

## Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode *Naïve Bayes*

Rado Yendra<sup>1</sup>, Laradea Marifni<sup>2</sup>, Irma Suryani<sup>3</sup>

<sup>1,2,3</sup> Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau  
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293  
Email: radoyendra@uin-suska.ac.id, marifnidea@gmail.com

### ABSTRAK

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tetapi minim informasi. Data mining merupakan cara untuk menemukan informasi dengan mencari pola atau aturan tertentu dari data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan informasi atau fitur-fitur seperti jenis kelamin, nilai Seleksi Kompetensi Dasar (SKD), nilai Seleksi Kompetensi Bidang (SKB) 1, nilai Seleksi Kompetensi Bidang (SKB) 2, nilai Seleksi Kompetensi Bidang (SKB) 3, dan keterangan lulus atau tidak lulus seleksi sebagai sumber datanya, diharapkan dapat menghasilkan informasi tentang klasifikasi seleksi penerimaan Calon Pegawai Negeri sipil (CPNS) 2017. Metode yang digunakan adalah metode *Naïve Bayes* yaitu metode yang digunakan untuk memprediksi berbasis probabilitas. Penelitian ini bertujuan untuk mendapatkan distribusi yang tepat untuk fitur-fitur data kontinu. Dari hasil pengujian yang dilakukan dengan menggunakan data sebanyak 284 dengan 227 sebagai data training dan 57 sebagai data testing baik untuk fitur kontinu menggunakan distribusi normal maupun distribusi gamma, akurasi untuk distribusi normal adalah 81% dengan jumlah data yang tepat sebanyak 46 dan yang tidak tepat sebanyak 11. Sedangkan akurasi distribusi gamma sebesar 70% dengan data yang tepat sebanyak 40 dan data yang tidak tepat sebanyak 17.

**Kata Kunci:** Data Mining, *Naïve Bayes*, CPNS, Distribusi Normal, Distribusi Gamma.

### ABSTRACT

*The rapid growth of accumulated data has created a condition rich in data but minimal information. Data mining is a way to find information by looking for certain patterns or rules from large amounts of data that are expected to overcome these conditions. By utilizing information or features such as gender, Selection of Basic Competence (SKD), Selection of Competency (SKB) 1, Selection of Competency (SKB) 2, Selection of Competency (SKB) 3, and admission pass or not pass the selection as its data source, it is hoped that it can produce information about the prospective civil servant (CPNS) 2017. The method used is the Naïve Bayes method, which is a method used to predict probability based. This research aims to get the right distribution for continuous data features. From the results of testing carried out using 284 data with 227 as training data and 57 as testing data both for continuous features using normal distribution and gamma distribution, the accuracy for normal distribution is 81% with the right amount of data as much as 46 and that which is incorrect 11. While the accuracy of the gamma distribution is 70% with the right data as many as 40 and incorrect data as much as 17.*

**Keywords:** Data Mining, *Naïve Bayes*, CPNS, Normal Distribution, Gamma Distribution

## Pendahuluan

Instansi pemerintah di Indonesia secara berkala menyelenggarakan rekrutmen dan seleksi Calon Pegawai Negeri Sipil (CPNS). Hampir tiap tahun pemerintah membuka formasi CPNS dan tidak pernah sepi pelamar. Hasil penelitian Dr. Madeline Kusharwanti (2008) menyatakan bahwa proses penerimaan dan seleksi CPNS di Indonesia dinilai masih sangat buruk dan menimbulkan kerawanan terjadinya Korupsi Kolusi dan Nepotisme (KKN), sehingga kualitas Sumber Daya Manusia (SDM) yang diterima sebagai Pegawai Negeri Sipil (PNS) masih rendah, padahal baik buruknya suatu pemerintahan sangat tergantung pada baik buruknya mesin birokrasi sebagai penyelenggara pemerintahan. Sementara itu, birokrasi pemerintah sangat bergantung pada PNS sebagai aparatur penyelenggara pemerintah. Birokrasi yang baik membutuhkan sosok PNS yang profesional, yang mempunyai sikap dan perilaku yang penuh kesetiaan, ketaatan, disiplin, bermoral, bermental baik, akuntabel dan memiliki kesadaran yang tinggi terhadap tanggung jawab sebagai pelayan publik yang baik.

Berdasarkan hal yang telah diuraikan tersebut, bisa dilihat bahwa birokrasi pemerintah membutuhkan PNS yang handal dan berkualitas untuk sistem pemerintahan yang lebih baik. Dan juga berdasarkan penelitian oleh Rendra [15] yang berjudul “Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma *Decision Tree C4.5*”. Penulis tertarik untuk mengklasifikasi suatu permasalahan sistem data mining yang mudah digunakan dan dapat menyajikan informasi relevan dan akurat yaitu mengklasifikasi data mining untuk menampilkan informasi seleksi penerimaan Calon Pegawai Negeri Sipil 2017 pada Kementerian Riset, Teknologi, dan Pendidikan Tinggi menggunakan metode *Naïve Bayes*.

Metode *Naïve Bayes Classifier* menggunakan prinsip dasar probabilitas sebagai acuannya. Metode ini bertujuan untuk mengklasifikasikan kelas satu objek dengan cara menghitung probabilitas dari masing-masing karakteristik objek tersebut berdasarkan kumpulan data yang tersedia.

## Metode dan Bahan Penelitian

### 1. Data Mining

Menurut Han dan Kamber menjelaskan bahwa data mining merupakan pemilihan atau “menggali” pengetahuan dari jumlah data yang banyak [8]. Berbeda dengan Segall, Guha dan Nonis yang menjelaskan bahwa data mining disebut penemuan pengetahuan atau menemukan pola yang tersembunyi dalam data. Data mining adalah proses menganalisis data dari perspektif yang berbeda dan meringkas menjadi informasi yang berguna” [17].

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan [11].

1. Deskripsi
2. Klasifikasi
3. Estimasi
4. Prediksi
5. Klasterisasi
6. Asosiasi

### 2. Klasifikasi

Nugroho & Subanar [13] mengatakan bahwa klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan yaitu:

1. Proses Training

Pada proses ini akan digunakan data training set atau data sampel yang telah diketahui label-label atau atribut dari data sampel tersebut untuk membangun model.

2. Proses Testing

Pada proses testing ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses training maka dibangun data yang disebut dengan data testing untuk mengklasifikasi label-labelnya.

3. Metode Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [18]. Berikut disajikan Persamaan dari Teorema Bayes pada Persamaan 1.

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} \tag{1}$$

dimana:

- $X$  : Data dengan kelas yang belum diketahui
- $Y$  : Data kelas spesifik
- $P(Y|X)$  : Probabilitas  $Y$  berdasarkan kondisi  $X$
- $P(Y)$  : Probabilitas  $Y$
- $P(X|Y)$  : Probabilitas  $X$  berdasarkan kondisi  $Y$
- $P(X)$  : Probabilitas  $X$

Pada sebuah dataset, setiap setiap baris atau dokumen  $I$  diasumsikan sebagai vektor dari nilai-nilai atribut dimana tiap nilai-nilai menjadi peninjauan atribut  $X_j$  ( $j \in [1, m]$ ). Setiap baris mempunyai label kelas  $Y_i \in \{Y_1, Y_2, \dots, Y_k\}$  sebagai nilai variabel kelas  $Y$ , sehingga untuk melakukan klasifikasi dapat dihitung nilai probabilitas ( $Y = Y_i | X = X_j$ ), dikarenakan pada Naïve Bayes diasumsikan setiap atribut saling bebas, maka persamaan yang didapat adalah sebagai berikut:

$$P(Y = Y_i | X = X_j) = \frac{P(Y = Y_i)P(X = X_j | Y = Y_i)}{P(X = X_j)} \tag{2}$$

$$P(Y_1, Y_2, \dots, Y_k | X_1, X_2, \dots, X_m) = \frac{P(Y_1, Y_2, \dots, Y_k)P(X_1, X_2, \dots, X_m | Y_1, Y_2, \dots, Y_k)}{P(X_1, X_2, \dots, X_m)} \tag{3}$$

Maka Persamaan (3) bisa ditulis menjadi Persamaan (4) berikut [14]:

$$P(Y_i | X_j) = \frac{P(Y_i) \prod_{j=1}^m P(X_j | Y_i)}{P(X_j)} \tag{4}$$

4. Peubah Acak dan Distribusi Peubah Acak

Definisi 1:

Peubah acak ialah suatu fungsi yang mengaitkan suatu bilangan real pada setiap unsur dalam ruang sampel,[20]

Suatu peubah acak diskrit dapat dinyatakan sebagai berikut:

$$F(x) = \sum p_{x_x}(x) \tag{5}$$

Suatu peubah acak kontinu dapat dinyatakan sebagai berikut:

$$f(x) = \int_{-\infty}^{\infty} f_x(x)dx \tag{6}$$

5. Ekspektasi dan Variansi

Definisi 2 [22]:

Misalkan  $X$  suatu peubah acak dengan distribusi peluang  $f(x)$ , maka nilai harapannya atau rata-rata  $X$  ialah:

$$E(X) = \mu = \sum_x x f(x) \quad (7)$$

**Definisi 3** [22]:

Misalkan  $X$  adalah suatu peubah acak dengan distribusi peluang  $f(x)$  dan rata-rata  $\mu$ , maka :

$$Var(X) = \sigma^2 = E[(X - \mu)^2] = \sum (x - \mu)^2 f(x), \quad \text{bila } X \text{ diskrit dan} \quad (8)$$

$$Var(X) = \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \quad \text{bila } X \text{ kontinu.} \quad (9)$$

## 6. Distribusi Normal , Fungsi Gamma dan Distribusi Gamma

Persamaan matematika distribusi peluang peubah normal kontinu bergantung pada dua parameter  $\mu$  dan  $\sigma$  yaitu rata-rata dan simpangan bakunya. Variabel random  $X$  dalam distribusi normal mempunyai fungsi densitas peluang (fdp) sebagai berikut:

$$f(x_j, \mu, \sigma) = P(X_j|Y_i) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_j - \mu)^2}{2\sigma^2}} \quad (10)$$

Adapun persamaan yang digunakan untuk menghitung nilai rata-rata (*mean*) dapat dilihat sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (11)$$

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (12)$$

Fungsi gamma didefinisikan

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \text{untuk } \alpha > 0 \quad (13)$$

**Definisi 4** [22]:

Perubah acak kontinu  $X$  berdistribusi gamma dengan parameter  $\alpha$  dan  $\beta$  , jika fungsi padatnya berbentuk [20]:

$$f(x_j, \alpha, \beta) = P(X_j|Y_i) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x_j^{\alpha-1} e^{-\frac{x_j}{\beta}} \quad (14)$$

**Teorema 1** [22]:

Bila  $X$  berdistribusi gamma  $X \sim G(x|\alpha, \beta, 0)$  maka rata-rata dan variansi distribusi gamma adalah:

$$\mu = E(X) = \alpha\beta \quad (15)$$

$$\sigma^2 = \alpha\beta^2 \quad (16)$$

$$E(X^2) = \beta^2\alpha^2 + \alpha\beta^2 \quad (17)$$

Maka berdasarkan dari Persamaan (15), (16) dan (17) diperoleh persamaan untuk menentukan parameter-parameter dari distribusi gamma yang disajikan pada Persamaan (18) dan (19).

$$\alpha = \frac{\mu}{\beta} \quad (18)$$

$$\beta = \frac{\frac{1}{n} \sum (x_i - \mu)^2}{\bar{X}} \quad (19)$$

## Hasil Dan Pembahasan

### 1. Perhitungan Naïve Bayes

Perhitungan naïve bayes dilakukan dengan menghitung sebanyak 284 data yang dibagi menjadi 227 data training dan 57 data testing.

## 1.1 Perhitungan Data Training

Data training dihitung secara manual menggunakan persamaan-persamaan yang disajikan sebelumnya. Untuk fitur-fitur data yang kontinu dihitung menggunakan distribusi normal dan distribusi gamma. Berikut disajikan Tabel data training pada Tabel 1.

**Tabel 1. Data Training**

NO	JK	SKD	SKB1	SKB2	SKB3	KETERANGAN
1	LK	72,6	46	93,5	84	L
2	PR	71,2	39	92,417	93,7	L
3	LK	73	75	79,5	81,5	L
4	PR	75	55	78,667	81,5	L
5	LK	76,2	58	73,083	68,1	TL
6	LK	75,8	48	73,167	65,1	TL
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
226	LK	70,8	41	81,667	78,8	L
227	PR	70,2	49	63,333	76,8	L

Berdasarkan Tabel 1 terdapat 2 fitur data kategori yaitu JK dan Keterangan, dan 4 fitur data kontinu yaitu SKD, SKB1, SKB2, dan SKB3. Oleh karena itu 4 fitur kontinu tersebut akan dihitung menggunakan distribusi normal dan distribusi gamma.

### 1. Distribusi Normal

#### a) Mean dan deviasi standard untuk fitur SKD

$$\mu_L = \frac{72,6 + 71,2 + \dots + 70,2}{109} = 70,288$$

$$\mu_{TL} = \frac{76,2 + 75,8 + \dots + 63,2}{118} = 69,375$$

$$\sigma^2_L = \frac{(72,6 - 70,288)^2 + (71,2 - 70,288)^2 + \dots + (70,2 - 70,288)^2}{109 - 1} = 17,11$$

$$\sigma_L = \sqrt{17,11} = 4,14$$

$$\sigma^2_{TL} = \frac{(76,2 - 69,375)^2 + (75,8 - 69,375)^2 + \dots + (63,2 - 69,375)^2}{118 - 1} = 20,5$$

$$\sigma_{TL} = \sqrt{20,5} = 4,53$$

#### b) Mean dan deviasi standard untuk fitur SKB 1

$$\mu_L = \frac{46 + 39 + \dots + 49}{109} = 60,44$$

$$\mu_{TL} = \frac{58 + 48 + \dots + 62}{118} = 54,3$$

$$\sigma^2_L = \frac{(46 - 60,44)^2 + (39 - 60,44)^2 + \dots + (49 - 60,44)^2}{109 - 1} = 241,77$$

$$\sigma_L = \sqrt{241,77} = 15,55$$

$$\sigma^2_{TL} = \frac{(58 - 54,3)^2 + (48 - 54,3)^2 + \dots + (62 - 54,3)^2}{118 - 1} = 159,12$$

$$\sigma_{TL} = \sqrt{159,12} = 12,61$$

**c) Mean dan deviasi standard untuk fitur SKB 2**

$$\mu_L = \frac{93,5 + 92,417 + \dots + 63,33}{109} = 81,74$$

$$\mu_{TL} = \frac{73,083 + 73,167 + \dots + 66,67}{118} = 76,13$$

$$\sigma^2_L = \frac{(93,5 - 81,74)^2 + (73,167 - 81,74)^2 + \dots + (63,33 - 81,74)^2}{109 - 1} = 41,55$$

$$\sigma_L = \sqrt{41,55} = 6,45$$

$$\sigma^2_{TL} = \frac{(73,083 - 76,13)^2 + (73,167 - 76,13)^2 + \dots + (66,67 - 76,13)^2}{118 - 1} = 50,69$$

$$\sigma_{TL} = \sqrt{50,69} = 7,12$$

**d) Mean dan deviasi standard untuk fitur SKB 3**

$$\mu_L = \frac{84 + 93,7 + \dots + 76,8}{109} = 81,3$$

$$\mu_{TL} = \frac{68,1 + 65,1 + \dots + 83,8}{118} = 76,38$$

$$\sigma^2_L = \frac{(84 - 81,3)^2 + (93,7 - 81,3)^2 + \dots + (76,8 - 81,3)^2}{109 - 1} = 66,73$$

$$\sigma_L = \sqrt{66,73} = 8,2$$

$$\sigma^2_{TL} = \frac{(68,1 - 76,38)^2 + (65,1 - 76,38)^2 + \dots + (83,8 - 76,38)^2}{118 - 1} = 64,76$$

$$\sigma_{TL} = \sqrt{64,76} = 8,05$$

**2. Distribusi Gamma**

**a)  $\alpha$  dan  $\beta$  untuk fitur SKD**

$$\beta_L = \frac{\frac{1}{109} [(72,6 - 70,288)^2 + (71,2 - 70,288)^2 + \dots + (70,2 - 70,288)^2]}{70,288}$$

$$\beta_L = 0,241$$

$$\alpha_L = \frac{70,288}{0,241} = 291,494$$

$$\beta_{TL} = \frac{\frac{1}{118} [(76,2 - 69,375)^2 + (75,8 - 69,375)^2 + \dots + (63,2 - 69,375)^2]}{69,375}$$

$$\beta_{TL} = 0,293$$

$$\alpha_{TL} = \frac{69,375}{0,293} = 236,889$$

**b)  $\alpha$  dan  $\beta$  untuk fitur SKB 1**

$$\beta_L = \frac{\frac{1}{109} [(46 - 60,44)^2 + (39 - 60,44)^2 + \dots + (49 - 60,44)^2]}{60,44}$$

$$\beta_L = 3,963$$

$$\alpha_L = \frac{60,44}{3,963} = 15,25$$

$$\beta_{TL} = \frac{\frac{1}{118} [(58 - 54,3)^2 + (48 - 54,3)^2 + \dots + (62 - 54,3)^2]}{54,3}$$

$$\beta_{TL} = 2,91$$

$$\alpha_{TL} = \frac{54,3}{2,91} = 18,686$$

**c)  $\alpha$  dan  $\beta$  untuk fitur SKB 2**

$$\beta_L = \frac{\frac{1}{109} [(93,5 - 81,74)^2 + (73,167 - 81,74)^2 + \dots + (63,33 - 81,74)^2]}{81,74}$$

$$\beta_L = 0,504$$

$$\alpha_L = \frac{81,74}{0,504} = 162,288$$

$$\beta_{TL} = \frac{\frac{1}{118} [(73,083 - 76,13)^2 + (73,167 - 76,13)^2 + \dots + (66,67 - 76,13)^2]}{76,13}$$

$$\beta_{TL} = 0,66$$

$$\alpha_{TL} = \frac{76,13}{0,66} = 115,299$$

**d)  $\alpha$  dan  $\beta$  untuk fitur SKB 3**

$$\beta_L = \frac{\frac{1}{109} [(84 - 81,3)^2 + (93,7 - 81,3)^2 + \dots + (76,8 - 81,3)^2]}{81,3}$$

$$\beta_L = 0,813$$

$$\alpha_L = \frac{81,3}{0,813} = 99,968$$

$$\beta_{TL} = \frac{\frac{1}{118} [(68,1 - 76,38)^2 + (65,1 - 76,38)^2 + \dots + (83,8 - 76,38)^2]}{76,38}$$

$$\beta_{TL} = 0,841$$

$$\alpha_{TL} = \frac{76,38}{0,841} = 90,851$$

Sedangkan untuk probabilitas fitur Jenis Kelamin dan Keterangan disajikan pada Tabel 2 dan Tabel 3.

**Tabel 2. Probabilitas Kemunculan Nilai Untuk Fitur Jenis Kelamin**

Jenis Kelamin	Jumlah kejadian "Dipilih"		Probabilitas	
	L	TL	L	TL
Laki-Laki	59	42	59/109	42/118
Perempuan	50	76	50/109	76/118
Jumlah	109	118	1	1

**Tabel 3. Probabilitas kemunculan Nilai Untuk Fitur Keterangan Lulus / Tidak Lulus**

Keterangan	Jumlah kejadian "Dipilih"		Probabilitas	
	L	TL	L	TL
Jumlah	109	118	109/227	118/227

## 1.2 Perhitungan Data Testing

Setelah melakukan perhitungan pada data training, selanjutnya adalah melakukan perhitungan data testing berdasarkan hasil perhitungan pada data training. Berikut disajikan Tabel data testing pada Tabel 4.

**Tabel 4. Data Testing**

NO	JK	SKD	SKB1	SKB2	SKB3	KETERANGAN
1	PR	65	45	71,667	77	TL
2	PR	69,8	43	61,667	77,4	TL
3	PR	74,2	85	85,417	90	L
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
56	LK	69,8	44	78,75	83,8	TL
57	LK	69,2	50	86,917	86,8	L

Berdasarkan Tabel 4 dilihat bahwa jumlah data testing sebanyak 57, dan akan dilakukan perhitungan untuk semua data testing tersebut.

### 1. Distribusi Normal

#### a) Untuk Data Testing no. 1

$$P(\text{SKD} = 65|\text{L}) = \frac{1}{\sqrt{2\pi} \cdot 4,14} e^{-\frac{(65-70,288)^2}{2(17,11)}} = 0,04259$$

$$P(\text{SKD} = 65|\text{TL}) = \frac{1}{\sqrt{2\pi} \cdot 4,53} e^{-\frac{(65-69,375)^2}{2(20,5)}} = 0,055183$$

$$P(\text{SKB} 1 = 45|\text{L}) = \frac{1}{\sqrt{2\pi} \cdot 15,55} e^{-\frac{(45-60,44)^2}{2(241,77)}} = 0,015671$$

$$P(\text{SKB} 1 = 45|\text{TL}) = \frac{1}{\sqrt{2\pi} \cdot 12,61} e^{-\frac{(45-54,3)^2}{2(159,12)}} = 0,024112$$

$$P(\text{SKB} 2 = 71,667|\text{L}) = \frac{1}{\sqrt{2\pi} \cdot 6,45} e^{-\frac{(71,667-81,741)^2}{2(41,55)}} = 0,0182703$$

$$P(\text{SKB} 2 = 71,667|\text{TL}) = \frac{1}{\sqrt{2\pi} \cdot 7,12} e^{-\frac{(71,667-76,126)^2}{2(50,692)}} = 0,046037$$

$$P(\text{SKB} 3 = 77|\text{L}) = \frac{1}{\sqrt{2\pi} \cdot 8,17} e^{-\frac{(77-81,302)^2}{2(66,374)}} = 0,042402$$

$$P(\text{SKB} 3 = 77|\text{TL}) = \frac{1}{\sqrt{2\pi} \cdot 8,05} e^{-\frac{(77-76,38)^2}{2(64,763)}} = 0,0491172$$

Maka:

$$\text{Nilai L} = (0,04259) \times (0,015671) \times (0,0182703) \times (0,042402) \times \frac{50}{109} \times \frac{109}{227} = 0,000000114$$

$$\text{Nilai TL} = (0,055183) \times (0,024112) \times (0,046037) \times (0,0491172) \times \frac{76}{118} \times \frac{118}{227} = 0,000001007$$

Nilai probabilitas dapat dihitung:

$$\text{Probabilitas L} = \frac{0,000000114}{0,000000114+0,000016458} = 0,101586$$

$$\text{Probabilitas TL} = \frac{0,000001007}{0,000001007+0,000000114} = 0,898414$$

Klasifikasi:  
Tidak Lulus



**e) Untuk Data Testing no.57**

$$P(\text{SKD} = 69,2|\text{L}) = \frac{1}{\sqrt{2\pi} 4,14} e^{-\frac{(69,2-70,288)^2}{2(17,11)}} = 0,0931$$

$$P(\text{SKD} = 69,2|\text{TL}) = \frac{1}{\sqrt{2\pi} 4,53} e^{-\frac{(69,2-69,375)^2}{2(20,5)}} = 0,08664$$

$$P(\text{SKB 1} = 50|\text{L}) = \frac{1}{\sqrt{2\pi} 15,55} e^{-\frac{(50-60,44)^2}{2(241,77)}} = 0,0205$$

$$P(\text{SKB 1} = 50|\text{TL}) = \frac{1}{\sqrt{2\pi} 12,61} e^{-\frac{(50-54,3)^2}{2(159,12)}} = 0,2987$$

$$P(\text{SKB 2} = 86,917|\text{L}) = \frac{1}{\sqrt{2\pi} 6,45} e^{-\frac{(86,917-81,741)^2}{2(41,55)}} = 0,04482$$

$$P(\text{SKB 2} = 86,917|\text{TL}) = \frac{1}{\sqrt{2\pi} 7,12} e^{-\frac{(86,917-76,126)^2}{2(50,692)}} = 0,01778$$

$$P(\text{SKB 3} = 86,8|\text{L}) = \frac{1}{\sqrt{2\pi} 8,17} e^{-\frac{(86,8-81,302)^2}{2(66,374)}} = 0,03885$$

$$P(\text{SKB 3} = 86,8|\text{TL}) = \frac{1}{\sqrt{2\pi} 8,05} e^{-\frac{(86,8-76,38)^2}{2(64,763)}} = 0,0216$$

Maka:

$$\text{Nilai L} = (0,0931) \times (0,0205) \times (0,04482) \times (0,03885) \times \frac{59}{109} \times \frac{109}{227} = 0,00000086$$

$$\text{Nilai TL} = (0,08664) \times (0,2987) \times (0,01778) \times (0,0216) \times \frac{42}{118} \times \frac{118}{227} = 0,00000018$$

Nilai probabilitas dapat dihitung:

$$\text{Probabilitas L} = \frac{0,000000863}{0,000000863+0,000000184} = 0,8243$$

$$\text{Probabilitas TL} = \frac{0,000000184}{0,000000184+0,000000863} = 0,1757$$

Klasifikasi:

**2. Distribusi Gamma**

**a) Untuk Data Testing no.1**

$$P(\text{SKD} = 65|\text{L}) = \frac{1}{(0,241)^{291,5} \Gamma(291,5)} (65)^{291,5-1} e^{-\frac{65}{0,241}} = 0,04394$$

$$P(\text{SKD} = 65|\text{TL}) = \frac{1}{(0,293)^{236,89} \Gamma(236,89)} (65)^{236,89-1} e^{-\frac{65}{0,293}} = 0,05775$$

$$P(\text{SKB 1} = 45|\text{L}) = \frac{1}{(3,963)^{15,25} \Gamma(15,25)} (45)^{15,25-1} e^{-\frac{45}{3,963}} = 0,01884$$

$$P(\text{SKB 1} = 45|\text{TL}) = \frac{1}{(2,91)^{18,69} \Gamma(18,69)} (45)^{18,69-1} e^{-\frac{45}{2,91}} = 0,02799$$

$$P(\text{SKB 2} = 71,667|\text{L}) = \frac{1}{(0,504)^{162,29} \Gamma(162,288)} (71,67)^{162,29-1} e^{-\frac{71,667}{0,504}} = 0,018483$$

$$P(\text{SKB 2} = 71,667|\text{TL}) = \frac{1}{(0,66)^{115,3} \Gamma(115,3)} (71,67)^{115,3-1} e^{-\frac{71,667}{0,66}} = 0,048616$$

$$P(\text{SKB 3} = 77|\text{L}) = \frac{1}{(0,813)^{99,97} \Gamma(99,97)} (77)^{99,97-1} e^{-\frac{77}{0,813}} = 0,044767$$

$$P(\text{SKB 3} = 77|\text{TL}) = \frac{1}{(0,841)^{90,851} \Gamma(90,851)} (77)^{90,851-1} e^{-\frac{77}{0,841}} = 0,049192$$

Maka:

$$\text{Nilai L} = (0,04394) \times (0,01884) \times (0,018483) \times (0,044767) \times \frac{50}{109} \times \frac{109}{227} = 0,00000015$$

$$\text{Nilai TL} = (0,05775) \times (0,02799) \times (0,048616) \times (0,049192) \times \frac{76}{118} \times \frac{118}{227} = 0,00000129$$

Nilai probabilitas dapat dihitung:

$$\left. \begin{aligned} \text{Probabilitas L} &= \frac{0,00000015}{0,00000015+0,00000129} = 0,10443 \\ \text{Probabilitas TL} &= \frac{0,00000129}{0,00000129+0,00000015} = 0,89557 \end{aligned} \right\} \begin{array}{l} \text{Klasifikasi:} \\ \text{Tidak Lulus} \end{array}$$

e) Untuk Data Testing no. 57

$$\begin{aligned} P(\text{SKD} = 69,2|\text{L}) &= \frac{1}{(0,241)^{291,5}\Gamma(291,5)} (69,2)^{291,5-1} e^{-\frac{69,2}{0,241}} = 0,056623 \\ P(\text{SKD} = 69,2|\text{TL}) &= \frac{1}{(0,293)^{236,89}\Gamma(236,89)} (69,2)^{236,89-1} e^{-\frac{69,2}{0,293}} = 0,08764 \\ P(\text{SKB 1} = 50|\text{L}) &= \frac{1}{(3,963)^{15,25}\Gamma(15,25)} (50)^{15,25-1} e^{-\frac{50}{3,963}} = 0,02395 \\ P(\text{SKB 1} = 50|\text{TL}) &= \frac{1}{(2,91)^{18,69}\Gamma(18,69)} (50)^{18,69-1} e^{-\frac{50}{2,91}} = 0,03228 \\ P(\text{SKB 2} = 86,917|\text{L}) &= \frac{1}{(0,504)^{162,288}\Gamma(162,288)} (86,917)^{162,288-1} e^{-\frac{86,917}{0,504}} = 0,06231 \\ P(\text{SKB 2} = 86,917|\text{TL}) &= \frac{1}{(0,66)^{115,3}\Gamma(115,3)} (86,917)^{115,3-1} e^{-\frac{86,917}{0,66}} = 0,01707 \\ P(\text{SKB 3} = 86,8|\text{L}) &= \frac{1}{(0,813)^{99,97}\Gamma(99,97)} (86,8)^{99,97-1} e^{-\frac{86,8}{0,813}} = 0,03689 \\ P(\text{SKB 3} = 86,8|\text{TL}) &= \frac{1}{(0,841)^{90,851}\Gamma(90,851)} (86,8)^{90,851-1} e^{-\frac{86,8}{0,841}} = 0,02015 \end{aligned}$$

Maka:

$$\begin{aligned} \text{Nilai L} &= (0,056623) \times (0,02395) \times (0,06231) \times (0,03689) \times \frac{59}{109} \times \frac{109}{227} = 0,00000081 \\ \text{Nilai TL} &= (0,08764) \times (0,03228) \times (0,01707) \times (0,02015) \times \frac{42}{118} \times \frac{118}{227} = 0,00000018 \end{aligned}$$

Nilai probabilitas dapat dihitung:

$$\left. \begin{aligned} \text{Probabilitas L} &= \frac{0,00000081}{0,00000081+0,00000018} = 0,81816 \\ \text{Probabilitas TL} &= \frac{0,00000018}{0,00000018+0,00000081} = 0,18184 \end{aligned} \right\} \begin{array}{l} \text{Klasifikasi:} \\ \text{Lulus} \end{array}$$

## 2 Analisis Data

Untuk membuktikan keakuratan algoritma naïve bayes, perlu dilakukan perhitungan akurasi. Perhitungan akurasi pada sistem klasifikasi naïve bayes dapat diketahui dengan cara membandingkan hasil klasifikasi pengujian pada sistem dengan hasil klasifikasi yang sudah ada sebelumnya. Berikut disajikan hasil klasifikasi naïve bayes dengan fitur kontinu menggunakan distribusi normal dan distribusi gamma pada Tabel 5 dan Tabel 6.

**Tabel 5. Hasil Klasifikasi NBC (Naïve Bayes Classifier) dengan Fitur Kontinu menggunakan Distribusi Normal**

NO	LULUS	TIDAK LULUS	L	TL	Hasil	Target
1	0,000000114	0,000001007	0,101586	0,898414	TL	TL
2	0,000000006	0,000000216	0,027468	0,972532	TL	TL
3	0,000000146	0,000000008	0,948995	0,051005	L	L
4	0,000001395	0,000001268	0,523823	0,476177	L	L
5	0,000000785	0,000000551	0,587508	0,412492	L	L

6	0,000000930	0,000000939	0,497783	0,502217	TL	TL
7	0,000000589	0,000000116	0,835981	0,164019	L	L
8	0,000000321	0,000000284	0,530228	0,469772	L	L
9	0,000000852	0,000000366	0,699856	0,300144	L	L
10	0,000000426	0,000000287	0,597669	0,402331	L	TL
11	0,000000103	0,000000691	0,130259	0,869741	TL	TL
12	0,000000466	0,000001520	0,234628	0,765372	TL	TL
13	0,000000055	0,000000397	0,122198	0,877802	TL	TL
14	0,000000130	0,000001018	0,113237	0,886763	TL	TL
15	0,000000014	0,000000330	0,040717	0,959283	TL	TL
16	0,000000968	0,000000340	0,740248	0,259752	L	L
17	0,000000235	0,000000139	0,628452	0,371548	L	L
18	0,000000604	0,000000521	0,537063	0,462937	TL	TL
19	0,000000170	0,000001302	0,115493	0,884507	TL	TL
20	0,000000238	0,000000325	0,422945	0,577055	TL	TL
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
53	0,000001410	0,000000695	0,669714	0,330286	L	TL
54	0,000001269	0,000000677	0,652068	0,347932	L	TL
55	0,000001157	0,000001345	0,462449	0,537551	TL	TL
56	0,000000941	0,000000598	0,611722	0,388278	L	TL
57	0,000000863	0,000000184	0,824268	0,175732	L	L

**Tabel 6. Hasil Klasifikasi NBC (*Naïve Bayes Classifier*) dengan Fitur Kontinu menggunakan Distribusi Gamma**

NO	LULUS	TIDAK LULUS	L	TL	Hasil	Target
----	-------	----------------	---	----	-------	--------

1	0,00000015	0,00000129	0,10443	0,89557	TL	TL
2	0,00000000	0,00000023	0,01592	0,98408	TL	TL
3	0,00000012	0,00000001	0,92434	0,07566	L	L
4	0,00000146	0,00000105	0,5798	0,4202	L	L
5	0,00000086	0,00000052	0,6235	0,3765	L	L
6	0,00000120	0,00000099	0,54694	0,45306	L	TL
7	0,00000059	0,00000010	0,85169	0,14831	L	L
8	0,00000037	0,00000027	0,57586	0,42414	L	L
9	0,00000080	0,00000029	0,73197	0,26803	L	L
10	0,00000047	0,00000028	0,62651	0,37349	L	TL
11	0,00000011	0,00000075	0,1253	0,8747	TL	TL
12	0,00000060	0,00000170	0,26181	0,73819	TL	TL
13	0,00000006	0,00000046	0,11018	0,88982	TL	TL
14	0,00000016	0,00000123	0,11242	0,88758	TL	TL
15	0,00000001	0,00000040	0,02869	0,97131	TL	TL
16	0,00000110	0,00000034	0,76546	0,23454	L	L
17	0,00000029	0,00000017	0,6294	0,3706	L	L
18	0,00000070	0,00000054	0,56281	0,43719	L	TL
19	0,00000020	0,00000149	0,11843	0,88157	TL	TL
20	0,00000027	0,00000036	0,42938	0,57062	TL	TL
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
53	0,00000098	0,00000054	0,64387	0,35613	L	TL
54	0,00000159	0,00000082	0,66017	0,33983	L	TL

55	0,00000144	0,00000080	0,64098	0,35902	L	TL
56	0,00000094	0,00000058	0,62044	0,37956	L	TL
57	0,00000081	0,00000018	0,81816	0,18184	L	L

Berdasarkan Tabel 5 dan Tabel 6 bisa dilihat hasil klasifikasi dengan fitur kontinu menggunakan distribusi normal dan distribusi gamma. Pada kolom pertama yaitu kolom nomor, terdapat warna merah pada beberapa nomor yang artinya nomor yang berwarna merah menunjukkan jenis kelamin laki-laki dan warna hitam untuk jenis kelamin perempuan. Dan juga pada kolom hasil dan target masing-masing kolom terdapat cell yang berwarna kuning yang menunjukkan bahwa kolom hasil dan target terdapat perbedaan hasil klasifikasi.

Pada umumnya perbedaan hasil atau selisih yang diperoleh dapat digunakan untuk mengevaluasi kinerja dari sistem yang digunakan. Berikut disajikan tingkat akurasi sistem klasifikasi NBC (*Naïve Bayes Classifier*) dengan fitur kontinu menggunakan distribusi normal dan distribusi gamma pada Persamaan 20 dan Persamaan 21.

1. Distribusi Normal

$$\text{Akurasi Distribusi Normal} = \frac{46}{57} \times 100\% = 81\% \quad (20)$$

2. Distribusi Gamma

$$\text{Akurasi Distribusi Gamma} = \frac{40}{57} \times 100\% = 70\% \quad (21)$$

Berdasarkan Persamaan 20 dan Persamaan 21 bisa dilihat bahwa distribusi normal memiliki akurasi sebesar 81% dimana hasil data yang tepat sebanyak 46 dan hasil data yang tidak tepat sebanyak 11. Sedangkan distribusi gamma memiliki akurasi sebesar 70% dimana hasil data yang tepat sebanyak 40 dan hasil data yang tidak tepat sebanyak 17. Hal ini menunjukkan bahwa distribusi normal memiliki tingkat akurasi lebih tinggi dibandingkan distribusi gamma. Oleh karena itu untuk klasifikasi NBC (*Naïve Bayes Classifier*) dengan fitur-fitur data yang kontinu lebih tepat menggunakan distribusi normal daripada distribusi gamma.

### Kesimpulan

Berdasarkan hasil penelitian, dapat diambil kesimpulan bahwa klasifikasi seleksi penerimaan CPNS menggunakan metode Naïve Bayes menghasilkan akurasi yang baik dengan perbandingan data training dan testing adalah 80% : 20%. Pengujian perbandingan akurasi metode Naïve Bayes dimana fitur kontinu menggunakan distribusi normal dan distribusi gamma menghasilkan akurasi distribusi normal lebih tinggi daripada distribusi gamma, yaitu 81% untuk distribusi normal dan 70% untuk distribusi gamma. Oleh karena itu distribusi normal lebih tepat digunakan untuk menghitung fitur-fitur data kontinu dalam metode Naïve Bayes karena lebih akurat dibandingkan distribusi gamma.

### Daftar Pustaka

- [1] A. Yusnita, R. Handini, Sistem Pendukung Keputusan Menentukan Lokasi Rumah Makan yang Strategis menggunakan Metode Niave Bayes, *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012)*, Semarang, 2012.
- [2] Aljumah A, Siddiqui M., Hypertension Interventions using Classification based Data Mining, *Research Journal of Applied Sciences, Engineering and Technology*, Vol. 7, pp. 3593-3602, 2014.

- [3] Dangare S, Apte S., Data Mining Approach For Prediction of Heart Disease Using Neural Networks, *International Journal of Computer Engineering and Technology (IJCET)*, Vol. 3, pp. 30-40, 2012.
- [4] Davinder Kaur et. al., Review of Decision Tree Data Mining Algorithms: ID3 and C4.5, *International Conference on Information Technology and Computer Science*, pp. 5-8, 2015.
- [5] Dudewicz, Edward J and Mishra, Satya N., *Statistika Matematika Modern*, ITB, Bandung, 1995.
- [6] Fadlina, Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori, *Informasi dan Teknologi (INTI)*, Vol. 3, No. 1, hal. 144-154, Mei 2014.
- [7] Gumilang, I. C., Sudjalwo, dan Rakhmadi, A., Prediksi Persediaan Obat Dengan Metode Naïve Bayes (Studi Kasus: Apotek Saputra), *Universitas Muhammadiyah Surakarta*, Surakarta, 2014.
- [8] Han, J., and Kamber, M., *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann, San Francisco, 2011.
- [9] Kusri, luthfi taufiq Emha, *Algoritma Data Mining*, Penerbit Andi, Yogyakarta, 2009.
- [10] Larose, D. T., *Discovering Knowledge In Data*, John Willey and Sons, Inc., 2005.
- [11] Meilani, B. D., dan Susanti, N., Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes, *JURNAL LINK*, Vol. 21, No. 2, hal. 1-6, September 2014.
- [12] Misbahussurur, Ahmad, Estimasi Parameter Distribusi Gamma dengan Metode Maximum Likelihood, *Tugas Akhir Mahasiswa Universitas Islam Negeri Maulana Malik Ibrahim*, Malang. 2009.
- [13] Nugroho, A., dan Subanar. Klasifikasi Naïve Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil. *Berkala MIPA*. Vol. 23, No. 3, halaman 297-308, September 2013.
- [14] Prasetyo, Eko, *Data Mining dan Konsep Aplikasi Menggunakan Matlab*, Penerbit Andi, Yogyakarta. 2012.
- [15] Rendra, Kumara Graha, Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma *Decision Tree C4.5*, Skripsi Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, 2015.
- [16] Romli, Lili, Masalah Reformasi Birokrasi, *Jurnal kebijakan dan Manajemen PNS*, Yogyakarta. 2007.
- [17] Richard S. Segall, Gauri S. Guha, and Sarath A. Nonis, Data Mining of Environmental Stress Tolerances On Plants, *Kybernetes*, Vol. 37 Issue: 1, pp. 127-148, 2008.
- [18] Saleh, Alfa, Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, *Citec Journal*, Vol.2, No.3, 2015.
- [19] Seema Sharma et. al., Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies, *International Journal of Computer Applications*, Vol. 82, pp. 20-27, Nov. 2013.
- [20] Sneha K. Dehankar et. al., Web Page Classification using Apriori Algorithm and Naïve Base Classifier, *International Journal of Advanced Research in Computer Science and Management Studies*, Vol. 3, pp. 527-533, Apr. 2015.
- [21] Umar Husei, *Metode Penelitian, Aplikasi dalam Pemasaran*, Jakarta, 1997.
- [22] Walpole, Ronald E. dan Myers Raymond H., *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan Terjemahan RK Sembiring.*, ITB, Bandung, 1995.
- [23] Wasiati, H., dan Wijayanti, D., Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: PT. Karyatama Mitra Sejati Yogyakarta), *Indonesian Journal on Networking and Security*, Vol. 3, No 2, hal. 45-51, 2014.