

DAFTAR PUSTAKA

- Barros, R. C. (2011). *Towards the automatic design of decision tree induction algorithms*. 8.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *ArXiv:1603.02754 [Cs]*. <https://doi.org/10.1145/2939672.2939785>
- Cost, S., & Salzberg, S. (1993). A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning*, 10(1), 57–78. <https://doi.org/10.1007/BF00993481>
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, 233–240. <https://doi.org/10.1145/1143844.1143874>
- Dewi, M. C. (2015). *Faktor-Faktor yang Menyebabkan Obesitas pada Anak*. 4.
- Diana, R., Yuliana, I., & Yasmin, G. (2013). *FAKTOR RISIKO KEGEMUKAN PADA WANITA DEWASA INDONESIA*. 8, 8.
- Dietterich, T. G. (2000a). *Ensemble Methods in Machine Learning*. 15.
- Dietterich, T. G. (2000b). Ensemble Methods in Machine Learning. Dalam G. Goos, J. Hartmanis, & J. van Leeuwen, *Multiple Classifier Systems* (Vol. 1857, hlm. 1–15). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45014-9_1
- Fithriasari, K., Hariastuti, I., & Wening, K. S. (2020). Handling Imbalance Data in Classification Model with Nominal Predictors. *International Journal of Computing Science and Applied Mathematics*, 6(1), 33. <https://doi.org/10.12962/j24775401.v6i1.6643>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Gorunescu, F. (2011). *Data mining: Concepts, models and techniques*. Springer.
- Haibo He, Yang Bai, Garcia, E. A., & Shutao Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress*

- on *Computational Intelligence*), 1322–1328.
<https://doi.org/10.1109/IJCNN.2008.4633969>
- Hanif, I. (2019). *Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction*. 21.
- Low, S., Chin, M. C., & Deurenberg-Yap, M. (2009). *Review on Epidemic of Obesity*. 38(1), 10.
- Norton, K. (1996). *Anthropometrica: A Textbook of Body Measurement for Sports and Health Courses*. Univesity of New South Wales Press.
- Rokach, L., & Maimon, O. (2005). *Top-Down Induction of Decision Trees Classifiers—A Survey*. 35(4), 12.
- Shakya, R. (2019). *Application of Machine Learning Techniques in Credit Card Fraud Detection*. <https://doi.org/10.34917/14279175>
- Yap, B. W., Rani, K. A., Rahman, H. A. A., Fong, S., Khairudin, Z., & Abdullah, N. N. (2014). An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. Dalam T. Herawan, M. M. Deris, & J. Abawajy (Ed.), *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (Vol. 285, hlm. 13–22). Springer Singapore. https://doi.org/10.1007/978-981-4585-18-7_2
- Zhang, L., & Zhan, C. (2017). Machine Learning in Rock Facies Classification: An Application of XGBoost. *International Geophysical Conference, Qingdao, China, 17-20 April 2017*, 1371–1374. <https://doi.org/10.1190/IGC2017-351>
- Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms* (0 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b12207>

LAMPIRAN

Lampiran 1 Konfigurasi hyperparameter model XGBoost

Hyperparameter	Nilai	Keterangan
Eta	0.1, 0.3	Nilai learning rate
Gamma	0	Nilai minimum loss reduction untuk splitting pada node tree berikutnya
Max_depth	4, 6, 8, 10	Mengatur tingkat kedalaman suatu tree untuk mengendalikan kompleksitas model dalam menghindari <i>overfitting</i>
Min_child_weight	1	Nilai bobot minimal yang dibutuhkan child
Colsample_bytree	1	Rasio kolom pada saat membangun pohon
Sub_sample	1	Rasio jumlah sampel yang digunakan untuk latih nilai 0.5 berarti menggunakan setengah dari data evaluasi sebelumnya secara acak dalam membuat tree baru
Reg_lambda	1	Nilai regularisasi
Nrounds	100, 300, 500	Jumlah iterasi

Lampiran 2 Hasil pengujian model XGBoost yang dibangun dari konfigurasi hyperparameter

Nama model	Nrounds	Eta	Max depth	Min child weight	Sub sample	Gamma	Colsample bytree	LDK
Model 01	100	0,1	4	1	1	1	1	69,371%
Model 02	100	0,3	4	1	1	1	1	69,437%
Model 03	100	0,1	6	1	1	1	1	69,415%
Model 04	100	0,3	6	1	1	1	1	69,279%
Model 05	100	0,1	8	1	1	1	1	69,310%
Model 06	100	0,3	8	1	1	1	1	68,673%
Model 07	100	0,1	10	1	1	1	1	68,969%
Model 08	100	0,3	10	1	1	1	1	67,951%
Model 09	300	0,1	4	1	1	1	1	69,436%
Model 10	300	0,3	4	1	1	1	1	69,360%
Model 11	300	0,1	6	1	1	1	1	69,314%
Model 12	300	0,3	6	1	1	1	1	68,862%

Model 13	300	0,1	8	1	1	1	1	68,854%
Model 14	300	0,3	8	1	1	1	1	67,937%
Model 15	300	0,1	10	1	1	1	1	68,164%
Model 16	300	0,3	10	1	1	1	1	66,888%
Model 17	500	0,1	4	1	1	1	1	69,434%
Model 18	500	0,3	4	1	1	1	1	69,300%
Model 19	500	0,1	6	1	1	1	1	69,167%
Model 20	500	0,3	6	1	1	1	1	68,620%
Model 21	500	0,1	8	1	1	1	1	68,589%
Model 22	500	0,3	8	1	1	1	1	67,501%
Model 23	500	0,1	10	1	1	1	1	67,664%
Model 24	500	0,3	10	1	1	1	1	66,522%

Lampiran 3 Hasil pengujian model XGBoost dengan ADASYN-N yang dibangun dari konfigurasi hyperparameter

Nama model	Nrounds	Eta	Max depth	Min child weight	Sub sample	Gamma	Colsample bytree	LDK
Model 01	100	0,1	4	1	1	1	1	73,651%
Model 02	100	0,3	4	1	1	1	1	73,655%
Model 03	100	0,1	6	1	1	1	1	73,625%
Model 04	100	0,3	6	1	1	1	1	73,441%
Model 05	100	0,1	8	1	1	1	1	73,468%
Model 06	100	0,3	8	1	1	1	1	72,944%
Model 07	100	0,1	10	1	1	1	1	73,080%
Model 08	100	0,3	10	1	1	1	1	72,098%
Model 09	300	0,1	4	1	1	1	1	73,672%
Model 10	300	0,3	4	1	1	1	1	73,570%
Model 11	300	0,1	6	1	1	1	1	73,493%
model012	300	0,3	6	1	1	1	1	73,077%
Model 13	300	0,1	8	1	1	1	1	73,007%
Model 14	300	0,3	8	1	1	1	1	72,148%
Model 15	300	0,1	10	1	1	1	1	72,229%
Model 16	300	0,3	10	1	1	1	1	70,943%
Model 17	500	0,1	4	1	1	1	1	73,639%
Model 18	500	0,3	4	1	1	1	1	73,501%
Model 19	500	0,1	6	1	1	1	1	73,360%
Model 20	500	0,3	6	1	1	1	1	72,820%
Model 21	500	0,1	8	1	1	1	1	72,700%
Model 22	500	0,3	8	1	1	1	1	71,630%
Model 23	500	0,1	10	1	1	1	1	71,710%
Model 24	500	0,3	10	1	1	1	1	70,540%

Lampiran 4 Contoh perhitungan jarak MVDM

Contoh perhitungan jarak antar kategori pada setiap variabel menggunakan persamaan (2.9):

a. Untuk X1 (2 Kategori)

$$\begin{aligned}\delta(x_{11}, y_{21}) &= \sum_{i=1}^2 \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right| \\ &= \left| \left(\frac{311270}{342994} - \frac{294592}{361983} \right) \right| + \left| \left(\frac{31724}{342994} - \frac{67391}{361983} \right) \right| \\ &= 0.1874\end{aligned}$$

b. Untuk X2 (4 Kategori)

$$\begin{aligned}\delta(x_{12}y_{22}) &= \sum_{i=1}^2 \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right| \\ &= \left| \left(\frac{311270}{342994} - \frac{294592}{361983} \right) \right| + \left| \left(\frac{31724}{342994} - \frac{67391}{361983} \right) \right| \\ &= 0.2145\end{aligned}$$

$$\begin{aligned}\delta(x_{12}y_{32}) &= \sum_{i=1}^2 \left| \frac{C_{1i}}{C_1} - \frac{C_{3i}}{C_3} \right| \\ &= \left| \left(\frac{311270}{342994} - \frac{102429}{118444} \right) \right| + \left| \left(\frac{31724}{342994} - \frac{16015}{118444} \right) \right| \\ &= 0.0904\end{aligned}$$

$$\begin{aligned}\delta(x_{12}y_{42}) &= \sum_{i=1}^2 \left| \frac{C_{1i}}{C_1} - \frac{C_{4i}}{C_4} \right| \\ &= \left| \left(\frac{311270}{342994} - \frac{14593}{15388} \right) \right| + \left| \left(\frac{31724}{342994} - \frac{795}{15388} \right) \right| \\ &= 0.0767\end{aligned}$$

$$\begin{aligned}\delta(x_{22}y_{32}) &= \sum_{i=1}^2 \left| \frac{C_{2i}}{C_2} - \frac{C_{3i}}{C_3} \right| \\ &= \left| \left(\frac{294592}{361983} - \frac{102429}{118444} \right) \right| + \left| \left(\frac{67391}{361983} - \frac{16015}{118444} \right) \right| \\ &= 0.1241\end{aligned}$$

$$\begin{aligned}\delta(x_{22}y_{42}) &= \sum_{i=1}^2 \left| \frac{C_{2i}}{C_2} - \frac{C_{4i}}{C_4} \right| \\ &= \left| \left(\frac{294592}{361983} - \frac{14593}{15388} \right) \right| + \left| \left(\frac{67391}{361983} - \frac{795}{15388} \right) \right| \\ &= 0.2912\end{aligned}$$

$$\begin{aligned}\delta(x_{32}y_{42}) &= \sum_{i=1}^2 \left| \frac{C_{3i}}{C_3} - \frac{C_{4i}}{C_4} \right| \\ &= \left| \left(\frac{102429}{118444} - \frac{14593}{15388} \right) \right| + \left| \left(\frac{16015}{118444} - \frac{795}{15388} \right) \right| \\ &= 0.1671\end{aligned}$$

c. Untuk X11 (2 Kategori)

$$\begin{aligned}\delta(x_{111}, y_{211}) &= \sum_{i=1}^2 \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right| \\ &= \left| \left(\frac{24112}{34958} - \frac{581750}{670019} \right) \right| + \left| \left(\frac{10846}{34958} - \frac{88269}{670019} \right) \right| \\ &= 0.1874\end{aligned}$$

Untuk memudahkan perhitungan berikutnya, hasil perhitungan MVDM disajikan dalam bentuk tabel berikut.

X1	C1	C2
C1	0	0.1874
C2	0.1874	0

X2	C1	C2	C3	C4
C1	0	0.2145	0.0904	0.0767
C2	0.2145	0	0.1241	0.2912
C3	0.0904	0.1241	0	0.1671
C4	0.0767	0.2912	0.1671	0

X3	C1	C2
C1	0	0.1548
C2	0.1548	0

X4	C1	C2
C1	0	0.1062
C2	0.1062	0

X5	C1	C2
C1	0	0.0849
C2	0.0849	0

X6	C1	C2	C3	C4	C5	C6	C7	C8
C1	0	0.022 3	0.050 6	0.082 2	0.087 6	0.107 6	0.124 6	0.214 8
C2	0.022 3	0	0.028 3	0.059 8	0.065 3	0.085 2	0.102 3	0.192 4
C3	0.050 6	0.028 3	0	0.031 5	0.037	0.056 9	0.073 9	0.164 1
C4	0.082 2	0.059 8	0.031 5	0	0.005 4	0.025 4	0.042 4	0.132 6
C5	0.087 6	0.065 3	0.037	0.005 4	0	0.02	0.037	0.127 2
C6	0.107 6	0.085 2	0.056 9	0.025 4	0.02	0	0.017	0.107 2
C7	0.124 6	0.102 3	0.073 9	0.042 4	0.037	0.017	0	0.090 2
C8	0.214 8	0.192 4	0.164 1	0.132 6	0.127 2	0.107 2	0.090 2	0

X7	C1	C2	C3	C4	C5	C6	C7	C8
C1	0	0.004 7	0.001 7	0.000 8	0.004 7	0.014 5	0.011 1	0.059

C2	0.004 7	0	0.003	0.005 5	3E-06	0.009 7	0.006 4	0.063 7
C3	0.001 7	0.003	0	0.002 5	0.003	0.012 7	0.009 4	0.060 7
C4	0.000 8	0.005 5	0.002 5	0	0.005 5	0.015 2	0.011 9	0.058 2
C5	0.004 7	3E-06	0.003	0.005 5	0	0.009 7	0.006 4	0.063 7
C6	0.014 5	0.009 7	0.012 7	0.015 2	0.009 7	0	0.003 3	0.073 4
C7	0.011 1	0.006 4	0.009 4	0.011 9	0.006 4	0.003 3	0	0.070 1
C8	0.059	0.063 7	0.060 7	0.058 2	0.063 7	0.073 4	0.070 1	0

X8	C1	C2	C3	C4	C5	C6
C1	0	0.0200	0.0023	0.0102	0.0237	0.0499
C2	0.0200	0	0.0223	0.0099	0.0037	0.0299
C3	0.0023	0.0223	0	0.0124	0.026	0.0522
C4	0.0102	0.0099	0.0124	0	0.0135	0.0397
C5	0.0237	0.0037	0.026	0.0135	0	0.0262
C6	0.0499	0.0299	0.0522	0.0397	0.0262	0

X9	C1	C2	C3	C4	C5	C6
C1	0	0.0103	0.0146	0.0238	0.0581	0.0613
C2	0.0103	0	0.0043	0.0135	0.0479	0.051
C3	0.0146	0.0043	0	0.0092	0.0435	0.0467
C4	0.0238	0.0135	0.0092	0	0.0343	0.0375
C5	0.0581	0.0479	0.0435	0.0343	0	0.0032
C6	0.0613	0.051	0.0467	0.0375	0.0032	0

X10	C1	C2	C3	C4	C5	C6
C1	0	0.0017	0.0306	0.0449	0.0781	0.0892
C2	0.0017	0	0.0289	0.0432	0.0763	0.0874
C3	0.0306	0.0289	0	0.0142	0.0474	0.0585
C4	0.0449	0.0432	0.0142	0	0.0332	0.0443
C5	0.0781	0.0763	0.0474	0.0332	0	0.0111
C6	0.0892	0.0874	0.0585	0.0443	0.0111	0

X11	C1	C2
C1	0	0.3570
C2	0.3570	0

Hitung jarak VDM dari amatan x dan y menggunakan persamaan (2.8):

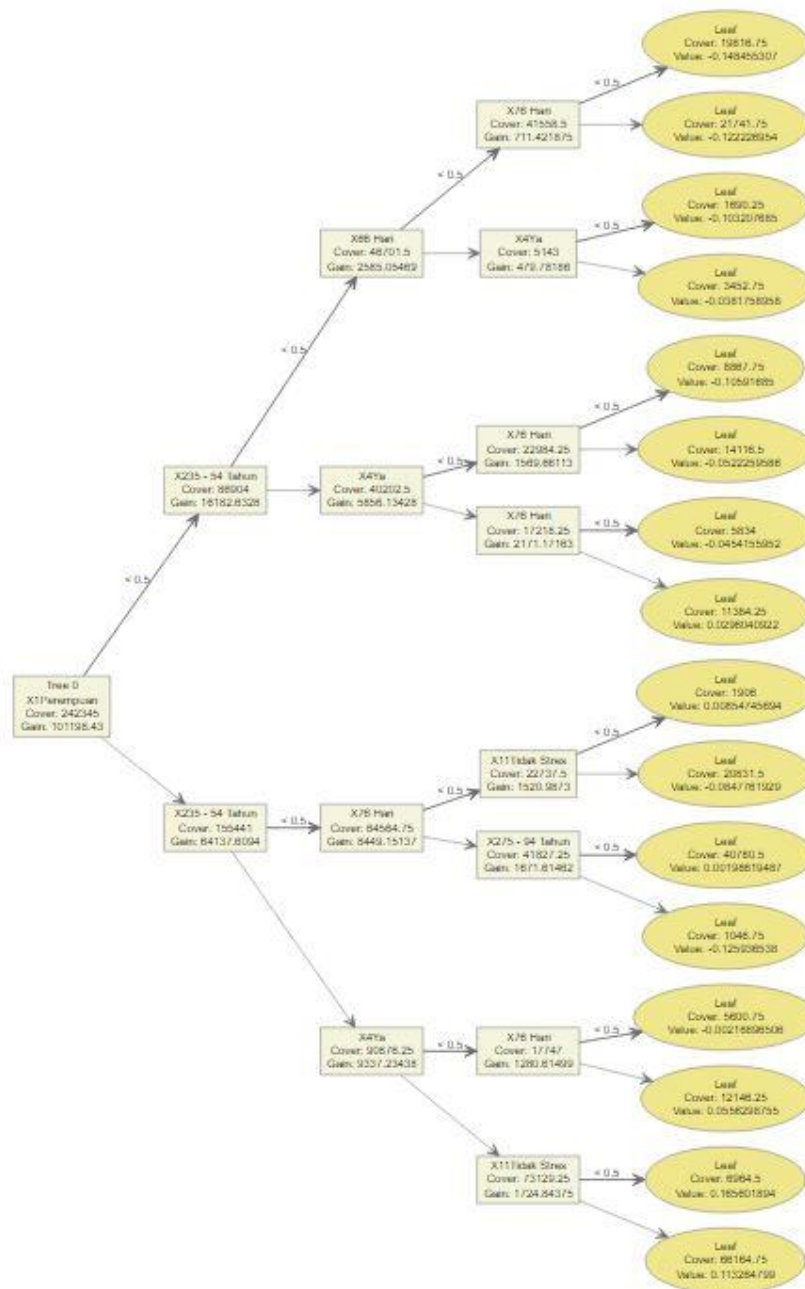
Instance	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X	2	2	2	2	1	4	8	2	5	3	2
Y	1	2	2	2	1	5	8	3	3	3	2
Distance	0.1874	0	0	0	0	0.0054	0	0.0223	0.0435	0	0

$$\begin{aligned} \Delta(x, y) &= 0.1874 + 0 + 0 + 0 + 0 + 0 + 0.0054 + 0 + 0.0223 + 0.0435 + 0 + 0 \\ &= 0.2586 \end{aligned}$$

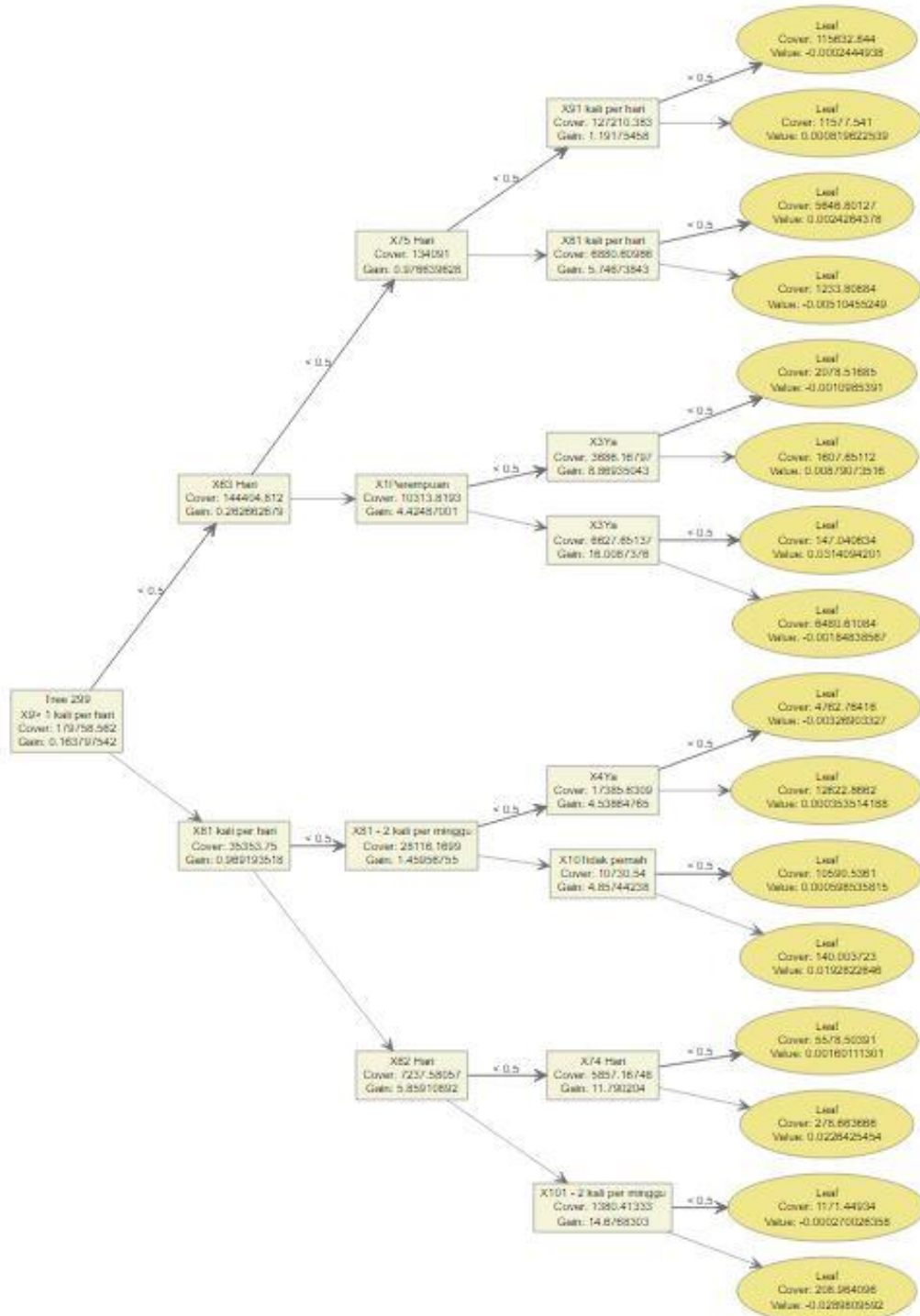
Lampiran 5 Bentuk pohon dan *Feature importance* model 09 XGBoost dengan ADASYN-N berdasarkan nilai *gain*

Berikut bentuk pohon ke-1 dan pohon ke-300 dari 300 pohon yang dibentuk berdasarkan model 09 XGBoost dengan ADASYN-N serta tingkat kepentingan fiturnya (*feature importance*):

- Pohon ke-1



- Pohon ke-300



- Struktur 300 pohon dari model

Pohon	Node	ID	Fitur	Split	Yes	No	Gain	Cover
0	0	0-0	X1Perempuan	0.5	0-1	0-2	101198.4	242345
0	1	0-1	X235 - 54 Tahun	0.5	0-3	0-4	16182.63	86904
0	2	0-2	X235 - 54 Tahun	0.5	0-5	0-6	64137.61	155441
0	3	0-3	X66 Hari	0.5	0-7	0-8	2585.055	46701.5
0	4	0-4	X4Ya	0.5	0-9	0-10	5856.134	40202.5
0	5	0-5	X76 Hari	0.5	0-11	0-12	8449.151	64564.75
0	6	0-6	X4Ya	0.5	0-13	0-14	9337.234	90876.25
0	7	0-7	X76 Hari	0.5	0-15	0-16	711.4219	41558.5
0	8	0-8	X4Ya	0.5	0-17	0-18	479.7819	5143
0	9	0-9	X76 Hari	0.5	0-19	0-20	1569.661	22984.25
0	10	0-10	X76 Hari	0.5	0-21	0-22	2171.172	17218.25
0	11	0-11	X11Tidak Stres	0.5	0-23	0-24	1520.987	22737.5
0	12	0-12	X275 - 94 Tahun	0.5	0-25	0-26	1671.615	41827.25
0	13	0-13	X76 Hari	0.5	0-27	0-28	1280.615	17747
0	14	0-14	X11Tidak Stres	0.5	0-29	0-30	1724.844	73129.25
...
...
299	0	299-0	X9> 1 kali per hari	0.5	299-1	299-2	0.163798	179758.6
299	1	299-1	X63 Hari	0.5	299-3	299-4	0.262663	144404.8
299	2	299-2	X81 kali per hari	0.5	299-5	299-6	0.969194	35353.75

299	3	299-3	X75 Hari	0.5	299-7	299-8	0.97664	134091
299	4	299-4	X1Perempuan	0.5	299-9	299-10	4.42487	10313.82
299	5	299-5	X81 - 2 kali per minggu	0.5	299-11	299-12	1.459568	28116.17
299	6	299-6	X62 Hari	0.5	299-13	299-14	5.859109	7237.581
299	7	299-7	X91 kali per hari	0.5	299-15	299-16	1.191755	127210.4
Pohon	Node	ID	Fitur	Split	Yes	No	Gain	Cover
299	8	299-8	X81 kali per hari	0.5	299-17	299-18	5.746738	6880.61
299	9	299-9	X3Ya	0.5	299-19	299-20	8.86935	3686.168
299	10	299-10	X3Ya	0.5	299-21	299-22	16.00874	6627.651
299	11	299-11	X4Ya	0.5	299-23	299-24	4.538648	17385.63
299	12	299-12	X10Tidak pernah	0.5	299-25	299-26	4.857442	10730.54
299	13	299-13	X74 Hari	0.5	299-27	299-28	11.7902	5857.167
299	14	299-14	X101 - 2 kali per minggu	0.5	299-29	299-30	14.67683	1380.413
299	0	299-0	X9> 1 kali per hari	0.5	299-1	299-2	0.163798	179758.6

- Tingkat kepentingan fitur berdasarkan jumlah nilai gain masing-masing fitur yang telah distandarisasi

No	Fitur	Gain
1	Jenis Kelamin Perempuan	37,13%
2	Umur 35 - 54 Tahun	31,36%
3	Aktifitas Berat	7,5%
4	Makan Sayur 6 Hari	5,81%
5	Stress	3,4%
6	Makan Buah 6 Hari	3%
7	Aktifitas Sedang	1,58%
8	Merokok	1,53%
9	Umur 55 - 74 Tahun	1,52%
10	Makan Buah 7 Hari	1,49%
11	Umur 75 - 94 Tahun	0,96%
12	Makan Berlemak lebih dari sekali per Hari	0,79%
13	Tidak Pernah Makan Buah	0,51%
No	Fitur	Gain
14	Tidak Pernah Makan Berlemak	0,49%
15	Makan Buah 2 Hari	0,37%
16	Makan Asin sekali per Hari	0,25%
17	Makan Asin 3 - 6 per Minggu	0,22%
18	Makan Manis 1 - 2 per Minggu	0,22%
19	Makan Asin lebih dari sekali per Hari	0,21%
20	Makan Asin 1 - 2 kali per Minggu	0,2%
21	Makan Berlemak 1 - 2 per Minggu	0,13%
22	Makan Berlemak 3 - 6 per Minggu	0,13%
23	Tidak Pernah Makan Asin	0,12%
24	Makan Manis sekali per Hari	0,12%
25	Makan Berlemak sekali per Hari	0,12%
26	Makan Manis lebih dari sekali per Hari	0,11%

27	Makan Manis 3 - 6 kali per Minggu	0,10%
28	Makan Buah 3 Hari	0,08%
29	Makan Sayur 4 Hari	0,08%
30	Tidak Pernah Makan Manis	0,08%
31	Makan Buah 4 Hari	0,08%
32	Makan Sayur 5 Hari	0,08%
33	Makan Buah 5 Hari	0,07%
34	Makan Sayur 2 Hari	0,07%
35	Makan Sayur 3 Hari	0,05%
36	Tidak Pernah Makan Sayur	0,03%
37	Makan Sayur 7 Hari	0,02%

Lampiran 6 Contoh Pembentukan Model Pohon *XGBoost*

Contoh pembentukan model *XGBoost* dengan 2 pohon klasifikasi dan 3 variabel acak menggunakan data (X, Y) berikut.

Instance	X1	X2	X3	Y
1	1	1	2	0
2	1	2	2	1
3	1	2	1	1
4	2	1	1	0
5	2	2	2	1
6	2	3	2	1
7	1	3	1	0

- A. Membangun model pohon klasifikasi *XGBoost* ke-1
2. Menentukan probabilitas awal F_0

Ditentukan probabilitas awal pada pembentukan pohon klasifikasi ke-1 yaitu 0,5. Nilai F_0 diperoleh berdasarkan probabilitas kelas Y yaitu 0 dan 1.

3. Menghitung nilai residual

$$Residual = Y - F_0$$

Instance (I)	X1	X2	X3	Y	Probabilitas \hat{Y}	Residual
1	1	1	2	0	0,5	-0,5
2	1	2	2	1	0,5	0,5
3	1	2	1	1	0,5	0,5
4	2	1	1	0	0,5	-0,5
5	2	2	2	1	0,5	0,5
6	2	3	2	1	0,5	0,5
7	1	3	1	0	0,5	-0,5

4. Membangun model 1 berdasarkan pohon XGBoost

Latih model pertama M1 menggunakan data $(X, residual)$.

a. Menentukan simpul 1

Menentukan nilai *similarity weight* dan *gain* untuk menetapkan simpul.

$$Similarity\ weight\ (SW) = \frac{\sum Residual^2}{\sum(F_0(1 - F_0) + \lambda)}$$

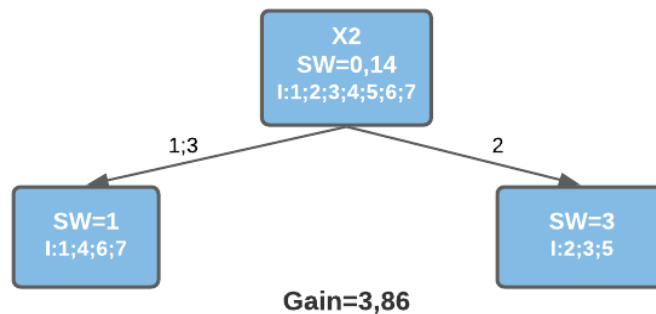
$$Gain = SW_{kiri} + SW_{kanan} - SW_{awal}$$

$$Cover = \sum (F_0(1 - F_0))$$

Variabel	Partisi	Total Instance	Output		Total Residual	SW	Cover
			1	0			
Total		7	4	3	0,5	0,14	1,75
X1	= 1	4	2	2	0	0	1
	= 2	3	2	1	0,5	0,33	0,75
X2	= 1	3	1	2	-0,5	0,33	0,50
	= 2	2	0	2	-1	2	0,75

	= 3	2	1	1	0	0	0,50
	= 1;2	5	3	2	0,5	0,20	1,25
	= 1;3	4	1	3	-1	1	1
	= 2;3	5	4	1	1,5	1,80	1,25
X3	= 1	3	1	2	-0,5	0,33	0,75
	= 2	4	3	1	0,5	0,25	1

Variabel	Partisi	Gain
X1	1 2	0,19
X2	1 2;3	3,66
X2	1;2 3	0,06
X2	1;3 2	3,86
X3	1 2	0,44



Variabel yang dipilih untuk dijadikan simpul adalah X2 berdasarkan nilai Gain = 3,86 dengan partisi 1;3 dan 2 sebagai cabangnya.

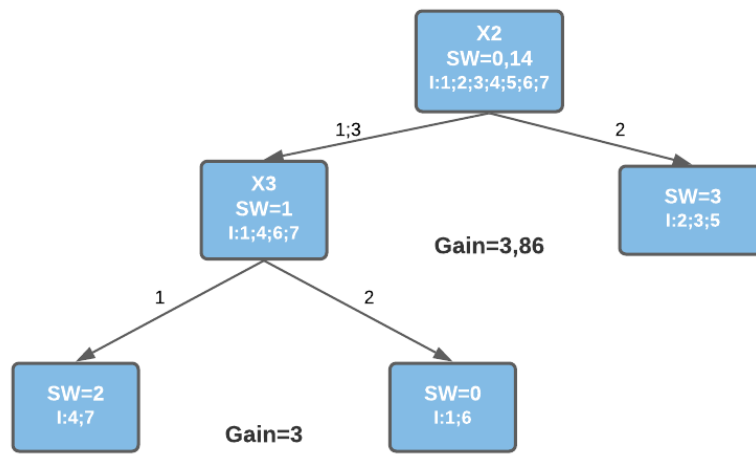
b. Menentukan simpul 1.1 berdasarkan partisi = 1;3

Instance	X1	X2	X3	Y	Probabilitas (\hat{Y})	Residual
1	1	1	2	0	0,5	-0,5
4	2	1	1	0	0,5	-0,5
6	2	3	2	1	0,5	0,5
7	1	3	1	0	0,5	-0,5

Variabel	Partisi	Total Instance	Output		Total Residual	SW	Cover
			1	0			
Total		4	1	3	-1	1	1
X1	= 1	2	0	2	-1	2	0,5

	= 2	2	1	1	0	0	0,5
X3	= 1	2	0	2	-1	2	0,5
	= 2	2	1	1	0	0	0,5

Variabel	Partisi		Gain
X1	1	2	1
X3	1	2	3



Variabel yang dipilih untuk dijadikan simpul adalah X3 berdasarkan nilai Gain = 3 dengan partisi 1 dan 2 sebagai cabangnya.

c. Menentukan simpul 1.1.1 berdasarkan partisi = 1

Instance	X1	X2	X3	Y	Probabilitas (\hat{Y})	Residual
4	2	1	1	0	0,5	-0,5
7	1	3	1	0	0,5	-0,5

Variabel	Partisi	Total Instance	Output		Total Residual	SW	Cover
			1	0			
Total		2	0	2	-1	2	0,5
X1	= 1	1	0	1	-0.5	1	0,25
	= 2	1	0	1	-0.5	1	0,25

Variabel	Partisi		Gain
X1	1	2	0

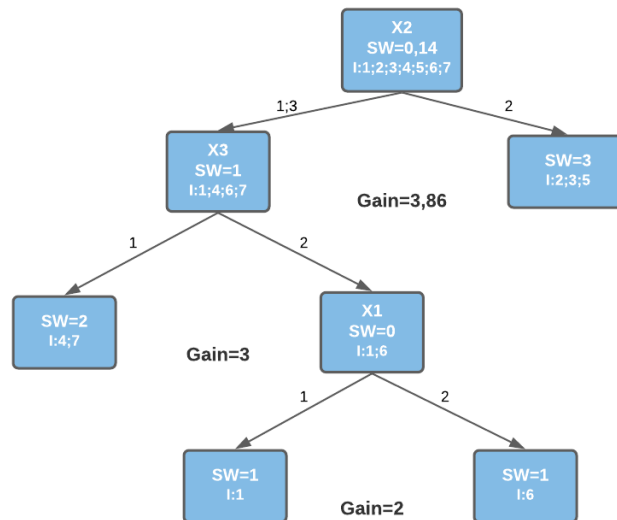
Percabangan berhenti karena nilai Gain = 0.

d. Menentukan simpul 1.1.2 berdasarkan partisi = 2

Instance	X1	X2	X3	Y	Probabilitas (\hat{Y})	Residual
1	1	1	2	0	0,5	-0,5
6	2	3	2	1	0,5	0,5

Variabel	Partisi	Total Instance	Output		Total Residual	SW	Cover
			1	0			
Total		2	1	1	0	0	0,5
X1	= 1	1	0	1	-0,5	1	0,5
	= 2	1	1	0	0,5	1	0,5

Variabel	Partisi	Gain
X1	1 2	2



Variabel yang dipilih untuk dijadikan simpul adalah X1 berdasarkan nilai Gain = 2 dengan partisi 1 dan 2 sebagai cabangnya.

e. Menentukan simpul 1.2

Instance	X1	X2	X3	Y	Probabilitas (\hat{Y})	Residual
----------	----	----	----	---	-------------------------------	----------

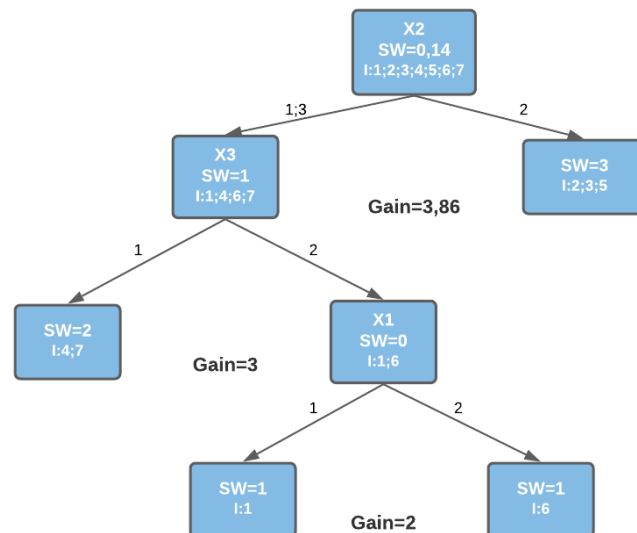
2	1	2	2	1	0,5	0,5
3	1	2	1	1	0,5	0,5
5	2	2	2	1	0,5	0,5

Variabel	Partisi	Total Instance	Output		Total Residual	SW	Cover
			1	0			
Total		3	3	0	1,5	3	0,75
X1	= 1	2	2	0	1	2	0,5
	= 2	1	1	0	0	1	0,25
X3	= 1	1	1	0	0,5	1	0,25
	= 2	2	2	0	1	2	0,5

Variabel	Partisi	Gain
X1	1 2	0
X3	1 2	0

Percabangan berhenti karena nilai Gain = 0.

f. Model pohon klasifikasi XGBoost ke-1



$$\text{Nilai Output Daun} = \frac{\sum \text{Residual}}{\sum (F0_i(1 - F0_i) + \lambda)}$$

Instance	X1	X2	X3	Y	Probabilitas (\hat{Y})	Residual	Output daun M1
1	1	1	2	0	0,5	-0,5	-2
2	1	2	2	1	0,5	0,5	1,67
3	1	2	1	1	0,5	0,5	1,67
4	2	1	1	0	0,5	-0,5	-1,5
5	2	2	2	1	0,5	0,5	1,67
6	2	3	2	1	0,5	0,5	2
7	1	3	1	0	0,5	-0,5	-1,5

B. Membangun model pohon klasifikasi XGBoost ke-2

1. Mentransformasi probabilitas awal F_0 untuk menentukan base model untuk pohon klasifikas XGBoost ke-2

$$\log(odds) = \log\left(\frac{p}{1-p}\right)$$

$$\log(odds) = \log\left(\frac{0.5}{1-0.5}\right)$$

$$\log(odds) = 0$$

Diperoleh base model = 0

2. Menentukan probabilitas F_1 dan nilai residual masing-masing instance

$$F_1 = \text{base model} + \text{Learning rate } \varepsilon(\text{Output daun M1})$$

Setelah menentukan F_1 , dilakukan transformasi probabilitas terhadap

F_1 menggunakan *logistic function*:

$$\text{Sigmoid } F_1 = \frac{1}{1 + e^{-F_1}}$$

Instance	X1	X2	X3	Y	Output daun M1	Base model	Learning rate ε	F1	Probabilitas	Residual
1	1	1	2	0	-2	0	0,1	-0,2	0,45	-0,45
2	1	2	2	1	1,67	0	0,1	0,17	0,54	0,46
3	1	2	1	1	1,67	0	0,1	0,17	0,54	0,46
4	2	1	1	0	-1,5	0	0,1	- 0,15	0,46	-0,46
5	2	2	2	1	1,67	0	0,1	0,17	0,54	0,46
6	2	3	2	1	2	0	0,1	0,2	0,55	0,45

7	1	3	1	0	-1,5	0	0,1	-	0,46	-0,46
									0,15	

3. Membangun model M2 berdasarkan pohon XGBoost

Latih model kedua M2 menggunakan data $(X, residual)$.

a. Menentukan simpul 1

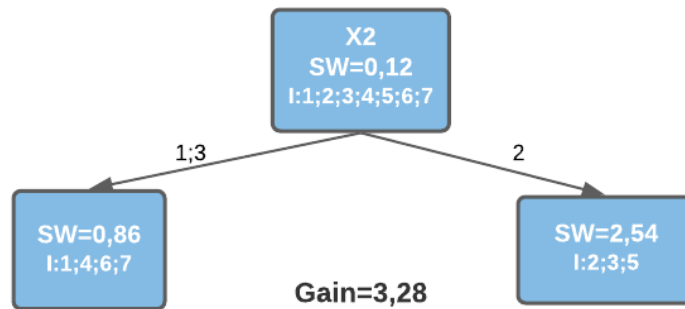
Menentukan nilai *similarity weight* dan *gain* untuk menetapkan simpul.

$$Similarity\ weight\ (SW) = \frac{\sum Residual^2}{\sum (F1_{i-1}(1 - F1_{i-1}) + \lambda)}$$

$$Gain = SW_{Kiri} + SW_{Kanan} - SW_{awal}$$

Variabel	Partisi	Total Instance	Output		Total Residual	SW
			1	0		
Total		7	4	3	0,45	0,12
X1	= 1	4	2	2	0	0
	= 2	3	2	1	0,45	0,27
X2	= 1	3	0	2	-0,91	1,68
	= 2	2	3	0	1,38	2,54
	= 3	2	1	1	-0,01	0
	= 1;2	5	3	2	0,46	0,17
	= 1;3	4	1	3	-0,93	0,86
	= 2;3	5	4	1	1,36	1,50
X3	= 1	3	1	2	-0,47	0,29
	= 2	4	3	1	0,92	0,85

Variabel	Partisi	Gain
X1	1 2	0,15
X2	1 2;3	3,06
X2	1;2 3	0,06
X2	1;3 2	3,28
X3	1 2	1,02



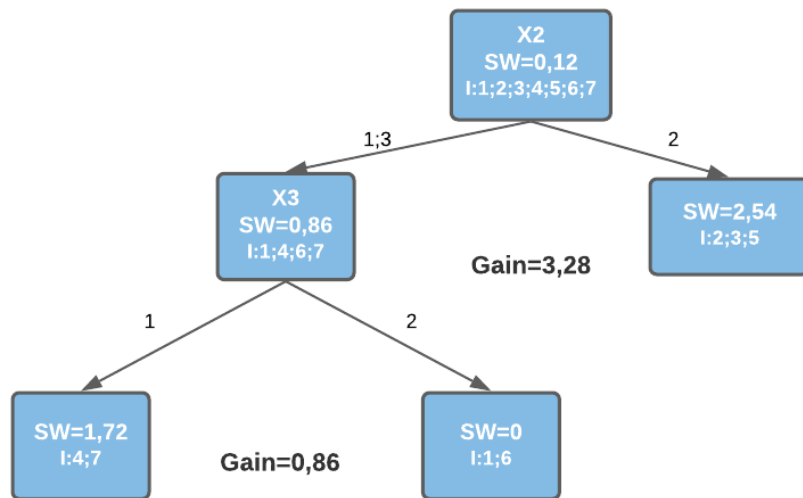
Variabel yang dipilih untuk dijadikan simpul adalah X2 berdasarkan nilai Gain = 3,28 dengan partisi 1;3 dan 2 sebagai cabangnya.

b. Menentukan simpul 1.1 berdasarkan partisi 1;3

Instance	X1	X2	X3	Y	Sigmoid F1	Residual
1	1	1	2	0	0,45	-0,45
4	2	1	1	0	0,46	-0,46
6	2	3	2	1	0,55	0,45
7	1	3	1	0	0,46	-0,46

Variabel	Partisi	Total Instance	Output		Total Residual	SW
			1	0		
Total		4	1	3	-0,93	0,86
X1	= 1	2	0	2	-0,91	1,68
	= 2	2	1	1	-0,01	0
X3	= 1	2	0	2	-0,93	1,72
	= 2	2	1	1	0	0

Variabel	Partisi	Gain
X1	1 2	0,82
X3	1 2	0,86



Variabel yang dipilih untuk dijadikan simpul adalah X3 berdasarkan nilai Gain = 0,86 dengan partisi 1 dan 2 sebagai cabangnya.

c. Menentukan simpul 1.1.1 berdasarkan partisi = 1

Instance	X1	X2	X3	Y	Sigmoid F1	Residual
4	2	1	1	0	0,46	-0,46
7	1	3	1	0	0,46	-0,46

Variabel	Partisi	Total Instance	Output		Total	SW
			1	0		
Total		2	0	2	-0,93	1,72
X1	= 1	1	0	1	-0,46	0,86
	= 2	1	0	1	-0,46	0,86

Variabel	Partisi	Gain
X1	1 2	0

Percabangan berhenti karena nilai Gain = 0.

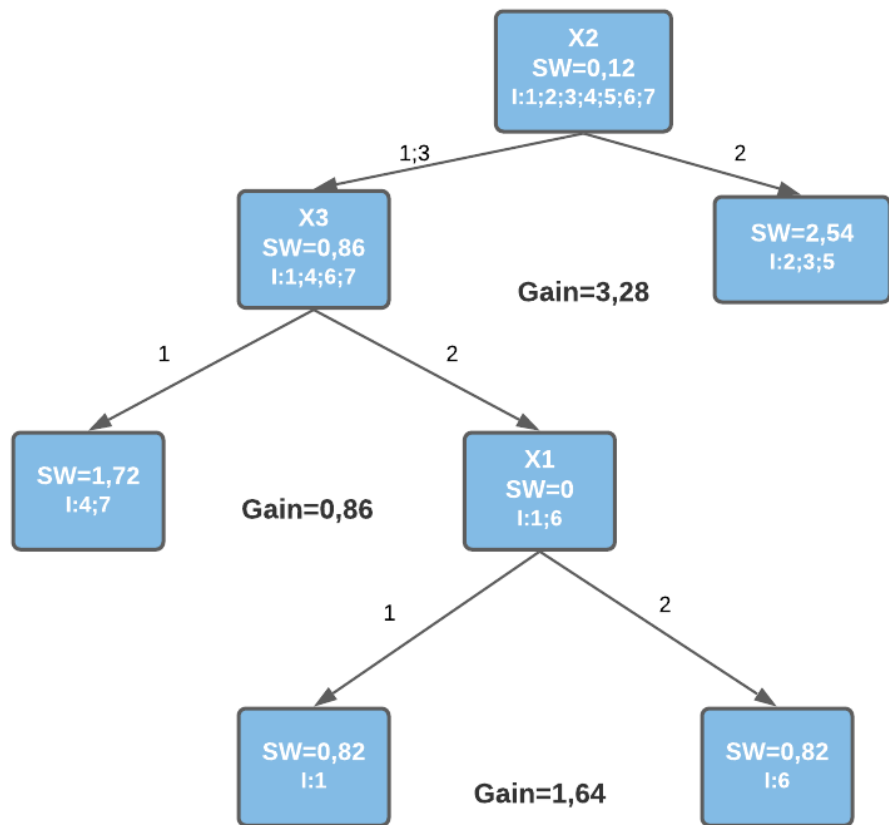
d. Menentukan simpul 1.1.2 berdasarkan partisi = 2

Instance	X1	X2	X3	Y	Sigmoid F1	Residual
1	1	1	2	0	0,45	-0,45

6	2	3	2	1	0,55	0,45
---	---	---	---	---	------	------

Variabel	Partisi	Total Instance	Output		Total	SW
			1	0		
Total		2	1	1	0	0
X1	= 1	1	0	1	-0,45	0,82
	= 2	1	1	0	0,45	0,82

Variabel	Partisi	Gain
X1	1 2	1,64



Variabel yang dipilih untuk dijadikan simpul adalah X1 berdasarkan nilai Gain = 1,64 dengan partisi 1 dan 2 sebagai cabangnya.

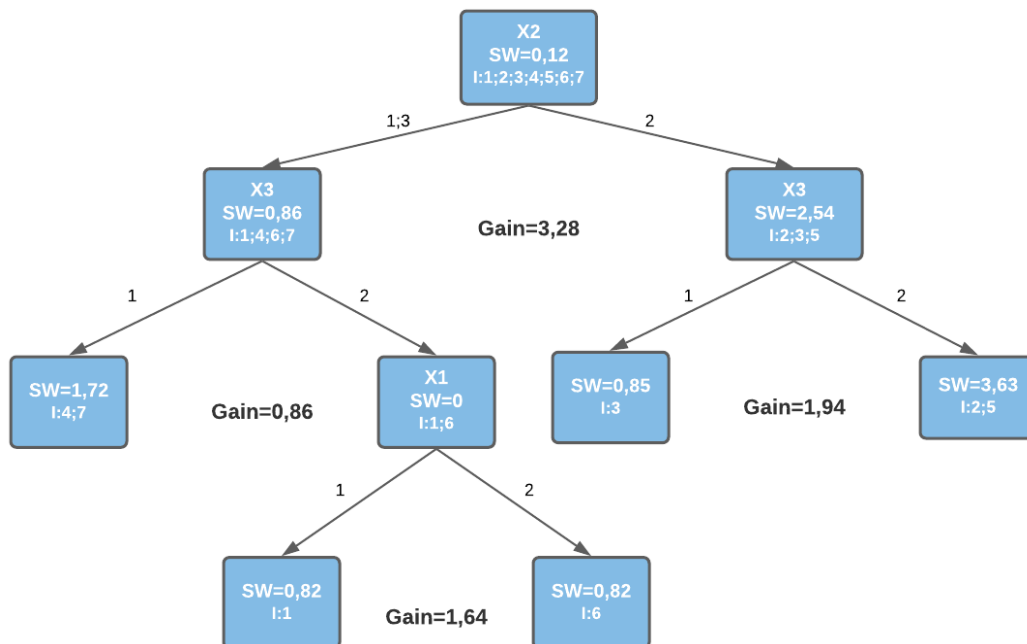
e. Menentukan simpul 1.2

Instance	X1	X2	X3	Y	Probabilitas	Residual
----------	----	----	----	---	--------------	----------

2	1	2	2	1	0,54	0,46
3	1	2	1	1	0,54	0,46
5	2	2	2	1	0,54	0,46

Variabel	Partisi	Total Instance	Output		Total Residual	SW
			1	0		
Total		3	3	0	1,38	2,54
X1	= 1	2	2	0	0,92	1,69
	= 2	1	1	0	0,46	0,85
X3	= 1	1	1	0	0,46	0,85
	= 2	2	2	0	0,92	3,63

Variabel	Partisi	Gain
X1	1 2	0
X3	1 2	1,94



Variabel yang dipilih untuk dijadikan simpul adalah X3 berdasarkan nilai Gain = 1,94 dengan partisi 1 dan 2 sebagai cabangnya.

f. Menentukan simpul 1.2.2 berdasarkan partisi = 2

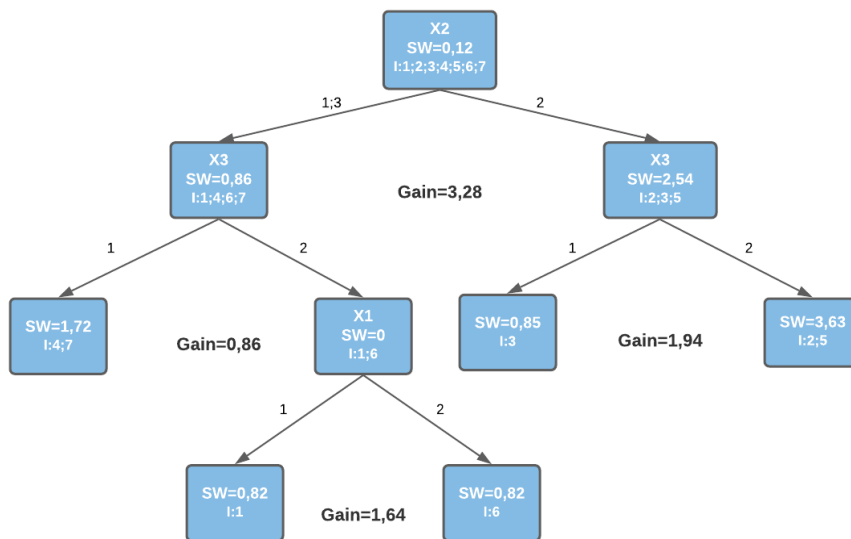
Instance	X1	X2	X3	Y	Probabilitas	Residual
2	1	2	2	1	0,54	0,46
5	2	2	2	1	0,54	0,46

Variabel	Partisi	Total Instance	Output		Total	SW
			1	0		
Total		2	2	0	0,92	1,69
X1	= 1	0	0	0	0,46	0,85
	= 2	2	2	0	0,46	0,85

Variabel	Partisi	Gain
X1	1 2	0

Percabangan berhenti karena nilai Gain = 0.

g. Model pohon klasifikasi XGBoost ke-2



Instance	Y	\hat{Y} M0	Residual M0	Output daun M1	\hat{Y} M1	Residual M1	Output daun M2	\hat{Y} M2	Residual M2
1	0	0,5	-0,5	-2	0,45	-0,45	-1,82	0,41	-0,41
2	1	0,5	0,5	1,67	0,54	0,46	1,53	0,58	0,42
3	1	0,5	0,5	1,67	0,54	0,46	1,85	0,59	0,41
4	0	0,5	-0,5	-1,50	0,46	-0,46	-1,39	0,43	-0,43

5	1	0,5	0,5	-1,67	0,54	0,46	1,38	0,58	0,42
6	1	0,5	0,5	2	0,55	0,45	1,82	0,59	0,41
7	0	0,5	-0,5	-1,50	0,46	-0,46	-1,39	0,43	-0,43

C. Kinerja Klasifikasi

Hasil klasifikasi menggunakan model dapat dilihat dengan menguji model pada data uji. Model yang diperoleh akan diuji menggunakan data berikut.

Instance	X1	X2	X3	Y
1	2	3	1	1
2	1	1	2	0
3	2	2	2	0
4	1	3	1	0
5	2	1	2	1

diperoleh:

Pohon Klasifikasi	Instance	X1	X2	X3	Y	\hat{Y} M0	Residual M0	Output daun M1	\hat{Y} M1	Residual M1
1	1	2	3	1	1	0,5	0,5	0	0,5	0,5
	2	1	1	2	0	0,5	-0,5	-2	0,45	-0,45
	3	2	2	2	0	0,5	-0,5	-2	0,45	0,45
	4	1	3	1	0	0,5	-0,5	-2	0,45	0,45
	5	2	1	2	1	0,5	0,5	2	0,55	0,45

Pohon Klasifikasi	Instance	X1	X2	X3	Y	Output daun M2	\hat{Y} M2	Residual M2
2	1	2	3	1	1	0,1	0,503	0,497
	2	1	1	2	0	-1,82	0,406	-0,406
	3	2	2	2	0	-1,82	0,406	-0,406
	4	1	3	1	0	0,1	0,453	-0,453
	5	2	1	2	1	1,82	0,594	0,406

Dari hasil diatas, diperoleh matriks konfusi dengan *threshold* = 0,5 sebagai berikut:

Matriks Konfusi		Prediksi	
		1	0
Aktual	1	2	0
	0	0	0

0	0	3
---	---	---

Berdasarkan matriks konfusi diatas, diperoleh akurasi model adalah 100%.

D. Tingkat Kepentingan Fitur

Pohon Klasifikasi	Fitur	Gain
1	X2	3,86
1	X3	3
1	X1	2
2	X2	2,8
2	X3	3,28
2	X1	1,64

Berdasarkan tabel diatas, tingkat kepentingan fitur dengan standarisasi terhadap gain sebagai berikut:

Rank	Fitur	Gain
1	X2	6,66
2	X3	6,28
3	X1	3,64

Berdasarkan tabel diatas, diperoleh kesimpulan X2 sebagai fitur yang paling penting dengan nilai gain sebesar 6,66.