

# NAÏVE BAYES CLASSIFICATION UNTUK PENENTUAN KELAYAKAN DONOR DARAH

**Sandi Fajar Rodiyansyah**

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Majalengka  
email : galuh29@gmail.com

## Abstrak

Sebelum seseorang melakukan donor darah biasanya tenaga kesehatan melakukan pemeriksaan usia, berat badan, tekanan darah, HB, denyut nadi dan suhu tubuh. Berdasarkan kriteria-kriteria tersebut tenaga kesehatan menentukan apakah calon pendonor layak atau tidak melakukan donor darah. Pada penelitian ini, dilakukan implementasi naïve bayes classification untuk penentuan kelayakan donor darah tersebut. Data training yang digunakan pada penelitian ini sejumlah 25 sampel yang dibagi menjadi 2 class yaitu label “layak” sebanyak 60% dan “tidak layak” sebanyak 40%. Setelah dilakukan pengujian, penentuan kelayakan donor darah dengan menggunakan naïve bayes classification memperoleh akurasi sebesar 88%.

**Kata Kunci :** donor darah, naïve bayes classifier

## 1. PENDAHULUAN

Donor darah merupakan proses pengambilan darah dari seseorang secara sukarela untuk disimpan di bank darah yang kemudian dapat digunakan untuk transfusi darah bagi pasien yang membutuhkan. Selama ini, Palang Merah Indonesia (PMI) masih menggunakan metode penentuan kelayakan calon pendonor dengan cara konvensional. Oleh karena itu, penelitian ini mencoba mengimplementasikan algoritma naïve bayes classification untuk menentukan seorang calon pendonor termasuk layak atau tidak layak melakukan donor darah.

Algoritma naïve bayes classification digunakan atas dasar bahwa algoritma ini adalah algoritma yang sederhana namun memiliki akurasi yang tinggi (Rish, 2006).

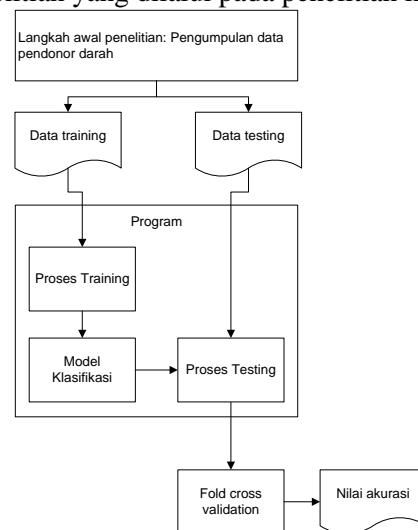
## 2. METODE PENELITIAN

Penelitian diawali dengan mengumpulkan data calon pendonor darah. Data ini digunakan untuk proses *training* dan *testing* program yang akan dibuat. Adapun data yang dikumpulkan antara lain data nama, usia, berat badan, tekanan darah, HB, denyut nadi dan suhu tubuh serta kelayakan donor darah.

Setelah data terkumpul selanjutnya dilakukan implementasi algoritma naïve bayes classification dengan perangkat lunak Rapid Miner. Pada tahap ini, seluruh tahapan

implementasi dan pengujian algoritma naïve bayes classification diimplementasikan.

Setelah implementasi algoritma naïve bayes selesai, kemudian dilakukan pengujian dengan membagi data yang terkumpul menjadi dua bagian yaitu *data training* dan *data testing*. *Data training* digunakan untuk memberikan pengetahuan kepada program sehingga terbentuk model klasifikasi. Setelah model klasifikasi terbentuk maka dilakukan pengujian model tersebut dengan menggunakan *data testing* untuk mengetahui tingkat akurasi proses klasifikasi tersebut. Gambar 1 merupakan ilustrasi metode penelitian yang dilalui pada penelitian ini.



Gambar 1 : Alur Penelitian

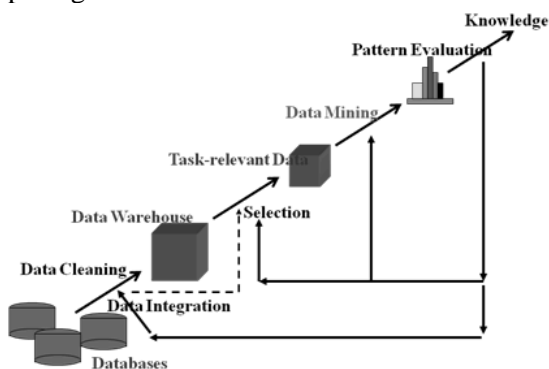
### 3. DONOR DARAH

Donor darah adalah proses pengambilan darah dari seseorang secara sukarela untuk disimpan dalam bank darah di Palang Merah Indonesia untuk kemudian digunakan dalam transfusi darah. (PMI, 2001).

Pada dasarnya melakukan aktivitas donor darah membawa simbiosis mutualisme, sebab setiap tetes darah yang disumbangkan tidak hanya dapat memberikan kesempatan hidup bagi yang menerima tetapi juga memberikan manfaat kesehatan bagi pendonornya.

### 4. DATA MINING

*Data mining* merupakan suatu metode untuk menemukan pengetahuan dalam suatu tumpukan data yang cukup besar. *Data mining* adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru dan bermanfaat dan akhirnya dapat ditemukan suatu corak atau pola dalam data tersebut. (Han dan Kamber, 2006). *Data mining* adalah bagian integral dari *knowledge discovery database* (KDD) yang merupakan proses keseluruhan mengubah data mentah menjadi pola-pola data yang menarik yang merupakan informasi yang dibutuhkan oleh pengguna sebagai pengetahuan, seperti yang ditunjukkan pada gambar 2.



Gambar 2 : Proses penemuan pengetahuan dalam database/KDD (Han & Kamber, 2006)

Terdapat empat tugas utama data mining yang terlihat pada gambar 3 yaitu :

#### 1. *Predictive modelling*

*Predictive modelling* digunakan untuk membangun sebuah model untuk target variabel sebagai fungsi dari *explanatory* variabel. *Explanatory* variabel dalam hal ini merupakan semua atribut yang digunakan untuk melakukan prediksi, sedangkan

variable target merupakan atribut yang akan diprediksi nilainya. *Predictive modelling* dibagi menjadi dua tipe yaitu : *Classification* yang digunakan untuk memprediksi nilai dari target variabel yang diskrit dan regresi yang digunakan untuk memprediksi nilai dari target variabel yang kontinyu.

#### 2. *Association analysis*

*Association analysis* adalah penemuan *association rule* yang menunjukkan pola-pola yang sering muncul dalam data. Terdapat nilai *support* dan *confidence* yang dapat menunjukkan seberapa besar suatu rule dapat dipercaya. *Support* adalah ukuran dimana seberapa besar tingkat dominasi suatu item atau itemset terhadap keseluruhan transaksi. Sedangkan *confidence* adalah ukuran yang menunjukkan hubungan antara dua item secara conditional.

*Association analysis* digunakan untuk menemukan aturan-aturan asosiasi yang memperlihatkan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

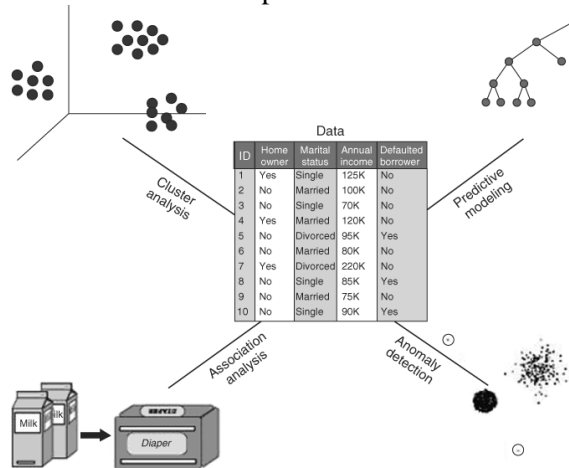
#### 3. *Cluster analysis*

Tidak seperti klasifikasi yang menganalisa kelas data objek yang mengandung label. *Clustering* digunakan untuk menganalisa objek dari data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data *training* karena belum diketahui sebelumnya. *Clustering* merupakan proses mengelompokkan sekumpulan objek yang sangat mirip.

#### 4. *Anomaly detection/outlier mining*

Sebuah database dapat mengandung data objek yang tidak sesuai atau menyimpang dari model data. Data objek ini disebut outlier. Banyak metode data mining yang menghilangkan outlier ini. Padahal, pada beberapa aplikasi seperti *fraud detection*, kejadian yang jarang terjadi tersebut justru lebih menarik untuk dianalisa daripada kejadian yang sering terjadi. Analisa dari outlier data tersebut disebut sebagai outlier mining. Outlier mining juga sering disebut dengan *anomaly detection* yang merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. Anomaly dapat dideteksi dengan

menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.



Gambar 3 Empat tugas utama data mining (Tan, dkk, 2006)

### 5. NAÏVE BAYES CLASSIFIER

Salah satu metode klasifikasi yang dapat digunakan adalah metode *naive bayes classifier* (NBC). Kelebihan NBC adalah sederhana tetapi memiliki akurasi yang tinggi (Rish, 2006). 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.

Pada *naive bayes classification* setiap data direpresentasikan dalam pasangan atribut  $\langle a_1, a_2, a_3, \dots, a_n \rangle$  dimana  $a_1$  adalah atribut pertama dan  $a_2$  adalah atribut kedua dan seterusnya, sedangkan  $V$  adalah himpunan kelas. (Manning, dkk. 2008). Pada saat klasifikasi, metode ini akan menghasilkan kategori/kelas yang paling tinggi probabilitasnya ( $V_{MAP}$ ) dengan memasukan atribut  $\langle a_1, a_2, a_3, \dots, a_n \rangle$ . Adapun rumus  $V_{MAP}$  dapat dilihat pada persamaan 1.

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad 1$$

Sementara itu, teorema Bayes menyatakan

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad 2$$

Dengan menggunakan teorema Bayes ini, maka persamaan 2 dapat ditulis menjadi

$$= \operatorname{argmax}_{v_j \in V} \frac{P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad 3$$

$P(a_1, a_2, a_3, \dots, a_n)$  nilainya konstan untuk semua  $v_j$  sehingga persamaan 3 dapat dinyatakan juga menjadi persamaan 4.

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, a_3, \dots, a_n | v_j) \quad 4$$

$$\times P(v_j)$$

*Naive bayes classifier* menyederhanakan hal ini dengan mengasumsikan bahwa didalam setiap kategori, setiap atribut bebas bersyarat satu sama lain. (Tan, dkk, 2006). Dengan kata lain

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad 5$$

Kemudian apabila persamaan 5 disubstitusikan ke persamaan 4, maka akan menghasilkan

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j) \times \prod_i P(a_i | v_j) \quad 6$$

### 6. PEMBAHASAN

Data yang dikumpulkan memiliki atribut nama, tekanan darah, HB, denyut nadi dan suhu tubuh serta kelayakan donor darah. Sebagian data sampel dari penelitian ini dapat dilihat pada tabel 1

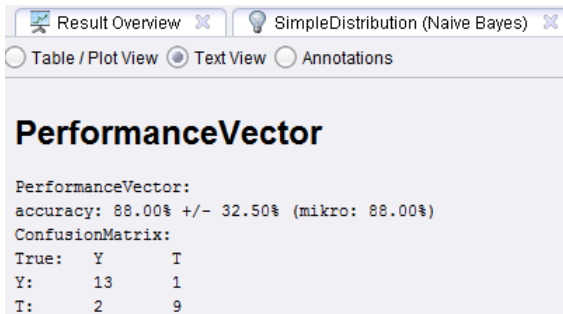
Tabel 1 : Data Sampel

ID	Usia	BB	TD	HB	Nadi	Suhu	Class
1	30	76	128/87	13	62	35,5	L
2	34	58	91/82	15	76	36,1	TL
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
25	17	57	90/88	11	59	36,3	L

Setelah *naive bayes classifier* selesai diimplementasikan selanjutnya dilakukan pengujian program tersebut. Berdasarkan gambar 1, pengujian program dilakukan dengan menggunakan teknik *fold cross validation*.

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.

Dengan demikian, setiap data sampel akan menjadi data *training* dan data *testing* secara bergantian. Gambar 4 merupakan hasil implementasi *naive bayes classifier* pada Rapid Miner.



Gambar 4 : Hasil program

Dari hasil 5 fold cross validation yang telah dilakukan pada perangkat lunak Rapid Miner diperoleh hasil akurasi sebesar 88%.

## 7. KESIMPULAN

Naïve bayes classifier dalam melakukan pengklasifikasian data calon pendonor darah untuk menentukan kelayakan donor darah merupakan solusi yang efektif. Karena sebelum melakukan pengklasifikasian data calon pendonor darah sistem melakukan *learning* dengan menggunakan *data training*. Selanjutnya, setelah sistem melakukan *learning* sistem kemudian dapat mengklasifikasikan data calon pendonor darah secara otomatis. Dengan demikian, proses penentuan kelayakan donor darah dapat dilakukan secara cepat.

Diperlukan data training yang valid dan akurat agar memastikan sistem dapat bekerja secara akurat.

## 8. REFERENSI

Han, J., & Kamber, M., 2006, *Data Mining: Concepts and Techniques 2e*, Morgan Kaufmann Publishers, San Francisco,

Rish, I., 2006, An empirical study of The Naive Bayes Classifier, *International Joint Conference on Artificial Intelligence*, California, 41-46.

Tan, P. N., Steinbach, M., & Kumar, V., 2006, *Introduction to Data Mining*, Pearson Education, Boston.

Tim Penyusun PMI. 2001. *Pedoman Pelayanan Transfusi Darah Modul I*. Jakarta: Unit Transfusi Darah PMI Pusat.