

# ANALISA PERBANDINGAN ALGORITME NAÏVE BAYES DAN DECISION TREE PADA KLASIFIKASI DATA TRANSFUSI DARAH

**Rini Indrayani**

Fakultas Ilmu Komputer  
 Universitas Amikom Yogyakarta  
 Yogyakarta, Indonesia  
 rini.i@amikom.ac.id

## Abstrak

Donor darah merupakan proses pengambilan darah dari pendonor yang telah dinyatakan layak, ditinjau dari berbagai faktor. Penyakit yang diderita, usia, berat badan, tekanan darah, kadar hemoglobin, dan interval waktu donor merupakan aspek-aspek yang menjadi pertimbangan saat uji kelayakan. Karena pentingnya uji kelayakan tersebut, berbagai penelitian terkait uji kelayakan pendonor dilakukan menggunakan klasifikasi data mining dengan berbagai metode. Tantangan dari berbagai penelitian yang dilakukan adalah menemukan metode paling tepat dengan nilai akurasi dan presisi yang tinggi. Penelitian ini menggunakan 748 data set donor darah yang diproses menggunakan metode klasifikasi Naïve Bayes dan Decision Tree sebagai metode yang dipilih dengan berbagai aspek kelayakan donor sebagai atribut. Hasil analisa menunjukkan bahwa metode Decision Tree menunjukkan akurasi yang lebih tinggi sebesar 77.8075 % dibandingkan metode klasifikasi Naïve Bayes yang menunjukkan nilai akurasi sebesar 75.4011 %. Selain nilai-nilai tersebut, *Decision Tree* juga unggul secara visual dimana *Decision Tree* memiliki *output* sebuah model pohon yang menunjukkan relasi antar atribut. Sedangkan Naïve Bayes secara visual hanya menunjukkan plot penyebaran data yang sulit dipahami bagi pengguna awam WEKA.

Kata Kunci : donor darah, Naïve Bayes, *Decision Tree*.

## Abstract

*Blood donation is the process of taking blood from donors who have been declared eligible, in terms of various factors. The illness, age, body weight, blood pressure, hemoglobin level, and time interval are*

*aspects that are taken into consideration during the feasibility test. Due to the importance of the feasibility test, various studies related to the feasibility test of donors were carried out using classification of data mining by using various methods. The challenge of various studies is finding the most appropriate method with high accuracy and precision values. This study used 748 blood donor data sets that were processed using the Naïve Bayes classification method and Decision Tree as the chosen method with various aspects of donor feasibility as attributes. The results of the analysis show that the Decision Tree method shows a higher accuracy of 77.8075% compared to the Naïve Bayes classification method which shows an accuracy value of 75.4011%. In addition to these values, the Decision Tree is also superior visually where the Decision Tree has the output of a tree model that shows the relationship between attributes. Whereas Naïve Bayes only shows a plot of data distribution that is difficult to understand for anyone that new to WEKA.*

Keywords : blood donors, Naïve Bayes, *Decision Tree*.

## I. PENDAHULUAN

Data mining memiliki empat properti penting yang menjadi kekuatan dan tantangan dalam pemrosesannya yaitu *classic V's*: (i) *Volume*- dengan teknologi perangkat keras dan media penyimpanan yang semakin berkembang, jejak digital dari data-data dokumen teks, gambar, video, dan lain-lain dari waktu lampau hingga saat ini dapat tersedia. Analisis terkait data-data antar waktu tersebut akan membutuhkan proses analisis yang tidak singkat, namun data mining dapat mengatasi berbagai jenis data tersebut untuk dianalisis dengan mekanisme tertentu. (ii) *Variety*- jenis-jenis dan ragam bentuk

**Rini Indrayani**

Jurnal Ilmiah Teknologi Informasi Terapan  
 Volume V, No. 1, 15 Desember 2018

data, yaitu mulai dari data teks di jejaring sosial, data numerik dari rekaman stasiun geofisika, data citra satelit yang memiliki ketajaman tinggi, hingga data berformat audio dapat menjadi data untuk bahan analisis data mining. (iii) *Velocity*- proses analisis yang cepat menjadi tuntutan tersendiri seiring dengan kebutuhan bisnis yang semakin menuntut kecepatan dan efisiensi. (iv) dan “V” yang paling penting dari properti yang lain yaitu *Veracity*- dapat dimaknai sebagai “*trustworthiness*” dari data.

Kehandalan dari sebuah metode analisis pada proses data mining mempunyai peranan penting dalam setiap penelitian. Hasil analisis suatu proses data mining dapat memberikan informasi atau bahkan pengetahuan yang dapat berguna bagi kehidupan sehari-hari. Namun tidak semua mekanisme atau metode data mining dapat diimplementasikan terhadap semua jenis dan tipe data. Terdapat kondisi dimana suatu metode akan sangat baik jika digunakan untuk tipe data tertentu dibandingkan dengan metode lain atau jenis data lain.

Penelitian ini menggunakan data donor darah yang termasuk dalam kategori data medis. Data medis memiliki tingkat sensitivitas terkait “*veracity*” atau tingkat *trustworthiness* yang tinggi dimana data tersebut digunakan untuk menganalisa suatu hal yang berkaitan dengan sebuah pengetahuan untuk kepentingan hidup manusia. Oleh karena itu dibutuhkan ketelitian dan akurasi yang baik untuk dapat menyimpulkan hasil analisis yang menggunakan data-data medis seperti data donor darah.

Donor darah adalah proses pengambilan darah dari seseorang secara sukarela untuk disimpan dalam bank untuk kemudian digunakan dalam transfusi darah. Setiap calon pendonor darah terlebih dahulu melakukan berbagai tes kelayakan untuk mengetahui kualitas darah dan status kelayakan sebelum melakukan donor darah. Berbagai aspek seperti penyakit yang diderita, usia, berat badan, tekanan darah, dan kadar hemoglobin menjadi pertimbangan penting. Bahkan terdapat ketetapan interval waktu donor yaitu minimal 12 minggu atau 3 bulan sejak donor darah sebelumnya (maksimal 5 kali dalam 1 tahun). Upaya untuk memudahkan penilaian terkait kelayakan calon pendonor terus dilakukan melalui berbagai penelitian dengan memanfaatkan bidang ilmu data mining khususnya teknik klasifikasi. *Data mining* merupakan suatu cara dalam penggalian informasi dari sejumlah data yang biasanya tersimpan

dalam repositori dengan menggunakan teknologi pengenalan pola, statistik dan teknik matematika.

Penelitian mengenai potensi donor darah telah dilakukan oleh Mahmud menggunakan algoritma C4.5 dan Fuzy Tahani dengan menggunakan variable golongan darah, jenis kelamin, usia, berat badan, tekanan darah, kadar HB (*hemoglobin*), dan kadar HTC (*hematocit*). Adapun target variabel outputnya adalah atribut Donor Lagi yang memuat nilai ‘ya’ atau ‘tidak’ berdasarkan nilai ambang batas jumlah donor, bernilai ‘ya’ apabila jumlah donor  $\geq 2$  dan bernilai ‘tidak’ apabila jumlah donor  $< 2$ . Pada penelitian ini menghasilkan akurasi rata-rata sebesar 89,64% dengan menggunakan dataset yang berasal dari PMI Kota Malang.

Sandi melakukan analisis untuk menentukan kelayakan calon pendonor darah menggunakan algoritma Naïve Bayes dengan atribut nama, tekanan darah, HB (Hemoglobin), denyut nadi dan suhu tubuh serta kelayakan donor darah. Pada pengujian ini, sebanyak 20% dari jumlah data sampel secara bergantian dijadikan data uji sebanyak 5 kali terhadap 80% data sampel lainnya yang dijadikan data *training*. Nilai akurasi diperoleh dari rata-rata nilai akurasi dari 5 kali pengujian tersebut.

Erraguntla melakukan penelitian terkait keamanan proses donor darah di Amerika Serikat. Penelitian dimulai dengan mengumpulkan data terkait pengorganisasian data, informasi pendonor, dan atribut yang berpotensi mempengaruhi reaksi aktivitas donor. Penelitian ini menggunakan *Multivariate Logistic Regression* (LR) sebagai teknik utama untuk menganalisis faktor-faktor yang mempengaruhi reaksi donor. Teknik yang selalu digunakan secara umum untuk analisis data mining adalah *Decision Tree*, *Support Vector Machine* (SVM), dan Bayes, namun LR lebih sering digunakan dalam ranah medis [sitasi Erraguntla]. Berbagai jenis data reaksi diolah menjadi informasi yang dikaitkan dengan atribut data kesehatan pendonor. Setelah itu dilakukan evaluasi untuk meningkatkan kualitas kategori pendonor untuk mengurangi jumlah reaksi donor yang tidak diinginkan.

Penelitian lainnya juga dilakukan oleh Darwiche menggunakan algoritma MLP dan SVM. Dataset yang digunakan berasal dari *UCI Machine Learning* adapun variabel yang digunakan adalah kebaruan variabel (*Recency*), frekwensi (*Frequency*), jumlah darah yang diambil (*Monetary*), lama menjadi

pedonor(*Time*), dan variabel targetnya adalah menyumbang atau tidak menyumbang (*Donation or non donation*) berdasarkan apakah pada bulan maret 2007 pendonor tersebut melakukan donor atau tidak. Dari penelitian ini dihasilkan *sensitivity* 65.8% dan *specificity* 78.2% dari 600 data *training* dan 148 data *testing* .

Dari berbagai penelitian yang telah dilakukan, belum terdapat penelitian yang menguji dataset donor darah dari Blood Transfusion Service Center menggunakan perbandingan algoritma Naïve Bayes dan Decision Tree dengan metode validasi 10-fold cross validation. Oleh karena itu paper ini membahas tentang analisis menggunakan metode Naïve Bayes dan Decision Tree untuk mengetahui metode terbaik untuk klasifikasi dataset donor darah.

## II. MATERIAL AND METHOD

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri dapat berupa aturan jika-maka (*if-then*), berupa pohon keputusan (*decision tree*) dan jaringan saraf tiruan (*neural network*).

Teknik klasifikasi dibagi ke dalam beberapa jenis berdasarkan model pendekatan yang dilakukan seperti model linier, model statistika, dan model aturan (*rules*). Proses klasifikasi pada bentuk klasifikasi menggunakan aturan(*rules*), dapat dilakukan menggunakan sejumlah aturan yang berbentuk jika-maka. Hasil klasifikasi merupakan sebuah kesimpulan dan bentuk data yang dapat dijadikan data masukan untuk jenis *rules*. Data tersebut adalah data kategorial yang bertipe nominal. Bentuk klasifikasi selanjutnya adalah klasifikasi dengan bentuk data berupa dokumen, dimana proses klasifikasinya menggunakan kemunculan kata-kata pada dokumen tersebut untuk mengidentifikasi jenis atau kelompok dokumen tersebut. Bentuk klasifikasi yang terakhir adalah bentuk model linier. Pendekatan model linier digunakan untuk mengelompokan data berjenis numeris. Proses pengelompokan data dilakukan dengan mengimplementasikan regresi untuk setiap kelasnya.

Penelitian ini menggunakan data set transfusi darah dari Blood Transfusion Service Center Hsin-

Chu City di Taiwan yang dipublikasikan pada portal UCI Machine Learning Repository. Data set ini menggunakan 5 atribut data yaitu *Recency*, *Frequency*, *Monetary*, *Time*, dan *Class*. *Recency* merupakan interval jumlah bulan sejak terakhir kali melakukan donasi. *Frequency* merupakan jumlah total donasi. *Monetary* adalah total darah yang disumbangkan dalam ukuran cc. Atribut *Time* merupakan keterangan jumlah bulan sejak donasi pertama, sedangkan atribut *Class* merupakan keterangan mengenai status donasi yaitu apakah pendonor melakukan donasi pada bulan Maret 2007. Atribut *Class* menggunakan bilangan biner dimana angka 1 merupakan representasi dari status “ya” dan angka 0 merupakan representasi dari status “tidak”. 5 sampel data yang diambil dari 748 data set yang digunakan ditunjukkan pada tabel 1.

Data yang digunakan pada penelitian ini tidak memiliki *missing values* dan tidak memerlukan tahap *preprocessing* sebelum dilakukan proses data mining lanjutan. Sehingga tahapan proses data mining pada penelitian ini adalah melakukan proses analisis klasifikasi terhadap data untuk setiap metodenya yang kemudian hasil keluaran dari masing-masing metodenya dianalisis untuk kemudian disimpulkan.

Data set yang digunakan diproses menggunakan tools WEKA Versi 3.8 dengan 2 algoritme klasifikasi yaitu Naïve Bayes dan Decision Tree. Kedua algoritme tersebut masing-masing memiliki kelebihan dan kekurangan. Penelitian ini menguji kinerja kedua algoritme tersebut menggunakan 748 data set transfusi darah. Beberapa sampel data transfusi darah disajikan dalam tabel 1.

Tabel 1. Sampel Data Transfusi Darah

Recency	Frequency	Monetary	Time	Class
2	50	12500	98	1
0	13	3250	28	1
1	16	4000	35	1
2	20	5000	45	1
1	24	6000	77	0

### I.1 WEKA Environment

Weka adalah aplikasi *open source* yang digunakan untuk keperluan *data mining* berbasis Java. WEKA memiliki berbagai *source code* implementasi dari berbagai algoritme dari *machine learning* yang dapat digunakan untuk melakukan generalisasi pada data set yang digunakan sebagai *input*. Aplikasi WEKA merupakan perangkat lunak yang canggih dan memiliki akurasi tinggi dalam melakukan tugas komputasi, namun pengetahuan *user* tetap menjadi hal yang paling penting untuk melakukan analisa terhadap hasil generalisasi yang dilakukan oleh WEKA. Paduan antara kemampuan komputasi WEKA yang akurat dan pengetahuan *user* yang mumpuni dapat menghasilkan pengetahuan baru untuk kontribusi dunia penelitian. *User Interface* WEKA ditunjukkan oleh gambar 1.



Gambar 1. WEKA Environment

Pada gambar 1 tombol di atas dapat digunakan untuk menjalankan aplikasi:

1. Explorer digunakan untuk melakukan eksplorasi terhadap data yang digunakan.
2. Experimenter merupakan tombol untuk fungsi pengujian statistik pada WEKA
3. KnowledgeFlow merupakan tombol untuk fungsi pengetahuan pendukung
4. Simple CLI merupakan *user interface* berbasis *command-line* sebagai media bagi user untuk mengeksekusi perintah WEKA di lingkungan sistem operasi yang tidak menyediakan *user interface* WEKA secara langsung.

### I.2 Algoritma Naïve Bayes Classifier

Algoritme Naïve Bayes Classifier (NBC) merupakan algoritme model probabilistik grafis berdasarkan teorema Bayes dengan asumsi bahwa nilai atribut pada kelas tertentu independen dari nilai-nilai atribut lainnya. Kelebihan NBC adalah sederhana tetapi memiliki akurasi yang tinggi. NBC menggunakan teori probabilitas sebagai dasar teori. Ada dua tahap pada proses klasifikasi, tahap pertama adalah pelatihan terhadap himpunan contoh (*training example*). Sedangkan tahap kedua adalah proses klasifikasi data yang belum diketahui kategorinya. Teorema NBC digambarkan pada persamaan 1.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad \dots [1]$$

Dimana

$P(c/x)$  : adalah probabilitas posterior dari *class* (target) yang diberikan prediktor (*attribute*).

$P(c)$  : adalah probabilitas kejadian awal dari *class*.  $P(x/c)$  adalah probabilitas dari kemungkinan situasi yang terjadi atas *class*.

$P(x)$  : adalah probabilitas atas kejadian awal prediktor. Dalam notasi ini  $P(c/x)$  berarti peluang kejadian  $c$  bila  $x$  terjadi, dan

$P(x/c)$  : peluang kejadian  $x$  bila  $c$  terjadi.

Kaitan antara *Naive Bayes* dengan klasifikasi, korelasi hipotesis, dan bukti klasifikasi adalah hipotesis dalam teorema *Bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi sedangkan bukti merupakan fitur-fitur yang menjadikan masukan dalam model klasifikasi. Jika  $X$  adalah vektor masukan yang berisi fitur dan  $C$  adalah label kelas.

*Bayes* pada umumnya mudah dihitung untuk fitur bertipe kategori, namun untuk fitur dengan tipe numerik (kontinu) membutuhkan perlakuan khusus sebelum diproses yaitu:

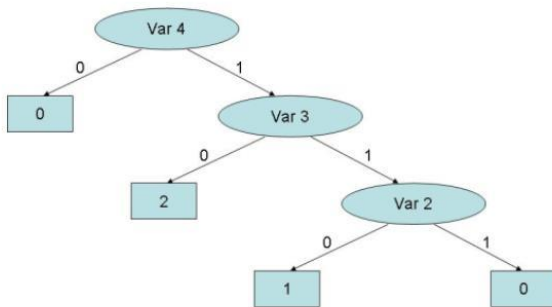
1. Melakukan pengelompokan pada nilai-nilai dengan interval tertentu, dengan mentransformasikan nilai numerik ke dalam bentuk ordinal berdasarkan kategori rentang nilai.

2. Merubah bentuk distribusi peluang ke dalam kelompok-kelompok tertentu untuk memperkirakan parameter distribusi dengan data pelatihan menggunakan konsep transformasi data.

### I.3 Algoritme Decision Tree

Pendekatan “*divide-and-conquer*” untuk proses pembelajaran mesin dari suatu himpunan yang berisi sejumlah data yang saling bebas secara alami membentuk sebuah representasi yang dinamakan pohon keputusan (*Decision Tree*).

*Decision tree* adalah algoritme yang populer digunakan untuk masalah pengklasifikasian. Algoritme *Decision Tree* secara visual mirip sebuah pohon dengan cabang-cabang dan ranting-rantingnya. Konsep *entropy* digunakan untuk menentukan pembentukan cabang-cabang dan ranting-ranting dari setiap atribut. Semakin rendah nilai *entropy* maka semakin kuat peluang atribut tersebut untuk menjadi akar.



Gambar 2. Contoh Bentuk Umum Decision tree

Model visual dari *Decision tree* dapat dilihat pada Gambar 2 dimana bentuk elips adalah variabel fitur/atribut bentuk sangkar adalah variabel output dari keputusannya.

Pada decision tree terdapat 3 jenis node, yaitu:

- a. Root Node, merupakan node paling atas, pada node ini tidak ada input sehingga memungkinkan keadaan tidak mempunyai output atau mempunyai output lebih dari satu.
- b. Internal Node, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.

- c. Leaf node atau terminal node, merupakan node akhir dimana hanya terdapat satu input dan tidak mempunyai output.

Rumus yang digunakan untuk menghitung *Entropy* sampel S adalah sebagai berikut [12]:

$$E(S) = \sum_{i=1}^c - p_i \log_2 p_i \quad \dots[2]$$

S adalah himpunan (dataset) kasus. *c* adalah banyaknya partisi S dan *p<sub>i</sub>* adalah probabilitas solusi positif (mendukung) ataupun negatif (tidak mendukung) pada data sample untuk kriteria tertentu.

Salah satu algoritma pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). Prosedur algoritma ID3 memiliki input berupa sampel training, label training, dan atribut. ID3 merupakan metode terdahulu dari algoritma C4.5. Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*.

*Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan sejumlah *field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma C4.5 adalah algoritma hasil pengembangan dari algoritma ID3. Beberapa perbaikan dari algoritma ID3 menjadi algoritma C4.5 yaitu [8]:

1. Dapat mengatasi *missing value*
2. Dapat mengatasi data kontinu
3. Pruning
4. Adanya aturan

Penelitian ini menggunakan algoritme pengembangan dari ID3 yaitu C4.5 memanfaatkan perangkat lunak *open source* WEKA. Aplikasi *open source* WEKA menggunakan nama J48 untuk algoritme C4.5.

### III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 748 data set yang diproses menggunakan algoritma Naïve Bayes dan Decision Tree. Metode validasi yang digunakan adalah metode 10-fold *cross validation*. *Cross*

validation merupakan metode statistik yang mengevaluasi dan membandingkan algoritme pembelajaran dengan cara membagi data set menjadi dua segmen: segmen pertama digunakan untuk proses belajar atau model latihan sedangkan segmen lainnya digunakan untuk validasi model.

WEKA Environment yang digunakan dalam penelitian ini berhasil memproses kedua algoritme menggunakan data set transfusi darah. Hasil dari proses kedua algoritme menggunakan WEKA disajikan dalam tabel 2.

Tabel 2. Perbandingan Kinerja Algoritme

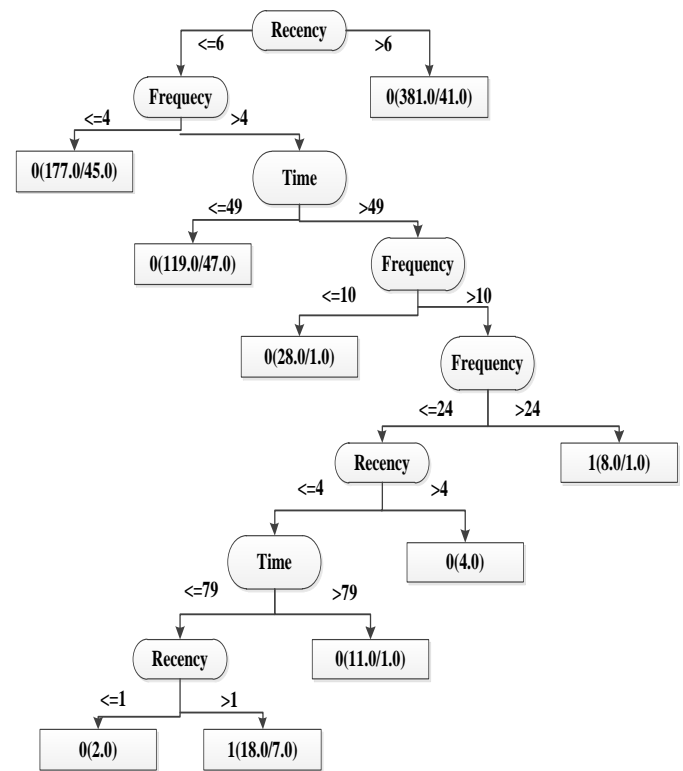
		Naïve Bayes	C4.5
Correctly Instances	Classified	75.4011 %	77.8075 %
Incorrectly Instances	Classified	24.5989 %	22.1925 %
	Kappa statistic	0.1593	0.3424
	Recall (Sensitivity)	0.754	0.778
	Precision	0.710	0.764

Pada Tabel II, akurasi algoritme NBC adalah 75.4011%, sedangkan C4.5 memiliki akurasi yang lebih tinggi yaitu sebesar 77.8075%. Nilai Kappa statistic yang dimiliki oleh C4.5 juga lebih tinggi dari NBC. Kappa adalah ukuran koreksi antara klasifikasi dan class yang benar. Precision merupakan nilai perbandingan antara kasus dengan hasil positif yang benar. Recall atau sensitivity merupakan proporsi kasus positif yang diidentifikasi dengan benar. Secara keseluruhan, algoritma Decision Tree menunjukkan kinerja yang lebih baik dari Naïve Bayes untuk data set donor darah yang digunakan baik dari segi akurasi maupun presisi.

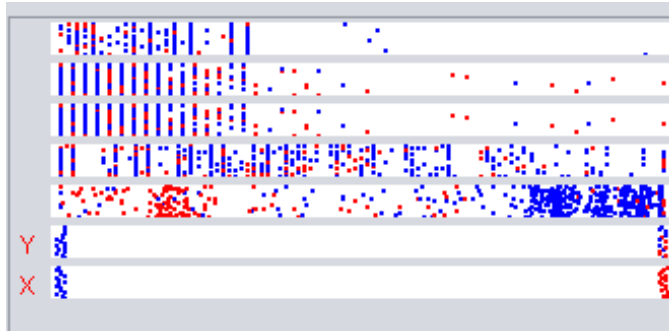
Selain unggul dari segi akurasi, Decision Tree juga memiliki kelebihan dari segi visual output. Gambar 3 menunjukkan visual hasil klasifikasi J48 sebagai salah satu fasilitas yang disediakan oleh WEKA. Gambar 3 menunjukkan bahwa atribut Monetary yaitu jumlah volume darah yang disumbangkan dalam satuan cc tidak memiliki relasi atau pengaruh terhadap nilai atribut class. Sementara node pertama atau atribut

paling berpengaruh sesuai dengan model pohon yang digambarkan adalah atribut Recency yaitu interval jumlah bulan sejak terakhir kali melakukan donasi.

Pada visualisasi classifier error metode Naïve Bayes di WEKA, terlihat plot penyebaran data yang ditandai dengan titik-titik biru dan merah seperti pada gambar 4. Namun bagi pengguna WEKA yang tidak memiliki pengetahuan membaca plot penyebaran data akan mengalami kesulitan dalam menerjemahkan plot. Berbeda dengan Decision Tree yang memiliki visual lebih baik, pengguna akan lebih kesulitan menggunakan algoritme Naïve Bayes secara visual.



Gambar 3. Model Decision Tree Dataset Donor Darah



**Gambar 4. Classifier Error Dataset Donor Darah metode Naïve Bayes**

#### IV. KESIMPULAN

Perbandingan kinerja algoritma Naïve Bayes dan Decision Tree menggunakan WEKA 3.8 dapat dilihat dari beberapa nilai yang dihasilkan antara lain *correctly classified instance* (akurasi), *incorrectly classified*, *kappa statistic*, *recall*, dan *precision*. Keseluruhan nilai yang menggambarkan kinerja metode klasifikasi tersebut menunjukkan bahwa algoritme *Decision Tree* khususnya C4.5 menunjukkan performa yang lebih baik dengan akurasi sebesar 77.8075% dibandingkan metode Naïve Bayes dengan akurasi sebesar 75.4011%. Selain nilai-nilai tersebut, *Decision Tree* juga unggul secara visual dimana *Decision Tree* memiliki *output* sebuah model pohon yang menunjukkan relasi antar atribut. Sedangkan Naïve Bayes secara visual hanya menunjukkan plot penyebaran data yang sulit dipahami bagi pengguna awam WEKA. Oleh karena itu, untuk klasifikasi dataset donor darah dari Blood Transfusion Service Center Hsin-Chu City di Taiwan dianjurkan menggunakan algoritme *Decision Tree*.

#### REFERENCES

- P. Jiang, F. Liu, J. Wang, and Y. Song, "Cuckoo search-designated fractal interpolation functions with winner combination for estimating missing values in time series," *Appl. Math. Model.* 40, vol. 40, pp. 9692–9718, 2016.
- Tim Penyusun PMI. *Pedoman Pelayanan Transfusi Darah Modul I*. Jakarta: Unit Transfusi Darah PMI Pusat. 2001.

- Daniel T Larose, *Data Mining Methods and Models*. Hoboken, New Jersey, United State of America: John Wiley & Sons, Inc., 2006.
- F. Tahani, M. Yunus, H. S. Dahlan, and B. Santoso, "SPK Pemilihan Calon Pendoron Darah Potensial," vol. 8, no. 1, pp. 47–54, 2014.
- S. F. Rodiyansyah, P. Studi, T. Informatika, F. Teknik, and U. Majalengka, "Naïve Bayes Classification Untuk Penentuan Kelayakan Donor Darah," pp. 156–159, 2016.
- M. Erraguntla, P. Tomasulo, K. Land, B. Whitaker, R. Mayer, and S. Khaire, "Data Mining to Improve Safety of Blood Donation Process," *2014 47th Hawaii Int. Conf. Syst. Sci.*, pp. 789–795, 2014.
- Darwiche, M., Feuilloy, M., Bousaleh, G., & Schang, D. (2010). *Prediction of blood transfusion donation. IEEE*, 978. 2010.
- M. A. H. Ian H. Witten, Eibe Frank, *DATA MINING Practical Machine LEarning Tools and Techniques*, 3rd ed. Burlington: Morgan Kaufmann, 2011.
- E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, and I. H. Witten, "Chapter 62 WEKA A Machine Learning Workbench for Data Mining", 2009.
- R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2<sup>nd</sup> Edition)*. Wiley-Interscience, 2000.
- Rish, I., An empirical study of The Naive Bayes Classifier, *International Joint Conference on Artificial Intelligence*, California, 41-46. 2006.
- Y. G. Jung, K. T. Kim, B. Lee, and H. Y. Youn, "Enhanced Naive Bayes Classifier for Real-time Sentiment Analysis with SparkR," pp. 141–146, 2016.
- Z. Lijuan, "The Application of Decision Tree C4 . 5 Algorithm to Soil Quality Grade Forecasting Model," vol. 1, no. 2, pp. 552–555, 2016.
- D. M. W. Powers, "Evaluation: From Precision, Recall And F-Measure To ROC, Informedness, Markedness & Correlation," vol. 2, no. 1, pp. 37–63, 2011.