

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai performansi hasil klasifikasi dari metode Support Vector Machine (SVM) & SVM + Adaboost Feature Selection *Forward Sequential Select (FSS)* pada studi simulasi dilanjutkan dengan hasil klasifikasi pada penerapan data publik. Pada bab ini juga dilakukan evaluasi mengenai efek dari penggunaan seleksi fitur *Selection Forward Sequential Select (FSS)* pada klasifikasi menggunakan SVM pada data publik.

4.1 Dataset

Proses validasi dilakukan untuk menemukan dan mengkonversi data agar sesuai dengan metode algoritma data mining, serta untuk mencapai akurasi dan kinerja yang optimal. Dalam dataset yang digunakan, validasi data dilakukan dengan menghapus entri yang tidak lengkap atau kosong (null). Setelah itu, dilakukan seleksi atribut untuk menentukan atribut mana yang relevan dari dataset yang digunakan dalam menganalisis kerusakan jalan. Berikut ini disajikan beberapa data training seperti yang tercantum dalam Tabel 4.2.

Tabel 4.1. Potongan Dataset

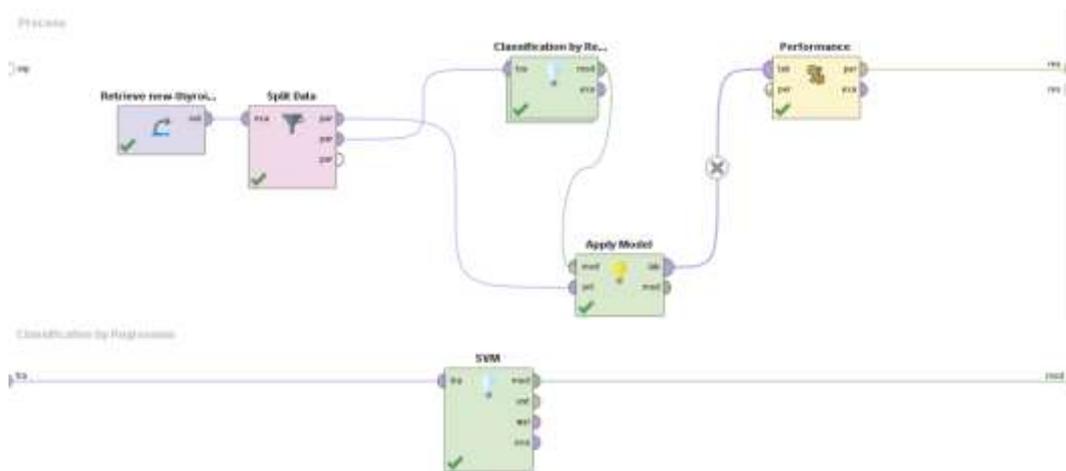
age	sex	On thyroxine	Query On thyroxine	On antithyroid medication	Sick	pregnant	Thyroid surgery	I131 treatment	Query hypothyroid	Query hyperthyroid
41	F	f	f	F	F	f	F	f	f	f
23	F	f	f	F	F	f	F	f	f	f
46	M	f	f	F	F	f	F	f	f	f
70	F	t	f	F	F	f	F	f	f	f

Data akan dijelajahi secara teliti guna menentukan variabel mana yang perlu dieksplorasi, dengan tujuan menghindari kemiripan dan redundansi yang tidak diperlukan dalam proses teknik data mining. Pada penanganan situasi yang muncul dalam kumpulan data publik, dalam tahap seleksi data, kita dapat memanfaatkan perangkat lunak RapidMiner untuk mengimpor data. Namun, proses ini mungkin menghadapi kendala karena adanya beberapa kesalahan dalam data. Secara otomatis, program akan memberi tahu pengguna bahwa dataset

memerlukan langkah-langkah preprocessing. Setelah mengidentifikasi kesalahan dalam data, kita dapat melakukan seleksi dengan menerapkan filter untuk menghapus data yang bermasalah tersebut. Hal ini bertujuan agar data dapat diproses dengan lancar dalam sistem

1. Pengujian 1 Menggunakan Algoritma Support Vector Machine (SVM)

Pada algoritma *support vector machine (svm)*, record yang sudah di import ke rapid miner akan digunakan untuk menentukan pola pohon keputusan, hasil dari pola pohon keputusan dapat dilihat pada gambar 4.2 berikut



Gambar 3. Tampilan Proses *Support Vector Machine (SVM)*

Pengujian yang dilakukan menggunakan matriks kebingungan (confusion matrix) terdiri dari akurasi, presisi, recall AUC pada sebuah dataset Hasil pengujian matriks dapat dilihat pada tabel di bawah ini:

Tabel 4.2 Hasil Proses menggunakan Algoritma *support vector machine (svm)*

No	Keterangan	Nilai <i>support vector machine (svm)</i>
1	<i>Accuracy</i>	83,63%
2	<i>Precision</i>	100%
3	<i>Recall / Confusion Matrix</i>	46,15%
4	AUC	0,500

Hasil Akurasi *support vector machine (svm)* dan Split Validation, pada pembagian data sampling 80% dan data testing 20% maka didapatkan akurasi sebesar 83,63%

$$\text{Menghitung akurasi : Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)}$$

$$\begin{aligned} \text{Akurasi} &= \frac{(352 + 559 + 256)}{(352 + 559 + 256 + 11 + 3 + 4 + 7 + 5 + 5)} \\ &= \frac{1167}{1202} \\ &= 97,09\% \end{aligned}$$

$$\text{Menghitung Precision, Rumus: } P:\textit{precision} = \frac{TP}{(TP+FP)}$$

$$\begin{aligned} \text{Presisi} &= \frac{559}{(559+3+5)} \\ &= \frac{559}{567} \\ &= 98,59\% \end{aligned}$$

$$\text{Menghitung Recall, Rumus : } \textit{Recall} = \frac{TP}{(TP+FN)}$$

$$\begin{aligned} \text{Reccal} &= \frac{559}{(559+11+5)} \\ &= \frac{559}{575} \\ &= 97,22\% \end{aligned}$$

Perhitungan F1 Score

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precission}) / (\text{Recall} + \text{Precission})$$

$$\text{dalam kasus di atas, F1 Score} = 2 * (46,15\% * 100\%) / (46,15\% + 100,00\%)$$

$$= 9,230\% / 146,15\% = 0,063\%$$

Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,500 seperti gambar dibawah ini.



Gambar 4. Nilai AUC Algoritma *support vector machine (svm)*

Pengujian yang dilakukan menggunakan matriks kebingungan (confusion matrix) terdiri dari akurasi, presisi, dan recall pada sebuah dataset Hasil pengujian matriks kebingungan untuk dataset yang diproses dapat dilihat pada tabel di bawah ini

Tabel 4. 3 Hasil dari SVM Split Validation (80:20)

No	Keterangan	Hasil Nilai <i>support vector machine (svm)</i>
1	Accuracy	83,63%
2	Precision	100,00%
3	Recall / Confusion Matrix	95,91%
4	AUC	0,500

Dari tabel yang diberikan, tampaknya ini adalah hasil evaluasi kinerja dari sebuah model Support Vector Machine (SVM). Berikut adalah analisis singkat dari setiap metrik yang tercantum dalam tabel:

1. Accuracy (Akurasi): Akurasi merupakan persentase prediksi yang benar dari total sampel yang dievaluasi. Dalam kasus ini, akurasi model SVM mencapai 83,63%. Ini menunjukkan bahwa sekitar 83,63% dari

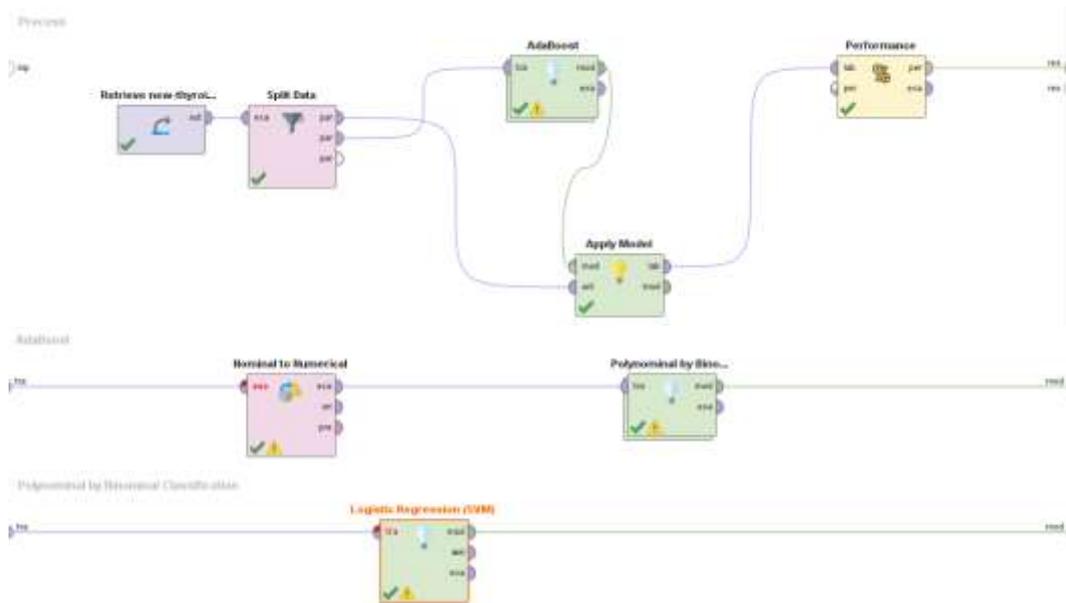
keseluruhan data yang dievaluasi diprediksi dengan benar oleh model SVM.

2. Precision (Presisi): Presisi mengukur seberapa baik model dalam mengidentifikasi instance positif yang sebenarnya dari semua instance yang diprediksi sebagai positif. Dalam konteks ini, nilai presisi sebesar 100,00% menunjukkan bahwa semua instance yang diprediksi positif oleh model SVM adalah benar-benar positif.
3. Recall / Confusion Matrix (Sensitivity): Recall, juga dikenal sebagai Sensitivitas, adalah rasio instance positif yang diidentifikasi dengan benar oleh model terhadap semua instance positif yang sebenarnya. Nilai Recall sebesar 95,91% menunjukkan bahwa model mampu mengidentifikasi sebagian besar instance positif yang sebenarnya.
4. AUC (Area Under the Curve): AUC mengacu pada Area di Bawah Kurva Receiver Operating Characteristic (ROC). ROC adalah kurva yang mengilustrasikan kinerja model klasifikasi di berbagai threshold. Dalam hal ini, nilai AUC sebesar 0,500 menunjukkan kinerja yang kurang baik, karena nilainya hanya sedikit di atas 0,5 yang menandakan klasifikasi acak.

Dari analisis ini, meskipun nilai presisi dan recall cukup baik, namun akurasi dan AUC menunjukkan ada potensi untuk peningkatan kinerja model SVM. Hal ini bisa disebabkan oleh beberapa faktor seperti ketidakseimbangan kelas dalam data, fitur yang tidak representatif, atau mungkin memerlukan penyetelan parameter yang lebih baik untuk meningkatkan kinerja model. Evaluasi lebih lanjut dan eksperimen mungkin diperlukan untuk memperbaiki kinerja model.

2. Pengujian 2 Menggunakan Algoritma Support Vector Machine (SVM) dan Adaboost

Pada algoritma *support vector machine (svm)* dan *Adaboost*, record yang sudah di import ke rapid miner akan digunakan untuk menentukan pola pohon keputusan, hasil dari pola pohon keputusan dapat dilihat pada gambar 4.7 berikut



Gambar 5. Proses menggunakan Algoritma *support vector machine (svm)* dan *Adaboost*

Hasil Akurasi *support vector machine (svm).adaboost* dan Split Validation, pada pembagian data sampling 80% dan data testing 20% maka didapatkan akurasi sebesar 88,89%.

Tabel 4. 4 Hasil dari *support vector machine (SVM)* dan Adaboost Split Validation (80:20)

No	Keterangan	Nilai Algoritma <i>support vector machine (svm)</i> dan <i>Adaboost</i>
1	<i>Accuracy</i>	88,89%
2	<i>Precision</i>	97,14%
3	<i>Recall / Confusion Matrix</i>	65,38%
4	AUC	0,823

Menghitung akurasi : Akurasi = $\frac{(TN+TP)}{(TN+FN+FP+TP)}$

$$\begin{aligned}
 \text{Akurasi} &= \frac{(352 + 559 + 256)}{(352 + 559 + 256 + 11 + 3 + 4 + 7 + 5 + 5)} \\
 &= \frac{1167}{1202} \\
 &= 97,09\%
 \end{aligned}$$

Menghitung Precision, Rumus: $P:precision = \frac{TP}{(TP+FP)}$

$$\begin{aligned} \text{Presisi} &= \frac{559}{(559+3+5)} \\ &= \frac{559}{567} \\ &= 98,59\% \end{aligned}$$

Menghitung Recall, Rumus : $Recall = \frac{TP}{(TP+FN)}$

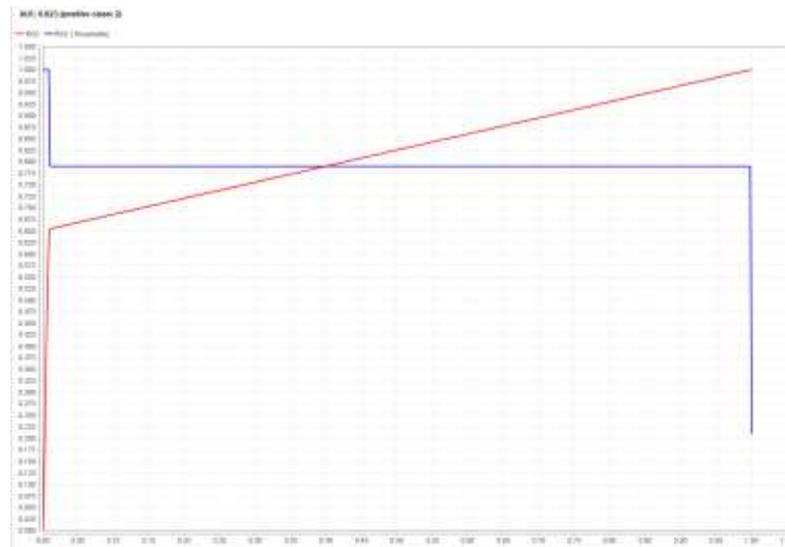
$$\begin{aligned} \text{Reccal} &= \frac{559}{(559+11+5)} \\ &= \frac{559}{575} \\ &= 97,22\% \end{aligned}$$

Perhitungan F1 Score

F1 Score = $2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$

dalam kasus di atas, F1 Score = $2 * (65,38\% * 97,14\%) / (65,38\% + 97,14\%) =$
 $12,702\% / 162,52\% = 0,078\%$

Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,823 seperti gambar dibawah ini.



Gambar 6. Nilai AUC Algoritma *support vector machine (svm)* dan *adaboost*

Pengujian yang dilakukan menggunakan matriks kebingungan (*confusion matrix*) terdiri dari akurasi, presisi, dan recall pada sebuah dataset Hasil pengujian matriks kebingungan untuk dataset yang diproses dapat dilihat pada tabel di bawah ini

Tabel 4. 3 Hasil dari *support vector machine (SVM)* dan *Adaboost Split Validation (80:20)*

No	<i>support vector machine dan Adaboost</i>	Hasil Nilai
1	Accuracy	88,89%
2	Precision	97,14%
3	Recall / Confusion Matrix	65,38%
4	AUC	0,823

Dari tabel yang diberikan, tampaknya ini adalah hasil evaluasi kinerja dua model: Support Vector Machine (SVM) dan Adaboost. Berikut adalah analisis singkat dari setiap metrik yang tercantum dalam tabel:

Support Vector Machine (SVM):

1. Accuracy (Akurasi): Akurasi model SVM mencapai 88,89%. Ini menunjukkan bahwa sekitar 88,89% dari keseluruhan data yang dievaluasi diprediksi dengan benar oleh model SVM.
2. Precision (Presisi): Presisi model SVM sebesar 97,14%. Ini menunjukkan bahwa 97,14% dari prediksi positif model SVM adalah benar-benar positif.
3. Recall / Confusion Matrix (Sensitivity): Recall model SVM sebesar 65,38%. Ini menunjukkan bahwa model SVM dapat mengidentifikasi sekitar 65,38% dari instance positif yang sebenarnya.
4. AUC (Area Under the Curve): AUC model SVM sebesar 0,823. Ini menunjukkan kinerja yang baik, karena nilai AUC mendekati 1, yang menandakan bahwa model memiliki kemampuan yang baik untuk membedakan antara kelas positif dan negatif.

Adaboost:

1. Accuracy (Akurasi): Akurasi model Adaboost mencapai 88,89%, yang sama dengan model SVM. Ini menunjukkan bahwa keduanya memiliki tingkat akurasi yang sama.
2. Precision (Presisi): Presisi model Adaboost sebesar 97,14%, sama dengan model SVM. Ini menunjukkan bahwa 97,14% dari prediksi positif model Adaboost adalah benar-benar positif.
3. Recall / Confusion Matrix (Sensitivity): Recall model Adaboost sebesar 65,38%, juga sama dengan model SVM. Ini menunjukkan bahwa keduanya memiliki tingkat recall yang sama.
4. AUC (Area Under the Curve): AUC model Adaboost sebesar 0,823, juga sama dengan model SVM. Ini menunjukkan bahwa keduanya memiliki kinerja yang mirip dalam hal kemampuan membedakan antara kelas positif dan negatif.

Dari analisis ini, kedua model (SVM dan Adaboost) memiliki kinerja yang serupa dalam hal akurasi, presisi, recall, dan AUC. Meskipun demikian, perbedaan mungkin muncul dalam konteks komputasi dan kemampuan adaptasi terhadap

berbagai jenis data. Evaluasi lebih lanjut dan eksperimen mungkin diperlukan untuk memilih model yang paling sesuai dengan kebutuhan spesifik aplikasi.

4.2 Evaluasi

Dari perbandingan kedua tabel hasil nilai Support Vector Machine (SVM) dan kombinasi SVM dengan Adaboost, beberapa perbedaan dan kesamaan terlihat dalam kinerja keduanya:

Perbandingan Kinerja:

1. Accuracy (Akurasi):

- SVM: 83,63%
- SVM + Adaboost: 88,89%
- Akurasi model SVM + Adaboost lebih tinggi daripada SVM saja, menunjukkan bahwa gabungan SVM dengan Adaboost dapat meningkatkan kemampuan prediksi model.

2. Precision (Presisi):

- SVM: 100,00%
- SVM + Adaboost: 97,14%
- Presisi SVM adalah 100%, namun presisi SVM + Adaboost sedikit lebih rendah. Meskipun demikian, presisi keduanya tetap tinggi, menunjukkan kemampuan model dalam mengidentifikasi instance positif dengan benar.

3. Recall / Confusion Matrix (Sensitivity):

- SVM: 95,91%
- SVM + Adaboost: 65,38%
- Perbedaan yang signifikan terlihat di sini. Recall SVM + Adaboost (65,38%) jauh lebih rendah dibandingkan dengan SVM (95,91%). Hal ini menunjukkan bahwa kombinasi SVM dengan Adaboost mungkin tidak seefektif SVM dalam mengidentifikasi instance positif yang sebenarnya.

4. AUC (Area Under the Curve):

- SVM: 0,500
- SVM + Adaboost: 0,823

- AUC SVM + Adaboost jauh lebih tinggi dibandingkan dengan SVM saja. Hal ini menunjukkan bahwa SVM + Adaboost memiliki kemampuan yang lebih baik dalam membedakan antara kelas positif dan negatif.

Dari hasil evaluasi maka SVM + Adaboost memiliki akurasi yang lebih tinggi dan AUC yang lebih baik dibandingkan dengan SVM saja. Namun, SVM memiliki recall yang jauh lebih tinggi dibandingkan dengan SVM + Adaboost. Hal ini menunjukkan bahwa SVM lebih baik dalam mengidentifikasi instance positif yang sebenarnya. Presisi kedua model tetap tinggi, tetapi sedikit lebih rendah untuk SVM + Adaboost. Evaluasi lebih lanjut diperlukan untuk memahami mengapa recall dari SVM + Adaboost jauh lebih rendah dari pada SVM saja, meskipun akurasi dan AUC lebih tinggi. Ini mungkin disebabkan oleh perbedaan dalam penanganan class imbalance atau karakteristik data yang berbeda.