

Metode *Classification And Regression Trees* untuk Pengklasifikasian Faktor-Faktor yang Mempengaruhi Pengangguran Terbuka di Provinsi Jawa Barat Tahun 2020

Seruni Purwanti Dj*, Anneke Iswani Achmad

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*serunipurwanti20@gmail.com, annekeiswani11@gmail.com

Abstract. CART (*Classification And Regression Trees*) is a classification method that uses historical data to build decision trees. The CART method is used to form a classification tree using the gini index value obtained from the probability value of each node candidate. The CART method used aims to look at the factors that influence open unemployment in West Java Province in 2020. In this study, the independent variables that affect open unemployment are gender (X1), age (X2), education (X3), and marital status (X4), while for the status variable in the household (X5) from the classification tree results indicate that this variable is not a differentiating variable. From the results of the classification tree research formed, the results obtained for respondents with low and high levels of education who are male, both married and unmarried are included in the working classification, while for respondents with low and high education levels, if a woman is not married, then she works, while if she is married, she is considered unemployed. And for respondents with secondary education level, both male and female, who are married or not included in the unemployment classification. In this study, the classification of open unemployment resulted in an accuracy of classification on the testing data of 69.79%.

Keywords: *Open Unemployment, Labor Force, Decision Tree, Gini Index, Probability, Classification And Regression Trees.*

Abstrak. CART (*Classification And Regression Trees*) adalah metode klasifikasi yang menggunakan data historis untuk membangun pohon keputusan. Metode CART digunakan untuk membentuk pohon klasifikasi dengan menggunakan nilai indeks gini yang didapat dari nilai probabilitas setiap calon simpul. Metode CART yang digunakan bertujuan untuk melihat faktor-faktor yang mempengaruhi pengangguran terbuka di Provinsi Jawa Barat tahun 2020. Dalam penelitian ini, variabel bebas yang berpengaruh terhadap pengangguran terbuka yaitu jenis kelamin (X₁), usia (X₂), pendidikan (X₃), dan status pernikahan (X₄), sementara untuk variabel status dalam rumah tangga (X₅) dari hasil pohon klasifikasi menunjukkan bahwa variabel tersebut bukanlah sebagai variabel pembeda. Dari hasil penelitian pohon klasifikasi yang terbentuk, maka didapatkan hasil untuk responden dengan tingkat pendidikan rendah serta tinggi yang berjenis kelamin laki-laki, baik yang sudah menikah maupun yang belum menikah termasuk ke dalam klasifikasi bekerja, sedangkan untuk responden dengan tingkat pendidikan rendah dan tinggi, yang berjenis kelamin perempuan jika belum menikah maka ia bekerja, sedangkan yang sudah menikah maka ia termasuk pengangguran. Dan untuk responden dengan tingkat pendidikan menengah dengan jenis kelamin laki-laki maupun perempuan, yang sudah menikah maupun belum termasuk ke dalam klasifikasi pengangguran. Dalam penelitian ini pula, pengklasifikasian pengangguran terbuka menghasilkan ketepatan klasifikasi pada data testing sebesar 69,79%.

Kata Kunci: *Pengangguran Terbuka, Angkatan Kerja, Pohon Keputusan, Indeks Gini, Probabilitas, Classification And Regression Trees.*

A. Pendahuluan

Pengklasifikasian adalah salah satu metode statistik yang digunakan untuk mengelompokkan atau mengklasifikasikan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, baik pada bidang sosial, akademik, kesehatan, maupun pada bidang lainnya. Masalah-masalah tersebut dapat diselesaikan dengan metode klasifikasi, namun pemilihan metode yang tepat perlu diperhatikan agar dapat menyelesaikan masalah tersebut dengan baik. (Amaliyyah, 2021).

Metode klasifikasi dengan pendekatan nonparametrik yang sering digunakan adalah *Decision Tree* (Pohon keputusan). Pohon keputusan adalah sebuah metode berstruktur pohon yang bertujuan untuk melihat hubungan antar variabel respon dengan variabel penjelasnya. Menurut Cho dan Ngai (2003), *decision tree* atau pohon keputusan adalah suatu metode yang digunakan untuk mengklasifikasi suatu objek yang ditampilkan melalui gambar daun dan cabang. Daun atau *leaf* pada *decision tree* diidentifikasi oleh suatu kelas, sementara cabangnya (*internal node*) digunakan untuk mewakili suatu kondisi dari atribut objek yang terukur. Atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu yaitu akar (*root*) yang merupakan level *node* teratas dari sebuah pohon keputusan. Pohon keputusan sendiri memiliki beberapa metode dimana metode yang populer salah satunya adalah metode CART.

Classification and Regression Trees (CART) adalah suatu metode algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan yang dikembangkan untuk melakukan analisis klasifikasi pada variabel respon baik yang nominal, ordinal maupun kontinu. CART dapat menyeleksi interaksi dari variabel-variabel yang paling penting dalam menentukan hasil atau variabel prediktor, dengan tujuan utamanya yaitu untuk mendapatkan suatu kelompok data yang akurat dari suatu pengklasifikasian. (Breiman *et al.*, 1993).

Dengan jumlah penduduk yang semakin bertambah setiap tahunnya, menyebabkan tingginya angka pengangguran sulit untuk dihindari bagi suatu negara, karena jumlah angkatan kerja yang ikut bertambah tetapi tidak diiringi dengan tersedianya lapangan kerja yang memadai. Pengangguran ialah suatu kondisi dimana seseorang tidak memiliki pekerjaan, yang sedang aktif mencari pekerjaan, atau bekerja kurang dari waktu kerja. (Rafiq *et al.*, 2008)

Pada penelitian ini akan digunakan metode *Classification and Regression Tree* (CART) pada kasus pengangguran terbuka di Jawa Barat tahun 2020 dengan faktor-faktor yang diduga yaitu jenis kelamin, usia, tingkat pendidikan, status pernikahan dan status dalam rumah tangga.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “bagaimana klasifikasi faktor-faktor yang mempengaruhi pengangguran terbuka di Provinsi Jawa Barat dengan menggunakan metode CART (Classification and Regression Tree)”. Selanjutnya, tujuan dalam penelitian ini berdasarkan identifikasi masalah diatas, dengan menggunakan metode *Classification And Regression Trees* (CART) maka tujuan yang ingin dicapai yaitu mengetahui klasifikasi faktor-faktor yang mempengaruhi pengangguran terbuka di Provinsi Jawa Barat tahun 2020.

B. Metodologi Penelitian

Jenis penelitian yang dilakukan dalam skripsi ini adalah penelitian terapan. Dimana akan dilakukan pengamplikan metode CART pada data pengangguran terbuka di Provinsi Jawa Barat yang berasal dari hasil survei angkatan kerja nasional (SAKERNAS) pada tahun 2020 yang bersumber langsung dari Badan Pusat Statistik yang berada di Provinsi Jawa Barat.

Data pada penelitian ini merupakan data sekunder, dimana variabel terikat (Y) yang digunakan adalah pengangguran terbuka di Provinsi Jawa Barat tahun 2020, serta variabel bebas yaitu jenis kelamin (X_1), usia (X_2), tingkat pendidikan (X_3), status pernikahan (X_4), dan status dalam rumah tangga (X_5).

Langkah Analisis Data

Penelitian ini menggunakan bantuan aplikasi R-Studio untuk menganalisis data.

Tahapan analisis yang dilakukan adalah sebagai berikut:

1. Membagi data yang akan digunakan menjadi dua bagian, yaitu data *training* dan data *testing*. Pada penelitian ini di tentukan data *training* sebesar 90% dan data *testing* sebesar 10%.

2. Membentuk pohon klasifikasi dengan metode CART, dimana tahapan yang akan dilakukan adalah sebagai berikut:
 - a. Melakukan pemilihan pemilah.
 - b. Hitung probabilitas dan indeks gini pada setiap calon simpul.
 - c. Melakukan pengurutan simpul menggunakan kriteria *Goodness of split*.
 - d. Membentuk pohon klasifikasi maksimal
 - e. Menghentikan pembentukan pohon apabila penurunan simpul hanya terdapat satu cabang untuk mendapatkan simpul terminal.
 - f. Memberikan penandaan label kelas pada simpul terminal berdasarkan data terbesar dari tiap kelas yang ada pada variabel terikat
 - g. Menentukan pohon klasifikasi maksimal.
3. Mencari nilai *complexity* parameter.
4. Melakukan pemangkasan pohon klasifikasi.
5. Menentukan pohon klasifikasi optimal menggunakan metode *Cross Validation v-fold*.
6. Menghitung ketepatan pohon klasifikasi menggunakan *sensitivity*, *specificity* serta akurasi.

C. Hasil Penelitian dan Pembahasan

Berikut adalah penelitian mengenai klasifikasi faktor-faktor yang mempengaruhi pengangguran terbuka menggunakan metode *Classification And Regression Trees*. Hasil pengujian yang didapatkan adalah sebagai berikut:

1. Karakteristik Pengangguran Terbuka Jawa Barat 2020
 Karakteristik Pengangguran Terbuka Jawa Barat 2020. Hasil survei angkatan kerja nasional (SAKERNAS) tahun 2020 memberikan informasi bahwa dari 14.469 terdapat 4.977 angkatan kerja adalah pengangguran terbuka dengan 1.545 dari 4.977 pengangguran terbuka berjenis kelamin laki-laki dan sebanyak 3.532 berjenis kelamin perempuan. Mayoritas pengangguran di Provinsi Jawa Barat memiliki pendidikan tinggi yaitu sebesar 3.697 atau 89 persen dan sebanyak 1.233 atau 25 persen dari pendidikan menengah serta 47 atau sebesar 1 persen dari pendidikan rendah. Dari 4.977 pengangguran di Provinsi Jawa Barat sebanyak 4.784 atau 96 persen tergolong dalam usia yang produktif. Hasil SAKERNAS Provinsi Jawa Barat tahun 2020 menyebutkan bahwa 64 persen dari 4.977 pengangguran terbuka berstatus sudah menikah, dan 36 persen belum menikah. Serta dari angkatan kerja yang termasuk pengangguran terbuka di Provinsi Jawa Barat tahun 2020, sebanyak 2.008 responden merupakan kepala keluarga dan sebanyak 2.969 merupakan bukan kepala keluarga.
2. Hasil Klasifikasi dengan Metode Cart
 Data learning digunakan untuk pembentukan pohon klasifikasi sedangkan data testing digunakan untuk validasi model.
 Kombinasi data learning dan testing yang digunakan yaitu data learning 95 persen dan testing 5 persen. Yaitu dari 14.459 angkatan kerja menggunakan data learning sebanyak 13.022 responden dan data testing sebanyak 1.477 responden.

Pemilihan pemilah

Tabel 1. Calon Simpul Kiri dan Calon Simpul Kanan

No	Calon Simpul Kanan	Calon Simpul Kiri
1	Jenis Kelamin = Laki-laki	Jenis Kelamin = Perempuan
2	Usia = Produktif	Usia = Tidak Produktif
3	Tingkat Pendidikan = Rendah	Tingkat Pendidikan = Lainnya
4	Tingkat Pendidikan = Menengah	Tingkat Pendidikan = Lainnya
5	Tingkat Pendidikan = Tinggi	Tingkat Pendidikan = Lainnya
6	Status Pernikahan = Kawin	Status Pernikahan = Belum Kawin

7	Status Dalam Rumah Tangga = Kepala Keluarga	Status Dalam Rumah Tangga = Bukan Kepala Keluarga
---	---	---

Dari Tabel 1 di atas, dapat dilihat pada kolom calon simpul kanan terdapat tingkat jenis kelamin laki-laki, usia produktif, tingkat pendidikan rendah, tingkat pendidikan menengah dan tinggi, tingkat status pernikahan kawin. Dan pada simpul kanan juga terdapat status dalam rumah tangga yakni kepala keluarga. Sedangkan pada calon simpul kiri telah terlampir pada Tabel 1.

Probabilitas setiap calon simpul

Tabel 2. Probabilitas Setiap Calon simpul

Calon Simpul	Variabel	p_R	p_L	Kelas	$p(j t_R)$	$p(j t_L)$
1	Jenis Kelamin	0,49762	0,50238	Bekerja	0,775617	0,461679
				Pengangguran terbuka	0,224382	0,538520
2	Usia	0,96744	0,03256	Bekerja	0,620257	0,455188
				Pengangguran terbuka	0,379742	0,544811
3	Pendidikan Rendah	0,00822	0,99178	Bekerja	0,560747	0,381726
				Pengangguran terbuka	0,439252	0,618273
4	Pendidikan Menengah	0,10429	0,89571	Bekerja	0,679012	0,320988
				Pengangguran terbuka	0,907953	0,092047
5	Pendidikan Tinggi	0,8875	0,11250	Bekerja	0,680107	0,12628
				Pengangguran terbuka	0,319893	0,873720
6	Status Pernikahan	0,77062	0,22938	Bekerja	0,680419	0,407432
				Pengangguran terbuka	0,319581	0,592568
7	Status dalam Rumah Tangga	0,40539	0,59461	Bekerja	0,619625	0,383443
				Pengangguran terbuka	0,380375	0,616557

Sumber: Data Penelitian yang Sudah Diolah, 2015.

Dalam mencari indeks gini dan *Goodness of split*, diperlukan nilai probabilitas yang terdapat pada setiap calon simpul. Maka dari tabel 2 di atas, didapatkan nilai probabilitas untuk setiap calon simpul.

Indeks Gini

Tabel 3. Indeks Gini

No	Calon Simpul Kanan	Calon Simpul Kiri	Indeks Gini
1	Jenis Kelamin = Laki-laki	Jenis Kelamin = Perempuan	0,84510
2	Usia = Produktif	Usia = Tidak Produktif	0,96706
3	Tingkat Pendidikan = Rendah	Tingkat Pendidikan = Lainnya	0,96464
4	Tingkat Pendidikan = Menengah	Tingkat Pendidikan = Lainnya	0,60305
5	Tingkat Pendidikan = Tinggi	Tingkat Pendidikan = Lainnya	0,65579

6	Status Pernikahan = Kawin	Status Pernikahan = Belum Kawin	0,91776
7	Status Dalam Rumah Tangga = Kepala Keluarga	Status Dalam Rumah Tangga = Bukan Kepala Keluarga	0,94421

Berdasarkan Tabel 3 diatas, diketahui bahwa nilai indeks gini terkecil adalah calon simpul pertama dengan calon simpul kanan yaitu tingkat pendidikan menengah dan calon simpul kiri yaitu tingkat pendidikan lainnya (rendah dan tinggi) sebesar 0,32597.

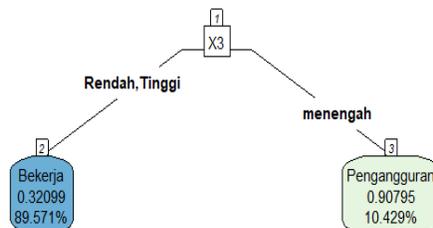
Goodness of split

Tabel 4 *Goodness of split*

Calon Simpul	$\phi(s, t)$
1	0,04934
2	0,00036
3	0,000054
4	0,06436
5	0,06125
6	0,02635
7	0,0000045

Berdasarkan hasil perhitungan *Goodness of split* dari tabel 4 diatas, yang telah dilakukan, diperoleh bahwa calon simpul tertinggi terdapat pada calon simpul pertama sebesar 0,06436. Sehingga calon simpul pertama tersebut menjadi simpul pertama.

Penentuan simpul pertama



Gambar 1 Pemecahan simpul pertama

Dapat dilihat berdasarkan Gambar 1 diperoleh hasil simpul pertama yaitu variabel X3, dimana dalam gambar tersebut untuk simpul kedua merupakan responden yang memiliki tingkat pendidikan rendah dan tinggi, sedangkan simpul ketiga merupakan responden yang memiliki tingkat pendidikan menengah. Pada Gambar 4.7 diketahui probabilitas responden berpendidikan rendah dan tinggi memiliki jumlah pengangguran sebesar 0,32099 sedangkan probabilitas responden berpendidikan menengah memiliki jumlah pengangguran sebesar 0,90795. Pada Gambar 1 juga diketahui, proporsi responden dengan tingkat pendidikan rendah dan tinggi yaitu sebesar 89,571% sedangkan proporsi responden dengan tingkat pendidikan menengah sebesar 10,429%. Proses tersebut akan terus berlanjut sampai mendapatkan simpul akhir.

Penandaan label kelas

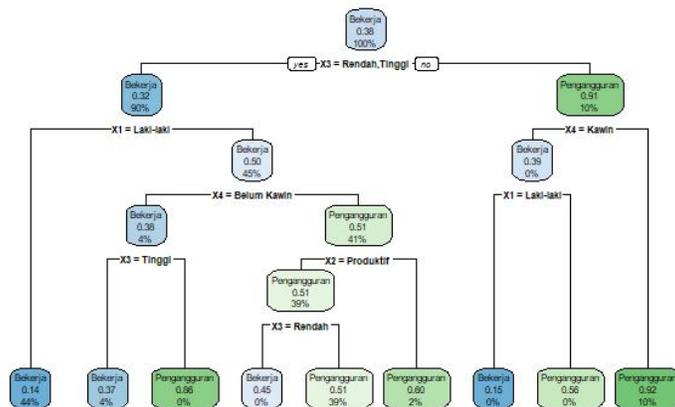
Penandaan label kelas ini ditentukan berdasarkan pada aturan jumlah terbanyak pada simpul terminal, yang dimaksudkan untuk mengetahui bawa simpul tersebut, apakah diberi label kelas “Pengangguran terbuka” atau kelas “Bekerja”. Dimana hasilnya dimuat sebagai berikut:

Tabel 5 Penandaan Label Kelas Pada Setiap Simpul

No	Simpul	p_R	p_L	Kelas	$p(j t_R)$	$p(j t_L)$
1	Jenis Kelamin	0,49762	0,50238	Bekerja	0,775617	0,461679
				Pengangguran terbuka	0,224382	0,538520
2	Usia	0,96744	0,03256	Bekerja	0,620257	0,455188
				Pengangguran terbuka	0,379742	0,544811
3	Pendidikan	0,00822	0,99178	Bekerja	0,560747	0,381726
				Pengangguran terbuka	0,439252	0,618273
4	Pendidikan	0,10429	0,89571	Bekerja	0,679012	0,320988
				Pengangguran terbuka	0,907953	0,092047
5	Pendidikan	0,8875	0,11250	Bekerja	0,680107	0,12628
				Pengangguran terbuka	0,319893	0,873720
6	Status Pernikahan	0,77062	0,22938	Bekerja	0,680419	0,407432
				Pengangguran terbuka	0,319581	0,592568
7	Status dalam Rumah Tangga	0,40539	0,59461	Bekerja	0,619625	0,383443
				Pengangguran terbuka	0,380375	0,616557

Setelah melakukan perhitungan untuk simpul terminal pertama, dapat dilihat variabel tingkat pendidikan (X3) untuk simpul kanan masuk pada kelas pengangguran terbuka karena peluangnya lebih besar dibandingkan peluang bekerja. Sebaliknya, pada simpul kiri termasuk klasifikasi bekerja karena peluangnya yang lebih besar. Maka dari itu, tingkat pendidikan menengah memiliki label kelas pengangguran terbuka, sedangkan tingkat pendidikan rendah dan tinggi memiliki label bekerja.

Pohon Klasifikasi Maksimal



Gambar 2 Pohon klasifikasi Maksimal

Berdasarkan Gambar 2 di atas diperoleh simpul 1 adalah tingkat pendidikan. Simpul kedua terdiri dari tingkat pendidikan rendah, tinggi. Simpul ketiga adalah tingkat pendidikan

menengah. Setelah simpul ketiga melakukan proses pemecahan, maka simpul ketiga akan bercabang menjadi simpul keempat yaitu status tidak bekerja. Begitu pun seterusnya.

Pemangkasan Pohon Klasifikasi

Proses pemangkasan ini dilakukan dengan melihat nilai *complexity* parameter minimum. Nilai *complexity parameter* di dapatkan berdasarkan hasil output dari *R-Studio*, yang dicari dari nilai *Cross Validation V-fold*.

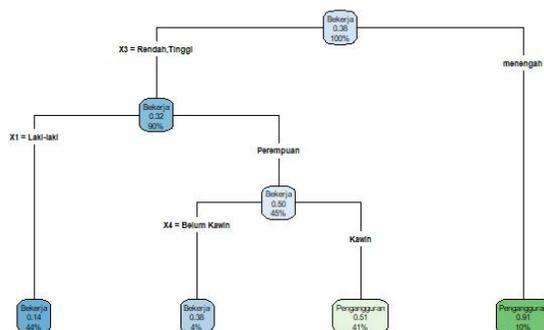
Tabel 6 Complexity Parameter

No.	Simpul	Complexity Parameter
1	Pendidikan Rendah	0,005625879
2	Usia	0,010247137
3	Pendidikan Tinggi	0,220012055

Berdasarkan pada Tabel 6 di atas sudah didapatkan nilai *complexity* parameter. Sehingga nilai *complexity* parameter yang terkecil atau nilai *complexity* parameter minimum sebesar 0,005625879.

Penentuan Pohon Klasifikasi Optimal

Pada Gambar 2 sama sekali tidak memperhatikan nilai *complexity* parameter atau nilai *complexity* parameter = 0. Setelah dicari nilai *Cross Validation v-fold*, di dapatkan bahwa nilai *complexity* parameter yang terkecil atau nilai *complexity* parameter *minimum* sebesar 0,005625879. Sehingga diperoleh hasil sebagai berikut:



Gambar 3 Pohon Klasifikasi Optimal

Pada Gambar 3 tersebut, simpul 1 adalah tingkat. Simpul kedua terdiri dari tingkat pendidikan rendah, tinggi. Simpul ketiga adalah tingkat pendidikan menengah. Setelah simpul ketiga melakukan proses pemecahan, maka simpul ketiga akan bercabang menjadi simpul keempat dan kelima yang masing- masing yaitu jenis kelamin laki-laki, dan jenis kelamin perempuan. Begitu pun seterusnya.

Ketepatan Klasifikasi

Tabel 7 Ketepatan Klasifikasi

Observasi	Prediksi		Total
	Pengguguran	Bukan Pengguguran	
Pengguguran	565	301	866
Bukan Pengguguran	136	445	581

Total	701	746	1.447
-------	-----	-----	-------

$$sensitivity = \frac{n_{00}}{N_0} = \frac{565}{866} = 0,6524$$

$$specificity = \frac{n_{11}}{N_1} = \frac{445}{581} = 0,7659$$

$$Akurasi = \frac{n_{00}+n_{11}}{N} = \frac{565 + 445}{1.447} = 0,6979$$

Setelah dilakukan perhitungan, didapatkan nilai sensitivity yang digunakan untuk mengukur ketepatan klasifikasi pada jumlah pengangguran sebesar 0,6524 atau 65,24% dan nilai specificity yang digunakan untuk mengukur ketepatan klasifikasi pada jumlah bukan pengangguran sebesar 0,7659 atau 76,59%. Sedangkan nilai akurasi yang digunakan untuk mengukur tingkat ketepatan klasifikasi untuk data testing sebesar 0,6979 atau 69,79%, yang artinya bahwa pohon optimal yang terbentuk mampu mengklasifikasikan data baru sebesar 69,79%.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

Metode CART dapat diterapkan untuk mengklasifikasikan faktor yang mempengaruhi pengangguran terbuka dari angkatan kerja kelompok umur 15-64 tahun di Indonesia. Hasil pengklasifikasian tersebut selanjutnya dapat digunakan untuk memberikan gambaran mengenai faktor apa saja yang berpengaruh terhadap klasifikasi pengangguran terbuka, sehingga meminimalisir terjadinya lonjakan pengangguran terbuka untuk masa yang akan datang. Dalam penelitian ini, variabel bebas yang berpengaruh terhadap pengangguran terbuka yaitu jenis kelamin (X1), pendidikan (X3), dan status pernikahan (X5), sementara untuk variabel usia (X2) dan status dalam rumah tangga (X5) dari hasil pohon klasifikasi menunjukkan bahwa variabel tersebut bukanlah sebagai variabel pembeda. Dari hasil penelitian pohon klasifikasi yang terbentuk, maka didapatkan hasil untuk responden dengan tingkat pendidikan rendah serta tinggi yang berjenis kelamin laki-laki baik yang sudah menikah maupun yang belum menikah termasuk ke dalam klasifikasi bekerja, sedangkan untuk responden dengan tingkat pendidikan rendah dan tinggi, yang berjenis kelamin perempuan, jika belum menikah maka ia bekerja, sedangkan yang sudah menikah maka ia termasuk pengangguran. Dan untuk responden dengan tingkat pendidikan menengah dengan jenis kelamin laki-laki maupun perempuan yang sudah menikah maupun belum termasuk ke dalam klasifikasi pengangguran. Dalam penelitian ini pula, pengklasifikasian jumlah pengangguran terbuka menghasilkan ketepatan klasifikasi pada data testing sebesar 69,79%.

Acknowledge

Penelitian ini dapat terlaksana dengan baik atas bantuan dari berbagai pihak, maka dari itu penulis mengucapkan terima kasih kepada Allah SWT, kedua orang tua yang selalu memberikan doa dan dukungan, Ibu Anneke Iswani Achmad, Dra., M.Si. yang telah memberikan arahan dan masukan, para dosen Statistika Unisba yang telah memberikan ilmu pengetahuannya, Dennis Ikhsan sebagai orang terkasih yang telah menemani penyusun, mendengarkan keluh kesah, menyemangati serta membantu dalam penyusunan penelitian ini, dan teman-teman yang selalu membantu serta menemani penulis hingga dapat menyelesaikan perkuliahan tepat pada waktunya.

Daftar Pustaka

- [1] Amaliyyah, R. (2021). Penggunaan Metode Classification and Regression Tree (CART) Untuk Klasifikasi Angkatan Kerja di kota Medan Tahun 2019 No 6.
- [2] Badan Pusat Statistik Indonesia. (2020). Catalog : 1101001. Statistik Indonesia 2020, 1101001,790.
<https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/> statistik-

indonesia-2020.html

- [3] Breiman L., Friedman J.H., Olshen R.A., & Stone C.J. (1984). *Classification And Regression Trees*. Wadsworth and Brooks, Monterey.
- [4] Cho, V. & Ngai, E. 2003. Data Mining for Selection of Insurance Sales Agents. *Expert Systems*, 20, 123-132, doi: 10.1111/1468-0394.00235.
- [5] Rafiq, M., Ahmad, I., Ullah, A., & Khan, Z. (2008). Determinants of Unemployment: A Case Study of Pakistan Economy. *Abasyn Journal of Social Sciences*, 3(1), 17–24.
- [6] Irawadi, Jody Alwin, Sunendiari, Siti. (2021). *Penerapan dan Perbandingan Tiga Metode Analisis Pohon Keputusan pada Klasifikasi Penderita Kanker Payudara*, *Jurnal Riset Statistika*, 1(1), 19-27.