e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

KLASIFIKASI DAN PREDIKSI TINGKAT PENGANGGURAN TERBUKA DI INDONESIA MENGGUNAKAN METODE CLASSIFICATION AND REGRESSION TREE (CART)

Risma Yulistiani[1], Nanda Cahaya Putra[2], Qahtan Said[3], Iin Ernawati[*]
Program Studi S-1 Informatika, Fakultas Ilmu Komputer Universitas
Pembangunan Nasional Veteran Jakarta
Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
yulistianir@gmail.com

Abstrak. Salah satu permasalahan di Indonesia ialah tingginya tingkat pengangguran terbuka, dan dibutuhkan solusi untuk mengatasinya. Beberapa variabel atau faktor yang mempengaruhi tingginya tingkat pengangguran dianalisis dengan menggunakan sebuah metode untuk menghasilkan pola data sehingga diperoleh karakteristik atau ciri data berkaitan dengan tingginya tingkat pengangguran di Indonesia. CART merupakan metode mengolah data dalam bentuk algoritma dengan tujuan menghasilkan pola-pola data yang dibutuhkan. Hasil yag diperoleh berupa akurasi sebesar 91.17% berdasarkan fungsi *recall, precision, accuracy* dan *error ratio*, dengan nilai masing-masing yaitu 96.87%, 93.93% dan 8.83% maka pola yg ditemukan menunjukkan bahwa penyebab paling berpengaruh terhadap tingkat pengangguran secara berurutan adalah rata-rata lama sekolah, kemiskinan dan Angka Partisipasi Sekolah (APS).

Kata Kunci: Pengangguran Terbuka, CART, klasifikasi, prediksi

1 Pendahuluan

Pengangguran diartikan sebagai individu yang menjadi bagian dari angkatan kerja dan sedang mencari pekerjaan dengan besaran upah tertentu, namun tidak memiliki pekerjaan yang sesuai dengan keinginannya [1]. Pengangguran menjadi salah satu permasalahan utama di Indonesia yang harus diatasi oleh pemerintah. Dalam upaya mengurangi tingginya angka pengangguran di Indonesia, pemerintah harus melakukan penanggulangan secara menyeluruh dan menyangkut seluruh faktor penyebab pengangguran di Indonesia. Pada Agustus 2019 sebanyak 126,51 juta orang adalah penduduk yang memiliki pekerjaan dan sebanyak 7,05 juta orang adalah pengangguran. Jumlah penduduk bekerja bertambah sebanyak 2,50 juta orang dan pengangguran meningkat 50 ribu orang dari tahun 2018[2].

Penelitian terhadap pengangguran dan *Classification and Regression Tree* (CART) sudah pernah dilakukan sebelumnya, diantaranya penelitian[3] melakukan perbandingan metode *backpropagation* dengan metode *Genetic-Based Backpropagation* untuk memprediksi tingkat pengangguran terbuka di Indonesia. Hasil dari penelitian tersebut menunjukkan bahwa metode *backpropagation* memiliki rata-rata nilai *Average Forecast Error Rate* (AFER) sebesar 4.715198444% dan metode *Genetic-Based Backpropagation* memiliki AFER sebesar 3.877514478%. Penelitian[4] membangun sistem untuk memprediksi tingkat pengangguran di provinsi Maluku dengan metode *Adaptive Neuro Fuzzy Inference System* (ANFIS). Penelitian tersebut menggunakan data yang digunakan yaitu angka angkatan kerja dari tahun 2001—2017 meliputi: jumlah angkatan kerja,jumlah penduduk usia kerja, dan jumlah pekerja. Hasil yang diperoleh dari penelitian tersebut memiliki rata-rata *error* sebesar 4,49%. Penelitian[5] melakukan penelitian untuk mengklasifikasi pengangguran di provinsi Sulawesi Utara menggunakan metode *Classification and Regression Tree (CART).* Penelitian tersebut menggunakan data sekunder dan terdiri atas beberapa variabel, yaitu status pengangguran, *gender*, tingkatan pendidikan, usia, pengalaman pelatihan kerja, status dalam rumah tangga, klasifikasi tempat tinggal dan status perkawinan. Hasil dari penelitian tersebut memiliki akurasi sebesar 78,9% dan nilai *specifity* 87,16% dan faktor yang mempengaruhi pengangguran terbuka di provinsi Sulawesi Utara yaitu jenis kelamin, pendidikan, usia, status dalam rumah tangga dan status perkawinan.

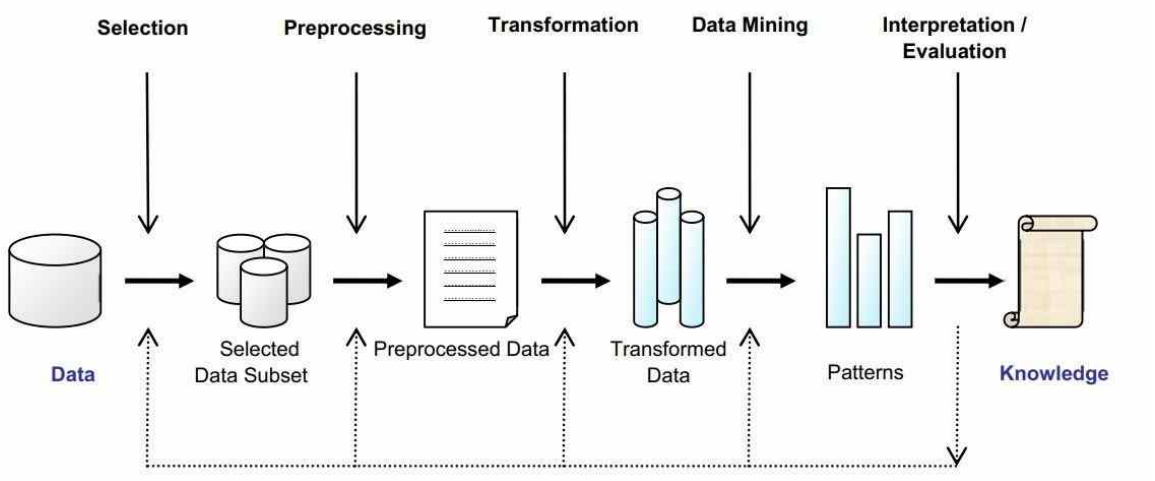
e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Dari beberapa penelitian diatas, penlitian tingkat pengangguran hanya mencakup tingkat provinsi dan belum ada yang mencakup ruang lingkup negara Indonesia. Tujuan dari penelitian ini, untuk memberikan informasi terkait faktor-faktor yang berpengaruh terhadap tingkat pengangguran di Indonesia yang dapat digunakan oleh pemerintah dalam upaya mengatasi pengangguran di Indonesia. Metode yang diterapkan dalam penelitian ini adalah *decision tree* menggunakan algoritma CART. Klasifikasi *decision tree* adalah salah satu metode yang terkenal dalam membuat keputusan dari suatu kasus yang ada dan menjadi salah satu teknik yang terkenal dalam *data mining.* Kasus-kasus yang memiliki dimensi besar dapat diselesaikan menggunakan *decision tree* tanpa perlu dilakukan proses pengelolaan pengetahuan sebelumnya [6].

Penelitian ini terdiri atas beberapa tahapan untuk melakukan pemodelan dan pengklasifikasian pada data*set* yang sudah dikumpulkan.Tahapan yang dilakukan diantaranya melakukan praproses data dengan mengisi *missing value* pada data*set,* integrasi data kemudian melakukan normalisasi data. Tahapan selanjutnya melakukan pemodelan data menggunakan metode *decision tree.*

2 Metodologi Penelitian

Penelitian ini terdiri atas beberapa tahapan yang dilakukan berdasarkan *Knowledge Discovery in* Data*bases* (KDD). *Knowledge Discovery in Databases* merupakan suatu proses untuk mendapatkan informasi yang pent ing serta menentukan pola-po la yang terdapat dalam data. Infor masi ini terdapat dalam *database* berukuran besar yang sebelumnya tidak diketahui [7]. Adapun tahapan-tahapan tersebut sebagai berikut.



Gambar 1. *Knowledge Discovery in Databases (sumber: Kavakiotis, 2017)*

2.1 Identifikasi Masalah

Pada penelitian ini, permasalahannya adalah faktor-faktor apa saja yang berpengaruh terhadap tingkat pengangguran di seluruh Provinsi yang terdapat di Indonesia dan mencari faktor yang paling berpengaruh pada tingkat pengangguran di Indonesia.

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Klasifikasi dan Prediksi Tingkat Pengangguran Terbuka di Indonesia Menggunakan Metode Classification and Regression Tree (CART)

2.2 Pengumpulan Data

Penelitian ini menggunakan data sekunder yang didapatkan melalui situs Badan Pusat Statistik (www.bps.go.id). Data yang tersaji meliputi lima atribut faktor penyebab pengangguran, dan satu variabel terikat yang terdiri dari 34 provinsi dari tahun 2010—2018. Variabel-variabel yang digunakan dalam penilitian ini yaitu:

Table 1. *Atribut yang digunakan dalam penelitian*

No	Atribut	Kode
1	Persentase Indeks Pembangunan Manusia	X1
2	Persentase Penduduk Miskin	X2
3	Besaran Upah Minimum Provinsi (dalam Rupiah)	X3
4	Persentase Angka Partisipasi Sekolah (rata-rata umur 6-24 tahun)	X4
5	Persentase Rata-rata Lama Sekolah	X5
6	Persentase Pengangguran Terbuka	Y

2.3 Praproses Data

Dalam praproses data dilakukan beberapa tahapan, yaitu mengisi *missing value,* integrasi data, normalisasi data dan memberi label pada data. Dalam *record* data terdapat beberapa *missing value*, terutama pada data provinsi Kalimantan Utara, hal tersebut disebabkan karena Provinsi Kalimantan Utara baru dimekarkan pada tahun 2012. Sedangkan data yang digunakan pada penelitian ini dari tahun 2010—2018. Untuk mengisi *missing value* ini dilakukan dengan cara menghitung nilai rata-rata (*mean)* pada tahun yang terdapat *missing value.* Rumus untuk menghitung nilai rata-rata ini menurut penelitian [8] adalah:

$$\bar{x} = \frac{\sum x}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Keterangan :
$\bar{x}$ = nilai rata-rata
x = nilai setiap data
n = jumlah data

Setelah mengisi *missing value,* tahapan selanjutnya melakukan integrasi data. Tahapan ini dilakukan karena data didapatkan dari beberapa data*set* yang berbeda, sehingga data tersebut perlu digabungkan menjadi satu data*set*. Tahapan selanjutnya adalah melakukan normalisasi data dengan mengubah *range* nilai setiap variabel. Normalisasi data diperlukan karena variabel UMP memiliki range nilai yang jauh berbeda dengan variabel lainnya, sehingga data dinormalisasi menggunakan metode *Min-Max*, range yang digunakan ialah 0-1. Normalisasi min-max melakukan transformasi linear pada data asli. Normalisasi Min-max memetakan nilai d dari P ke d ′ dalam rentang [new_min (p), new_max (p)] [9]. Adapun rumus *Min-Max*, sebagai berikut:

$$x' = \frac{[x - min(x)] * [new\_max(x) - new\_min(x)]}{[max(x) - min(x)]} + new\_min(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Proses selanjutnya ialah proses pemberian label atau kelas, pemberian label dilakukan agar data dapat di bagi berdasarkan kelas. Label kelas yang digunakan adalah membagi data menjadi dua kelas, yaitu “Rendah” dan “Tinggi”. Pembentukkan kelas ini didapatkan dari pembentukkan interval pada variabel *y*. Interval terbentuk data nilai maksimal dikurangi nilai minimal, lalu dibagi *n* sesuai jumlah kelas yang diinginkan, pada penelitian ini, ditentukan *n* adalah 2 karena kelas yang diinginkan sebanyak dua kelas [10]. Interval dapat dihitung menggunakan rumus sebagai berikut.

$$i = \frac{x_i x - min\ (R)}{} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Setelah didapatkan interval, maka label kelas dapat dirumuskan sebagai berikut:

- IF nilai_variabel y ≤ 0.5, maka label yang diberikan adalah Rendah atau tingkat pengangguran terbuka rendah
- IF nilai_variabel y > 0.5, maka label yang diberikan adalah Tinggi atau tingkat pengangguran terbuka tinggi.

2.4 Pembagian Data

Setelah melakukan tahapan praproses, data yang dihasilkan selanjutnya dibagi menjadi data *training* dan data *testing*. Dalam penelitian ini, data *training* menggunakan *record* yang diambil dari tahun 2010—2017, sebanyak 272 data. Sedangkan data *testing* diambil dari record tahun 2018 atau sebanyak 34 data.

2.5 Pemodelan

2.5.1 Clasification and Regression Trees (CART)

Pemodelan dilakukan dengan metode CART (Classification and Regression Trees), menggunakan data *training*. Model yang didapatkan kemudian diuji lagi menggunakan data *testing*, untuk melihat seberapa baik model yang dibuat. Pada penelitian ini kami menggunakan CART karena menurut penelitian[11] algoritma CART dapat menghasilkan aturan yang mudah dimengerti dan dijelaskan. Algoritma ini juga memiliki mekanisme bawaan untuk melakukan pemilihan atribut. Algoritma *binery recursive portioning* digunakan untuk membentuk pohon keputusan. Pemilihan digunakan untuk membagi data menjadi dua kelompok, yaitu kelompok simpul kiri dan simpul kanan. Pemilihan dilakukan pada masing-masing simpul hingga mendapatkan suatu simpul terminal/ akhir[5]. Data *training* yang bersifat *heterogen* digunakan untuk membentuk pohon keputusan.

2.5.2 Pruning

Selanjutnya dilakukan *pruning* untuk memotong bagian dari pohon keputusan, *pruning* merupakan bagian dari proses *decision tree* untuk mengurangi ukurannya karena terdapat *node* yang merupakan *outlier* maupun hasil dari *noise* data. Penerapan *pruning* dapat meningkatkan akurasi dari klasifikasi data[12].

$$\alpha = \frac{r(t) - r(T_t)}{|n_{T_t}| - 1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Keterangan:

- r(t) = rata-rata *error* dari *node* t
- $r(T_t)$ = rata-rata *error* dari *subtree* T_t
- n_{Tt} = jumlah *leaf node* pada pohon T_t

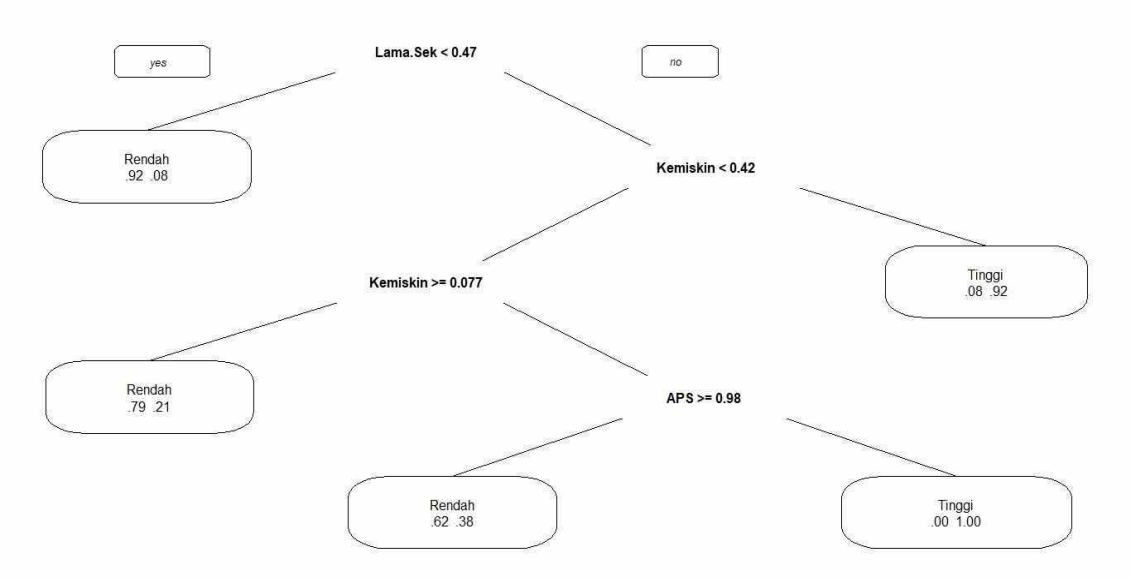
3 Hasil dan Pembahasan

3.1 Pengujian Model

Setelah dilakukan pemodelan data menggunakan data *training*, akan didapatkan pohon keputusan yang nantinya akan digunakan untuk menentukan *rules* dalam mengklasifikasikan data. *Rules* yang diperoleh digunakan sebagai aturan untuk memprediksi kelas dari data *testing* yang akan diuji. Berikut hasil pohon keputusan yang didapatkan dari proses pemodelan:

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Klasifikasi dan Prediksi Tingkat Pengangguran Terbuka di Indonesia Menggunakan Metode Classification and Regression Tree (CART)



Gambar 2. *Hasil Pemodelan* Data *Training*

Dari gambar pemodelan yang ada pada **Gambar 2**, diperoleh *rules* sebagai berikut:

- ◆◆1 = *I◆* (◆◆5 < 0.47) = ◆◆◆◆ 1 *R◆◆◆◆◆h* (◆◆◆◆◆◆◆◆*a◆◆*
  ◆◆◆◆◆◆*a◆i◆i◆a◆i* ◆◆◆◆◆◆◆*h* 92%)
- ◆◆2 = *I◆* ((◆◆5 ≥ 0.47) ∧ (0.077 ≤ ◆◆2 < 0.42)) =
  ◆◆◆◆ 2 *R◆◆◆◆◆h* (◆◆◆◆◆◆◆◆*a◆◆*
  ◆◆◆◆◆◆*a◆◆◆◆i◆a◆i* ◆◆◆◆◆◆◆*h* 79%)
- ◆◆3 = *I◆* ((◆◆5 ≥ 0.47) ∧ (0.077 > ◆◆2 < 0.42) ∧ (◆◆4 ≥ 0.98) =
  ◆◆◆◆ 3 *R◆◆◆◆◆h* (◆◆◆◆◆◆◆◆*a◆◆*
  ◆◆◆◆◆◆*a◆◆◆◆i◆a◆i* ◆◆◆◆◆◆◆*h* 62%)
- ◆◆4 = *I◆* ((◆◆5 ≥ 0.47) ∧ (0.077 > ◆◆2 < 0.42) ∧ (◆◆4 < 0.98)) =
  ◆◆◆◆ 4 *◆◆i◆◆◆i* (◆◆◆◆◆◆◆◆*a◆◆*
  ◆◆◆◆◆◆*a◆i◆i◆◆◆◆i ◆i◆◆◆i* 100%)
- ◆◆5 = *I◆* ((◆◆5 ≥ 0.47) ∧ (◆◆2 ≥ 0.42)) =
  ◆◆◆◆ 5 *◆◆i◆◆◆i* (◆◆◆◆◆◆◆◆*a◆◆*
  ◆◆◆◆◆◆*a◆i◆i◆◆◆◆i ◆i◆◆◆i* 92%)

3.2 Hasil Prediksi Pengujian Data *Testing*

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Berdasarkan *rules* yang diperoleh, kemudian model diuji menggunakan data *testing*, hal ini dilakukan untuk melihat seberapa baik *rules* dan model klasifikasi dalam memprediksi data yang belum memiliki kelas atau label. Hasil yang diperoleh dari prediksi data, kemudian dibandingkan dengan nilai aktual atau nilai pada data yang sebenarnya. Hasil prediksi dapat dilihat pada **Tabel 2**, dimana nilai atau data sebenarnya dinotasikan dengan *y*, sedangkan nilai prediksi dinotasikan dengan *y'*. Berikut dibawah ini adalah hasil prediksi terhadap tahun 2018 pada 34 provinsi.

Table 2. *Hasil Prediksi tahun 2018*

Provinsi	y	y'
Aceh	Rendah	Rendah
Sumatera Utara	Rendah	Rendah
Sumatera Barat	Rendah	Rendah
Riau	Rendah	Rendah
Jambi	Rendah	Rendah
Sumatera Selatan	Rendah	Rendah
Bengkulu	Rendah	Rendah
Lampung	Rendah	Rendah
Kep. Bangka Belitung	Rendah	Rendah
Kepulauan Riau	Rendah	Rendah
Dki Jakarta	Rendah	Rendah
Jawa Barat	Tinggi	Rendah
Jawa Tengah	Rendah	Rendah
D I Yogyakarta	Rendah	Rendah
Jawa Timur	Rendah	Rendah
Banten	Tinggi	Rendah
Bali	Rendah	Rendah
Nusa Tenggara Barat	Rendah	Rendah
Nusa Tenggara Timur	Rendah	Rendah
Kalimantan Barat	Rendah	Rendah
Kalimantan Tengah	Rendah	Rendah
Kalimantan Selatan	Rendah	Rendah
Kalimantan Timur	Rendah	Rendah
Kalimantan Utara	Rendah	Rendah
Sulawesi Utara	Rendah	Rendah
Sulawesi Tengah	Rendah	Rendah
Sulawesi Selatan	Rendah	Rendah
Sulawesi Tenggara	Rendah	Rendah
Gorontalo	Rendah	Rendah
Sulawesi Barat	Rendah	Rendah
Maluku	Rendah	Tinggi
Maluku Utara	Rendah	Rendah
Papua Barat	Rendah	Rendah
Papua	Rendah	Rendah

3.3 Evaluasi Model

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Evaluasi model yang digunakan adalah dengan menghitung nilai *Confusion Matrix* pada model. *Confusion Matrix* menunjukkan nilai prediksi dan nilai actual, menggunakan tabel *matrix* [11]. Hasil dari nilai tersebut dapat dilihat pada Table 3. seperti berikut.

Table 3. *Confusion Matrix*

		Prediksi	
		Rendah	Tinggi
Aktual	Rendah	31	1
	Tinggi	2	0

Pada *confusion matrix*, nilai akurasi dapat dihitung menggunakan rumus sebagai berikut.

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

Dimana:
TP = total nilai data positif yang diprediksi benar TN = total nilai data negatif yang diprediksi benar FN = total nilai data negatif yang diprediksi salah FP = total nilai data positif yang diprediksi salah

Sedangkan untuk menentukan nilai kesalahan atau *error rate*, dapat digunakan rumus sebagai berikut

$$\text{Error rate} = \frac{\text{Jumlah data salah}}{\text{Jumlah seluruh data}} \times 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Untuk menghitung kualitas kelengkapan hasil relevan yang ditampilkan pada model, dapat dihitung dengan rumus:

$$Recall = TP/(TP + FN) \quad \ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots (7)$$

Sedangkan untuk menghitung kualitas seberapa bagus system atau model yang dibuat [13], dihitung berdasarkan rumus berikut:

$$Precision = TP/(TP + FP)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

Hasil evaluasi dengan menggunakan *confusion matrix* menghasilkan nilai akurasi yang didapat adalah sebesar 91.17% dan *error ratio* sebesar 8.83%, serta mampu menampilkan recall atau hasil yang relevan dari model sebesar 96.87% dan dihasilkan kualitas model dari perhitungan *precision* sebesar 93.93%. Sehingga dapat disimpulkan bahwa model yang dibuat berhasil menentukan faktor penyebab pengangguran terbesar dan memprediksi kelas pada proses pengujian model.

4. Penutup

4.1 Kesimpulan

e-ISBN 978-623-93343-0-7
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 28 Januari 2020

Hasil pemodelan yang didapatkan dalam penelitian ini menunjukan faktor yang paling mempengaruhi tingkat pengangguran terbuka di Indonesia, secara berurutan adalah rata-rata lama sekolah, kemiskinan dan Angka Partisipasi Sekolah (APS), sedangkan Upah Minimum Provinsi (UMP) dan Indeks Pembangunan Manusia (IPM) kurang signifikan mempengaruhi tingkat pengangguran terbuka di Indonesia, sehingga tidak berpengaruh pada model. Dengan kata lain, dalam penelitian ini hanya tiga variabel bebas yang berpengaruh pada tingkat pengangguran terbuka di Indonesia. Model yang didapatkan dari klasifikasi menggunakan metode *decision tree* dengan algoritma *Classification and Prediction* (CART) berhasil memprediksi kelas dari tingkat pengangguran terbuka pada data *testing* sebesar 91.17% dan *error ratio* sebesar 8.83%.

4.2 Saran

Dalam penelitian yag sudah dilakukan ini masih terdapat *error ratio* sebesar 8.83%, hasil penelitian menunjukkan bahwa model yang didapatkan masih memiliki kekurangan yang cukup besar, sehingga diharapkan pada penelitian berikutnya mengenai tingkat pengangguran terbuka di Indonesia dapat menggunakan metode yang lain, untuk mendapat hasil yang lebih baik. Selain itu, dari hasil yang diperoleh dapat diambil kesimpulan bahwa faktor yang berpengaruh dari kelima variabel adalah rata-rata lama sekolah, kemiskinan dan angka partisipasi sekolah, diharapkan dengan adanya penelitian ini, pemerintah maupun masyarakat mampu lebih menyadari pentingnya Pendidikan dan memberikan penanganan lebih serius dalam menaikan tingkat rata-rata sekolah di Indonesia.

Referensi

[1] M. R. Muslim, “Pengangguran Terbuka dan Determinannya,” *Ekon. dan Stud. Pembang.*, vol. 15, pp. 171–181, 2014.
[2] B. P. Statistik, “Berita Resmi Statistik.” pp. 1–20, 2019.
[3] D. P. Adwandha, D. E. Ratnawati, and P. P. Adikara, “Prediksi Jumlah Pengangguran Terbuka di Indonesia menggunakan Metode Genetic-Based Backpropagation,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 4, pp. 341–351, 2017.
[4] D. L. Rahakbauw, M. I. Tanassy, and B. P. Tomasouw, “Sistem Prediksi Tingkat Pengangguran Di Provinsi Maluku Menggunakan Anfis (Adaptive Neuro Fuzzy Inference System),” *Barekeng J. Ilmu Mat. Dan Terap.*, vol. 12, no. 2, pp. 099–106, 2018.
[5] F. E. Pratiwi, F. E. Pratiwi, and I. Zain, “Der Ratgeber ... Landesverband Norden im Installateur - und Klempner-Gewerbe, Hamburg,” *J. Sains dan Seni ITS*, vol. 3, no. 1, pp. D54–D59, 2014.
[6] R. A. Putranto, T. Wuryandari, and Sudarno, “Perbandingan Analisis Klasifikasi Antara Decision Tree dan Support Vector Machine Muticlass Untuk Penentu Jurusan Pada Siswa SMA,” *Gausian*, vol. 4, no. Data Mining, pp. 1007–1016, 2015.
[7] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research,” *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017.
[8] N. C. Putra, I. Yunizar, and N. Aji, “Keterkaitan Antara Variabel dan Prediksi Nilai Indeks Pembangunan Manusia (IPM) di Indonesia Menggunakan Regresi Linier,” no. 2015, pp. 12–21, 2018.
[9] J. Y. Kumar and B. S. Kumar, “Min max normalization based data perturbation method for privacy protection,” *Int. J. Comput. Commun. Technol.*, vol. 2, no. 8, pp. 45–50, 2011.
[10] I. W. Santiyasa, “Statistika dasar,” 2007.
[11] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, “Confusion matrix-based feature selection,” *CEUR Workshop Proc.*, vol. 710, pp. 120–127, 2011.
[12] M. Budi, R. Karyadin, and S. Wijaya, “Perbandingan Algoritme Pruning pada Decision Tree yang Dikembangkan dengan Algoritme CART,” *J. Ilm. Ilmu Komput.*, vol. 8, no. 2, pp. 7–13, 2010.
[13] R. Trevethan, “Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice,” *Front. Public Heal.*, vol. 5, no. November, pp. 1–7, 2017.