

Classification of Unemployment at West Sumatra Province in 2021 using Algorithm Classification and Regression Tree

Nur Fadillah, Syafriandi Syafriandi*, Nonong Amalita, Dony Permana

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

*Corresponding author: syafriandi.math@fmipa.unp.ac.id

Submitted : 08 Mei 2024

Revised : 31 Mei 2024

Accepted : 31 Mei 2024

ABSTRACT

Unemployment is a problem that often occurs in developing countries. This is caused by the imbalance between the number of labor force and the number of working population. According to the Central Bureau of Statistics, West Sumatra Province in 2021 is the eighth province with a high open unemployment rate of 6,52%, which is higher than the average Indonesian open unemployment rate of 6,49%. The increase in unemployment has occurred from 2017 to 2021 which is caused by educated unemployment. This is due to the habit of job seekers who tend to pick and choose the types of jobs available, while business needs are very limited. The problem of unemployment will get higher if it is not resolved. As a result, unemployment can lead to poverty and other social problems. In this study, CART analysis is used to classify unemployment in West Sumatra Province in 2021 which aims to determine the factors that affect unemployment. CART is a decision tree that shows the relationship between the response variable and one or more predictor variables. The purpose of CART analysis is to obtain the right data group for classification purposes. Based on the analysis obtained, the variables that affect unemployment in West Sumatra Province in 2021 are marital status, gender, household status, education level, age, and place of residence with an accuracy value of 71,73%.

Keywords: *Unemployment, Classification, CART analysis*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Memajukan kesejahteraan umum dan mencerdaskan kehidupan bangsa adalah tujuan dari pembangunan nasional. Sebagai negara yang berkembang, Indonesia berupaya meningkatkan kesejahteraan demi tercapainya tujuan tersebut. Pengangguran adalah salah satu tantangan yang dihadapi Indonesia. Kesejahteraan masyarakat dapat terpengaruh oleh tingkat pengangguran yang tinggi. Menurut Swaramarinda (2014) meningkatnya pengangguran disebabkan oleh pertumbuhan kesempatan kerja yang lebih lambat daripada jumlah pekerja yang terus berkembang.

Menurut Badan Pusat Statistik, Provinsi Sumatera Barat pada tahun 2021 merupakan provinsi kedelapan yang memiliki tingkat pengangguran terbuka tinggi yaitu sebesar 6,52% yang lebih tinggi dibandingkan dengan rata-rata tingkat pengangguran terbuka Indonesia yaitu 6,49%. Peningkatan pengangguran sudah terjadi sejak tahun 2017 hingga 2021 yang disebabkan oleh pengangguran yang berpendidikan. Meningkatnya tingkat pengangguran terbuka dari kelompok angkatan kerja terdidik ini menjadi salah satu isu besar saat ini, tidak hanya di Sumatera Barat namun juga nasional. Hal ini disebabkan oleh kecenderungan pencari kerja untuk memilih-milih jenis pekerjaan yang tersedia saat jumlah lapangan kerja yang terbatas. Untuk itu pemerintah harus cepat tanggap untuk mengeluarkan berbagai kebijakan dan program yang dapat meminimalisir terjadinya pengangguran, agar tidak meningkatnya angka kemiskinan dan masalah sosial lainnya (Ishak, 2018).

Penelitian mengenai pengangguran telah dilakukan oleh peneliti sebelumnya. Menurut Pratiwi dan Zain (2014) tentang klasifikasi pengangguran terbuka menggunakan CART di Provinsi Sumatera Utara, dengan variabel pencari klasifikasi adalah jenis kelamin, pendidikan terakhir, usia, status dalam rumah tangga, status perkawinan, pengalaman pelatihan kerja, status perkawinan, dan klasifikasi tempat tinggal. Pada penelitian ini, indeks gini digunakan sebagai pemilah. Variabel status dalam rumah tangga sangat dominan dalam pengelompokkan dan merupakan pemilah dalam pembentukan pohon optimal. Variabel yang berpengaruh terhadap pengangguran di Provinsi Sumatera Utara adalah status dalam rumah tangga, jenis kelamin, usia, pendidikan terakhir dan status perkawinan.

Salah satu metode klasifikasi yang dapat mengklasifikasikan data pengangguran dalam bentuk visual adalah pohon klasifikasi (Helena dkk., 2019). Dengan menggunakan aturan keputusan sederhana, pohon klasifikasi membagi kumpulan data yang besar menjadi kumpulan data yang lebih kecil secara berurutan (Gordon, 2004:166). Pada pohon klasifikasi terdapat berbagai macam algoritma yang digunakan seperti algoritma ID3, algoritma C4.5, algoritma CART, algoritma CHAID, dan lainnya.

Pada penelitian ini digunakan algoritma CART (*Classification and Regression Tree*) untuk melakukan klasifikasi dalam menentukan penyebab terjadinya pengangguran. Algoritma CART pertama kali di gagaskan oleh Leo Braiman, Jerome Freidman, Richard Olshen, dan Charles Stone pada tahun 1984. Analisis CART menghasilkan kumpulan data dengan karakteristik klasifikasi yang tepat (HG, 2017). Menurut Steinberg dan Philip (1995:24-25), CART memiliki kelebihan yaitu tidak terikat oleh asumsi-asumsi dan efektif digunakan pada data yang memiliki dimensi besar atau memiliki banyak variabel prediktor serta kemudahan dalam menginterpretasikan hasil. Berdasarkan penjelasan tersebut, maka dilakukan penelitian yang berjudul “Klasifikasi Pengangguran di Provinsi Sumatera Barat Tahun 2021 Menggunakan Algoritma CART”. Penelitian ini bertujuan untuk mengidentifikasi semua variabel yang dapat mempengaruhi pengangguran di Provinsi Sumatera Barat tahun 2021.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

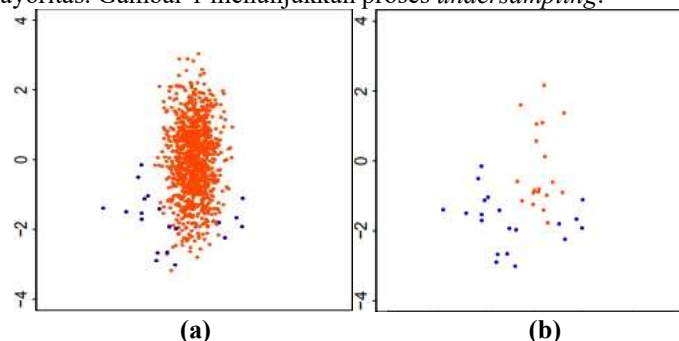
Pada penelitian ini, jenis data yang digunakan yaitu data sekunder. Data tersebut diperoleh dari data Survei Angkatan Kerja Nasional (SAKERNAS) tahun 2021 di Provinsi Sumatera Barat. Kegiatan SAKERNAS ini dilaksanakan oleh Badan Pusat Statistik Kabupaten/Kota yang ada di Provinsi Sumatera Barat. Variabel yang digunakan pada penelitian ini adalah jenis kelamin (X_1), tingkat pendidikan (X_2), usia (X_3), status dalam rumah tangga (X_4), pengalaman pelatihan kerja (X_5), status perkawinan (X_6), tempat tinggal (x_7) dan status kerja (Y).

B. Teknik Analisis Data

Dalam penelitian ini, *software R Studio* digunakan untuk melakukan analisis CART. Tahapan analisis CART adalah sebagai berikut.

1. Deskripsi data
Deskripsi data adalah rangkuman informasi dari kumpulan data penelitian untuk memberikan gambaran yang jelas.
2. Penanganan data

Teknik pengambilan sampel merupakan salah satu solusi untuk mengatasi kasus ketidakseimbangan berdasarkan tingkatan data (Choirunnisa dan Lianto, 2018). Secara umum teknik pengambilan sampel dibedakan menjadi tiga metode yaitu (1) *Undersampling*; (2) *Oversampling*; dan (3) Kombinasi antara *Undersampling* dan *Oversampling*. Pada penelitian ini digunakan metode *undersampling* untuk penanganan data. *Undersampling* merupakan pendekatan yang sederhana dan cepat dibandingkan dengan *oversampling* (Kaur dan Gosain, 2018). Menurut Lunardo dkk. (2014) proses *undersampling* adalah proses pengambilan data secara acak pada kelas mayoritas sehingga kelas mayoritas memiliki proporsi yang sama dengan kelas minoritas. Namun, *undersampling* memiliki kelemahan yaitu pengambilan data secara acak pada kelas mayoritas dapat menyebabkan kehilangan beberapa data yang merupakan informasi berguna yang terkandung dalam kelas mayoritas. Gambar 1 menunjukkan proses *undersampling*.



Gambar 1. (a) Data tidak Seimbang dan (b) Proses *Undersampling*

Berdasarkan Gambar 1 dapat diketahui bahwa pada gambar (a) data tidak seimbang karena proporsi kelas mayoritas lebih banyak dibandingkan proporsi kelas minoritas. Dengan demikian, maka dilakukan proses *undersampling* untuk memperoleh data yang seimbang yaitu proporsi kelas mayoritas dan proporsi kelas minoritas sama banyak seperti pada gambar (b).

3. Pembentukan pohon klasifikasi menggunakan langkah-langkah berikut.

a. Pemilahan pemilah

Langkah pertama dalam pembentukan pohon klasifikasi adalah menentukan calon simpul kiri dan calon simpul kanan menggunakan rumus pemilah berikut.

$$1) \text{ Variabel prediktor kontinu} : b-1 \text{ pemilahan} \quad (1)$$

$$2) \text{ Variabel prediktor kategori nominal dan ordinal} : 2^{L-1}-1 \text{ pemilahan} \quad (2)$$

dengan:

b : Banyak data dalam sebuah variabel

L : Banyaknya kategori dalam sebuah variabel

Metode pemilahan pemilah dapat diukur menggunakan *impurity measure* $i(t)$, yang merupakan pengukuran tingkat heterogenitas kelas dari simpul tertentu dalam pohon klasifikasi yang dapat membantu dalam menentukan pemilah yang paling cocok (Pratiwi dan Zain, 2014). Beberapa fungsi *impurity measure* adalah (1) Indeks Gini; (2) Indeks Informasi; (3) Indeks Twoing; dan (4) Indeks Entropi

Berdasarkan keempat fungsi *impurity measure* maka digunakan fungsi indeks gini untuk pemilahan pemilah. Menurut Breiman dkk (1993:103) indeks gini merupakan pemilahan pemilah yang sering digunakan karena lebih sederhana dan sesuai dengan berbagai kasus. Fungsi indeks gini adalah sebagai berikut.

$$i(t) = 1 - \sum_{j=1} p^2(j|t) \quad (3)$$

Dimana $i(t)$ adalah fungsi heterogenitas indeks gini dan $p(j|t)$ adalah proporsi kelas ke j pada simpul t . Pemilahan yang terpilih disebut dengan simpul. Kemudian simpul melakukan pemilahan lagi hingga diperoleh *terminal nodes*. Selanjutnya, dengan menggunakan rumus berikut dapat menentukan kriteria *goodness of split* untuk mengevaluasi pemilah oleh pemilah s pada simpul t .

$$\emptyset(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (4)$$

Dimana,

$\emptyset(s, t)$: Kriteria *goodness of split*

$i(t)$: Fungsi heterogenitas pada simpul t

$i(t_L)$: Fungsi heterogenitas pada simpul anak kiri

$i(t_R)$: Fungsi heterogenitas pada simpul anak kanan

P_L : Probabilitas pengamatan pada simpul kiri

P_R : Probabilitas pengamatan pada simpul kanan

Berdasarkan tahapan tersebut, simpul terbaik dihasilkan dari nilai $\emptyset(s, t)$ lebih tinggi karena dapat mengurangi tingkat heterogenitas yang lebih tinggi.

b. Penentuan simpul terminal

Untuk penghentian pertumbuhan pohon, maka ketika simpul t sudah tercapai secara maksimal sehingga tidak ada lagi penurunan heterogenitas yang signifikan, maka t tidak terpecah dan akan menjadi simpul terminal (Breiman dkk., 1993:26).

c. Penandaan label kelas

Penandaan label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah anggota kelas terbanyak, yaitu:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (5)$$

Keterangan:

$P(j|t)$: Proporsi kelas j pada *node* t

$N_j(t)$: Jumlah pengamatan kelas j pada *node* t

$N(t)$: Jumlah pengamatan pada *node* t

j_0 : Nilai dugaan kesalahan pengklasifikasian simpul t terbesar

4. Ukuran ketepatan klasifikasi

Untuk mengevaluasi proses klasifikasi maka dilakukan pengujian untuk proses prediksi kebenaran suatu objek. Proses evaluasi kinerja dapat diukur menggunakan nilai akurasi (*accuracy*) atau tingkat kesalahan (*error rate*). Pengujian tersebut dapat menggunakan tabel *confusion matrix* (Kantardzic, 2003:143). *Confusion matrix* adalah suatu tabel yang menunjukkan jumlah data uji yang benar diklasifikasikan dan yang salah diklasifikasikan. Tabel *confusion matrix* untuk klasifikasi dapat ditunjukkan pada Tabel 1.

Tabel 1. Confusion Matrix

| Klasifikasi Kelas | Kelas Prediksi |
|-------------------|----------------|
|-------------------|----------------|

| | | | |
|------------|---|----|----|
| | | 0 | 1 |
| Kelas Asli | 0 | TP | FN |
| | 1 | FP | TN |

Dimana,

0 : Bekerja

1 : Pengangguran

TP : Jumlah kelas bekerja dengan hasil prediksi kelas bekerja

TN : Jumlah kelas pengangguran dengan hasil prediksi kelas pengangguran

FP : Jumlah kelas pengangguran dengan hasil prediksi kelas bekerja

FN : Jumlah kelas bekerja dengan hasil prediksi kelas pengangguran

Hasil prediksi *accuracy* menunjukkan seberapa baik model dalam mengklasifikasikan data secara benar dan hasil prediksi *error rate* menunjukkan seberapa baik model dalam mengklasifikasikan data secara salah. Oleh karena itu, pengukuran nilai *accuracy* dapat dituliskan dengan persamaan berikut.

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

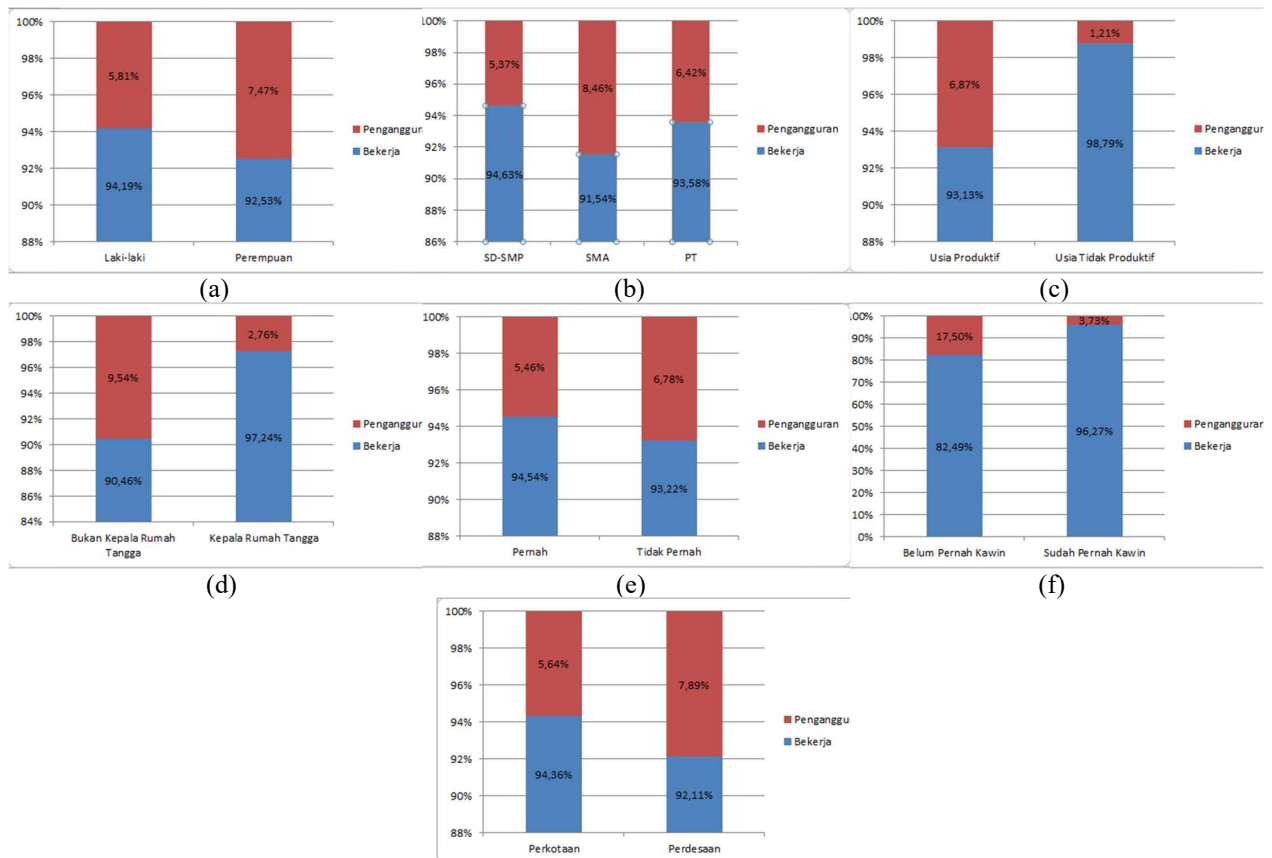
Sedangkan untuk menghitung *error rate* dapat menggunakan persamaan berikut.

$$Error\ rate = \frac{FP+FN}{TP+FN+FP+TN} \times 100\% \quad (7)$$

III. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Berdasarkan data angkatan kerja tahun 2021 terdapat 18.767 responden yang terdiri dari bekerja sebanyak 17.543 responden dan pengangguran sebanyak 1.224 responden. Pada penelitian ini terdapat tujuh variabel yang diduga berpengaruh terhadap pengangguran yaitu jenis kelamin, tingkat pendidikan, usia, status dalam rumah tangga, pengalaman pelatihan kerja, status perkawinan dan tempat tinggal. Deskripsi data dari masing-masing variabel disajikan dalam Gambar 2.



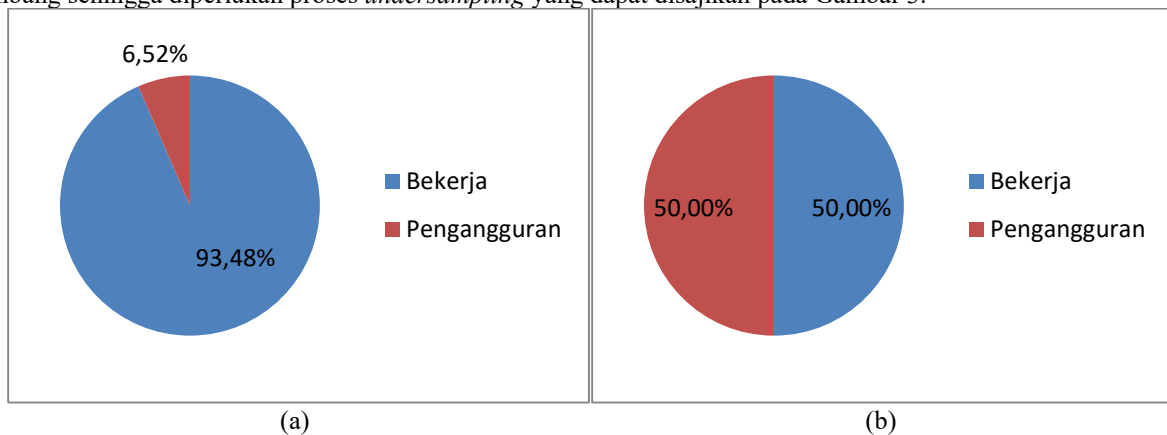
(g)

Gambar 2. (a) Jenis Kelamin, (b) Tingkat Pendidikan, (c) Usia, (d) Status dalam Rumah Tangga, (e) Pengalaman Pelatihan Kerja, (f) Status Perkawinan, dan (g) Tempat Tinggal

Berdasarkan Gambar 2, dapat dilihat bahwa pada gambar (a) menunjukkan bahwa tidak ada hubungan antara jenis kelamin dan status kerja karena persentase laki-laki dan perempuan yang bekerja dan pengangguran hampir sama. Gambar (b) persentase orang yang memiliki tingkat pendidikan SD-SMP, SMA dan Perguruan Tinggi yang menganggur hampir sama dengan persentase orang yang bekerja. Hasil ini menunjukkan bahwa tidak ada hubungan antara tingkat pendidikan dan status kerja. Gambar (c) menunjukkan bahwa tidak ada hubungan antara usia dan status kerja, karena usia produktif dan usia tidak produktif pada status kerja yang bekerja hampir sama, dan sebaliknya. Gambar (d) menunjukkan bahwa tidak ada hubungan antara status dalam rumah tangga dan status kerja, karena persentase bukan kepala rumah tangga dan kepala rumah tangga pada status kerja yang pengangguran memiliki persentase yang hampir sama, begitu juga untuk yang bekerja. Gambar (e) menunjukkan bahwa tidak ada hubungan antara pengalaman pelatihan kerja dan status kerja, karena persentase yang pernah dan tidak pernah mengikuti pelatihan kerja baik bekerja maupun pengangguran hampir sama. Gambar (f) menunjukkan bahwa status perkawinan belum pernah kawin dan sudah pernah kawin pada status kerja pengangguran memiliki persentase yang tidak seimbang, begitu juga untuk bekerja. Hasil ini menunjukkan ada hubungan antara status perkawinan dan status kerja. Gambar (g) menunjukkan bahwa persentase orang yang bekerja dan pengangguran di tempat tinggal perkotaan dan perdesaan hampir sama. Hasil ini menunjukkan tidak ada hubungan antara tempat tinggal dan status kerja.

B. Pembentukan Pohon Klasifikasi

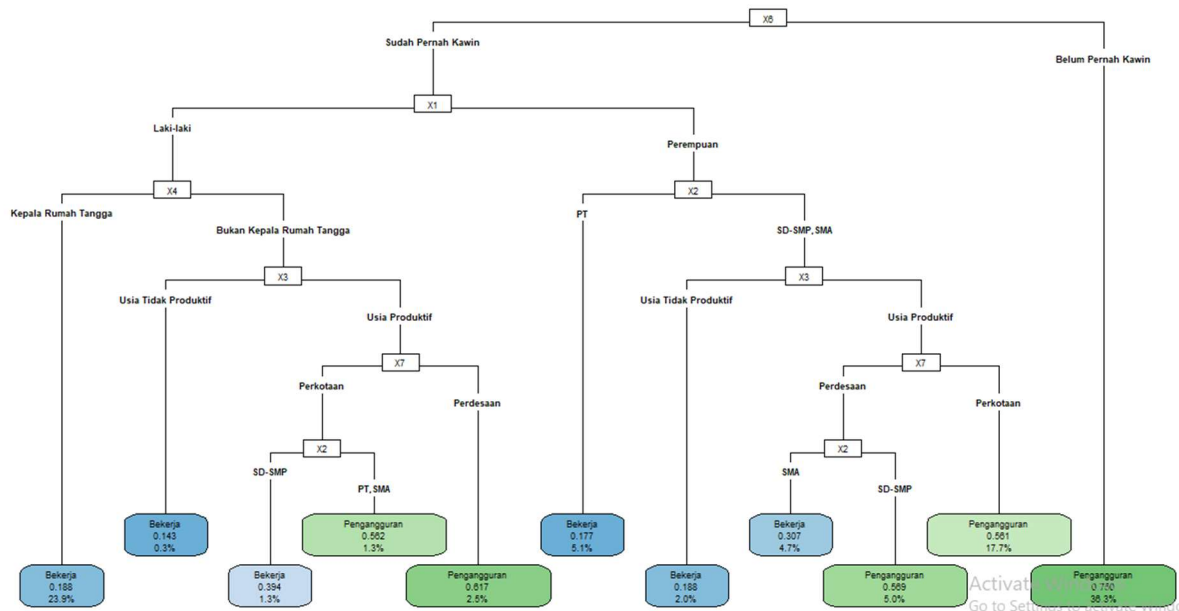
Berdasarkan data angkatan kerja tahun 2021 terdapat 18.767 responden yang terdiri dari bekerja sebanyak 17.543 responden dan pengangguran sebanyak 1.224 responden. Artinya data antara bekerja dan pengangguran tidaklah seimbang sehingga diperlukan proses *undersampling* yang dapat disajikan pada Gambar 3.



Gambar 3. (a) Proporsi Status Kerja dan (b) Proporsi Status Kerja setelah dilakukan Proses *Undersampling*

Berdasarkan Gambar 3 dapat dilihat bahwa pada Gambar (a) status kerja yang bekerja dan pengangguran memiliki proporsi yang tidak seimbang yaitu bekerja sebanyak 93,48% dan pengangguran sebanyak 6,52%. Untuk penanganan data tersebut dapat digunakan proses *undersampling* seperti yang disajikan pada Gambar (b), dimana status kerja yang bekerja dan pengangguran sudah seimbang yaitu masing-masing sebesar 50%.

Langkah selanjutnya dilakukan pembentukan pohon klasifikasi menggunakan analisis CART. Pohon klasifikasi terdiri dari tiga simpul yaitu simpul utama, simpul dalam, dan simpul terminal. Untuk mendapatkan simpul tersebut dilakukan pemilahan pemilah pada variabel prediktor terlebih dahulu yang dilanjutkan dengan menghitung nilai indeks gini dari pemilahan yang ada. Kemudian hasil perhitungan indeks gini digunakan untuk menentukan kriteria *goodness of split* yang paling tinggi sehingga diperoleh pohon klasifikasi seperti pada Gambar 4.



Gambar 4. Pohon Klasifikasi

Berdasarkan hasil analisis data pada Gambar 4 diperoleh aturan keputusan kriteria bekerja dan pengangguran yang disajikan pada Tabel 3.

Tabel 2. Aturan Keputusan Kriteria Status Kerja

| Kriteria Status Kerja | Aturan |
|-----------------------|--|
| Bekerja | <ol style="list-style-type: none"> 1. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, dan status dalam rumah tangga kepala rumah tangga. 2. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, dan tingkat pendidikan perguruan tinggi. 3. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, dan usia tidak produktif. 4. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, tingkat pendidikan SD-SMA sederajat, dan usia tidak produktif. 5. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, usia produktif, tempat tinggal perdesaan, dan tingkat pendidikan SMA sederajat. 6. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, tempat tinggal perkotaan, dan tingkat pendidikan SD-SMP sederajat. |
| Pengguguran | <ol style="list-style-type: none"> 1. Responden dengan status perkawinan belum pernah kawin. 2. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, dan tempat tinggal perdesaan. 3. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, tingkat pendidikan SD-SMA sederajat, usia produktif, dan tempat tinggal perdesaan. 4. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, tempat tinggal perkotaan, dan tingkat pendidikan SMA sederajat hingga perguruan tinggi. 5. Responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, usia produktif, tempat tinggal perdesaan, dan tingkat pendidikan SD-SMP sederajat. |

Berdasarkan Tabel 2 dapat disimpulkan bahwa atribut yang berpengaruh terhadap pengangguran di Provinsi Sumatera Barat tahun 2021 adalah variabel status perkawinan, jenis kelamin, status dalam rumah tangga, tingkat pendidikan, usia dan tempat tinggal.

Model pohon keputusan yang telah diperoleh dilakukan validasi sistem untuk mengukur kinerjanya dengan menggunakan *confusion matrix* yang disajikan pada Tabel 3.

Tabel 3. Ketepatan Klasifikasi Status Kerja

| Klasifikasi Kelas | Kelas Prediksi | |
|-------------------|----------------|--------------|
| | Bekerja | Pengangguran |
| Bekerja | 722 | 502 |
| Pengangguran | 190 | 1034 |

Berdasarkan Tabel 3 dapat dihitung nilai *accuracy* dan *error rate* menggunakan persamaan (6) dan (7) berikut.

$$Accuracy = \frac{722+1034}{722+502+190+1034} \times 100\% = 71,73\%$$

$$Error\ rate = \frac{190+502}{722+502+190+1034} \times 100\% = 28,27\%$$

Dari perhitungan tersebut dapat diperoleh nilai akurasi sebesar 71,73% artinya pohon klasifikasi dapat mengklasifikasikan data baru sebesar 71,73% dan tingkat kesalahan sebesar 28,27%.

IV. KESIMPULAN

Berdasarkan analisis yang dilakukan, maka diperoleh kesimpulan bahwa faktor-faktor yang mempengaruhi pengangguran di Provinsi Sumatera Barat tahun 2021 adalah status perkawinan, jenis kelamin, status dalam rumah tangga, tingkat pendidikan, usia dan tempat tinggal. Hasil pengklasifikasian tersebut berupa aturan berdasarkan model pohon keputusan yang menunjukkan bahwa responden akan bekerja jika (1) responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, dan status dalam rumah tangga kepala rumah tangga; (2) responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, dan tingkat pendidikan perguruan tinggi; (3) responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, dan usia tidak produktif; (4) responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, tingkat pendidikan SD-SMA sederajat, dan usia tidak produktif; (5) responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, usia produktif, tempat tinggal perdesaan, dan tingkat pendidikan SMA sederajat; dan (6) responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, tempat tinggal perkotaan, dan tingkat pendidikan SD-SMP sederajat. Sedangkan responden akan menganggur jika (1) responden dengan status perkawinan belum pernah kawin; (2) responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, dan tempat tinggal perdesaan; (3) responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, tingkat pendidikan SD-SMA sederajat, usia produktif, dan tempat tinggal perdesaan; (4) responden dengan status perkawinan sudah pernah kawin, jenis kelamin laki-laki, status dalam rumah tangga bukan kepala rumah tangga, usia produktif, tempat tinggal perkotaan, dan tingkat pendidikan SMA sederajat hingga perguruan tinggi; dan (5) responden dengan status perkawinan sudah pernah kawin, jenis kelamin perempuan, usia produktif, tempat tinggal perdesaan, dan tingkat pendidikan SD-SMP sederajat.

Pada penelitian ini diperoleh nilai akurasi yang cukup besar, maka disarankan untuk penelitian selanjutnya menggunakan analisis klasifikasi yang lain, seperti CHAID, Random Forest, SVM dan sebagainya agar diperoleh nilai akurasi yang besar. Selain itu, dapat dipertimbangkan lagi dalam penambahan faktor-faktor yang berpengaruh terhadap pengangguran.

DAFTAR PUSTAKA

- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1993). *Classification and Regression Trees*. Chapman And Hall: New York.
- Choirunnisa, S., & Lianto, J. (2018). "Hybrid Method of Undersampling and Oversampling for Handling Imbalanced Data", in *2018 International Seminal on Reseach of Information Technology and Intelligent Systems (ISRITI)*, hal. 276-280.

- Helena, K., Susanti, Y. and Respatiwan, R. (2019). "Penerapan Metode CHAID dan CART pada Klasifikasi Status Kerja di Kabupaten Brebes", *In Seminar Nasional Sains & Entrepreneurship*, Vol. 1, No. 1.
- HG, I. R. (2017). "Telaah Faktor-Faktor yang Mempengaruhi Status Gizi Balita di Kota Padang Berdasarkan Berat Badan Per Tinggi Badan Menggunakan Metode CART", *EKSAKTA: Berkala Ilmiah Bidang MIPA*, Vol. 18, No. 2, hal. 86-99.
- Ishak, K. (2018). "Faktor-faktor yang Mempengaruhi Pengangguran dan Inflasi terhadap Indeks Pembangunan di Indonesia", *Jurnal Ilmiah Ekonomi Kita*, Vol. 7, No. 1, hal. 22-38.
- Kantardzic, M. (2003). *Data Mining: concept, Models, Methods, and Algorithms*. John Wiley & Sons.
- Kaur, P., & Gosain, A. (2018). "Comparing the Behavior of Oversampling and Undersampling Approach of Class Imbalance Learning by Combining Class Imbalance Problem with Noise", *in ICT Based Innovations: Proceedings of CSI 2015*, hal. 23-30.
- Lunardon, N., Menardi, G., & Torelli, N. (2014). "ROSE: a package for binary imbalanced learning", *R Journal*, Vol. 6, No.1.
- Pratiwi, F. E., dan Zain, I. (2014). "Klasifikasi Pengangguran Terbuka Menggunakan CART (*Classification and Regression Tree*) di Provinsi Sulawesi Utara", *Jurnal Sains dan Seni Pomits*, Vol. 3, No. 1, hal 2337-3520.
- Steinberg, D. and P. Colla. (1995). *CART: Tree-Structured Non-Parametric Data Analysis*. San Diego, CA: Salford Systems.
- Swaramarinda, D. R. (2014). "Analisis Dampak pengangguran Terhadap Kemiskinan Di DKI Jakarta", *Jurnal Pendidikan Ekonomi dan Bisnis (JPEB)*, Vol. 2, No.2, hal 63-70.