

Klasifikasi Persalinan Prematur Menggunakan Perbandingan Algoritma C4.5 dan Random Forest

Handini Arga Damar Rani ^{✉1}, Mohammad Burhan Hanif ², Afis Pratama ³, R. Irlanto Sudomo ⁴
Universitas Ivet, Indonesia^{1,3,4}, Universitas Semarang²

Info Articles

Keywords:
Klasifikasi; C4.5;
Random Forest;
Persalinan Prematur;

Abstrak

Persalinan prematur adalah kondisi kelahiran bayi kurang dari 37 minggu dengan berat janin kurang dari 2500 gram. Dimana tingkat persalinan bayi premature yang berujung pada kematian meningkat dari tahun 2000 sampai 2014 sebanyak 8.5%. Klasifikasi untuk penanganan persalinan premature secara dini merupakan salah satu solusi yang sering diteliti. Beberapa algoritma klasifikasi yang sering digunakan adalah C4.5 dan Random Forest. Dari kedua algoritma tersebut akan dipilih mana algoritma terbaik dengan melihat pada tingkat akurasi yang paling tinggi yang akan dipilih. Dari hasil perhitungan algoritma Random Forest didapatkan akurasi sebesar 99.38% dengan AUC 0.988. Sedangkan capaian dari algoritma C4.5 sebesar 98.75% dengan AUC 0.5. Maka dari hasil itu dapat dibandingkan dengan nilai akurasi kedua algoritma disimpulkan bahwa algoritma Random Forest lebih baik akurasinya untuk mengatasi data persalianan premature jika disbanding dengan algoritma C4.5.

Abstract

Premature labor is a condition of the birth of a baby less than 37 weeks with a fetal weight of less than 2500 grams. Where the rate of premature infant delivery leading to death increased from 2000 to 2014 by 8.5%. Classification for early handling of premature labor is one of the solutions that is often studied. Some of the classification algorithms that are often used are C4.5 and Random Forest. From the two algorithms, the best algorithm will be selected by looking at the highest level of accuracy to be selected. From the calculation results of the Random Forest algorithm, an accuracy of 99.38% was obtained with an AUC of 0.988. Meanwhile, the achievement of the C4.5 algorithm is 98.75% with an AUC of 0.5. Therefore, from the results that can be compared with the accuracy values of the two algorithms, it is concluded that the Random Forest algorithm is better accurate to overcome premature data when compared to the C4.5 algorithm

[✉] Alamat Korespondensi:
E-mail: hanifburhan@usm.ac.id

PENDAHULUAN

Persalinan prematur adalah kondisi kelahiran bayi dimana berat janin kurang dari 2500 gram dengan usia kehamilan kurang dari 37 minggu (Mustika & Fika Minata, 2021). Persalinan prematur merupakan kondisi yang berbahaya karena akan meningkatkan resiko kemungkinan komplikasi medis bagi bayi terutama pada masalah pernafasan terhadap bayi yang akan mengalami kondisi penyakit yang dinamakan ikterus (Irianti et al., 2019). Kejadian bayi persalinan prematur di dunia ini mempunyai peningkatan yang cukup tinggi dimana pada tahun 2000 hingga 2014 terdapat 8.5% tingkat kenaikannya dan tingkat kematian bayi meninggal sebesar 35% (Chawanpaiboon et al., 2019). Di Indonesia sendiri pada tahun 2018 WHO mencatat ada sebanyak 675.700 terjadinya kelahiran bayi prematur (Ananda Ika Nuriza, Semi Na'im, 2020).

Persalinan bayi prematur merupakan sumber utama dari terjadinya morbiditas dan mortalitas neonatal yang di prosentasekan sebanyak 60-80 % diseluruh dunia dan ini cukup memprihatinkan (Yetti Anggraini, 2014). Banyak sekali faktor pemicu kenapa persalinan prematur bisa terjadi hal ini dipacu oleh penyakit lain seperti penyakit tekanan darah baik darah tinggi maupun darah rendah atau penyakit menular lainnya, selain itu dipengaruhi juga oleh faktor usia bayi dan kandungan serta masih banyak faktor lainnya (Carolin Bunga Tiara, 2019). Dengan melihat dampak dari persalinan bayi prematur ini kiranya kita perlu untuk melakukan penjegahan sejak awal agar persalinan bayi prematur ini tidak terjadi. Dengan adanya banyak data faktor penyebab persalianan prematur kita dapat mengklasifikasinya agar kita bisa mendeteksi kemungkinan lahir bayi prematur sejak dini.

Klasifikasi penanganan persalinan bayi prematur dalam ilmu data mining sudah banyak dilakukan oleh beberapa peneliti terdahulu seperti yang diteliti oleh anas dan utomo. Mereka melakukan penelitian penanganan masalah data persalinan dengan menggunakan beberapa algoritma *supervise learning* yaitu C4.5, Naïve Bayes, Support Vector Machine, Neural Networks dan K- Nearest Neighbor. Dari beberapa algoritma tersebut dilakukanlah perbandingan dan dipilih mana yang terbaik yang bisa menangani data persalinan bayi prematur. Hasil dari perbandingan beberapa metode tersebut memperoleh bahwa algortima yang cocok untuk penangan data persalinan bayi prematur adalah alogritma C4.5. Dengan segala kelemahan yang ada pada algortima C4.5 ini mampu menangi data persalinan bayi dengan sangat baik dibanding algoritma lainnya karena memiliki beberapa kelebihan dibanding algoritma lainnya (Gading Pertiwi & Pujiyanto, 2020).

Penelitian data persalinan bayi prematur juga dilakukan oleh mohammad burhan hanif dan tim. Penelitian yang dilakukan oleh hanif ini membandingkan algrpitma C4.5 dengan algoritma naïve bayes. Algoritma naïve bayes mempunyai kelebihan dalam penangan set attribute data persalinan bayi prematur. Sedangkan algoritma C4.5 memiliki kemampuan yang baik juga dalam penangan data persalinan bayi prematur. Dari kedua algoritma tersebut maka di ujicoba untuk memilih algoritma terbiak dalam penangan bayi prematur. Dari hasil perbandingan tersebut terdapat nilai akurasi algoritma C4.5 sebanyak 98.75% dengan AUC 0.5. Sedangkan nilai akurasi dari algoritma naïve bayes yaitu

sebesar 81.88% dan AUC 0.94. Dari perbandingan tersebut munculah hasil terbaik yaitu algoritma C4.5 dalam penanganan data persalinan bayi prematur (Hanif et al., 2022).

Penelitian lain tentang persalinan prematur dilakukan oleh jefi. Untuk penelitiannya kali ini jefi menggunakan algoritma C4.5 yang dioptimasi menggunakan metode particle swarm optimization dalam penanganan data persalinan bayi pada klinik umi. Hasil dari penelitiannya ini diperoleh besaran nilai akurasi sebesar 94.30% dengan nilai AUC yaitu 0.986. Penelitian ini berhasil memecahkan masalah yang ada pada data persalinan bayi yang diperoleh di klinik umi (Jefi, 2019).

Penelitian relevan tentang persalinan bayi prematur dilakukan juga oleh irzan fajari nurahmadan beserta tim. Penelitiannya berfokus dengan klasifikasi denyut jantung janin bayi untuk mencegah terjadinya persalinan bayi prematur. Beberapa algoritma klasifikasi yang dipakai yaitu menggunakan algoritma Random Forest. Hasil pencapai nilai akurasi algoritma Random Forest sebesar 94.36% lebih besar dibanding algoritma lain dengan parameter tes size sebanyak 0.2 (Nurahmadan et al., 2021).

Penelitian relevan tentang persalinan bayi prematur juga dilakukan oleh mambang dan agus byna. Penelitiannya bercerita tentang klasifikasi kecemasan ibu hamil untuk mencegah persalinan bayi prematur. Salah satu algoritma klasifikasi yang digunakan dalam penelitian ini yaitu Random Forest yang mempunyai keunggulan dalam pohon keputusan yang menggunakan vector acak. Hasil pengukuran dari algoritma random forest ini diperoleh akurasi sebesar 62.6%. Hasil ini dinilai lebih baik dari pada hasil algoritma klasifikasi lain (Mambang & Byna, 2017).

Persalinan prematur mempunyai banyak dampak yang berbahaya seperti infeksi endometrium dan luka episiotomi yang mengalami kelambatan dalam fase penyembuhan. Sedangkan dampak yang lebih berbahaya adalah terhadap bayi itu sendiri yang bisa menyebabkan resiko kematian lebih tinggi. Maka dari itu klasifikasi resiko persalinan bayi prematur haruslah dilakukan sejak awal mungkin (Mustika & Fika Minata, 2021).

Perhitungan untuk mengklasifikasi suatu data pada data mining ada beberapa metode yang umum digunakan. Dalam dunia medis metode algoritma decision tree C4.5 biasa dan sering digunakan. Hal ini disebabkan karena algoritma decision tree C4.5 ini memiliki kelebihan dalam hal klasifikasi yaitu nilai akurasi yang cukup tinggi. Selain nilai akurasi yang tinggi, kinerja komputasi atau perhitungan algoritma decision tree C4.5 ini juga tergolong cepat (Xu et al., 2021). Dengan kemampuan kecepatan perhitungan komputasi seperti itu algoritma C4.5 sangat bisa dipakai untuk klasifikasi persalinan prematur.

Dalam klasifikasi bidang penyakit dan kesehatan ada algoritma lain selain algoritma C4.5 yang kerap digunakan yaitu salah satunya algoritma Random Forest. Algoritma Random Forest merupakan bentuk umum dari prinsip ansambel acak dari suatu pohon keputusan (Wuryani & Agustian, 2021). Random Forest ini mempunyai kelebihan yaitu dapat mengatasi data yang resisting outlier, efisien dalam penyimpanan data juga handal dalam mengatasi missing value (Supriyadi et al., 2020). Random Forest ini juga memiliki keunggulan yaitu dapat memilih fitur – fitur terbaik dan membuang fitur yang jelek sehingga tingkat akurasi dan performa akan meningkat (Devella et al., 2020).

Algoritma C4.5 dan Algoritma Random Forest sering digunakan dalam dunia kesehatan hal ini maka pantas juga di terapkan dalam kasus klasifikasi persalinan prematur. Dengan keunggulan kedua algoritma itu maka perlu dilakukan komparasi antar algoritma dalam menentukan klasifikasi data persalinan prematur. Dengan komparasi algoritma C4.5 dan Random Forest ini diharapkan dapat muncul algoritma yang terbaik sehingga mampu membantu tenaga medis dalam mengatasi masalah klasifikasi persalinan prematur.

METODE

Tahapan metode dalam penelitian ini memiliki beberapa proses, yang pertama diawali dengan pengumpulan data. Tahapan selanjutnya adalah dilakukannya pre-prosesing data dengan menyesuaikan fitur pada posisinya. Tahapan berikutnya adalah tahap klasifikasi dengan membandingkan kedua metode yaitu metode C4.5 akan di komparasi dengan metode Random Forest. Setelah tahapan itu dilanjutkan dengan pengujian model dimana model akan diuji kelayakan kinerjanya. Langkah terakhir yaitu evaluasi dan validasi hasil untuk kedua model algoritma tersebut. Tahapan model secara mendetail dijelaskan berikut ini :

1. Pengumpulan Data

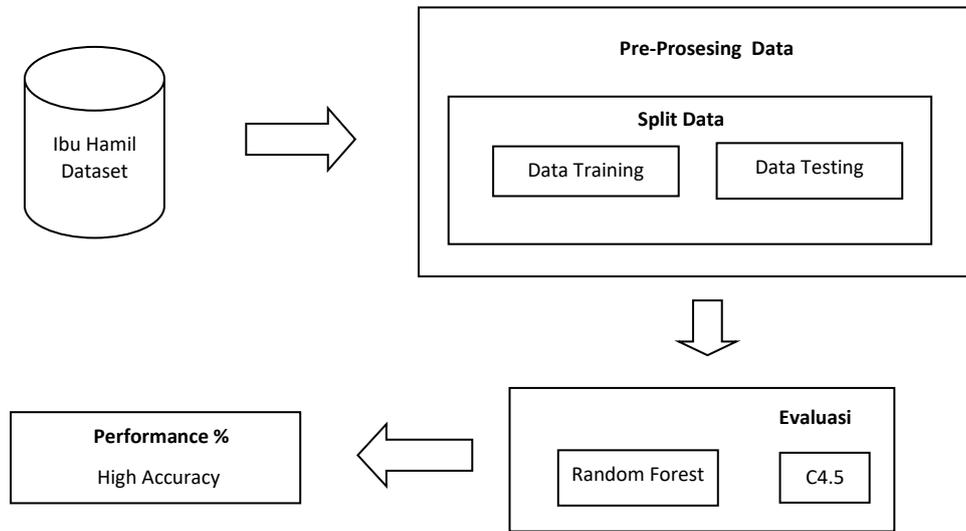
Pengumpulan data adalah penentuan data yang dipakai. Untuk penelitian yang dilakukan menggunakan data sekunder hasil observasi serta hasil wawancara di klinik bersalin pada kota kudas. Jumlah baris data sebanyak 160 dengan 11 fitur yang didalamnya terdapat 1 kelas fitur. Jenis fiturnya terdiri dari tanggal, no_rm, nama, nama_suami, alamat, tekanan darah, berat badan, usia ibu bayi, letak janin, tfu dan kelas.

2. Pre-Prosesing Data

Pre-prosesing data yaitu mengolah data mentah untuk dilakukan beberapa persiapan sebelum digunakan. Pengolahan biasanya terkait dengan data cleaning missal dari data noise atau missing value. Pengolahan lain bisa juga dengan data integration yaitu menyambungkan antar data. Pengolahan lain bisa untuk seleksi data yaitu memilih dan memilih data yang akan dipakai. Selain itu tahapan ini di gunakan juga untuk pemisahan data latih dan juga data uji.

3. Metode yang diusulkan

Metode yang diusulkan dalam penelitian ini secara garis besar akan membandingkan kinerja dari dua algoritma klasifikasi. Fokus utama algoritma terletak pada algoritma decision tree C4.5 dengan algoritma Random Forest tentunya dengan penggunaannya terhadap data persalinan bayi prematur. Dalam tahapan metode akan mengacu pada kerangka kerja model yang akan didetailkan pada gambar 1 berikut ini :



Gambar 1. Kerangka Model kerja yang diusulkan

4. Pengujian Model

Pengujian model terhadap kedua algoritma yaitu algoritma decision tree C4.5 dengan algoritma Random Forest menitik beratkan pada capaian kinerja yang dihasilkan. Salah satu capaian kinerja yang dihasilkan yang akan di uji adalah besaran nilai akurasi yang dihasilkan dari algoritma decision tree C4.5 dengan algoritma Random Forest.

5. Evaluasi dan Validasi

Evaluasi dan Validasi yang dalam hal ini merupakan tahap akhir dari proses penelitian. Evaluasi dan validasi yang dilakukan akan menggunakan confusion matrik untuk menghitung akurasi algoritma. Hasil dari akurasi yang diperoleh merupakan nilai perbandingan antara algoritma decision tree C4.5 dengan algoritma Random Forest.

HASIL DAN PEMBAHASAN

Hasil dan pembahasan akan disesuaikan dengan apa yang dijelaskan pada tahapan metodologi penelitian. Tahapan langkah demi langkah akan dilaksanakan secara runtut dan terperinci seperti berikut ini :

1. Data Penelitian

Dataset dalam penelitian ini menggunakan data ibu hamil dengan 11 fitur yang didalamnya terdapat 1 label. Sedangkan baris datanya sebanyak 160 baris data yang terpampang tampak pada tabel 1 berikut ini :

Tabel 1. Data Ibu Hamil

No	Tanggal	No_rm	Nama	Suami	TFU	Kelas
1	6-1-2012	Pasien_1	Mu'alifatu	Hadi	Sesuai	Normal
2	6-1-2012	Pasien_2	Yunita	Efendi	Sesuai	Normal
3	6-1-2012	Pasien_3	Elis	Yanto	Tidak Sesuai	Resiko
4	6-1-2012	Pasien_4	Miftakul	Deni	Sesuai	Normal
5	6-1-2012	Pasien_5	Suyanti	Agus	Sesuai	Normal
6	6-1-2012	Pasien_6	Faidloh	Arifin	Sesuai	Resiko
7	6-1-2012	Pasien_7	Amelia	Fais	Sesuai	Resiko
8	6-1-2012	Pasien_8	Nanik	Saikun	Sesuai	Normal
9	6-1-2012	Pasien_9	Suhartatik	Mustaqim	Sesuai	Normal
....
....
158	9-10-2012	Pasien_158	Fitriyah	Santosa	Sesuai	Normal
159	9-10-2012	Pasien_159	Kurniasih	Badrudin	Tidak sesuai	Resiko
160	9-10-2012	Pasien_160	Yuniatun	Aji	Sesuai	Resiko

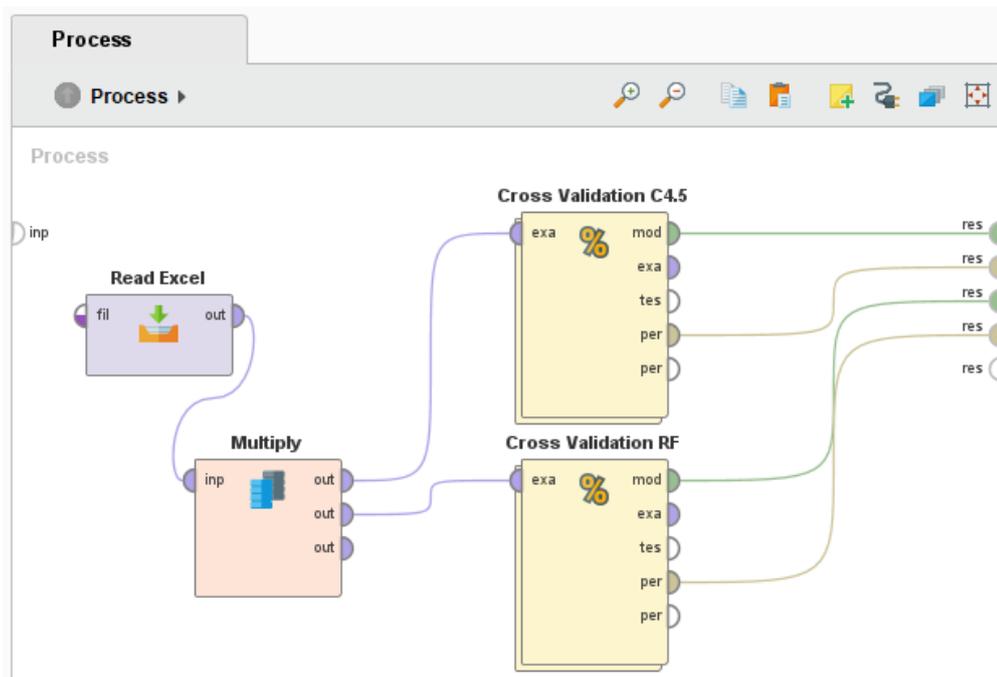
Setelah data sudah di dapatkan yaitu data ibu hamil dengan banyak baris data sebesar 160 . Maka langkah berikutnya yaitu melakukan pre-prosesing data. Tahapan preprosesing data ini akan melakukan pembersihan data di lanjut dengan integrasi data dan terakhir seleksi data.

2. Penerapan metode dan pengujian model

Selanjutnya setelah data yang sudah kita preprosesing sudah siap kita gunakan maka kita sudah bisa memakainya untuk di terapkan kepada kedua algoritma. Algoritma pertama yang akan menerapkan data tersebut adalah algoritma decision tree C4.5 dengan fasenya sebagai berikut:

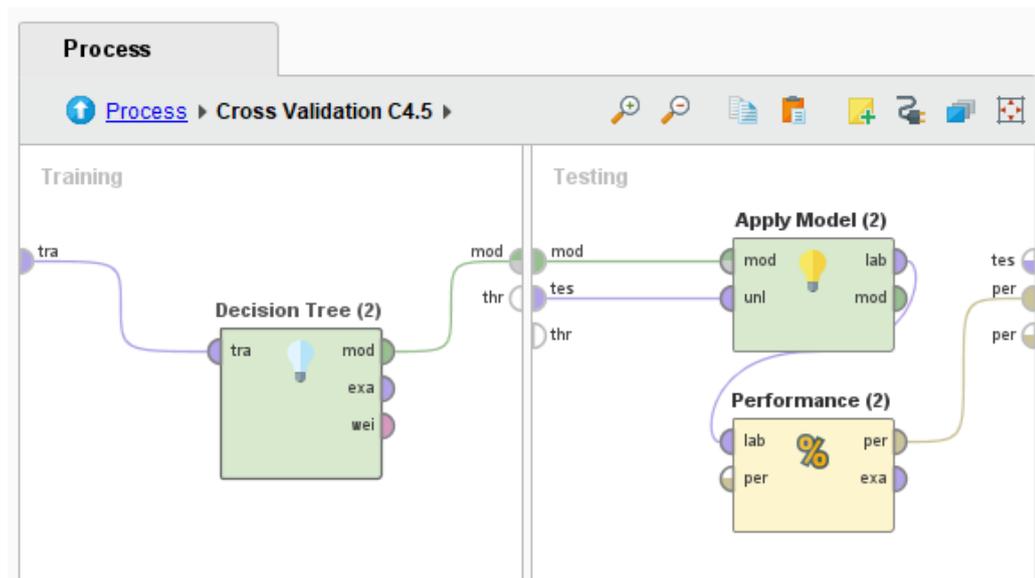
- a. Menghitung hasil dari penjumlahan data dengan dilandasi banyaknya atribut hasil dengan mengacu pada syarat yang ditetapkan.
- b. Setelahnya atribut akan dipakai sebagai node dengan melihat nilai gian yang dirasa paling tinggi.
- c. Menentukan percabangan yang akan dipakai untuk semua anggota node.
- d. Memilah mana anggota node yang bernilai 0 maka inilah yang akan dijadikan daun pohon keputusan.
- e. Melakukan pengulangan langkah sebelumnya sampai semua nilai entropy dari Node bernilai nol.
- f. Apabila nilai entropynya lebih dari nol, maka proses sebelumnya akan dilakukan berulang dari awal hingga semua Node bernilai nol.

Perbandingan algoritma C4.5 dengan algoritma Random Forest ini akan dilakukan dengan bantuan software rapidminer seperti nampak pada gambar 2 berikut ini:



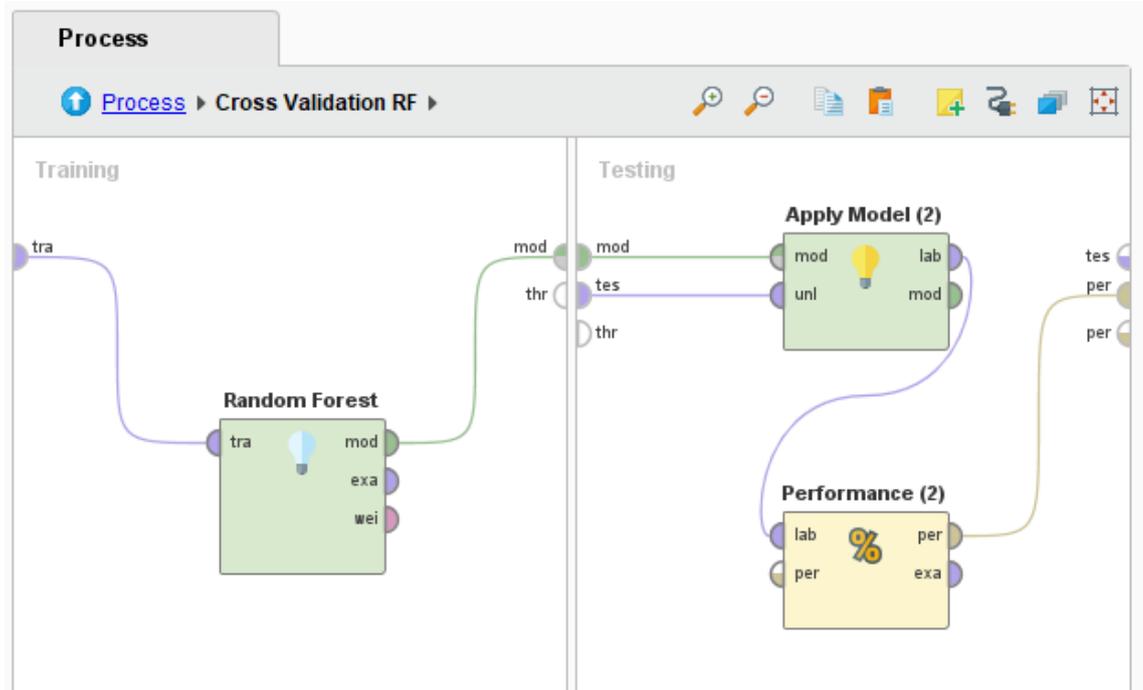
Gambar 2. Model perbandingan algoritma

Pada gambar 2 model perbandingan algoritma menggunakan satu dataset yang akan langsung di gunakan. Dari dataset itu selanjutnya dipakai untuk klasifikasi algoritma decision tree C4.5 dan juga algoritma Random Forest. Penggunaan satu dataset yang langsung di gunakan oleh kedua algoritma tersebut berfungsi untuk menjaga konsistenitas dataset yg sama. Sedangkan gambaran lebih rinci model training dan testing algoritma decision tree diperlihatkan pada gambar 3 berikut ini:



Gambar 3. Model proses algoritma C4.5

Model training dan testing secara rinci pada algoritma Random Forest juga bisa terlihat pada gambar 4 berikut ini .



Gambar 4. Model proses algoritma Random Forest

3. Evaluasi dan Validasi Model

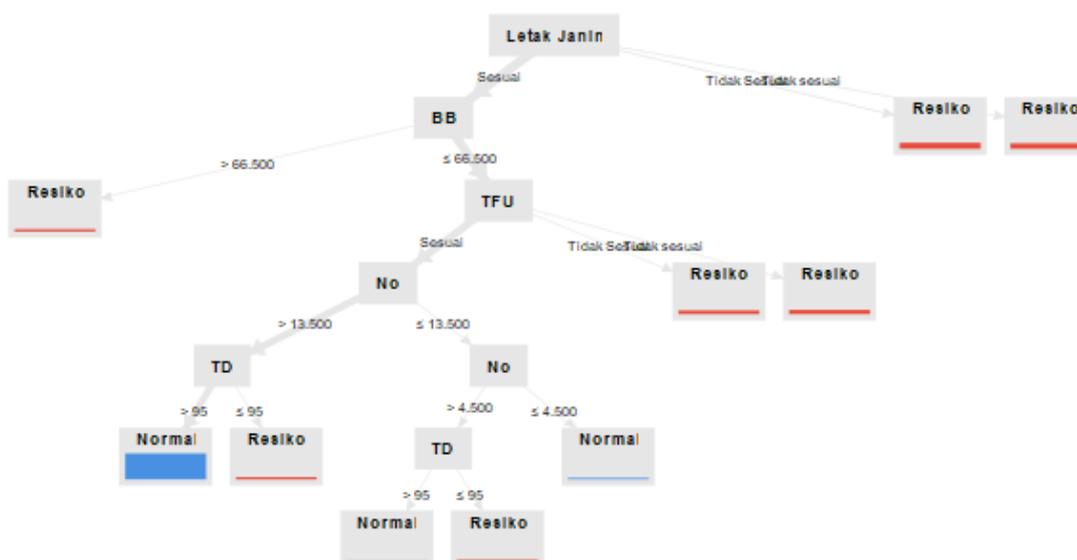
Hasil pengukuran dan validasi memakai metode cross-validation yang sudah ada di aplikasi rapidminer. Besaran nilai k adalah 10 yang di gunakan untuk pengujian cross validation. Sedangkan untuk pembagian dataset latih dan dataset uji perbandinganya yaitu 9 berbandi 1. Selanjutnya iterasi uji perulangan dilakukan sebanyak 10 kali sesuai dengan jumlah k yang sudah ditentukan diawal.

Proses iterasi perulangan akan menghasilkan nilai performa dari masing – masing metode klasifikasi yang digunakan. Untuk metode algoritma decision tree C4.5 selain didapati performa dihasilkan pohon keputusan seperti gambar 5.



Gambar 5. Pohon Keputusan algoritma C4.5

Sedangkan untuk metode algoritma Random Forest menghasilkan luaran berupa pohon keputusan juga sesuai dengan gambar 6.



Gambar 6. Pohon Keputusan algoritma Random Forest

Tabel confusion matriks dihasilkan dari cross-validation yang didalamnya mengandung nilai akurasi, juga recall, dan AUC. Hasil nilai confusion matriks algoritma C4.5 ter gambar pada tabel 2 sedangkan untuk algoritma Random Forest terlihat pada tabel 3.

Tabel 2. Confusion Matriks C4.5

		Actual Class	
		+ (Positive)	- (Negative)
Prediction class	+ (positive)	93	2
	- (negative)	0	65

Tabel 3. Confusion Matriks Random Forest

		Actual Class	
		+ (Positive)	- (Negative)
Prediction class	+ (positive)	93	1
	- (negative)	0	66

Dari tabel confusion matriks pada tabel 2 dan tabel 3 selanjutnya di mencari nilai dari akurasi, presisi, recall dan AUC nya. Hasil perhitungan nilai didapati pada tabel 4.

Tabel 4. Hasil Perhitungan Akurasi Metode Klasifikasi

Metode	Akurasi	AUC
C4.5	98.75 %	0.500
Random Forest	99.38 %	0.988

SIMPULAN

Hasil dari proses perhitungan akhir didapatkan besaran akurasi sebanyak 98.75% pada metode algoritma decision tree C4.5. Sedangkan besaran akurasi untuk algoritma random Forest sebesar 99.38%. Untuk nilai AUC dari algoritma C4.5 didapatkan nilai 0.5 sedangkan AUC untuk algoritma Random Forest sebesar 0.988. Jadi hasil akhir dapat dikatakan bahwa dalam klasifikasi persalinan bayi prematur pada data private ibu hamil diklinik kodus algoritma Random Forest lebih tinggi akurasinya dibanding algoritma C4.5.

Untuk penelitian lebih lanjut dapat olah lagi dengan menerapkan banyak dataset public pada algoritma klasifikasi. Karena bisa saja dengan penggunaan dataset yang beragam akan diperoleh data akurasi yang sama. Oleh sebab itu dengan digunakannya banyak dataset diharapkan perbandingan algoritma klasifikasi Random Forest dengan algoritma C4.5 dapat diukur secara lebih tepat.

DAFTAR PUSTAKA

- Ananda Ika Nuriza, Semi Na'im, A. H. (2020). Hubungan Preeklampsia dengan Persalinan prematur di RSUD Jombang Bulan Juni Tahun 2019. *Jurnal Akademika Husada*, II(1), 58–71.

- Carolin Bunga Tiara, W. I. (2019). Faktor-Faktor Yang Berhubungan Dengan Kejadian Persalinan Preterm Di Rumah Sakit Muhammadiyah Taman Puring Kebayoran Baru Jakarta Selatan Periode Januari-Juni Tahun 2017. *Jurnal Ilmu Keperawatan Dan Kebidanan Nasional*, 1(1), 12.
- Chawanpaiboon, S., Vogel, J. P., Moller, A. B., Lumbiganon, P., Petzold, M., Hogan, D., Landoulsi, S., Jampathong, N., Kongwattanakul, K., Laopaiboon, M., Lewis, C., Rattanakanokchai, S., Teng, D. N., Thinkhamrop, J., Watananirun, K., Zhang, J., Zhou, W., & Gülmezoglu, A. M. (2019). Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *The Lancet Global Health*, 7(1), e37–e46. [https://doi.org/10.1016/S2214-109X\(18\)30451-0](https://doi.org/10.1016/S2214-109X(18)30451-0)
- Devella, S., Yohannes, Y., & Rahmawati, F. N. (2020). Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 310–320. <https://doi.org/10.35957/jatisi.v7i2.289>
- Gading Pertiwi, A., & Pujiyanto, U. (2020). *Metode-metode data mining untuk penyelesaian masalah kehamilan dan persalinan*. 2(1), 11.
- Hairul Umam, M., Wahanggara, V., Cahyanto, T. A., & Muharom, L. A. (2017). Analisis Perbandingan Algoritma C4.5 Dan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Prodi Teknik Informatika Universitas Muhammadiyah Jember). *Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember*, 1310651100, 1–9.
- Hanif, M. B., Arga, H., Rani, D., Rifai, A., & Guntoro, G. (2022). Komparasi Metode Naive Bayes dan C4 . 5 Pada Klasifikasi Persalinan Prematur. *JOINED JOURNAL*, 5(1), 54–65.
- Irianti, A. P., Mudrikatin, S., & Sari, W. A. (2019). Hubungan Persalinan Prematur dengan Kejadian Ikterus di RSUD Jombang Bulan Januari-Juni Tahun 2019. *Jurnal Akademia Husada*, 1(2), 14–26.
- Jefi. (2019). Prediksi Bayi Lahir Secara Prematur Dengan Menggunakan Metode C . 45 Berbasis Particle Swarm Optimization Pada Klinik Umi. *Indonesian Journal on Networking and Security*, 8(2), 1–7.
- Mambang, & Byna, A. (2017). Analisis Perbandingan Algoritma C.45, Random Forest Dengan Chaid Decision Tree Untuk Klasifikasi Tingkat Kecemasan Ibu Hamil. *Semnasteknomedia Online*, 5(1), 103–108.
- Mustika, E., & Fika Minata. (2021). Analisis Hubungan Faktor Maternal Dan Penyakit Kronik Pada Persalinan Prematur. *Jurnal Kesehatan Dan Pembangunan*, 11(21), 19–27. <https://doi.org/10.52047/jkp.v11i21.94>
- Nurahmadan, I. F., Agusta, A., Winarno, P. A., Sazali, B. H., Thurfah, Y., & Rosaliah, A.

- (2021). Perbandingan Algoritma Machine Learning Untuk Klasifikasi Denyut Jantung Janin. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, April, 733–740.
- Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 67–75. <https://doi.org/10.51903/e-bisnis.v13i2.247>
- Syamsul, B., Dwi, M., & Rahmi, H. (2018). Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, B24–B31.
- Tyas, S. J. S., Febianah, M., Solikhah, F., Kamil, A. L., & Arifin, W. A. (2021). Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan. *Jurnal Teknologi Informasi Dan Komunikasi*, 8(1), 86–99.
- Wuryani, N., & Agustian, S. (2021). Random Forest Classifier untuk Deteksi Penderita COVID-19 berbasis Citra CT Scan. *Jurnal Teknik Komputer AMIK BSI*, 7(2), 187–193. <https://doi.org/10.31294/jtk.v4i2>
- Xu, Z., Shen, D., Nie, T., Kou, Y., Yin, N., & Han, X. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. *Information Sciences*, 572, 574–589. <https://doi.org/10.1016/j.ins.2021.02.056>
- Yetti Anggraini. (2014). Hubungan Antara Persalinan Prematur dengan Hiperbilirubin pada Neonatus. *Jurnal Kesehatan*, 5(2), 109–112.