

PREDIKSI PENYEBAB UTAMA KEMISKINAN DI INDONESIA DENGAN ALGORITMA *DECISION TREE* C5.0

Velia Rahmadi¹, Risma Yulistiani², Ramadhani Sheffi Tiara Gultom³, Mayanda Mega Santoni⁴
Program Studi S-1 Informatika, Fakultas Ilmu Komputer
Universitas Pembangunan Veteran Jakarta, DKI Jakarta
Jl. RS Fatmawati, Pondok Labu, Cilandak, Jakarta Selatan, DKI Jakarta 12450, Indonesia
veliarhmd@gmail.com

Abstrak. Indonesia merupakan negara yang memiliki jumlah penduduk paling banyak keempat di dunia. Dan termasuk ke dalam negara yang masih berkembang. Banyaknya jumlah penduduk tersebut tidak sesuai dengan fasilitas yang diberikan untuk masyarakat yang tinggal di Indonesia. Sehingga menimbulkan masalah sosial berupa kemiskinan. Kemiskinan terjadi di berbagai provinsi di Indonesia, terdapat beberapa faktor yang menyebabkan kemiskinan. Dari berbagai faktor tersebut, penulis ingin melakukan prediksi penyebab utama kemiskinan yang terjadi di Indonesia sehingga dapat mempermudah Pemerintah dalam mengatasi permasalahan kemiskinan. Model dibuat dengan algoritma *Decision Tree* C5.0 yang diharapkan mampu membuat model klasifikasi yang akurat untuk digunakan sebagai prediksi penyebab utama kemiskinan di Indonesia

Kata Kunci: *Decision Tree*, Indonesia, Kemiskinan

1. Pendahuluan

Indonesia merupakan negara yang memiliki jumlah penduduk paling banyak keempat di dunia dari laporan Departemen Populasi Divisi Urusan Sosial dan Ekonomi Perserikatan Bangsa-Bangsa (PBB) pada Juni 2017. Sebagai negara yang belum seabad merdeka, masalah kependudukan menjadi masalah yang besar bagi pemerintah. Tingkat kemiskinan dan kesejahteraan masyarakat menjadi masalah utama yang hingga saat ini masih sangat berpengaruh bagi kemajuan negara. Menurut [1], Meningkatkan kualitas hidup masyarakat Indonesia terdapat dalam nawacita agenda pembangunan tahun 2015-2019. Adapun kemiskinan diartikan sebagai suatu keadaan dimana seseorang tidak sanggup memelihara dirinya sendiri sesuai dengan taraf kehidupan kelompok dan juga tidak mampu memanfaatkan tenaga mental, maupun fisiknya dalam kelompok tersebut.

Kemiskinan yang terjadi di Indonesia tidak dapat dihindari. Karena kebutuhan hidup semakin tinggi namun tidak diimbangi dengan peningkatan kualitas SDM dibidang Pendidikan membuat indeks pembangunan manusia semakin rendah[2]. Oleh karena itu, banyak ditemukan masyarakat yang tidak memiliki pekerjaan[1]. Dan juga menimbulkan masalah baru yaitu rendahnya angka partisipasi sekolah[3]. Kemudian di pemukiman padat penduduk, akses air bersih masih sulit ditemukan[4] dan sanitasi yang kurang layak[5], tingginya keluhan kesehatan dari masyarakat[6] masih menjadi faktor penentu kemiskinan.

Teknik data mining dapat dimanfaatkan untuk melakukan prediksi tingkat kemiskinan yang terjadi di Indonesia. Salah satu metode data mining yang banyak digunakan untuk melakukan prediksi yakni metode pohon keputusan (*Decision Tree*). Metode pohon keputusan merupakan metode yang mudah untuk dipahami, mudah untuk menginterpretasi model yang dihasilkan, serta dapat melakukan prediksi dalam penentuan keputusan. Pada penelitian [7] analisis kemiskinan dengan algoritma J48 *Decision Tree* menghasilkan model prediksi dengan akurasi 88.6% dengan menggunakan data Indeks Pembangunan Manusia dari tahun 2010-2017.

Pada penelitian ini, data berasal dari BPS (Badan Pusat Statistik) dengan atribut 6 penyebab kemiskinan dari tahun 2007-2017. Metode data mining yang digunakan adalah metode *Decision Tree* yakni algoritma C5.0. Hasil akhir dari penelitian ini yaitu sebuah model prediksi yang diharapkan dapat memberikan saran kepada Pemerintah

tentang faktor utama kemiskinan yang harus diselesaikan terlebih dahulu. Sehingga Pemerintah dapat melakukan aksi yang tepat sasaran dalam mengatasi masalah kemiskinan di Indonesia.

2. Landasan Teori

2.1 Data Mining

Data mining adalah proses mendapatkan pola penting pada sekumpulan data dengan menggunakan teknik statistik dan matematika[8]. Data mining juga memiliki banyak fungsi yang dapat digunakan dalam kasus tertentu untuk menyelesaikan masalah. Fungsi data mining diantaranya; *classification, clustering, association, regression, dan forecasting.*

2.2 Decision Tree

Decision Tree atau pohon keputusan adalah salah satu teknik klasifikasi. Dengan menggunakan pohon keputusan dapat memberikan kemudahan terhadap gambaran hubungan antara faktor-faktor yang terjadi pada suatu masalah. Pohon keputusan merupakan salah satu metode klasifikasi yang bisa melakukan prediksi dengan menggunakan struktur pohon. Struktur pohon tersebut dapat dijadikan aturan yang berguna untuk melakukan prediksi[9].

2.3 Algoritma C5.0

Algoritma C5.0 adalah pembaharuan dari algoritma ID3 dan C4.5 yang mengandalkan nilai informasi gain tertinggi untuk root bagi node selanjutnya. Algoritma C5.0 dirumuskan dengan semua data menjadi akar dari pohon keputusan dan atribut akan menjadi pembagi dari sampel tersebut. Berikut ini adalah rumus pengukuran atribut:

$$I(A) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Dimana:

(1)

Info(D) = informasi untuk klasifikasi label kelas tuple di D
 P_i = peluang bukan nol pada sebuah tuple acak di D
 Log = basis 2 (dalam bit)

Selain itu, Info(D) dapat dikatakan sebagai entropy. Nilai entropy yang dihasilkan untuk menghasilkan tuple dari D berdasarkan partisi oleh A yaitu:

$$I(A) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \log_2 \left(\frac{|D_i|}{|D|} \right)$$

Untuk mendapatkan nilai information gain pada atribut A selanjutnya digunakan rumus:

$$IG(A) = I(D) - I(A|D) \tag{3}$$

Gain (A) akan menunjukkan banyaknya cabang yang akan diperoleh pada A. Atribut A dengan information gain tertinggi. Atribut pada node N berdasarkan informasi, gain(A).

2.4 Pruning

Pruning merupakan proses yang digunakan untuk menghilangkan beberapa cabang yang tidak diperlukan pada struktur pohon keputusan. Nilai error pruning dapat menggunakan rumus:

$$E = \frac{z + \sqrt{z^2 + 4cn}}{2(1+c)}$$

Dimana:

r = nilai perbandingan error rate

n = total sampel

c = confidence level

$z = \Phi^{-1}(c)$

2.5 Confusion Matrix

Confusion Matrix adalah metode yang digunakan untuk evaluasi dengan tabel matrix. Hasil evaluasi tersebut menggunakan confusion matrix untuk menghasilkan nilai akurasi. Berikut adalah rumus akurasi:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \quad (5)$$

Dimana:

TP = jumlah data positif yang terklasifikasi dengan benar oleh sistem

TN = jumlah data negative yang terklasifikasi dengan benar oleh sistem

FN = jumlah data negative namun terklasifikasi salah oleh sistem

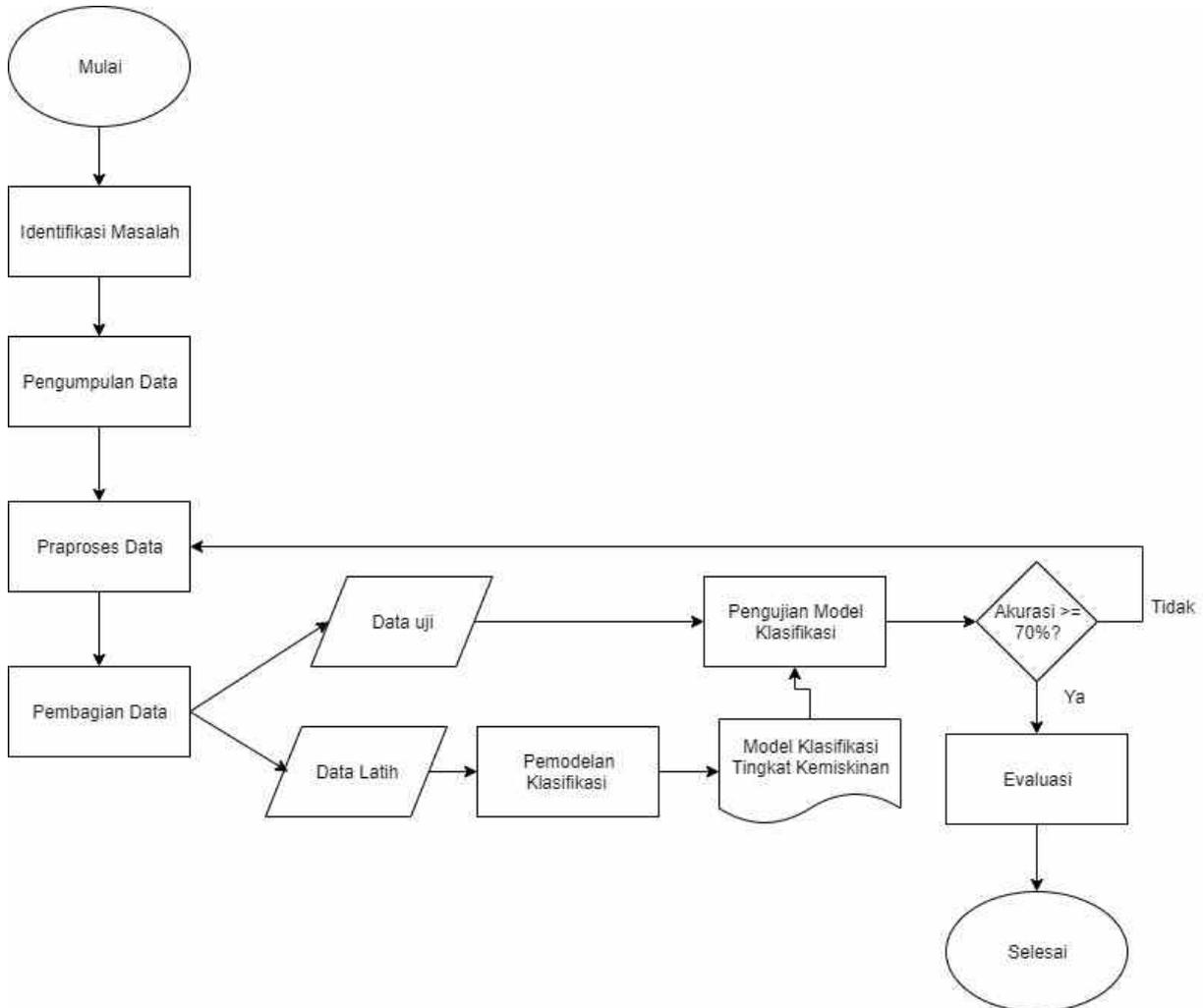
FP = jumlah data positif namun terklasifikasi salah oleh sistem

Laju Error yang dapat dihitung adalah:

$$Error\ Rate = \frac{FP + FN}{TP + FP + FN + TN} \times 100\% \quad (6)$$

3. Metode Penelitian

Pada penelitian dilakukan beberapa tahapan seperti identifikasi masalah, pengumpulan data, kemudian dilanjutkan dengan praproses data, selanjutnya pembagian data menjadi dua kelompok yaitu data latih (*training*) dan data uji (*testing*). Data latih dilakukan untuk mendapatkan model klasifikasi tingkat kemiskinan menggunakan algoritma C5.0. Pengujian model klasifikasi ini dilakukan pada data pengujian untuk mendapatkan tingkat akurasi. Tahapan praproses sampai dengan pengujian model klasifikasi akan dilakukan terus menerus hingga diperoleh tingkat akurasi lebih dari 70%. Pada tahapan terakhir dilakukan evaluasi pada model klasifikasi. Keseluruhan tahapan pada penelitian dapat dilihat pada Gambar 1.



Gambar 1 Perancangan Sistem

3.1 Identifikasi Masalah

Pada penelitian ini penulis mengangkat tema tentang kemiskinan yang terjadi di Indonesia. Dari banyaknya faktor penyebab dari kemiskinan yang terjadi. Penulis ingin melakukan prediksi faktor penyebab utama yang menyebabkan tingginya tingkat kemiskinan di Indonesia. Sehingga Pemerintah mendapatkan informasi faktor penyebab kemiskinan yang paling berpengaruh dan dapat dilakukan penyelesaian dengan cepat dan efisien.

3.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dari situs badan pusat statistik Indonesia (<http://www.bps.go.id>). Data yang didapatkan meliputi 6 atribut (variabel bebas) yang menjadi faktor kemiskinan di setiap provinsi Indonesia dalam jangka waktu 2007-2017. Kemudian digunakan data persentase penduduk miskin sebagai kelas (variabel terikat). Data yang digunakan dapat dilihat pada tabel 1.

Tabel 1 Atribut data penelitian

No	Atribut	Kode
1	Persentase Keluhan Kesehatan	X1
2	Persentase Angka Partisipasi Sekolah	X2
3	Persentase Indeks Pembangunan Manusia	X3
4	Persentase Pengangguran Terbuka	X4
5	Persentase Sumber Air Minum	X5
6	Persentase Kelayakkan Sanitasi	X6
7	Persentase Penduduk Miskin (kelas)	Y

3.3 Praproses Data

Pada tahap praproses data yang telah dikumpulkan dilakukan proses integrasi data yaitu menyatukan data ke dalam file format .csv untuk mempermudah proses data mining. Setelah itu, dilakukan pengisian data kosong (*missing values*) dengan menghitung nilai rata-rata (*mean*):

$$\frac{\sum x_i}{n}$$

Dimana :

x_i = nilai *mean*

x = nilai masing-masing data

n = jumlah data

Selanjutnya dilakukan perhitungan nilai interval dari setiap variabel untuk menentukan kelas.

Tabel 2 Interval pada data kemiskinan

	X1	X2	X3	X4	X5	X6	Y
Nilai Minimum	15.49	43.58	54.45	0	22.32	14.98	3.48
Nilai Maksimum	49.66	87.61	80.06	16.11	93.4	91.12	40.78
Interval	11.39	14.67	8.53	5.37	23.69	25.38	12.43

Setelah didapat interval seperti tabel diatas, label dirumuskan sebagai berikut:

1. If nilai_atribut = 0, maka label nilai atribut 0 atau tidak ada informasi
2. If nilai_atribut ≤ batas atas (ba) 1, maka nilai atribut 3 atau tingkat kemiskinan rendah
3. If nilai_atribut ≤ batas atas (ba) 2, maka nilai atribut 2 atau tingkat kemiskinan sedang
4. If nilai_atribut ≤ batas atas (ba) 3, maka nilai atribut 1 atau tingkat kemiskinan tinggi

Tahap terakhir yaitu transformasi data yang berguna melabeli setiap atribut dengan nilai yang sudah dikategorikan berdasarkan tingkat kemiskinan. Label ini didasari interval dari setiap variable yang digunakan pada tabel 2.

3.4 Pembagian data

Pada tahap ini data dibagi menjadi data latih (*training*) dan data uji (*testing*) untuk dilakukan proses klasifikasi. Data *training* digunakan untuk membangun pohon keputusan dan data *testing* untuk menguji model. Pembagian dilakukan dengan metode K-Fold Validation. Dengan K= 10 yang artinya proses pembagian data terjadi sebanyak 10 lipatan untuk data latih dan data uji.

3.5 Pengujian Model

Setelah tahap pembagian data dilakukan pembuatan model klasifikasi yang diharapkan mampu memiliki akurasi yang tertinggi. Fungsi dari Akurasi untuk menunjukkan tingkat kebenaran pengklasifikasian data terhadap kelas yang sebenarnya. Semakin tinggi tingkat akurasi maka semakin rendah kesalahan klasifikasi. Apabila akurasi dari model terbaik kurang dari 70% maka akan dilakukan praproses data kembali untuk mendapatkan akurasi yang lebih tinggi.

		Folds									
		1	2	3	4	5	6	7	8	9	10
Iteration	1	0.8378 378	0.702 7027	0.611 1111	0.783 7838	0.837 8378	0.666 6667	0.756 7568	0.666 6667	0.81 081	0.7 027
	2	0.6756 757	0.810 8108	0.666 6667	0.864 8649	0.783 7838	0.666 6667	0.837 8378	0.722 2222	0.75 675	0.6 756
	3	0.7297 297	0.756 7568	0.694 4444	0.702 7027	0.756 7568	0.833 3333	0.756 7568	0.694 4444	0.78 378	0.6 756
	4	0.7837 838	0.756 7568	0.694 4444	0.729 7297	0.783 7838	0.722 2222	0.810 8108	0.777 7778	0.72 972	0.7 027
	5	0.7567 568	0.756 7568	0.722 2222	0.729 7297	0.675 6757	0.694 4444	0.756 7568	0.666 6667	0.78 378	0.7 837
	6	0.7837 838	0.675 6757	0.777 7778	0.783 7838	0.729 7297	0.861 1111	0.675 6757	0.833 3333	0.59 459	0.8 108
	7	0.8108 108	0.702 7027	0.722 2222	0.810 8108	0.756 7568	0.638 8889	0.810 8108	0.777 7778	0.75 675	0.7 297
	8	0.7297 297	0.702 7027	0.722 2222	0.756 7568	0.729 7297	0.777 7778	0.756 7568	0.666 6667	0.70 270	0.7 567
	9	0.7297 297	0.810 8108	0.777 7778	0.783 7838	0.756 7568	0.722 2222	0.648 6486	0.722 2222	0.75 675	0.6 756
	10	0.7837 838	0.702 7027	0.75 5946	0.594 5946	0.783 7838	0.694 4444	0.837 8378	0.75 270	0.70 270	0.7 027

Gambar 2 Akurasi berdasarkan setiap fold dan iterasi semua fold

Metode yang dipilih dalam melakukan pengujian adalah K-fold Validation dengan banyak fold adalah k=10. Kemudian dilakukan iterasi sebanyak 10 kali untuk mencari akurasi data maksimal. data tersebut disimpan dalam array 10*10, yang mana data dihitung dari nilai rata-rata setiap fold dikali dengan nilai rata-rata setiap iterasi. Dari hasil pengujian tersebut, didapat akurasi terbesar sebanyak 86% dari iterasi kedua pada posisi fold keempat. Karena jumlah akurasi diatas 70%, dapat dikatakan penambangan data berhasil dan model yang didapatkan dapat digunakan untuk prediksi. Berikut adalah tabel *confusion matrix* yang diperoleh dari pengujian tersebut:

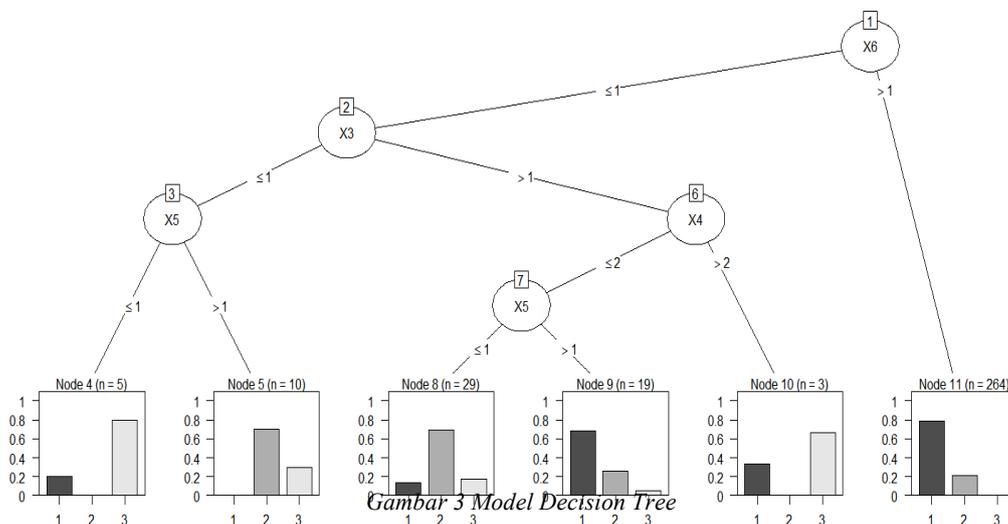
Tabel 3 Confusion Matrix

P	1	2	3
1	28	0	1
2	2	3	0

3	1	1	1
---	---	---	---

Dari tabel diatas, dapat dikatakan prediksi rendah yang sesuai fakta sejumlah 28, prediksi rendah yang salah terklasifikasi ada 1. Prediksi sedang sesuai dengan yang seharusnya ada 3, yang salah terklasifikasi ada 2. Prediksi tinggi terklasifikasi salah merata dengan klasifikasi yang sesuai.

4. Evaluasi



Gambar 3 merupakan struktur model pohon keputusan hasil dari klasifikasi faktor penyebab kemiskinan di Indonesia. Dari ke-6 atribut yang digunakan pada penelitian, atribut yang memiliki pengaruh terbesar adalah X6 yaitu Sanitasi Layak. Diikuti dengan indeks pembangunan manusia (X3), sumber air minum (X5) dan pengangguran terbuka (X4). Sedangkan, atribut keluhan kesehatan (X1) dan angka partisipasi sekolah (X2) tidak terlalu mempengaruhi hasil prediksi. Karena pada *Decision Tree* tersebut X6 berada di root berarti dapat disimpulkan tingkat kemiskinan di provinsi-provinsi Indonesia disebabkan oleh banyaknya sanitasi yang tidak layak.

Untuk mengetahui rules dari *Decision Tree* yaitu:

$$R1 = (IF(X6 \leq 1) \wedge (X3 \leq 1) \wedge (X5 \leq 1)) \rightarrow \text{NODE4}$$

$$R2 = (IF(X6 \leq 1) \wedge (X3 \leq 1) \wedge (X5 > 1)) \rightarrow \text{NODE5}$$

$$R3 = (IF(X6 \leq 1) \wedge (X3 > 1) \wedge (X4 \leq 2) \wedge (X5 \leq 1)) \rightarrow \text{NODE8}$$

$$R4 = (IF(X6 \leq 1) \wedge (X3 > 1) \wedge (X4 \leq 2) \wedge (X5 > 1)) \rightarrow \text{NODE9}$$

$$R5 = (IF(X6 \leq 1) \wedge (X3 > 1) \wedge (X4 > 2)) \rightarrow \text{NODE10}$$

$$R6 = (IF(X6 > 1)) \rightarrow \text{NODE11}$$

5. Kesimpulan

Penggunaan metode Pohon Keputusan yakni algoritma C5.0 pada tingkat kemiskinan berhasil memprediksi tingkat kemiskinan di Indonesia sesuai dengan faktor-faktor yang ada. Hasil prediksi yang diperoleh dilakukan percobaan sebanyak 10 kali iterasi. Percobaan dilakukan dengan 10 cross fold validation dengan 90% sebagai data latih dan 10% sebagai data uji. Hasil percobaan memperlihatkan dari 10 kali percobaan didapatkan rata-rata akurasi pohon keputusan sebesar 86%, dengan model terbaik yaitu pada model ke 4 di iterasi kedua Selain itu, aktivitas yang paling mempengaruhi prediksi tingkat kemiskinan provinsi-provinsi di Indonesia berdasarkan hasil klasifikasi yang dilakukan dari iterasi pertama sampai sepuluh yaitu atribut X6 (persentase sanitasi layak) karena atribut tersebut menghasilkan nilai information gain tertinggi dari pohon keputusan yang terbentuk.

6. Dokumentasi

6.1 Data Mentah

PROVINSI	TAHUN	X1	X2	X3	X4	X5	X6	Y
ACEH	2007	40.81	72.79	70.35	10.27	32.85	30.65	26.65
	2008	36.8	72.73	70.76	9.2	28.02	40.17	23.53
	2009	35.28	72.74	71.31	9.31	30.6	42.03	21.8
	2010	35.09	73.53	67.09	8.6	29.02	45.17	20.98
	2011	31.18	72.14	67.45	8.62	57.21	50.1	19.57
	2012	31.06	74.59	67.81	7.94	56.98	52.53	18.58
	2013	28.94	74.7	68.3	8.34	62.41	53.47	17.72
	2014	30.55	80.89	68.81	6.75	61.1	33.68	16.98
	2015	27.92	81.43	69.45	7.73	61.23	54.68	17.11
	2016	25.78	81.82	70	8.13	63.31	62.68	16.43
2017	24.85	82.15	70.6	7.39	64.85	63.38	15.92	
SUMATERA UTARA	2007	25.4	65.87	72.78	10.63	49.85	49.11	13.9
	2008	25.19	65.87	73.29	9.55	49.52	52.87	12.55
	2009	29.11	66.34	73.8	8.25	51.04	51.92	11.51
	2010	26.68	66.94	67.09	8.01	46.06	57.1	11.31
	2011	25.6	67.1	67.34	7.47	62.29	56.47	11.33
	2012	20.73	69.86	67.74	6.43	65.48	59.7	10.41
	2013	21.3	71.24	68.36	6.09	67.81	61.92	10.39
	2014	23.55	75.78	68.87	5.95	67.13	66.92	9.85

6.2 Data Hasil Transformasi

1	X1	X2	X3	X4	X5	X6	Y
2	3	2	2	2	1	1	2
3	2	2	2	2	1	1	2
4	2	2	2	2	1	2	2
5	2	3	2	2	1	2	2
6	2	2	2	2	2	2	2
7	2	3	2	2	2	2	2
8	2	3	2	2	2	2	2
9	2	3	2	2	2	1	2
10	2	3	2	2	2	2	2
11	1	3	2	2	2	2	2
12	1	3	2	2	2	2	2
13	1	2	3	2	2	2	1
14	1	2	3	2	2	2	1
15	2	2	3	2	2	2	1
16	1	2	2	2	2	2	1
17	1	2	2	2	2	2	1
18	1	2	2	2	2	2	1
19	1	2	2	2	2	2	1
20	1	3	2	2	2	3	1

6.3 Screenshot Code

```
#inisiasi variabel p_all<-
list() conf_mat_folds<-
list() y_train_all<-list()
y_test_all<-list()
model_folds<-list()
tmp_mean<-array(0, dim=c(10,10))
tmp_model<-list()
model_iter<-list()
conf_iter<-list()

#definisikan pengulangan n-iterasi n=10
for(i in 1:n){
#split data berdasarkan k=10
folds<- split(data, cut(sample(1:nrow(data)),10))
#definisikan class berdasarkan 6 atribut
form<- "y ~ x1 + x2 + x3 + x4 + x5 + x6"
#hitung error perhitungan prediksi
err.c50<-rep(NA, length(folds))
#pengulangan k sebanyak folds
for(k in 1:length(folds)){
#pembagian data train dan data test
train<-ldply(folds[-k], data.frame)
test<-ldply(folds[k], data.frame)
#algoritma c.50
model<-c5.0(as.formula(form), train) p<-
predict(model.newdata=test) conf.mat<-
table(test$y, p)
err.c50[k]<-1 - sum(diag(conf.mat))//sum(conf.mat)
#akurasi
accuracy<-sum(diag(conf.mat))//sum(conf.mat)

y_train_all[[k]]<-train
y_test_all[[k]]<-test
model_folds[[k]]<-model
p_all[[k]]<-p tmp_mean[i,k]<-
accuracy
conf_mat_folds[[k]]<-conf.mat
}
#definisikan model folds confusion matrix
model_iter[[i]]<-model_folds conf_iter[[i]]<-
conf_mat_folds
}
#mencari index akurasi tertinggi
max<-which(tmp_mean == max(tmp_mean), arr.ind = true)
x = as.numeric(max[1])
y = as.numeric(max[2])
#tampilkan confusion matrix
conf_iter[[x]][[y]]
max(tmp_mean)
#plot plot(model_iter[[a]]
[[b]])
```

7. Referensi

- [1] R. Probosiwi, “Pengguguran Dan Pengaruhnya Terhadap Tingkat Kemiskinan,” *Balai Besar Penelit. Dan Pengemb. Pelayanan Kesejaht. Sos.*, No. 1, Pp. 89–100, 2016.
- [2] A. K. Prasetyoningrum, “Analisis Pengaruh Indeks Pembangunan Manusia (Ipm), Pertumbuhan Ekonomi, Dan Pengangguran Terhadap Kemiskinan Di Indonesia,” *Equilib. J. Ekon. Syariah*, Vol. 6, No. 2, P. 217, 2018.
- [3] P. Karini, “Pengaruh Tingkat Kemiskinan Terhadap Angka Partisipasi Sekolah Usia 16 —18 Tahun Di Provinsi Kepulauan Bangka Belitung,” *Al-Ishlah J. Pendidik.*, Vol. 10, No. 1, Pp. 103–115, 2018.
- [4] H. S. Putra And N. Rianto, “Pengaruh Akses Air Bersih Terhadap Kemiskinan Di Indonesia : Pengujian Data Rumahtangga,” Pp. 65–76, 2016.
- [5] E. T. Adhi, “Pelayanan Sanitasi Buruk Akar Dari Kemiskinan,” *J. Anal. Sos.*, Vol. 14, Pp. 76–88, 2009.
- [6] A. Nurmalita And Suryandari, “Pengaruh Pertumbuhan Ekonomi, Pendidikan, Dan Kesehatan Terhadap Tingkat Kemiskinan Di Provinsi Daerah Istimewa Yogyakarta Tahun 2004 -2014,” 2017.
- [7] F. J. Kaunang, “Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan Di Indonesia,” *Cogito Smart J.*, Vol. 4, No. 2, P. 348, 2019.
- [8] M. Winny Amelia, A. S. . Lumenta, And A. Jacobus, “Prediksi Masa Studi Mahasiswa Dengan Menggunakan Algoritma Naïve Bayes,” *J. Tek. Inform.*, Vol. 11, No. 1, 2017.
- [9] I. M. Sudarma, “Implementasi Algoritma C5 . 0 Pada Penilaian,” *Maj. Ilm. Teknol. Elektro*, Vol. 17, No. 3, Pp. 1–6, 2018.