

**KLASIFIKASI CALON PENDONOR DARAH  
MENGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**(STUDI KASUS : Calon Pendonor Darah di Kota Semarang)**



---

**SKRIPSI**

---

**Disusun Oleh :**

**DHIMAS BAYUSUSETYO**

**NIM. 24010212130081**

**DEPARTEMEN STATISTIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**SEMARANG**

**2017**

**KLASIFIKASI CALON PENDONOR DARAH  
MENGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**(STUDI KASUS : Calon Pendonor Darah di Kota Semarang)**

**Disusun Oleh :  
DHIMAS BAYUSUSETYO  
NIM. 24010212130081**

Tugas Akhir sebagai salah satu syarat untuk memperoleh  
gelar Sarjana Sains pada Departemen Statistika

**DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2017**

## HALAMAN PENGESAHAN I

Judul : Klasifikasi Calon Pendoror Darah Menggunakan Metode *Naive Bayes Classifier*

(Studi Kasus: Calon Pendoror Darah di Kota Semarang)

Nama : DhimasBayususetyo

NIM : 24010212130081

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 27 Maret 2017 dan dinyatakan lulus pada tanggal 27 Maret 2017.

Semarang, 30 Maret 2017

**Mengetahui,**

**Ketua Departemen Statistika**

**Fakultas Sains dan Matematika Undip**

I



**Dr. Tarno, M.Si.**

**NIP. 196307061991021001**

**Panitia Penguji Ujian Tugas Akhir**

**Ketua,**



**Dr. Rukun Santoso, M.Si.**

**NIP. 196502251992011001**

## HALAMAN PENGESAHAN II

Judul : Klasifikasi Calon Pendoror Darah Menggunakan Metode *Naïve Bayes Classifier*

(Studi Kasus: Calon Pendoror Darah di Kota Semarang)

Nama : DhimasBayususetyo

NIM : 24010212130081

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 27 Maret 2017.

Semarang, 30 Maret 2017

Pembimbing I

Pembimbing II

Dr. Rukun Santoso, M.Si.  
NIP. 196502251992011001

Dr. Tarno, M.Si.  
NIP. 196307061991021001

## KATA PENGANTAR

Puji Syukur penulis ucapkan kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan Tugas Akhir dengan judul “Klasifikasi Calon Pendonor Darah Menggunakan Metode *Naive Bayes Classifier* (Studi Kasus: Calon Pendonor Darah di Kota Semarang)”. Begitu banyak pihak yang telah membantu, oleh karena itu rasa hormat dan terima kasih penulis ingin sampaikan kepada:

1. Bapak Dr. Tarno, M.Si. selaku Ketua Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro sekaligus dosen pembimbing II.
2. Bapak Dr. Rukun Santoso, M.Si. selaku dosen pembimbing I.
3. Bapak dan Ibu dosen Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro yang telah memberikan ilmu selama proses belajar di Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis dalam penulisan Tugas Akhir ini.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan dalam kesempatan berikutnya.

Semarang, 29 Maret 2017

Penulis

## ABSTRAK

Klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Ada beberapa metode yang termasuk dalam metode klasifikasi, salah satunya adalah *Naïve Bayes*. *Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat. Pada penelitian ini dilakukan perbaikan terhadap metode *Naïve Bayes* dalam menghitung peluang bersyarat dari setiap fiturnya menggunakan dua pendekatan, yaitu pendekatan fungsi densitas normal dan pendekatan selisih peluang kumulatif. Kedua pendekatan tersebut digunakan untuk melakukan klasifikasi calon pendonor darah di Kota Semarang. Variabel prediktor yang digunakan adalah kadar hemoglobin, tensi atas, tensi bawah, dan berat badan. Hasil dari penelitian ini menunjukkan bahwa kedua pendekatan tersebut sama baiknya dalam melakukan klasifikasi dengan nilai *Matthews Correlation Coefficient* (MCC) sebesar 0,8985841 atau mendekati +1.

**Kata Kunci** : Klasifikasi, *Naïve Bayes*, Fungsi Densitas, Peluang Kumulatif, Donor Darah, *Matthews Correlation Coefficient* (MCC).

## ABSTRACT

Classification is the process of finding a model or function that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. There are some methods that are included in the classification methods, one of them is *Naïve Bayes*. *Naïve Bayes* is a prediction technique that based simple probabilistic are based on the application of Bayes theorem with strong independence assumption. On this study carried out correction to the *Naïve Bayes* method in calculating the conditional probability of each feature using two approaches, normal density function and cumulative distribution function approaches. These two approaches are used to classify prospective blood donors in Semarang City. The predictor variables used are hemoglobin level, upper blood pressure, lower blood pressure, and weight. The result of this study shows that both approaches have the same *Matthews Correlation Coefficient* (MCC) values, 0,8985841 or close to +1. It means that both approaches equally well doing classification.

**Keywords** : Classification, Naïve Bayes, Normal Density Function, Cumulative Distribution Function, Blood Donors, *Matthews Correlation Coefficient* (MCC).

## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN I .....	ii
HALAMAN PENGESAHAN II .....	iii
KATA PENGANTAR .....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	4
BAB II TINJAUAN PUSTAKA	
2.1 Fungsi Darah .....	5
2.2 Donor Darah .....	6
2.3 Pendorong Darah .....	6
2.4 Syarat-Syarat Menjadi Pendorong Darah .....	7
2.4.1 Usia .....	8
2.4.2 Hemoglobin .....	8
2.4.3 Berat Badan .....	9
2.4.4 Tensi Atas/Sistole .....	9
2.4.5 Tensi Bawah/Diastole .....	9

2.5	Klasifikasi .....	10
2.6	Probabilitas .....	10
2.6.1	Kejadian Saling Bebas dan Kejadian Saling Asing .....	10
2.6.2	Teorema Bayes .....	12
2.7	Fungsi Densitas Normal .....	13
2.8	Klasifikasi <i>Naïve Bayes</i> .....	16
2.9	Karakteristik <i>Naïve Bayes</i> .....	23
2.10	Pemilihan Fitur Berbasis Statistik .....	23
2.11	Teknik Validasi Model .....	26
2.12	Ketepatan Klasifikasi .....	27

### BAB III METODOLOGI PENELITIAN

3.1	Sumber Data .....	29
3.2	Variabel Penelitian .....	29
3.3	Tahapan Analisis .....	30

### BAB IV HASIL DAN PEMBAHASAN

4.1	Deskripsi Data .....	34
4.2	Hubungan antara Fitur dengan Status Calon Pendoror .....	36
4.2.1	Hubungan Fitur Hemoglobin dengan Status Calon Pendoror .....	36
4.2.2	Hubungan Fitur Tensi Atas dengan Status Calon Pendoror .....	36
4.2.3	Hubungan Fitur Tensi Bawah dengan Status Calon Pendoror .....	37
4.2.4	Hubungan Fitur Berat Badan dengan Status Calon Pendoror .....	37
4.2.5	Hubungan Fitur Usia dengan Status Calon Pendoror..	38
4.3	Pemilihan Fitur Berbasis Statistik .....	38
4.3.1	Uji Homogenitas untuk Hemoglobin .....	39
4.3.2	Uji Homogenitas untuk Tensi Atas .....	39
4.3.3	Uji Homogenitas untuk Tensi Bawah .....	40

4.3.4	Uji Homogenitas untuk Berat Badan .....	41
4.3.5	Uji Homogenitas untuk Usia .....	42
4.3.6	Uji <i>Welch's t-test</i> untuk Hemoglobin .....	43
4.3.7	Uji <i>Welch's t-test</i> untuk Tensi Atas .....	44
4.3.8	Uji <i>Welch's t-test</i> untuk Tensi Bawah .....	45
4.3.9	Uji <i>Welch's t-test</i> untuk Berat Badan .....	47
4.3.10	Uji t untuk Usia .....	48
4.4	Teknik Validasi Model .....	49
4.5	Proses Klasifikasi dengan <i>Naïve Bayes</i> .....	49
4.5.1	Klasifikasi <i>Naïve Bayes</i> dengan Pendekatan Densitas Normal .....	51
4.5.2	Klasifikasi <i>Naïve Bayes</i> dengan Pendekatan Selisih Peluang Kumulatif .....	58
4.6	Ketepatan Klasifikasi .....	66
 BAB V PENUTUP		
5.1	Kesimpulan .....	69
5.2	Saran .....	69
DAFTAR PUSTAKA .....		70
LAMPIRAN .....		72

## DAFTAR TABEL

	Halaman
Tabel 1. Matriks Konfusi .....	28
Tabel 2. Proporsi Label Kelas Calon Pendonor .....	34
Tabel 3. Fitur-Fitur Calon Pendonor Darah .....	35
Tabel 4. Hemoglobin dengan Status Calon Pendonor .....	36
Tabel 5. Tensi Atas dengan Status Calon Pendonor .....	36
Tabel 6. Tensi Bawah dengan Status Calon Pendonor .....	37
Tabel 7. Berat Badan dengan Status Calon Pendonor .....	37
Tabel 8. Usia dengan Status Calon Pendonor .....	38
Tabel 9. Data Latih .....	50
Tabel 10. Data Uji .....	50
Tabel 11. Jumlah Dat Latih per Kelas .....	51
Tabel 12. Matriks Konfusi dari Fungsi Densitas .....	67
Tabel 13. Matriks Konfusi dari Peluang Kumulatif .....	67

## DAFTAR GAMBAR

	Halaman
Gambar 1. $P(a \leq X \leq b) =$ luas wilayah yang berada di interval 1 dan 2 ..	14
Gambar 2. $P(X = x) \approx$ luas daerah yang diarsir dengan $x = 2$ .....	15
Gambar 3. Kurva Pendekatan Fungsi Densitas .....	21
Gambar 4. Kurva Pendekatan Selisih Peluang Normal .....	22
Gambar 5. Diagaram Alir Klasifikassi <i>Naïve Bayes</i> .....	33

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Calon Pendorong Darah Kota Semarang 16-27 September 2016.....	73
Lampiran 2. Uji <i>Welch's t-test</i> untuk Setiap Fitur .....	74
Lampiran 3. Data Latih dan Data Uji Calon Pendorong Darah untuk Proses Klasifikasi .....	76
Lampiran 4. Hasil Klasifikasi Menggunakan Metode <i>Naïve Bayes</i> dengan Pendekatan Fungsi Densitas .....	80
Lampiran 5. Hasil Klasifikasi Menggunakan Metode <i>Naïve Bayes</i> dengan Pendekatan Selisih Peluang Kumulatif.....	83
Lampiran 6. Sintaks Klasifikasi Menggunakan Metode <i>Naïve Bayes</i> dengan Pendekatan Fungsi Densitas .....	86
Lampiran 7. Sintaks Klasifikasi Menggunakan Metode <i>Naïve Bayes</i> dengan Pendekatan Selisih Peluang Kumulatif.....	88

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Donor darah adalah proses menyalurkan darah atau unsur-unsur darah dari satu orang ke sistem peredaran orang lainnya. Donor darah berhubungan dengan kondisi medis seperti kehilangan darah dalam jumlah besar disebabkan trauma, operasi, syok dan tidak berfungsinya organ pembentuk sel darah merah (Depkes RI, 2009).

Aktifitas donor darah merupakan kewajiban setiap masyarakat sebagai wujud kepedulian terhadap orang lain. Banyak orang yang tidak tahu tentang manfaat donor darah bagi kesehatan. Bahkan ada juga orang yang enggan mendonorkan darah karena khawatir terhadap efek samping yang ditimbulkannya. Padahal dengan melakukan donor darah, maka sel-sel darah di dalam tubuh menjadi lebih cepat terganti dengan yang baru. Apabila mendonorkan darah tiga bulan sekali, maka kesehatan tubuh tetap terjaga. Selain bermanfaat untuk membantu orang lain, donor darah juga membuat tubuh kita menjadi lebih sehat (Depkes RI, 2009).

Dengan meningkatnya permintaan suplai darah di masyarakat, persediaan darah yang mencukupi sangat dibutuhkan. Meskipun demikian, perekrutan dan pemeliharaan pendonor darah tetap sebagai tantangan utama bagi organisasi donor darah. Pendonor harus terlebih dahulu menjalani pemeriksaan kesehatan, baik pengukuran tekanan darah, golongan darah, kadar hemoglobin (Hb) maupun konsultasi medis. Sebagian calon pendonor mungkin berkeinginan untuk

mendonorkan darahnya, tetapi itu semua tergantung dengan hasil pemeriksaan, sehingga ada yang memenuhi persyaratan untuk mendonorkan darah dan ada yang terpaksa kecewa (Masser, 2008).

Pada tugas akhir ini digunakan metode *Naïve Bayes* dalam mengklasifikasi dan memprediksi seseorang apakah bisa mendonorkan darahnya atau tidak, berdasarkan kadar hemoglobin, tensi atas, tensi bawah, berat badan, dan usia yang dimilikinya sebagai variabel pendukung. Variabel pendukung ini sebagai informasi pendukung dalam menentukan seseorang bisa mendonorkan darahnya atau tidak. Menurut Han dan Kamber (2006), klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya atau dapat memprediksi kecenderungan data-data yang muncul dimasa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data.

Ada beberapa metode yang termasuk dalam metode klasifikasi, salah satunya adalah *Naïve Bayes*. Menurut Prasetyo (2013), *Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat. Meskipun asumsi independensi cukup sulit dipenuhi dalam dunia nyata, akan tetapi dalam praktiknya *Naïve Bayes* masih berfungsi dengan baik. *Naïve Bayes* masuk ke dalam algoritma kategori *eager learner*. Algoritma ini melakukan proses pelatihan yang cukup memakan waktu tetapi proses prediksinya berjalan cepat. *Naïve Bayes* dapat diterapkan pada data fitur kategorik maupun kontinu. Pada data fitur yang

bersifat kontinu, *Naïve Bayes* mengasumsikan data kontinu ke dalam distribusi tertentu dan memperkirakan parameter distribusi dengan data latih. Biasanya digunakan distribusi Gaussian untuk menghitung probabilitas bersyarat dari fitur kontinu pada sebuah kelas. Parameter distribusi Gaussian adalah mean dan standar deviasi (Han dan Kamber, 2006).

Pada penelitian ini dalam menghitung nilai probabilitas bersyarat dari fitur kontinu pada sebuah kelas digunakan selisih peluang kumulatif dengan interval tertentu. Setelah itu hasil klasifikasi menggunakan pendekatan fungsi densitas normal dibandingkan dengan hasil klasifikasi menggunakan pendekatan selisih peluang kumulatif. Selain itu penelitian mengenai klasifikasi pendonor darah juga sebelumnya belum pernah dilakukan, sehingga membuat peneliti tertarik menggunakan studi kasus ini. Peneliti mengambil topik mengenai “Klasifikasi calon pendonor darah menggunakan Metode *Naïve Bayes Classifier*” dengan studi kasus data calon pendonor darah di PMI Kota Semarang.

## 1.2 Rumusan Masalah

Permasalahan yang akan dibahas pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana ketepatan model klasifikasi data calon pendonor darah menggunakan metode *Naïve Bayes* dengan pendekatan fungsi densitas normal?
2. Bagaimana ketepatan model klasifikasi data calon pendonor darah menggunakan metode *Naïve Bayes* dengan pendekatan selisih peluang kumulatif?

3. Bagaimana perbandingan nilai ketepatan model klasifikasi data calon pendonor darah menggunakan metode *Naïve Bayes* dengan pendekatan fungsi densitas normal dan selisih peluang kumulatif?

### **1.3 Batasan Masalah**

Permasalahan pada tugas akhir ini dibatasi untuk calon pendonor darah di Kota Semarang yang mendaftarkan diri ke PMI *mobile unit*. Pengolahan tersebut diberikan batasan masalah pada pembahasan tentang perbandingan nilai ketepatan model klasifikasi menggunakan metode *Naïve Bayes* dengan dua pendekatan yaitu fungsi densitas normal dan selisih peluang kumulatif.

### **1.4 Tujuan Penelitian**

Berdasarkan rumusan permasalahan yang telah diuraikan sebelumnya, maka tujuan yang ingin dicapai dalam penelitian ini adalah mengukur ketepatan model klasifikasi data calon pendonor darah menggunakan metode *Naïve Bayes* dengan dua pendekatan, fungsi densitas normal dan selisih peluang kumulatif, serta membandingkan nilai ketepatan model klasifikasi status calon pendonor dari kedua pendekatan tersebut.