

ANALISA POTENSI PENDONOR DARAH POTENSIAL PADA UNIT
TRANSFUSI DARAH PALANG MERAH INDONESIA KOTA DEPOK

PENULIS¹⁾Devi Puspita Sari, ²⁾Sfenrianto, ³⁾Yuli Prasetya, ⁴⁾Rino Subekti

ABSTRAK
Persediaan darah merupakan faktor yang sangat penting dalam Unit Transfusi Darah PMI. Hal ini untuk memudahkan dan memfasilitasi masyarakat yang membutuhkan darah, sehingga stok darah di PMI dapat menurun tajam, ada beberapa faktor yang menyebabkannya. Oleh karena itu diperlukan penelitian untuk menjawab permasalahan tersebut dengan mengolah dataset menggunakan algoritma K-Nearest Neighbour, Naïve Bayes dan C4.5. Pada penelitian ini difokuskan pada pencarian algoritma terbaik. Pada penelitian ini difokuskan pada pencarian algoritma terbaik berdasarkan nilai akurasi dan Under Curve Area, dimana hasil penelitian akan menemukan keakuratan pendonor darah tetap, sehingga meminimalisir terjadinya kekurangan darah.

Kata Kunci Donor Darah, Algoritma, Akurasi

AFILIASI
Program Studi^{1,4)}Teknik Informatika, Fakultas Ilmu Komputer
²⁾Manajemen Sistem Informasi
³⁾Sistem Informasi, Fakultas Ilmu Komputer
Nama Institusi^{1,3,4)}Institut Bisnis dan Informatika (IBI) Kosgoro 1957
²⁾Universitas Bina Nusantara
Alamat Institusi^{1,3,4)}Jl. M. Kahfi II No. 33, Jagakarsa, Jakarta Selatan, DKI Jakarta
²⁾Jl. Kebon Jeruk Raya, No.27, Jakarta Barat, DKI Jakarta

KORESPONDENSI
Penulis
Email
Devi Puspita Sari
devipuspitasari0903@gmail.com

LICENSE
 This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

1.1 Latar Belakang

Palang Merah Indonesia (PMI) Kota Depok merupakan salah satu lembaga yang berperan dalam penyediaan darah, di samping institusi resmi yang ditunjuk oleh Kementerian Kesehatan. Kegiatan transfusi darah dilaksanakan oleh Unit Transfusi Darah (UTD) sebagai bagian dari tugasnya dalam memberikan pelayanan darah kepada masyarakat. Permintaan darah tersebut umumnya berasal dari rumah sakit maupun Bank Darah Rumah Sakit (BDRS), baik melalui kerja sama resmi dengan perjanjian *Memorandum of Understanding* (MoU) maupun tanpa adanya kerja sama formal.

Ketersediaan darah menjadi faktor krusial dalam meningkatkan mutu layanan kesehatan serta menyelamatkan nyawa pasien. Berdasarkan rekomendasi *World Health Organization* (WHO), darah yang aman, berkualitas, dan berisiko rendah sebaiknya diperoleh dari pendonor sukarela dibandingkan dari donor pengganti. Donor sukarela adalah individu yang mendonorkan darah tanpa ditujukan kepada penerima tertentu, sementara donor pengganti merupakan pendonor yang menyumbangkan darahnya untuk orang tertentu yang telah ditentukan sebelumnya. Rendahnya tingkat kesadaran masyarakat untuk berdonor menjadi salah satu penyebab terbatasnya stok darah di PMI. Oleh karena itu, salah satu strategi yang penting adalah membangun kelompok pendonor tetap yang rutin menyumbangkan darah secara berkala[1].

1.2 Donor Darah

Donor darah merupakan kegiatan pengambilan darah dari seseorang, baik secara sukarela maupun sebagai donor pengganti, untuk kemudian disimpan di bank darah sebagai persediaan dan digunakan dalam kebutuhan transfusi. Kegiatan ini dapat dilakukan secara rutin di Unit Transfusi Darah (UTD) PMI, baik di tingkat pusat maupun daerah, dalam periode tertentu. Selain itu, donor darah juga kerap diselenggarakan melalui kegiatan khusus yang diadakan oleh instansi pemerintah maupun swasta.

1.3 Jenis Golongan Darah

Golongan darah merupakan sistem klasifikasi darah dalam suatu kelompok yang ditentukan berdasarkan ada atau tidaknya antigen yang diwariskan pada permukaan membran sel darah merah. Perbedaan ini muncul karena variasi jenis karbohidrat dan protein yang terdapat pada membran sel darah merah. Dua sistem penggolongan darah yang paling utama adalah sistem ABO dan Rhesus (Rh). Secara umum, penentuan golongan darah pada manusia didasarkan pada keberadaan aglutinogen dan aglutinin di dalam darah. Berdasarkan keberadaan antigen A dan B pada permukaan sel darah merah, terdapat empat jenis golongan darah, yaitu:

- 1) **Golongan darah A:** memiliki antigen A pada membran sel darah merah serta membentuk antibodi terhadap antigen B.
- 2) **Golongan darah B:** memiliki antigen B pada membran sel darah merah serta membentuk antibodi terhadap antigen A.
- 3) **Golongan darah AB:** memiliki antigen A dan B sekaligus pada membran sel darah merah, namun tidak menghasilkan antibodi terhadap keduanya.
- 4) **Golongan darah O:** tidak memiliki antigen A maupun B pada membran sel darah merah, tetapi menghasilkan antibodi terhadap antigen A dan B.

1.4 Data Mining

Data Mining terdiri dari beberapa algoritma yang memungkinkan seseorang memperoleh wawasan dan pengetahuan mendasar dari data massif. Data Mining adalah “penggabungan bidang konsep interdisipliner dari wilayah sejenis seperti *System Database* (sistem database), *statistic* (statistik), *Machine Learning* (mesin pembelajaran), dan *Pattern Recognition* (pengenalan pola)” [2].

Klasifikasi merupakan proses analisis data yang bertujuan menghasilkan model untuk merepresentasikan kelompok-kelompok (kelas) yang terdapat di dalam data. Model tersebut dikenal sebagai classifier, yang berfungsi untuk menyusun dan mengidentifikasi kelas-kelas dalam data. Terdapat berbagai algoritma yang dapat digunakan dalam klasifikasi, di antaranya Decision Tree C4.5, Naïve Bayes, dan K-Nearest Neighbour (KNN).

1) **Decision Tree**

Decision Tree digunakan untuk mempelajari klasifikasi dan prediksi pola dari data dan menggambarkan relasi dari variabel atribut x dan variabel target y dalam bentuk pohon [3]. *Decision Tree* adalah struktur yang menyerupai *flowchart* dimana setiap *internal node* (*node* yang bukan *leaf* atau bukan *node* terluar) merupakan pengujian terhadap variabel atribut, tiap cabangnya merupakan hasil dari pengujian tersebut, sedangkan *node* terluar yakni *leaf* menjadi labelnya [4].

2) **Naïve Bayes**

Naïve Bayes adalah metode klasifikasi berbasis probabilistik yang sederhana, dengan cara menghitung probabilitas melalui frekuensi serta kombinasi nilai dalam suatu dataset. Algoritma ini didasarkan pada Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen atau tidak saling bergantung terhadap nilai yang diberikan oleh variabel kelas. Secara lain, *Naïve Bayes* dapat dipahami sebagai teknik klasifikasi yang menggunakan pendekatan probabilitas statistik, yang pertama kali diperkenalkan oleh ilmuwan asal Inggris, Thomas Bayes, untuk memprediksi kemungkinan berdasarkan data sebelumnya.

Bentuk umum dari teori *Naïve Bayes* sebagai berikut: $P(H|X) = (P(X|H)P(H))/P(X)$

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data X merupakan suatu *class* spesifik

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

$P(H)$: Probability hipotesis H (*prior probability*)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada Hipotesis H

$P(X)$: Probabilitas dari X

3) **K-Nearest Neighbour**

Algoritma *K-Nearest Neighbour* (K-NN) merupakan metode klasifikasi yang tidak membangun model dari data latih, melainkan hanya menyimpan dataset tersebut. Proses K-NN berlangsung ketika terdapat *test case* baru, di mana prediksi dilakukan dengan mencari kasus yang paling mirip pada data latih yang tersedia [5]. Dalam penerapan klasifikasi, hasil prediksi umumnya diperoleh dengan memilih sejumlah tetangga terdekat, sehingga nilai K biasanya ditentukan sebagai bilangan ganjil untuk menghindari hasil seri. Kinerja algoritma ini sangat bergantung pada ukuran kesamaan antar kasus, yang umumnya didefinisikan melalui suatu metrik pada ruang input berdasarkan variabel prediktor. Metrik yang paling sering digunakan adalah fungsi jarak Euclidean, yang dirumuskan sebagai berikut:

$$d(X_i, X_j) = \sqrt{\sum_{k=1}^p (X_{i,k} - X_{j,k})^2}$$

P : Jumlah Prediktor

X_i & X_j : Dua Pengamatan

K : Pilihan Jumlah

4) **CRISP-DM**

CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan suatu metodologi yang dirancang untuk menyelesaikan permasalahan penelitian secara sistematis. Pendekatan ini lahir dari kolaborasi beberapa perusahaan besar, antara lain Daimler-Benz, HRA, NCR Corp., dan SPSS Inc., yang mulai dikembangkan sejak tahun 1999 [6].



Gambar 1. Proses CRISP-DM

II. METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian secara umum dapat diartikan sebagai sebuah usaha untuk mencari pengetahuan. Pencarian pengetahuan ini melalui metode pencarian solusi dari permasalahan secara objektif dan sistematis [7].

2.2 Metode Pemilihan Populasi dan Sampel

Populasi merupakan keseluruhan objek atau subjek yang memiliki karakteristik serta kualitas tertentu yang s itu, sampel adalah bagian dari populasi yang mencerminkan jumlah dan karakteristik populasi tersebut. Penelitian biasanya menggunakan sampel agar hasil yang diperoleh dapat digeneralisasikan terhadap populasi secara keseluruhan. Oleh karena itu, sampel yang dipilih harus benar-benar representatif. Pada penelitian ini, sampel yang digunakan adalah data pendonor rutin yang dipilih secara acak dari database pendonor tahun 2018. Data ini merupakan data internal PMI yang dijadikan dasar penelitian, sebagaimana ditunjukkan pada tabel 1 berikut.

Tabel 1. Data Jumlah Sampel Donor

Laki-Laki	Perempuan	Total Sample
90	60	150

2.3 Metode Pengumpulan Data

Metode pengumpulan data umumnya dibedakan menjadi dua sumber, yaitu data primer dan data sekunder. Data primer merupakan data yang dikumpulkan langsung oleh peneliti untuk pertama kalinya guna memperoleh gambaran nyata mengenai kondisi yang diteliti, misalnya melalui observasi, wawancara, maupun kuesioner. Sementara itu, data sekunder adalah data yang sudah pernah dikumpulkan dan diolah oleh pihak lain, baik yang telah dipublikasikan maupun belum. Contohnya dapat berupa dokumen, literatur, buku, jurnal, maupun sumber informasi lain yang relevan dengan permasalahan penelitian.

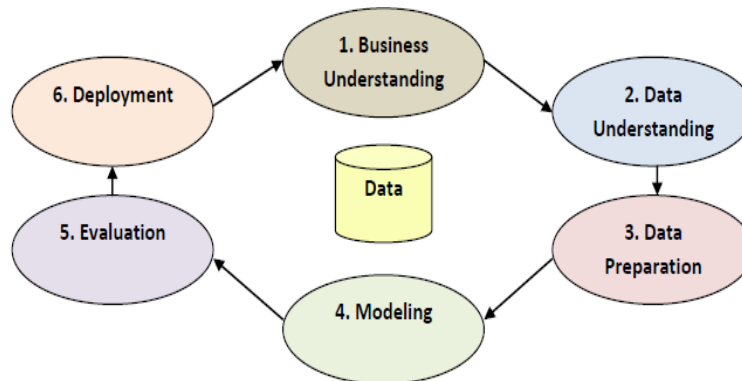
2.4 Instrumen Penelitian

Dalam penelitian ini, instrumen yang digunakan meliputi data sekunder berupa catatan pendonor di Unit Transfusi Darah Palang Merah Indonesia (PMI) Kota Depok sepanjang tahun 2018. Data tersebut dimanfaatkan sebagai instrumen utama untuk menganalisis potensi jumlah pendonor yang rutin melakukan donor darah. Informasi disajikan dalam bentuk tabulasi berdasarkan model dan variabel, dengan total 150 pendonor, terdiri atas 90 laki-laki dan 60 perempuan. Untuk proses analisis, penelitian ini menggunakan perangkat lunak *RapidMiner*.

2.5 Kerangka Kerja Penelitian

Metodologi penelitian merupakan suatu pendekatan yang digunakan untuk menyelesaikan permasalahan penelitian secara sistematis. Melalui metodologi ini, peneliti dapat memahami tahapan umum yang biasanya ditempuh dalam mengkaji masalah riset (Kothari, 2004). Dalam penelitian ini, tahapan yang diacu adalah kerangka kerja CRISP-DM, yang menguraikan langkah-langkah penelitian secara bertahap agar hasil yang diperoleh sesuai dengan tujuan. Pendekatan CRISP-DM juga membantu meminimalisasi kemungkinan terjadinya kesalahan dalam proses penelitian dengan menyediakan alur kerja yang terstruktur.

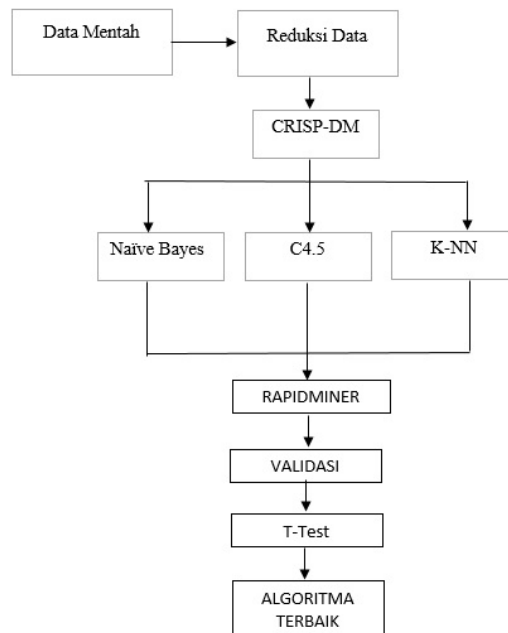
2.6 Tahapan Analisa



Gambar 2. Konsep Model CRISP-DM

- 1) *Business Understanding*
Pada tahap *Business Understanding* atau biasa disebut dengan *Organizational Understanding*, karena semua detail dari permasalahan akan dijawab ditahapan ini untuk memecahkan dari permasalahan yang akan menggunakan data *mining*.
- 2) *Data Understanding*
Pada data *understanding* dimana mulai melakukan proses pengumpulan data yang akan dibuthkan dan menganalisa data yang sudah dikumpulkan.
- 3) *Data Preparation*
Pada tahapan data *preparation* yaitu data akan dilakukan penggabungan atau data akan dikurangi sesuai dengan kebutuhan yang akan digunakan untuk tujuan konsistensi.
- 4) *Modelling*
Pada tahapan *modelling* yaitu penerapan dari algoritma untuk mencari, mengidentifikasi, dan menampilkan pola atau pesan apapun yang terdapat pada data tersebut.
- 5) *Evaluation*
Semua analisa data memiliki potensi positif palsu yaitu sebuah harapan berdasarkan data yang sebenarnya tidak benar, bahkan jika model tidak menghasilkan *false positive* namun model tidak menemukan pola menarik. Tahap evaluasi secara khusus menentukan berharganya dari sebuah model yang telah digunakan.
- 6) *Deployment*
Jika telah berhasil mengidentifikasi pertanyaan, menyiapkan data yang dapat menjawab pertanyaan dan menciptakan sebuah model yang lolos uji menjadi menarik dan berguna, maka telah sampai pada titik yang benar menggunakan hasil yaitu deployment. Dimana tahapan ini mencakup pembuatan otomasi model, bertemu dengan konsumen dari output yang dihasilkan dari data *mining*.

2.7 Langkah-langkah Penelitian



Gambar 3. Langkah-langkah Penelitian

Pada langkah-langkah penelitian di atas dapat dijelaskan sebagai berikut:

- 1) Tahap pertama mengumpulkan data pendonor darah melalui database yang tersedia atau formulir donor peserta donor darah.
- 2) Data direduksi terlebih dahulu kemudian disesuaikan dengan format dan kebutuhan pengolahan data.
- 3) Dengan menggunakan metode CRISP-DM data training di uji dengan algoritma *Naïve Bayes*, *Decision Tree*, dan *K-Nearest Neighbour* untuk mendapatkan dataset dari data tersebut.
- 4) Data diuji dengan menggunakan validasi untuk mendapatkan akurasi dan nilai AUC yang terbaik dari ketiga algoritma tersebut.
- 5) Hasil dari algoritma tersebut dikomparasi guna mendapatkan hasil algoritma yang terbaik.
- 6) Algoritma terpilih digunakan sebagai acuan untuk pedoman menganalisa pendonor darah yang berpotensi untuk mendonorkan darahnya kembali, sehingga didapatkan pendonor tetap.

III. HASIL DAN PEMBAHASAN

3.1 Data Preparation

Tabel 2. Pembagian Atribut dan Label

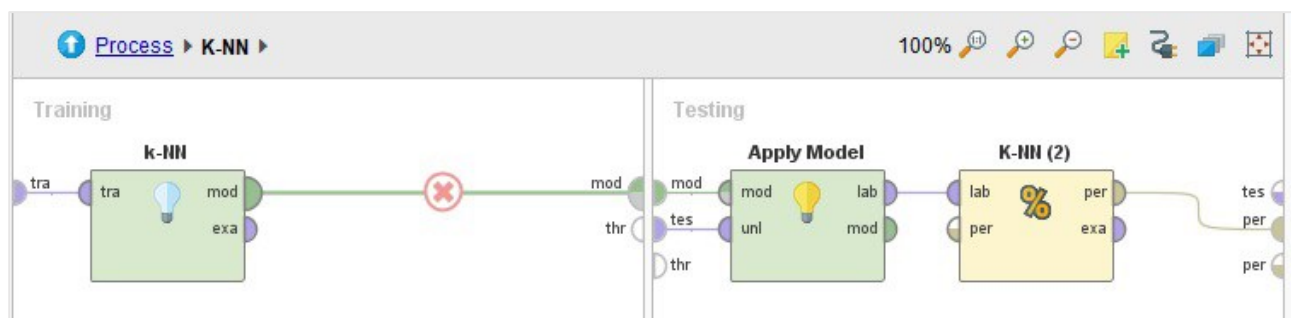
Nama	Tipe	Jenis
Nama Pendonor	Polinomial	Atribut
Jenis Kelamin	Binomial	Atribut
Usia	Integer	Atribut
Kelompok Umur	Polinomial	Atribut
HB	Numeric	Atribut
Kategori HB	Polinomial	Atribut
BB	Integer	Atribut
Gol. Darah	Polinomial	Atribut
Tensi Darah	Polinomial	Atribut
Kategori Tensi	Polinomial	Atribut
MCU	Binomial	Label

A	B	C	D	E	F	G	H	I	J
Nama	Jenis Ke...	Usia	Kelampo...	HB	Kategori ...	BB	Gol. Darah	Tensi Da...	Kategori ...
Lidya Oct...	Peremp...	35.000	Dewasa	14.500	Normal	70.000	A	130/80	Normal
Shelly P...	Peremp...	23.000	Remaja	15.600	Sedang	58.000	O	110/70	Normal
Jodie. S	Laki-Laki	27.000	Dewasa	15.600	Sedang	58.000	O	120/80	Normal
Noviana ...	Peremp...	35.000	Dewasa	14.100	Normal	77.000	O	120/70	Normal
Yoelia P...	Peremp...	27.000	Dewasa	12.700	Cukup	70.000	O	100/70	Cukup
Lusia Sri...	Peremp...	42.000	Dewasa	13.500	Normal	60.000	O	90/70	Rendah
Harry Ku...	Laki-Laki	29.000	Dewasa	17.000	Maksimal	72.000	O	110/70	Normal
Diana Ut...	Peremp...	24.000	Remaja	13.400	Normal	78.000	AB	130/80	Normal
Yeni Yusra	Peremp...	31.000	Dewasa	13.300	Normal	80.000	O	120/80	Normal
Yudi Arie...	Laki-Laki	34.000	Dewasa	14.700	Normal	71.000	B	130/70	Normal
Alief M U...	Laki-Laki	36.000	Dewasa	16.200	Sedang	122.000	A	160/100	Tinggi
Rizki Raf...	Laki-Laki	23.000	Remaja	15.400	Sedang	74.000	O	120/70	Normal
Fariz Adn...	Laki-Laki	28.000	Dewasa	15.000	Sedang	58.000	O	120/80	Normal

Gambar 4. Data Preparation

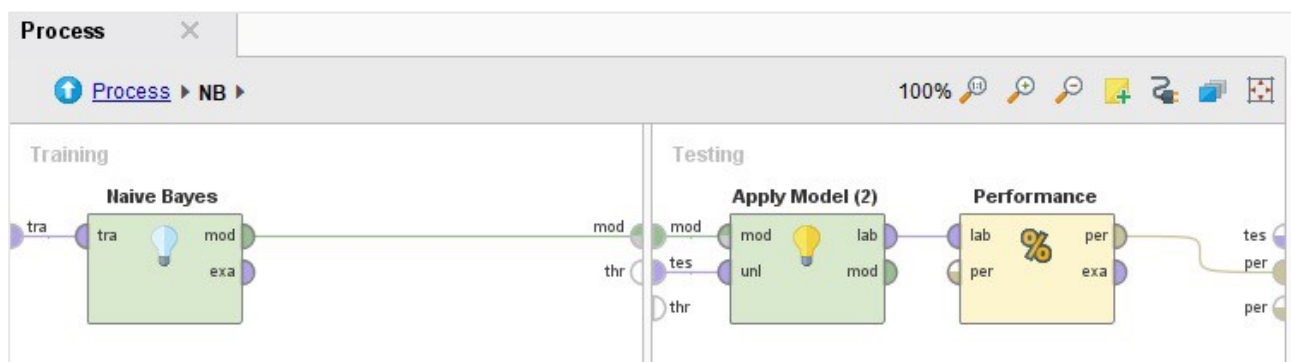
Gambar 4 dataset yang akan digunakan untuk pengolahan dengan menggunakan rapidminer dan data yang digunakan berjumlah 100 *record*.

3.2 Modelling



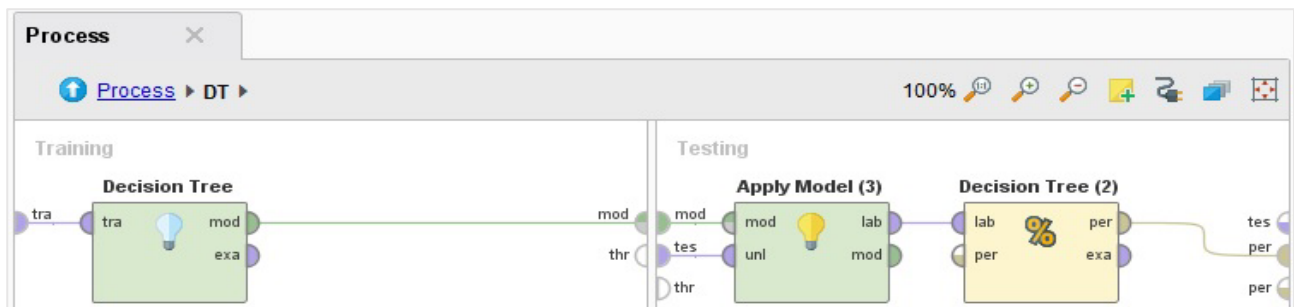
Gambar 5. Desain Proses 10-Fold Cross Validation K-Nearest Neighbour

Pada gambar 5 adalah desain proses didalam 10-Fold Cross Validation untuk *K-Nearest Neighbour* digunakan untuk melakukan training terhadap *dataset* masukan kemudian untuk *testing* dengan *apply model* dan *performance* yang bertujuan untuk mengetahui nilai *accuracy* dan *auc* dari *K-Nearest Neighbour*.



Gambar 6. Desain Proses 10-Fold Cross Validation Naïve Bayes

Pada gambar 6 adalah desain proses didalam 10-Fold Cross Validation untuk *Naïve Bayes* digunakan untuk melakukan *training* terhadap *dataset* masukan kemudian untuk *testing* dengan *apply model* dan *performance* yang bertujuan untuk mengetahui nilai *accuracy* dan *auc* dari *Naïve Bayes*.



Gambar 7. Desain Proses 10-Fold Cross Validation Decision Tree C4.5

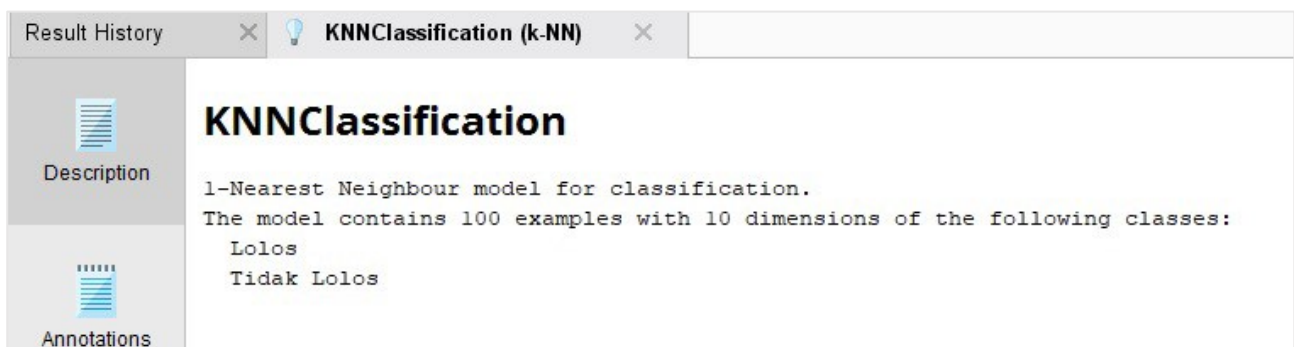
Pada gambar 7 adalah desain proses didalam 10-Fold Cross Validation untuk Decision Tree digunakan untuk melakukan *training* terhadap *dataset* masukan kemudian untuk *testing* dengan apply model dan *performance* yang bertujuan untuk mengetahui nilai *accuracy* dan *auc* dari Decision Tree.

Cross Validation melakukan proses 10-Fold Cross Validation untuk 3 (tiga) algoritma yang digunakan. Setiap percobaan akan mendapatkan nilai *accuracy* dan nilai *kurva* ROC – AUC. Hasil dari *accuracy* disajikan dalam bentuk kurva.

3.3 K-Nearest Neighbour

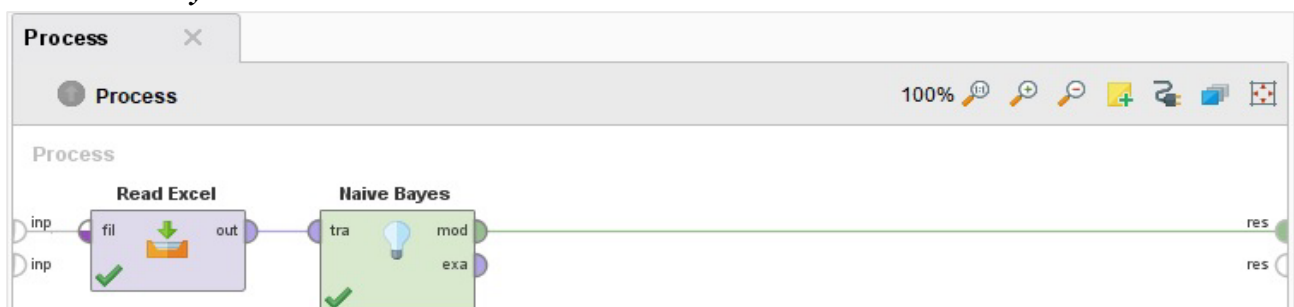


Gambar 8. Proses K-Nearest Neighbour

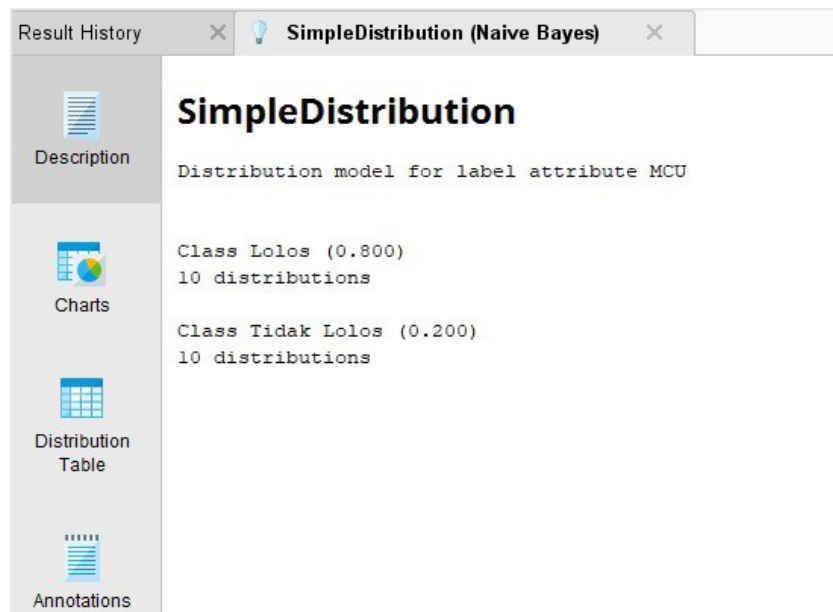


Gambar 9. Klasifikasi K-Nearest Neighbour

3.4 Naïve Bayes

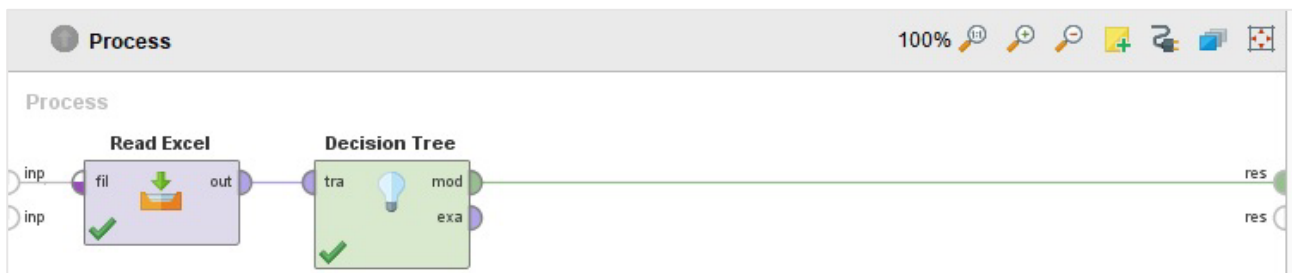


Gambar 10. Proses Naïve Bayes

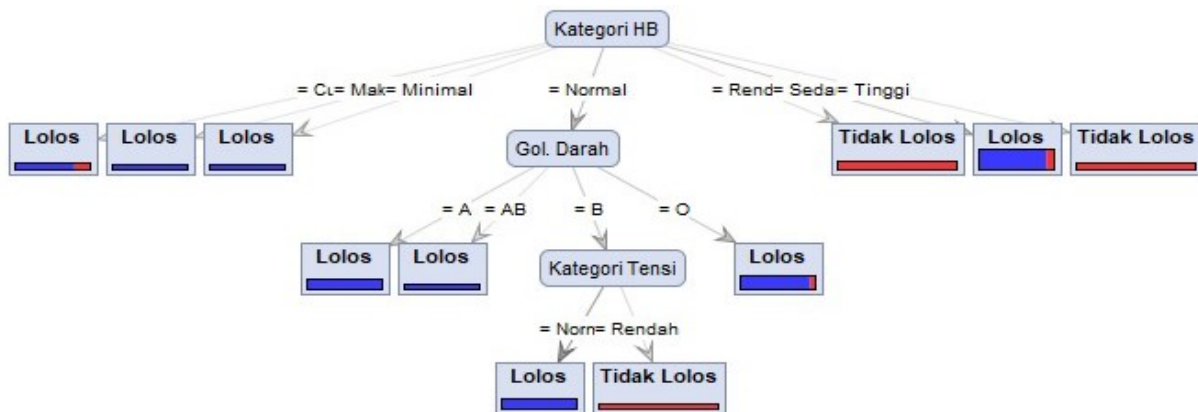


Gambar 11. Simple Distribution Naïve Bayes

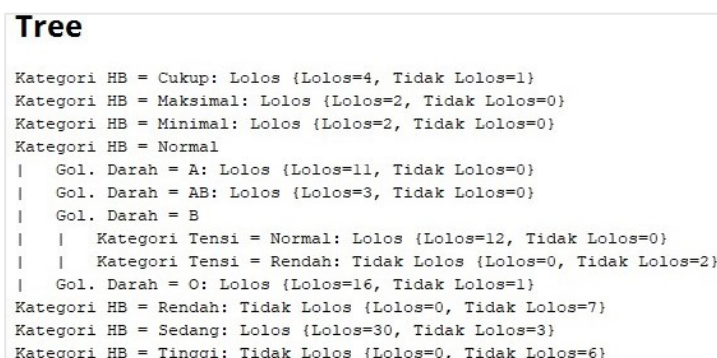
3.5 Decision Tree C4.5



Gambar 12. Proses Decision Tree C4.5



Gambar 13. Model Decision Tree C4.5



Gambar 14. Deskripsi Decision Tree

Dari *rule* yang terlihat pada pohon keputusan dan deskripsi *Decision Tree* terlihat pola analisa seleksi donor Unit Transfusi Darah PMI Kota Depok Tahun 2018. Menggunakan label MCU (Medical Check Up) dapat diketahui model seleksi donor. Adapun *rule* atau model yang didapat adalah:

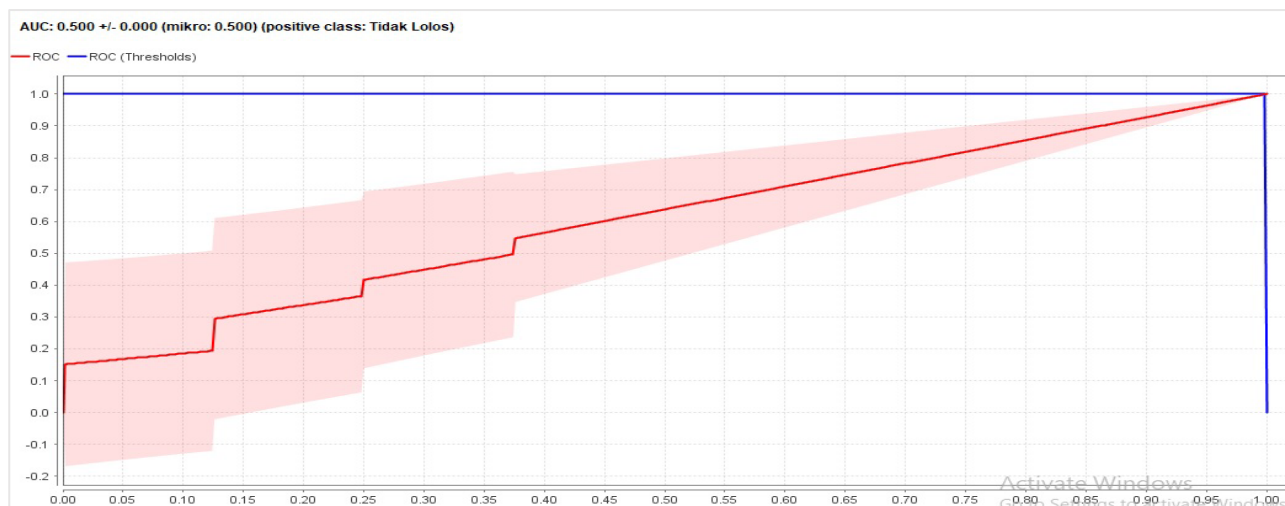
- 1) Apabila calon pendonor memiliki kadar HB dengan kategori normal maka dinyatakan lolos untuk pemeriksaan selanjutnya, kadar HB dengan kategori normal yaitu 13.0 sampai dengan 16.9.
- 2) Apabila calon pendonor memiliki kadar HB dengan kategori minimal maka dinyatakan lolos untuk pemeriksaan berikutnya, kadar HB dengan kategori cukup yaitu 12.7 sampai dengan 12.9.
- 3) Apabila calon pendonor memiliki kadar HB dengan kategori minimal maka dinyatakan lolos untuk pemeriksaan berikutnya, kadar HB dengan kategori minimal yaitu 12.6.
- 4) Apabila calon pendonor memiliki kadar HB dengan kategori rendah maka dinyatakan tidak lolos untuk pemeriksaan berikutnya, kadar HB dengan kategori rendah yaitu dibawah 12.6.
- 5) Apabila calon pendonor memiliki kadar HB dengan kategori maksimal maka dinyatakan lolos untuk pemeriksaan berikutnya, kadar HB dengan kategori maksimal yaitu 17.0.
- 6) Apabila calon pendonor memiliki kadar HB dengan kategori tinggi akan dinyatakan tidak lolos untuk pemeriksaan berikutnya, kadar HB dengan kategori tinggi yaitu lebih dari 17.0.
- 7) Dari pemeriksaan HB calon pendonor akan ditentukan oleh pemeriksaan tekanan darah (tensi), jika dikategorikan tensi normal maka calon pendonor tersebut bisa menyumbangkan darahnya karena lolos medical check up. Tensi normal yaitu 110/70 sampai dengan 150/90.
- 8) Jika tekanan darah dikategorikan tensi rendah maka calon pendonor tersebut tidak dapat menyumbangkan darahnya karena tidak lolos medical check up. Tensi rendah yaitu dibawah 110/70.
- 9) Jika tekanan darah dikategorikan tensi tinggi maka calon pendonor tersebut tidak dapat menyumbangkan darahnya karena tidak lolos medical check up. Tensi tinggi yaitu diatas 150/90.

Bentuk *rule* atau model yang terbentuk dari pohon keputusan dan deskripsi *Decision Tree* adalah model yang selalu dilakukan pada tahap awal seleksi donor pada Unit Transfusi Darah PMI Kota Depok.

3.6 Akurasi dan AUC *K-Nearest Neighbour*

accuracy: 79.00% +/- 10.44% (mikro: 79.00%)		
	true Lolos	true Tidak Lolos
pred. Lolos	72	13
pred. Tidak Lolos	8	7
class recall	90.00%	35.00%

Gambar 15. Hasil Akurasi *K-Nearest Neighbour* dengan Akurasi 79.00%

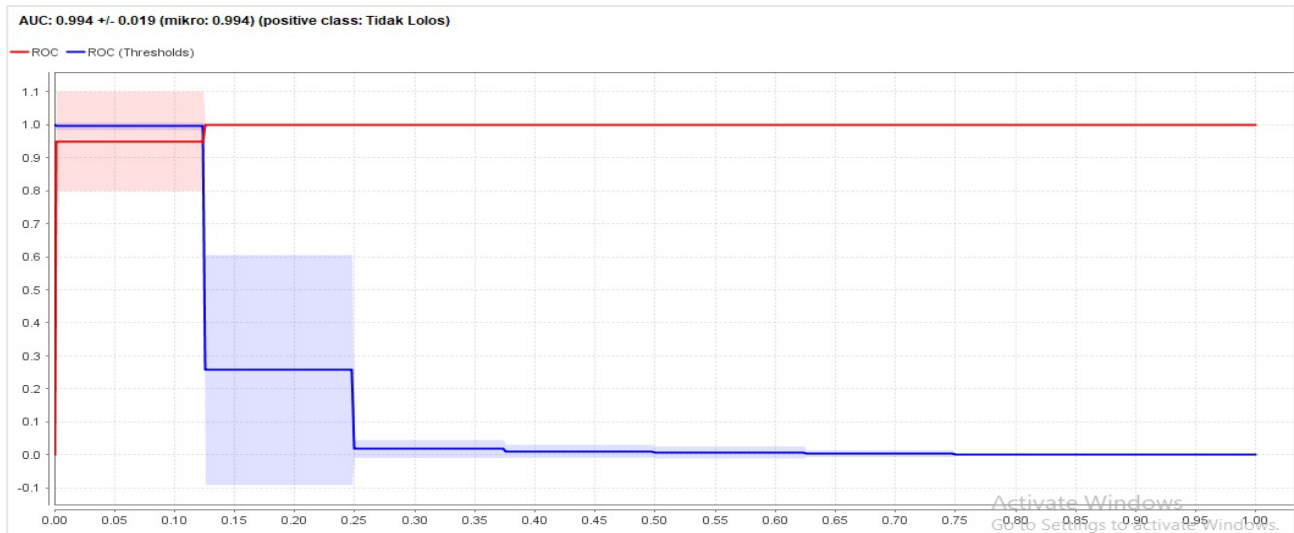


Gambar 16. Hasil AUC *K-Nearest Neighbour* Sebesar 0.500

3.7 Akurasi dan AUC Naïve Bayes

accuracy: 97.00% +/- 4.58% (mikro: 97.00%)			
	true Lolos	true Tidak Lolos	class precision
pred. Lolos	77	0	100.00%
pred. Tidak Lolos	3	20	86.96%
class recall	96.25%	100.00%	

Gambar 17. Hasil Akurasi *Naïve Bayes* dengan Akurasi 97.00%

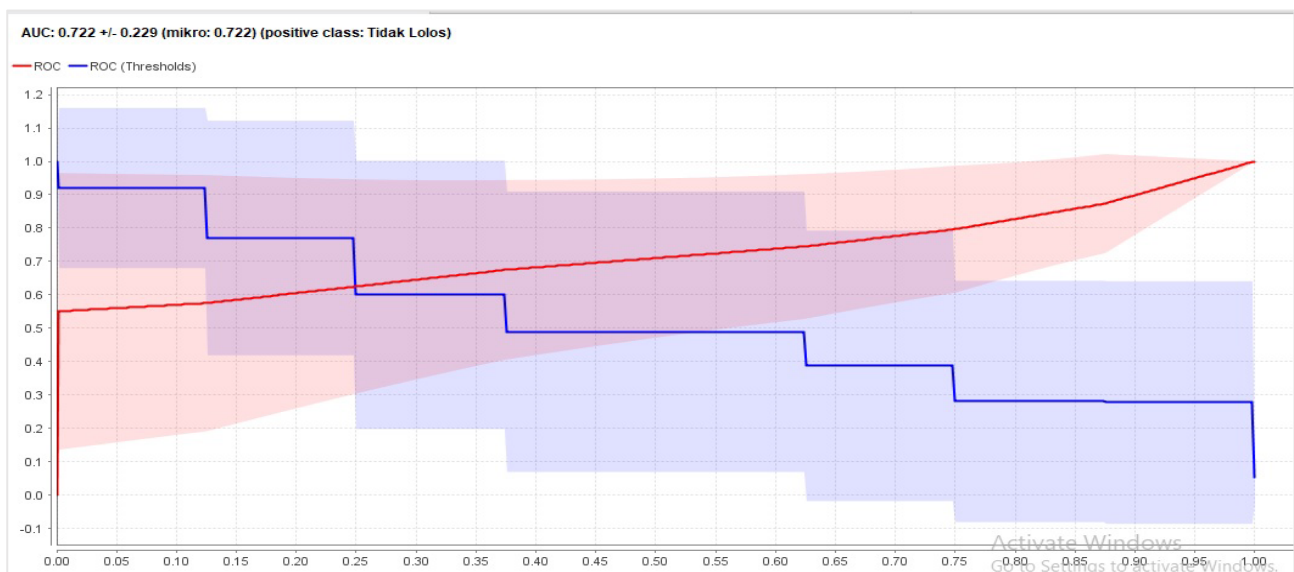


Gambar 18. Hasil AUC *Naïve Bayes* sebesar 0.994

3.8 Akurasi dan AUC Decision Tree C4.5

accuracy: 90.00% +/- 7.75% (mikro: 90.00%)			
	true Lolos	true Tidak Lolos	class precision
pred. Lolos	80	10	88.89%
pred. Tidak Lolos	0	10	100.00%
class recall	100.00%	50.00%	

Gambar 19. Hasil Akurasi *Decision Tree* dengan Akurasi Sebesar 90.00%



Gambar 20. Hasil AUC *Decision Tree* Sebesar 0.722

3.9 Hasil

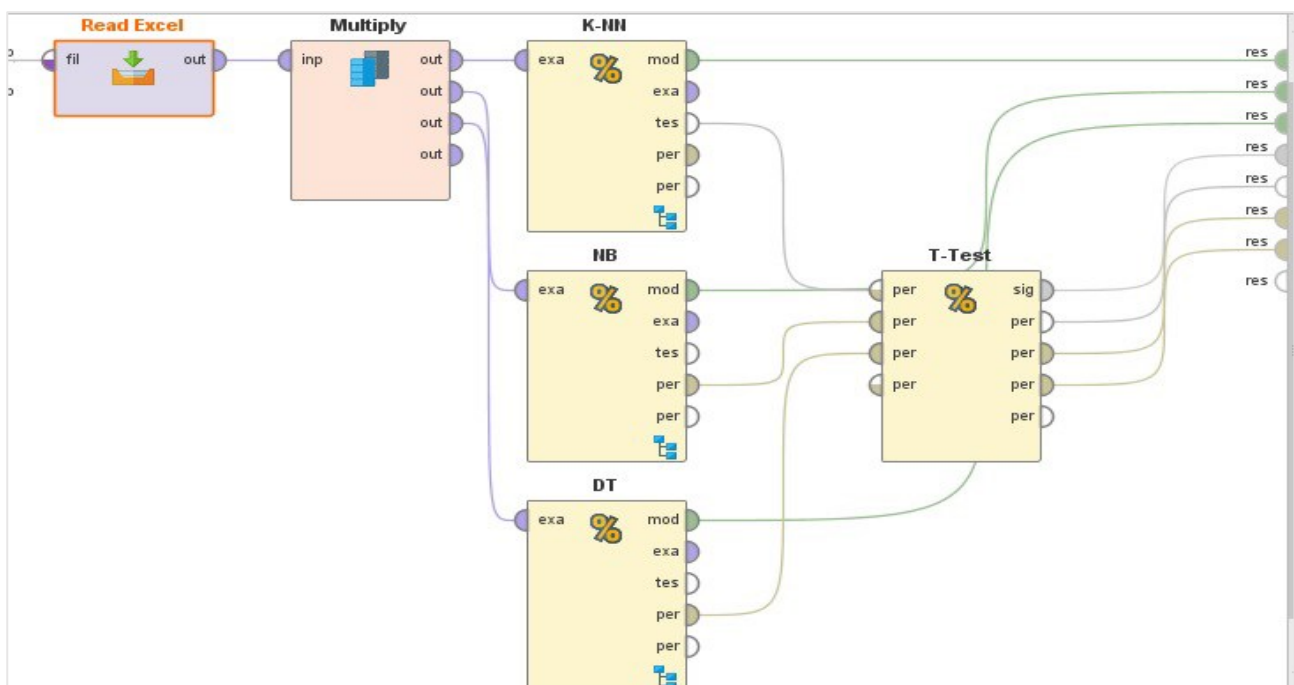
Tabel 3. Komparasi Nilai Akurasi dan AUC

	K-NN	Naïve Bayes	C4.5
AKURASI	79.00%	97.00%	90.00%
AUC	0.500	0.994	0.722

Pada sisi akurasi tabel 3 menggambarkan bahwa hasil perbandingan nilai Akurasi, menunjukan bahwa algoritma *Naïve Bayes* mendapatkan nilai tertinggi yaitu 97.00%, diikuti dengan C4.5 dengan nilai 90.00%, dan *K-Nearest Neighbour* dengan nilai 79.00%. Dari ketiga algoritma yang diuji berdasarkan nilai akurasi, maka didapatkan bahwa algoritma *Naïve Bayes* memperoleh nilai terbaik dari sisi akurasinya.

3.10 Uji Beda

Karena dari hasil pemilihan algoritma belum dapat ditentukan secara mutlak dan langsung tentang mana algoritma terbaik, maka dilakukan uji beda guna mendapatkan algoritma terbaik dari ketiga algoritma yang diuji. Ketiga algoritma yang akan digunakan akan di uji dengan operator *T-test* dari *Rapidminer* sehingga dapat menghasilkan penilaian tentang algoritma terbaik yang akan digunakan. Hal ini dapat dilihat pada desain uji beda 21 dibawah ini:



Gambar 21. Proses T-Test K-NN, Naïve Bayes dan C4.5

Setelah proses Rapid Miner dijalankan maka didapatkan hasil T-test seperti pada gambar 22 dibawah ini:

A	B	C	D
	0.790 +/- 0.104	0.970 +/- 0.046	0.900 +/- 0.077
0.790 +/- 0.104		0.000	0.015
0.970 +/- 0.046			0.024
0.900 +/- 0.077			

Gambar 22. Hasil Uji Beda

Dari hasil uji beda terlihat bahwa algoritma *Naïve Bayes* mempunyai nilai terbaik yaitu 0.000 dibandingkan dengan algoritma *K-Nearest Neighbour* dan *Decision Tree*. Dengan demikian dari uji beda dapat ditarik kesimpulan bahwa algoritma *Naïve Bayes* adalah algoritma terbaik yang dapat digunakan pada penelitian ini. Hal tersebut ditunjukan pula bahwa algoritma *Naïve Bayes* mempunyai nilai probabilitas terbaik diantara ketiga algoritma yang ada karena memiliki nilai $\alpha \leq 0.05$, yang berarti terkecil dibandingkan dengan kedua algoritma lainnya.

IV. KESIMPULAN

Dalam penelitian ini dilakukan pengujian model dengan membandingkan tiga metode data *mining* yaitu algoritma *K-Nearest Neighbour*, *Naïve Bayes* & *Decision Tree*. Dari hasil evaluasi dan validasi dapat diketahui:

- 1) Algoritma *Naïve Bayes* memiliki nilai akurasi tertinggi yaitu sebesar 97.00% lebih baik dibandingkan dengan *K-Nearest Neighbour* yang memiliki nilai akurasi 79.00% dan *Decision Tree* memiliki nilai akurasi 90.00%. Sehingga secara akurasi algoritma *Naïve Bayes* adalah yang memiliki nilai akurasi terbaik.
- 2) Untuk nilai AUC yang memiliki nilai paling tinggi adalah algoritma *Naïve Bayes* dengan nilai AUC 0.994 diikuti oleh algoritma *Decision Tree C4.5* dengan nilai AUC 0.722 dan algoritma *K-Nearest Neighbour* dengan nilai AUC 0.500. sehingga secara nilai AUC algoritma *Naïve Bayes* adalah algoritma yang memiliki nilai AUC terbaik.
- 3) Berdasarkan uji beda yang menguji ketiga algoritma tersebut didapatkan hasil untuk algoritma *Naïve Bayes* dengan nilai sebesar 0.000 ketika diuji, sehingga setelah melalui uji beda didapat kesimpulan bahwa algoritma *Naïve Bayes* adalah algoritma yang terbaik karena memiliki nilai $\alpha < 0.05$.

REFERENSI

- [1] H. Wahono and D. Riana, "Prediksi Calon Pendorong Darah Potensial dengan Algoritma *Naïve Bayes*, *K-Nearest Neighbors* dan *Decision Tree C4.5*," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, pp. 7-14, 2020.
- [2] A. Astuti, N. W. Utami, dan I. G. Juliharta, "Classification of Blood Donor Data Using C4.5 and K-Nearest Neighbor Method (Case Study: UTD PMI Bali Province)," *PILAR Nusa Mandiri: Journal of Computing and Information System*, vol. 18, no. 1, pp. 9-16, 2022.
- [3] K. J. Atmaja, I. B. Anandita, dan N. K. Dewi, "Penerapan Data Mining untuk Memprediksi Potensi Pendorong Darah Menjadi Pendorong Tetap Menggunakan Metode *Decision Tree C4.5*," *Jurnal Ilmu Komputer dan Sains Terapan*, vol. 7, no. 2, pp. 101-108, 2017.
- [4] M. Yunus, H. S. Dahlan, dan P. B. Santoso, "SPK Pemilihan Calon Pendorong Darah Potensial
- [5] S. Nurkasanah, C. C. Dewi, E. J. Qosiani, D. A. Kholifatin, dan E. S. Anggita, "Analysis of voluntary blood donors' characteristics during the third-peak COVID-19 in Bojonegoro Regency, Indonesia," *Journal of Community Empowerment for Health*, vol. 6, no. 1, pp. 13-17, 2023.
- [6] D. C. McElfresh et al., "Matching Algorithms for Blood Donation," preprint arXiv, 2021.
- [7] F. Angellia, W. Cahya, R. Ramadhan, and I. Hariyansah, "Sosialisasi Penggunaan Virtual Meeting Untuk Mendukung Kegiatan Bisnis UMKM di Masa Pandemi Covid 19," *J. Pengabd. Teratai*, vol. 1, no. 2, pp. 206–213, 2020, doi: 10.55122/teratai.v1i2.153.
- [8] B. Firmansyah, "Implementasi Teknologi Qr-Code Sebagai Pencarian Data Ruangan Pada Ibi Kosgoro 1957 Berbasis Android," *JunifJurnal Nas. Inform.*, vol. 1, no. 1, pp. 30–42, 2020.
- [9] S. Hidayat, W. Cahya, and H. Rifiyanti, "Bimbingan Teknis Pengelolaan Konten Pemasaran Digital Pada Lembaga Bimbingan Belajar Ahe Tanah Baru," vol. 4, no. 2, pp. 189–194, 2023.