

**PENGLASIFIKASIAN STATUS KERJA PADA ANGKATAN
KERJA DI KOTA PADANG MENGGUNAKAN *RANDOM
OVERSAMPLING* DENGAN METODE *CLASSIFICATION
AND REGRESSION TREE***

TUGAS AKHIR

Diajukan sebagai salah satu persyaratan untuk memperoleh gelar Ahli Madya



Oleh
RAHMI NOFITA
NIM 20037053

**PROGRAM STUDI DIPLOMA III STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

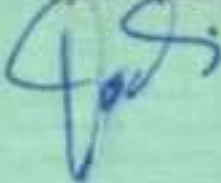
PERSETUJUAN TUGAS AKHIR

PENGLASIFIKASIAN STATUS KERJA PADA ANGKATAN KERJA DI KOTA PADANG MENGGUNAKAN *RANDOM OVERSAMPLING* DENGAN METODE *CLASSIFICATION AND REGRESSION TREE*

Nama : Halimi Nofita
NIM : 20037053
Program Studi : D3 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

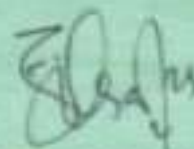
Padang, 25 Agustus 2023

Mengetahui
Kepala Departemen Statistika



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

Disetujui Oleh:
Pembimbing



Zilrahmi, S.Pd., M.Si
NIP. 198911062019032009

PENGESAHAN LULUS UJIAN TUGAS AKHIR




Nama : Rahmi Nofita
NIM : 20037053
Program Studi : D3 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PENGKLASIFIKASIAN STATUS KERJA PADA ANGGARAN KERJA DI KOTA PADANG MENGGUNAKAN *RANDOM OVERSAMPLING* DENGAN METODE *CLASSIFICATION AND REGRESSION TREE*

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Tugas Akhir
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 25 Agustus 2023

Tim Penguji

	Nama	Tanda Tangan
Ketua	: Zilrahmi, S.Pd., M.Si	
Anggota	: Dr. Yenni Kurniawati, M.Si	
Anggota	: Dima Fitria, M.Si	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Rahmi Nofita
NIM : 20037053
Program Studi : D3 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, Tugas Akhir saya dengan judul "**Pengklasifikasian Status Kerja pada Angkatan Kerja Di Kota Padang Menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree***" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Kepala Departemen Statistika,



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

Saya yang menyatakan,



Rahmi Nofita
NIM. 20037053

ABSTRAK

Rahmi Nofita : Pengklasifikasian Status Kerja pada Angkatan Kerja di Kota Padang Menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*

Pertumbuhan penduduk Kota Padang terus mengalami peningkatan. Seiring dengan banyaknya jumlah penduduk, Kota Padang menjadi wilayah di Provinsi Sumatera Barat yang memiliki jumlah pengangguran tertinggi. Kondisi status angkatan kerja di Kota Padang sangat tidak seimbang (*imbalance*) sehingga perlu dilakukan teknik *imbalance* data agar dapat meningkatkan kinerja hasil klasifikasi. Penelitian ini bertujuan untuk mengetahui bentuk pohon klasifikasi CART, mengetahui karakteristik status angkatan kerja di Kota Padang dan mengetahui tingkat ketepatan klasifikasi status angkatan kerja di Kota Padang.

Penelitian ini merupakan penelitian terapan. Data yang digunakan adalah data Survei Angkatan Kerja Nasional (SAKERNAS) tahun 2022 yang diperoleh dari BPS Provinsi Sumatera Barat dengan variabel yang diperhatikan adalah klasifikasi tempat tinggal, status dalam rumah tangga, jenis kelamin, usia, status perkawinan, tingkat pendidikan, dan pelatihan kerja. Jumlah data yang digunakan sebanyak 1.249 orang responden. Analisis yang digunakan adalah CART dengan terlebih dahulu dilakukan penanganan *imbalance* dengan metode *random oversampling*.

Hasil penelitian dengan menggunakan *random oversampling* untuk mengatasi *imbalance* data pada CART diperoleh pohon maksimal yang terbentuk sebanyak 8 simpul terminal dengan 4 label kelas bekerja dan 4 label kelas pengangguran. Karakteristik yang paling dominan dalam mencirikan orang bekerja di Kota Padang adalah bukan kepala rumah tangga, berpendidikan dasar/menengah, berada pada usia produktif, jenis kelamin laki-laki, dan bertempat tinggal diperkotaan. Karakteristik yang paling dominan dalam mencirikan orang pengangguran di Kota Padang adalah berpendidikan dasar/menengah dan berstatus sebagai kepala rumah tangga. Nilai sensitivitas yang diperoleh sebesar 51,34% sehingga ketepatan klasifikasi dalam memprediksi pengangguran sebagai pengangguran sudah lebih tepat.

Kata Kunci: Angkatan Kerja, CART, Padang, *Random Oversampling*

ABSTRACT

Rahmi Nofita : *Classification of Work Status in the Labor Force in Padang City Using Random Oversampling with the Classification and Regression Tree Method*

The population growth of Padang City continues to increase. Along with the large population, Padang City is the region in West Sumatra Province that has the highest number of unemployed. The condition of the labor force status in Padang City is very unbalanced so it is necessary to use imbalance data techniques in order to improve the performance of classification results. This research aims to determine the shape of the CART classification tree, determine the characteristics of the labor force status in Padang City and determine the level of accuracy of the labor force status classification in Padang City.

This research is applied research. The data used is the 2022 National Labor Force Survey (SAKERNAS) data obtained from BPS West Sumatra Province with the variables considered being residence classification, household status, gender, age, marital status, education level and job training. The amount of data used was 1,249 respondents. The analysis used is CART with imbalance handling first using the random oversampling method.

The results of research using random oversampling to overcome imbalance data in CART obtained a maximum tree formed of 8 terminal nodes with 4 working class labels and 4 unemployed class labels. The most dominant characteristics in characterizing working people in Padang City are not being the head of the household, having primary/secondary education, being of productive age, being male, and living in an urban area. The most dominant characteristics in characterizing unemployed people in Padang City are primary/secondary education and status as head of household. The sensitivity value obtained was 51.34% so that the classification accuracy in predicting unemployment as unemployed was more precise.

Keywords: *Labor Force, CART, Padang, Random Oversampling*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas berkas rahmat dan karunianya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul **“Pengklasifikasian Status Kerja pada Angkatan Kerja Di Kota Padang Menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*”**. Tugas Akhir ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi DIII Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

Peneliti ingin mengucapkan terimakasih kepada semua pihak yang telah memberi saran dan idenya sehingga Tugas Akhir ini dapat diselesaikan, yaitu:

1. Bapak Dodi Vionanda, M.Si., Ph.D sebagai Kepala Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
2. Ibu Dra. Nonong Amalita, M.Si., Sekretaris Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
3. Ibu Dr. Yenni Kurniawati, M.Si sebagai Koordinator Program Studi DIII Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang dan sekaligus dosen penguji.
4. Ibu Zilrahmi, S.Pd., M.Si sebagai dosen pembimbing Tugas Akhir dan penasehat akademik yang telah memberikan bimbingan, saran, dan nasehat dalam menyelesaikan Tugas Akhir ini, sekaligus sebagai tenaga pengajar Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Ibu Dina Fitria, M.Si sebagai dosen penguji yang telah memberikan masukan, saran dan nasehat dalam pengerjaan Tugas Akhir ini.

6. Bapak dan Ibu dosen Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang
7. Teristimewa untuk kedua orang tua dan adik yang telah memberikan semangat, kasih sayang, dan doa kepada peneliti dalam pengerjaan Tugas Akhir ini.
8. Terima kasih kepada sahabat dan teman-teman Diploma III Statistika 2020 yang telah membantu dan memberikan semangat dalam menyelesaikan Tugas Akhir ini.

Semoga semua bimbingan, bantuan, dan doanya mendapat balasan amalan ibadah dari Allah SWT. Dalam penyusunan Tugas Akhir ini penulis telah berusaha semaksimal mungkin memberikan yang terbaik, tetapi penulis menyadari bahwa penulisan Tugas Akhir ini belum sempurna, karena keterbatasan pengetahuan dan kemampuan yang dimiliki. Untuk itu kepada pembaca diharapkan kritik dan saran untuk penyempurnaan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi peneliti dan pembaca umumnya. Aamiin

Padang, Agustus 2023

Rahmi Nofita

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	5
C. Rumusan Masalah	5
D. Tujuan Penelitian.....	6
E. Manfaat Penelitian	6
BAB II KERANGKA TEORITIS	8
A. Angkatan Kerja.....	8
B. Status Kerja	9
C. Faktor-faktor yang diduga Mempengaruhi Status Kerja pada Angkatan Kerja .	9
D. <i>Preprocessing</i> Data	13
E. Metode <i>Classification and Regression Tree (CART)</i>	15
F. <i>Random Oversampling</i>	21
G. Ukuran Ketepatan Klasifikasi	23
BAB III METODOLOGI PENELITIAN	25
A. Jenis Penelitian.....	25
B. Jenis dan Sumber Data	25
C. Variabel-variabel dalam Penelitian	25
D. Teknik Analisis Data.....	26
BAB IV HASIL DAN PEMBAHASAN	28
BAB V PENUTUP	45
DAFTAR PUSTAKA	47
LAMPIRAN	50

DAFTAR GAMBAR

Gambar	Halaman
1. Persentase Status Kerja pada Angkatan Kerja Kota Padang pada Tahun 2022 ..	3
2. Struktur pohon klasifikasi CART	16
3. Proses <i>Random Oversampling</i>	22
4. Deskripsi Data Status Kerja Tahun 2022	29
5. Status Kerja Berdasarkan Klasifikasi Tempat Tinggal	29
6. Status Kerja Berdasarkan Status dalam Rumah Tangga	30
7. Status Kerja Berdasarkan Jenis Kelamin	31
8. Status Kerja Berdasarkan Usia	31
9. Status Kerja Berdasarkan Status Perkawinan	32
10. Status Kerja Berdasarkan Tingkat Pendidikan	33
11. Status Kerja Berdasarkan Pelatihan Kerja	34
12. Analisis CART Sebelum <i>Random Oversampling</i>	36
13. Jumlah Data Setelah <i>Oversampling</i>	38
14. Diagram Pohon Klasifikasi Setelah ROS	38

DAFTAR TABEL

Tabel	Halaman
1. Kriteria Perkotaan dan Pedesaan.....	10
2. <i>Confusion Matrix</i>	24
3. Variabel dalam Penelitian	26
4. Pemilihan Pemilah	35
5. <i>Confusion Matrix</i> pada CART	37
6. <i>Confusion Matrix</i> Setelah <i>Random Oversampling</i>	40
7. Perbandingan <i>Confusion Matrix</i> Analisis CART dengan ROS-CART	43

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Status Kerja di Kota Padang.....	50
2. Sintaks Deskripsi Data.....	53
3. Sintaks CART.....	54
4. <i>Confusion Matrix</i> Menggunakan CART.....	57
5. Sintaks <i>Random Oversampling</i> (ROS).....	57
6. Sintaks CART Setelah <i>Random Oversampling</i> (ROS).....	57
7. <i>Confusion Matrix</i> Setelah <i>Random Oversampling</i>	59

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Angkatan kerja adalah penduduk usia kerja (15 tahun ke atas) yang bekerja, atau punya pekerjaan namun sementara tidak bekerja dan pengangguran. Status kerja pada angkatan kerja digolongkan menjadi dua kategori yaitu bekerja dan tidak bekerja (pengangguran). Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu. Pengangguran adalah mereka yang tidak punya pekerjaan, sedang mencari pekerjaan, sedang mempersiapkan usaha, tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan, atau punya pekerjaan tetap namun belum mulai bekerja (BPS, 2020).

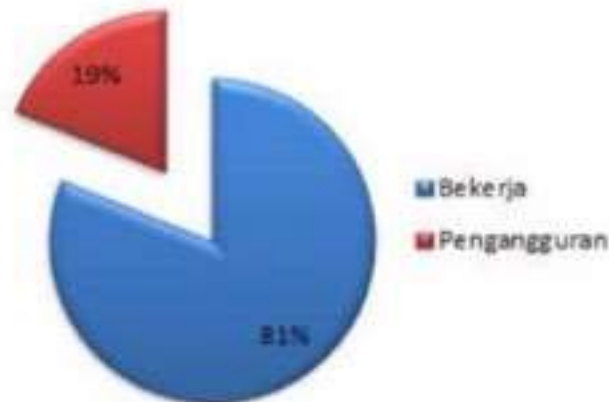
Penelitian sebelumnya yang dilakukan oleh Pratiwi & Zain (2014) menunjukkan bahwa beberapa faktor yang mempengaruhi status kerja di Provinsi Sulawesi Utara diantaranya jenis kelamin, tingkat pendidikan, usia, status dalam rumah tangga, status pelatihan kerja, status perkawinan, dan klasifikasi tempat tinggal. Berdasarkan faktor-faktor tersebut, untuk mengetahui karakteristik dari status kerja, maka perlu dilakukan pengklasifikasian (Fajriati, 2022). Pengklasifikasian adalah salah satu metode statistik yang digunakan untuk mengelompokkan atau mengklasifikasikan suatu data berdasarkan karakteristik tertentu yang disusun secara sistematis (Djuniar & Achmad, 2022).

Berdasarkan data Badan Pusat Statistik (BPS) Provinsi Sumatera Barat Tahun 2022, Kota Padang memiliki jumlah penduduk terbanyak di Sumatera Barat. Jumlah penduduk Kota Padang terus mengalami penambahan. Dibandingkan dengan hasil sensus penduduk tahun 2010 jumlah penduduk Kota Padang telah bertambah sekitar 82.474 orang atau meningkat 10 persen selama sepuluh tahun terakhir (BPS, 2022). Seiring dengan banyaknya jumlah penduduk, Kota Padang menjadi wilayah di Provinsi Sumatera Barat yang memiliki jumlah pengangguran tertinggi.

Decision tree (pohon keputusan) sering digunakan untuk menentukan keputusan suatu kasus. Salah satu metode yang terdapat pada *Decision Tree* adalah *Classification and Regression Tree* (CART). Menurut Yasa (2019) CART merupakan salah satu metode atau algoritma yang digunakan untuk mengeksplorasi data dengan penggambaran pohon keputusan. Metode ini dikembangkan oleh Leo Breiman, Jeremoe H. Friedman, Richard A. Olshen dan Charles J. Stone sekitar tahun 1980-an. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penjelasan lebih lanjut dari suatu pengklasifikasian.

Menurut Sartono dan Syafitri (2010), CART memberikan kemudahan untuk menginterpretasikan hasil analisis dan memberikan dugaan dengan tingkat kesalahan yang kecil. CART dapat digunakan pada data yang berjumlah besar dan kekar terhadap pencilan. Salah satu masalah yang sering terjadi ketika menggunakan CART adalah adanya ketidakseimbangan data yang mengakibatkan tingkat akurasi kecil.

Gambar 1 menyajikan persentase status kerja di Kota Padang pada Agustus 2022.



Gambar 1. Persentase Status Kerja pada Angkatan Kerja Kota Padang pada Tahun 2022

Berdasarkan Gambar 1, dapat dilihat perbandingan persentase status kerja di Kota Padang pada Tahun 2022. Status kerja yang bekerja sebanyak 81% dan yang status kerja yang pengangguran sebanyak 19%. Sehingga terlihat data tersebut tidak seimbang. Oleh karena itu, perlu dilakukan penyeimbangan data antara kelas minor dan kelas mayor dengan suatu metode agar dapat meningkatkan hasil akurasi (Nabila dkk, 2022).

Penelitian terdahulu mengenai status angkatan kerja di Kota Padang pernah dilakukan oleh Yulianti (2022). Metode yang di gunakan adalah *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi *imbalance data* pada CART. Peneliti menyebutkan bahwa pohon klasifikasi terbaik adalah pohon klasifikasi dengan SMOTE yang dapat meningkatkan nilai sensitivitas sebesar 43,1%. Penelitian ini menghasilkan ketepatan sensitivitas yang masih kecil. Sehingga ketepatan pengklasifikasian orang yang pengangguran sebagai pengangguran masih kurang tepat. Sehingga peneliti menyarankan untuk

penelitian selanjutnya menggunakan teknik *imbalance* data yang lain untuk meningkatkan kinerja hasil klasifikasi.

Umumnya terdapat dua pendekatan yang dapat dilakukan untuk menangani permasalahan data tidak seimbang, yakni pendekatan level data dan pendekatan level algoritma. Pendekatan pada level data dilakukan dengan menyeimbangkan distribusi kelas mayoritas dan minoritas dengan teknik pengambilan sampel seperti *undersampling*, *oversampling*, maupun kombinasi dari kedua metode tersebut. Sedangkan pendekatan pada level algoritma dilakukan dengan memodifikasi dan mengoptimalkan kinerja algoritma pembelajaran mesin (Santoso dkk, 2017).

Menurut He dkk (2008) pengambilan sampel pada data telah terbukti efektif untuk menangani data tidak seimbang. Pengambilan sampel merupakan pendekatan pada level data. Keuntungan dari pendekatan level data ini yakni tidak bergantung pada pengklasifikasi yang digunakan dan pendekatan ini dinilai lebih fleksibel (Ramyachitra & Manikandan, 2014). Untuk mengatasi masalah ketidakseimbangan kelas (*class imbalance*), maka pada penelitian ini akan dilakukan pengukuran kinerja pendekatan level data yang dilakukan dengan *resampling*, yaitu teknik *random oversampling* (ROS).

Random Oversampling merupakan metode Pemerataan data minoritas sehingga menjadi sebanyak data mayoritas (Triyanto & Kusumaningrum, 2017). Penelitian yang dilakukan Nugraha & Sabaruddin (2021) untuk mengatasi ketidakseimbangan data menggunakan teknik *resampling* dengan hasil analisis menunjukkan bahwa dengan teknik *resampling* yang dilakukan yaitu *undersampling* dan *oversampling* memberikan hasil yang baik. Menurut Santoso

dkk (2017) metode *oversampling* umumnya memberikan hasil yang lebih baik dibandingkan metode *undersampling*. Pada penelitian ini digunakan metode *oversampling* untuk mengatasi masalah ketidakseimbangan data.

Berdasarkan permasalahan yang telah diuraikan sebelumnya, akan dilakukan penanganan masalah ketidakseimbangan pada pengklasifikasian data status kerja pada angkatan kerja di Kota Padang menggunakan *oversampling* dengan variabel yang digunakan sama dengan penelitian sebelumnya. Oleh karena itu, peneliti tertarik untuk melakukan penelitian yang berjudul "Pengklasifikasian Status Kerja pada Angkatan Kerja di Kota Padang Menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*".

B. Batasan Masalah

Untuk memperjelas arah dari penelitian ini diperlukan adanya batasan masalah agar tidak menimbulkan kekeliruan. Batasan masalah pada penelitian ini yaitu penelitian hanya berfokus pada klasifikasi status kerja pada angkatan kerja di Kota Padang menggunakan data sekunder yang diperoleh dari data SAKERNAS Tahun 2022.

C. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah

1. Bagaimana bentuk pohon klasifikasi status kerja pada angkatan kerja di Kota Padang menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*

2. Apa saja karakteristik masyarakat yang bekerja dan pengangguran pada angkatan kerja di Kota Padang menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*
3. Bagaimana tingkat akurasi pohon klasifikasi status kerja pada angkatan kerja di Kota Padang menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*

D. Tujuan Penelitian

Tujuan dalam penelitian ini sebagai berikut.

1. Untuk mengetahui bentuk pohon klasifikasi status kerja pada angkatan kerja di Kota Padang menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*
2. Untuk mengetahui karakteristik masyarakat yang bekerja dan pengangguran pada angkatan kerja di Kota Padang menggunakan metode *Random Oversampling* dengan Metode *Classification and Regression Tree*
3. Untuk mengetahui tingkat akurasi pohon klasifikasi status kerja pada angkatan kerja di Kota Padang menggunakan *Random Oversampling* dengan Metode *Classification and Regression Tree*

E. Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi penulis, dapat menambah pemahaman dan pengetahuan tentang penerapan metode *Random Oversampling* dengan Metode *Classification and Regression Tree*

2. Bagi pemerintah, dapat memberikan informasi tambahan dalam pengambilan keputusan dan kebijakan dalam mengatasi masalah pengangguran di Kota Padang.
3. Bagi pencari kerja, dapat memberikan gambaran mengenai kriteria/karakteristik orang yang bekerja di Kota Padang