

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini penulis akan membahas mengenai implementasi data yang digunakan. Data tersebut akan diimplementasikan menggunakan rapidminer menggunakan K-Nearest Neighbor dan Adaboost, data yang telah disiapkan akan dimuat ke dalam perangkat lunak RapidMiner. Kemudian, algoritma K-NN dan Adaboost akan diterapkan pada data tersebut untuk membangun model prediksi. Proses ini akan melibatkan beberapa tahap seperti pemrosesan data, pembagian data menjadi data latih dan data uji, pelatihan model, dan evaluasi performa model.

4.1. Persiapan Data

Dataset yang digunakan menggunakan dataset dari kaggle <https://www.kaggle.com/code/cahyaalkahfi/klasifikasi-obesitas-dengan-keras-r/input> dengan jumlah sebanyak 2111 data yang menggunakan tools rapid miner. Atribut yang digunakan sebanyak 16 atribut yaitu Gender, Age, Height, Weight, family_history_with_overweight, FAVC (Frequency of consumption of high caloric food /Frekuensi konsumsi makanan berkalori tinggi), FCVC (Frequency of consumption of vegetables/ Frekuensi konsumsi sayuran), NCP (Number of main meals/ Jumlah makanan utama), CAEC (Consumption of food between meals/ Konsumsi makanan di antara waktu makan), SMOKE, CH2O(Consumption of water daily/ Konsumsi air setiap hari), SCC (Calories consumption monitoring/ Pemantauan konsumsi kalori), FAF (Physical activity frequency/ Frekuensi aktivitas fisik), TUE (Time using technology devices/ Waktu menggunakan perangkat teknologi), CALC (Calories consumption monitoring/ Pemantauan konsumsi kalori), MTRANS (Transportation used/ Transportasi yang digunakan). Metode yang digunakan dalam penelitian ini Adaboost dan K-nn. Kumpulan data ini digunakan untuk memprediksi apakah seorang pasien kemungkinan besar menderita obesitas. Data obesitas terdapat pada gambar 4.1 dibawah ini:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Gender	Age	Height	Weight	family_hi:FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad		
2	Female	21.0	1.62	64.0	yes	no	2.0	3.0	Sometime	no	2.0	no	0.0	1.0	no	Public_Transportation	Normal	
3	Female	21.0	1.52	56.0	yes	no	3.0	3.0	Sometime	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transportation	Normal	
4	Male	23.0	1.8	77.0	yes	no	2.0	3.0	Sometime	no	2.0	no	2.0	1.0	Frequently	Public_Transportation	Normal	
5	Male	27.0	1.8	87.0	no	no	3.0	3.0	Sometime	no	2.0	no	2.0	0.0	Frequently	Walking	Overweight	
6	Male	22.0	1.78	89.8	no	no	2.0	1.0	Sometime	no	2.0	no	0.0	0.0	Sometimes	Public_Transportation	Overweight	
7	Male	29.0	1.62	53.0	no	yes	2.0	3.0	Sometime	no	2.0	no	0.0	0.0	Sometimes	Automobile	Normal	
8	Female	23.0	1.5	55.0	yes	yes	3.0	3.0	Sometime	no	2.0	no	1.0	0.0	Sometimes	Motorbike	Normal	
9	Male	22.0	1.64	53.0	no	no	2.0	3.0	Sometime	no	2.0	no	3.0	0.0	Sometimes	Public_Transportation	Normal	
10	Male	24.0	1.78	64.0	yes	yes	3.0	3.0	Sometime	no	2.0	no	1.0	1.0	Frequently	Public_Transportation	Normal	
11	Male	22.0	1.72	68.0	yes	yes	2.0	3.0	Sometime	no	2.0	no	1.0	1.0	no	Public_Transportation	Normal	
12	Male	26.0	1.85	105.0	yes	yes	3.0	3.0	Sometime	no	3.0	no	2.0	2.0	Sometimes	Public_Transportation	Obesity	
13	Female	21.0	1.72	80.0	yes	yes	2.0	3.0	Frequent	no	2.0	yes	2.0	1.0	Sometimes	Public_Transportation	Overweight	
14	Male	22.0	1.65	56.0	no	no	3.0	3.0	Sometime	no	3.0	no	2.0	0.0	Sometimes	Public_Transportation	Normal	
15	Male	41.0	1.8	99.0	no	yes	2.0	3.0	Sometime	no	2.0	no	2.0	1.0	Frequently	Automobile	Obesity	
16	Male	23.0	1.77	60.0	yes	yes	3.0	1.0	Sometime	no	1.0	no	1.0	1.0	Sometimes	Public_Transportation	Normal	
17	Female	22.0	1.7	66.0	yes	no	3.0	3.0	Always	no	2.0	yes	2.0	1.0	Sometimes	Public_Transportation	Normal	
18	Male	27.0	1.93	102.0	yes	yes	2.0	1.0	Sometime	no	1.0	no	1.0	0.0	Sometimes	Public_Transportation	Overweight	
19	Female	29.0	1.53	78.0	no	yes	2.0	1.0	Sometime	no	2.0	no	0.0	0.0	no	Automobile	Obesity	
20	Female	30.0	1.71	82.0	yes	yes	3.0	4.0	Frequent	yes	1.0	no	0.0	0.0	no	Automobile	Overweight	
21	Female	23.0	1.65	70.0	yes	no	2.0	1.0	Sometime	no	2.0	no	0.0	0.0	Sometimes	Public_Transportation	Overweight	
22	Male	22.0	1.65	80.0	yes	no	2.0	3.0	Sometime	no	2.0	no	3.0	2.0	no	Walking	Overweight	
23	Female	52.0	1.69	87.0	yes	yes	3.0	1.0	Sometime	yes	2.0	no	0.0	0.0	no	Automobile	Obesity	

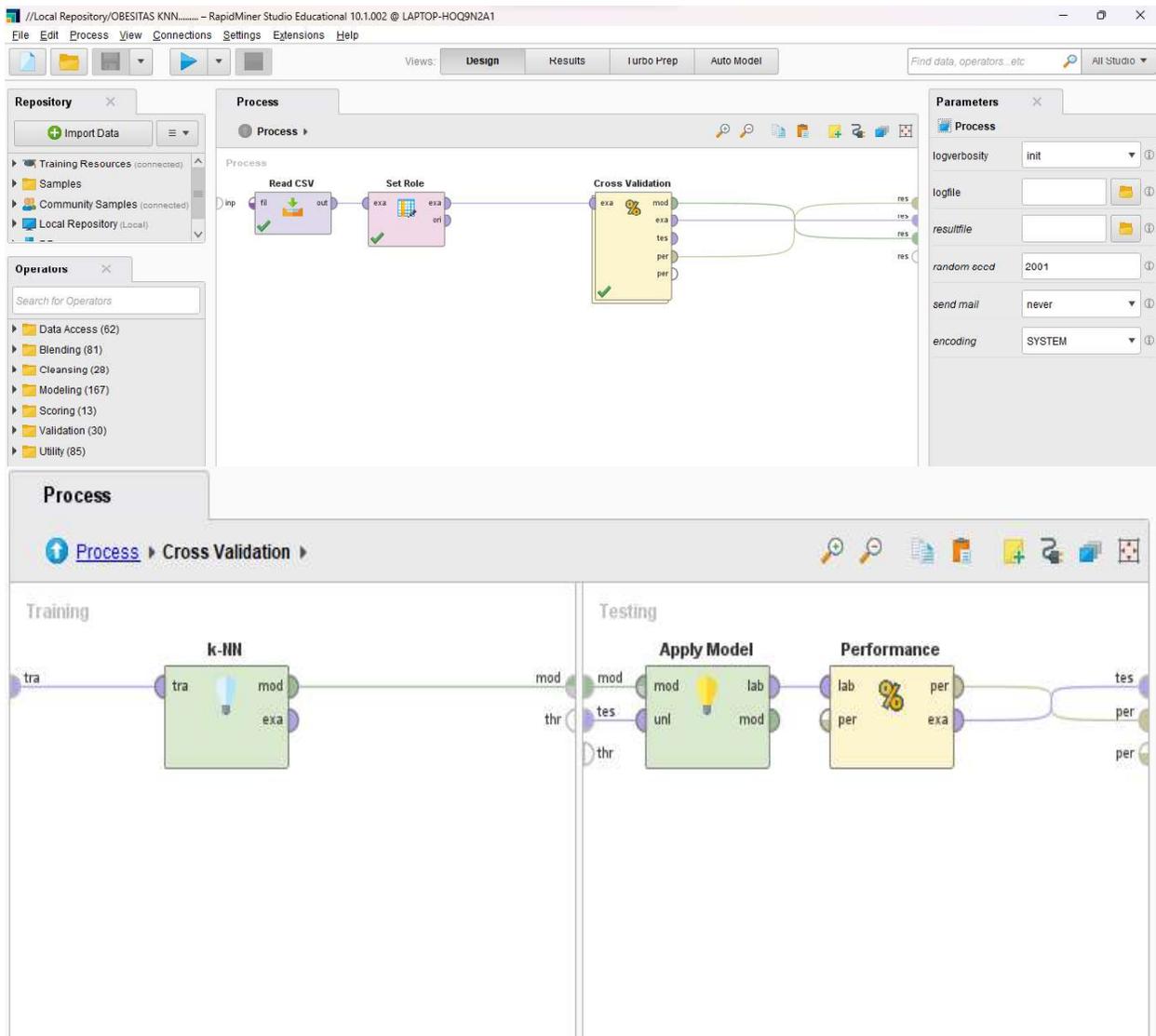
Gambar 4.1 Dataset Penyakit Obesitas

4.2. Pemodelan

Pemilihan dan penerapan teknik pemodelan yang sesuai merupakan langkah penting dalam tahap analisis data. Hal ini dilakukan untuk memastikan bahwa metode yang digunakan dapat memberikan hasil prediksi yang akurat dan relevan dengan tujuan penelitian. Dalam konteks penelitian ini, fokusnya adalah pada prediksi terkait penyakit obesitas. Teknik pemodelan yang digunakan adalah data mining, yang merupakan proses ekstraksi pola tersembunyi yang signifikan atau pengetahuan yang berguna dari dataset yang besar. Dalam data mining, terdapat berbagai macam algoritma yang dapat digunakan untuk membangun model prediksi, salah satunya adalah K-NN, dan optimasi menggunakan adabost.

4.2.1 Penelitian Menggunakan Algoritma K-Nearest Neighbor

Penerapan data pada Rapidminer untuk Prediksi Penyakit Obesitas menggunakan algoritma K-Nearest Neighbor ditunjukkan pada Gambar 4.2 dibawah ini:



Gambar 4.2 Penerapan Data Prediksi Obesitas Menggunakan Algoritma K-Nearest Neighbor pada Rapidminer

Pada gambar 4.2, data yang telah disiapkan diimplementasikan pada aplikasi RapidMiner untuk melakukan prediksi. Proses ini melibatkan eksperimen menggunakan metode cross validation. Cross validation merupakan sebuah teknik yang umum digunakan dalam evaluasi model di bidang data mining dan machine learning. Teknik ini secara langsung membagi dataset menjadi dua bagian utama, yaitu data training dan data testing. Dalam penelitian ini, data yang digunakan adalah data supervised, yang berarti kita memiliki label atau target yang ingin

diprediksi, yaitu status obesitas seseorang berdasarkan atribut-atribut tertentu. Algoritma yang dipilih untuk prediksi adalah K-Nearest Neighbor (K-NN).

Pada proses eksperimen menggunakan cross validation, data dibagi menjadi beberapa subset atau lipatan (folds) pada penelitian ini menggunakan Folds 10. Kemudian, model K-NN dilatih menggunakan data dari lipatan-lipatan tersebut secara bergantian, dengan satu lipatan sebagai data testing dan lipatan lainnya sebagai data training. Proses ini dilakukan berulang-ulang hingga semua lipatan menjadi data testing satu kali. Tujuan dari cross validation adalah untuk mengukur kinerja model secara keseluruhan dengan mempertimbangkan variasi yang mungkin terjadi dalam pembagian data. Hasil dari eksperimen tersebut dapat dilihat pada gambar 4.3.

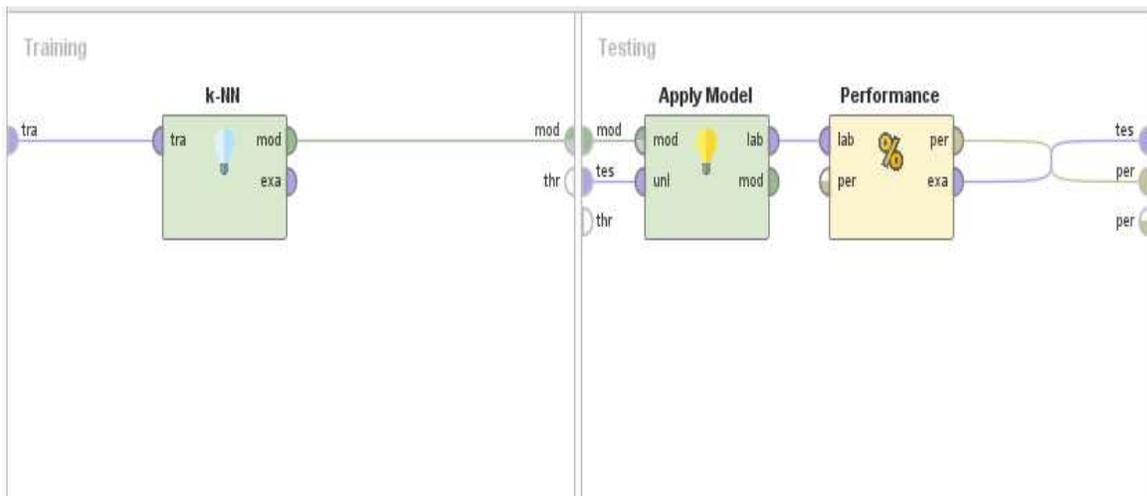
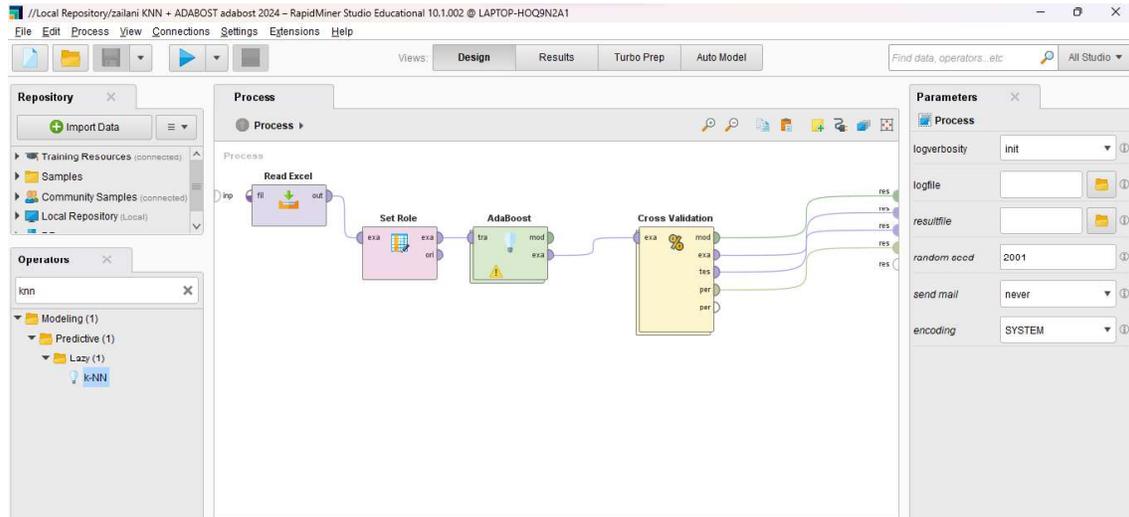
	true Normal	true Overweight	true Obesity	true Insufficient	class precision
pred. Normal	241	251	281	97	27.70%
pred. Overweight	29	288	80	5	71.64%
pred. Obesity	7	23	603	4	94.66%
pred. Insufficient	10	18	8	166	82.18%
class recall	83.97%	49.66%	62.04%	61.03%	

Gambar 4.3 Confusion Matrix Hasil Prediksi Obesitas Menggunakan Algoritma K-Nearest Neighbor pada Rapidminer

Gambar 4.3 adalah confusion matrix yang menunjukkan hasil eksperimen, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 61,49 % class recall pada true normal 83,97 % , class recall pada true Overweight 49,66%, class recall pada true Obesity 62,04%, class recall pada true insufficient 61,03%, dan class presisi pada pred normal 27,70 % , class presisi pada pred Overweight 71,64%, class presisi pada pred Obesity 94,66%, class presisi pada pred insufficient 82,18%.

4.2.2 Penelitian Menggunakan Adabost dan Algoritma K-Nearest Neighbor

Metode yang digunakan menggunakan metode adabost dengan algoritma K-nn. Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar 4.4 di bawah ini:



Gambar 4.4 Penerapan Data Prediksi Obesitas Menggunakan Metode Adabost dan Algoritma K-Nearest Neighbor pada Rapidminer

Pada gambar 4.4 Di dalam proses penelitian ini, data yang telah diolah terlebih dahulu diterapkan pada aplikasi RapidMiner, suatu platform yang digunakan untuk melakukan analisis data secara efisien. Penggunaan data tersebut bertujuan untuk memprediksi kemungkinan obesitas dengan menerapkan teknik cross validation, sebuah metode yang digunakan untuk memastikan bahwa model yang dibangun dapat diterapkan dengan baik pada data yang baru. Dalam eksperimen ini, tidak hanya satu, tetapi dua algoritma digunakan, yakni Adaboost dan K-Nearest Neighbor (K-NN). Adaboost dipilih untuk meningkatkan performa model dengan menggabungkan beberapa model lemah menjadi satu yang lebih kuat, sedangkan K-NN digunakan sebagai algoritma dasar untuk memperoleh hasil prediksi. Selanjutnya, hasil eksperimen yang dihasilkan dari proses ini akan ditampilkan dalam gambar 4.5, yang akan memberikan gambaran tentang seberapa baik model yang telah dibangun dalam memprediksi obesitas.

accuracy: 83.75% +/- 3.27% (micro average: 83.75%)

	true Normal	true Overweight	true Obesity	true Insufficient	class precision
pred. Normal	247	42	2	26	77.92%
pred. Overweight	35	422	82	27	74.56%
pred. Obesity	1	79	882	2	91.49%
pred. Insufficient	4	37	6	217	82.20%
class recall	86.06%	72.76%	90.74%	79.78%	

Gambar 4.5 Confusion Matrix Hasil Prediksi Obsitas Menggunakan Algoritma Adaboost dan K-Nearest Neighbor pada Rapidminer

Gambar 4.5 adalah confusion matrix yang menunjukkan hasil eksperimen, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 83,75 % class recall pada true normal 88,06 % , class recall pada true Overweight 72,76%, class recall pada true Obesity 90,74 % , class recall pada true insufficient 79,78%, dan class presisi pada pred normal 77,92 % , class presisi pada pred Overweight 74,56 % , class presisi pada pred Obesity 91,49%, class presisi pada pred insufficient 82,20%.

4.3. Perbandingan Penelitian Tidak Menggunakan Adabost dan Menggunakan Adabost

Perbandingan hasil eksperimen yang tidak menggunakan metode Adabost dan eksperimen yang menggunakan Adabost ditunjukkan pada tabel dibawah ini:

Tabel 4.1 Perbandingan Hasil Akurasi Penggunaan Adabost.

Penelitian	Tingkat Akurasi
K-Nearest Neighbor	61,49 %
K-Nearest Neighbor dan Adabost	83,75 %

Hasil dari eksperimen menunjukkan bahwa penggunaan metode K-Nearest Neighbor (K-NN) dan Adaboost berhasil menghasilkan tingkat akurasi yang tinggi, yaitu sebesar 83,75%. Hal ini menandakan bahwa model yang dibangun mampu memprediksi obesitas dengan tingkat keakuratan yang signifikan. Dalam eksperimen ini, terdapat peningkatan akurasi sebesar 22,26% dari hasil sebelumnya. Peningkatan tersebut menunjukkan bahwa penggunaan kedua metode tersebut secara bersamaan mampu meningkatkan performa model secara signifikan dibandingkan dengan menggunakan metode yang satu saja.

Peningkatan akurasi yang signifikan ini dapat menjadi indikasi bahwa kombinasi antara K-Nearest Neighbor dan Adaboost mampu mengatasi kelemahan atau keterbatasan masing-masing algoritma saat digunakan secara individual. Dengan demikian, hasil eksperimen ini memberikan bukti bahwa penggunaan ensemble learning dengan menggabungkan beberapa model dapat meningkatkan performa prediksi secara keseluruhan.

4.4 Evaluasi

Pemodelan yang dilakukan dengan menggunakan algoritma k-nearest neighbor (K-NN) awalnya menghasilkan tingkat akurasi yang kurang memuaskan. Ini menunjukkan bahwa dalam kasus ini, K-NN mungkin tidak mampu mengidentifikasi pola-pola yang kompleks dalam data dengan baik, atau mungkin tidak sesuai dengan struktur data yang digunakan. Sebagai hasilnya, prediksi obesitas yang dihasilkan oleh K-NN tidak memiliki tingkat keakuratan yang diharapkan. Namun, setelah penggunaan algoritma Adaboost, terjadi peningkatan yang signifikan dalam akurasi prediksi obesitas. Adaboost merupakan sebuah metode ensemble learning yang mampu meningkatkan kinerja model dengan mengkombinasikan beberapa model lemah menjadi satu model yang lebih kuat. Dengan menggabungkan beberapa model klasifikasi lemah dari K-NN, Adaboost berhasil meningkatkan kemampuan prediktif model secara keseluruhan, sehingga menghasilkan prediksi obesitas yang lebih akurat.

Selain meningkatkan akurasi, kinerja algoritma dalam penelitian ini juga menunjukkan performa yang sangat baik sekali. Performa yang sangat baik ini bisa dilihat dari peningkatan signifikan dalam akurasi setelah menggunakan Adaboost. Hal ini menunjukkan bahwa Adaboost mampu mengatasi kelemahan atau keterbatasan yang dimiliki oleh K-NN, dan secara keseluruhan, model yang dibangun mampu memberikan prediksi obesitas yang lebih andal dan akurat. Dengan demikian, penggunaan Adaboost sebagai metode ensemble learning berhasil meningkatkan akurasi prediksi obesitas, serta menunjukkan performa yang sangat baik dalam penelitian ini. Hal ini menegaskan pentingnya pemilihan metode dan algoritma yang tepat dalam membangun model prediktif yang efektif dan handal.