

Komparasi Metode Naive Bayes dan C4.5 Pada Klasifikasi Persalinan Prematur

Mohammad Burhan Hanif^{✉1}, Handini Arga Damar Rani², Ahmad Rifai³, Gallet Guntoro Setiaji⁴

Universitas Semarang, Indonesia^{1,3,4}

Prodi Pendidikan Informatika, Fakultas Sains dan Teknologi, Universitas IVET, Indonesia²

Info Articles

Keywords:
Klasifikasi; C4.5; Naive Bayes; Persalinan Prematur;

Abstrak

Persalinan prematur adalah persalinan diluar masa lahir bayi yang menyebabkan kematian bagi bayi serta komplikasi terhadap ibu bayi. Juga menjadi beban tenaga medis dengan tren peningkatan sebanyak 8%. Klasifikasi data mining hadir sebagai pemecah masalah deteksi pencegahan awal persalianan premature. Dengan memanfaatkan algoritma klasifikasi C4.5 dan algoritma naïve bayes yang dianggap baik secara kinerja. Untuk memilih algoritma terbaik dalam klasifikasi persalinan prematur maka harus diukur dengan baik kinerjanya. Dari hasil pebandingan algoritma naïve bayes dengan algoritma C4.5 didapatkan akurasi sebesar 98.75% dengan AUC 0.5. Sedangkan capaian dari algoritma naïve bayes sebesar 81.88% dan AUC 0.945. Maka dari hasil perbandingan nilai akurasi kedua algoritma tersebut disimpulkan bahwa algoritma C4.5 mampu lebih unggul dalam penanganan data persalianan premature dibandingkan dengan algoritma naïve bayes.

Abstract

Preterm labor is a delivery outside the baby's birth period that causes death for the baby as well as complications to the baby's mother. It is also a burden on medical personnel with an increasing trend of as much as 8%. Data mining classification is present as a problem solver for the early prevention of premature memorization. By utilizing the C4.5 classification algorithm and the naïve bayes algorithm which are considered good in performance. To choose the best algorithm in the classification of preterm labor it must be well measured its performance. From the results of comparing the naïve bayes algorithm with the C4.5 algorithm, an accuracy of 98.75% was obtained with an AUC of 0.5. Meanwhile, the achievement of the naïve bayes algorithm was 81.88% and the AUC was 0.945. Therefore, from the results of the comparison of the accuracy values of the two algorithms, it is concluded that the C4.5 algorithm is able to be superior in handling premature memorization data compared to the naïve bayes algoritma

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

PENDAHULUAN

Persalinan prematur merupakan suatu persalinan diluar waktu normal dimana usia dari kehamilan diperkirakan terjadi sebelum 37 minggu dan lebih dari 20 minggu (Panada Sedianing Drastita et al., 2022). Kejadian persalinan prematur ini dapat menjadi penyebab utama kematian neonatus dimana angka kejadian persalinan prematur ini cukup tinggi (Ida Rahmawati et al., 2021). Kejadian ini terlihat dari tahun 2000 sampai 2014 dinilai tingkat kenaikan sebesar 8.5% dimana sebesar 35% kematian bayi baru lahir terjadi ketika proses kelahirannya melalui persalinan premature (Chawanpaiboon et al., 2019).

Persalinan prematur rentan sekali terhadap banyak komplikasi terhadap bayi seperti pendarahan intraventrikular, *respiratory distress syndrome* (RDS), *displasia bronkopulmoner* (BPD), *patem duktus arteriosus* (PDA), *retinopathy of prematurity* (ROP), dan *necrotizing entero cilitis* (NEC) dan masih banyak lagi (Loviana et al., 2019). Faktor penyebab terjadi persalinan prematur ada banyak diantaranya seperti penyakit menular, hipertensi, usia kandungan, usia ibu bayi, anemia, dan pendarahan antepartum (Carolin Bunga Tiara, 2019). Dengan adanya banyaknya data faktor penyebab terjadinya persalinan premature hal ini kita bisa memanfaatkan untuk mencari sebuah pola klasifikasi data mining untuk deteksi dini persalinan premature. Klasifikasi dini persalina premature ini diharapkan dapat menurunkan tingkat persalinan prematur sejak awal.

Klasifikasi persalinan premature dalam data mining sudah terlebih dahulu diteliti oleh banyak peneliti sebelumnya diantaranya seperti yang dilakukan oleh ari puspita. Penelitian ini melakukan prediksi kelahiran bayi premature dengan memanfaatkan bidang ilmu data mining. Dalam penelitian ini menggunakan algoritma C4.5 dan optimasi particle swarm optimization untuk melakukan klasifikasi data bayi premature. Hasil yang didapat bahwa akurasi klasifikasi dari persalinan premature sebesar 93.60% dengan nilai AUC sebesar 0.946. Dalam penelitian ini algortima C4.5 dianggap berhasil melakukan klasifikasi terhadap data persalinan premature (Puspita, 2018).

Penelitian lain yang relevan dengan klasifikasi persalinan bayi prematur juga dilakukan oleh jefi. Penelitanya berkaitan dengan penggunaan metode algoritma C4.5 di kombinasikan dengan metode particle swarm optimization yang di terapkan pada data private di sebuah pelayanan kesehatan bernama klinik umi. Pencapaian nilai akurasi terhadap data private klinik umi menunjukkan angka 94.30% dengan nilai AUC sebesar 0.986. penelitian ini dianggap berhasil secara akurat dapat digunakan untuk data private pada sebuah layanan kesehatan klinik umi (Jefi, 2019).

Peneliti selanjutnya di lakukan oleh Irmawati Carolina dan Kresna Ramanda. Penilitannya mengerjakan tentang klasifikasi kelahiran prematur menggunakan algoritma C4.5 . penelitian ini menggunakan data dari RSUPN Cipto Mangunkusumo yang memiliki 750 baris data dan memiliki 11 fitur. Dari hasil percobaan penelitian yang dilakukan didapatkan hasil yang memuaskan berupa besaran akurasi sebanyak 88.80% dengan nilai precision sebesar 89.81% dan nilai recall sebanyak 87.40%.

Sedangkan nilai AUC didapatkan sebesar 0.948 dimana pencapaian algoritma ini berhasil melakukan klasifikasi terhadap data pasien RSUPN (Carolina & Kresna, 2018).

Penelitian persalinan prematur juga dilakukan oleh handini arga damar rani, et.al. Penelitiannya tentang pembuatan sistem prediksi kondisi kelahiran bayi menggunakan metode klasifikasi naive bayes. Menggunakan data private yang diperoleh dari hasil wawancara dan observasi terhadap 165 orang ibu hamil. Dari data tersebut diolah menggunakan algoritma klasifikasi naive bayes yang menghasilkan nilai akurasi sebanyak 92.73% . Sedangkan nilai precision sebesar 83.33%, dengan hasil nilai tersebut algoritma naive bayes berhasil digunakan dalam prediksi persalinan prematur (Arga et al., 2020).

Penelitian persalinan prematur dilakukan kembali pada tahun berikutnya oleh handini arga damar rani. Masih menggunakan data private ibu hamil sebanyak 165 orang dan algoritma naive bayes. Namun kali ini penerapan algoritma klasifikasi naive bayes akan di tingkatkan dengan memanfaatkan algoritma particle swarm optimization. Hasil dari percobaan menggunakan optimasi ini didapatkan besaran akurasi sebanyak 91.82% . Hal ini menjadikan algoritma naive bayes dengan particle optimization lebih baik dibandingkan penggunaan naive bayes tanpa particle swarm optimization (Rani, 2021).

Dampak dari persalinan prematur sangat merugikan bagi ibu dan bayi itu sendiri. Dampak untuk ibu bisa menyebabkan infeksi endometrium yang berimbas pada sepsis dan juga penyembuhan luka episiotomi yang lambat. Sedangkan pada bayi itu sendiri bisa mempunyai resiko kematian lebih tinggi 70 kali dibanding persalinan normal (Mustika & Fika Minata, 2021). Oleh sebab itu resiko persalinan prematur ini harus diklasifikasikan sedini mungkin supaya dapat menghindari dampak persalina prematur yang sangat merugikan.

Klasifikasi dalam bidang data mining mempunyai cara perhitungan dengan berbagai metode. Salah satu metode klasifikasi yang sering digunakan dalam dunia medis adalah algoritma decision tree C4.5. Algoritma C4.5 ini memiliki keunggulan dalam hal klasifikasi yaitu berupa pencapaian nilai akurasi yang tinggi. Selain mempunyai tingkat akurasi yang tinggi algoritma C4.5 ini juga memiliki kecepatan yang tinggi juga dalam hal komputasi (Xu et al., 2021). Dengan kemampuan seperti itu algoritma C4.5 sangat cocok digunakan untuk klasifikasi persalinan prematur.

Algoritma klasifikasi data mining tidak hanya berupa algoritma C4.5, selain itu ada juga algoritma naive bayes. Algoritma naive bayes didasarkan pada sekala probabilitas atau harapan keanggotaan terhadap class tertentu, dimana nilai class fitur bersifat independen terhadap fitur lain (Syamsul et al., 2018). Algoritma naive bayes ini memiliki keunggulan mudah diimplementasikan pada banyak kasus, tidak membutuhkan data latih yang banyak dan memberikan hasil yang sangat baik (Tyas et al., 2021).

Algoritma C4.5 sangat unggul dalam kecepatan komputasi sedangkan algoritma naive bayes memiliki kelebihan dalam hasil serta implementasi yang mudah (Hairul Umam et al., 2017). Kedua algoritma tersebut juga bisa diterapkan dalam berbagai

kasus klasifikasi terutama kasus persalinan prematur. Dengan kelebihan masing – masing algoritma tersebut hendaknya perlu menentukan algoritma terbaik yang bisa digunakan. Untuk itu perlu dilakukan komparasi metode naïve bayes dan C4.5 untuk klasifikasi persalinan prematur yang akan sangat bermanfaat bagi dunia medis.

METODE

Metode penelitian ini mempunyai tahapan tahapan sesuai dengan urutan model yang sudah di tentukan. Tahapan awal yang harus dikerjakan pertama kali adalah pengumpulan data. Tahapan urut berikutnya yaitu melakukan pre-prosesing yaitu menempatkan fitur sesuai dengan posisinya. Tahapan selanjutnya adalah menerapkan kedua metode yang akan di bandingkan yaitu metode C4.5 dengan metode naïve bayes. Setelah model metode diterapkan maka akan dilakukan pengujian model apakah model metode berjalan dengan baik atau tidak. Selanjutnya untuk langkah terakhir dilakukanlah evaluasi dan validasi hasil dari kedua model tersebut. Begitulah tahapan yang akan dilankukan pada penelitan ini, untuk lebih rincinya akan di jelaskan sebagai berikut :

1. Pengumpulan Data

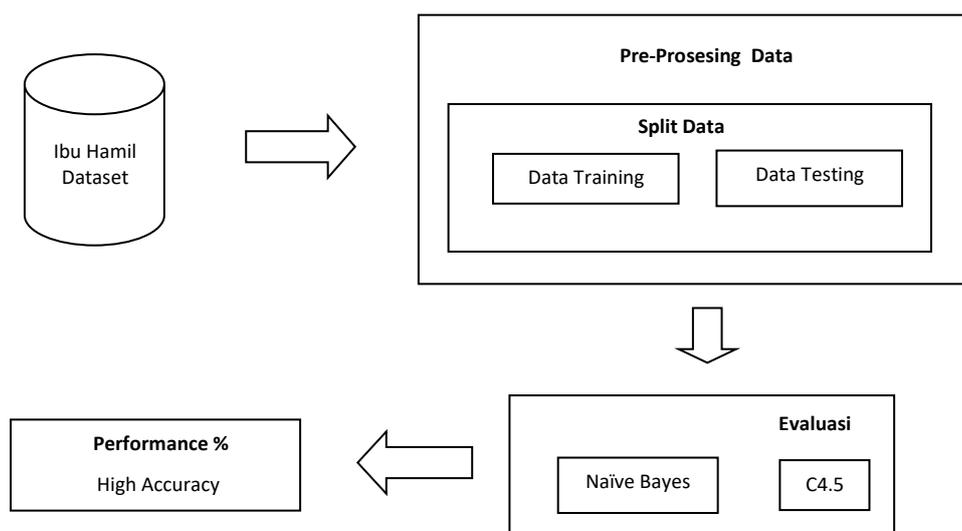
Tahapan pengumpulan data adalah fase dimana menentukan data mana yang akan digunakan dalam penelitian. Didalam penelitian kali ini data yang digunakan adalah data sekunder, diambil dari hasil observasi dan wawancara sebuah klinik bersalin di kota kudas. Data yang diperoleh mempunyai 160 baris data dan memiliki 10 fitur dan 1 fitur sebagai class, sehingga total fitur yang ada sebanyak 11 fitur. Fitur – fitur tersebut antara lain adalah tanggal, no_rm, nama, nama_suami, alamat, td (tekanan darah), bb (berat badan), usia ibu bayi, letak janin, tfu dan kelas.

2. Pre-Prosesing Data

Tahapan pre-prosesing adalah perlakuan awal ketika data mentah sudah di dapatkan. Pada tahap pre-prosesing ini dilakukan langkah penanganan data awal seperti pembersihan data (data cleaning), data integration, seleksi data. Dan bisa juga pada tahapan inilah dilakukan proses pemisahan antara data latih dan data uji yang nantinya akan diterapkan pada model yang sudah ditentukan.

3. Metode yang diusulkan

Pada penelitian ini metode yang diusulkan yaitu ditekankan pada penggunaan 2 algortima yang akan dibandingkan kinerjanya. Dua algoritma tersebut yaitu algoritma decision tree C4.5 dengan algoritma naïve bayes yang akan digunakan untuk klasifikasi persalinan prematur. Tentunya penerapan kedua metode algoritma tersebut sesuai dengan kerangka kerja tahapan pemodelan seperti pada gambar 1 berikut ini:



Gambar 1. Kerangka Model kerja yang diusulkan

4. Pengujian Model

Pengujian model yang dilakukan adalah melihat capaian kinerja perbandingan antara kedua metode yang dipakai yaitu metode naïve bayes dengan metode C4.5. Dari hasil pengujian ini maka akan terlihat seberapa besar akurasi yang dihasilkan antara metode naïve bayes dengan hasil dari metode C4.5.

5. Evaluasi dan Validasi

Tahpan akhir adalah evaluasi dan validasi hasil disini akan menggunakan confusion matrik dalam perhitungan nilai akurasi. Nilai akurasi inilah yang nantinya sebagai acuan perbandingan kinerja dari metode naïve bayes dengan metode C4.5.

HASIL DAN PEMBAHASAN

Hasil pembahasan dalam penelitian ini diambil sesuai paparan metode penelitian yang sudah dijelaskan dari langkah pertama hingga akhir dengan rincian sebagai berikut:

1. Data Penelitian

Data yang dipakai adalah data ibu hamil sebanyak 11 fitur dengan 160 baris data seperti tampak pada tabel 1

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

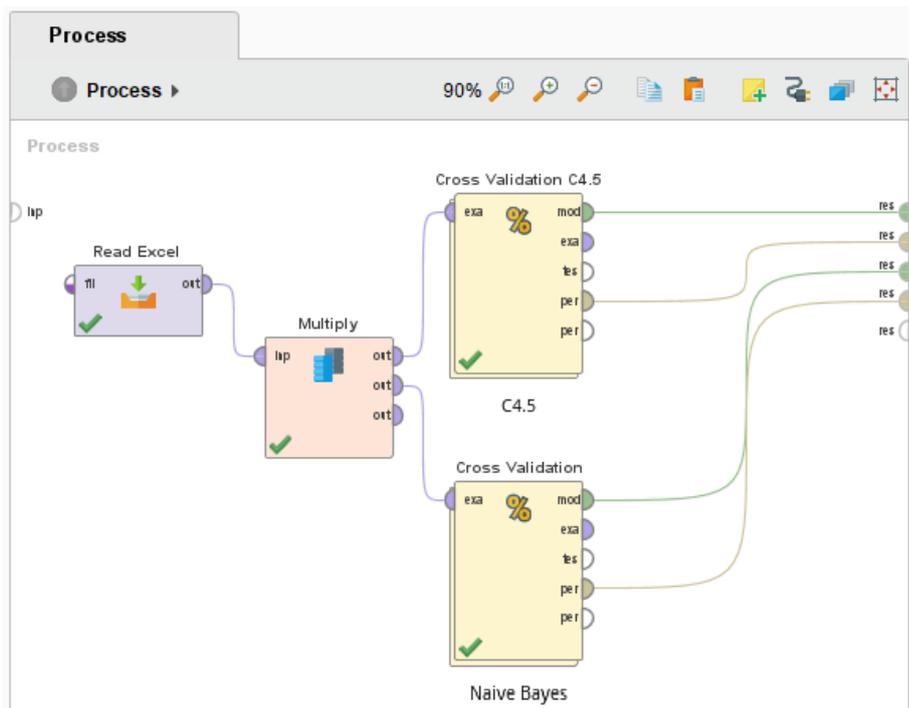
Dari dataset kehamilan ibu sebanyak 160 baris data langkah selanjutnya adalah preprosesing. Tahapan pre-prosesing ini akan dilakukan pembersihan data (data cleaning), data integration, dan seleksi data.

2. Penerapan metode dan pengujian model

Setelah data matang di peroleh selanjutnya penerapan perbandingan dua metode yaitu metode naïve bayes yang di bandingkan dengan metode C4.5. Proses pohon keputusan metode Decision tree C4.5 memiliki beberapa fase :

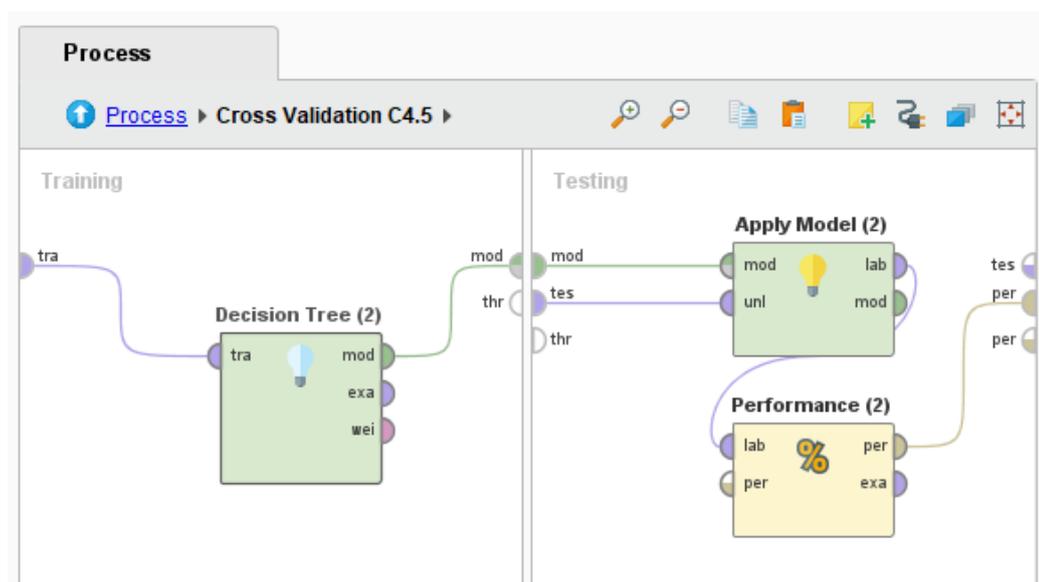
- a. Menghitung hasil penjumlahan data yang didasarkan oleh banyaknya atribut hasil dengan ketentuan syarat yang sudah ditetapkan.
- b. Selanjutnya atribut tersebut akan dipakai sebagai node dengan nilai gian yang paling tinggi.
- c. Menentukan percabangan kepada semua anggota node.
- d. Memilih anggota node yang bernilai 0 untuk dijadikan daun pohon keputusan.
- e. Ulangi langkah sebelumnya sampai keseluruhan nilai entropy anggota dari Node bernilai nol itu artinya proses berhenti.
- f. Jika nilai entropy lebih dari nol yang asalnya dari salah satu anggota node maka proses sebelumnya akan dikerjakan ulang dari awal sampai semua Node bernilai nol seluruhnya.

Perbandingan pemodelan algoritma C4.5 dengan algoritma Naïve Bayes dilakukan melalui bantuan software rapidminer seperti gambar 2 berikut ini



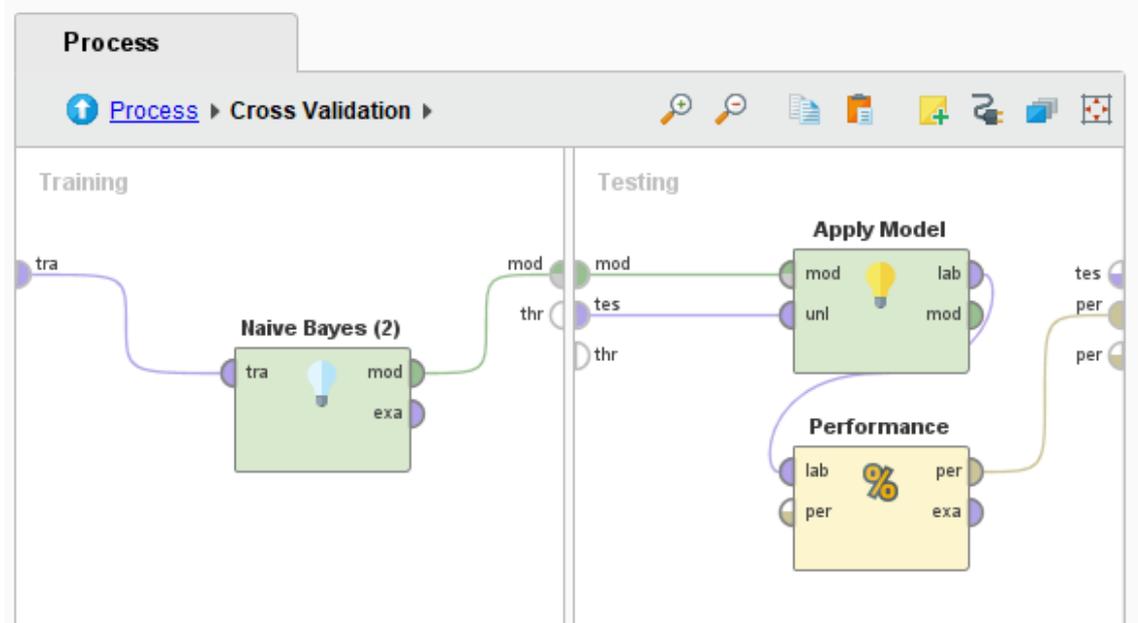
Gambar 2. Model perbandingan algoritma

Model perbandingan algoritma sesuai gambar 2 ini langsung menggunakan satu dataset yang sama yang akan dipakai pada dual algoritma klasifikasi yaitu naïve bayes dengan algoritma C4.5. Hal ini dilakukan untuk menjaga konsistensi dataset yang akan dipakai dipastikan sama. Sedangkan rincian proses model training dan testing pada algoritma C4.5 di tunjukkan pada gambar 3.



Gambar 3. Model proses algoritma C4.5

Sedangkan untuk rincian proses model training dan testing dengan algoritma naïve bayes sesuai dengan representasi gambar 4.

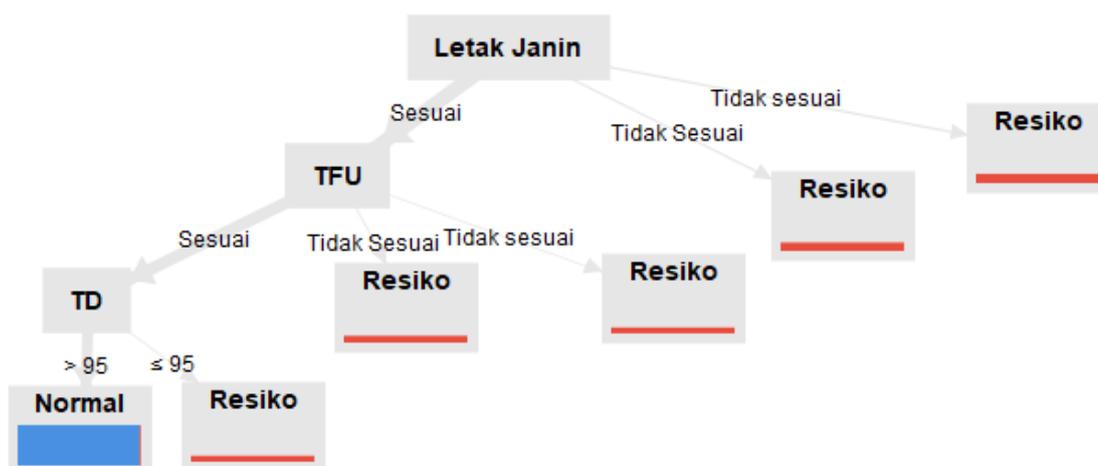


Gambar 4. Model proses algoritma Naïve Bayes

3. Evaluasi dan Validasi Model

Pengukuran hasil dan validasi model diukur menggunakan metode cross-validation yang terdapat pada aplikasi rapidminer. Proses cross validation ini menggunakan nilai k 10 untuk pengujian. Dimana pembagian dataset latih dan uji perbandinganya 9 untuk data latih dan 1 untuk data uji. Sedangkan perulangan iterasi uji juga dilakukan susai jumlah k yaitu sebanyak 10 kali iterasi.

Dari proses itersai itu didapati luaran berupa performa dari metode klasifikasi yang dipakai. Selain performa didalam algoritma decision tree C4.5 juga menghasilkan pohon keputusan seperti nampak pada gambar 5.



Gambar 5. Pohon Keputusan algoritma C4.5

Sedangkan untuk algoritma naïve bayes menghasilkan Simple Distribution sesuai dengan gambar 6.

SimpleDistribution

Distribution model for label attribute Kelas

Class Normal (0.581)
10 distributions

Class Resiko (0.419)
10 distributions

Gambar 6. Simple Distribution algoritma Naïve Bayes

Cross-validation juga menghasilkan tabel confusion matriks yang didalamnya terdapat hasil perhitungan nilai dari akurasi, presisi, recall, dan AUC. Hasil pengukuran nilai confusion matriks algoritma C4.5 sesuai dengan tabel 2 sedangkan untuk algoritma naïve bayes Nampak 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 Naïve Bayes

		Actual Class	
		+ (Positive)	- (Negative)
Prediction class	+ (positive)	69	5
	- (negative)	24	62

Dari tabel confusion matriks pada tabel 2 dan tabel 3 selanjutnya di hitung hasil nilai akurasi, presisi, recall dan AUC nya. Hasil perhitungan nilai tersebut Nampak pada tabel 4.

Tabel 4. Hasil Perhitungan Akurasi Metode Klasifikasi

Metode	Akurasi	AUC
C4.5	98.75 %	0.500
Naïve Bayes	81.88 %	0.945

SIMPULAN

Dalam penelitian ini nampak hasil perhitungan akurasi dari algoritma decision tree C4.5 berada pada angka 98.75% dengan nilai AUC sebesar 0.5. Sedangkan hasil dari algoritma Naïve Bayes sebesar 81.88% dengan nilai AUC sebesar 0.945. Jadi dapat disimpulkan bahwa pada kasus klasifikasi persalinan prematur pada data private ibu hamil algoritma C4.5 lebih unggul dari pada algoritma Naïve Bayes.

Untuk penelitian kedepanya dapat menerapkan banyak dataset public. Karena belum tentu dengan penggunaan dataset yang berbeda akan menghasilkan data perbandingan akurasi yang sama. Sehingga dengan menggunakan banyak dataset maka perbandingan algoritma klasifikasi naive bayes dengan algoritma C4.5 dapat diukur secara presisi.

DAFTAR PUSTAKA

- Arga, H., Rani, D., Hanif, M. B., & Zuhri, S. (2020). *Sistem prediksi kondisi kelahiran bayi menggunakan klasifikasi naïve bayes*. 3.
- 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.
- Carolina, I., & Kresna, R. (2018). Klasifikasi kelahiran prematur menggunakan algoritma c4.5. *Seminar Nasional Teknologi*, 668–672.
- 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)
- 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.
- Ida Rahmawati, Mutiara, V. siska, Absari, N., & Andini, P. (2021). Faktor-Faktor Yang Berhubungan dengan Persalinan Prematur. *Professional Health Journal*, 2(2), 112–121. <https://doi.org/10.54832/phj.v2i2.143>
- 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.
- Loviana, N., Darsini, N., & Aditiawarman, A. (2019). Faktor Yang Berhubungan Dengan Kejadian Persalinan Prematur Di Rsud Dr Soetomo. *Indonesian Midwifery and Health Sciences Journal*, 3(1), 85–97. <https://doi.org/10.20473/imhsj.v3i1.2019.85-97>
- 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>
- Panada Sedianing Drastita, Hardianto, G., Fitriana, F., & Utomo, M. T. (2022). Faktor Risiko Terjadinya Persalinan Prematur. *Oksitosin : Jurnal Ilmiah Kebidanan*, 9(1), 40–50. <https://doi.org/10.35316/oksitosin.v9i1.1531>
- Puspita, A. (2018). Prediksi Kelahiran Bayi Secara Prematur Dengan Menggunakan Algoritma C . 45. *Jurnal Teknik Informatika Stmik Antar Bangsa*, II(1), 11–16.
- Rani, H. A. D. (2021). Optimasi Particle Swarm Optimization Pada Naïve Bayes Untuk Prediksi Kondisi Kelahiran Bayi. *Jurnal Dialektika Informatika (Detika)*, 2(1), 28–33. <https://doi.org/10.24176/detika.v2i1.6964>
- 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.

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>