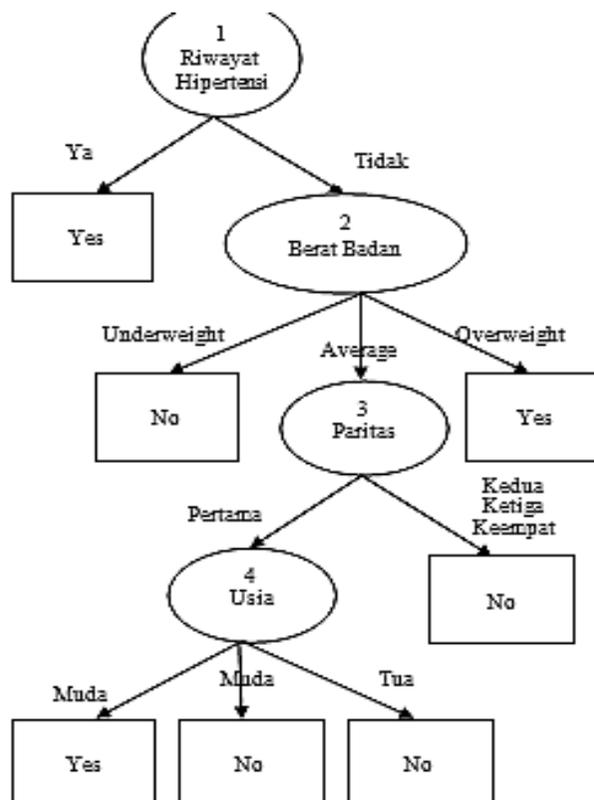
P_i = Proporsi dari S_i terhadap S

Adapun atribut-atribut yang terdapat pada *data training* seperti yang ditunjukkan pada Gambar 6.



Gambar 6. Atribut pada data training

Dengan menggunakan algoritma C4.5 pada data training, akan terbentuk decision tree (pohon keputusan) seperti yang ditunjukkan pada Gambar 7.



Gambar 7. Decision tree

3. HASIL DAN PEMBAHASAN

Dalam teknik *decision tree*, pola atau informasi yang dihasilkan dari proses *data mining* adalah berupa *rules* yang didapat dari *decision tree* yang telah dibangun. Berikut ini adalah *rules* yang dihasilkan dari *decision tree* seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Rules

Rules	
1	<i>IF</i> Riwayat Hipertensi = Ya <i>THEN</i> Hipertensi = <i>Yes</i>
2	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Underweight</i> <i>THEN</i> Hipertensi = <i>No</i>
3	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Overweight</i> <i>THEN</i> Hipertensi = <i>Yes</i>
4	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Average</i> ^ Paritas = Kedua ^ Paritas = Ketiga ^ Paritas = Keempat <i>THEN</i> Hipertensi = <i>No</i>
5	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Average</i> ^ Paritas = Pertama ^ Usia = Muda <i>THEN</i> Hipertensi = <i>Yes</i>
6	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Average</i> ^ Paritas = Pertama ^ Usia = Muda <i>THEN</i> Hipertensi = <i>No</i>
7	<i>IF</i> Riwayat Hipertensi = Tidak ^ Berat Badan = <i>Average</i> ^ Paritas = Pertama ^ Usia = Tua <i>THEN</i> Hipertensi = <i>No</i>

Pada *rules* pertama, dapat diartikan bahwa jika seorang ibu hamil memiliki riwayat penyakit hipertensi, maka kemungkinan besar ia akan menderita penyakit hipertensi dalam kehamilan.

Pada *rules* yang kedua, dapat diartikan bahwa jika seorang ibu hamil tidak memiliki penyakit hipertensi dalam kehamilan dan ia memiliki berat badan yang dikategorikan *underweight*, maka kemungkinan besar ia tidak akan menderita penyakit hipertensi dalam kehamilan.

Pada *rules* yang ketiga, dapat diartikan bahwa jika seorang ibu hamil tidak memiliki riwayat penyakit hipertensi namun ia memiliki berat badan yang dikategorikan *overweight*, maka kemungkinan ia bisa menderita penyakit hipertensi dalam kehamilan.

Pada *rules* yang keempat, dapat diartikan bahwa jika seorang ibu hamil tidak memiliki riwayat penyakit hipertensi, berat badannya dikategorikan *average*, dan paritas pada saat kehamilan adalah yang kedua, ketiga, keempat, maka kemungkinan ia tidak menderita penyakit hipertensi dalam kehamilan.

Pada *rules* yang kelima, dapat diartikan bahwa jika seorang ibu hamil tidak memiliki riwayat penyakit hipertensi, berat badannya dikategorikan *average*, paritas pada saat kehamilan adalah yang pertama, dan usianya muda, maka kemungkinan ia bisa menderita penyakit hipertensi dalam kehamilan. Sedangkan pada *rules* yang keenam, memiliki kasus yang sama, namun bisa saja seorang ibu hamil tersebut tidak menderita penyakit hipertensi dalam kehamilannya. Untuk mendapatkan prediksi yang akurat, diperlukan campur tangan seorang pakar.

Pada *rules* yang ketujuh, dapat diartikan bahwa jika seorang ibu hamil tidak memiliki riwayat penyakit hipertensi, berat badannya dikategorikan *average*, paritas pada saat

kehamilan adalah yang pertama, dan usianya tua, maka kemungkinan ia tidak menderita penyakit hipertensi dalam kehamilannya.

Dari rules tersebut, dapat diprediksi apakah seorang ibu hamil dapat menderita penyakit hipertensi atau tidak. Namun, pada tahapan akhir dari proses *Knowledge Discovery in Database* (KDD), perlu dilakukan proses *evaluation/interpretation*. Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Dalam hal ini, proses *evaluation/interpretation* dilakukan dengan menggunakan WEKA. Adapun fitur WEKA yang digunakan adalah *Supplied Test Set*. *Supplied test set* adalah menguji seberapa baik prediksi yang dihasilkan dari *data training*. Sehingga, akan didapatkan hasil *Supplied Test Set* seperti yang ditunjukkan pada Tabel 2 berikut ini.

Tabel 2. Hasil *supplied test set*

<i>Instances</i>	<i>Correctly Classified</i>	<i>Incorrectly Classified</i>
286	265 (92.6573%)	21 (7.3427%)

Hasil dari *Supply Test Set* menunjukkan bahwa terdapat 92.6573% atau 265 *instance correctly classified* (prediksi benar) dan 7.3427% atau 21 *instance incorrectly classified* (prediksi salah) dari *data training* yang digunakan yang berjumlah 286 *instance*.

4. SIMPULAN

Implementasi *data mining* dengan teknik *decision tree* menggunakan algoritma C4.5 dapat menghasilkan informasi berupa prediksi penyakit hipertensi dalam kehamilan, dimana dari *data training* yang digunakan dengan jumlah 286 *instance* dapat dibangun sebuah *decision tree* yang menghasilkan *rules* yang bisa digunakan untuk memprediksi penyakit hipertensi dalam kehamilan. Dari *decision tree* yang dibangun, menunjukkan bahwa atribut yang menjadi faktor pendukung seorang ibu hamil bisa menderita penyakit hipertensi dalam kehamilannya, yaitu berdasarkan usia, berat badan, riwayat hipertensi, dan paritas.

Setelah mendapatkan *decision tree* dan *rules* yang dapat memprediksi penyakit hipertensi dalam kehamilan, dilakukan evaluasi dengan *supplied test set* menggunakan WEKA dihasilkan kesalahan (*error*) 7.3427% dan tingkat akurasi 92.6573%. *Data training* yang berjumlah 286 *instances*, hal ini menunjukkan bahwa terdapat 265 *instances* yang akurat dan 21 *instances* yang *error* atau prediksinya salah.

5. REFERENSI

- [1] Han J dan Kamber M. 2006. *Data mining Concepts and Techniques 2nd Edition*. The Morgan Kaufmann Publisher, San Fransisco.

- [2] Pramudiono, I. 2006. Apa itu *Data Mining*? Dalam <http://datamining.japati.net/cgibin/indodm.cgi>. Diakses tanggal 26 November 2015.
- [3] Sirait, Ana Maria. 2012. Prevalensi Hipertensi pada Kehamilan di Indonesia. *Buletin Penelitian Sistem Kesehatan*, Vol. 15 (2), April 2012: 103–109, Jakarta.
- [4] Yudasmar, Kusuma. 2013. *Hipertensi dalam Kehamilan*. (<http://www.balipost.com>, diakses pada tanggal 17 November 2015).