

Analisis Performa Algoritma Naïve Bayes Untuk Penentuan Kelayakan Pendonor Darah

Funny Farady Coastera¹, Mochammad Yusa², Nadiza Lediwara³, Julia Purnama Sari⁴

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu,

⁴Program Studi Sistem Informasi, Fakultas Teknik, Universitas Bengkulu,

Jl. WR Supratman Kandang Limun Sungai Serut Bengkulu, 38371

e-mail: ^{1*}ffaradyc@unib.ac.id, ²mochammad.yusa@unib.ac.id,

³nadizalediwara@gmail.com, ⁴juliapurnamasari@unib.ac.id

Abstrak

Donor darah merupakan proses tranfusi darah dari pendonor ke dalam bank darah. Dalam proses pendonoran darah, calon pendonor harus dalam kondisi yang benar-benar masuk dalam kondisi paling aman. Jika salah menentukan kriteria indikator calon pendonor darah maka akan berakibat fatal bahkan dapat menimbulkan kematian. Dalam era kecanggihan, data mining sangat bermanfaat bagi untuk beberapa bidang termasuk dalam membantu pendonoran darah. Penelitian ini bertujuan untuk mengukur kinerja dari algoritma Naive Bayes yang diimplementasikan pada dataset pendonoran darah. Penelitian dimulai dengan pengumpulan sample pendonor darah baik kriteria yang boleh atau tidak boleh. Kemudian dilanjutkan dengan proses pembersihan data (data preprocessing). Setelah didapatkan dataset yang memenuhi kriteria, tahapan selanjutnya adalah mengimplementasi algoritma supervised learning Naive Bayes. Dalam penelitian ini juga lebih mengeksplor data atribut yang bersifat kontinyu. Atribut-atribut yang dipertimbangkan dalam penelitian ini adalah berat badan, kadar hemoglobin, umur, tekanan darah sistolik, distolik. Metode validasi pengukuran kinerja model yang digunakan dalam penelitian ini adalah k-fold cross validation dengan nilai k=10. Akurasi, Presisi, dan Nilai Recall menjadi parameter kinerja model algoritma yang diukur dalam penelitian ini. Hasil dari penelitian ini menunjukkan bahwa akurasi model algoritma Naive Bayes tergolong excellent yaitu sebesar 80%, nilai recall rata-rata kedua label adalah 80,61%, dan nilai precision dari algoritma ini sebesar 80,401%.

Kata kunci— Donor Darah, Algoritma Naive Bayes, Akurasi, Precisi, Recall

Abstract

Blood donation represents a process of moving human blood to the blood bank in case of helping people. In blood transfers, the blood donors should be in safe criteria because if the committee has fault to decide, there will be upcoming death. In sophisticated era, data mining can be a good solution to help blood transferring committee in case of determining advisability of the blood donors. This research objects to evaluate the Naive Bayes Algorithm that is implemented to Blood Donor Dataset. The research start with collecting blood donors from Indonesian Blood Voluenter's committee. Then the activity is continued with the process of cleaning up the data. After obtaining a dataset that meets the criteria, the next step is to implement the Naive Bayes supervised learning algorithm to the final dataset. In this study, we explore continuous attribute data. The considered attributes in this study are body weight, hemoglobin level, age, systolic, and diastolic blood pressure. The validation method of measuring the performance of the used model in this study is k-fold cross validation with a value of $k = 10$. Accuracy, Precision, and Recall value are the measured performance parameters of the algorithm model in this study. The results of this study indicate that the accuracy of the Naive Bayes algorithm model is classified as excellent at 80%, the average recall value of the two labels is 80.61%, and the precision value of this algorithm is 80.401%.

Keywords— Blood Donors, Naive Bayes Algorithm, Accuracy, Recall, Precision

1. PENDAHULUAN

Donor darah merupakan proses tranfusi darah yang dilakukan pada seseorang. Orang yang mendonor disebut pendonor. PMI selaku organisasi yang menangani proses tranfusi darah harus benar-benar memastikan persediaan darah untuk setiap daerah. Dalam hal ini proses sumber daya dan rekrutmen harus benar-benar tervalidasi dan memiliki komunikasi yang baik[1]. Pendonor darah harus memenuhi syarat dalam proses tranfusi dara ke bank darah[2]. Jika syarat tersebut tidak dilakukan dengan benar maka proses donor darah tersebut akan berakibat fatal seperti terjangkit penyakit bahkan bisa mengakibatkan kematian[3]. Sehingga penentuan calon pendonor menjadi sesuatu yang sangat krusial dalam proses pendonoran darah.

Dalam era big data, teknik data mining merupakan lambing darisebuahkecanggihan yang dapatmemberikaninformasidaritumpukan data yang besar. Salah satuteknikklasifikasi yang sangatcanggihadalah Naive Bayes. Naive Bayes merupakan salah satu algoritma klasifikasi. Algoritma Klasifikasi telah banyak dimanfaatkan untuk beberapa bidang kehidupan sehari-hari. Misalnya di bidangpendidikan[4][5], pembangunan[6], perbankan[7][8], kesehatan[9], Manajemenrumah tangga[10], dan lain-lain.

Metode Naive Bayes memiliki nilai akurasi yang cukup baik. Menurut penelitian[11], Naive Bayes memiliki performa yang baik dan cepat jika diimplementasikan pada jumlah data yang besar. Namun tipe data nominal yang digunakan pada penelitian tersebut tidak variatif dan cenderung monoton dan jumlah atributnya sedikit. Hal terkait Performa Algoritma Naive Bayes tersebut dikuatkan oleh beberapa penelitian lain seperti referensi [12] dan [13] yang menyebutkan bahwa Naive Bayes Classifier cenderung kuat untuk data yang kecil dan monoton. Pada kasus Pendonor Darah, cukup bervariasi sehingga penelitian ini akan menganalisis kinerja dari Algoritma Naive Bayes dengan data yang sedikit namun tipe data bervariasi sehingga metode yang akan diimplementasikan nantinya akan memiliki performa yang lebih baik pula.

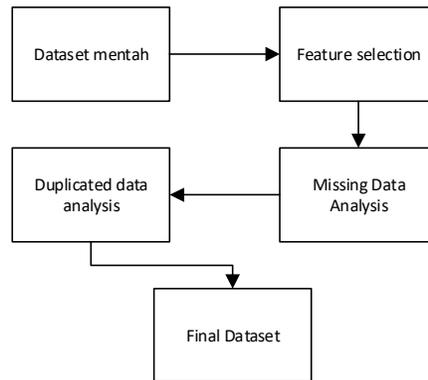
Penentuan klasifikasi layak atau tidak calon pendonor dalam mendonor sangatlah penting. Dirasa penting karena jika terdapat kesalahan dalam penentuan tentunya akan berakibat fatal baik untuk institusi maupun calon pendonor darah. Dengan berkembangnya metode algoritma yang banyak tentunya dalam pembuatan keputusan juga perlu digunakan algoritma yang layak agar dapat meminimalisir kejadian yang tidak diinginkan. Penelitian ini akan membahas terkait dengan analisis performa algoritma Naive Bayes yang merupakan tahapan pra-implementasi dalam tahap pembangunan sistem informasi.

2. METODE PENELITIAN

Penelitian yang dilakukan adalah penelitian experimental. Penelitian yang dilakukan secara garis besar menganalisis performa dari algoritma Naive Bayes. Penelitian ini bertujuan untuk menganalisis kemampuan kinerja Algoritma Naive Bayes jika diimplementasikan pada klasifikasi pendonor darah. Gambar 1 merupakan diagram keseluruhan dari penelitian yang akan dilakukan. Proses pertama adalah proses pengumpulan dataset. Dataset yang diambil adalah dataset yang berasal dari PMI dan divalidasi oleh salah satu anggota PMI cabang Bengkulu. Untuk tahapan penelitian dibagi menjadi dua bagian besar yaitu Tahapan Data Preprocessing dan Eksperimental.

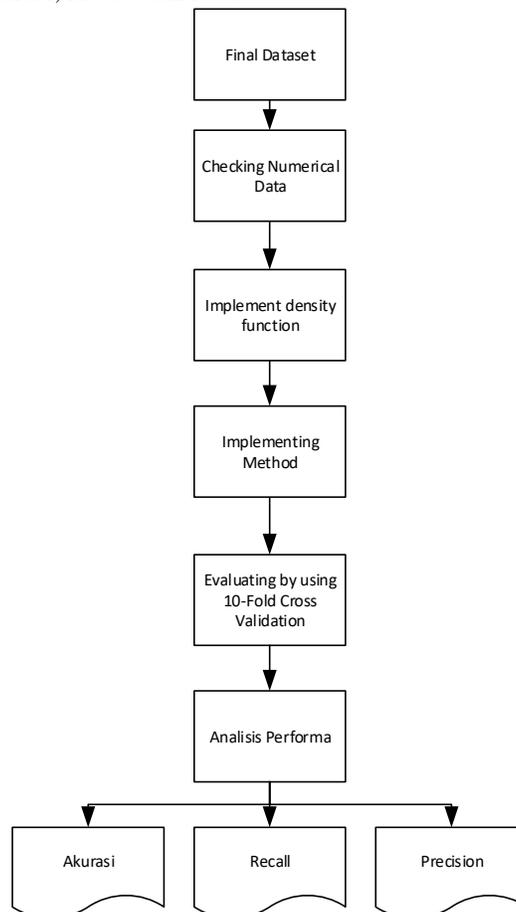
Tahapan data preprocessing ditunjukkan pada langkah-langkah penelitian (Gambar 1) yang dimulai dengan mengumpulkan data mentah dari PMI. Data yang didapatkan adalah sebanyak 100 baris data. Data tersebut juga divalidasi secara manual oleh tim PMI untuk menentukan layak atau tidaknya pendonor melakukan donasi. Setelah data divalidasi kemudian tahap selanjutnya adalah melakukan pemilihan atribut yang berpengaruh terhadap proses klasifikasi pendonor (*feature selection*). Tahapan dilanjutkan dengan memeriksa duplikasi data yang di dapat. Selanjutnya setelah dilakukan pemeriksaan, aktivitas dilanjutkan dengan proses analisis

data yang hilang (*missing values*). setelah proses tersebut dilakukan maka tahapan data preparation sudah dilaksanakan.



Gambar1. Diagram alur Tahap Data Preprocessing

Gambar 2 menunjukkan tahap implementasi model algoritma terhadap dataset PMI yang sudah melalui tahapan data preprocessing. Langkah pertama adalah mengecek atribut atau fitur yang berjenis numerical. Jika terdapat data yang berjenis *numerical* maka atribut tersebut akan ditangani dengan fungsi densitas dengan mencari nilai mean dari setiap atribut numerical dan *standard deviation*-nya. Kemudian dengan menguji langsung dataset dengan mengimplementasi model Naive Bayes dan mengvalidasi dengan metode *10-fold cross validation* untuk mendapatkan tingkat akurasi, *Recall* dan *Precision*.



Gambar2. Diagram alur Eksperimen

Pengujian validasi data yang digunakan dalam eksperimen ini adalah *10-fold cross-validation* yang artinya dataset akan dibagi menjadi 10 subset yaitu $(x_1, x_2, \dots, x_{10})$ dimana n merupakan nilai subset. Kemudian dari kesepuluh subset tersebut akan dilakukan iterasi sebanyak 10 kali dengan data training dan testing yang secara bergantian. Misalnya x_1 menjadi data training maka x_2, x_3, \dots, x_{10} akan menjadi data testing. Setelah selesai iterasi pertama maka x_2 menjadi bergantian menjadi data training dan $x_1, x_3, x_4, \dots, x_{10}$ akan menjadi data testing. Iterasi dilakukan sebanyak jumlah subset midalkan *10-fold* berarti akan ada pembagian menjadi 10 subset dan 10 iterasi [14].

Setelah dilakukan iterasi maka akan didapatkan prediksi hasil dari implementasi model. Hasil validasi direpresentasikan dengan *confusion matrix* (Gambar 3).

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) <i>Corect result</i>	FP (False Positive) <i>Unexpected result</i>
	FALSE	FN (False Negative) <i>Missing result</i>	TN (True Negative) <i>Corect absence of result</i>

Gambar3. Confusion Matriks

Accuracy dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi[14]. Rumus *Accuracy* dapat didefinisikan pada rumus (3) berikut ini:

$$A = \frac{T + T}{P} \quad (3)$$

Dimana, TP merupakan jumlah *record/tuple* positif yang dilabelkan secara benar oleh model algoritma klasifikasi, TN merupakan jumlah *record/tuple* negatif yang dilabelkan secara salah oleh model algoritma klasifikasi, dan P merupakan total semua record yang dievaluasi. Semakin tinggi level *accuracy*-nya, maka dapat dikatakan semakin efektif model algoritma klasifikasi tersebut[16].

Untuk pengukuran precision dan value dapat digunakan pada rumus berikut (4,5):

$$P = \frac{T}{T + F} \quad (4)$$

$$R = \frac{T}{T + F} \quad (5)$$

Precision value dalam klasifikasi merupakan tingkat kedekatan pengukuran kuantitas terhadap nilai yang sebenarnya. Sedangkan *recall* merupakan digunakan untuk mengukur tingkat keberhasilan sistem klasifikasi berdasarkan jumlah *universe label* dari sebuah algoritma *supervised learning*.

2.1. Teknik Klasifikasi Data Mining

Menurut referensi [14], klasifikasi adalah penugasan data mining yang memperkerjakan sebuah objek ke salah satu kategori yang sudah didefinisikan berdasarkan atribut-objeknya. Input yang dimasukkan pada problem adalah sebuah dataset yang disebut training set yang terdiri dari sejumlah sampel yang mempunyai jumlah atribut-atribut. Atribut-atribut tersebut juga bersifat berkelanjutan ketika values dari atribut dapat dihitung atau bersifat kategorikal ketika nilai atribut atau values dari atribut tidak dapat dihitung.

Salah satu atribut kategorikal disebut class-label atau atribut yang telah diklasifikasikan. Tujuannya adalah untuk menggunakan training-set dalam membangun sebuah model dari class-label berdasarkan pada atribut-atribut lain agar model tersebut dapat digunakan untuk mengklasifikasikan data baru bukannya dari data-set yang sudah di-training. Teknik klasifikasi telah dipelajari secara ekstensif pada ilmu statistik, *machine-learning*, *neural networks*, dan sistem pakar [15].

Setiap teknik memiliki kelebihan dan kekurangannya sendiri. Data dengan profil tertentu mungkin paling optimal jika diklasifikasi dengan teknik tertentu, atau dengan kata lain, profil data tertentu dapat mendukung termanfaatkannya kelebihan dan teknik ini.

2.2. Algoritma Naive Bayes

Klasifikasi Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Menurut Wu dan Kumar bahwa Naive Bayes merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining. Naive Bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training [6]. Teorema Naive Bayes dapat dirumuskan sebagai berikut (1):

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \tag{1}$$

Dimana X merupakan data dengan class yang akan diklasifikasikan, H merepresentasikan hipotesis data yang ditentukan. P(H|X) adalah probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas). P(H) adalah probabilitas hipotesis H (prior probabilitas). P(X|H) mengindikasikan kemungkinan X berdasarkan kondisi pada hipotesis H dan P(X) merepresentasikan probabilitas dari X.

Untuk jenis data yang bersifat kontinu, Naive Bayes mengimplementasikan fungsi Densitas Gauss yang dirumuskan dalam rumus (2) berikut:

$$P(X|Y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

P(X|Y) merupakan probabilitas yang dicari untuk atribut X yang merepresetasikan label Y. Sedangkan μ merupakan mean yang menyatakan rata-rata atribut kontinu dan σ adalah standar baku deviasi yang menyatakan varian untuk seluruh atribut.

2.3. Dataset

Dataset yang digunakan pada penelitian ini adalah dataset acak yang didasarkan pada atribut-atribut yang mempengaruhi atau faktor-faktor yang menjadi pertimbangan PMI dalam mengklasifikasikan kelayakan seorang pendonor. Dataset ini terdiri dari 9 atribut/feature yang meliputi 8 atribut terikat dan 1 atribut label atau prediktor. Jumlah record yang digunakan dalam penelitian ini adalah sejumlah 100 records yang sudah tervalidasi oleh salah satu anggota PMI cabang Bengkulu. Atribut-atribut tersebut dapat di lihat pada Tabel 1.

Tabell. Atribut Dataset

No	Nama Atribut	Jenis Data	Bentuk Data
1	Umur	Numerical	19,20,67,... dalam satuan tahun
2	Bb atau Berat Badan	Numerical	51,49,67,... Dalam satuan Kilogram

No	Nama Atribut	Jenis Data	Bentuk Data
3	Kadar Hemoglobin	Numerical	12.05, 15.00, 15.01,.... Dalam satuan gram
4	Jenis Kelamin	Binominal	L atau P
5	Tekanan Darah Sistolik	Numerical	100, 110, 120, 130,... Dalam satuan mmHg
6	Tekanan Darah Distolik	Numerical	60, 70, 80, 90,... Dalam satuan mmHg
7	Jumlah Denyut Nadi Permenit	Polinomial	0-49, 50-100, >100
8	Histori Donor dalam 3 bulan terakhir	Binominal	Ya untuk relawan yang sudah mendonor dalam kurun waktu 3 bulan terakhir dan “Belum” untuk yang belum mendonorkan darah dalam kurun waktu 3 bulan terakhir
9	Klasifikasi Donor	Binominal	Boleh atau Tidak Merupakan label atau prediktor dalam klasifikasi

3. HASIL DAN PEMBAHASAN

3.1. Tahapan Persiapan Data (Data Preparation)

Setelah didapatkan data dan dianalisis seperti yang ditunjukkan pada Gambar 1, maka diputuskan bahwa semua fitur/atribut yang didapat dari PMI terkait Klasifikasi pendonor darah digunakan. Dari kesembilan atribut tersebut atribut denyut nadi tidak dapat dijadikan parameter dalam perhitungan klasifikasi. Hal tersebut disebutkan langsung oleh narasumber dikarenakan denyut nadi manusia yang cenderung berubah-ubah seiring dengan kegiatan yang dilakukan sebelumnya. Kemudian atribut lainnya yang tidak dijadikan pertimbangan klasifikasi adalah atribut rentang waktu donor dalam tiga bulan terkahir karena menurut narasumber PMI bahwa jika sudah melakukan donor dalam waktu 3 bulan terakhir sudah dipastikan bahwa calon pendonor tersebut tidak dapat mendonor kembali. Sehingga ada dua atribut yang tidak dipilih dalam penelitian ini. Tabel 3 menunjukkan daftar atribut yang digunakan setelah dilakukannya tahap *feature selection*.

Tabel2. Final Dataset

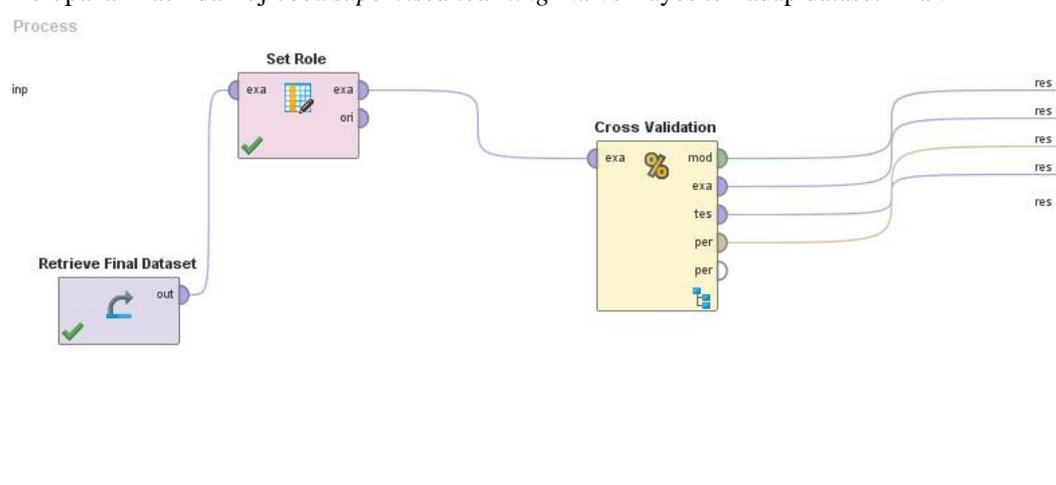
No	Nama Atribut	Jenis Data	Bentuk Data
1	Umur	Numerical	19,20,67,... dalam satuan tahun
2	Bb atau Berat Badan	Numerical	51,49,67,.... Dalam satuan Kilogram
3	Kadar Hemoglobin	Numerical	12.05, 15.00, 15.01,.... Dalam satuan gram
4	Jenis Kelamin	Binominal	L atau P
5	Tekanan Darah Sistolik	Numerical	100, 110, 120, 130,... Dalam satuan mmHg
6	Tekanan Darah Distolik	Numerical	60, 70, 80, 90,... Dalam satuan mmHg
7	Klasifikasi Donor	Binominal	Boleh atau Tidak

Merupakan label atau prediktor dalam klasifikasi

Kemudian setelah dilakukan seleksi fitur, langkah selanjutnya adalah *missing value analysis*. Hasil pemeriksaan data yang hilang, menunjukkan bahwa tidak ada data yang kosong. Langkah terakhir setelah dilakukan pemeriksaan, juga menunjukkan hasil bahwa tidak ditemukannya duplikasi baris data sehingga semua data di dalam dataset tergolong unik. Maka dapat disimpulkan secara sementara bahwa terdapat 7 fitur/atribut dan 100 baris data yang akan menjadi final dataset yang akan digunakan untuk proses pengujian model.

3.2. Implementasi Algoritma dan Validasi

Setelah tahap preprocessing dilakukan kemudian langkah selanjutnya adalah tahapan uji coba (eksperimental). Pada tahapan ini ujicoba pertama yang dilakukan adalah mengimplementasi model Naive Bayes dengan mengabaikan atribut yang berisikan jenis data *numerical*. Gambar 4 merupakan hasil dari ujicoba *supervised learning* Naive Bayes terhadap dataset final.



Gambar4. Implementasi Naive Bayes pada Machine Learning

Proses ujicoba dimulai dari mengimport data yang sudah diproses pada tahapan data preprocessing. Kemudian ditambahkan fungsi *set role* untuk menentukan atribut yang menjadi label/class. Pada penelitian ini atribut yang menjadi parameter prediktor model adalah atribut klasifikasi donor. Validasi model menggunakan cross validation dengan nilai fold sama dengan 10. Kemudian setelah dilakukan eksperimen didapatkan hasil *mean* dan *standart deviation*(Tabel 3) yang digunakan untuk menghitung probabilitas dari setiap atribut dengan perhitungan Rumus (2).

Tabel3. Mean dan Standard Deviation Atribut Kontinyu

No	Atribut	Parameter	Label Tidak	Label Boleh
1	bb	mean	59,6809	59,4906
2	bb	standard deviation	15,9743	11,9028
3	distolik	mean	79,8085	80,3585
4	distolik	standard deviation	10,1224	7,9617
5	kadar_hb	mean	13,4164	15,6311
6	kadar_hb	standard deviation	2,4388	1,6526
7	sistolik	mean	117,3191	119,8491
8	sistolik	standard deviation	20,4377	14,3693

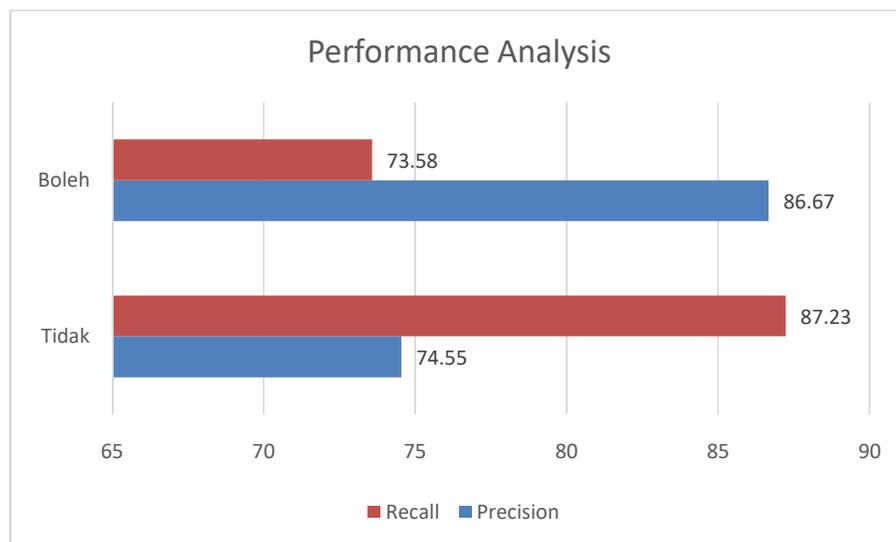
9	umur	mean	30,5106	27,4340
10	umur	standard deviation	18,5930	12,5940

Hasil dari perhitungan mean dan standard deviation tersebut kemudian digunakan untuk menghitung semua probabilitas yang memungkinkan berdasarkan rumus (1). Hasil dari perhitungan model Naive Bayes tersebut direpresentasikan dalam bentuk Matriks Confusi seperti yang ditunjukkan pada Tabel 4.

Tabel4. Hasil Matrik Confusi Naive Bayes

	True. Tidak	True. Boleh
Pred. Tidak	41	14
Pred. Boleh	6	39

Dari evaluasi kinerja yang ditunjukkan pada Tabel 4, nilai akurasi dari Algoritma Naive Bayes yang memiliki nilai data *numerical* dan kemungkinan probabilitas menggunakan fungsi densitas dan satu atribut nominal yaitu sebesar 80%. Nilai kinerja tersebut didapat dari jumlah yang terprediksi dengan benar yaitu sebanyak 80 tuple/record dibagi dengan total keseluruhan data yaitu 100 record menggunakan rumus (3). Hasil tersebut dapat digolongkan sebagai akurasi yang berada di kluster sangat baik. Sedangkan performa nilai precision dan recall dapat dilihat pada Gambar 5.



Gambar5. Recall dan Precision dalam persen

Gambar 5 menunjukkan hasil precision dan recall berdasarkan rumus (4) dan (5). Hasil menunjukkan bahwa parameter kinerja *Precision* yang didapat dari hasil eksperimen ini cukup baik yaitu untuk prediksi “Tidak” tergolong baik yaitu di nilai 74,55% dan *class precision* untuk prediksi “Boleh” sebesar 86,67% sehingga rata class precision dari kedua label tersebut 80,61%. Kemudian untuk label “Tidak” nilai recallnya adalah 87,23% sedangkan untuk label “Boleh” nilai recallnya adalah 73,58% atau rata-rata nilai recall dari kedua label adalah 80,405% yang mana hasil tersebut mengindikasikan bahwa tingkat recallnya tergolong sangat baik. Berdasarkan hasil dari pengujian tersebut maka Algoritma Naive Bayes juga memiliki performa yang sangat baik dalam penanganan data yang bersifat kontinyu.

4. KESIMPULAN

Berdasarkan hasil dari ujicoba implementasi model algoritma Naive Bayes terhadap dataset klasifikasi kelayakan calon pendonor darah PMI, performa model tergolong sangat baik yaitu berada di jumlah 80% dari 100 records. Tahap *data preprocessing* melibatkan 7 atribut yang menjadi indikator penilaian dalam hal mempertimbangkan atribut umur, berat badan, kadar hemoglobin, tekanan darah sistolik dan distolik, serta label atau klasifikasi calon pendonor. Ditinjau dari performa Recall dan Precisionnya, Algoritma Naive Bayes juga memiliki nilai kinerja yang sangat baik yaitu 80,61% untuk nilai *precision* dan 80,405% untuk nilai *Recall*.

5. SARAN

Model algoritma pada dasarnya memiliki performa yang baik dan cepat. Namun pada penelitian yang sudah dilakukan performa model Algoritma Naive Bayes masih belum maksimal karena lima dari tujuh atribut yang diuji pada penelitian ini memiliki jenis data yang bersifat kontinyu. Untuk penelitian selanjutnya, kami harap dapat melakukan penyesuaian lebih komprehensif terkait dengan penanganan jenis data kontinyu.

UCAPAN TERIMA KASIH

Ucapan Terima Kasih kami haturkan kepada Rektor, Dekan Fakultas Teknik, Ketua Program Studi Informatika Universitas Bengkulu, dan segenap rekan-rekan yang telah berkontribusi baik materil dan non materil guna terselesainya paper ini.

DAFTAR PUSTAKA

- [1] R. Dewi, M. Kartasurya and A. Mawarni, "Analisis Kebijakan Donor Darah Dan Implementasi Program Rekrutmen Donor Di Unit Donor Darah (UDD PMI) Kota Pontianak," *Jurnal Manajemen Kesehatan Indonesia*, vol. 4, no. 2, pp. 109-117, 2017.
- [2] K. RI, "Permenkes RI Nomor 83 Tahun 2014 tentang Unit Transfusi Darah, Bank Darah Rumah Sakit, dan Jejaring Pelayanan Transfusi Darah," Kementerian Kesehatan Republik Indonesia, Jakarta, 2014.
- [3] Admin, "Pemerintah Kota Medan," Pemerintah Kota Medan, 16 Desember 2018. [Online]. Available: <https://pemkomedan.go.id/artikel-18369-dampak-buruk-transfusi-darah.html>. [Accessed 03 Agustus 2019].
- [4] M. Ridwan, H. Suyono and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 7, no. 1, pp. 59-64, 2013.
- [5] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Jurnal Dinamik*, vol. 18, no. 1, 2013.
- [6] A. Mukminin and D. Riana, "Komparasi Algoritma C4. 5, Naive Bayes Dan Neural Network Untuk Klasifikasi Tanah," *Jurnal Informatika*, vol. 4, no. 1, 2017.
- [7] B. Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI-Jurnal Teknik Informatika*, vol. 5, no. 2, 2013.
- [8] D. S. Hidayatullah, R. B. Widhakdo, A. Arivianto, M. L. Taqwim and M. Yusa, "Evaluasi Performa Algoritma Decision Tree C4. 5 Pada Dataset Bank Marketing Dengan Telepon Langsung," *Technomedia Journal*, vol. 3, no. 2, pp. 185-196, 2018.
- [9] A. M. Zamani, B. Amaliah and A. Munif, "Implementasi Algoritma Genetika pada Struktur

- Backpropagation Neural Network untuk Klasifikasi Kanker Payudara," *Jurnal Teknik ITS*, vol. 1, no. 1, pp. A222-A227, 2012.
- [10] A. Saleh, "Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga," *Creative Information Technology Journal*, vol. 2, no. 3, pp. 207-217, 2015.
- [11] M. Yusa, and E. Utami, "Classifiers evaluation: Comparison of performance classifiers based on tuples amount," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Yogyakarta, 2017.
- [12] A. A. M. Syarli, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmu Komputer*, vol. 2, no. 1, 2016.
- [13] A. P. Fadillah and B. Hardiyana, "PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK PEMILIHAN KONSENTRASI MATA KULIAH," *Jurnal Teknologi dan Informasi (JATI)*, vol. 8, no. 2, pp. 1-9, 2018.
- [14] J. Han and M. P. J. Kamber, *Data Mining: Concept and Techniques*, Burlington: Elsevier, 2012.
- [15] Mitchell, *Machine Learning*, 1Ed, Indian: Mcgraw Hill, 2013.
- [16] P. Mittal and N. S. Gill, "A Comparative Analysis Of Classification Techniques On Medical Data Sets," *JRET: International Journal of Research in Engineering and Technology*, vol. 3, no. 6, pp. 454-460, 2014.
-