

Implementasi *Decision Tree* untuk Prediksi Kelahiran Bayi Prematur

Decision Tree Implementation for Premature Birth Prediction

Putri Lailatul Rosida¹, Mieke Nurmalasari^{2*}, Hosizah³, Dewi Krismawati⁴

^{1,2,3} Program Studi Manajemen Informasi Kesehatan, Fakultas Ilmu-Ilmu Kesehatan, Universitas Esa Unggul, Jakarta, Indonesia

⁴ Direktorat Analisis dan Pengembangan Statistik, Badan Pusat Statistik, Indonesia

*E-mail: mieke@esaunggul.ac.id

Abstrak

Kelahiran bayi dini di Indonesia merupakan kasus yang mempunyai angka kejadian sangat tinggi. Menurut data KEMENKES tahun 2021, menunjukkan presentasi kelahiran bayi prematur di Indonesia yaitu 84%. Jumlah kematian bayi di Indonesia masih relatif tinggi dibandingkan dengan negara-negara ASEAN lainnya. Tujuan penelitian ini adalah melakukan prediksi kelahiran bayi prematur dengan implementasi *decision tree*, dengan jenis penelitian analisis prediktif. Populasi dalam penelitian ini yaitu pasien ibu hamil dengan jumlah sampel 350 data pasien ibu hamil meliputi variabel yang diteliti Usia, IMT, Keputihan/Infeksi, Riwayat Keguguran, Riwayat Prematur dan Jarak Kehamilan. Prediksi dilakukan dengan membagi dua data training sebanyak 245 dan data testing sebanyak 105. Hasil yang diperoleh adalah variabel Indeks Massa Tubuh (IMT) menjadi faktor yang paling berisiko terjadinya kelahiran bayi prematur dan model *decision tree* menghasilkan AUC 91,7% dapat disimpulkan bahwa *decision tree* memiliki nilai akurasi *good classification*.

Kata kunci: *Decision Tree*, Kelahiran Prematur, Prediksi.

Abstract

The early birth of baby in Indonesia is a case that has a very high incidence rate. According to data from the Ministry of Health in 2021, the presentation of premature babies in Indonesia is 84%. The number of infant deaths in Indonesia is still relatively high compared to other ASEAN countries. The purpose of this study was to predict the birth of premature babies with the implementation of *decision tree*, with this type of predictive analysis research. The population in this study is pregnant women patients with a sample of 350 pregnant women patient data covering the variables studied Age, BMI, Vaginal Discharge, History of Miscarriage, History of Prematurity and Pregnancy Spacing. The prediction was made by halving the training data by 245 and the testing data by 105. The results obtained are the variable Body Mass Index (BMI) is the riskiest factor for premature birth The *decision tree* model yields an AUC of 91.7%, it can be concluded that the *decision tree* has a good classification accuracy value.

Keywords: *Decision Tree*, Prediction, Preterm Birth.

Naskah diterima 09 Mei 2024; direvisi 12 Jul. 2024; dipublikasikan 01 Okt. 2024.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



I. PENDAHULUAN

Kelahiran bayi prematur di Indonesia masih menjadi salah satu kasus yang sangat tinggi, dari data Badan Pusat Statistik Provinsi DKI Jakarta Tahun 2018 menunjukkan persentase kelahiran bayi prematur di DKI Jakarta yaitu 1.381, pada Tahun 2019 sebanyak 1.530 dan Tahun 2020 sebanyak 2.145. Persalinan prematur yaitu kelahiran terjadi sebelum memasuki minggu ke 37 kehamilan, bayi yang lahir kurang bulan mempunyai berat badan kurang dari 2,5 kg. Usia kehamilan disebut *aterm* atau cukup bulan apabila umur kehamilan ibu memasuki 37 sampai 42 minggu [1]. Apabila umur kehamilan kurang dari 37 minggu maka disebut *preterm* atau kurang bulan. Apabila umur kehamilan lebih dari 42 minggu disebut *postterm* atau *serotinus*. Menentukan diagnosis persalinan prematur sering kali sulit, karena kontraksi yang muncul pada kehamilan prematur tidak selalu menandakan ancaman persalinan yang sebenarnya. Jika tidak segera ditangani, persalinan mungkin tidak akan terjadi. Beberapa kriteria yang dapat digunakan untuk mendiagnosis ancaman persalinan prematur meliputi: kontraksi berulang setidaknya setiap 7-8 menit sekali, atau 2-3 kali dalam 10 menit; nyeri punggung bawah; perdarahan bercak; perasaan tekanan di daerah serviks; pemeriksaan serviks

menunjukkan pembukaan setidaknya 2 cm dan penipisan 50-80%; presentasi janin rendah hingga mencapai *spina isiadika*. Selaput ketuban pecah dapat merupakan tanda awal terjadinya persalinan prematur, Terjadi pada usia kehamilan 22- 37 minggu [2]. Bayi yang lahir kurang bulan mengakibatkan organ pada bayi tidak bekerja optimal dan berisiko mengalami masalah kesehatan yang lebih serius dibandingkan bayi-bayi yang lahir pada waktu yang dijadwalkan [3]. Secara umum bayi yang lahir kurang bulan akan mengalami beberapa masalah pasca lahir dibandingkan bayi yang lahir secara normal. Belum cukupnya masa gestasi mengakibatkan tidak matangnya semua sistem organ, contohnya pada sistem pernafasan, sistem pencernaan, sistem penyerapan, sistem peredaran darah, serta sistem saraf pusat. Oleh karena itu anak yang lahir kurang bulan memiliki tingkat cerebral palsy dan defisit sensoriknya lebih tinggi serta kurangnya kemampuan belajar dibandingkan anak yang lahir cukup bulan [4].

Persalinan prematur adalah suatu kondisi yang disebabkan oleh berbagai faktor. Gabungan faktor obstetrik, sosiodemografi, dan medis berkontribusi terhadap terjadinya persalinan prematur. Penyebab persalinan prematur mencakup iatrogenik (20%), infeksi (30%), ketuban pecah dini pada kehamilan prematur (20-25%), dan persalinan prematur spontan (20-25%). Berdasarkan penelitian, faktor risiko persalinan prematur meliputi jarak antar persalinan yang pendek (kurang dari 60 bulan), riwayat persalinan prematur sebelumnya, ras/etnik, usia ibu yang ekstrim (kurang dari 20 tahun atau lebih dari 40 tahun), malnutrisi ibu dan stres kronis, infeksi, status sosioekonomi rendah, merokok (termasuk perokok pasif), konsumsi alkohol atau kokain, faktor plasenta, dan kehamilan ganda. [5]. Persalinan prematur merupakan ancaman serius karena merupakan salah satu penyebab mortalitas dan morbiditas neonatal di Indonesia. Menurut data Kemenkes tahun 2021 menunjukkan presentasi kelahiran bayi prematur di Indonesia yaitu 84%. Jumlah kematian bayi di Indonesia masih relatif tinggi dibandingkan dengan negara-negara ASEAN lainnya [6]. Oleh karena itu perlu untuk kita mengetahui faktor-faktor apa saja yang mempengaruhi kelahiran bayi prematur dengan memprediksi mengenai masa mendatang diupayakan untuk mengurangi angka kematian ibu dan bayi.

Berbagai faktor yang menjadi penyebab kelahiran prematur mencakup usia ibu, indeks massa tubuh (IMT), jarak kehamilan, riwayat keguguran, riwayat prematur, keputihan, dan riwayat merokok. Usia ibu memengaruhi risiko kelahiran prematur, dengan risiko lebih tinggi pada ibu yang lebih muda dari 20 tahun dan lebih tua dari 35 tahun *previa* [1]. Indeks Massa Tubuh (IMT) juga berperan penting, di mana hasil kehamilan lebih baik pada ibu dengan IMT normal [7]. Jarak kehamilan yang terlalu pendek, khususnya kurang dari dua tahun, meningkatkan risiko kelahiran prematur karena tubuh ibu tidak memiliki cukup waktu untuk pulih sepenuhnya [8]. Riwayat keguguran dapat mempengaruhi kehamilan berikutnya, meningkatkan risiko persalinan prematur dan berat badan lahir rendah (BBLR) [9]. Wanita dengan riwayat kelahiran prematur sebelumnya memiliki kemungkinan yang lebih besar untuk mengalami kelahiran prematur pada kehamilan berikutnya [10]. Selain itu, keputihan yang disebabkan oleh infeksi bakteri selama kehamilan dapat memicu kontraksi uterus dan persalinan prematur [7]. Terakhir, merokok selama kehamilan merupakan faktor risiko utama yang dapat menyebabkan kelahiran prematur dan komplikasi lainnya pada bayi. Merokok juga dapat menyebabkan solusio plasenta, plasenta previa, dan pecahnya selaput ketuban [11]. Oleh karena itu, penting bagi ibu hamil untuk menghindari penggunaan tembakau dan faktor-faktor risiko lainnya untuk mencegah kelahiran prematur.

Penambangan data adalah proses menggali informasi berharga dari dataset yang ada. Ini melibatkan eksplorasi data untuk menemukan pola-pola tersembunyi, mengidentifikasi wawasan yang berarti, dan melakukan prediksi. Dengan demikian, penambangan data merupakan usaha untuk mengekstrak pengetahuan yang terdapat dalam basis data yang besar, membuka pemahaman yang tersembunyi dari informasi tersebut. Kemajuan dalam berbagai bidang ilmu seperti ilmu pengetahuan, bisnis, dan sebagainya telah menghasilkan peningkatan dalam kumpulan data yang tersedia. Penambangan data sering kali dianggap sebagai bagian dari proses penemuan pengetahuan dalam basis data (*Knowledge Discovery in Database/KDD*), yang mencari wawasan yang bermakna dari informasi yang ada. Pemilihan metode yang tepat sangat tergantung pada tujuan dan langkah-langkah yang diperlukan dalam KDD. Metode yang dilakukan dalam proses mining yaitu menggunakan model *decision tree* dengan algoritma C4.5. Salah satu teknik data mining yang dikenal sebagai klasifikasi merupakan teknik yang digunakan untuk memprediksi atau pengklasifikasian kategori dari sekumpulan atribut data [12]. Metode klasifikasi yang paling mudah dipahami oleh manusia yaitu pohon keputusan (*decision tree*). Terdapat empat pelompokan dalam data mining yaitu klasifikasi, asosiasi, clustering dan prediksi [13]

Decision tree adalah metode klasifikasi yang mengilustrasikan struktur pohon di mana setiap cabang mewakili sebuah variabel, cabang-cabangnya menunjukkan nilai-nilai variabel tersebut, dan daun-daunnya mewakili setiap kelas. Tujuan dari penggunaan *decision tree* adalah untuk memberikan gambaran atau prediksi serta membuat keputusan yang sederhana. Sejumlah penelitian telah dilakukan untuk mengembangkan algoritma C4.5 agar mampu menangani masalah seperti missing value, data kontinu, dan proses pruning.

Algoritma C4.5 menerima sampel-sampel latihan (training samples) dan sampel-sampel di mana sampel latihan tersebut telah diuji kebenarannya. Sedangkan, sampel-sampel tersebut mewakili data bidang yang akan digunakan sebagai parameter dalam proses klasifikasi data. [14]. *Decision tree* merupakan salah satu algoritma supervised machine learning yang digunakan dalam teknik prediksi. Supervised learning melibatkan penggunaan data yang telah diberi label, di mana setiap sampel data memiliki asosiasi dengan label atau hasil yang sudah diketahui sebelumnya. Tujuan dari supervised learning adalah untuk melatih model komputer agar dapat memahami pola-pola dalam data dan melakukan prediksi yang akurat terhadap data yang belum dikenal. Dengan menggunakan algoritma-algoritma supervised learning, model dapat digunakan untuk mengklasifikasikan data ke dalam kategori yang sesuai atau untuk melakukan prediksi numerik [15]. Dalam pohon keputusan, terdapat tiga jenis node yang berbeda: (1) Root node, yang merupakan node teratas dan tidak memiliki input, serta bisa tidak memiliki output atau memiliki lebih dari satu output, (2) Internal node, yang merupakan node percabangan di mana hanya terdapat satu input dan minimal dua output, (3) Leaf node atau terminal node, yang merupakan node terakhir di mana hanya terdapat satu input dan tidak memiliki output.

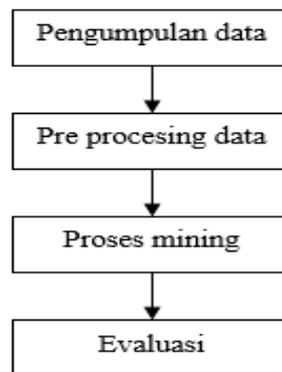
Algoritma C4.5 yaitu pembaruan dari algoritma ID3 yang dikembangkan oleh J. Ross Quinlan. Algoritma C4.5 dapat membentuk struktur pohon yang bermanfaat untuk *classification* [13]. Algoritma C4.5 sering diandalkan sebagai solusi untuk memecahkan masalah dalam teknik klasifikasi. Hasil dari algoritma C4.5 berupa decision tree, mirip dengan teknik klasifikasi lainnya. *Decision tree* adalah struktur yang memungkinkan pembagian dataset besar menjadi subset yang lebih kecil dengan menerapkan serangkaian aturan keputusan [16]. Setiap tahap pembagian menghasilkan subset yang serupa satu sama lain. Algoritma C4.5 merupakan salah satu algoritma yang umum digunakan, terutama dalam pembelajaran mesin, dan telah mengalami berbagai perbaikan dari versi sebelumnya, seperti ID3. Salah satu perbaikan utamanya adalah metode pemangkasan yang lebih canggih [15].

Penelitian sebelumnya telah mengungkap faktor-faktor yang berpengaruh terhadap kelahiran bayi prematur diantaranya adalah yang dilakukan oleh Rahmat Fitra A dan M Ihsan Jambak (2023) menggunakan metode C4.5 untuk klasifikasi tindakan persalinan pada pasien ibu bersalin, dari hasil pengujiannya dapat disimpulkan bahwa ada beberapa faktor yang mempengaruhi persalinan prematur, seperti faktor utama penentu karakteristik dalam menentukan tindakan persalinan adalah Volume Ketuban, Preeklampsia, Presentasi Bayi, Riwayat SC, Disproporsi Kepala Panggul (DKP), Plasenta Previa, Lilitan Tali Pusat, Fetal Distress, Tunggal/Ganda, KPSW/KPD, Faktor Lain-lain, dan Usia Kehamilan, dalam penelitian tersebut dihasilkan Penambahan data dapat diimplementasikan dengan baik menggunakan algoritma C4.5 [17]. Penelitian yang dilakukan oleh Hamsir Saleh (2020) telah mengidentifikasi ada lima faktor penyebab stunting yaitu jenis kelamin, berat badan, tinggi badan, fasilitas kesehatan, asi eksklusif, PHBS, jamban, SPAL dalam penelitiannya menggunakan metode Algoritma C4.5 untuk mengetahui faktor utama penyebab stunting. Hasil yang didapatkan yaitu diperoleh gain ratio tertinggi pada atribut pelayanan kesehatan dengan nilai 0.8889 [18]. Selanjutnya penelitian yang dilakukan oleh Sunanto (2022) menerapkan algoritma C4.5 untuk membuat model prediksi pasien yang mengidap penyakit diabetes dengan hasil perhitungan confusion matrix yang mendapatkan hasil accuracy sebesar 95,51% [19].

Berdasarkan observasi awal di RSUD Kembangan dalam laporan tahunan kasus kelahiran prematur cukup berisiko. Kasus kelahiran prematur terjadi karena beberapa faktor, implementasi data mining ini diharapkan dapat membantu rumah sakit dalam menentukan faktor apa saja yang menjadi penyebab kelahiran bayi prematur dan dapat memprediksi berapa banyak kasus kelahiran bayi prematur di masa mendatang. Dalam penelitian ini data mining yang digunakan yaitu algoritma *decision tree*. Berdasarkan penelitian terdahulu membandingkan algoritma *decision tree* dengan algoritma lain, hasilnya menunjukkan algoritma *decision tree* ini mampu menghasilkan prediksi dengan tingkat akurasi tinggi dan mudah dipahami oleh karena itu algoritma *decision tree* merupakan metode yang tepat untuk memprediksi dengan kumpulan data yang besar dan rumit. Pada penelitian ini variabel yang dipilih berdasarkan teori Dr. Hermanto Tri Joewono, dr., Sp. OG (K) dalam buku acuan persalinan kurang bulan (2020) dan penyesuaian variabel di RSUD Kembangan oleh karena itu penelitian ini melibatkan lebih banyak faktor dalam memprediksi kelahiran prematur dari pada penelitian sebelumnya yang belum memasukkan variable Indeks Massa Tubuh (IMT).

II. METODE PENELITIAN

Dalam penelitian ini pengolahan data menggunakan analisis prediktif dilakukan untuk melihat kondisi diwaktu yang akan datang dan berbagai kemungkinan yang dapat terjadi dengan teknik klasifikasi menggunakan *decision tree* data mining. Klasifikasi data mining mengelompokkan beberapa variabel dan diberikan label. Menggunakan bantuan metode *Knowledge Discovery in Database* (KDD) beberapa tahapan atau langkah untuk data mining seperti di bawah ini [20]:



Gambar 1. KDD

Pengumpulan Data

Tahap awal yang dilakukan dalam penelitian ini adalah menyiapkan atau pengumpulan data mentah 3 tahun terakhir dimana data diperoleh dari laporan tahunan di rumah sakit. Setelah dilakukan tahap seleksi data yang didapat adalah data pasien ibu hamil dengan jumlah data sebanyak 350, terdiri dari 6 variabel atau atribut. Variabel tersebut ada yang tergolong variable dependen yaitu variabel yang dijadikan sebagai variabel target adalah kelahiran prematur dan variabel independen yaitu variabel yang dijadikan sebagai faktor penyebab kelahiran prematur. Adapun variabel independennya adalah Usia, Indeks Masa Tubuh (IMT), Keputihan, Riwayat Keguguran, Riwayat Prematur, dan Jarak Kehamilan

Pre Processing

Pada proses ini sampel data dibagi menjadi 2, yaitu data training 70% dan data testing 30%. Dalam penelitian ini data training dihasilkan data training 70% yaitu 245 dari 350 data dan data testing 30% yaitu 105 dari 350 data. Pada tahap ini dataset tidak mengalami *missing value* oleh karena itu dapat dilakukan tahap *transformation*. Pada penelitian ini tahap *transformation* yang dilakukan yaitu mengubah bentuk data, data yang masih berupa angka yang ditransformasikan kedalam bentuk kategorikal seperti pada variabel usia dan variabel IMT (Indeks Masa Tubuh).

Proses Mining

Pada tahap ini yaitu melakukan penerapan teknik algoritma decision tree. Proses ini adalah tahapan yang menentukan model pohon keputusan yang sesuai dengan data training. Hasil penelitian dilakukan dengan menggunakan tools orange dengan jumlah data sebanyak 350 data ibu hamil. Penggunaan data mining ini bertujuan untuk menentukan prediksi bayi lahir secara prematur dengan menggunakan metode C.45.

Evaluasi

Pada penelitian ini tahap evaluasi yaitu untuk mengevaluasi hasil dari algoritma *Decision tree* C4.5. Evaluasi menggunakan *Confusion Matrix* agar didapat nilai akurasi, presisi, dan recall. *Confusion Matrix* merupakan sebuah hasil evaluasi dari sebuah klasifikasi data mining yang diwujudkan dalam sebuah tabel. *Confusion matrix* adalah metode yang banyak dipakai untuk menghitung nilai akurasi. Pengukuran kinerja menggunakan *confusion matrix* memiliki empat istilah sebagai gambaran dari hasil klasifikasi. Adapun keempat istilah tersebut, yaitu:

1. False Positive (FP), yaitu data negatif tapi terprediksi sebagai data positif.
2. False Negative (FN), yaitu data positif yang terprediksi sebagai data negatif.
3. True Positive (TP), yaitu data positif yang terprediksi benar.
4. True Negative (TN), yaitu data negatif yang terprediksi dengan benar

Sensitivitas atau Recall adalah rasio prediksi benar positif dipadukan dengan keseluruhan data yang benar positif atau mengukur proporsi positif asli yang diramal secara benar sebagai positif. Dalam sensitivitas berkaitan dengan kecakapan pengujian untuk mengenali hasil yang positif dari sejumlah data yang seharusnya positif. Untuk menghitung sensitivitas atau recall menggunakan persamaan di bawah ini:

$$\text{Sensitivasi} = \frac{TP}{TP+FN} \quad (1)$$

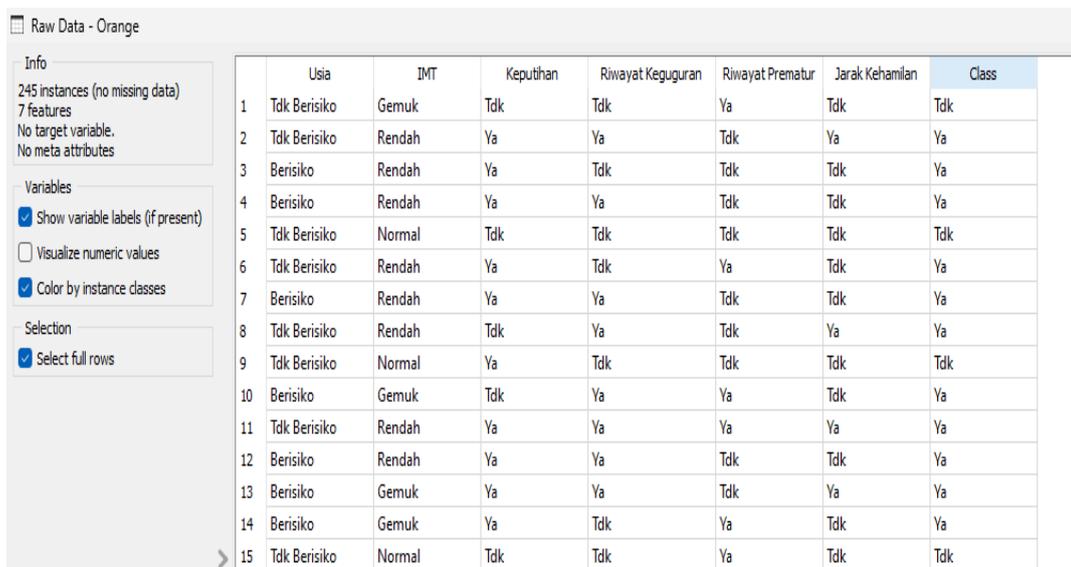
Precision adalah rasio antara jumlah prediksi benar positif dengan total hasil yang diprediksi positif. *Precision* adalah metrik yang menggambarkan kemampuan sistem dalam menghasilkan data yang relevan.

Dalam konteks data mining, precision mengukur jumlah data yang terklasifikasi sebagai positif dengan benar (true positive) dibagi dengan total jumlah data yang diklasifikasikan sebagai positif. Persamaan untuk menghitung *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

III. HASIL DAN PEMBAHASAN

Pengumpulan data yang dilakukan pada penelitian ini adalah mengumpulkan data mentah yang diperoleh dari data pasien ibu hamil. Pada penelitian ini data yang diambil merupakan data 3 tahun terakhir yaitu pada tahun 2021-2022, setelah dilakukan pengumpulan keseluruhan data kemudian dilakukan tahap seleksi sehingga diperoleh 350 data pasien ibu hamil dengan 6 atribut yaitu Usia, IMT, Keputihan, Riwayat Keguguran, Riwayat Prematur, dan Jarak Kehamilan dengan variabel target kelahiran bayi Prematur. Pada penelitian ini yang dijelaskan pada gambar 2 merupakan tahap *preprocessing* dan *selection* data. Pada tahap ini dataset tidak mengalami *missing value*. Pada proses ini, sampel data dibagi menjadi dua bagian, yaitu data training sebesar 70% dan data testing sebesar 30%. Dalam penelitian ini, dari total 350 data, data training terdiri dari 245 data (70%) dan data testing terdiri dari 105 data (30%).



	Usia	IMT	Keputihan	Riwayat Keguguran	Riwayat Prematur	Jarak Kehamilan	Class
1	Tdk Berisiko	Gemuk	Tdk	Tdk	Ya	Tdk	Tdk
2	Tdk Berisiko	Rendah	Ya	Ya	Tdk	Ya	Ya
3	Berisiko	Rendah	Ya	Tdk	Tdk	Tdk	Ya
4	Berisiko	Rendah	Ya	Ya	Tdk	Tdk	Ya
5	Tdk Berisiko	Normal	Tdk	Tdk	Tdk	Tdk	Tdk
6	Tdk Berisiko	Rendah	Ya	Tdk	Ya	Tdk	Ya
7	Berisiko	Rendah	Ya	Ya	Tdk	Tdk	Ya
8	Tdk Berisiko	Rendah	Tdk	Ya	Tdk	Ya	Ya
9	Tdk Berisiko	Normal	Ya	Tdk	Tdk	Tdk	Tdk
10	Berisiko	Gemuk	Tdk	Ya	Ya	Tdk	Ya
11	Tdk Berisiko	Rendah	Ya	Ya	Ya	Ya	Ya
12	Berisiko	Rendah	Ya	Ya	Tdk	Tdk	Ya
13	Berisiko	Gemuk	Ya	Ya	Tdk	Ya	Ya
14	Berisiko	Gemuk	Ya	Tdk	Ya	Tdk	Ya
15	Tdk Berisiko	Normal	Tdk	Tdk	Ya	Tdk	Tdk

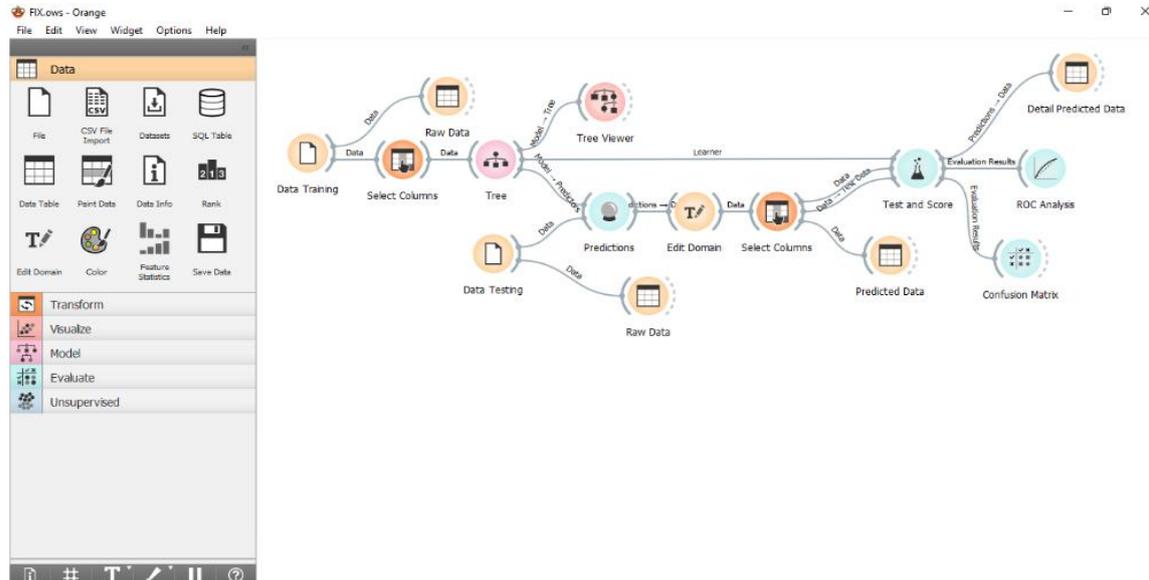
Gambar 2. *Preprocessing* dan *Selection* Data

Selanjutnya dilakukan tahap Transformasi data, yaitu langkah mengubah format dan menggabungkan data ke dalam bentuk yang spesifik untuk meningkatkan akurasi dan efisiensi algoritma. Sebelum diterapkan dalam data mining, diperlukan format data yang sesuai. Transformasi pada penelitian ini dapat dilihat pada tabel 1, yaitu tahap yang mengubah bentuk data, data yang masih berupa angka akan ditransformasikan kedalam bentuk kategorikal pada masing-masing atribut.

TABEL 1
ATRIBUT YANG DIGUNAKAN

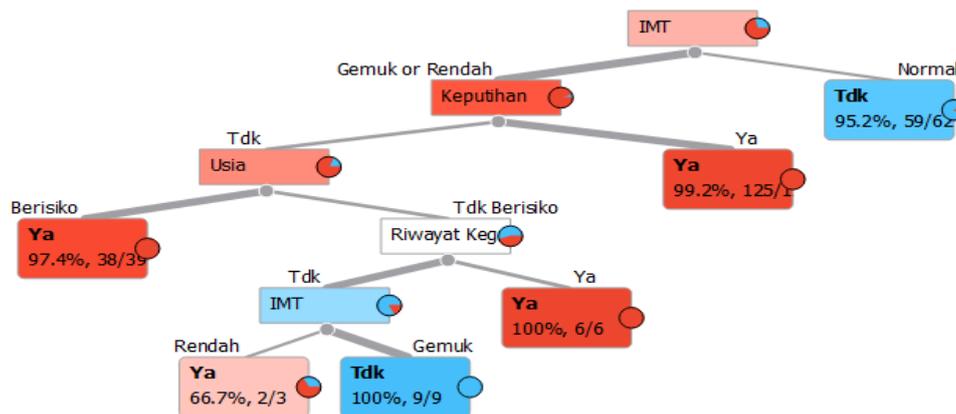
No.	Atribut	Nilai
1.	Usia	Berisiko dan Tidak Berisiko
2.	Indek Massa Tubuh (IMT)	Gemuk, Normal dan Rendah
3.	Keputihan	Ya dan Tidak
4.	Riwayat Keguguran	Ya dan Tidak
5.	Riwayat Prematur	Ya dan Tidak
6.	Riwayat Jarak Kehamilan ≥ 2 Tahun	Ya dan Tidak

Pada gambar 3 menunjukkan alur dan proses pemodelan dengan *decision tree* dengan menggunakan aplikasi Orange Data Mining. Pemodelan *decision tree* dilakukan dengan enam atribut yang merupakan faktor-faktor yang dapat menyebabkan kelahiran prematur.



Gambar 3. Pemodelan *decision tree*

Gambar 4 dapat dilihat bahwa IMT merupakan *root/akar* pohon keputusan yang telah dibuat. IMT menjadi root atau akar dari pohon keputusan dalam gambar 4 karena IMT memberikan pembagian terbaik dalam hal pengukuran ketidakpastian atau *impurity* pada data (memiliki *information gain* tertinggi). Arti dari warna merah mutlak menjelaskan bahwa tidak terdapat faktor risiko yang dapat mengakibatkan bayi prematur. Warna biru mutlak menjelaskan bahwa terdapat faktor lain yang dapat mengakibatkan bayi prematur. Berikut merupakan hasil dari pohon keputusan yang dibuat menggunakan metode *decision tree*.



Gambar 4. Pohon keputusan *decision tree*

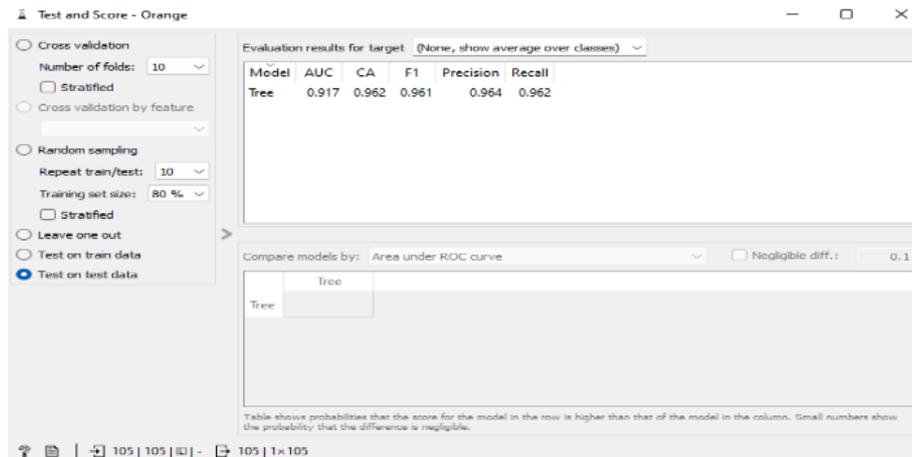
Gambar 4 menjelaskan semua kasus sudah masuk dalam kelas, oleh karena itu terbentuk beberapa aturan dalam *tree*. Aturan yang terbentuk dijelaskan sebagai berikut:

1. Jika IMT = Gemuk atau Rendah, Keputihan = Tdk, Usia = Berisiko maka *class* = Prematur
2. Jika IMT = Gemuk atau Rendah, Keputihan = Tdk, Usia = Tidak Berisiko, Riwayat Keguguran = Tdk, IMT= Rendah maka *class* = Prematur
3. Jika IMT = Gemuk atau Rendah, Keputihan = Tdk, Usia = Tidak Berisiko, Riwayat Keguguran = Tdk, IMT= Gemuk maka *class* = Tidak prematur
4. Jika IMT = Gemuk atau Rendah, Keputihan = Tdk, Usia = Tidak Berisiko, Riwayat Keguguran = Ya maka

class = Prematur

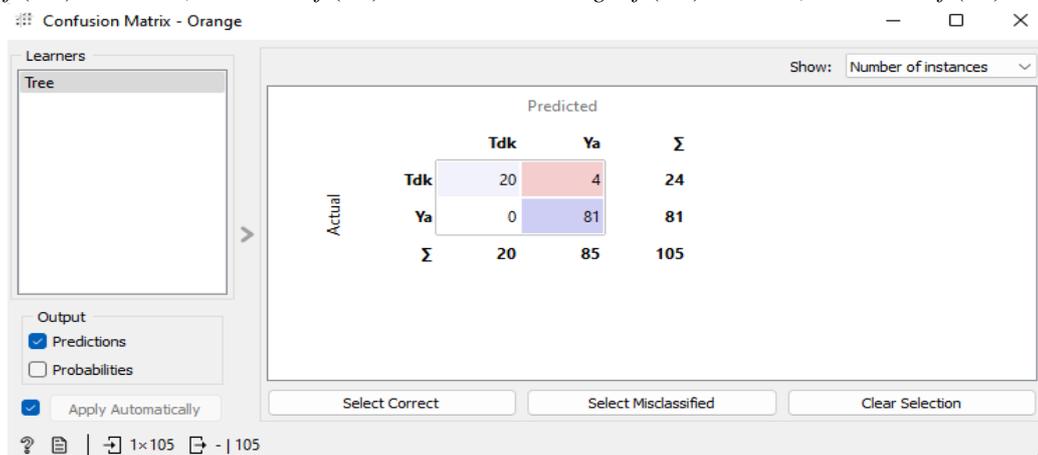
5. Jika IMT = Gemuk atau Rendah, Keputihan = Ya maka *class* = Prematur
6. Jika IMT = Normal maka *class* = Tidak prematur

Evaluasi hasil Pengujian tingkat akurasi dengan prediksi *decision tree* dapat dilihat dari hasil test and score dengan perhitungan *accuracy*, *precision*, *recall*. Hasil penelitian menunjukkan bahwa nilai test and score dapat dilihat pada gambar 5 yang diperoleh menggunakan metode *decision tree* yaitu nilai AUC 0,917 (91,7%) artinya sangat baik karena hampir mencapai 100%, nilai CA 0,962 (96,2%), nilai F1 0,961 (96,1%), nilai precision 0,964 (96,4%), dan nilai recall 0,962 (96,2%) seperti Gambar 5.



Gambar 5. Test and score orange data mining

Evaluasi *decision tree* dapat dilihat dengan nilai *Confusion matrix*. Gambar 6 menjelaskan nilai, *True Negatif* (TN) adalah 81, *True Positif* (TP) adalah 20, *False Negatif* (FN) adalah 0, *False Positif* (FP) adalah 4.



Gambar 6. Confusion Matrix Decision Tree

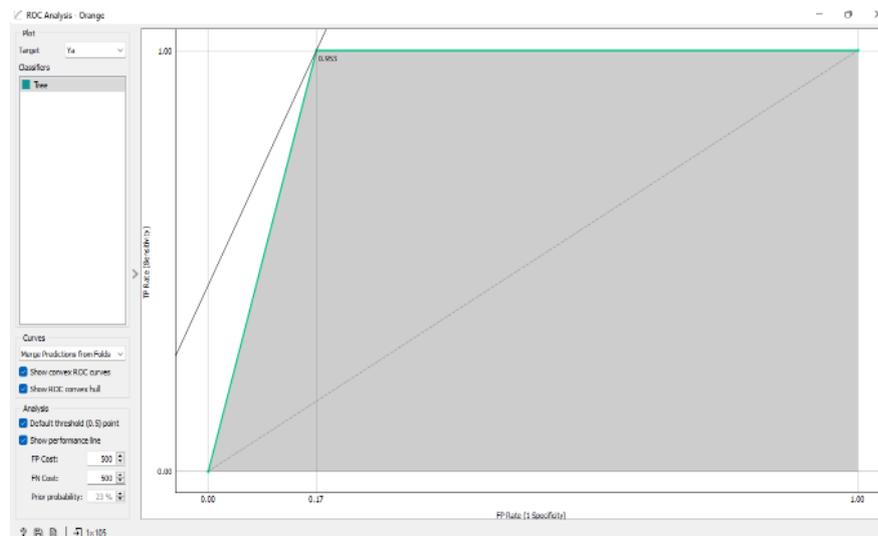
Berdasarkan *confusion matrix* pada gambar 6 maka nilai *Accuracy*, *Precision*, dan *Recall* sebagai berikut:

$$Accuracy = \frac{20+81}{20+81+4+0} \times 100\% = 96,2$$

$$Precision = \frac{20}{20+4} \times 100\% = 83,3$$

$$Recall = \frac{20}{20+0} \times 100\% = 100$$

Dari hasil perhitungan, dapat disimpulkan bahwa model *decision tree* memiliki tingkat *Accuracy* yang tinggi, yaitu 96,2%. Pada *Precision* memiliki nilai 83,3% dan pada *Recall* memiliki nilai 100%. Selain menggunakan confusion matrix, pengujian akurasi dapat dilakukan dengan melihat ROC curve. ROC curve adalah sebuah kurva yang digunakan untuk menilai sebuah model yang membedakan contoh positif dan negatif, dan untuk mengidentifikasi ambang batas terbaik dalam memisahkannya. ROC curve membandingkan nilai *true positive rate* dengan nilai *false positive rate*. Pada gambar 7 dapat dilihat bahwa hasil analisis ROC untuk prediksi kelahiran bayi prematur adalah 0,953. Model *decision tree* dalam memprediksi kelahiran bayi prematur memiliki nilai akurasi yang baik karena kurvanya mendekati titik 1.



Gambar 7. ROC Curve

IV. KESIMPULAN

Hasil prediksi dari model algoritma C4.5 berbasis *Decision tree* menunjukkan tingkat akurasi 91,7%. Dengan tingkat akurasi yang telah tercapai *good classification* artinya *decision tree* adalah model yang cukup efektif. Meskipun tingkat akurasinya tinggi, namun perlu peningkatan dan penyempurnaan model untuk meningkatkan prediksi kelahiran bayi prematur. Dalam konteks ini disarankan adanya penambahan atribut lain seperti riwayat merokok dan malnutrisi. Algoritma C4.5 bisa dilakukan perbandingan dengan model yang lain guna mendukung pengujian data yang tersedia, agar dapat tingkat akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] W. Aprilia, "Perkembangan pada masa pranatal dan kelahiran," *Yaa Bunayya J. Pendidik. Anak Usia Dini*, vol. 4, no. 1, pp. 40–55, 2020. Available: <https://jurnal.umj.ac.id/index.php/YaaBunayya/article/download/6684/4246>
- [2] F. O. Shariff *et al.*, "G5P4A0 Hamil 28 Minggu dengan Partus Prematurus Imminens (PPI) G5P4A0 28 Weeks Pregnant with Imminens Premature Parturition," vol. 14, no. April 2024, pp. 773–780, 2022, doi: 10.53089/medula.v14i4.1020.
- [3] M. Mirawati, Pratiwi Puji Lestari, Rr. Sri Nuriaty Masdiputri, and Mahfuzhah Deswita Puteri, "Faktor yang Berhubungan dengan Berat Bayi Lahir Rendah," *Sehat Rakyat J. Kesehat. Masy.*, vol. 2, no. 1, pp. 91–98, Feb. 2023, doi: 10.54259/sehatrakyat.v2i1.1502.
- [4] M. Warliani, Nilla Mayasari, and Ferius Soewito, "Mengenal Masalah Oromotor pada Bayi Prematur." *J. Indones. Med. Assoc.*, vol. 70, no. 12, pp. 278–286, 2021, doi: 10.47830/jinma-vol.70.12-2020-157.
- [5] Nesty Ladziina Sheliha, "Faktor – faktor yang mempengaruhi kejadian persalinan prematur," *UNISA Digit.*, p. 9, 2020. Available: <http://digilib.unisayogya.ac.id/5211/#>
- [6] F. Rachman, "Kematian Bayi Di Indonesia Akibat Lahir Prematur," 2024, 2022. <https://validnews.id/nasional/kemenkes-84-kematian-bayi-di-indonesia-akibat-lahir-prematur>
- [7] U. Hasdiana, *Gizi Pada Ibu Hamil*, vol. 11, no. 1. 2022. [Online]. Available: <http://link.springer.com/10.1007/978-3-319-59379-1%0A>
- [8] A. Usman, Rosdiana, and A. Misnawati, "Faktor Risiko Kejadian Persalinan Prematur Di Rumah Sakit Umum Polewali Tahun 2021," *J. Kesehat. Lentera Acitya*, vol. 8, no. 2, pp. 63–68, 2021. Available:

- <https://lppmfatimaparepare.org/index.php/acitya/article/view/73/60>
- [9] I. Rahim, R. Fitriani, A. W. Gama, A. Rahman, and Z. Alwi, "Analisis Faktor Risiko Kejadian Persalinan Prematur Di RSUD Haji Makassar Tahun 2021," *J. Kedokt. dan Kesehat.*, vol. 19, no. 2, p. 132, 2023, doi: 10.24853/jkk.19.2.132-145.
- [10] P. Andika, D. A. Ningsi, and R. M. Sari, "Faktor-Faktor Yang Mempengaruhi Persalinan Prematur Di Rsud Dr.M.Yunus Bengkulu," *PREPOTIF J. Kesehat. Masy.*, vol. 7, no. April, pp. 944–950, 2023. Available: <https://journal.universitaspahlawan.ac.id/index.php/prepotif/article/view/13361/11027>
- [11] S. E. D. Simamora and S. Ronoatmodjo, "Program Magister Epidemiologi, Fakultas Kesehatan Masyarakat, Universitas Indonesia 2 Departemen Epidemiologi, Fakultas Kesehatan Masyarakat, Universitas Indonesia Kampus Baru UI Depok , Jawa Barat – 16424 Indonesia," *J. Kesehat. Reproduksi*, vol. 11, no. 1, pp. 89–100, 2020, doi: 10.22435/kespro.v11i1.2753.89-100.
- [12] A. Y. Simanjuntak, I. S. E. S. Simatupang, and A. Anita, "Implementasi Data Mining Menggunakan Metode Nave Bayes Classifier Untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan," *J. Sci. Soc. Res.*, vol. 5, no. 1, p. 85, 2022, doi: 10.54314/jssr.v5i1.804.
- [13] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/37650>
- [14] C. O. S. Patricia, "Memprediksi Risiko Penyakit Jantung Menggunakan Metode *Decision tree* Dengan Validasi Silang," vol. 3, no. 2, p. 6, 2021. Available: <https://repository.unsri.ac.id/41280/>
- [15] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," *BIIKMA Bul. Ilm. Ilmu Komput. dan Multimed.*, vol. 99, no. 99, pp. 216–219, 2022. Available: <https://jurnalmahasiswa.com/index.php/biikma/article/view/101/101>
- [16] A. Purwanto, A. Primajaya, and A. Voutama, "Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 4, p. 390, 2020, doi: 10.26418/justin.v8i4.41959.
- [17] R. Fitra Arkamil and M. I. Jambak, "Klasifikasi Tindakan Persalinan Pada Pasien Ibu Bersalin Menggunakan Metode *Decision tree* C4.5," *Media Online*, vol. 4, no. 1, pp. 513–523, 2023, doi: 10.30865/klik.v4i1.1168.
- [18] H. Saleh, "Analisa Faktor Penyebab Stunting Menggunakan Algoritma C4.5," *Sci. Comput. Sci. Informatics*, vol. 3, no. 1, pp. 11–17, 2020. Available: <http://jurnal.untad.ac.id/jurnal/index.php/scientico/article/view/15904#>
- [19] N. Sunanto and G. Falah, "Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 208–216, Jul. 2022, doi: 10.36341/rabit.v7i2.2435.
- [20] S. S. Mukrimaa *et al.*, "Data Mining Algoritma C4.5," *J. Penelit. Pendidik. Guru Sekol. Dasar*, vol. 6, no. August, p. 128, 2016. Available: http://lib.unnes.ac.id/33080/6/Buku_Data_Mining.PDF