



Konferensi Internasional tentang Tren yang Muncul dalam Teknik, Sains, dan Teknologi (ICETEST - 2015)

Kajian C.5 Algoritma Klasifikasi Pohon Keputusan untuk Prediksi Risiko Selama Kehamilan

Lakshmi.B.Na*, Dr.Indumathi.T.Sb, Dr.Nandini Ravic

^a Sarjana Ph.D, Pusat Penelitian PG, VIAT, VTU RRC, Muddenahalli, Chikkaballapura-562101, Karnataka, Profesor dan Koordinator

^b India, Pusat Penelitian PG, VIAT, VTU RRC, Muddenahalli, Chikkaballapura-562101, Karnataka, India Dr.Nandini Ravi, MBBS, MD(Obs&Gyn), Panti

^c Jompo Dhruva, Hosakote, Bangalore, Karnataka, India

Abstrak

Komplikasi saat hamil ternyata menjadi masalah utama bagi wanita di era sekarang. Wanita hamil harus dilindungi dari komplikasi yang timbul selama masa kehamilan, suatu tahap di mana setiap wanita mengalami banyak perubahan fisiologis, terkadang menyebabkan masalah kesehatan yang parah yang menyebabkan kematian ibu dan janin. Intervensi teknologi di bidang diagnosis medis sebagian besar dapat membantu menemukan solusi untuk masalah ini guna melindungi ibu hamil, sehingga pada gilirannya mengurangi kematian ibu dan janin secara luas. Metode Klasifikasi Pohon Keputusan adalah metode yang populer digunakan yang algoritmanya paling cocok dalam diagnosis medis. Algoritma C4.5 Decision Tree adalah salah satu classifier yang populer dan efektif digunakan untuk klasifikasi data kehamilan dalam penelitian ini. Tujuan utama dari makalah ini adalah untuk menunjukkan pentingnya standarisasi parameter yang dipilih untuk pengumpulan data dalam studi, membandingkan hasil yang diperoleh dari pengklasifikasi C4.5 pada dataset yang tidak standar dan standar dan menganalisis kinerja algoritma C4.5 di hal akurasi prediksi ketika diterapkan pada database yang dibuat dari data kehamilan yang dikumpulkan dan distandarisasi.

© 2016 The Authors. Diterbitkan oleh Elsevier Ltd. Ini adalah artikel akses terbuka di bawah lisensi CC BY-NC-ND

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review di bawah tanggung jawab panitia penyelenggara ICETEST – 2015.

Peer-review di bawah tanggung jawab panitia penyelenggara ICETEST – 2015

Kata kunci: kehamilan; model pembelajaran; Algoritma klasifikasi C4.5; komplikasi kehamilan; kelainan; kumpulan data standar; kumpulan data yang tidak terstandarisasi.

1. Perkenalan

Komplikasi kehamilan yang menyebabkan peningkatan angka kematian ibu dan janin telah membuka cakrawala baru bagi para peneliti untuk bekerja dan memberikan solusi yang layak untuk melindungi ibu hamil dari kemungkinan kematian. Kehamilan merupakan keadaan yang sulit dalam kehidupan setiap wanita yang menyebabkan berbagai perubahan kesehatan yang menyebabkan komplikasi akut yang terkadang mengakibatkan kematian ibu dan janin. Perubahan yang disebabkan oleh parameter fisiologis adalah alasan utama timbulnya komplikasi.

* Lakshmi.BN Telp: +91-8147949107

Email-id: keerthisri.20@gmail.com

Perubahan parameter fisiologis jika tidak diperhatikan oleh praktisi medis atau wanita hamil dapat memperburuk ke tingkat situasi darurat. Ketika situasi darurat tercipta, kesehatan wanita hamil menjadi rumit dan sulit untuk dinormalisasi dan dalam kasus seperti itu, nyawa ibu atau anak hilang atau keduanya hilang. Menurut sebuah penelitian, setiap hari sekitar 800 wanita meninggal karena berbagai penyebab kehamilan dan persalinan. Sekitar 99% dari semua kematian ibu ini terjadi di negara berkembang. Sebuah penelitian pada tahun 2013 menyebutkan bahwa 2.89.000 ibu hamil meninggal selama dan setelah kehamilan dan persalinan. Hampir semua kematian ini terjadi di lingkungan dengan sumber daya rendah dan sebagian besar dapat dicegah [1][2]. Banyak dari penyebab ini dapat dicegah atau dikontrol jika wanita hamil ini dapat diperingatkan sebelum situasinya menjadi lebih parah dan dengan demikian melindungi wanita hamil dan bayinya dari komplikasi lebih lanjut yang dapat menyebabkan kematian. Beberapa parameter fisiologis seperti tekanan darah, kadar glukosa darah dan berat badan yang perubahannya selama kehamilan dapat menyebabkan komplikasi dapat diidentifikasi dan komplikasi lebih lanjut yang disebabkan oleh perubahan ini dapat dicegah[1]. Tujuan dari makalah ini adalah untuk menggunakan berbagai jenis database untuk klasifikasi, untuk menganalisis dan memprediksi tingkat risiko di setiap contoh. Di sini, dalam makalah ini kami telah menganalisis database yang tidak terstandarisasi dengan semua parameter yang dipilih untuk penelitian dan juga menganalisis database dengan parameter yang dipilih setelah standarisasi, sebuah prosedur yang memberikan parameter yang tepat, valid, andal, dan akurat untuk dipelajari. Algoritma C4.5 adalah salah satu algoritma yang paling efisien, kuat dan populer dari metode Decision Tree Induction yang digunakan untuk tugas klasifikasi data medis, oleh karena itu algoritma ini dipilih diantara semua algoritma metode Decision Tree Induction lainnya. Algoritma ini memberikan hasil yang lebih baik bila dibandingkan dengan algoritma lain dan meninjau banyak pekerjaan dari peneliti ditemukan bahwa C4.5 mengungguli banyak algoritma klasifikasi lainnya bila diterapkan pada data medis. Bagian kedua menjelaskan tentang kehamilan dan komplikasi yang ditimbulkan selama kehamilan. Bagian ketiga menyajikan beberapa studi terkait. Bagian keempat memberikan penjelasan rinci tentang analisis algoritma pohon keputusan C4.5 dan prosedur kerjanya, yang digunakan untuk memprediksi status kesehatan wanita selama kehamilan. Bagian terakhir menyajikan hasil yang diperoleh dari analisis pada data standar dan data tidak standar, membandingkan hasil yang diperoleh dari analisis dan menyimpulkan bagaimana standarisasi parameter mempengaruhi hasil yang diperoleh dari algoritma C4.5 untuk memprediksi status kesehatan dari mengumpulkan data kehamilan.

2. Survei Literatur

Jay Gholap bertujuan untuk memprediksi kelas kesuburan tanah menggunakan algoritma pohon keputusan dalam penambangan data dan berfokus pada penyediaan kinerja algoritma pohon keputusan J48(C4.5) dengan bantuan teknologi meta seperti pemilihan dan peningkatan atribut. [3]. Aniket Ray et al menyajikan komite ansambel-hierarkis dari rambur acak yang menggunakan faktor risiko yang dicatat saat lahir untuk memprediksi risiko berkembangnya Retinopati prematuritas. [4]. Karpagavalli S et al mempresentasikan penerapan 3 algoritma pembelajaran terawasi C4.5 decision tree classifier, Naive Bayes dan Multilayer perceptron di lingkungan WEKA pada dataset penilaian pra-operasi. Akurasi prediksi pengklasifikasi dievaluasi menggunakan validasi silang 10 kali lipat dan hasilnya dibandingkan. [5]. Brijain R Patel dkk memberikan fokus pada berbagai algoritma pohon keputusan, karakteristik, tantangan, kelebihan dan kekurangannya. [6]. T.Miranda Lakshmi et al membandingkan dan menganalisis kinerja algoritma pohon keputusan ID3, C4.5 dan CART menggunakan data siswa kualitatif. [7]. Reza Entezari-Maleki dkk membandingkan keefektifan tujuh metode klasifikasi Decision Tree (DT), k-Nearest Neighbor (k NN), Logistic Regression (LogR), Naïve Bayes (NB), C4.5, Support Vector Machine (SVM) dan Linear Classifier (LC) berkaitan dengan metrik AUC. [8]. Sharareh R. Niakan et al telah menerapkan dan membandingkan teknik pembelajaran mesin pada awalnya untuk memprediksi hasil terapi tuberkulosis. Setelah analisis fitur, enam algoritma pohon keputusan (DT), jaringan saraf tiruan (ANN), regresi logistik (LR), fungsi basis radial (RBF), jaringan Bayesian (BN), dan mesin vektor dukungan (SVM) dikembangkan dan divalidasi. [9]. Masud Karim telah menyelidiki dua teknik penambangan data naïve bayes dan algoritma pohon keputusan C4.5 dan diimplementasikan untuk memprediksi apakah klien akan berlangganan deposito berjangka. [10].

3. C4.5 Algoritma Klasifikasi Pohon Keputusan

Klasifikasi teknik yang paling cocok di bidang diagnosis medis untuk penemuan otomatis informasi yang valid, baru, tidak diketahui, berguna dan dapat dipahami dari database yang sangat besar yang terdiri dari data yang diperoleh dari beberapa aplikasi. Di antara semua metode klasifikasi, pengklasifikasi yang paling cocok untuk penelitian ini yang terdiri dari data medis adalah algoritma klasifikasi pohon keputusan C4.5. Algoritme pohon keputusan sangat kuat, populer, berbasis logika, mudah ditafsirkan, lurus ke depan, dan dapat diterapkan secara luas untuk beberapa masalah

dalam penambahan data. Algoritma ini adalah algoritma off-the-shelf yang memberikan kinerja yang sangat baik dan mudah dimengerti. Algoritma Klasifikasi Pohon Keputusan C4.5 dikembangkan oleh Ross Quinlan sebagai perluasan dari algoritma ID3 yang juga dikembangkan olehnya. Pengklasifikasi ini membangun pohon keputusan sebagai model pembelajaran dari sampel data. Pendekatan bagi dan taklukkan diikuti untuk konstruksi model pohon keputusan menggunakan ukuran yang disebut perolehan informasi untuk memilih atribut dari dataset untuk pohon tersebut. Pertimbangan tes yang mungkin dipilih dengan n hasil yang membagi kumpulan data L dengan sampel pelatihan menjadi himpunan bagian $\{L_1, L_2, L_3, \dots, L_n\}$. Distribusi kelas dalam L dan himpunan bagiannya L_i adalah satu-satunya informasi yang tersedia untuk konstruksi pohon.

Anggap P sebagai sembarang himpunan sampel, maka $\text{freq}(C_i, P)$ adalah jumlah total sampel dalam P milik C_i dan $|P|$ dilambangkan dengan jumlah sampel di P . Entropi untuk himpunan P diberikan oleh

$$\text{Keterangan}(P) = - \sum_{i=1}^k \frac{\text{freq}(C_i, P)}{|P|} \log_2 \left(\frac{\text{freq}(C_i, P)}{|P|} \right) \quad (1)$$

Isi informasi dari L dapat diukur dengan menghitung $\text{Info}(L)$, isi informasi total dari L dapat dihitung setelah L dibagi sehubungan dengan hasil dari atribut yang dipilih katakanlah z . Jumlah bobot entropi dari setiap subet memberikan total konten informasi dari L .

$$\text{Info}_z(L) = \sum_{i=1}^n \frac{|L_i|}{|L|} \text{Info}(L_i) \quad (2)$$

Keuntungan diberikan oleh

$$\text{Perolehan}(z) = \text{Info}(L) - \text{Info}_z(L) \quad (3)$$

Memberikan informasi yang diperoleh dengan membagi L sehubungan dengan uji pada z . Hal ini dilakukan untuk pemilihan atribut z dengan perolehan informasi tertinggi. Tiga kasus dasar untuk algoritma C4.5 dipertimbangkan (1) Jika semua sampel dalam kumpulan data termasuk dalam kelas yang sama, simpul daun dibuat untuk pohon keputusan yang memilih kelas tersebut. (2) Jika tidak ada perolehan informasi yang disediakan oleh fitur/atribut apa pun, simpul keputusan dibuat di atas pohon dengan nilai yang diharapkan dari kelas. (3) Jika kelas dari instance yang tidak terlihat ditemui, simpul keputusan dibuat tinggi di atas pohon dengan nilai yang diharapkan dari kelas tersebut.

4. Metodologi Bagian

metodologi memberikan rincian teknik dan alat yang digunakan untuk pengumpulan data, pemilihan responden dan metode yang digunakan untuk menganalisis kasus data yang dikumpulkan. Bagian singkat tentang prosedur ilmiah diikuti untuk karya ini untuk menarik kesimpulan yang rasional, logis dan bermakna.

4.1. Pemilihan Parameter

Seperangkat atribut untuk penelitian diidentifikasi dari survei literatur dan saran dari para ahli.

Survei literatur yang luas dilakukan untuk mengidentifikasi masalah yang dihadapi ibu hamil dalam hal perubahan yang disebabkan oleh ibu hamil dan serangkaian parameter dipilih. Parameter ini menjadi sasaran analisis lebih lanjut oleh empat anggota komite doktor untuk menyarankan satu set parameter yang valid. Berdasarkan saran dari anggota komite, total dua belas parameter dipertimbangkan seperti yang ditunjukkan pada tabel 1.

Tabel 1 Atribut dan Deskripsi Ibu Hamil SI.

Tabel 1. Atribut dan Deskripsi ibu hamil G1.			
Tidak	Deskripsi Atribut		Tipe data
1	Usia	Usia ibu hamil dalam tahun	Numerik
2	Paritas Kehamilan	Angka kehamilan saat ini	Nomor
3	Riwayat Pre-eklampsia	Apakah ibu hamil pernah mengalami keadaan pre-eklampsia sebelumnya kehamilan (Ya/Tidak)	Tekstual
4	Sejarah Gestasional Diabetes	Apakah ibu hamil tersebut mengalami keadaan diabetes gestasional pada kehamilan sebelumnya (Ya/Tidak)	Tekstual
5	Ibu/Saudari dengan pre eklampsia	Apakah ibu/saudara ibu hamil mengalami pre eklampsia menyatakan kehamilannya (Ya/Tidak)	Tekstual
6	Ibu/Kakak dengan	Apakah ibu/kakak ibu hamil mengalami masa kehamilan	Tekstual

	Diabetes Gestasional	diabetes menyatakan kehamilan mereka (Ya / Tidak)	
7	Negara-	Status kesehatan ibu hamil saat ini (Hipertensi/Normal/Kegemukan/Kekurangan Berat Badan)	Tekstual
8	Trimester	Trimester kehamilan saat ini (2/3)	Numerik
	Bulan Sekarang	Bulan kehamilan saat ini (4/5/6/7/8/9)	Numerik
9 10	Tekanan darah	Tekanan darah saat ini tercatat dalam mm/Hg	Numerik
11	Kehadiran Kehamilan	Kadar Gula Darah saat ini tercatat dalam mg/dl	Numerik
	Diabetes		
12	Pertambahan Berat Badan	Pertambahan berat badan saat ini dari bulan sebelumnya dalam kg	Numerik

Atribut ini adalah variabel input yang diidentifikasi untuk penelitian dan hasilnya adalah tingkat risiko dalam kehamilan yang diwakili oleh nada status sebagai Risiko Ringan, Risiko Tinggi, atau Tanpa Risiko. Semua parameter saling terkait, kelainan pada satu parameter akan menyebabkan banyak komplikasi parah selama kehamilan. Oleh karena itu efisiensi setiap parameter penting untuk kesehatan ibu hamil.

4.2. Pengumpulan data:

Pengumpulan data dilakukan di distrik Bangalore, negara bagian Karnataka untuk penelitian ini dan tiga pusat kesehatan dari distrik Bangalore diidentifikasi untuk pengumpulan data. Izin yang diperlukan diperoleh dan izin dari wanita hamil yang berkunjung untuk konsultasi bulanan juga diperoleh untuk mengumpulkan data penelitian. Jadwal wawancara dikembangkan untuk pengumpulan yang menggabungkan semua atribut penelitian yang dipilih. Jadwal wawancara yang dikembangkan ini digunakan untuk pengumpulan data dari ibu hamil yang berkunjung ke puskesmas yang teridentifikasi untuk pemeriksaan bulanan. Populasi penelitian terdiri dari 600 ibu hamil yang datang ke puskesmas untuk pemeriksaan. Gambar 1 menunjukkan jadwal wawancara yang diikuti untuk pengumpulan data. Alat ukur yang berbeda digunakan untuk mengukur parameter fisiologis penelitian. Nilai-nilai yang dikumpulkan oleh seorang ahli medis dimasukkan ke dalam dataset.

Parameters		P_ID	P_ID 1	P_ID 2	P_ID 3	P_ID 4	P_ID 5 P_ID n
Personal Details								
a	Age							
b	Height							
c	Weight on conceiving							
d	BMI							
e	Present trimester							
f	Present month							
g	Present weight							
Health information								
a.	Health state before pregnancy							
b.	Blood Pressure							
c.	Gestational Diabetic presence							
d.	Weight Gain							
Personal History								
a.	Pregnancy Parity							
b.	History of Pre-eclampsia							
c.	History of Gestational Diabetes							
Family History								
a.	Mother/ Sister with Pre-eclampsia							
b.	Mother/ Sister with Gestational Diabetes							

Gambar 1: Jadwal wawancara untuk pengumpulan data

4.3. Pembuatan Dataset Kehamilan:

Dataset kehamilan disusun berdasarkan data kehamilan yang diperoleh dari pengumpulan data primer dengan menggunakan jadwal wawancara.

x Data Kehamilan Tidak Standar: Data kehamilan tidak standar adalah data kehamilan yang diperoleh dari jadwal wawancara.

Kumpulan data diperiksa untuk sampel data yang tidak normal, rawan kesalahan, hilang, dan tidak relevan. Jika sampel data yang tidak relevan diidentifikasi, sampel tersebut diganti dengan nilai yang valid atau dihapus dari kumpulan data. Dataset tersebut kemudian dipartisi menjadi dataset training dan testing. Partisi dataset dilakukan untuk menganalisis dataset kehamilan menggunakan klasifikasi C4.5

algoritma. Dataset dibagi secara acak dan sekitar 370 sampel data merupakan data pelatihan dan 230 sampel data lainnya merupakan data uji.

Dataset Kehamilan Standar: Standardisasi adalah prosedur statistik yang diikuti oleh banyak peneliti di berbagai bidang. Standardisasi parameter studi menyediakan kumpulan parameter yang valid, tepat, andal, bermakna, dan akurat yang mengarah ke peningkatan kinerja oleh pengklasifikasi yang dipilih.

Prosedur standardisasi diikuti karena kehamilan adalah masalah medis yang sensitif dan parameter yang mempengaruhi kehamilan dapat ditentukan dengan lebih baik oleh ahli medis, yang dapat menyediakan seperangkat parameter yang andal yang digunakan untuk penelitian. Sebanyak 40 juri yang memiliki pengetahuan dan ahli di bidang ginekologi dan kebidanan dipilih secara acak. Para juri ini diminta untuk memberikan pendapat mereka tentang kesesuaian parameter luas yang dipilih untuk standardisasi. Mereka diminta untuk menggunakan skala relevansi peringkat 3 poin seperti yang disarankan oleh Gulford (1954) yang terdiri dari skor Paling Relevan, Relevan, dan Tidak Relevan yang digunakan untuk tanggapan dan skor masing-masing 2,1 dan 0. Tanggapan yang diterima dari para ahli dikenai analisis relevansi komponen ke- i (R_i) dengan menggunakan rumus yang disarankan oleh Gulford (1954)

$$RC_i = \frac{\text{Total score of all the judges on } i^{\text{th}} \text{ component}}{\text{Maximum score on the continuum} \cdot \text{Total number of judges}} \quad (4)$$

Parameter menjadi sasaran uji relevansi dan parameter yang tidak relevan dikeluarkan dari daftar parameter yang dipertimbangkan untuk penelitian. Dari dua belas parameter yang dipilih untuk penelitian, enam parameter dianggap tidak relevan oleh 40 ahli medis dan karenanya dihapus. Satu set enam parameter dipertimbangkan untuk penelitian yang memiliki persentase relevansi lebih dari 70. Parameter penelitian yang tersisa tercantum bersama dengan peringkatnya pada tabel 2.

Tabel 2: Urutan peringkat parameter

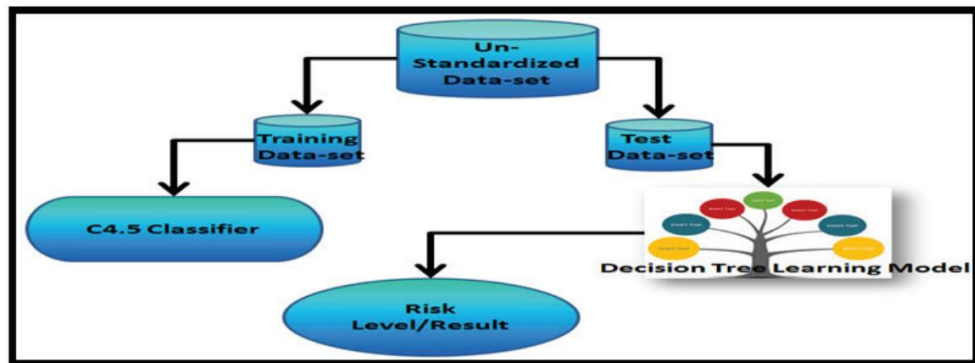
Parameter/Atribut	Peringkat
Negara	1 2
Tekanan darah	3 4
Tingkat Glukosa Darah	5 6
Bobot	
Trimester	
Bulan	

Di sini standardisasi adalah metode yang paling tepat untuk pemilihan parameter karena masalah yang dipertimbangkan untuk analisis adalah masalah medis yang sensitif dan sensitif serta memerlukan keahlian dalam menentukan parameter untuk penelitian. Dataset dibagi secara acak dan sekitar 370 sampel data merupakan data pelatihan dan 230 sampel data lainnya merupakan data uji.

4.4. Analisis Data:

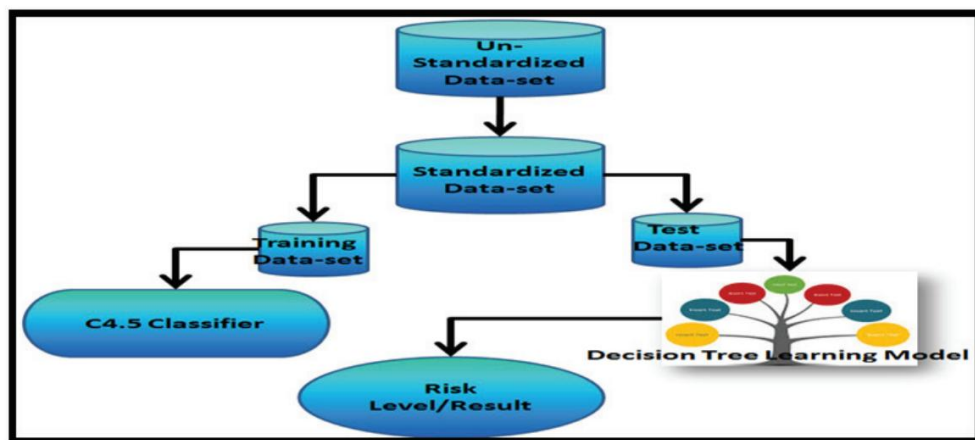
Dataset pelatihan dan pengujian dianalisis menggunakan algoritma klasifikasi pohon keputusan C4.5, dataset pelatihan yang terdiri dari 370 sampel data digunakan untuk membangun model pembelajaran pohon keputusan dan dataset pengujian digunakan untuk mengevaluasi model pembelajaran yang dibangun.

C4.5 Pengklasifikasi pada Kumpulan Data Tidak Standar: Kumpulan data kehamilan yang tidak standar dibagi menjadi kumpulan data pelatihan dan pengujian. Dataset pelatihan memiliki 370 sampel data dengan semua 12 nilai dan hasil parametrik, tingkat risiko yang sesuai dengan setiap sampel data. Dataset uji memiliki 230 sampel data dengan semua 12 nilai parametrik untuk setiap sampel dan tidak ada hasil atau tingkat risiko yang sesuai untuk setiap sampel data. Algoritma klasifikasi C4.5 diterapkan pada dataset pelatihan untuk membangun model pembelajaran pohon keputusan untuk dataset yang tidak standar. Untuk memperoleh dataset uji model pohon keputusan digunakan untuk menganalisis kinerja model pembelajaran dalam hal akurasi dalam prediksi risiko selama kehamilan dari dataset kehamilan yang tidak standar. Gambar 3 menunjukkan penerapan classifier C4.5 pada dataset pelatihan yang tidak dibakukan dan evaluasi model pembelajaran yang dibangun menggunakan dataset uji yang tidak dibakukan.



Gambar 3: C4.5 Classifier pada Un-Standardized Dataset

x Pengklasifikasian C4.5 pada Kumpulan Data Standar: Setelah standarisasi parameter studi, sampel data yang sesuai dengan parameter tersebut merupakan kumpulan data kehamilan standar. Dataset kehamilan standar dibagi secara acak menjadi 370 dataset pelatihan dan 230 dataset pengujian. Dataset pelatihan memiliki 370 sampel data dengan 6 nilai parametrik terpilih dan hasilnya, tingkat risiko yang sesuai dengan setiap sampel data. Dataset uji memiliki 230 sampel data dengan 6 nilai parametrik terpilih untuk setiap sampel dan tidak ada hasil atau tingkat risiko yang sesuai untuk setiap sampel data. Algoritma klasifikasi C4.5 diterapkan pada dataset pelatihan untuk membangun model pembelajaran pohon keputusan untuk dataset standar. Untuk mendapatkan dataset uji model pohon keputusan digunakan untuk menganalisis kinerja model pembelajaran dalam hal akurasi dalam prediksi risiko selama kehamilan menggunakan dataset standar. Gambar 4 menunjukkan penerapan classifier C4.5 pada dataset pelatihan standar dan evaluasi model pembelajaran yang dibangun menggunakan dataset uji standar.



Gambar 4: C4.5 pada Dataset Standar

Tujuan dari proses evaluasi ini adalah untuk mengetahui keakuratan model pembelajaran dalam memprediksi dan mengklasifikasikan ketika mengalami dua jenis database yang berbeda. Pengetahuan yang diperoleh dari proses evaluasi ini membantu dalam mengidentifikasi dampak parameter terhadap keakuratan suatu algoritma untuk memprediksi risiko yang diinduksi selama kehamilan pada wanita.

5. Hasil

Model pembelajaran yang dihasilkan oleh algoritma C4.5 mudah dipahami dan tidak memerlukan pakar domain. Teknik ini efisien, kuat, dan populer dalam mengklasifikasikan data dan memprediksi masing-masing risiko yang diinduksi dalam kehamilan. Studi ini juga menyoroti sedikit peningkatan yang diperoleh dalam akurasi ketika parameter standar dipilih untuk dianalisis melalui pengklasifikasi C4.5.

Results SetMAIN C4.5		
=====		
Correctly Classified Instances	152	66.087 %
Incorrectly Classified Instances	78	33.913 %
Kappa statistic	0.2152	
Mean absolute error	0.2154	
Root mean squared error	0.431	
Relative absolute error	74.2838 %	
Root relative squared error	110.5209 %	
Total Number of Instances	230	

Gambar 5: Hasil C4.5 pada data uji yang tidak standar

Gambar 5 menunjukkan ringkasan yang diperoleh dari analisis C4.5 pada dataset uji. Dari 230 sampel data 152 sampel data diklasifikasikan dengan benar atau tingkat risiko yang sesuai untuk sampel data yang diprediksi dengan benar dan 78 sampel data yang salah diklasifikasikan atau tingkat risiko yang diprediksi secara tidak tepat.

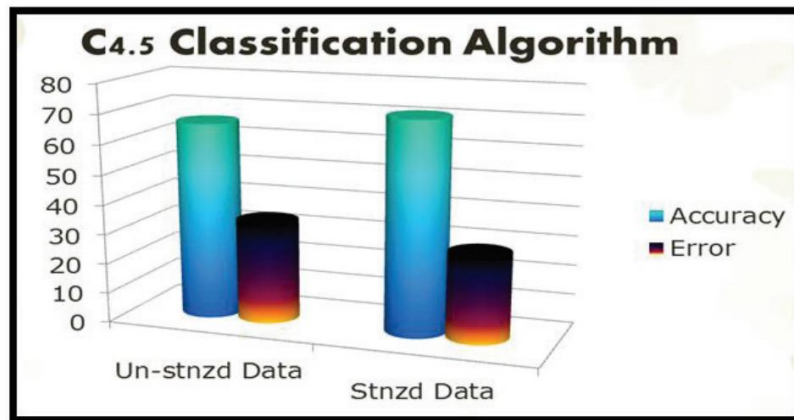
Results SetSTD C4.5		
=====		
Correctly Classified Instances	164	71.3043 %
Incorrectly Classified Instances	66	28.6957 %
Kappa statistic	0	
Mean absolute error	0.2982	
Root mean squared error	0.3861	
Relative absolute error	99.3727 %	
Root relative squared error	99.9959 %	
Total Number of Instances	230	

Gambar 6: Hasil C4.5 pada data uji Standar

Gambar 5 menunjukkan ringkasan yang diperoleh dari analisis C4.5 pada dataset uji. Dari 230 sampel data 152 sampel data diklasifikasikan dengan benar atau tingkat risiko yang sesuai untuk sampel data yang diprediksi dengan benar dan 78 sampel data yang salah diklasifikasikan atau tingkat risiko yang diprediksi secara tidak benar. Ringkasan analisis yang diekstraksi untuk kedua dataset melalui model pembelajaran yang dihasilkan oleh classifier C4.5 menggunakan WEKA toolkit adalah seperti yang ditunjukkan pada gambar 3 dan 4. Persentase akurasi untuk dataset tidak standar dan dataset standar masing-masing mencapai 66,087% dan 71,3043% dan persentase kesalahan masing-masing adalah 33,913% dan 28,6957%.

6. Kesimpulan

Dari hasil analisis pada kedua dataset dapat disimpulkan bahwa classifier pohon keputusan C4.5 memiliki potensi akurasi yang lebih besar untuk memprediksi tingkat risiko selama kehamilan ketika diterapkan pada data standar daripada data yang tidak standar. Gambar 7 menunjukkan grafik akurasi dan error yang diperoleh classifier C4.5 pada dataset yang tidak terstandarisasi dan terstandarisasi. Grafik menunjukkan bahwa pengklasifikasi C4.5 memberikan kinerja yang lebih baik dengan standarisasi daripada dengan dataset yang tidak standar.



Gambar 7 Grafik yang menunjukkan kinerja pengklasifikasi C4.5 pada data kehamilan

Dalam studi ini, kinerja pengklasifikasi C4.5 dievaluasi untuk akurasi dan dampak standardisasi pada akurasi dipahami dan diidentifikasi untuk memberikan prediksi risiko kehamilan yang lebih baik. Makalah ini memberikan pentingnya klasifikasi dan prediksi dalam menambang data terkait kehamilan. Teknik klasifikasi lain juga dapat digunakan untuk menganalisis data kehamilan tetapi untuk penelitian ini pengklasifikasi C4.5 dipertimbangkan untuk masalah ini karena kekuatan, popularitas dan efisiensinya serta kelezatan masalah kehamilan. Gambar 7 memberikan akurasi dan persentase kesalahan algoritma C4.5 ketika diterapkan pada data kehamilan standar dan tidak standar. Jadi sebagai pernyataan penutup classifier C4.5 menawarkan kinerja yang lebih baik jika standarisasi diikuti sebelum analisis seperti yang terlihat dari hasil yang diperoleh dari kedua dataset. Oleh karena itu, dari penelitian ini makalah menunjukkan keuntungan yang ditawarkan oleh algoritma klasifikasi pohon keputusan C4.5 untuk pemantauan kesehatan berbasis prediksi pada wanita hamil dan pentingnya standardisasi. Penelitian lebih lanjut dilakukan untuk meningkatkan akurasi bahkan lebih baik setelah prosedur standarisasi untuk memberikan ibu hamil tingkat risiko yang tepat untuk memberi mereka masa kehamilan yang aman dan sehat.

Referensi

- [1]. United Nations Mortality Estimation Inter-agency Group, terdiri dari perwakilan dari World Health Organization (WHO), United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA), United Nations Population Division, Bank Dunia dan Wikipedia.
- [2]. Organisasi Kesehatan Dunia (November 2014). "Lembar Fakta kelahiran prematur N°363".who.int.Diakses pada 6 Mar 2015.
- [3]. Aplikasi Pembelajaran Mesin dalam Prediksi dan Prognosis Kanker oleh Joseph A. Cruz, David S. Wishart, Departemen Ilmu Biologi dan Ilmu Komputasi, Universitas Alberta Edmonton, AB, Kanada T6G 2E8 dalam Informatika Kanker 2006: 2 59– 78
- [4]. Pohon Keputusan bab 9 oleh Lior Rokach, Departemen Teknik Industri, Universitas Tel-Aviv dan Oded Maimon, Departemen Teknik Industri, Universitas Tel-Aviv
- [5]. Pendekatan Pembelajaran Mesin untuk Prediksi Risiko Anestesi Pra Operasi oleh Karpagavalli S1, Jamuna KS2, dan Vijaya MS2,2 GR Sekolah Teknologi Komputer Terapan Govindarajulu, PSGR Krishnammal College for Women, Coimbatore, India, Email: karpagam@grgsact.com dalam Jurnal Internasional Terbaru Tren Teknik, Vol. 1, No. 2, Mei 2009
- [6]. Penambangan Data Prediktif untuk Diagnosis Medis: Tinjauan Prediksi Penyakit Jantung oleh Jyoti Soni Ujma Ansari Dipesh Sharma, Mahasiswa, M.Tech (CSE). Professor Reader, Raipur Institute of Technology, Raipur, Chhattisgarh, India dalam International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, Maret 2011
- [7]. Penelitian Sistem Pakar Berbasis Pengetahuan Digunakan dalam Diagnosis Maternitas, Lu Binjie, Ph.D. Universitas Donghua, Departemen TI Rumah Sakit Renji berafiliasi dengan Fakultas Kedokteran Universitas Shanghai Jiaotong, Shanghai, China, lu-bj@126.com pada tahun 2010 Internasional Konferensi tentang Pemodelan Aplikasi dan Sistem Komputer (ICCA SM 2010, 978-1-4244-7237-6/\$26.00 C 2010 IEEE [8]. Pembelajaran Mesin yang Diawasi: Tinjauan Teknik Klasifikasi, SB Kotsiantis, Departemen Ilmu dan Teknologi Komputer, Universitas of Peloponnese, Greece, End of Karaiskaki, 22100 , Tripolis GR.Informatica 31 (2007) 249-268 [9]. Performance Tuning Algoritma J48 Menggunakan Metode Genetic Algorithm pada Data Breast Cancer [10]. Perbandingan Eksperimental Naive Bayesian dan Penyaringan Anti-Spam Berbasis Kata Kunci dengan Pesan Email Pribadi, Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinou dan Constantine D. Spyropoulos, Rekayasa Perangkat Lunak dan Pengetahuan Laboratorium, Lembaga Informatika dan Telekomunikasi, Pusat Penelitian Ilmiah Nasional "Demokritos", 153 10 Ag.