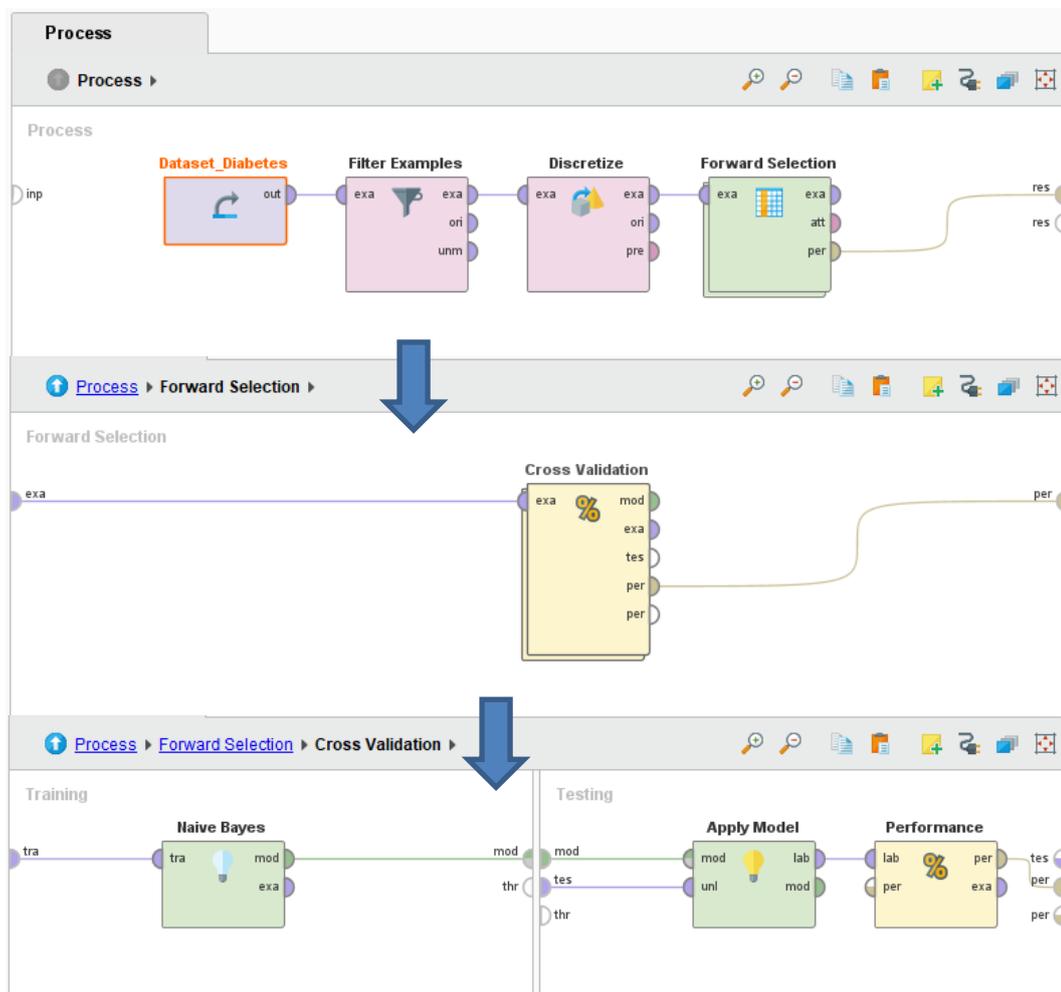


BAB IV

HASIL DAN PEMBAHASAN

4.1. Implementasi Algoritma *Naïve Bayes* menggunakan *Feature Selection*

Berikut merupakan implementasi algoritma *naïve bayes* yang dioptimasi menggunakan *feature forward selection* menggunakan tool rapidminer



Gambar 4.1 Optimasi Algoritma *Naïve Bayes* Menggunakan *Forward Selection*

Pada gambar 4.1 terdapat beberapa operator yang digunakan, diantaranya yaitu :

- Operator *retrieve*: digunakan untuk memanggil dataset diabetes.

Dataset yang digunakan pada penelitian ini berasal dari situs UCI Machine Learning dengan judul dataset *Early stAge diabetes risk prediction*. dataset dengan jumlah record 520 dan 17 atribut meliputi *Age*, *gender*, *polyuria*, *polydipsia*, *sudden weight loss*, *weakness*, *polyphagia*, *genital thrush*, *visual blurring*, *itching*, *Irritability*, *delayed healing*, *partial paresis*, *muscle stiffness*, *alopecia*, *obesity*, dan *class* sebagai atribut penentu klasifikasi. Deskripsi variabel tersebut dapat dilihat pada tabel di bawah ini.

Tabel 4.1 Deskripsi Variabel Dataset Penyakit Diabetes

No	Atribut	Value
1	Age	1. Mature = 21-45, 2. Old = 46-65
2	Gender	1. Male, 2. Female
3	Polyuria	1. Yes, 2. No
4	Polydipsia	1. Yes, 2. No
5	sudden weight loss	1. Yes, 2. No
6	weakness	1. Yes, 2. No
7	Polyphagia	1. Yes, 2. No
8	Genital thrush	1. Yes, 2. No
9	visual blurring	1. Yes, 2. No
10	Itching	1. Yes, 2. No
11	Irritability	1. Yes, 2. No
12	delayed healing	1. Yes, 2. No
13	partial paresis	1. Yes, 2. No
14	muscle stiffness	1. Yes, 2. No
15	Alopecia	1. Yes, 2. No
16	Obesity	1. Yes, 2. No
17	class	1. Positive, 2. Negative

- Operator *Filter Examples*: digunakan untuk menyeleksi umur pada dataset diabetes. Berdasarkan informasi dari *UCI machine learning*, *Age* yang digunakan yaitu dari usia dari 20-65 tahun, sehingga setelah dilakukan

Filter Examples diperoleh data yang siap digunakan yaitu sebanyak 468 data. Dengan jumlah *class Positive* sebanyak 284 data dan *class Negative* sebanyak 184 data.

- Operator *Discretize*: digunakan untuk mengelompokkan umur berdasarkan rentang umur. Pada penelitian ini umur dikelompokkan menjadi dua yaitu *Mature* dari umur 21-45 tahun dan *Old* dari umur 46-65 tahun.
- Operator *Forward selection*: digunakan untuk optimasi *naïve bayes* menggunakan metode *forward selection*. Dari 16 atribut yang ada, setelah diseleksi menggunakan *forward selection* diperoleh sebanyak 4 atribut yang berpengaruh. Atribut tersebut adalah *polyuria* , *gender*, *alopecia*, dan *Irritability*.

Tabel 4.2 Hasil Bobot Atribut Menggunakan *Feature Forward Selection*

No	Atribut	Bobot
1	<i>Age</i>	0
2	<i>Gender</i>	1
3	<i>Polyuria</i>	1
4	<i>Polydipsia</i>	0
5	<i>sudden weight loss</i>	0
6	<i>weakness</i>	0
7	<i>Polyphagia</i>	0
8	<i>Genital thrush</i>	0
9	<i>visual blurring</i>	0
10	<i>Itching</i>	0
11	<i>Irritability</i>	1
12	<i>delayed healing</i>	0
13	<i>partial paresis</i>	0
14	<i>muscle stiffness</i>	0
15	<i>Alopecia</i>	1
16	<i>Obesity</i>	0

Tabel 4.3 Atribut Terpilih Yang Memiliki Pengaruh Signifikan

No	Atribut	Bobot
1	<i>Gender</i>	1
2	<i>Polyuria</i>	1
3	<i>Irritability</i>	1
4	<i>Alopecia</i>	1

Dari hasil *feature forward selection*, atribut yang memiliki nilai bobot sama dengan 1 (satu) artinya atribut yang memiliki pengaruh signifikan, sedangkan atribut yang memiliki nilai bobot sama dengan 0 (nol) artinya atribut yang tidak berpengaruh dan bahkan dapat mengurangi tingkat akurasi. Dalam hal ini atribut yang nilai bobot sama dengan nol dan akan dihilangkan adalah *age, polydispia, sudden weight loss, weakness, polyphagia, Genital thrush, visual blurring, Itching, delayed healing, partial paresis, muscle stiffness, obesity*. Dari hasil *feature forward selection* pada dataset diabetes dapat diprediksi secara optimal menggunakan 4 atribut yaitu *gender, polyuria, irritability, alopecia*.

- Operator *Cross validation*: digunakan untuk membagi data *training* dan data *testing*. Teknik *cross validation* yang digunakan pada penelitian ini yaitu dengan teknik 10 validasi.
- Operator *Naïve bayes*: digunakan sebagai algoritma untuk menghitung prediksi.
- Operator *Apply Model*: digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data *training* pada data *testing*.
- Operator *Performance*: digunakan untuk mengevaluasi kinerja model klasifikasi dengan kriteria yang diberikan yaitu *accuracy, precision* , dan *recall* . Berikut adalah hasil perhitungan menggunakan *Confusion matrix* yang diimplementasikan pada tool rapidminer

PerformanceVector (Performance) ×

Criterion: accuracy (selected), precision, recall, AUC (optimistic), AUC (pessimistic)

Table View (selected) Plot View

accuracy: 92.10% +/- 3.33% (micro average: 92.09%)

	true Positive	true Negative	class precision
pred. Positive	265	18	93.64%
pred. Negative	19	166	89.73%
class recall	93.31%	90.22%	

Gambar 4.2 hasil kinerja algoritma *naïve bayes* menggunakan *forward selection*

PerformanceVector (Performance) ×

Criterion: accuracy, precision (selected), recall, AUC (optimistic), AUC (pessimistic)

Table View (selected) Plot View

precision: 89.92% +/- 5.56% (micro average: 89.73%) (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	265	18	93.64%
pred. Negative	19	166	89.73%
class recall	93.31%	90.22%	

Gambar 4.3 Nilai *precision* algoritma *naïve bayes* menggunakan *forward selection*

PerformanceVector (Performance) ×

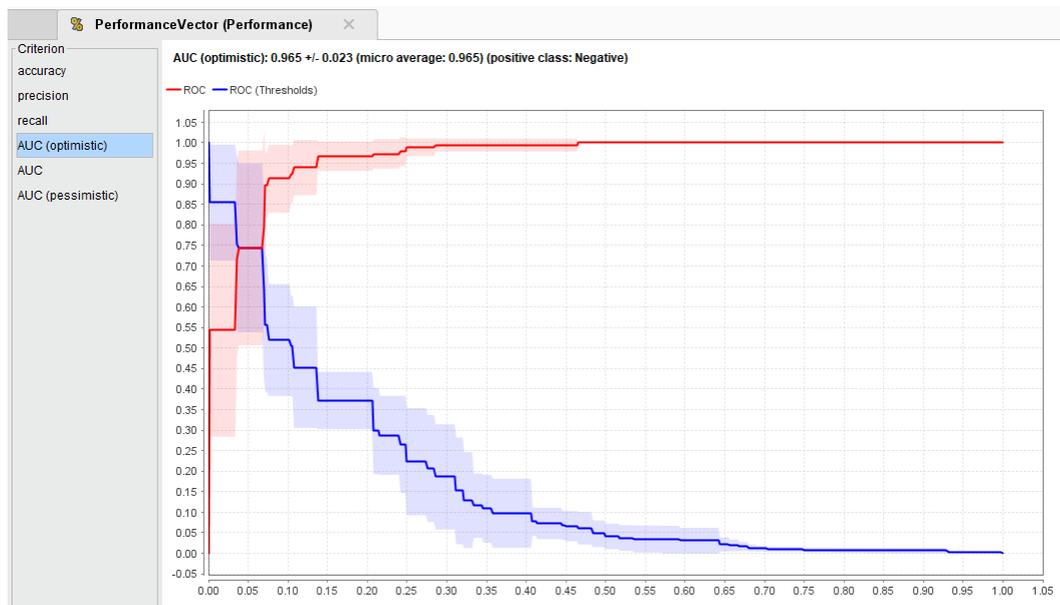
Criterion: accuracy, precision, recall (selected), AUC (optimistic), AUC (pessimistic)

Table View (selected) Plot View

recall: 90.26% +/- 4.80% (micro average: 90.22%) (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	265	18	93.64%
pred. Negative	19	166	89.73%
class recall	93.31%	90.22%	

Gambar 4.4 Nilai *recall* algoritma *naïve bayes* menggunakan *forward selection*



Gambar 4.5 grafik AUC *naïve bayes* menggunakan *forward selection*

Pada gambar 4.2 merupakan hasil kinerja model *naïve bayes* menggunakan *forward selection* menggunakan confusion matrik, berdasarkan 468 data pasien diabetes diperoleh hasil 265 pasien yang diprediksi diabetes dan positive diabetes, dan 18 pasien yang diprediksi diabetes ternyata *Negative* atau tidak menderita penyakit diabetes. Kemudian 166 pasien diprediksi *Negative* dan benar *Negative* atau tidak menderita penyakit diabetes, namun 19 pasien diprediksi *Negative* ternyata *Positive* diabetes. Dengan demikian nilai akurasi yang didapat dari uji model *naïve bayes* menggunakan *forward selection* sebesar 92,10%, dengan nilai precision 89,92%, dan nilai recall 90,26%. Sedangkan untuk grafik AUC (*area under curve*) pada gambar 4.5 garis horizontal menunjukkan *false Positive* dan garis vertical menunjukkan *false Negative* dari model *naïve bayes* menggunakan *forward selection* sebesar 0,965.

4.2. Evaluasi Performance

Berikut adalah perbandingan dari peneliti sebelumnya yang memprediksi penyakit diabetes tanpa menggunakan *feature selection* dengan peneliti sekarang yang menerapkan *feature selection* menggunakan metode *forward selection*.

Tabel 4.3 Perbandingan Performance Metode

<i>Feature selection</i>	Akurasi <i>Naïve bayes</i>	Precision	Recall	AUC
Peneliti sebelumnya : tanpa <i>feature selection</i> (Ridwan, A., 2020)	90,20%	82,35%	87,50%	0.955
Peneliti Sekarang : <i>feature selection</i> menggunakan metode <i>Forward selection</i>	92,10%.	89,92%	90,26%	0, 965

Berdasarkan tabel 4.3 dapat disimpulkan bahwa adanya peningkatan akurasi algoritma *naïve bayes* yang dioptimasi menggunakan *feature selection* yang belum diterapkan pada peneliti sebelumnya. Dan dari eksperimen yang dilakukan menggunakan *feature forward selection* mengalami kenaikan dengan nilai akurasi 92,10%, nilai *precision* 89,92%, nilai *recall* 90,26% dan nilai AUC sebesar 0,965 dengan kategori *excellent Classification*.