

Penerapan Metode *Chi-Squared Automatic Interaction Detection* pada Segmentasi Karakteristik yang Mempengaruhi Pengangguran di Provinsi Banten Tahun 2022

Fahry Al Hayyu Fikham*, Anneke Iswani Achmad

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

*fahryalf157@gmail.com, annekeiswani11@gmail.com

Abstract. This study will discuss the application of the *Chi-Squared Auto Interaction Detection* (CHAID) method for segmenting characteristics that affect the unemployed workforce in Banten Province in 2022. The CHAID method was introduced by Dr. G. V Kass in 1980. In the CHAID method, segmentation will be formed which divides a sample into two or more different groups/categories based on the *chi-square* test statistics, then the segmentation results will be displayed in a tree diagram. In this study, the variables of gender, age, education level, marital status, status in the household and regional classification will be used as independent variables with labor force status as the dependent variable, where all data are categorical types which are suitable for use in CHAID analysis. This study uses secondary data sourced from Badan Pusat Statistik (BPS), using data from the result of Survei Hasil Angkatan Kerja Nasional (SAKERNAS). The results of the CHAID segmentation in this study resulted in 8 different segments with 3 of them producing a classification of the workforce with unemployed status. Then, based on the CHAID tree diagram, it was found that there were 5 (five) independent variables, namely gender, age, education level, household status and regional classification which had a significant association with labor force status which resulted in a classification accuracy of 70.6%.

Keywords: *Segmentation, Labor Force, Unemployment.*

Abstrak. Dalam penelitian ini akan membahas tentang penerapan metode *Chi-Squared Auto Interaction Detection* (CHAID) pada segmentasi karakteristik yang mempengaruhi angkatan kerja berstatus pengangguran di Provinsi Banten tahun 2022. Metode CHAID diperkenalkan pertama kali oleh Dr. G. V Kass pada tahun 1980. Pada metode CHAID akan membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok/kategori yang berbeda berdasarkan dengan statistik uji *chi-square* kemudian hasil segmentasi akan ditampilkan dalam diagram pohon. Pada penelitian ini akan menggunakan variabel jenis kelamin, usia, tingkat pendidikan, status pernikahan, status dalam rumah tangga dan klasifikasi wilayah sebagai variabel independen dengan status angkatan kerja sebagai variabel dependen, dimana semua datanya merupakan tipe kategorik yang cocok digunakan pada analisis CHAID. Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS), yaitu data hasil Survei Angkatan Kerja Nasional (SAKERNAS). Hasil segmentasi CHAID pada penelitian ini menghasilkan 8 segmentasi yang berbeda dengan 3 segmentasi diantaranya menghasilkan klasifikasi angkatan kerja yang berstatus pengangguran. Kemudian berdasarkan diagram pohon CHAID didapatkan bahwa terdapat 5 (lima) variabel independen yaitu jenis kelamin, usia, tingkat pendidikan, status dalam rumah tangga dan klasifikasi wilayah yang mempunyai asosiasi yang signifikan dengan status angkatan kerja yang menghasilkan ketepatan klasifikasi yaitu sebesar 70,6%..

Kata Kunci: *Segmentasi, Angkatan Kerja, Pengangguran.*

A. Pendahuluan

Klasifikasi merupakan metode yang prosesnya melibatkan pengelompokan data menurut kemiripan dalam data. Klasifikasi telah banyak digunakan di berbagai bidang seperti *e-commerce*, akademik, segmentasi pasar, penelitian lingkungan, ilmu kedokteran, sosial, ketenagakerjaan dan lain-lain. Klasifikasi merupakan suatu pekerjaan mengevaluasi atau menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia (Fitrianto dkk., 2022).

Metode klasifikasi yang digunakan untuk membuat keputusan dengan pendekatan nonparametrik adalah *decision tree* atau pohon keputusan. Menurut Ye (2014) pohon keputusan digunakan untuk mempelajari klasifikasi dan prediksi pola dari data dan menggambarkan relasi atau hubungan dari variabel independen dan variabel dependen dalam bentuk pohon. Pohon keputusan adalah metode yang digunakan untuk mengelompokkan suatu objek yang ditampilkan melalui gambar daun dan cabang. Daun pohon keputusan ditentukan oleh suatu kelas, sementara cabangnya digunakan untuk mewakili suatu kondisi dari atribut objek yang terukur. Pohon keputusan memiliki beberapa metode, dimana salah satu metode yang digunakan untuk data kategorik salah satunya adalah metode *Chi-Squared Automatic Interaction Detection* (CHAID).

Metode CHAID membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kategori yang berbeda berdasarkan dengan statistik uji *chi-square*. Proses ini kemudian diteruskan dengan membagi kategori-kategori tersebut menjadi kategori yang lebih kecil berdasarkan variabel-variabel independen yang lain. Proses tersebut terus berlanjut hingga memenuhi salah satu kriteria *stopping* (Kunto dan Hasana, 2006).

Diagram pohon CHAID memiliki keunggulan yang memungkinkan sampel untuk dibagi menjadi maksimum 15 kategori (tergantung berapa banyak kategori yang dimiliki variabel independen terbaik) di titik manapun di pohon (Struhl, 1992). Selain itu metode CHAID lebih mudah dan cepat digunakan dibandingkan metode lainnya, serta *node* dan cabang yang dihasilkan oleh CHAID didasarkan pada hasil nilai *chi-square*, sehingga *node* pada diagram pohon saling berhubungan atau terkait dan lebih mudah untuk mengetahui variabel independen mana yang paling signifikan (Nisbet dkk., 2009).

Sedangkan kekurangan dari metode CHAID adalah data atau variabel yang digunakan harus berupa data kategorik, jika data asli berupa rasio atau skala, maka peneliti perlu mengkategorikan ke dalam kategori tertentu dan hasil CHAID terkadang *multiple split*, sehingga sulit merealisasikan dalam kondisi sebenarnya (Perez dan Cejas, 2016).

Ketenagakerjaan adalah salah satu hal yang sangat esensial dalam usaha memajukan perekonomian suatu bangsa. Indonesia sebagai negara yang berkembang merupakan satu dari banyak negara yang memiliki masalah mengenai ketenagakerjaan (Fadilah & Kudus, 2023). Masalah ketenagakerjaan adalah masalah yang sangat luas dan kompleks. Salah satunya adalah masalah pengangguran pada angkatan kerja di Indonesia yang muncul sebagai imbas dari jumlah angkatan kerja yang tidak seimbang dengan jumlah permintaan lapangan pekerjaan yang tersedia. Dalam masalah ketenagakerjaan hal ini menunjukkan bahwa semakin tinggi angka pengangguran maka akan menambah beban perekonomian daerah dan mengurangi kesejahteraan rakyat (Hadi, 2009).

Provinsi Banten merupakan provinsi yang mempunyai peluang yang tinggi dalam investasi sektor industri karena mempunyai infrastruktur yang memadai dalam mendukung sektor industri serta di Provinsi Banten memiliki beberapa kawasan industri besar yang tersebar di wilayahnya. Dengan melihat peluang sektor industri yang sangat besar yang seharusnya dapat meningkatkan penyerapan tenaga kerja akan tetapi dalam kenyataannya BPS mencatat bahwa Provinsi Banten menduduki peringkat ketiga sebagai provinsi dengan Tingkat Pengangguran Terbuka (TPT) tertinggi di Indonesia yaitu sebesar 8,09% pada periode Agustus tahun 2022. Dimana angka TPT merupakan salah satu tolak ukur dari permasalahan pengangguran di Indonesia yang menggambarkan proporsi angkatan kerja yang tidak memiliki pekerjaan dan secara aktif mencari dan bersedia untuk bekerja.

Penelitian ini akan menggunakan metode *Chi-Squared Auto Interaction Detection* (CHAID) pada pengangguran di Provinsi Banten periode Agustus 2022 dengan karakteristik yang diduga mempengaruhi pengangguran yaitu jenis kelamin, usia, tingkat pendidikan, status pernikahan, status dalam rumah tangga dan klasifikasi wilayah.

Berdasarkan penjelasan pada latar belakang, maka perumusan masalah dalam penelitian ini adalah “Bagaimana segmentasi karakteristik yang mempengaruhi pengangguran di Provinsi Banten tahun 2022 dengan menggunakan metode CHAID”. Selanjutnya, tujuan dari penelitian ini adalah untuk mengetahui segmentasi karakteristik yang mempengaruhi pengangguran di Provinsi Banten tahun 2022 dengan menggunakan metode CHAID

B. Metodologi Penelitian

Jenis penelitian yang dilakukan dalam skripsi ini adalah penelitian terapan. Dimana akan dilakukan penerapan metode CHAID pada data angkatan kerja di Provinsi Banten yang berasal dari hasil SAKERNAS pada tahun 2022 yang berasal dari Badan Pusat Statistik. Sampel pada penelitian ini berjumlah sebanyak 14211 orang. Pada penelitian ini akan menggunakan variabel jenis kelamin (X1), usia (X2), tingkat pendidikan (X3), status pernikahan (X4), status dalam rumah tangga (X5) dan klasifikasi wilayah (X6) sebagai variabel independen dengan status angkatan kerja (Y) sebagai variabel dependen.

Metode Analisis Data

Chi-Squared Auto Interaction Detection (CHAID)

Metode CHAID (*Chi-Squared Automatic Interaction Detection*) diperkenalkan pertama kali oleh Dr. G. V Kass pada tahun 1980 di sebuah artikel berjudul “An Exploratory Technique for Investigating Large Quantities of Categorical Data” dalam buku *Applied Statistics* (Kass, 1980). CHAID digunakan untuk menduga sebuah variabel tunggal sebagai variabel dependen, yang didasarkan pada sejumlah variabel-variabel lain sebagai variabel-variabel independen. CHAID merupakan suatu metode iteratif yang menguji satu-persatu variabel independen yang digunakan dalam klasifikasi, dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya (Gallagher, 2000).

Variabel Pada Analisis CHAID

Menurut Gallagher (2000), variabel independen dalam metode CHAID dapat dibedakan menjadi 3 jenis, yaitu:

1. Variabel Monotonik, adalah variabel independen yang berskala ordinal, dimana kategori-kategori di dalamnya dapat digabungkan jika berurutan.
2. Variabel Bebas, adalah variabel independen yang berskala nominal, dimana kategori-kategori di dalamnya dapat digabungkan meskipun tidak berurutan.
3. Variabel Mengambang (*Floating*), adalah variabel monotonik yang mengandung nilai kategori yang tidak jelas posisinya dalam urutan.

Uji Chi – Square

Statistik uji *chi-square* dapat digunakan untuk mengetahui independensi (kebebasan) antara dua variabel. Pada CHAID akan menguji signifikansi *chi-square* antara setiap variabel independen dan dependennya. Uji *chi-square* merupakan uji non parametrik. Hipotesis pada pengujian statistik *chi-square* adalah sebagai berikut:

- H0 : Tidak terdapat asosiasi antara variabel pertama dan variabel kedua
 H1 : Terdapat asosiasi antara variabel pertama dan variabel kedua

Untuk menguji hipotesis di atas, digunakan statistik uji *chi-square* sebagai berikut.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad \dots (1)$$

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}} \quad \dots (2)$$

Keterangan:

χ^2 : Nilai *chi-square*

E_{ij} : Frekuensi harapan pengamatan yang termasuk dalam kategori ke-*i* dari variabel pertama dan kategori ke-*j* dari variabel kedua.

r : Jumlah kategori dalam variabel pertama.

c : Jumlah kategori dalam variabel kedua.

n_{ij} : Banyaknya pengamatan yang termasuk dalam kategori ke-*i* dari variabel pertama dan kategori ke-*j* dari variabel kedua.

$n_{i.}$: Banyaknya pengamatan yang termasuk dalam kategori ke-*i* dari variabel pertama.

$n_{.j}$: Banyaknya pengamatan yang termasuk dalam kategori ke-*j* dari variabel kedua.

$n_{..}$: Banyaknya seluruh pengamatan.

Menurut Daniel (1978) kriteria pengambilan keputusan dalam uji *chi-square* yaitu H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2 = \chi_{\alpha; df=(r-1)(c-1)}^2$ dengan kriteria lain yaitu membandingkan nilai signifikansi (*p-value*) dengan taraf signifikansi (α). Dengan kriteria H_0 ditolak jika *p-value* (*sig.*) $< \alpha$. . Selanjutnya membuat kesimpulan yang dibuat berdasarkan hasil dari penaksiran apakah H_0 diterima atau ditolak.

Metode CHAID menggunakan statistik *chi-square* dalam dua tahap algoritma CHAID. Yang pertama, statistik *chi-square* digunakan untuk menentukan apakah kategori-kategori dalam sebuah variabel independen dapat digabungkan menjadi satu. Kedua, ketika semua variabel independen sudah diringkas kategorinya menjadi bentuk yang signifikan dan tidak mungkin digabung lagi, statistik *chi-square* digunakan untuk menentukan variabel independen mana yang paling signifikan untuk membagi atau membedakan kategori-kategori dalam variabel dependen.

Koreksi Bonferroni

Koreksi Bonferroni adalah banyaknya cara yang mungkin yang mana *c* kategori dapat digabungkan menjadi *r* kategori. Dengan demikian nilai *p-value* dari uji *chi-square* untuk independensi yang baru merupakan perkaliannya dengan pengali Bonferroni sesuai dengan jenis variabelnya (Gallagher, 2000). Gallagher (2000) menyebutkan bahwa pengali Bonferroni untuk masing-masing jenis variabel independen adalah sebagai berikut:

1. Variabel Monotonik

$$M = \binom{c-1}{r-1} \quad \dots (3)$$

2. Variabel Bebas

$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-i)^c}{i! (r-i)!} \quad \dots (4)$$

3. Variabel Mengambang

$$M = \binom{c-2}{r-2} + r \binom{c-2}{r-1} \quad \dots (5)$$

Keterangan:

- M : Pengali Bonferroni
 c : Banyaknya kategori variabel independen awal
 r : Banyaknya kategori variabel independen setelah penggabungan

Algoritma CHAID

Secara garis besar algoritma ini dapat dibagi menjadi tiga tahap, yaitu penggabungan (*merging*), pemisahan (*splitting*) dan penghentian (*stopping*). Diagram pohon dimulai dari *root node* (simpul akar) melalui tiga tahap tersebut pada setiap simpul (*node*) yang terbentuk dan dilakukan secara berulang hingga memenuhi salah satu kriteria *stopping* (Aniswita, 2023).

Penggabungan (*Merging*)

Tahap *merging* dilakukan pada variabel independen yang mempunyai kategori lebih dari 2 (Siswantining dkk., 2018). Pada tahap ini diperiksa uji signifikansi *chi-square* dari masing-masing kategori variabel independen terhadap variabel dependen. Apabila terdapat pasangan kategori yang tidak signifikan maka gabungkan menjadi kategori baru dan lakukan koreksi Bonferroni, jika signifikan maka pasangan kategori tidak digabung kemudian lanjutkan ke tahap selanjutnya tanpa dilakukan koreksi Bonferroni.

Pemisahan (*Splitting*)

Pada tahap ini akan memilih variabel independen yang akan digunakan sebagai pemisah simpul. Pemilihan dilakukan dengan mengurutkan *p-value* (menggunakan *p-value* terkoreksi Bonferroni jika melalui proses penggabungan) dan *chi-square* pada setiap variabel independen. Kemudian pilih variabel independen yang memiliki nilai *chi square* terbesar dan *p-value* terkecil.

Penghentian (*Stopping*)

Proses penggabungan dan pemisahan terus dilakukan hingga memenuhi salah satu syarat *stopping* berikut:

1. Tidak ada lagi variabel independen yang signifikan menunjukkan asosiasi terhadap variabel dependen.
2. Jika kedalaman pohon mencapai batas kedalaman pohon maksimal. Kedalaman pohon maksimum pada IBM *Statistics SPSS* adalah 3 yang sudah ditetapkan secara otomatis oleh program *SPSS*, hal ini mencegah pertumbuhan pohon yang terlalu besar dan dapat menyebabkan hasil pohon kurang optimal (Hasibuan, 2019).
3. Jika ukuran dari simpul anak kurang dari nilai ukuran simpul anak minimum yang telah ditentukan dengan nilai ukuran simpul anak minimum pada penelitian ini adalah 20.

Ukuran Ketepatan Klasifikasi

Untuk mengevaluasi hasil ketepatan klasifikasi yaitu dengan cara menghitung akurasi klasifikasi. Pada penelitian ini evaluasi akurasi klasifikasi ialah dengan menghitung nilai APER (*apparent error rate*) dan *total accuracy rate*. APER merupakan proporsi observasi yang diprediksi secara tidak benar oleh fungsi klasifikasi, sedangkan *total accuracy rate* merupakan proporsi observasi yang diprediksi secara benar oleh fungsi klasifikasi dan berguna untuk mengukur tingkat ketepatan klasifikasi (Johnson dan Winchern, 2007).

Ukuran-ukuran tersebut dapat dihitung dengan mudah dari matriks konfusi yang menunjukkan keanggotaan observasi dan prediksi. Matriks konfusi dapat dilihat pada tabel 1 berikut.

Tabel 1. Matriks Konfusi.

OBSERVASI Y	PREDIKSI Y		TOTAL
	1	2	
1	n_{11}	n_{12}	$n_{1.} = n_{11} + n_{12}$
2	n_{21}	n_{22}	$n_{2.} = n_{21} + n_{22}$
TOTAL	$n_{.1} = n_{11} + n_{21}$	$n_{.2} = n_{12} + n_{22}$	N

Dengan menggunakan tabel diatas, maka nilai APER dan *total accuracy rate* dapat dihitung dengan persamaan sebagai berikut.

$$APER = \frac{n_{12} + n_{21}}{N} \quad \dots (2.8)$$

$$Total Accuracy Rate = 1 - APER = \frac{n_{11} + n_{22}}{N} \quad \dots (2.9)$$

Keterangan:

- n_{11} : Jumlah subjek dari y_1 tepat diklasifikasikan sebagai y_1
 n_{12} : Jumlah subjek dari y_1 salah diklasifikasikan sebagai y_2
 n_{21} : Jumlah subjek dari y_2 salah diklasifikasikan sebagai y_1
 n_{22} : Jumlah subjek dari y_2 tepat diklasifikasikan sebagai y_2
 N : Jumlah keseluruhan observasi

Semakin besar *total accuracy rate* atau semakin kecil nilai APER, maka dapat dikatakan bahwa tingkat hasil klasifikasi akan semakin baik

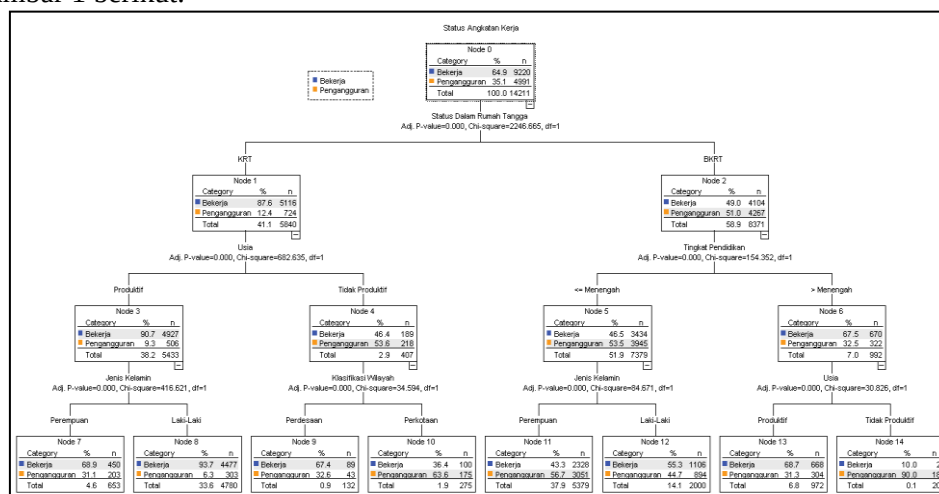
C. Hasil Penelitian dan Pembahasan

Pembentukan Segmentasi pada Analisis CHAID

Berdasarkan hasil algoritma CHAID terdapat 5 (lima) variabel independen yang mempunyai asosiasi yang signifikan terhadap variabel dependen atau status angkatan kerja, yaitu jenis kelamin, usia, tingkat pendidikan, status dalam rumah tangga dan klasifikasi wilayah, sedangkan status pernikahan tidak mempunyai asosiasi yang signifikan terhadap variabel status angkatan kerja. Variabel status dalam rumah tangga merupakan variabel independen terbaik yang digunakan untuk membagi variabel dependennya yaitu status angkatan kerja pada kedalaman pohon pertama.

Kemudian pada kedalaman kedua variabel usia merupakan variabel independen yang paling signifikan untuk memisah kategori pada node ke-1 dan pada node ke-2 dibagi lagi berdasarkan variabel tingkat pendidikan dimana pada tahap ini terjadi penggabungan pada kategori dasar dan menengah menjadi satu kategori. Selanjutnya pada kedalaman ketiga, variabel jenis kelamin, klasifikasi wilayah dan usia merupakan variabel yang paling menunjukkan kesignifikansian terhadap variabel dependen lalu pertumbuhan pohon berhenti pada kedalaman ketiga ini yang kemudian menghasilkan terminal simpul (terminal node) sebanyak delapan node akhir yang menjadi segmentasi angkatan kerja.

Adapun hasil diagram pohon CHAID pada segmentasi angkatan kerja yang dapat dilihat pada gambar 1 berikut.



Gambar 1. Diagram Pohon CHAID Segmentasi Angkatan Kerja di Provinsi Banten

Berdasarkan gambar di atas terlihat bahwa terdapat 8 segmentasi hasil klasifikasi CHAID pada status angkatan kerja, dimana 5 segmen yaitu pada node 7, 8, 9, 12 dan 13 menghasilkan klasifikasi angkatan kerja dengan status bekerja dan 3 segmen lainnya yaitu pada node 10, 11 dan 14 menghasilkan klasifikasi angkatan kerja dengan status pengangguran.

Dari hasil segmentasi tersebut angkatan kerja dengan segmentasi berstatus sebagai kepala rumah tangga, berada pada usia yang produktif dan berjenis kelamin laki-laki merupakan segmentasi dengan persentase terbesar yaitu 93,7% yang diklasifikasi sebagai angkatan kerja dengan status bekerja.

Sedangkan angkatan kerja dengan segmentasi berstatus sebagai bukan kepala rumah tangga, dengan tingkat pendidikan tinggi serta berada pada usia yang tidak produktif merupakan segmentasi dengan persentase terbesar yaitu 90% yang diklasifikasi sebagai angkatan kerja dengan status pengangguran.

Evaluasi Ketepatan Klasifikasi

Setelah pohon CHAID terbentuk dan pelabelan kelas dilakukan, maka dilakukan evaluasi dari hasil klasifikasi CHAID, maka evaluasi ketepatan klasifikasi pada penelitian ini dapat dilihat pada tabel 2 berikut

Tabel 2. Evaluasi Ketepatan Klasifikasi CHAID.

Observasi	Prediksi		Total
	Bekerja	Pengangguran	
Bekerja	6790	2430	9220
Pengangguran	1747	3244	4991
Total	8537	5674	14211

Berdasarkan tabel di atas, dapat diketahui nilai APER dan *total accuracy rate*, kemudian diperoleh hasil sebagai berikut:

$$APER = \frac{n_{12}+n_{21}}{N} = \frac{2430+1747}{14211} = 0,294$$

$$Total Accuracy Rate = \frac{n_{11}+n_{22}}{N} = \frac{6790 + 3244}{14211} = 0,706$$

Setelah dilakukan perhitungan, didapatkan nilai APER (*apparent rate error*) dan *total accuracy rate* untuk mengukur ketepatan klasifikasi pada angkatan kerja. Diperoleh nilai APER sebesar 0,294 atau 29,4%, artinya besarnya proporsi observasi yang diprediksi secara tidak benar oleh fungsi klasifikasi yaitu 0,294 atau 29,4%. Sedangkan nilai *total accuracy rate* yang diperoleh sebesar 0,706 atau 70,6%, artinya besarnya proporsi observasi yang diprediksi secara benar oleh fungsi klasifikasi yaitu 0,706 atau 70,6%. Dari hasil tersebut juga dapat dikatakan bahwa pohon hasil klasifikasi yang terbentuk mampu mengklasifikasi sampel dengan benar sebesar 70,6%.

D. Kesimpulan

Berdasarkan hasil dan pembahasan sebelumnya dapat disimpulkan bahwa metode CHAID menghasilkan 8 segmentasi yang berbeda dengan 3 segmentasi diantaranya menghasilkan klasifikasi angkatan kerja yang berstatus pengangguran. Hasil segmentasi tersebut selanjutnya dapat digunakan untuk memberikan gambaran mengenai karakteristik apa saja yang berpengaruh terhadap angkatan kerja yang berstatus sebagai pengangguran, sehingga meminimalisir terjadinya lonjakan pengangguran pada angkatan kerja untuk masa yang akan datang.

Pada hasil diagram pohon CHAID yang terbentuk juga terlihat bahwa terdapat 5 (lima) variabel independen yaitu jenis kelamin (X1), usia (X2), tingkat pendidikan (X3), status dalam

rumah tangga (X5) dan klasifikasi wilayah (X6) yang mempunyai asosiasi yang signifikan dengan variabel dependennya yaitu status angkatan kerja (Y).

Acknowledge

Puji dan syukur penulis panjatkan kehadiran Allah SWT karena berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan penelitian ini. Terimakasih kepada ayah, ibu, kakak-kakak dan keluarga yang selalu mendo'akan dan memberi dukungan baik moral maupun materi kepada penulis. Terimakasih kepada Ibu Anneke Iswani Achmad, Dra. M.Si. yang telah memberi bimbingan kepada penulis hingga penelitian ini selesai. Dosen-dosen Program Studi Statistika Universitas Islam Bandung yang telah banyak memberikan ilmu pengetahuan. Sahabat dan teman-teman serta semua pihak yang telah hingga penelitian ini selesai.

Daftar Pustaka

- [1] Aniswita, Fitri, H., Efendi, R., Sastri, L., & Rahmadani, P. (2023). *Determinant of the Study Period for Mathematics Education Students at IAIN Bukittinggi Based on CHAID Method*. Jurnal Iqra': Kajian Ilmu Pendidikan, 8(1), 9-23.
- [2] Daniel, Wayne, W. (1978). *Statistik Nonparametrik Terapan*. Terjemahan oleh Alex Tri Kantjono W. (1989). Jakarta: Gramedia.
- [3] Fitrianto, A., Wan Muhamad, W.Z.A., & Susetyo, B. (2022). *Development Of Direct Marketing Strategy for Banking Industry: The Use of a Chi-Squared Automatic Interaction Detector (CHAID) In Deposit Subscription Classification*. Journal of Socioeconomics and Development, 5(1), 64-75.
- [4] Gallagher, C.A., H. M. Monroe, & J.L. Fish. (2000). *An Iterative Approach to Classification Analysis* (online), (<https://www.casact.org/pubs/dpp/dpp90/90dpp237.pdf>, diakses pada 11 Mei 2023).
- [5] Hadi, Setia Tunggal (2009). *Pengantar Hukum Ketenagakerjaan Indonesia*. Jakarta: Harvarindo.
- [6] Hasibuan, E.A., & Harahap, A.N. (2019). *Aplikasi Metode CHAID dalam Menganalisis Kecendrungan Penelitian Skripsi Mahasiswa pada Program Studi Pendidikan Matematika*. Edumatika: Jurnal Riset Pendidikan Matematika. 2(2), 89-97.
- [7] Johnson, R. A. & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis - 6th edition*. New Jersey: Prentice Hall.
- [8] Kass, G. V. (1980). *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. Applied Statistics, 29 (2), 119-127.
- [9] Kunto, Y. S. & S. N. Hasana (2006). *Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar (Studi Kasus Pada Koperasi Syari'ah Alhidayah)*. Jurnal Manajemen Pemasaran. 1(2), 88-98.
- [10] Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Canada: Elsevier Inc.
- [11] Perez, F.M.D., & M. B. Cejas. (2016) *CHAID Algorithm as An Appropriate Analytical Method for Tourism Market Segmentation*. Journal of Destination Marketing & Management. 5, 275-282.
- [12] Siswantining, T., Maghfiroh, Kamelia, T., & Sarwinda, D. (2018). *Implementation of Chi Square Automatic Interaction Detection (CHAID) Method to Identify Type 2 Diabetes Mellitus in Tuberculosis Patient. A Case Study in Cipto Mangunkusumo Hospital*. Journal of Physics: Conference Series.
- [13] Struhl, Steven. (1992). *Classification Tree Methods: AID, CHAID and CART* (online), (<https://www.quirks.com/articles/classification-tree-methods-aid-chaid-and-cart>, diakses pada 28 Mei 2023).

- [14] Ye, N. (2014). *Data Mining Theories, Algorithms, and Examples*. Boca Raton: Taylor & Francis Group, LLC.
- [15] Fadilah, T. S., & Kudus, A. (2023). Penerapan Metode Regresi Kernel Smoothing untuk Imputasi Data Lama Waktu Terinfeksi Covid-19. *Jurnal Riset Statistika*, 51–60. <https://doi.org/10.29313/jrs.v3i1.1802>