JOINED JOURNAL
JOURNAL OF INFORMATICS EDUCATION

Analisis Faktor-Faktor Penentuan Jenis Persalinan Menggunakan Metode *K-Nn* Dan *Naive Bayes*

Kristianto Oggie Ibrahim[1], Diwahana Mutiara Candrasari Hermanto[2], Rosa Ratri Kusuma Hariningsih[3]

Prodi Pendidikan Informatika, Sekolah Tinggi Ilmu Komputer Yos Sudarso Purwokerto, Indonesia

Info Articles

Keywords:
Cross Validation; Confusion Matrix; K-Nearest Neighbor; Naïve Bayes

Abstrak

Tujuan dari penelitian ini adalah mengklasifikasikan jenis persalinan normal atau *caesar* menggunakan algoritma *K-NN* dan *Naïve Bayes*. Data yang digunakan adalah data tentang usia ibu, *hemoglobin*, usia kehamilan, masalah kehamilan. Hasil dari penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan jenis persalinan normal atau caesar. Dengan metode validasi *K-Fold Cross Validation* dan *Confusion Matrix* untuk perhitungan missclassifikasi dimana akan membuat kita mengetahui seberapa buruk model kita dalam melakukan prediksi, untuk menghitung tingkat akurasi dari metode klasifikasi yang kita gunakan. Hasil menunjukan bahwa Dalam perhitungan validasi *Cross Validation* didapat nilai presentasi akurasi sebesar 63,57% untuk data training dan 95,48% untuk data testing dan pada *Confusion Matrix* didapat nilai presentasi akurasi sebesar 66,2% untuk data training dan 96,18% untuk data testing.

Abstract

The purpose of this research is to classify the type of normal or caesarean delivery using the K-NN and Naïve Bayes algorithms. The data used are data on maternal age, hemoglobin, gestational age, pregnancy problems. The results of this study show that the Naïve Bayes algorithm is able to classify the type of normal or cesarean delivery. With the validation method K-Fold Cross Validation and Confusion Matrix for the calculation of misclassification which will make us know how bad our model is in making predictions, to calculate the accuracy level of the classification method we use. The results show that in the calculation of Cross Validation validation, the accuracy presentation value is 63.57% for training data and 95.48% for testing data and in Confusion Matrix, the accuracy presentation value is 66.2% for training data and 96.18% for testing data.

✉ Alamat Korespondensi:
E-mail: candrasari5860@stikomyos.ac.id

p-ISSN 2621-9484
e-ISSN 2620-8415

PENDAHULUAN

Rumah Sakit adalah institusi pelayanan kesehatan bagi masyarakat dengan karateristik tersendiri yang dipengaruhi oleh perkembangan ilmu pengetahuan kesehatan, kemajuan teknologi, dan kehidupan sosial ekonomi masyarakat yang harus tetap mampu meningkatkan pelayanan yang lebih bermutu dan terjangkau oleh masyarakat agar terwujud derajat kesehatan yang setinggi-tingginya. Rumah Sakit Bunda Purwokerto adalah rumah sakit yang terkenal dengan rumah sakit persalinan yang sangat baik di kota Purwokerto dengan biayanya yang terjangkau dengan pelayanan yang sangat ramah. Rumah sakit Bunda terkenal dengan rumah sakit bersalin dengan adanya alat-alat bersalin yang lengkap selain itu banyak dokter anak dan dokter persalinan yang sangat terkenal di kota purwokerto.Tidak hanya layanan yang di berikan selain pemeriksaan kehamilan dan persalinan tersedia juga pelayanan psikologi untuk ibu dan anak.(Ilyas, 2017)

Ibu hamil melakukan persalinan normal/*caesar* ditentukan oleh indikator yang tersedia agar keselamatan ibu dan bayi tidak terancam. Persalinan diartikan sebagai proses pengeluaran hasil konsepsi atau yang biasa kita sebut sebagai janin atau kandungan. Umumnya seorang ibu akan merasa bahagia dan senang sebelum proses persalinan setelah penantian panjang. Sebagian akan merasa takut dan gelisah, baik senang maupun gelisah hal tersebut merupakan hal yang normal setelah seorang ibu mengandung 9 bulan.

Proses persalinan juga menjadi proses yang melelahkan, baik bagi sang ibu maupun sang ayah karena diperlukan kesabaran dalam menjalani prosesnya. Ada banyak hal yang harus diketahui dan dilakukan untuk memastikan bahwa sang ibu dan si kecil berada dalam kondisi sehat sebelum dan setelah persalinan. Tak hanya itu saja, metode persalinan juga harus diketahui agar ibu bisa mempersiapkan segala hal dengan baik nantinya. Persalinan merupakan titik tertinggi dari seluruh persiapan yang telah dipersiapkan. Beberapa wanita akan menyambut persalinan dengan gembira. Kebanyakan wanita ingin melahirkan / melakukan persalinan dengan normal hanya saja banyak hal yang dapat menjadi halangan untuk bisa melahirkan normal. dengan demikian saya membuat proposal ini untuk membantu penentuan persalinan normal atau *Caesar* sesuai dengan indikator-indikator yang sudah ada dengan menggunakan metode *K-NN* dan *Naïve bayes.*

Dengan metode algoritma *K-NN* dan *Naive Bayes* kita dapat memprediksi proses persalinan juga menjadi proses yang melelahkan, baik bagi sang ibu maupun sang ayah karena diperlukan kesabaran dalam menjalani prosesnya. Ada banyak hal yang harus diketahui dan dilakukan untuk memastikan bahwa sang ibu dan si kecil berada dalam kondisi sehat sebelum dan setelah persalinan. Tak hanya itu saja, metode persalinan juga harus diketahui agar ibu bisa mempersiapkan segala hal dengan baik nantinya. Persalinan merupakan titik tertinggi dari seluruh persiapan yang telah dipersiapkan. Beberapa wanita akan menyambut persalinan dengan gembira. Kebanyakan wanita ingin melahirkan / melakukan persalinan dengan normal hanya saja banyak hal yang dapat menjadi halangan untuk bisa melahirkan normal. dengan demikian saya membuat proposal ini untuk membantu penentuan persalinan normal atau *Caesar* sesuai dengan indikator-indikator yang sudah ada dengan menggunakan metode *K-*

NN dan *Naïve bayes.*penentuan persalinan jenis apa pada ibu hamil melalui indikator-indikator yang tersedia agar hasil yang di peroleh lebih akurat. Algoritma *Naive Bayes* merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. Ciri utama dr *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.(Nurfalinda, Nola Ritha, n.d.)

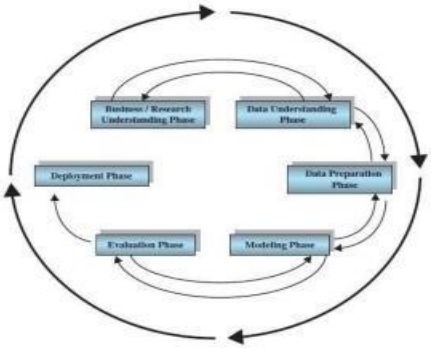
METODE

1. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD). (Telaumbanua et al., 2019)

2. *CRIPS-DM*

CRISP- DM menyediakan standar proses untuk *data mining* yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada bisnis atau pada unit peneltian. CRISP-DM membandingkan metodologi *data mining* lain lebih lengkap dan terdokumentasi dengan baik. Setiap fase terstruktur dan terdefinisi dengan jelas sehingga mudah diaplikasikan bahkan bagi pemula sekalipun. (Feblian & Daihani, 2017)



Gambar 1 Crisp-DM

3. Klasifikasi

Metode klasifikasi merupakan sebuah proses untuk menentukan model yang menjelaskan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Beberapa metode klasifikasi yang sering di gunakan dalam kehidupan sehari-hari diantaranya adalah *K-NN, ID3,* Algoritma *C4.5*, *Naive bayes*, *Neural network* dan lain sebagainya. (Annur, 2018)

4. DBMS (Database Management System)

Sistem pemrosesan basis data dimaksudkan untuk mengatasi kelemahan- kelemahan yang ada pada system pemrosesan berkas. Sistem ini dikenaldengan sebutan DBMS. (Setyawati, E., Wijoyo, H, & Soeharmoko, 2020)

5. Metode *K-NN*

K-NN (*K-Nearest Neighbor*) adalah metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor)terdekatnya dalam data pelatihan. (Bode, 2017)

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \cdots \cdots \cdots (1)$$

Keterangan:

x = sampel data

y = data uji

D = jarak

$$fknn(x') = \frac{1}{K}\sum_{i \in N_k(x^F)} y_i \cdots \cdots \cdots (2)$$

(x') = perkiraan atau estimasi

K = jumlah tetangga terdekat

$N_k(x')$ = tetangga terdekat

y_i = output tetangga terdekat

6. Metode *Naïve Bayes*

Bayesian Classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decesion tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasai dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode *Bayes* merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema *Bayes*, kemudian menggunkan teorema ini untuk melakukan klasifikasi dalam *Data Mining*.(Nurfalinda, Nola Ritha, n.d.)

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \cdots \cdots \cdots (3)$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X meruapakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (posterior prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

7. K-fold Cross Validation

K-Fold Cross Validation menurut Anam & Santoso (2018) adalah teknik untuk

memvalidasi nilai akurasi metode yang diterapkan berdasarkan data set.Menyatakan bahwa metode *K-Fold Cross Validation* membagi himpunan data D secara acak menjadi *k-fold* (sub himpunan) yang saling bebas f_1, f_2, ...,f_k, sehingga setiap *fold* berisi $1/k$ bagian data. Selanjutnya dapat membangun himpunan data: D_1, D_2, ..., D_k, yang masing-masing berisi (k-1) *fold* untuk *training* data, 1-fold untuk *testing* data. Pada umumnya metode *K-Fold Cross Validation* menggunakan 10 kali iterasi (k=10) dengan tujuan untuk memperoleh akurasi dengan bias dan variansi yang cukup rendah. (Etriyanti et al., 2020)

8. Confusion Matrix

Tujuan *Confusion Matrix* menganalisa kualitas kinerja model klasifikasi dalam mengenali variabel dari seluruh kelas. *Confussion Matrix* berisi informasi mengenai kelas sebenarnya dan kelas prediksi dari suatu proses klasifikasi

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) *Corect result*	FP (False Positive) *Unexpected result*
	FALSE	FN (False Negative) *Missing result*	TN (True Negative) *Corect absence of result*

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Gambar 2 *Confusion Matrix*

TP *(True Positive)* : Jumlah variabel positif yang dilabeli dengan benar oleh *classifier*, sebagai contoh variabel dengan label status kelulusan = tepat waktu
TN *(True Negative)* : Jumlah variabel negatif yang dilabeli dengan benar oleh *classifier*
FP *(False Positive)* : Jumlah variabel negatif yang salah dilabeli oleh *classifier*
FN *(False Negative)* : Jumlah variabel positif yang salah dilabeli oleh *classifier*
P : Jumlah sampel positif

9. ISO 9126

ISO 9126 adalah standar internasional untuk uji kualitas evaluasi perangkat lunak. Gagasan mendasar di balik model ini adalah menentukan dan mengevaluasi kualitas produk perangkat lunak berdasarkan kualitas perangkat lunak internal dan eksternal dan hubungannya dengan atribut. Keuntungan utama dari model ini adalah model dapat diterapkan pada kualitas produk perangkat lunak. (Triyadi, 2020)

HASIL DAN PEMBAHASAN

1. Deskripsi Data

Data yang di peroleh keseluruhan sebanyak 850 data. Sedangkan setelah dalam proses preprosessing data dalam penelitian ini berjumblah sebanyak 700 data. 100 data di gunakan untuk training dan 700 data untuk data testing dengan 4 indikator, yaitu indikator tersebut yaitu usia, hb, usia kehamilan, masalah kehamilan. Dapat dilihat pada tabel 1:

Tabel 1 Data *K-NN*

Pasien	usia	Hb(g/dl)	Usia kehamilan	Masalah kehamilan
Pasien 1	3	1	3	1
Pasien 2	1	1	1	2
Pasien 3	2	1	1	2
Pasien 4	1	3	1	2
Pasien 5	1	3	1	2
⋮	⋮	⋮	⋮	⋮
Pasien 800	1	3	2	2

2. Proses Data Training Menggunakan Metode K-NN dan Naïve Bayes

Dalam proses ini data yang digunakan sebanyak 100 data untuk data training yang akan di olah menggunkan *K-NN* terlebih dahulu untuk menentukan kelayakan sebuah data. Berikut nilai K proses *K-NN* pada tabel 2 :

Tabel 2 Nilai K proses K-NN

X_1	X_2	X_3	X_4
157	182	138	178

Dihasilkan 71 data layak dan 29 data tidak layak. Yang dimana data layak akan dilanjutkan pada proses *Naïve Bayes*. Dapat di lihat pada tabel 3:

Tabel 3 Data Layak *K-Nn*

No	Pasien	X_1	X_2	X_3	X_4
1	Pasien 2	1	1	1	2
2	Pasien 3	2	1	1	2
3	Pasien 4	1	3	1	2
4	Pasien 5	1	3	1	2
5	Pasien 6	1	3	1	2
⋮	⋮	⋮	⋮	⋮	⋮
71	Pasien 100	2	1	1	2

Proses pemberian keterangan normal caesar dimana jika seseorang pasien memiliki masalah persalian dinyatakan caesar kecuali hb rendah. Dapat dilihat pada tabel 4:

Tabel 4 *Klasifikasi Naive Bayes*

No	Pasien	X_1	X_2	X_3	X_4	Keterangan
1	Pasien 2	1	1	1	2	Normal
2	Pasien 3	2	1	3	2	Caesar
3	Pasien 4	1	3	1	2	Caesar
4	Pasien 5	1	3	1	2	Caesar
5	Pasien 6	1	3	1	2	Normal
….						
71	Pasien 100	2	1	1	2	Caesar

Pada tabel 5 didapatkan hasil probabilitas persalinan secara normal sebesar 0,451 dan caesar sebesar 0,549.

Tabel 5 Probabilitas Awal

Jumlah Data	Jumlah Normal	Jumlah Caesar	Probabilitas Awal	
			Normal	Caesar
71	32	39	0,451	0,549

Menghitung probabilitas tiap label lalu dari hasil perhitungan tersebut akan di ambil hasil yang terbesar.

Tabel 6 Penentuan Jenis Persalinan

Kriteria	Normal	Caesar
1	25	25
2	22	18
2	29	35
1	11	3
Total	175450	47250
Probabilitas	79076,0563	25954,2254
terbesar	79076,0563	

Pada tabel 6 menghitung total probababilitas jenis persalinan secara normal sebesar 79076,0563 dan caesar sebesar 25954,2254 karena label kelas normal memiliki nilai yang lebih besar maka data yang di ambil adlah label normal dengan nilai sebesar 79076,0563.

3. **Proses Data Testing Menggunakan Metode K-NN**

Dalam proses ini data yang digunakan sebanyak 700 data untuk data training yang akan di olah menggunkan *K-NN* terlebih dahulu untuk menentukan kelayakan sebuah data. Berikut nilai Kproses *K-NN* pada tabel 7:

Tabel 7 Nilai K Proses *K-Nn*

X_1	X_2	X_3	X_4
1,71	1,98	1,44	1,73

Dihasilkan 419 data layak dan 281 data tidak layak. Yang dimana data layak akan dilanjutkan pada proses *Naïve Bayes*. Dapat dilihat pada tabel 8:

Tabel 8 Data Layak *K-Nn*

Pasien	X1	X2	X3	X4
Pasien 2	1	1	1	2
Pasien 3	2	1	1	2
Pasien 4	1	3	1	2
Pasien 5	1	3	1	2
Pasien 6	1	3	1	1
Pasien 9	1	1	1	1
⋮	⋮	⋮	⋮	⋮

Pasien 700	1	3	2	2

Proses pemberian keterangan normal caesar dimana jika seseorang pasien memiliki masalah persalian di nyatakan caesar kecuali hb rendah. Dapat dilihat pada tabel 9:

Tabel 9 Klasifikasi *Naive bayes*

No	Pasien	X1	X2	X3	X4	Keterangan
1	Pasien 1	3	1	3	1	Normal
2	Pasien 2	1	1	1	2	Normal
3	Pasien 3	2	1	1	2	Caesar
4	Pasien 4	1	3	1	2	Normal
5	Pasien 5	1	3	1	2	Normal
⋮	⋮	⋮	⋮	⋮	⋮	⋮
419	Pasien 700	1	3	2	2	Caesar

Pada tabel 10 didapatkan hasil probabilitas persalinan secaranormal sebesar 0,451 dan caesar sebesar 0,549.

Tabel 10 Probabilitas Awal

Jumlah Data	Jumlah Normal	Jumlah Caesar	Probabilitas Awal	
			Normal	Caesar
419	182	230	0,451	0,549

Menghitung probabilitas tiap label lalu dari hasil perhitungan tersebut akan di ambil hasil yang terbesar.

Tabel 11 Penentuan Jenis Persalinan

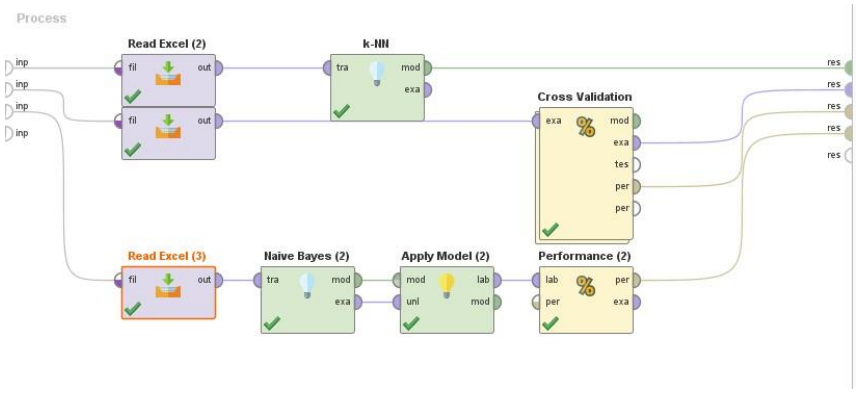
Kriteria	Normal	Caesar
1	6	36

2	129	26
2	170	178
1	150	221
Total	19737000	36820368
Probabilitas	8902847,2554	20211657,8520
Terbesar	20211657,8520	

Pada tabel 11 menghitung total probababilitas jenis persalinan secara normal sebesar 8902847,2554 dan caesar sebesar 20211657,8520 karena label kelas normal memiliki nilai yang lebih besar maka data yang di ambil adalah label normal dengan nilai sebesar 20211657,8520.

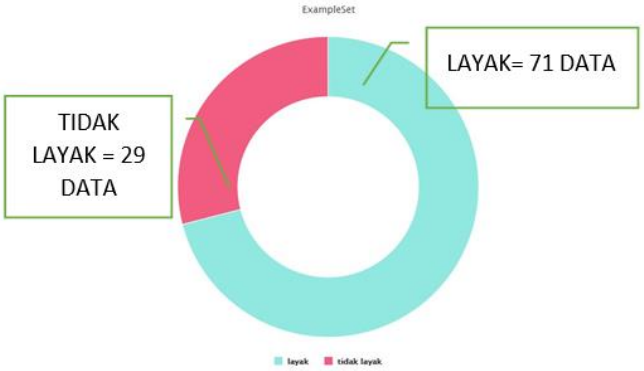
4. **Proses Data Training Rapidminer Metode K-NN dan Naïve Bayes**

Proses data training dalam gambar menggunakan data training sebanyak 100 data dilakukan dengan 2 metode yaitu K-NN dan NaïveBayes.



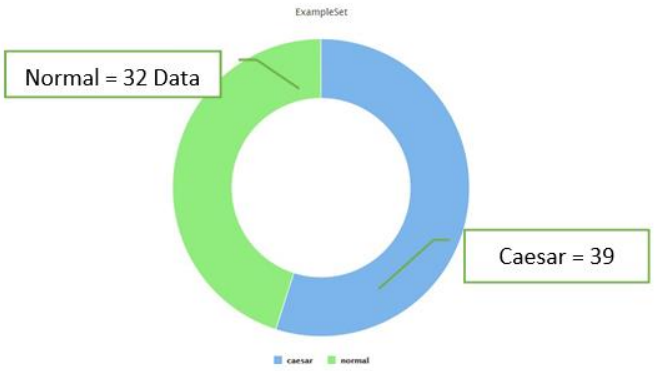
Gambar 3 Proses Training

Data training pada excel dimasukan kedalam fungsi rapiminer yaitu read excel, lalu dilanjutkan dengan proses *K-NN* proses tersebut menghasilkaan data layak sebanyak 71 data dan data tidak layaksebanyak 29 data. Dapat dilihat pada gambar 4:



Gambar 4 Hasil *K-NN*

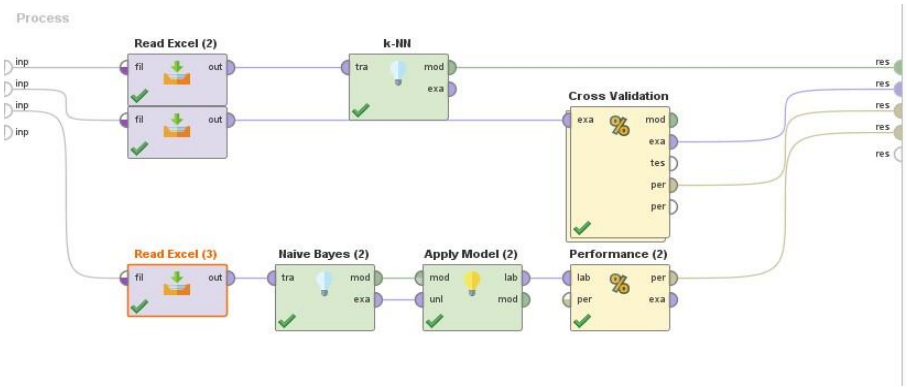
Setelah proses *K-NN* lalu dilanjutkan pada proses *Naïve Bayes* denganmenggunakan data layak yang sudah melalui proses *K-NN* lalu di masukan dalam fungsi Rapidminer read excel lalu proses naïve bayes maka dihasilkan 32 data normal dan 39 data caesar. Dapat dilihat padagambar 5:



Gambar 5 Hasil *Naive Bayes*

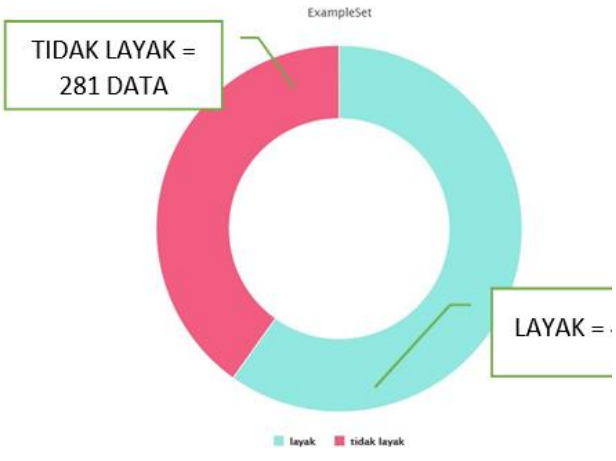
5. **Proses Data Testing Rapidminer Metode K-NN dan Naïve Bayes**

Proses data Testing dalam gambar menggunakan data testing sebanyak 700 data dilakukan dengan 2 metode yaitu K-NN dan NaïveBayes.



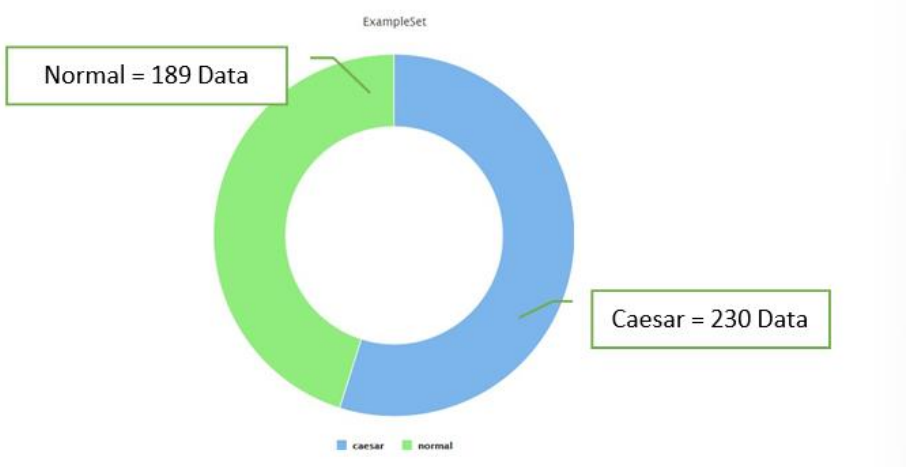
Gambar 6 Proses K-NN testing

Data testing pada excel dimasukan kedalam fungsi rapiminer yaitu read excel, lalu dilanjutkan dengan proses *K-NN* proses tersebut menghasilkaan data layak sebanyak 419 data dan data tidak layak sebanyak 281 data. Dapat dilihat pada gambar 7:



Gambar 7 Hasil *K-NN*

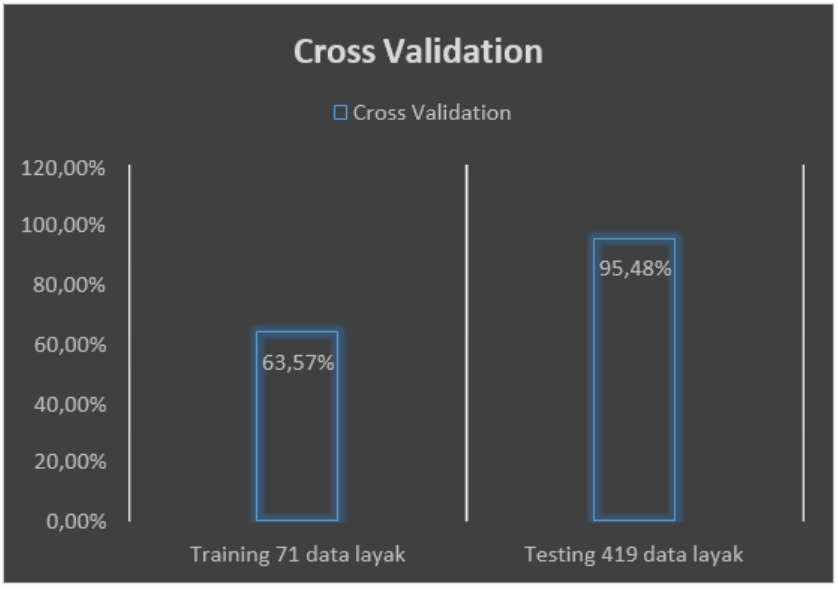
Setelah proses *K-NN* lalu dilanjutkan pada proses *Naïve Bayes* denganmenggunakan data layak yang sudah melalui proses *K-NN* lalu di masukan dalam fungsi Rapidminer read excel lalu proses *Naïve Bayes* maka dihasilkan 189 data normal dan 230 data caesar. Dapat dilihat pada gambar 8:



Gambar 8 Hasil *Naive Bayes*

6. **Cross Validation**

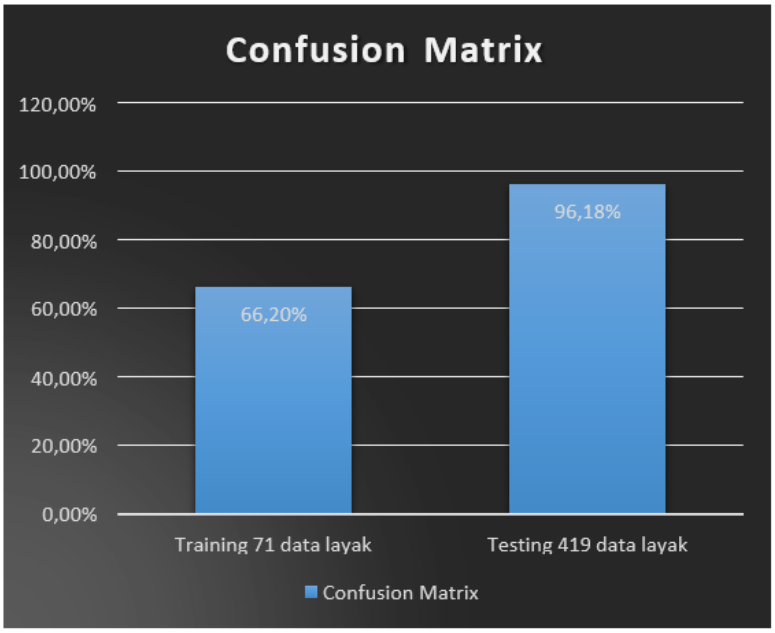
Dalam penelitian ini, *Cross Validation* digunakan untuk mengukur seberapa besar tingkat akurasi sebuah prediksi. Data training 71(layak)memiliki tingkat akurasi sebesar 63,57% dan data testing 419(layak) memiliki tingkat akurasi sebesar 95,48%



Gambar 9. Histogram *Cross Validation*

7. **Confusion Matrix**

Dalam penelitian ini, *confusion matrix* digunakan untuk mengukur seberapa besar tingkat akurasi sebuah prediksi. Data training 71(layak) memiliki tingkat akurasi sebesar 66,2% dan data testing 419(layak) memiliki tingkat akurasi sebesar 96,18%.



Gambar 10. Histogram *Confusion Matrix*

8. **Uji Manfaat**

Dalam proses ini dihasilkan uji manfaat analisis data penentuan jenis persalinan yang diperoleh dari jawaban kuesioner uji manfaat oleh 1responden :

Tabel 1. Hasil Uji Manfaat

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
3	4	4	2	4	3	3	2	3	3	3	3

4	3	3	3	3	3	3	3	3	3	4	4
3	4	3	3	3	4	3	3	4	3	3	4
3	3	4	4	4	3	4	4	3	3	4	3
4	4	4	4	4	4	3	3	4	4	3	3
4	4	3	2	3	3	4	3	4	4	3	4
4	3	4	3	3	4	3	3	4	2	3	3
3	3	4	4	3	3	4	3	3	2	2	4
2	4	4	3	4	4	4	4	2	4	4	4
3	4	2	3	2	4	3	3	3	4	3	4

Dari data tersebut menghasilkan uji manfaat sebagai berikut yang dapat dilihat pada tabel 2, tabel 3, tabel 4 dan tabel 5 :

Tabel 2. Uji Manfaat *Usability*

USABILITY				
	P1	P2	P3	TOTAL
STS	0	0	0	0
TS	10	0	10	6,666667
S	50	40	30	40
SS	40	60	60	53,33333
	100	100	100	86,66667

Tabel 3. Uji Manfaat *Learnability*

LEARNABILITY				
	P4	P5	P6	TOTAL
STS	0	0	0	0
TS	20	10	0	10
S	50	50	50	50
SS	30	40	50	40
	100	100	100	80

Tabel 4. Uji Manfaat *Efficiency*

EFFICIENCY				
	P7	P8	P9	TOTAL

STS	0	0	0	0
TS	0	10	10	6,666667
S	60	70	50	60
SS	40	20	40	33,33333
	100	100	100	86,66667

Tabel 5. Uji Manfaat *Acceptability*

	ACCEPTABILITY			
	P10	P11	P12	TOTAL
STS	0	0	0	0
TS	20	10	0	10
S	40	60	40	46,66667
SS	40	30	60	43,33333
	100	100	100	80

Dalam proses uji manfaat melalui kuesioner menghasilkan rata-rata responden sebesar 83,333%. Uji manfaat dinyatakan berhasil apabila jawaban setuju > 70% dan gagal apabila jawaban < 70%. Apabila uji manfaat dinyatakan gagal, maka akan dilakukan kembali pengujian hingga uji manfaat dinyatakan berhasil.

ANALISA HASIL PENGUJIAN

Dari pengujian dalam metode *K-NN* dan *Naive Bayes* dengan data training 100 data dan data testing 700 data yang telah dilakukan dalam proses perhitungan *Excel* dan *RapidMiner* dihasilkan:

1. Data training yang dilakukan dalam proses Excel menggunakan proses *K-NN* menggunakan 100 data yang dihasilkan 71 data dinyatakan layak sehingga data tersebut digunakan dalam proses *Naive Bayes* .
2. Data testing yang dilakukan dalam proses Excel menggunakan proses *K- NN* menggunakan 700 data dihasilkan 419 data dinyatakan layak sehingga data tersebut digunakan dalam proses *Naive Bayes*.
3. Data training yang dilakukan dalam proses *RapidMiner* menggunakan proses *K-NN* menggunakan 100 data yang dihasilkan 71 data dinyatakan layak sehingga data tersebut digunakan dalam proses *Naive Bayes* Data testing yang dilakukan dalam proses *RapidMiner* menggunakan proses *K-NN* menggunakan 700 data dihasilkan 419 data dinyatakan layak sehingga data tersebut digunakan dalam proses *Naive Bayes.*
4. Dengan menggunakan perhitungan pengujian *Cross validation* didapatkan nilai persentase untuk data training sebesar 63,57% sedangkan nilai persentase optimal untuk data testing sebesar 95,48%.
5. Dengan menggunakan perhitungan pengujian *Confusion Matrix* didapatkan nilai persentase untuk data training sebesar 66,2% sedangkan nilai persentase optimal untuk

data testing sebesar 96,18%.

SIMPULAN

Berdasarkan seluruh hasil tahapan penelitian dan pembahasan, diperolehkesimpulan sebagai berikut :

1. Dalam metode *K-NN* menghasilkan data layak sebanyak 71 untuk proses training dan 419 untuk proses testing.
2. Dalam metode *Naïve Bayes* dihasilkan nilai presentase akurasi pada penentuan jenis persalinan menurut sample yang sudah ada. Pada data training dihasilkan presentase sebesar 75% untuk normal dan pada datatesting dihasilkan sebesar 69% untuk caesar.
3. Dalam pengujian validasi menggunakan *Cross Validation* didapatkan nilai persentase untuk data training sebesar 63,57% sedangkan nilai persentase optimal untuk data testing sebesar 95,48%.
4. Dalam pengujian validasi *Confusion Matrix* didapatkan nilai persentase untuk data training sebesar 66,2% sedangkan nilai persentase optimal untuk data testing sebesar 96,18%.
5. Dalam uji manfaat dalam penelitian ini didapatkan tingkat pemahaman dari pengguna mengenai analisis data pelanggan potensial menggunakan metode *KNN* dan *Naïve Bayes* sebesar 83,333%.

DAFTAR PUSTAKA

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, *10*(2), 160–165.

Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, *9*(2), 188–195.

Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, *13*(1), 56–67.

Feblian, D., & Daihani, D. U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X. *Jurnal Teknik Industri*, *6*(1), 1–12.

Ilyas, M. M. (2017). Pelayanan Pendaftaran Pasien Rawat Jalan Di Rumah Sakit. *Prosiding Seminar Nasional Darmajaya*, *1*(1), 477–486.

Kusrini & Luthfi, E. T. (2009). Alogritma Data Mining. Penerbit Andi. Yogkarta.

Nugraha, F. F., Sunandar, I., & Juliane, C. (2022). Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika dan Sistem Informasi*, *9*(4), 2862 – 2869.

Nurfalinda, Nola Ritha, M. (n.d.). *1 , 2 , 3*. 1–8.

M. J. Zaki, W. Meira Jr & W. Meira (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.

Puteri, Q. A., Sagirani, T., & Lemantara, J. (2023). Perbandingan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) untuk Mengetahui Keakuratan Diagnoas Penyakit Diabetes. *Jurnal Nasional Teknologi dan Sistem Informasi, 9*(03), 248 – 254.

Rani, H. A. D., Zuhri, S., & Fuji, S. (2020). Sistem Prediksi Kondisi Kelahiran Bayi menggunakan Klasifikasi Naïve Bayes. *Joined Journal (Journal of Informatics Education)*, 3(2), 48.

Setyawati, E., Wijoyo, H, & Soeharmoko, N. (2020). Relational Database Management System (RDBMS). *Building and Maintaining a Data Warehouse*, 43–51.

Shedriko., Firdaus M. (2022). Penentuan Klasifikasi dengan CRISP-DM dalam Memprediksi Kelulusan Mahasiswa pada Suatu Mata Kuliah. *Seminar Nasional Riset dan Teknologi (SEMNAS RISTEK).* 826 – 831.

Telaumbanua, F. D., Hulu, P., Nadeak, T. Z., Lumbantong, R. R., & Dharma, A. (2019). *Penggunaan Machine Learning*. *3*, 57–64.

Triyadi, T. (2020). Aplikasi Monitoring Server dan Analisis Kualitas Menggunakan Model ISO 9126. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, *4*(3), 30