

## PREDIKSI *MISSING IMPUTATION* UNTUK DATA PENYEBARAN DEMAM BERDARAH *DENGUE* MENGGUNAKAN *NAÏVE BAYES*

Aji Seto Arifianto, S.ST, M.T<sup>1)</sup>, Didit Rahmat Hartadi, S.Kom, M.T<sup>2)</sup>,  
Anik Nur Novitasari E S, A.Md<sup>3)</sup>

<sup>1),2)</sup>Dosen Teknologi Informasi Politeknik Negeri Jember  
Jalan Mastrip Kotak Pos 164 Jember, Jawa Timur 68121

<sup>1)</sup>[ajiseto@gmail.com](mailto:ajiseto@gmail.com) <sup>2)</sup>[didit.r.hartadi@gmail.com](mailto:didit.r.hartadi@gmail.com)

<sup>3)</sup>Mahasiswa, Politeknik Negeri Jember

\*<sup>3)</sup>[aniknurnovitasari@gmail.com](mailto:aniknurnovitasari@gmail.com)

### ABSTRACT

*Dengue fever is a disease caused by dengue virus by Aedes Aegypti mosquitoes intermediary. There are four dengue virus serotypes, called DEN-1, DEN-2, DEN-3, and DEN-4, each of them can lead to dengue fever, either mild or fatal. Based on a survey of public health in Jember, recorded during January 2015 there were 300 cases of dengue fever, and seven of them died. Previous research using fuzzy method to determined the potential spread of Dengue in Jember. Parameters that used is the amount of rain in a period, the amount of rain in one month, free amount of larva and house index. Data that used is taken from 2009 until 2012 from 31 districts. The weaknesses in this study were not contain a way to resolve the imputation of missing data. In fact, survey data is often incomplete. Based on these issues, it will be created a prediction system of imputation of missing data on the spread of dengue by using Naïve Bayes method. The data refers to the prediction of the missing data and the data were used as training data, so it can be processed further. This research has been successfully predicting missing data imputation using existing data, so that the missing data can be completed with high degree of accuracy.*

**Keywords:** dengue fever, missing of imputation, naïve bayes.

### PENDAHULUAN

Demam Berdarah *Dengue* atau disingkat DBD adalah penyakit yang ditularkan oleh nyamuk *Aedes Aegypti* melalui virus yang dibawanya [1]. Penyakit ini ditemukan di daerah tropis, subtropis dan menjangkit luas di Asia Tenggara, termasuk Indonesia. Terdapat 4 jenis virus dengue yaitu DEN 1, DEN 2, DEN 3 dan DEN 4, semua jenis virus tersebut menjadi penyebab penyakit demam berdarah, baik ringan maupun fatal. Berdasarkan hasil survei ke Dinas Kesehatan (Dinkes) Kabupaten Jember tercatat selama Januari 2015 ada 7 orang diantaranya meninggal dunia dari 300 kasus pasien DBD. Pekan pertama bulan Februari 2015 sudah terjadi 42 kasus. Hal ini menunjukkan sedemikian cepat penyebaran wabah DBD.

Sebuah penelitian tentang metode fuzzy yang digunakan untuk memprediksi potensi penyebaran Demam berdarah dengan akurasi mencapai 75% [2]. Namun dalam penelitian selama ini terdapat banyak sekali kasus hilangnya nilai pada dataset untuk

atribut tertentu. Permasalahan hilangnya nilai pada data ini lebih sering disebut dengan *missing imputation*. Penyebab terjadinya *missing imputation* adalah tidak adanya respon terhadap unit maupun item, hal ini merupakan permasalahan yang terjadi pada sebagian survei dan menjadikan hasil dari penelitian kurang baik. Dalam penelitian ini dataset penyebaran demam berdarah *dengue* di Kabupaten Jember dengan metode *Naïve Bayes* dan data yang akan digunakan adalah data faktor-faktor penyebaran penyakit terhadap demam berdarah *dengue* yaitu curah hujan (CH), hari hujan (HH), angka bebas jentik (ABJ) dan *house indeks* (HI). Data yang digunakan mulai tahun 2009 sampai dengan tahun 2012 dari 31 kecamatan di Kabupaten Jember.

### TINJAUAN PUSTAKA

#### A. Demam Berdarah

Menurut Soengeng Soegijanto dalam bukunya yang berjudul Demam Berdarah *Dengue* edisi 2 menyatakan bahwa

“Penyakit demam berdarah dengue (DBD) merupakan penyakit yang disebabkan oleh virus dengue dan ditularkan oleh nyamuk *Aedes Aegypti*”. Penyakit ini umumnya mewabah, dapat menyerang berbagai kelompok usia masyarakat baik orang dewasa maupun anak-anak, bahkan seringkali berakhir dengan kematian. Gejala umum penderita Demam Berdarah antara lain demam hingga mencapai 40°C, nyeri di mata, sendi, otot dan tulang, mimisan/gusi berdarah, terkadang disertai batuk, pilek, mual, muntah dan gatal pada telapak kaki [3].

Tiga faktor yang berpengaruh dalam penularan DBD yaitu manusia, virus dan factor perantara. Masa inkubasi **instrinsik** (dalam tubuh manusia) virus *dengue* dalam darah selama 4-7 hari mulai 1-2 hari sebelum orang tersebut mengalami demam. Virus *dengue* yang sudah ada didalam tubuh manusia (misal si A) yang mengalami viremia (adanya virus di dalam darah) kemudian digigit oleh nyamuk *Aedes Aegypti*, selanjutnya virus akan berkembang biak dan menyebar ke seluruh tubuh bagian tubuh nyamuk, dan juga kelenjar saliva. Kira-kira satu minggu setelah menghisap darah penderita (masa inkubasi **ekstrensik**). Virus ini akan tetap berada dalam tubuh nyamuk sepanjang hidupnya. Ketika nyamuk tersebut menggigit orang lain (misal si B) maka akan terjadi penularan. Seseorang yang di dalam darahnya memiliki virus *dengue* (infeksi) merupakan sumber penularan DBD [4].

**B. Metode Naïve Bayes**

Teorema *Bayes* awal dikemukakan oleh orang kebangsaan Inggris pada tahun 1763 yang bernama Thomas Bayes. Teorema tersebut kemudian disempurnakan oleh Laplace. *Bayes* merupakan metode statistik sederhana yang mempertimbangkan nilai probabilitas. *Bayes* kemudian berkembang untuk menyelesaikan kasus-kasus klasifikasi (*Bayes Classifier*). Kemudian berkembang istilah Naïve Bayes yang selanjutnya dapat digunakan untuk memprediksi data yang hilang.

*Naïve Bayes* menggunakan model fitur independen-si, maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data berkaitan atau tidak dengan fitur lain dalam data yang sama [5]. Prediksi *Bayes* didasarkan pada teorema

*Bayes* dengan menggunakan persamaan sebagai berikut:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Penjelasan dari formula tersebut adalah sebagai berikut:

- a. P(H|E) adalah Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika (*evidence*) E terjadi.
- b. P(E|H) adalah Probabilitas sebuah kejadian E terjadi tanpa memandang *evidence* apapun.
- c. P(E) adalah Probabilitas awal (priori) kejadian E terjadi tanpa memandang hipotesis/*evidence* lain.

Ada beberapa hal yang penting dari aturan *Bayes* tersebut, yaitu:

- a. sebuah probabilitas awal/prior H atau P(H) adalah probabilitas dari suatu hipotesis sebelum diamati.
- b. sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah diamati.

Kaitan antara *Naïve Bayes* dengan topik klasifikasi adalah bahwa hipotesis dalam teorema *Bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan parameternya merupakan fitur-fitur yang menjadi masukkan atau *variabels input*. Jika A adalah variabel masukkan yang berisi fitur dan B adalah label kelas, maka *Naïve Bayes* ditulis dengan notasi P(A|B). Selama proses pelatihan harus dilakukan *learning* probabilitas akhir P(B|A) pada model untuk kombinasi A dan B berdasarkan informasi yang didapat dari data training. Dengan model tersebut, suatu data uji A' dapat diklasifikasi dengan mencari nilai B' dengan memaksimalkan nilai P(A'|B') yang didapat. Formulasi *Naïve Bayes* dengan menggunakan persamaan (2.4) untuk klasifikasi adalah:

$$P(B|A) = \frac{P(A) \prod_{i=1}^q P(A_i|B)}{P(A)}$$

P(B|A) adalah probabilitas data dengan variabel A pada kelas B. P(B) adalah probabilitas awal B dari kelas B.  $\prod_{i=1}^q P(A_i|B)$  adalah probabilitas independen kelas B dari semua fitur dalam vektor A. Nilai P(A) konstan sehingga dalam perhitungan prediksi tinggal menghitung bagian P(B)  $\prod_{i=1}^q P(A_i|B)$  dengan memilih yang terbesar sebagai kelas yang terpilih sebagai hasil prediksi. Sementara

probabilitas independen  $\prod_{i=1}^q P(A_i|B)$  tersebut merupakan pengaruh semua fitur dari data tersebut setiap kelas B, yang dinotasikan dengan.

$$P(A|B = y) = \prod_{i=1}^q P(A_i|B = y)$$

Setiap set fitur  $A = \{A_1, A_2, A_3, \dots, A_q\}$  terdiri atas q atribut (q dimensi).

Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut metode *Naïve Bayes* bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*). *Naïve Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi, dapat digunakan jika mendapati atribut yang tidak relevan. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

### C. Metode *Naïve Bayes* untuk data Kontinyu

Umumnya, Bayes mudah dihitung untuk fitur bertipe **kategoris** seperti "gender" dengan nilai pria dan wanita. Namun untuk fitur dengan tipe numerik (**kontinyu**) ada perlakuan khusus sebelum dimasukkan dalam *Naïve Bayes*. Caranya adalah:

- a. Melakukan diskretisasi pada setiap fitur kontinyu dan mengganti nilai fitur kontinyu tersebut dengan nilai internal diskret. Pendekatan ini dilakukan dengan mentransformasikan fitur kontinyu kedalam fitur ordinal.
- b. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinyu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi yang umum digunakan untuk mempresentasikan probabilitas bersyarat dari fitur kontinyu pada sebuah kelas  $P(X_i|Y)$  adalah Gaussian, sedangkan distribusi Gaussian dikarakteristikan dengan dua parameter: *mean*,  $\mu$  dan varian,  $\sigma^2$ . Untuk setiap kelas  $y_j$  untuk fitur  $X_i$  adalah: Mean( $\mu$ ) merupakan rata-rata yang diperoleh dari attribute yang dimiliki data kontinyu. Didalam mean ( $\mu$ ) terdapat rumus sebagai berikut:

$$\text{Mean}(\mu) = \frac{\sum_{i=1}^n x_i}{n}$$

Sedangkan untuk parameter variansi ( $\sigma^2$ ) didapat dari rumus sebagai berikut:

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

Dari parameter-parameter yang sudah kita temukan akan dapat fungsi densitas mengekspresikan probabilitas relatif. Dengan fungsi densitas probabilitasnya adalah sebagai berikut:

$$\varphi_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu$  dan  $\sigma$  dapat diestimasi dari data, untuk setiap kelas. Gunakan  $\varphi_{\mu, \sigma}(x)$  untuk  $P(X|C)$ .

### METODE PENELITIAN

Pelaksanaan penelitian ini mengacu pada langkah-langkah sebagai berikut:

#### 1. Studi Literatur

Studi literatur dilakukan menggali sebanyak mungkin informasi yang relevan terkait teori dan metode yang dipakai untuk memecahkan masalah *missing imputation* dengan fokus perspektif persebaran demam berdarah melalui berbagai referensi yang relevan. Referensi tersebut berisikan tentang:

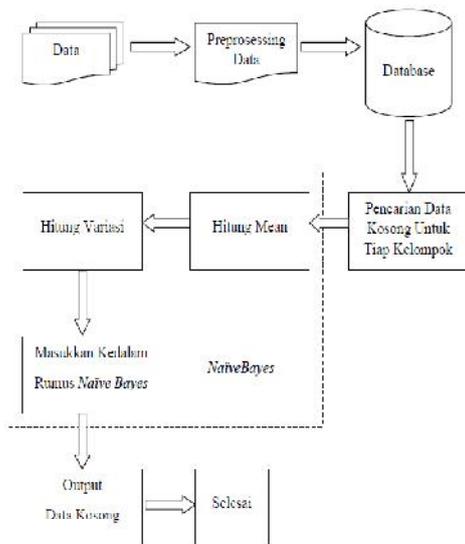
- a) Demam berdarah (penyebab dan penyebarannya)
- b) *Naïve Bayes*
- c) Pengolahan data

#### 2. Pengambilan dan Pengolahan Data

Tahap ini dilakukan dengan mengambil data dari Dinas Kesehatan Kabupaten Jember, proses ini dilakukan berkerjasama dengan peneliti dari Politeknik Elektronika Negeri Surabaya yang berjudul "*A Prediction System of Dengue Fever Using Monte Carlo Method*" [6]. Pengolahan data yang dimaksud adalah proses memetakan data berdasarkan beberapa kategori, untuk mengetahui informasi tentang tahun, curah hujan, hari hujan, angka bebas jentik setiap tahunnya. Setelah data diolah akan digunakan pada tahap analisis/desain.

#### 3. Desain

Proses desain merupakan proses penggambaran, perancangan dan pembuatan Penerapan *Naïve Bayes* Untuk Prediksi *Missing Imputation* Pada Data Persebaran Demam Berdarah *Dengue*.



Gambar 1. Desain Sistem

**DATA**

Bahan yang digunakan dalam pembuatan Penerapan *Naive Bayes* Untuk Prediksi *Missing Imputation* adalah data Persebaran Demam Berdarah *Dengue* di Kabupaten Jember sebagai berikut:

- a. Data Curah Hujan setiap bulan dengan satuan millimeter (mm<sup>3</sup>) yang diambil dari BAPEDA Jember berdasarkan stasiun pengukur pada setiap wilayah di Kabupaten Jember antara tahun 2009 sampai dengan tahun 2012.

Tabel 1. Contoh Data yang Kosong

Kecamatan		GMKS
Bulan	Jan	100
	Feb	100
	Mar	10
	Apr	100
	Mei	100
	Jun	100
	Jul	
	Agt	
	Sep	100
	Okt	100
	Nov	100
	Des	100

- b. Data Hari Hujan setiap bulan dengan suatu hari (h) yang diambil dari Dinas Pengairan Kabupaten Jember berdasarkan stasiun pengukuran hujan

antara tahun 2009 sampai dengan tahun 2012

- c. Data Angka Bebas Jentik dengan satuan persen (%) yang diambil berdasarkan puskesmas di setiap kecamatan pada data Dinas Kesehatan Kabupaten Jember antara tahun 2009 sampai dengan tahun 2012.
- d. Data *House Index* dengan satuan persen (%) yang diambil berdasarkan data dari puskesmas di setiap kecamatan yang berada dalam lingkup Dinas Kesehatan Kabupaten Jember antara tahun 2009 sampai dengan tahun 2012.

**ALUR SISTEM**

Pada bagian alur program menjelaskan beberapa tahapan seperti pada gambar di bawah ini:

- a. Input Data *Missing Imputation*

Pada tahap input pengguna dapat memilih variabel-variabel faktor penyebaran DBD yang telah ada dalam aplikasi ini, seperti Curah Hujan (CH), Hari Hujan (HH), House Indeks (HI) dan Angka Bebas Jentik (ABJ).

Tabel 2. Contoh Data yang Diganti dengan -1

Kecamatan		GMKS
Bulan	Jan	100
	Feb	100
	Mar	10
	Apr	100
	Mei	100
	Jun	100
	Jul	-1
	Agt	-1
	Sep	100
	Okt	100
	Nov	100
	Des	100

Keterangan:

-1 = Data Kosong

- b. Proses *Naive Bayes*

Pada tahap ini proses terbagi menjadi 3 bagian, yaitu Pra Proses, Proses dan Pasca Proses.

- 1) Pra Proses

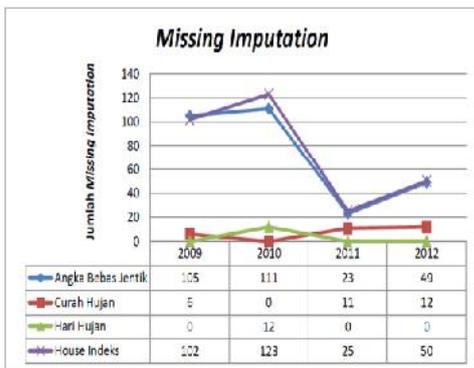
Setelah user memilih variabel apa yang akan di tindak lanjuti maka user dapat memilih tahun dan kecamatan lalu menekan tombol Search.

2) Proses

Tahap setelah melakukan pra proses pengguna menekan tombol proses. Dimana tombol tersebut terdapat perhitungan yang menggunakan metode *Naïve Bayes* untuk mendapatkan nilai *missing imputation*. Proses penghitungan *Naïve Bayes* memiliki beberapa langkah dengan menggunakan data pada Tabel 1 dan Tabel 2.

**HASIL DAN PEMBAHASAN**

Data tersebut terdiri dari empat faktor yaitu curah hujan, hari hujan, house indeks, dan angka bebas jentik. Data diambil dari 31 kecamatan di Kabupaten Jember dalam rentang tahun 2009 sampai tahun 2012. Total data 5.952 yang terdiri dari 629 data *missing imputation*. Seperti disajikan dalam Gambar 2.



Gambar 2. Jumlah *Missing Data*

a) Mencari Nilai *Mean*

Nilai *mean* didapat dari menjumlah data dari setiap variabel kecuali data kosong (-1) dan dibagi banyaknya jumlah data yang bukan termasuk masuk data kosong. Menggunakan persamaan *mean* (2.5) seperti perhitungan di bawah ini.

$$\text{Mean} = \frac{100+100+10+100+100+100+100+100+100+100}{10} = \frac{910}{10} = 91$$

b) Mencari Nilai Variansi

Nilai Variansi didapat dari hasil nilai varian yang sudah diproses dan hasil dijadikan bilangan bulat menggunakan persamaan variansi (2.6) dengan hasil perhitungan di antaranya:

$$\begin{aligned} \sigma^2 &= \frac{(100-91)^2+(100-91)^2+(10-91)^2+(100-91)^2+(100-91)^2+(100-91)^2+(100-91)^2+(100-91)^2+(100-91)^2+(100-91)^2}{10 \cdot (10-1)} \\ &= \frac{(9)^2+(9)^2+(81)^2+(9)^2+(9)^2+(9)^2+(9)^2+(9)^2+(9)^2+(9)^2}{90} \\ &= \frac{81+81+6561+81+81+81+81+81+81+81}{90} \\ &= \frac{7290}{90} = 81 \end{aligned}$$

$$\sigma = \sqrt{81} = 9$$

c) Mencari Nilai *Missing Imputation*

Dari nilai tiap variabel diproses kedalam rumus *Naïve Bayes*, kecuali data kosong dengan menggunakan persamaan *Naïve Bayes* (2.6) dengan menghasilkan hasil nilai *Missing Imputation*.

Bulan Januari

$$\begin{aligned} \text{NB}_{\text{Januari}} &= \frac{1}{\sqrt{2 \cdot \pi} (9)} e^{-\frac{(100-91)^2}{2 \cdot 81}} = \frac{81}{162} \\ &= 0,5 \\ &= \frac{1}{\sqrt{2 \cdot 3,14} (9)} e^{0,5} \\ &= \frac{1}{\sqrt{56,52}} e^{0,5} \\ &= \frac{1}{7,517} e^{0,5} \\ &= 0,133 e^{0,5} = 0,066 \end{aligned}$$

Bulan Februari

$$\begin{aligned} \text{NB}_{\text{Februari}} &= \frac{1}{\sqrt{2 \cdot \pi} (9)} e^{-\frac{(100-91)^2}{2 \cdot 81}} = \frac{81}{162} \\ &= 0,5 \\ &= \frac{1}{\sqrt{2 \cdot 3,14} (9)} e^{0,5} \\ &= \frac{1}{\sqrt{56,52}} e^{0,5} \\ &= \frac{1}{7,517} e^{0,5} \\ &= 0,133 e^{0,5} = 0,066 \end{aligned}$$

Bulan Maret

$$\begin{aligned} \text{NB}_{\text{Maret}} &= \frac{1}{\sqrt{2 \cdot \pi} (9)} e^{-\frac{(10-91)^2}{2 \cdot 81}} = \frac{6561}{162} \\ &= 40,5 \\ &= \frac{1}{\sqrt{2 \cdot 3,14} (9)} e^{40,5} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{56,52}} e^{40,5} \\
 &= \frac{1}{7,517} e^{40,5} \\
 &= 0,133 e^{40,5} = 3,280
 \end{aligned}$$

Bulan April

$$\begin{aligned}
 NB_{\text{April}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan Mei

$$\begin{aligned}
 NB_{\text{Mei}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan Juni

$$\begin{aligned}
 NB_{\text{Juni}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan September

$$\begin{aligned}
 NB_{\text{September}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan Oktober

$$\begin{aligned}
 NB_{\text{Oktober}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan November

$$\begin{aligned}
 NB_{\text{Desember}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Bulan Desember

$$\begin{aligned}
 NB_{\text{Desember}} &= \frac{1}{\sqrt{2 \cdot \pi (9)}} e^{-\frac{(100-91)^2}{2 \cdot 91}} = \frac{81}{162} \\
 &= 0,5 \\
 &= \frac{1}{\sqrt{2 \cdot 3,14 (9)}} e^{0,5} \\
 &= \frac{1}{\sqrt{56,52}} e^{0,5} \\
 &= \frac{1}{7,517} e^{0,5} \\
 &= 0,133 e^{0,5} = 0,066
 \end{aligned}$$

Setelah proses perhitungan tiap variabel kemudian hasil di jumlahkan sehingga mendapatkan nilai *missing imputation*.

$$\begin{aligned}
 MI &= 0,066+0,066+3,280+0,066+ \\
 &\quad 0,066+0,066+0,066+0,066+0,066 \\
 &= 3,874
 \end{aligned}$$

Perhitungan akan berulang hingga semua data kosong yang ada lengkap. Perulangan dimulai dari perhitungan rumus *Naïve Bayes* dengan cara menghitung kembali data yang ada, kecuali data *missing imputation*.

Keterangan:

- Mean( $\mu$ ) = Nilai rata-rata dari setiap kecamatan
- Variansi( $\sigma^2$ ) = Nilai tengah dari setiap kecamatan

Variasi( $\sigma$ ) = Nilai tengah yang telah diakar  
 $NB_{\text{Bulan}}(\varphi_{\mu,\sigma}(x))$  = Nilai *Naïve Bayes* tiap bulan  
 $MI$  = Nilai *Missing Imputation*

- d) Pasca Proses  
 Maka perhitungan selesai dan menghasilkan data lengkap.
- e) Output Hasil Data *Missing Imputation*  
 Menampilkan hasil data *missing imputation* yang diperoleh dari perhitungan menggunakan metode *Naïve Bayes*.

#### KESIMPULAN

Berdasarkan hasil dari pelaksanaan penelitian, dapat diambil kesimpulan antara lain, data diperoleh mengandung missing imputation dan perlu dilakukan prediksi untuk mengisi kekosongan tersebut. Proses mendapatkan nilai-nilai yang kosong menggunakan metode *Naïve Bayes* untuk data kontinu dengan beberapa tahapan yaitu menghitung nilai mean, nilai variansi dan kedua nilai dimasukkan kedalam rumus *Naïve Bayes*. Untuk penelitian selanjutnya perlu dilakukan juga perhitungan untuk mengukur seberapa baik hasil yang didapatkan dari prediksinya dengan berbagai metode perhitungan toleransi *error*.

#### DAFTAR PUSTAKA

- [1] Syukur Nugroho. 2010. Sistem Informasi Peringatan Dini dan Peramalan Penderita Demam Berdarah di Sura-karta. Universitas 11 Maret Surakarta :Skripsi Santoso, Djoko. 2013. Wawancara Klasifikasi Hari Hujan Bulanan. Jl. Brawijaya 63
- [2] Dia Bitari Mei Yuana. 2014. *Model Potensi Penyebaran Penyakit Demam Berdarah Dengue Di Kabupaten Jember*. Politeknik Negeri Jember: Tugas Akhir.
- [3] Soegijanto, Soegeng. 2006. *Demam Berdarah Dengue*. Surabaya. Airlangga Unversitasy Press.
- [4] Novelani, Bonita A. 2007. *Studi Habitat dan Perilaku Menggigit Nyamuk Aedes serta Kaitannya dengan Kasus Demam Berdarah di Kelurahan Utan Kayu Utara*. Jakarta Timur, Indonesia: ITB
- [5] Arfiana, F. 2014. *Pengertian Naïve Bayes*. Website: <http://www.google.co.id/gwsrd=cr,ssl&ei=xDdVVelD4uruQTo7Y04Cw#q=pengertian+naive+bayes> -. Diakses pada tanggal 15 Mei 2015 pukul 07.09 WIB.
- [6] Mochammad Choirur Roziqin, Achmad Basuki, Tri Harsono. 2016. *A Prediction System of Dengue Fever Using Monte Carlo Method*. EMITTER International Journal of Engineering Technology Vol. 4, No. 1, June 2016 ISSN: 2443-1168