Prediksi Calon Pendoror Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5

Hermanto Wahono*, Dwiza Riana

Program Studi Ilmu Komputer, STMIK Nusa Mandiri, Jakarta, Indonesia

Email: ¹*mr.h3rm4n@gmail.com, ²dwiza@nusamandiri.ac.id

Email Penulis Korespondensi mr.h3rm4n@gmail.com

Submitted 13-01-2020; Accepted 21-01-2020; Published 15-02-2020

Abstrak

Donor darah merupakan proses pengambilan darah dari pendonor yang dinyatakan layak, ditinjau dari berbagai faktor antara lain usia, berat badan, tekanan darah, kadar hemoglobin, dan status donor yang menjadi pertimbangan saat uji kelayakan. Penelitian ini dilakukan untuk menemukan metode paling tepat dengan nilai akurasi dan nilai Area Under Curve (AUC) yang tinggi menggunakan 3710 dataset donor darah dari PMI Kota Bekasi, diproses menggunakan metode algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5. Hasil analisa menunjukkan bahwa algoritma Decision Tree C4.5 menunjukkan akurasi yang lebih tinggi sebesar 93.83% dibandingkan algoritma Naïve Bayes yang menunjukkan nilai akurasi sebesar 85.15% dan algoritma K-Nearest Neighbors dengan nilai akurasi sebesar 84.10%. Selain nilai-nilai tersebut, Decision Tree C4.5 juga unggul secara visual dimana Decision Tree memiliki output sebuah model pohon yang menunjukkan hubungan atribut dan memiliki nilai AUC sebesar 0.978, Naïve Bayes dengan nilai AUC sebesar 0.927 dan K-Nearest Neighbors dengan nilai AUC sebesar 0.816

Kata Kunci: Donor Darah, Naive Bayes, K-Nearest Neighbors, Decision Tree C4.5

Abstract

Blood donation is a process of taking blood from donors that is declared feasible, in terms of various factors including age, weight, blood pressure, hemoglobin levels, and donor status which are taken into consideration during the feasibility test. This study was conducted to find the most appropriate method with high accuracy and Area Under Curve (AUC) values using 3710 blood donor datasets from the Bekasi City PMI, processed using the Naïve Bayes algorithm method, K-Nearest Neighbors and Decision Tree C4.5. The analysis shows that the Decision Tree C4.5 algorithm shows higher accuracy of 93.83% compared to Naïve Bayes algorithm which shows an accuracy value of 85.15% and the K-Nearest Neighbors algorithm with an accuracy value of 84.10%. In addition to these values, Decision Tree C4.5 is also visually superior where the Decision Tree has an output model tree that shows attribute relationships and has an AUC value of 0.978, Naïve Bayes with an AUC value of 0.927 and K-Nearest Neighbors with an AUC value of 0.816.

Keywords: Blood Donation, Naive Bayes, K-Nearest Neighbors, Decision Tree C4.5

1. PENDAHULUAN

Palang Merah Indonesia (PMI) merupakan sebuah organisasi perhimpunan nasional di Indonesia yang bergerak dalam bidang sosial kemanusiaan, salah satu tugas dari PMI adalah melayani permintaan darah oleh masyarakat yang membutuhkan. PMI didirikan oleh Pemerintah kolonial Belanda dengan nama *Het Nederland-Indische Rode Kruis* (NIRK) yang kemudian namanya menjadi *Nederlands Rode Kruis Afdeling Indie* (NERKAI) pada tanggal 21 Oktober 1873. PMI menjadi organisasi kemanusiaan yang berstatus badan hukum, diundangkan dengan Undang-Undang nomor 1 tahun 2018 tentang Kepalangmerahan guna menjalankan kegiatan Kepalangmerahan sesuai dengan Konvensi Jenewa Tahun 1949, dengan tujuan untuk mencegah dan meringankan penderitaan dan melindungi korban tawanan perang dan bencana, tanpa membedakan agama, bangsa, suku bangsa, warna kulit, jenis kelamin, golongan, dan Pandangan Politik [1].

Pelayanan darah adalah upaya pelayanan kesehatan yang memanfaatkan darah manusia sebagai bahan dasar dengan tujuan kemanusiaan dan tidak untuk tujuan komersial. Peraturan Pemerintah No.7 tahun 2011 tentang Pelayanan Darah menyebutkan penyelenggaraan donor darah dan pengolahan darah dilakukan oleh Unit Donor Darah (UDD) yang diselenggarakan oleh organisasi sosial dengan tugas pokok dan fungsinya di bidang Kepalangmerahan atau dalam hal ini Palang Merah Indonesia (PMI). Lebih lanjut, baik dalam UU No. 36/2009 tentang Kesehatan maupun Peraturan Pemerintah No.7/2011 tentang Pelayanan Darah, dinyatakan bahwa Pemerintah bertanggung jawab atas pelaksanaan pelayanan darah yang aman, mudah diakses, dan sesuai dengan kebutuhan masyarakat. Tanggung jawab pemerintah dan pemerintah daerah (Pemda) meliputi pengaturan, pembinaan, pengawasan dan pendanaan pelayanan darah untuk kepentingan pelayanan kesehatan. Sesuai penjelasan UU No. 36/2009 tentang Kesehatan Pasal 90 dan PP No. 7/2011 tentang Pelayanan Darah Pasal 46, jaminan pendanaan pemerintah diwujudkan dalam bentuk pemberian subsidi kepada UDD dari APBN, APBD dan bantuan lainnya [2].

Dalam rangka pemenuhan darah yang sangat penting dalam menyelamatkan nyawa seseorang, PMI terus mengampanyekan donor darah sebagai bagian dari gaya hidup (lifestyle). Setiap tahunnya, PMI menargetkan hingga 4,5 juta kantong darah sesuai dengan kebutuhan darah nasional, disesuaikan dengan standar Lembaga Kesehatan Internasional (WHO) yaitu 2% dari jumlah penduduk untuk setiap harinya [3].

Hal yang perlu diperhatikan, darah merupakan salah satu media yang digunakan untuk penularan berbagai penyakit, maka sebaiknya darah transfusi berasal dari donor resiko rendah. Menurut rekomendasi *World Health Organization* (WHO) darah beresiko rendah yang aman dan berkualitas adalah darah yang berasal dari pendonor sukarela, daripada berasal dari donor pengganti [4]. Donor sukarela adalah pendonor yang memberikan darahnya dimana darah tersebut tidak diperuntukkan bagi orang tertentu sebagai pengguna darah yang didonorkan sedangkan yang maksud dengan donor

pengganti ialah pendonor yang memberikan darahnya ditujukan untuk orang tertentu yang telah ditunjuk oleh pendonor [5] Namun minimnya kesadaran masyarakat untuk menjadi pendonor sukarela hal ini menjadi salah satu faktor yang mengakibatkan persediaan darah di Unit Donor Darah (UDD) menjadi rendah.

Pada penelitian sebelumnya, penentuan calon pendonor darah masih bersifat manual dan belum terkomputerisasi secara utuh. Penentuan calon pendonor darah berdasarkan kriteria tertentu. Dalam menghadapi tantangan tersebut beberapa penelitian telah dilakukan untuk menentukan pendonor darah yang potensial untuk mendonorkan darahnya kembali. Sudah banyak literatur mengenai pembahasan prediksi calon pendonor darah yang menerapkan beberapa metode, antara lain [6] dengan judul penelitian: *Blood Donor Classification Using Neural Network and Decision Tree Techniques*, [7] dengan judul penelitian SPK Pemilihan Calon Pendonor Darah Potensial dengan Algoritma C4.5 dan Fuzzy Tahani, [8] dengan judul pembuatan aplikasi data mining untuk pola permintaan darah di UDD (unit donor darah) PMI Kota Surabaya menggunakan metode apriori, [9] melakukan penelitian tentang Komparasi Algoritma Neural Network, K-Nearest Neighbor Dan Naive Baiye Untuk Memprediksi Pendonor Darah Potensial, [10] melakukan penelitian penentuan calon pendonor darah menggunakan algoritma *naive bayes classification* (studi kasus PMI Semarang), [11] melakukan penelitian Komparasi Akurasi Algoritma C4.5 dan *Naive Bayes* untuk Prediksi Pendonor Darah Potensial dengan Dataset RFMTC, [12] melakukan penelitian dan pembuatan *Application of blood donor routine detector using K-Nearest neighbors*

Dari beberapa penelitian yang telah dilakukan, penggunaan beberapa metode algoritma telah dilakukan uji nilai akurasi, presisi dan penarikan kembali (*recall*). Dari penelitian yang telah dijelaskan diatas maka algoritma *Neural Network* memiliki nilai akurasi, presisi, dan penarikan kembali untuk mengklasifikasikan donor darah lebih baik daripada nilai-nilai pada model algoritma *Decision Tree* seperti pada penelitian yang dilakukan oleh [6] dan [9]. Namun pada penelitian yang dilakukan oleh [7] penggunaan algoritma *Decision Tree* C4.5 memperoleh nilai akurasi yang lebih tinggi dibandingkan dengan algoritma *Fuzzy Tahani*.

Untuk mengatasi permasalahan tersebut, kehadiran cabang ilmu baru di bidang komputer *data mining* tak luput perhatian dalam dunia sistem informasi kesehatan. *Data mining* merupakan langkah penting dalam penemuan pengetahuan dari serangkaian dataset dalam ukuran yang besar. *Data Mining* adalah serangkaian data untuk dihasilkan informasi yang lebih berguna, disiplin ilmu ini mengkaji berbagai metode yang umum digunakan untuk melakukan pengolahan data, salah satu metode pengolahan dalam prosesnya adalah klasifikasi data. Klasifikasi data biasa digunakan pada sejumlah data yang telah di ketahui data induknya, untuk kemudian dijadikan data *training*/data model yang hasilnya akan menjadi keputusan prediksi dari sejumlah data yang serupa namun belum lengkap pada salah satu atributnya [13].

2. METODE PENELITIAN

2.1 Data Mining

Data Mining merupakan proses penggalian dan pertambangan pengetahuan dari sejumlah data yang besar, database atau repository database lainnya. Tujuan utama dari penambangan data ini untuk menemukan pengetahuan baru yang tersembunyi dari database tersebut.[14]

Data mining adalah suatu rangkaian proses yang dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara *knowledge base*.

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Tahapan proses yang umumnya dilakukan oleh *data mining* antara lain: deskripsi, prediksi, estimasi, klasifikasi, clustering dan asosiasi [15]. Menurut [16] proses *data mining* dijelaskan sebagai berikut:

a. Deskripsi

Deskripsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (*knowledge*) pada sistem. Tugas deskriptif merupakan tugas *data mining* yang sering dibutuhkan pada teknik postprocessing untuk melakukan validasi dan menjelaskan hasil dari proses *data mining*. Postprocessing merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

b. Prediksi

Prediksi memiliki kemiripan dengan klasifikasi, akan tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang.

c. Estimasi

Estimasi hampir sama dengan prediksi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi.

d. Klasifikasi

Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya.

e. *Clustering*

Clustering merupakan pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama. Sebuah kluster adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu cluster dan semakin besar perbedaan tiap cluster maka kualitas analisis kluster semakin baik.

f. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analysis). Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

2.2 Decision Tree

Pohon (tree) adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root node), simpul percabangan/ internal (branch/ internal node) dan simpul daun (leaf node). Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [17].

2.2 Naïve Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes.

Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.[18]

Langkah-langkah algoritma *Naive Bayes* dapat diuraikan sebagai berikut : [18]

- Baca data *training*
- Cari nilai probabilitasnya dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut
- Mendapatkan nilai dalam tabel probabilitas.

Klasifikasi *Bayes* didasarkan pada teorema *Bayes*, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas *Bayes* (1702-1761), yaitu:

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)} \quad (1)$$

Keterangan:

- y = data dengan kelas yang belum diketahui
- x = hipotesis data y merupakan suatu kelas spesifik
- $P(x|y)$ = probabilitas hipotesis x berdasarkan kondisi y
- $P(x)$ = probabilitas hipotesis x
- $P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x
- $P(y)$ = probabilitas dari y

Naïve Bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. [19]

Metode ini penting karena beberapa alasan, hal ini sangat mudah untuk dibangun, tidak perlu ada yang rumit untuk parameter estimasi skema yang berulang. Ini berarti dapat segera diterapkan untuk dataset yang besar. Sangat mudah untuk menafsirkan, sehingga pengguna yang tidak terampil dalam teknologi *classifier* dapat memahami dalam membuat klasifikasi. Ini mungkin bukan *classifier* terbaik dalam setiap diberikan aplikasi, tetapi biasanya dapat diandalkan untuk menjadi kuat dan melakukan *processing data* dengan sangat baik [20].

2.3 K-Nearest Neighbors

Algoritma *K-Nearest Neighbors* (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma K-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik. K-NN juga merupakan contoh teknik lazy learning, yaitu teknik yang menunggu sampai pertanyaan (query) datang agar sama dengan data training [21].

2.4 RapidMiner

RapidMiner adalah platform perangkat lunak data *science* yang dikembangkan oleh perusahaan dengan nama yang sama yang menyediakan lingkungan terpadu untuk persiapan data (*data preparation*), pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*).

Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, *prototyping* yang cepat, dan pengembangan aplikasi dan mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi model dan pengoptimalan.

RapidMiner dikembangkan dengan model open core. *RapidMiner Studio Free Edition*, yang terbatas pada 1 prosesor logis dan 10.000 baris data. Didalam aplikasi *Rapidminer* terdapat istilah operator. Operator di *RapidMiner* berupa Blok bangunan yang dikelompokkan berdasarkan fungsinya dan digunakan untuk membuat proses *RapidMiner*. Operator memiliki *port input* dan *output*; Tindakan yang dilakukan pada *input* pada akhirnya mengarah pada apa yang dipasok ke *output*. Parameter operator mengendalikan tindakan tersebut. Ada lebih dari 1500 operator yang ada di *RapidMiner* [22].

3. ANALISA DAN PEMBAHASAN

3.1 Business Understanding

Pada tahap *Business Understanding* atau biasa disebut dengan *Organizational Understanding* adalah untuk memahami tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

3.2 Data Understanding

Pada tahapan ini peneliti melakukan penelitian data pendonor pada UTD PMI Kota Bekasi berdasarkan data pendonor darah pada bulan Oktober 2019, kemudian menyusunnya berdasarkan atribut. Tahap ini adalah mengumpulkan data primer dari Unit Transfusi Darah (UTD) PMI Kota Bekasi berdasarkan data pendonor darah dengan jumlah 3710 orang.

3.3 Data Preparation

Pada tahap peneliti melakukan kegiatan untuk membangun *dataset* akhir. Data yang akan diproses pada tahap pemodelan dari data mentah untuk diseleksi berdasarkan delapan atribut *predictor* dan satu atribut hasil agar dapat diketahui status pendonor darah yang berhasil atau gagal dan nilai kategori dapat dilihat pada tabel berikut ini

Tabel 1. Representasi Atribut Prediksi Pendonor Darah

Atribut	Deskripsi	Satuan
Umur	Umur	Tahun
Gol	Golongan Darah	A, B, AB, O
JK	Jenis kelamin	Pria, Wanita
Status Donor	Status Donor	Baru, Ulang
Donor	Frekuensi Donor	Bulan
HB	Ukuran Hemoglobin	gram
Tensi	Tensi darah pendonor	mmHg
Berat Badan	Berat Badan pendonor	kg

Setelah atribut dan nilai kategori ditentukan, selanjutnya dataset yang diperoleh akan dilakukan validasi berdasarkan atribut yang ada untuk dilakukan eksperimen lebih lanjut dengan menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbors*, dan *Decision Tree C4.5*. Hasil dari penelitian ini berupa hasil proses pengolahan kualitatif dan kuantitatif yang telah dikumpulkan dengan perhitungan berdasarkan model yang diusulkan. Hasil validasi data awal dapat dilihat pada tabel sampel dataset berikut ini.

Tabel 2. Dataset Pendonor Darah

Umur	Gol	JK	Status Donor	Donor Ke-	HB	Tensi	Berat Badan	Kesimpulan
40	B+	Wanita	Baru	4	9,2	120/80	50	Batal
21	A+	Wanita	Baru	3	9,9	110/70	54	Batal
26	O+	Wanita	Baru	3	10,1	110/80	56	Batal
20	O+	Wanita	Baru	1	10,2	100/80	50	Batal
30	O+	Wanita	Baru	1	10,4	110/70	62	Batal
46	B+	Pria	Baru	4	10,4	140/80	75	Batal
29	AB+	Pria	Baru	11	10,5	120/80	75	Batal
21	O+	Pria	Baru	15	10,7	120/80	75	Batal

22	B+	Wanita	Baru	2	10,7	110/80	74	Batal
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Pria	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Pria	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Pria	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
17	B+	Wanita	Baru	1	12,5	120/80	50	Berhasil
....								
....								

3.4 Modelling

Pada tahapan ini, *dataset* yang telah dibuat pada tahap sebelumnya digunakan sebagai masukan untuk algoritma klasifikasi. Pada penelitian ini akan digunakan empat jenis algoritma, yaitu *Naïve Bayes*, *K-Nearest Neighbors*, dan *Decision Tree C4.5*. Berikut ini desain proses yang digunakan pada penelitian ini.

Gambar 1. Desain Proses Perbandingan *Naïve Bayes*, *K-Nearest Neighbors* dan *Decision Tree C4.5*

3.5 Evaluation

Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*). Pada tahap ini pula harus ditentukan penggunaan hasil proses data mining. Berdasarkan hasil pengujian algoritma yang telah dilakukan maka algoritma *Decision Tree C4.5* memiliki nilai akurasi tertinggi dan nilai *Area Under Curve* (AUC) tertinggi dibandingkan dengan algoritma *Naïve Bayes* dan *K-Nearest Neighbors*.

3.5 Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus berupa laporan sederhana sehingga dapat digunakan oleh pengguna untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat. Untuk mengetahui proses dan hasil yang akan dihasilkan pada penelitian, berikut ini dijelaskan proses eksperimen dan pengujian agar diperoleh nilai akurasi, presisi, recall dan AUC dari algoritma yang digunakan pada penelitian ini.

3.6 Hasil Pengolahan Data

Setelah melalui proses pengolahan data dengan menggunakan aplikasi RapidMiner seperti yang diproses pada desain gambar, maka dapat diperoleh hasil akurasi, presisi, recall dan AUC dari masing-masing algoritma yang digunakan. Nilai akurasi dan AUC digunakan sebagai awalan untuk melakukan perbandingan pencarian algoritma terbaik yang akan dihasilkan. Perbandingan hasil perhitungan nilai akurasi dan AUC untuk metode algoritma *Naïve Bayes*, *K-Nearest Neighbors* dan *Decision Tree C4.5* dapat disimpulkan pada tabel 3 dibawah ini.

Tabel 3. Komparasi Nilai Akurasi, Presisi, Recall dan AUC

Algoritma	<i>Naïve Bayes</i>	K-NN	<i>Decision Tree C4.5</i>
Akurasi	85.15 %	84.10 %	93.83 %
AUC	0.927	0.816	0.978

Berdasarkan tabel 3 yang berisi komparasi nilai akurasi, presisi, recall dan AUC untuk algoritma *Naïve Bayes*, *K-Nearest Neighbors* (K-NN) dan *Decision Tree C4.5* menunjukkan bahwa algoritma *Decision Tree C4.5* mendapatkan nilai tertinggi dari sisi akurasi yaitu sebesar 93.83 %, selanjutnya diikuti dengan *Naïve Bayes* sebesar 85.15 % dan *K-Nearest Neighbors* sebesar 84.10 %. Dari ketiga algoritma tersebut dapat disimpulkan bahwa algoritma *Decision Tree C4.5* memperoleh nilai tertinggi dari nilai akurasinya.

Untuk nilai AUC pada tabel 4.4 menggambarkan bahwa hasil perbandingan nilai AUC yang menunjukkan bahwa algoritma *Decision Tree C4.5* mendapatkan nilai tertinggi yaitu 0.978, selanjutnya diikuti dengan *Naïve Bayes* dengan nilai AUC sebesar 0.927 dan *K-Nearest Neighbors* dengan nilai AUC sebesar 0.816. Dari ketiga metode yang dibandingkan algoritma *Decision Tree C4.5* mendapatkan nilai AUC lebih tinggi dibandingkan algoritma yang lain. Namun ketiga algoritma tersebut termasuk kedalam klasifikasi baik dikarenakan memiliki nilai diantara 0.80 – 0.90 (Gorunescu, 2011).

4. IMPLEMENTASI

4.1 Pengembangan Aplikasi Model Algoritma *Decision Tree C4.5*

Berdasarkan hasil eksperimen didapatkan model terbaik yaitu dari metode algoritma *Decision Tree C4.5*, yang kemudian dikembangkan menjadi sebuah aplikasi. Pengujian model aplikasi menggunakan salah satu *sample* pada dataset pendonor darah seperti tabel berikut ini:

Tabel 4. *Sample Data Set* Pendonor Darah

NIK	3275051006830020
Nama	Hermanto
Tempat Lahir	Jakarta
Tanggal Lahir	10-Juni-1983
Jenis Kelamin	Pria
No Hp	0812111699090
Status Donor	Ulang
Golongan Darah	O
Rhesus	+
HB	15
Berat Badan	73
Tensi Darah	120/80
Hasil/kesimpulan	Berhasil

Untuk tampilan aplikasi dapat dilihat pada gambar 2 berikut ini

Gambar 2. Tampilan Aplikasi Model Algoritma *Decision Tree C4.5* dan Hasil Pengujiannya

4.2 Uji Beda

Dari hasil pengujian proses algoritma sebelumnya, belum dapat menentukan secara mutlak dan secara langsung untuk algoritma mana yang terbaik. Oleh karena itu perlu dilakukan uji beda untuk mendapatkan algoritma terbaik dari ketiga algoritma yang telah diuji sebelumnya. Ketiga algoritma yang digunakan akan diuji dengan operator *T-Test* dari aplikasi

Rapidminer sehingga dapat dihasilkan penilaian tentang algoritma yang terbaik yang akan digunakan. Desain uji beda dapat dilihat pada gambar 3 dibawah ini.

Gambar 3. T-Test Algoritma

Setelah algoritma diproses dengan operator T-test dengan menggunakan aplikasi Rapidminer maka dihasilkan nilai seperti pada gambar 4 berikut ini.

A	B	C	D
	0.851 +/- 0.016	0.938 +/- 0.015	0.841 +/- 0.014
0.851 +/- 0.016		0.000	0.137
0.938 +/- 0.015			0.000
0.841 +/- 0.014			

Gambar 4. Hasil Uji Beda Algoritma

Dari gambar tabel hasil uji beda dapat terlihat bahwa algoritma *Decision Tree C4.5* memiliki nilai terbaik yaitu 0,000 dibandingkan dengan algoritma *K-Nearest Neighbors* dan *Naïve Bayes*. Dengan demikian dari hasil uji beda tersebut dapat ditarik kesimpulan bahwa *Decision Tree C4.5* adalah algoritma terbaik yang dapat digunakan pada penelitian ini. Hal ini ditunjukkan pula algoritma *Decision Tree C4.5* memiliki nilai probabilitas terbaik diantara ketiga algoritma yang ada karena memiliki nilai $\alpha \leq 0,05$ terkecil dibandingkan dengan kedua algoritma yang lainnya.

5. KESIMPULAN

Dalam penelitian ini telah dilakukan pengujian model dengan membandingkan tiga algoritma yaitu Algoritma *Naïve Bayes*, *K-Nearest Neighbors* (K-NN) dan *Decision Tree C4.5*. Dari hasil pengujian tersebut dapat ditarik kesimpulan sebagai bahan evaluasi sebagai berikut :

1. Algoritma *Decision Tree C4.5* memperoleh nilai akurasi tertinggi yaitu sebesar 93,83%, lebih baik dibandingkan dengan *Naïve Bayes* yang memperoleh nilai akurasi sebesar 85,15% dan *K-Nearest Neighbors* memperoleh nilai akurasi sebesar 84,10%
2. Untuk nilai Area Under Curve (AUC) yang memiliki nilai paling tinggi adalah algoritma *Decision Tree C4.5* dengan nilai AUC sebesar 0.978 kemudian diikuti dengan algoritma *Naïve Bayes* dengan nilai AUC sebesar 0.927 dan algoritma *K-Nearest Neighbors* memperoleh nilai AUC sebesar 0.816. Sehingga berdasarkan nilai AUC tertinggi algoritma *Decision Tree C4.5* adalah algoritma yang memiliki nilai AUC terbaik.
3. Berdasarkan uji beda yang menguji ketiga algoritma tersebut didapatkan hasil untuk algoritma *Decision Tree C4.5* dengan nilai 0.000 ketika diuji dengan operator T-Test pada aplikasi Rapidminer. Sehingga setelah melalui uji beda diperoleh kesimpulan bahwa algoritma *Decision Tree C4.5* adalah algoritma yang terbaik karena memiliki nilai $\alpha < 0,05$.

REFERENCES

- [1] PMI, "Sejarah PMI," *Palang Merah Indonesia*, 2011. [Online]. Available: <http://www.pmi.or.id/index.php/tentang-kami/sejarah->

- pmi.html?showall=1&limitstart=. [Accessed: 10-Nov-2019].
- [2] Kemenkumham RI, "Peraturan Pemerintah No. 7 Tahun 2011," 2011.
 - [3] Pusdatin Kemkes RI, "Pelayanan Darah Di Indonesia," *Kementerian Kesehatan RI*, 2014.
 - [4] WHO, "World Blood Donor Day 2008," *World Health Organization*, 2008. [Online]. Available: <https://www.who.int/worldblooddonorday/archives/2008/en/>.
 - [5] American Cancer Society, "Blood Transfusion and Donation," *American Cancer Society*, 2017. [Online]. Available: <https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/blood-transfusion-and-donation/donating-blood.html>. [Accessed: 07-Apr-2017].
 - [6] W. Boonyanusith and P. Jittamai, "Blood donor classification using neural network and decision tree techniques," *Lect. Notes Eng. Comput. Sci.*, vol. 1, pp. 499–503, 2012.
 - [7] M. Yunus, H. Dachlan, and P. Santoso, "SPK Pemilihan Calon Pendorong Darah Potensial Dengan Algoritma C4.5 Dan Fuzzy Tahani," *J. EECCIS*, 2014.
 - [8] B. D. Meilani and C. Utomo, "Aplikasi Data Mining Untuk Pola Permintaan Darah Di Udd (Unit Donor Darah) Pmi Kota Surabaya," *Pros. Semin. Nas. Manaj. Teknol. XXII*, pp. 1–7, 2015.
 - [9] W. E. Susanto and D. Riana, "Komparasi Algoritma Neural Network, K-Nearest Neighbor Dan Naive Baiyes Untuk Memprediksi Pendorong Darah Potensial," *J. Speed - Sentra Penelit. Eng. dan Edukasi*, vol. 8, no. 3, pp. 18–27, 2016.
 - [10] A. Kurniawan, "(Studi Kasus Pmi Semarang)," pp. 1–29, 2010.
 - [11] W. E. Susanto and C. Agustina, "Komparasi Akurasi Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Pendorong Darah Potensial Dengan Dataset Rfmte," *Semin. Nas. Ilmu Komput. (SNIK 2016)*, no. Snik, pp. 16–21, 2016.
 - [12] Y. Nurdiansyah, P. Pandunata, N. D. Prasetyo, A. Trihartono, F. G. Putrianti, and F. Wijayanto, "Application of blood donor routine detector using K-Nearest neighbors," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 293, no. 1, 2019.
 - [13] G. Suwardika, "Pengelompokan Dan Klasifikasi Pada Data Hepatitis Dengan Menggunakan Support Vector Machine (SVM), Classification And Regression Tree (Cart) Dan Regresi Logistik Biner," *J. Educ. Res. Eval.*, vol. 1, no. 3, p. 183, 2017.
 - [14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
 - [15] D. T. Larose, *Data Mining Methods and Models*. 2006.
 - [16] Priyadharsini.C and D. A. S. Thanamani, "An Overview of Knowledge Discovery Databaseand Data mining Techniques," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 1571–1578, 2014.
 - [17] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksieran," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9–13, 2016.
 - [18] Bustami, "Penerapan Algoritma Naive Bayes," *J. Inform.*, 2014.
 - [19] Aswendy, "Analisis data iklim indonesia menggunakan aplikasi weka dengan metode klasifikasi," *J. Teknol. Rekayasa*, vol. Volume 21, pp. 217–228, 2016.
 - [20] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *None*, vol. 13, no. 1, pp. 60–66, 2016.
 - [21] Y. Wu, K. Ianakiev, and V. Govindaraju, "Improved k-nearest neighbor classification," *Pattern Recognit.*, vol. 35, no. 10, pp. 2311–2318, 2002.
 - [22] B. Setywan, "VISUALISASI DASHBOARD POWER BI DAN PERAMALAN JUMLAH KASUS DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN METODE ARTIFICIAL NEURAL NETWORK," Institut Teknologi Sepuluh Nopember Surabaya, 2017.