

ANALISIS PERBANDINGAN ALGORITMA C.45, RANDOM FOREST DENGAN CHAID DECISION TREE UNTUK KLASIFIKASI TINGKAT KECEMASAN IBU HAMIL

Mambang¹⁾, Agus Byna²⁾

¹⁾ STIKES SARI MULIA Banjarmasin

²⁾AKBID SARI MULIA Banjarmasin

Jln. Pramuka No. 02 Banjarmasin, 70123

Email : mbgche@yahoo.co.id¹⁾

Abstrak

Kecemasan dalam menghadapi proses persalinan pada ibu hamil menjelang persalinan dapat mempengaruhi proses persalinan tersebut, banyak faktor yang membuat ibu hamil merasa cemas, faktor-faktor yang mempengaruhi kecemasan ibu hamil dalam menghadapi persalinan tersebut dapat diketahui dari tingkatannya. Tingkat kecemasan dalam menghadapi persalinan dapat di klasifikasi menggunakan pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang berguna untuk mengeksplorasi data untuk menemukan hubungan tersembunyi antar atribut data.

Pada penelitian ini metode yang diusulkan untuk klasifikasi tingkat kecemasan ibu hamil adalah dengan metode Decision Tree dengan membandingkan kinerja algoritma C4.5, RF(Random Forest) dengan CHAID. Algoritma tersebut nantinya akan dibandingkan untuk klasifikasi error dan tingkat akurasi.

Hasil akurasi yang didapatkan menggunakan algoritma pohon keputusan C4.5, RF dengan CHAID yang menggunakan pruning=3 menghasilkan akurasi yang lebih baik yaitu berada pada angka 64% dan 62,67%, sedangkan yang menggunakan pre pruning menghasilkan akurasi lebih rendah.

Kata kunci: Kecemasan, persalinan, pohon, keputusan

1. Pendahuluan

Tingkat kecemasan setiap individu berbeda-beda, begitu juga dengan kecemasan yang dialami oleh ibu hamil menjelang persalinan. Tidak mudah bagi setiap ibu hamil dalam melewati proses persalinan, setiap proses persalinan yang dilalui dapat berbeda-beda yang dirasakan oleh para perempuan bahkan yang sudah pernah melahirkan lebih dari sekali. Pembicaraan di masyarakat juga berperan dalam menimbulkan rasa cemas yang dialami perempuan hamil, walaupun secara alamiah perasaan cemas merupakan proses alamiah yang terjadi pada setiap perempuan yang akan melewati proses persalinan. Perasaan cemas tersebut merupakan

reaksi normal terhadap situasi yang sangat menekan kehidupan seseorang, dan karena itu berlangsung tidak lama. Penting sekali untuk mengingat bahwa kecemasan bisa muncul sendiri atau bergabung dengan gejala-gejala lain dari berbagai gangguan emosi[1]. Faktor-faktor penting dalam persalinan adalah: *power* seperti HIS, kontraksi otot dinding perut, kontraksi diaphragma pelvis atau kekuatan mengejan, ketegangan dan kontraksi ligamentum; *Pasanger* antara lain seperti janin dan plasenta; *Pasage* yakni jalan lahir lunak dan jalan lahir tulang; serta psikis ibu bersalin[2]. Dukungan suami dan anggota keluarga yang lain untuk mendampingi ibu selama bersalin dan kelahiran sangat mempengaruhi psikis ibu bersalin. Dengan menghargai keinginan ibu untuk didampingi dapat membantu kenyamanan ibu. Menurut Stuart [3] Kecemasan atau ansietas adalah kekhawatiran yang tidak jelas dan menyebar, yang berkaitan dengan perasaan tidak pasti dan tidak berdaya. Keadaan emosi ini tidak memiliki objek spesifik. Ansietas dialami secara subjektif dan dikomunikasikan secara interpersonal. Kecemasan yang dialami ibu hamil menjelang persalinan juga berkaitan dengan gangguan non organik, beberapa diantaranya penyakit dan komplikasi tidak semata-mata ditimbulkan oleh gangguan organik namun dapat ditimbulkan atau diperberat oleh gangguan psikologik. Latar belakang timbulnya penyakit dan komplikasi dapat dijumpai dalam berbagai tingkat ketidakmatangan dalam perkembangan emosional dan psikoseksual dalam rangka kesanggupan seseorang dalam menyesuaikan diri dengan situasi tertentu yang sedang dihadapi, dalam hal ini khususnya kehamilan, persalinan dan nifas[2]. Metode klasifikasi *Machine Learning* telah lama digunakan di bidang data mining dan banyak bidang lain dari ilmu komputer. Sebagai algoritma klasifikasi yang melibatkan metode dari statistik, kecerdasan buatan dan manajemen database, juga dapat di kategorikan sebagai kunci elemen dalam interpretasi data dan visualisasi data[4]. Decision tree atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplor data dan menemukan hubungan yang tersembunyi[5]. Metode decision tree memiliki beberapa algoritma yang berbeda

cara perhitungannya untuk membuat suatu pohon keputusan. Random forest merupakan klasifikasi yang terdiri dari beberapa pohon keputusan. Setiap pohon keputusan dibangun dengan menggunakan vektor acak. Pendekatan umum yang digunakan untuk menyisipkan vektor acak dalam pembentukan pohon[6]. CHAID (Chi-Squared Automatic Interaction Detector)[7] adalah metode klasifikasi yang digunakan untuk menemukan hubungan antara variabel menggunakan *chi-square* berbasis kriteria untuk menentukan perpecahan multi-level. Manfaat penggunaan CHAID adalah bahwa secara visual lebih ekspresif dan lebih mudah ditafsirkan dibandingkan dengan jenis lain dari pengklasifikasi.

Dalam penelitian ini model klasifikasi yang digunakan untuk menentukan apakah umur, tingkat pendidikan, pekerjaan dan paritas ibu hamil berpengaruh dengan kecemasan dan membentuk kategori tingkat kecemasan dari ibu hamil.

Tujuan dari penelitian ini yaitu untuk menerapkan metode klasifikasi pohon keputusan dengan mengklasifikasi tingkat kecemasan ibu hamil berdasarkan umur, tingkat pendidikan, pekerjaan dan paritas dari ibu hamil. Sehingga dapat membantu meningkatkan hasil akurasi pada proses klasifikasi sehingga memperoleh klasifikasi tingkat kecemasan ibu hamil menjadi 3 kategori, yaitu kecemasan berat, kecemasan sedang dan kecemasan ringan.

Dataset yang digunakan dalam penelitian ini merupakan data kesehatan pada Puskesmas Gambut, untuk menentukan tingkat kecemasan ibu hamil di daerah Gambut, yaitu salah satu kecamatan dari kabupaten Banjar provinsi Kalimantan Selatan. Pengklasifikasian tersebut dilakukan berdasarkan data yang diperoleh dari tenaga kesehatan yaitu salah satu Bidan Desa Puskesmas Gambut.

2. Pembahasan

2.1 Pengumpulan dan pengolahan data awal

Penelitian ini dilakukan di di Wilayah Kerja Puskesmas Gambut, hal ini berhubungan dengan Luas Wilayah : 160 KM². Topografi : berada di ketinggian $\pm 0,375$ meter dari permukaan laut, dengan keadaan tanah berbentuk rawa dataran rendah.

Pengolahan awal data yang mengidentifikasi kecemasan ibu hamil menjelang persalinan di wilayah kerja Puskesmas Gambut berdasarkan penelitian tentang kecemasan ibu hamil diketahui beberapa faktor pada kecemasan ibu hamil menjelang persalinan, yaitu:

a. Umur ibu hamil

Usia seorang wanita pada saat hamil sebaiknya tidak terlalu muda dan tidak terlalu tua. Umur yang kurang dari 20 tahun atau lebih dari 35 tahun, berisiko tinggi untuk melahirkan. Kesiapan seorang perempuan untuk hamil harus siap fisik, emosi, psikologi, sosial dan ekonomi[8]. Umur ibu hamil dikategorikan menjadi 3 yaitu:

≥ 35 tahun

20 – 34 tahun

≤ 19 tahun

b. Paritas ibu hamil

Paritas adalah jumlah janin dengan berat badan lebih dari atau sama dengan 500 gram yang pernah dilahirkan hidup maupun mati. Bila berat badan tak diketahui maka dipakai umur kehamilan, yaitu 24 minggu [9]. Penggolongan paritas bagi ibu yang masih hamil atau pernah hamil berdasarkan jumlahnya menurut WHO, yaitu :

Primigravida adalah wanita hamil untuk pertama kali

Multigravida adalah wanita yang pernah hamil beberapa kali, dimana kehamilan tersebut tidak lebih dari 5 kali

Grandemultigravida adalah wanita yang pernah hamil lebih dari 5 kali

c. Pekerjaan

Pekerjaan ibu hamil dikategorikan menjadi 2 yaitu:

-Bekerja

-Tidak Bekerja

d. Pendidikan

Pendidikan ibu hamil dikategorikan menjadi 3 yaitu:

Pendidikan dasar, Pendidikan menengah dan Pendidikan tinggi

Menurut UU Sisdiknas No. 20 Tahun 2005, Jenjang pendidikan formal terdiri atas pendidikan dasar, pendidikan menengah dan pendidikan tinggi.

a) Pendidikan Dasar, terdiri dari :

1) Sekolah Dasar/Madrasah Ibtidaiyah

2) SMP/MTs

b) Pendidikan Menengah, terdiri dari :

1) SMA dan MA

2) SMK dan MAK

c) Pendidikan Tinggi, terdiri dari :

1) Akademi

2) Institut

3) Sekolah tinggi

4) Universitas

2.2 Metode

Metode yang diusulkan dalam penelitian ini adalah klasifikasi tingkat kecemasan ibu hamil dengan algoritma Random Forest dan CHAID

2.2.1 Random Forest (RF)

Random Forest (RF) adalah klasifikasi yang terdiri dari beberapa pohon keputusan. Setiap pohon keputusan dibangun dengan menggunakan vektor acak. Pendekatan umum yang digunakan untuk menyisipkan vektor acak dalam pembentukan pohon adalah memilih nilai F acak, seperti F atribut (fitur) masukan untuk dibagi pada setiap node di pohon keputusan yang akan dibentuk. Dengan memilih nilai acak F maka seharusnya tidak harus memeriksa semua atribut yang ada, hanya melihat F dipilih atribut. Parameter yang digunakan untuk mengatur kekuatan RF dalam pemilihan nilai F dan jumlah tree yang akan dibangun [6]. Jika nilai F terlalu kecil, maka pohon tersebut memiliki kecenderungan untuk memiliki korelasi yang sangat kecil, dan itu berlaku sama untuk sebaliknya. Oleh karena itu, nilai F dapat ditentukan dengan rumus:

$$F = \text{Log}_2(M + 1) \quad (1)$$

Dari rumus tersebut di mana M adalah jumlah total fitur. Selain pemilihan atribut, juga dilakukan secara acak ketika memilih set pelatihan. *Bootstrap aggregating* adalah teknik yang dapat digunakan untuk membentuk sampel bootstrap. Setiap pohon keputusan dibangun dengan menggunakan sampel bootstrap dari data dan kandidat atribut untuk dibagi di setiap node yang berasal dari himpunan atribut acak dari hasil data yang dipilih dan disimpan. Berikut adalah Tahapan algoritma RF:

1. Pilih nilai n yang menunjukkan jumlah pohon
2. Menghasilkan sampel n bootstrap dengan teknik training set.
3. Pada setiap node pada tree, pilih nilai F yang diperoleh dari persamaan (1).
4. Ambil set sebanyak atribut F yang akan menjadi kandidat atribut untuk membelah setiap node. Atribut yang menjadi simpul berikutnya ditentukan berdasarkan kriteria tertentu (berdasarkan algoritma pohon keputusan yang dipilih). Selama proses pembentukan pohon, nilai F adalah konstan.
5. RF terus terbentuk tanpa pemangkasan apapun. Hal ini ditunjukkan untuk menghilangkan bias dalam persentase hasil prediksi.
6. Hasil prediksi yang diperoleh dari model (frekuensi yang paling sering muncul) dari masing-masing pohon keputusan di RF.

2.2.2 CHAID (Chi-Squared Automatic Interaction Detector)

CHAID adalah sebuah metode untuk mengklasifikasikan data kategori dimana tujuan prosedurnya adalah untuk membagi rangkaian data menjadi subgrup-subgrup berdasarkan pada variabel dependennya [7]. Hasil dari pengklasifikasian dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Menurut [9], CHAID merupakan suatu teknik iteratif yang menguji satu-

persatu variabel independen yang digunakan dalam klasifikasi, dan menyusunnya berdasarkan pada tingkat signifikansi statistik chi-square terhadap variabel dependennya.

Algoritma CHAID digunakan untuk melakukan pemisahan dan penggabungan kategori-kategori dalam variabel yang dipakai dalam analisisnya. Secara garis besar algoritma ini dapat dibagi menjadi tiga tahap, yaitu Penggabungan (*merging*), Pemisahan (*splitting*) dan Penghentian (*stopping*). Diagram pohon dimulai dari root node (node akar) melalui tiga tahap tersebut pada setiap simpul (node) yang terbentuk dan secara berulang.

Tahap pertama algoritma CHAID adalah penggabungan. Pada tahap ini akan diperiksa signifikansi dari masing-masing kategori variabel independen terhadap variabel dependen. Tahap penggabungan untuk setiap independen dalam menggabungkan kategori-kategori nonsignifikan adalah sebagai berikut:

A. Tahap Penggabungan

1. Bentuk tabel kontingensi dua arah untuk masing-masing variabel independen dengan variabel dependennya.
2. Hitung statistik chi-square untuk setiap pasang kategori yang dapat dipilih untuk digabungkan menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi $2 \times j$ yang dibentuk oleh sepasang kategori tersebut dengan variabel dependennya yang mempunyai sebanyak kategori. Langkah uji chi square adalah sebagai berikut:

- a. Perumusan hipotesisnya
- b. H_0 : tidak terdapat hubungan antara kategori i pada variabel independen dan kategori j pada variabel dependen
- c. H_1 : terdapat hubungan kategori i pada variabel independen dan kategori j pada variabel dependen
- d. Besaran-besaran yang diperlukan Menghitung E_{ij}
- e. Statistik uji

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

- f. Kriteria Pengujian H_0 di tolak, jika $\chi^2 \geq \chi^2_{\alpha; (b-1)(k-1)}$ (3)

- g. Kesimpulan Penaksiran dari H_0 diterima atau ditolak.

3. Untuk masing-masing nilai chi-square berpasangan, hitung p-value berpasangan bersamaan. Diantara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu

pasangan yang mempunyai nilai chi-square berpasangan terkecil dan pvalue terbesar) menjadi sebuah kategori tunggal dan kemudian lanjutkan ke langkah nomor 4.

4. Periksa kembali kesignifikansian kategori baru setelah digabungkan dengan kategori lainnya dalam variabel independen. Jika masih ada pasangan yang belum signifikan, ulangi langkah nomor 3. Jika semua signifikan lanjutkan ke langkah selanjutnya.

5. Hitung p-value terkoreksi Bonferroni didasarkan pada tabel yang telah digabung

B. Tahap Pemisahan

Pilih variabel independen yang memiliki nilai χ^2 hitung signifikan terbesar. Pembagian kelompok atau pemisahan kelompok dilakukan dengan variabel kelompok tersebut (Gunakan masing – masing kategori variabel independen tersebut, yang telah digabung secara optimal, untuk menentukan sub pembagian dari kelompok induk menjadi sub kelompok yang baru). Jika tidak ada variabel independen dengan nilai χ^2 hitung yang signifikan pembagian kelompok atau pemisahan kelompok tidak dapat dilakukan.

C. Tahap Penghentian

Kembali ke langkah penggabungan untuk menganalisis sub kelompok berikutnya. Hentikan ketika semua sub kelompok telah dianalisis dan juga telah berisi pengamatan dengan jumlah sedikit.

2.2.3 C4.5 Decision Tree

Tahapan Algoritma Decision Tree C4.5 [11]:

- 1) Menyiapkan data training.
- 2) Menentukan akar dari pohon dihitung dengan menggunakan persamaan 1:

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi \tag{4}$$

- 3) Hitung nilai Gain dengan menggunakan persamaan 2:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{5}$$

- 4) Ulangi langkah ke-2 hingga semua tupel terpartisi
- 5) Proses partisi pohon keputusan akan berhenti saat:
 - Semua tupel dalam node N mendapat kelas yang sama
 - Tidak ada atribut di dalam tupel yang dipartisi lagi
 - Tidak ada tupel di dalam cabang yang kosong

2.3 Pengujian Model

Pada pengujian model menggunakan RF, C.45 dan CHAID yang di simulasikan menggunakan rapidminer di dapat hasil diagram pohon keputusan sebagai berikut:

2.3.1 Pengujian RF

Dataset yang digunakan untuk pengujian dengan algoritma RF untuk menghasilkan tingkat akurasi menggunakan 3 kriteria, yaitu Gini Index, Gain Rasio dan Information gain dengan jumlah prepruning yang

telah di tentukan, dalam penelitian ini menggunakan 0 dan 3 jumlah prepruning, berikut akurasi yang didapatkan dapat dilihat pada table 1.

Tabel 1. Tabel hasil pengujian akurasi

Algoritma	Criterion	No Pre Pruning	Akurasi
RF	Gain Ratio	No	64,00%
RF	Gain Ratio	Yes	59.00%
RF	Information Gain	No	63,33%
RF	Information Gain	Yes	56.33%
RF	Gini Index	Yes	42.67%
CHAID	Gain Rasio	No	62,67%
CHAID	Gain Rasio	Yes	41,33%
C4.5	Gain Rasio	No	64%
C4.5	Gain Rasio	Yes	35,67%

Dari hasil pengujian menggunakan algoritma RF, banyaknya tree =10, nilai confidence=0,25, minimal gain =0,1 , minimal leaf size=2 minimal size for split=4.

Kriteria gain rasio dengan jumlah prepruning 3 mendapatkan hasil terbaik dari pengujian table diatas dengan akurasi sebesar 64,00%. Sedangkan pengujian yang menggunakan *no pre pruning* mendapatkan hasil akurasi yang rendah.

2.4 Hasil

Pada pengujian model menggunakan C4.5, RF dan CHAID diatas di dapat tingkat akurasi yang terbaik dengan pruning=3 sebagai berikut:

Tabel 2. Tabel hasil pengujian akurasi

Algoritma	Criterion	Akurasi
RF	Gain Ratio	64,00%
CHAID	Information Gain	62,67%
C4.5	Gain Ratio	64.00%

2.5 Evaluasi

Pengukuran yang digunakan untuk memvalidasi hasil pengujian menggunakan RMSE dapat dilihat pada table berikut ini:

Tabel 3. Tabel hasil Pengukuran

Algoritma	Criterion	RMSE
RF	Gain Ratio	0.584
RF	Gain Ratio	0.538
RF	Information Gain	0.588

RF	Information Gain	0.583
RF	Gini Index	0.616
CHAID	Gain Rasio	0.548
CHAID	Gain Rasio	0.591
C4.5	Gain Rasio	0.588
C4.5	Gain Rasio	0.638

3. Kesimpulan

Dari hasil pengujian dan validasi hasil akurasi yang didapatkan menggunakan algoritma pohon keputusan C4.5, RF dengan CHAID yang menggunakan pruning=3 menghasilkan akurasi yang lebih baik yaitu berada pada angka 64% dan 62%, sedangkan yang menggunakan pre pruning menghasilkan akurasi lebih rendah.

Pada pengujian training dan testing yang dilakukan dapat disimpulkan bahwa metode Decision tree RF, C4.5 dan CHAID dapat diterapkan. Setiap akurasi akan berbeda nilainya tergantung dari pruning atau tidak menggunakan prepruning. Akurasi yang lebih baik didapat dari pruning= 3. Kriteria yang digunakan juga dapat mempengaruhi hasil akurasi dan pengukuran untuk memvalidasi dataset.

Berikut tabel perbandingannya:

Tabel 4. Tabel hasil perbandingan

Algoritma	Criterion	RMSE	Akurasi
RF	Gain Ratio	0.584	64,00%
RF	Gain Ratio	0.538	59.00%
RF	Information Gain	0.588	63,33%
RF	Information Gain	0.583	56.33%
RF	Gini Index	0.616	42.67%
CHAID	Gain Rasio	0.548	62,67%
CHAID	Gain Rasio	0.591	41,33%
C4.5	Gain Rasio	0.588	64.00%
C4.5	Gain Rasio	0.638	35,67%

4. Saran

Saran yang dapat diberikan dalam pengujian menggunakan metode Decision tree Random forest(RF), C4.5 dengan CHAID adalah sebagai berikut:

1. Tingkat akurasi yang didapatkan hanya berada pada kisaran 64%, nilai akurasi dapat ditingkatkan lagi dengan mengoptimasi dataset yang digunakan untuk mengklasifikasi tingkat kecemasan ibu hamil menjelang persalinan.
2. Dataset yang digunakan dapat diuji kembali dengan metode dan algoritma yang berbeda, sehingga dapat menghasilkan nilai akurasi dan pengukuran yang lebih baik lagi.

Daftar Pustaka

- [1] Savitri, R.. 2003. Kecemasan Bagaimana Mengatasi Penyebabnya. Jakarta: Pustaka Populer Obor.
- [2] Yeyeh. 2009. Asuhan Kebidanan II (Persalinan). Jakarta: Trans Info Media
- [3] Stuart, G.W. dan Sundeen, S.J. 1998, Buku Saku Keperawatan Jiwa, Jakarta, EGC.
- [4] Berry, M. J., & Linoff, G. S.2004. Data Mining Techniques for marketing, sale, customer relationship management. Wiley Publishing, Inc.
- [5] Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques., 2005.
- [6] Ron Appel, Thomas Fuchs, Piotr Dollar and Pietro Perona, Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28, "Quickly Boosting Decision Trees - Pruning Underachieving Features Earl.
- [7] Lehmann, T. & Eherler, D. 2001. Responder Profiling with CHAID and Dependency Analysis..
- [8] Ruswana.2006. Tanya Jawab Seputar Kehamilan. Bhuana Ilmu Populer. Jakarta
- [9] Gallagher, C.A. 2000. An Iterative Approach to Classification Analysis.
- [10] Siswosudarmo, R., 2008. Obstetri Fisiologi Yogyakarta: Pustaka Cendekia
- [11]Wu, X., & Kumar, V. 2009. The Top Ten Algorithm In Data Mining. London: Taylor & Francis Group

Biodata Penulis

Mambang, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK INDONESIA Banjarmasin, lulus tahun 2008. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas Dian Nuswantoro Semarang, lulus tahun 2013. Saat ini menjadi Dosen di STIKES Sari Mulia Banjarmasin.

Agus Byna, memperoleh gelar Sarjana Komputer (S.Kom), STMIK Bandung, lulus tahun 2012. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas Dian Nuswantoro Semarang, lulus tahun 2015. Saat ini menjadi Dosen di Akademi Kebidanan Sari Mulia Banjarmasin.

