

Klasifikasi Penderita Penyakit Anemia dengan Metode NBC menggunakan R Programming

Adnan Sauddin

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, adnan.sauddin@uin-alauddin.ac.id

Risnawati Ibtnas

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, risnawati.ibnas@uin-alauddin.ac.id

ABSTRAK, Era data digital menghasilkan data yang sangat besar. Menggali informasi yang termuat dalam data yang demikian sangat sulit jika dilakukan secara manual. Perkembangan aplikasi pengolahan data yang bersifat *opensource* sangat membantu setiap analis data dalam melakukan pekerjaannya. *R Programming* merupakan salah satu aplikasi yang dikembangkan dan penggunaannya sangat luas. Pada artikel ini akan membahas bagaimana memodelkan dataset penderita penyakit anemia yang dimaksudkan untuk memahami tingkat keakuratan diagnosis bahwa seseorang mengidap penyakit anemia atau tidak yang didasarkan pada ciri-ciri sebagaimana dimaksudkan pada dataset. Hasilnya menunjukkan bahwa tingkat keakuratan model adalah 99%

Kata Kunci: *pengangguran, terdidik, pemuda, regresi logistik, Banten*

1. PENDAHULUAN

Anemia merupakan salah satu penyakit yang terkait dengan keadaan gizi dari seseorang yang diakibatkan oleh kurangnya HB dalam darah. Individu yang rentan terserang penyakit ini diantaranya, ibu hamil dan menyusui, remaja [3], [4], [8], [13], [14], [19], [21], [22], [28], [32], [33], [34], [36], [37], [39], [40], [44]).

Klasifikasi penyakit anemia dengan naive bayes adalah untuk meningkatkan kualitas dan efektivitas dalam memprediksi kategori atau kelas dari data penderita anemia. Klasifikasi adalah proses mengelompokkan data berdasarkan karakteristik tertentu, seperti kategori atau kelas. Contohnya adalah mengklasifikasikan penderit penyakit anemia sebagai anemia dan tidak anemia, mengklasifikasikan gambar sebagai manusia atau hewan, atau mengklasifikasikan pasien sebagai pasien dengan penyakit tertentu atau tidak.

Salah satu metode klasifikasi yang banyak digunakan adalah metode naive bayes. Metode ini didasarkan pada teorema Bayes, yaitu sebuah teorema statistik yang digunakan untuk

menghitung probabilitas suatu kejadian berdasarkan probabilitas kejadian terkait. Metode naive bayes memiliki beberapa keunggulan, di antaranya adalah mudah diimplementasikan, cepat dalam menghasilkan prediksi, dan dapat digunakan dalam berbagai jenis data.

Namun, meskipun metode Naive Bayes memiliki banyak keunggulan, terdapat juga beberapa kelemahan yang perlu diperhatikan dalam penelitian klasifikasi [7], [15], [16], [20], [24], [25], [31]. Kelemahan-kelemahan tersebut di antaranya adalah asumsi bahwa variabel-variabel yang digunakan dalam model naive bayes bersifat independen secara statistik, sehingga metode ini tidak cocok untuk data yang memiliki ketergantungan antar variabel. Selain itu, metode naive bayes juga dapat menghasilkan akurasi yang rendah jika data yang digunakan tidak sesuai dengan asumsi-asumsi yang digunakan dalam metode ini [1], [2], [5], [6], [10], [12], [17], [20], [26], [27], [35], [38], [43]. Dalam penelitian klasifikasi dengan Naive Bayes, perlu dilakukan evaluasi yang baik untuk mengetahui seberapa baik model yang digunakan dalam memprediksi kategori atau kelas dari sebuah data. Evaluasi ini dapat dilakukan dengan menggunakan metrik-metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score. Dengan melakukan evaluasi yang baik, dapat diketahui kelemahan dan kelebihan dari model naive bayes dan dapat dilakukan perbaikan dan pengembangan model yang lebih baik dan efektif [7], [9], [11], [16], [18], [23], [24], [29], [30], [31], [41], [42].

2. TINJAUAN PUSTAKA

Naive Bayes

Machine learning telah menjadi skill yang paling populer. Hal yang penting untuk ketahui terkait machine learning adalah algoritma yang

digunakan dalam machine learning. Salah satu algoritma yang populer digunakan adalah Naive Bayes. Naive Bayes merupakan algoritma *Supervised Machine Learning* yang didasarkan pada teorema Bayes yang digunakan untuk menyelesaikan masalah pengklasifikasian dengan pendekatan peluang. Pendekatan tersebut didasarkan pada ide bahwa variabel prediktor dalam model *machine learning* adalah independen dari setiap variabel yang lain. Artinya, bahwa kejadian dari suatu model bergantung pada himpunan variabel independen yang tidak ada hubungannya satu dengan yang lain.

Dalam masalah dunia nyata, variabel prediktor tidak selalu independen satu sama lain, selalu ada korelasi di antara keduanya. Karena Naive Bayes menganggap setiap variabel prediktor tidak tergantung pada variabel lain dalam model, itu disebut 'Naif'.

Prinsip Matematika pada Naive Bayes

Prinsip pada *Naive Bayes* adalah teorema Bayes juga dikenal dengan aturan Bayes. Teorema Bayes digunakan untuk menghitung peluang bersyarat, yang merupakan peluang suatu peristiwa yang terjadi berdasarkan informasi dimasa lalu. Secara matematis, teorema Bayes dituliskan sebagai berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

Dimana,

$P(A|B)$: Peluang bersyarat terjadinya peristiwa A , dimana diketahui peristiwa B telah terjadi

$P(A)$: Peluang terjadinya peristiwa A

$P(B)$: Peluang terjadinya peristiwa B

$P(B|A)$: Peluang bersyarat terjadi peristiwa B , dimana diketahui peristiwa A telah terjadi

Secara formal, terminologi dari teorema Bayesian diuraikan sebagai berikut:

A merupakan preposisi dan B merupakan bukti (*evidence*). $P(A)$ menggambarkan peluang

sebelumnya (prior-Probability) dari preposisi. $P(B)$ menggambarkan peluang sebelumnya (prior-probability) dari bukti (Evidence) $P(A|B)$ disebut posterior (sebelumnya) $P(B|A)$ merupakan likelihood (Peluang atau kemungkinan)

Oleh karena itu, teorema Bayes dapat diringkas sebagai berikut:

$$\begin{aligned} \text{Posterior} &= \frac{(\text{Likelihood}) \times (\text{Proposition prior probability})}{\text{Evidence prior probability}} \\ (2.2) \end{aligned}$$

Dapat juga dituliskan dalam bentuk:

Diberikan suatu hipotesism H dan bukti (evidence) E , Teorema Bayes menyatakan bahwa hubungan antara peluang hipotesis sebelum diperoleh bukti $P(H)$ dan peluang hipotesis setelah mendapat bukti adalah:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (2.3)$$

Pada bagian selanjutnya akan dijelaskan penurunan rumus dari teorema Bayes

Teorema Bayes untuk Algoritma Naive Bayes

Persamaan (2.1) dan (2.3) mempunyai satu variabel prediktor, namun, dalam penerapan duni nyata, terdapat lebih dari satu variabel prediktor dan untuk masalah klasifikasi, terdapat lebih dari satu kelompok. Kelompok-kelompok tersebut digambarkan sebagai, C_1, C_2, \dots, C_k dan variabel prediktor digambarkan sebagai suatu vektor, x_1, x_2, \dots, x_n .

Tujuan algoritma Naive Bayes adalah untuk mengukur peluang bersyarat dari suatu peristiwa dengan suatu fitur vektor x_1, x_2, \dots, x_n yang beada dalam kelompok tertentu dari C_i ,

Dengan mengoperasikan persamaan berikut, diperoleh:

$$\begin{aligned} P(x_1, x_2, \dots, x_n|C_i) &= P(x_1, x_2, \dots, x_n, C_i) \\ &= P(x_1|x_2, \dots, x_n, C_i) \cdot P(x_2, \dots, x_n, C_i) \\ &= P(x_1|x_2, \dots, x_n, C_i) \cdot P(x_2|x_3, \dots, x_n, C_i) \cdot P(x_3, \dots, x_n, C_i) \end{aligned} \quad (2.4)$$

$$P(x_1, x_2, \dots, x_n|C_i) = P(x_1|x_2, \dots, x_n, C_i) \dots P(x_{n-1}|x_n, \dots, x_n, C_i) \cdot P(x_n|C_i) \cdot P(C_i)$$

Namun demikian, peluang bersyarat, $P(x_j|C_i)$ karena setiap variabel prediktor adalah sebagai contoh, independen di Naive Bayes. Persamaan $P(x_j|x_{j+1}, \dots, x_n, C_i)$ dijumlahkan menjadi umumnya menjadi:

$$P(C_i|x_1, x_2, \dots, x_n) = \left(\prod_{j=1}^{j=n} P(x_j|C_i) \right) \cdot \frac{P(C_i)}{P(x_1, x_2, \dots, x_n)}; \quad \text{untuk: } 1 < i < k \quad (2.5)$$

Perhatikan $P(x_1, x_2, \dots, x_n)$ merupakan konstan untuk semua kelompok, oleh karena itu diperoleh

$$P(C_i|x_1, x_2, \dots, x_n) \propto \left(\prod_{j=1}^{j=n} P(x_j|C_i) \right) \cdot P(C_i); \quad \text{untuk: } 1 < i < k \quad (2.6)$$

Pada contoh sederhana berikut akan dijelaskan bagaimana Naive Bayes bekerja? Perhatikan contoh berikut; data yang memuat 1500 pengamatan dengan kelompok:

- Cat
- Parrot
- Turtle

Variabel prediktor merupakan variabel dengan tipe data kategori dengan dua nilai, benar dan salah:

- Swim
- Wings
- Green Color
- Sharp Teeth

Dari tabel di atas, diuraikan sebagai berikut:

Kelompok jenis kucing menunjukkan bahwa:

- dari 500, 450 (90%) kucing bisa berenang
- 0 ekor kucing tidak mempunyai sayap
- 0 ekor kucing bukan berwarna merah
- seluruh kucing- (500) mempunyai gigi yang tajam

Kelas dari jenis burung beo menunjukkan bahwa:

- 50 (10%) dari burung beo mempunyai sayap

- semua (500) burung beo mempunyai sayap
- dari 500, 400 (80%) burung beo berwarna hijau
- tidak ada burung beo yang mempunyai gigi
- Kelompok jenis Penyu menunjukkan:
 - semua (500) penyu dapat berenang
 - tidak ada (0) penyu yang mempunyai sayap
 - dari 500, 100 (20%) penyu berwarna hijau
 - 50 dari 500 (10%) penyu mempunyai gigi yang tajam

Selanjut, dengan data tersebut, kita akan klasifikasi pengamatan berikut ke dalam dua kategori (Kucing, burung beo atau Penyu) dengan menggunakan Naive Bayes Classifier.

Tujuannya adalah untuk memprediksi apakah kucing, burung beo atau penyu yang didefinisikan sebagai variabel prediktor (Berenang, sayap, hijau, gigi yang tajam).

Penyelesaian dari masalah ini adalah, kita akan menggunakan pendekatan Naive Bayes,

$$P(H|Multiple Evidences) = \frac{P(C_1|H) \times P(C_2|H) \times \dots \times P(C_n|H) \times P(H)}{P(Multiple Evidences)} \quad (2.7)$$

Pada pengamatan, variabel berenang dan hijau adalah *true* dan kejadiannya dapat berupa; Cat, Parrot, Turtle. Untuk Kategori kucing:

$$\begin{aligned} P(Cat|Swim, Green) &= \frac{P(Swim|Cat) \times P(Green|Cat) \times P(Cat)}{P(Swim, Green)} \\ &= \frac{0.9 \times 0 \times 0.333}{P(Swim, Green)} \\ &= 0 \end{aligned}$$

Burung Beo:

$$\begin{aligned}
 P(\text{Parrot}|\text{Swim}, \text{Green}) &= \frac{P(\text{Swim}|\text{Parrot}) \times P(\text{Green}|\text{Parrot}) \times P(\text{Parrot})}{P(\text{Swim}, \text{Green})} \\
 &= \frac{0.1 \times 0.80 \times 0.333}{P(\text{Swim}, \text{Green})} \\
 &= \frac{0.0264}{P(\text{Swim}, \text{Green})}
 \end{aligned}$$

Penyu:

$$\begin{aligned}
 P(\text{Turtle}|\text{Swim}, \text{Green}) &= \frac{P(\text{Swim}|\text{Turtle}) \times P(\text{Green}|\text{Turtle}) \times P(\text{Turtle})}{P(\text{Swim}, \text{Green})} \\
 &= \frac{1 \times 0.2 \times 0.333}{P(\text{Swim}, \text{Green})} \\
 &= \frac{0.0666}{P(\text{Swim}, \text{Green})}
 \end{aligned}$$

Dari semua hasil perhitungan di atas penyebutnya sama, sebagai contoh- $P(\text{Swim}, \text{Green})$. Nilai dari $P(\text{Penyu}|\text{Berenang}, \text{Hijau})$ lebih besar dari $P(\text{Burung Beo}|\text{Berenang}, \text{Hijau})$, oleh karena itu kita dapat secara tepat memprediksi kelas dari binatang sebagai penyu.

Anemia pada Remaja dan Faktor Penyebabnya

Anemia adalah kondisi dimana jumlah sel darah merah dalam tubuh berkurang atau jumlah sel darah merah berada di bawah batas normal. Ini karena kekurangan hemoglobin (protein kaya zat besi), yang memengaruhi produksi sel darah merah. Hal ini membuat oksigen sulit mencapai sel dan jaringan tubuh. Masalah kesehatan atau penyakit anak muda, termasuk anemia, seringkali membuat khawatir para orang tua. Selain itu, anak anemia tampak lebih mudah lelah dan lesu. Namun, mengutip publikasi Anak Sehat, disebutkan bahwa pertumbuhan yang cepat adalah penyebab utama anemia pada orang muda. Pada usia ini anak-anak paling rentan mengalami anemia.

Beberapa orang dengan anemia tidak menunjukkan tanda atau gejala. Namun, anak-anak mungkin mengalami gejala anemia berikut:

1. Kulit tampak pucat.
2. mengalami perubahan suasana hati.
3. Terlihat sangat lelah.
4. Kepala terasa sangat pusing.
5. Detak jantung lebih cepat dari biasanya.
6. Penyakit kuning (kulit dan mata menguning).

Ketika remaja mengalami anemia berat, mereka mungkin juga mengalami tanda dan gejala lain, termasuk:

1. Sesak napas.
2. Tangan dan kaki bengkak.
3. Pusing disertai sakit kepala.
4. Memiliki sindrom kaki gelisah.

Berikut ini adalah penyebab paling umum dari anemia:

1. Kurang Asupan Zat besi; Kurangnya penyerapan zat besi Penyebab paling umum dari anemia pada anak-anak adalah kekurangan zat besi, baik dari makanan maupun dari suplemen makanan. Apalagi anemia lebih sering terjadi pada wanita muda karena mereka merasakan awal menstruasi saat pubertas. Perlu diketahui bahwa kebutuhan zat besi pada masa remaja adalah sekitar 8-15 mg per hari. Itu sebabnya Anda juga perlu berhati-hati saat remaja mulai menurunkan berat badan. Berikan perhatian khusus pada makanan dan diet, karena ini juga dapat memicu anemia.
2. Anemia akibat perdarahan; Ini bisa disebabkan oleh pendarahan akibat cedera, menstruasi yang berat, gangguan pencernaan, dan masalah kesehatan lainnya. Oleh karena itu, anemia remaja lebih sering terjadi pada anak perempuan karena mereka mengalami menstruasi setiap bulan.
3. Sel darah merah yang rusak; Ini adalah kondisi yang juga bisa disebut anemia hemolitik. Kondisi ini terjadi ketika

sistem kekebalan tubuh menghancurkan sel darah merah sendiri. Perhatikan bahwa ini juga jenis yang disebabkan oleh kelainan sel darah merah keturunan. Misalnya, anemia sel sabit juga merupakan talasemia.

4. Produksi sel darah merah terlalu lambat; Ada beberapa hal yang menjadi penyebab anemia pada remaja ini, seperti: Anemia aplastik, ketika tubuh berhenti memproduksi sel darah merah yang disebabkan oleh infeksi atau penyakit. Kekurangan vitamin B12 dari makanan, suplemen makanan, sehingga tubuh tidak dapat menyerap vitamin ini.

3. METODOLOGI

Data dan Referensi

Tahapan Klasifikasi Penderita Anemia

- a. Berikut adalah tahapan klasifikasi penyakit anemia dengan metode naive bayes menggunakan R Programming:
- b. Persiapan data
 - Mengumpulkan data anemia yang akan digunakan dalam analisis.
 - Memeriksa data untuk memastikan konsistensi dan integritas data.
- c. Pra-pemrosesan data
 - Menghapus nilai yang hilang atau missing values.
 - Mengubah data kategorikal ke dalam format numerik.
- d. Pembagian data
 - Membagi data menjadi dua set, yaitu set latihan dan set uji.
 - Set latihan digunakan untuk melatih model dan set uji digunakan untuk menguji performa model.
- e. Pelatihan model
 - Membuat model naive bayes dengan menggunakan fungsi 'naiveBayes' pada package 'e1071'.
 - Melatih model dengan menggunakan set latihan.
- f. Pengujian model
 - Menggunakan set uji untuk menguji performa model.
 - Menghitung akurasi model dengan membandingkan hasil prediksi dengan nilai aktual.

- g. Evaluasi model
 - Menggunakan beberapa metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score untuk mengevaluasi performa model.
 - Memperbaiki model jika diperlukan dan mengulang tahap pelatihan dan pengujian model.

4. PEMBAHASAN

Hasil Penelitian

Preparation Data

```
#describe(data)
```

Langkah 4: Data Cleaning

Ketika menganalisis struktur dataset, kita periksa bahwa nilai minimum untuk Glukosa, Tekanan Darah, Skinthickness, Insulin, and BMI semuanya nol. Hal ini bukan tidak masuk kerangka konsep terkait dengan nilai Glucose, blood pressure, etc karena unsur-unsur tersebut tidak mungkin bernilai nol. Karena itu kita akan lakukan treatment terhadap pengamatan 'missing'.

Kode berikut kita gunakan untuk menetapkan nilai nol kenilai NA:

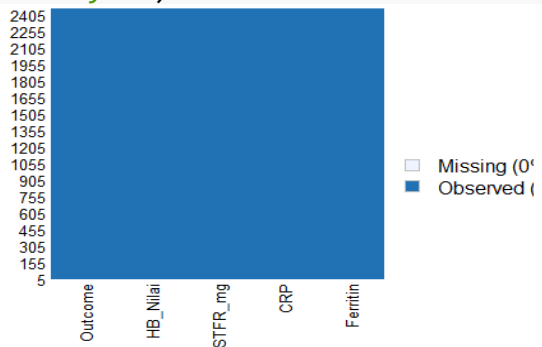
Konversi Nilai 0 sebagai nilai NA

```
anemia[, 1:3][anemia[,1:3] == 0] <- NA
```

Bagaimana cara memeriksa nilai missing, kita visualisasi data:

Visualisasi data missing

```
missmap(anemia,main="Plot Data Missing - Naive Bayes")
```



Gambar 4.1 Plot Data Missing-Naive Baye

Grafik diatas menunjukkan bahwa dataset memuat banyak nilai missing. Hilangkan semua nilai missing, gunakan paket mice di R.

Menggunakan mice package dalam memprediksi nilai missing

```
mice_mod <- mice(anemia[, c("Ferritin",
"CRP", "STFR_mg")], method='rf')
##
##          iter      imp    variable
##          1         1         1
##          1         1         2
##          1         1         3
##          1         1         4
##          1         1         5
##          2         2         1
##          2         2         2
##          2         2         3
##          2         2         4
##          2         2         5
##          3         3         1
##          3         3         2
##          3         3         3
##          3         3         4
##          3         3         5
##          4         4         1
##          4         4         2
##          4         4         3
##          4         4         4
##          4         4         5
##          5         5         1
##          5         5         2
##          5         5         3
##          5         5         4
##          5         5
mice_complete <- complete(mice_mod)
```

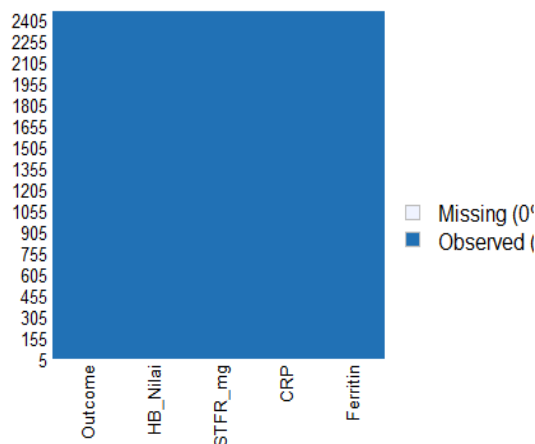
Transfer nilai missing yang diprediksi ke dalam kumpulan data utama

```
anemia$Ferritin <- mice_complete$Ferritin
anemia$CRP <- mice_complete$CRP
anemia$STFR_mg <- mice_complete$STFR_mg
```

Untuk mengecek apakah masih ada nilai yang missing, gunakan plot *missmap*:

```
missmap(anemia, main="Menggunakan Mice Package dalam R - Naive Bayes")
```

Output pada gambar 4.2 terlihat baik, tidak data yang *missing*



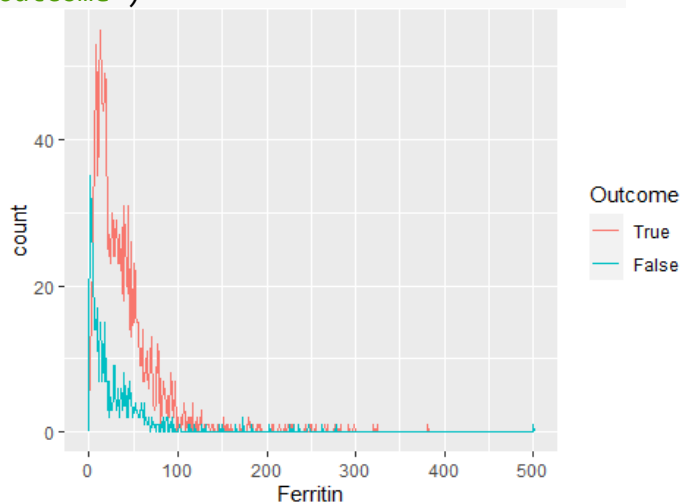
Gambar 4.2 Hasil penggunaan package Mice dalam R

Step 5: Exploratory Data Analysis

Selanjutnya dilakukan eksplorasi data dengan visualisasi secara berpasangan untuk melihat keadaan setiap variabel dengan lebih baik, bagian ini merupakan tahapan penting untuk memahami signifikansi dari setiap variabel predictor.

Data Visualization

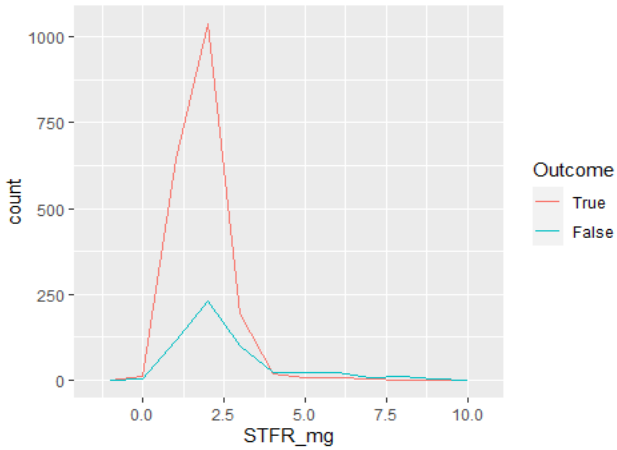
```
Visual 1
par(mfrow=c(1,2))
ggplot(anemia, aes(x = Ferritin, color = Outcome))
geom_freqpoly(binwidth = 1)
labs(title="Age Distribution by Outcome")
```



Gambar 4.3 Plot Distribusi Usia untuk visual 1

```
visual 2
par(mfrow=c(1,2))
ggplot(anemia, aes(x = STFR_mg, color = Outcome))
```

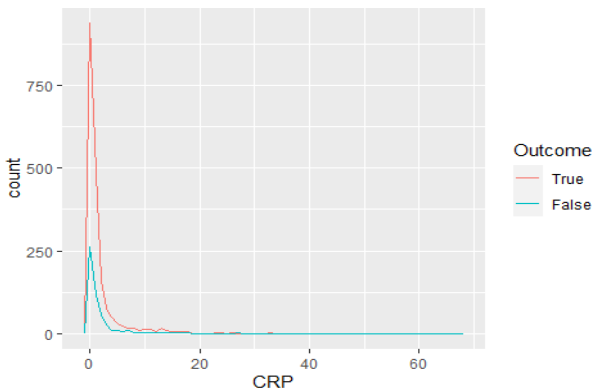
```
geom_freqpoly(binwidth = 1) +
labs(title="Age Distribution by Outcome")
```



Gambar 4.4 Plot Distribusi Usia untuk visual 2

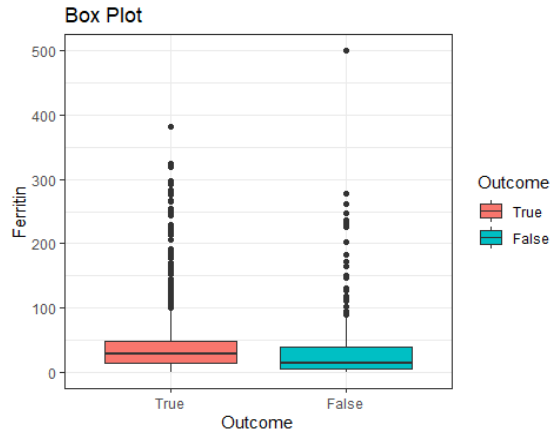
visual 3

```
par(mfrow=c(1,2))
ggplot(anemia, aes(x = CRP, color = Outcome)) +
geom_freqpoly(binwidth = 1) +
labs(title="Age Distribution by Outcome")
```



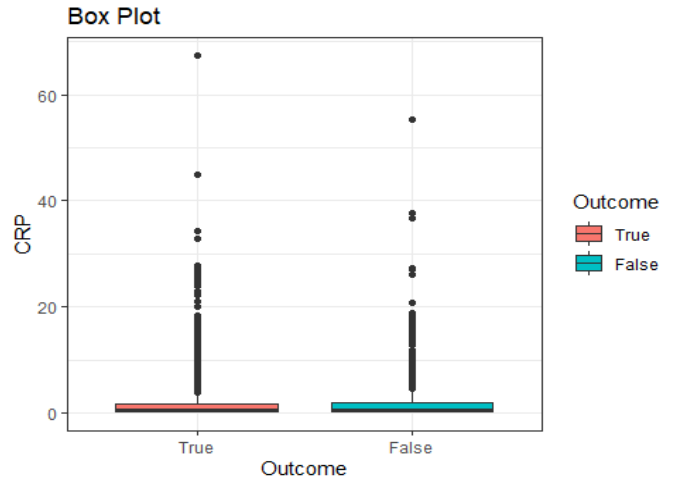
Gambar 4.5 Plot Distribusi Usia untuk visual 3

```
ggplot(anemia, aes(x=Outcome, y=Ferritin, fill = Outcome)) +
geom_boxplot() +
theme_bw()+
ggtitle("Box Plot")
```



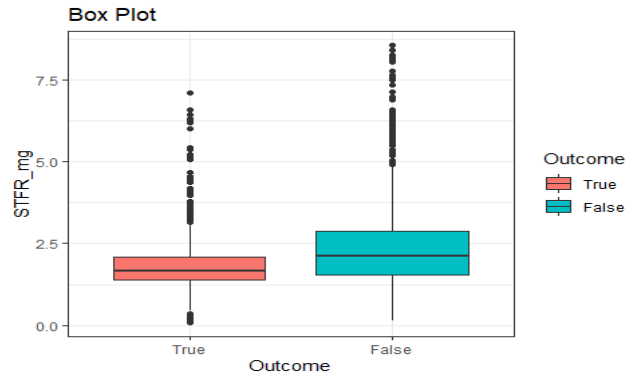
Gambar 4.6 Box Plot 1

```
ggplot(anemia, aes(x=Outcome, y=CRP, fill = Outcome)) +
geom_boxplot() +
theme_bw()+
ggtitle("Box Plot")
```



Gambar 4.7 Box Plot 2

```
ggplot(anemia, aes(x=Outcome, y=STFR_mg, fill = Outcome)) +
geom_boxplot() +
theme_bw()+
ggtitle("Box Plot")
```



Gambar 4.8 Box Plot 3

Step 6: Data Modelling

Pada tahapan ini kita mulai dengan membagi data menjadi dua bagian, yaitu data training dan data test:

Data training: merupakan bagian data yang digunakan untuk mengembangkan modle dan melakukan pelatihan model. Data testing merupakan bagian data yang akan digunakan untuk mengevaluasi efisiensi model.

Membangun model

- split data into training and test data sets

```
indxTrain <- createDataPartition(y = anemia$Outcome, p = 0.75, list = FALSE)
training <- anemia[indxTrain,]
testing <- anemia[-indxTrain,]
#Check dimensions of the split >
prop.table(table(data$Outcome)) * 100
prop.table(table(training$Outcome)) * 100
##
##      True      False
## 77.94595 22.05405
prop.table(table(testing$Outcome)) * 100
##
##      True      False
## 78.04878 21.95122
```

Membandingkan hasil dari training dan pengujian, kita buat variabel terpisah yang menyimpan nilai dari variabel respona:

```
x = training[, -9]
y = training$Outcome
selanjutnya aktifkan paket yang memuat fungsi Naive Bayes.
```

```
library(e1071)
```

setelah mengaktifkan paket “e1071”, kode berikut akan membuat model Naive Bayes dengan menggunakan data training

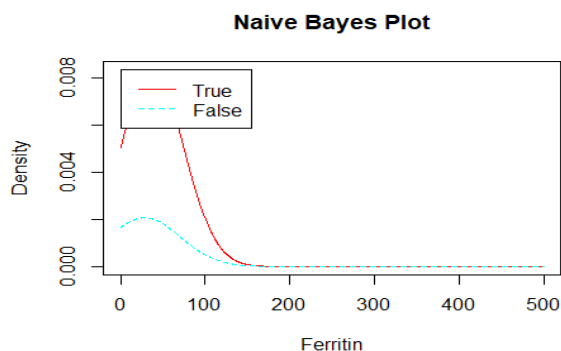
```
model = train(x,y,'nb',trControl=trainControl(method='cv',number=10))
```

Selanjutnya membuat model prediksi dengan menggunakan Naive Bayes Classifier.

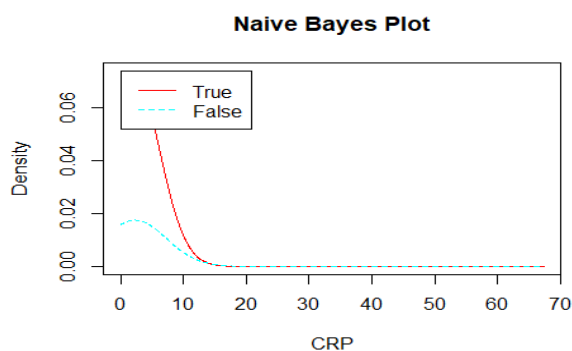
Step 7: Model Evaluation

Memeriksa efisiensi metode, dilakukan pengujian dataset test pada model sebelumnya, setelah itu kita akan mengevaluasi akurasi dari model dengan menggunakan Confusion matrix.

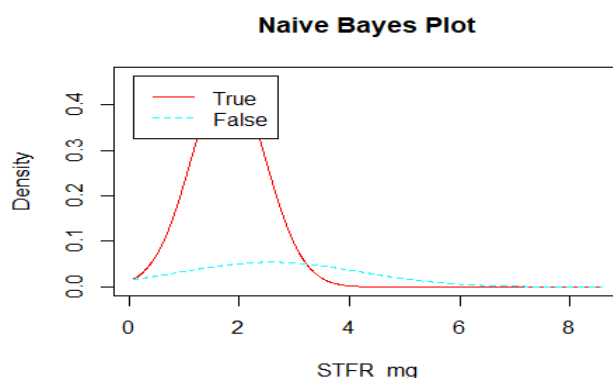
```
mN <- NaiveBayes(Outcome ~ ., data = anemia)
plot(mN)
```



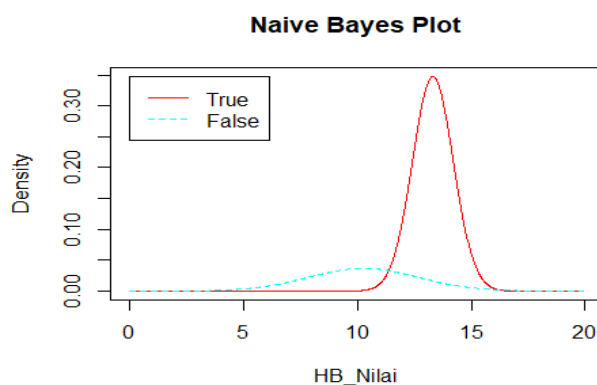
Gambar 4.9 Plot Ferritin-Density untuk model mN



Gambar 4.10 Plot CRP-Density model mN

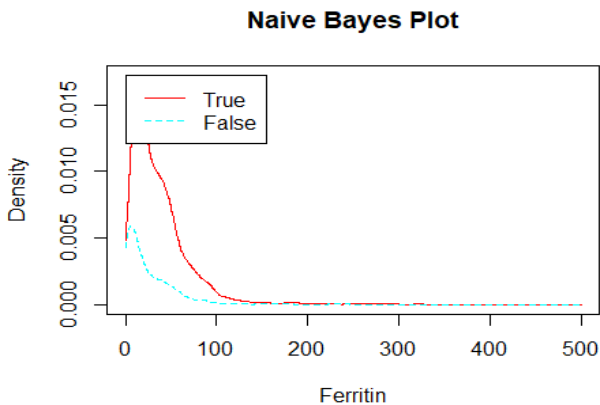


Gambar 4.11 Plot STFR_mg-Density model mN

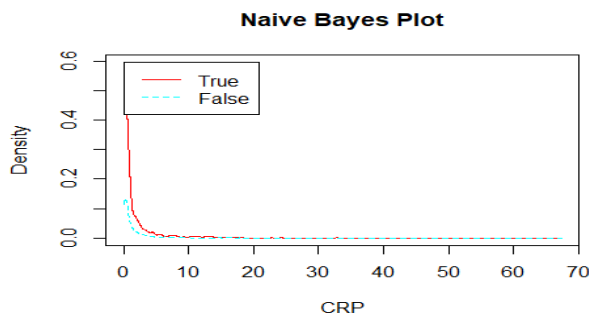


Gambar 4.12 Plot Nilai HB-Density model mN

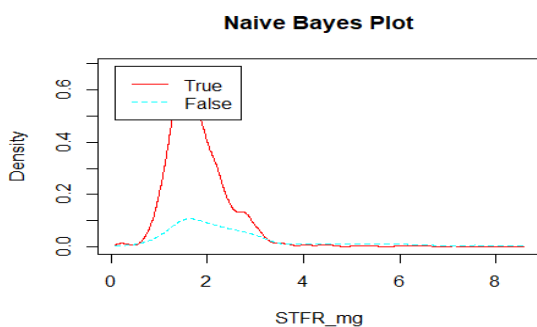

```
mK <- NaiveBayes(Outcome ~ ., data = anemia, usekernel = TRUE)
plot(mK)
```



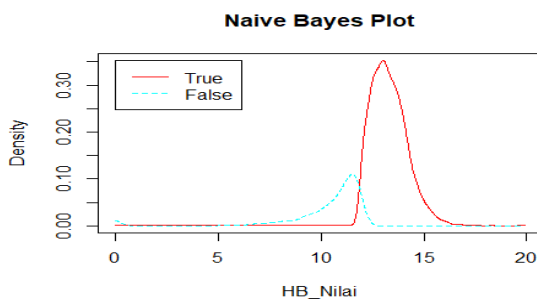
Gambar 4.13 Plot Nilai Ferritin-Density untuk model mK



Gambar 4.14 Plot Nilai CRP-Density untuk model mK



Gambar 4.15 Plot Nilai STFR_mg-Density untuk model mK



Gambar 4.16 Plot Nilai Nilai HB-Density untuk model mK

Model Evaluation - Predict testing set

```
confusionMatrix(Predict,
testing$Outcome )
## Confusion Matrix and Statistics
##
##           Reference
## Prediction True False
##      True    479     0
##      False     1    135
##
##              Accuracy : 0.9984
##              95% CI : (0.991,
1)
##      No Information Rate : 0.7805
##      P-Value [Acc > NIR] : <2e-16
##
##              Kappa : 0.9953
##
##      Mcnemar's Test P-Value : 1
##
##              Sensitivity : 0.9979
##              Specificity : 1.0000
##              Pos Pred Value : 1.0000
##              Neg Pred Value : 0.9926
##              Prevalence : 0.7805
##              Detection Rate : 0.7789
##              Detection Prevalence : 0.7789
##              Balanced Accuracy : 0.9990
##
##              'Positive' Class : True
##
```

Output diatas menunjukkan bahwa kita mengembangkan Naive Bayes classifier yang dapat memprediksi apakah seseorang mengidap anemia atau tidak, dengan akurasi sekitar 99%.

5. KESIMPULAN DAN SARAN

Bahwa untuk melakukan analisis data dengan metode Naive Bayes Classification dengan bantuan *R Programming* tahapannya sebagaimana dijelaskan pada bagian hasil. Adapun rangkuman kode R yang dapat digunakan adalah

```
# memuat library e1071
library(e1071)
```

1. Membaca file data

```
anemia <- read.csv("anemia.csv",
header = TRUE)
```
2. Membagi data menjadi set latihan dan set uji

```
set.seed(123)
```

```

train.index <-
sample(1:nrow(anemia), 0.7 *
nrow(anemia))
train <- anemia[train.index, ]
test <- anemia[-train.index, ]

```

3. Membuat model Naive Bayes

```
nb <- naiveBayes(Status ~ ., data
= train)
```
4. Melakukan prediksi pada set uji

```
predictions <- predict(nb,
test[, -9])
```
5. Menghitung akurasi model

```
accuracy <- sum(predictions ==
test$Status) / nrow(test)
cat("Akurasi model:", accuracy,
"\n")
```
6. Menghitung metrik evaluasi

```
library(caret)
confusionMatrix(predictions,
test$Status)
```

6. DAFTAR PUSTAKA

- [1] Alrawashdeh, M. J. (2022). Improved Naive Bayes Classification for Joint Investment Plan. *WSEAS Transactions on Mathematics*, 21. <https://doi.org/10.37394/23206.2022.21.6>
- [2] Bradley, E. (1968). *Multiple Classification Arbitrary Experimental Analysis for Arrangements*. 10(1), 13–27.
- [3] Cao, G., Wang, Y., Wu, Y., Jing, W., Liu, J., & Liu, M. (2022). Prevalence of anemia among people living with HIV: A systematic review and meta-analysis. *EclinicalMedicine*, 44. <https://doi.org/10.1016/j.eclinm.2022.101283>
- [4] Chaparro, C. M., & Suchdev, P. S. (2019). Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. In *Annals of the New York Academy of Sciences* (Vol. 1450, Issue 1). <https://doi.org/10.1111/nyas.14092>
- [5] Chhikara, R. S., Register, D. T., & Company, L. E. (1979). *Classification Method for Partitioning of a Large Multidimensional Mixed Data Set*. 2(4).
- [6] Classification, C. (1964). *Notes Cross Classification*. 6(3).
- [7] Cox, D. R. (1960). Serial Sampling Acceptance Schemes Derived From Bayes's Theorem. *Technometrics*, 2(3), 353–360. <https://doi.org/10.1080/00401706.1960.10489902>
- [8] Currie, A. R., Cockerill, D., Diez-Padrisa, M., Haining, H., Henriquez, F. L., & Quinn, B. (2022). Anemia in salmon aquaculture: Scotland as a case study. In *Aquaculture* (Vol. 546). <https://doi.org/10.1016/j.aquaculture.2021.737313>
- [9] Diggle, P., & Lophaven, S. (2006). Bayesian Geostatistical Design. *Scandinavian Journal of Statistics*, 33(1), 53–64. <https://doi.org/10.1111/j.1467-9469.2005.00469.x>
- [10] Donnellan, E., Aslan, S., Fastrich, G. M., & Murayama, K. (2022). How Are Curiosity and Interest Different? Naive Bayes Classification of People's Beliefs. In *Educational Psychology Review* (Vol. 34, Issue 1). <https://doi.org/10.1007/s10648-021-09622-9>
- [11] Draper, N. (1971). *Bayesian Estimation of the Binomial Parameter*. 13(3), 667–673.
- [12] Dunn, J. (1971). *Some Expected Values for Probabilities of Correct Classification in Discriminant Analysis*. 13(2).
- [13] Fernández-Plaza, S., & Gómez, S. V. (2021). Anemia ferropénica. *Pediatrica Integral*, 25(5).
- [14] Fitriany, J., & Saputri, A. I. (2018). ANEMIA DEFISIENSI BESI. *AVERROUS: Jurnal Kedokteran Dan Kesehatan Malikussaleh*, 4(2). <https://doi.org/10.29103/averrous.v4i2.1033>
- [15] Foo, L. K., Chua, S. L., & Ibrahim, N. (2022). Attribute weighted naïve bayes classifier. *Computers, Materials and Continua*, 71(1). <https://doi.org/10.32604/cmc.2022.022011>
- [16] Gaver, P., & Postgraduate, N. (1987). *Robust Empirical Event Rates Bayes Analyses of*. 1979.
- [17] Goodman, L. E. O. A. (1971). *The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications*. 13(1), 33–61.

- [18] Handcock, M. S. (1993). *A Bayesian Analysis of Kriging*. 1989.
- [19] Igbinsola, I., Berube, C., & Lyell, D. J. (2022). Iron deficiency anemia in pregnancy. In *Current Opinion in Obstetrics and Gynecology* (Vol. 34, Issue 2).
<https://doi.org/10.1097/GCO.00000000000000772>
- [20] Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). Naive Bayes Classification. In *Bayes Rules!*
<https://doi.org/10.1201/9780429288340-14>
- [21] Kędziora-Kornatowska, K., Mađra-Gackowska, K., & Gackowski, M. (2019). Anemia. In *Encyclopedia of Biomedical Gerontology* (pp. 222–228). Elsevier.
<https://doi.org/10.1016/B978-0-12-801238-3.11352-2>
- [22] Kurkina, N. v., Gorshenina, E. I., Chegodaeva, L. v., & Polagimova, A. v. (2021). Anemia of Chronic Diseases. *Klinicheskaya Onkogematologiya/Clinical Oncohematology*, 14(3).
<https://doi.org/10.21320/2500-2139-2021-14-3-347-354>
- [23] Laboratories, S. N. (1987). *Bayesian Analysis Risk Analyses of Learning*. 29(2).
- [24] Lamotte, R. (1978). *Bayes Linear Estimators*. 20(3).
- [25] Limits, B. C., & Density, P. (1968). *Bayesian Confidence Limits for Reliability of Bayesian Confidence Limits for Reliability Terminated at First Failure*. 10(1).
- [26] Link, R. F. (1965). *Short-Cut Multiple Comparisons for Balanced s in gl e and Double Classifications : Part I , Results **. 7(2).
- [27] Low, L. Y. (1976). *Some Properties of Variance Component Estimators in the Two-Way Classification*. 18(1).
- [28] Manrai, M., Dawra, S., Srivastava, S., Kapoor, R., & Singh, A. (2022). Anemia in cirrhosis: An underestimated entity. *World Journal of Clinical Cases*, 10(3).
<https://doi.org/10.12998/wjcc.v10.i3.777>
- [29] Mazzuchi, A. (1998). *Bayesian Computations for a Class of Reliability Growth Models*. 40(1).
- [30] Miller, B. (1980). *Bayesian Analysis of the Two-Parameter Distribution*. 22(March 1979).
- [31] Miller, R. W. (1989). *Parametric Empirical Tolerance Intervals Bayes*. 1964.
- [32] Montañó-Figueroa, E., & Gómez-Almaguer, D. (2022). Anemia aplásica. *Gaceta de México*, 157(93).
<https://doi.org/10.24875/gmm.m21000472>
- [33] Montañó-Figueroa, E. H., & Gómez-Almaguer, D. (2021). Aplastic anemia. *Gaceta Medica de Mexico*, 157.
<https://doi.org/10.24875/GMM.M21000472>
- [34] Natekar, P., Deshmukh, C., Limaye, D., Ramanathan, V., & Pawar, A. (2022). A micro review of a nutritional public health challenge: Iron deficiency anemia in India. In *Clinical Epidemiology and Global Health* (Vol. 14).
<https://doi.org/10.1016/j.cegh.2022.100992>
- [35] Pertiwi, M. W., Kusmira, M., Rezkiani, R., Simpony, B. K., Apriyani, Y., Iskandar, I. D., Wibisono, T., & Amirulloh, I. (2022). Naïve Bayes Classification Model for the Producer Price Index Prediction. *SISTEMASI*, 11(1).
<https://doi.org/10.32520/stmsi.v11i1.1669>
- [36] Samson, K. L. I., Fischer, J. A. J., & Roche, M. L. (2022). Iron Status, Anemia, and Iron Interventions and Their Associations with Cognitive and Academic Performance in Adolescents: A Systematic Review. In *Nutrients* (Vol. 14, Issue 1).
<https://doi.org/10.3390/nu14010224>
- [37] Shah, T., Khaskheli, M. S., Ansari, S., Lakhani, H., Shaikh, F., Zardari, A. A., Warsi, J., Rind, N. A., Rind, K. H., & Shar, A. H. (2022). Gestational Anemia and its effects on neonatal outcome, in the population of Hyderabad, Sindh, Pakistan. *Saudi Journal of Biological Sciences*, 29(1).
<https://doi.org/10.1016/j.sjbs.2021.08.053>

- [38] Snapinn, S. M., & Knoke, J. D. (1985). *An Evaluation of Smoothed Classification Error - Rate Estimators*. May.
- [39] Stanley, A. Y., Wallace, J. B., Hernandez, A. M., & Spell, J. L. (2022). Anemia in pregnancy: Screening and clinical management strategies. *MCN The American Journal of Maternal/Child Nursing*, 47(1). <https://doi.org/10.1097/NMC.00000000000000787>
- [40] Tao, Z., Xu, J., Chen, W., Yang, Z., Xu, X., Liu, L., Chen, R., Xie, J., Liu, M., Wu, J., Wang, H., & Liu, J. (2021). Anemia is associated with severe illness in COVID-19: A retrospective cohort study. *Journal of Medical Virology*, 93(3), 1478–1488. <https://doi.org/10.1002/jmv.26444>
- [41] Techniques, E. A. (1996). *Bayesian Analysis of Ordered Categorical Data From Industrial Experiments*. 38(1).
- [42] Thompson, W. E., & Chang, E. Y. (1975). *Bayes Confidence Limits for Reliability of Redundant Systems Mkz (%)*. 17(1), 89–93.
- [43] Van Herwerden, D., O'Brien, J. W., Choi, P. M., Thomas, K. v., Schoenmakers, P. J., & Samanipour, S. (2022). Naive Bayes classification model for isotopologue detection in LC-HRMS data. *Chemometrics and Intelligent Laboratory Systems*, 223. <https://doi.org/10.1016/j.chemolab.2022.104515>
- [44] Weiss, G., Ganz, T., & Goodnough, L. T. (2019). Anemia of inflammation. In *Blood* (Vol. 133, Issue 1). <https://doi.org/10.1182/blood-2018-06-856500>